Föreläsning 2, Matematisk statistik 7.5hp för E
Slumpvariabel

Stas Volkov

***

Grundläggande begrepp

- **Utfall** – resultatet av ett slumpmässigt försök. Bet. $\omega_1, \omega_2, \ldots$
- **Händelse** – en samling av ett eller flera utfall. Bet. $A, B, \ldots$
- **Utfallsrum** – mängden av möjliga utfall. Bet. Ω

Kolmogorovs axiomsystem

$$0 \leq P(A) \leq 1$$

En sannolikhet är ett tal mellan 0 och 1

$$P(\Omega) = 1$$

Sannolikheten att **något** skall hända är 1

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

Om A_1, A_2, ... är **oförenliga**

Några viktiga samband

Additionssatsen:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Betingad sannolikhet:

$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ defineras bara om $P(B) > 0$

Total sannolikhet:

$P(A) = \sum_{i=1}^{n} P(A \mid H_i) \cdot P(H_i)$ om $H_i \cap H_j = \emptyset,\ i \neq j$ och $\bigcup_{i=1}^{n} H_i = \Omega$

Bayes sats:

$P(H_i \mid A) = \frac{P(A|H_i) \cdot P(H_i)}{P(A)}$

Oberoende:

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

Stokastisk variabel

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

En stokastisk variabel är

- **Diskret** – om den kan anta ett ändligt antal (ex $1, 3, \pi$), eller uppräkneligt oändligt (ex $0, 1, 2, \ldots$).
  Ex:
  - X = **Antal** sönderfallande partiklar i ett radioaktivt ämne under 1 s.
  - Y = **Antal** personer av 100 som svarar ja på frågan ”Röstade du för republikanska partiet i mellanvalet igår?”
- **Kontinuerlig** – om den kan anta alla reella tal i ett intervall, typsikt resultatet av en mätning.
  Ex:
  - X = Hastigheten hos nästa bil som passerar.
  - Y = Temperaturen i morgon.

Sannolikhetsfunktion

För en **diskret** s.v. X definieras **sannolikhetsfunktionen** som

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

Några egenskaper:

- $0 \leq p_X(k) \leq 1$, eftersom de är sannolikheter
- $P(a \leq X \leq b) = \sum_{k=a}^{b} p_X(k)$
- $\sum_{\text{alla } k} p_X(k) = 1$. Slh att X skall anta **något** värde är 1.

Observera att en stokastisk variabel jämförd med ett tal, dvs. $\{X = k\}$ är en händelse.

Exempel: Keno-3

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

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

Två vinstnr ger 5 kr och 3 vinstnr ger 90 kr.

a) Vad är sannolikheten att vinna något?

b) Vad är sannolikheten att man fått två vinstnummer under förutsättning att man vunnit?

Binomialfördelning

Beteckning: $X \in \mathrm{Bin}(n, p)$ där $0 < p < 1$ och $n \in \{1, 2, \ldots\}$

Förekomst: Ett slumpmässigt försök med en händelse A där $P(A) = p$ upprepas n oberoende gånger,
$X =$ Antal ggr A inträffar.

Sannolikhetsfunktion:

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

Ex: X=Antal sexor på tio tärningskast. $X \in \mathrm{Bin}(10, 1/6)$

Poissonfördelning

Beteckning: $X \in Po(\mu)$ där $\mu > 0$

Förekomst: Räknar antal händelser.

Sannolikhetsfunktion:

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

Ex: Bilar som passerar på en väg. Antal radioaktiva sönderfall.

ffg-fördeling

Beteckning: $X \in \text{ffg}(p)$ där $0 < p < 1$

Förekomst: Försöket med händelsen A upprepas oberoende, där $P(A) = p$.
X=Antal försök **tills** A inträffar **f**ör **f**örsta **g**ången.

Sannolikhetsfunktion:

$$p_X(k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

Ex: Antal slantsinglingar **t.o.m.** första krona. $X \in \text{ffg}(1/2)$.

Geometrisk fördelning

Beteckning: $Y \in \mathrm{Ge}(p)$ där $0 < p < 1$

Förekomst: Försöket upprepas.
Y=Antal försök **innan** A inträffar första gången (dvs $Y = X - 1$), där $P(A) = p$.

Sannolikhetsfunktion:

$$p_Y(k) = p(1-p)^k, \quad k = 0, 1, \ldots$$

Ex: Antal tärningskast **innan** första sexa. $X \in \mathrm{Ge}(1/6)$.

