Föreläsning 1: Repetition och punktskattningar

Johan Thim (johan.thim@liu.se)

29 augusti 2018

1 Repetition

Vi repeterar lite kort de viktigaste begreppen från sannolikhetsläran. Materialet kan återfinnas i mer fullständig form (med exempel och mer diskussion) i föreläsningsanteckningarna för kursen TAMS79.

1.1 Sannolikhetsteorins grunder

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.

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 kommer komplementet $A^* \in \mathcal{F}$.

(iii) om $A_1, A_2, \ldots \in \mathcal{F}$ så är unionen $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.

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 för många tillämpningar.

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; $P\colon \mathcal{F} \to [0,1]$. Sannolikhetsmåttet P *måste* uppfylla Kolmogorovs axiom:

(i) $0 \leq P(A) \leq 1$ för varje $A \in \mathcal{F}$.

(ii) $P(\Omega) = 1$.

(iii) Om $A \cap B = 0$ så gäller att $P(A \cup B) = P(A) + P(B)$.

Oberoende

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

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 $\mathscr{B}$. Denna algebra brukar kallas för **Borel-σ-algebran** på $\mathbf{R}$.

Algebran $\mathscr{B}$ 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 uppräkneliga 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; $X : \Omega \to \mathbf{R}$.
Mer precist så kräver vi att $X^{-1}(B) \in \mathcal{F}$ för alla $B \in \mathscr{B}$. 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. **Bilden** av en delmängd A av Ω betecknas med $X(A)$, och

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

Mängden $X(A)$ är alltså värdemängden för X på mängden A.
Om $X(\Omega)$ är ändlig, eller bara har uppräkneligt många värden, så kallar vi X för en **diskret stokastisk variabel**. Annars kallas vi X för **kontinuerlig**.

Uttrycket $X(\omega)$ är alltså det siffervärde vi sätter på ett visst utfall $\omega \in \Omega$. 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.
För variabler i högre dimension fokuserar vi 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. I $\mathbf{R}^2$ är Borelfamiljen $\mathscr{B}$ den minsta σ-algebra som innehåller alla öppna rektanglar $(a, b) \times (c, d)$. Generaliserar naturligt till högre dimensioner.

Flerdimensionell stokastisk variabel

Definition. En tvådimensionell stokastisk variabel är en reell-vektorvärd funktion (X, Y) definierad på ett utfallsrum Ω. (X, Y) avbildar alltså olika utfall på reella vektorer; $(X, Y) : \Omega \to \mathbf{R}^2$. Vi kräver att $(X, Y)^{-1}(B) \in \mathcal{F}$ för alla $B \in \mathscr{B}$. Algebran $\mathcal{F}$ är mängden av alla tillåtna händelser. Om (X, Y) bara antar ändligt eller uppräkneligt 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!

Vad måste jag förstå av all matematiska?

- En stokastisk variabel är en *snäll* funktion från Ω till $\mathbf{R}^n$.
- Utfallsrummet Ω kan vara abstrakt, e.g., $\Omega = \{\text{Krona, Klave}\}$.
- Det är mängden $X(\Omega)$ som består av siffror (vektorer 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.
- Urbilden $X^{-1}(B)$ av en delmängd $B \subset \mathbf{R}^n$ består av *alla* utfall $\omega \in \Omega$ så att siffran (vektorn) $X(\omega)$ ligger i mängden B.

1.2 Beskrivningar av stokastiska variabler

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

Sannolikhetsfunktion

Definition. Sannolikhetsfunktionen $p_X \colon X(\Omega) \to [0,1]$ 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) $\displaystyle\sum_{k \in X(\Omega)} p_X(k) = 1.$

(iii) Om $A \subset X(\Omega)$ så är $P(X \in A) = \displaystyle\sum_{k \in A} p_X(k).$

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 enligt sambandet $F_X(x) = P(X \leq x)$ för alla $x \in \mathbf{R}$.

Det följer från definitionen att följande påståenden gäller.

Egenskaper hos fördelningsfunktionen

(i) $F_X(x) \to \begin{cases} 0, & x \to -\infty, \\ 1, & x \to +\infty. \end{cases}$

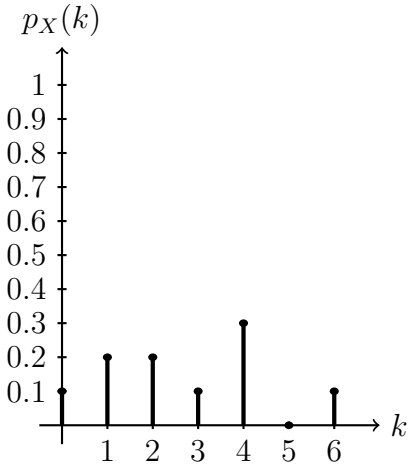
(ii) $F_X(x)$ är icke-avtagande och högerkontinuerlig.

(iii) $F_X(x) = \sum_{\{k \in X(\Omega): k \leq x\}} p_X(k)$.

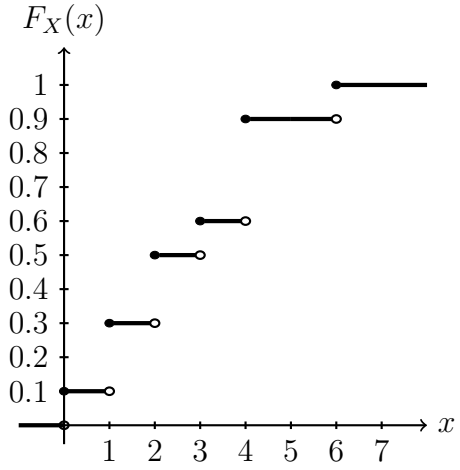
(iv) $P(X > x) = 1 - F_X(x)$.

