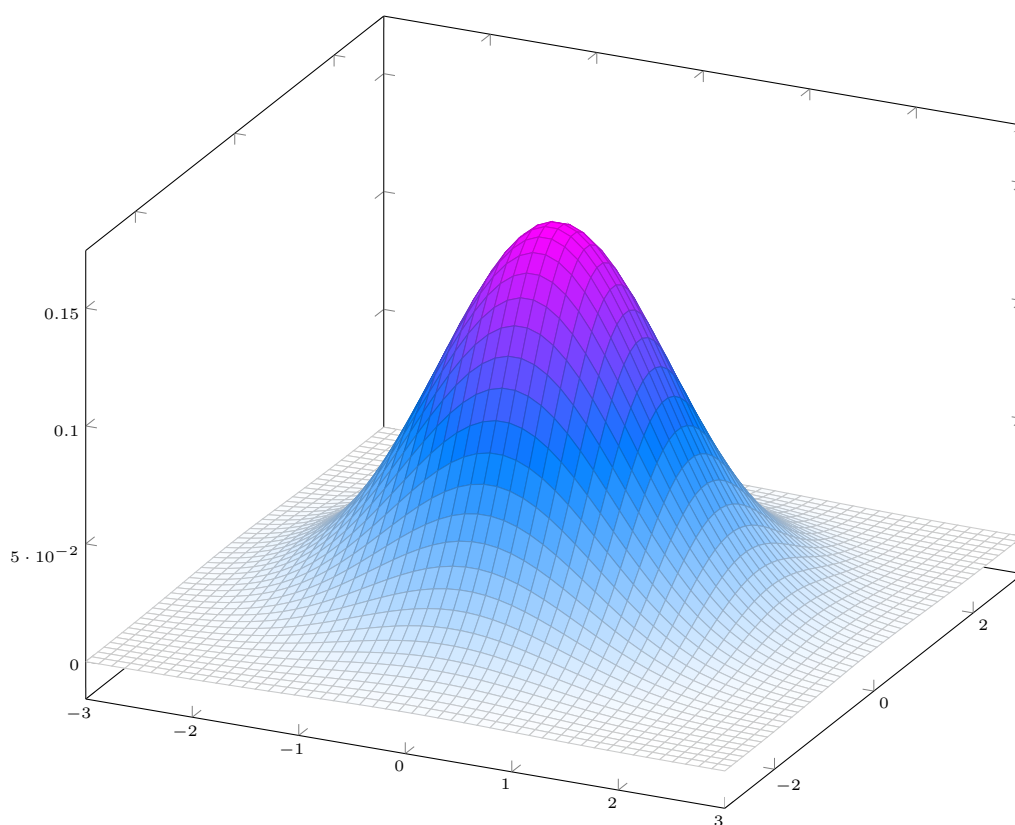


TAMS79: Matematisk statistik vt 2013

Föreläsningsanteckningar

Johan Thim, MAI



TAMS79: Föreläsning 1

Grundläggande begrepp

Johan Thim*

13 januari 2013

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.

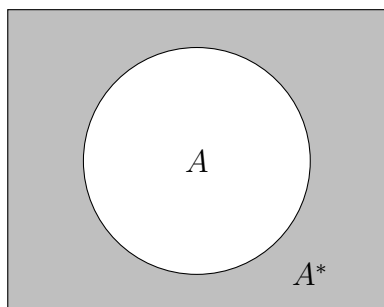


Exempel

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.

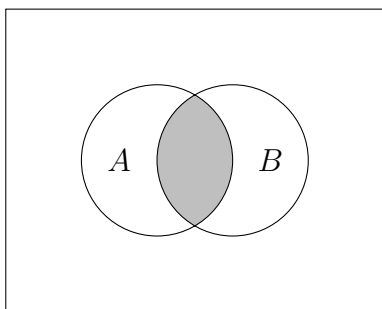
1.2 Mängdlära

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 .

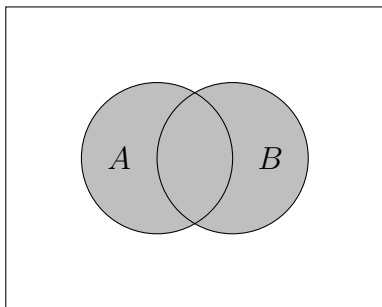
*jothi@mai.liu.se



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

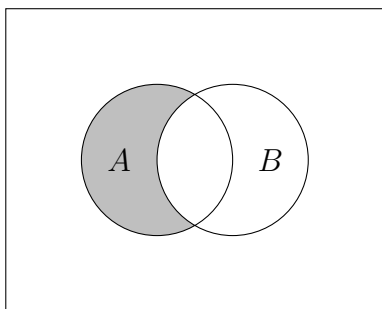
Om $A \cap B = \emptyset$ (tomma mängden) så kallas A och B för **oförenliga** (eller **disjunkta**). Två oförenliga händelser kan *ej* inträffa samtidigt.

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



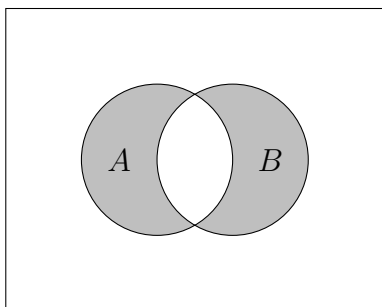
Unionen $A \cup B$ mellan händelserna $A, B \subset \Omega$. Händelsen att *någon* av A och B inträffar (eller båda).

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



Symmetrisk 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 Sannolikhet

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 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; $P: \mathcal{F} \rightarrow [0, 1]$. 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 $(\Omega, \mathcal{F}, P)$ 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!

1.3.1 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

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



Två tärningar

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

Lösning: Det finns 36 st möjliga utfall. Det första kastet ger sex möjligheter, och för vart och ett av dessa utfall får vi sex nya möjligheter vid andra kastet. Av dessa fall är följande

”gynsamma”:

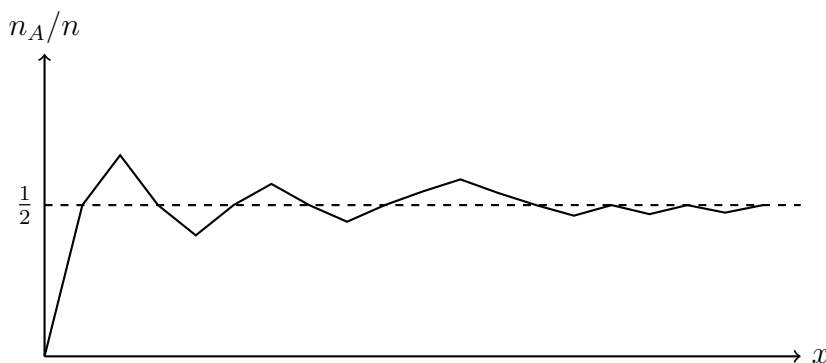
$$C = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

Vi representerar kasten som två koordinater, den första är från tärning ett och den andra från tärning två. Enligt den klassiska sannolikhetsdefinitionen får vi

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

1.3.2 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 singlar 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?).

1.4 Oberoende



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.

1.5 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.

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 att 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) att det finns

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

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.

1.6 Betingad sannolikhet



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
Returtothepit.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 $B = \text{Returtothepit.com}$. 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 $A = \text{första kolumnen}$): $P(B|A) = \frac{199}{246}$. Använder vi definitionen istället blir det

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

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

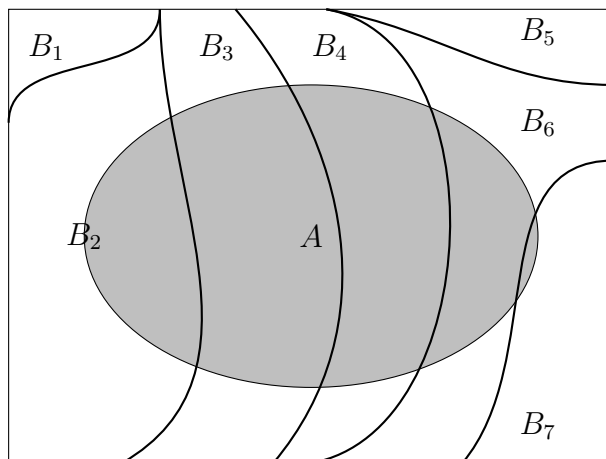
Sats. Låt $A, B_1, B_2, \dots, B_n \subset \Omega$ vara händelser sådana att:

- (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 $i \neq j$.



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.



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

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

Den andra frågan kan vi svara på mha Bayes sats:

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



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

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

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?

TAMS79: Föreläsning 2

Stokastiska variabler

Johan Thim*

13 januari 2013

2.1 Endimensionella stokastiska variabler

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 .

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 uppräknligt 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.

*jothi@mai.liu.se

Termen variabel är egentligen lite olycklig då våra stokastiska variabler är funktioner, men det är en gammal tradition som lever kvar. Ett par exempel kan vara på sin plats.



Exempel

- (i) Kasta en tärning. Utfallsrummet $\Omega = \{\square, \blacksquare, \boxtimes, \boxplus, \boxminus, \boxdot\}$. 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 .
- 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 .

2.2 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 $p_X: X(\Omega) \rightarrow [0, 1]$ 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 vid ett tärningskast till exempel).

Vissa egenskaper gäller för 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 $F_X(x) = P(X \leq x)$ för alla $x \in \mathbf{R}$.

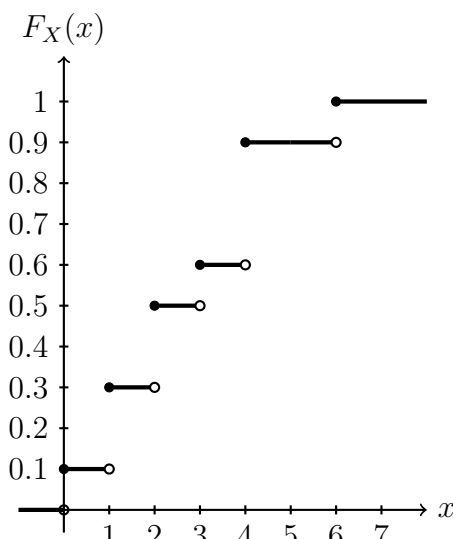
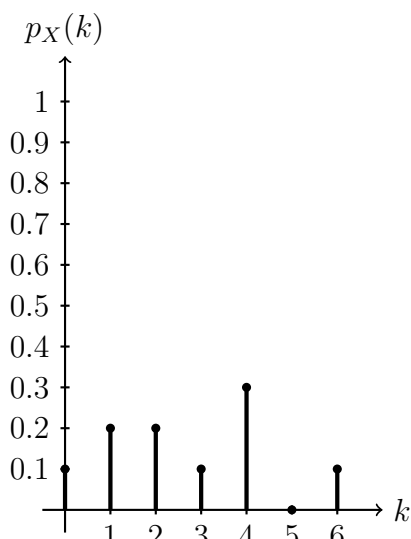
Det följer från definitionen att följande påståenden gäller.



Egenskaper hos fördelningsfunktionen

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



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 $p = 0.3$.



Binomialfördelning

Vi kallar X 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 Hypergeometriskt fördelningen.



Hypergeometrisk fördelning

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

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

Vi kallar X för **Hypergeometriskt** fördelad.



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 $X = 1, 2, \dots, m$, och $p_X(k) = \frac{1}{m}$ för varje $k = 1, 2, \dots, m$, så kallar vi X för **likformigt** fördelad.



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. Mängden $A = \{0, 3, 6, 9\}$ är 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 $p = 2/6 = 1/3$ (gynsamma/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,

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

En nära besläktad fördelning är den geometriska.



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!

TAMS79: Föreläsning 3

Kontinuerliga stokastiska variabler

Johan Thim*

13 januari 2013

3.1 Kontinuerliga stokastiska variabler



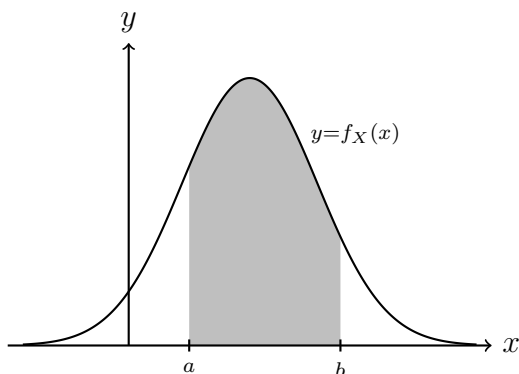
Täthetsfunktion

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

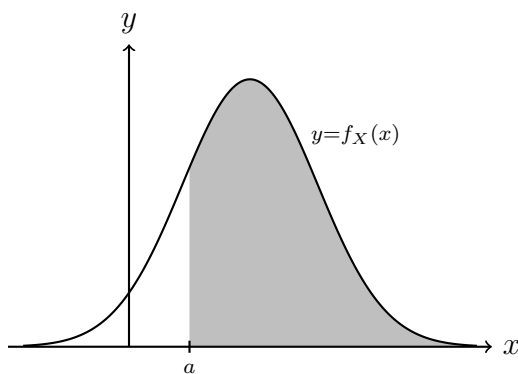
$$P(a < X < b) = \int_a^b f_X(x) dx$$

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

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) $\int_{-\infty}^{\infty} f_X(x) dx = 1$.
- (iii) $f_X(x)$ anger hur mycket sannolikhetsmassa det finns per längdenhet i punkten x .

Definitionen kanske ser oskyldig ut, men här finns det både hundar och ugglor begravda i mossen. Problemet ligger i integralbegreppet och hur generella händelser vi vill tillåta. I grundanalysen introducerar man Riemann-integralen, men tyvärr räcker den inte riktigt till

*jothi@mai.liu.se

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 slumpstal mellan noll och ett är irrationellt eller inte. Täthetsfunktionen skulle 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), där det visar sig att integralen av f mycket riktigt blir ett. Detta ligger utanför ramarna för denna kurs. 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.



Strikt olikhet eller inte?

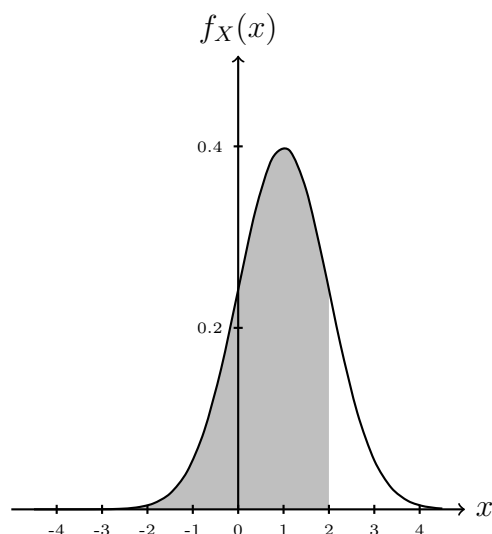
Om X är en kontinuerlig variabel, så är $P(X < x) = P(X \leq x)$. Detta följer från att integralen inte gör någon skillnad på om ändpunkten är med eller ej. Vi kan till och med definiera om funktionen i uppräknligt många punkter (även mer, men det kräver lite mått-teori för att definiera) utan att ändra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

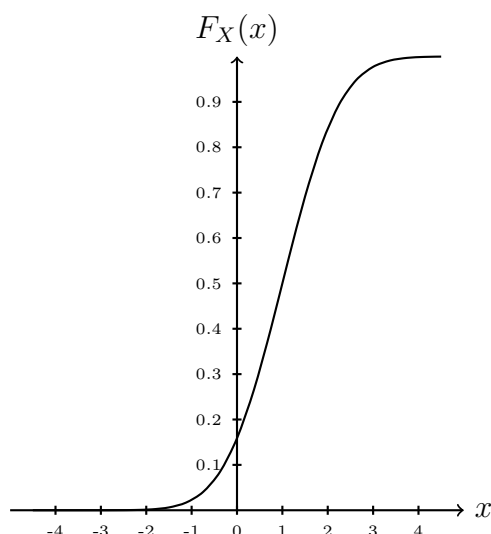
$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt, \quad x \in \mathbf{R}.$$

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

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 :

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

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 :

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

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

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

3.2 Vanliga kontinuerliga fördelningar

Analogt med diskreta variabler definieras de kontinuerliga ofta från sina respektiva 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.)



Normalfördelning

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

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

Vi kommer att studera normalfördelningen i mer detalj senare. Det är antagligen den viktigaste fördelningen ni kommer att stöta på. Fler vanliga fördelningar följer.