Täthetsfunktion

En **kontinuerlig** s.v. X har i stället en **täthetsfunktion** $f_X(x)$.

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

Några egenskaper:

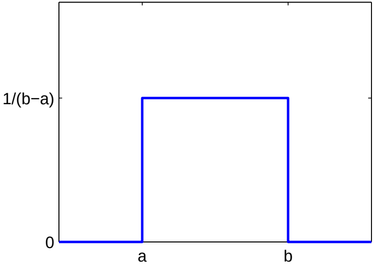
- $f_X(x) \geq 0$
- $P(a \leq X \leq b) = P(a < X < b) = \int_a^b f_X(x)\,dx$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$. Slh att X skall anta **något** värde är 1.

Rektangel- eller likformig fördelning

Beteckning: $X \in R(a,b)$ eller $X \in U(a,b)$ (engelska *Uniform*)

Täthetsfunktion:

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

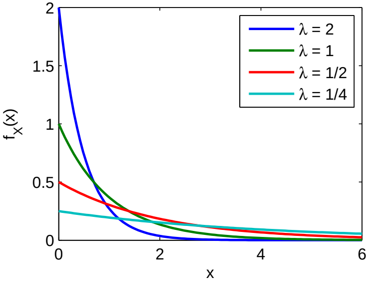


Exponentialfördelning

Beteckning: $X \in Exp(\lambda)$ (eller $X \in \Gamma(1, \lambda)$) där $\lambda > 0$

Täthetsfunktion:

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



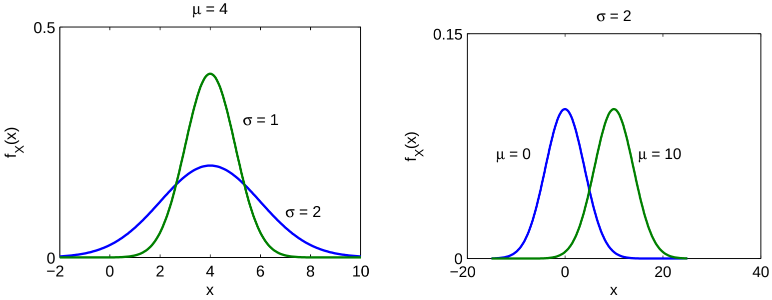
OBS! Ibland (t.ex i Matlab) används bet. $Exp(\mu)$ där $\mu = 1/\lambda$.

Normalfördelning

Beteckning: $X \in \mathrm{N}(\mu, \sigma)$ där $\mu \in \mathbb{R}, \sigma > 0$

Täthetsfunktion:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$



Fördelningsfunktion

För att räkna ut sannolikheter behöver man summera $p_X(k)$ eller integrera $f_X(x)$. Det kan därför vara användbart att ha en **fördelningsfunktion** (borde heta *kumulativ fördelningsfunktion*)

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

Några egenskaper:

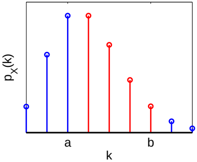
- $0 \leq F_X(x) \leq 1$, eftersom det är en sannolikhet
- $F_X(x)$ är växande (icke strängt).

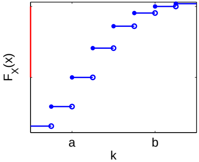
Diskret	Kontinuerlig
$F_X(x) = \sum_{k \leq x} p_X(k)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$
$p_X(k) = F_X(k) - F_X(k-1)$	$f_X(x) = \frac{d}{dx} F_X(x)$

Fördelningsfunktion

Diskret

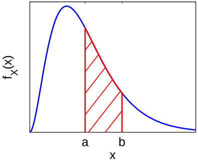
$$P(a < X \leq b) = \sum_{k=a+1}^{b} p_X(k) \qquad P(a < X \leq b) = F_X(b) - F_X(a)$$

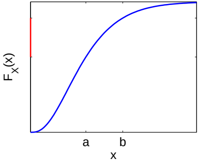




Kontinuerligt

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

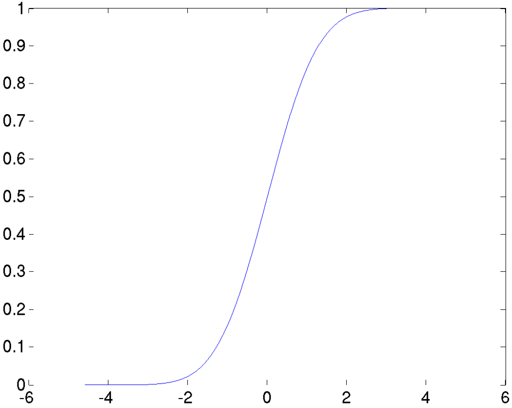




Fördelningsfunktion N(0,1)

