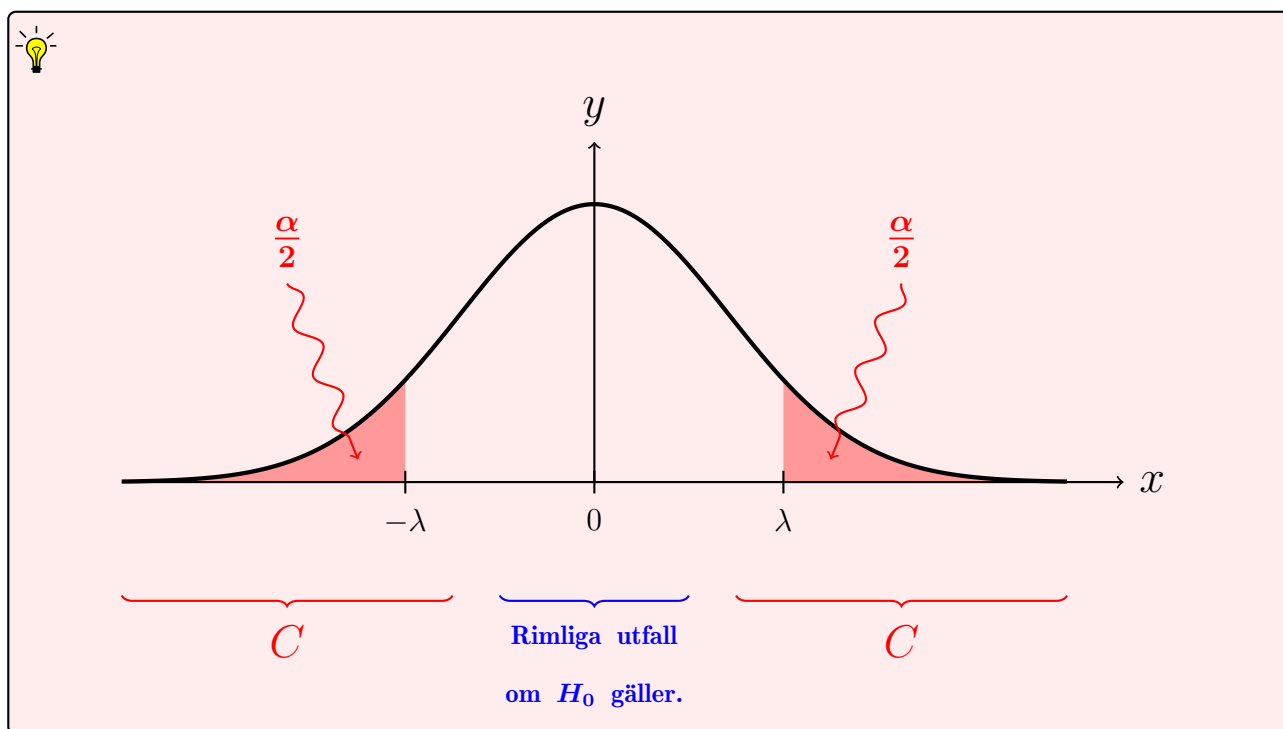


TAIU06/9GMA05: Matematisk statistik VT 2022

Föreläsningsanteckningar

Johan Thim, MAI

(johan.thim@liu.se)



Innehåll

1	Grundläggande begrepp	7
1.1	Begrepp	7
1.2	Mängdlära	7
1.3	Händelser	8
1.4	Sannolikhet	9
1.4.1	Klassiska definitionen av sannolikhet	10
1.4.2	Frekvenstolkning	11
1.5	Beroende händelser	12
1.6	Kombinatorik	12
1.7	Betingad sannolikhet	14
1.7.1	Lagen om total sannolikhet	15
1.7.2	Bayes sats	16
2	Stokastiska variabler	19
2.1	Endimensionella stokastiska variabler	19
2.2	Diskreta stokastiska variabler	20
2.3	Vanliga diskreta fördelningar	22
2.4	Kontinuerliga stokastiska variabler	25
2.5	Oberoende stokastiska variabler	27
2.6	($\star\star$)Vad är mer exakt en stokastisk variabel?	29
2.7	(\star)Kontinuerliga s.v. med täthetsfunktion	29
2.8	($\star\star\star$)Singulära stokastiska variabler?	30
3	Stokastiska variabler (forts)	33
3.1	Väntevärde	33
3.1.1	Funktioner och väntevärden	34
3.2	Varians och standardavvikelse	35
3.3	Räknelagar	36
3.4	Vanliga kontinuerliga fördelningar	37
3.5	Median	38
3.6	Normalfördelning	40
3.7	($\star\star$)Väntevärde för geometrisk fördelning	42
4	Normalfördelning och CGS	43
4.1	Normalfördelningen (forts)	43
4.2	Linjärkombinationer av normalfördelade variabler	45
4.3	De stora talens lag	47
4.4	Centrala gränsvärdessatsen	49
4.5	Approximation av binomialfördelning	52

4.5.1	Poissonapproximation	53
4.6	Approximation av Poissonfördelning	53
4.7	(★)Poissonfördelning	54
5	Flerdimensionella s.v.	57
5.1	Flervariabelanalys	57
5.1.1	Funktioner av flera variabler	57
5.1.2	Integration av funktioner av flera variabler	58
5.1.3	Iteration i 2D	58
5.2	Flerdimensionella stokastiska variabler	61
5.3	Väntevärden	67
5.4	Kovarians	69
5.5	Vad innebär korrelationen grafiskt?	71
5.6	Funktioner av stokastiska variabler	72
5.7	(★★)Formell definition av 2D-stokastisk variabel	75
5.8	(★)Bökigt exempel	75
5.9	(★★)Normalfördelning	75
6	Statistisk inferens	77
6.1	Representation av stickprov	77
6.1.1	Två-dimensionell data	79
6.2	Punktskattningar	82
6.2.1	Vanliga punktskattningar	84
6.3	Vilka skattningar är bra?	84
6.3.1	Effektivitet – jämförelse mellan skattningar	86
6.4	Momentmetoden	86
7	Punktskattningar och konfidensintervall	89
7.1	Vanliga punktskattningar	89
7.2	Metoder för att hitta punktskattningar	90
7.2.1	Momentmetoden	91
7.2.2	MK-skattning	91
7.2.3	ML-skattning	93
7.3	Flera stickprov	94
7.4	Medelfel	95
7.5	Intervallskattningar	95
7.6	K.I. för μ (σ känd)	97
7.7	(★★)ML-skattning för normalfördelning	100
8	Konfidensintervall (forts)	101
8.1	χ^2 - och t -fördelning	101
8.2	Konfidensintervall för μ i normalfördelning	102
8.2.1	Konfidensintervall för μ när σ är känd	102
8.2.2	Okänd varians	103
8.3	Prediktionsintervall	105
8.4	Skillnad mellan parametrar	106
8.5	Linjärkombinationer av normalfördelningar	106
8.5.1	Känd varians	107
8.5.2	Okända men likadana varianser ($\sigma_1 = \sigma_2$)	108

8.5.3	Okända varianser ($\sigma_1 \neq \sigma_2$)	109
8.6	Enkelsidiga konfidensintervall	110
8.7	Stickprov i par	111
8.8	Konfidensintervall via CGS	112
8.9	Stickprov för andel	112
8.10	Jämförelse av två andelar	113
8.11	($\star\star$) χ^2 -fördelningen	114
8.12	(\star) t -fördelningen	116
8.12.1	t -fördelningens kvantiler	117
8.13	(\star)Vektorer med stokastiska variabler	119
8.14	(\star)Cochrans sats	119
8.15	($\star\star$) t - och χ^2 -fördelning; Gossets sats	120
8.16	(\star)Bonus: Gammafunktionen	122
9	Hypotestester	123
9.1	Hypotestest	123
9.1.1	Teststorhet och kritiskt område	123
9.1.2	Styrka, fel och p-värde	125
9.2	Hypotestest för Binomialfördelning	127
9.3	Normalapproximation – Generellt	129
9.4	Test för skillnad i andel	129
9.5	Poissonapproximation	131
9.6	Hypotestest för väntevärde vid normalfördelning	132
9.6.1	Känd varians	132
9.6.2	Okänd varians	133
9.7	Hypotestester och konfidensintervall	134
9.8	Generellt om hypotestester	135
9.9	Flera stickprov	135
9.9.1	$\sigma_X = \sigma_Y = \sigma$ okänd	136
9.10	(\star)Hypotestest för Poissonfördelning	137
9.11	(\star)Plåtfabriken	138
10	χ^2-test och linjär regression	143
10.1	Det grundläggande χ^2 -testet	143
10.2	Test av given diskret fördelning	144
10.3	Test för kontinuerlig fördelning	145
10.4	Skattade storheter	147
10.5	Homogenitetstest	150
10.6	Skattningar för kovarians och korrelation	152
10.7	Enkel linjär regression	152
10.8	Variansanalys	155
10.8.1	SS_{TOT}	155
10.8.2	SS_R och SS_E	156
10.9	Hypotestester och konfidensintervall	157
10.9.1	Skattning av σ^2	157
10.9.2	Enskilda koefficienter	157
10.9.3	Hypotestest: $H_0 : \beta_i = 0$	158
10.9.4	Formler utan matriser	158
10.10	Ett löst exempel	160

10.11(★)Bevis för vissa resultat i linjär regression	164
10.12(★)Regressionsanalysens huvudsats	165
10.13(★) Bevis av χ^2 -testet	168

Kapitel 1

Grundläggande begrepp

1.1 Begrepp

Ett **slumpförsök** är ett försök där resultatet ej kan förutsägas deterministiskt. Slumpförsöket har olika möjliga **utfall**. Vi låter **Utfallsrummet** Ω vara mängden av alla möjliga utfall. En **händelse** är en delmängd av Ω , dvs en mängd av utfall. Men, alla möjliga delmängder av Ω behöver inte vara *tillåtna* händelser. För att precisera detta kräver vi att mängden av alla händelser (detta är alltså en mängd av mängder) är en så kallad σ -algebra. Vi definierar detta begrepp lite senare.

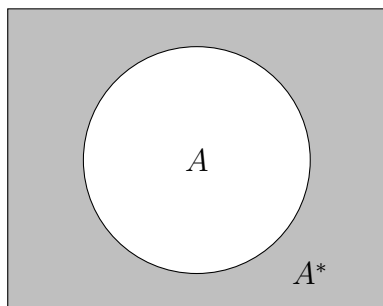


Exempel

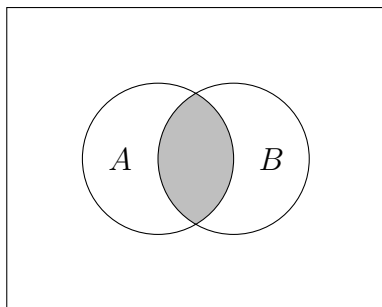
1. Myntkast: $\Omega = \{ \text{Krona, Klave} \}$.
2. Tärning (T-6): $\Omega = \{1, 2, 3, 4, 5, 6\}$. $A = \{1, 3, 5\}$ och $B = \{4\}$ är exempel på händelser.
3. Tiden till bilen går sönder: $\Omega = [0, \infty[= \{x \in \mathbf{R} : x \geq 0\}$. T ex $A = [0, 10[$ är händelsen att bilen går sönder innan 10 tidsenheter gått.

1.2 Mängdlära

En mängd M är en samling element utan ordning. Antalet element (kardinaliteten) i mängden betecknar vi $|M|$. Om mängden är ändlig är detta alltså bara hur många element som finns i mängden. Om mängden innehåller oändligt många element blir begreppet lite krångligare. En delmängd $A \subset M$ innehåller endast element från M (möjligen alla värden i M , eller inga). Mängder illustreras ofta med Venn-diagram. Nedan är hela rektangeln utfallsrummet Ω och de skuggade områdena olika delmängder.



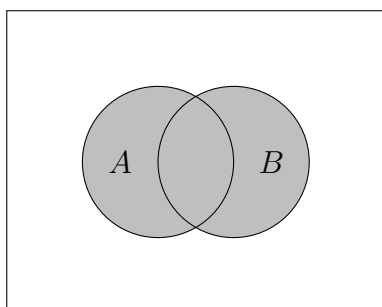
A är en händelse och A^* dess komplement. Komplementet A^* är händelsen att A **inte** inträffar. Komplementet A^* består av *alla* utfall (i Ω) som *inte* finns i händelsen A .



Snittet $A \cap B$ mellan händelserna $A, B \subset \Omega$. Händelsen att *både* A och B inträffar.

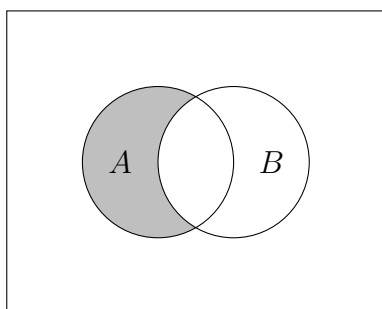
Om $A \cap B = \emptyset$ (tomma mängden) så kallas A och B för **oförenliga** (eller **disjunkta**). Två oförenliga händelser kan *ej* inträffa samtidigt.

Observera att $A \cap A^* = \emptyset$



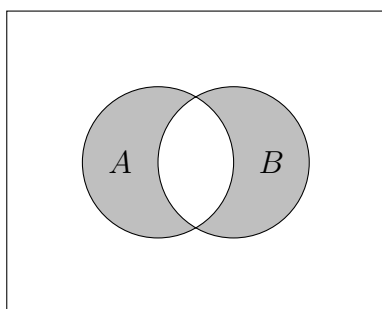
Unionen $A \cup B$ mellan händelserna $A, B \subset \Omega$. Händelsen att *någon* av A och B inträffar (eller båda).

Observera att $A \cup A^* = \Omega$ (hela utfallsrummet).



Skillnaden $A \setminus B = A \cap B^*$. Alla utfall i A utom de som även ligger i B . Händelsen att A inträffar men *inte* B .

Observera att $A^* = \Omega \setminus A$.



Symmetriska skillnaden: $A \Delta B = (A \setminus B) \cup (B \setminus A)$. Händelsen att en av A och B inträffar, men *inte* båda. Exklusivt eller.

1.3 Händelser

Händelser skulle som sagt vara element i en σ -algebra, vilket är ett objekt som definieras enligt följande.



σ -algebra

Definition. \mathcal{F} är en σ -algebra på Ω om \mathcal{F} består av delmängder av Ω så att

- (i) $\Omega \in \mathcal{F}$.
- (ii) om $A \in \mathcal{F}$ så är $A^* \in \mathcal{F}$.
- (iii) om $A_1, A_2, \dots \in \mathcal{F}$ så är unionen $A_1 \cup A_2 \cup \dots \in \mathcal{F}$.

Det enklaste exemplet på en σ -algebra är $\mathcal{F} = \{\Omega, \emptyset\}$, dvs endast hela utfallsrummet och den tomma mängden. Av förklarliga skäl kommer vi inte så långt med detta. Ett annat vanligt exempel är att \mathcal{F} består av *alla* möjliga delmängder till Ω ; skrivs ibland $\mathcal{F} = 2^\Omega$, och kallas potensmängden av Ω . Denna konstruktion är lämplig när vi har diskreta utfall. Om Ω består av ett kontinuum så visar det sig dock att 2^Ω blir alldeles för stor.



Exempel

Vi kastar en 6-sidig tärning och låter utfallen beskrivas av $\Omega = \{1, 2, 3, 4, 5, 6\}$. Som σ -algebra \mathcal{F} brukar vi normalt sett använda 2^Ω som består av alla möjliga delmängder av Ω :

\emptyset ,
 $\{1\}, \{2\}, \dots, \{6\}$,
 $\{1, 2\}, \{1, 3\}, \dots, \{1, 6\}, \{2, 3\}, \{2, 4\}, \dots, \{2, 6\}, \dots, \{5, 6\}$,
 $\{1, 2, 3\}, \{1, 2, 4\}, \dots, \{1, 2, 6\}, \{2, 3, 4\}, \{2, 3, 5\}, \{2, 3, 6\}, \dots, \{4, 5, 6\}$,
 $\{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 3, 6\}, \dots, \{3, 4, 5, 6\}$,
 $\{1, 2, 3, 4, 5\}, \dots, \{2, 3, 4, 5, 6\}$,
 $\{1, 2, 3, 4, 5, 6\}$.

De tillåtna händelserna är alltså alla möjliga mängder av utfall vid ett tärningskast. Händelsen $A = \{1, 3, 5\}$ att få ett udda utfall finns med. Händelsen $B = \{3, 6\}$ att få en trea eller sexa finns med. Och så vidare. Alla möjligheter.

1.4 Sannolikhet

Nästa naturliga fråga är givetvis hur vi introducerar begreppet sannolikhet. Vi söker en funktion P som definierar sannolikheter för alla händelser i \mathcal{F} ; detta är alltså en funktion definierad på \mathcal{F} . I fallet med tärningen så tror jag de flesta naturligt skulle tycka att

$$P(A) = \frac{|A|}{|\Omega|}, \quad A \in \mathcal{F},$$

är en vettig definition. Den innebär att vi helt enkelt tar kvoten mellan antalet utfall i A och delar med totala antalet möjliga utfall i Ω (detta är den *klassiska definitionen av sannolikhet* som vi återkommer till). Eftersom vi antagit att tärningen är rättvis så borde alla utfall vara lika troliga, så definitionen ovan vore naturlig. Detta innebär till exempel att sannolikheten att få ett udda utfall blir

$$P(\{1, 3, 5\}) = \frac{|\{1, 3, 5\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{3}{6} = \frac{1}{2}$$

och sannolikheten att få en trea eller sexa blir

$$P(\{3, 6\}) = \frac{|\{3, 6\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{2}{6} = \frac{1}{3}.$$

Vi ser vidare att

$$P(\emptyset) = \frac{|\{\emptyset\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{0}{6} = 0 \quad \text{och} \quad P(\Omega) = \frac{|\{1, 2, 3, 4, 5, 6\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{6}{6} = 1.$$

De två sista egenskaperna känns det rimligt att de borde gälla för alla sätt att mäta sannolikhet. Så vilka krav måste vi ställa på P ?



Kolmogorovs Axiom: Sannolikhetsmått

Definition. Ett sannolikhetsmått på en σ -algebra \mathcal{F} över ett utfallsrum Ω tilldelar ett tal mellan noll och ett, en sannolikhet, för varje händelse som är definierad (dvs tillhör \mathcal{F}). Formellt är P en mängdfunktion; $P: \mathcal{F} \rightarrow [0, 1]$. Sannolikhetsmättet P *måste* uppfylla Kolmogorovs axiom:

- (i) $0 \leq P(A) \leq 1$ för varje $A \in \mathcal{F}$.
- (ii) $P(\Omega) = 1$.
- (iii) Om $A \cap B = \emptyset$ så gäller att $P(A \cup B) = P(A) + P(B)$.

Formellt är det alltid en trippel (Ω, \mathcal{F}, P) när vi diskuterar sannolikhet, men vi låter ofta \mathcal{F} vara underförstådd.

Masstolkning: Ibland tolkas $P(A)$ som händelsen A 's sannolikhetsmassa. Ger en intuitiv bild av sannolikhetsfördelning mellan händelser (ofta grafiskt). Rita proportionerliga Venn-diagram! Följder av dessa axiom innefattar följande.



Egenskaper för sannolikhetsmättet

- (i) $P(A^*) = 1 - P(A)$;
- (ii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- (iii) $P(A \cup B) \leq P(A) + P(B)$ (Booles olikhet);
- (iv) $P(\emptyset) = 0$;
- (v) $P(A \setminus B) = P(A) - P(A \cap B)$.

Bevis? Rita Venn-diagram!

1.4.1 Klassiska definitionen av sannolikhet

Ett försök där varje utfall har samma sannolikhet säges ha **likformig** sannolikhetsfördelning. Om Ω är ändlig, säg $|\Omega| = m$, så är $P(\{\omega\}) = 1/m$ för varje utfall $\omega \in \Omega$.

För en likformig sannolikhetsfördelning på Ω så gäller för en händelse $A \subset \Omega$ att

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{"gynnsamma utfall"}}{\text{"möjliga utfall"}}.$$



Två tärningar

Vi kastar två rättvisa tärningar. Låt C vara händelsen att poängsumman blir sju. Vad är sannolikheten för C ?

Lösning: Låt

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \{(1, 1), (1, 2), \dots, (6, 6)\}.$$

Vi representerar alltså kasten som två koordinater, den första är från tärning ett och den andra från tärning två. Det finns 36 st möjliga utfall i Ω : det första kastet ger sex möjligheter, och för vart och ett av dessa utfall får vi sex nya möjligheter vid andra kastet. Vi väljer $\mathcal{F} = 2^\Omega$ (det typiska vid ändliga diskreta situationer).

De ”gynnsamma” utfallen samlas i händelsen

$$C = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

Observera att ordningen i talparen spelar roll. Enligt den klassiska sannolikhetsdefinitionen får vi

$$P(C) = \frac{|C|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

Detta gäller under förutsättning att alla utfall är lika troliga (med andra ord att tärningarna är rättvisa).



Vad måste framgå i en lösning?

Det kommer inte vara ett krav att specificera \mathcal{F} för att lösa en uppgift utan vi låter det vara underförstått, om inte själva frågan handlar om just σ -algebran.

En lösning på föregående exempel skulle mer kompakt kunna skrivas enligt följande.

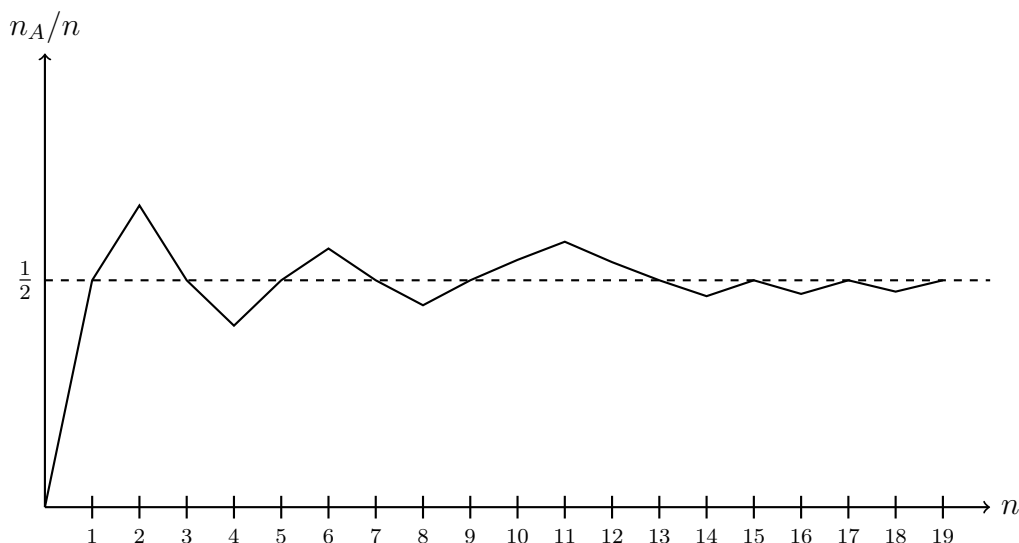
Lösning (kort version): Vi representerar utfallet av de två tärningskasterna som (x_1, x_2) , där x_1 är resultatet av det första kastet och x_2 resultatet av det andra. Det finns 36 möjliga sådana par. Av dessa är de 7 paren

$$(1, 6), \quad (2, 5), \quad (3, 4), \quad (4, 3), \quad (5, 2), \quad (6, 1)$$

de gynnsamma utfallen. Alltså ges den eftersökta sannolikheten — enligt den klassiska definitionen av sannolikhet — av $\frac{6}{36} = \frac{1}{6}$.

1.4.2 Frekvenstolkning

Vi upprepar ett försök n gånger och räknar antalet n_A gånger som händelsen A inträffar. Den relativa frekvensen definieras som n_A/n . Om $n \rightarrow \infty$ förefaller det rimligt att $n_A/n \rightarrow P(A)$. Detta kallas frekvenstolkningen av sannolikhet. Som exempel, låt oss singla slant många gånger och räkna antalet kronor ($A = \{\text{Krona}\}$) och plotta den relativa frekvensen:



Om myntet är symmetriskt förväntar vi oss att $n_A/n \rightarrow 1/2$ (eller hur?).

1.5 Beroende händelser



Oberoende

Definition. Två händelser A och B kallas **oberoende** om $P(A \cap B) = P(A)P(B)$.

Observera att denna likhet *inte* gäller i allmänhet. Ta till exempel händelsen $A = \{\text{Krona}\}$ och $B = \{\text{Klave}\}$ vid ett myntkast. Klart att $A \cap B = \emptyset$ så $P(A \cap B) = 0$. Men

$$P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \neq 0.$$

Självklart är A och B inte oberoende. Observera även att om vi pratar om tre eller fler händelser blir definitionen av oberoende krångligare; se boken.

1.6 Kombinatorik

Multiplikationsprincipen: Om vi har en tvåstegsprocess av valmöjligheter, där vi i första steget har n_1 möjligheter och i det andra n_2 möjliga val, så finns det totalt sätt $n_1 \cdot n_2$ kombinationer.

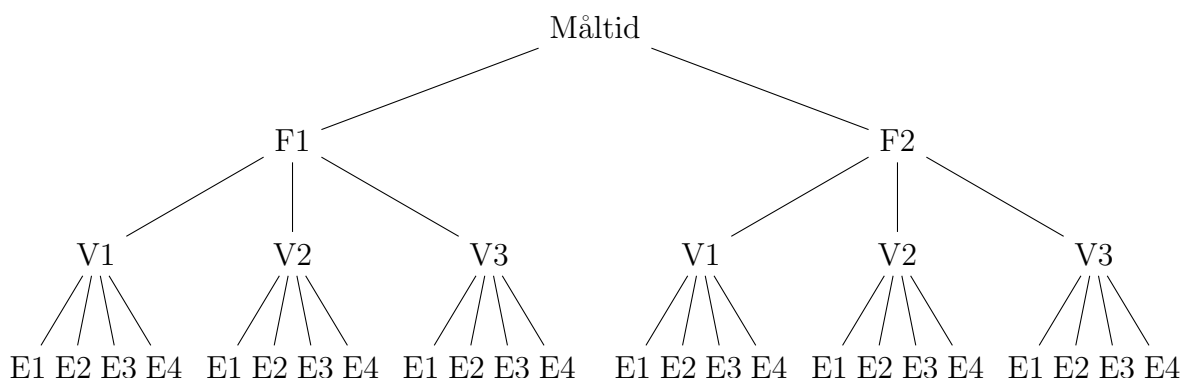


Exempel

En tre-rätters meny har 2 förrätter, 3 varmrätter, och 4 efterätter. Hur många olika måltider kan man beställa om man vill ha förrätt, varmrätt och efterätt?

Enligt multiplikationsprincipen blir det $2 \cdot 3 \cdot 4 = 24$ olika måltider.

Man kan illustrera multiplikationsprincipen med hjälp av trädidiagram. I figuren nedan väljer vi på nivå 1 mellan två förrätter (F1 och F2). I nästa nivå väljer vi mellan 3 varmrätter (V1, V2 och V3). I det sista steget väljer vi mellan fyra efterätter. Varje väg genom trädet ger en unik måltid. Hur många sådana vägar finns det? Det är bara att räkna ihop hur många "löv" det finns på den sista nivån, vilket blir precis 24 st.



Något lite krångligare? Vi utnyttjar multiplikationsprincipen för att reda ut följande scenario.



Inbrottstjuven

Inbrottstjuven Ivar försöker öppna 10 st dörrar, D_1, D_2, \dots, D_{10} , och Ivar har 80% sannolikhet att lyckas med varje dörr. Vad är sannolikheten att exakt sex st dörrar blir öppnade?

Lösning: Vi antar att öppnandet av olika dörrar är oberoende av varandra (är det rimligt?). En viss följd av resultat, t ex $D_1 = Y, D_2 = N, D_3 = Y, \dots, D_{10} = N$ (med sex st Y , öppna dörrar, och fyra st N , misslyckade försök), har eftersom händelserna är oberoende sannolikheten

$$\begin{aligned} P(D_1 = Y, \dots, D_{10} = N) &= P(D_1 = Y)P(D_2 = N) \cdots P(D_{10} = N) \\ &= 0.8 \cdot 0.2 \cdots 0.2 \\ &= 0.8^6 \cdot 0.2^4 \approx 4.194 \cdot 10^{-4}. \end{aligned}$$

Hur många sådana följder finns det? Vi har tio dörrar och skall välja ut sex st som öppnas:

Dörr 1	Dörr 2	Dörr 3	Dörr 4	Dörr 5	Dörr 6
10	9	8	7	6	5

Dörr 1 kan vi välja på 10 olika sätt. När vi sedan väljer dörr 2 finns det bara 9 kvar att välja på. Och så vidare. Ordningen på dörrarna är nu fixerad, och vi får (från multiplikationsprincipen)

$$10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 151200$$

sådana val.

När de sex dörrarna är valda kan vi variera ordningen mellan dessa 6 på $6!$ olika sätt:

Dörr 1	Dörr 2	Dörr 3	Dörr 4	Dörr 5	Dörr 6
6	5	4	3	2	1

Vi kan nu ta bort ”multipla” dörrval (de kombinationer som bara skiljer sig åt med i vilken ordning sex st specifika dörrar ligger):

$$\frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{6!} = \frac{10!}{6! \cdot 4!} = \binom{10}{6}.$$

Vi får alltså en binomialkoefficient!

Eftersom de olika sekvenserna av dörrval är oförenliga händelser (två olika val ger olika dörrsekvenser) får vi sannolikheten

$$\binom{10}{6} 0.8^6 0.2^4 \approx 0.088.$$

Det är alltså ungefär 8.8% chans att exakt sex stycken dörrar blir öppnade.

1.7 Betingad sannolikhet

Tänk om vi skulle vilja räkna ut sannolikheten för en händelse A men att vi i förväg redan vet att en händelse — som påverkar sannolikheten för A — inträffar. Tänk om vi vill veta sannolikheten att ett tärningskast blev en 3:a men att vi av någon anledning vet att utfallet är udda innan vi behöver svara på frågan. Hur hanterar vi den här typen av information?



Betingad sannolikhet

Definition. Låt $P(B) > 0$. Den betingade sannolikheten $P(A|B)$ för händelsen A , givet att händelsen B inträffar, definieras som $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Om A och B är oberoende och $P(B) \neq 0$ ser vi att

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

Rimligt?



Exempel

Alla som lyssnar på hårdrock i någon form har säkert funderat över vilken av Slayer-låtarna *Angel of Death* och *Raining Blood* som är bäst^a. Examinator funderade över detta och samlade in följande siffror på internet:

	<i>Angel of Death</i>	<i>Raining Blood</i>	Summa
Returntothepit.com	199	173	372
MetalStorm.net	47	43	90
Summa	246	216	462

^aSjälvklart är *Angel of Death* den bästa av dessa två, men det är inte poängen!

Låt $A = \text{Angel of Death}$ och $B = \text{Returntothepit.com}$. Från tabellen erhåller vi

$$P(A) = \frac{246}{462}, \quad P(B) = \frac{372}{462} \quad \text{och} \quad P(B \cap A) = \frac{199}{462}.$$

Vi kan direkt beräkna $P(B|A)$ genom att titta enbart i första kolumnen (det vi menar med sannolikhet betingad på $A =$ första kolumnen): $P(B|A) = \frac{199}{246}$. Använder vi definitionen istället blir det

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{199/462}{246/462} = \frac{199}{246}.$$

1.7.1 Lagen om total sannolikhet

Ibland hjälper det att dela upp ett problem i mindre bitar där vi enklare kan finna sannolikheterna. Lagen om total sannolikhet ger oss en enkel möjlighet att pussla ihop dessa bitar igen efteråt.



Lagen om total sannolikhet

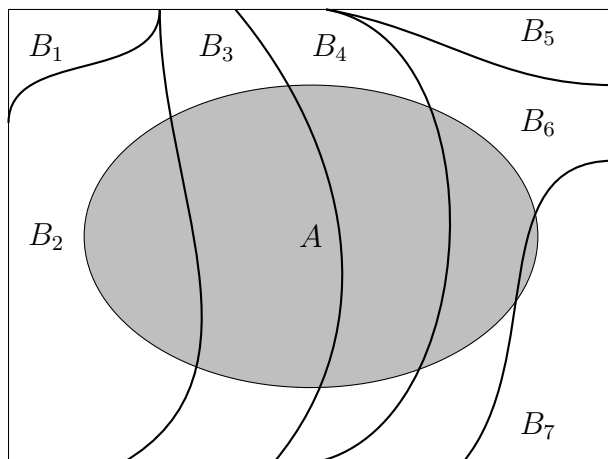
Sats. Låt $A, B_1, B_2, \dots, B_n \subset \Omega$ vara händelser sådana att:

- (i) $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$;
- (ii) $P(B_k) \neq 0$ för alla $k = 1, 2, \dots, n$;
- (iii) $B_i \cap B_j = \emptyset$ om $i \neq j$.

Då gäller lagen om total sannolikhet:

$$P(A) = \sum_{k=1}^n P(A|B_k)P(B_k).$$

Figuren nedan visar ett exempel på hur situationen skulle kunna se ut.



Beviset? Ganska enkelt:

$$\sum_{k=1}^n P(A|B_k)P(B_k) = \sum_{k=1}^n \frac{P(A \cap B_k)}{P(B_k)} P(B_k) = \sum_{k=1}^n P(A \cap B_k) = P(A),$$

där vi använt definitionen av betingad sannolikhet samt Kolmogorovs tredje axiom. Händelserna $A \cap B_k$ är disjunkta för $k = 1, 2, \dots, n$ eftersom $B_i \cap B_j = \emptyset$ om $i \neq j$.

1.7.2 Bayes sats



Bayes sats

Sats. Med samma villkor som för lagen om total sannolikhet gäller

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{k=1}^n P(A | B_k)P(B_k)}$$

för varje $i = 1, 2, \dots, n$.

Bayes sats är en följd av lagen om total sannolikhet samt definitionen av betingad sannolikhet.

Anmärkning. Satsen är även sann om vi har uppräknligt många händelser B_1, B_2, \dots som uppfyller kraven i satserna. Beviset fungerar som skrivet.



Exempel

Tre maskiner tillverkar prylar. Maskin 1 står för 60%, M-2 för 25% och M-3 för 15%. Av de tillverkade prylarna är 5, 3 respektive 2% felaktiga från de olika maskinerna.

Hur stor är sannolikheten att en på måfå vald enhet är trasig? Om en enhet visar sig vara trasig, hur stor är sannolikheten att den kommer från maskin ett?

Lösning: Enligt lagen om total sannolikhet får vi

$$\begin{aligned} P(\text{trasig pryl}) &= P(M_1) \cdot P(\text{trasig} | M_1) + P(M_2) \cdot P(\text{trasig} | M_2) \\ &\quad + P(M_3) \cdot P(\text{trasig} | M_3) \\ &= 0.6 \cdot 0.05 + 0.25 \cdot 0.03 + 0.15 \cdot 0.02 = 0.0405. \end{aligned}$$

Den andra frågan kan vi svara på mha Bayes sats:

$$P(M_1 | \text{trasig}) = \frac{P(M_1) \cdot P(\text{trasig} | M_1)}{P(\text{trasig})} = \frac{0.6 \cdot 0.05}{0.0405} \approx 0.741.$$



Sensitivitet och specificitet

Ett forensiskt test för narkotikapåverkan har sensitivitet 0.9999 (positivt utslag vid påverkan) och specificitet 0.995 (negativt utslag om inte påverkad). Antag att den forensiska analytikern får tillbaka beskedet "positivt utslag" vid en undersökning. Vad är sannolikheten att personen i fråga faktiskt var påverkad?

Betrakta två grupper. Om personen kommer från grupp ett bedöms sannolikheten att personen är påverkad till 20%, och i grupp två bedöms motsvarande sannolikhet till 0.1%.

Låt A vara händelsen att testet är positivt och B sannolikheten att personen är påverkad. Vi låter $P(B) = p$. Bayes sats medför att

$$\begin{aligned} P(B | A) &= \frac{P(B)P(A | B)}{P(B)P(A | B) + P(B^*)P(A | B^*)} = \frac{p \cdot 0.9999}{p \cdot 0.9999 + (1 - p)(1 - P(A^* | B^*))} \\ &= \frac{p \cdot 0.9999}{p \cdot 0.9999 + (1 - p)0.05} = \frac{1}{1 + \left(\frac{1}{p} - 1\right) \frac{0.05}{0.9999}} \end{aligned}$$

Här har vi utnyttjat att $P(A \cap B^*) = P(A) - P(A \cap B)$.

Om $p = 0.2$ så är $P(B \mid A) \approx 0.83$ och om $p = 0.001$ så är $P(B \mid A) = 0.020$.

Observera att det alltså *inte* är 99.999% chans att positivt utslag innebär påverkan. Vad skulle krävas för att detta skulle gälla?

Kapitel 2

Stokastiska variabler

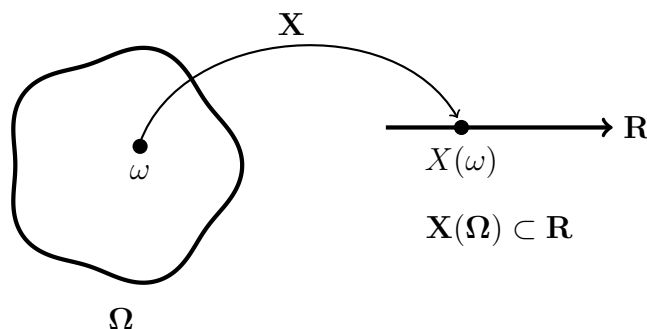
2.1 Endimensionella stokastiska variabler

När vi arbetar med slumpförsök så introducerar vi ofta något som brukar kallas för en stokastisk variabel (eller slumpvariabel). Detta gör att vi slipper arbeta direkt med det underliggande utfallsrummet (som kan vara abstrakt eller bökigt) och den där konstiga σ -algebran. Så vad är då en stokastisk variabel? En lite handviftande definition följer (se sista avsnittet på föreläsningen för en mer ordentlig definition även om den biten faller utanför denna kurs ramar).



Stokastisk variabel

Definition. En **stokastisk variabel** är en reellvärd funktion definierad på ett utfallsrum Ω . Funktionen X avbildar alltså olika utfall på reella tal; $X: \Omega \rightarrow \mathbf{R}$.



Vi sammanfattar några följder och gör ett par följddefinitioner.

- (i) Uttrycket $X(\omega)$ (funktionsvärdet för X i punkten ω) är alltså det siffrvärde vi sätter på ett visst utfall $\omega \in \Omega$.
- (ii) **Bilden** av en delmängd A av Ω betecknas med $X(A)$, så med andra ord:

$$X(A) = \{x \in \mathbf{R} : X(\omega) = x \text{ för något } \omega \in A\}.$$

- (iii) Mängden $X(A)$ är alltså värdemängden för X på mängden A . Av tradition brukar vi använda stora bokstäver för att beteckna stokastiska variabler. Termen variabel är egentligen lite olycklig då våra stokastiska variabler är funktioner, men det är en gammal tradition som lever kvar.

- (iv) Om $X(\Omega)$ är ändlig, eller bara har uppräknligt många värden, så kallar vi X för en **diskret stokastisk variabel**. Annars kallar vi X för **kontinuerlig**.

Ett par exempel kan vara på sin plats.



Exempel

- (i) Kasta en tärning. Utfallsrummet $\Omega = \{\square, \square, \square, \square, \square, \square\}$. Låt X vara antalet ögon vid ett tärningskast. X antar värdena 1, 2, 3, 4, 5, 6, så X är diskret.
- (ii) Handla tacosås på måfå. $\Omega = \{\text{Het}, \text{Medel}, \text{Mild}\}$. Låt $X = 1$ om såsen är mild, $X = 5$ om såsen är medel och $X = 10$ om såsen är het. X är diskret.
- (iii) Låt X vara livslängden för ett kylskåp. Då kan X (teoretiskt) anta alla värden i intervallet $[0, \infty[$, så X är kontinuerlig.
- (iv) Låt Ω bestå av alla möjliga färger på gräset. Låt $X = \lambda$ vara motsvarande våglängd för färgen (kontinuerlig variabel). Låt $Y = 1$ om färgen är grön och $Y = 0$ annars (diskret variabel).

Definitionen ovan är av ganska teknisk karaktär, så vad måste man ta med sig för att kunna tillgodogöra sig resten av denna kurs?



Vad måste jag förstå av all matematiska?

- En stokastisk variabel är en *snäll* funktion från Ω till \mathbf{R} .
- Utfallsrummet Ω kan vara abstrakt, e.g., $\Omega = \{\text{Krona}, \text{Klave}\}$.
- Det är mängden $X(\Omega)$ som består av siffror.
- Ibland finns en naturlig koppling mellan Ω och $X(\Omega)$, säg om vi kastar en tärning och räknar antalet ögon vi får.
- En händelse är en *snäll* delmängd av Ω .
- Om A är en händelse så är $X(A)$ *värdemängden* för funktionen X med A som definitionsmängd. Speciellt så är $X(\Omega)$ alla möjliga värden vi kan få från variabeln X .

2.2 Diskreta stokastiska variabler

Om $X(\Omega)$ är ändlig eller uppräknligt oändlig så kallade vi X för diskret. En sådan variabel kan vi karaktärisera med en så kallad **sannolikhetsfunktion**.



Sannolikhetsfunktion

Definition. Sannolikhetsfunktionen $p_X: X(\Omega) \rightarrow [0, 1]$ för en diskret stokastisk variabel definieras av $p_X(k) = P(X = k)$ för alla $k \in X(\Omega)$.

Den vanligaste situationen vi stöter på är att utfallsrummet är numrerat med heltal på något sätt så att p_X är en funktion definierad för (en delmängd av) heltal (när det finns en naturlig koppling mellan Ω och $X(\Omega)$). Ibland är vi slarviga och tänker oss att $p_X(k) = 0$ för siffror k som ej är möjliga ($p_X(-1) = 0$ om X är antal ögon vi ett tärningskast till exempel).

Vissa egenskaper gäller för alla alla sannolikhetsfunktioner:



Egenskaper hos sannolikhetsfunktionen

- (i) $p_X(k) \geq 0$ för alla $k \in X(\Omega)$.
- (ii) $\sum_{k \in X(\Omega)} p_X(k) = 1$.
- (iii) Om $A \subset X(\Omega)$ så är $P(X \in A) = \sum_{k \in A} p_X(k)$.

En sannolikhetsfunktion är alltså *aldrig* negativ, om vi summerar över alla möjliga värden (alla $k \in X(\Omega)$) så måste summan bli ett, och om vi är ute efter sannolikheten att få vissa värden på X så summerar vi sannolikheten för vart och ett av dessa värden!



Fördelningsfunktion

Definition. Fördelningsfunktionen $F_X(x)$ för en stokastisk variabel X definieras av för alla tal $x \in \mathbf{R}$ av sambandet $F_X(x) = P(X \leq x)$.

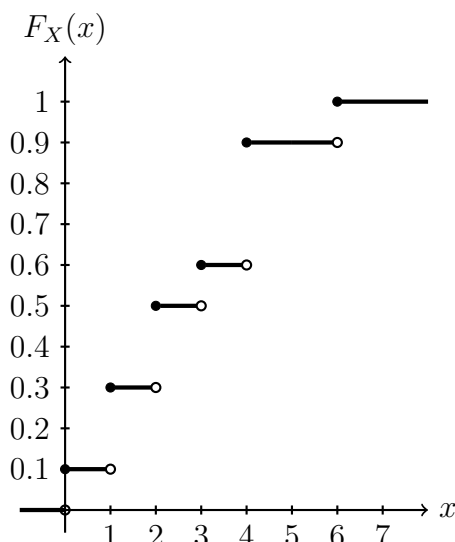
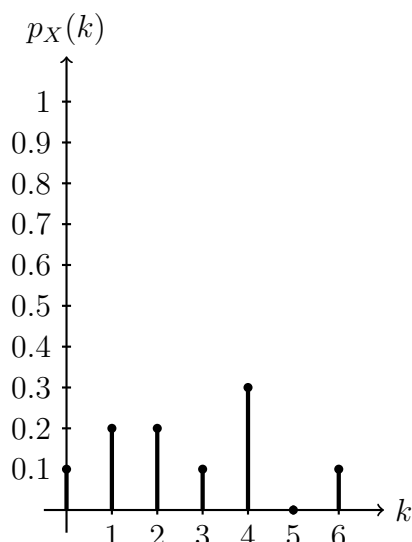
Det följer från definitionen att följande påståenden gäller.



Egenskaper hos fördelningsfunktionen

- (i) $F_X(x) \rightarrow \begin{cases} 0, & x \rightarrow -\infty, \\ 1, & x \rightarrow +\infty. \end{cases}$
- (ii) $F_X(x)$ är icke-avtagande och högerkontinuerlig.
- (iii) $F_X(x) = \sum_{\{k \in X(\Omega): k \leq x\}} p_X(k)$.
- (iv) $P(X > x) = 1 - F_X(x)$.
- (v) $F_X(k) - F_X(k-1) = p_X(k)$ för $k \in X(\Omega)$.

Exempel på hur en sannolikhetsfunktion och motsvarande fördelningsfunktion kan se ut:



Sannolikhetsfunktion $p_X(k) = P(X = k)$.

Fördelningsfunktion $F_X(x) = P(X \leq x)$.

2.3 Vanliga diskreta fördelningar

Det räcker med sannolikhetsfunktionen för att karakterisera en diskret variabel, så vi sammanfattar några av de vanligaste fallen. Vi antar genomgående att $0 < p < 1$. En av de enklaste fördelningarna vi stöter på är Bernoullifördelningen, eller 2-punkts fördelningen.



Tvåpunktsfördelning (Bernoullifördelning)

Den stokastiska variabeln X kan anta två värden: a och b . Vi kallar X för **tvåpunktsfördelad**, $X \sim \text{Be}(p)$, om $p_X(a) = p$ och $p_X(b) = 1 - p$.



Exempel

Slantsingling med osymmetriskt mynt. Låt $X = -1$ vid krona och $X = 1$ vid klave. Till exempel kan vi ha $p_X(-1) = P(X = -1) = 0.4$ och $p_X(1) = P(X = 1) = 0.6$.

Vad händer om vi betraktar summan av oberoende Bernoullivariabler?



Exempel

En händelse har 30% sannolikhet. Vi upprepar försöket 10 gånger, oberoende av varandra. Vad blir sannolikheten att:

1. Händelsen inträffar exakt k gånger;
2. Händelsen inträffar högst 1 gång;
3. Händelsen inträffar minst 2 gånger.

Lösning: Låt X vara antalet gånger händelsen inträffar. Då är $X = 0, 1, \dots, 10$ möjliga värden.

1. Enligt exemplet från föregående föreläsning (inbrottstjuven) måste

$$P(X = k) = \binom{10}{k} 0.3^k \cdot 0.7^{10-k}, \quad k = 0, 1, \dots, 10.$$

Observera att $P(X = k) = p_X(k)$, så uttrycket ovan är sannolikhetsfunktionen för X .

$$2. P(X \leq 1) = \sum_{k=0}^1 p_X(k) = \binom{10}{0} 0.3^0 \cdot 0.7^{10} + \binom{10}{1} 0.3^1 \cdot 0.7^9 \approx 0.149.$$

$$3. P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) \approx 0.851.$$

Vi säger att X är binomialfördelad med parametrarna $n = 10$ och $p = 0.3$.

Generellt så kallar vi en stokastisk variabel för Binomialfördelad om vi oberoende upprepar ett experiment med två möjliga utfall ett förutbestämt antal gånger n och räknar antalet gånger ett av utfallen inträffar (vilket vid varje försök sker med konstant sannolikhet p).



Binomialfördelning

Vi kallar X för binomialfördelad med parametrarna n och p om X har sannolikhetsfunktionen

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

och vi skriver $X \sim \text{Bin}(n, p)$.

Antag att vi har en större mängd med N element där $N = v + s$ består av två olika sorters element. Låt $p = v/N$ vara andelen v -märkta element. Om vi på måfå plockar ut n stycken element från N , hur många är v -märkta? Svaret kommer i form av den Hypergeometrisk fördelningen.



Hypergeometrisk fördelning

Vi skriver $X \sim \text{Hyp}(N, n, p)$ om

$$p_X(k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots, \min\{Np, n\}.$$

Vi kallar X för **Hypergeometriskt** fördelad.

Varför blir det så? Det är bara multiplikationsprincipen *in action*. Vi väljer k stycken av de Np v -märkta kulorna och $n - k$ stycken av de $N(1 - p)$ s -märkta kulorna (vilket ger de gynnsamma utfallen). Totalt sett väljer vi n stycken kulor från de N som finns (det totala antalet). Vi räknar allt utan ordning och använder den klassiska definitionen av sannolikhet.

**Exempel**

Ur en grupp bestående av 100 studenter väljer vi 20 på måfå. Den stora gruppen består till 60% av kvinnor. Vad är sannolikheten att vi har precis elva kvinnor i den mindre gruppen?

Lösning: Låt X vara antalet kvinnor i den mindre mängden. Det följer att $X \sim \text{Hyp}(100, 20, 0.6)$, så sannolikheten vi söker kan beräknas enligt

$$P(X = 11) = p_X(11) = \frac{\binom{60}{11} \binom{40}{9}}{\binom{100}{20}} \approx 0.175.$$

**Likformig (rektangel-) fördelning**

Om X antar ändligt många värden, säg $X \in E = \{1, 2, \dots, m\}$, och vi definierar $p_X(k) = \frac{1}{m}$ för varje $k = 1, 2, \dots, m$, så kallar vi X för **likformigt** fördelad (på mängden E).

**Exempel**

Låt $\Omega = \{0, 1, 2, \dots, 9\}$ och låt X vara likformigt fördelad på Ω . Vidare, låt $Y(x) = 0$ om x är jämnt delbart med 3, annars är $Y = 1$. Bestäm p_X och p_Y .

Lösning: Vi har $p_X(k) = 1/10$ om $k = 0, 1, 2, \dots, 9$ och $p_X(k) = 0$ annars. Vi låter mängden $A = \{0, 3, 6, 9\}$ bestå av de tal som är delbara med 3 och $B = \{1, 2, 4, 5, 7, 8\}$ de som inte är delbara med tre. Klassiska definitionen på sannolikhet ger $P(A) = 4/10$ och $P(B) = 6/10$. Variabeln Y blir Bernoullifördelad med $p = 2/5$ (med $a = 0$ och $b = 1$).

**För-första-gången-fördelning**

Vi skriver $X \sim \text{Ffg}(p)$ om $p_X(k) = (1 - p)^{k-1}p$, $k = 1, 2, \dots$. Vi kallar X för **För-första-gången-fördelad**.

Ett slumpförsök har två olika utfall, säg A och B , med sannolikheterna p respektive $1 - p$. Vi upprepar försöket oberoende tills dess att händelsen A inträffar för första gången. Antalet försök X till och med att A inträffar för första gången är $\text{Ffg}(p)$ -fördelad. Om $X = k$ innebär det att A inträffade för första gången vid den k :te upprepningen, och att i de $k - 1$ första försöken inträffade B . Alltså måste $P(X = k) = P(B)^{k-1}P(A) = (1 - p)^{k-1}p$ eftersom försöken är oberoende.

**Exempel**

Låt oss kasta en 6-sidig tärning tills dess att vi för första gången får en 1:a eller 3:a. Låt X vara antalet kast. Händelsen att få en 1:a eller 3:a vid ett kast är $p = 2/6 = 1/3$ (gynnsamma/möjliga). Då blir alltså $X \sim \text{Ffg}(1/3)$ -fördelad. Vad är sannolikheten att det tar fyra eller fler kast innan vi får en 1:a eller 3:a för första gången?

Lösning: Som bekant är $X \sim \text{Ffg}(1/3)$, så

$$P(X \geq 4) = 1 - \sum_{k=1}^3 P(X = k) = 1 - \left(\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3} + \left(\frac{2}{3} \right)^2 \cdot \frac{1}{3} \right) = \frac{8}{27}.$$

Alternativt (om man tidigare arbetat med geometriska serier),

$$P(X \geq 4) = \sum_{k=4}^{\infty} P(X = k) = \frac{1}{3} \sum_{k=4}^{\infty} \left(\frac{2}{3} \right)^{k-1} = \frac{2^3}{3^4} \sum_{k=0}^{\infty} \left(\frac{2}{3} \right)^k = \frac{2^3}{3^4} \cdot \frac{1}{1 - 2/3} = \frac{8}{27}.$$

En nära besläktad fördelning är den geometriska. Vi kan tänka oss att vi räknar antalet misslyckade försök innan en händelse inträffar för första gången.



Geometrisk fördelning

Vi skriver $X \sim \text{Geo}(p)$ om $p_X(k) = (1-p)^k p$, $k = 0, 1, 2, \dots$. Vi kallar X för **Geometriskt** fördelad.

En annan diskret fördelning vi kommer att stöta på framöver är Poissonfördelningen.



Poissonfördelning

Vi skriver $X \sim \text{Po}(\mu)$ om $p_X(k) = \frac{\mu^k}{k!} e^{-\mu}$, $k = 0, 1, 2, \dots$ och $\mu > 0$. Vi kallar X för **Poisson-fördelad**.

För vissa fördelningar och parametervärden har vi tabeller av sannolikheter att tillgå, speciellt för Poisson- och Binomialfördelning. Studera formelsamlingen! Se till att ni lär er känna igen och skilja de olika fördelningarna åt. De flesta kommer dyka upp i andra sammanhang och andra kurser senare i utbildningen.

2.4 Kontinuerliga stokastiska variabler

Vi kommer nu att utveckla teori för kontinuerliga stokastiska variabler som motsvarar den vi tog fram i det diskreta fallet. Åtminstone i de fall där det finns en så kallad täthetsfunktion. Så vi börjar med att definiera detta begrepp.

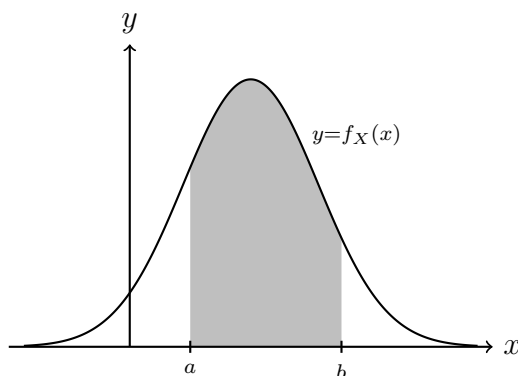


Täthetsfunktion

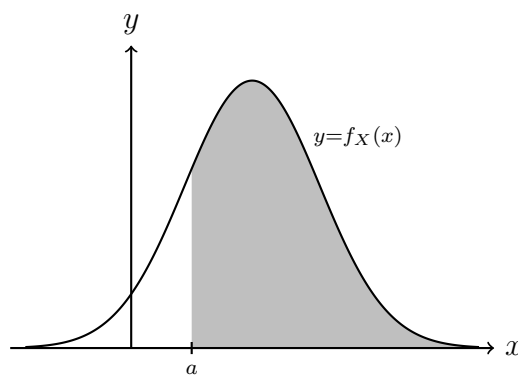
Definition. Om det finns en icke-negativ *integrerbar* funktion f_X så att

$$P(a < X < b) = \int_a^b f_X(x) dx$$

för alla intervall $(a, b) \subset \mathbf{R}$, kallar vi f_X för variabelns **täthetsfunktion**.



Skuggad area: $P(a \leq X \leq b)$.



Skuggad area: $P(X > a) = \int_a^\infty f_X(x) dx$.

Det är inte på något sätt självklart att det finns en täthetsfunktion även om variabeln inte är diskret. Men i de fall denna funktion existerar så gäller alltid vissa egenskaper.



Egenskaper hos täthetsfunktionen

- (i) $f_X(x) \geq 0$ för alla $x \in \mathbf{R}$.
- (ii) $\int_{-\infty}^{\infty} f_X(x) dx = 1$.
- (iii) $f_X(x)$ anger hur mycket sannolikhetsmassa det finns per längdenhet i punkten x .



Strikt olikhet eller inte?

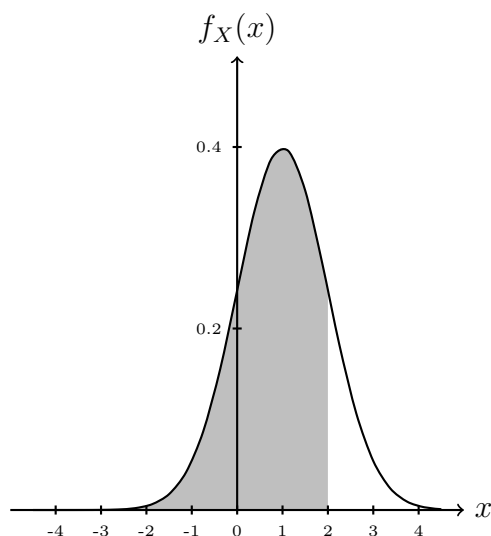
Om X är en kontinuerlig variabel, så är $P(X < x) = P(X \leq x)$. Detta följer från att integralen inte gör någon skillnad på om ändpunkten är med eller ej. Vi kan till och med definiera om funktionen i uppräknligt många punkter (även mer, men det kräver lite måttteori för att definiera) utan att ändra sannolikheten. Detta gäller dock *absolut inte* i det diskreta fallet.

Vi definierar **fördelningsfunktionen** $F_X(x)$ på samma sätt som i det diskreta fallet, och finner att

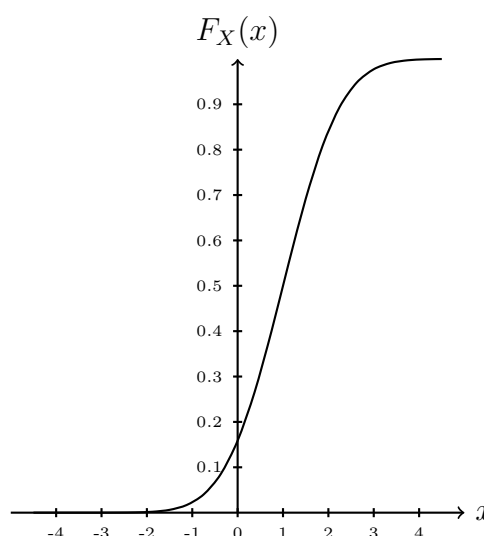
$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt, \quad x \in \mathbf{R}.$$

Fördelningsfunktionen uppfyller (i)–(iii) från det diskreta fallet, och i alla punkter där $f_X(x)$ är kontinuerlig gäller dessutom att $F'_X(x) = f_X(x)$. Det sista är i princip analysens huvudsats. Man kan fundera över hur pass diskontinuerlig f_X skulle kunna vara, men som exemplet ovan visar finns det inte så mycket begränsningar på det. I denna kurs kommer dock de flesta kontinuerliga fördelningar ha täthetsfunktioner som är kontinuerliga för det mesta.

Exempel på hur en täthetsfunktion och motsvarande fördelningsfunktion kan se ut:



Täthet: Hur "sannolikhetsmassan" är fördelad. Skuggad area är $P(X \leq 2) = F_X(2)$.



Fördelningsfunktionen är växande och gränsvärdena mot $\pm\infty$ verkar stämma!



Exempel

Låt $f_1(x) = x^2 + bx$ och $f_2(x) = c(x^3 + x)$ båda för $x \in [0, 2]$. Om det går, bestäm konstanterna b och c så dessa blir täthetsfunktioner och beräkna sannolikheten att respektive variabel är ≤ 1 .