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 i minuter. Det visar 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:

$$(i) \quad P(T \leq 80) = \int_{-\infty}^{80} f_T(x) dx = \frac{1}{100} \int_0^{80} e^{-x/100} dx = \frac{1}{100} \left[\frac{e^{-x/100}}{-1/100} \right]_0^{80} = 1 - e^{-4/5}.$$

Sannolikheten blir alltså ca 55.1%.

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

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

Observera att denna sannolikhet är samma 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 levt tidigare. Kanske inte alltid rimligt för komponenter?

3.3 Funktioner av stokastiska variabler

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? Vi belyser med ett par exempel.



Exempel

Låt $X \sim \text{Exp}(1)$ och definiera $Z = 5X + 2$. Vad blir $f_Z(z)$?

Lösning: Fördelningsfunktionen för Z kan beräknas genom

$$F_Z(z) = P(Z \leq z) = P(5X + 2 \leq z) = P\left(X \leq \frac{z-2}{5}\right) = F_X\left(\frac{z-2}{5}\right).$$

Vidare får vi då

$$f_Z(z) = F'_Z(z) = F'_X\left(\frac{z-2}{5}\right) \frac{1}{5} = f_X\left(\frac{z-2}{5}\right) \frac{1}{5} = \begin{cases} \frac{1}{5} e^{-(z-2)/5}, & z \geq 2, \\ 0, & z < 2. \end{cases}$$



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 $f_X(x) = 0$ för $x \leq 0$. 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

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

TAMS79: Föreläsning 4

Flerdimensionella stokastiska variabler

Johan Thim*

20 januari 2013

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. 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-vektorstvärdfunktion (X, Y) definierad på ett utfallsrum Ω . (X, Y) avbildar alltså 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.

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:

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

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.

*jothi@mai.liu.se



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

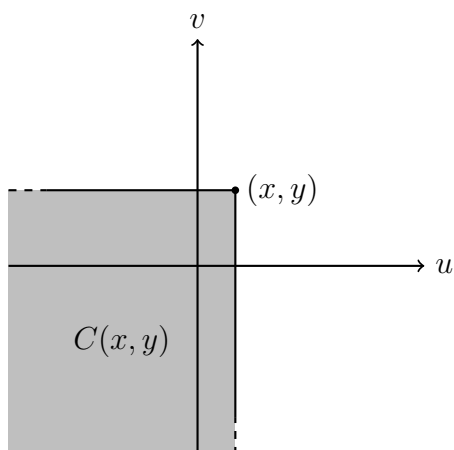
$$F_{X,Y}(x, y) = \sum_{j \leq x} \sum_{k \leq y} p_{X,Y}(j, k).$$

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

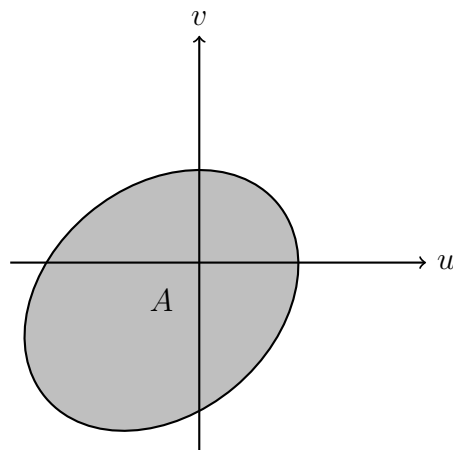
$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) \, du \, dv,$$

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. Arealen 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

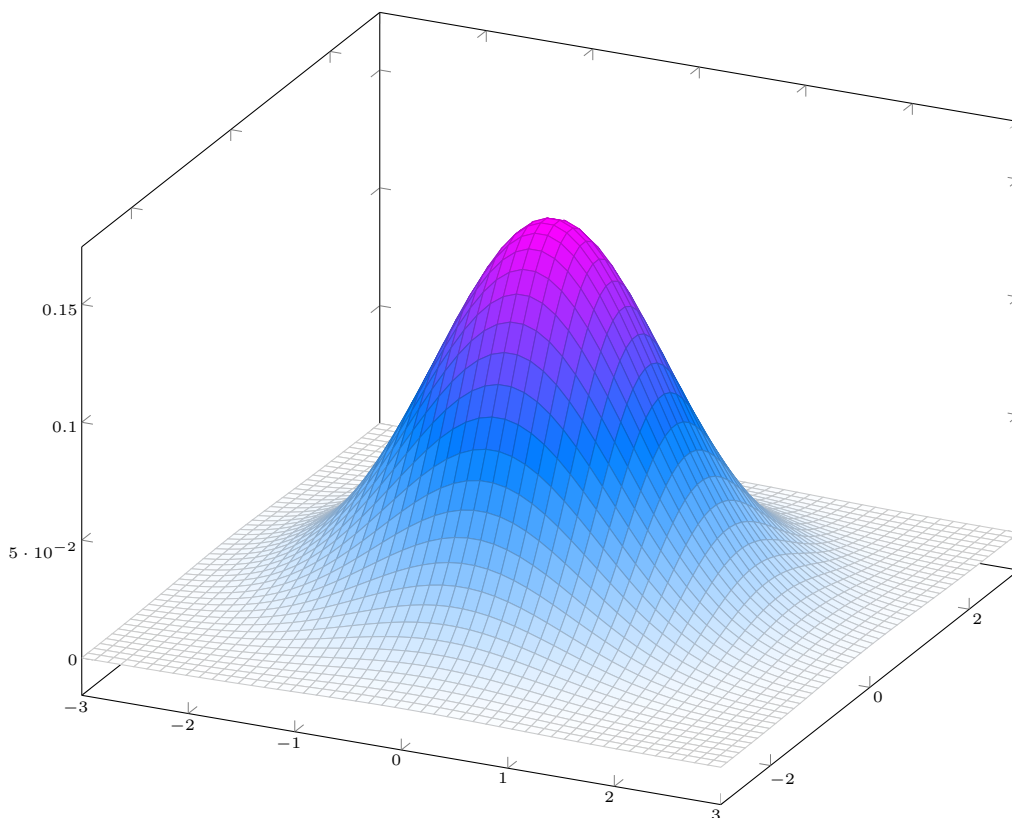
- (i) $p_{X,Y}(j, k) \geq 0$ för alla (j, k) .
- (ii) $\sum_j \sum_k p_{X,Y}(j, k) = 1$.
- (iii) Om $A \subset \mathbf{R}^2$ så är $P(X \in A) = \sum_{(j,k) \in A} p_{X,Y}(j, k)$.



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 $A \in \mathcal{B}$ (så $A \subset \mathbf{R}^2$ är *snäll*) så är $P((X, Y) \in A) = \iint_A f_{X,Y}(x, y) dx dy$.
- (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. Sannolikheten att $X > Y$ kan vi enkelt beräkna genom att titta på hur stor del av A där detta villkor är uppfyllt, vilket är triangeln med hörn i $(0, 0)$, $(2, 0)$ och $(2, 2)$. Vi erhåller alltså $P(X > Y) = \frac{\text{area triangel}}{\text{area rektangel}} = \frac{2}{6} = \frac{1}{3}$.



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

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



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

$$\begin{aligned} \sum_{j=0}^3 p_X(j) &= c(1+2+3+4) = 1, \\ \sum_{k=0}^1 p_Y(k) &= 2c(3+2) = 1, \end{aligned}$$

så p_X och p_Y är sannolikhetsfunktioner. Vi testar 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

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

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

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

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

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

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

- (iv) Enligt definitionen av betingad sannolikhet,

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

- (v) 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.

(vi) 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

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

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

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

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

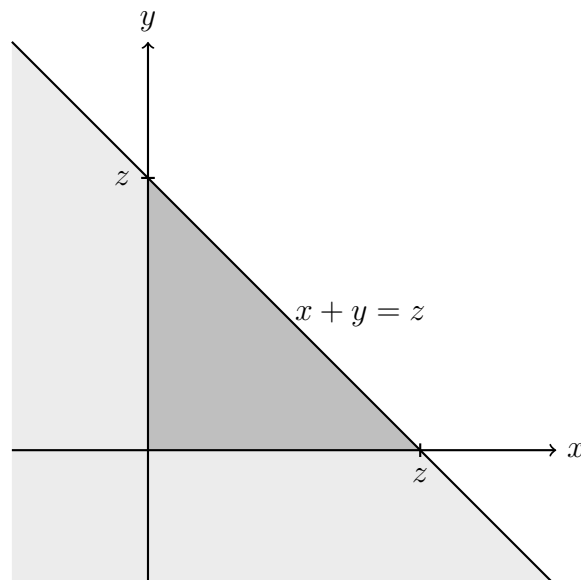
Övriga kombinationer av x och y kommer att ge $\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = 0$.



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?

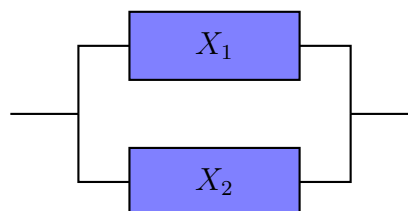
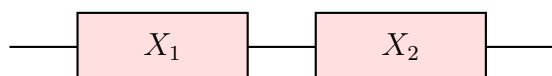
Låt $Y = \max(X_1, X_2)$. Då blir

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

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 komponenter. 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.

TAMS79: Föreläsning 5

Väntevärde och varians

Johan Thim*

22 januari 2013



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 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).



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:

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

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:

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

*jothi@mai.liu.se

5.1 Varians och standardavvikelse

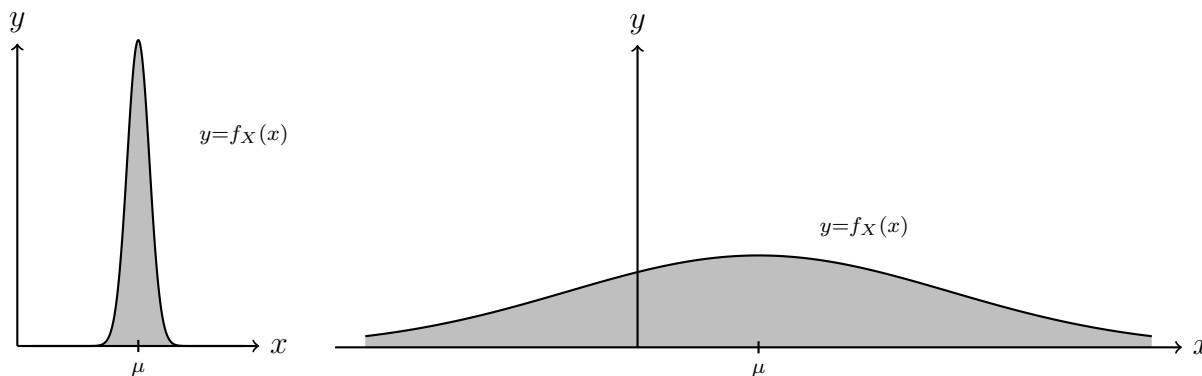


Varians och standardavvikelse

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.



Litet $D(X) = \sigma$.

Stort $D(X) = \sigma$.



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 ser snarlika ut. Sannolikheten är mycket mer utspridd för X_2 .



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 $V(X) = \frac{1-p}{p^2}$.

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 $x > 0$,

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

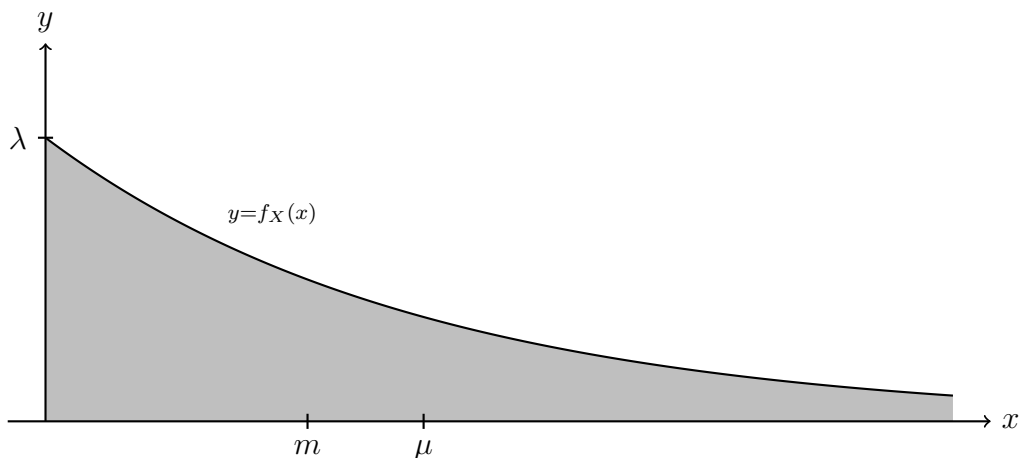
Medianen finner vi ur ekvationen $F_X(m) = 1/2$, dvs

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

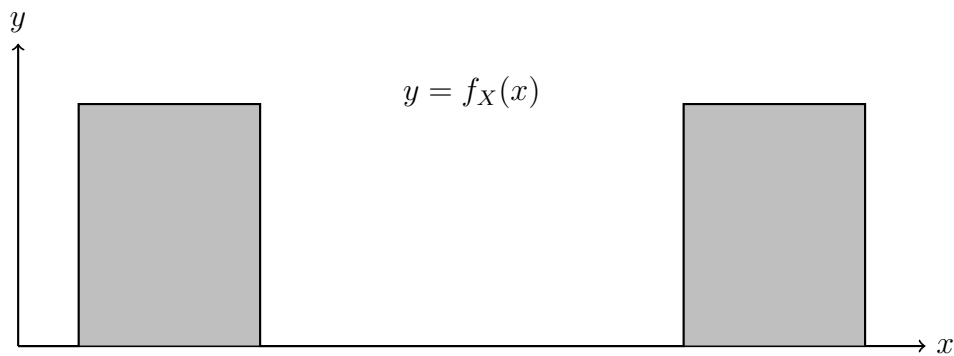
Jämför detta med väntevärdet för X :

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

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



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



Vart är medianen??

5.2 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 $E(XY) = E(X)E(Y)$.

För variansen kan vi visa följande:



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 linjärkombinationer: $V(aX \pm bY) = a^2V(X) + b^2V(Y)$. Vi kommer *aldrig* att bilda skillnader mellan varianser!

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



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 $D(2X - 3Y)$.

Då 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 $V(X + Y)$.

Vi börjar med att bestämma $p_Z(k)$. Man kan göra detta på samma sätt som i exemplet från föregående föreläsning (med summor istället för integraler), men då vi har så få möjliga värden på variablerna är det enklare att ställa upp en tabell.

$Z = X + Y = n$	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:

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

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

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

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



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 $f_Y(y)$:

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

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?

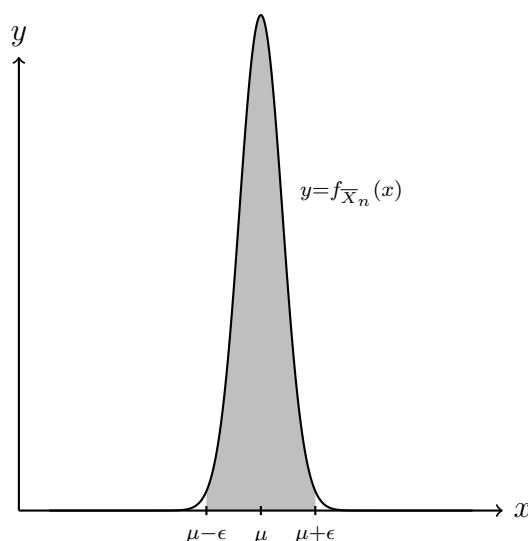
5.3 De stora talens lag



Sats. Låt $X_1, X_2, X_3, \dots$ vara en oändlig följd av oberoende och likafördelade stokastiska variabler sådana att $E(X_i) = \mu$ för alla $i = 1, 2, \dots$ och μ är ändlig. Definiera $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ (medelvärdet av de n första variablerna i följd). 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 :



Beviset följer från kända olikheter. Låt X vara en stokastisk variabel med $E(X) = \mu$ och $V(X) = \sigma^2$, och låt $k > 0$. Tjebysjovs (Chebychevs) olikhet säger att

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

För att bevisa denna olikhet kan man använda Markovs olikhet (som gäller om X är icke-negativ):

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

Kan du pussla ihop beviset för de stora talens lag från dessa?



