Föreläsning 3, Matematisk statistik 7.5 hp för E
Flerdim

Stas Volkov

∗∗∗

Stokastisk variabel

En **stokastisk variabel** eller **slumpvariabel** är ett **tal** vars värde styrs av slumpen (en funktion $\Omega \to R$).
Bet. $X, Y, \ldots$.
Kan vara **diskret** eller **kontinuerlig**.

En stokastisk variabel beskrivs av:

Sannolikhetsfunktion För en diskret s.v. X

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

Täthetsfunktion För en kontinuerlig s.v. X har vi $f_X(x)$.

$$P(X \in A) = \int_A f_X(x)\,dx$$

Fördelningsfunktion Summa av $p_X(k)$ eller integral av $f_X(x)$.

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

Standardfördelningar & kvantiler

- Diskret fördelning
  - Binomialfördelning
  - Poissonfördelning
  - ffg-fördelning
  - Geometrisk fördelning
- Kontinuerlig fördelning
  - Rektangel- eller likformig fördelning
  - Exponentialfördelning
  - Normalfördelning

α-kvantil, x_α

En **kvantil**, x_α, till en s.v. X är en gräns som överskrids med slh α. Den fås som lösning till någon av följande ekvationer.

$$F_X(x_\alpha) = 1 - \alpha \iff \int_{-\infty}^{x_\alpha} f_X(x)\,dx = 1 - \alpha \iff \int_{x_\alpha}^{\infty} f_X(x)\,dx = \alpha$$

Exempel: Keno-3 (igen)

I Keno-3 väljs 3 av 70 nr. Vid dragning väljs 20 av dessa 70 ut som vinstnummer. Låt X = Antal vinstnr man prickar in.
Sannolikhetsfunktionen för X är

j	0	1	2	3
$p_X(j)$	0.36	0.45	0.17	0.02

Två vinstnr ger 5 kr och 3 vinstnr ger 90 kr.
Sätt Y = Vinsten (kr). Hur ser fördelningen för Y ut?

k	0	5	90
$p_Y(k)$	0.81	0.17	0.02

Transformation av stokastiska variabler

Givet en s.v. X. Vilken fördelning får $Y = g(X)$?
Om X (och därmed Y) är diskret kan man räkna ut sannolikhetsfunktionen

$$p_Y(k) = \sum_{j;g(j)=k} p_X(j)$$

dvs $P(Y = k)$ fås genom att ”lägga ihop $p_X(j)$ för alla j sådana att $g(j) = k$”.

Metod om X (och Y) är kontinuerlig:

1. Sätt upp $F_Y(y) = P(Y \leq y)$.
2. Stoppa in $Y = g(X)$ och uttryck $F_Y(y)$ som funktion av $F_X(\cdot)$.
3. Derivera för att få $f_Y(y)$ som funktion av $f_X(\cdot)$

Exempel

1. Vilken täthetsfunktion har $Y = 2 + 3X$ om X har täthet $f_X(x)$?
2. Vilken täthetsfunktion har $Y = \pi X^2$ om $X \in R(-1, 1)$?
3. Vilken fördelning får man om man stoppar in en kont. s.v. X i sin egen fördelningsfunktion $F_X(x)$?
4. Vilken fördelning får man om man stoppar in $X \in R(0, 1)$ i inversen till fördelningsfunktionen för en kont s.v. Y?
   dvs $Z = F_Y^{-1}(X)$.

Lösning till (2)

Vilken täthetsfunktion har $Y = \pi X^2$ om $X \in R(-1, 1)$?

Antag $0 \le y \le \pi$. Då

$$P(Y \le y) = P(X^2 \le y/\pi) = P(-\sqrt{y/\pi} \le X \le \sqrt{y/\pi})$$
$$= \int_{-\sqrt{y/\pi}}^{\sqrt{y/\pi}} \frac{1}{2} dx = \sqrt{y/\pi}$$

eftersom $f_X(x) = 1/2$ för $-1 \le x \le 1$.