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

1.3 Kontinuerliga stokastiska variabler

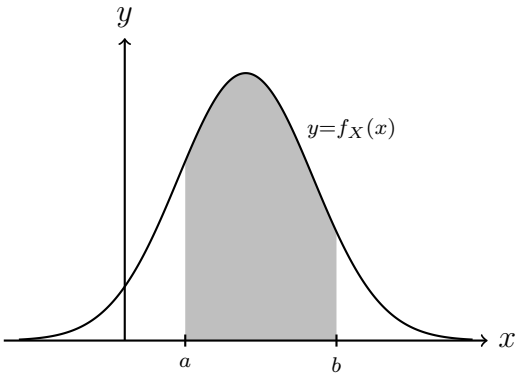
Tähetsfunktion

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

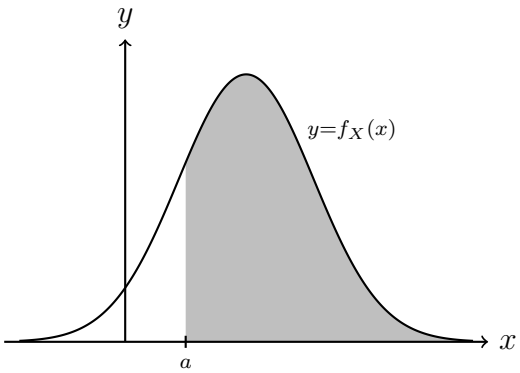
$$P(a < X < b) = \int_a^b f_X(x)\,dx$$

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

Exempel:



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



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

Egenskaper hos täthetsfunktionen

(i) $f_X(x) \geq 0$ för alla $x \in \mathbf{R}$.

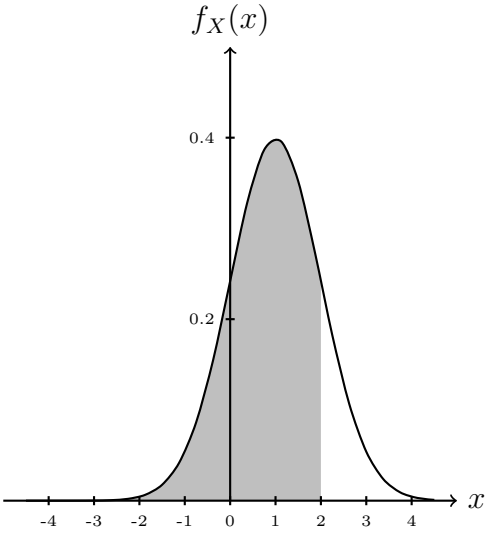
(ii) $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

(iii) $f_X(x)$ anger hur mycket sannolikhetsmassa det finns per längdenhet i punkten x.

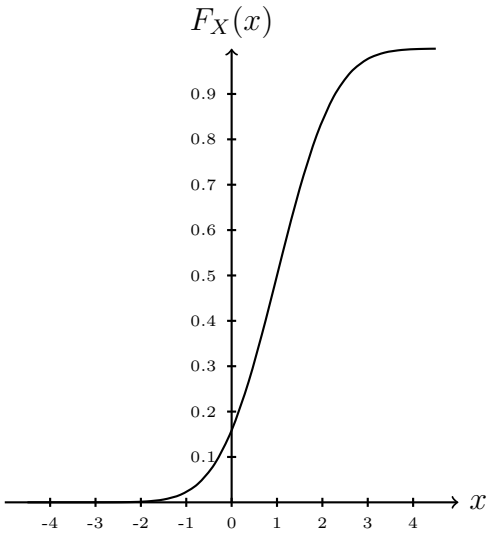
Vi definierar **fördelningsfunktionen** $F_X(x)$ på samma sätt som i det diskreta fallet, och finner att

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

Fördelningsfunktionen uppfyller (i)–(iii) från det diskreta fallet, och i alla punkter där $f_X(x)$ är kontinuerlig gäller dessutom att $F_X'(x) = f_X(x)$.
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 gränsvärderna mot $\pm\infty$ verkar stämma!

Väntevärde

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

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

för kontinuerliga och diskreta variabler.

Andra vanliga beteckningar: μ eller μ_X. Väntevärdet är ett lägesmått som anger vart sannolikhetsmassan har sin tyngdpunkt (jämför med mekanikens beräkningar av tyngdpunkt).

1.4 Högre dimensioner

Sannolikhetsfunktionen för en diskret 2D-variabel ges av $p_{X,Y}(j,k) = P(X = j, Y = k)$.

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 $P(X \in A) = \sum_{(j,k)\in A} \sum p_{X,Y}(j,k)$.

Analogt med en dimension kan man introducera begreppet täthetsfunktion för en kontinuerlig 2D-variabel.