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!).

TAMS79: Föreläsning 6

Normalfördelningen

Johan Thim*

14 februari 2013



Normalfördelning

Definition. Låt $\mu \in \mathbf{R}$ och $\sigma > 0$. Om X är en stokastisk variabel med täthetsfunktion

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

så kallar vi X för normalfördelad med parametrarna μ och σ^2 . Vi skriver $X \sim N(\mu, \sigma^2)$.

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 .



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

*jothi@mai.liu.se

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

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

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

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



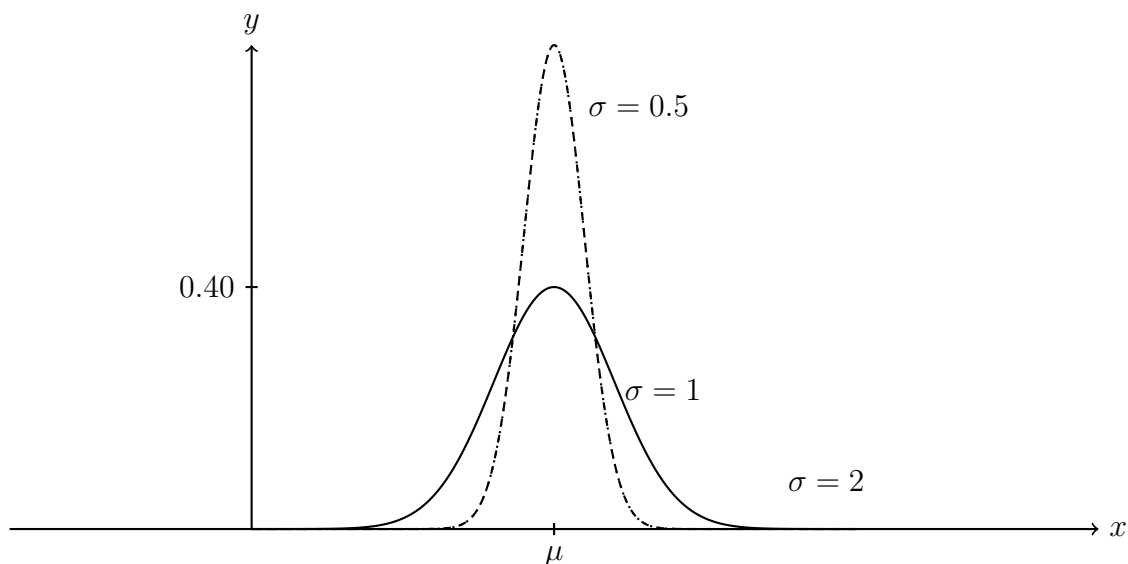
$$X \sim N(\mu, \sigma^2)$$

Om $X \sim N(\mu, \sigma^2)$ 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.

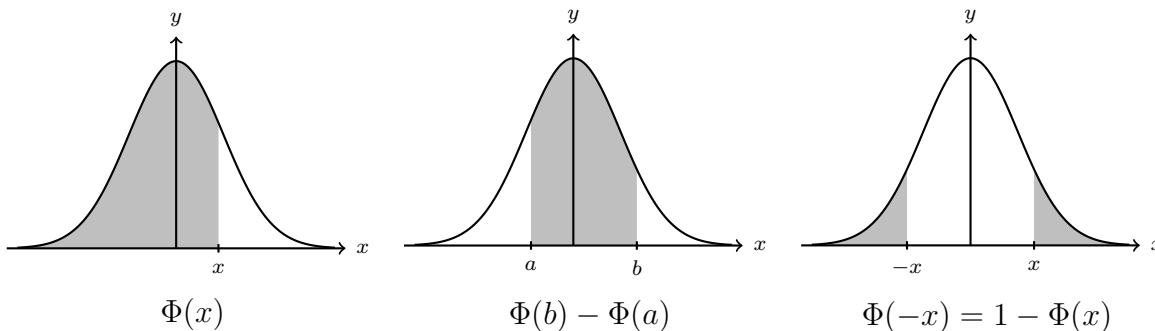




Bråk av tabell för $\Phi(x)$

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 $P(0 < X \leq 1)$.

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

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

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

Alltså har vi även

$$f_X(x) = \varphi\left(\frac{x - \mu}{\sigma}\right) \quad \text{och} \quad F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$



Exempel

Låt $X \sim N(1, 4)$, där $\sigma^2 = 4$. 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)

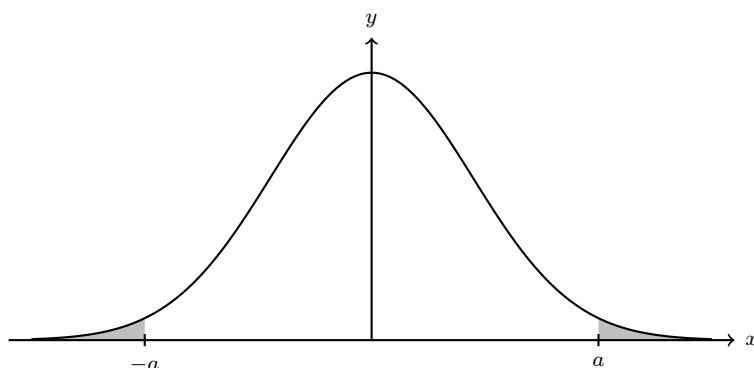
$$\begin{aligned} P(|X-2| < 3) &= P(-3 < X-2 < 3) = P\left(\frac{-1-1}{2} < \frac{X-1}{2} \leq \frac{5-1}{2}\right) \\ &= \Phi(2) - \Phi(-1) = \Phi(2) - 1 + \Phi(1) \approx 0.8186. \end{aligned}$$



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 $P(X \leq a) = 0.975$.

6.1 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 $X_1 \sim N(\mu_1, \sigma_1^2)$ och $X_2 \sim N(\mu_2, \sigma_2^2)$ vara oberoende och låt $a, b \in \mathbf{R}$. Då gäller att

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

Beviset är inte trivialt. 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 $X_k \sim N(\mu, \sigma^2)$ för $k = 1, 2, \dots, n$. Då gäller följande:

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

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 $X \sim N(10, 3^2)$ och $Y \sim N(21, 6^2)$ vara oberoende. Vad är sannolikheten att Y är mer än dubbelt så stor som X ?

Lösning: Låt $W = Y - 2X \sim N(21 - 20, 6^2 + 4 \cdot 3^2) = N(1, 72)$. 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 $T \sim N(20, 8)$.

- (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 $T_1 + T_2 \sim N(40, 16)$, 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 $\bar{T} \sim N(20, 8/n)$. Vi vill att

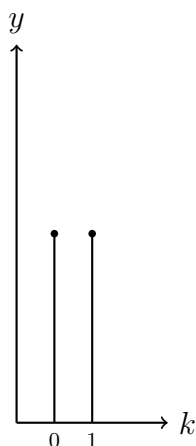
$$\begin{aligned} 0.95 &= P(\bar{T} > 19) = P\left(\frac{\bar{T} - 20}{\sqrt{8/n}} > \frac{19 - 20}{\sqrt{8/n}}\right) = 1 - P(Z \leq -1/\sqrt{8/n}) \\ &= 1 - \Phi(-1/\sqrt{8/n}) = 1 - (1 - \Phi(1/\sqrt{8/n})) = \Phi(\sqrt{n/8}), \end{aligned}$$

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.

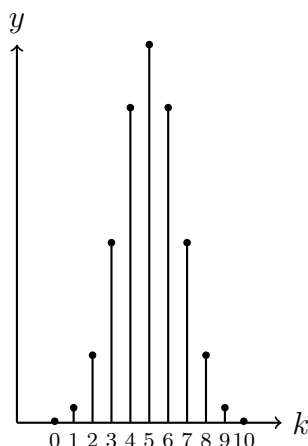
6.2 Centrala gränsvärdessatsen

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.

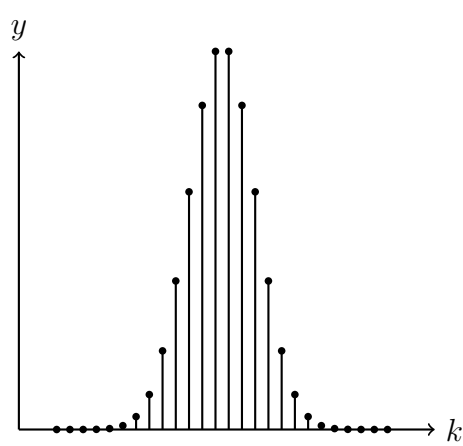
Vi betraktar ett exempel. Låt oss utföra det klassiska experimentet med slantsingling och räkna antalet X kronor vid ett visst antal, säg n , kast. Från tidigare exempel (inbrottstjuven) så vet vi att $X \sim \text{Bin}(n, p)$, där $p = 1/2$ om myntet är rättvist. En binomialfördelad variabel kan ses 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 $p = 0.5$.



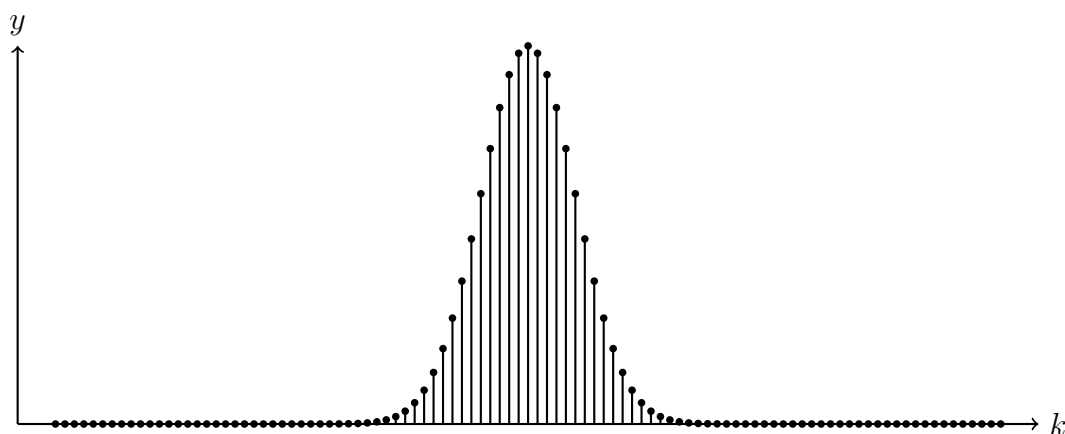
Med $n = 1$.



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

(i) summan $X = \sum_{k=1}^n X_k$ uppfyller

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

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

(ii) medelvärdet $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ uppfyller

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

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 $X \stackrel{\text{appr.}}{\sim} N(n\mu, n\sigma^2)$ och $\bar{X} \stackrel{\text{appr.}}{\sim} N(\mu, \sigma^2/n)$; 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 $X \overset{\text{appr.}}{\sim} N(50, 50/3)$. 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 åttatimmars 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 $X \overset{\text{appr.}}{\sim} N(28 \cdot 15, 28 \cdot 15^2) = N(420, 6300)$. 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`).

TAMS79: Föreläsning 7

Approximationer och stokastiska processer

Johan Thim*

23 januari 2013

7.1 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 $X \stackrel{\text{appr.}}{\sim} N(np, np(1 - p))$ om $np(1 - p) \geq 10$.

Beviset för väntevärde och varians följer av argumentet i föregående föreläsning angående **CGS**. 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 gro). Vad är sannolikheten att högst 180 stycken inte gro?

Låt X vara antalet frön som inte gro. 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 $X \stackrel{\text{appr.}}{\sim} N(200, 160)$. Vi beräknar

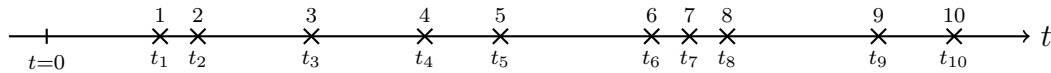
$$P(X \leq 180) \approx \Phi((180 - 200)/\sqrt{160}) = \Phi(-1.5811) = 1 - \Phi(1.5811) = 0.057.$$

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

*jothi@mai.liu.se

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

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

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 $X \stackrel{\text{appr.}}{\sim} \text{Po}(np)$.

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 bökigare. Vi kan inte direkt räkna ut $E(X^2)$, utan tar till tricket

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

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 $V(X) = E(X^2) - E(X)^2 = \mu$.



Exempel

Sågaren Sverker sågar ut brädor som har normalfördelad längd $L \sim N(200, 50)$ ($\sigma^2 = 50$), 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. Verklighet sannolikhet blir 42.15%. Normalapproximation skulle i fallet ge 31%, vilket är alldeles för lågt.



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 faltningsatsen. 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 $X \stackrel{\text{appr.}}{\sim} N(\mu, \mu)$.



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.

7.3 Exponentialfördelning

Vi fortsätter med situationen för Poissonfördelningen, men istället för att räkna antalet händelser räknar vi tiden mellan händelser. Det visar sig nämligen att tiden är exponentialfördelad. Varför? Betrakta följande. Låt T_1 vara tiden till den första händelsen. Detta innebär att $P(T_1 > t) = P(X(t) = 0)$, dvs att inga händelser inträffat på tiden t . Men $P(X(t) = 0) = e^{-\lambda t}$. Vi betraktar fördelningsfunktionen för T_1 :

$$F_{T_1}(t) = P(T_1 \leq t) = 1 - P(T_1 > t) = 1 - e^{-\lambda t}, \quad t \geq 0.$$

Vi kan nu derivera fram täthetsfunktionen: $f_{T_1}(t) = F'_{T_1}(t) = \lambda e^{-\lambda t}$ för $t \geq 0$. Detta är precis täthetsfunktionen för en $\text{Exp}(\lambda)$ -fördelad variabel!



Exponentialfördelning

Sats. En stokastisk variabel X med täthetsfunktion $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$, kallas exponentialfördelad. Väntevärde och varians ges av $E(X) = \lambda^{-1}$ och $V(X) = \lambda^{-2}$.

En intressant egenskap hos exponentialfördelningen är att den är minnesslös:

$$\begin{aligned} P(X > x + x_0 \mid X > x_0) &= \frac{P(\{X > x + x_0\} \cap \{X > x_0\})}{P(X > x_0)} = \frac{P(X > x + x_0)}{P(X > x_0)} = \frac{e^{-\lambda(x_0+x)}}{e^{-\lambda x_0}} \\ &= e^{-\lambda x} = P(X > x). \end{aligned}$$

7.4 Stokastiska processer



Stokastisk process

Definition. En **stokastisk process** är en familj $\{X_t\}_{t \in I}$ av stokastiska variabler $X_t: \Omega \rightarrow E$, där t är ett index i indexmängden I och processen tar värden i **tillståndsrummet** E .

Ibland skriver vi $\{X(t)\}_{t \in I}$ istället om det inte finns risk för missförstånd.

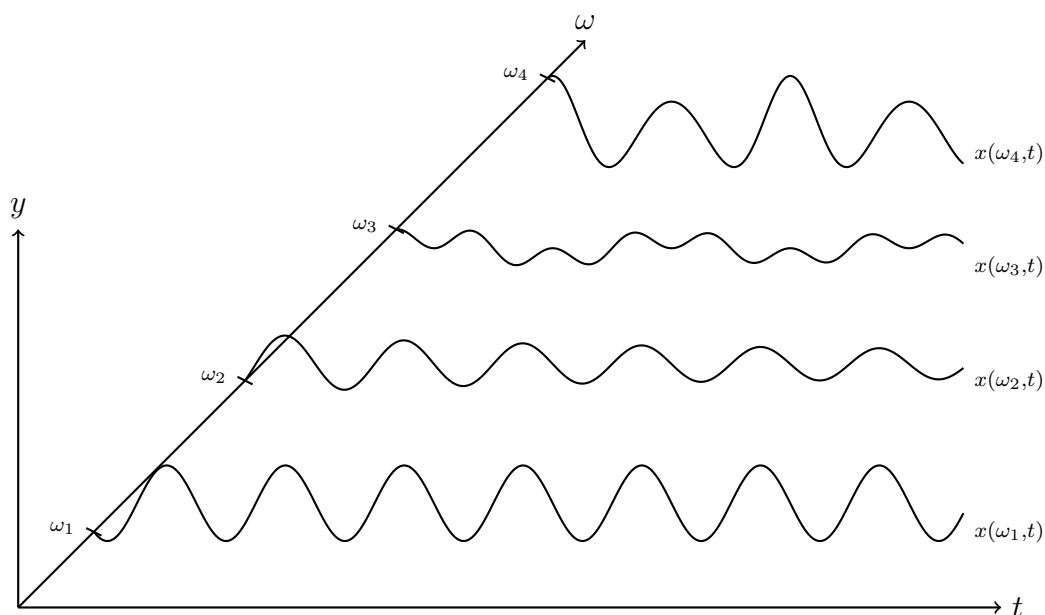
Följande kategorisering kan göras. Processen $\{X_t\}_{t \in I}$ säges

- (i) ha **diskret tid** om I är uppräknelig, typiskt $I = \{0, 1, 2, 3, \dots\}$;
- (ii) ha **kontinuerlig tid** om I är ett kontinuum, typiskt $I \subset [0, \infty)$ ett intervall;
- (iii) vara **diskret** om tillståndsrummet E är uppräkneligt: $E = \{E_0, E_1, E_2, \dots\}$;
- (iv) vara **kontinuerlig** om tillståndsrummet E är ett kontinuum, typiskt $E \subset [0, \infty)$ igen.