Lösning: Vi börjar med f_1 :

$$1 = \int_0^2 (x^2 + bx) dx = \frac{8}{3} + 2b \quad \Rightarrow \quad b = -5/6.$$

Men om b är negativ kommer $f_1(x)$ att vara negativ för x nära noll (x^2 termen går mot noll snabbare än x). Detta kan alltså *inte* vara en täthetsfunktion. Vi testar f_2 :

$$1 = c \int_0^2 (x^3 + x) dx = 6c \quad \Rightarrow \quad c = 1/6.$$

Det är även klart att $f_2(x) \geq 0$ för alla $x \in [0, 2]$. Med $c = 1/6$ är alltså f_2 en täthetsfunktion. Den eftersökta sannolikheten kan beräknas enligt

$$P(X \leq 1) = \int_0^1 \frac{1}{6} (x^3 + x) dx = \frac{1}{8}.$$

2.5 Oberoende stokastiska variabler

Precis som när vi pratade om oberoende händelser kallar vi två stycken stokastiska variabler X och Y för oberoende om

$$P(X \leq x \text{ och } Y \leq y) = P(X \leq x)P(Y \leq y) \quad \text{för alla } x, y \in \mathbf{R}.$$

Vi betraktar ett exempel med kontinuerliga variabler. Det fungerar analogt i det diskreta fallet.

**Exempel**

Låt X vara en s.v. med täthetsfunktionen $f_X(x) = 2e^{-2x}$, $x \geq 0$ och låt Y vara en s.v. med täthetsfunktionen $f_Y(x) = 3e^{-3x}$, $x \geq 0$. Antag att X och Y är oberoende och ställ upp ett uttryck för $P(X \leq x, Y \leq y)$.

Lösning. Vi har täthetsfunktionerna så för $x, y \geq 0$:

$$P(X \leq x) = \int_0^x 2e^{-2t} dt = [-e^{-2t}]_0^x = 1 - e^{-2x}$$

och

$$P(Y \leq y) = \int_0^y 3e^{-3t} dt = [-e^{-3t}]_0^y = 1 - e^{-3y}.$$

Eftersom X och Y är oberoende så gäller att

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = (1 - e^{-2x})(1 - e^{-3y})$$

om $x, y \geq 0$. Om $x < 0$ eller $y < 0$ blir sannolikheten 0. Hur skulle man kunna räkna ut sannolikheten om X och Y är beroende? Omöjligt att svara på utan att veta hur beroendet ser ut!

Om vi har flera stokastiska variabler så utvidgar vi begreppet oberoende variabler naturligt. Vi kallar X_1, X_2, \dots, X_n för oberoende om

$$P(X_1 \leq x_1 \text{ och } X_2 \leq x_2 \text{ och } \dots \text{ och } X_n \leq x_n) = P(X_1 \leq x_1)P(X_2 \leq x_2) \cdots P(X_n \leq x_n)$$

för alla $x_1, x_2, \dots, x_n \in \mathbf{R}$. Vi återkommer till karaktärisering av oberoende variabler senare i kursen.

2.6 (★★) Vad är mer exakt en stokastisk variabel?

För att kunna precisera vad för slags funktion (för det är en funktion) en stokastisk variabel är, behöver vi diskutera öppna mängder på den reella axeln \mathbf{R} .



Definition. Den minsta (minst antal element) σ -algebran på \mathbf{R} som innehåller *alla* öppna intervall betecknar vi med \mathcal{B} . Denna algebra brukar kallas för **Borel- σ -algebran** på \mathbf{R} .

Algebran \mathcal{B} innehåller alltså alla mängder av typen $(a, b) \subset \mathbf{R}$, $(-\infty, c) \cup (d, \infty) \subset \mathbf{R}$, komplement av sådana mängder, samt alla uppräknliga unioner av mängder av föregående typ. Detta är ganska tekniskt, och inget vi kommer att arbeta med direkt. Men för att få en korrekt definition behövs begreppet.



Stokastisk variabel

Definition. En **stokastisk variabel** är en reellvärd funktion definierad på ett utfallsrum Ω . Funktionen X avbildar alltså olika utfall på reella tal; $X: \Omega \rightarrow \mathbf{R}$.

Mer precist så kräver vi att $X^{-1}(B) \in \mathcal{F}$ för alla $B \in \mathcal{B}$. Mängden $X^{-1}(B)$ definieras som $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ och kallas för **urbilden** av B . Mängden består alltså av alla $\omega \in \Omega$ som avbildas in i B .

Varför kravet att urbilden $X^{-1}(B)$ skall tillhöra de tillåtna händelserna? Det faller sig ganska naturligt, då $X^{-1}(B)$ är precis de utfall i Ω som avbildas in i mängden B . Således vill vi gärna att denna samling utfall verkligen utgör en händelse, annars kan vi inte prata om någon sannolikhet för denna samling utfall.

Det viktigaste att ta med sig från denna definition är att urbilden $X^{-1}(B)$ av en delmängd $B \subset \mathbf{R}$ består av *alla* utfall $\omega \in \Omega$ så att siffran $X(\omega)$ ligger i mängden B och att om B är en snäll mängd så måste $X^{-1}(B)$ också vara en snäll mängd.

2.7 (★) Kontinuerliga s.v. med täthetsfunktion

Definitionen kanske ser oskyldig ut, men här finns det både hundar och ugglor begravda i mossen (som säkert ligger i Danmark). Problemet ligger i integralbegreppet och hur generella händelser vi vill tillåta. I grundanalysen introducerar man Riemann-integralen, men tyvärr räcker den inte riktigt till för allt. Betrakta följande funktion: $f(x) = 0$ om $x < 0$, $x > 1$, eller om x är rationell (dvs ett bråk p/q av heltal p och q). I övriga punkter är $f(x) = 1$ (dvs på alla irrationella punkter i intervallet $[0, 1]$).

Man kan tänka sig den stokastiska variabeln X som indikerar om ett slumptal mellan noll och ett är irrationellt eller inte. Kanske skulle täthetsfunktionen då ges av $f(x)$ ovan, men är detta verkligen en täthetsfunktion? Den är icke-negativ, så den biten är OK. Men har den "area" ett?? Av nödvändighet kommer alla undertrappor till $f(x)$ på $[0, 1]$ att vara identiskt lika med noll, och på samma sätt är alla övertrappor identiskt lika med ett. Vi kan alltså aldrig approximera funktionen med över- och undertrappor. Således är $f(x)$ *inte* Riemann-integrerbar. Så hur löser man detta? Med ett nytt integralbegrepp (Lebesgueintegralen) smidigt nog, där det visar sig att integralen av f mycket riktigt blir ett.

Lebesgueintegralen konstrueras på ett annorlunda sätt i jämförelse med Riemannintegralen. Istället för att bara stycka upp definitionsområdet (dvs x -axeln) i finare och finare likadana

bitar och försöka approximera integralen med arean av rektanglar (över- och undertrappor), så styckar vi istället upp värdemängden. Genom att approximera funktionen med så kallade *enkla funktioner* – funktioner som är konstant på ett ändligt antal mätbara mängder och lika med noll annars – så kan man komma åt betydligt fler funktioner. Mätbarheten här blir i en sådan här kurs med avseende på det sannolikhetsmått man är intresserad av, så sannolikheten kommer in på ett väldigt naturligt sätt. Detta ligger dock utanför ramarna för denna kurs. Men termen *integrerbar* i definitionen syftar på denna ”nya” typ av integral.

För *snälla* funktioner (funktioner som till exempel bara har uppräknligt många diskontinuiteter) så sammanfaller de båda integralbegreppen. Vi kommer alltså inte att fundera så mycket mer på detta.

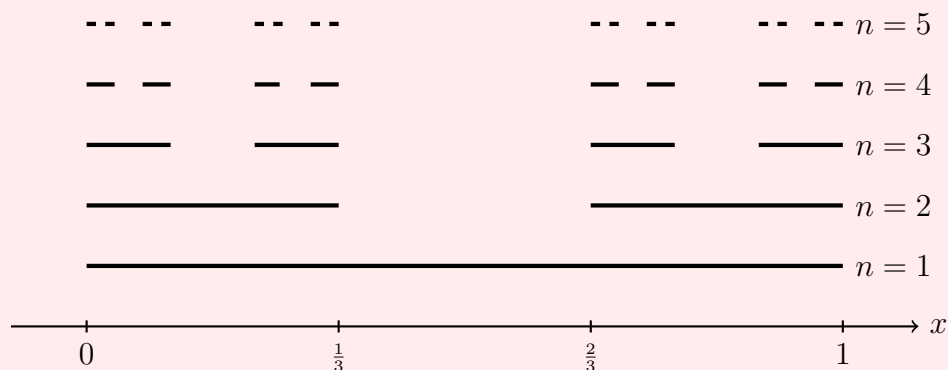
2.8 (★★★)Singulära stokastiska variabler?

Vi har delat upp klassen av stokastiska variabler i två delar. En del vi kallat diskret, där vi använt oss av en sannolikhetsfunktion $P(X = k)$, och en kontinuerlig där vi *krävt* att det finns en täthetsfunktion. Detta krav är emellertid inte alltid uppfyllt. Det finns icke-diskreta stokastiska variabler som inte har någon täthetsfunktion. Variabler av denna typ brukar sägas vara singulära. Ibland så kräver man till och med definitionsmissigt att en stokastisk variabel ska ha en täthetsfunktion för att kallas kontinuerlig. Så vad gör man åt situationen att det verkar finnas andra djur i djungeln? Dessa blir tyvärr svåra att hantera med den teori vi har tillgång till just nu, men låt oss titta på ett välkänt exempel som åtminstone visar på troligheten att singulära fördelningar finns.



Cantormängden

Låt oss börja med något kul, nämligen en tämligen sönderstyckad delmängd av intervallet $[0, 1]$. Processen fungerar enligt följande. Vi delar $C_1 = [0, 1]$ i tre delar och tar bort den öppna mittersta delen. Vi kallar de punkter som är kvar (dvs $[0, 1/3] \cup [2/3, 1]$) för C_2 . Vi delar nu de två intervallen i C_2 i tre likadana delar vardera och tar bort mittensegmenten. Kalla den nya mängden C_3 . Sen upprepar vi processen gång på gång och kallar mängden som är kvar vid steg k för C_k .



Vi noterar att $C_1 \supset C_2 \supset C_3 \supset \dots$ och definierar $C = \bigcap_{k=1}^{\infty} C_k$. Detta är **Cantormängden**.

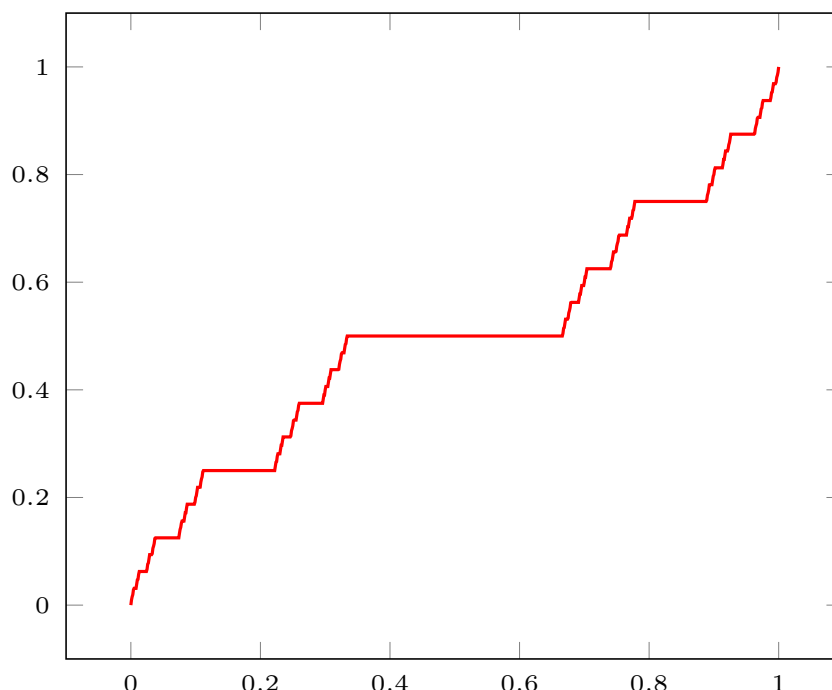
Så vad ska vi med denna mängd till? Givetvis att konstruera något intressant motexempel.

Man kan se att om vi representerar ett tal $x \in [0, 1]$ i basen 3, dvs vi skriver $x = \sum_{k=1}^{\infty} x_k 3^{-k}$

där varje $x_k = 0, 1$ eller 2 , så kommer $x_k \neq 1$ för alla k om $x \in C$. Vi tar hela tiden bort mittendelen, vilket gör att vi aldrig får motsvarande siffra i expansionen i basen 3. Den så kallade Cantorfördelningen erhåller vi nu om vi definierar fördelningsfunktionen enligt följande:

$$F(x) = \frac{1}{2} \sum_{k=1}^{\infty} \frac{x_k}{2^k}, \quad x \in C.$$

För $x < 0$ definierar vi $F(x) = 0$ och för $x > 1$ definierar vi $F(x) = 1$. Vi täpper till hålen vi lämnat i $[0, 1]$ genom att observera att $F(x_1) \leq F(x_2)$ för $x_1 \leq x_2$, så F är växande. Vi låter $F(x) = \sup\{F(y) : y \in C, y < x\}$ för $x \in [0, 1] \setminus C$, vilket definierar $F(x)$ för alla $x \in \mathbf{R}$ så att F är kontinuerlig. Observera att denna definition innebär att $F(x)$ är konstant på de linjesegment som vi tog bort när vi konstruerade C . Vi ser också att $F(x) \rightarrow 1$ då $x \rightarrow 1$ och $F(x) \rightarrow 0$ då $x \rightarrow 0$. Denna funktion brukar gå under namnet *the Devil's staircase* – Djävulens trappa – just för att den har ett par roliga egenskaper som går lite mot intuitionen.



Detta är alltså en fördelningsfunktion. Så vad var poängen med detta? Jo, poängen är den att en stokastisk variabel X med fördelningsfunktionen F inte är diskret men saknar täthetsfunktion! Yikes. Att X inte är diskret följer av att F är kontinuerlig (det blir alltid hopp i F om variabeln är diskret). Så varför finns ingen täthetsfunktion? Det följer av att $F(x)$ är konstant nästan överallt. När vi konstruerade C så tog vi hela tiden bort tredjedelar av linjesegment, så om vi tittar på figuren ovan där vi konstruerade Cantormängden så ser vi att vi tar bort längderna

$$\frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \cdots = \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k = \frac{1}{3} \cdot \frac{1}{1 - 2/3} = 1.$$

Huh? Då är ju inget kvar? Nja, inte direkt. Dels är argumentet ovan lite handviftande och dels så kan vi fortfarande ha kvar en synnerligen sönderstyckad mängd där varje punkt är hyfsat isolerad så att det inte blir några intervall. Faktum är att Cantormängden faktiskt inte är uppräkningsbar (så betydligt större än till exempel $[0, 1] \cap \mathbf{Q}$). Men viktigast för oss, derivatan

av fördelningsfunktionen måste vara noll nästan överallt eftersom funktionen är konstant där. Därför kan vi inte ha någon täthetsfunktion.

Så vad var poängen med allt det där? En poäng är att vi egentligen borde sträva efter att formulera saker utan att explicit välja summa eller integral av täthetsfunktion. Kan vi formulera och bevisa resultat där vi endast använder sannolikhetsmåttet (och fördelningsfunktionen) så får vi resultat som gäller generellt.

Kapitel 3

Stokastiska variabler (forts)

3.1 Väntevärde



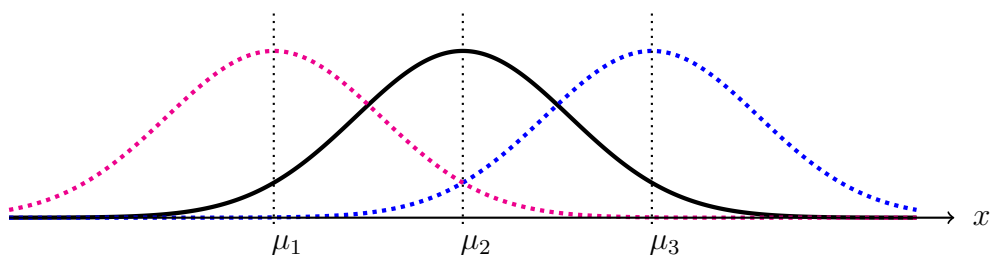
Väntevärde

Definition. Väntevärdet $E(X)$ av en stokastisk variabel X definieras som

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad \text{respektive} \quad E(X) = \sum_k k p_X(k)$$

för kontinuerliga (om täthetsfunktion finns) och diskreta variabler.

Andra vanliga beteckningar: μ eller μ_X . Väntevärdet är ett lägesmått som anger var sannolikhetsmassan har sin tyngdpunkt (jämför med mekanikens beräkningar av tyngdpunkt). Om fördelningen är symmetrisk blir det i mitten, men är fördelningen skev blir det annorlunda. I figuren nedan ser vi tre täthetsfunktioner som har samma form men olika väntevärden. De är helt enkelt translationer av samma funktion i detta fall.



Exempel

Vid tillämpningar tolkas ofta väntevärdet som just det *förväntade värdet* för en stokastisk variabel.

- (i) Om X är koncentrationen i en flaska salpetersyra så är $E(X)$ den koncentration vi förväntar oss när vi tar ned flaskan från hyllan.
- (ii) Om X är antalet kast med en tärning innan vi får en 6:a för första gången så är $E(X)$ det förväntade antalet kast innan vi ser den första 6:an.

Observera att väntevärdet är ett reellt tal, så det kan mycket väl vara så att en variabel (då oftast en diskret sådan) inte *kan* anta sitt väntevärde.



Exempel

Kasta en 4-sidig tärning och låt X vara utfallet 1, 2, 3 eller 4. Beräkna $E(X)$.

Lösning. Vi antar att tärningen är ärlig så $p_X(k) = 1/4$ för $k = 1, 2, 3, 4$. Då blir

$$E(X) = \sum_{k=1}^4 k p_X(k) = \frac{1 + 2 + 3 + 4}{4} = \frac{5}{2}.$$

Det förväntade resultatet är alltså 2.5. Knappast ett resultat vi förväntar oss vid ett enskilt kast!



Vad är *egentligen* ett väntevärde?

Vi har gjort en definition av begreppet väntevärde ovan, så det är den som gäller. Men åtminstone följande tolkningar eller alternativa definitioner finns.

- (i) Ett sannolikhetsviktat medelvärde av de värden X kan anta.
- (ii) Integralen av X med avseende på sannolikhetsmåttet: $\int_{\Omega} X(\omega) dP(\omega)$ (vad nu detta betyder).
- (iii) Masscentrum för sannolikhetsfördelningen.
- (iv) Det värde X hamnar på i snitt vid väldigt många upprepningar.

Vad punkt (ii) betyder kräver mer analys än vi har tillgång till.

3.1.1 Funktioner och väntevärden

Vad händer om vi har en funktion av en stokastisk variabel, säg att $Y = g(X)$, där vi känner till fördelningen för X och hur funktionen g ser ut? Om g är snäll så ger detta upphov till en ny stokastisk variabel och ibland kan man explicit härleda fördelning för denna (vi återkommer till detta senare), men faktum är att vi kan hantera funktioner av stokastiska variabler på ett smidigare sätt om vi endast behöver väntevärdet.

Följande sats (ofta känd som *the law of the unconscious statistician*) visar hur detta fungerar, men resultatet behöver egentligen lite diskussion (samt ett bevis).



Väntevärde och funktioner av stokastiska variabler

Sats. Låt $Y = g(X)$ där g är snäll. Då gäller

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \text{och} \quad E(Y) = \sum_k g(k) p_X(k)$$

för Y kontinuerlig (med täthetsfunktion) respektive diskret.

**Exempel**

Låt X vara utfallet vid ett tärningskast med en symmetrisk 4-sidig tärning. Beräkna $E(X^2)$.

Lösning. Vi söker $E(X^2)$, så $g(t) = t^2$ är funktionen som transformerar $Y = g(X)$. Enligt satsen ovan blir då

$$E(X^2) = \sum_{k=1}^4 k^2 p_X(k) = \frac{1^2 + 2^2 + 3^2 + 4^2}{4} = \frac{15}{2}.$$

Notera speciellt att $E(X^2) \neq E(X)^2 = (5/2)^2$.

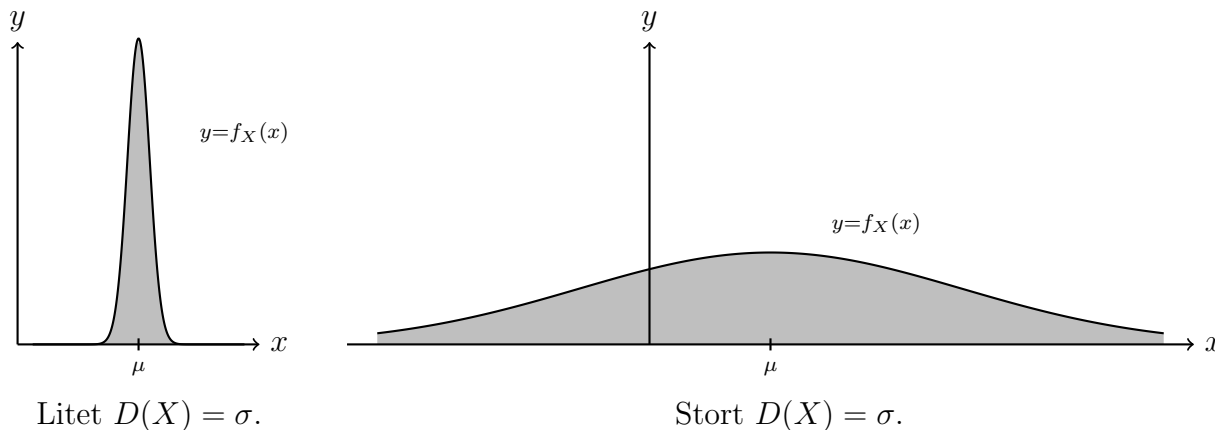
3.2 Varians och standardavvikelse

**Varians och standardavvikelse**

Definition. Låt X vara en stokastisk variabel med $|E(X)| < \infty$. **Variansen** $V(X)$ definieras som $V(X) = E((X - E(X))^2)$. **Standardavvikelsen** $D(X)$ definieras som $D(X) = \sqrt{V(X)}$.

Andra vanliga beteckningar för standardavvikelsen: σ , σ_X , $\sigma(X)$.

Variansen är ett spridningsmått. Stor varians (eller standardavvikelse) betyder att sannolikhetsfördelning har stor spridning. Många värden är troliga. Liten varians betyder att fördelningen är centrerad, hög sannolikhet att hamna kring en viss punkt; se figuren nedan.

**Steiners sats**

Sats. $V(X) = E(X^2) - E(X)^2$.

**Exempel**

Låt X_1 anta värdena $\{-1, 1\}$ med $p_{X_1}(-1) = p_{X_1}(1) = 1/2$ och låt X_2 anta värdena $\{-10, 10\}$ med $p_{X_2}(-10) = p_{X_2}(10) = 1/2$. Beräkna $V(X_1)$ och $V(X_2)$.

Lösning: Det är klart att $E(X_1) = E(X_2) = 0$ (varför?), så

$$V(X_1) = E(X_1^2) - E(X_1)^2 = \frac{1}{2}(-1)^2 + \frac{1}{2}1^2 - 0 = 1$$

och

$$V(X_2) = E(X_2^2) - E(X_2)^2 = \frac{1}{2}(-10)^2 + \frac{1}{2}(10)^2 - 0 = 50 + 50 = 100.$$

Tydligt att X_2 har mycket större varians även om fördelningarna kan tyckas se snarlika ut, men sannolikheten är mycket mer utspridd för X_2 .

3.3 Räknelagar

För väntevärdet gäller bland annat följande regler.



Linjäritet och oberoende produkt

Låt X och Y vara stokastiska variabler. Då gäller

- (i) $E(aX + b) = aE(X) + b$ för alla $a, b \in \mathbf{R}$;
- (ii) $E(aX + bY) = aE(X) + bE(Y)$ för alla $a, b \in \mathbf{R}$;
- (iii) Om X och Y är oberoende gäller $E(XY) = E(X)E(Y)$.

Vi visar dessa räkneregler genom att sätta in allt i definitionen och sedan utnyttja att både integralen och summan är linjära operationer (så vi kan dela upp över plus-tecknet och bryta ut konstanter).

För variansen kan vi visa istället visa följande. Här blir beviset lite bökgigare på grund av att vi inte längre arbetar med en så kallad linjär operator, men tanken är snarlik (sätt in i definitionen och se vad som händer).



Låt X och Y vara stokastiska variabler. Då gäller

- (i) $V(aX + b) = a^2V(X)$ för alla $a, b \in \mathbf{R}$;
- (ii) $V(aX \pm bY) = a^2V(X) + b^2V(Y) + 2ab(E(XY) - E(X)E(Y))$ för alla $a, b \in \mathbf{R}$;
- (iii) Om X och Y är oberoende gäller $V(aX \pm bY) = a^2V(X) + b^2V(Y)$.



Varianser adderas alltid!

Observera att det *alltid* blir ett plustecken mellan varianserna för linjär-kombinationer:

$$V(aX \pm bY) = a^2V(X) + b^2V(Y).$$

Vi kommer *aldrig* att bilda skillnader mellan varianser!

Det följer att standardavvikelsen för en linjärkombination $aX + bY$ av två oberoende stokastiska variabler ges av $\sigma_{aX+bY} = \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}$.

3.4 Vanliga kontinuerliga fördelningar

Analogt med diskreta variabler definieras de kontinuerliga ofta från sina respektive täthetsfunktioner. Vi definierar några av de vanligaste. Det finns många andra fördelningar som ofta används, men dessa är de vi kommer att använda mest. Se boken för fler exempel (Gammafördelning, Weibullfördelning, χ^2 -fördelning, t-fördelning mfl.)



Likformig fördelning

Variabeln X kallas **likformigt** fördelad (eller **rektangel-**), $X \sim U(a, b)$ eller $X \sim \text{Re}(a, b)$, om

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{övriga } x \end{cases}$$



Exempel

Ubbe håller upp Whisky i sitt glas. Vätskenivån är likformigt fördelad mellan två och fem fingrar. Vad är sannolikheten att Ubbe håller upp mindre än 3.2 fingrar?

Lösning: Låt $X \sim \text{Re}(2, 5)$ vara vätskenivån. Vi söker $P(X < 3.2)$:

$$P(X < 3.2) = \int_2^{3.2} \frac{1}{5-2} dx = \frac{1}{3} (3.2 - 2) = 0.4.$$



Exponentialfördelning

Variabeln X kallas **exponentialfördelad** med parametern $\lambda > 0$, $X \sim \text{Exp}(\lambda)$, om

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x), & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Parametern λ tolkas ibland som *intensiteten*.



Exempel

Låt X vara väntetiden i en telefonkö (minuter). Av någon anledning har det visat sig att X har en täthetsfunktion $f_X(x) = c e^{-0.05x}$ för $x \geq 0$, där c är en konstant.

- (i) Bestäm c så att f_X blir en täthetsfunktion.
- (ii) Vad är sannolikheten att få vänta i mer än 50 minuter vid ett samtal?
- (iii) Om man ringer 10 olika (oberoende) samtal, vad är sannolikheten att högst ett av dessa har en väntetid på över 50 minuter?

Lösning:

$$(i) \quad 1 = \int_{-\infty}^{\infty} f_X(x) dx = c \int_0^{\infty} e^{-0.05x} dx = \frac{c}{-0.05} [e^{-0.05x}]_0^{\infty} = \frac{c}{20}, \text{ så } c = 1/20.$$

$$(ii) \quad P(X > 50) = \int_{50}^{\infty} f_X(x) dx = \frac{1}{20} \left[\frac{e^{-0.05x}}{-0.05} \right] = e^{-5/2} \approx 0.082.$$

(iii) Varje samtal har sannolikheten $e^{-5/2}$ att ha mer än 50 minuters väntetid. Antalet Y av tio stycken samtal som har mer än 50 minuters väntetid blir alltså Binomialfördelad med $n = 10$ och $p = e^{-5/2}$. Vi erhåller

$$\begin{aligned} P(Y \leq 1) &= \sum_{k=0}^1 \binom{10}{k} (e^{-5/2})^k (1 - e^{-5/2})^{10-k} \\ &= (1 - e^{-5/2})^{10} + 10e^{-5/2}(1 - e^{-5/2})^9 \approx 0.804. \end{aligned}$$

**Exempel**

En komponent (som inte åldras) antas ha en livslängd T som är $\text{Exp}(1/100)$ -fördelad (enhet: dagar).

- (i) Vad är sannolikheten att komponenten går sönder innan 80 dagar?
- (ii) Givet att komponenten överlevt 80 dagar, vad är sannolikheten att den klarar 100 dagar?

Lösning:

$$(i) \quad P(T \leq 80) = \int_{-\infty}^{80} f_X(x) dx = \frac{1}{100} \int_0^{80} e^{-x/100} dx = \frac{1}{100} \left[\frac{e^{-x/100}}{-1/100} \right]_0^{80} = 1 - e^{-4/5}. \text{ Sannolikheten blir alltså ca } 55.1\%.$$

(ii) Här använder vi definitionen av betingad sannolikhet och erhåller

$$\begin{aligned} P(T \geq 100 \mid T \geq 80) &= \frac{P(\{T \geq 100\} \cap \{T \geq 80\})}{P(T \geq 80)} = \frac{P(T \geq 100)}{P(T \geq 80)} \\ &= \frac{\frac{1}{100} \int_{100}^{\infty} e^{-x/100} dx}{\frac{1}{100} \int_{80}^{\infty} e^{-x/100} dx} = \frac{e^{-100/100}}{e^{-80/100}} = e^{-1/5} \approx 0.8187. \end{aligned}$$

Detta är ett exempel på en *betingad fördelning*. Observera även att denna sannolikhet är densamma som

$$P(T \geq 20) = \frac{1}{100} \int_{20}^{\infty} e^{-x/100} dx = e^{-1/5}.$$

Detta gäller generellt för exponentialfördelningen. Sannolikheten att komponenten klarar 20 dagar är oberoende av hur länge den levtt tidigare. Kanske inte alltid rimligt för komponenter?

3.5 Median

Ett annat lägesmått än väntevärdet är medianen.



Median

Definition. En **median** för en stokastisk variabel X är ett tal $m \in \mathbf{R}$ så att

$$P(X \leq m) = P(X \geq m) = \frac{1}{2}.$$

Observera att medianen inte behöver vara entydig!



Medianen för en Exponentialfördelning

Låt $X \sim \text{Exp}(\lambda)$. Beräkna medianen och väntevärdet för X .

Lösning: Vi räknar ut fördelningsfunktionen för X . Om $x > 0$,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}.$$

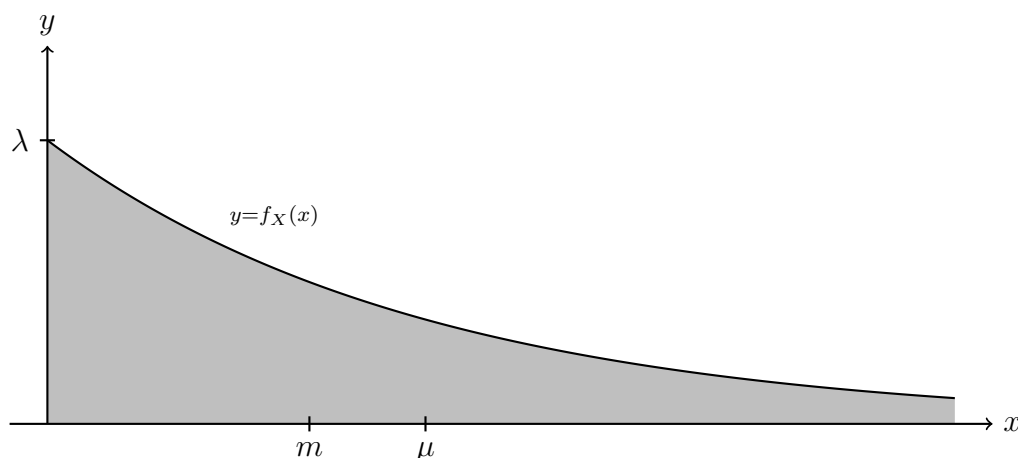
Medianen finner vi ur ekvationen $F_X(m) = 1/2$, dvs

$$1 - e^{-\lambda m} = \frac{1}{2} \quad \Leftrightarrow \quad e^{-\lambda m} = \frac{1}{2} \quad \Leftrightarrow \quad m = \frac{\ln 2}{\lambda}.$$

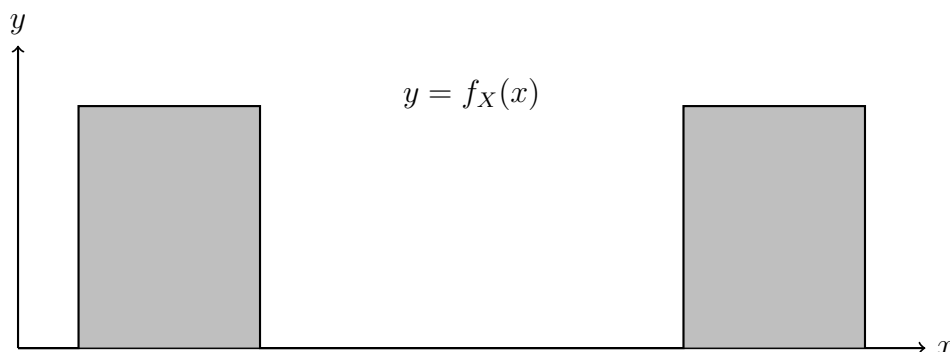
Jämför detta med väntevärdet för X :

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = 0 - 0 + \left[-\frac{e^{-\lambda x}}{\lambda} \right]_0^{\infty} = \frac{1}{\lambda}.$$

Medianen och väntevärdet behöver alltså *inte* vara samma sak!



Ett annat exempel, där medianen *inte* är entydigt definierad:



Var är medianen??

3.6 Normalfördelning



Normalfördelning

Variabeln X kallas **normalfördelad** med parametrarna μ och σ , $X \sim N(\mu, \sigma)$, om

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbf{R}.$$

Om $\mu = 0$ och $\sigma = 1$ kallar vi X för **standardiserad**, och i det fallet betecknar vi täthetsfunktionen med

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbf{R}.$$

Fördelningsfunktionen för en normalfördelad variabel ges av

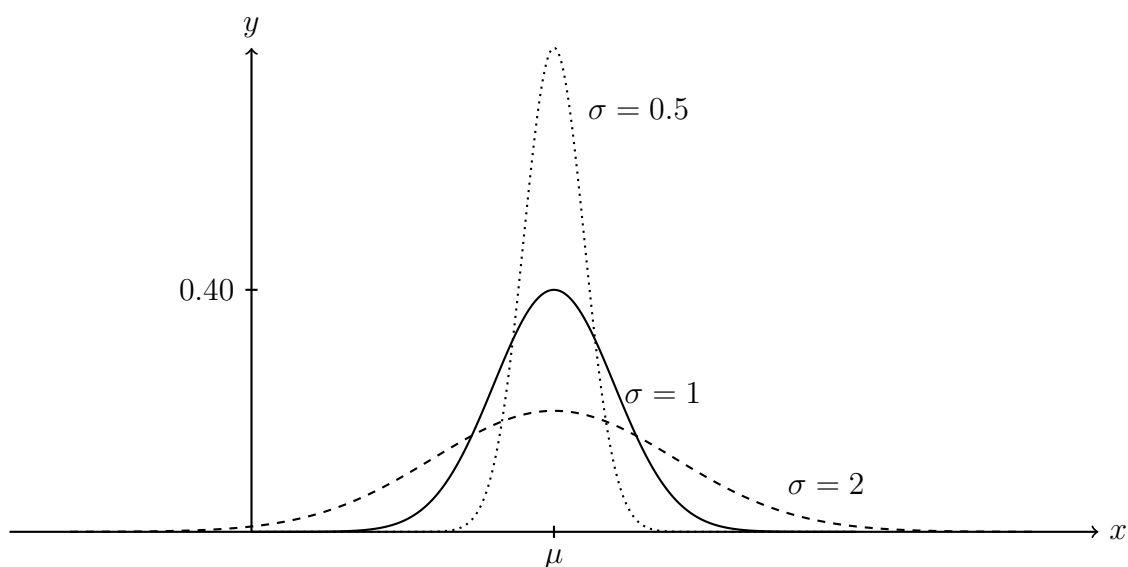
$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du, \quad x \in \mathbf{R},$$

och även här döper vi speciellt den standardiserade fördelningsfunktionen till

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{u^2}{2}\right) du, \quad x \in \mathbf{R}.$$

Är det någon som kommer ihåg vad som hände när man försökte integrera e^{x^2} i envariabelanalysen? Man kan visa att det inte går att uttrycka den primitiva funktionen i elementära funktioner, utan man definierar helt enkelt en *ny* funktion utifrån den bestämda integralen (slå upp *erf*-funktionen i något matematiskt uppslagsverk). Vad detta innebär för oss är att vi kommer att använda tabell för att beräkna numeriska värden för uttryck som innehåller funktionen Φ .

Vi kommer visa att $E(X) = \mu$ och $V(X) = \sigma^2$ om $X \sim N(\mu, \sigma)$.





Standardavvikelse eller varians?

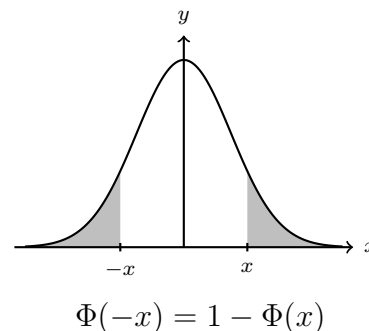
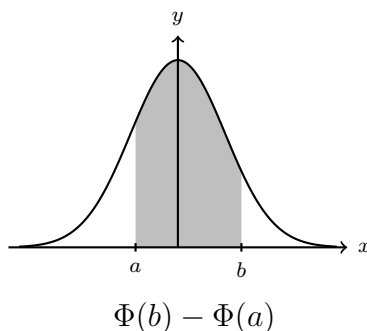
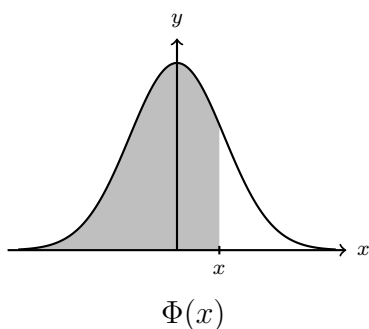
I kursboken (Blom et al) används beteckningen $X \sim N(\mu, \sigma)$, så precis som i vårt fall ovan är den andra parametern är alltså standardavvikelsen σ , inte variansen σ^2 . Varför ta upp detta? I mycket av litteraturen så används variansen som andra parameter. Var försiktig när ni slår upp saker eller använder färdiga formler!



Bruk av tabell för $\Phi(x)$

Låt $X \sim N(0, 1)$. Då gäller

- (i) $P(X \leq x) = \Phi(x)$ för alla $x \in \mathbf{R}$;
- (ii) $P(a \leq X \leq b) = \Phi(b) - \Phi(a)$ för alla $a, b \in \mathbf{R}$ med $a \leq b$;
- (iii) $\Phi(-x) = 1 - \Phi(x)$ för alla $x \in \mathbf{R}$.



Exempel

Låt $X \sim N(0, 1)$. Bestäm $P(X \leq 1)$, $P(X < 1)$, $P(X \leq -1)$, samt $P(0 < X \leq 1)$.

Lösning. Direkt ur tabell, $P(X \leq 1) = \Phi(1) \approx 0.8413$. Eftersom X är kontinuerlig kvittar det om olikheterna är strikta eller inte, så $P(X < 1) = P(X \leq 1) = \Phi(1)$ igen. Vidare har vi

$$P(X \leq -1) = \Phi(-1) = 1 - \Phi(1) = 0.1587$$

och $P(0 < X \leq 1) = \Phi(1) - \Phi(0) = 0.8413 - 0.5 = 0.3413$.

3.7 (★★) Väntevärde för geometrisk fördelning

Följande avsnitt kräver att man arbetat en del med potensserier (kapitel 10 i envariabelboken).



Exempel

Låt X anta värdena $0, 1, 2, \dots$ med sannolikheterna $p_X(k) = 2^{-k-1}$. Beräkna $E(X)$ och $V(X)$.

Lösning: Till att börja med kan vi kontrollera att p_X verkligen är en sannolikhetsfunktion. Klart att $p_X(k) \geq 0$, och summan nedan är geometrisk så

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} 2^{-k-1} = 2^{-1} \sum_{k=0}^{\infty} 2^{-k} = 2^{-1} \cdot \frac{1}{1-1/2} = 1.$$

Vi beräknar väntevärdet. Låt $q \in]0, 1[$ så $p_X(k) = (1-q)q^k$ med $q = 1/2$. Vi kan beräkna summan av kq^k genom följande manöver:

$$\sum_{k=0}^{\infty} kq^k = \sum_{k=1}^{\infty} kq^k = q \sum_{k=1}^{\infty} kq^{k-1} = q \sum_{k=0}^{\infty} \frac{d}{dq} q^k = q \frac{d}{dq} \sum_{k=0}^{\infty} q^k = q \frac{d}{dq} \frac{1}{1-q} = \frac{q}{(1-q)^2}.$$

Alltså blir

$$E(X) = (1-q) \sum_{k=0}^{\infty} kq^k = \frac{q}{1-q} \quad \text{och} \quad E(X) = 1 \text{ om } q = \frac{1}{2}.$$

För att beräkna $E(X^2)$ kikar vi på motsvarande kalkyl för andraderivatans:

$$q^2 \sum_{k=0}^{\infty} \frac{d^2}{dq^2} q^k = q^2 \sum_{k=0}^{\infty} (k^2 - k)q^{k-2} = \sum_{k=0}^{\infty} (k^2 - k)q^k = \sum_{k=0}^{\infty} k^2 q^k - \frac{q}{(1-q)^2}.$$

Således,

$$\sum_{k=0}^{\infty} k^2 q^k = \frac{q}{(1-q)^2} + q^2 \frac{d^2}{dq^2} \sum_{k=0}^{\infty} q^k = \frac{q}{(1-q)^2} + \frac{2q^2}{(1-q)^3},$$

vilket medför att

$$E(X^2) = (1-q) \sum_{k=0}^{\infty} k^2 q^k = \frac{q}{1-q} + \frac{2q^2}{(1-q)^2}$$

så

$$V(X) = E(X^2) - E(X)^2 = \frac{q}{1-q} + \frac{q^2}{(1-q)^2} = \frac{q}{(1-q)^2}.$$

och med $q = 1/2$ får vi $V(X) = 2$. Den observante läsaren kanske känner igen sannolikhetsfunktionen vi arbetar med då det är den geometriska fördelningen $X \sim \text{Geo}(1-q)$. Så vad vi visat ovan är följande:



Geometrisk fördelning

Sats. Om $X \sim \text{Geo}(p)$ så är $E(X) = \frac{1-p}{p}$ och $V(X) = \frac{1-p}{p^2}$.

Kapitel 4

Normalfördelning och CGS

Vi kommer nu fortsätta att studera normalfördelningen. Detta är en mycket viktig fördelning i tillämpningar och vi kommer i avsnittet med centrala gränsvärdessatsen (CGS) att belysa fenomenet att i princip alla summor av många stokastiska variabler tenderar att bli normalfördelade.

4.1 Normalfördelningen (forts)

Eftersom det är bökigt att hantera integraler av uttryck som innehåller $\exp(-x^2)$ så fokuserar vi istället på hur vi kan transformera problemet till $N(0, 1)$ -fallet och där helt enkelt använda tabell för att beräkna sannolikheter.



Standardisering av variabel

Om X är en stokastisk variabel med $E(X) = \mu$ och $V(X) = \sigma^2$, så är $Z = (X - \mu)/\sigma$ en stokastisk variabel med $E(Z) = 0$ och $V(Z) = 1$. Vi kallar Z för **standardiserad**.

Beviset för att $E(Z) = 0$ följer direkt från linjäriteten hos väntevärdet. Variansen kan ses genom följande argument:

$$\begin{aligned} V((X - \mu)/\sigma) &= E((X - \mu)^2/\sigma^2) - 0^2 = \frac{1}{\sigma^2} (E(X^2) - 2\mu E(X) + \mu^2) \\ &= \frac{1}{\sigma^2} (E(X^2) - E(X)^2) = \frac{\sigma^2}{\sigma^2} = 1. \end{aligned}$$



Standardiserad normalfördelning

Sats. Om $X \sim N(0, 1)$ så är $E(X) = 0$ och $V(X) = 1$.

Eftersom $e^{-x^2/2}$ går mot noll väldigt snabbt, så kommer

$$\int_{-\infty}^{\infty} |x\varphi(x)| dx < \infty.$$

Alltså måste

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\varphi(x) dx = 0$$

eftersom $x \mapsto x\varphi(x)$ är en udda funktion. På liknande sätt ser vi att

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left(x e^{-x^2/2} \right) dx \\ &= \frac{1}{\sqrt{2\pi}} \left(\left[x(-e^{-x^2/2}) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2} dx \right) = 0 + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1, \end{aligned}$$

där vi partialintegrerat och utnyttjat att $\varphi(x)$ är en täthetsfunktion så den sista integralen blir ett.

Ett korollarium av satsen ovan (gör ett variabelbyte i integralerna) är följande.



$$X \sim \mathbf{N}(\mu, \sigma)$$

Om $X \sim \mathbf{N}(\mu, \sigma)$ så är $E(X) = \mu$ och $V(X) = \sigma^2$.



Standardisering av normalfördelning

Sats. $X \sim \mathbf{N}(\mu, \sigma) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \sim \mathbf{N}(0, 1)$.

Bevis. Antag att $Z \sim \mathbf{N}(0, 1)$. Eftersom

$$F_X(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

och $\Phi'(x) = \varphi(x)$ då φ är kontinuerlig, så följer det att

$$f_X(x) = F'_X(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbf{R},$$

vilket är precis hur vi definierat $N(\mu, \sigma)$ tidigare.

Omvänt, om $X \sim \mathbf{N}(\mu, \sigma)$ så är

$$F_Z(z) = P(Z \leq z) = P(\mu + \sigma Z < \mu + \sigma z) = F_X(\mu + \sigma z),$$

så

$$f_Z(z) = f_X(\mu + \sigma z)\sigma = \varphi(z),$$

dvs $Z \sim \mathbf{N}(0, 1)$. □



Exempel

Låt $X \sim \mathbf{N}(1, 2)$. Bestäm $P(X \leq 1)$, $P(X \leq -1)$, $P(0 < X \leq 1)$, samt $P(|X - 2| < 3)$.

$$(i) \quad P(X \leq 1) = P\left(\frac{X-1}{2} \leq \frac{1-1}{2}\right) = \Phi(0) = \frac{1}{2}.$$

$$(ii) \quad P(X \leq -1) = P\left(\frac{X-1}{2} \leq \frac{-1-1}{2}\right) = \Phi(-1) = 1 - \Phi(1) \approx 0.1587.$$

(iii)

$$\begin{aligned} P(0 < X \leq 1) &= P\left(\frac{0-1}{2} < \frac{X-1}{2} \leq \frac{1-1}{2}\right) = \Phi(0) - \Phi(-1/2) \\ &= 1/2 - (1 - \Phi(1/2)) = -1/2 + \Phi(1/2) \approx 0.1915. \end{aligned}$$

(iv)

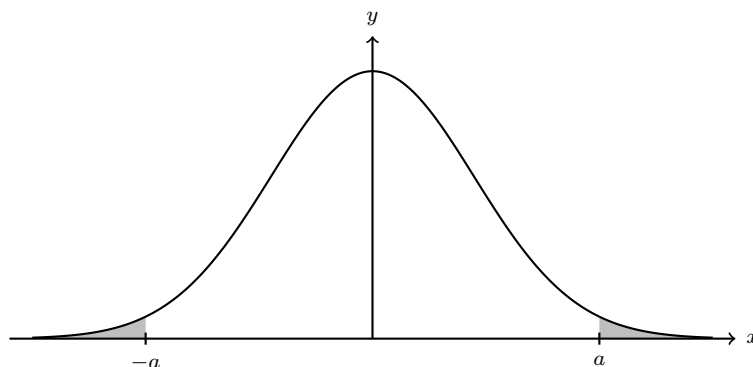
$$\begin{aligned} P(|X-2| < 3) &= P(-3 < X-2 < 3) = P\left(\frac{-1-1}{2} < \frac{X-1}{2} \leq \frac{5-1}{2}\right) \\ &= \Phi(2) - \Phi(-1) = \Phi(2) - 1 + \Phi(1) \approx 0.8186. \end{aligned}$$



Exempel

Låt $X \sim N(0, 1)$. Hitta ett tal a så att $P(|X| > a) = 0.05$.

Situationen ser ut som i bilden nedan. De skuggade områdena utgör tillsammans 5% av sannolikhetsmassan, och på grund av symmetri måste det vara 2.5% i varje "svans".



Om vi söker talet a , och vill använda funktionen $\Phi(x) = P(X \leq x)$, måste vi söka det tal som ger $\Phi(a) = 0.975$ (dvs de 2.5% i vänstra svansen tillsammans med de 95% som ligger i den stora kroppen). Detta gör vi genom att helt enkelt leta efter talet 0.975 i tabellen över $\Phi(x)$ värden. Där finner vi att $a = 1.96$ uppfyller kravet att $P(X \leq a) = 0.975$.

4.2 Linjärkombinationer av normalfördelade variabler

Det är inte på något sätt uppenbart att summan av två likafördelade variabler har samma fördelning. Oftast är det inte ens sant. Men, just normalfördelningen har precis denna trevliga egenskap!



Summa av normalfördelade variabler

Sats. Låt $X_1 \sim N(\mu_1, \sigma_1)$ och $X_2 \sim N(\mu_2, \sigma_2)$ vara oberoende och låt $a, b \in \mathbf{R}$. Då gäller att

$$aX_1 + bX_2 \sim N(a\mu_1 + b\mu_2, \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}).$$

Beviset är inte triviellt. Att $E(aX_1 + bX_2) = a\mu_1 + b\mu_2$ och $V(aX_1 + bX_2) = a^2\sigma_1^2 + b^2\sigma_2^2$ är enkelt att se. Detta gäller oavsett vad variablerna har för slags fördelning (enbart egenskaper för väntevärde och varians). Det faktum att summan blir normalfördelad kräver ett djupare argument; se boken (man använder faltningssatsen).

Satsen ovan generaliserar direkt till flera variabler. Vi har följande användbara specialfall.



Summor och medelvärde

Sats. Låt X_1, X_2, \dots, X_n vara oberoende och $X_k \sim N(\mu, \sigma)$ för $k = 1, 2, \dots, n$. Då gäller följande:

$$X := \sum_{k=1}^n X_k \sim N(n\mu, \sigma\sqrt{n}) \quad \text{och} \quad \bar{X} := \frac{1}{n} \sum_{k=1}^n X_k \sim N(\mu, \sigma/\sqrt{n}).$$

Den sista likheten är intressant, då det innebär att ju fler ”likadana” variabler vi tar med i ett medelvärde, desto *mindre* blir variansen. Till exempel får vi alltså säkrare resultat ju fler mätningar vi gör (något som känns intuitivt korrekt). Det är dock mycket viktigt att variablerna är *oberoende*. Annars gäller inte satsen! Vi bildar aldrig heller några skillnader mellan varianser, utan det som gör att variansen minskar med antalet termer är faktorn $1/n$ i medelvärdet:

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} V\left(\sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n V(X_k) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

eftersom variablerna är oberoende och $V(X_k) = \sigma^2$ för alla k .



Exempel

Låt $X \sim N(10, 3)$ och $Y \sim N(21, 6)$ vara oberoende. Vad är sannolikheten att Y är mer än dubbelt så stor som X ?

Lösning: Låt $W = Y - 2X \sim N(21 - 20, \sqrt{6^2 + 4 \cdot 3^2}) = N(1, \sqrt{72})$. Vi söker alltså

$$\begin{aligned} P(Y > 2X) &= P(Y - 2X > 0) = P(W > 0) = 1 - P(W \leq 0) = 1 - \Phi\left(\frac{0 - 1}{\sqrt{72}}\right) \\ &= 1 - (1 - \Phi(0.12)) = 0.5478. \end{aligned}$$

**Exempel**

Låt T vara livslängden för en viss sorts lysrör (enhet: månader) och $T \sim N(20, \sqrt{8})$.

- (i) Vad är sannolikheten att man klarar 43 månader om man har två (oberoende) lysrör och byter direkt det första går sönder?
- (ii) Hur många lysrör måste man skaffa för att medellivslängden ska vara mer än 19 månader med sannolikhet 95%?

Lösning:

- (i) Vi har två lysrör, T_1 och T_2 . Vi söker sannolikheten att $T_1 + T_2 \geq 43$. Satsen ovan visar att $T_1 + T_2 \sim N(40, 4)$, så

$$\begin{aligned} P(T_1 + T_2 \geq 43) &= 1 - P(T_1 + T_2 < 43) = 1 - P\left(\frac{T_1 + T_2 - 40}{\sqrt{16}} < \frac{43 - 40}{\sqrt{16}}\right) \\ &= 1 - \Phi(3/\sqrt{16}) \approx 1 - \Phi(0.75) \approx 0.2266. \end{aligned}$$

- (ii) Låt \bar{T} vara medelvärdet av n stycken lysrör. Det följer att $\bar{T} \sim N(20, \sqrt{8}/\sqrt{n})$. Vi vill att

$$\begin{aligned} 0.95 &= P(\bar{T} > 19) = P\left(\frac{\bar{T} - 20}{\sqrt{8/n}} > \frac{19 - 20}{\sqrt{8/n}}\right) = 1 - P(Z \leq -1/\sqrt{8/n}) \\ &= 1 - \Phi(-1/\sqrt{8/n}) = 1 - \left(1 - \Phi(1/\sqrt{8/n})\right) = \Phi(\sqrt{n/8}), \end{aligned}$$

där $Z = \frac{\bar{T} - 20}{\sqrt{8/n}} \sim N(0, 1)$. Ur tabell finner vi då att $\sqrt{n/8} = 1.645$, eller ekvivalent, att $n \approx 21.6$. Således behövs åtminstone 22 stycken lysrör.

4.3 De stora talens lag

Vi kommer nu betrakta en fundamental situation i sannolikhetslära. Vi låter X_1, X_2, \dots vara en oändlig följd av oberoende och likafördelade stokastiska variabler. Vi låter $E(X_i) = \mu$ och $V(X_i) = \sigma^2$ (så vi antar att variansen är ändlig just nu). I vanlig ordning definierar vi det aritmetiska medelvärdet \bar{X}_n av de n första variablerna i följderna som

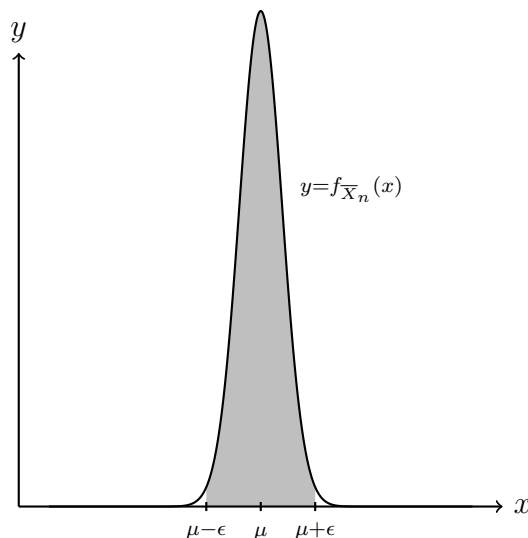
$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

**De stora talens lag (svag formulering)**

Sats. För varje $\epsilon > 0$ gäller att

$$P(|\bar{X}_n - \mu| < \epsilon) \rightarrow 1 \text{ då } n \rightarrow \infty.$$

En tolkning av satsen är att det aritmetiska medelvärdet av en följd oberoende och likafördelade variabler kommer att ha sin sannolikhetsmassa koncentrerad kring väntevärdet μ :



Variansen för \bar{X}_n är som bekant

$$V(\bar{X}_n) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

eftersom variablerna är oberoende (och vi antagit ändlig varians). Så då $n \rightarrow \infty$ ser vi att variansen för medelvärdet går mot noll. Konvergensen i de stora talens lag förefaller alltså rimlig. Ett mer ordentligt bevis följer från kända olikheter, så låt oss formulera dessa.



Markovs olikhet

Sats. Om X är en icke-negativ stokastisk variabel med ändligt väntevärde så gäller att

$$P(X \geq a) \leq \frac{E(X)}{a}, \quad a > 0.$$

Bevis. För det kontinuerliga fallet med täthetsfunktion, eftersom $a > 0$ och $f_X(x) \geq 0$,

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \geq \int_a^{\infty} x f_X(x) dx \geq a \int_a^{\infty} f_X(x) dx = aP(X \geq a),$$

så följer att $P(X \geq a) \leq \frac{E(X)}{a}$. Det diskreta fallet hanteras analogt (gör det!) □



Tjebysjovs (Chebychevs) olikhet

Sats. Låt X vara en stokastisk variabel med ändligt väntevärde $E(X) = \mu$ och ändlig varians $V(X) = \sigma^2$, och låt $k > 0$. Då gäller att

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Bevis. Eftersom $(X - \mu)^2$ är en icke-negativ stokastisk variabel och $E(X - \mu) = 0$, så gäller enligt Markovs olikhet att

$$\begin{aligned} P(|X - \mu| \geq k\sigma) &= P((X - \mu)^2 \geq k^2\sigma^2) \leq \frac{E((X - \mu)^2)}{k^2\sigma^2} \\ &= \frac{V(X - \mu)}{k^2\sigma^2} = \frac{1}{k^2}, \end{aligned}$$

där den näst sista likheten är Steiners sats. □

En följd av denna olikhet är att vi får en direkt uppskattning av hur mycket sannolikhetsmassa som finns i intervall av typen $(\mu - k\sigma, \mu + k\sigma)$. Vi kan till exempel se att det finns minst 50% av sannolikhetsmassan om $k = \sqrt{2}$, minst 75% om $k = 2$ och minst 96% om $k = 5$.



Vanligt missförstånd

Låt oss (oberoende) kasta en sex-sidig balanserad tärning 1800 gånger. Vi förväntar oss att medelvärdet ligger nära 3.5. Antag att medelvärdet blev 4.0. Betyder detta enligt satsen ovan att vi kommer att få fler resultat 1, 2, 3 än 4, 5, 6 om vi kastar tärningen 1800 gånger till? Svaret är nej. De olika kasten anses oberoende, och kan därför inte påverkas av tidigare utfall. Så hur kan då satsen ovan gälla? Faktum är att det inte behöver vara fler låga resultat vid kommande upprepningar, det räcker med att medelvärdet av de nya resultaten är mindre än 4.0 för att vi ska hamna närmare 3.5 *totalt* sett.

Det lönar sig alltså inte att satsa mer pengar bara för att man förlorat så många gånger på rad (om händelserna är oberoende, annars kan lite vad som helst inträffa!).