Egenskaper hos den simultana täthetsfunktionen

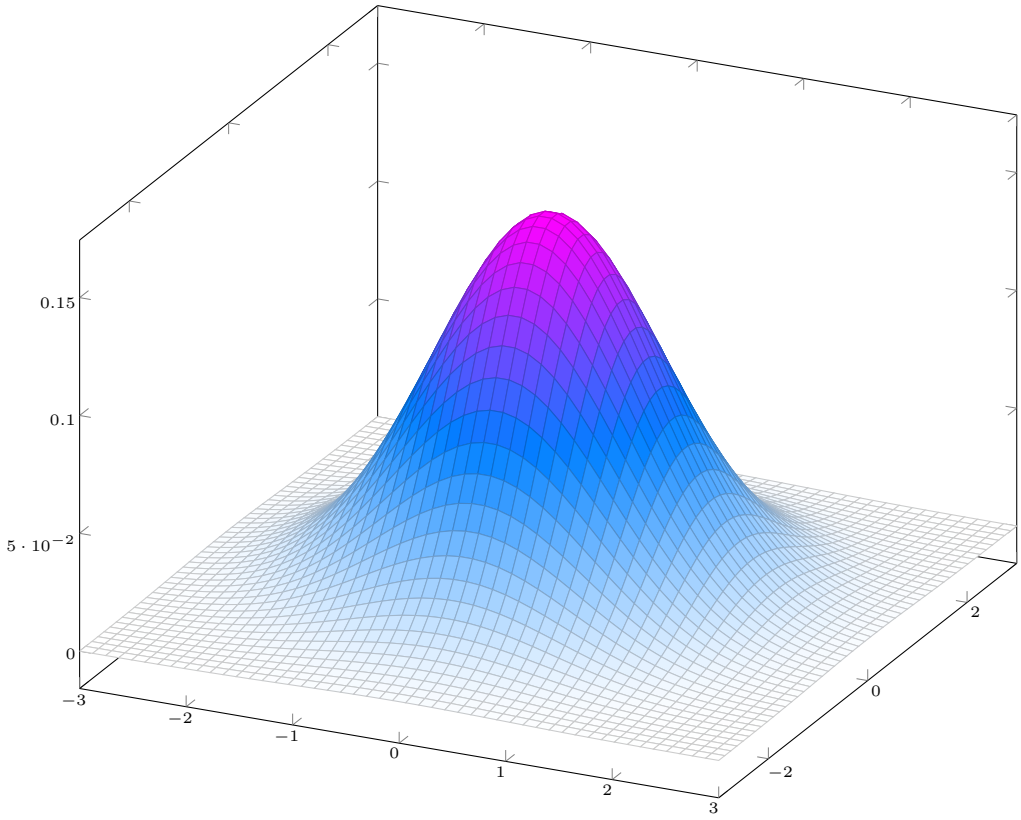
(i) $f_{X,Y}(x,y) \geq 0$ för alla $(x,y) \in \mathbf{R}^2$.

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy = 1$.

(iii) Om $A \in \mathscr{B}$ (så $A \subset \mathbf{R}^2$ är *snäll*) så är $P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dxdy$.

(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.



Väntevärde och funktioner av stokastiska variabler

Sats. Låt $Y = g(X)$ och $W = h(X_1, X_2, \ldots, X_n)$. I de kontinuerliga fallen blir

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

och

$$E(W) = \int \cdots \int_{\mathbf{R}^n} h(x_1, x_2, \ldots, x_n) f_{(x_1,\ldots,x_n)}(x_1, x_2, \ldots, x_n)\, dx_1 \cdots dx_n.$$

Det diskreta fallet är analogt.

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 $D(X) = \sqrt{V(X)}$.

Steiners sats

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

För väntevärdet gäller bland annat följande regler.

Linjäritet och oberoende produkt

Låt $X_1, X_2, \ldots, X_n$ vara stokastiska variabler. Då gäller

(i) $E\left(\sum_{k=1}^{n} c_k X_k\right) = \sum_{k=1}^{n} c_k E(X_k)$ för alla $c_1, c_2, \ldots, c_n \in \mathbf{R}$;

(ii) Om X_i och X_j är oberoende gäller $E(X_i X_j) = E(X_i)E(X_j)$.

(iii) $V(aX_i + b) = a^2 V(X_i)$ för alla $a, b \in \mathbf{R}$;

(iv) $V(aX_i \pm bX_j) = a^2 V(X_i) + b^2 V(X_j) + 2ab(E(X_i X_j) - E(X_i)E(X_j))$ för alla $a, b \in \mathbf{R}$;

(v) Om $X_1, X_2, \ldots, X_n$ är *oberoende* stokastiska variabler är $V\left(\sum_{k=1}^{n} c_k X_k\right) = \sum_{k=1}^{n} c_k^2 V(X_k)$ för alla $c_1, c_2, \ldots, c_n \in \mathbf{R}$;

Varianser adderas alltid!

Observera att det *alltid* blir ett plustecken mellan varianserna: $V(aX \pm bY) = a^2 V(X) + b^2 V(Y)$. Vi kommer *aldrig* att bilda skillnader mellan varianser!

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

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

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

Motsvarande gäller om (X, Y) är diskret:

$$p_X(j) = \sum_{k} p_{X,Y}(j, k) \qquad \text{och} \qquad p_Y(k) = \sum_{j} p_{X,Y}(j, k).$$

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

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:

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

2 Normalfördelning

Normalfördelningen är så viktig att den får ett eget avsnitt. Se till att ni verkligen kommer ihåg hur man hanterar normalfördelning, mycket är vunnet senare om detta maskineri sitter bra.

Normalfördelning

Variabeln X kallas **normalfördelad** med parametrarna μ och σ, $X \sim \mathrm{N}(\mu, \sigma^2)$, om

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

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

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

Standardisering av variabel

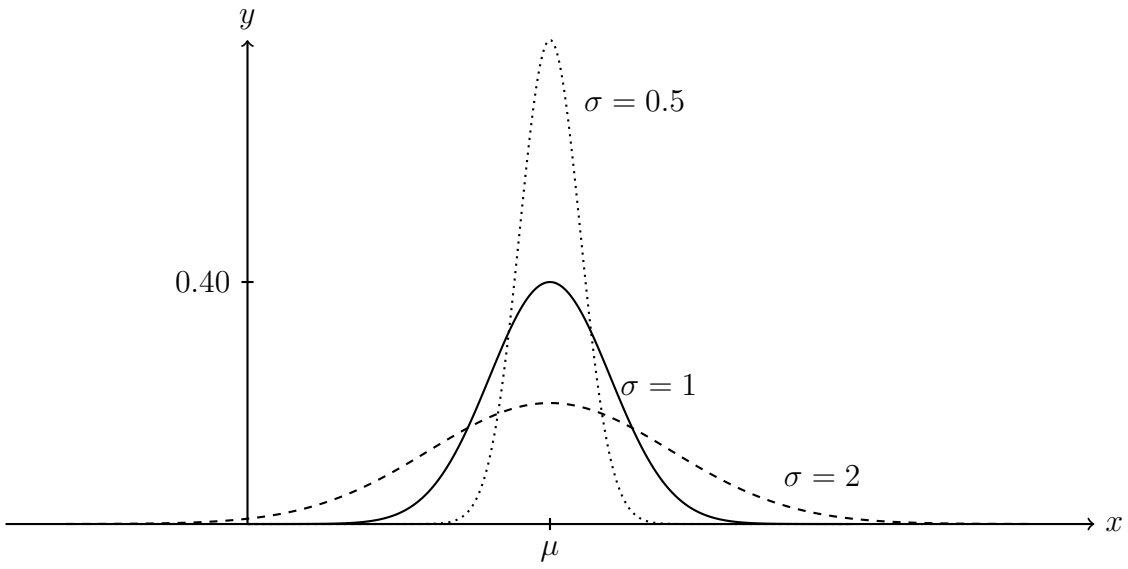
Om X är en stokastisk variabel med $E(X) = \mu$ och $V(X) = \sigma^2$, så är $Z = (X - \mu)/\sigma$ en stokastisk variabel med $E(Z) = 0$ och $V(Z) = 1$. Vi kallar Z för **standardiserad**.

