Föreläsning 2: Stokastiska variabler

Johan Thim (johan.thim@liu.se)

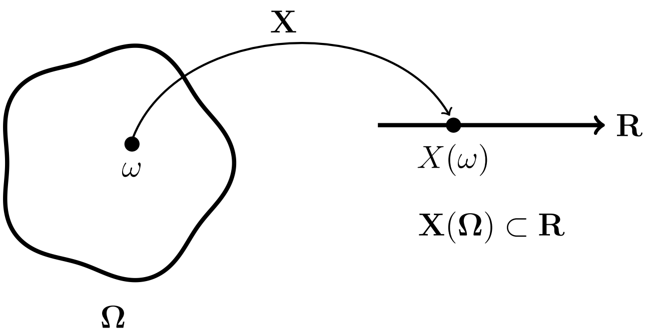
January 15, 2022

1 Endimensionella stokastiska variabler

När vi arbetar med slumpförsök så introducerar vi ofta något som brukar kallas för en stokastik 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; $X : \Omega \to \mathbf{R}$.



Vi sammanfattar några följder och gör ett par följddefinitioner.

(i) Uttrycket $X(\omega)$ (funktionsvärdet för X i punkten ω) är alltså det siffervärde 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 uppräkneligt 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 $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$. 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 möjlig 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 Diskreta 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 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) $F_X(x) \to \begin{cases} 0, & x \to -\infty, \\ 1, & x \to +\infty. \end{cases}$

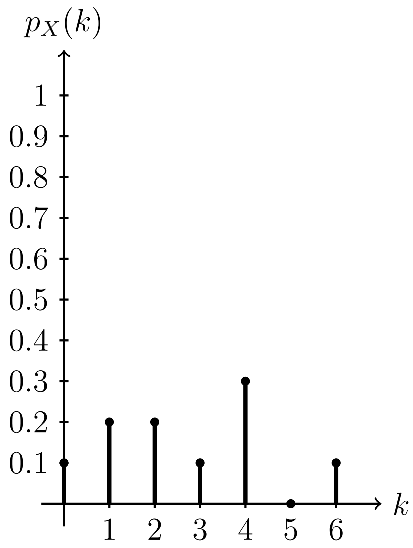
(ii) $F_X(x)$ är icke-avtagande och högerkontinuerlig.

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

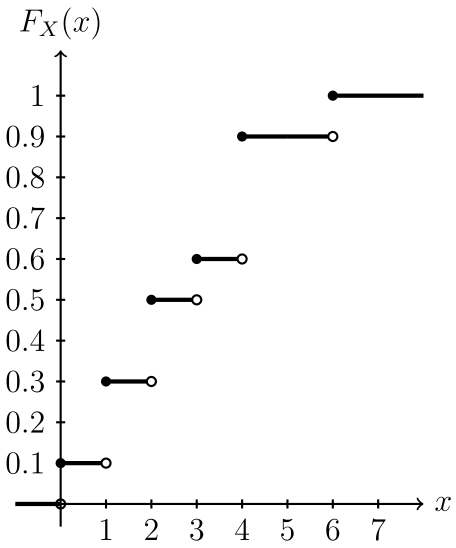
(iv) $P(X > x) = 1 - F_X(x)$.

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

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**, $X \sim \mathrm{Be}(p)$, om $p_X(a) = p$ och $p_X(b) = 1 - p$.

Exempel

Slantsingling med osymmetriskt mynt. Låt $X = -1$ vid krona och $X = 1$ vi 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 $X = 0, 1, \ldots, 10$ möjliga värden.

1. Enligt exemplet från föregående föreläsning (inbrottstjuven) måste

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

   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 förrsök sker med konstant sannolikhet p).

Binomialfördelning

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

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

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 Hypergeometriska fördelningen.

Hypergeometrisk fördelning

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

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

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 $X \in E = \{1, 2, \ldots, m\}$, och vi definierar $p_X(k) = \dfrac{1}{m}$ för varje $k = 1, 2, \ldots, m$, 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 $p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \dots$. 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 försöker oberoende tills dess att händelsen A inträffar för första gången. Antalet försök X till och med att A inträffar för första gången är 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 $p_X(k) = (1-p)^k p, \quad k = 0, 1, 2, \dots$. Vi kallar X för **Geometriskt** fördelad.