Bevis av de stora talens lag. I princip följer detta direkt av olikheterna ovan. Låt $\epsilon > 0$. Vi ser direkt att

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0,$$

då $n \rightarrow \infty$. □

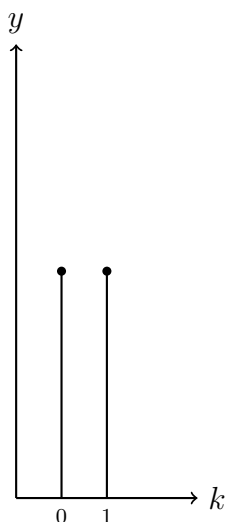
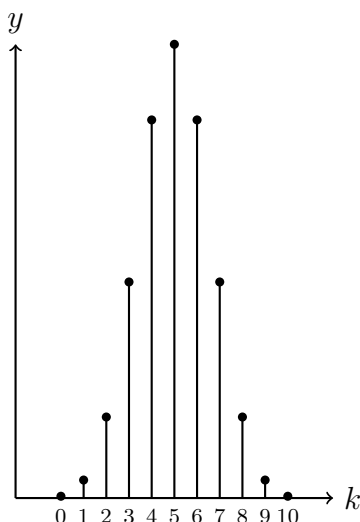
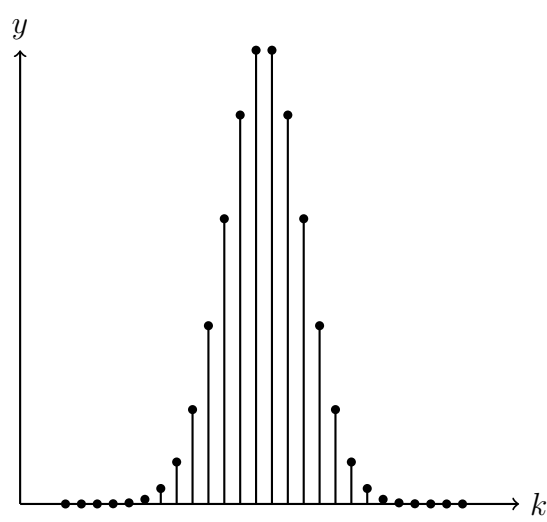
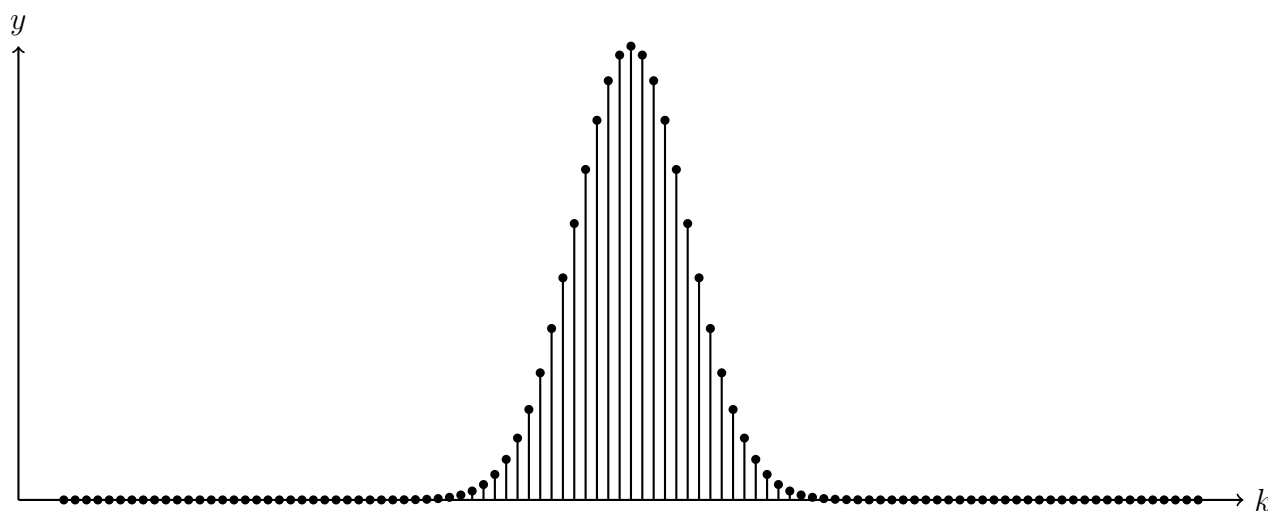
4.4 Centrala gränsvärdessatsen

Alla vägar leder till Rom. Eller åtminstone: alla fördelningar leder till normalfördelning? Faktum är att det är precis det den centrala gränsvärdessatsen säger: summan av ett stort antal oberoende och likafördelade stokastiska variabler är approximativt normalfördelad.

Vi betraktar ett exempel. Låt oss utföra det klassiska experimentet med slantsingling och räkna antalet X kronor vid ett visst antal, säg n , kast. Från tidigare exempel (inbrottstjuven) så vet vi att $X \sim \text{Bin}(n, p)$, där $p = 1/2$ om myntet är rättvist. En binomialfördelad variabel kan ses som en summa av oberoende Bernoulli-fördelade variabler X_k , en variabel för varje försök (slantsingling), där $X_k = 0$ om försök nr k "misslyckas" (klave), och $X_k = 1$ om försök k lyckas

(krona). Alltså kan vi skriva $X = \sum_{k=1}^n X_k$. Varje X_k har sannolikhetsfunktionen $p_{X_k}(1) = p$

och $p_{X_k}(0) = 1 - p$. Med andra ord, binomialfördelningen kan ses som en summa av oberoende och likafördelade variabler. Om bara n är tillräckligt stort borde vi i så fall närma oss normalfördelningen. Hur stort? Vi skisserar några fall när n blir större och större och $p = 0.5$.

Med $n = 1$.Med $n = 10$.Med $n = 25$.Med $n = 100$.

Här ser vi ganska tydligt att ju större n blir, desto mer lik blir sannolikhetsfördelning en normalfördelningskurva. Följande sats verkar alltså rimlig (åtminstone i Binomialfallet).



Centrala gränsvärdessatsen (CGS)

Sats. Låt X_1, X_2, \dots vara en oändlig följd av likafördelade och oberoende stokastiska variabler.

Vidare, låt $E(X_k) = \mu$ och $V(X_k) = \sigma^2$ för $k = 1, 2, \dots$. Då gäller att $Y_n = \sum_{k=1}^n X_k$ konvergerar i fördelning enligt

$$P\left(a < \frac{Y_n - n\mu}{\sigma\sqrt{n}} < b\right) \rightarrow \Phi(b) - \Phi(a) \text{ då } n \rightarrow \infty$$

för alla $a, b \in \mathbf{R}$ med $a < b$. Vi säger att X är **asymptotiskt** normalfördelad.

Den kanske vanligaste situationen (åtminstone i denna kurs) kommer att vara att den summa vi är intresserade av är ett medelvärde, så vi formulerar detta specialfall separat.



Sats. Med samma förutsättningar som ovan så uppfyller medelvärdet $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ att

$$P\left(a < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < b\right) \rightarrow \Phi(b) - \Phi(a) \text{ då } n \rightarrow \infty$$

för alla $a, b \in \mathbf{R}$ med $a < b$.

Beviset för satsen faller utanför ramen för denna kurs. Se, till exempel, Rick Durrett: *Probability: Theory and Examples* eller Allan Gut: *An Intermediate Course in Probability*.
Så hur använder vi CGS?



Approximation via CGS

Med samma beteckningar och förutsättningar som ovan så är

$$P(X \leq x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right) \quad \text{och} \quad P(\bar{X} \leq x) \approx \Phi\left(\frac{x - \mu}{\sigma/\sqrt{n}}\right), \quad x \in \mathbf{R},$$

om n är stort. Oftast brukar $n \geq 30$ duga, men skeva fördelningar kräver större n . Vi skriver $X \stackrel{\text{appr.}}{\sim} N(n\mu, \sqrt{n}\sigma)$ och $\bar{X} \stackrel{\text{appr.}}{\sim} N(\mu, \sigma/\sqrt{n})$; variablerna är **approximativt** normalfördelade.



Exempel

Vad är sannolikheten att summan av 50 stycken slumpstal mellan 0 och 2 överstiger 53?

Lösning: Vi antar att slumptalen är likformigt fördelade, så varje slumpstal $X_k \sim \text{Re}(0, 2)$, och att slumptalen är oberoende av varandra. Det råder likformig fördelning, så $E(X_k) = 1$ och $V(X_k) = 1/3$. Varför? Enkelt att se från definitionen:

$$E(X_k) = \int_0^2 x \frac{1}{2} dx = \left[\frac{x^2}{4} \right]_0^2 = 1$$

och

$$V(X_k) = \int_0^2 x^2 \frac{1}{2} dx - 1^2 = \left[\frac{x^3}{6} \right]_0^2 - 1 = 1/3.$$

Så vi har en summa av 50 stycken likformigt fördelade variabler X_k med samma väntevärde och varians. Låt $X = \sum_{k=1}^{50} X_k$. CGS implicerar att $X \stackrel{\text{appr.}}{\sim} N(50, \sqrt{50/3})$. Alltså erhåller vi

$$P(X > 53) = 1 - P(X \leq 53) \approx 1 - \Phi(3/\sqrt{50/3}) = 1 - \Phi(0.7348) \approx 0.2312.$$

Det är alltså ca 23% chans att summan överstiger 53.



Exempel

Antag att samtalstiderna till 1177 är oberoende och exponentialfördelade med väntevärde 15 minuter. Om en sjuksköterska förväntas svara på 28 samtal under ett åtta-timmars pass, vad är sannolikheten att hon lyckas?

Lösning: Låt $X_k \sim \text{Exp}(1/15)$ vara tiden för samtal k , $k = 1, 2, \dots, 28$. Den totala tiden för 28 samtal ges av $X = \sum_{k=1}^{28} X_k$. Faktum är att man kan visa att X blir gamma-fördelad (se Blom et al.), men den fördelningen är ganska bölig att arbeta med. Vad säger CGS? Vi har kring 30 stycken samtal, så $X \stackrel{\text{appr.}}{\sim} N(28 \cdot 15, \sqrt{28 \cdot 15}) = N(420, \sqrt{6300})$. Alltså är

$$P(X \leq 8 \cdot 60) \approx \Phi\left(\frac{480 - 420}{\sqrt{6300}}\right) \approx \Phi(0.76) = 0.7764.$$

Nästan 80% chans alltså! Hur bra stämmer då detta? Man kan härleda att X i själva verket har fördelningen $X \sim \Gamma(28, 1/15)$, så $P(X \leq 480) = 0.7838$ (MATLAB, `gamcdf`).

4.5 Approximation av binomialfördelning

Vi har redan stött på denna fördelning flera gånger. Situationen är att ett slumpförsök har två möjliga utfall, ett med sannolikhet p och det andra med $1 - p$. Vi upprepar försöket oberoende n gånger, och räknar antalet X gånger det första utfallet inträffar. Vi kallar X för binomialfördelad med parametrarna n och p , och skriver $X \sim \text{Bin}(n, p)$.



Binomialfördelning

Sats. Om $X \sim \text{Bin}(n, p)$, så är $E(X) = np$ och $V(X) = np(1 - p)$. Vidare gäller att vi kan approximera $X \stackrel{\text{appr.}}{\sim} N(np, \sqrt{np(1 - p)})$ om $np(1 - p) \geq 10$.

Beviskiss. Vi skriver X som en summa av n oberoende Bernoulli-variabler $X_k \sim \text{Be}(p)$, så väntevärdet $E(X) = \sum_{k=1}^n E(X_k) = np$, eftersom $E(X_k) = 0 \cdot (1 - p) + 1 \cdot p = p$. På samma sätt, $V(X) = np(1 - p)$ eftersom $V(X_k) = 0^2 \cdot (1 - p) + 1^2 \cdot p - p^2 = p(1 - p)$.

När det gäller approximationen till normalfördelning så följer detta av CGS. Att just kravet $np(1 - p) \geq 10$ ger en bra approximation kräver en lite djupare analys av på vilket sätt sannolikheterna konvergerar.



Exempel

Vi sår 1000 stycken frön som har en grobarhet på 80% (sannolikheten att ett frö gror). Vad är sannolikheten att högst 180 stycken inte gror?

Lösning. Låt X vara antalet frön som inte gror. Vi antar att olika frön är oberoende av varandra. Då är $X \sim \text{Bin}(1000, 0.2)$. Eftersom

$$np(1 - p) = 1000 \cdot 0.2 \cdot 0.8 = 160 \gg 10,$$

så är $X \stackrel{\text{appr.}}{\sim} N(200, \sqrt{160})$. Vi beräknar

$$P(X \leq 180) \approx \Phi((180 - 200)/\sqrt{160}) = \Phi(-1.5811) = 1 - \Phi(1.5811) = 0.057.$$

Det är alltså ca 6% sannolikhet. Verklig sannolikhet (MATLAB `binocdf(180, 1000, 0.2)`) är 6.02%.

4.5.1 Poissonapproximation

I vissa lägen så fungerar det dåligt med normalapproximationen (ofta då sannolikheten är nära 0 eller 1). Ibland kan då följande approximation fungera (se avsnittet i slutet av föreläsningen för argument kring varför detta är sant).



Approximation: Binomial till Poisson

Sats. Om $X \sim \text{Bin}(n, p)$ med $n \geq 10$ och $p \leq 0.1$, så är $X \stackrel{\text{appr.}}{\sim} \text{Po}(np)$.



Exempel

Sågaren Sverker sågar ut brädor som har normalfördelad längd $L \sim N(200, \sqrt{50})$ ($\sigma^2 = 50$), enhet: cm. Om Sverker en vacker dag sågar upp 300 brädor (oberoende av varandra), vad är sannolikheten att färre än 5 stycken är kortare än 185 cm?

Lösning: Vi räknar först ut sannolikheten p att en bräda är kortare än 185 cm:

$$p = P(L < 185) = P\left(\frac{L - 200}{\sqrt{50}} < \frac{-15}{\sqrt{50}}\right) = \Phi(-2.12) = 1 - \Phi(2.12) = 0.0170.$$

Låt X vara antalet brädor av 300 som är kortare än 185 cm. Det följer att $X \sim \text{Bin}(300, p)$. Sannolikheten p är alltså liten, och $300p(1-p) = 5.01$ är betydligt mindre än 10, så normalapproximation fungerar antagligen inte. Men Poissonapproximation borde fungera bra då $p \ll 0.1$ och $n = 300 \gg 10$. Alltså är $X \stackrel{\text{appr.}}{\sim} \text{Po}(300 \cdot 0.0170) = \text{Po}(5.10)$. Ur tabell (interpolation mellan $\text{Po}(5.0)$ och $\text{Po}(5.2)$):

$$P(X \leq 4) \approx \frac{0.4405 + 0.4061}{2} = 0.4233.$$

Alltså ungefär 42% chans. Verklig sannolikhet blir 42.15%. Normalapproximation skulle i fallet ge 31%, vilket är alldeles för lågt.

4.6 Approximation av Poissonfördelning



Approximation av Poissonfördelning

Sats. Låt $X \sim \text{Po}(\mu)$ med $\mu \geq 15$. Då är $X \stackrel{\text{appr.}}{\sim} N(\mu, \sqrt{\mu})$ ($V(X) = \mu$).



Exempel

Låt X vara antal paket i en datakö under en sekund. Mätningar har visat att en vettig modell är $X \sim \text{Po}(250)$. Beräkna approximativt $P(X < 240)$.

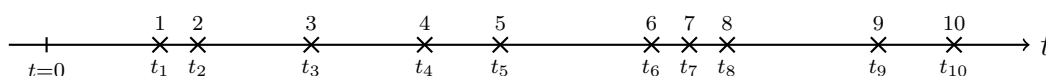
Lösning: Eftersom väntevärdet $\mu = 250 \gg 15$ så kan vi normalapproximera. Då blir

$$P(X < 240) = P(X \leq 239) \approx \Phi\left(\frac{239 - 250}{\sqrt{250}}\right) = \Phi(-0.6957) = 1 - \Phi(0.6957) = 0.2433.$$

Exakt värde: 0.2552.

4.7 (★)Poissonfördelning

Antag att vi har en situation där händelser inträffar oberoende av varandra med en konstant intensitet λ , det vill säga, på t tidsenheter inträffar i genomsnitt λt händelser. Denna typ av situation brukar ofta modularas med hjälp av *Poisson*-fördelningen. Om $X(t)$ är antalet händelser i tidsintervallet $[0, t]$, så säger vi att $X(t)$ är Poissonfördelad med väntevärde $\mu = \lambda t$.



Händelser (markerade med kryss och numrerade) i tidsintervallet $[0, t]$. Tiderna t_k är tidpunkten för händelsen k .



Poissonfördelning

Sats. Vi kallar X för Poissonfördelad med parametern μ , $X \sim \text{Po}(\mu)$, om sannolikhetsfunktionen ges av

$$p_X(k) = P(X = k) = \frac{\mu^k e^{-\mu}}{k!}, \quad k = 0, 1, 2, \dots$$

Variabeln X har $E(X) = V(X) = \mu$ (samma väntevärde som varians, parametern μ).

Hur hänger situationen ovan ihop med definitionen av p_X ? Vi fixerar tiden t och delar in intervallet $[0, t]$ i n lika stora delar, där vi väljer n så stort att det högst finns en händelse i varje delintervall. Vi introducerar en sannolikhet p som är sannolikheten att ett visst delintervall innehåller en händelse. Det är samma p för alla delintervall och sambandet $np = \lambda t$ måste gälla. Eftersom händelserna är oberoende måste $X(t) \sim \text{Bin}(n, p)$. Egenskaper för binomialfördelningen medför att $E(X(t)) = np = \lambda t$.

Vi börjar med att betrakta fallet med noll händelser i intervallet $[0, t]$, det vill säga, händelsen att $X(t) = 0$:

$$P(X(t) = 0) = \binom{n}{0} p^0 (1-p)^{n-0} = \left(1 - \frac{\lambda t}{n}\right)^n = \exp\left(n \ln\left(1 - \frac{\lambda t}{n}\right)\right) \rightarrow e^{-\lambda t},$$

då $n \rightarrow \infty$. Här har vi utnyttjat standardgränsvärdet $s^{-1} \ln(1+s) \rightarrow 1$ då $s \rightarrow 0$.

I det allmänna fallet kan vi visa att

$$P(X(t) = k) \rightarrow \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad n \rightarrow \infty.$$

För att se detta, låt k vara fix och betrakta

$$\begin{aligned} P(X(t) = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{(\lambda t)^k}{k!} \cdot \frac{\left(1 - \frac{\lambda t}{n}\right)^n}{\left(1 - \frac{\lambda t}{n}\right)^k} \rightarrow 1 \cdot \frac{(\lambda t)^k}{k!} \cdot \frac{e^{-\lambda t}}{1}, \quad n \rightarrow \infty. \end{aligned}$$

Vad detta innebär är att om $X_n \sim \text{Bin}(n, \lambda/n)$, så kommer $X_n \xrightarrow{D} X \sim \text{Po}(\lambda)$. Detta ger oss även en användbar approximationssats för binomialfördelningen.



Approximation: Binomial till Poisson

Sats. Om $X \sim \text{Bin}(n, p)$ med $n \geq 10$ och $p \leq 0.1$, så är $X \overset{\text{appr.}}{\sim} \text{Po}(np)$.

Att p_X verkligen är en sannolikhetsfunktion följer från Maclaurinutveckling av e^x :

$$\sum_{k=0}^{\infty} p_X(k) = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{-\mu} \cdot e^{\mu} = 1.$$

Låt oss även härleda väntevärde och varians. För väntevärdet:

$$\sum_{k=0}^{\infty} k p_X(k) = e^{-\mu} \left(0 + \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} \right) = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = \mu.$$

Variansen är lite bökgigare. Vi kan inte direkt räkna ut $E(X^2)$, utan tar till omskrivningen

$$E(X^2) = E(X(X-1)) + E(X).$$

Alltså,

$$\begin{aligned} E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1)p_X(k) = e^{-\mu} \left(\sum_{k=2}^{\infty} \frac{\mu^k}{(k-2)!} \right) = \mu^2 e^{-\mu} \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} \\ &= \mu^2 e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = \mu^2, \end{aligned}$$

så $E(X^2) = \mu^2 + \mu$, vilket medför att $V(X) = E(X^2) - E(X)^2 = \mu$.



Addition av oberoende Poissonfördelade variabler

Sats. Låt $X \sim \text{Po}(\mu_1)$ och $Y \sim \text{Po}(\mu_2)$ vara oberoende. Då är $X + Y \sim \text{Po}(\mu_1 + \mu_2)$.

Satsen förefaller intuitivt att vara rimlig. Vi lägger helt enkelt ihop händelserna från två liknande processer, det förväntade antalet blir nu $\mu_1 + \mu_2$, och på grund av beteendet hos var och en tippar vi att summan fungerar på samma sätt. Formellt kan vi visa satsen medelst den så kallade faltningssatsen. Den simultana sannolikhetsfunktionen för (X, Y) ges av produkten $p_X(i)p_Y(j)$, och vi söker sannolikhetsfunktionen för $Z = X + Y$. Alltså,

$$p_Z(k) = P(X + Y = k) = \sum_{i+j=k} p_X(i)p_Y(j) = \sum_{i=0}^k p_X(i)p_Y(k-i).$$

Dubbelsumman blir en enkelsumma eftersom vi bara summerar över ”diagonalen” (när k är fixt och $i + j = k$). Sen är $p_X(i) = 0$ då $i < 0$ och $p_Y(k-i) = 0$ då $i > k$ så det räcker att summera från $i = 0$ till $i = k$. Vidare,

$$\begin{aligned} p_Z(k) &= \sum_{i=0}^k e^{-\mu_1} \frac{\mu_1^i}{i!} e^{-\mu_2} \frac{\mu_2^{k-i}}{(k-i)!} = \frac{e^{-(\mu_1+\mu_2)}}{k!} \sum_{i=0}^k \frac{k!}{(k-i)!i!} \mu_1^i \mu_2^{k-i} \\ &= \frac{e^{-(\mu_1+\mu_2)}}{k!} \sum_{i=0}^k \binom{k}{i} \mu_1^i \mu_2^{k-i} = \frac{e^{-(\mu_1+\mu_2)}}{k!} (\mu_1 + \mu_2)^k, \end{aligned}$$

där vi utnyttjat binomialsatsen i sista steget. Detta uttryck är inget annat än sannolikhetsfunktionen för en $\text{Po}(\mu_1 + \mu_2)$ -fördelad variabel, vilket var precis det vi ville visa!

Denna sats kan vi använda för att dela upp en $\text{Po}(\mu)$ fördelad variabel i $\lfloor \mu \rfloor$ stycken oberoende variabler med väntevärde ett, och en liten svans (med längd $\mu - \lfloor \mu \rfloor$). På detta sätt kan man visa följande sats.



Approximation av Poissonfördelning

Sats. Låt $X \sim \text{Po}(\mu)$ med $\mu \geq 15$. Då är $X \stackrel{\text{appr.}}{\sim} N(\mu, \sqrt{\mu})$ ($V(X) = \mu$).

Kapitel 5

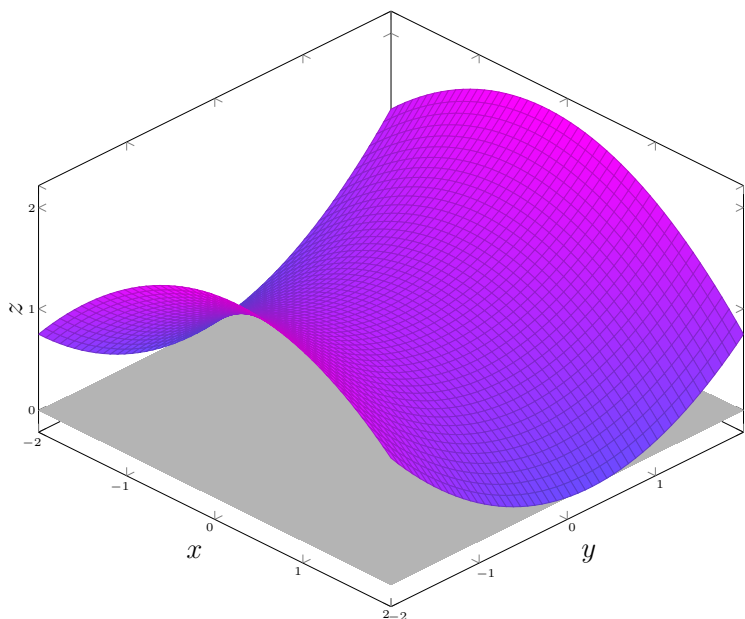
Flerdimensionella s.v.

5.1 Flervariabelanalys

Precis som i linjär algebra så betecknar vi vektorer enligt $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbf{R}^n$. Har vi bara två eller tre variabler använder vi ofta uttrycken (x, y) respektive (x, y, z) istället av tradition.

5.1.1 Funktioner av flera variabler

Så vi behöver en gnutta flervariabelanalys för att kunna hantera det som komma skall. För att komma åt de bitar vi behöver så måste vi generalisera integralbegreppet till högre dimensioner. Som exempel betraktar vi situationen med en funktion som beror på två variabler (säg x och y) och som ger ett reellt tal som "utdata." Mer kompakt uttryckt så skriver vi att $f : \mathbf{R}^2 \rightarrow \mathbf{R}$. Typiskt för oss är att vi tänker oss att $z = f(x, y)$, där vi tolkar funktionsvärdet som höjden ovanför (eller under om vi har negativt tecken) xy -planet. Skulle man försöka rita en figur skulle vi kunna få något enligt nedan.



Det gråa området där $z = 0$ (under själva grafen) är det område där (x, y) varierar över. Generellt sett så behöver inte detta vara en rektangel utan kan vara i princip vilket hyfsat snällt två-dimensionellt objekt som helst. I varje punkt (x, y) i detta område har vi alltså en "höjd" $z = f(x, y)$. Här kan vi givetvis analysera hur denna funktion beter sig och introducera

begrepp som kontinuitet, derivata etc i flera variabler, men det blir lite för omfattande. Vad vi däremot kommer behöva är ett volymsbegrepp. Tanken är att vi skulle vilja räkna ut volymen mellan xy -planet och ytan $z = f(x, y)$.

5.1.2 Integration av funktioner av flera variabler

Om vi jämför med det endimensionella fallet, där $y = f(x)$, så approximerade vi arean av ett område $0 \leq y \leq f(x)$ med rektanglar och summerade arean av dessa för att konstruera Riemannintegralen av $f(x)$ (bläddra tillbaka till kapitel 6 i envariabelboken). Man kan göra motsvarande manöver i högre dimensioner, men istället för rektanglar får vi nu summera volymen på rätblock (staplar) som får mindre och mindre tvärsnittsarea. För en funktion av två variabler, där $z = f(x, y)$, så ter sig detta ganska naturligt rent geometriskt. Vi staplar in legobitar under ytan så högt som möjligt utan att sticka ut för mycket från ytan $z = f(x, y)$. Vi betecknar denna volym med

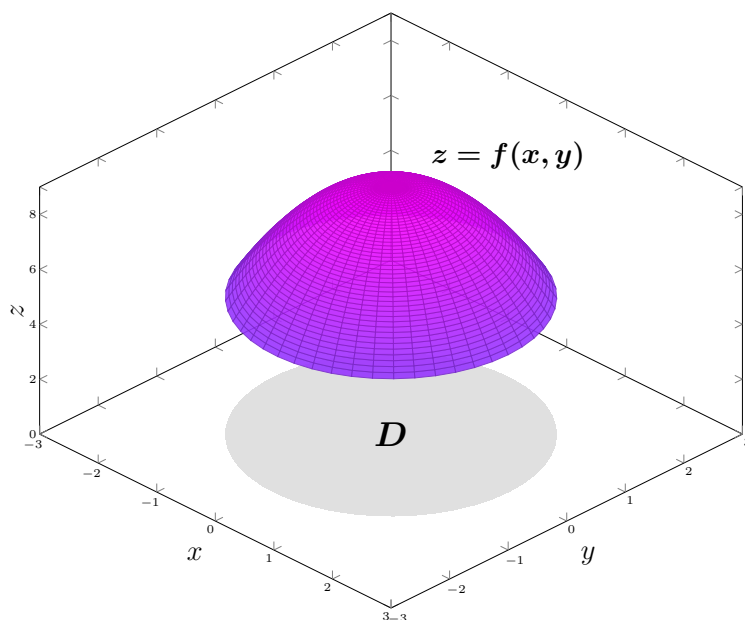
$$V = \iint_D f(x, y) \, dx \, dy$$

där $D \subset \mathbf{R}^2$ är det område som vi låter (x, y) variera över.

Detta generaliserar sedan naturligt till \mathbf{R}^n (med skillnaden att det nu blir aningen svårare att visualisera). Konstruktionen kräver egentligen en hel del bevis (samt egentligen ett annat integralbegrepp då Riemannintegralen är lite problematisk), men vi skippar det och fokuserar istället på hur vi rent analytiskt hanterar integralerna om vi vill räkna ut "volymen."

5.1.3 Iteration i 2D

Låt oss fokusera på fallet i två dimensioner.

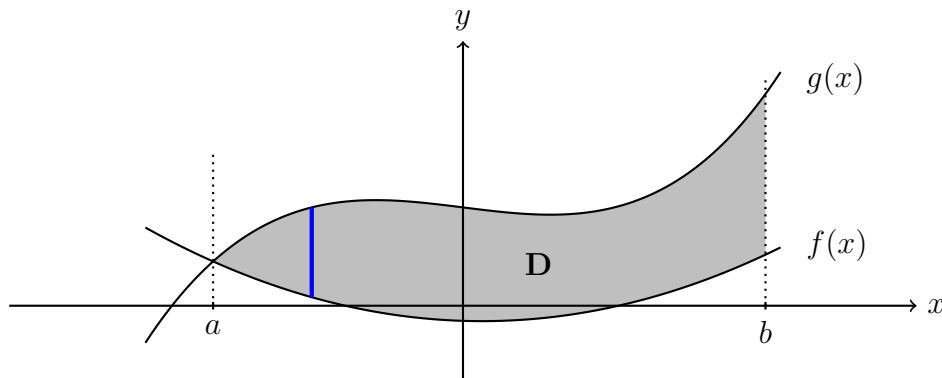


Den funktion som integreras ligger i en dimension ovanför (tänkt att du har en z -axel som pekar rakt upp ur sidan och att höjden ovanför sidan är funktionsvärdet $z = f(x, y)$ i punkten (x, y)). Typiskt skulle det kunna se ut enligt ovan. Volymen mellan ytan $z = f(x, y)$ och xy -planet när $(x, y) \in D$ betecknar vi nu med

$$V = \iint_D f(x, y) \, dx \, dy.$$

Hur räknar vi ut detta? Generellt så kan det vara svårt att hitta ett analytiskt uttryck, men om D uppfyller vissa krav kan vi *iterera* integralerna och räkna ut en i taget. Betrakta följande specialfall där vi kan stänga in y -koordinaten mellan funktioner av x :

$$D = \{(x, y) \in \mathbf{R}^2 : a \leq x \leq b, f(x) \leq y \leq g(x)\}.$$



Vi tänkar oss nu att vi styckar upp D i "stolpar" som är parallella med y -axeln, så att om $a \leq x \leq b$ så täcker dessa stolpar precis området D (se den blåa linjen i figuren ovan). Det går i detta fall att visa att

$$\iint_D f(x, y) dx dy = \int_a^b \left(\int_{f(x)}^{g(x)} f(x, y) dy \right) dx$$

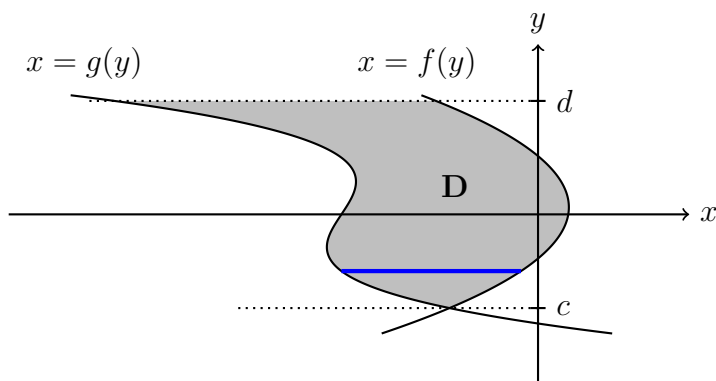
Motsvarande kan göras om området har formen

$$D = \{(x, y) \in \mathbf{R}^2 : c \leq y \leq d, f(y) \leq x \leq g(y)\}$$

så

$$\iint_D f(x, y) dx dy = \int_c^d \left(\int_{f(y)}^{g(y)} f(x, y) dx \right) dy.$$

Stolparna är nu parallella med x -axeln istället.



Notera speciellt att det är konstanter i gränserna på de yttre integralerna i båda fallen. Vi får *inte* ha några x eller y löst hängande när vi är färdiga. Skulle det inte gå att få till ett av fallen kan man behöva stycka upp området. Det fungerar utan problem: om $D = D_1 \cup D_2$ så gäller att

$$\iint_D f(x, y) dx dy = \iint_{D_1} f(x, y) dx dy + \iint_{D_2} f(x, y) dx dy,$$

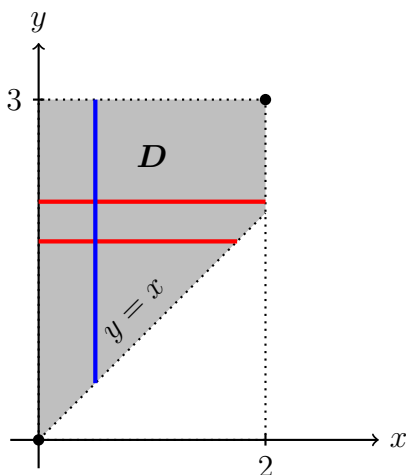
så vi kan hantera en del av området i taget.



Exempel

Beräkna $\iint_D (xy + 2) dx dy$ om D är den del av rektangeln med hörn i $(0,0)$ och $(2,3)$ där $y > x$.

Lösning. Låt oss börja med att rita en figur så vi ser vad D är.



Det skuggade området är det område D vi vill integrera över. Vi har två alternativ. Endera så itererar vi med de röda linjerna eller längs de blå. Vi ställer upp båda fallen som övning (de kommer alltid att ge samma svar).

Alternativ 1. Om vi väljer de röda linjerna ser vi att längden på dessa betar sig olika beroende på om $0 \leq y \leq 2$ eller om $2 < y \leq 3$. Om $0 \leq y \leq 2$ så sträcker sig x från 0 till linjen $y = x$, så $0 \leq x \leq y$. Om $2 < y \leq 3$ sträcker sig x från 0 till 2. Därför delar vi upp i två delar:

$$\begin{aligned} \iint_D (xy + 2) dx dy &= \int_0^2 \left(\int_0^y (xy + 2) dx \right) dy + \int_2^3 \left(\int_0^2 (xy + 2) dx \right) dy \\ &= \int_0^2 \left[\frac{x^2}{2} y + 2x \right]_{x=0}^y dy + \int_2^3 \left[\frac{x^2}{2} y + 2x \right]_{x=0}^2 dy \\ &= \int_0^2 \left(\frac{y^3}{2} + 2y - 0 \right) dy + \int_2^3 \left(\frac{2^2}{2} y + 4 - 0 \right) dy \\ &= \left[\frac{y^4}{8} + y^2 \right]_0^2 + [y^2 + 4y]_2^3 = \frac{2^4}{8} + 2^2 + 3^2 + 12 - (2^2 + 8) = 15. \end{aligned}$$

Alternativ 2. Om vi väljer de blåa linjerna ser vi att för $0 \leq x \leq 2$ så sträcker sig y från linjen $y = x$ till $y = 3$. Vi behöver alltså inte dela upp integralen:

$$\begin{aligned} \iint_D (xy + 2) dx dy &= \int_0^2 \left(\int_x^3 (xy + 2) dy \right) dx = \int_0^2 \left[x \frac{y^2}{2} + 2y \right]_{y=x}^3 dx \\ &= \int_0^2 \left(\frac{9x}{2} + 6 - \left(\frac{x^3}{2} + 2x \right) \right) dx = \int_0^2 \left(\frac{5x - x^3}{2} + 6 \right) dx \\ &= \left[\frac{5x^2}{4} - \frac{x^4}{8} + 6x \right]_0^2 = \frac{20}{4} - 2 + 12 = 15. \end{aligned}$$

Vilket alternativ var enklast?

5.2 Flerdimensionella stokastiska variabler

Vi fokuserar på två-dimensionella variabler. Det är steget från en dimension till två som är det svåraste. Generaliseringar till högre dimensioner följer utan problem i de flesta fall.



Stokastisk variabel

Definition. En tvådimensionell stokastisk variabel är en reell-vektorvärd funktion (X, Y) (två komponenter) definierad på ett utfallsrum Ω . Alltså avbildar (X, Y) olika utfall på reella vektorer; $(X, Y): \Omega \rightarrow \mathbf{R}^2$. Om (X, Y) bara antar ändligt eller uppräknligt många värden så kallar vi (X, Y) för en diskret stokastisk variabel. Om varken X eller Y är diskret kallar vi (X, Y) för kontinuerlig.

Definitionen är analog med envariabelfallet. Observera dock följande: en situation som kan uppstå är att vi får ”halvdiskreta” variabler med ena variabeln diskret och den andra kontinuerlig! Inträffar inte ofta i denna kurs, men det kan vara värt att beakta.



Exempel

- (i) Låt (X, Y) vara vikten X och längden Y hos en person. Variabeln är kontinuerlig och $\Omega = (X(\Omega), Y(\Omega)) = [0, \infty)^2 = [0, \infty) \times [0, \infty)$.
- (ii) Låt (X, Y) vara resultaten av ett tärningskast (X) respektive ett myntkast (Y), där vi representerar krona med 1 och klave med -1 . Variabeln är diskret och:

$$\Omega = \{ \square, \square\square, \square\square\square, \square\square\square\square, \square\square\square\square\square, \square\square\square\square\square\square \} \times \{ \text{Krona, Klave} \},$$

$$(X(\Omega), Y(\Omega)) = \{ 1, 2, 3, 4, 5, 6 \} \times \{ -1, 1 \}.$$

I det 2-dimensionella fallet är vi nu intresserade av delmängder i planet. För att kunna prata om en fördelningsfunktion introducerar vi mängden

$$C(x, y) = \{(u, v) \in \mathbf{R}^2 : u \leq x \text{ och } v \leq y\}.$$

Vi ser att $P((X, Y) \in C(x, y)) = P(X \leq x, Y \leq y)$ och gör följande definition.



Fördelningsfunktion

Definition. Funktionen $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ kallas fördelningsfunktionen för den stokastiska variabeln (X, Y) .

Om (X, Y) är diskret låter vi

$$p_{X,Y}(j, k) = P(X = j, Y = k).$$

Detta är **sannolikhetsfunktionen** för (X, Y) . Fördelningsfunktionen ges då av

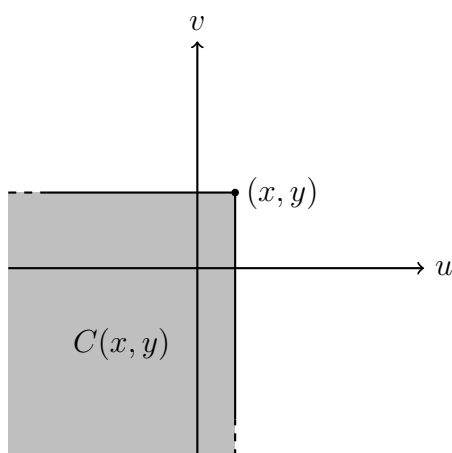
$$F_{X,Y}(x, y) = \sum_{j \leq x} \sum_{k \leq y} p_{X,Y}(j, k).$$

Om det finns en icke-negativ integrerbar funktion $f_{X,Y}$ så att

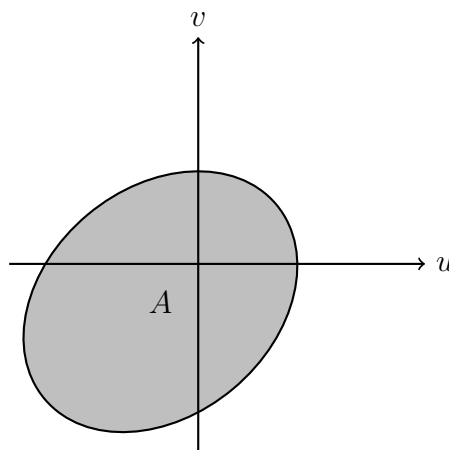
$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) \, du \, dv,$$

så kallar vi $f_{X,Y}$ för variabelns **simultana täthetsfunktion**. Detta är typfallet för att (X, Y) är en kontinuerlig tvådimensionell stokastisk variabel.

Sannolikheten $F_{X,Y}(x, y)$ och sannolikheten för en mer generell delmängd $A \subset \mathbf{R}^2$ kan grafiskt illustreras genom figuren nedan. Observera att sannolikheten inte är den skuggade arean, utan volymen som uppstår när vi har en funktionsyta definierad ovanför det skuggade området. Arean symboliserar alltså ett integrations- eller summationsområde. Vi ”summerar” (via en integral eller summa) sedan sannolikhetsvärden för de intressanta värdena för variabeln.



Den halvoändliga rektangeln $C(x, y)$.



$P((X, Y) \in A)$ är sannolikheten att få ett resultat (x, y) som ligger i mängden A .



Egenskaper hos sannolikhetsfunktionen

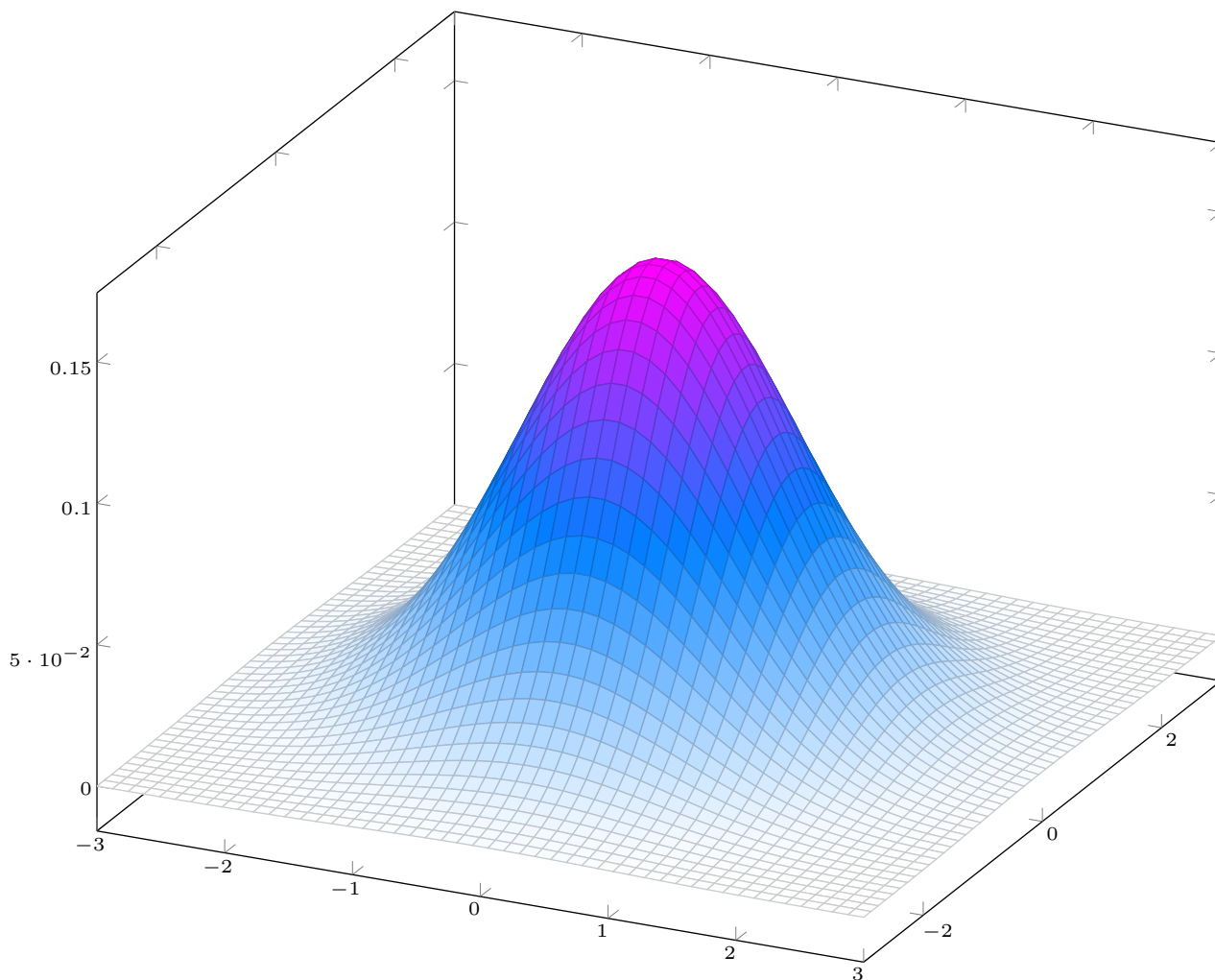
- (i) $p_{X,Y}(j, k) \geq 0$ för alla (j, k) .
- (ii) $\sum_j \sum_k p_{X,Y}(j, k) = 1$.
- (iii) Om $A \subset \mathbf{R}^2$ så är $P((X, Y) \in A) = \sum_{(j,k) \in A} p_{X,Y}(j, k)$.



Egenskaper hos den simultana täthetsfunktionen

- (i) $f_{X,Y}(x, y) \geq 0$ för alla $(x, y) \in \mathbf{R}^2$.
- (ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$.
- (iii) Om $A \subset \mathbf{R}^2$ är snäll så är $P((X, Y) \in A) = \int \int_A f_{X,Y}(x, y) dx dy$.
- (iv) Talet $f_{X,Y}(x, y)$ anger hur mycket sannolikhetsmassa det finns per areaenhet i punkten (x, y) .

Exempel på hur en tvådimensionell täthetsfunktion kan se ut. Det är nu volymen, inte arean, som ska vara ett.



Exempel

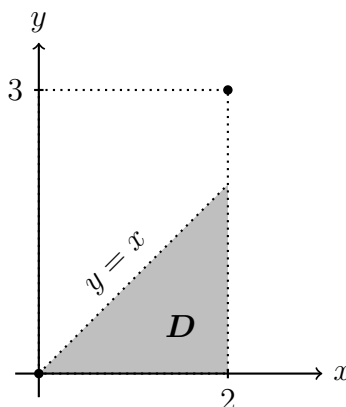
Det enklaste exemplet på en 2D-täthetsfunktion är den likformiga fördelningen. Låt A vara rektangeln med hörn i $(0,0)$ och $(2,3)$ och låt (X,Y) vara likformigt fördelad på A . Då ges $f_{X,Y}$ av $f_{X,Y}(x,y) = 1/6$ om $(x,y) \in A$ och $f_{X,Y}(x,y) = 0$ för övrigt. Hitta sannolikheten att $X > Y$.

Lösning. Det är alltid bra att försöka rita en figur, så vi gör detta parallellt med att vi betraktar arean av de delar vi är intresserad av.

Vi finner alltså denna sannolikhet genom att titta på hur stor del av A där detta villkor (dvs att $x > y$) är uppfyllt, vilket är triangeln med hörn i $(0,0)$, $(2,0)$ och $(2,2)$. Vi erhåller alltså

$$P(X > Y) = \frac{\text{area triangel}}{\text{area rektangel}} = \frac{2}{6} = \frac{1}{3}.$$

Testa att ställa upp en integral. Blir svaret detsamma?





Definition. De **marginella** täthetsfunktionerna f_X och f_Y för X och Y i en kontinuerlig stokastisk variabel (X, Y) ges av

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \quad \text{och} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx.$$

Motsvarande gäller om (X, Y) är diskret:

$$p_X(j) = \sum_k p_{X,Y}(j, k) \quad \text{och} \quad p_Y(k) = \sum_j p_{X,Y}(j, k).$$

Man kan även definiera marginella fördelningsfunktioner genom

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) \quad \text{och} \quad F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y).$$

Så vad är egentligen dessa *marginella* funktioner? Det vi gör är att vi summerar alla möjligheter för den variabel vi inte är intresserade av och på det sättet skapar något som bara beror på en variabel. Detta leder också till följande sats.



Oberoende variabler

Sats. Om (X, Y) är en stokastisk variabel med simultan täthetsfunktion $f_{X,Y}$ gäller att X och Y är oberoende om och endast om $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. För en diskret variabel är motsvarande villkor $p_{X,Y}(j, k) = p_X(j)p_Y(k)$.

Det är även sant att (i båda fallen) X och Y är oberoende om och endast om

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$



Exempel

Låt (X, Y) ha sannolikhetsfunktionen $p_{X,Y}(j, k) = c(jk + k^3)$ för $j = 0, 1, 2, 3$ och $k = 0, 1$. Annars är $p_{X,Y} = 0$.

- (i) Vad är c ?
- (ii) Bestäm $p_X(j)$ och $p_Y(k)$. Är X och Y oberoende?
- (iii) Beräkna $P(X \leq 2, Y \geq 0.5)$.

Lösning:

- (i) Vi måste välja $c > 0$ för att $p_{X,Y}$ ska kunna vara en sannolikhetsfunktion. För att finna c summerar vi över alla j och k :

$$\sum_{j=0}^3 \sum_{k=0}^1 c(jk + k^3) = \sum_{j=0}^3 c(j+1) = c(1+2+3+4) = 10c,$$

så $c = 1/10$ är nödvändigt och tillräckligt.

(ii) De marginella sannolikhetsfunktionerna fås ur definitionen enligt

$$p_X(j) = \sum_{k=0}^1 c(jk + k^3) = c(j+1), \quad j = 0, 1, 2, 3.$$

$$p_Y(k) = \sum_{j=0}^3 c(jk + k^3) = c(k^3 + k + k^3 + 2k + k^3 + 3k + k^3) = 2c(3k + 2k^3), \quad k = 0, 1.$$

Vi testar även att

$$\sum_{j=0}^3 p_X(j) = c(1 + 2 + 3 + 4) = 1,$$

$$\sum_{k=0}^1 p_Y(k) = 2c(3 + 2) = 1,$$

så p_X och p_Y är sannolikhetsfunktioner. Vi undersöker oberoendet:

$$p_X(j)p_Y(k) = 2c^2(j+1)(3k+2k^3).$$

Här kan man kanske direkt tro att X och Y är beroende, men det skulle vara en gissning. Vi undersöker explicit:

$p_{X,Y}$	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$p_X p_Y$	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$k = 0$	0	0	0	0	$k = 0$	0	0	0	0
$k = 1$	1/10	2/10	3/10	4/10	$k = 1$	$\frac{1 \cdot 2 \cdot 5}{100}$	$\frac{2 \cdot 2 \cdot 5}{100}$	$\frac{3 \cdot 2 \cdot 5}{100}$	$\frac{4 \cdot 2 \cdot 5}{100}$

Vi ser här att alla siffror matchar varandra och att därmed $p_{X,Y}(j, k) = p_X(j)p_Y(k)$ för *alla* j och k . Detta trots att uttrycken såg väldigt olika ut från början. Var försiktiga med att dra slutsatser utan att undersöka ordentligt!

(iii) Vi kan numrera de tillåtna värdena (där $X = j \leq 2$ och $Y = k \geq 0.5$) på (j, k) : $(0, 1)$, $(1, 1)$, $(2, 1)$. Sannolikheten blir då

$$P(X \leq 2, Y \geq 0.5) = p_{X,Y}(0, 1) + p_{X,Y}(1, 1) + p_{X,Y}(2, 1) = c(1 + 2 + 3) = 6/10.$$



Exempel

Låt (X, Y) ha den simultana täthetsfunktionen

$$f_{X,Y}(x, y) = \begin{cases} c(x^2y + xy^2), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{annars.} \end{cases}$$

Lös följande problem.

(i) Bestäm c ;

(ii) Vad är sannolikheten för händelsen att $X > 1/2$ och att $Y < 1/2$? Det vill säga, beräkna sannolikheten $P(X > 1/2, Y < 1/2)$;

Lösning:

(i) Vi räknar ut följande

$$1 = \iint_{\mathbf{R}^2} f_{X,Y}(x,y) dx dy = \int_0^1 \int_0^1 c(x^2y + xy^2) dx dy = c \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2} \right) dy = \frac{c}{3},$$

så $c = 3$ är nödvändigt för att få en täthetsfunktion.

(ii) Vi har

$$P(X > 1/2 \text{ och } Y < 1/2) = \int_{1/2}^1 \int_0^{1/2} 3(x^2y + xy^2) dy dx = \frac{5}{32}.$$

**Exempel**

Med samma täthetsfunktion som i föregående exempel, lös följande problem.

- (i) Beräkna $P(X > 1/2)$;
- (ii) Beräkna $P(Y < 1/2 \mid X > 1/2)$;
- (iii) Beräkna $f_X(x)$ och $f_Y(y)$. Är X och Y oberoende?
- (iv) Beräkna $F_{X,Y}(x, y)$.
- (v) Vad blir $\frac{\partial^2}{\partial_x \partial_y} F_{X,Y}(x, y)$?

Lösning:(i) Vi har endast ett krav på x , så vi integrerar över alla y :

$$P(X > 1/2) = \int_{1/2}^1 \int_0^1 3(x^2y + xy^2) dy dx = \frac{13}{16}.$$

(ii) Enligt definitionen av betingad sannolikhet,

$$P(Y < 1/2 \mid X > 1/2) = \frac{P(X > 1/2, Y < 1/2)}{P(X > 1/2)} = \frac{5/32}{13/16} = \frac{5}{26}.$$

(iii) De marginella tätheterna beräknas direkt från definitionen:

$$f_X(x) = \int_0^1 3(xy^2 + x^2y) dy = x + \frac{3}{2}x^2, \quad 0 \leq x \leq 1,$$

$$f_Y(y) = \int_0^1 3(xy^2 + x^2y) dx = y + \frac{3}{2}y^2, \quad 0 \leq y \leq 1.$$

Vi ser tydligt att $f_X(x)f_Y(y) \neq f_{X,Y}(x, y)$ för många val av punkter (x, y) ; ta till exempel $(x, y) = (1, 1)$. Variablerna är beroende.

(iv) Låt $(x, y) \in [0, 1] \times [0, 1]$. Då blir

$$F_{X,Y}(x, y) = \int_0^x \int_0^y 3(uv^2 + u^2v) dv du = \frac{x^3y^2 + x^2y^3}{2}.$$

Om $x < 0$ eller $y < 0$, måste $F_{X,Y}(x, y) = 0$, och om $x > 1$ och $y > 1$, så måste $F_{X,Y}(x, y) = 1$. Övriga fall täcks av

$$F_{X,Y}(x, y) = \frac{y^2 + y^3}{2}, \quad 0 \leq y \leq 1 \text{ och } x > 1,$$

$$F_{X,Y}(x, y) = \frac{x^3 + x^2}{2}, \quad 0 \leq x \leq 1 \text{ och } y > 1.$$

(v) Om $0 \leq x \leq 1$ och $0 \leq y \leq 1$,

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = \frac{\partial}{\partial x} \left(\frac{2x^3y + 3x^2y^2}{2} \right) = \frac{6x^2y + 6xy^2}{2} = f_{X,Y}(x, y).$$

Övriga kombinationer av x och y kommer att ge $\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = 0$.

5.3 Väntevärden

Väntevärden av funktioner av flerdimensionella stokastiska variabler fungerar analogt med tidigare. Låt oss formulera en sats.



Väntevärde och funktioner av stokastiska variabler

Sats. Låt $Y = g(X)$ och $W = h(U, V)$. I de kontinuerliga fallen blir

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \text{och} \quad E(W) = \iint_{\mathbf{R}^2} h(x, y) f_{U,V}(x, y) dx dy,$$

och om X, U, V är diskreta:

$$E(Y) = \sum_k g(k) p_X(k) \quad \text{och} \quad E(W) = \sum_j \sum_k h(j, k) p_{U,V}(j, k).$$

Det kontinuerliga fallet (med täthetsfunktion) kommer vi inte åt på något annat sätt än att egentligen ta satsen ovan som definition. Om vi tänker på punkt (ii) i rutan föregående satsen, så förefaller det ganska rimligt. Sammansättningen $g(Y)$ till exempel är en ny stokastisk variabel och då skulle

$$E(g(Y)) = \int_{\Omega} g(Y)(\omega) dP(\omega),$$

vilket är precis hur satsen tolkas. I det diskreta fallet kan vi faktiskt producera ett bevis:

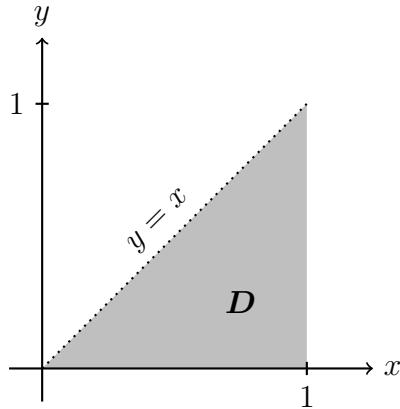
$$\begin{aligned} E(g(X)) &= \sum_k k P(g(X) = k) = \sum_k k \sum_{m: g(m)=k} P(X = m) \\ &= \sum_m \sum_{k: g(m)=k} k P(X = m) = \sum_m P(X = m) \sum_{k: g(m)=k} k = \sum_m g(m) P(X = m), \end{aligned}$$

om vi antar att serien är absolutkonvergent i det oändliga fallet så vi kan byta summationsordningen. Rent praktiskt kan beräkningarna gå till på följande sätt.

**Exempel**

Låt $f_{X,Y}(x, y) = 2$ om $0 < y < x < 1$ och $f_{X,Y}(x, y) = 0$ för övrigt. Bestäm väntevärdet $E(XY + Y^2X)$.

Lösning: Vi använder satsen ovan:



$$\begin{aligned}
 E(XY + Y^2X) &= \iint_{\mathbf{R}^2} (xy + y^2x) f_{X,Y}(x, y) dx dy \\
 &= 2 \int_0^1 \int_0^x (xy + y^2x) dy dx \\
 &= 2 \int_0^1 \left[\frac{xy^2}{2} + \frac{xy^3}{3} \right]_0^x dx \\
 &= \int_0^1 \left(x^3 + \frac{2x^4}{3} \right) dx \\
 &= \frac{1}{4} + \frac{2}{15} = \frac{23}{60}.
 \end{aligned}$$

**Exempel**

Låt X och Y vara oberoende med $E(X) = 2$, $E(X^2) = 8$, $E(Y) = -1$, $V(Y) = 2$. Beräkna $E(XY)$, $E(X^2Y^2)$ och $D(2X - 3Y)$.

Lösning. Här behöver vi inte använda satsen direkt. Eftersom variablerna är oberoende blir $E(XY) = E(X)E(Y) = 2 \cdot (-1) = -2$, och

$$E(X^2Y^2) = E(X^2)E(Y^2) = 8E(Y^2) = 8(V(Y) + E(Y)^2) = 8 \cdot 3 = 24.$$

Vidare erhåller vi

$$V(2X - 3Y) = 2^2V(X) + (-3)^2V(Y) = 4(8 - 2^2) + 9 \cdot 2 = 34,$$

så $D(2X - 3Y) = \sqrt{34}$.

**Exempel**

Låt X och Y vara oberoende likafördelade variabler med $p_X(k) = p_Y(k) = 1/4$ om $k = 1, 2$ och $p_X(k) = p_Y(k) = 1/2$ om $k = 0$. Bestäm sannolikhetsfunktionen för $Z = X + Y$, väntevärdet $E(X + Y)$ samt variansen $V(X + Y)$.

Lösning. Vi börjar med att bestämma $p_Z(k)$.