$$X \sim \mathbf{N}(\mu, \sigma^2)$$

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

Standardavvikelse eller varians?

I kursboken (Blom et al) används beteckningen $X \sim N(\mu, \sigma)$, så den andra parametern är alltså standardavvikelsen σ, inte variansen σ^2 som vi använt ovan.



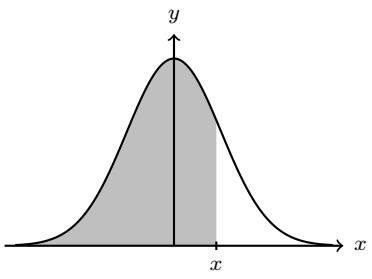
Bruk av tabell för $\Phi(x)$

Låt $X \sim \mathrm{N}(0,1)$. Då gäller

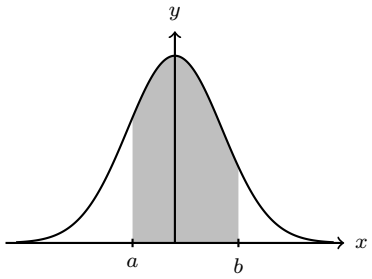
(i) $P(X \leq x) = \Phi(x)$ för alla $x \in \mathbf{R}$;

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

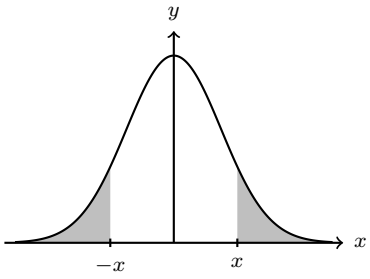
(iii) $\Phi(-x) = 1 - \Phi(x)$ för alla $x \in \mathbf{R}$.



$\Phi(x)$



$\Phi(b) - \Phi(a)$



$\Phi(-x) = 1 - \Phi(x)$

Exempel

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

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

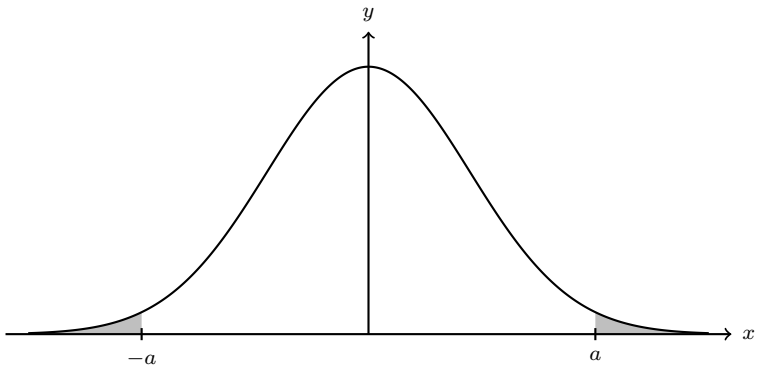
Standardisering av normalfördelning

Sats. $X \sim \mathrm{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad Z = \dfrac{X - \mu}{\sigma} \sim \mathrm{N}(0, 1)$.

Exempel

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

Situationen ser ut som i bilden nedan. De skuggade områderna 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 $P(X \leq a) = 0.975$.

Summor och medelvärde

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

$$X := \sum_{k=1}^{n} X_k \sim \mathrm{N}(n\mu, n\sigma^2) \qquad \text{och} \qquad \overline{X} := \frac{1}{n}\sum_{k=1}^{n} X_k \sim \mathrm{N}(\mu, \sigma^2/n).$$

Mer generellt, om $X_k \sim \mathrm{N}(\mu_k, \sigma_k^2)$ och $c_0, c_1, \ldots, c_n \in \mathbf{R}$ är

$$c_0 + \sum_{k=1}^{n} c_k X_k \sim \mathrm{N}\left(c_0 + \sum_{k=1}^{n} c_k \mu_k, \sum_{k=1}^{n} c_k^2 \sigma_k^2\right).$$

Likheten för medelvärdet 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:

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

eftersom variablerna är oberoende och $V(X_k) = \sigma^2$ för alla k.
Nottera även att satsen faktiskt säger att summan av normalfördelade variabler fortfarande är normalfördelad, något som inte gäller vilken fördelning som helst (se faltningssatsen).

Statistisk inferens

"We have such sights to show you"
–Pinhead

3 Begrepp

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 vi diskuterar.