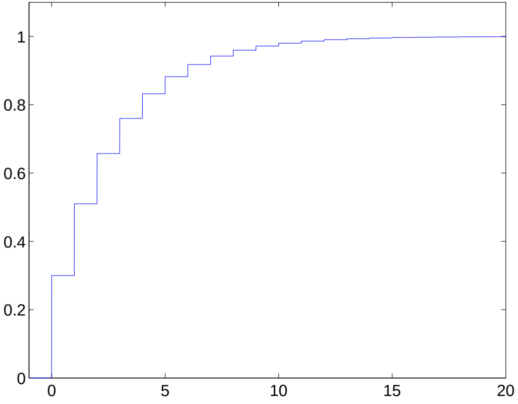
$$F_X(x) = \Phi(x)$$

där $\Phi(x)$ räknas ut numeriskt eller fås från tabell.



Fördelningsfunktion Ge(0.3)

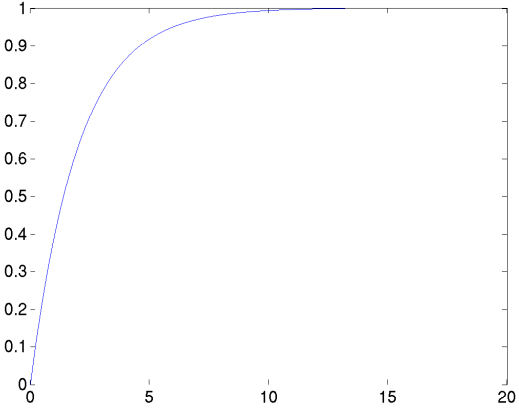
$$F_X(x) = \begin{cases} 1 - (1 - 0.3)^{\lfloor x \rfloor + 1} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



...stegfunktion

Fördelningsfunktion Exp(0.5)

$$F_X(x) = \begin{cases} 1 - e^{-0.5x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



Exempel — Glödlampa

Låt X = Livslängden hos en glödlampa i år. Antag att fördelningen för X beskrivs av följande täthetsfunktion

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

a) Beräkna $F_X(x)$ samt skissa den och $f_X(x)$.

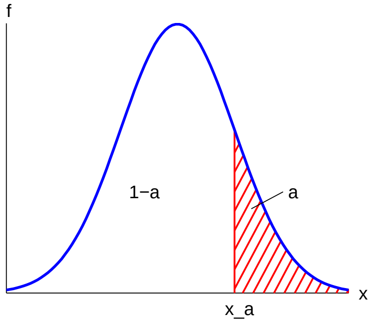
b) Beräkna sannolikheten att lampan håller minst två år.

c) Om vi sett att lampan lyst ett år, vad är sannolikheten att den lyser två år till?

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.

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



Eller direkt ur $x_\alpha = F_X^{-1}(1-\alpha)$

Exempel — Glödlampa (forts.)

Låt X = Livslängden hos en glödlampa i år. Antag att fördelningen för X beskrivs av följande täthetsfunktion

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

a) Beräkna kvantilen x_α som funktion av α.

b) Beräkna numeriskt de tre kva**r**tilerna $x_{0.25}$, $x_{0.50}$ och $x_{0.75}$. ($x_{0.50}$ kallas även **median**)

c) Gör en konkret tolkning av $x_{0.25}$.

Sammanfattning

En stokastisk variabel X är

diskret	om den kan anta ett *ändligt* eller *uppräkneligt* antal värden, typiskt några positiva heltal;
kontinuerlig	om den kan anta alla reella tal i ett intervall.

En stokastisk variabel X beskrives av:

- Sannolikhetsfunktion $p_X(k) = P(X = k)$ om X är diskret.
- Täthetsfunktion $f_X(x)$ om X är kontinuerlig.
- Fördelningsfunktion $F_X(x) = P(X \leq x)$ i alla fall; bra att räkna ut sannolikheter med.

En α-kvantil, x_α, är en gräns som överskrids med sannolikhet α.