$Z = X + Y = n$	Par (j, k) så $j + k = n$	Total sannolikhet
-1	–	0
0	(0, 0)	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$
1	(0, 1), (1, 0)	$2 \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{4}$
2	(0, 2), (1, 1), (2, 0)	$2 \cdot \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} = \frac{5}{16}$
3	(1, 2), (2, 1)	$2 \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{8}$
4	(2, 2)	$\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$
5	–	0

Två sätt att beräkna väntevärdet:

$$E(Z) = E(X + Y) = E(X) + E(Y) = 2E(X) = 2 \sum_{k=0}^2 kp_X(k) = 2 \left(0 + \frac{1}{4} + \frac{2}{4} \right) = \frac{3}{2},$$

$$E(Z) = \sum_{k=0}^4 kp_Z(k) = \frac{1}{4} + \frac{10}{16} + \frac{3}{8} + \frac{4}{16} = \frac{24}{16} = \frac{3}{2}.$$

Möjligen kan man tycka att det andra alternativet ser enklare ut, men då har man glömt bort allt arbete som gick åt till att skapa tabellen ovan. Det första alternativet är nästan alltid det enklaste. Variansen beräknar man enklast enligt följande

$$V(Z) = V(X + Y) = [\text{oberoende variabler}] = V(X) + V(Y) = 2V(X) = \frac{1}{8}$$

eftersom vi vet att

$$V(X) = E(X^2) - E(X)^2 = \sum_{k=0}^2 k^2 p_X(k) - \frac{9}{16} = \left(0 + \frac{1}{4} + \frac{4}{4} \right) - \frac{9}{16} = \frac{1}{16}.$$

5.4 Kovarians

Vi har betraktat variansen för stokastiska variabler, men kan man säga någon om variationen mellan två variabler utan att stirra sig blind på sannolikhets- eller täthetsfunktionen? Visst kan man det, och svaret kommer i form av kovarians eller korrelation



Definition. Låt X och Y vara två stokastiska variabler sådana att $E(X) = \mu_X$, $V(X) = \sigma_X^2$, $E(Y) = \mu_Y$ samt $V(Y) = \sigma_Y^2$. Kovariansen $C(X, Y)$ definieras enligt

$$C(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

och korrelationen mellan X och Y enligt

$$\rho(X, Y) = \frac{C(X, Y)}{\sigma_X \sigma_Y}.$$

Både kovarians och korrelation är ett mått på linjärt beroende mellan X och Y där korrelationen är normerad så det går att jämföra olika fall.



Okorrelerad

Definition. Om $C(X, Y) = 0$ kallas X och Y för okorrelerade.

Kovariansen uppfyller följande egenskaper.

- (i) $C(X, Y) = E(XY) - E(X)E(Y)$.
- (ii) Om X och Y är oberoende så är $C(X, Y) = 0$.
- (iii) $|\rho(X, Y)| \leq 1$ med likhet om och endast om det finns ett linjärt samband mellan X och Y .

- (iv) $C(X, X) = V(X)$.
- (v) $C\left(a_0 + \sum_{i=1}^m a_i X_i, b_0 + \sum_{j=1}^n b_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j C(X_i, Y_j)$.

Bevis.

- (i) Detta följer från i princip samma argument som Steiners sats. Låt oss expandera definition av kovarians:

$$C(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y = E(XY) - \mu_X \mu_Y,$$

vilket är precis vad vi ville visa.

- (ii) Eftersom X och Y är oberoende, så följer det att $E(XY) = E(X)E(Y)$. Detta leder givetvis med föregående punkt i tanken till att $C(X, Y) = 0$.
- (iii) Väntevärdet av en kvadrat är givetvis icke-negativt, så

$$\begin{aligned} 0 &\leq E\left(\left(\frac{X - \mu_X}{\sigma_X} \pm \frac{Y - \mu_Y}{\sigma_Y}\right)^2\right) \\ &= \frac{1}{\sigma_X^2} E((X - \mu_X)^2) + \frac{1}{\sigma_Y^2} E((Y - \mu_Y)^2) \pm 2 \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} \\ &= 2(1 \pm \rho(X, Y)). \end{aligned}$$

Alltså måste $|\rho(X, Y)| \leq 1$. Dessutom ser vi på köpet att om $\rho(X, Y) = \pm 1$ så måste väntevärdet av kvadraten vara lika med noll, så

$$\frac{X - \mu_X}{\sigma_X} \pm \frac{Y - \mu_Y}{\sigma_Y} = 0 \quad \Leftrightarrow \quad Y = \mu_Y \pm \frac{\sigma_Y(X - \mu_X)}{\sigma_X} \quad \Leftrightarrow \quad Y = aX + b.$$

Således är beroendet mellan X och Y i form av en rät linje. En lite kvalifikation är nödvändig: det är en integral som blir noll och vi drar slutsatsen att integranden är noll. Detta är inte helt självklart. Nu råkar vi veta att integranden är icke-negativ, så vi har ingen negativ area som spökar. Men vi kan fortfarande modifiera integranden här och där utan att ändra integralen, så viss försiktighet är nödvändig. Formellt heter detta att likheten gäller *nästan överallt*. Men vi är tillbaka till Lebesgue-integralen nu, så låt oss lämna området kvickt.

- (iv) Vi ser direkt att

$$C(X, X) = E(X^2) - E(X)^2 = V(X)$$

vilket är variansen enligt Steiners sats.

- (v) Ohyggligt åbåke som lämnas som övning.. skämt åsido så kommer ni stöta på detta igen, men då med metoder från linjär algebra i bagaget så saker och ting faktiskt går att hantera utan tårar.



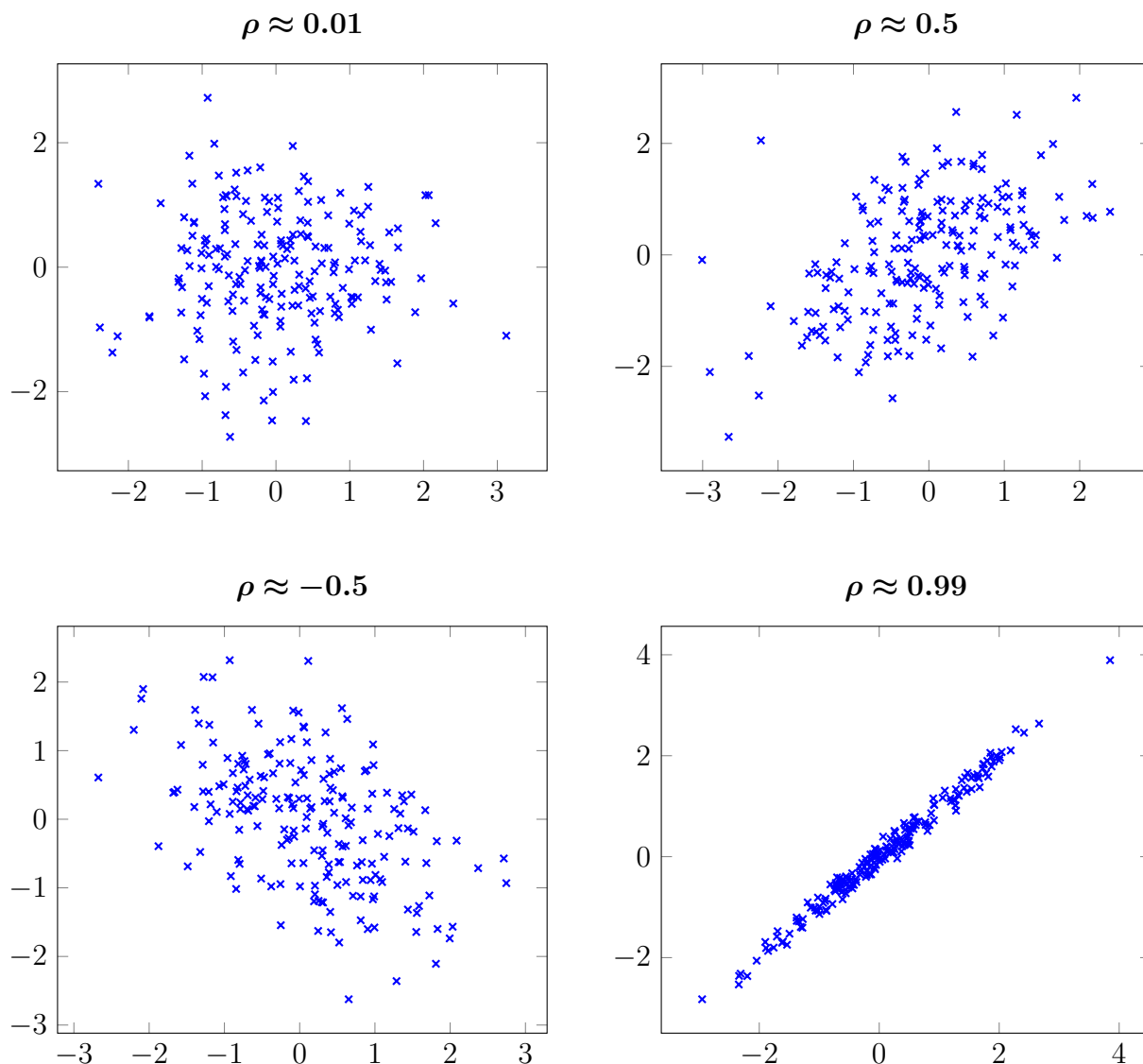
Observera att $C(X, Y) = 0$ *inte* nödvändigtvis innebär oberoende. Låt till exempel X vara rektangelfördelad enligt $X \sim \text{Re}(-1, 1)$ och definiera $Y = X^2$. Uppenbarligen beroende variabler, men

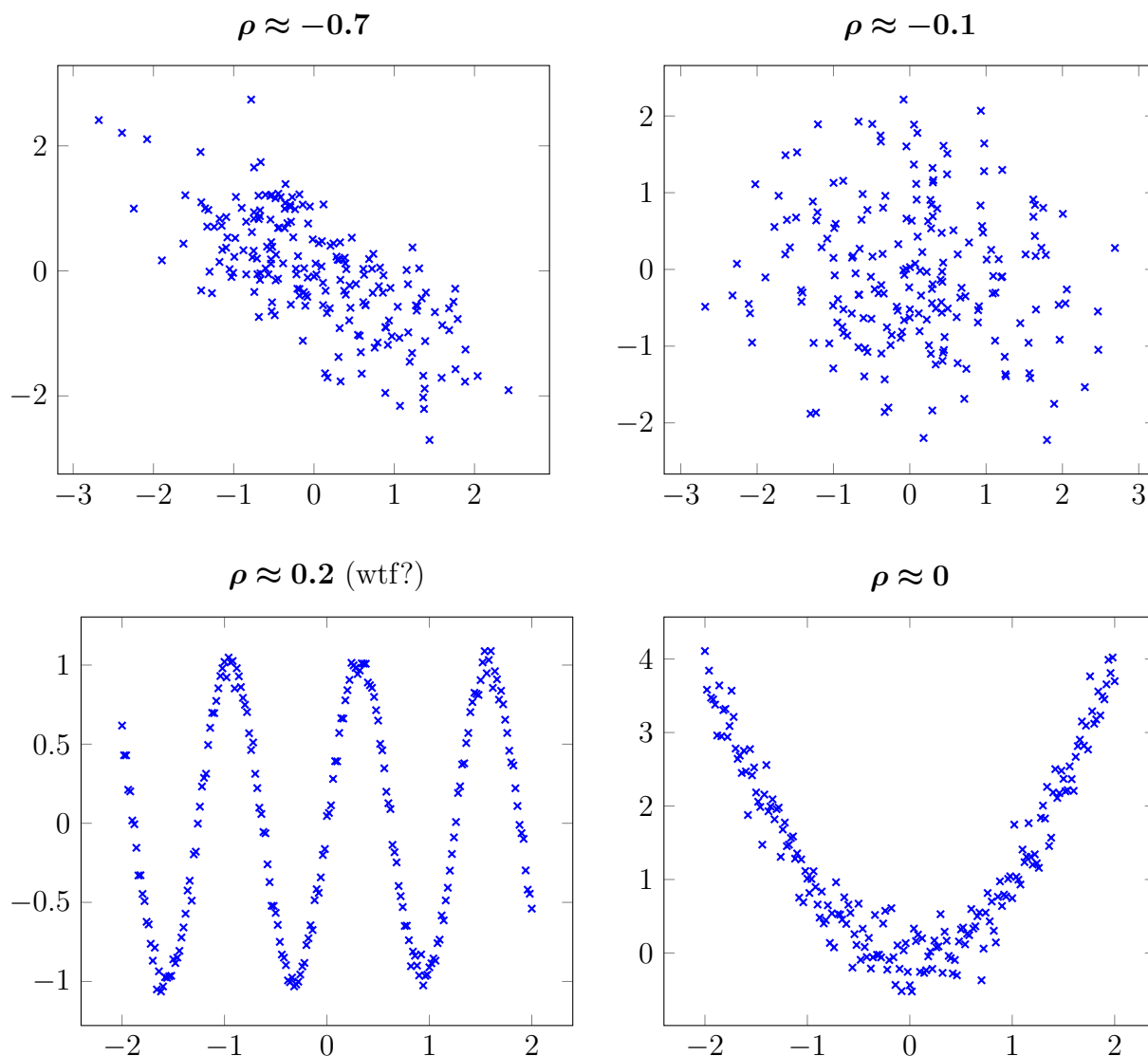
$$C(X, Y) = E(XY) - E(X)E(Y) = E(X^3) - 0 \cdot E(Y) = E(X^3) = \int_{-1}^1 x^3 \cdot \frac{1}{2} dx = 0,$$

så X och Y är okorrelerade.

5.5 Vad innebär korrelationen grafiskt?

Korrelation används ofta för att avgöra om det finns ett linjärt samband mellan x och y -värden när vi på något sätt fått en samling data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Även detta återkommer i nästa del av kursen!





5.6 Funktioner av stokastiska variabler



Exempel

Om $X \sim \text{Exp}(\lambda)$, vad får $Y = e^X$ för täthetsfunktion?

Lösning: Vad blir f_Y ? Vi ser att $Y > 0$ från definitionen så $f_Y(y) = 0$ för $y \leq 0$. Vi ställer upp $F_Y(y)$ för $y > 0$:

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \log y) = F_X(\log y), \quad y > 0.$$

Vi deriverar fram $f_Y(y) = F'_Y(y) = \frac{1}{y} \cdot f_X(\log y)$. Vi vet att $f_X(x) = \lambda e^{-\lambda x}$ för $x > 0$ och om $x \leq 0$ blir $f_X(x) = 0$. Eftersom $\log y < 0$ då $0 < y < 1$ får vi två fall:

$$f_Y(y) = \begin{cases} \frac{\lambda}{y} e^{-\lambda \log y}, & y > 1, \\ 0, & y \leq 1, \end{cases} = \begin{cases} \lambda y^{-1-\lambda}, & y > 1, \\ 0, & y \leq 1. \end{cases}$$

**Exempel**

Om $X \sim \text{Re}(-1, 2)$, vad får $Y = X^2$ för täthetsfunktion?

Lösning: Om $y < 0$ så måste $F_Y(y) = P(Y \leq y) = 0$ ($Y = X^2$ kommer aldrig att vara negativ). Låt oss anta att $y \geq 0$. Vi beräknar

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Vi antar att f_Y är kontinuerlig och deriverar fram ett uttryck:

$$f_Y(y) = F'_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})), & y > 0, \\ 0, & y \leq 0. \end{cases}$$

Vidare vet vi att $f_X(x) = 1/3$ om $-1 \leq x \leq 2$ och $f_X(x) = 0$ annars, så

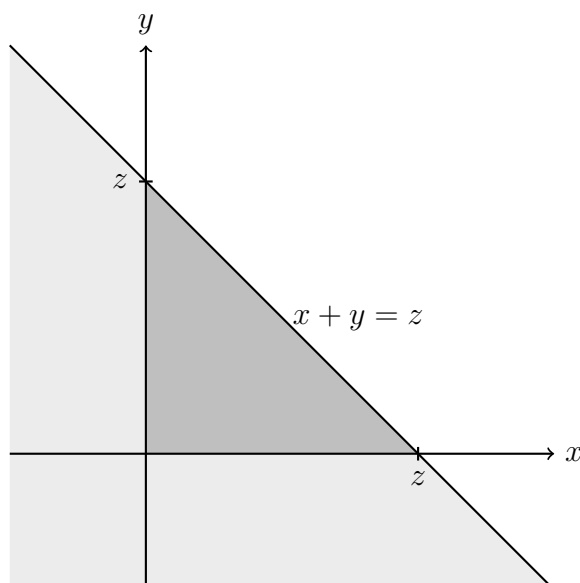
$$f_Y(y) = \begin{cases} 0, & y < 0, \\ \frac{1}{3\sqrt{y}}, & 0 \leq y < 1, \\ \frac{1}{6\sqrt{y}}, & 1 \leq y < 4, \\ 0, & y \geq 4. \end{cases}$$

Om vi har den simultana täthets- eller sannolikhetsfunktionen så kan vi hitta fördelningar för funktioner av flera stokastiska variabler. Låt oss betrakta ett par vanliga exempel.

**Exempel**

Låt $X \sim \text{Exp}(1)$ och $Y \sim \text{Exp}(2)$ vara oberoende. Beräkna täthetsfunktionen för $Z = X + Y$.

Klart att $f_{X,Y}(x,y) = f_X(x)f_Y(y) = 2 \cdot 1 \cdot e^{-x-2y}$ för $x \geq 0$ och $y \geq 0$ (annars $f_{X,Y} = 0$). Vi söker $f_Z(z)$. Det är klart att om $z < 0$ så är $f_{X,Y}(x,y) = 0$. Antag att $z \geq 0$. Vi ställer upp fördelningsfunktionen $F_Z(z)$ för Z , och för det behöver vi ha klart för oss vilket område vi integrerar över:



Fördelningsfunktionen blir nu, för $z \geq 0$,

$$\begin{aligned} F_Z(z) &= P(X + Y \leq z) = \int_{-\infty}^z \int_{-\infty}^{z-x} f_{X,Y}(x, y) dy dx \\ &= 2 \int_0^z e^{-x} \int_0^{z-x} e^{-2y} dy dx = \int_0^z e^{-x} (1 - e^{-2(z-x)}) dx \\ &= \int_0^z (e^{-x} - e^{-2z+x}) dx = 1 + e^{-2z} - 2e^{-z}, \end{aligned}$$

och vi kan derivera fram $f_Z(z) = F'_Z(z) = 2e^{-z} - 2e^{-2z}$. Kontrollera att $f_Z(z) \geq 0$ samt att $\int_0^\infty f_Z(z) dz = 1$.



Faltningssatsen

Sats. Om X och Y är oberoende kontinuerliga stokastiska variabler så ges täthetsfunktionen f_Z för $Z = X + Y$ av

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx, \quad z \in \mathbf{R}.$$

Motsvarande gäller för diskreta variabler:

$$p_Z(k) = \sum_j p_X(j) p_Y(k - j).$$



Min och max

Vad blir fördelningsfunktionerna för maximum respektive minimum av två oberoende stokastiska variabler?

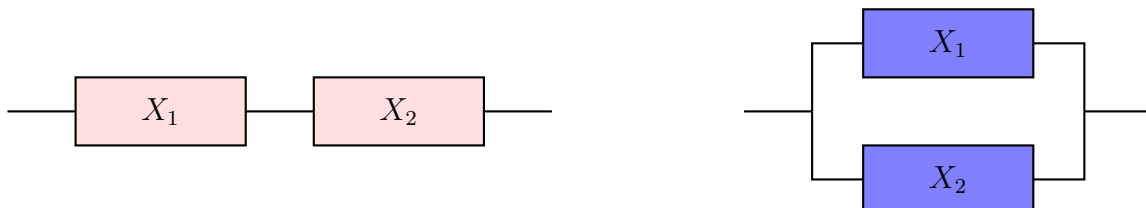
Låt $Y = \max(X_1, X_2)$. Då blir

$$\begin{aligned} F_Y(y) &= P(\max(X_1, X_2) \leq y) = P(X_1 \leq y \text{ och } X_2 \leq y) = P(X_1 \leq y) P(X_2 \leq y) \\ &= F_{X_1}(y) F_{X_2}(y). \end{aligned}$$

För $Y = \min(X_1, X_2)$ så erhåller vi

$$\begin{aligned} F_Y(y) &= P(\min(X_1, X_2) \leq y) = 1 - P(\min(X_1, X_2) > y) = 1 - P(X_1 > y \text{ och } X_2 > y) \\ &= 1 - P(X_1 > y) P(X_2 > y) = 1 - (1 - P(X_1 \leq y))(1 - P(X_2 \leq y)) \\ &= 1 - (1 - F_{X_1}(y))(1 - F_{X_2}(y)). \end{aligned}$$

En vanlig situation där dessa uttryck dyker upp är i parallell- och seriekoppling av system. Schematiskt kan man beskriva dessa situationer med blockscheman.



Seriekoppling. Båda kanalerna måste fungera. Ger minimum av livslängderna.

Parallellkoppling. Räcker att en kanal fungerar. Ger maximum av livslängderna.

5.7 (★★)Formell definition av 2D-stokastisk variabel

I \mathbf{R}^2 är Borelfamiljen \mathcal{B} den minsta σ -algebra som innehåller alla öppna rektanglar $(a, b) \times (c, d)$.



Stokastisk variabel

Definition. En tvådimensionell stokastisk variabel är en reell-vektorvärd funktion (X, Y) (två komponenter) definierad på ett utfallsrum Ω . Alltså avbildar (X, Y) olika utfall på reella vektorer; $(X, Y): \Omega \rightarrow \mathbf{R}^2$. Vi kräver att $(X, Y)^{-1}(B) \in \mathcal{F}$ för alla $B \in \mathcal{B}$. Algebran \mathcal{F} är mängden av alla tillåtna händelser. Om (X, Y) bara antar ändligt eller uppräknligt många värden så kallar vi (X, Y) för en diskret stokastisk variabel. Om varken X eller Y är diskret kallar vi (X, Y) för kontinuerlig.

5.8 (★)Bökigt exempel



Två sätt att räkna ut $E(g(X))$

Låt $\Theta \sim \text{Re}(0, \pi/4)$ vara likformigt fördelad och definiera $Y = \cos \Theta$. Vad blir $E(Y)$?

Lösning: Det enklaste sättet är att använda satsen från början av föreläsningen:

$$E(Y) = E(\cos \Theta) = \int_{-\infty}^{\infty} \cos \theta f_{\Theta}(\theta) d\theta = \frac{4}{\pi} \int_0^{\pi/4} \cos \theta d\theta = \frac{4}{\pi} \left(\frac{\sqrt{2}}{2} - 0 \right) = \frac{2\sqrt{2}}{\pi}.$$

Den andra varianten börjar med beräkning av täthetsfunktionen för $Y = \cos \Theta$. Vi ställer upp fördelningsfunktionen först:

$$F_Y(y) = P(Y \leq y) = P(\cos \Theta \leq y) = P(\Theta \geq \arccos y) = 1 - P(\Theta < \arccos y) = 1 - F_{\Theta}(\arccos y).$$

Här har vi utnyttjat att $\cos \theta$ är avtagande för $\theta \in [0, \pi/4]$. Vi kan nu derivera fram $f_Y(y)$:

$$\begin{aligned} f_Y(y) &= F'_Y(y) = -F'_{\Theta}(\arccos y) \cdot \frac{-1}{\sqrt{1-y^2}} = \frac{f_{\Theta}(\arccos y)}{\sqrt{1-y^2}} \\ &= \begin{cases} \frac{4}{\pi} \frac{1}{\sqrt{1-y^2}}, & 0 \leq \arccos y \leq \frac{\pi}{4} \\ 0, & \text{för övrigt.} \end{cases} \Leftrightarrow \begin{cases} \frac{\sqrt{2}}{2} \leq y \leq 1 \\ \text{för övrigt.} \end{cases} \end{aligned}$$

Vi kan nu beräkna $E(Y)$ enligt definitionen:

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \frac{4}{\pi} \int_0^{\sqrt{2}/2} \frac{y}{\sqrt{1-y^2}} dy = \frac{4}{\pi} \left[-\sqrt{1-y^2} \right]_0^{\sqrt{2}/2} = \frac{2\sqrt{2}}{\pi}.$$

Vilket metod tycker du är enklast?

5.9 (★★)Normalfördelning

Täthetsfunktionen φ föreslagen för normalfördelningen $N(0, 1)$ är en täthetsfunktion. Beviset är så roligt att jag inte kan låta bli, men det kräver lite flervariabelanalys. Vi behöver visa att

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

Sättet vi kommer åt det hela är genom att titta på kvadraten och tolka denna som en itererad dubbelintegral:

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right)^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right) \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy\right) = \iint_{\mathbf{R}^2} e^{-(x^2+y^2)/2} dx dy \\ &= \int_0^{\infty} \int_0^{2\pi} r e^{-r^2/2} d\theta dr = 2\pi \left[-e^{-r^2/2}\right]_0^{\infty} = 2\pi.\end{aligned}$$

Kapitel 6

Statistisk inferens

Vi är nu redo för att dyka ned i statistisk inferensteori! I sedvanlig ordning börjar vi med att definiera lite begrepp så vi är överens om vad det är vi diskuterar.



Stickprov

Definition. Låt de stokastiska variablerna X_1, X_2, \dots, X_n vara oberoende och ha samma fördelningsfunktion F . Följden X_1, X_2, \dots, X_n kallas ett **slumpmässigt stickprov** (av F). Ett **stickprov** x_1, x_2, \dots, x_n består av **observationer** av variablerna X_1, X_2, \dots, X_n . Samtliga möjliga observationer brukar kallas **populationen**. Vi säger att **stickprovsstorleken** är n .



Exempel

Antag att vi kastar en perfekt tärning 5 gånger och att dessa kast är oberoende. Före kasten representerar den stokastiska variabeln X_k resultatet vid kast k där alla X_k har samma fördelning; vi vet ännu inte vad resultatet blir, men känner sannolikhetsfördelningen. Följden X_k , $k = 1, 2, \dots, 5$ är det *slumpmässiga stickprovet*.

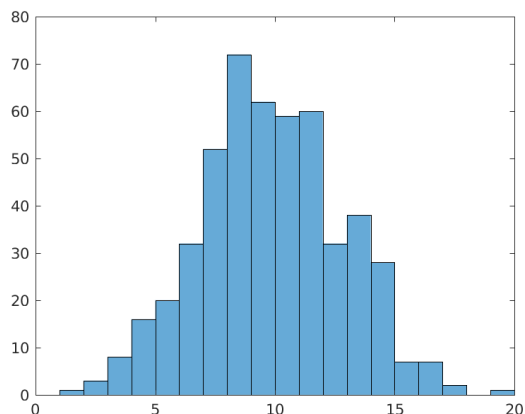
Efter kasten har vi erhållit observationer x_1, x_2, \dots, x_5 av det slumpmässiga stickprovet. Detta är vårt *stickprov* och består alltså av utfallen vid kasten. Dessa observationer tillhör *populationen*. *Stickprovsstorleken* är 5.

Värt att notera är att språkbruket ibland är slarvigt där både stickprov och slumpmässigt stickprov används för att beskriva både följderna av stokastiska variabler och följderna av observationer (utfall). Det viktiga är att hålla koll på vad ni själva menar när ni genomför analyser.

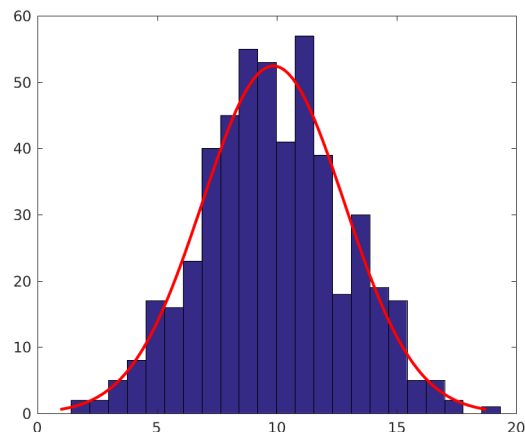
6.1 Representation av stickprov

Man kan representera statistiska data på en hel drös olika sätt med allt från tabeller till stolpdiagram till histogram till lådplottar. Läs avsnittet i boken om detta. Vi nöjer oss med att titta lite närmare på de verktyg vi kommer använda oss av i kursen. Ett mycket vanligt sätt att visualisera fördelningen för en mängd data är med hjälp av histogram. Vi genererar lite normalfördelad slumpdata i MATLAB och renderar ett histogram.

```
>> U = normrnd(10,3,500,1);
>> histogram(U);
```

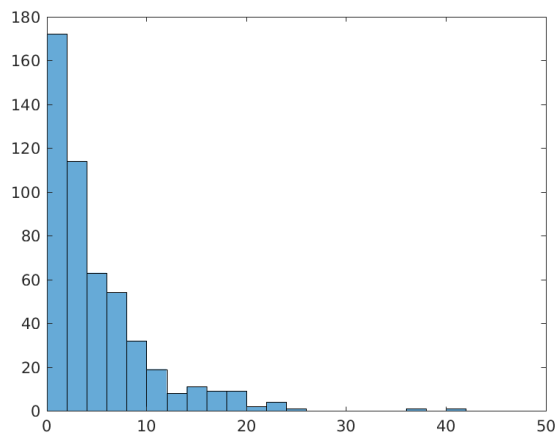


```
>> U = normrnd(10,3,500,1);
>> histfit(U)
```



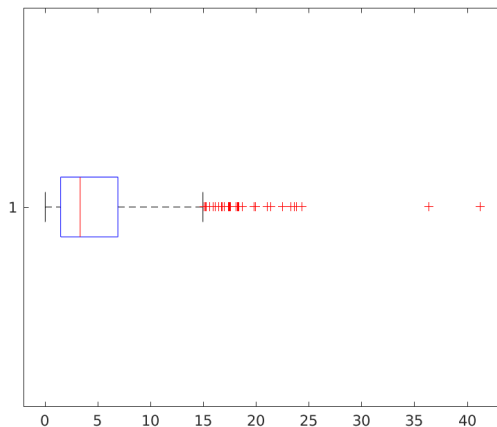
Om vi testar med exponentialfördelning istället blir resultatet enligt nedan.

```
>> U = exprnd(10,500,1);
> histogram(U);
```

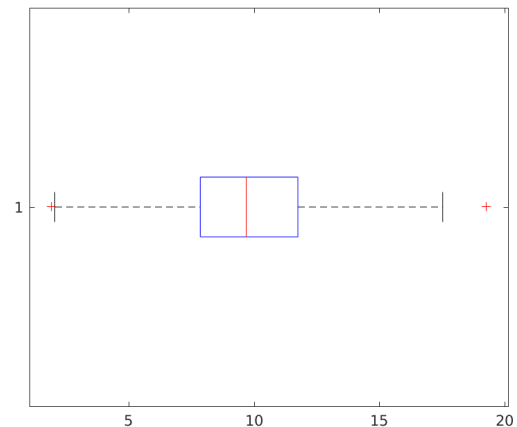


Ett lådagram (boxplot) representerar också materialet, kanske på ett enklare sätt för den oinsatte i sannolikhetsfördelningar. Lådan innehåller 50% av resultaten och den vänstra lådkanten är den undre kvartilen (25% till vänster om den) och den högra är den övre kvartilen (med 25% till höger om den). Medianen markeras med ett streck i lådan. Maximum och minimum markeras med små vertikala streck i slutet på en horisontell linje genom mitten på lådan. Värden som bedöms vara uteliggare markeras med kryss längs samma centrumlinje.

```
>> U = exprnd(5,500,1);
>> boxplot(U, 'orientation', 'horizontal');
```



```
>> U = normrnd(10,3,500,1);
>> boxplot(U, 'orientation', 'horizontal');
```

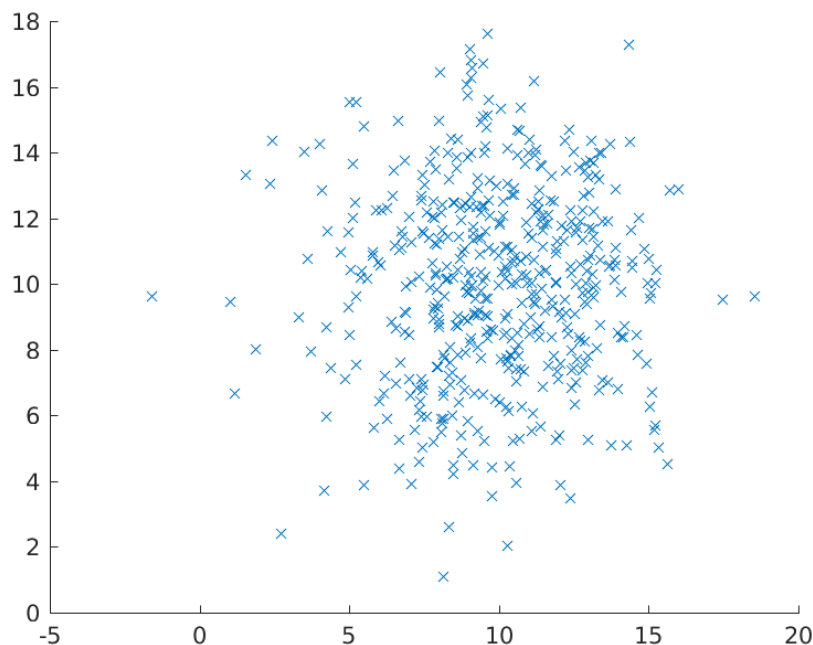


Vi kan tydligt se skillnad på hur mätvärden är spridda. Jämför även med motsvarande histogram ovan.

6.1.1 Två-dimensionell data

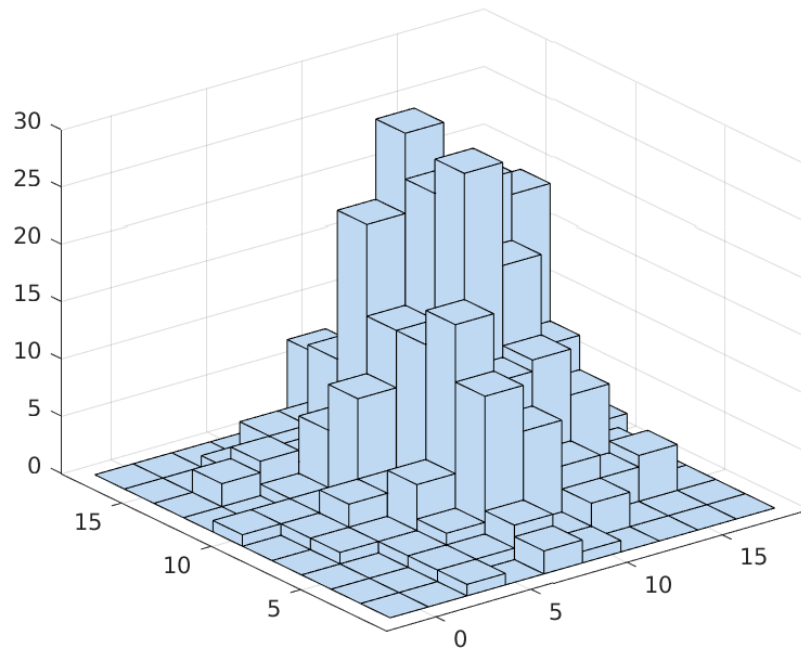
Vi kan även ha mätvärden i form av punkter (x, y) och den vanligaste figuren i dessa sammanhang är ett **spridningsdiagram** (scatter plot) där man helt enkelt plottar ut punkter vid varje koordinat (x_i, y_i) .

```
>> U = normrnd(10,3,500,2);
>> scatter(U(:,1), U(:,2), 'x');
```



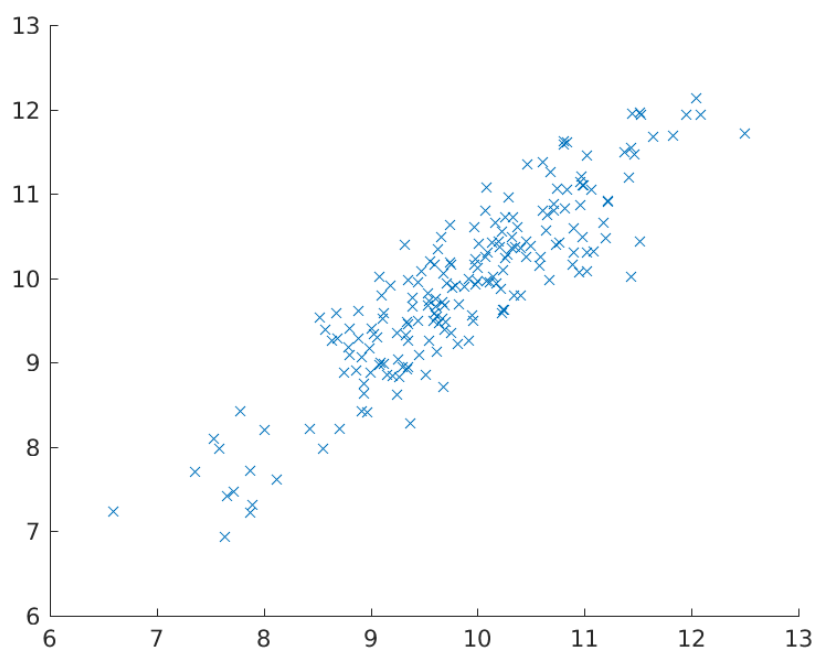
Vi ser från figuren att värdena verkar vara centrerade kring (10,10) och att det verkar föreligga någon form av cirkulär symmetri. Stämmer det för den bivariata normalfördelningen när komponenterna är oberoende? Vi kan även rendera ett två-dimensionellt histogram.

```
>> hist3(U);
```



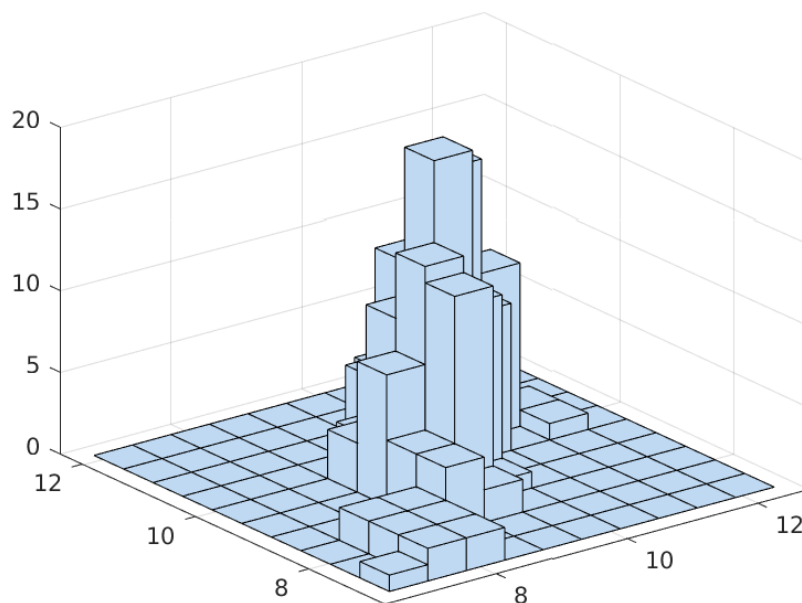
Vad gäller om variablerna i den bivariata normalfördelningen inte är oberoende?

```
>> mu = [10 10]; rho = 0.90; s1 = 1; s2 = 1;
>> Sigma = [s1*s1 s1*s2*rho; s1*s2*rho s2*s2]
>> R = chol(Sigma);
>> z = repmat(mu, 200, 1) + randn(200,2)*R;
>> scatter(z(:,1),z(:,2), 'x');
```



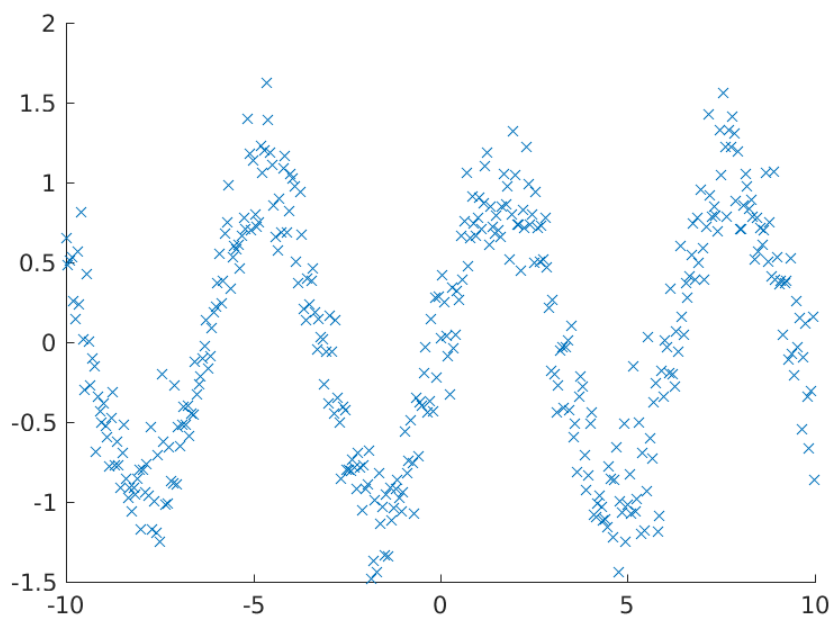
Värdena verkar fortfarande vara centrerade kring (10,10) (i någon mening) men symmetrin verkar nu utdragen diagonalt. Stämmer det för en bivariat normalfördelningen med korrelationen 0.90? Ett histogram kan genereras som ovan.

```
>> hist3(z);
```



Något konstigare? Visst.

```
>> x = (-10:0.05:10); y = sin(x) + normrnd(0,0.25,size(x));
>> scatter(x,y,'x');
```



Vi kan tydligt urskilja sinus-termen och något slags brus som gör att det inte blir en perfekt linje. Kan man få bort bruset?

6.2 Punktskattningar

Antag att en fördelning beror på en okänd parameter θ . Med detta menar vi att fördelningens täthetsfunktion (eller sannolikhetsfunktion) beror på ett okänt tal θ , och skriver $f(x; \theta)$ respektive $p(k; \theta)$ för att markera detta. Om vi har ett stickprov från en fördelning med en okänd parameter, kan vi *skatta* den okända parametern? Med andra ord, kan vi göra en "gissning" på det verkliga värdet på parametern θ ?



Punktskattning

Definition. En **punktskattning** $\hat{\theta}$ av parametern θ är en funktion (ibland kallad **stickprovsfunktion**) av de observerade värdena x_1, x_2, \dots, x_n :

$$\hat{\theta} = g(x_1, x_2, \dots, x_n).$$

Vi definierar motsvarande **stickprovsvariabel** $\hat{\Theta}$ enligt

$$\hat{\Theta} = g(X_1, X_2, \dots, X_n).$$

Det är viktigt att tänka på att $\hat{\theta}$ är en siffra, beräknad från de observerade värdena, medan $\hat{\Theta}$ är en stokastisk variabel. Som vanligt använder vi stora bokstäver för att markera att vi syftar på en stokastisk variabel. Sambandet mellan $\hat{\theta}$ och $\hat{\Theta}$ är alltså att $\hat{\theta}$ är en observation av den stokastiska variabeln $\hat{\Theta}$. Förutom detta dras vi fortfarande med det okända talet θ , som inte är stokastiskt, utan endast en okänd konstant.



Exponentialfördelning

Betrakta en exponentialfördelning med okänt väntevärde. Formeln är välkänd: för alla $x \geq 0$ gäller att $f(x; \mu) = \mu^{-1} \exp(-\mu^{-1}x)$. Parametern θ är alltså väntevärdet μ i detta fall. Ibland använder man exponentialfördelningen för att beskriva elektriska komponenters livslängd, och genom att betrakta ett stickprov kan man då uppskatta livslängden för en hel tillverkningsomgång.



Stokastiskt eller ej?

Var noggran med att tydligt visa och göra skillnad på vad som är stokastiskt eller inte i din redovisning! Vi har tre storheter:

- (i) θ – verkligt värde. Okänt. Deterministiskt.
- (ii) $\hat{\theta}$ – skattat värde. Känt (beräknat från stickprovet). Deterministiskt.
- (iii) $\hat{\Theta}$ – stickprovsvariabeln. Denna är stokastisk!

Sannolikheter som beräknas bör använda sig av $\hat{\Theta}$ då $\hat{\Theta}$ beskriver variationen hos $\hat{\theta}$ för olika stickprov. Om bara $\hat{\theta}$ och θ ingår är sannolikheten alltid noll eller ett (varför?).

Så om vi har ett stickprov från en fördelning som beror på en okänd parameter, hur hittar vi skattningsfunktionen g ? Fungerar vad som helst?

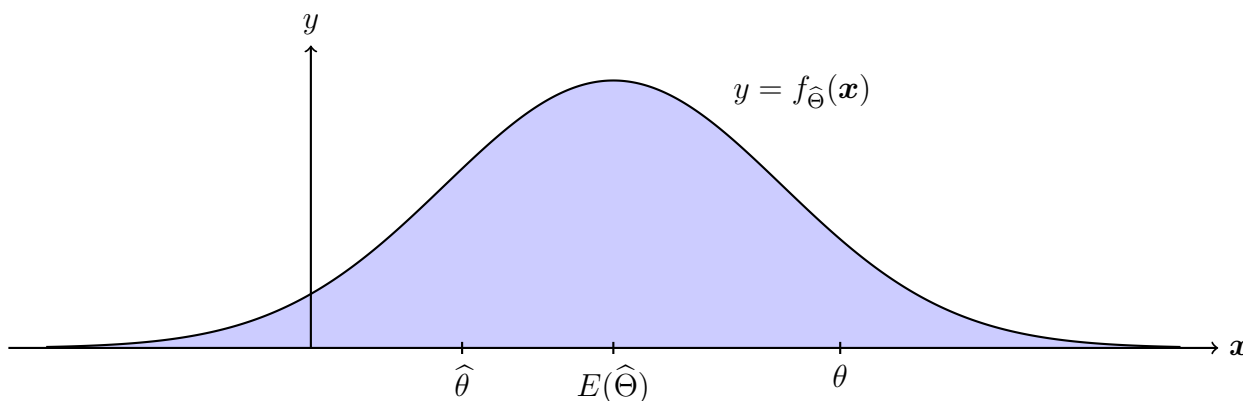


Exempel

Vid fyra dagar på en festival gjordes ljudnivåmätningar vid lunchtid. Följande mätdata erhöles: 107dB, 110dB, 117dB, 101dB. Vi antar att mätningarna är observationer av oberoende och likafördelade variabler med okänt väntevärde μ . Hur hittar vi en skattning $\hat{\mu}$?

- (i) $\hat{\mu} = 100\text{dB}$ är en skattning.
- (ii) $\hat{\mu} = 107\text{dB}$ (den första dagen) är en skattning.
- (iii) $\hat{\mu} = (107 + 110 + 117 + 101)/4 = 108.75\text{dB}$ (medelvärde) är en skattning.
- (iv) $\hat{\mu} = \min\{107, 110, 117, 101\} = 101\text{dB}$ är en skattning.

Så svaret är i princip ”ja,” alla värden $\hat{\theta}$ som är tillåtna i modellen vi betraktar är punktskattningar. Hur väljer vi då den bästa, eller åtminstone en bra, punktskattning? Stickprovsvariabeln $\hat{\Theta}$ är en stokastisk variabel, så normalt sett har den en täthetsfunktion (alternativt sannolikhetsfunktion). Vi skisserar en tänkbar täthetsfunktion för $\hat{\Theta}$ (tänk dock på att $\hat{\Theta}$ typiskt är en flerdimensionell stokastisk variabel då $\hat{\Theta} = g(X_1, X_2, \dots, X_n)$).



Vi vet att $\hat{\theta}$ beräknas från observerade siffror, så $\hat{\theta}$ kan hamna lite var som helst. Dessutom vet vi inte om väntevärdet $E(\hat{\Theta})$ sammanfaller med det okända värdet θ . Så hur kan vi då avgöra om en punktskattning är bra eller inte? Det finns två viktiga kriterier: *väntevärdesriktighet* och *konsistens*. Vi återkommer till dessa nästa föreläsning.



Väntevärdesriktig skattning

Definition. Stickprovsvariabeln $\hat{\Theta}$ kallas **väntevärdesriktig** (vvr) om $E(\hat{\Theta}) = \theta$.

Om en punktskattning inte är väntevärdesriktig pratar man ibland om ett systematiskt fel. Vi definierar detta som skillnaden $E(\hat{\Theta}) - \theta$. En väntevärdesriktig skattning $\hat{\Theta}$ har alltså inget systematiskt fel; i ”medel” kommer den att hamna rätt (tänk på de stora talens lag).



Systematiskt fel; bias

Definition. Om $E(\hat{\Theta}) - \theta \neq 0$ så säger vi att $\hat{\Theta}$ har ett systematiskt fel (ett bias).

6.2.1 Vanliga punktskattningar

Vissa punktskattningar är så vanliga att de ha fått egna namn. Vi vill ofta skatta medelvärdet som positionsmått och stickprovsstandardavvikelsen är ett vanligt mått på spridningen.



Stickprovsmedelvärde

Definition. Stickprovsmedelvärdet $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ är en observation av $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ som skattar väntevärdet.



Stickprovsvarians och stickprovsstandardavvikelse

Definition. Stickprovsvariansen, definierad av $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$, är en observation av $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ som skattar variansen. Stickprovsstandardavvikelsen $s = \sqrt{s^2}$ är en skattning av σ .

Varför $n-1$? Vi återkommer till det nästa föreläsning.

6.3 Vilka skattningar är bra?

När vi har ett stickprov från en fördelning som beror på en okänd parameter så fungerar alltså i princip vad som helst som skattning på parametern. Funktionen g är således godtycklig. Vi vet att $\hat{\theta}$ beräknas från observerade siffror, så $\hat{\theta}$ kan hamna lite var som helst. Dessutom vet vi inte om väntevärdet $E(\hat{\theta})$ sammanfaller med det okända värdet θ . Så hur kan vi då avgöra om en punktskattning är bra eller inte? Det finns två viktiga kriterier: *väntevärdesriktighet* som vi såg ovan och *konsistens*.

Om en punktskattning inte är väntevärdesriktig pratar man ibland om ett systematiskt fel. Vi definierar detta som skillnaden $E(\hat{\theta}) - \theta$. En väntevärdesriktig skattning $\hat{\theta}$ har alltså inget systematiskt fel; i "medel" kommer den att hamna rätt (tänk på de stora talens lag). Vi vill också gärna ha egenskapen att en punktskattning blir bättre ju större stickprov vi använder.



Konsistent skattning

Definition. Antag att vi har en punktskattning $\hat{\theta}_n$ för varje stickprovsstorlek n . Om det för varje $\epsilon > 0$ gäller att

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0,$$

så kallar vi denna punktskattning för *konsistent*.

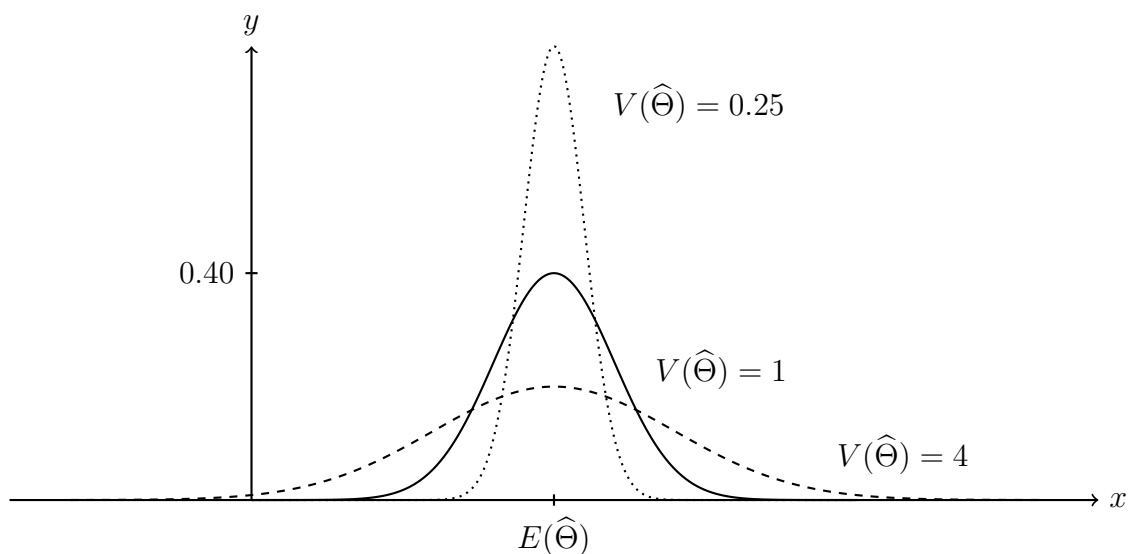
Teknisk definition, men innebörden bör vara klar. När stickprovsstorleken går mot oändligheten så är sannolikheten att skattningen befinner sig nära det okända värdet stor. Villkoret för konsistens kan vara lite jobbigt att arbeta med så följande sats är ofta användbar för att kontrollera konsistens.



Ett kriterium för konsistens

Om $E(\hat{\Theta}_n) = \theta$ för alla n och $\lim_{n \rightarrow \infty} V(\hat{\Theta}_n) = 0$ så är skattningen konsistent.

Bevisskiss: Här använder vi Tjebysjovs olikhet: om $a > 0$ och X är en stokastisk variabel så att $E(X) = \mu$ och $V(X) = \sigma^2 < \infty$, så gäller $P(|X - \mu| > a\sigma) \leq \frac{1}{a^2}$. Om vi låter $a = \epsilon/\sigma_n$ för fixt $\epsilon > 0$ så erhåller vi $P(|\hat{\Theta} - \theta| > \epsilon) \leq \frac{\sigma_n^2}{\epsilon^2} \rightarrow 0$ då $n \rightarrow \infty$, eftersom $\sigma_n^2 = V(\hat{\Theta}_n) \rightarrow 0$ då $n \rightarrow \infty$. \square



Mindre varians för $\hat{\Theta}$ medför att sannolikhetsmassan är mer centrerad kring väntevärdet (vid symmetrisk fördelning).



Exempel

Betrakta exemplet med ljudnivåerna igen, vi hade följande mätdata: 107dB, 110dB, 117dB, 101dB. Vi undersöker skattningarna lite närmare.

- (i) $\hat{\mu} = 100\text{dB}$ är en fix siffra och kan varken vara väntevärdesriktig eller konsistent. Dålig skattning.
- (ii) Den första siffran är en observation av den första variabeln X_1 i stickprovet. Alltså är $\hat{M} = X_1$. Eftersom $E(\hat{M}) = E(X_1) = \mu$ så är skattningen väntevärdesriktig. Med konstant varians oavsett stickprovsstorlek kan den dock inte vara konsistent.
- (iii) Medelvärdet är både väntevärdesriktigt och konsistent; se nästa avsnitt!

När det gäller det sista förslaget så blir det lite klurigare när vi bildar minimum av observationerna. Om vi undersöker ett specialfall där variablerna är exponentialfördelade, säg till exempel att $X_i \sim \text{Exp}(\mu)$, så kan man faktiskt visa att (inte självklart)

$$\hat{M} = \min\{X_1, X_2, X_3, X_4\} \sim \text{Exp}(\mu/4).$$

Således erhåller vi att $E(\widehat{M}) = \mu/4 \neq \mu$. Alltså är inte detta en väntevärdesriktig skattning av μ . Notera att vi här behövde använda fördelningen för variablerna X_i för att kunna svara på frågan dessutom. Går det att korrigera skattningen (så att den blir väntevärdesriktig) i detta fall?

6.3.1 Effektivitet – jämförelse mellan skattningar

Så om vi har två olika stickprovsvariabler $\widehat{\Theta}$ och Θ^* , hur avgör vi vilken som är ”bäst”? Om båda är väntevärdesriktiga och konsistenta, kan man säga att en är bättre?



Effektivitet

Definition. En skattning $\widehat{\Theta}$ kallas *effektivare* än en skattning Θ^* om $V(\widehat{\Theta}) \leq V(\Theta^*)$.

Den stickprovsvariabel med minst varians kallas alltså mer effektiv, och med mindre varians känns det rimligt att kalla den skattningen bättre (om den är någorlunda väntevärdesriktig).

6.4 Momentmetoden

Så kan man systematiskt finna lämpliga skattningar på något sätt om man känner till viss information om fördelningen? Svaret är ja, det finns många sådana metoder. Bland annat momentmetoden, MK-metoden (minsta kvadrat), och kanske den vanligaste, ML-skattningar (maximum likelihood). Vi börjar med att betrakta momentmetoden.



Momentmetoden (för en parameter)

Definition. Låt $E(X_i) = \mu(\theta)$ för alla i . Momentskattningen $\widehat{\theta}$ av θ fås genom att lösa ekvationen $\mu(\widehat{\theta}) = \bar{x}$.



Exempel

Låt x_1, x_2, \dots, x_n vara ett stickprov från en fördelning med täthetsfunktionen $f(x; \theta) = \theta e^{-\theta x}$ för $x \geq 0$. Använd momentmetoden för att punktskatta θ .

Lösning: Vi börjar med att beräkna väntevärdet, det vill säga funktionen $\mu(\theta)$. Alltså,

$$\mu(\theta) = \int_0^\infty x\theta e^{-\theta x} dx = \left[x\theta \frac{e^{-\theta x}}{-\theta} \right]_0^\infty + \int_0^\infty e^{-\theta x} = \theta^{-1}.$$

Vi löser nu ekvationen $\mu(\widehat{\theta}) = \bar{x}$, och erhåller då att

$$\widehat{\theta}^{-1} = \bar{x} \quad \Leftrightarrow \quad \widehat{\theta} = \frac{1}{\bar{x}},$$

så länge $\bar{x} \neq 0$. Momentskattningen av θ ges alltså av $\widehat{\theta} = (\bar{x})^{-1}$. Vad händer om $\bar{x} = 0$?

Om man har flera parametrar då? Här visar det sig varför metoden ovan kallas för just *momentmetoden*.



Moment

Definition. Låt X vara en stokastisk variabel X . För $k = 1, 2, \dots$ definierar vi momenten m_k för X enligt $m_k = E(X^k)$.

Det första momentet m_1 är alltså inget annat än väntevärdet för X .



Momentskattning med flera parametrar

Definition. Låt $X \sim F(x; \theta_1, \theta_2, \dots, \theta_j)$ bero på j okända parametrar $\theta_1, \theta_2, \dots, \theta_j$ och definiera $m_i(\theta_1, \theta_2, \dots, \theta_j) := E(X^i)$, $i = 1, 2, \dots$. Momentskattningarna för θ_k , $k = 1, 2, \dots, j$, ges av lösningen till ekvationssystemet

$$m_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_j) = \frac{1}{n} \sum_{k=1}^n x_k^i, \quad i = 1, 2, \dots, j.$$

Observera att det inte är säkert att en lösning finns eller att lösningen är entydig i de fall den existerar. Vidare kan det även inträffa att lösningen hamnar utanför det område som är tillåtet för parametern (i vilket fall vi givetvis inte kan använda den).



Exempel

Låt $X_k \sim N(\mu, \sigma^2)$, $k = 1, 2, \dots, n$ vara ett stickprov. Hitta momentskattningarna för μ och σ^2 .

Lösning. Vi vet att $E(X) = \mu$ och $E(X^2) = V(X) + E(X)^2 = \sigma^2 + \mu^2$, så

$$\begin{cases} \hat{\mu} = \bar{x}, \\ \hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{n} \sum_{k=1}^n x_k^2. \end{cases}$$

Således erhåller vi direkt att $\hat{\mu} = \bar{x}$. För $\hat{\sigma}^2$ är

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2 = \dots = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Nästan stickprovsvariansen alltså.



Vektornotation för parametrar

Definition. När vi har en fördelning som beror på flera parametrar, säg $\theta_1, \theta_2, \dots, \theta_j$, så skriver vi ibland $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_j) \in \mathbf{R}^j$ som en j -dimensionell vektor. Notationen blir då mer kompakt. Bokstäver typsatta i fet stil indikerar oftast en vektor i denna kurs.

Kapitel 7

Punktskattningar och konfidensintervall

7.1 Vanliga punktskattningar

Vi stötte på medelvärdet och stickprovsvariansen under föregående föreläsning. Dessa skattningar är vettiga skattningar av väntevärdet och variansen i meningen att de är väntevärdesriktiga och konsistenta.