Processer kan alltså vara både kontinuerliga eller diskreta, och ha kontinuerlig eller diskret tid. Speciellt kan vi till exempel ha en diskret process med kontinuerlig tid (tillståndsrummet uppräkningsbart men tiden ett kontinuum).

Anmärkning: eftersom X_t formellt sett är en stokastisk variabel så borde vi kräva att $E \subset \mathbf{R}$ eftersom vi definierat termen stokastisk variabel på det sättet. Vi kommer i de närmaste avsnitten att tillåta tillståndsrummet att vara mer abstrakt, säg $E = \{\text{solen skiner, det är natt}\}$ till exempel. Detta underlättar i våra tilltänkta tillämpningar.

För varje $\omega \in \Omega$ kan vi prata om **realiseringen** av processen i form av en tidsfunktion $x(\omega, t)$ som avbildar indexmängden I in i tillståndsrummet E ; $x(\omega, \cdot): I \rightarrow E$. Med andra ord, $x(\omega, t)$ ger värdet av utfallet vid tiden t .



Några realiseringar vid fixerade utfall ω_i för en tidskontinuerlig kontinuerlig stokastisk process.



Exempel

- (i) $X_t = s_t + N_t$ där $N_t \sim (0, \sigma^2)$ för alla t är oberoende. Detta är en signal s_t vi ofta är intresserade av som är störd av brus (som är normalfördelat i detta fall).
- (ii) $X_t = A \cos 2t + Be^{-t}$ där A och B är stokastiska variabler. Vad händer när t går mot ∞ ?

Det finns mycket teori om stokastiska processer; hur man klassificerar dem (stationär? svagstationär? etc) och hur man studerar beteendet. Detta ligger utanför ramen på denna kurs, men vi skall punktstudera vissa specialfall den närmaste tiden. Först ut är den så kallade Poissonprocessen.

7.5 Poissonprocessen

Vi återkommer nu till poissonfördelning, men på ett lite annorlunda sätt. Låt $X(t)$ räkna antalet impulser i intervallet $[0, t]$. Vi ställer följande krav.



Poissonprocess

Definition. En stokastisk process $\{X(t)\}_{t \in [0, \infty)}$ är en **Poissonprocess** med intensitet λ om och endast om

- (i) Antalet impulser i disjunkta tidsintervall är oberoende stokastiska variabler;
- (ii) $P(X(t+h) - X(t) = 1) = \lambda h + o(h)$;
- (iii) $P(X(t+h) - X(t) \geq 2) = o(h)$.

Notationen $o(h)$ betyder något som går mot noll snabbare än h . Om $f(h) = o(h)$ innebär det att $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$. Kommer ni ihåg stora *ordo* från resttermer i Maclaurinutvecklingar i analysen? Vi hanterar lilla *ordo* på liknande sätt och kan göra analoga "förenklingar."

Så hur hänger den här definitionen ihop med Poissonfördelningen?

Låt $p_n(t) = P(X(t) = n)$ och anta att $h > 0$ är litet. Eftersom $[0, t]$ och $[t, t+h]$ är disjunkta tidsintervall så är variablerna $X(t)$ och $X(t+h) - X(t)$ oberoende, och $X(t+h)$ är alltså summan av dessa två oberoende variabler. Händelsen att $X(t+h) = n$ kan nu delas upp i unionen

$$\{X(t+h) = n\} = \cup_{k=0}^n (\{X(t) = n-k\} \cap \{X(t+h) - X(t) = k\}),$$

där händelserna i varje snitt är oberoende! Vidare medför våra villkor att

$$P(X(t+h) - X(t) = 0) = 1 - \lambda h + o(h).$$

Alltså blir

$$\begin{aligned} p_n(t+h) &= P(X(t) = n)P(X(t+h) - X(t) = 0) \\ &\quad + P(X(t) = n-1)P(X(t+h) - X(t) = 1) \\ &\quad + \sum_{k=2}^n P(X(t) = n-k)P(X(t+h) - X(t) = k) \\ &= p_n(t)(1 - \lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h)) + \sum_{k=2}^n p_{n-k}(t)o(h) \\ &= p_n(t)(1 - \lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h)) + o(h). \end{aligned}$$

Vi stuvlar om lite och finner att

$$\frac{p_n(t+h) - p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t) + \frac{o(h)}{h}.$$

Vi låter $t \rightarrow 0^+$ och erhåller då att högerderivatan uppfyller

$$(p_n)'_+(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n = 1, 2, \dots$$

Speciellt gäller $p'_0(t) = -\lambda p_0(t)$, $t \geq 0$, för $n = 0$. Med det naturliga begynnelsevillkoret $p_0(0) = 1$ har denna differentialekvation lösningen $p_0(t) = e^{-\lambda t}$. Om vi löser differentialekvationen för $n \geq 1$ erhåller vi

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda}, \quad n = 0, 1, 2, 3, \dots$$

Detta är sannolikhetsfunktionen för Poissonfördelningen! Vi har med andra ord precis visat att $X(t) \sim \text{Po}(\lambda t)$ för alla $t > 0$.

Från detta och tidigare härledda resultat för Poissonfördelningen har vi nu följande resultat för Poissonprocessen.



Egenskaper hos Poissonprocessen

Låt $X(t)$ vara en Poissonprocess med intensitet $\lambda > 0$.

- (i) Vid tiden $t > 0$ gäller att $X(t) \sim \text{Po}(\lambda t)$.
- (ii) Tiden T mellan två händelser är exponentialfördelad: $T \sim \text{Exp}(\lambda)$.
- (iii) Tider mellan olika händelser är oberoende stokastiska variabler.
- (iv) Sammanslagningen av två oberoende Poissonprocesser med intensiteter λ_1 respektive λ_2 är en Poissonprocess med intensitet $\lambda_1 + \lambda_2$.
- (v) Om $p_1, p_2 \in [0, 1]$ med $p_1 + p_2 = 1$, så kan $X(t)$ delas upp i två oberoende Poissonprocesser med intensiteter $p_1\lambda$ respektive $p_2\lambda$.

TAMS79: Föreläsning 8

Markovkedjor

Johan Thim*

5 februari 2013

8.1 Markovkedjor

Vi ska nu betrakta en speciell tidsdiskret diskret stokastisk process, nämligen Markovkedjan. Vi börjar med en definition.



Markovkedja

Definition. En diskret Markovprocess $\{X_n\}_{n \in \{0,1,2,\dots\}}$ med diskret tid kallas för en **Markovkedja**. Kravet på processen är att

$$P(X_n = x_n \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n \mid X_{n-1} = x_{n-1})$$

för alla $n \geq 2$ och $x_0, x_1, \dots, x_n \in E$.

Med andra ord så har historien ingen betydelse för processen. Endast det tillstånd kedjan just nu befinner sig i har någon inverkan på framtiden.

Om Markovkedjan har ett ändligt tillståndsrum E säger vi att Markovkedjan är **ändlig**. Vidare har kedjan **ingångssannolikheter** $p_j(0) = P(X_0 = E_j)$ för alla $j \geq 0$. Funktionerna $p_j(n)$ beskriver generellt sannolikhetsfördelningen för tiden n , dvs $p_j(n) = P(X_n = E_j)$ är sannolikheten att kedjan vid tiden n befinner sig i tillstånd E_j . Vi kallar dessa sannolikheter för **absoluta sannolikheter**.

Kedjan kallas **tidshomogen** om $P(X_{n+1} = E_j \mid X_n = E_i) = p_{ij}$ är konstant för alla $n \geq 0$. Detta innebär att vi *alltid* har samma sannolikhet för övergångar mellan olika tillstånd i ett steg. Vi tänker oss p_{ij} som sannolikheten att flytta oss från tillståndet E_i till E_j i ett steg. Från detta följer att vi fullständigt kan karaktärisera en tidshomogen Markovkedja utifrån **övergångsmatrisen** $P = (p_{ij})$,

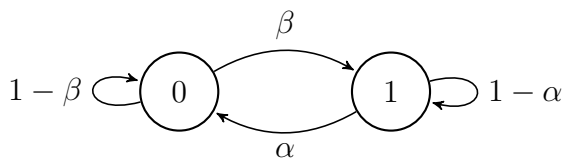
$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

där talet på rad i och kolonn j representerar övergångssannolikheten p_{ij} att kedjan går från tillstånd E_i till E_j med ett steg. Denna matris har egenskapen att varje radsumma blir ett, dvs att $\sum_j p_{ij} = 1$ för varje i . Matriser som har denna egenskap tillsammans med kravet att

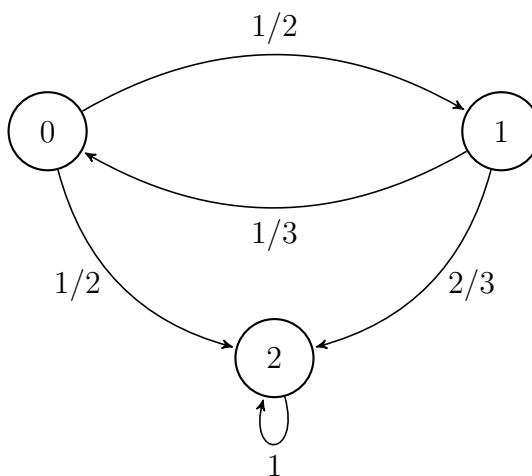
*jothi@mai.liu.se

Vi kommer att anta tidshomogenitet om inget annat anges. Ofta beskrivs Markovkedjor med diagram; vi listar några exempel tillsammans med respektive övergångsmatris.

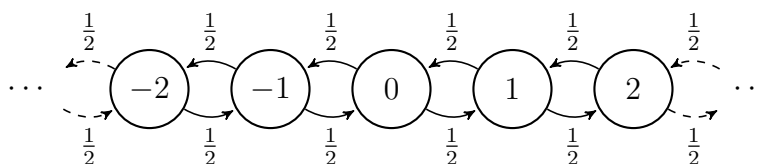
$$P = \begin{pmatrix} 1 - \beta & \beta \\ \alpha & 1 - \alpha \end{pmatrix}$$



$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 0 & 2/3 \\ 0 & 0 & 1 \end{pmatrix}$$



[illegible]



För $i, j, n \geq 0$ låter vi $p_{ij}(n) = P(X_{n+k} = E_j \mid X_k = E_i)$, vilket är sannolikheten att med precis n steg förflytta sig från E_i till E_j på något sätt. Om kedjan är tidshomogen är detta uttryck oberoende av k , så vi kan lika gärna skriva

$$p_{ij}(n) = P(X_n = E_j \mid X_0 = E_i), \quad i, j, n \geq 0.$$

2



Övergångsmatrix av ordning n

Sats. För alla heltal $n \geq 2$ gäller att $P(n) = P^n$.

Beviset för satsen kan göras till exempel med induktion på n . Vi visar basfallet $n = 2$:

$$\begin{aligned} p_{ij}(2) &= P(X_2 = E_j \mid X_0 = E_i) \\ &= \sum_k P(X_2 = E_j \mid X_1 = E_k, X_0 = E_i) \cdot P(X_1 = E_k \mid X_0 = E_i) \\ &= \sum_k P(X_2 = E_j \mid X_1 = E_k) \cdot P(X_1 = E_k \mid X_0 = E_i) \\ &= \sum_k p_{kj}(1) \cdot p_{ik}(1), \end{aligned}$$

vilket är elementet på rad i och kolonn j i matrisen P^2 . Med samma teknik kan man visa att $P(n) = P \cdot P(n-1)$, så det sökta resultatet följer från induktionsaxiomet.



Chapman-Kolmogorov

Sats. Om $m, n \geq 0$ gäller att $P(m+n) = P(m)P(n) = P^{m+n}$.

8.3 Sannolikhetsfördelningar

Om vi vet ingångssannolikheterna $p_j(0)$ och övergångsmatrisen P för en tidshomogen Markovkedja, kan vi direkt beskriva fördelningen av sannolikheter efter ett godtyckligt antal steg n genom ekvationen

$$p(n) = p(0)P^n, \quad n \geq 0,$$

där vi låter $p(n)$ vara radvektorn $(p_0(n), p_1(n), p_2(n), \dots)$ för $n = 0, 1, 2, \dots$. Observera att $p(0)$ är radvektorn med ingångssannolikheter.

Allmänt kallar vi en vektor $\underline{\pi} = (\pi_0, \pi_1, \pi_2, \dots)$ för en **sannolikhetsvektor** om $\sum_k \pi_k = 1$ och $\pi_k \geq 0$ för alla k .



Stationär fördelning

Definition. Låt $\underline{\pi}$ vara en sannolikhetsvektor. Om $\underline{\pi}$ uppfyller ekvationen $\underline{\pi} = \underline{\pi}P$ kallar vi $\underline{\pi}$ för en **stationär fördelning**.

En stationär fördelning $\underline{\pi}$ uppfyller $\underline{\pi} = \underline{\pi}P$, vilket medför att om vi låter vektorn $p(0)$ av ingångssannolikheter ges av $p(0) = \underline{\pi}$, så erhåller vi $p(n) = p(0)P^n = \underline{\pi}P^n = \underline{\pi}$ för alla n . Därav ordet stationär.



Asymptotisk fördelning

Definition. Om $\lim_{n \rightarrow \infty} p(n) = \underline{\pi}$ för någon sannolikhetsvektor $\underline{\pi}$ som inte beror på ingångssannolikheter (vektorn $p(0)$), så säger vi att $\underline{\pi}$ är kedjans **asymptotiska fördelning**.



Asymptotisk fördelning vs. stationär fördelning

Om en tidshomogen Markovkedja har en asymptotisk fördelning $\underline{\pi}$ så fås denna genom att lösa ekvationen $\underline{\pi} = \underline{\pi}P$.

En asymptotisk fördelning är alltså en stationär fördelning. Varför?

Eftersom $p(n) = p(0)P^n \rightarrow \underline{\pi}$ enligt antagande (asymptotisk fördelning existerar), kan vi i gränsvärde erhålla

$$\begin{array}{ccc} p(0)P^{n+1} & = & (p(0)P^n)P \\ \downarrow & & \downarrow \\ \underline{\pi} & = & \underline{\pi}P \end{array}$$

då $n \rightarrow \infty$. Observera att motsatsen *inte* gäller. En stationär fördelning behöver inte vara den asymptotiska fördelningen.



Exempel

Med $\alpha, \beta \in (0, 1)$, låt en Markovkedja ha övergångsmatrisen $P = \begin{pmatrix} 1-\beta & \beta \\ \alpha & 1-\alpha \end{pmatrix}$. Finn alla stationära fördelningar och bestäm (om den finns) den asymptotiska fördelningen.