Som resultat, $f_Y(y) = \frac{d}{dy} P(Y \le y) = \begin{cases} \frac{1}{2\sqrt{\pi y}}, & 0 \le y \le \pi; \\ 0, & \text{för övrigt.} \end{cases}$

Inversmetoden

Hur drar vi slumptal från en godtycklig kontinuerlig s.v. med fördelningsfunktion $F_Y(y)$?

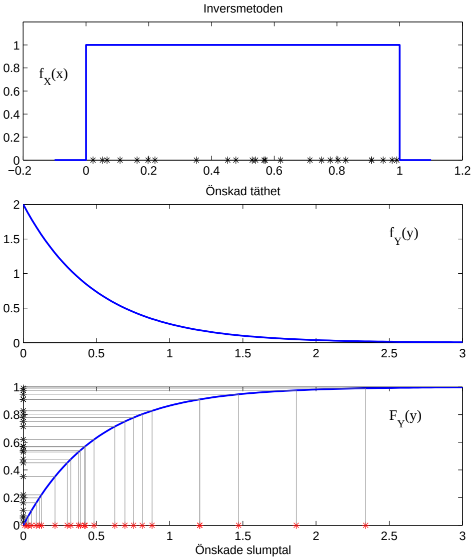
- Låt $X \in R(0,1)$, dvs $F_X(x) = x,\ 0 \leq x \leq 1$.
- Finn inversen $F_Y^{-1}(y)$, till fördelningsfunktionen $F_Y(y)$ för Y.
- Bestäm fördelningsfunktionen för $Z = F_Y^{-1}(X)$

$$
\begin{aligned}
F_Z(z) =& P(Z \leq z) = P(F_Y^{-1}(X) \leq z) = P(F_Y(F_Y^{-1}(X)) \leq F_Y(z)) \\
=& P(X \leq F_Y(Z)) = F_X(F_Y(z)) = F_Y(z)
\end{aligned}
$$

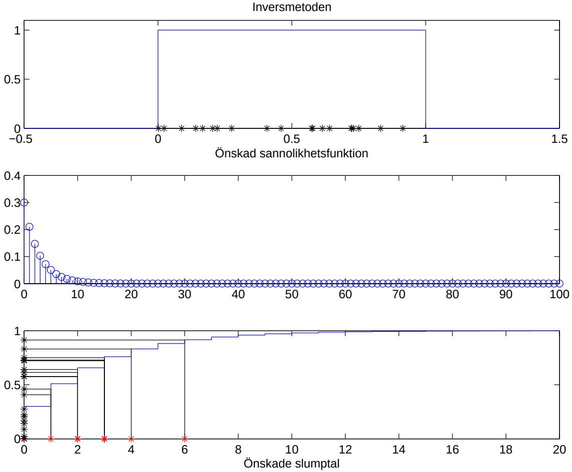
För att dra slumptal från en fördelning med fördelningsfunktion $F_Y(y)$:

1. Räkna ut $F_Y^{-1}(y)$
2. Dra slumptal från en $R(0,1)$-fördelning.
3. Beräkna $F_Y^{-1}(y)$ för varje slumptal.

Kontinuerlig fördelning $Exp(1/2)$



Diskret fördelning $\mathsf{Ge}(0.3)$



Tvådim. stokastisk variabel (X, Y)

Simultan fördelningsfunktion: $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$

Simultan sannolikhetsfunktion: $p_{X,Y}(j, k) = P(X = j, Y = k)$

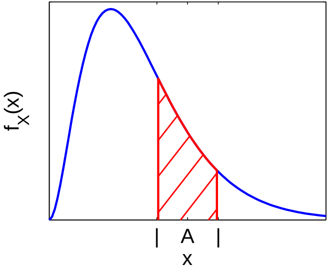
Simultan täthetsfunktion: $f_{X,Y}(x, y) = \dfrac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$

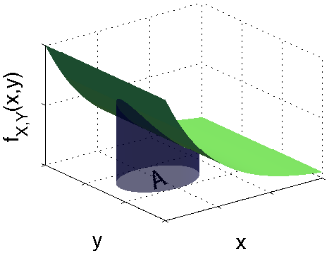
Några egenskaper:

- $P[(X, Y) \in A] = \sum_{(j,k)\in A} p_{X,Y}(j, k)$
- $P[(X, Y) \in A] = \iint_A f_{X,Y}(x, y)\, dxdy$
- $p_X(j) = \sum_k p_{X,Y}(j, k)$ Marginell slh.funkt. för X
- $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$ Marginell täthet för Y

Sannolikheten att hamna i A ges av integralen av tätheten över A

$$P(X \in A) = \int_A f_X(x)\,dx \qquad P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dxdy$$

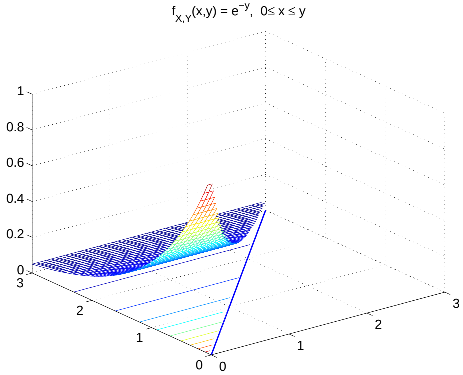




Ex: Radioaktivt sönderfall

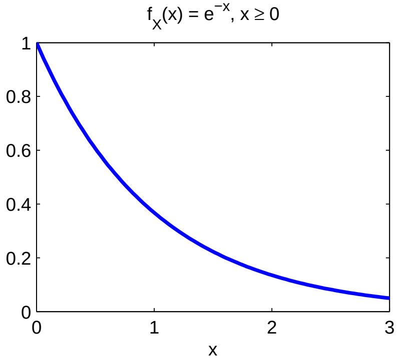
Vid observation av sönderfall beskrivs den simultana fördelningen för tidpunkten för första, X, och andra, Y, sönderfallet av

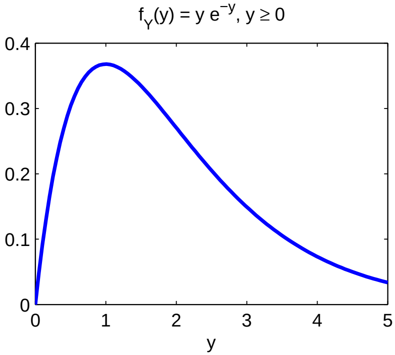
$$f_{X,Y}(x,y) = e^{-y},\ 0 \le x \le y$$



Ex: Om $f_{X,Y}(x,y) = e^{-y},\ 0 \le x \le y$

- Vad blir $f_X(x)$ och $f_Y(y)$?





Oberoende stokastiska variabler

Oberoende

Händelserna A och B är oberoende $\iff$ $P(A \cap B) = P(A) \cdot P(B)$

X och Y kallas **oberoende** stokastiska variabler om
$P(\{X \in C\} \cap \{Y \in D\}) = P(X \in C) \cdot P(Y \in D)$
för **alla** $C \subseteq \mathbb{R}$ och $D \subseteq \mathbb{R}$

$$\iff$$

$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$ för **alla** (x, y)

$\iff$	$\iff$
$p_{X,Y}(j, k) = p_X(j) \cdot p_Y(k)$ för **alla** (j, k)	$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ för **alla** (x, y)

Betingade fördelningar

Betingad slh

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

- Betingad sannolikhetsfunktion för X givet att $Y = k$

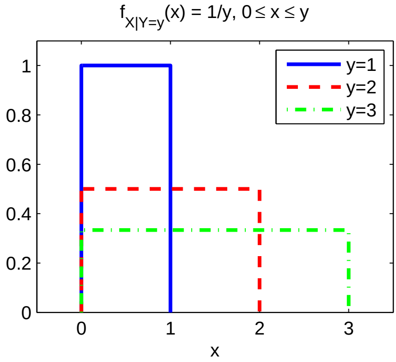
$$p_{X|Y=k}(j) = \frac{p_{X,Y}(j,k)}{p_Y(k)}$$

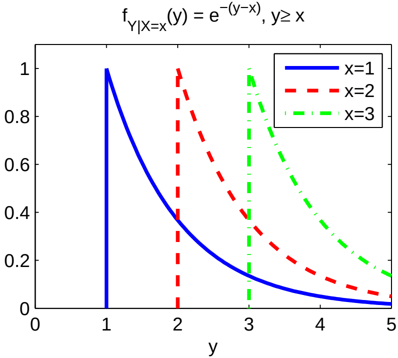
- Betingad täthetsfunktion för X givet att $Y = y$

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Ex (forts): Om $f_{X,Y}(x,y) = e^{-y},\ 0 \leq x \leq y$

- Vad blir $f_{Y|X=x}(y)$ och $f_{X|Y=y}(x)$?