Medelvärde

Medelvärdet $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ är en väntevärdesriktig och konsistent skattning av väntevärdet.

Bevis: Variablerna X_k är oberoende och likafördelade. Låt $E(X_i) = \mu$ och $V(X_i) = \sigma^2$ för alla i . Eftersom väntevärdesoperatoren är linjär så gäller att

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{n\mu}{n} = \mu.$$

Alltså är \bar{X} en väntevärdesriktig skattning av μ .

Då variablerna är oberoende kan vi göra en liknande kalkyl för variansen:

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Här ser vi att $V(\bar{X}) \rightarrow 0$ då $n \rightarrow \infty$, så enligt satsen från föregående föreläsning är skattningen konsistent.



Stickprovsvarians

Stickprovsvariansen $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ är en väntevärdesriktig skattning av variansen.

Bevis: Detta bevis är lite bökgigare, men följer samma princip.

$$\begin{aligned} E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) &= \frac{1}{n-1} E\left(\sum_{i=1}^n X_i^2 - 2X_i\bar{X} + \bar{X}^2\right) \\ &= \frac{1}{n-1} \sum_{i=1}^n (E(X_i^2) - 2E(X_i\bar{X}) + E(\bar{X}^2)). \end{aligned}$$

Vi vet att $E(\bar{X}) = \mu$ och att $V(\bar{X}) = \sigma^2/n$. Steiners formel säger att $E(Y^2) = V(Y) + E(Y)^2$ för en stokastisk variabel Y , vilket vi kan utnyttja för att skriva

$$E(X_i^2) = V(X_i) + E(X_i)^2 = \sigma^2 + \mu^2 \quad \text{samtidigt} \quad E(\bar{X}^2) = \sigma^2/n + \mu^2.$$

Vidare så ser vi att

$$E(X_i\bar{X}) = E\left(X_i \frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_i X_k)$$

och eftersom $E(X_i X_k) = E(X_i)E(X_k) = \mu^2$ om $i \neq k$ (eftersom dessa variabler är oberoende) och $E(X_i^2) = \sigma^2 + \mu^2$ (då $i = k$) kan vi skriva

$$E(X_i\bar{X}) = ((n-1)\mu^2 + \sigma^2 + \mu^2)/n = \mu^2 + \sigma^2/n.$$

Vi återgår till det sökta väntevärdet:

$$E(S^2) = \frac{1}{n-1} \sum_{i=1}^n (\sigma^2 + \mu^2 - 2(\mu^2 + \sigma^2/n) + \sigma^2/n + \mu^2) = \frac{n\sigma^2 - n\sigma^2/n}{n-1} = \sigma^2.$$

Alltså är S^2 en väntevärdesriktig skattning (av σ^2). Värt att notera är att $S = \sqrt{S^2}$ **inte** är en väntevärdesriktig skattning av σ (men den används oftast ändå!).

7.2 Metoder för att hitta punktskattningar

Vi har slarvat lite i definitionen av punktskattningar när det gäller vilka värden på den okända parametern θ som är tillåtna. Vi inför begreppet **parameterrum**.



Parameterrum

Definition. Vi låter Ω_θ beteckna **parameterrummet** av alla tillåtna värden på parametern θ .

Parameterrummet är alltså en delmängd av \mathbf{R}^p där p är antalet parametrar (tänk på att θ kan vara en vektor $\theta = (\theta_1, \theta_2, \dots, \theta_p)$).



Exempel

- (i) Om $X \sim N(\mu, \sigma^2)$ kan vi tänka oss $\theta = (\mu, \sigma^2)$, i vilket fall parameterrummet kan representeras som $\mathbf{R} \times (0, \infty)$.
- (ii) Om $X \sim \text{Bin}(n, p)$ där n är fixerad är parameterrummet $\Omega_p = [0, 1]$.

Skulle vi med någon metod hitta en skattning som faller utanför parameterrummet måste den förkastas. Så åter till frågan hur vi hittar skattningar mer systematiskt.

7.2.1 Momentmetoden

Vi såg momentmetoden i förra föreläsningen. Låt oss endast repetera vad den gick ut på.



Momentskattning med flera parametrar

Definition. Låt $X \sim F(x; \theta_1, \theta_2, \dots, \theta_j)$ bero på j okända parametrar $\theta_1, \theta_2, \dots, \theta_j$ och definiera $m_i(\theta_1, \theta_2, \dots, \theta_j) := E(X^i)$, $i = 1, 2, \dots$. Momentskattningarna för θ_k , $k = 1, 2, \dots, j$, ges av lösningen till ekvationssystemet

$$m_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_j) = \frac{1}{n} \sum_{k=1}^n x_k^i, \quad i = 1, 2, \dots, j.$$

7.2.2 MK-skattning

Minsta kvadrat-metoden har vi egentligen stött på i tidigare kurser, mer specifikt när vi hittade approximativa lösningar till överbestämda ekvationssystem. Faktum är att vi kommer att upprepa den proceduren senare i denna kurs i samband med linjär regression.

Låt x_1, x_2, \dots, x_n vara observationer av oberoende stokastiska variabler X_1, X_2, \dots, X_n sådana att $E(X_k) = \mu_k(\theta)$ och $V(X_k) = \sigma^2$ för $k = 1, 2, \dots, n$ (alltså samma varians men potentiellt olika väntevärden).



Minsta kvadrat-skattning

Definition. Minsta kvadrat-skattningen för θ ges av den vektor $\hat{\theta}$ som minimerar

$$Q(\hat{\theta}) = \sum_{k=1}^n \left(x_k - \mu_k(\hat{\theta}) \right)^2.$$



Exempel

Låt X_1, \dots, X_n vara ett slumpmässigt stickprov från en fördelning F . Hitta MK-skattningen för väntevärdet μ .

Lösning. Vi ställer upp funktionen

$$Q(\mu) = \sum_{k=1}^n (x_k - \mu)^2, \mu \in \mathbf{R}.$$

Vi söker nu det värde $\hat{\mu}$ som minimerar Q . Enklast är att ta till envariabelanalysen och derivera och söka efter stationära punkter:

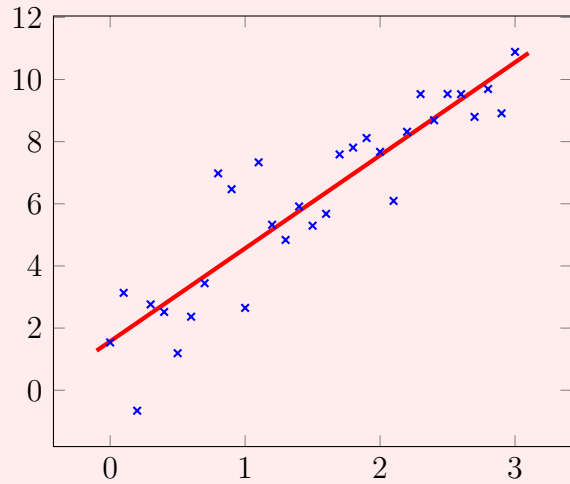
$$0 = Q'(\mu) = -2 \sum_{k=1}^n (x_k - \mu) \Leftrightarrow n\mu = \sum_{k=1}^n x_k \Leftrightarrow \mu = \frac{1}{n} \sum_{k=1}^n x_k = \bar{x}.$$

Är detta ett minimum? Eftersom $Q''(\bar{x}) = 2n > 0$ är det mycket riktigt ett minimum. Den eftersökta MK-skattningen av väntevärdet är alltså $\hat{\mu} = \bar{x}$.



Enkel linjär regression

Antag att vi gjort mätningar y_k på något vid vissa värden x_k , $k = 1, 2, \dots, n$ och att ett spridningsdiagram visar något i stil med figuren till höger. Det förefaller rimligt att det föreligger ett approximativt linjärt samband. Kan vi hitta en linje som passar in i mätserien? Vi söker alltså en linje $y = \beta_0 + \beta_1 x$ som i någon mening approximerar mätresultaten. I vilken mening? Där finns flera sätt, men det vanligaste är nog att minimera kvadraten i felen.



Lösning. Vi betraktar varje punkt (x_k, y_k) som att x_k är fixerad och y_k är en observation av en stokastisk variabel $Y = \beta_0 + \beta_1 x_k + \epsilon_k$ där ϵ_k är oberoende stokastiska variabler med $E(\epsilon_k) = 0$ och $V(\epsilon_k) = \sigma^2$. Detta är den typiska modellen vid linjär regression. Konstanterna β_0 och β_1 är okända och det är dessa vi vill bestämma. Eftersom

$$E(Y_k) = \beta_0 + \beta_1 x_k \quad \text{och} \quad V(Y_k) = \sigma^2$$

så blir

$$Q(\beta_0, \beta_1) = \sum_{k=1}^n (y_k - E(Y_k))^2 = \sum_{k=1}^n (y_k - \beta_0 - \beta_1 x_k)^2.$$

Minimering av denna funktion med avseende på β_0 och β_1 ger skattningarna $\hat{\beta}_0$ och $\hat{\beta}_1$. Jakten på minimum sker nog enklast med lite flervariabelanalys:

$$\mathbf{0} = \nabla Q = (Q'_{\beta_0}, Q'_{\beta_1}) = -2 \sum_{j=1}^n (y_j - \beta_0 - \beta_1 x_j, x_j(y_j - \beta_0 - \beta_1 x_j))$$

så

$$n\beta_0 + \beta_1 \sum_{j=1}^n x_j = \sum_{j=1}^n y_j \quad \Leftrightarrow \quad \beta_0 + \beta_1 \bar{x} = \bar{y}$$

och

$$\beta_0 \sum_{j=1}^n x_j + \beta_1 \sum_{j=1}^n x_j^2 = \sum_{j=1}^n x_j y_j \quad \Leftrightarrow \quad n\beta_0 \bar{x} + \beta_1 \sum_{j=1}^n x_j^2 = \sum_{j=1}^n x_j y_j.$$

Första ekvationen ger att $\beta_0 = \bar{y} - \beta_1 \bar{x}$, så

$$n\bar{x}\bar{y} - \beta_1 n\bar{x}^2 + \beta_1 \sum_{j=1}^n x_j^2 = \sum_{j=1}^n x_j y_j$$

vilket om vi löser ut β_1 leder till

$$\beta_1 = \frac{\sum_{j=1}^n x_j y_j - n\bar{x}\bar{y}}{\sum_{j=1}^n x_j^2 - n\bar{x}^2} = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}.$$

7.2.3 ML-skattning

Låt X_1, X_2, \dots, X_n vara oberoende stokastiska variabler med täthets- eller sannolikhetsfunktioner $f_i(x; \theta)$ respektive $p_i(k; \theta)$. Vi antar att samtliga endera är kontinuerliga eller diskreta. Det typiska är att alla variablerna har samma fördelning, men det är inget nödvändigt krav för metoden (däremot förenklar det så klart). Samtliga fördelningar beror dock på en och samma parameter θ som kan vara vektorvärd.



ML-skattning

Definition. ML-skattningen för θ är det värde som gör att **likelihood-funktionen** $L(\theta)$ maximeras, där

$$L(\theta) = \prod_{k=1}^n f_k(x_k; \theta) = f_1(x_1; \theta) \cdot f_2(x_2; \theta) \cdots f_n(x_n; \theta)$$

i det kontinuerliga fallet och

$$L(\theta) = \prod_{k=1}^n p_k(x_k; \theta) = p_1(x_1; \theta) \cdot p_2(x_2; \theta) \cdots p_n(x_n; \theta)$$

i det diskreta fallet.

Så vad är då ML-skattningen? Ganska enkelt är det den skattning som gör att det stickprov vi observerat är det mest troliga. Eftersom vi antar att variablerna som stickprovet är observationer av är oberoende ges den simultana täthets- eller sannolikhetsfunktionen av produkten av de marginella, så vi väljer helt enkelt den skattning som maximerar den simultana tätheten/sannolikheten.

Ofta när man arbetar med ML-skattningar nyttjar man den så kallade log-likelihood-funktionen:

$$l(\theta) = \ln L(\theta).$$

Denna funktion bevarar de flesta av de egenskaper vi är intresserade av eftersom \ln är strängt växande och $L(\theta) \in [0, 1]$. Specifikt så har $L(\theta)$ och $l(\theta)$ samma extrempunkter.



Exempel

Låt x_1, x_2, \dots, x_n vara ett stickprov av en exponentialfördelning med okänd intensitet θ . Hitta ML-skattningen för θ .

Lösning. Täthetsfunktionen ges av $f(x) = \theta e^{-\theta x}$, $x \geq 0$, så

$$L(\theta) = \prod_{k=1}^n \theta e^{-\theta x_k} = \theta^n \exp\left(-\theta \sum_{k=1}^n x_k\right) \Rightarrow l(\theta) = \ln L(\theta) = n \ln \theta - \theta \sum_{k=1}^n x_k.$$

Vi undersöker var det finns extrempunkter och finner att

$$0 = l'(\theta) = \frac{n}{\theta} - \sum_{k=1}^n x_k \Leftrightarrow \theta = \frac{n}{\sum_{k=1}^n x_k} = \frac{1}{\bar{x}},$$

under förutsättning att $\bar{x} \neq 0$. Är detta ett maximum? Använd det ni lärt er i envariabelanalysen! Till exempel ser vi att

$$l''(\theta) = -\frac{n}{\theta^2},$$

så $l''(\theta) < 0$ för alla $\theta > 0$. Således är det ett maximum vi funnit.



Exempel

Låt $X \sim \text{Bin}(n, p)$ med p okänd och låt x vara en observation av X . Hitta ML-skattningen för p .

Lösning. Sannolikhetsfunktionen ges av $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$, så

$$\begin{aligned} L(p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ \Rightarrow l(p) &= C(n, x) + x \ln p + (n-x) \ln(1-p), \end{aligned}$$

där $C(n, x)$ är en konstant (med avseende på p). Parameterrummet ges av $\Omega_p = (0, 1)$. Vi deriverar och erhåller att

$$0 = l'(p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{(1-p)x - (n-x)p}{p(1-p)} = \frac{x - np}{p(1-p)} \Leftrightarrow x = np \Leftrightarrow p = \frac{x}{n}.$$

ML-skattningen är således $\hat{p} = \frac{x}{n}$ om detta är ett maximum. Vi kontrollerar:

	\hat{p}		
$l'(p)$	+	0	-
$l(p)$	\nearrow	max	\searrow

Vad skulle hända om observationen blev $x = 0$ (eller $x = n$)?

7.3 Flera stickprov; sammanvägd variansskattning

Antag att vi har två stickprov x_1, x_2, \dots, x_m och y_1, y_2, \dots, y_n från normalfördelningar med olika väntevärde men samma varians. ML-skattningarna för respektive väntevärde blir $\hat{\mu}_1 = \bar{x}$ respektive $\hat{\mu}_2 = \bar{y}$. För standardavvikelsen kan man visa att den **sammanvägda variansskattningen** (*pooled variance*) blir

$$s^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{n+m-2},$$

där s_1^2 och s_2^2 är stickprovsvarianserna för respektive stickprov. Formeln generaliserar naturligt till fler stickprov. Vi kan även direkt se att

$$E(S^2) = \frac{1}{m+n-2} ((m-1)E(S_1^2) + (n-1)E(S_2^2)) = \frac{1}{m+n-2} ((m+n-2)\sigma^2) = \sigma^2,$$

så skattningen är väntevärdesriktig.

7.4 Medelfel

Vi har använt variansen $V(\hat{\Theta})$ (eller standardavvikelsen $D(\hat{\Theta})$) för att jämföra olika skattningar (effektivitet och konsistens). Mindre varians betyder helt enkelt att skattningen i någon mening är bättre. Detta är ett problem då dessa storheter i allmänhet inte är kända. Vad vi gör är att vi helt enkelt skattar de okända storheterna i $D(\hat{\Theta})$ och kallar resultatet för medelfelet.



Medelfel

Definition. En skattning $d = d(\hat{\Theta})$ av standardavvikelsen $D(\hat{\Theta})$ kallas för skattningens **medelfel**.

Vi ersätter alltså helt enkelt okända storheter i $V(\hat{\Theta})$ med skattningar. Givetvis påverkar detta precisionen och sättet vi väljer att ersätta de okända storheterna har inverkan på resultatet.



Exempel

Om X_1, \dots, X_n är ett slumpmässigt stickprov av en $N(\mu, \sigma^2)$ -fördelning där både μ och σ^2 är okända kan vi uppskatta μ med medelvärdet $\hat{M} = \bar{X}$. Således är $D(M) = \frac{\sigma}{\sqrt{n}}$, men då σ är okänd behöver vi skatta σ med något. Förslagsvis med stickprovsstandardavvikelsen s , vilket ger medelfelet

$$d(\hat{M}) = \frac{s}{\sqrt{n}}.$$

Detta är inte på något sätt unikt. En annan skattning av σ ger ett annat medelfel. Med det sagt är detta ett ganska naturligt val för medelfelet.

Ett annat vanligt exempel är vid skattningar av andel. Ofta gör vi som i följande exempel.



Exempel

Ett annat vanligt exempel är när p ska skattas i binomialfördelning. Låt $X \sim \text{Bin}(n, p)$. Vi vet att $V(X) = np(1-p)$ så om vi skattar p med $\hat{P} = \frac{X}{n}$ erhåller vi att $D(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}$. Eftersom p är okänd känner vi inte denna storhet exakt, men medelfelet skulle bli

$$d(\hat{P}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

7.5 Intervallskattningar

Vi har nu studerat hur man mer eller mindre systematiskt kan hitta skattningar för okända parametrar när vi har stickprov från en fördelning som beror på parametern. Den naturliga följdfrågan är givetvis hur ”bra” skattningen är. Vi har vissa mått i form av väntevärdesriktighet, konsistens och effektivitet, men går det att säga något med en given sannolikhet? Kan vi hitta ett intervall som med en viss given sannolikhet måste innehålla den okända parametern?



Konfidensintervall

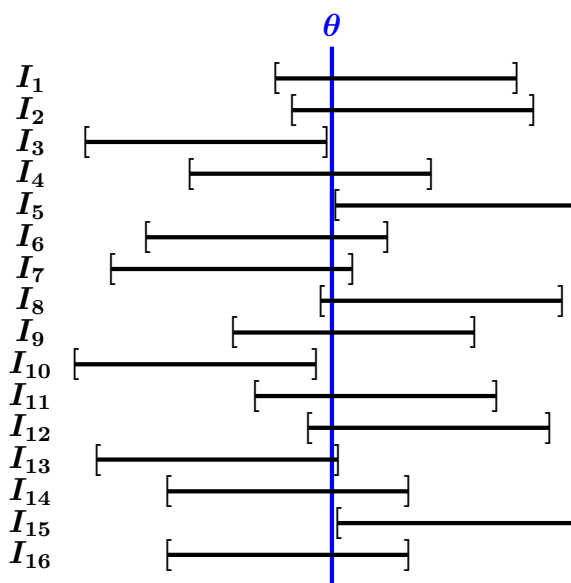
Definition. Låt x_1, \dots, x_n vara ett stickprov av en fördelning som beror på en okänd parameter θ och låt $\alpha \in [0, 1]$. Ett intervall $I_\theta^{1-\alpha} = (\hat{\theta}_L, \hat{\theta}_U)$ kallas för ett **konfidensintervall** för θ med **konfidensgrad** $1 - \alpha$ om

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha.$$

Gränserna $\hat{\theta}_L = a(x_1, \dots, x_n)$ och $\hat{\theta}_U = b(x_1, \dots, x_n)$ är skattningar som beräknas från stickprovet. Dessa ändpunkter kallas **konfidensgränser**.

Så hur fungerar detta i praktiken? Säg att vi har tillgång till 100 olika stickprov $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)})$ från en och samma fördelning som beror på samma okända parameter θ . Vi hittar konfidensintervall för alla 100 stickproven med konfidensgrad $1 - \alpha$. Då kommer $100 \cdot (1 - \alpha)$ av dessa intervall att innehålla θ (i snitt).

Av de 16 intervall till höger är det 4 som inte innehåller det verkliga värdet på θ . Så med andra ord verkar det som att ungefär $12/16 = 3/4$ av intervallen innehåller det verkliga värdet på θ . Detta innebär att konfidensgraden vid skattningen kanske är ungefär 75%.



Notera att det är gränserna i konfidensintervallet som är stokastiska variabler (eller skattningar därav). Storheten θ är okänd (och behöver inte ens ligga i intervallet).



Inga intervall är mer värda

Vi kan inte säga att till exempel I_9 är ett "bättre" intervall än I_{13} , utan det är en binär fråga: gäller det att $\theta \in I_k$ eller inte.

Något som kommer bli viktigt är följande definition från sannolikheteorin.



Kvantil

Definition. En α -kvantil λ_α för en stokastisk variabel X är ett tal λ_α sådant att

$$P(X > \lambda_\alpha) = \alpha.$$

Vi finner ofta kvantiler i tabell endera genom en explicit kvantiltabell eller genom att söka upp sannolikheten $1 - \alpha$ och identifiera (approximativt) vilket värde på x som gör att vi erhåller $F(x) = 1 - \alpha$, där F är fördelningsfunktionen. Saknar vi tabell får vi istället lösa ekvationen

$$1 - \alpha = \int_{-\infty}^{\lambda_\alpha} f_X(x) dx.$$

Observera att svaret inte nödvändigtvis är entydigt.

Så då kommer vi till nästa rimliga fråga: hur hittar vi systematiskt konfidensintervall med given konfidensgrad?



Konstruktion av konfidensintervall

1. Ställ upp en lämplig skattningsvariabel $\hat{\Theta}$ för θ . Här kan vi använda de metoder vi tagit fram tidigare (moment-, MK- och ML-skattningar till exempel).
2. Konstruera en hjälpvariabel H (teststorhet) utifrån $\hat{\Theta}$. Hjälpvariabeln får endast innehålla kända storheter utöver θ (och om θ förekommer flera gånger kan vi behöva skatta bort en del instanser för att få något användbart).
3. Stäng in hjälpvariabeln i ett intervall $I = (c, d)$ så att $P(c < H < d) = 1 - \alpha$.
4. Lös ut θ ur olikheten $c < H < d$:

$$c < H < d \Leftrightarrow a(X_1, \dots, X_n) < \theta < b(X_1, \dots, X_n)$$

vilket ger att $P(a(X_1, \dots, X_n) < \theta < b(X_1, \dots, X_n)) = 1 - \alpha$.

5. Ersätt de stokastiska storheterna X_1, \dots, X_n med observationerna x_1, \dots, x_n vilket ger intervallet

$$I_\theta^{1-\alpha} = (a(x_1, \dots, x_n), b(x_1, \dots, x_n)).$$

7.6 Konfidensintervall för μ vid normalfördelning (σ känd)

Låt x_1, x_2, \dots, x_n vara ett stickprov från en $N(\mu, \sigma^2)$ -fördelning där vi känner σ och vill hitta ett konfidensintervall för μ . En punktskattning för väntevärdet ges av

$$\widehat{M} = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \sim N(\mu, \sigma^2/n).$$

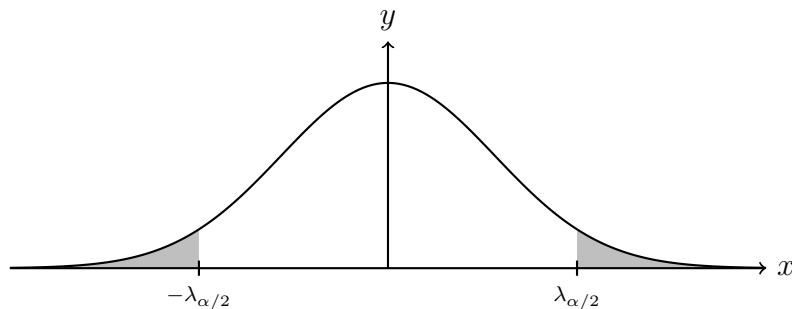
Vi skapar testvariabeln

$$Z = \frac{\widehat{M} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Det följer då att vi kan välja ett tal $\lambda_{\alpha/2}$ så att

$$P(-\lambda_{\alpha/2} < Z < \lambda_{\alpha/2}) = 1 - \alpha. \quad (7.1)$$

Talet $\lambda_{\alpha/2}$ är $\alpha/2$ -kvantilen för en $N(0, 1)$ -fördelning och ges av $\lambda_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. Eftersom vi saknar explicit uttryck för denna invers är det enklast (utan dator åtminstone) att slå i tabell. Standardtabell som finns i formelsamlingen enligt nedan (vi får utnyttja symmetri för att finna sannolikheter mindre än 0.5).



Det skuggade områdena är sannolikheten att $P(Z < -\lambda_{\alpha/2}) + P(Z > \lambda_{\alpha/2})$.

Vi löser ut μ ur olikheten i sannolikhetsmåttet i ekvation (7.1) ovan:

$$\begin{aligned} -\lambda_{\alpha/2} < Z < \lambda_{\alpha/2} &\Leftrightarrow -\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \widehat{M} - \mu < \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ &\Leftrightarrow \widehat{M} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \widehat{M} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Om vi ersätter \widehat{M} med den observerade punktskattningen $\hat{\mu} = \bar{x}$ (medelvärde av observationerna) så får vi ett konfidensintervall

$$I_{\mu} = \left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

med konfidensgrad $1 - \alpha$.



Exempel

Vid en mätning av en process fick man följande mätdata:

6.04 4.96 4.93 3.40 7.04 4.73 3.57 7.70 4.55 3.82

Antag att mätningarna är ett stickprov på en normalfördelad variabel $X \sim N(\mu, \sigma = 2)$ (man tycker sig veta så pass mycket om processen att standardavvikelsen anses vara känd). Beräkna ett 99% konfidensintervall för väntevärdet μ .

Lösning. Vi betraktar siffrorna som ett stickprov från oberoende s. v. $X_j \sim N(\mu, \sigma = 2)$. Vi punktskattar väntevärdet μ med

$$\widehat{M} = \bar{X} = \frac{1}{10} \sum_{j=1}^{10} X_j \sim N(\mu, 2/\sqrt{10}).$$

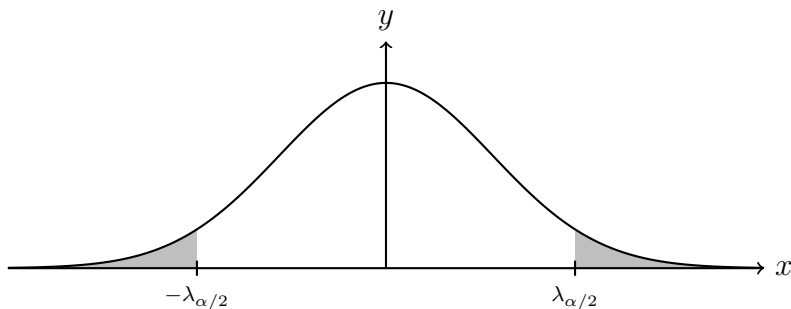
Vi skapar testvariabeln

$$Z = \frac{\widehat{M} - \mu}{\sigma/\sqrt{10}} \sim N(0, 1).$$

Det följer då att

$$P(-\lambda_{\alpha/2} < Z < \lambda_{\alpha/2}) = 1 - \alpha, \quad (7.2)$$

och då vi söker ett 99% konfidensintervall så är $\alpha = 0.01$ och $\lambda_{0.005} \approx 2.575$ (det sista ur tabell).



Det skuggade området är $\alpha \cdot 100\%$ av sannolikhetsmassan jämt fördelad på svansarna. Vi löser ut μ ur olikheten i sannolikhetsmåttet i ekvation (7.2) ovan och erhåller att

$$\widehat{M} - \frac{2.575 \cdot 2}{\sqrt{10}} < \mu < \widehat{M} + \frac{2.575 \cdot 2}{\sqrt{10}}.$$

Om vi ersätter \widehat{M} med den observerade punktskattningen $\widehat{\mu} = 5.074$ (medelvärdet av observationerna) så får vi ett konfidensintervall $I_\mu = (3.45, 6.70)$ med konfidensgrad 99%.

Så vad gör man om σ inte är känd? Ja då skattar man helt enkelt σ , men detta påverkar fördelningen. Vi återkommer till detta!

7.7 (★★)ML-skattning för normalfördelning



Exempel

Låt x_1, x_2, \dots, x_n vara ett stickprov från $N(\mu, \sigma^2)$ där både μ och σ^2 är okända. Hitta ML-skattningarna för μ och σ^2 .

Lösning. Vi har nu två okända parametrar och likelihoodfunktionen ges av

$$L(\mu, v) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x_k - \mu)^2}{2v}\right) = \frac{1}{(2\pi v)^{n/2}} \exp\left(-\frac{1}{2v} \sum_{k=1}^n (x_k - \mu)^2\right),$$

där $v = \sigma^2$, så

$$l(\mu, v) = \text{konstant} - \frac{n}{2} \ln v - \frac{1}{2v} \sum_{k=1}^n (x_k - \mu)^2.$$

Parameterrummet ges av $\Omega_{\mu, v} = \mathbf{R} \times (0, \infty)$ och vi vill maximera $l(\mu, v)$. Stationära punkter finner vi där $\nabla l(\mu, v) = (0, 0)$, så vi beräknar de partiella derivatorerna:

$$l'_\mu(\mu, v) = \frac{1}{v} \sum_{k=1}^n (x_k - \mu) = \frac{n}{v} (\bar{x} - \mu)$$

och

$$l'_v(\mu, v) = -\frac{n}{2v} + \frac{1}{2v^2} \sum_{k=1}^n (x_k - \mu)^2.$$

Det är tydligt att $\mu = \bar{x}$ och

$$\frac{n}{2v} = \frac{1}{2v^2} \sum_{k=1}^n (x_k - \mu)^2 \quad \Leftrightarrow \quad v = \frac{1}{n} \sum_{k=1}^n (x_k - \mu)^2,$$

så $\nabla l = 0$ precis då

$$\mu = \bar{x} \quad \text{och} \quad v = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Är detta ett maximum? Vi undersöker närmare:

$$H(\mu, v) = \begin{pmatrix} l''_{\mu\mu} & l''_{\mu v} \\ l''_{v\mu} & l''_{vv} \end{pmatrix} = \begin{pmatrix} -\frac{n}{v} & -\frac{n}{v^2} (\bar{x} - \mu) \\ -\frac{n}{v^2} (\bar{x} - \mu) & \frac{n}{2v^2} - \frac{1}{v^3} \sum_{k=1}^n (x_k - \mu)^2 \end{pmatrix},$$

där vi låter $SS = \sum_{k=1}^n (x_k - \mu)^2$ och i punkten $(\mu, v) = \left(\bar{x}, \frac{1}{n} SS\right)$ blir

$$H\left(\bar{x}, \frac{1}{n} SS\right) = \begin{pmatrix} -\frac{n^2}{SS} & 0 \\ 0 & \frac{n^3}{2SS^2} - \frac{n^3}{SS^3} SS \end{pmatrix} = \begin{pmatrix} -\frac{n^2}{SS} & 0 \\ 0 & -\frac{n^3}{2SS^2} \end{pmatrix},$$

vilket är en negativt definit matris, så detta är ett maximum.

Vi vet sedan tidigare att skattningen för v behöver ha faktorn $1/(n-1)$ för att vara väntevärdesriktig, så ML-skattningen av σ^2 är således inte väntevärdesriktig.

Kapitel 8

Konfidsensintervall (forts)

8.1 χ^2 - och t -fördelning

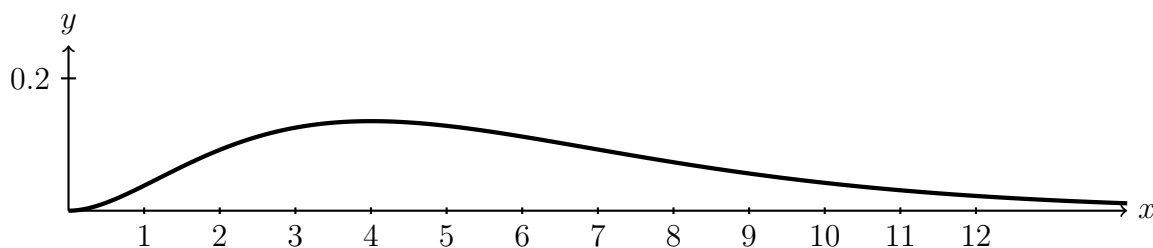
För att hantera situationen när vi inte känner till variansen för den normalfördelning som vi tagit stickprovet ifrån så måste vi introducera ett par nya fördelningar. Först och främst formulerar vi följande sats där χ^2 -fördelningen finns i mer detaljer i avsnitt 8.11. Fördelningen uppstår alltså när vi summerar kvadrater av oberoende normalfördelningar.



Sats. Om X_1, X_2, \dots, X_n är oberoende och $X_k \sim N(0, 1)$ så är

$$\sum_{k=1}^n X_k^2 \sim \chi^2(n).$$

Vi kan skissa ett exempel på hur täthetsfunktionen ser ut (här med $n = 6$). Utseendet beror givetvis på parametern n (som brukar kallas för antalet frihetsgrader).



Så varför är vi intresserade av denna kvadratsumma? Tänk på hur vi definierat stickprovsvariansen. Man kan visa (se avsnitt 8.14) följande sats.



Sats. Låt X_1, X_2, \dots, X_n vara oberoende likafördelade stokastiska variabler där $X_k \sim N(\mu, \sigma)$ för $k = 1, 2, \dots, n$. Då gäller att

$$\frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \bar{X})^2 \sim \chi^2(n-1).$$

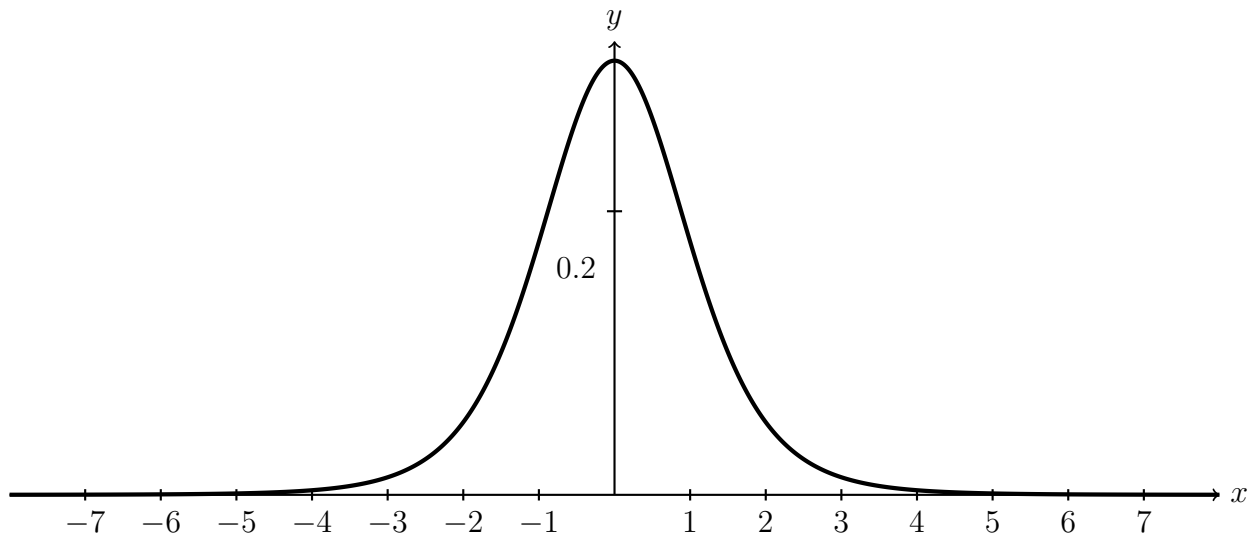
Anledningen till att vi får $n - 1$ trots att det är n termer i summan beror på att medelvärdet \bar{X} inte är oberoende av X_1, X_2, \dots, X_n . Det faktum att det blir just $n - 1$ är inte självklart (se avsnitt 8.14).

Med dessa resultat på plats så introducerar vi en till ny fördelning: t -fördelningen. Vi nöjer oss med att karaktärisera den som den fördelning som uppstår om kvoten $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ bildas. Se avsnitt 8.12 för mer detaljer.



Sats. Om X_1, X_2, \dots, X_n är oberoende och likafördelade stokastiska variabler med fördelningen $N(\mu, \sigma)$ så är $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1)$, där S^2 är stickprovsvariansen.

Vi kan även skissa ett exempel på hur täthetsfunktionen ser ut för t -fördelningen ($n = 6$). Likt χ^2 -fördelningen så beror utseendet givetvis på parametern n (som även här brukar kallas för antalet frihetsgrader).



Tätheten är symmetrisk kring origo och påminner en hel del om normalfördelning. Faktum är att om $n \rightarrow \infty$ så får vi tillbaka normalfördelningen.

8.2 Konfidensintervall för μ i normalfördelning

8.2.1 Konfidensintervall för μ när σ är känd

Vi tog fram detta resultat förra föreläsningen så låt oss bara kortfattat beskriva vad vi gjorde. Låt x_1, x_2, \dots, x_n vara ett stickprov från en $N(\mu, \sigma)$ -fördelning där vi känner σ och vill hitta ett konfidensintervall för μ . En punktskattning för väntevärdet ges av

$$\widehat{M} = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \sim N(\mu, \sigma/\sqrt{n}).$$

Vi skapar testvariabeln

$$Z = \frac{\widehat{M} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Det följer då att vi kan välja ett tal $\lambda_{\alpha/2}$ så att $P(-\lambda_{\alpha/2} < Z < \lambda_{\alpha/2}) = 1 - \alpha$. Om man löser ut μ ur olikheten i sannolikhetsmåttet finner vi till slut intervallet, med konfidensgrad $1 - \alpha$,

$$I_\mu = \left(\bar{x} - \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right),$$

där vi ersatt \widehat{M} med den observerade punktskattningen $\widehat{\mu} = \bar{x}$ (medelvärde av observationerna) för att få ett konfidensintervall; se föregående föreläsning för detaljerna.

8.2.2 Okänd varians

Låt x_1, x_2, \dots, x_n vara ett stickprov från en $N(\mu, \sigma)$ -fördelning där vi *inte* vet vad σ är och vi vill hitta ett konfidensintervall för μ . En punktskattning för väntevärdet ges av

$$\widehat{M} = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \sim N(\mu, \sigma/\sqrt{n}).$$

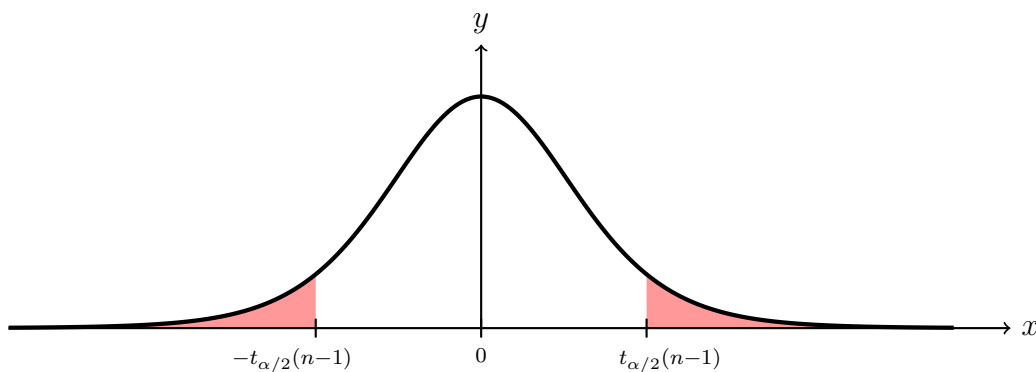
Eftersom σ är okänd behöver vi en skattning och förslagsvis väljer vi stickprovsstandardavvikelsen. Vi skapar sedan testvariabeln

$$T = \frac{\widehat{M} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

där faktumet att T är t -fördelad följer från Gossets sats. Det följer då att vi kan välja ett tal $t_{\alpha/2}$ så att

$$P(-t_{\alpha/2}(n-1) < T < t_{\alpha/2}(n-1)) = 1 - \alpha. \quad (8.1)$$

Talet $t_{\alpha/2}(n-1)$ är $\alpha/2$ -kvantilen för en $t(n-1)$ -fördelning (vi finner denna i tabell).



Vi löser ut μ ur olikheten i sannolikhetsmåttet i ekvation (8.1) ovan:

$$\begin{aligned} -t_{\alpha/2}(n-1) < T < t_{\alpha/2}(n-1) &\Leftrightarrow -t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} < \widehat{M} - \mu < t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \\ &\Leftrightarrow \widehat{M} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} < \mu < \widehat{M} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \end{aligned}$$

Om vi ersätter \widehat{M} med den observerade punktskattningen $\widehat{\mu} = \bar{x}$ (medelvärde av observationerna) och S med stickprovsstandardavvikelsen så får vi ett konfidensintervall

$$I_\mu = \left(\bar{x} - t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}} \right)$$

med konfidensgrad $1 - \alpha$.

**Exempel**

Samma exempel som tidigare där man vid en mätning av en process fick man följande mätdata:

6.04 4.96 4.93 3.40 7.04 4.73 3.57 7.70 4.55 3.82

En som arbetar med processen håller inte med om att standardavvikelsen kan antas vara given, utan tycker att man måste skatta den utifrån datan. Hjälp personen i fråga med att ställa upp ett 99% konfidsensintervall för väntevärdet μ då mätningarna är ett stickprov på en normalfördelad variabel $X \sim N(\mu, \sigma)$ och σ är okänd.

Lösning. Vi betraktar siffrorna som ett stickprov från oberoende s. v. $X_j \sim N(\mu, \sigma)$. Vi punktskattar med $\widehat{M} = \bar{X} \sim N(\mu, \sigma/\sqrt{10})$ som tidigare och skattar σ med s , där

$$s^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2 \approx 2.0842$$

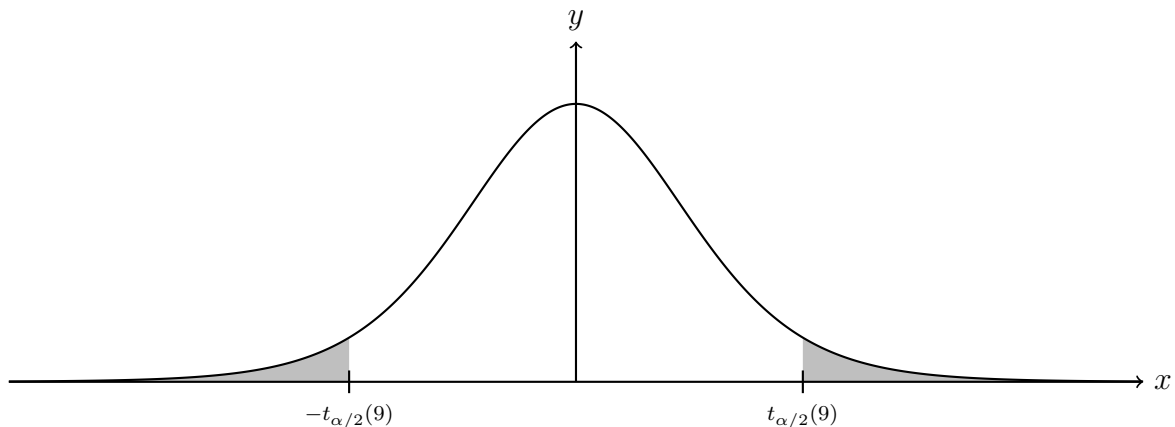
är stickprovsvariansen. Vi skapar testvariabeln

$$T = \frac{\widehat{M} - \mu}{S/\sqrt{10}} \sim t(9).$$

Som i förra deluppgiften följer det att

$$P(-t_{\alpha/2}(9) < T < t_{\alpha/2}(9)) = 1 - \alpha,$$

där $t_{\beta}(9)$ är kvantilerna till $t(9)$ -fördelningen, $\beta \in [0, 1]$.



Ur tabell finner vi $t_{0.005}(9) = 3.25$. Genom att lösa ut μ ur olikheten i sannolikhetsmåttet får vi

$$\widehat{M} - \frac{3.25 \cdot S}{\sqrt{10}} < \mu < \widehat{M} + \frac{3.25 \cdot S}{\sqrt{10}}.$$

Om vi ersätter \widehat{M} med de observerade punktskattningarna $\hat{\mu} = 5.074$ (medelvärdet av observationerna) och $s = \sqrt{2.0842} = 1.444$ (stickprovsstandardavvikelsen) så får vi ett konfidsensintervall $I_{\mu} = (3.59, 6.56)$ med konfidsensgrad 99%.

8.3 Prediktionsintervall

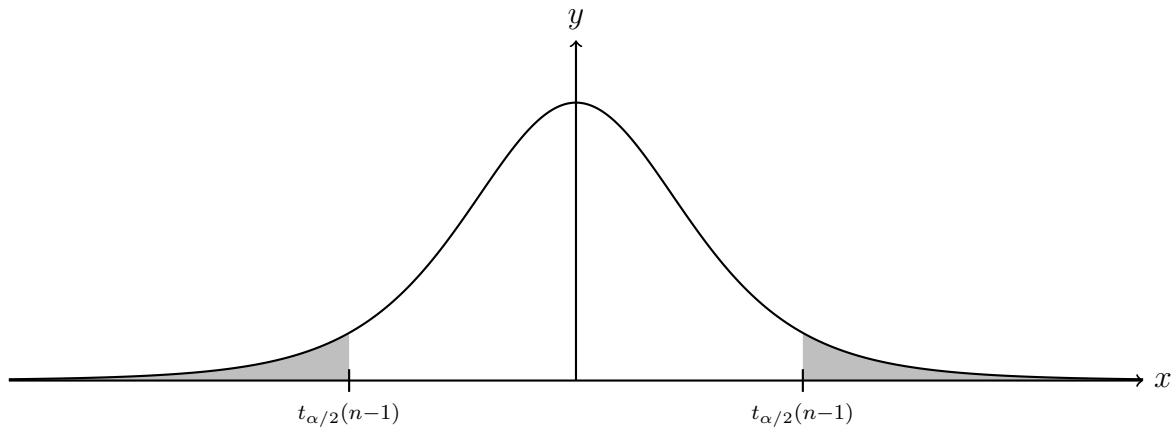
Vi har hittat konfidensintervall för väntevärdet, men kan man säga något om var en enskild observation hamnar? Låt X_1, X_2, \dots, X_n vara ett slumpmässigt stickprov från $N(\mu, \sigma)$. Vi vill stänga in en enskild observation av en variabel X_0 (som antages vara oberoende) från denna fördelning. Givetvis vill vi utnyttja stickprovet, så vi betraktar variabeln $X_0 - \bar{X}$ som är normalfördelad med

$$E(X_0 - \bar{X}) = 0 \quad \text{och} \quad V(X_0 - \bar{X}) = \sigma^2 + \frac{\sigma^2}{n}.$$

Alltså kommer

$$T = \frac{X_0 - \bar{X}}{S\sqrt{1 + \frac{1}{n}}} \sim t(n-1),$$

eftersom S^2 fortfarande är $\chi^2(n-1)$ -fördelad. Vi kan på samma sätt som tidigare stänga in denna variabel med sannolikhet $1 - \alpha$,



och sedan lösa ut X_0 :

$$\begin{aligned} -t_{\alpha/2}(n-1) &< \frac{X_0 - \bar{X}}{S\sqrt{1 + \frac{1}{n}}} < t_{\alpha/2}(n-1) \\ \Leftrightarrow \quad \bar{X} - t_{\alpha/2}(n-1) S\sqrt{1 + \frac{1}{n}} &< X_0 < \bar{X} + t_{\alpha/2}(n-1) S\sqrt{1 + \frac{1}{n}}. \end{aligned}$$

Vi ersätter nu \bar{X} med det observerade medelvärdet \bar{x} och S med stickprovsstandardavvikelsen s och får då intervallet

$$I_{X_0} = \left(\bar{x} - t_{\alpha/2}(n-1) s\sqrt{1 + \frac{1}{n}}, \bar{x} + t_{\alpha/2}(n-1) s\sqrt{1 + \frac{1}{n}} \right).$$

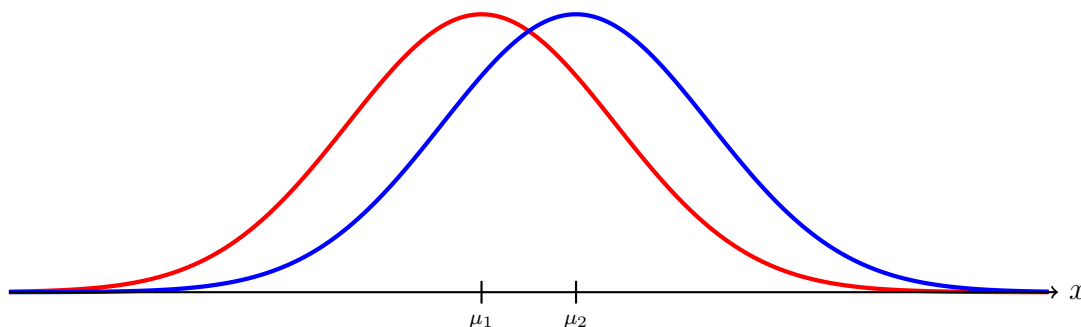
Om vi jämför intervallen för väntevärde respektive predikterat värde ser vi att $I_{X_0} \subset I_\mu$ alltid gäller med den metod vi använt ovan. Ett prediktionsintervall blir alltså alltid bredare om vi utgår från samma stickprov och samma konfidensgrad.



Se upp med om en fråga ställs angående konfidsensintervall för väntevärde eller ett prediktionsintervall. Det är helt olika frågor! Svarar du med ett konfidsensintervall för väntevärdet när det efterfrågas ett prediktionsintervall blir det noll poäng.

8.4 Skillnad mellan parametrar

Vi kommer nu fortsätta med att konstruera konfidsensintervall och vi kommer betrakta lite olika situationer där vi börjar med att titta på framför allt skillnader mellan olika mätningar. En rimlig fråga är om det föreligger någon skillnad mellan till exempel väntevärden för två stycken stickprov. Antag att vi har två slumpmässiga stickprov från två normalfördelningar. Vi vet inte direkt om fördelningarna har samma parametrar, så situationen skulle kunna se ut enligt följande.



Hur avgör vi om till exempel $\mu_1 = \mu_2$? Eller snarare om det är så att $\mu_1 \neq \mu_2$? Eller kanske om $\mu_2 > \mu_1$? Går det att avgöra om varianserna skiljer sig åt? Vad gör vi om inte stickprovet är från en normalfördelning?

8.5 Linjärkombinationer av normalfördelningar

Låt X_1, \dots, X_m och Y_1, \dots, Y_n vara oberoende slumpmässiga stickprov från $N(\mu_1, \sigma_1)$ respektive $N(\mu_2, \sigma_2)$. Om c_1 och c_2 är konstanter, kan vi hitta ett konfidsensintervall för linjärkombinationen $c_1\mu_1 + c_2\mu_2$? Svaret beror på vilka antaganden vi gör. Vi börjar med att hitta en lämplig stokastisk storhet. Vi ser att

$$E(c_1\bar{X} + c_2\bar{Y}) = c_1\mu_1 + c_2\mu_2 \quad \text{och} \quad V(c_1\bar{X} + c_2\bar{Y}) = c_1^2 \frac{\sigma_1^2}{m} + c_2^2 \frac{\sigma_2^2}{n},$$

så eftersom vi har oberoende normalfördelade variabler gäller att

$$Z = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_1 + c_2\mu_2)}{\sqrt{c_1^2 \frac{\sigma_1^2}{m} + c_2^2 \frac{\sigma_2^2}{n}}} \sim N(0, 1). \quad (8.2)$$

Om vi känner σ_1 och σ_2 räcker detta för att ställa upp ett resultat.

8.5.1 Känd varians

**Kända varianser**

Antag att följande värden är uppmätta.

x_i	47.7	55.6	51.3	46.1	54.9			
y_i	29.2	47.8	30.9	37.7	27.9	40.1	41.5	40.9

Låt x_i vara observationer av stokastiska variabler $X_i \sim N(\mu_1, 4)$ och y_i observationer av stokastiska variabler $Y_i \sim N(\mu_2, 9)$, där samtliga variabler är oberoende. Ange ett 95% konfidensintervall för $\mu_1 - 2\mu_2$.

Lösning: Låt $W = \bar{X} - 2\bar{Y}$. Varför? Denna storhet har egenskapen att

$$E(W) = E(\bar{X}) - 2E(\bar{Y}) = \mu_1 - 2\mu_2,$$

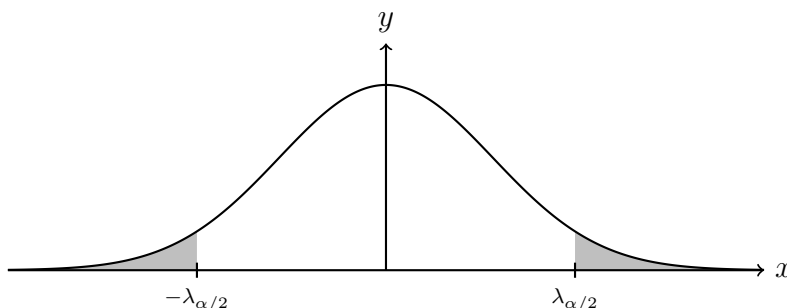
vilket är precis vad vi är intresserade av. Vidare är

$$V(W) = V(\bar{X}) + (-2)^2 V(\bar{Y}) = \frac{4^2}{5} + 4 \frac{9^2}{8} = 43.7.$$

En lämplig teststorhet ges av

$$Z = \frac{W - (\mu_1 - 2\mu_2)}{\sqrt{V(W)}} \sim N(0, 1).$$

Dags för en obligatorisk principfigur!



Eftersom

$$P(-\lambda_{\alpha/2} < Z < \lambda_{\alpha/2}) = 1 - \alpha$$

kan vi ur olikheten lösa ut sambandet

$$W - \lambda_{\alpha/2} \sqrt{V(W)} < \mu_1 - 2\mu_2 < W + \lambda_{\alpha/2} \sqrt{V(W)}.$$

Vi skattar W med $w = \bar{x} - 2\bar{y} = 51.12 - 2 \cdot 37 = -22.88$. Ur tabell finner vi att $\Phi(1.96) = 0.95 + 0.025 = 0.975$, så $\lambda_{\alpha/2} = 1.96$. Alltså blir intervallet

$$\begin{aligned} I_{\mu_1 - 2\mu_2} &= (-22.88 - 1.96 \cdot \sqrt{43.7}, -22.88 + 1.96 \cdot \sqrt{43.7}) \\ &= (-35.84, -9.92). \end{aligned}$$

Vad säger detta oss? Jo, att med 95% säkerhet så ligger det verkliga värdet för $\mu_1 - 2\mu_2$ i intervallet $(-35.84, -9.92)$. Till exempel ser vi att noll inte finns med i intervallet, så det måste vara så att $2\mu_2 > \mu_1$ med hög säkerhet!

8.5.2 Okända men likadana varianser ($\sigma_1 = \sigma_2$)

Så om vi inte känner till vad varianserna är behöver vi skatta dessa. Om vi dessutom antar att $\sigma_1 = \sigma_2$ får vi ett enklare resultat, så vi börjar med det. Om vi nyttjar att $\sigma_1 = \sigma_2 = \sigma$ i ekvation (8.2) erhåller vi att

$$Z = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_1 + c_2\mu_2)}{\sigma \sqrt{\frac{c_1^2}{m} + \frac{c_2^2}{n}}} \sim N(0, 1).$$

Men vi vet fortfarande inte vad σ är, så vi ersätter σ med stickprovsstandardavvikelsen s . Eftersom vi har två stickprov viktar vi ihop dessa på sedvanligt sätt:

$$s^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}.$$

Motsvarande stickprovsvariabel S^2 uppfyller som bekant att $\frac{(m+n-2)S^2}{\sigma^2} \sim \chi^2(m+n-2)$ och enligt Gossets sats blir

$$T = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_1 + c_2\mu_2)}{S \sqrt{\frac{c_1^2}{m} + \frac{c_2^2}{n}}} \sim t(m+n-2).$$



Okänd varians

Samma siffror som i exemplet ovan, men nu vet vi inte vad standardavvikelserna är. Antag att de är lika, dvs att $\sigma_1 = \sigma_2 = \sigma$. Finn ett 95% K.I. för $\mu_1 - \mu_2$ (inte samma uttryck som sist!). Kan du säga något om påståendet att $\mu_1 > \mu_2$?

Lösning: Vi antar alltså här att $X_i \sim N(\mu_1, \sigma)$ och $Y_i \sim N(\mu_2, \sigma)$. Vi kan skatta varianserna för varje serie med de vanliga stickprovsvarianserna, så

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{och} \quad s_2^2 = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2$$

är kända storheter. Dessa viktras ihop enligt

$$s^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}.$$

Det följer nu att

$$T = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_1 + c_2\mu_2)}{S \sqrt{\frac{c_1^2}{n} + \frac{c_2^2}{m}}} \sim t(n+m-2).$$

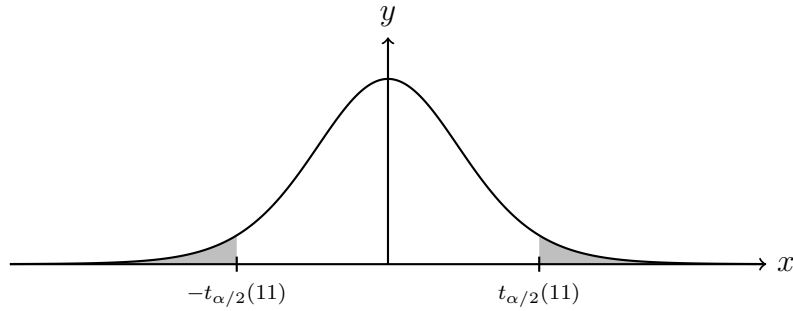
Låt $k := \sqrt{\frac{c_1^2}{n} + \frac{c_2^2}{m}}$. Snarlikt med fallet där vi kände varianserna kan vi stänga in T :

$$P(-t_{\alpha/2}(n+m-2) < T < t_{\alpha/2}(n+m-2)) = 1 - \alpha$$

där vi ur olikheten kan lösa ut sambandet

$$T - t_{\alpha/2}(n+m-2) \cdot S \cdot k < c_1\mu_1 + c_2\mu_2 < T + t_{\alpha/2}(n+m-2) \cdot S \cdot k.$$

Vi har $n = 5$ och $m = 8$, så $m+n-2 = 11$ frihetsgrader. Ur tabell finner vi att $t_{0.025}(11) = 2.20$.



Vi kan räkna ut stickprovsvarianserna för x_i och y_i separat (med formel eller miniräknare). Vi erhåller $s_1^2 = 17.822$ och $s_2^2 = 49.009$ (små bokstäver, ej stokastiskt!). Den sammanvägda standardavvikelsen blir då

$$s = \sqrt{\frac{4s_1^2 + 7s_2^2}{11}} = 6.1374.$$

Vidare är $c_1 = 1$ och $c_2 = -1$, så

$$k = \sqrt{\frac{c_1^2}{n} + \frac{c_2^2}{m}} = \sqrt{\frac{1}{5} + \frac{1}{8}} = 0.5701.$$

Alltså blir

$$t_{0.025}(11)s\sqrt{\frac{c_1^2}{n} + \frac{c_2^2}{m}} = 2.20 \cdot 6.1374 \cdot 0.5701 = 7.6976.$$

Vi kan också räkna ut att $\bar{x} - \bar{y} = 14.12$, så det sökta intervallet ges av

$$\begin{aligned} I_{\mu_1 - \mu_2} &= (14.12 - 7.70, 14.12 + 7.70) \\ &= (6.42, 21.82). \end{aligned}$$

Vi ser att noll ej ingår i intervallet, så det förligger troligt att $\mu_1 > \mu_2$.