Lösning: Vi börjar med att skriva ut ekvationen $\underline{\pi} = \underline{\pi}P$:

$$\begin{cases} \pi_0 = \pi_0(1 - \beta) + \pi_1\alpha, \\ \pi_1 = \pi_0\beta + \pi_1(1 - \alpha), \\ 1 = \pi_0 + \pi_1, \end{cases} \Leftrightarrow \begin{cases} \pi_0 = \frac{\alpha}{\alpha + \beta}, \\ \pi_1 = \frac{\beta}{\alpha + \beta}, \end{cases}$$

där vi utnyttjat att $\pi_0 + \pi_1 = 1$ (annars får vi oändligt många lösningar). Lösningen är entydig så det finns bara en stationär fördelning, så om en asymptotisk fördelning existerar måste den sammanfalla med vektorn ovan. Så hur visar vi att en asymptotisk fördelning existerar? Ett sätt är att helt enkelt räkna ut gränsvärdet. Det visar sig ganska bökigt att räkna ut $P^2, P^3, \dots$ (testa), så vi använder lite linjär algebra. Man kan nästan gissa att egenvärdena till P är $\lambda = 1, 1 - \alpha - \beta$. Lös annars sekularekvationen $\det(P - \lambda I) = 0$. Motsvarande egenvektorer kan beräknas till

$$v_1 = s \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{och} \quad v_2 = s \begin{pmatrix} -\beta \\ \alpha \end{pmatrix}, \quad s \in \mathbf{R} \setminus \{0\}.$$

Vi skapar en transformationsmatris T med v_1 och v_2 som kolonner. Matrisen T och dess invers ges av

$$T = \begin{pmatrix} 1 & -\beta \\ 1 & \alpha \end{pmatrix} \quad \text{och} \quad T^{-1} = \frac{1}{\alpha + \beta} \begin{pmatrix} \alpha & \beta \\ -1 & 1 \end{pmatrix}.$$

Vi diagonaliserar P genom $P = TDT^{-1}$ där D är diagonalmatrisen med egenvärdena på diagonalen. Då erhåller vi

$$\begin{aligned} P^n &= (TDT^{-1})^n = TD^nT^{-1} = T \begin{pmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{pmatrix}^n T^{-1} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} \alpha + \beta(1 - \alpha - \beta)^n & \beta - \beta(1 - \alpha - \beta)^n \\ \alpha - \alpha(1 - \alpha - \beta)^n & \beta + \alpha(1 - \alpha - \beta)^n \end{pmatrix} \\ &\rightarrow \frac{1}{\alpha + \beta} \begin{pmatrix} \alpha & \beta \\ \alpha & \beta \end{pmatrix}, \quad n \rightarrow \infty, \end{aligned}$$

eftersom $|1 - \alpha - \beta| < 1$. Alltså måste

$$p(n) = p(0)P^n \rightarrow \underline{\pi} = \begin{pmatrix} \frac{\alpha}{\alpha + \beta} & \frac{\beta}{\alpha + \beta} \end{pmatrix}, \quad n \rightarrow \infty,$$

oberoende av ingångssannolikheten $p(0)$. Detta är alltså den asymptotiska fördelningen $\underline{\pi}$.

Detta är en ganska lång beräkning för att komma fram till att den stationära fördelning i detta fall är den asymptotiska. Det finns många satser som garanterar existens av asymptotiska fördelningar. En av dessa (utan bevis) är följande.



Existens av asymptotisk fördelning

Sats. Om man i en ändlig kedja kan finna ett $n > 0$ så att alla element i någon kolonn av matrisen P^n är positiva, så existerar en asymptotisk fördelning.

I exemplet ovan kan vi alltså direkt säga att en asymptotisk fördelning existerar, och eftersom kedjan bara har en stationär fördelning måste denna även vara den asymptotiska!

8.4 Vidare klassificering av Markovkedjor

Vi kallar ett tillstånd E_j **tillgängligt** från tillstånd E_i om $p_{ij}(n) > 0$ för något $n \geq 0$; vi skriver $E_i \rightarrow E_j$. Om både $E_j \rightarrow E_i$ och $E_i \rightarrow E_j$ säger vi att tillstånden **kommunicerar** och skriver $E_i \leftrightarrow E_j$. Kommunikation är en ekvivalensrelation på mängden av alla tillstånd. Med detta menar vi att

- (i) $E_i \leftrightarrow E_i$ för alla tillstånd E_i ;
- (ii) $E_i \leftrightarrow E_j$ medför att $E_j \leftrightarrow E_i$;
- (iii) Om $E_i \leftrightarrow E_j$ och $E_j \leftrightarrow E_k$ så gäller att $E_i \leftrightarrow E_k$.

En ekvivalensrelation på en mängd delar in mängden i så kallade ekvivalensklasser. Vi säger att två tillstånd som kommunicerar med varandra tillhör samma **klass**. En Markovkedja där alla tillstånd kommunicerar med varandra (vi har endast en klass) kallas **irreducibel**. Finns det flera klasser kallar vi kedjan **reducibel**.

Låt $f_{ii}^{(n)}$ vara sannolikheten att kedjan efter n steg för första gången sedan starten i E_i åter igen befinner sig i E_i . Vi låter $f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$ vara sannolikheten att kedjan någon gång återvänder till tillstånd E_i . Om $f_{ii} = 1$ kallar vi tillståndet E_i **beständigt** och om $f_{ii} < 1$ kallas E_i **obeständigt**. Om ett tillstånd är obeständigt finns en positiv sannolikhet $1 - f_{ii}$ att kedjan *aldrig* återvänder till E_i .

Varje tillstånd i en Markovkedja är alltså beständigt eller obeständigt.

Om E_i är beständigt, låt den stokastiska variabeln T_i vara antalet steg innan kedjan återvänder till E_i . Om $E(T_i) = \sum_{n=1}^{\infty} n f_{ii}^{(n)} < \infty$ (det förväntade antalet steg är ändligt) kallar vi E_i för **positivt beständigt**. Om $E(T_i) = \infty$ kallas tillståndet **noll-beständigt**.

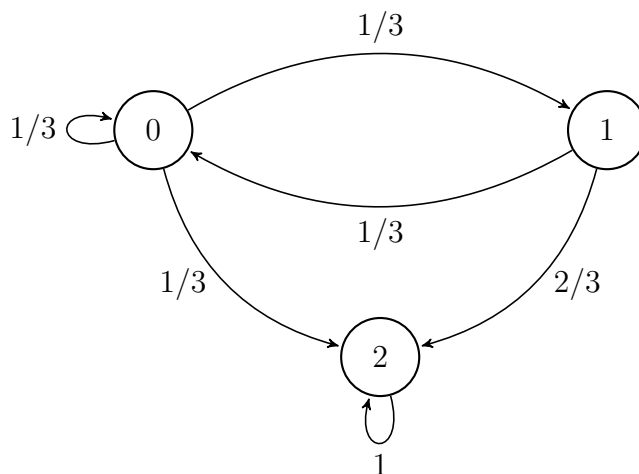
Det är även sant att dessa beständighetsegenskaper för tillstånd måste delas av hela klasser i en Markovkedja. Om ett tillstånd i klassen till exempel är positivt beständigt så är alla tillstånd i klassen det.



Exempel

Låt $P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 1 \end{pmatrix}$ vara övergångsmatrisen för kedjan X_n . Rita upp Markovkedjan och klassificera alla tillstånd. Beräkna alla stationära fördelningar och den asymptotiska fördelningen (om den existerar).

Lösning: Figur:



Vi undersöker nu vilka tillstånd som är beständiga. Först ut är E_0 :

$$f_{00} = \sum_{n=1}^{\infty} f_{00}^{(n)} = \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} = \frac{4}{9}$$

Tillståndet är obeständigt. Vi fortsätter med E_1 :

$$f_{11} = \sum_{n=1}^{\infty} f_{11}^{(n)} = 0 + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} + \cdots = \frac{1}{9} \sum_{k=0}^{\infty} \left(\frac{1}{3}\right)^k = \frac{1}{6}.$$

Även E_1 är obeständigt. Om vi tittar närmare på E_2 ser vi att tillståndet är absorberande (och därmed beständigt); $f_{22} = 1$. Detta tillstånd är även positivt beständigt (självklart).

Vi löser ekvationen $\underline{\pi} = \underline{\pi}P$ för att hitta stationära vektorer:

$$\begin{cases} \frac{\pi_0}{3} + \frac{\pi_1}{3} &= \pi_0 \\ \frac{\pi_0}{3} &= \pi_1 \\ \frac{\pi_0}{3} + \frac{2\pi_1}{3} + \pi_2 &= \pi_2 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{cases} \Leftrightarrow \begin{cases} \pi_0 &= 0 \\ \pi_1 &= 0 \\ \pi_2 &= 1 \end{cases}$$

Inte så förvånande resultat med tanke på att tillstånd två är absorberande.

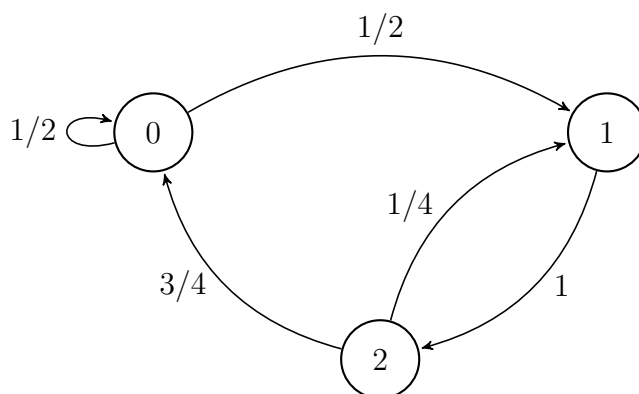
I matrisen P är alla element i kolonn tre skilda från noll, så enligt satsen ovan existerar en asymptotisk fördelning, som måste vara den unika stationära lösningen vi fann ovan.



Exempel

Låt $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}$ vara övergångsmatrisen för kedjan X_n . Visa att P är en övergångsmatris, rita upp Markovkedjan och klassificera alla tillstånd. Bestäm alla stationära fördelningar och finn den asymptotiska fördelningen om denna existerar.

Lösning: Eftersom alla rader summerar till ett och alla element ligger i $[0, 1]$ så är P en stokastisk matris, och därmed en övergångsmatris för en Markovkedja. Vilken? Vi ritar en figur:



Vi undersöker nu vilka tillstånd som är beständiga. Först ut är E_0 :

$$\begin{aligned}
 f_{00} &= \sum_{n=1}^{\infty} f_{00}^{(n)} = \frac{1}{2} + 0 + \frac{1}{2} \cdot 1 \cdot \frac{3}{4} + 0 + \frac{1}{2} \cdot 1 \cdot \frac{1}{4} \cdot 1 \cdot \frac{3}{4} + 0 + \frac{1}{2} \cdot 1 \cdot \frac{1}{4} \cdot 1 \cdot \frac{1}{4} \cdot 1 \cdot \frac{3}{4} + \dots \\
 &= \frac{1}{2} + \frac{3}{8} \sum_{k=0}^{\infty} \left(\frac{1}{4}\right)^k = \frac{1}{2} + \frac{3}{8} \cdot \frac{4}{3} = 1.
 \end{aligned}$$

Eftersom termer i en geometrisk serie avtar så snabbt, så följer det även att

$$E(T_0) = \sum_{n=1}^{\infty} n f_{00}^{(n)} < \infty.$$

Tillståndet E_0 är alltså positivt beständigt. Vi fortsätter med E_1 :

$$\begin{aligned}
 f_{11} &= \sum_{n=1}^{\infty} f_{11}^{(n)} = 0 + 1 \cdot \frac{1}{4} + 1 \cdot \frac{3}{4} \cdot \frac{1}{2} + 1 \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \dots \\
 &= \frac{1}{4} + \frac{3}{4} \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{4} + \frac{3}{4} \cdot 1 = 1.
 \end{aligned}$$

På samma sätt som ovan så avtar termerna i den geometriska serien så snabbt att

$$E(T_1) = \sum_{n=1}^{\infty} n f_{11}^{(n)} < \infty.$$

Tillståndet E_1 är alltså positivt beständigt. Det visar sig att f_{22} blir numeriskt likadan (betrakta figuren ovan) så även detta tillstånd är positivt beständigt. Alltså samma för alla tillstånd. Är denna kedja irreducibel?

För att hitta alla stationära tillstånd löser vi ekvationssystemet $\underline{\pi} = \underline{\pi}P$, dvs

$$\begin{cases} \frac{\pi_0}{2} + \frac{3\pi_2}{4} = \pi_0 \\ \frac{\pi_0}{2} + \frac{\pi_2}{4} = \pi_1 \\ \pi_1 = \pi_2 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases} \Leftrightarrow \begin{cases} \pi_0 = \frac{3}{7} \\ \pi_1 = \frac{2}{7} \\ \pi_2 = \frac{2}{7} \end{cases}$$

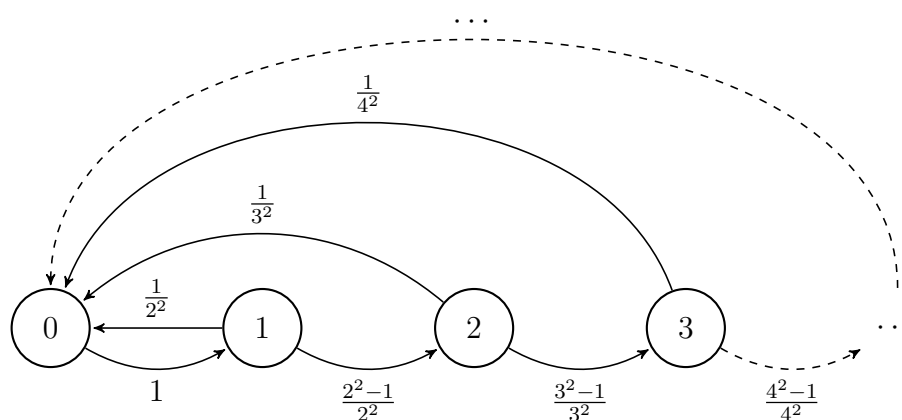
Eftersom till exempel P^2 har hela kolonn ett skild från noll så vet vi att en asymptotisk fördelning existerar, och denna måste vara den stationära fördelningen ovan (eftersom det bara finns en).



Exempel

Låt X_n vara en Markovkedja med tillstånden $E_0, E_1, \dots$ och övergångssannolikheterna $p_{i0} = \frac{1}{i^2}$ för $i \geq 1$, $p_{01} = 1$ samt $p_{i,i+1} = 1 - \frac{1}{i^2}$ för $i \geq 1$. Övriga $p_{ij} = 0$. Avgör om E_0 är beständigt.

Lösning:



Vi ser att $f_{00}^{(1)} = 0$ och att för $n \geq 2$,

$$\begin{aligned} f_{00}^{(n)} &= \frac{1}{n^2} \prod_{k=2}^{n-1} \frac{k^2 - 1}{k^2} = \frac{1}{n^2} \prod_{k=2}^{n-1} \frac{(k-1)(k+1)}{k \cdot k} = \frac{1}{n^2} \cdot \frac{1 \cdot 2 \cdots (n-2) \cdot 3 \cdot 4 \cdots n}{2 \cdot 3 \cdots n \cdot 2 \cdot 3 \cdots n} \\ &= \frac{1}{n^2} \cdot \frac{1}{2} \cdot \frac{(n-2)! \cdot n!}{n! \cdot n!} = \frac{1}{2n^2} \cdot \frac{n}{n-1} = \frac{1}{2n(n-1)}. \end{aligned}$$

För att räkna ut f_{00} delar vi upp $f_{00}^{(n)}$ i partialbråk och utnyttjar att summan blir en teleskopsumma:

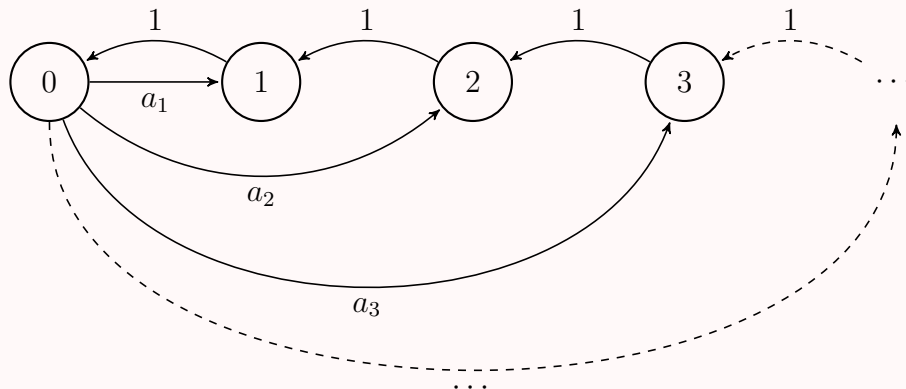
$$\begin{aligned} f_{00} &= \sum_{n=1}^{\infty} f_{00}^{(n)} = \sum_{n=2}^{\infty} \frac{1}{2n(n-1)} = \frac{1}{2} \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) = \frac{1}{2} \lim_{k \rightarrow \infty} \sum_{n=2}^k \left(\frac{1}{n-1} - \frac{1}{n} \right) \\ &= \frac{1}{2} \lim_{k \rightarrow \infty} \left(1 - \frac{1}{k} \right) = \frac{1}{2}. \end{aligned}$$

Tillståndet E_0 blir alltså obeständigt!