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

Poissonfördelning

Vi skriver $X \sim \mathrm{Po}(\mu)$ om $p_X(k) = \dfrac{\mu^k}{k!}e^{-\mu}$, $k = 0, 1, 2, \ldots$ 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.

4 Kontinuerliga stokastiska variabler

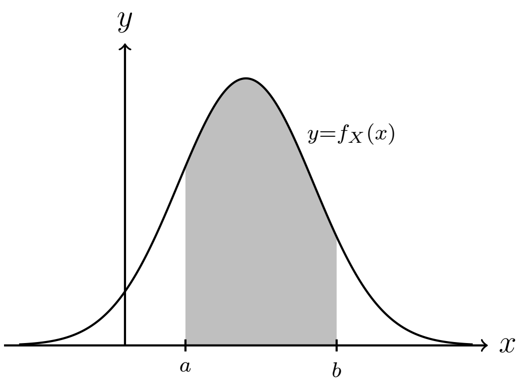
Vi kommer nu att utveckla teori för kontinuerliga stokastiska variabler som motsvarar den vi tog fram i det diskreta fallet. Åtminstone i de fall där det finns en så kallad täthetsfunktion. Så vi börjar med att definiera detta begrepp.

Täthetsfunktion

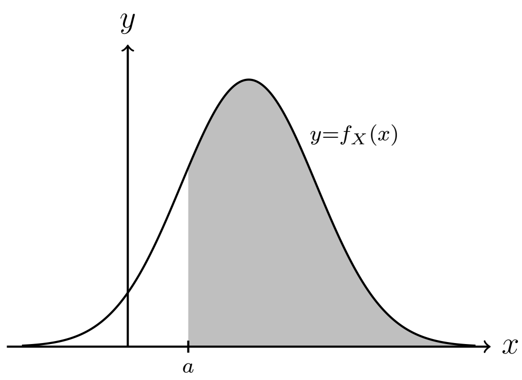
Definition. Om det finns en icke-negativ *integrerbar* funktion f_X så att

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

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

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

Strikt olikhet eller inte?

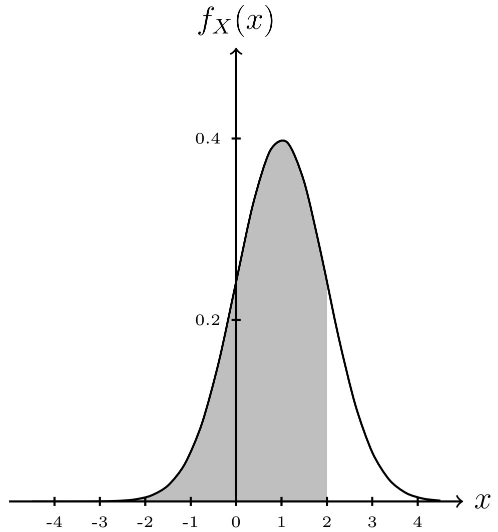
Om X är en kontinuerlig variabel, så är $P(X < x) = P(X \leq x)$. Detta följer från att integralen inte gör någon skillnad på om ändpunkten är med eller ej. Vi kan till och med definiera om funktionen i uppräkneligt många punkter (även mer, men det kräver lite måttteori för att definiera) utan att ändra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

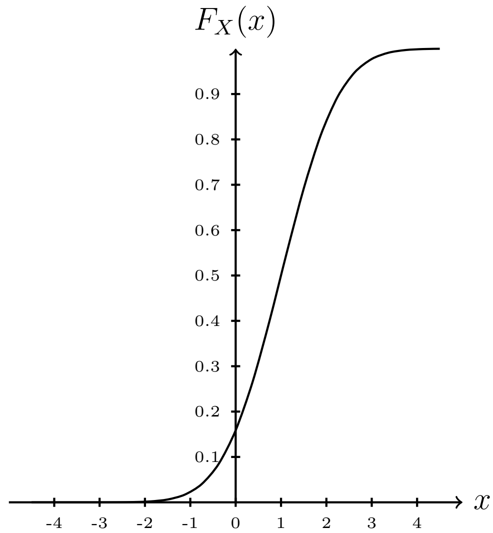
$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt, \qquad x \in \mathbf{R}.$$

Fördelningsfunktionen uppfyller (i)–(iii) från det diskreta fallet, och i alla punkter där $f_X(x)$ är kontinuerlig gäller dessutom att $F_X'(x) = f_X(x)$. 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:

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

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:

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

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

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

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 $x, y \geq 0$:

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

och

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

Eftersom X och Y är oberoende så gäller att

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

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 $X_1, X_2, \ldots, X_n$ för oberoende om

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

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

6 $(\star\star)$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 $\mathscr{B}$. 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.