8.5.3 Okända varianser ($\sigma_1 \neq \sigma_2$)

Ha ha. Well.. vi har inget användbart exakt samband, men det finns metoder för att hantera även denna situation. Dessa metoder ligger utanför denna kurs, men det kanske kan vara intressant att ha hört talas om dem. Problemet ligger i att uppskatta frihetsgraden ν för $t(\nu)$ -fördelningen. Man kan visa (Welch-Satterthwaite-ekvationen) att

$$\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \stackrel{\text{appr.}}{\sim} \chi^2(\nu), \quad \text{där } \nu = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2 \bigg/ \left(\frac{1}{n_1 - 1} \frac{s_1^4}{n_1^2} + \frac{1}{n_2 - 1} \frac{s_2^4}{n_2^2} \right).$$

Därför kan vi till exempel använda att

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \stackrel{\text{appr.}}{\sim} t(\nu)$$

för att ställa upp ett konfidensintervall för $\mu_1 - \mu_2$.

8.6 Enkelsidiga konfidensintervall

De konfidensintervall vi arbetat med har varit tvåsidiga i den meningen att båda gränserna har varit observationer av stokastiska variabler. Det innebär att vi lagt ut den osäkerhet vi tillåter på båda "svansarna" i fördelningen. Men det är givetvis inte nödvändigt. Kanske är vi bara intresserade av gränsen åt ena hållet?

Typexemplet är konfidensintervall för variansen (även om vi inte hinner dit i denna kurs). Att variansen är liten brukar inte vara något större bekymmer, så vi lägger allt krut på att hålla koll på gränsen uppåt. Men det kan även handla om väntevärdet (eller ett predikerat värde). Kanske mäter vi något där vi inte får överstiga en viss nivå. Kanske en situation där det inga problem är om koncentrationen av något skadligt ämne är låg, men ett betydligt större problem om koncentrationen är hög?

Så hur åstadkommer vi detta? Vi betraktar ett exempel.



Exempel

Belinda är en hobbykemist som experimenterar med organiska peroxider. Hon försöker syntetisera hexametyltriperoxidiamin (HMTD) med två snarlika metoder. Den första använder citronsyra medan den andra använder isättika. Belinda är intresserad om utbytet blir bättre med citronsyra för att se om det är värt det extra besväret då denna metod producerar mer värme och kräver större försiktighet. Hon har gjort 10 experiment för varje metod och utbytet (beräknat som en kvot med mängden hexametylendiamin som används) i procent avrundat till heltal kan ses nedan. Vi antar att mätningar är normalfördelade och att olika tillverkningsomgångar är oberoende. Vi antar också att variansen är densamma för båda metoderna (rimligt?).

	Utbyte										\bar{x}	s
Citronsyra	55	36	55	64	53	58	55	45	51	40	51.2	8.5088
Isättika	50	38	39	40	27	54	47	40	53	35	42.3	8.5641

Utför ett test med åtminstone ett konfidensintervall för att se om metoden med citronsyra ger ett bättre utbyte än isättika. Använd konfidensgraden 90%.

Lösning. Modellen för citronsyran är att $X_i \sim N(\mu_1, \sigma)$ och för isättikan gäller $Y_i \sim N(\mu_2, \sigma)$ (samma varians). Alla variabler antas vara oberoende.

Vi viktat ihop varianserna:

$$s^2 = \frac{9s_1^2 + 9s_2^2}{18} = \frac{1}{2} (s_1^2 + s_2^2).$$

Det följer nu av Cochrans and Gossets satser att

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{10} + \frac{1}{10}}} \sim t(18),$$

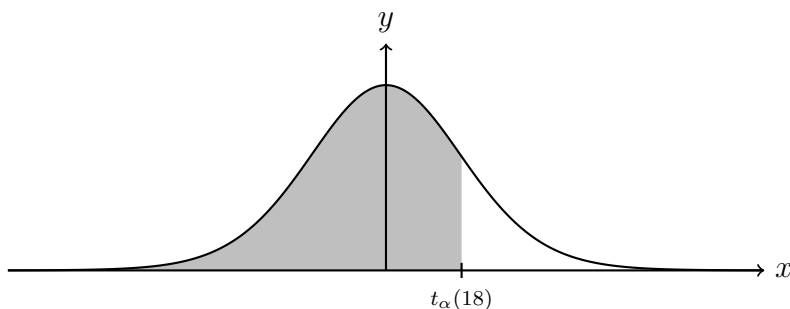
och vi vill att

$$P(T < t_\alpha(18)) = 1 - \alpha,$$

där vi då kan lösa olikheten för att finna att

$$\bar{X} - \bar{Y} - t_\alpha(18) \cdot \frac{S}{\sqrt{5}} < \mu_1 - \mu_2.$$

Vi använder ett enkelsidigt intervall eftersom vi endast vill undersöka om $\mu_1 > \mu_2$. Från en tabell finner vi att $t_{0.10}(18) = 1.3304$.



Som en observation av S använder vi $\sqrt{s^2}$, så

$$t_{0.10}(18) \frac{s}{\sqrt{5}} = 1.3304 \cdot 3.8176 = 5.0790.$$

Eftersom $\bar{x} - \bar{y} = 8.9$ så ges därmed det eftersökta intervallet av

$$\begin{aligned} I_{\mu_1 - \mu_2} &= (8.9 - 5.0790, \infty) \\ &= (3.821, \infty). \end{aligned}$$

Här ser vi att 0 inte är med i intervallet, så vi kan säga att citronsyran verkar ge bättre utbyte (det är rimligt att $\mu_1 > \mu_2$) på den här signifikansnivån.

8.7 Stickprov i par

Om stickproven X_1, \dots, X_m och Y_1, \dots, Y_n inte är oberoende får vi problem. Åtminstone om inte beroendet är känt. Låt oss betrakta ett vanligt förekommande exempel, nämligen stickprov i par. Av nödvändighet är då $m = n$ så stickproven har samma storlek. Vi tänker oss att x_k är observationer från $X_k \sim N(\mu_k, \sigma_1)$ och $Y_k \sim N(\mu_k + \Delta, \sigma_2)$. Typexemplet är när vi mäter något före och efter en förändring.

Bilda nu ett "nytt" stickprov Z_k av oberoende variabler:

$$Z_k = Y_k - X_k \sim N(\Delta, \sigma),$$

för något σ . Vi är nu tillbaka där vi var föregående föreläsning, så de tekniker vi utvecklade där fungerar även nu.



Exempel

Preparat mot (h)järnbrist. Mätningar (någon enhet) före och efter behandling.

Person	1	2	3	4	5	6	7	8	9
Före	15.8	12.1	18.2	9.4	11.8	16.6	13.7	13.5	17.5
Efter	14.8	12.4	18.3	9.5	12.2	15.6	13.4	14.4	16.0

Bestäm ett 99% KI av den genomsnittliga effekten hos preparatet. Kan du styrka funktionen?

Lösning: Låt x_i vara värde före behandling för person i och y_i motsvarande värde efter behandling. Vi antar att olika personer är oberoende, att x_i är observationer av $X_i \sim N(\mu_i, \sigma_1)$ och att y_i är observationer av $Y_i \sim N(\mu_i + \Delta, \sigma_2)$. Bilda $Z_i = Y_i - X_i \sim N(\Delta, \sigma)$. Vi har nu en enda serie $z_i = y_i - x_i$ som ges enligt

$$z_i \mid -1.0 \quad 0.3 \quad 0.1 \quad 0.1 \quad 0.4 \quad -1.0 \quad -0.3 \quad 0.9 \quad 0.5$$

Vi räknar ut $s = 0.7886$ och $\bar{z} = 0.2222$. Vidare är $n - 1 = 8$ och $\alpha = 0.01$, så från tabell finner vi att $t_{\alpha/2}(8) = t_{0.005}(8) = 3.36$. Alltså:

$$I_\Delta = (0.222 - 3.36 \cdot 0.7886/\sqrt{9}, 0.222 + 3.36 \cdot 0.7886/\sqrt{9}) = (-0.66, 1.11).$$

Eftersom nollan finns med kan vi inte förkasta att $\Delta = 0$ (med 99% säkerhet). Preparatet kan alltså vara verkningslöst.

8.8 Konfidensintervall via CGS

Så vad gör vi om stickprovet inte är från en normalfördelning?

8.9 Stickprov för andel



Exempel

Ett företag som sysslar med opinionsanalys väljer slumpmässigt ut 400 vuxna i Sverige och frågar om de har åsikt A. Av dessa svarar 80 ja (alla svarar). Bestäm ett approximativt 95% konfidensintervall för andelen av den stora populationen som håller åsikt A.

Lösning. Vi låter X vara antalet som svarar ja. Då är egentligen $X \sim \text{Hyp}(N, 400, p)$, där N är antalet vuxna i Sverige (rimligen ca 8 miljoner). Då $400 \ll 8000000$ är det helt rimligt att anta att $X \stackrel{\text{appr.}}{\sim} \text{Bin}(400, p)$. Vi vill skatta den okända andelen p och väljer som skattningsvariabel

$$\hat{P} = \frac{X}{400}$$

Vi har observerat att $\hat{p} = 80/400 = 0.2$.

Binomialfördelningen är lite jobbig eftersom den är diskret, så vi försöker oss på en approximation. Eftersom

$$400 \cdot \hat{p} \cdot (1 - \hat{p}) = 400 \cdot 0.2 \cdot 0.8 = 64$$

är ordentligt större än 10 är det rimligt att approximera binomialfördelningen med normalfördelning. Alltså,

$$\hat{P} \stackrel{\text{appr.}}{\sim} N(p, \sqrt{p(1-p)/400}).$$

Låt oss bilda

$$Z = \frac{\hat{P} - p}{\sqrt{\hat{p}(1-\hat{p})/400}} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Observera att vi ersatt med det skattade värdet på p i kvadratroten (men **inte** i täljaren). Vi nyttjar här alltså medelfelet d , dvs

$$d(\hat{P}) = \sqrt{\hat{p}(1-\hat{p})/400} = 0.02.$$

Vi kan nu räkna precis som om vi känner standardavvikelsen exakt, så om vi söker ett approximativt 95% K.I. erhåller vi

$$I_p = (0.2 - 1.96 \cdot 0.02, 0.2 + 1.96 \cdot 0.02) = (0.16, 0.24).$$

8.10 Jämförelse av två andelar



Exempel

Anonyme Alva har tillgång till två maskiner för att pressa piller. I dessa pressar Alva en perfekt homogeniserad blandning av obsykra bensodiazepiner med en tillsats av fentanyl för lite extra skjuts. Vid uppmätning fann man att Maskin 1 producerade 20 defekta enheter av 400 (där totala mängden fentanyl blir farligt hög för opiatnaiva individer), och att Maskin 2 producerade 60 defekta enheter av 600. Är det någon skillnad på andelen felaktiga piller producerade med de olika maskinerna? Svara med approximativt signifikansnivån 5%.

Lösning. Modell: Låt X vara antal defekta enheter från Maskin 1 och Y antal defekta enheter från Maskin 2. Under lämpligt oberoendeantagande vet vi att $X \sim \text{Bin}(400, p_1)$ och $Y \sim \text{Bin}(600, p_2)$ där p_1 och p_2 är de verkliga felsannolikheterna. Vi skattar lämpligen med

$$\widehat{P}_1 = \frac{X}{400} \quad \text{och} \quad \widehat{P}_2 = \frac{Y}{600}.$$

Vi har observerat att $\widehat{p}_1 = 20/400 = 0.05$ och $\widehat{p}_2 = 60/600 = 0.10$. Alltså är $\widehat{p}_1 - \widehat{p}_2 = -0.05$. Är detta signifikant? För att svara på frågan behöver vi räkna lite sannolikheter. Eftersom både $n_1\widehat{p}_1(1 - \widehat{p}_1)$ och $n_2\widehat{p}_2(1 - \widehat{p}_2)$ är mycket större än 10 är det rimligt att approximera binomialfördelningen med normalfördelning. Alltså,

$$\widehat{P}_1 \stackrel{\text{appr.}}{\sim} N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{400}}\right) \quad \text{och} \quad \widehat{P}_2 \stackrel{\text{appr.}}{\sim} N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{600}}\right).$$

Då följer det att

$$\widehat{P}_1 - \widehat{P}_2 \stackrel{\text{appr.}}{\sim} N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{400} + \frac{p_2(1-p_2)}{600}}\right).$$

Vi bildar nu

$$Z = \frac{\widehat{P}_1 - \widehat{P}_2 - (p_1 - p_2)}{\sqrt{\widehat{p}_1(1 - \widehat{p}_1)/400 + \widehat{p}_2(1 - \widehat{p}_2)/600}} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Observera att vi ersatt med skattade värden på p_1 och p_2 i kvadratroten (men **inte** i täljaren). Det blir fortfarande approximativt (men lite sämre så klart) normalfördelat, men underlättar mycket för beräkningar. Vi har

$$\sqrt{\widehat{p}_1(1 - \widehat{p}_1)/400 + \widehat{p}_2(1 - \widehat{p}_2)/600} = 0.0164.$$

Vi kan nu räkna precis som om vi känner standardavvikelsen exakt, så om vi söker ett approximativt 95% K.I. erhåller vi

$$I_{p_1-p_2} = (-0.05 - 1.96 \cdot 0.0164, -0.05 + 1.96 \cdot 0.0164) = (-0.08, -0.02).$$

Endast negativa värden, så $p_1 < p_2$ med hög sannolikhet! Maskin 2 är antagligen sämre.

8.11 $(\star\star)\chi^2$ -fördelningen

En situation som dyker upp frekvent i statistik inferens är summor av kvadrater av normalfördelade variabler, så en naturlig fråga är så klart vilken fördelning en sådan summa får (åtminstone då variablerna antas vara oberoende). Svaret fås i form av χ^2 -fördelningen.



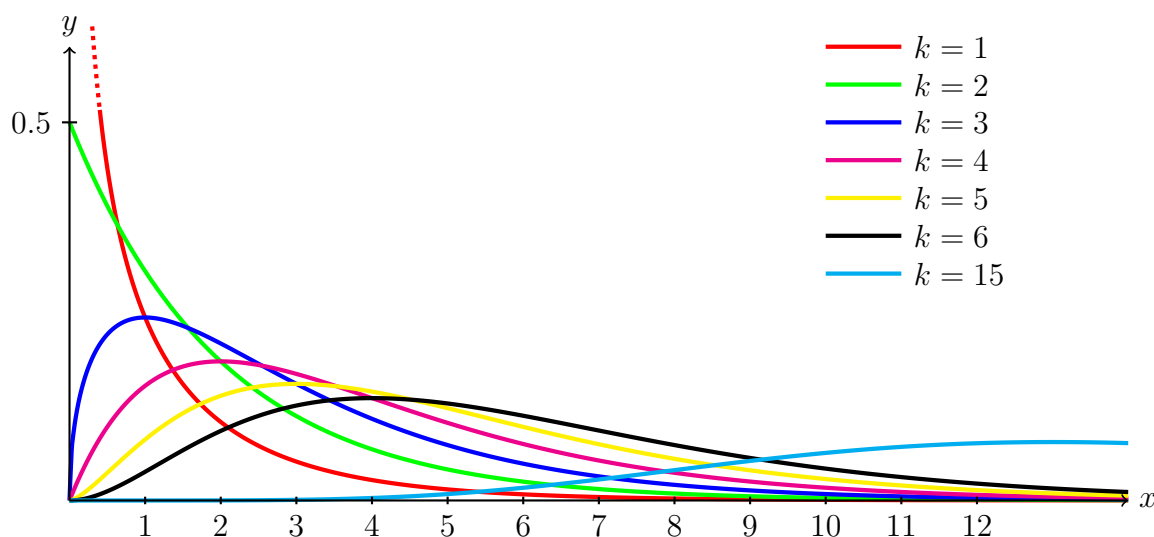
χ^2 -fördelning

Definition. Om X är en stokastisk variabel med täthetsfunktionen

$$f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad x \geq 0 \text{ om } k > 1,$$

kallar vi X för $\chi^2(k)$ -fördelad med k frihetsgrader, där $k = 1, 2, \dots$

Här är Γ gamma-funktionen¹ och $\Gamma(n) = (n-1)!$ och $\Gamma(n+1/2) = \frac{(2n)!}{4^n n!} \sqrt{\pi}$ om $n \in \mathbb{N}$.



Om $X \sim \chi^2(k)$ är $E(X) = k$ och $V(X) = 2k$.

Bevis. Låt täthetsfunktionen skrivas $f(x) = cx^{k/2-1}e^{-x/2}$. Då gäller att

$$\begin{aligned} E(X) &= c \int_0^\infty x^{k/2} e^{-x/2} dx = c \left([-2x^{k/2} e^{-x/2}]_0^\infty + 2 \int_0^\infty \frac{k}{2} x^{k/2-1} e^{-x/2} dx \right) \\ &= k \int_0^\infty f(x) dx = k. \end{aligned}$$

På samma sätt följer att

$$E(X^2) = c \int_0^\infty x^{k/2+1} e^{-x/2} dx = 2 \left(\frac{k}{2} + 1 \right) c \int_0^\infty x^{k/2} e^{-x/2} dx = (k+2)E(X) = k^2 + 2k,$$

så $V(X) = E(X^2) - E(X)^2 = k^2 + 2k - k^2 = 2k$. □

¹Se avsnitt 8.16 nedan för mer detaljer.



Sats. Om $X \sim \chi^2(\nu_1)$ och $Y \sim \chi^2(\nu_2)$ är oberoende så är $X + Y \sim \chi^2(\nu_1 + \nu_2)$.

Bevis. Enklast är att betrakta Fouriertransformen för täthetsfunktionen (alternativt den närbesläktade **karaktäristiska funktionen** definierad enligt $E(e^{itX})$). Det är nämligen så att

$$\mathcal{F}(f_X)(t) = (1 + 2it)^{-\nu_1/2}, \quad \mathcal{F}(f_Y)(t) = (1 + 2it)^{-\nu_2/2}$$

och

$$\mathcal{F}(f_X * f_Y) = \mathcal{F}(f_X)\mathcal{F}(f_Y) = (1 + 2it)^{-(\nu_1 + \nu_2)/2},$$

så $f_{X+Y} \sim \chi^2(\nu_1 + \nu_2)$. □



Sats. Om X_1, X_2, \dots, X_n är oberoende och $X_k \sim N(0, 1)$ så är

$$\sum_{k=1}^n X_k^2 \sim \chi^2(n).$$

Bevis. Eftersom variablerna är oberoende ges den simultana täthetsfunktionen av

$$f(x_1, \dots, x_n) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} = \frac{1}{2^{n/2} \pi^{n/2}} \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_n^2)\right).$$

Vi söker fördelningen för $Z = X_1^2 + \dots + X_n^2$, så låt oss ställa upp fördelningsfunktionen:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = \int_{x_1^2 + \dots + x_n^2 \leq z} f(x_1, \dots, x_n) dx_1 dx_2 \dots dx_n \\ &= \frac{1}{2^{n/2} \pi^{n/2}} \int_{S^{n-1}} \int_0^{\sqrt{z}} r^{n-1} e^{-r^2/2} dr dS \\ &= \frac{1}{2^{n/2} \pi^{n/2}} \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^{\sqrt{z}} r^{n-1} e^{-r^2/2} dr \\ &= \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^z t^{n/2-1} e^{-t/2} dt, \end{aligned}$$

där S^{n-1} är enhetssfären i \mathbf{R}^n och dS är ytmåttet på S^{n-1} . Då enhetssfären har ytmåttet $|S^{n-1}| = \frac{2\pi^{n/2}}{\Gamma(n/2)}$ följer likheten ovan efter ett variabelbyte i sista integralen (låt $t = r^2$).

Analysens huvudsats medför nu att (för $z > 0$) att

$$f_Z(z) = F'_Z(z) = \frac{1}{2^{n/2} \Gamma(n/2)} z^{n/2-1} e^{-z/2}.$$

För $z < 0$ är givetvis $f_Z(z) = 0$ (varför?). □

8.12 (★) t -fördelningen



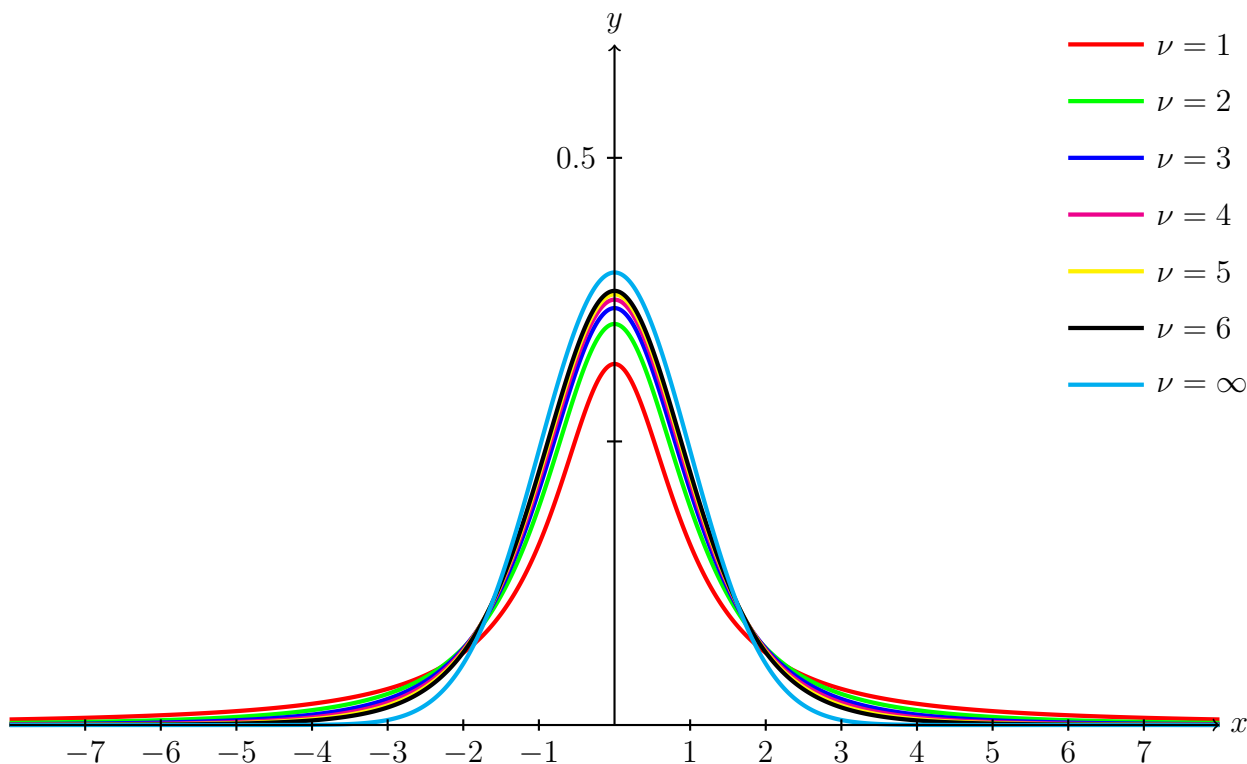
t -fördelning

Definition. Om X är en stokastisk variabel med täthetsfunktionen

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbf{R} \text{ och } \nu > 0,$$

kallar vi X för $t(\nu)$ -fördelad med ν frihetsgrader.

Denna fördelning är symmetrisk och om antalet frihetsgrader går mot oändligheten konvergerar täthetsfunktionen mot täthetsfunktionen för normalfördelning.



Sats. Om $X \sim t(\nu)$ är $E(X) = 0$ (om $\nu > 1$) och $V(X) = \nu/(\nu - 2)$ (om $\nu > 2$).

Bevis. Om $\nu > 1$ är integralen $E(X)$ absolutkonvergent (visa det) och då integranden är udda blir således $E(X) = 0$. För att beräkna $E(X^2)$ låter vi $c_\nu = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)}$.

Om $\nu > 2$ ser vi genom partialintegration att

$$\begin{aligned} E(X^2) &= c_\nu \int_{-\infty}^{\infty} x \cdot x \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx = c_\nu \int_{-\infty}^{\infty} \frac{\nu}{\nu-1} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu-1}{2}} dx \\ &= c_\nu \frac{\nu}{\nu-1} \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu-1}{2}} dx \\ &= \frac{c_\nu}{c_{\nu-2}} \frac{\nu^{3/2}}{(\nu-1)\sqrt{\nu-2}} c_{\nu-2} \int_{-\infty}^{\infty} \left(1 + \frac{u^2}{\nu-2}\right)^{-\frac{\nu-1}{2}} du = \frac{c_\nu}{c_{\nu-2}} \frac{\nu^{3/2}}{(\nu-1)\sqrt{\nu-2}} \end{aligned}$$

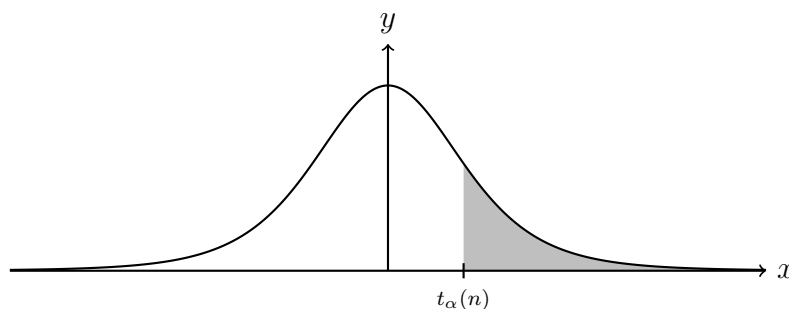
där vi bytte variabel så $x\sqrt{\nu-2} = u\sqrt{\nu}$ och utnyttjade att integralen som dök upp är precis integralen av täthetsfunktionen för en $t(\nu-2)$ -fördelad variabel (om $\nu > 2$). Vi förenklar uttrycket och finner att

$$\begin{aligned} \frac{c_\nu}{c_{\nu-2}} \frac{\nu^{3/2}}{(\nu-1)\sqrt{\nu-2}} &= \frac{\Gamma\left(\frac{\nu+1}{2}\right) \Gamma\left(\frac{\nu-2}{2}\right) \sqrt{(\nu-2)\pi}}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{\nu-1}{2}\right)} \frac{\nu^{3/2}}{(\nu-1)\sqrt{\nu-2}} \\ &= \frac{\frac{\nu-1}{2} \Gamma\left(\frac{\nu-1}{2}\right) \Gamma\left(\frac{\nu-2}{2}\right)}{\frac{\nu-2}{2} \Gamma\left(\frac{\nu-2}{2}\right) \Gamma\left(\frac{\nu-1}{2}\right)} \frac{\nu}{\nu-1} = \frac{\nu}{\nu-2}, \end{aligned}$$

där vi nyttjat att $\Gamma(z+1) = z\Gamma(z)$. Eftersom $E(X) = 0$ följer det nu att $V(X) = E(X^2)$. \square

8.12.1 t -fördelningens kvantiler

Kvantilerna för t -fördelningen är de tal $t_\alpha(n)$ sådana att $P(T > t_\alpha(n)) = 1 - \alpha$. Det vill säga gränser $t_\alpha(n)$ sådana att för $T \sim t(n)$ gäller att andelen α av sannolikhetsmassan ligger till höger om $t_\alpha(n)$. Eftersom gränserna är jobbiga att räkna fram för hand brukar vi använda tabellverk enligt nedan (studera även formelsamlingen).



$\alpha \backslash n$	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
70	1.294	1.667	1.994	2.381	2.648	3.211	3.435
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
90	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

8.13 (★)Vektorer med stokastiska variabler

Låt $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ vara en vektor vars komponenter är stokastiska variabler. Vi strävar efter att skriva vektorer som kolonnvektorer. Det faller sig naturligt att definiera väntevärdet av \mathbf{X} genom

$$E(\mathbf{X}) = (E(X_1), E(X_2), \dots, E(X_n)).$$

På samma sätt definierar vi väntevärdet av en matris av stokastiska variabler. Variansen blir lite konstigare så vi introducerar den så kallade kovariansmatrisen mellan två vektorer (av samma dimension). Låt $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ och definiera $C(\mathbf{X}, \mathbf{Y})$ enligt

$$C(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} C(X_1, Y_1) & C(X_1, Y_2) & \cdots & C(X_1, Y_n) \\ C(X_2, Y_1) & C(X_2, Y_2) & \cdots & C(X_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(X_n, Y_1) & C(X_n, Y_2) & \cdots & C(X_n, Y_n) \end{pmatrix}$$

där $C(X_i, Y_j) = E(X_i Y_j) - E(X_i)E(Y_j)$ är kovariansen mellan X_i och Y_j .

En stor anledning att blanda in vektorer och matriser är givetvis att få tillgång till maskineriet från linjär algebra. Kovariansen (en matris) mellan två vektorer \mathbf{X} och \mathbf{Y} kan då lite mer kompakt skrivas

$$C(\mathbf{X}, \mathbf{Y}) = E(\mathbf{X}\mathbf{Y}^T) - E(\mathbf{X})E(\mathbf{Y})^T,$$

där $(\cdot)^T$ innebär transponering. En produkt $A = \mathbf{x}\mathbf{y}^T$ brukar kallas för den yttre produkten och består av element $(a)_{ij} = x_i y_j$, $i, j = 1, 2, \dots, n$. Detta är alltså *inte* skalärprodukten $(\mathbf{X}^T \mathbf{Y})$. Låt $A, B \in \mathbf{R}^{n \times n}$ vara matriser. Då är $A\mathbf{X}$ en linjärkombination av X_1, X_2, \dots, X_n och $B\mathbf{Y}$ en linjärkombination av Y_1, Y_2, \dots, Y_n . Dessutom kan *alla* linjärkombinationer skrivas på detta sätt. Vidare gäller nu tack varje linjäriteten att

$$E(A\mathbf{X}) = AE(\mathbf{X}) \quad \text{och} \quad C(A\mathbf{X}, B\mathbf{Y}) = A\mathbf{X}(B\mathbf{Y})^T = A\mathbf{X}\mathbf{Y}^T B^T.$$

8.14 (★)Cochrans sats



Sats. Låt X_1, X_2, \dots, X_n vara oberoende likafördelade stokastiska variabler där $X_k \sim N(\mu, \sigma)$ för $k = 1, 2, \dots, n$. Då gäller att

$$\frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \bar{X})^2 \sim \chi^2(n-1).$$

Bevis. Låt $Y_k = X_k - \mu$ så att $Y_k \sim N(0, \sigma)$. Vi ser att

$$\sum_{k=1}^n (X_k - \bar{X})^2 = \sum_{k=1}^n (Y_k - \bar{Y})^2.$$

Låt J vara $n \times n$ -matrisen vars samtliga element är 1 och låt $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$. Då kan vi skriva

$$\begin{pmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{pmatrix} = I\mathbf{Y} - \frac{1}{n}J\mathbf{Y} = Q\mathbf{Y},$$

där $Q = I - (1/n)J$. Låt $P = I - Q = (1/n)J$. Då är

$$P + Q = I, \quad P^2 = P^T = P, \quad Q^2 = Q^T = Q \quad \text{samtidigt} \quad PQ = QP = 0.$$

Matriserna P och Q representerar alltså ortogonala projektioner på \mathbf{R}^n och av naturliga skäl är $\text{rank}(P) = 1$ så $\text{rank}(Q) = n - 1$ (eftersom $P + Q = I$).

Vidare gäller att

$$C(P\mathbf{Y}, Q\mathbf{Y}) = PC(\mathbf{Y}, \mathbf{Y})Q^T = P\sigma^2 I Q^T = PQ^T = PQ = 0$$

då $E(\mathbf{Y}) = 0$ och $C(Y_i, Y_j) = \sigma^2$ om $i = j$ och $C(Y_i, Y_j) = 0$ då $i \neq j$ eftersom olika Y_k är oberoende. Således är $Y_i - \bar{Y}$ och \bar{Y} oberoende stokastiska variabler (eftersom kovariansen noll mellan normalfördelade variabler är ekvivalent med oberoende). Eftersom $\text{rank}(Q) = n - 1$ så kan vi representera $Q\mathbf{Y}$ i en ortogonal bas så att

$$\frac{1}{\sigma^2} \sum_{k=1}^n (Y_i - \bar{Y})^2 = \frac{1}{\sigma^2} (Q\mathbf{Y})^T Q\mathbf{Y} = \frac{1}{\sigma^2} \mathbf{Y}^T Q\mathbf{Y} = Z_1^2 + Z_2^2 + \cdots + Z_{n-1}^2,$$

där $Z_k \sim N(0, 1)$ och dessa variabler är oberoende. Vi kan nu nyttja den tidigare satsen om att summan av n stycken kvadrater av $N(0, 1)$ -fördelade variabler är $\chi^2(n)$ -fördelad för att dra slutsatsen att $\frac{1}{\sigma^2} \mathbf{X}^T Q \mathbf{X} \sim \chi^2(n - 1)$. \square

8.15 (★★) t - och χ^2 -fördelning; Gossets sats

Det finns givetvis en anledning till att vi studerar just dessa två fördelningar. William Gosset bevisade nämligen följande sats.



Sats. Låt $Z \sim N(0, 1)$ och $V \sim \chi^2(\nu)$ vara oberoende. Då är $\frac{Z}{\sqrt{V/\nu}} \sim t(\nu)$.

Bevis. Eftersom Z och V är oberoende ges den simultana täthetsfunktionen av

$$f(z, v) = \frac{1}{\sqrt{2\pi} 2^{\nu/2} \Gamma(\frac{\nu}{2})} e^{-z^2/2} v^{\nu/2-1} e^{-v/2}, \quad z \in \mathbf{R}, v \geq 0.$$

Låt $T = \frac{Z}{\sqrt{V/\nu}}$ och $c = \frac{1}{\sqrt{2\pi} 2^{\nu/2} \Gamma(\frac{\nu}{2})}$. Vi söker täthetsfunktionen f_T för T . Betrakta

$$P(T \leq t) = \iint_{z/\sqrt{v/\nu} \leq t} f(z, v) dz dv = c \iint_{z/\sqrt{v/\nu} \leq t} e^{-z^2/2} v^{\nu/2-1} e^{-v/2} dz dv.$$

Vi gör ett variabelbyte,

$$\begin{cases} u\sqrt{\frac{v}{\nu}}, \\ w = v \end{cases} \Rightarrow \frac{d(z, v)}{d(u, w)} = \begin{vmatrix} \sqrt{\frac{w}{\nu}} & \frac{u}{2\nu\sqrt{\frac{w}{\nu}}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{w}{\nu}},$$

så integralen blir

$$\begin{aligned} c \int \int_{u \leq t} \sqrt{\frac{w}{\nu}} e^{-u^2 w / (2\nu)} w^{\nu/2-1} e^{-v/2} dz dv &= \frac{c}{\sqrt{\nu}} \int_{-\infty}^t \int_0^\infty w^{\frac{\nu+1}{2}-1} \exp\left(-\frac{w}{2} \left(1 + \frac{u^2}{\nu}\right)\right) dw du \\ &= \frac{c}{\sqrt{\nu}} \int_{-\infty}^t \int_0^\infty \frac{r^{\frac{\nu+1}{2}-1}}{\left(1 + \frac{u^2}{\nu}\right)^{\frac{\nu+1}{2}}} e^{-\frac{r}{2}} dr du, \\ &= \frac{c}{\sqrt{\nu}} \int_{-\infty}^t \left(1 + \frac{u^2}{\nu}\right)^{-\frac{\nu+1}{2}} \int_0^\infty r^{\frac{\nu+1}{2}-1} e^{-\frac{r}{2}} dr du, \end{aligned}$$

där vi gjorde ett variabelbyte $r = w \left(1 + \frac{u^2}{\nu}\right)$ i den innersta integralen och bröt ut den faktor som inte beror på u . Den innersta integralen är nu nästan (upp till normeringskonstanten) integralen av täthetsfunktionen för en $\chi^2(\nu+1)$ -variabel, så

$$\int_0^\infty r^{\frac{\nu+3}{2}-1} e^{-\frac{r}{2}} dr = 2^{(\nu+1)/2} \Gamma\left(\frac{\nu+1}{2}\right).$$

Således ges fördelningsfunktionen

$$F_T(t) = \frac{2^{(\nu+1)/2} \Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu} \sqrt{2\pi} 2^{\nu/2} \Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^t \left(1 + \frac{u^2}{\nu}\right)^{-\frac{\nu+1}{2}} du = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^t \left(1 + \frac{u^2}{\nu}\right)^{-\frac{\nu+1}{2}} du$$

vilket efter derivering ger täthetsfunktionen

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

vilket är precis täthetsfunktionen för en $t(\nu)$ -fördelad variabel. □



Sats. Om X_1, X_2, \dots, X_n är oberoende och likafördelade med fördelningen $N(\mu, \sigma)$ så är kvoten $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, där S^2 är stickprovsvariansen.

Bevis. Detta följer direkt från föregående resultat och Cochrans sats. Vi kan formulera T enligt

$$T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{1}{\frac{1}{\sigma} S} = \frac{Z}{\sqrt{\frac{V}{n-1}}},$$

där $Z \sim N(0, 1)$ och $V = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \bar{X})^2 \sim \chi^2(n-1)$ (med $S^2 = \frac{1}{n-1} V$). □

8.16 (★)Bonus: Gammafunktioner

För $z \in \mathbf{C}$ med $\operatorname{Re} z > 0$ är integralen

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$$

absolutkonvergent. Detta är ett sätt att definiera gammafunktionen på. Funktionen ovan går även att analytiskt utvidga till $\operatorname{Re} z \leq 0$ förutom för negativa heltal. Ifrån definitionen ovan kan vi medelst partialintegration erhålla att

$$\Gamma(z+1) = \int_0^\infty x^z e^{-x} dx = [-x^z e^{-x}]_0^\infty + z \int_0^\infty x^{z-1} e^{-x} dx = z\Gamma(z).$$

Eftersom $\Gamma(1) = 1$ visar denna likhet att

$$\Gamma(n) = (n-1)!$$

för alla positiva heltal. Gamma-funktionen utvidgar således fakultetet till alla komplexa z förutom negativa heltal. Kopplingen till normaliseringen av χ^2 -fördelningen är ganska naturlig. Vi ser att om $X \sim \chi^2(k)$ så är

$$\begin{aligned} \int_0^\infty f_X(x) dx &= \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^\infty x^{k/2-1} e^{-x/2} dx \\ &= \left/ \text{variabelbyte: } \begin{array}{l} u = x/2 \\ dx = 2du \end{array} \right/ = \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^\infty 2^{k/2} u^{k/2-1} e^{-u} du \\ &= \frac{1}{\Gamma(k/2)} \Gamma(k/2) = 1. \end{aligned}$$

Även identiteten $\Gamma(z+1) = z\Gamma(z)$ är det vi använde när vi beräknade $E(X)$ och $V(X)$ (gå tillbaka och studera partialintegrationen!).

För att identifiera $\Gamma(n+1/2)$ kan vi till exempel göra variabelbytet $u = \sqrt{x}$ och partialintegrera n gånger:

$$\begin{aligned} \Gamma(n+1/2) &= \int_0^\infty x^{n+1/2} e^{-x} dx = 2 \int_0^\infty u^{2n} e^{-u^2} du = 2 \int_0^\infty u^{2n-1} \cdot (ue^{-u^2}) du \\ &= - \left[u^{2n-1} e^{-u^2} \right]_0^\infty + (2n-1) \int_0^\infty u^{2n-3} \cdot (ue^{-u^2}) du \\ &= (2n-1) \left(-\frac{1}{2} \left[u^{2n-3} e^{-u^2} \right]_0^\infty + \frac{2n-3}{2} \int_0^\infty u^{2n-5} \cdot (ue^{-u^2}) du \right) \\ &= \frac{(2n-1)(2n-3)}{2} \left(-\frac{1}{2} \left[u^{2n-5} e^{-u^2} \right]_0^\infty + \frac{2n-5}{2} \int_0^\infty u^{2n-7} \cdot (ue^{-u^2}) du \right) \\ &= \dots = \frac{(2n-1)(2n-3) \cdots (2n-(2n-1))}{2^{n-1}} \int_0^\infty u^{2n-2n} e^{-u^2} du \\ &= \frac{(2n-1)(2n-3) \cdots (2n-(2n-1))}{2^{n-1}} \frac{\sqrt{\pi}}{2} = \frac{(2n)!}{4^n n!} \sqrt{\pi}, \end{aligned}$$

där vi i sista steget använde den välkända identiteten $\int_0^\infty e^{-u^2} du = \frac{\sqrt{\pi}}{2}$.

Det finns mer eleganta sätt att ta fram identiteten för $\Gamma(n+1/2)$ genom exempelvis Eulers reflektionsformel:

$$\Gamma(1-z)\Gamma(z) = \frac{\pi}{\sin(\pi z)}, \quad z \notin \mathbf{Z},$$

men den likheten är lite mer komplicerad att bevisa, så vi nöjer oss med ovanstående.

Kapitel 9

Hypotestester

Vi har nu studerat metoder för hur man hittar lämpliga skattningar av okända parametrar och även stängt in dessa skattningar i konfidensintervall för att ha kontroll på vad som är rimligt eller ej. Den sista frågan kan man närma sig på lite annorlunda (men egentligen mer naturligt sätt) genom så kallade hypotestester (ibland kallade signifikanstester).

9.1 Hypotestest

Ett hypotesttest i detta sammanhang består av en **nollhypotes** H_0 och en **mothypotes** H_1 . Typiskt är att nollhypotesen är något vi vill motbevisa (och därmed styrka att mothypotesen antagligen gäller). I denna kurs kommer vi oftast begränsa oss till så kallade enkla nollhypoteser och oftast av typen

$$H_0 : \theta = \theta_0.$$

Mothypotesen kan väljas på olika sätt beroende på vad vi vill visa. De vanligaste är av typerna

$$H_1 : \theta \neq \theta_0 \quad \text{eller} \quad H_1 : \theta > \theta_0 \quad \text{eller} \quad H_1 : \theta < \theta_0.$$

Det går att ha betydligt mer komplicerade nollhypoteser (och mothypoteser för den delen). Ett ganska vanligt exempel är $H_0 : X$ är normalfördelad eller något dylikt. I dessa fall är det svårare att hitta ett lämpligt test. För de enkla typen av nollhypoteser så finns det ganska naturliga teststorheter.

9.1.1 Teststorhet och kritiskt område

För att testa hypotesen behöver vi en teststorhet t som avgör hur ett stickprov ska behandlas. Denna storhet har analog funktion med de som användes när vi ställde upp konfidensintervall. Vi låter x_1, x_2, \dots, x_n vara ett stickprov från en fördelning F som beror på en okänd parameter θ . Motsvarande slumpmässiga stickprov betecknas X_1, X_2, \dots, X_n i vanlig ordning.



Teststorhet/Testvariabel

Definition. En funktion $t : \mathbf{R}^n \rightarrow \mathbf{R}$ given av $t(x_1, x_2, \dots, x_n)$ kallas **teststorhet** eller **testvariabel** och är en observation av den stokastiska variabeln $t(X_1, X_2, \dots, X_n)$.

För att avgöra om vi ska förkasta H_0 väljer vi en **signifikansnivå** α och bestämmer sedan ett kritiskt område C som är en delmängd av det område funktionen t varierar över (en del av värdemängden). Detta område beror på fördelningen F och den signifikansnivå vi vill utföra hypotestestet på.



Kritiskt område, signifikansnivå

Definition. Det **kritiska området** C är ett område så att H_0 förkastas om

$$t(x_1, \dots, x_n) \in C.$$

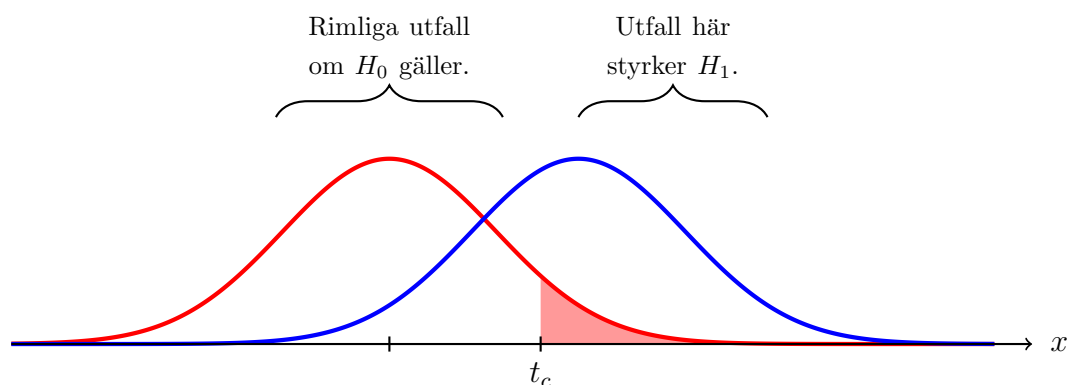
Om H_0 förkastas säger vi att H_1 är styrkt och drar slutsatsen att H_1 gäller. Sannolikheten

$$\alpha = P(t(X_1, \dots, X_n) \in C \mid H_0 \text{ är sann})$$

kallas för testets **signifikansnivå**.

Det kritiska området består alltså av värden som är för extrema för att vara troliga under förutsättningen att nollhypotesen gäller.

Låt oss ställa upp ett hypotestest för väntevärdet för fördelningen F enligt $H_0 : \mu = \mu_0$ mot $H_1 : \mu > \mu_0$. Vi vill således styrka att det verkliga väntevärdet är större än μ_0 .

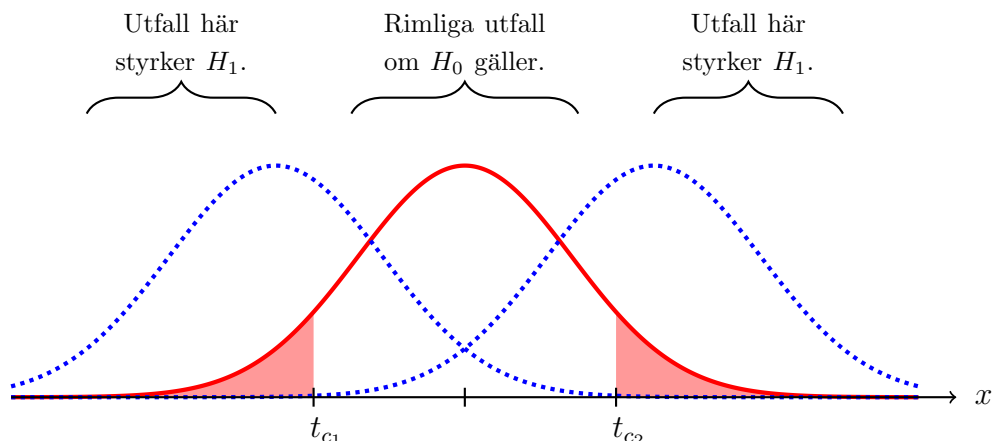


Den röda kurvan är täthetsfunktionen för $t(X_1, \dots, X_n)$ om H_0 skulle vara sann medan den blå är den verkliga täthetsfunktionen. Vi ser att observerade värden är betydligt rimligare i det kritiska området om den blå fördelningen gäller. Det kritiska området blir således

$$C = \{x \in \mathbf{R} : x > t_c\}.$$

Om $t > t_c$ så förkastar vi H_0 .

Om vi istället skulle testa $H_0 : \mu = \mu_0$ mot $H_1 : \mu \neq \mu_0$, vad blir skillnaden? Vi vill således i detta läge styrka att det verkliga väntevärdet är något annat än μ_0 (inte nödvändigtvis att det verkliga väntevärdet är större).



De blå kurvorna är potentiella verkliga fördelningar för $t(X_1, \dots, X_n)$ medan den röda fortfarande är fördelningen om H_0 skulle vara sann. Det kritiska området blir således

$$C = \{x \in \mathbf{R} : x > t_{c_2} \text{ eller } x < t_{c_1}\}.$$

Om $t > t_{c_2}$ eller om $t < t_{c_1}$ så förkastar vi H_0 . När vi vet mer om fördelningen för $t(X_1, \dots, X_n)$ kan vi under antagandet att H_0 stämmer hitta gränserna explicit.



Att ställa upp H_0 och H_1 ska göras *innan* stickprov observerats. Utgår man från mätdata för att hitta på sina hypoteser betar man sig bedrägligt.

9.1.2 Styrka, fel och p-värde

Så säg att vi har valt en teststorhet och ett kritiskt område. Vi har då en metod för att förkasta nollhypotesen om teststorheten sticker ut för mycket från vad som är förväntat om nollhypotesen är sann. Om vi då vet vad det verkliga värdet på parametern är, vad blir sannolikheten för att vi kommer att förkasta nollhypotesen? Idealiskt vore den sannolikheten 1 så fort parametern har ett annat värde än vad som angavs i nollhypotesen (dvs θ_0). Detta blir dock svårt att uppfylla, men hur ser sannolikheten ut för att korrekt förkasta H_0 om vi låter θ variera? Detta brukar kallas för testets **styrka**.



Styrka

Definition. Vi definierar **styrkefunktionen** $h(\theta)$ enligt

$$h(\theta) = P(H_0 \text{ förkastas} \mid \theta \text{ är det riktiga värdet}).$$

Sannolikheten $h(\theta)$ kallas för testets styrka i θ .

För ett bra hypotestest bör $h(\theta)$ vara stor för $\theta \in H_1$ och $h(\theta)$ liten för $\theta \in H_0$. Notera även att $h(\theta_0) = \alpha$. Vi kommer inte att gå in så mycket på styrkan hos tester i denna kurs men det kan vara bra att ha sett begreppet.

Uppenbarligen finns det en risk att vi tar fel beslut. Denna risk kan delas upp i två olika typer.



Fel av typ I och II

Definition. Att förkasta H_0 då H_0 är sann kallas **fel av typ I** och har sannolikheten α . Risken för ett fel av typ I är således signifikansnivån. Att *inte* förkasta H_0 då H_0 är falsk kallas för **fel av typ II**.

Antag att Bill har lite dålig koll och är ute och letar ormar. När han hittar ett djur så är Bills nollhypotes H_0 : Ormen är giftig.

Fel av typ I



Ormen är inte giftig!

Fel av typ II



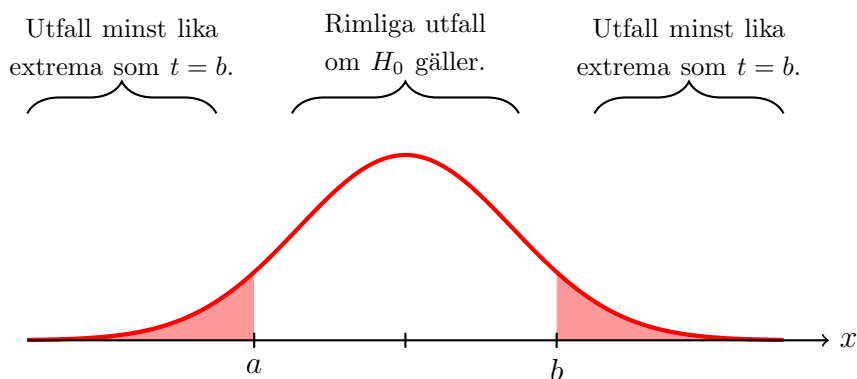
Ormen är giftig!



p -värde

Definition. För ett givet stickprov kan man för ett signifikanstest beräkna ett p -värde. Denna sannolikhet är den lägsta signifikansnivån på vilken vi skulle förkasta H_0 . Med andra ord är p sannolikheten att vi får ett minst lika extremt utfall som det givna stickprovet med antagandet att H_0 är sann.

Låt oss testa $H_0: \theta = \theta_0$ mot $H_1: \theta \neq \theta_0$. Om vi utifrån stickprovet beräknar teststorheten $t(x_1, \dots, x_n) = b$ så behöver vi alltså karakterisera alla utfall som är minst lika extrema om H_0 gäller. Nu blir vi beroende av hur fördelningen ser ut. Låt oss anta något symmetriskt.



Så p -värdet kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$p = P(t(X_1, \dots, X_n) \leq a) + P(t(X_1, \dots, X_n) \geq b) = 2P(t(X_1, \dots, X_n) \geq b),$$

där a måste väljas så vi har samma sannolikhetsmassa i båda "svansarna." Om fördelningen har en riktig skum uppsyn då? Ja, då blir det svårt. En variation vi kan hantera är om mothypotesen är av typen $H_1: \theta > \theta_0$ (till exempel) då vi endast har

$$p = P(t(X_1, \dots, X_n) \geq b)$$

eftersom utfall i vänstra svansen nu inte längre räknas som extrema. Utseendet på mothypotesen är alltså fundamentalt.



Märk väl att p -värdet inte säger någonting om huruvida H_0 är sann eller ej givet observationen av t . Det vi har är sannolikheten för ett lika extremt utfall *givet* att H_0 gäller. *Inte* tvärtom!

Alla principfigurer ovan har varit små söta symmetriska och kontinuerliga historier. Hur blir det vid andra typer av fördelningar?

9.2 Hypotestest för Binomialfördelning

Vi undersöker situationen med ett belysande exempel.



Exempel

Ett mynt kastas (oberoende) 30 gånger och vid 10 av dessa blir det en krona. Kan vi förkasta hypotesen att myntet är ärligt med signifikansnivå 5%?

Lösning. Vi vill testa om myntet är ärligt, så vi börjar med att ställa upp en modell. Låt X vara antalet krona vid 30 kast. Då är $X \sim \text{Bin}(n, p)$ där $n = 30$ och $p =$ sannolikheten för krona är okänd. Så en rimlig nollhypotes ges av

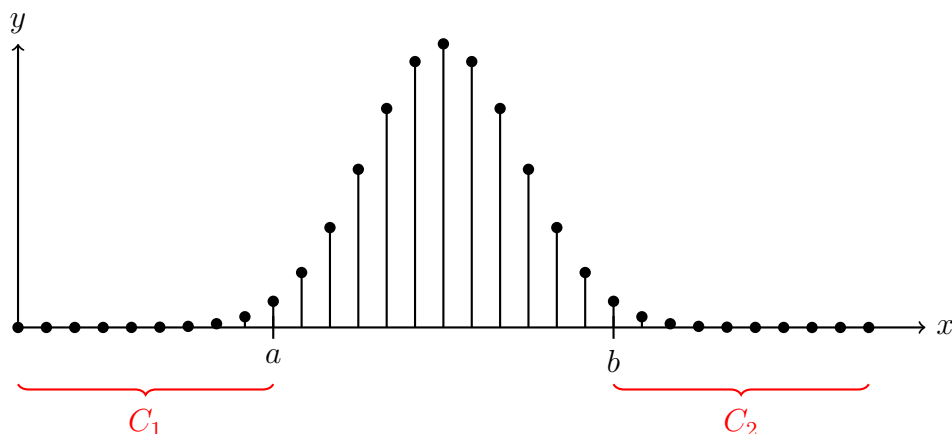
$$H_0 : p = \frac{1}{2}$$

och innan experimentet vet vi inte om mothypotesen bör vara $p < 1/2$ eller $p > 1/2$, så vi tar det säkra före det osäkra och väljer att testa mot

$$H_1 : p \neq \frac{1}{2}.$$

Givet att H_0 är sann så förväntar vi oss frekvensen $30 \cdot 0.5 = 15$ utfall som är krona. Är 10 signifikant mindre? Vi ställer upp det kritiska området:

$$C = \{x \in \mathbf{Z} : 0 \leq x \leq a \text{ eller } b \leq x \leq n\}$$



Hur hittar vi a och b ? Vi får helt enkelt testa oss fram (och använda tabeller). Eftersom

$$p(x) = \binom{30}{x} \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{30-x}$$

kan vi beräkna att

$$\sum_{x=0}^9 p(x) = 0.0214 \quad \text{och} \quad \sum_{x=0}^{10} p(x) = 0.0494$$

samt (känt redan pga symmetri då $p = 0.5$ men för fullständighetens skull):

$$\sum_{x=21}^{30} p(x) = 0.0214 \quad \text{och} \quad \sum_{x=20}^{30} p(x) = 0.0494.$$

Vi väljer $a = 9$ och $b = 21$. Då gäller att

$$\begin{aligned} P(X \in C | H_0) &= P(X \in C_1 | H_0) + P(X \in C_2 | H_0) \\ &= P(X \leq a) + P(X \geq b) = 0.0214 + 0.0214 = 0.0428 < 0.05. \end{aligned}$$

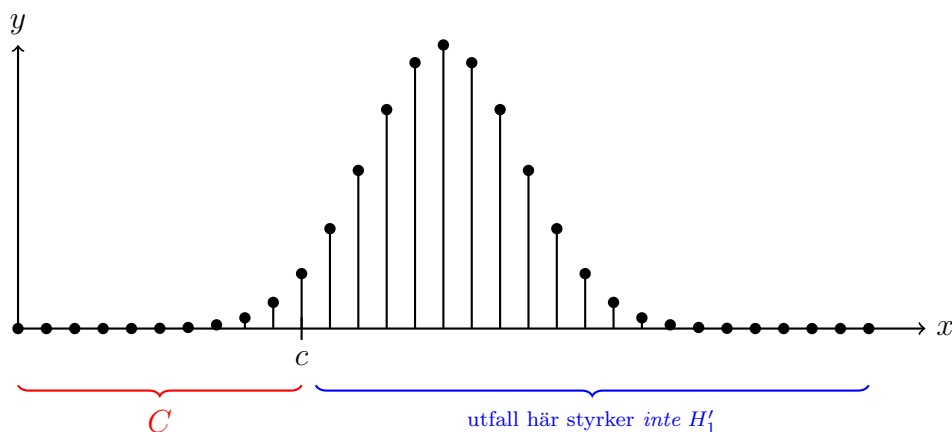
Detta är det största kritiska område vi kan få för att hålla signifikansnivån. Observera att vi alltså inte kan träffa $\alpha = 0.05$ exakt. Detta är typiskt vid diskreta fördelningar.

Eftersom $x = 10 \notin C$ kan vi inte dra någon slutsats, utan myntet kan mycket väl vara ärligt. Vi kan således *inte* förkasta H_0 (vilket inte på något sätt betyder att H_0 är sann).

Antag att vi istället vill testa mothypotesen H'_1 att myntet ger färre krona än klave. Vi har då

$$H'_1 : p < \frac{1}{2}.$$

Hur ser det kritiska området C ut?



Eftersom

$$\sum_{x=0}^{10} p(x) = 0.0494 \quad \text{och} \quad \sum_{x=0}^{11} p(x) = 0.1002$$

så ser vi att $c = 10$ är nödvändigt. Därmed blir

$$C = \{x \in \mathbf{Z} : 0 \leq x \leq 10\}$$

och vår observation $x = 10 \in C$. Alltså kan vi förkasta H_0 och anse att H'_1 är styrkt. Slutsats?

9.3 Normalapproximation – Generellt

När vi approximerar med normalfördelningen är tillvägagångssättet nästan alltid det samma. Vi har en punktskattning $\hat{\theta}$ där $\hat{\Theta} \stackrel{\text{appr.}}{\sim} N(\theta, D^2)$ och vi vill testa nollhypotesen $H_0 : \theta = \theta_0$. Som teststorhet använder vi då oftast

$$Z = \frac{\hat{\Theta} - \theta_0}{D} \quad \text{eller} \quad Z = \frac{\hat{\Theta} - \theta_0}{d}.$$

Den senare teststorheten då vi inte känner D exakt utan skattar med d . Vi förutsätter att d är en vettig skattning av D då H_0 är sann. Notera att i båda fallen kommer $Z \stackrel{\text{appr.}}{\sim} N(0, 1)$ om H_0 är sann. Vi använder alltså ingen t -fördelning här (det finns inget som säger att det skulle bli bättre i det generella fallet).

Hur det kritiska området ser ut beror på hur vi ställer upp mothypotesen. Om $H_1 : \theta \neq \theta_0$ får C utseendet $]-\infty, -a[\cup]a, \infty[$. Är mothypotesen enkelsidig blir det bara ett av intervallen (med annan parameter a). Talet a hittar vi i normalfördelningstabell.