Exempel

Låt X_n vara den oändliga Markovkedjan med följande tillståndsdigram:



Om $a_k = \frac{6}{k^2\pi^2}$ för $k = 1, 2, 3, \dots$, klassificera tillståndet E_0 .

Vi ser att $f_{00}^{(n)} = a_{n-1}$ om $n \geq 2$ och att $f_{00}^{(1)} = 0$. Vi undersöker om E_0 är beständigt:

$$\sum_{n=1}^{\infty} f_{00}^{(n)} = \sum_{n=2}^{\infty} a_{n-1} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} = 1$$

enligt välkänd matematisk formel. Tillståndet är alltså beständigt. Vidare,

$$\sum_{n=1}^{\infty} n f_{00}^{(n)} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n} \geq \frac{6}{\pi^2} \int_1^{\infty} \frac{1}{x} dx = \infty.$$

Alltså är E_0 noll-beständigt!

Vad hade hänt om $a_k = \frac{90}{k^4\pi^4}$? Man kan visa att $\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}$, så E_0 är beständigt. Blir det nollbeständigt? Faktiskt så blir det *inte* så eftersom

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \leq \int_1^{\infty} \frac{1}{(x+1)^3} dx < \infty.$$

Om $a_k = Ck^{-\alpha}$ för lämplig konstant $C > 0$, vart går gränsen på α för att E_0 skall vara positivt beständig?

TAMS79: Föreläsning 9

Markovprocesser

Johan Thim*

5 februari 2013

Vad händer om vi i en Markovkedja har kontinuerlig tid istället för diskreta steg? Detta är ett specialfall av en kategori stokastiska processer som kallas för **Markovprocesser**. Dessa definieras av följande villkor.



Markovprocess

Definition. En stokastisk process $\{X_t\}_{t \in I}$, som tar värden i $\mathbf{R}$, med kontinuerlig tid, kallas en **Markovprocess** om

$$P(X_{t_n} \leq x \mid X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) = P(X_{t_n} \leq x \mid X_{t_{n-1}} = x_{n-1})$$

för alla $x, x_1, x_2, \dots, x_{n-1}$ och alla växande följder $t_1 < t_2 < \dots < t_n$ av tider.

Slarvigt uttryckt innebär detta att det inte spelar någon roll *hur* processen har kommit till ett visst tillstånd, utan *endast* det nuvarande tillståndet bestämmer vad som händer närmast; precis som för Markovkedjor.



Markovegenskapen och exponentialfördelningen

Sats. Om en icke-negativ stokastisk variabel T har Markovegenskapen, d v s

$$P(T > t + a \mid T > a) = P(T > t), \quad a, t \in [0, \infty[,$$

så är T exponentialfördelad.

Bevis: Markovegenskapen implicerar att

$$P(T > t) = P(T > t + a \mid T > a) = \frac{P(T > t + a)}{P(T > a)},$$

så $P(T > t + a) = P(T > a)P(T > t)$ för alla $a, t \geq 0$. Låt $g(s) = \ln P(T > s)$ för $s \geq 0$. Då kommer g att vara högerkontinuerlig (varför?) och

$$g(x + y) = g(x) + g(y), \quad x, y \geq 0.$$

Låt $m, n \in \mathbf{N}$ med $n \neq 0$. Då måste

$$g(1) = g\left(\sum_{k=1}^n \frac{1}{n}\right) = \sum_{k=1}^n g\left(\frac{1}{n}\right) = ng\left(\frac{1}{n}\right) \Rightarrow g\left(\frac{1}{n}\right) = \frac{g(1)}{n}.$$

*jothi@mai.liu.se

Vidare ser vi då att

$$g\left(\frac{m}{n}\right) = \sum_{k=1}^m g\left(\frac{1}{n}\right) = \frac{m}{n}g(1).$$

Alltså har vi visat att $g(s) = sg(1)$ för alla positiva rationella tal s . Eftersom alla reella tal kan approximeras godtyckligt nära med rationella tal (vi kan välja en sekvens r_n så att $r_n \in \mathbf{Q}$ och $r_n \rightarrow s$ med $r_n \geq s$ för alla n) och g är högerkontinuerlig, följer det att $g(s) = sg(1)$ för alla $s > 0$. Detta implicerar direkt att

$$F_T(s) = 1 - P(T > s) = 1 - e^{g(s)} = 1 - e^{-\lambda s}, \quad s \geq 0,$$

där $\lambda = -\ln P(T > 1)$. Detta är fördelningsfunktionen för en exponentialfördelad variabel T .

9.1 Markovkedjor med kontinuerlig tid

Vi betraktar den så kallade **födelse-döds-processen**. Typexemplet är att låta $X(t)$ vara antalet kunder vid ett betjäningställe vid tiden t . Det följer att tillståndsrummet E lämpligen väljs som $E = \{0, 1, 2, \dots\}$ (ändligt eller oändligt). Tanken är att processen nu endast kan ändra sitt tillstånd med ett steg i vardera riktning, alternativt inte ändra sig alls. Vi preciserar i följande definition.



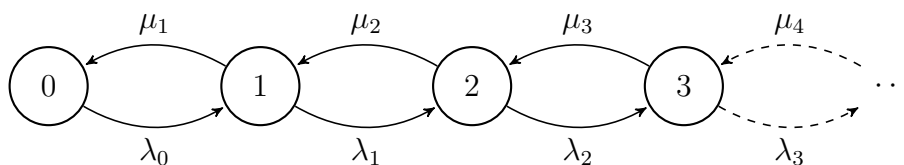
Födelse-döds-process

Definition. Den stokastiska processen $\{X(t)\}_{t \in [0, \infty[}$ är en **födelse-döds-process** om och endast om, i alla små tidsintervall $]t, t + h[$, följande gäller:

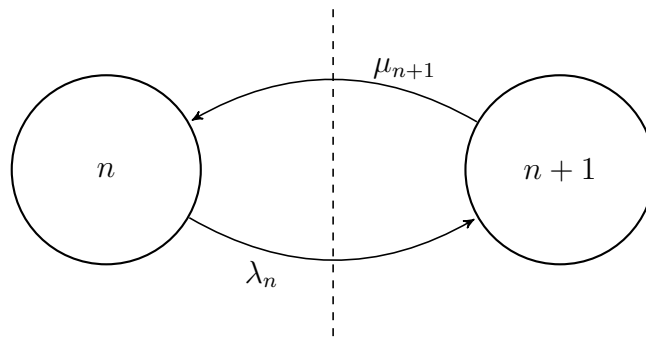
- (i) tillståndet ökar från k till $k + 1$ med sannolikhet $\lambda_k h + a(h)$,
- (ii) tillståndet minskar från k till $k - 1$ med sannolikhet $\mu_k h + b(h)$,
- (iii) tillståndet är oförändrat k med sannolikhet $1 - \lambda_k h - \mu_k h + c(h)$.

Vi kräver att $a(h) + b(h) + c(h) = 1$ och att $\mu_0 = 0$.

Vi kallar λ_k för **födelseintensiteterna** och μ_k för **dödsintensiteterna**.



Detta är en Markovkedja med kontinuerlig tid. Tyvärr kan vi inte lösa ut sannolikheter lika enkelt som för Poissonprocessen (det går att ställa upp liknande differentialekvationer, men förutom i specialfall blir dessa svårlösta). Vad vi kan göra är att med lite argumentation ta fram stationära sannolikheter. Om kedjan befinner sig i ett stationärt tillstånd måste det rimligen vara samma "flöde" in och ut ur ett tvärsnitt av grafen ovan. Vi kikar mellan tillstånd n och $n + 1$:



Antag att kedjan befinner sig i jämvikt (har uppnått ett stationärt tillstånd). Låt p_n vara sannolikheten att kedjan befinner sig i tillstånd n . Då måste $p_n \lambda_n = p_{n+1} \mu_{n+1}$ för att vi ska ha jämvikt. Således blir

$$p_{n+1} = \frac{\lambda_n}{\mu_{n+1}} p_n = \frac{\lambda_n}{\mu_{n+1}} \frac{\lambda_{n-1}}{\mu_n} p_{n-1} = \cdots = \frac{\lambda_n \lambda_{n-1} \cdots \lambda_0}{\mu_{n+1} \mu_n \cdots \mu_1} p_0.$$

Vidare vet vi att alla p_n måste summera till ett (kedjan måste finnas i något tillstånd), så

$$1 = p_0 + p_1 + \cdots = p_0 \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \cdots \right),$$

förutsatt att

$$1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \cdots < \infty.$$

Vi finner alltså att

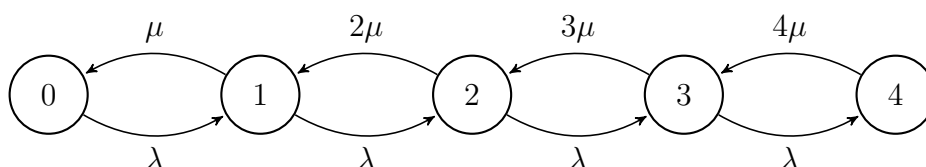
$$p_0 = \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \cdots \right)^{-1} \quad \text{och} \quad p_{n+1} = \frac{\lambda_n \lambda_{n-1} \cdots \lambda_0}{\mu_{n+1} \mu_n \cdots \mu_1} p_0.$$



Exempel

Biluthyraren Billy har fyra stycken likadana bilar och hyr ut i linjär taxa (så har du bilen i 28 timmar betalar du för $28/24 \approx 1.17$ dygn). Antag att kunder anländer som en Poissonprocess med intensitet $\lambda = 1.8$ per dygn. Varje uthyrning har exponentialfördelad utlåningstid med väntevärde 1.5 dygn och vi antar att olika uthyrningar är oberoende av varandra. Vad är sannolikheten att en kund förloras (på grund av att alla bilar är uthyrda)? Skulle man tjäna på att ha en extra bil för samma kostnad man betalar för att leasa övriga bilar?

Lösning: Vi tänker oss en födelse-döds process $X(t)$ med fem tillstånd (noll till alla fyra bilar uthyrda), där $\lambda_i = 1.8$, $\mu_i = i\mu$ för $i = 0, 1, 2, 3, 4$ och $\mu = \frac{1}{1.5}$. Vi skissar processen:



Vi ställer upp uttrycken för den stationära fördelningen:

$$p_0 = \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} + \frac{\lambda^3}{3!\mu^3} + \frac{\lambda^4}{4!\mu^4}\right)^{-1} \quad \text{och} \quad p_n = \frac{\lambda^n}{n!\mu^n} p_0.$$

Med siffror erhåller vi vektorn

$$p = (0.0779 \quad 0.2103 \quad 0.2839 \quad 0.2555 \quad 0.1725).$$

Det är alltså 17.3% risk att uthyraren förlorar en kund (detta inträffar då processen befinner sig i tillstånd 4 och alla bilar är uthyrda). Det förväntade antalet uthyrda bilar är

$$E(X(t)) = \sum_{k=0}^4 k p_k = 2.234$$

så förväntad vinst per tidsenhet blir $2.234\alpha - 4\beta$ där α är uthyrningspriset och β är kostnad för en bil.

Vi gör om samma beräkning med fem bilar istället. Det som ändras är att vi har ett extra tillstånd för fem bilar uthyrda. Vi finner nu att

$$p = (0.0712 \quad 0.1924 \quad 0.2597 \quad 0.2337 \quad 0.1578 \quad 0.0852).$$

Alltså är det nu 8.5% chans att vi förlorar en kund. Bättre, men är det värt priset? Det förväntade antalet uthyrda bilar är

$$E(X(t)) = \sum_{k=0}^5 k p_k = 2.47$$

så förväntad vinst per tidsenhet blir $2.47\alpha - 5\beta$ med α och β enligt ovan. Svaret på om det är värt priset att skaffa en bil till beror alltså på uthyrningspris och kostnad för bilarna, men vi har nu en uppskattning av på vilket sätt!

9.2 Mer om Exponentialfördelningen

Om vi summerar två oberoende variabler X och Y ges täthetsfunktionen för $Z = X + Y$ av

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx;$$

se föreläsning 4. Antag att $X, Y \sim \text{Exp}(\mu)$ är oberoende. Då blir

$$f_Z(z) = \int_0^z f_X(x) f_Y(z-x) dx = \mu^2 e^{-\mu z} \int_0^z e^{-\mu x + \mu x} dx = z\mu^2 e^{-\mu z}, \quad z \geq 0.$$

Genom induktion kan man visa att, om $W_1, W_2, \dots, W_{n+1} \sim \text{Exp}(\mu)$ är oberoende, så har $W = W_1 + W_2 + \dots + W_{n+1}$ täthetsfunktionen

$$f_W(w) = \frac{(\mu w)^n}{n!} \mu e^{-\mu w}, \quad w \geq 0.$$

Denna fördelning brukar kallas **Gammafördelningen**. Ofta ser man $W \sim \Gamma(n+1, \mu)$. Genom partialintegration (i n steg) kan vi räkna ut att

$$F_W(w) = \int_0^w f_W(t) dt = 1 - e^{-\mu w} \sum_{k=0}^n \frac{(\mu w)^k}{k!}, \quad w \geq 0.$$



Gammafördelning

Sats. Vi skriver att $X \sim \Gamma(\alpha, \mu)$, $\alpha, \mu > 0$, om

$$f_X(x) = \frac{\mu^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\mu x}, \quad x > 0.$$

Variabeln X uppfyller $E(X) = \frac{\alpha}{\mu}$ och $V(X) = \frac{\alpha}{\mu^2}$.

Här är $\Gamma(\alpha)$ gammafunktionen, och om $\alpha = n \geq 1$ är ett heltal kan den beräknas enligt $\Gamma(n) = (n-1)!$. Observera även att $\Gamma(1, \mu)$ är $\text{Exp}(\mu)$.



Summa av oberoende $\text{Exp}(\mu)$ -variabler

Sats. Om $W_i \sim \text{Exp}(\mu)$, $i = 1, 2, \dots, n$, är oberoende så är

$$W = W_1 + W_2 + \dots + W_n \sim \Gamma(n, \mu).$$

Ibland vill man även ta minimum av exponentialfördelade variabler, och då gäller följande.



Minimum av oberoende $\text{Exp}(\mu_k)$ -variabler

Sats. Antag att $X_i \sim \text{Exp}(\mu_i)$, $i = 1, 2, \dots, n$, är oberoende. Då gäller att

$$X = \min\{X_1, X_2, \dots, X_n\} \sim \text{Exp}(\mu_1 + \mu_2 + \dots + \mu_n).$$

Vi låter

$$X = \min\{X_1, X_2, \dots, X_n\}.$$

Eftersom variablerna X_i är oberoende erhåller vi

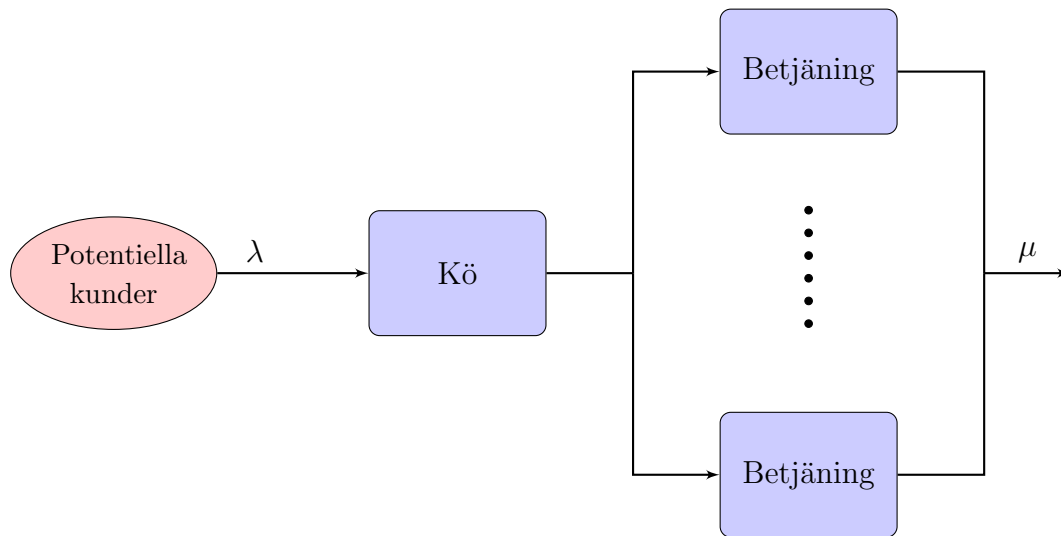
$$\begin{aligned} F_X(x) &= P(X \leq x) = 1 - P(X > x) = 1 - P(\min\{X_1, X_2, \dots, X_n\} > x) \\ &= 1 - P(X_1 > x)P(X_2 > x) \cdots P(X_n > x) = 1 - e^{-\mu_1 x} e^{-\mu_2 x} \cdots e^{-\mu_n x} \\ &= 1 - \exp\left(-x \sum_{k=1}^n \mu_k\right), \end{aligned}$$

vilket är fördelningsfunktionen för en exponentialfördelad variabel.