Stickprov

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

Exempel

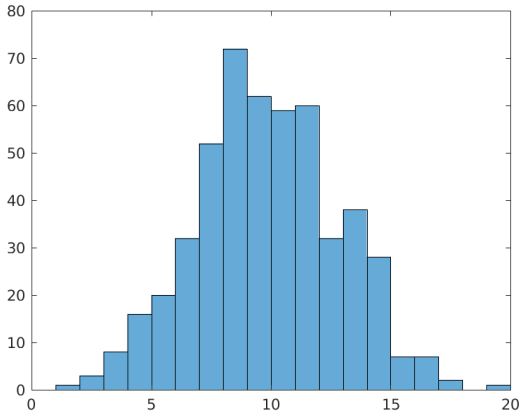
Antag att vi kastar en perfekt tärning 5 gånger och att dessa kast är oberoende. Före kasten representerar den stokastiska variabeln X_k resultatet vid kast k där alla X_k har samma fördelning; vi vet ännu inte vad resultatet blir, men känner sannolikhetsfördelningen. Följden X_k, $k = 1, 2, \ldots, 5$ är det *slumpmässiga stickprovet*.
Efter kasten har vi erhållit observationer $x_1, x_2, \ldots, x_5$ av det slumpmässiga stickprovet. Dett 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 följden av stokastiska variabler och följden av observationer (utfall). Det viktiga är att hålla koll på vad ni själva menar när ni genomför analyser.

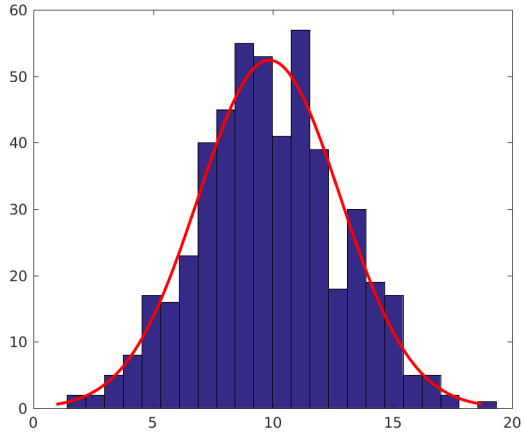
4 Representation av stickprov

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);
```

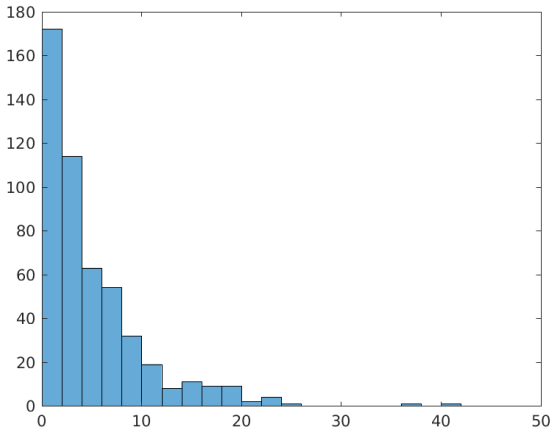



```
>> U = normrnd(10,3,500,1);
>> histfit(U)
```



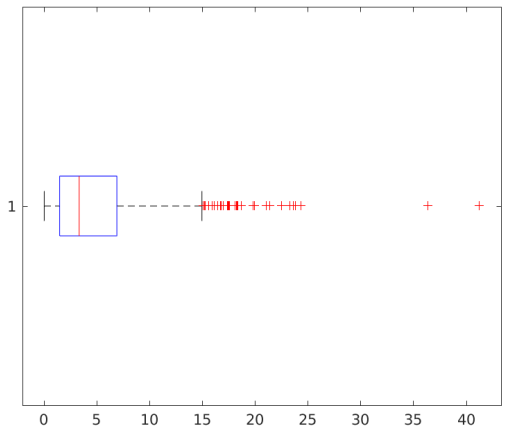
Om vi testar med exponentialfördelning istället blir resultatet enligt nedan.

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

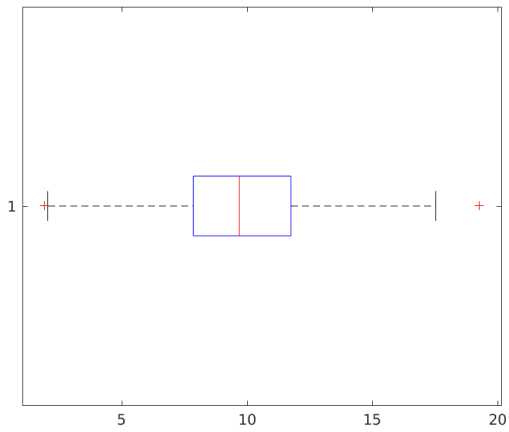


Ett lådagram (boxplot) representerar också materialet, kanske på ett enklare sätt för den oinsatte i sannolikhetsfördelningar. Lådan innehåller 50% av resultaten och den vänstra lådkanten är den undre kvartilen (25% till vänster om den) och den högra är den övre kvartilen (med 25% till höger om den). 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');
```

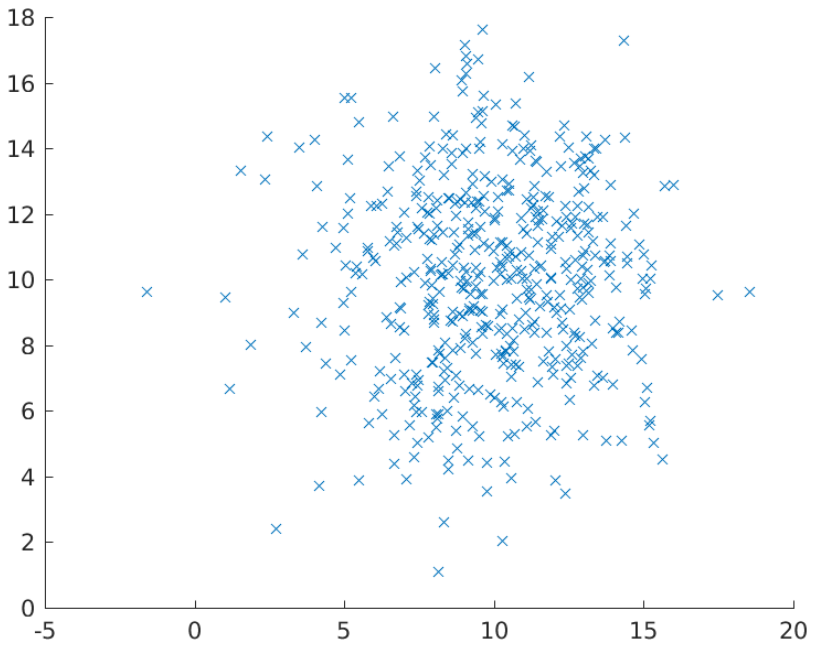


Vi kan tydligt se skillnad på hur mätvärden är spridda. Jämför även med motsvarande histogram ovan.