9.4 Test för skillnad i andel

En mycket vanlig situation är att vi vill undersöka om det föreligger någon skillnad i andel mellan två grupper. Antag att vi har x_1 som observation av $X_1 \sim \text{Bin}(n_1, p_1)$ och x_2 som observation av $X_2 \sim \text{Bin}(n_2, p_2)$ (vi antar oberoende).

Vi är intresserade av att testa hypotesen $H_0 : p_1 = p_2$ mot till exempel $H_1 : p_1 \neq p_2$. Om H_0 är sann så är en lämplig skattning av $p = p_1 = p_2$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

Faktum är att detta är ML-skattningen (om H_0 är sann) och därmed har den bra egenskaper såsom konsistens. Vad gäller fördelningen för \hat{P} blir den värre (vad händer om man summerar binomialfördelningar?). Men, om n_1 och n_2 är ganska stora och p inte är allt för nära ändpunkterna i $[0, 1]$, så kanske vi kan normalapproximera? Vi har redan gjort detta (se konfidensintervall för $p_1 - p_2$), men för fullständighetens skull låt oss repetera. Om H_0 är sann gäller att

$$E(\hat{P}) = \frac{n_1 p + n_2 p}{n_1 + n_2} = p$$

och

$$V(\hat{P}) = \frac{n_1 p(1-p) + n_2 p(1-p)}{(n_1 + n_2)^2} \rightarrow 0,$$

då $n_1 + n_2 \rightarrow \infty$, så skattningen av p är väntevärdesriktig och konsistent. För att testa H_0 använder vi $\hat{P}_1 - \hat{P}_2$, och om H_0 är sann så gäller att

$$\hat{P}_1 - \hat{P}_2 \stackrel{\text{appr.}}{\sim} N\left(0, \hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

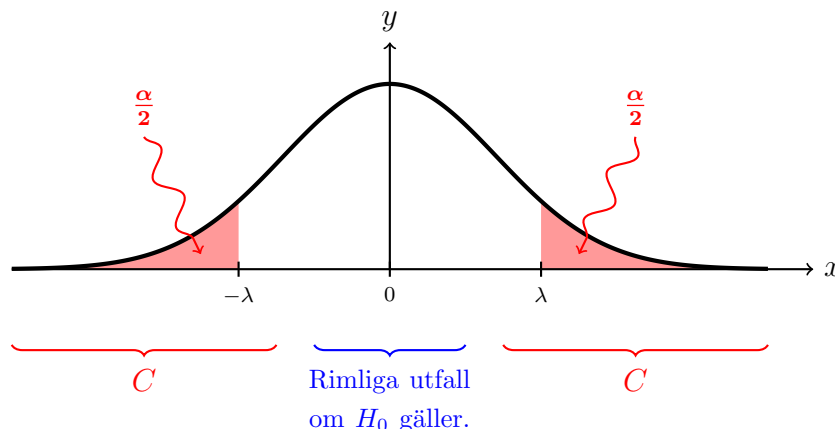
Eftersom vi inte känner p exakt använder vi skattningen \hat{p} ovan i uttrycket för variansen (eller vi ersätter standardavvikelsen med medelfelet). Vi kan även gå över i standardiserad form så vi känner igen oss:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Om $H_1 : p_1 \neq p_2$ så ges det kritiska området av

$$C = \{z \in \mathbf{R} : |z| > \lambda\}$$

för något lämpligt $\lambda = \Phi^{-1}(1 - \alpha/2)$ vi finner ur tabell (eller MATLAB).



Exempel

Två opinionsinstitut Analysera Mera AB och StickProvarna AB undersöker om befolkningen tycker att sommaren varit för varm. AM frågar 500 personer och andelen $p_1 = 0.7$ (350 st) håller med. SP frågar 400 personer och $p_2 = 0.8$ (320 stycken) håller med. Undersök om det finns någon signifikant skillnad mellan resultaten på signifikansnivån 5% (approximativt).

Lösning. Låt $H_0 : p_1 = p_2 = p$ och $H_1 : p_1 \neq p_2$. Om H_0 är sann väljer vi skattningen

$$\hat{p} = (350 + 320)/(500 + 400) = 0.744.$$

Med beteckningarna ovan gäller då (om H_0 är sann) att

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{500} + \frac{1}{400} \right)}} = \frac{\hat{P}_1 - \hat{P}_2}{0.0293} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Det är rimligt att approximera både \hat{P}_1 och \hat{P}_2 med normalfördelning eftersom både $500 \cdot 0.7 \cdot 0.3 \geq 10$ och $400 \cdot 0.8 \cdot 0.2 \geq 10$. Vi hittar det kritiska området

$$C = \{z \in \mathbf{R} : |z| > \lambda\}$$

där $\lambda = \Phi^{-1}(0.975) = 1.96$. Således ska – om H_0 är sann –

$$\left| \frac{\hat{p}_1 - \hat{p}_2}{0.0293} \right| > 1.96 \quad \Leftrightarrow \quad |\hat{p}_1 - \hat{p}_2| > 1.96 \cdot 0.0293 = 0.0573$$

för att vi ska förkasta H_0 . Med $\hat{p}_1 = 0.7$ och $\hat{p}_2 = 0.8$ ser vi att $0.1 > 0.0573$, så vi förkastar H_0 . Det är troligen en skillnad i resultaten.

Ett alternativ är att ställa upp konfidensintervallet $I_{p_1 - p_2}$ för $p_1 - p_2$ och sedan testa hypotesen genom att undersöka om $0 \in I_{p_1 - p_2}$. Skulle det vara så att 0:an ingår kan vi inte förkasta H_0 . Ligger intervallet helt på ena sidan 0 däremot så förkastar vi H_0 . Detta test är helt ekvivalent eftersom vi nyttjar samma testvariabel.

9.5 Poissonapproximation

Som bekant kan man även approximera binomialfördelning med Poissonfördelning om $n \geq 10$ och $p \leq 0.1$. Detta kan vara nödvändigt då p ligger nära 0 eller 1 så normalapproximation inte fungerar bra. Vi betraktar ett exempel.



Exempel

En leverantör av laboratorieutrustning hävdar att deras pipetter bara behöver kalibreras en gång per år och att risken för att en pipett faller utanför toleransnivån innan dess är 0.5% (vid normal användning). Laboratorieansvarig Laura (för ett stort laboratorium) tycker inte att det stämmer och har ett år efter inköpet och kontinuerligt användande av 1000 stycken behövt kalibrera om 11 st. Testa hypotesen att felrisken är 0.5% mot att den är högre på signifikansnivån 1% (approximativt).

Lösning. Den stokastiska variabeln X är antalet av de 1000 pipetterna som behövs kalibreras i förtid. Om vi antar att händelserna är oberoende (är det rimligt?) så är $X \sim \text{Bin}(1000, p)$ där p är felrisken. Låt $H_0 : p = 0.005$ och $H_1 : p > 0.005$. Vi kan använda $\hat{P} = \frac{X}{1000}$, men enklare är att direkt nyttja X . Om H_0 är sann så gäller att

$$X \stackrel{\text{appr.}}{\sim} \text{Po}(1000 \cdot 0.005) = \text{Po}(5).$$

Det kritiska området väljs som

$$C = \{z \in \mathbf{Z} : z > k\}$$

för något $k \in \mathbf{Z}$. Vi vill att

$$P(X \in C | H_0) \leq 0.01$$

och i tabell (eller med $\mathbf{k} = \text{poissinv}(0.99, 5)$ i MATLAB, vilket ger det minsta heltalet k så att $P(X \leq k) \geq 0.99$) finner vi att $k = 11$. Alltså gäller

$$P(X > 11 | H_0) < 0.01 \quad (\text{exakt värde: } 0.0055),$$

och Lauras observation $x = 11$ är alltså *inte* signifikant. Vi kan inte förkasta H_0 och säga att leverantören har fel.



Exempel

Laura är inte nöjd och kräver att examensarbetaren Audrey ska göra om hypotestestet och använda normalapproximation *som folk*. Motivera varför det inte är bra men utför testet. Undersök också hur hypotestestet blir om man inte approximerar för att hjälpa den stackars examensarbetaren att motivera.

Lösning. Vid normalapproximation kräver vi att $np(1-p) \geq 10$ och om vi väljer att skatta p med $\hat{p} = 10/1000 = 0.01$ hamnar vi precis kring den gränsen så osäkerheten är stor. Använder vi leverantörens $p = 0.005$ blir det betydligt under. Alltså inget att rekommendera. Men om vi envisas så skulle

$$X \stackrel{\text{appr.}}{\sim} N(1000p, 1000p(1-p)) \quad \text{som dålig approximation.}$$

Om vi antar att H_0 är sann skulle då

$$Z = \frac{X - 1000 \cdot 0.005}{\sqrt{1000 \cdot 0.005 \cdot (1 - 0.005)}} \stackrel{\text{appr.}}{\sim} N(0, 1),$$

återigen som en tveksam approximation. Kritiskt område ges av

$$0.01 = P(Z > \lambda) = 1 - \Phi(\lambda) \quad \Leftrightarrow \quad \lambda = \Phi^{-1}(0.99) = 2.3263$$

så

$$\frac{X - 5}{\sqrt{4.975}} > 2.3263 \quad \Leftrightarrow \quad X > 10.1888$$

och vi skulle därför låta C ges av $X \geq 11$, varvid resultatet $x = 11$ skulle verka signifikant.

Vi kan ställa upp ett exakt test genom att låta $H_1 : p > 0.005$ och välja

$$C = \{x \in \mathbf{Z} : x > k\}$$

för något $k \in \mathbf{Z}$. Precis som med Poissonapproximationen hittar vi k genom att i MATLAB använda `k = binoinv(0.99, 1000, 0.005)` vilket resulterar i $k = 11$. Alltså samma gräns som vi fick med Poissonapproximationen. Exakt värde här blir $P(X > k) = 0.0053$.

9.6 Hypotestest för väntevärde vid normalfördelning

Låt oss undersöka den generella situationen vid ett stickprov X_1, X_2, \dots, X_n från en normalfördelning $N(\mu, \sigma)$. Vi kan tänka oss att

$$X_i = \mu + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma), \quad i = 1, 2, \dots, n,$$

där ϵ_i är oberoende. Notera att samtliga variabler har samma varians. Som föregående avsnitt visade är det skillnad på när vi känner variansen exakt och när den behöver skattas.

Vi börjar med att testa

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu \neq \mu_0.$$

Givetvis kan man vilja testa mot $H'_1 : \mu > \mu_0$ eller $H''_1 : \mu < \mu_0$ också, vi kommer ta upp något sådant exempel senare.

9.6.1 Känd varians

Eftersom

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

kan vi när σ är känd direkt använda \bar{X} som teststorhet. Men för att göra det hela systematiskt och analogt med fallet då σ inte är känd skapar vi en testvariabel

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1),$$

där fördelningen gäller under förutsättning att H_0 är sann. Vad vi egentligen gör är att vi utnyttjar att \bar{X} är en skattning (konsistent och väntevärdesriktig) av (det okända) väntevärdet μ . Testet går ut på att se om det uppmätta värdet på skattningen sticker ut så mycket från vad som är förväntat att det gör H_0 orimlig.

Det kritiska området C ges av

$$P(Z \in C \mid H_0) = \alpha$$

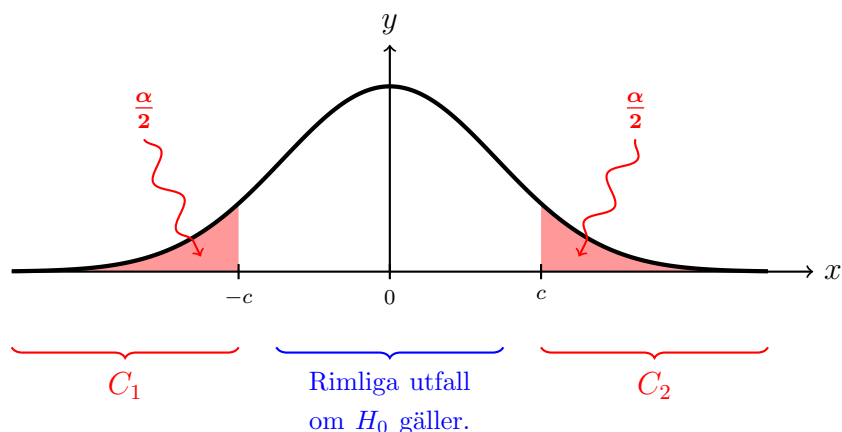
där vi av symmetriskäl (eftersom $Z \sim N(0, 1)$) kan – för något $c > 0$ – uttrycka C enligt

$$C = \{z \in \mathbf{R} : |z| > c\} = \{z \in \mathbf{R} : z > c \text{ eller } z < -c\}$$

Vi noterar att C består av två delar C_1 och C_2 där talen i C_1 är negativa och talen i C_2 är positiva. Återigen, av symmetriskäl måste

$$P(Z \in C_1) = P(Z \in C_2) = \frac{\alpha}{2}.$$

Gränsen hittar vi i tabell genom att leta reda på ett tal $c = \Phi^{-1}(1 - \alpha/2)$ (sitter du med MATLAB kan du använda `c = -norminv(alpha/2)`).



Approximativt test via CGS

Som vi såg på förra föreläsningen kan man använda approximationer för att utföra hypotestest. Om vi i vårt fall inte vet att X_i är normalfördelad kan vi ändå via centrala gränsvärdessatsen säga att

$$\bar{X} \stackrel{\text{appr.}}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

om $n \geq 30$ (lite beroende på hur skev fördelningen för X_i är). Som teststorhet använder vi sedan

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Notera att vi ersätter σ med s utan att förändra fördelningen (eftersom vi redan håller på med approximationer vet vi inte om det blir bättre med t -fördelningen). Faktum är att vi kan göra detta även om variablerna är lite beroende. Det finns flera varianter av CGS som kan hantera lite olika situationer.

9.6.2 Okänd varians

Om vi inte känner till σ så kan vi inte direkt använda \bar{X} som teststorhet och inte heller Z från föregående stycke fungerar utan problem (vad ska vi göra med den okända storheten σ ?). Visserligen kan vi se det approximativt via CGS, men vad som vore bättre är att ersätta σ^2 med stickprovsvariansen s^2 . Som vi sett tidigare leder det till t -fördelningen. Så, då gäller att

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

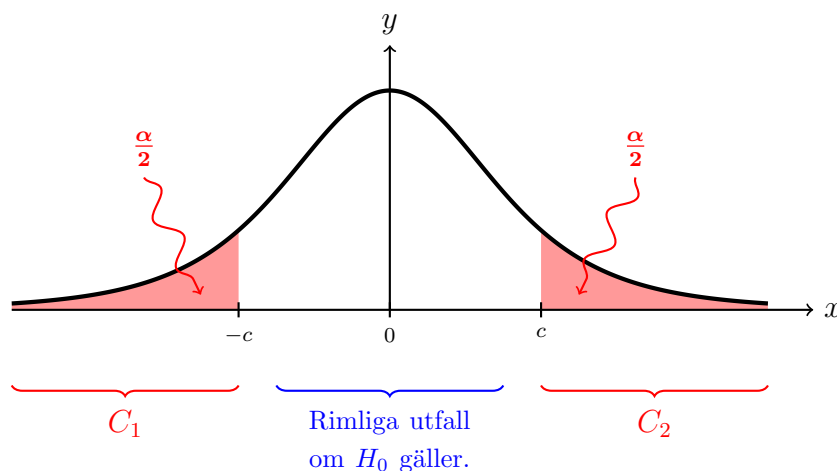
om H_0 är sann. Helt analogt med föregående situation erhåller vi nu

$$C = \{t \in \mathbf{R} : t > c \text{ eller } t < -c\}$$

där

$$P(T > c) = P(T < -c) = \frac{\alpha}{2}.$$

Gränsen hittar vi i tabell genom att leta reda på ett tal $c = F_T^{-1}(1 - \alpha/2)$ (sitter du med MATLAB kan du använda $c = -\text{tinv}(\alpha/2)$).



9.7 Hypotestester och konfidensintervall

Den observante läsaren har nog redan reflekterat över att det vi ägnat oss åt är ganska snarlikt de föregående föreläsningarna om konfidensintervall. Vi ställer upp liknande storheter (ja, identiska för det mesta) men istället för att stänga in något okänt i ett intervall så testar vi skattningen mot ett kritiskt område.

Ett annat sätt att testa hypoteserna på är att ställa upp konfidensintervall och sedan testa om intervallet täcker nollhypotesen eller ej.



Exempel

I ett husvagnsdrivet laboratorie kokar Janne ihop den suspekta kemikalien $C_{10}H_{15}N$. Varje vecka startar han med samma mängd utgångsmaterial och följer samma procedur. Janne har fått en ny köpare och har hävdade att han kan producera 500 gram i veckan. För att inte riskera problem med hälsan vill Janne testa hypotesen $H_0 : \mu = 500$ mot $H_1 : \mu > 500$ på signifikansnivån 1%. Under 16 veckor producerar han i snitt 525 gram med stickprovsstandardavvikelsen 30 gram.

Lösning. Vi antar normalfördelning och ställer upp ett enkelsidigt konfidensintervall för väntevärdet μ . Låt

$$T = \frac{\bar{X} - \mu}{S/\sqrt{16}} \sim t(15).$$

Då gäller att

$$P(T < t) = 0.99$$

om $t = 2.6025$ (ur tabell). Eftersom

$$T < t \Leftrightarrow \frac{\bar{X} - \mu}{S/4} < t \Leftrightarrow \frac{St}{4} > \bar{X} - \mu \Leftrightarrow \mu > \bar{X} - \frac{St}{4}.$$

så erhåller vi konfidensintervallet

$$I_\mu = \left(\bar{x} - \frac{st}{4}, \infty \right) = (505.48, \infty).$$

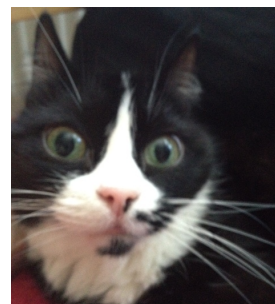
Om H_0 är sann så kommer $\mu = 500 \in I_\mu$ med sannolikheten 99%. Eftersom detta inte är sant kan vi förkasta H_0 . Ska Janne sitta lugnt i båten att han inte lovat för mycket? Mäter Jannes test rätt sak?

9.8 Generellt om hypotestester

Innan vi avslutar med en diskussion och tester vid flera stickprov tar vi och summerar lite att tänka på.

- (i) Formulera hypoteser *innan* du tittar på datan. Att du vill göra ett enkelsidigt test ska inte bero på hur datan ser ut. Av den anledningen är den vanligaste typen av tester två-sidiga.
- (ii) Men vissa situationer är alltid enkelsida på grund av konstruktion. Vi kommer se det i samband med regressionsanalysen där vi till exempel vet att varians minskar med fler förklaringsvariabler.
- (iii) Kom ihåg när hypotestestet ställs upp att det är *mothypotesen* vi vill styrka.
- (iv) Var *mycket* försiktig med tolkning av resultaten.

Att man inte förkastar H_0 betyder inte att H_0 gäller. Ett klassiskt exempel handlar om fyrbenta djur: låt H_0 : djuret har fyra ben och H_1 : djuret har inte fyra ben. Vi vill undersöka om en observation är en häst och testar H_0 mot H_1 . Bara för att vi inte kan förkasta H_0 när djuret är en katt betyder det inte att det är en häst, eller hur? Kanske ett urartat exempel, det kan vara betydligt mer diffust att läsa av resultaten rätt i andra fall.



- (v) Signifikansnivån kan också vara missvisande. Vid stora stickprov kan man ofta se en skillnad och förkasta H_0 även om skillnaden kanske inte spelar någon större roll i praktiska fall.

9.9 Flera stickprov

Vi kan givetvis betrakta flera stickprov samtidigt. Ofta är man intresserad av att testa om de har samma väntevärde och/eller samma varians. Men vi kan ställa upp test för linjärkombinationer av väntevärdena direkt.

Låt X_1, X_2, \dots, X_m och Y_1, Y_2, \dots, Y_n vara stickprov från $N(\mu_X, \sigma_X)$ respektive $N(\mu_Y, \sigma_Y)$. Vi kan ställa upp hypotestester för linjärkombinationen $c_1\mu_X + c_2\mu_Y$. Om varianserna är kända kan vi direkt använda att

$$Z = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_X + c_2\mu_Y)}{\sqrt{c_1^2\sigma_X^2/m + c_2^2\sigma_Y^2/n}} \sim N(0, 1),$$

alltså precis samma variabel vi såg när vi tog fram konfidensintervall för $c_1\mu_X + c_2\mu_Y$.

9.9.1 $\sigma_X = \sigma_Y = \sigma$ okänd

Helt analogt med motsvarande situation när vi tog fram konfidensintervall använder vi att

$$T = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_X + c_2\mu_Y)}{S\sqrt{c_1^2/m + c_2^2/n}} \sim t(m+n-2),$$

där S^2 är den sammanvägda variansskattningen. Vi betraktar ett exempel.



Exempel

Janne har fått konkurens av den före detta lärlingen Rossana som använder samma metod. Under 9 veckor producerar hon i snitt 600 gram med en stickprovsstandardavvikelse på 50 gram. Testa på signifikansnivån 1% hypotesen $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 < \mu_2$ med antagandet att variansen är densamma, där μ_1 är Jannes förväntade värde och μ_2 är Rossanas. Borde köparen byta leverantör?

Lösning. Vi formulerar om enligt $H_0 : \mu_2 - \mu_1 = 0$ mot $H_1 : \mu_2 - \mu_1 > 0$. Om H_0 är sann så gäller att

$$T = \frac{\bar{Y} - \bar{X}}{S\sqrt{1/9 + 1/16}} = \frac{\bar{Y} - \bar{X}}{0.4167 S} \sim t(23),$$

och det kritiska området blir

$$C = \{t \in \mathbf{R} : t > 2.4999\}$$

eftersom $P(T < 2.4999) = 0.99$. Med uppmätta siffrorna blir

$$t = \frac{600 - 525}{0.4167 \cdot s_p} = \frac{75}{0.4167 \cdot 38.16} = 4.7161,$$

där

$$s_p^2 = \frac{15 s_1^2 + 8 s_2^2}{23} = 38.16^2$$

är den sammanvägda variansskattningen. Eftersom $4.7161 \in C$ förkastar vi H_0 . Blir det samma resultat om vi testar mot $H_1 : \mu_1 \neq \mu_2$? (svar: ja, förkasta H_0 . Vad ändras?)

Vi kan givetvis ta fram ett konfidensintervall $I_{\mu_2 - \mu_1}$ och testa nollhypotesen genom att undersöka om $0 \in I_{\mu_2 - \mu_1}$ också (förkasta H_0 om $0 \notin I_{\mu_2 - \mu_1}$).

9.10 (★)Hypotestest för Poissonfördelning

Övriga diskreta fördelningar kan givetvis hanteras analogt med binomialexemplet i föregående avsnitt och med datorkraft är det inte större problem att räkna exakt i väldigt många fall. Men som vi kommer ihåg från tidigare kurser går det även att approximera flera diskreta fördelningar med normalfördelning om vissa förutsättningar är uppfyllda. Låt oss studera ett exempel med Poissonfördelning på två sätt.



Exempel

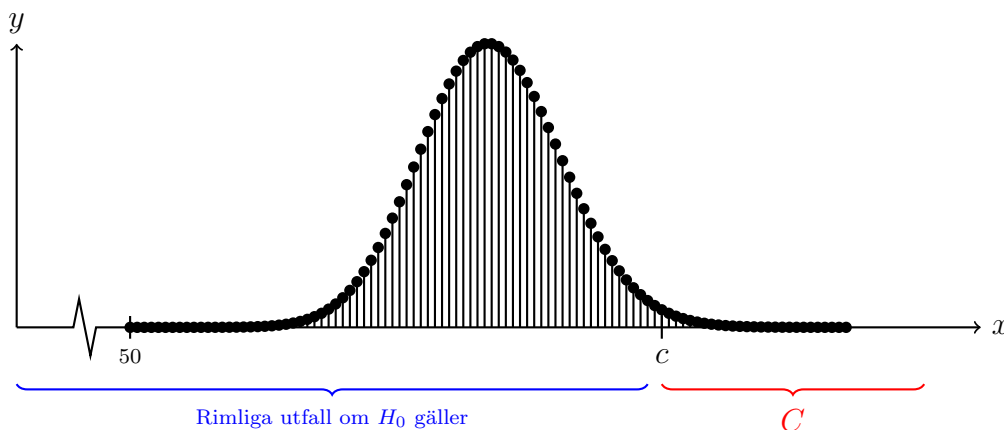
Antalet datapaket till en server kan betraktas som en Poissonprocess $X(t)$ med en okänd intensitet λ . För att kunna hantera överbelastning har man ett varningssystem som varnar om antalet paket överstiger en gräns N på två tidsenheter. Varningen sker alltså om intensiteten är större än väntat. Antag att $\lambda = 50$ (enhet: tusen paket). Det är dyrt att avbryta servicen så man vill högst tillåta felaktig varning med 1% risk.

Hitta gränsen N och avgör om man bör varna om $x = 120$ vid en mätning. Vad skulle p -värdet bli om $x = 130$?

Lösning. Det förväntade antalet paket är $\mu = E(X(t)) = \lambda t$, så om $\lambda = 50$ förväntar vi oss $\mu = 50 \cdot 2 = 100$ (tusen) paket. Låt

$$H_0 : \mu = 100 \quad \text{och} \quad H_1 : \mu > 100.$$

Vi söker det kritiska området C . En figur kan vara bra.



Låt $p(k)$, $k = 0, 1, 2, \dots$, vara sannolikhetsfunktionen för en $\text{Po}(100)$ -fördelad variabel. Ur tabell (eller med hjälp av matlab och funktionerna `poisspdf` eller `poisscdf`) kan vi finna att

$$\sum_{k=124}^{\infty} p(k) = 1 - \sum_{k=0}^{123} p(k) = 0.0112 \quad \text{och} \quad \sum_{k=125}^{\infty} p(k) = 0.0088.$$

Således blir det kritiska området

$$C = \{k \in \mathbf{Z} : k \geq 125\}.$$

Eftersom observationen $x = 120 \notin C$ så kan vi inte förkasta H_0 . Vi bör inte varna.

Vi beräknar p -värdet vid observationen $x = 130$ genom

$$p = P(X \geq 130 | H_0) = \sum_{k=130}^{\infty} p(k) = \left/ \text{tabell} \right/ = 0.0023.$$

Vi summerar alltså sannolikheterna för alla utfall som är minst lika extrema som $x = 130$.

Om vi stirrar lite på plotten ovan så ser den tämligen normalfördelad ut, eller hur? Det är ingen slump. Om $X \sim \text{Po}(\mu)$ med $\mu \geq 15$ så är $X \stackrel{\text{appr.}}{\sim} N(\mu, \sqrt{\mu})$ (variansen är μ). Vi kan använda detta för att hitta en approximativ gräns N . Låt $X \sim \text{Po}(100)$. Då gäller att

$$0.01 = P(X \geq N) = 1 - P(X < N) = 1 - P\left(\frac{X - 100}{\sqrt{100}} < \frac{N - 100}{\sqrt{100}}\right) = 1 - \Phi\left(\frac{N - 100}{10}\right).$$

Således är

$$\begin{aligned} 0.01 = 1 - \Phi\left(\frac{N - 100}{10}\right) &\Leftrightarrow 0.99 = \Phi\left(\frac{N - 100}{10}\right) \\ &\Leftrightarrow 2.3263 = \frac{N - 100}{10} \\ &\Leftrightarrow N = 23.263 + 100 = 123.263. \end{aligned}$$

Eftersom N måste vara ett heltal väljer vi $N = 124$. Även med halvtstegskorrigering hamnar vi inte på det exakta värdet, men det är tillräckligt nära för de flesta ändamål. Vi kan även återskapa kalkylen för p -värdet vid $x = 130$ enligt

$$p \approx 1 - \Phi\left(\frac{130 - 100}{10}\right) = 0.0013.$$

9.11 (★)Plåtfabriken

I detta avsnitt följer en sekvens av exempel som kanske illustrerar en del av föregående material.



Exempel

I en fabrik med mångårig erfarenhet tillverkar man material av en viss tjocklek. Man mäter med jämna mellanrum tjockleken på 9 nytillverkade material och testar om medelvärdet uppfyller $|\bar{x} - 5.0| > 0.05$. Om så är fallet stoppas tillverkningen och tekniker får gå igenom maskineriet. Ansvarig för metodutvecklingen vet av erfarenhet att $\sigma = 0.1$. Om vi antar normalfördelning, vad är bästa signifikansnivån för testet om $H_0 : \mu = 5$ testas med mothypotesen $H_1 : \mu \neq 5$?

Lösning. Vi antar att $X_i \sim N(\mu, \sigma) = N(\mu, 0.1)$, $i = 1, 2, \dots, 9$, är oberoende. För att testa H_0 ställer vi upp teststorheten

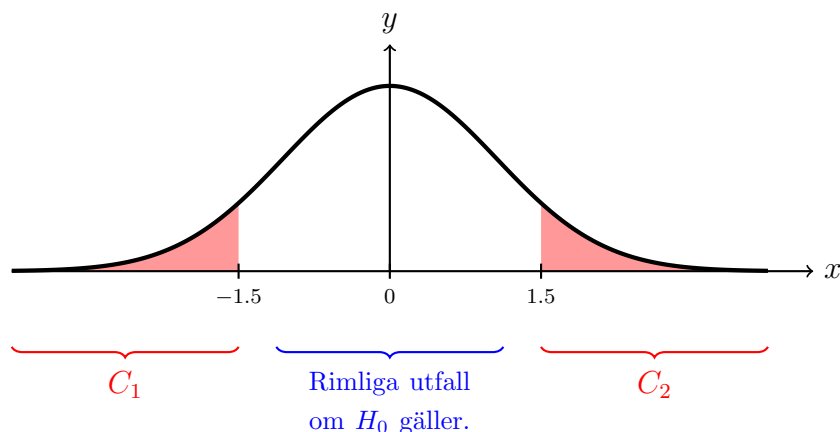
$$Z = \frac{\bar{X} - 5.0}{0.1/\sqrt{9}}.$$

Om H_0 är sann så är $Z \sim N(0, 1)$. Om vi jämför med fabriken test så ser vi att

$$\left| \frac{0.1}{3} Z \right| > 0.05 \Leftrightarrow |Z| > 1.5$$

ger det kritiska området

$$C = \{z \in \mathbf{R} : |z| > 1.5\}.$$



Så signifikansnivån α kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$\begin{aligned} p &= P(Z \leq -1.5) + P(Z \geq 1.5) = 2P(Z \leq -1.5) \\ &= 2\Phi(-1.5) = 2(1 - \Phi(1.5)) = 0.1336. \end{aligned}$$

Den bästa signifikansnivån vi kan välja är alltså $\alpha = 0.1336$.



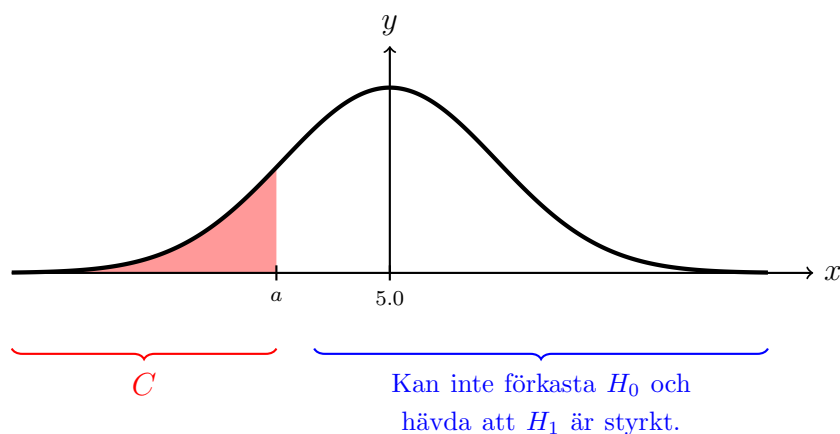
Exempel

Fabriken har fått en ny beställare som inte har något problem om materialet blir tjockare. Ansvarig tänker lite snabbt och ställer upp ett test med samma signifikansnivå för att endast testa att materialet inte blir för tunt. Hur ser testet ut nu och varför är detta antagligen inte vad man vill göra?

Lösning. Vi har fortfarande $H_0 : \mu = 5$ men mothypotesen ges nu av $H_1 : \mu < 5$. Vi kan använda samma teststorhet och om H_0 är sann så är

$$Z = \frac{\bar{X} - 5.0}{0.1/\sqrt{9}} \sim N(0, 1).$$

Vi söker en gräns a så att $\bar{X} < a$ med sannolikheten $\alpha = 0.1336$ om H_0 är sann.



Det är tydligt att

$$\bar{X} < a \quad \Leftrightarrow \quad Z = \frac{\bar{X} - 5.0}{0.1/3} < \frac{a - 5.0}{0.1/3},$$

så

$$\begin{aligned} 0.1336 &= P\left(Z < \frac{a - 5.0}{0.1/3}\right) \Leftrightarrow \frac{a - 5.0}{0.1/3} = \Phi^{-1}(0.1336) = -1.1095 \\ &\Leftrightarrow a - 5.0 = -0.0370. \end{aligned}$$

Det sökta värdet blir alltså $a = 4.9630$. Detta test blir alltså mer känsligt för att materialet är för tunnt än det föregående. Om den nya beställaren har samma tolerans för fel som de tidigare är det kanske mer strategiskt att istället sänka signifikansnivån till hälften.

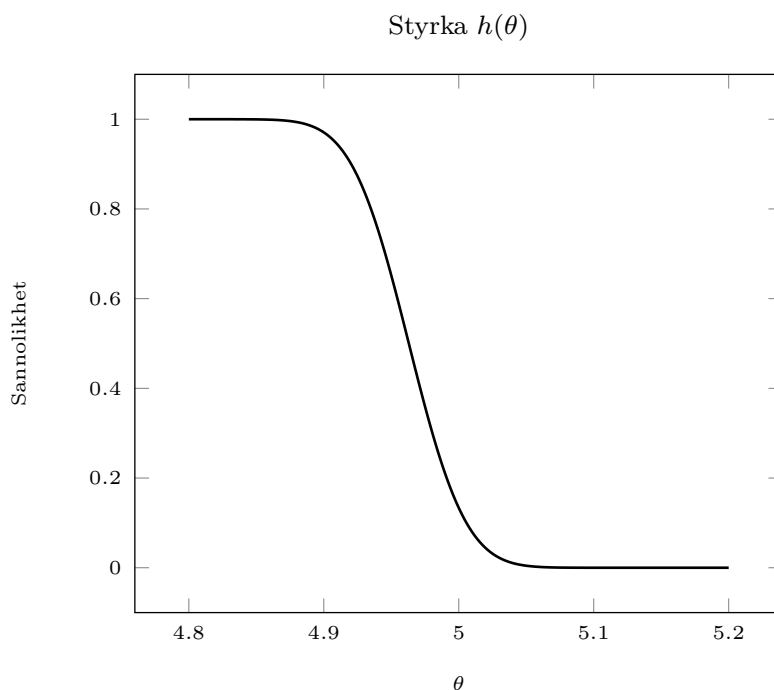


Exempel

Föregående hypotestest har en lite udda signifikansnivå. Hur ser styrkefunktionen ut?

Lösning. Styrkefunktionen definieras enligt

$$\begin{aligned} h(\theta) &= P(H_0 \text{ förkastas} \mid \mu = \theta) = P\left(\bar{X} < 4.9630 \mid \bar{X} \sim N\left(\theta, \frac{0.1^2}{9}\right)\right) \\ &= \Phi\left(\frac{4.9630 - \theta}{0.1/3}\right) = \Phi(148.89 - 30\theta). \end{aligned}$$



Exempel

En nyanställd i fabriken (med en kurs i statistisk inferens i bagaget) påtalar att det kanske är olämpligt att anta att variansen är känd och att man borde skatta den från mätningen. Vid en mätning fick man stickprovsvariansen 0.0144, vad ger testet $|\bar{x} - 5.0| > 0.05$ för signifikansnivå i denna situation?

Lösning. Vi testar således $H_0 : \mu = 5.0$ mot $H_1 : \mu \neq 5.0$ och som testvariabel blir

$$T = \frac{\bar{X} - 5.0}{S/\sqrt{9}} \sim t(8)$$

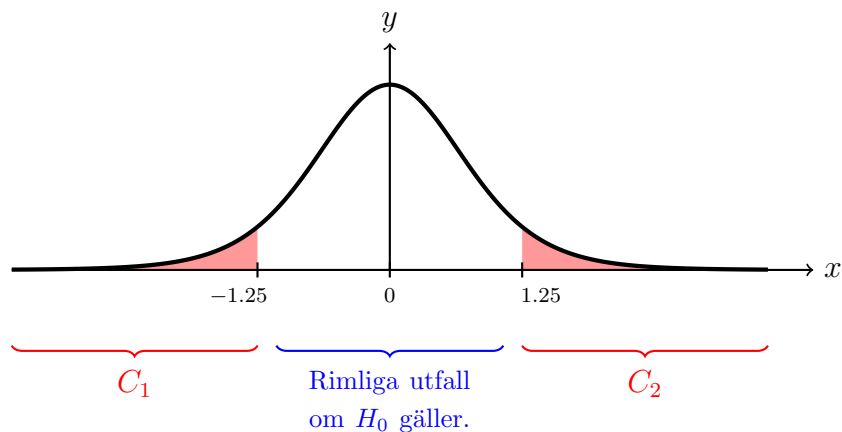
om H_0 är sann. Analogt med första exemplet måste då

$$\left| \frac{s}{3} t \right| > 0.05 \quad \Leftrightarrow \quad |t| > \frac{0.15}{s},$$

vilket ger det kritiska området

$$C = \left\{ t \in \mathbf{R} : |t| > \frac{0.15}{s} \right\} = \{t \in \mathbf{R} : |t| > 1.25\}$$

i vårt fall. Fördelningen för T är symmetrisk lik normalfördelningen, så situationen är snarlik.



Så p -värdet kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$\begin{aligned} p &= P(T \leq -1.25) + P(T \geq 1.25) = 2P(T \leq -1.25) \\ &= 2F_T(-1.25) = 2 \cdot 0.1233 = 0.2466. \end{aligned}$$

Här använde vi `tcdf(-1.25,8)` i MATLAB (vi har inga tabeller i formelsamlingen för att slå på t -fördelningar i "den riktningen").

Den bästa signifikansnivån vi kan välja är alltså i princip $\alpha = 0.25$. Mindre lyckat! Testet kanske behöver ändras.

Kapitel 10

Pearsons χ^2 -test och linjär regression

10.1 Det grundläggande χ^2 -testet

Antag att vi har följande situation

- (i) Vi har n stycken oberoende stokastiska variabler X_j med samma fördelning, där X_j har precis k möjliga utfall.
- (ii) Numrera utfallen enligt A_1, \dots, A_k och låt $p_j = P(A_j)$ vara respektive sannolikhet. Då är $p_1 + p_2 + \dots + p_k = 1$.
- (iii) Låt $Y_i, i = 1, 2, \dots, k$, vara antalet gånger händelsen A_i inträffar.

För att konkretisera en aning, tänk att vi har k stycken lådor A_j vi kastar bollar i. Experimentet är uppställt så att en kastad boll alltid hamnar i en låda. Vi låter p_j vara sannolikheten att en boll hamnar i låda A_j . Vi kastar n bollar (oberoende) och räknar sedan hur många bollar Y_j som det finns i varje låda. Givetvis kommer $Y_j \sim \text{Bin}(n, p_j)$, men variablerna Y_j är *inte* oberoende av varandra (antalet bollar i alla lådorna summerar till n).

Vad vi kommer göra är att betrakta uppdelningar av denna typ och ställa upp hypotestest där vi låter nollhypotesen H_0 ges av

$$H_0 : P(A_1) = p_1, P(A_2) = p_2, \dots, P(A_k) = p_k,$$

där p_1, p_2, \dots, p_k är sannolikheter så att $p_1 + \dots + p_k = 1$, och testar mot hypotesen

$$H_1 : \text{det finns något } j \text{ så att } P(A_j) \neq p_j.$$

Om H_0 är sann, så blir de förväntade frekvenserna $E(Y_j) = n \cdot p_j, j = 1, 2, \dots, k$. Låt oss definiera

$$q = \sum_{j=1}^k \frac{(y_j - np_j)^2}{np_j},$$

där y_j är observationen av Y_j . Ett stort värde på q borde rimligen indikera att H_0 inte gäller (åtminstone något p_j måste skilja sig markant från det förväntade värdet np_j).

Storheten q är en observation av den stokastiska variabeln

$$Q = \sum_{j=1}^k \frac{(Y_j - np_j)^2}{np_j} \stackrel{\text{appr.}}{\sim} \chi^2(k-1).$$

Att detta blir approximativt χ^2 -fördelat följer av följande sats.



Sats. Med beteckningarna ovan gäller att $\sum_{j=1}^k \frac{(Y_j - np_j)^2}{np_j}$ konvergerar till $X \sim \chi^2(k-1)$ i fördelning.



När duger approximationen?

Föregående sats gäller alltså *asymptotiskt* (då $n \rightarrow \infty$) och säger inget direkt om vad som gäller i det enskilda fallet. En tumregel är att vi vill ha $np_j \geq 5$ för $j = 1, 2, \dots, k$ för att vara ganska säkra på att approximationen är bra. Har vi lådor med väldigt få "bollar" i kan det hända att testet inte blir bra.

10.2 Test av given diskret fördelning

Låt X_1, X_2, \dots, X_n vara oberoende diskreta stokastiska variabler med $X_j \in A$ för någon diskret mängd A . Vi är intresserade av att testa om $X_j \sim F$ för någon given diskret fördelning med sannolikhetsfunktion $p(j)$, $j \in A$. Vi kommer använda nollhypotesen

$$H_0 : P(X = j) = p(j), \quad j \in A,$$

och testar den med mothypotesen

$$H_1 : P(X = j) \neq p(j) \text{ för något } j \in A.$$



Exempel

Den stokastiska variabeln X antar värden i mängden $\{0, 1, 2\}$. Vid 1250 observationer fann man att $X = 0$ 783 gånger, $X = 1$ 425 gånger samt $X = 2$ 42 gånger. Testa med signifikansnivån 1% om $X \sim \text{Bin}(2, 1/5)$.

Lösning. Vi låter $H_0 : X \sim \text{Bin}(2, 1/5)$. Om vi antar att H_0 är sann så gäller att

$$\begin{aligned} P(X = 0) &= \binom{2}{0} \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^2 = \frac{16}{25}, \\ P(X = 1) &= \binom{2}{1} \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^1 = \frac{8}{25}, \\ P(X = 2) &= \binom{2}{2} \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^0 = \frac{1}{25}. \end{aligned}$$

Kom ihåg att kontrollera att dessa summerar till 1, det är en billig kontroll på tentan. Utifrån detta kan vi beräkna de förväntade frekvenserna vid 1250 försök (om H_0 är sann):

$$np_j = \begin{cases} 800, & j = 0, \\ 400, & j = 1, \\ 50, & j = 2. \end{cases}$$

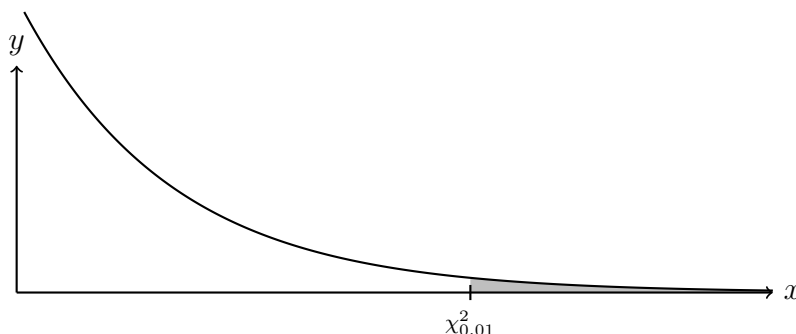
Testvariabeln q ges nu av

$$q = \sum_{j=0}^2 \frac{(x_j - np_j)^2}{np_j} = \frac{(783 - 800)^2}{800} + \frac{(425 - 400)^2}{400} + \frac{(42 - 50)^2}{50} \approx 3.2038.$$

Eftersom $k = 3$ är q en observation av $Q \stackrel{\text{appr.}}{\sim} \chi^2(2)$ om H_0 är sann. Vi finner att

$$0.01 = P(Q > \chi_{0.01}^2(2)) \Leftrightarrow \chi_{0.01}^2(2) = 9.21$$

ur tabell.



Eftersom $q = 3.2038 < 9.21$ kan vi inte förkasta H_0 . Fördelningen kan mycket riktigt vara binomialfördelning med $p = 1/5$.

10.3 Test för kontinuerlig fördelning

Om vi istället har en kontinuerlig situation där vi vill testa om mätdata följer en given fördelning F måste vi agera lite annorlunda. Vi skulle önska att ställa upp

$$H_0 : X \sim F$$

mot

$$H_1 : X \text{ har ej fördelningen } F.$$

Men detta blir lite för komplicerat i det generella fallet.

Istället gör vi så att vi diskretiserar det hela på något sätt. Vi gör oftast detta genom att skapa lådor i form av intervall och sedan undersöka hur många observationer som hamnar i varje delintervall. Detta gör att vi inte exakt testar om nollhypotesen ovan utan vi testar en svagare nollhypotes.

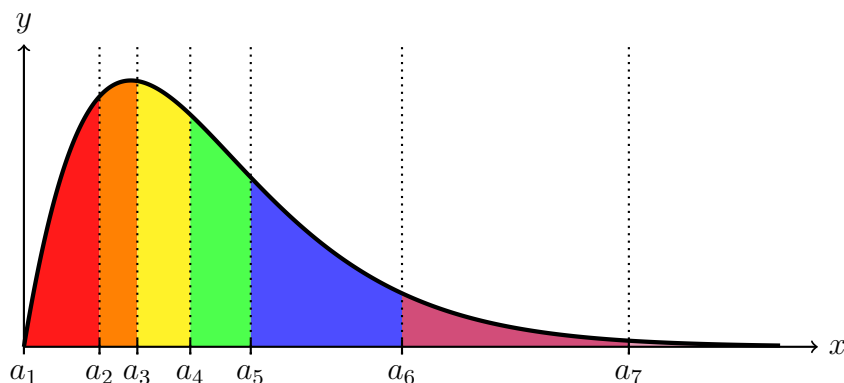
Låt X_i , $i = 1, 2, \dots, n$ vara oberoende och likafördelade variabler med täthetsfunktion $f(x)$. Vi väljer a_j , $j = 1, 2, \dots, k + 1$, så att

$$-\infty \leq a_1 < a_2 < \dots < a_k < a_{k+1} \leq \infty$$

och definierar $A_j = [a_j, a_{j+1}[$ för $j = 2, 3, \dots, k$ och låter typiskt $A_1 =]-\infty, a_2[$. Vi definierar sedan

$$p_j = P(X_i \in A_j) = \int_{a_j}^{a_{j+1}} f(x) dx.$$

Om f är en täthetsfunktion så blir nu $p_1 + p_2 + \dots + p_k = 1$ och vi har täckt alla möjligheter. Om stödet för f inte är hela \mathbf{R} modifierar vi naturligt definitionen (eller låter $f(x) = 0$ utanför sin definition). En tumregel för valet är att vi låter $k \approx n/10$. En annan tumregel är att välja intervallen så stora att alla p_j är ungefär lika stora.



Hypotesen vi kommer testa är

$$H_0 : P(X \in A_j) = p_j, \quad j = 1, 2, \dots, k,$$

mot

$$H_1 : P(X \in A_j) \neq p_j \text{ för något } j.$$

Skulle X ha rätt fördelning kommer H_0 att vara sann med stor sannolikhet, men om vi styrker H_0 innebär det inte nödvändigtvis att det är just den fördelning vi utgick från när vi ställde upp A_j som är den sanna (bara någon med motsvarande sannolikheter i uppdelningen). Vill man ha ett starkare resultat krävs andra metoder.



Exempel

Säljaren på ELFA hävdar bestämt att livslängden på en komponent är exponentialfördelad med väntevärde 2 år. Uttråkade pensionären Sture tror inte på det utan köper 50 stycken komponenter för att testa. Sture kopplar upp komponenterna och kikar till var 6:e månad för att se hur många som gått sönder.

Tid (mån)	< 6	< 12	< 18	< 24	< 30	< 36	< 42	< 48	< 54	< 60
Antal:	11	19	25	31	36	39	39	40	42	43

Undersök om antagandet är rimligt på approximativt 1% nivån.

Lösning. Vi kan organisera om datan mer användbart enligt hur många enheter som gick sönder under en viss tidsenhet. För att få ungefär jämnstora klasser så buntar vi ihop enligt följande.

Tid	Hur många dog
$I_1 = [0, 6)$	11
$I_2 = [6, 12)$	8
$I_3 = [12, 24)$	12
$I_4 = [24, 36)$	8
$I_5 = [36, \infty)$	11

Om vi antar H_0 så gäller att täthetsfunktionen för livslängden hos en komponent X ges av $f(x) = \mu^{-1} \exp(-\mu^{-1}x)$, så

$$P(a \leq X < b) = \int_a^b \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right) dx = \exp\left(-\frac{a}{\mu}\right) - \exp\left(-\frac{b}{\mu}\right).$$

Med siffrorna ovan ser vi att

$$P(X \in I_k) = \begin{cases} p_1 = 0.2212, & k = 1, \\ p_2 = 0.1723, & k = 2, \\ p_3 = 0.2387, & k = 3, \\ p_4 = 0.1447, & k = 4, \\ p_5 = 0.2231, & k = 5. \end{cases}$$

Teststorheten vi använder kommer nu ges av

$$q = \sum_{j=1}^5 \frac{(x_j - np_j)^2}{np_j} = \frac{(11 - 50 \cdot 0.2212)^2}{50 \cdot 0.2212} + \dots + \frac{(11 - 50 \cdot 0.2231)^2}{50 \cdot 0.2231} = 0.1276.$$

Om H_0 är sann så kommer q vara en observation av $Q \stackrel{\text{appr.}}{\sim} \chi^2(5-1) = \chi^2(4)$, så med det kritiska området $C = (0, c)$ där $c = 13.28$, ser vi att vi inte kan förkasta H_0 . Säljaren kan mycket väl ha rätt.

10.4 Skattade storheter

Normalt sätt kanske vi inte får exakt väntevärde (eller andra parametrar i fördelningen) utan dessa måste skattas innan vi kan utföra testet. Hur påverkar det fördelningen för teststorheten Q ? Svaret är enkelt: för varje skattning vi gör tappar vi en frihetsgrad, under förutsättningen att skattningen är vettig (ML-skattningar brukar bete sig bra). Bevis är däremot lite bökligare (å andra sidan får vi det första χ^2 -testet mer eller mindre på köpet). Får jag tid över kommer jag skriva ned det och uppdatera anteckningarna. Om vi antar att sannolikheterna p_j beror på okända $\theta = (\theta_1 \theta_2 \dots \theta_r)^T$, så gäller alltså att

$$Q = \sum_{j=1}^k \frac{(Y_j - n\hat{p}_j(\theta))^2}{n\hat{p}_j(\theta)} \stackrel{\text{appr.}}{\sim} \chi^2(k - r - 1),$$

under förutsättning att skattningarna som används beter sig tillräckligt bra.



Exempel

Linnea gör en signalbehandlingslaboration i matlab men hennes algoritm fungerar inte som planerat. Givetvis tycker Linnea att felet måste ligga i matlabs sätt att generera normalfördelade slumpstal. För att testa hypotesen att slumpstalen inte är normalfördelade genererar Linnea 1000 slumpstal och sorterar dessa i storleksordning följt av en klassindelning så det är precis 100 element i varje klass. Gränserna kan ses nedan.

Undre gräns	1.57	12.47	15.00	17.04	18.80	20.33	21.76	23.26	25.00	27.43
Övre gräns	12.46	14.98	17.03	18.77	20.32	21.75	23.25	24.99	27.42	40.80

Det beräknade medelvärdet är $\bar{x} = 20.14$ och stickprovsvariansen är $s^2 = 35.25$. Testa på nivån 5% om värdena är normalfördelade.

Lösning. Låt H_0 : datan kommer från $N(\mu, \sigma)$ och H_1 : datan är inte normalfördelad. Om vi använder $\bar{x} = 20.14$ som skattning för väntevärdet och $s = \sqrt{35.25} = 5.94$ som skattning för standardavvikelsen, så kan vi (om vi antar att H_0 är sann) beräkna sannolikheterna för en normalfördelad variabel Z att hamna i de olika klasserna enligt

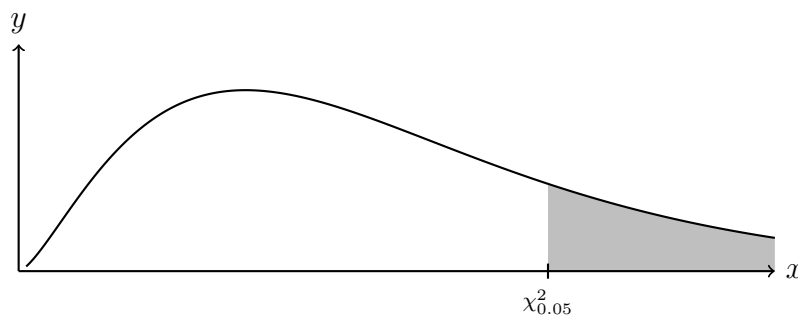
$$\begin{aligned} P(a \leq Z < b) &= P\left(\frac{a - \mu}{\sigma} \leq \frac{Z - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \\ &\approx \Phi\left(\frac{b - 20.14}{5.94}\right) - \Phi\left(\frac{a - 20.14}{5.94}\right). \end{aligned}$$

Resultatet kan beskådas nedan.

Intervall	Sannolikhet
$I_1 = (-\infty, 12.46)$	0.10
$I_2 = [12.47, 14.99)$	0.09
$I_3 = [15.00, 17.03)$	0.11
$I_4 = [17.04, 18.77)$	0.11
$I_5 = [18.78, 20.32)$	0.10
$I_6 = [20.33, 21.75)$	0.09
$I_7 = [21.76, 23.25)$	0.09
$I_8 = [23.26, 24.99)$	0.09
$I_9 = [25.00, 27.42)$	0.10
$I_{10} = [27.43, \infty)$	0.11

Vi ser redan nu att sannolikheterna väldigt nära hamnar runt 10-delar (vilket borde ske om normalfördelning gäller med tanke på konstruktionen). Men låt oss ställa upp teststorheten och se:

$$q = \sum_{j=1}^{10} \frac{(x_j - n\hat{p}_j)^2}{n\hat{p}_j} = \frac{(100 - 1000 \cdot 0.10)^2}{1000 \cdot 0.10} + \dots + \frac{(100 - 1000 \cdot 0.11)^2}{1000 \cdot 0.11} = 4.34.$$



Om H_0 är sann så är q en observation av $\chi^2(10 - 2 - 1) = \chi^2(7)$ eftersom vi skattar två parametrar. På nivån 0.1% så gäller att $P(Q > 14.07) = 0.05$, och då $4.34 < 14.07$ så kan vi inte förkasta nollhypotesen. Linnea har antagligen implementerat sin algoritm fel.

Att testa normalfördelning på detta sätt är inte helt lämpligt. Det finns betydligt bättre metoder som till exempel Kolmogorov-Smirnovs metod som istället baserar sig på den empiriska fördelningsfunktionen. Test av denna typ ger bättre resultat i allmänhet.

Låt oss avrunda med att betrakta ett annat diskret exempel som exemplifierar hur vi hanterar potentiellt oändligt många utfall och skattade parametrar.

**Exempel**

Astrid har slaktat 100 stycken brandvarnare och försöker mäta hur radioaktivt materialet är. I lämplig enhet finner hon följande frekvensdata från den 100 enheterna.

$X = i$	0	1	2	3	4	5	6	7
antal (y_i)	8	8	30	17	21	9	5	2

Testa $H_0 : X \sim \text{Po}(\mu)$ mot $H_1 : X$ är inte Poissonfördelad. på nivån 5%.

Lösning. Eftersom väntevärdet μ är okänt måste vi skatta detta och det gör vi enklast med

$$\hat{\mu} = \frac{1}{100} \sum_{i=0}^7 i y_i = \frac{0 \cdot 8 + 1 \cdot 8 + \dots + 7 \cdot 2}{100} = 2.92.$$

Detta är medelvärdet och det är en rimlig skattning av väntevärdet (undersök om det är ML-skattningen!). Om H_0 är sann kan vi räkna ut sannolikheterna för de olika observationerna med hjälp av sannolikhetsfunktionen för Poissonfördelning:

$$p_i = P(X = i) = e^{-\mu} \frac{\mu^i}{i!} \approx e^{-\hat{\mu}} \frac{\hat{\mu}^i}{i!} = \hat{p}_i$$

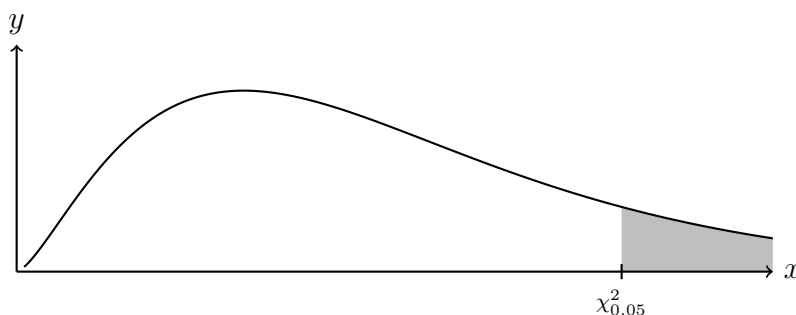
för $i = 1, 2, 3, \dots$ (oändligt många möjligheter). Vi måste slå ihop utfallen i svansen på fördelningen då sannolikheterna där blir för små ($np_i < 5$ och dessutom har vi inget test för oändligt många möjligheter) så vi testar att låta $i \geq 6$ vara den sista klassen och gör om tabellen:

$X = i$	0	1	2	3	4	5	≥ 6
antal (y_i)	8	8	30	17	21	9	5+2=7
$100\hat{p}_i$	5.4	15.8	23.0	22.4	16.3	9.5	7.61

Vi ser här att samtliga $n\hat{p}_i \geq 5$ (även om den första är ganska nära) så kraven är uppfyllda. Vi räknar ut teststorheten

$$q = \sum_{i=0}^6 \frac{(y_i - 100\hat{p}_i)^2}{100\hat{p}_i} = \frac{(8 - 5.4)^2}{5.4} + \dots + \frac{(7 - 7.61)^2}{7.61} = 9.91.$$

Om H_0 är sann så är q en observation av $Q \stackrel{\text{appr.}}{\sim} \chi^2(7 - 1 - 1) = \chi^2(5)$ eftersom vi har 7 klasser och skattar en parameter (med en vettig skattning). Ur tabell finner vi att $P(Q > \chi_{0.05}^2(5) = 11.07) = 0.05$ och eftersom $q < 11.07$ så kan vi *inte* förkasta H_0 . Det kan mycket väl vara så att observationerna kommer ifrån en Poissonfördelning.



10.5 Homogenitetstest

Det kan ofta vara intressant att avgöra om egenskaper skiljer sig åt mellan olika grupper. Låt oss sätta upp följande scenario. Vi har s stycken grupper eller serier av försök som vi är nyfikna på om de uppvisar samma sorts fördelning med avseende på en mängd egenskaper A_1, A_2, \dots, A_r . Vi kan då ställa upp datan enligt följande där siffrorna är absoluta frekvenser.