9.3 Köteori: terminologi och notation

Det naturligaste exemplet på ett kösystem är kanske en affär med ett visst antal kassor. Kunder kommer, ställer sig i kö, betjänas, och går därifrån. Men kömodeller existerar i väldigt många mer abstrakta tillämpningar. Till exempel använder en router för datapaket en kö för att ta mot paket samtidigt som paket vidarebefodras. Vi kommer mest att studera köer i termer av stationära fördelningar π .

Följande illustration beskriver situationen.



Kunder ankommer till ett kösystem med viss intensitet λ , väntar, betjänas och lämnar systemet med intensitet μ .

Vi kommer att beteckna olika kösystem med notationen **A/B/c/K/m/O**, där dessa bokstäver har följande betydelse:

- A**: Fördelning för tiden mellan kundankomster; typiskt **M** (Markov).
- B**: Fördelning för betjäningstiden; typiskt **M** (Markov).
- c**: Antal betjäningsställen.
- K**: Maximala antalet tillåtna kunder i systemet; typiskt ∞ .
- m**: Maximala antalet kunder i populationen; typiskt ∞ .
- O**: Betjäningsordning; typiskt FIFO (first-in-first-out).

Till exempel **M/M/1** är en kö med "Markovsk" ankomst- och betjäningstid (vilket innebär exponentialfördelade tider). Det innebär även att ankomsterna drivs av en Poissonprocess. Vidare har vi bara ett betjäningsställe. Om bokstäver är utelämnade i slutet antar vi de typiska värdena, så i vårt exempel är $K = m = \infty$ och ordningen FIFO.

För att beskriva trafiken genom systemet introducerar vi följande stokastiska variabler:

- $N_q(t)$ = antalet kunder i kön vid tiden t ,
- $N_s(t)$ = antalet kunder i betjäning vid tiden t ,
- $N(t)$ = totala antalet kunder i systemet vid tiden t ,
- W = kötiden för en kund,
- S = betjäningstiden för en kund,
- T = totala tiden i systemet för en kund.

Observera att $T = W + S$ samt $N(t) = N_q(t) + N_s(t)$. När vi pratar om kösystem använder vi vissa parametrar för att beskriva trafiken. Intensiteten λ anger **ankomstintensiteten** och $\mu = \frac{1}{E(S)}$ anger **betjäningsintensiteten**. Vi använder också **utnyttjandegraden**

$$\rho = \frac{\lambda E(S)}{c} = \frac{\lambda}{c\mu}.$$

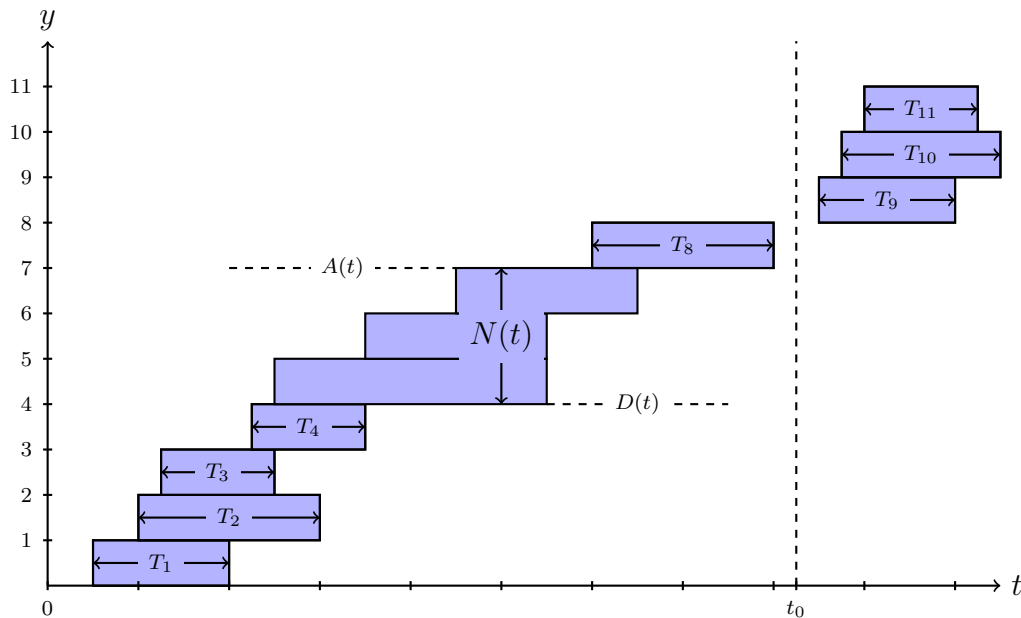


Little's formler

Sats. Med beteckningarna ovan gäller att

$$E(N(t)) = \lambda E(T) \quad \text{och} \quad E(N_q(t)) = \lambda E(W).$$

Båda dessa känns inte helt orimliga intuitivt. Det förväntade antalet kunder i systemet borde vara lika med ankomstintensiteten gånger den förväntade tiden varje kund befinner sig i systemet; Alltså $E(N(t)) = \lambda E(T)$. På liknande sätt förefaller Littles andra ekvation rimlig. Lite mer noggrant: vi inför beteckningarna T_i för tiden kund i är i systemet, och funktionerna $A(t)$ och $D(t)$ är antalet ankomna kunder respektive avgångna kunder vid tiden t . Det följer därmed att $N(t) = A(t) - D(t)$ är antalet kunder i systemet vid tiden t . Vi antar att systemet är tomt vid tiden $t = 0$ ($N(0) = 0$) och fixerar en tidpunkt t_0 då vi för enkelhetens skull antar att systemet är tomt igen: $N(t_0) = 0$.



En realisering av den stokastiska processen $\{N(t)\}$.

Den skuggade arean kan beräknas på flera sätt:

$$\int_0^{t_0} N(t) dt = \sum_{k=1}^{D(t_0)} T_k = \sum_{k=1}^{A(t_0)} T_k, \quad (1)$$

eftersom $N(t_0) = 0$ så $A(t_0) = D(t_0)$.

Det förefaller rimligt (behöver bevisas) att

$$\frac{1}{t_0} \int_0^{t_0} N(t) dt \rightarrow E(N(t)), \quad t_0 \rightarrow \infty.$$

På samma sätt borde också vara så att, om T är tiden för en kund i systemet,

$$E(T) = \lim_{t_0 \rightarrow \infty} \frac{1}{A(t_0)} \sum_{k=1}^{A(t_0)} T_k \quad \text{och} \quad \lambda = \lim_{t_0 \rightarrow \infty} \frac{A(t_0)}{t_0}.$$

Men detta innebär enligt (1) ovan att

$$E(N(t)) = \lambda E(T),$$

vilket är precis vad vi ville visa. Saken är biff! På liknande sätt kan Littles andra formel ”bevisas.”

Observera att vi inte gjort några antaganden kring fördelningar och turordning i systemet, så dessa formler gäller i väldigt generella fall. Vi kräver egentligen bara att systemet befinner sig i ett stationärt tillstånd. Kom bara ihåg att λ är den verkliga intensiteten in i systemet, så om man har ett system där kunder kan avvisas måste λ skalas om för att ta hänsyn till detta.



Exempel

Tre inköpare fyller på ett förråd med varor med intensiteterna 24, 48 respektive 36 varor per vecka. Förrådet är stort och blir aldrig fullt. Genom bokföring vet man att det finns i genomsnitt 2800 varor i förrådet. Chefen blir en dag orolig för att varorna ska bli för gamla och frågar statistikern i företaget hur lång tid i medel en vara ligger i förrådet.

Lösning: Frågan verkar vid första anblick svår att svara på, vi har ju nästan ingen information! Hur plockas varor ut (kö-ordning)? Hur ofta? Men faktum är att så länge vi antar att förrådet befinner sig i ett stationärt tillstånd, kan vi använda Littles formel. Total intensitet in i systemet är $\lambda = 24 + 48 + 36 = 108$ varor per vecka och $E(N(t)) = 2800$ varor. Vi erhåller alltså förväntad tid i lagret $E(T) = 2800/108 \approx 28$ veckor.

TAMS79: Föreläsning 10

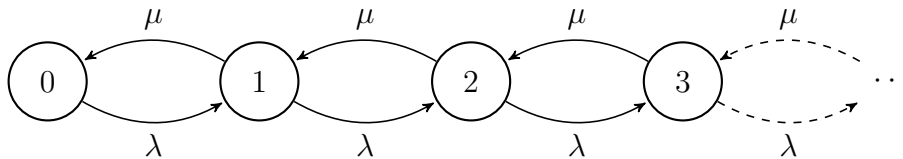
Köteori

Johan Thim*

10 mars 2013

10.1 M/M/1 kö

Ankomsttider och väntetider är exponentialfördelade och vi kan beskriva systemet som en födelse-döds-process där vi har konstant födelseintensitet λ (ny kund anländer) och konstant dödsintensitet μ (kund har fått betjäning och går sin väg). Kedjan är oändlig (finns inget max-antal kunder i kön). Figur:



Om

$$1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2} + \dots = \sum_{k=0}^{\infty} \rho^k = \frac{1}{1-\rho} = \frac{\mu}{\mu-\lambda} < \infty,$$

det vill säga om $\rho = \lambda/\mu < 1$ (dödsintensiteten större än födelseintensiteten), så uppfyller jämnviktsfördelningen $\underline{\pi}$ att

$$\pi_0 = 1 - \rho \quad \text{och} \quad \pi_k = \pi_0 \rho^k = (1 - \rho) \rho^k, \quad k \geq 1.$$

Detta känner vi igen som den geometriska fördelning. Vi skriver att $X \sim \text{Ge}(\rho)$ om

$$p_X(k) = (1 - \rho) \rho^k, \quad k = 0, 1, 2, 3, \dots$$

Vi sammanfattar lite information om den geometriska fördelningen.



Geometrisk fördelning

Sats. Om $X \sim \text{Ge}(\rho)$ med $0 < \rho < 1$ så är $E(X) = \frac{\rho}{1-\rho}$ och $V(X) = \frac{1-\rho}{\rho^2}$.

Faktum är att vi redan bevisat uttrycket för $E(X)$ och $V(X)$ i föreläsning fem.

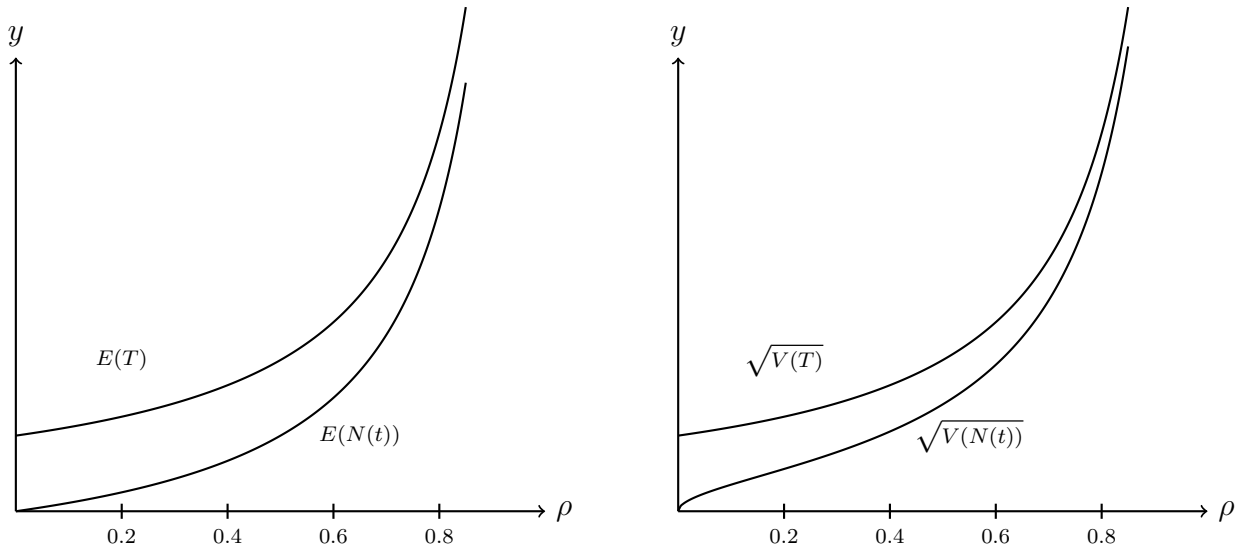
Vi har alltså $N(t) \sim \text{Ge}(\rho)$, så $E(N(t)) = \rho/(1-\rho)$. Little's sats implicerar att den förväntade totala tiden i systemet för en kund ges av

$$E(T) = \frac{E(N(t))}{\lambda} = \frac{1}{\mu - \lambda}.$$

*jothi@mai.liu.se

Även varianserna kan beräknas:

$$V(N(t)) = \frac{\rho}{(1-\rho)^2} \quad \text{och} \quad V(T) = \frac{1}{(\mu - \lambda)^2}.$$



Vi ser i figurerna ovan att systemet blir instabilt då nyttjandegraden ρ närmar sig ett. Väntevärdena växer kraftigt och även variationen i väntetid och antal personer i systemet ökar kraftigt, vilket gör förutsägelser opålitliga. Detta är ett fenomen i de flesta kösystem.

Vi kan även räkna ut förväntad kötid och förväntat antal kunder i kön:

$$E(W) = E(T) - E(S) = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)},$$

$$E(N_q(t)) = \lambda E(W) = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{1 - \rho}.$$

10.1.1 Väntetider

Antag att det är n personer i kön när nästa person anländer. Den totala tiden T i systemet för denna person är summan av $n + 1$ stycken oberoende variabler $T_i \sim \text{Exp}(\mu)$. Enligt föregående föreläsning ger detta upphov till Gamma-fördelningen, så

$$F_T(t) = \int_0^t f_T(x) dx = 1 - e^{-\mu t} \sum_{k=0}^n \frac{(\mu t)^k}{k!}, \quad t \geq 0.$$

Men det är inte alltid precis n kunder före oss när vi kommer, så hur finner vi fördelningen för T utan detta antagande? Svaret kommer i form av lagen om total sannolikhet:

$$\begin{aligned} F_T(t) &= P(T \leq t) = \sum_{n=0}^{\infty} P(N(t) = n) P(T \leq t \mid N(t) = n) \\ &= \sum_{n=0}^{\infty} (1 - \rho) \rho^n P(T \leq t \mid N(t) = n) = \sum_{n=0}^{\infty} (1 - \rho) \rho^n \left(1 - e^{-\mu t} \sum_{k=0}^n \frac{(\mu t)^k}{k!} \right) \\ &= 1 - e^{-\mu t} \sum_{n=0}^{\infty} (1 - \rho) \rho^n \sum_{k=0}^n \frac{(\mu t)^k}{k!} = 1 - e^{-\mu t} \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!} \sum_{n=k}^{\infty} (1 - \rho) \rho^n \\ &= 1 - e^{-\mu t} \sum_{k=0}^{\infty} \frac{(\rho \mu t)^k}{k!} = 1 - e^{-\mu t(1-\rho)} = 1 - e^{-t(\mu-\lambda)}. \end{aligned}$$

Vi har alltså visat att $T \sim \text{Exp}(\mu - \lambda)$, så speciellt är $E(T) = 1/(\mu - \lambda)$ (vilket vi redan visste).