4.1 Två-dimensionell data

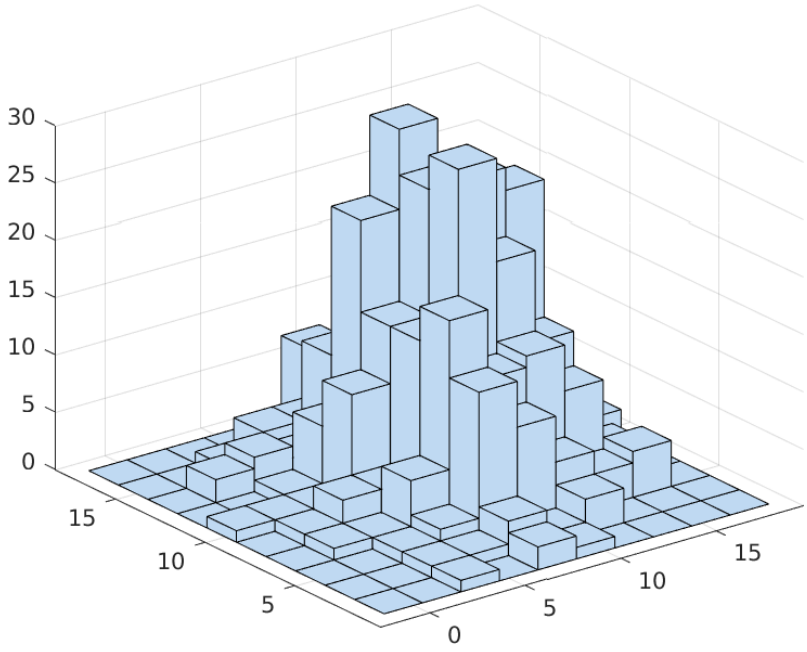
Vi kan även ha mätvärden i form av punkter (x, y) och den vanligaste figuren i dessa sammahang är ett **spridningsdiagram** (scatter plot) där man helt enkelt plottar ut punkter vid varje koordinat (x_i, y_i).

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



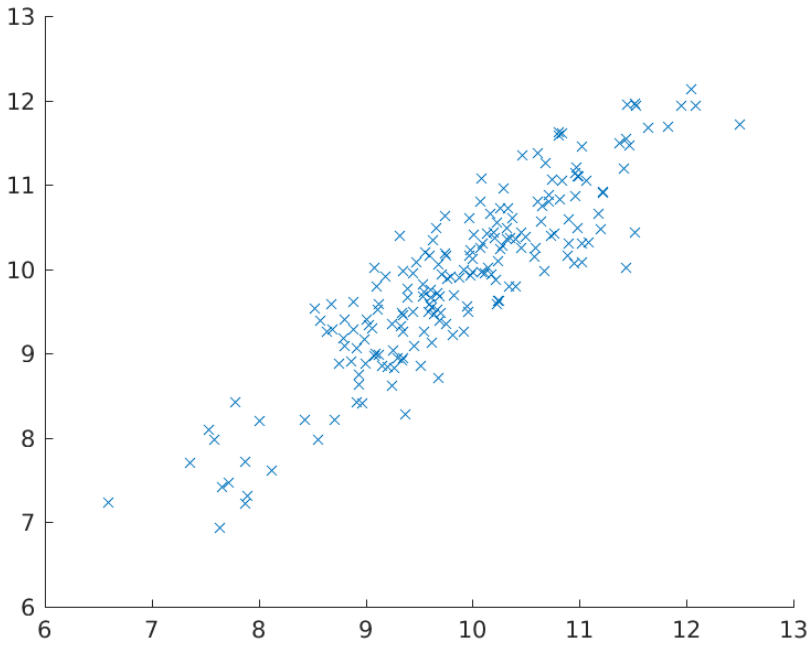
Vi ser från figuren att värdena verkar vara centrerade kring $(10, 10)$ och att det verkar föreligga någon form av cirkulär symmetri. Stämmer det för den bivarata normalfördelningen när komponenterna är oberoende? Vi kan även rendera ett två-dimensionellt histogram.

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



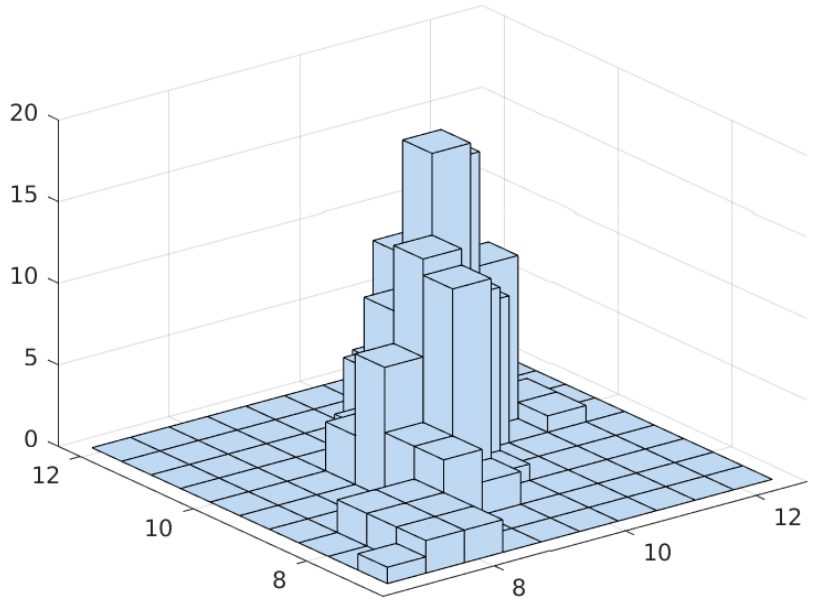
Vad gäller om variablerna i den bivariata normalfördelningen inte är oberoende?

```
>> 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');
```