Simulering av tvådimensionell fördelning med hjälp av betingning

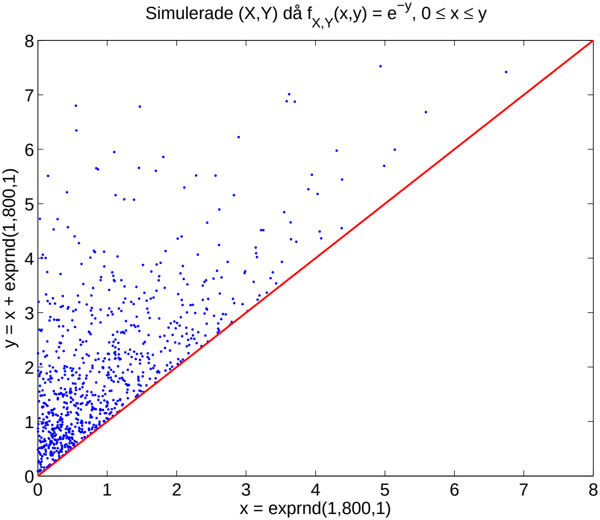
Vet man fördelningen för X och för $Y \mid X = x$ kan man först simulera X och sedan Y via $Y \mid X = x$.

Om $f_{X,Y}(x,y) = e^{-y},\ 0 \leq x \leq y$ blir (enligt ex. på tavlan)

$$f_X(x) = e^{-x},\ x \geq 0 \qquad \text{dvs } X \in Exp(1)$$
$$f_{Y|X=x}(y) = e^{-(y-x)},\ y \geq x \quad \text{dvs } Y \mid X = x \in \text{"}x + Exp(1)\text{"}$$

N stycken par (X, Y) kan då simuleras med

```
x = exprnd(1,N,1);
y = x + exprnd(1,N,1);
```


Simulerade (X,Y) då $f_{X,Y}(x,y) = e^{-y}$, $0 \le x \le y$
y = x + exprnd(1,800,1)
x = exprnd(1,800,1)
0
1
2
3
4
5
6
7
8

Satsen om total sannolikhet igen

Total slh

$$P(A) = \sum_i P(A \mid H_i) \cdot P(H_i)$$

- För sannolikhetsfunktioner

$$p_X(j) = \sum_k p_{X|Y=k}(j) \cdot p_Y(k)$$

- För täthetsfunktioner

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y=y}(x) \cdot f_Y(y)\, dy$$

Exempel: kombinera diskret och kontinuerligt

Vi är intresserade av Y = **antalet fel hos en komponent under 1 timme** men vi vet att felintensiteten varierar slumpmässigt från komponent till komponent. Vi antar att X = **felintensiteten (timme^{-1}) för en komponent** är $X \in Exp(1)$ och $Y \mid X = x \in Po(x)$, dvs

$$f_X(x) = e^{-x},\ x \geq 0 \text{ resp. } \quad p_{Y|X=x}(k) = e^{-x}\frac{x^k}{k!},\ k = 0, 1, 2, \ldots$$

Vilken fördelning har då Y?

$$p_Y(k) = \int_{-\infty}^{\infty} p_{Y|X=x}(k) \cdot f_X(x)\,dx = \int_0^{\infty} e^{-x}\frac{x^k}{k!} \cdot e^{-x}\,dx$$

$$= \int_0^{\infty} \frac{x^k}{k!} \cdot e^{-2x}\,dx = [y = 2x] = \frac{1}{2^{k+1}k!}\int_0^{\infty} y^k \cdot e^{-y}\,dy$$

$$= \frac{1}{2^{k+1}} = \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right)^k,\ k = 0, 1, 2, \ldots \quad \text{dvs } Ge(1/2).$$