På samma sätt kan man visa att $F_W(w) = P(W \leq w) = 1 - \rho e^{-w(\mu - \lambda)}$, $w \geq 0$. Notera att $F_W(0) = P(W = 0) = 1 - \rho$ är sannolikheten att kön är tom när en kund anländer. Vi kan även finna detta genom sannolikheten att betjäningen är upptagen:

$$P(N(t) > 0) = P(N(t) \geq 1) = 1 - \pi_0 = \rho,$$

så $\rho = \frac{\lambda}{\mu}$ är verkligen nytjandegraden.



Exempel

En router med stort arbetsminne (oändlig bufferkapacitet) och väldigt snabb processor har en utgående bandbredd på 100 Mbps (som routern kan fylla). Paket som kommer in till routern har en storlek som är exponentialfördelad med väntevärde 16 Kbit, och paket anländer enligt en Poissonprocess med $\lambda = 2000$ paket per sekund. Hitta en lämplig kö-modell, och svara på följande:

- (a) Förväntat antal paket i routern och förväntad svarstid.
- (b) Sannolikhet att ett paket tar mer än 0.5 ms att behandla.
- (c) Sannolikhet att routern är "idle" (inte har några paket i systemet).
- (d) Sannolikhet att routern har mer än ett paket i systemet.

Lösning: Vi väljer en M/M/1 kö med $\lambda = 2000$ paket/s och

$$\mu = \frac{100 \text{ Mbps}}{16 \text{ Kbit}} = \frac{100000}{16} = 6250 \text{ paket/s.}$$

Låt $\rho = \lambda/\mu = 0.32$.

- (a) $E(N(t)) = \frac{\rho}{1 - \rho} \approx 0.47$. Ett halvt paket i kön alltså. Svarstiden är tiden för paketet att ta sig genom systemet, vilket blir

$$E(T) = \frac{1}{\mu - \lambda} = 0.2353 \text{ ms.}$$

- (b) Sannolikheten att det tar mer än 0.5 ms för ett paket fås enligt

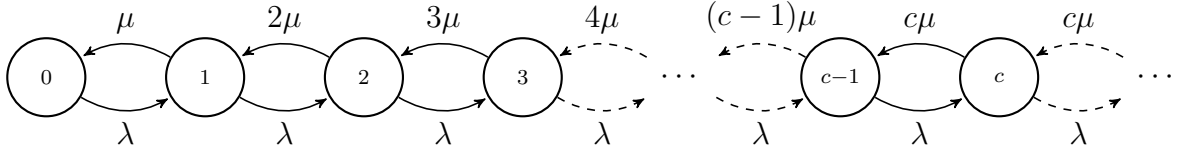
$$P(T > 0.5 \text{ ms}) = 1 - P(T \leq 0.5 \text{ ms}) = 1 - (1 - e^{-0.5(\mu - \lambda)/1000}) \approx 0.1194.$$

- (c) Sannolikheten att kön är tom ges av $P(N(t) = 0) = \pi_0 = 1 - \rho = 0.68$.

- (d) $P(N(t) > 1) = 1 - P(N(t) \leq 1) = 1 - \pi_0 - \pi_1 \approx 1 - 0.68 - 0.22 = 0.10$.

10.2 M/M/c kö

Ankomsttider och väntetider är exponentialfördelade och vi kan beskriva systemet som en födelse-döds-process där vi har konstant födelseintensitet λ (ny kund anländer) och en dödsintensitet som är μ vid en kund i systemet, 2μ om det är två, 3μ om det är tre och så vidare upp till c . För fler än c kunder är dödsintensiteten $c\mu$. Kedjan är oändlig (finns inget max-antal kunder i kön).



Låt $\rho = \frac{\lambda}{c\mu}$. För att kunna prata om en stationär fördelning är det tillräckligt att $\rho < 1$, för då är

$$\begin{aligned} \sum_{k=0}^{c-1} \frac{\lambda^k}{k! \mu^k} + \sum_{k=c}^{\infty} \frac{\lambda^k}{c! c^{k-c} \mu^k} &= 1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} + \frac{\lambda^3}{3!\mu^3} + \cdots + \frac{\lambda^c}{c!\mu^c} + \frac{\lambda^{c+1}}{c!c\mu^{c+1}} + \frac{\lambda^{c+2}}{c!c^2\mu^{c+2}} \cdots \\ &= \sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} + \frac{(c\rho)^c}{c!} \frac{1}{1-\rho} < \infty. \end{aligned}$$

Den stationära fördelningen ges nu av

$$\pi_0 = \left(\sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} + \frac{(c\rho)^c}{c!} \frac{1}{1-\rho} \right)^{-1} \quad (1)$$

och

$$\pi_k = \begin{cases} \pi_0 \frac{(c\rho)^k}{k!}, & 1 \leq k \leq c, \\ \pi_0 \frac{\rho^k c^c}{c!}, & k \geq c. \end{cases}$$

Sannolikheten att alla betjäningsställen är upptagna ges av uttrycket

$$C(c, \lambda/\mu) = \sum_{k=c}^{\infty} \pi_k = \frac{c^c \sum_{k=c}^{\infty} \rho^k}{c! \left(\sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} + \frac{(c\rho)^c}{c!} \cdot \frac{1}{1-\rho} \right)} = \frac{(c\rho)^c}{(c\rho)^c + c!(1-\rho) \sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!}}.$$

Denna formel är känd som **Erlang C-formel**. Ofta beräknas den enklast genom

$$C(c, \lambda/\mu) = \pi_0 \frac{(c\rho)^c}{c!} \frac{1}{1-\rho} = \frac{\pi_c}{1-\rho}.$$

Förväntat antal kunder ges av

$$E(N(t)) = \sum_{k=0}^{\infty} k \pi_k = \pi_0 \left(\sum_{k=0}^c k \frac{(c\rho)^k}{k!} + \sum_{k=c+1}^{\infty} \frac{k \rho^k c^c}{c!} \right).$$

Vi betraktar en del av högerledet i taget. Först,

$$\begin{aligned} \pi_0 \sum_{k=0}^c k \frac{(c\rho)^k}{k!} &= \pi_0 c\rho \sum_{k=1}^c \frac{(c\rho)^{k-1}}{(k-1)!} = \pi_0 c\rho \sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} = \pi_0 c\rho \left(\frac{1}{\pi_0} - \frac{(c\rho)^c}{c!} \frac{1}{1-\rho} \right) \\ &= c\rho - c\rho C(c, \lambda/\mu), \end{aligned}$$

om vi utnyttjar uttrycket för π_0 givet i (1) ovan. För den andra delen,

$$\begin{aligned}\pi_0 \sum_{k=c+1}^{\infty} \frac{k\rho^k c^c}{c!} &= \frac{\pi_c}{\rho^{c-1}} \sum_{k=c+1}^{\infty} k\rho^{k-1} = \frac{\pi_c}{\rho^{c-1}} \frac{d}{d\rho} \sum_{k=c+1}^{\infty} \rho^k \\ &= \frac{\pi_c}{\rho^{c-1}} \frac{d}{d\rho} \left(\rho^{c+1} \frac{1}{1-\rho} \right) = \frac{\pi_c}{1-\rho} \left((c+1)\rho + \frac{\rho^2}{1-\rho} \right) \\ &= \frac{\pi_c}{1-\rho} \left(c\rho + \frac{\rho}{1-\rho} \right).\end{aligned}$$

Vi summerar dessa två och finner att

$$E(N(t)) = c\rho - c\rho C(c, \lambda/\mu) + \frac{\pi_c}{1-\rho} \left(c\rho + \frac{\rho}{1-\rho} \right) = c\rho + \frac{\rho}{1-\rho} C(c, \lambda/\mu).$$

Little's sats kan nu användas för att ta fram följande uttryck för relevanta förväntade värden:

$$\begin{aligned}E(T) &= \frac{E(N(t))}{\lambda} = \frac{1}{c\mu - \lambda} C(c, \lambda/\mu) + \frac{1}{\mu}, \\ E(W) &= E(T) - E(S) = \frac{1}{c\mu - \lambda} C(c, \lambda/\mu), \\ E(N_q(t)) &= \lambda E(W) = \frac{\lambda/c}{1-\rho} C(c, \lambda/\mu).\end{aligned}$$

Man kan även härleda fördelningar för T och W likt förra fallet, men vi nöjer oss med de förväntade värdena.

En annan fördelning som dock kan vara intressant är avgångstiden X om alla betjäningar är upptagna. Vi har c stycken betjäningar och var och en har exponentialfördelad betjänings-tid S_i med intensitet μ . Variabeln X kan skrivas som minimum av $S_1, S_2, \dots, S_c$ och får då fördelningsfunktionen

$$F_X(x) = 1 - e^{-c\mu x}, \quad x \geq 0.$$

Detta är fördelningsfunktionen för en exponentialfördelad variabel med intensitet $c\mu$. Vi har alltså visat att $X \sim \text{Exp}(c\mu)$. Om bara k betjäningar är upptagna, $1 \leq k \leq c$, så är $X \sim \text{Exp}(k\mu)$.



Exempel

En M/M/3 kö har $\lambda = 3$ och $\mu = 2$. Vad är den förväntade kö-tiden? Vad är sannolikheten att en ny kund får stå i kö? Om alla betjäningar är fulla, vad är den förväntade tiden till en betjäning blir ledig?

Lösning: Vi har $c = 3$ och en Markov/Markov modell, så tider blir exponentialfördelade. Sannolikheten att vi får stå i kö ges av, med $c = 3$ och $\rho = 1/2$,

$$P(N(t) \geq c) = \sum_{k=c}^{\infty} \pi_k = C(c, \lambda/\mu) = 0.2368$$

eftersom

$$\pi_0^{-1} = \sum_{k=0}^2 \frac{(3\rho)^k}{k!} + \frac{(3\rho)^3}{3!} \cdot \frac{1}{1-\rho} = 4.75$$

och

$$C(c, \lambda/\mu) = \frac{\pi_c}{1-\rho} = \pi_0 \frac{(c\rho)^c}{c!} \frac{1}{1-\rho}.$$

Förväntad kötid: $E(W) = 0.0789$. Om alla betjäningar är fulla vet vi att fördelningen för tiden till dess att någon lämnar systemet är $\text{Exp}(c\mu)$, så den förväntade tiden blir $1/c\mu = 1/6$.

10.2.1 M/M/2

Om $c = 2$ har vi med $\rho = \lambda/2\mu$ specialfallet

$$1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} + \frac{\lambda^3}{2^2\mu^3} \cdots = 1 + 2 \sum_{k=1}^{\infty} \left(\frac{\lambda}{2\mu} \right)^k = 1 + 2 \frac{\rho}{1-\rho} = \frac{1+\rho}{1-\rho},$$

vilket är ändligt om $\lambda < 2\mu$. Jämnviktsfördelningen π uppfyller att

$$\pi_0 = \frac{1-\rho}{1+\rho} \quad \text{och} \quad \pi_k = 2\rho^k \cdot \frac{1+\rho}{1-\rho}, \quad k = 1, 2, 3, \dots$$

Formlerna för förväntade värden blir nu lite enklare:

$$\begin{aligned} E(N(t)) &= \frac{2\rho}{1-\rho^2}, & E(T) &= \frac{1/\mu}{1-\rho^2}, \\ E(W) &= \frac{\rho^2}{\mu(1-\rho^2)}, & E(N_q(t)) &= \frac{2\rho^3}{1-\rho^2}. \end{aligned}$$



Exempel

Den nuvarande databasservern har en betjäningsintensitet μ och en ankomstintensitet λ . Prestandan börjar bli för dålig och man ska uppgradera. Jämför följande val:

1. Köper en ny server med betjäningsintensitet 4μ .
2. Köper två stycken nya servrar med betjäningsintensitet 2μ och kör dessa med en gemensam kö.

Vilket alternativ ger bäst prestanda?

Lösning: Alternativ ett är en M/M/1 situation och alternativ två en M/M/2 situation. Vi jämför. Låt $\rho_1 = \frac{\lambda}{4\mu}$ och $\rho_2 = \frac{\lambda}{2 \cdot 2\mu}$. Numeriskt är alltså $\rho_1 = \rho_2$.

	$E(N(t))$	$E(T)$	$E(W)$
Alt. 1	$\frac{\rho_1}{1-\rho_1}$	$\frac{1}{4\mu(1-\rho_1)}$	$\frac{\rho_1}{4\mu(1-\rho_1)}$
Alt. 2	$\frac{2\rho_2}{1-\rho_2^2}$	$\frac{1}{2\mu(1-\rho_2^2)}$	$\frac{\rho_2^2}{2\mu(1-\rho_2^2)}$

Om vi jämför kvoten av de olika $E(N(t))$ för alternativen ser vi att

$$\frac{\text{Alt. 1}}{\text{Alt. 2}} = \frac{\frac{\rho}{1-\rho}}{\frac{2\rho}{1-\rho^2}} = \frac{1+\rho}{2} < 1,$$

Detta är en allmän princip. Om vi inte har ett problem som har en struktur som gör att parallell hantering snabbar upp så går det snabbare med ett dubbelt så snabbt enkelsystem som två långsammare.

och om $a = 1$ blir $E(N(t)) = K/2$.

Den verkliga intensiteten λ_{in} in i systemet ges av

$$\lambda_{\text{in}} = (1 - P(N(t) = K))\lambda = (1 - \pi_K)\lambda.$$

Intensiteten att kunder avvisas ges alltså av $\lambda\pi_K$.

Vi har även $E(T) = E(N(t))/\lambda_{\text{in}}$ och $E(W) = E(T) - E(S)$, samt att nyttjandegraden ρ ges av $\rho = \lambda_{\text{in}}/\mu = a(1 - \pi_K)$.



Exempel

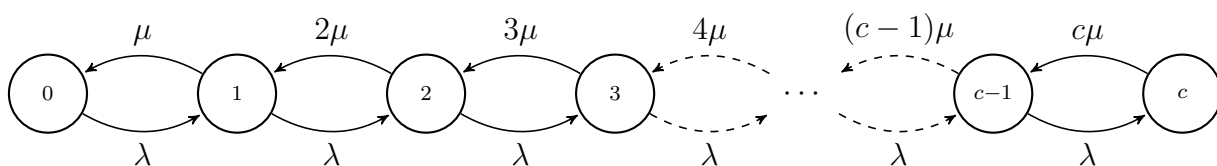
En dataswitch har förhållandet 0.5 mellan ankomstintensiteten λ och betjäningensintensiteten μ , dvs $a = \lambda/\mu = 0.5$. Hur stor buffert behövs för att sannolikheten för paketförlust blir $< 10^{-3}$?

Lösning: Sannolikheten för paketförlust är $P(N(t) = K) = \pi_K = \frac{(1-a)a^K}{1-a^{K+1}}$. Vi testar olika värden på K och finner att $K = 9$ stycken är den minsta längden på buffert vi kan acceptera. Sitter man vid en dator kan man använda MATLAB:

```
>> a = 0.5;
>> K = (1:1:10);
>> ((1-a).*a.^K) ./ (1 - a.^(K+1))
ans =
    0.3333    0.1429    0.0667    0.0323    0.0159    0.0079    0.0039    0.0020    0.0010    0.0005
>> ((1-a).*a.^K) ./ (1 - a.^(K+1)) < 10^(-3)
ans =
     0     0     0     0     0     0     0     0     1     1
```

10.4 M/M/c/c

Specialfallet när $K = c$ är intressant då det innebär att vi helt enkelt inte har någon kö, utan bara c betjäningstillfällen och kunder som anländer när det är fullt vid betjäningarna avvisas. Markovkedjan är ändlig, och kan ses nedan.



Eftersom kedjan är ändlig behövs inget extra villkor på λ och μ . Låt $a = \lambda/\mu$. Den stationära fördelningen ges av

$$\pi_0 = \left(\sum_{k=0}^c \frac{a^k}{k!} \right)^{-1} \quad \text{och} \quad \pi_k = \pi_0 \frac{a^k}{k!}, \quad 1 \leq k \leq c.$$

Det följer att

$$E(N(t)) = \sum_{k=0}^c k\pi_k = \pi_0 a \sum_{k=1}^c \frac{a^{k-1}}{(k-1)!} = \pi_0 a \sum_{k=0}^{c-1} \frac{a^k}{k!} = a \sum_{k=0}^{c-1} \pi_k = a(1 - \pi_c).$$

Det finns ingen kö, så $E(W) = 0$ och $E(T) = E(S) = \frac{1}{\mu}$.