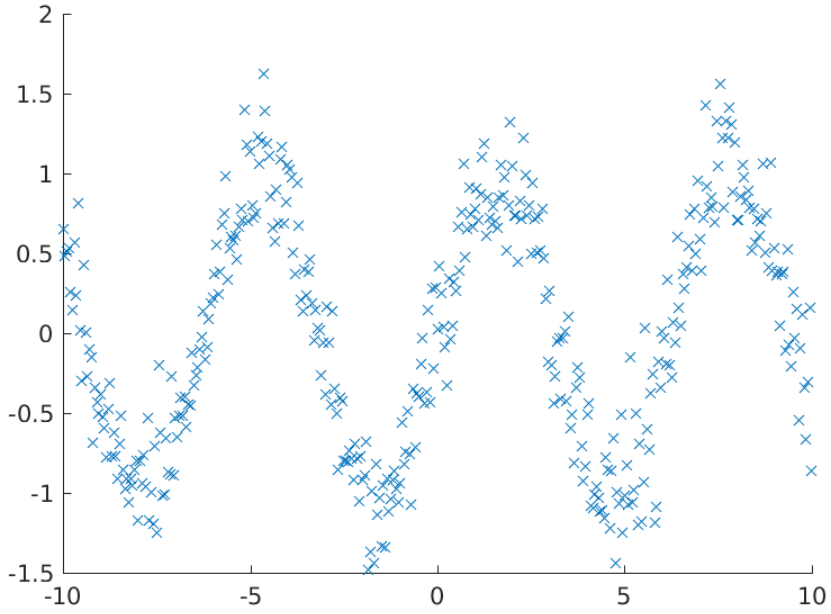
Värdena verkar fortfarande vara centrerade kring $(10, 10)$ (i någon mening) men symmetrin verkar nu utdragen diagonalt. Stämmer det för en bivarat normalfördelningen med korrelationen 0.90? Ett histogram kan genereras som ovan.

```
>> 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?

5 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** $\widehat{\theta}$ av parametern θ är en funktion (ibland kallad **stickprovsfunktion**) av de observerade värdena $x_1, x_2, \ldots, x_n$:

$$\widehat{\theta} = g(x_1, x_2, \ldots, x_n).$$

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

$$\widehat{\Theta} = g(X_1, X_2, \ldots, X_n).$$

Det är viktigt att tänka på att $\widehat{\theta}$ är en siffra, beräknad från de observerade värdena, medan $\widehat{\Theta}$ är en stokastisk variabel. Som vanligt använder vi stora bokstäver för att markera att vi syftar på en stokastisk variabel. Sambandet mellan $\widehat{\theta}$ och $\widehat{\Theta}$ är alltså att $\widehat{\theta}$ är en observation av den stokastiska variabeln $\widehat{\Theta}$. 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) $\widehat{\theta}$ – skattat värde. Känt (beräknat från stickprovet). Deterministiskt.

(iii) $\widehat{\Theta}$ – stickprovsvariabeln. Denna är stokastisk!

Sannolikheter som beräknas bör använda sig av $\widehat{\Theta}$ då $\widehat{\Theta}$ beskriver variationen hos $\widehat{\theta}$ för olika stickprov. Om bara $\widehat{\theta}$ 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 erhölls: 107dB, 110dB, 117dB, 101dB. Vi antar att mätningarna är observationer av oberoende och likafördelade variabler med okänt väntevärde μ. Hur hittar vi en skattning $\widehat{\mu}$?

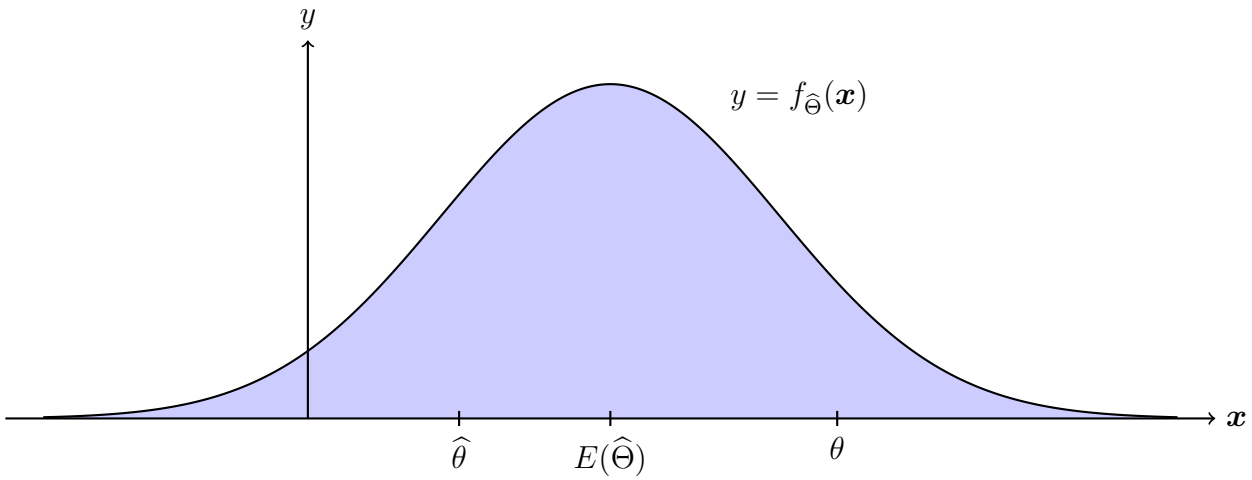
(i) $\widehat{\mu} = 100$dB är en skattning.

(ii) $\widehat{\mu} = 107$dB (den första dagen) är en skattning.

(iii) $\widehat{\mu} = (107 + 110 + 117 + 101)/4 = 108.75$dB (medelvärdet) är en skattning.

(iv) $\widehat{\mu} = \min\{107, 110, 117, 101\} = 101$dB är en skattning.

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



Vi vet att $\widehat{\theta}$ beräknas från observerade siffror, så $\widehat{\theta}$ kan hamna lite vart som helst. Dessutom vet vi inte om väntevärdet $E(\widehat{\Theta})$ 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 $\widehat{\Theta}$ kallas **väntevärdesriktig** (vvr) om $E(\widehat{\Theta}) = \theta$.

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

Systematiskt fel; bias

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

5.1 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 $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ är en skattning av stickprovsväntevärdet $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

Stickprovsvarians och stickprovsstandardavvikelse

Definition. Stickprovsvariansen $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ skattar $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
Stickprovsstandardavvikelsen skattar vi med med $s = \sqrt{s^2}$, dvs s är en skattning av S.

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