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.

7 $(\star)$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 definitionsmängden (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.

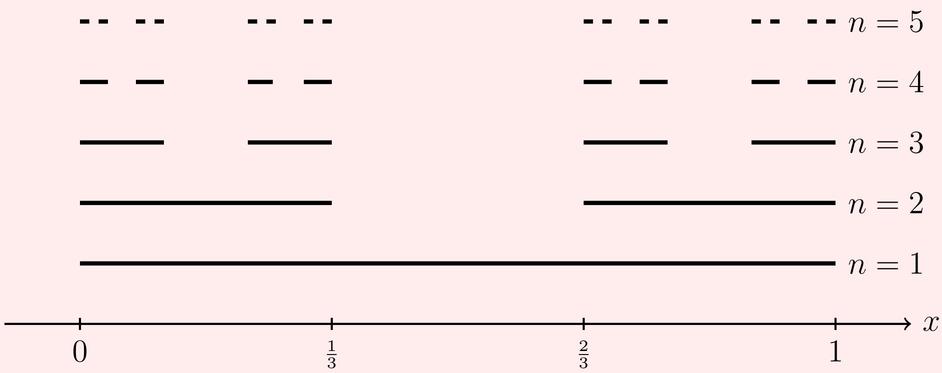
För *snälla* funktioner (funktioner som till exempel bara har uppräkneligt många diskontinuiteter) så sammanfaller de båda integralbegreppen. Vi kommer alltså inte att fundera så mycket mer på detta.

8 $(\star\star\star)$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 definitionsmässigt 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.

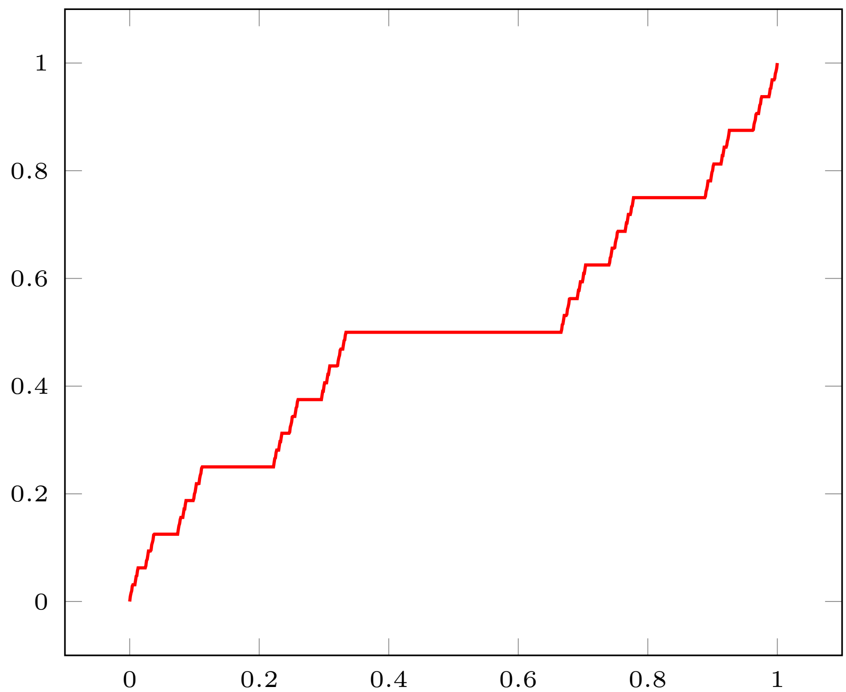


Vi noterar att $C_1 \supset C_2 \supset C_3 \supset \cdots$ och definierar $C = \cap_{k=1}^{\infty} C_k$. Detta är **Cantormängden**.

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 $F(x) \to 1$ då $x \to 1$ och $F(x) \to 0$ då $x \to 0$. 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 har 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.