	Egenskap 1	Egenskap 2	\dots	Egenskap r	Summa
Grupp 1	N_{11}	N_{12}	\dots	N_{1r}	G_1
Grupp 2	N_{21}	N_{22}	\dots	N_{2r}	G_2
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
Grupp s	N_{s1}	N_{s2}	\dots	N_{sr}	G_s
Summa	E_1	E_2	\dots	E_r	N

Om vi antar att grupperna är *homogena*, dvs att de uppvisar samma fördelning för egenskaperna, så är en bra skattning för sannolikheten p_j att ett objekt har egenskap j helt enkelt

$$\hat{p}_j = E_j/N.$$

Vi formar samma sorts teststorhet som vi gjort innan

$$Q = \sum_{i=1}^s \sum_{j=1}^r \frac{(N_{ij} - G_i \cdot \hat{P}_j)^2}{G_i \hat{P}_j} \stackrel{\text{appr.}}{\sim} \chi^2((r-1)(s-1)).$$

Att det blir just $(r-1)(s-1)$ kommer från de linjära restriktioner som trillar ut ur tabellen ovan. Vi kan se att

$$\sum_{j=1}^r \frac{(N_{ij} - G_i \cdot p_j)^2}{G_i p_j} \stackrel{\text{appr.}}{\sim} \chi^2(r-1)$$

enligt tidigare argument, men då antar vi att p_j är kända. Sedan summerar vi s sådana oberoende variabler, så resultatet blir $\chi^2(s(r-1))$ -fördelat. Men nu vill vi skatta p_j och för varje skattning tappar vi en frihetsgrad. Märk dock att vi inte skattar alla r stycken p_j , utan bara $r-1$ stycken då den sista ges av att summan måste bli ett. Vi tappar alltså $r-1$ frihetsgrader. Totalt sett har vi alltså $s(r-1) - (r-1) = (s-1)(r-1)$ frihetsgrader.

Så när gäller approximationen? Den gäller under förutsättning att $n_i \hat{p}_j$ är stora. En rimlig tumregel är att $n_i \hat{p}_j \geq 5$.



Exempel

Från en stor population frågar vi två grupper om de tycker det borde vara lagligt att kasta tallkottar på hundägare som inte håller sina hundar kopplade.

Grupp	Kottkastning är OK!	Nej man får inte kasta tallkottar på folk.
G_1	59	41
G_2	145	55

Testa på signifikansnivån 1% (approximativt) om det finns någon skillnad mellan vad grupperna tycker.

Lösning. Vi utför ett homogenitetstest.

Grupp	Kottekastning är OK!	Nej man får inte kasta tallkottar på folk.	Summa
G_1	59	41	100
G_2	145	55	200
$G_1 + G_2$	204	96	300
\hat{p}_j	0.68	0.32	1.0

Låt H_0 : Grupperna tycker likadant mot H_1 : Grupperna tycker olika. Vår observation av teststorheten ges av

$$q = \frac{(59 - 68)^2}{68} + \frac{(41 - 32)^2}{32} + \frac{(145 - 136)^2}{136} + \frac{(55 - 64)^2}{64} \approx 5.58.$$

Detta är en observation av $Q \stackrel{\text{appr.}}{\sim} \chi^2(1 \cdot 1)$. Ur tabell finner vi att $P(Q > 6.6349) = 0.01$. Eftersom $q < 6.6349$ så kan vi inte förkasta hypotesen att grupperna tycker lika.



Exempel

Alla som lyssnar på hårdrock i någon form har säkert funderat över vilken av Slayer-låtarna *Angel of Death* och *Raining Blood* som är bäst^a. Examinator funderade över om resultaten är homogena över några olika grupper och samlade in följande siffror på internet:

	<i>Angel of Death</i>	<i>Raining Blood</i>
Returntothepit.com	199	173
MetalStorm.net	47	43
RockBand.com	21	16
MetalRules.com	23	3

Utför ett homogenitetstest på nivån 5% för att se om man kan förkasta hypotesen att åsikterna är likafördelade i de fyra olika grupperna.

^aSjälvklart är *Angel of Death* den bästa av dessa två, men det är inte poängen!

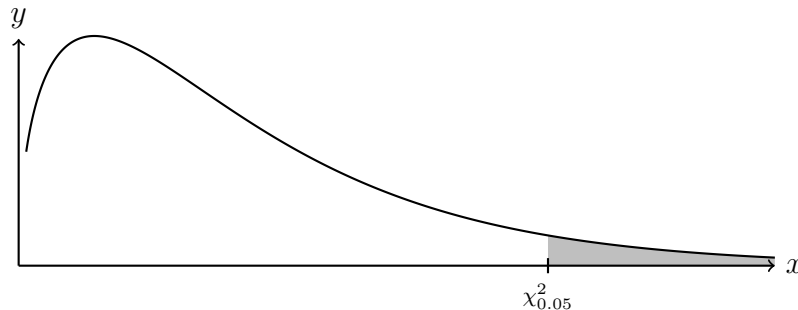
Lösning. Först kompletterar vi tabellen med all information som behövs:

	<i>Angel of Death</i>	<i>Raining Blood</i>	Summa (n_i)
Returntothepit.com	199	173	372
MetalStorm.net	47	43	90
RockBand.com	21	16	37
MetalRules.com	23	3	26
Summa	290	235	525
\hat{p}_j (skattat p_j)	$\hat{p}_1 = 0.552$	$\hat{p}_2 = 0.448$	1.00

Vi beräknar observationen q

$$\begin{aligned}
 q &= \sum_{i=1}^4 \sum_{j=1}^2 \frac{(x_{ij} - n_i \hat{p}_j)^2}{n_i \hat{p}_j} \\
 &= \frac{(199 - 372 \hat{p}_1)^2}{372 \hat{p}_1} + \frac{(173 - 372 \hat{p}_2)^2}{372 \hat{p}_2} + \frac{(47 - 90 \hat{p}_1)^2}{90 \hat{p}_1} + \frac{(43 - 90 \hat{p}_2)^2}{90 \hat{p}_2} \\
 &\quad + \frac{(21 - 37 \hat{p}_1)^2}{37 \hat{p}_1} + \frac{(16 - 37 \hat{p}_2)^2}{37 \hat{p}_2} + \frac{(23 - 26 \hat{p}_1)^2}{26 \hat{p}_1} + \frac{(3 - 26 \hat{p}_2)^2}{26 \hat{p}_2} \\
 &= 12.43
 \end{aligned}$$

Låt H_0 vara utsagan att favoriten bland de två låtarna är likadant fördelad i alla fyra serier. Det vill säga, att $P(\text{AoD favorit}) = p_1$ och $P(\text{RB favorit}) = p_2$ gäller i alla fyra serierna med samma sannolikheter p_j . Antag att H_0 är sann.



Vi förkastar H_0 om $Q > \chi^2_\alpha(3)$, d v s om den observerade testvariabeln hamnar utanför det skuggade området i figuren ovan. Med $\alpha = 0.05$ finner vi att $\chi^2_{0.05}(3) = 7.81$ (ur en tabell eller med matlab), så $Q > \chi^2_\alpha(3)$. Vi kan alltså förkasta hypotesen att alla grupperna tycker likadant (ganska tydligt från siffrorna att den fjärde raden skiljer sig markant från de andra).

Svar: Vi kan förkasta hypotesen om homogenitet på nivån 5%.

10.6 Skattningar för kovarians och korrelation

Om vi har ett stickprov (x_k, y_k) , $k = 1, 2, \dots, n$, där (X_k, Y_k) är stokastiska variabler med samma fördelning, så skattar vi kovariansen C med

$$\hat{c} = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$$

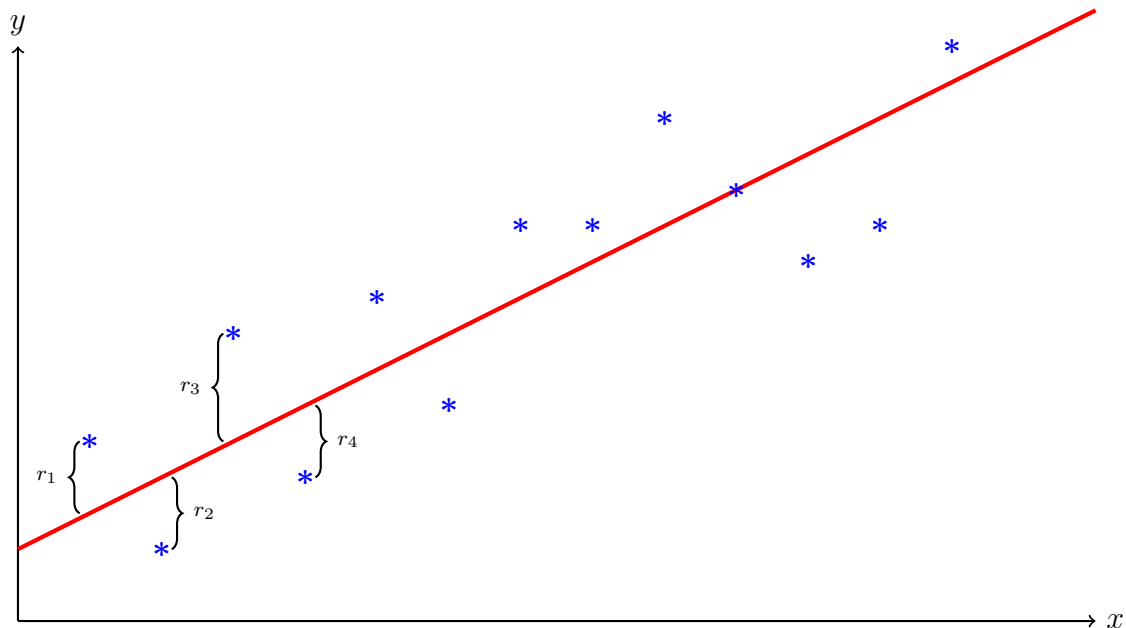
och korrelationen med

$$\hat{\rho} = \frac{\hat{c}}{s_x s_y} = \frac{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\left(\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2\right)^{1/2} \left(\frac{1}{n-1} \sum_{k=1}^n (y_k - \bar{y})^2\right)^{1/2}}.$$

Av tradition betecknar man ofta $\hat{\rho} = r$. En naturlig fråga i detta skede är om vi kan säga något om fördelningen för den skattade korrelationen under något lämpligt antagande om det slumpmässiga stickprovet. I vissa fall kan man det, men det är inget vi ger oss in på i denna kurs.

10.7 Enkel linjär regression

Vi återgår nu till ett exempel vi stött på redan vid ett flertal tillfällen (om inte annat så i tidigare kurser som linjär algebra), nämligen att anpassa en rät linje $y = \beta_0 + \beta_1 x$ efter mätdata (x_i, y_i) , $i = 1, 2, \dots, n$. Grafiskt illustrerat enligt nedan.



Målsättningen är att – givet en mätserie – hitta den linje som approximerar denna serie på lämpligt sätt. Det resulterar i ett par naturliga funderingar.

- (i) Hur hittar man en approximativ linje systematiskt?
- (ii) Om man upprepar försöket, får man samma linje?
- (iii) I vilken mening är linjen optimal? På vilket sätt mäter vi avvikelserna mellan linjen och mätserien?

Vi ska försöka svara på dessa frågor och för att göra det behöver vi ställa upp en modell.



Enkel linjär regression

Definition. Vi kommer betrakta följande modell: givet (x_j, y_j) , $j = 1, 2, \dots, n$, där vi betraktar x_j som fixerade och y_j som observationer av stokastiska variabler

$$Y_j = \beta_0 + \beta_1 x_j + \epsilon_j, \quad j = 1, 2, \dots, n,$$

där $\epsilon_j \sim N(0, \sigma)$ antas oberoende (och likafördelade). Den räta linjen $y = \beta_0 + \beta_1 x$ kallas **regressionslinjen**.

Vi använder beteckningen $\mu_j = \beta_0 + \beta_1 x_j = E(Y_j)$.

Är det givet att denna modell är sann? Nej, det är inte självklart utan hänger på vilka förutsättningar datan kommer från. Däremot tenderar modellen att fungera bra i de flesta fall om vi har en rimlig mängd observationer. Med notationen från figuren ser vi att

$$r_j = y_j - \mu_j$$

om vi tar hänsyn till tecknet (positivt tecken om y_j ligger ovanför regressionslinjen). Ett mått på hur väl linjen approximerar mätserien ges av kvadratsumman

$$\sum_{j=1}^n r_j^2 = \sum_{j=1}^n (y_j - \mu_j)^2.$$

Vi skulle kunna minimera denna summa med avseende på (β_0, β_1) , vilket ger MK-skattningar för β_0 och β_1 . Detta var det som hände i exemplet från föreläsning 7. Istället för att upprepa argumentet går vi över till den generella modellen för linjär regression.

Dubbelindexeringen är lite jobbig att arbeta med, så låt oss gå över till matrisnotation:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 & \beta_1 x_1 \\ \beta_0 & \beta_1 x_2 \\ \vdots & \vdots \\ \beta_0 & \beta_1 x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Vi låter

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \text{samt } \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

vilket leder till sambandet

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Således gäller att

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \quad \text{och} \quad C_{\mathbf{Y}} = \sigma^2 I_2,$$

där I_2 är den n -dimensionella enhetsmatrisen och $C_{\mathbf{Y}}$ är kovariansmatrisen för \mathbf{Y} . Vi söker MK-skattningen $\hat{\boldsymbol{\beta}}$ för $\boldsymbol{\beta}$, vilket vi kan erhålla genom att minimera den kvadratiske formen

$$Q(\beta_0, \beta_1) = \sum_{j=1}^n (y_j - \mu_j)^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

där $\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_n)^T$. En variant är att sätta igång och derivera, men lite mer elegant gäller följande sats (välkänd från linjär algebra).



Normalekvationerna

Sats. Om det $X^T X \neq 0$ så ges MK-skattningen av $\boldsymbol{\beta}$ enligt

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{y}.$$

Med förutsättningarna ovan gäller att

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (X^T X)^{-1}).$$

Detta vektorvärda samband innebär att $\hat{\beta}_0 \sim N(\beta_0, \sigma\sqrt{h_{00}})$ och $\hat{\beta}_1 \sim N(\beta_1, \sigma\sqrt{h_{11}})$, där

$$(X^T X)^{-1} = \begin{pmatrix} h_{00} & h_{01} \\ h_{10} & h_{11} \end{pmatrix}.$$



När det gäller $\hat{\beta}$ och dess komponenter kommer vi använda samma beteckningar för observationer av punktskattningen och den stokastiska variabeln. Skulle vi vara konsekventa borde den stokastiska variabeln betecknas \hat{B} , men av tradition görs inte så. Var observant!

10.8 Variansanalys

Vi låter

$$\hat{\mu}_j = \hat{\beta}_0 + \hat{\beta}_1 x_j, \quad j = 1, 2, \dots, n.$$

Ibland betecknas vektorn $\hat{\mu}$ som \hat{y} eftersom det i någon mening är en skattningen av y , men det blir lite olyckligt för det är inte y vi skattar. Vi ska nu studera hur bra den skattning vi tagit fram är.



Regressionsanalysens kvadratsummor

Definition. Vi definierar tre kvadratsummor som beskriver variationen hos y -värdena:

(i) Den **totala variationen** definieras enligt $SS_{\text{TOT}} = \sum_{j=1}^n (y_j - \bar{y})^2$.

(ii) Variationen som förklaras av x_1, x_2, \dots, x_k , definieras enligt $SS_{\text{R}} = \sum_{j=1}^n (\hat{\mu}_j - \bar{y})^2$.

(iii) Variationen som inte förklaras av regressionsmodellen är $SS_{\text{E}} = \sum_{j=1}^n (y_j - \hat{\mu}_j)^2$.

Ibland används beteckningarna Q_{TOT} för SS_{TOT} , Q_{REGR} för SS_{R} och Q_{RES} för SS_{E} . Hur hänger dessa summor ihop? Som tur är finns ett enkelt svar.

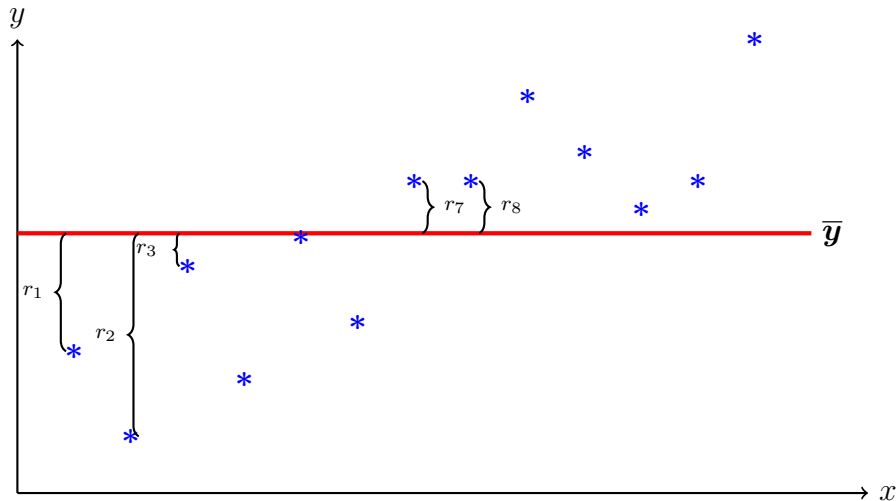


Sats. Den totala variationen SS_{TOT} kan delas upp enligt

$$SS_{\text{TOT}} = SS_{\text{R}} + SS_{\text{E}}.$$

10.8.1 SS_{TOT}

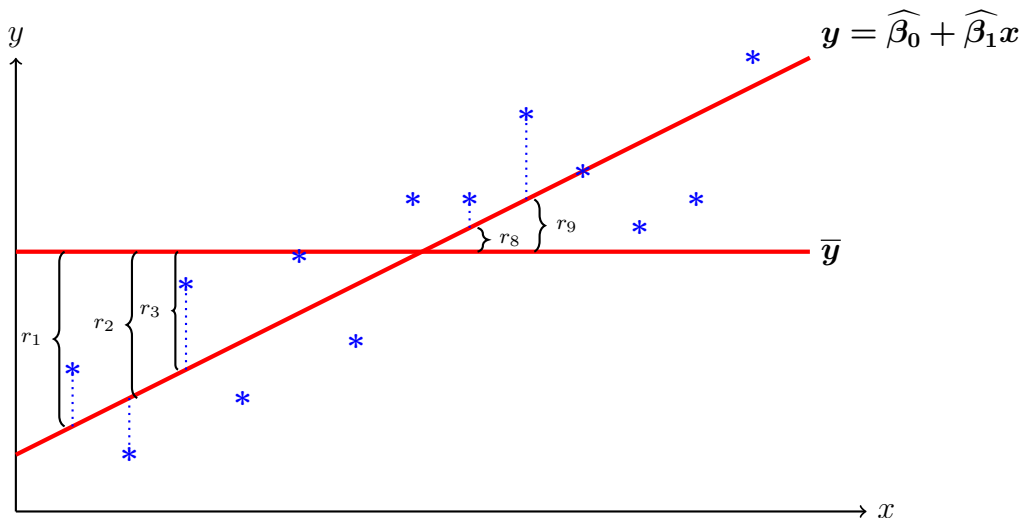
Storheten SS_{TOT} mäter den totala variation av mätvärden jämfört med mätvärdenas medelvärde. I fallet då vi använder modellen $Y = \beta_0 + \epsilon$ ger således SS_{TOT} hela felet i regressionen.



Vi ser att $SS_{TOT} = \sum_{j=1}^n r_j^2 = \sum_{j=1}^n (y_j - \bar{y})^2$.

10.8.2 SS_R och SS_E

Om vi istället använder modellen $y = \beta_0 + \beta_1 x + \epsilon$ blir summorna SS_E och SS_R relevanta (SS_{TOT} ser fortfarande ut som ovan). Låt oss först illustrera SS_R .

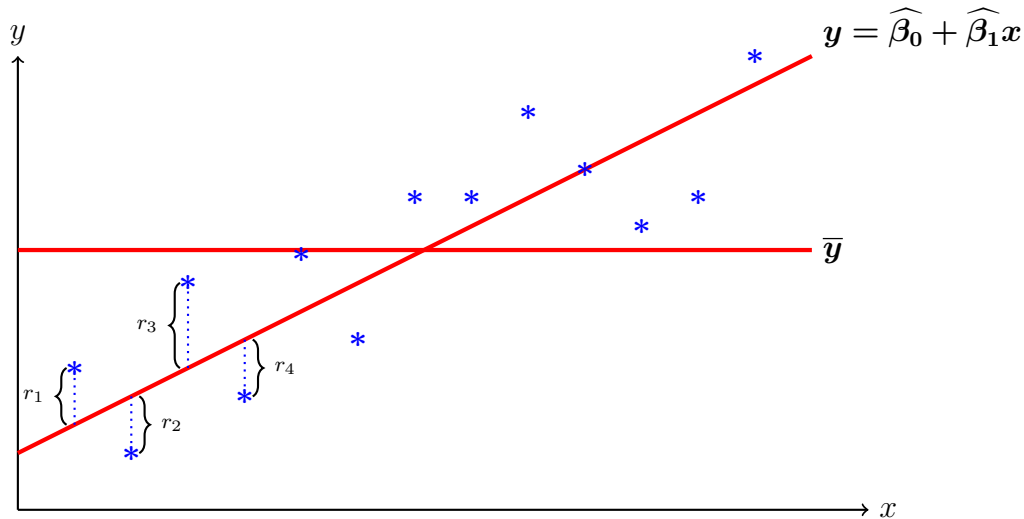


Vi ser att

$$SS_R = \sum_{j=1}^n r_j^2 = \sum_{j=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_j - \bar{y})^2$$

och SS_R mäter alltså hur mycket den skattade regressionslinjen (i mätpunkterna x_j) skiljer sig från medelvärdet av y -värdena.

När det gäller SS_E är det istället skillnaden mellan den skattade regressionslinjen och mätvärdena vi betraktar.



Kvadratsumman av dessa avvikelser blir

$$SS_E = \sum_{j=1}^n r_j^2 = \sum_{j=1}^n (y_j - (\hat{\beta}_0 + \hat{\beta}_1 x_j))^2.$$

10.9 Hypotestester och konfidensintervall

Vi har nu samlat på oss en ordentlig verktygslåda, så det kanske är dags att se hur vi använder de olika delarna.

10.9.1 Skattning av σ^2

Variansen σ^2 skattar vi med

$$s^2 = \frac{SS_E}{n-2}.$$

Denna skattning är väntevärdesriktig:

$$E(S^2) = \frac{\sigma^2}{n-2} E\left(\frac{SS_E}{\sigma^2}\right) = \sigma^2 \frac{n-2}{n-2} = \sigma^2$$

ty $\frac{1}{\sigma^2} SS_E \sim \chi^2(n-2)$. Det faktum att vi får en $\chi^2(n-2)$ -fördelning är inte självklart; se avsnitt 10.12.

10.9.2 Enskilda koefficienter

Med beteckningen

$$(X^T X)^{-1} = \begin{pmatrix} h_{00} & h_{01} \\ h_{10} & h_{11} \end{pmatrix}$$

så är $\hat{\beta}_i \sim N(\beta_i, \sigma \sqrt{h_{ii}})$. Om vi känner σ^2 kan vi använda att

$$Z = \frac{\hat{\beta}_i - \beta_i}{\sigma \sqrt{h_{ii}}} \sim N(0, 1)$$

för att testa hypotesen $H_0 : \beta_i = 0$ eller för att ställa upp konfidensintervall I_{β_i} .

Nu är det extremt sällan vi känner σ^2 exakt, men vi vet att $\hat{\beta}$ är oberoende av SS_E enligt huvudsatsen (åter igen refererar vi till avsnitt 10.12), så $\hat{\beta}_i$ är oberoende av SS_E . Vidare är $s^2 = SS_E/(n-2)$ en skattning av σ^2 vi känner väl, så

$$\frac{(\hat{\beta}_i - \beta_i)/(\sigma\sqrt{h_{ii}})}{\sqrt{((n-2)S^2/\sigma^2)/(n-2)}} = \frac{\hat{\beta}_i - \beta_i}{S\sqrt{h_{ii}}} \sim t(n-2)$$

enligt Gossets sats.

10.9.3 Hypotestest: $H_0 : \beta_i = 0$

Vi kan även utföra hypotestester för en enskild koefficient β_i för att se om den förklaringsvariabeln tillför något signifikant (givetvis går det att ställa upp konfidensintervall också för mer kvantitativt innehåll).

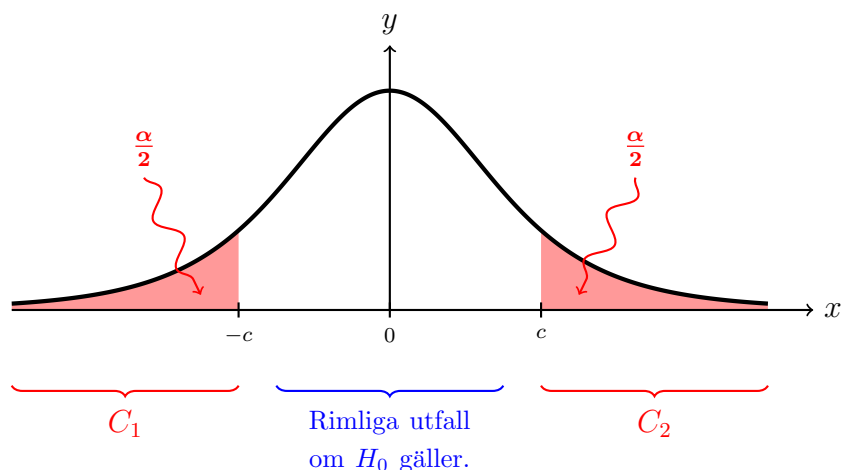
Låt $H_0 : \beta_i = 0$ och $H_1 : \beta_i \neq 0$. Vi vet enligt ovan att

$$T = \frac{\hat{\beta}_i - \beta_i}{S\sqrt{h_{ii}}} \sim t(n-2),$$

så det kritiska området finner vi enligt

$$C = \{t \in \mathbf{R} : |t| > c\}$$

för lämpligt tal $c > 0$ beroende på signifikansnivån och antalet frihetsgrader.



Gränsen hittar vi i tabell genom att leta reda på ett tal $c = F_T^{-1}(1 - \alpha/2)$ (sitter du med MATLAB kan du använda `c = -tinv(alpha/2)`).

10.9.4 Formler utan matriser

Vi diskuterade enkel linjär regression tidigare i samband med MK-skattningar (föreläsning 7). Låt oss visa att vi får samma resultat med den metod vi nu tagit fram (och fördelningar för ingående storheter på ett enkelt sätt).

Vi låter x_1, x_2, \dots, x_n vara fixerade tal och y_1, y_2, \dots, y_n vara observationer från

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

där $\epsilon_i \sim N(0, \sigma)$ är oberoende. Alltså precis den modell vi använt tidigare. Syntesmatrisen X ges av

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

så

$$X^T X = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix} \Rightarrow (X^T X)^{-1} = \frac{1}{n \sum_{i=1}^n x_i^2 - (n\bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}$$

om $\det(X^T X) \neq 0$. Således blir

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X^T X)^{-1} X^T \mathbf{y} \\ &= \frac{1}{n \sum_{i=1}^n x_i^2 - (n\bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} \\ &= \frac{1}{n \sum_{i=1}^n x_i^2 - (n\bar{x})^2} \begin{pmatrix} n\bar{y} \sum_{i=1}^n x_i^2 - n\bar{x} \sum_{i=1}^n x_i y_i \\ -n^2 \bar{x} \bar{y} + n \sum_{i=1}^n x_i y_i \end{pmatrix}, \end{aligned}$$

vilket ger att

$$\hat{\beta}_1 = \frac{\sum_{j=1}^n x_j y_j - n\bar{x}\bar{y}}{\sum_{j=1}^n x_j^2 - n\bar{x}^2} = \frac{\sum_{j=1}^n y_j(x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{\sum_{j=1}^n (y_j - \bar{y})(x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \quad \text{och} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Det följer nu att

$$\hat{\beta}_0 \sim N\left(\beta_0, \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}\right) \quad \text{och} \quad \hat{\beta}_1 \sim N\left(\beta_1, \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}\right).$$

Någonstans nu inser vi vilket kraftigt syntaktiskt verktyg matriser och linjär algebra är... Eftersom σ är okänd skattar vi σ^2 med

$$s^2 = \frac{\text{SS}_E}{n-2} = \frac{\sum_{j=1}^n (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_j)^2}{n-2}$$

där $S^2 \sim \chi^2(n-2)$. Vi kan skriva om SS_E genom att utnyttja att

$$\begin{aligned} \text{SS}_R &= \sum_{j=1}^n (\hat{\mu}_j - \bar{y})^2 = \hat{\beta}_1^2 \sum_{j=1}^n (x_j - \bar{x})^2 = \left(\frac{\sum_{j=1}^n (y_j - \bar{y})(x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \right)^2 \sum_{j=1}^n (x_j - \bar{x})^2 \\ &= r^2 \sum_{j=1}^n (y_j - \bar{y})^2 \end{aligned}$$

så

$$\text{SS}_E = \text{SS}_{\text{TOT}} - \text{SS}_R = (1 - r^2) \sum_{j=1}^n (y_j - \bar{y})^2.$$

Vi kan här se att $r = 1$ ger perfekt matchning så alla (x_j, y_j) ligger på en rät linje.

10.10 Ett löst exempel



Exempel

Vid ett experiment där man mäter hårdheten på stål som funktion av kolhalten finner man följande data.

Kolhalt (%) x_i	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4
Hårdhet (DPH) y_i	66	116	129	175	209	238	269	301

Använd modellen att y_i är oberoende observationer av $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \sim N(0, \sigma)$.

1. Vad blir ekvationen för den skattade regressionslinjen?
2. Skatta σ^2 med en väntevärdesriktig skattning.
3. Utför hypotestestet $H_0 : \beta_1 = 0$ med mothypotesen $H_1 : \beta_1 > 0$ på nivån 95%.
4. Beräkna ett 99% konfidensintervall för β_1 .
5. Beräkna ett 99% prediktionsintervall för hårdheten när kolhalten är 1.3%.

Typiskt när en liknande uppgift dyker upp på tentan så får man lite räknehjälp:

$$\bar{x} = 0.7, \quad \bar{y} = 187.875,$$

$$\sum_{i=1}^8 (x_i - \bar{x})^2 = 1.680, \quad \sum_{i=1}^8 (y_i - \bar{y})^2 = 45988.9, \quad \sum_{i=1}^8 (x_i - \bar{x})(y_i - \bar{y}) = 277.1,$$

			Variansanalys		
i	$\hat{\beta}_i$	$d(\hat{\beta}_i)$	Frihetsgrader		Kvadratsumma
0	72.42	4.44	REGR	1	45705.01
1	164.94	5.31	RES	6	283.87
			TOT	7	45988.88

$$(X^T X)^{-1} = \begin{pmatrix} 0.4167 & -0.4167 \\ -0.4167 & 0.5952 \end{pmatrix} = \begin{pmatrix} h_{00} & h_{01} \\ h_{10} & h_{11} \end{pmatrix}.$$

Ett par formler som kanske är användbara:

$$\hat{\beta}_i \sim N(\beta_i, \sigma \sqrt{h_{ii}}), \quad \frac{\hat{\beta}_i - \beta_i}{S \sqrt{h_{ii}}} \sim t(n-2), \quad S^2 = \frac{SS_E}{n-2}.$$

Lösning. Syntesmatrisen för vår modell finner vi enligt

$$X = \begin{pmatrix} 1.00 & 0 \\ 1.00 & 0.20 \\ 1.00 & 0.40 \\ 1.00 & 0.60 \\ 1.00 & 0.80 \\ 1.00 & 1.00 \\ 1.00 & 1.20 \\ 1.00 & 1.40 \end{pmatrix}$$

Vi kommer inte behöva använda denna matris direkt, utan kan använda räknehjälpen där vi fått

$$(X^T X)^{-1} = \begin{pmatrix} 0.4167 & -0.4167 \\ -0.4167 & 0.5952 \end{pmatrix}.$$

Om man inte tycker om den framställningen kan man direkt använda formlerna från formelsamlingen (som utgår från S_{xy} etc.). Ni kommer få siffror för att använda båda varianterna.

1. Vi har $\hat{\beta}$ direkt i tabell ovan, så

$$y = \beta_0 + \beta_1 x = 72.42 + 164.94x.$$

Alternativt får vi räkna ut koefficienterna med formelsamlingen:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{277.1}{1.68} = 164.94$$

och

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 72.42.$$

2. Den naturliga skattningen för σ^2 finner vi med

$$s^2 = \frac{SS_E}{n-2} = \frac{283.87}{6} = 47.31,$$

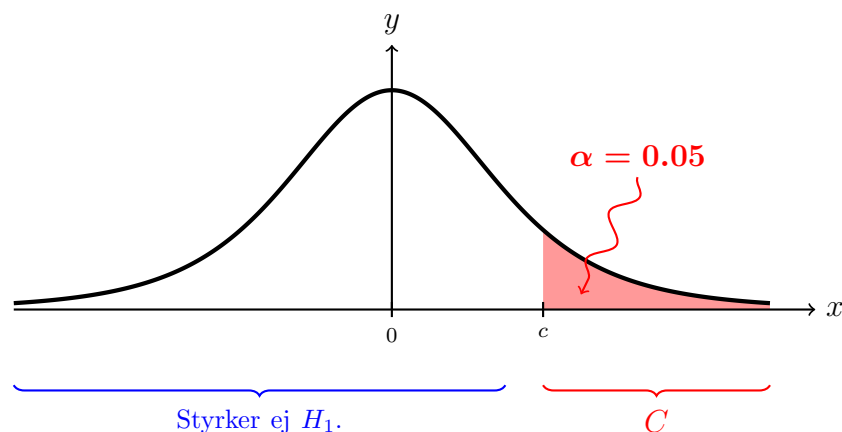
alternativt ur formelsamlingen (med $S_{yy} = SS_{TOT} = 45988.9$, $S_{xx} = 1.68$ och $S_{xy} = 277.1$)

$$s^2 = \frac{S_{yy} - S_{xy}^2/S_{xx}}{n-2} = \frac{45988.9 - 277.1^2/1.680}{6} = 47.32.$$

3. Antag att H_0 är sann. Enligt ovan så gäller då att

$$T = \frac{\hat{\beta}_1 - 0}{S\sqrt{h_{11}}} \sim t(8-2) = t(6).$$

Det kritiska området väljer vi som $C = [c, \infty[$ där c är ett tal så att $P(T > c) = 0.05$.



Ur tabell finner vi $c = 1.94$. Vi har som observation av T

$$t = \frac{164.94}{\sqrt{47.31} \sqrt{0.5952}} = 31.08 \in C$$

så observationen ligger i det kritiska området. Vi förkastar således H_0 och anser att H_1 är styrkt (dvs att $\beta_1 > 0$).

Alternativt så använder vi teststorheten ur formelsamlingen:

$$T = \frac{\hat{\beta}_1 - 0}{\hat{\sigma}/\sqrt{S_{xx}}} \sim t(6)$$

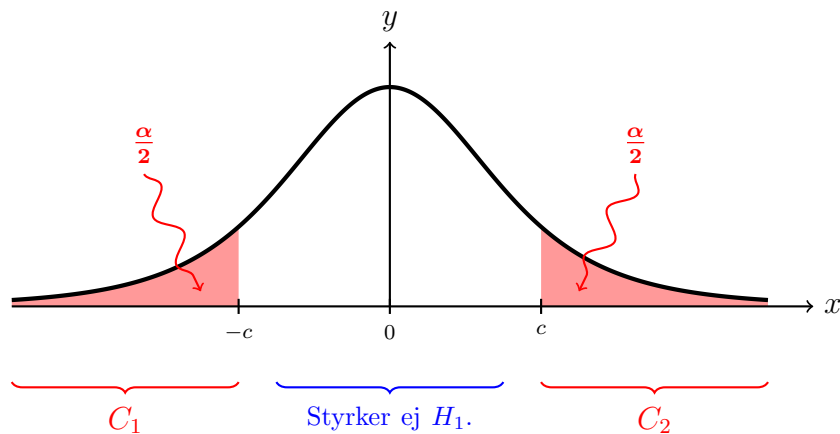
där vi som observation finner att

$$t = \frac{\hat{\beta}_1}{\sqrt{47.32}/\sqrt{1.68}} = \frac{164.94}{\sqrt{47.32}/\sqrt{1.68}} = 31.08.$$

4. För att ställa upp ett konfidensintervall för β_1 så använder vi samma teststorhet som ovan, så

$$T = \frac{\hat{\beta}_1 - \beta_1}{S\sqrt{h_{11}}} \sim t(8-2) = t(6) \quad \text{eller} \quad T = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{S_{xx}}} \sim t(8-2) = t(6).$$

Vi stänger in T så att $P(-c < T < c) = 0.99$ genom att lägga $\alpha/2 = 0.01/2 = 0.005$ i varje svans.



Genom att lösa ut $\hat{\beta}_1$ ur olikheten $-c < T < c$ så finner vi endera att

$$\hat{\beta}_1 - S \cdot c \sqrt{h_{11}} < \beta_1 < \hat{\beta}_1 + S \cdot c \sqrt{h_{11}}$$

eller att

$$\hat{\beta}_1 - S \cdot c / \sqrt{S_{xx}} < \beta_1 < \hat{\beta}_1 + S \cdot c / \sqrt{S_{xx}},$$

beroende på vilken framställning vi använder. Med $c = 3.71$ och $s = \sqrt{47.31} = 6.88$ så får vi intervallet

$$I_{\beta_1} =]145.3, 184.6[.$$

5. Prediktionsintervall (i samband med linjär regression) kommer inte att dyka upp på tentan, men för fullständighetens skull så gör man som i föreläsning 8 men vi behöver hantera lite kovariansproblem som ställer till det. Detta leder till en formel av följande typ:

$$\hat{\beta}_0 + \hat{\beta}_1 \cdot x \pm t_{\alpha/2}(n-2)s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}$$

där s skattas som tidigare och $x = 1.3$ i detta fall.

10.11 (★)Bevis för vissa resultat i linjär regression



Sats. Med förutsättningarna ovan gäller att

$$\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1}).$$

Bevis. Det faktum att $\hat{\beta}$ är normalfördelad följer direkt från faktumet att elementen i $\hat{\beta}$ är linjärkombinationer av normalfördelade variabler. Återstår att visa väntevärde och kovarians:

$$E(\hat{\beta}) = (X^T X)^{-1} X^T E(\mathbf{Y}) = (X^T X)^{-1} X^T X \beta = \beta$$

och

$$\begin{aligned} C_{\hat{\beta}} &= (X^T X)^{-1} X^T C_{\mathbf{Y}} ((X^T X)^{-1} X^T)^T = (X^T X)^{-1} X^T \sigma^2 I_n ((X^T X)^{-1} X^T)^T \\ &= (X^T X)^{-1} X^T \sigma^2 I_n (X^T)^T ((X^T X)^{-1})^T = \sigma^2 (X^T X)^{-1} X^T X ((X^T X)^T)^{-1} = \sigma^2 (X^T X)^{-1}. \end{aligned}$$

Således blir fördelningen precis som beskriven i satsen. \square



Sats. Den totala variationen SS_{TOT} kan delas upp enligt

$$SS_{\text{TOT}} = SS_{\text{R}} + SS_{\text{E}}.$$

Bevis. Vi vill uttrycka SS_{TOT} i termer av SS_{R} och SS_{E} , så

$$\begin{aligned} SS_{\text{TOT}} &= \sum_{j=1}^n (y_j - \bar{y})^2 = \sum_{j=1}^n (y_j - \hat{\mu}_j + \hat{\mu}_j - \bar{y})^2 \\ &= \sum_{j=1}^n (y_j - \hat{\mu}_j)^2 + 2 \sum_{j=1}^n (y_j - \hat{\mu}_j)(\hat{\mu}_j - \bar{y}) + \sum_{j=1}^n (\hat{\mu}_j - \bar{y})^2 \\ &= SS_{\text{R}} + 2 \sum_{j=1}^n (y_j - \hat{\mu}_j) \hat{\mu}_j - 2\bar{y} \sum_{j=1}^n (y_j - \hat{\mu}_j) + SS_{\text{E}}. \end{aligned}$$

Vi visar att de båda summorna i mitten summerar till noll. Eftersom $\hat{\mu} = X\hat{\beta}$ gäller det att

$$\begin{aligned} \sum_{j=1}^n (y_j - \hat{\mu}_j) \hat{\mu}_j &= \hat{\mu}^T (\mathbf{y} - \hat{\mu}) = (X\hat{\beta})^T (\mathbf{y} - X\hat{\beta}) = \hat{\beta}^T X^T (\mathbf{y} - X\hat{\beta}) \\ &= \hat{\beta}^T (X^T \mathbf{y} - X^T X (X^T X)^{-1} X^T \mathbf{y}) = \hat{\beta}^T (X^T \mathbf{y} - X^T \mathbf{y}) = 0. \end{aligned}$$

Med andra ord är $\mathbf{y} - \hat{\mu}$ vinkelrät mot $\hat{\mu}$.

Vidare vet vi att $\hat{\beta}$ minimerar $Q(\beta) = \sum_{j=1}^n (y_j - \mu_j)^2$, så

$$0 = \left. \frac{\partial}{\partial \beta_0} Q(\beta) \right|_{\beta=\hat{\beta}} = -2 \sum_{j=1}^n (y_j - \hat{\mu}_j) \quad \Leftrightarrow \quad \sum_{j=1}^n y_j = \sum_{j=1}^n \hat{\mu}_j,$$

vilket var precis det vi behövde. \square

10.12 (★)Regressionsanalysens huvudsats

Eftersom vi arbetar med matriser och MK-lösningen i princip är projektionen på ett underrum av \mathbf{R}^n så är följande resultat inte allt för förvånande.



Hatt-matrisen

Definition. Vi definierar $H = X(X^T X)^{-1} X^T$. Vi definierar även J som en matris vars samtliga element är 1. Vi skriver J_{nm} om vi vill markera dimensionen.

Varför kallar vi H för hatt-matrisen? Ganska enkelt:

$$H\mathbf{y} = X(X^T X)^{-1} X^T \mathbf{y} = X\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\mu}} = \hat{\mathbf{y}}.$$

Avbildningen sätter alltså hatten på!



Sats. Matriserna H , $I - H$ och $H - \frac{1}{n}J$ är projektionsmatriser P som uppfyller $P^2 = P^T = P$.

Bevis. Direkt från definitionen av H blir

$$H^2 = (X(X^T X)^{-1} X^T)(X(X^T X)^{-1} X^T) = X(X^T X)^{-1} X^T = H$$

och

$$H^T = (X(X^T X)^{-1} X^T)^T = X(X^T X)^{-T} X^T = X((X^T X)^T)^{-1} X^T = H.$$

Därför följer det att $(I - H)^2 = I^2 - 2H + H^2 = I - H$ och att $(I - H)^T = I^T - H^T = I - H$. För den sista operatören skriver vi

$$\left(H - \frac{1}{n}J\right)^2 = H^2 - \frac{1}{n}HJ - \frac{1}{n}JH + \frac{1}{n^2}J^2 = H - \frac{1}{n}HJ - \frac{1}{n}JH + \frac{1}{n}J$$

ty J^2 är matrisen med samtliga element lika med n . Vidare gäller att J kommuterar med alla kvadratiske matriser, så $JH = HJ$. Här kan vi se att $H\mathbf{1} = \mathbf{1}$ (där $\mathbf{1}$ är en vektor med samtliga element lika med 1) eftersom lösningen till regressionsproblemet om \mathbf{y} är konstant är just den konstanten (lösningen är exakt). Alltså blir

$$\left(H - \frac{1}{n}J\right)^2 = H - \frac{1}{n}J.$$

Givetvis gäller även att $\left(H - \frac{1}{n}J\right)^T = H - \frac{1}{n}J$. □

En följsats av detta är att vi kan skriva kvadratsummorna som kvadratiske former.



Sats. $SS_{\text{TOT}} = \mathbf{y}^T \left(I - \frac{1}{n}J\right) \mathbf{y}$, $SS_{\text{R}} = \mathbf{y}^T \left(H - \frac{1}{n}J\right) \mathbf{y}$, samt $SS_{\text{E}} = \mathbf{y}^T (I - H) \mathbf{y}$.

Innan vi ger oss in på huvudsatsen visar vi en hjälpsats från linjär algebra.



Sats. Låt $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)$ vara sådan att

$$\mathbf{x}^T \mathbf{x} = \sum_{j=1}^n x_j^2 = \mathbf{x}^T A \mathbf{x} + \mathbf{x}^T B \mathbf{x}$$

för $A, B \in \mathbf{R}^{n \times n}$ positivt semi-definita och symmetriska med $\text{rank}(A) = r$ och $\text{rank}(B) = n - r$ för något heltal r så att $0 < r < n$. Då finns en ON-matris C så att med $\mathbf{x} = C\mathbf{y}$ gäller att

$$\mathbf{x}^T A \mathbf{x} = \sum_{j=1}^r y_j^2 \quad \text{och} \quad \mathbf{x}^T B \mathbf{x} = \sum_{j=r+1}^n y_j^2.$$

Bevis. Eftersom A är positivt semi-definit har A endast icke-negativa egenvärden som vi ordnar enligt $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Vi vet även att $\text{rank}(A) = r$, så $\lambda_{r+1} = \cdots = \lambda_n = 0$. Vidare är A diagonaliserbar eftersom A är symmetrisk, så det finns en ON-matris C så att

$$C^T A C = D = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \vdots & \ddots & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Med $\mathbf{x} = C\mathbf{y}$ gäller då att

$$\mathbf{x}^T \mathbf{x} = (C\mathbf{y})^T (C\mathbf{y}) = \mathbf{y}^T C^T C \mathbf{y} = \mathbf{y}^T \mathbf{y}$$

och

$$\mathbf{x}^T A \mathbf{x} = (C\mathbf{y})^T A C \mathbf{y} = \mathbf{y}^T C^T A C \mathbf{y} = \mathbf{y}^T D \mathbf{y} = \sum_{j=1}^r \lambda_j y_j^2.$$

Vi har även

$$\mathbf{x}^T B \mathbf{x} = (C\mathbf{y})^T B C \mathbf{y} = \mathbf{y}^T C^T B C \mathbf{y}$$

och då blir

$$\sum_{j=1}^n y_j^2 = \mathbf{y}^T \mathbf{y} = \mathbf{x}^T \mathbf{x} = \mathbf{x}^T A \mathbf{x} + \mathbf{x}^T B \mathbf{x} = \sum_{j=1}^r \lambda_j y_j^2 + \mathbf{y}^T C^T B C \mathbf{y}$$

så

$$\mathbf{y}^T C^T B C \mathbf{y} = \sum_{j=1}^r (1 - \lambda_j) y_j^2 + \sum_{j=r+1}^n y_j^2.$$

Det faktum att $\text{rank}(B) = n - r$ visar att $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 1$ och vi erhåller identiteten

$$\mathbf{y}^T C^T B C \mathbf{y} = \sum_{j=r+1}^n y_j^2.$$

Alla beteckningar är nu ur vägen och vi är framme vid huvudresultatet.



Regressionsanalysens huvudsats

Sats. Med förutsättningarna ovan gäller följande för SS_E och SS_R sedda som stokastiska variabler.

$$(i) \quad \frac{SS_E}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^n (Y_j - \hat{\mu}_j)^2 \sim \chi^2(n-2).$$

$$(ii) \quad \text{Givet att } \beta_1 = 0 \text{ är } \frac{SS_R}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^n (\hat{\mu}_j - \bar{Y})^2 \sim \chi^2(1).$$

(iii) Både SS_R och $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{Y}$ är oberoende av SS_E .

Bevis. Eftersom H är en projektionsmatris har H endast egenvärdena $\lambda = 0$ och $\lambda = 1$. Således gäller det finurliga att matrisens rang är lika med summan av egenvärdena, vilka kan beräknas genom att ta spåret¹ av matrisen:

$$\text{rank}(H) = \text{tr}(H) = \text{tr}(X(X^T X)^{-1} X^T) = \text{tr}(X^T X (X^T X)^{-1}) = \text{tr}(I_2) = 2,$$

Det följer sedan att

$$\text{rank}(I - H) = n - 2.$$

Eftersom $HX = X$ så gäller att

$$\mathbf{Y}^T (I - H) \mathbf{Y} = \boldsymbol{\epsilon}^T (I - H) \boldsymbol{\epsilon}.$$

Detta är användbart eftersom $E(\boldsymbol{\epsilon}) = \mathbf{0}$. Enligt föregående hjälpsats gäller att sambandet

$$\boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \boldsymbol{\epsilon}^T (I - H) \boldsymbol{\epsilon} + \boldsymbol{\epsilon}^T H \boldsymbol{\epsilon}$$

medför att det finns en ON-matris C så att $\mathbf{Z} = C\boldsymbol{\epsilon}$ reducerar likheten till

$$\boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \sum_{j=1}^{n-2} Z_j^2 + \sum_{j=n-1}^n Z_j^2$$

där

$$\boldsymbol{\epsilon}^T (I - H) \boldsymbol{\epsilon} = \mathbf{Y}^T (I - H) \mathbf{Y} = SS_E = \sum_{j=1}^{n-2} Z_j^2.$$

Eftersom komponenterna i $\boldsymbol{\epsilon}$ är oberoende och $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I)$ följer det att

$$E(\mathbf{Z}) = C\mathbf{0} = \mathbf{0} \quad \text{och} \quad C_Z = C_{C\boldsymbol{\epsilon}} = \sigma^2 C^T I C = \sigma^2 I$$

så $\mathbf{Z} \sim N(\mathbf{0}, \sigma^2 I)$. Komponenterna i \mathbf{Z} är alltså oberoende och $Z_j \sim N(0, \sigma^2)$. Detta medför att

$$\frac{SS_E}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^{n-2} Z_j^2 \sim \chi^2(n-2).$$

¹se sista (bonus)avsnittet för lite detaljer kring spår av matriser.

Vi erhåller även att SS_E och $H\epsilon$ är oberoende (de delar inga variabler Z_j). Detta medför att SS_E och $\hat{\beta}$ är oberoende. Det faktum att SS_E och SS_R är oberoende följer av att kovariansen

$$\begin{aligned} C\left(\left(H - \frac{1}{n}J\right)\mathbf{Y}, (I - H)\mathbf{Y}\right) &= \left(H - \frac{1}{n}J\right) C(\mathbf{Y}, \mathbf{Y})(I - H)^T = \sigma^2 \left(H - \frac{1}{n}J\right) (I - H) \\ &= \sigma^2 \left(H - H^2 - \frac{1}{n}J + \frac{1}{n}JH\right) = \sigma^2 \left(-\frac{1}{n}J + \frac{1}{n}HJ\right) \\ &= \sigma^2 \left(-\frac{1}{n}J + \frac{1}{n}J\right) = 0, \end{aligned}$$

så dessa vektorer är okorrelerade och normalfördelade, så oberoende. Det följer direkt att SS_E och SS_R är oberoende (som funktioner av oberoende variabler).

Om vi fokuserar på fördelningen för SS_R så kan vi på samma sätt som ovan utnyttja att $(H - \frac{1}{n}J)$ är en projektionsmatris, så

$$\text{rank}\left(H - \frac{1}{n}J\right) = \text{tr}\left(H - \frac{1}{n}J\right) = \text{tr}(H) - \text{tr}\left(\frac{1}{n}J\right) = 1 + 1 - 1 = 2.$$

Matrisen J är synnerligen rangdefekt med $\text{rank}(J) = 1$ (eftersom alla kolonner är lika spänner vi bara upp ett en-dimensionellt rum).

Nu stöter vi på lite problem. Vi ser att

$$\left(H - \frac{1}{n}J\right)\mathbf{Y} = \left(I - \frac{1}{n}J\right)X\beta + \left(H - \frac{1}{n}J\right)\epsilon$$

där den första termen *inte* försvinner såvida inte $\beta_1 = 0$. Men under detta antagande har vi åter igen en situation där vi kan ”byta ut” \mathbf{Y} mot ϵ . Föregående hjälpsats visar – analogt med föregående argument – att

$$\mathbf{Y}^T \left(H - \frac{1}{n}J\right) \mathbf{Y} = \sum_{j=1}^1 Z_j^2,$$

där Z_j är oberoende och $Z_j \sim N(0, \sigma^2)$, vilket medför att $\frac{1}{\sigma^2}SS_R \sim \chi^2(k)$, under förutsättningen att $\beta_1 = \beta_2 = \dots = \beta_k = 0$. \square

Kommentar: utan villkoret $\beta_1 = \beta_2 = \dots = \beta_k = 0$ kan man fortfarande genomföra argumentet, men resultatet blir en icke-centrerad $\chi^2(k)$ -fördelning (något som inte ingår i kursen).

10.13 (★) Bevis av χ^2 -testet



Sats. Med beteckningarna ovan gäller att $\sum_{j=1}^k \frac{(Y_j - np_j)^2}{np_j} \xrightarrow{D} X$, där $X \sim \chi^2(k-1)$. Konvergensen är alltså i fördelning.

Bevis. Eftersom Y_j är binomialfördelad vet vi att $E(Y_j) = np_j$ och $V(Y_j) = np_j(1 - p_j)$, så de standardiserade variablerna

$$\frac{Y_j - np_j}{\sqrt{np_j(1 - p_j)}} \xrightarrow{D} \widetilde{Z}_j \sim N(0, 1),$$

för något \widetilde{Z}_j enligt centrala gränsvärdessatsen (CGS). Konvergensen är i meningen att fördelningsfunktionen $F_{n,j}(y) \rightarrow \Phi(y)$ för alla $y \in \mathbf{R}$. En följd av detta är att

$$\frac{Y_j - np_j}{\sqrt{np_j}} \xrightarrow{D} Z_j \sim N(0, 1 - p_j),$$

eftersom om $U_n \xrightarrow{D} U$ så gäller att $h(U_n) \xrightarrow{D} h(U)$ för alla kontinuerliga funktioner h (brukar kallas sannolikhetsteorins open mapping theorem). Anledningen till den sista manövern är att vi ska få det lite lättare att analysera beroendestrukturen hos Z_j , $j = 1, 2, \dots, k$. Eftersom väntevärdet är $E(Y_j) = np_j$ kommer

$$\begin{aligned} C\left(\frac{Y_i - np_i}{\sqrt{np_i}}, \frac{Y_j - np_j}{\sqrt{np_j}}\right) &= E\left(\frac{Y_i - np_i}{\sqrt{np_i}} \frac{Y_j - np_j}{\sqrt{np_j}}\right) = \frac{1}{n\sqrt{p_i p_j}} (E(Y_i Y_j) - 2n^2 p_i p_j + n^2 p_i p_j) \\ &= \frac{1}{n\sqrt{p_i p_j}} (E(Y_i Y_j) - n^2 p_i p_j) \end{aligned}$$

För att beräkna $E(Y_i Y_j)$ går vi tillbaka till variablerna X_i , $i = 1, 2, \dots, n$. Låt I_A beteckna indikatorfunktionen för mängden A . Detta innebär att

$$I_{A_j}(X_i) = \begin{cases} 1 & \text{om } X_i \in A_j, \\ 0 & \text{om } X_i \notin A_j. \end{cases}$$

Vi kan då skriva $Y_j = \sum_{i=1}^n I_{A_j}(X_i)$ och eftersom X_i är Bernoullifördelade (2-punktsfördelade) följer det att $E(I_{A_j}(X_i)) = p_j$. Vi har nu, för $i \neq j$,

$$\begin{aligned} E(Y_i Y_j) &= E\left(\left(\sum_{l=1}^n I_{A_i}(X_l)\right) \left(\sum_{m=1}^n I_{A_j}(X_m)\right)\right) = E\left(\sum_{l=1}^n \sum_{m=1}^n I_{A_i}(X_l) I_{A_j}(X_m)\right) \\ &= E\left(\sum_{l=1}^n I_{A_i}(X_l) I_{A_j}(X_l)\right) + E\left(\sum_{l=1}^n \sum_{\substack{m=1 \\ m \neq l}}^n I_{A_i}(X_l) I_{A_j}(X_m)\right) \\ &= 0 + \sum_{l=1}^n \sum_{\substack{m=1 \\ m \neq l}}^n E(I_{A_i}(X_l) I_{A_j}(X_m)) = \sum_{l=1}^n \sum_{\substack{m=1 \\ m \neq l}}^n p_i p_j = n(n-1)p_i p_j, \end{aligned}$$

eftersom $I_{A_i}(X_l) I_{A_j}(X_l) = 0$ (samma boll kan inte hamna i två lådor) samt att $I_{A_i}(X_l)$ och $I_{A_j}(X_m)$ är oberoende om $l \neq m$. Således blir

$$C\left(\frac{Y_i - np_i}{\sqrt{np_i}}, \frac{Y_j - np_j}{\sqrt{np_j}}\right) = -\sqrt{p_i p_j},$$

för $i \neq j$. Följaktligen måste således kovariansmatrisen för $\mathbf{Z} = (Z_1 \ Z_2 \ \cdots \ Z_k)^T$ ha utseendet

$$C_{\mathbf{Z}} = \begin{pmatrix} 1 - p_1 & -\sqrt{p_1 p_2} & -\sqrt{p_1 p_3} & \cdots & -\sqrt{p_1 p_k} \\ -\sqrt{p_2 p_1} & 1 - p_2 & -\sqrt{p_2 p_3} & \cdots & -\sqrt{p_2 p_k} \\ -\sqrt{p_3 p_1} & -\sqrt{p_3 p_2} & 1 - p_3 & \cdots & -\sqrt{p_3 p_k} \\ \vdots & \vdots & & \ddots & \vdots \\ -\sqrt{p_k p_1} & -\sqrt{p_k p_2} & -\sqrt{p_k p_3} & \cdots & 1 - p_k \end{pmatrix}.$$

vilket kan skrivas lite mer kompakt som $C_{\mathbf{Z}} = I - \mathbf{p}\mathbf{p}^T$, där $\mathbf{p} = (\sqrt{p_1} \ \sqrt{p_2} \ \cdots \ \sqrt{p_k})^T$. Denna omskrivning gör att vi enkelt kan se att

$$(I - \mathbf{p}\mathbf{p}^T)^2 = I - \mathbf{p}\mathbf{p}^T \quad \text{och} \quad (I - \mathbf{p}\mathbf{p}^T)^T = I - \mathbf{p}\mathbf{p}^T,$$

så $I - \mathbf{p}\mathbf{p}^T$ är en projektionsmatris och har därför egenvärdena $\lambda = 0$ och $\lambda = 1$. Vi har nu att $\mathbf{Z} \sim N(\mathbf{0}, C_{\mathbf{Z}})$. På samma sätt som i beviset av regressionsanalysens huvudsats ser vi att

$$\text{rank}(I - \mathbf{p}\mathbf{p}^T) = \text{tr}(I - \mathbf{p}\mathbf{p}^T) = k - 1,$$

så $\lambda = 0$ är ett enkelt egenvärde. Matrisen är symmetrisk och positivt semidefinit, så det finns en ON-matris C så att $C^T C_{\mathbf{Z}} C = \text{diag}(1, 1, \dots, 1, 0)$ blir en diagonalmatris. Om vi låter $\mathbf{W} = C\mathbf{Z}$ ser vi att $\mathbf{W} \sim N(\mathbf{0}, \text{diag}(1, 1, \dots, 1, 0))$ och att

$$\mathbf{Z}^T \mathbf{Z} = \mathbf{W}^T \mathbf{W} = \sum_{j=1}^{k-1} W_j^2,$$

där $W_j \sim N(0, 1)$ är oberoende. Denna summa är som bekant $\chi^2(k-1)$ -fördelad! □