6 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 $\widehat{\theta}$ beräknas från observerade siffror, så $\widehat{\theta}$ kan hamna lite vart som helst. Dessutom vet vi inte om väntevärdet $E(\widehat{\Theta})$ 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 $E(\widehat{\Theta}) - \theta$. En väntevärdesriktig skattning $\widehat{\Theta}$ 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 $\widehat{\Theta}_n$ för varje stickprovsstorlek n. Om det för varje $\epsilon > 0$ gäller att

$$\lim_{n\to\infty} P(|\widehat{\Theta}_n - \theta| > \epsilon) = 0,$$

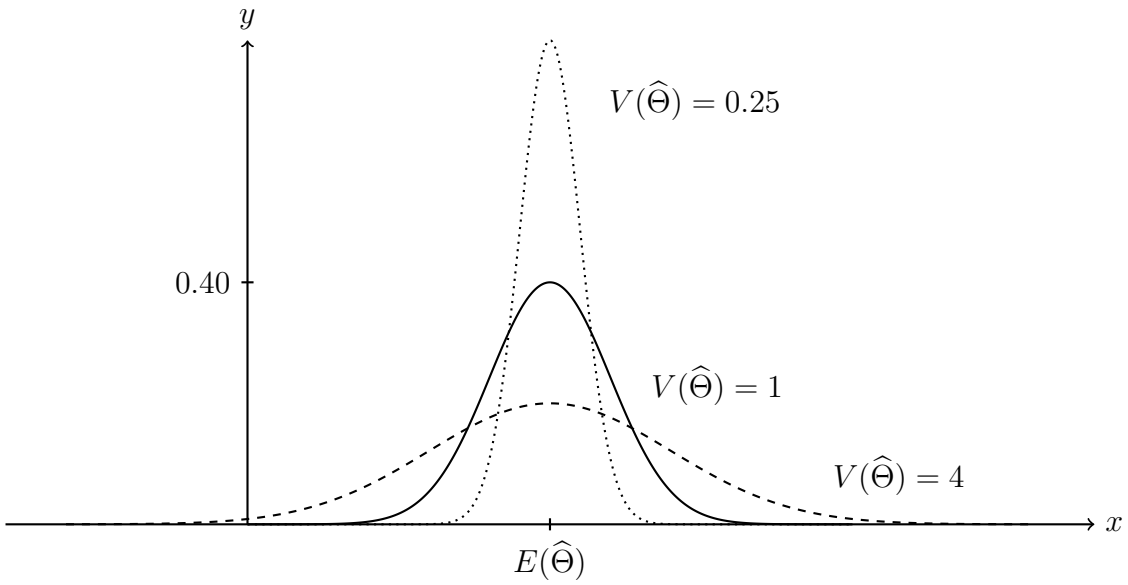
så kallar vi denna punktskattning för *konsistent*.

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 $E(\widehat{\Theta}_n) = \theta$ för alla n och $\lim_{n\to\infty} V(\widehat{\Theta}_n) = 0$ 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 $P(|\widehat{\Theta} - \theta| > \epsilon) \leq \frac{\sigma_n^2}{\epsilon^2} \to 0$ då $n \to \infty$, eftersom $\sigma_n^2 = V(\widehat{\Theta}_n) \to 0$ då $n \to \infty$. □



Mindre varians för $\widehat{\Theta}$ 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) $\widehat{\mu} = 100$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 $\widehat{M} = X_1$. Eftersom $E(\widehat{M}) = E(X_1) = \mu$ 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!

(iv) Här blir det lite klurigare när vi bildar minimum av observationerna. Vi undersöker ett specialfall där variablerna är exponentialfördelade, säg $X_i \sim \text{Exp}(\mu)$. Då gäller att (se avsnittet med kö-teori i TAMS79)

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

Således erhåller vi att $E(\widehat{M}) = \mu/4 \neq \mu$. Det finns alltså gott om fall då detta inte är en väntevärdesriktig skattning! Går det att korrigera skattningen?

6.1 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).

7 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 $\mu(\widehat{\theta}) = \overline{x}$.

Exempel

Låt $x_1, x_2, \ldots, x_n$ 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 $\mu(\widehat{\theta}) = \overline{x}$, och erhåller då att

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

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

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

Moment

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

$$m_i(\widehat{\theta_1},\widehat{\theta_2},\ldots,\widehat{\theta_j}) = \frac{1}{n}\sum_{k=1}^{n} x_k^i, \quad i = 1,2,\ldots,j.$$

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 $X_k \sim N(\mu,\sigma^2)$, $k = 1,2,\ldots,n$ 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

$$\begin{cases} \widehat{\mu} = \overline{x}, \\ \widehat{\sigma}^2 + \widehat{\mu}^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} x_k^2. \end{cases}$$

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

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \overline{x}^2 = \cdots = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2.$$

Nästan stickprovsvariansen alltså.

Vektornotation för parametrar

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