

SF1901: SANNOLIKHETSTEORI OCH
STATISTIK
FÖRELÄSNING 4
KONTINUERLIGA STOKASTISKA VARIABLER

Tatjana Pavlenko

8 september 2015



PLAN FÖR DAGENS FÖRELÄSNING

- ▶ Repetition av *diskreta stokastiska variabler*. Väntevärde och varians av diskreta s. v.
- ▶ Kontinuerlig stokastisk variabel (Kap. 3.5–3.6)
- ▶ Exempel på kontinuerliga fördelningarna.
- ▶ Väntevärde och varians av kontinuerliga s. v. (Kap. 5.3)
- ▶ Fördelningsfunktion (Kap. 3.7)
- ▶ Kvantiler (Kap. 3.7)
- ▶ Funktioner av stokastiska variabler. (Kap 3.10)



- ▶ **Def:** En stokastisk variabel, $X(\omega)$ är **diskret** om den endast kan anta ändligt eller uppräkneligt oändligt antal värden $\{k_1, k_2, \dots\}$, (syftar på heltal).
- ▶ **Def:** $p_X(k) = P(X = k)$, $k = k_1, k_2, \dots$ kallas för **sannolikhetsfunktionen** för en diskret s.v. X .
- ▶ Villkor:
 - ▶ $0 \leq p_X(k) \leq 1$ för alla k
 - ▶ $\sum_{\text{alla } k} p_X(k) = 1$
- ▶ Med hjälp av $p_X(k)$ har vi:
 - ▶ $P(a \leq X \leq b) = \sum_{k:a \leq k \leq b} p_X(k)$
 - ▶ $P(X \leq a) = \sum_{k:k \leq a} p_X(k)$
 - ▶ $P(X > a) = \sum_{k:k > a} p_X(k) = 1 - \sum_{k:k \leq a} p_X(k) = 1 - P(X \leq a)$



- ▶ **Def:** Väntevärde för en diskret s.v. X definieras av

$$\mu_X = E(X) = \sum_{\text{alla } k} kp_X(k).$$

- ▶ **Sats (5.1, kap. 5.2):** Låt X vara en s.v., $g(\cdot)$ är en reellvärd funktion och låt s.v. Y vara definierad av $Y = g(X)$. Då gäller att

$$E(Y) = \sum_{\text{alla } k} g(k)p_X(k).$$

- ▶ **Tolkning:** Man får väntevärdet för den nya s.v. $Y = g(X)$ genom att för varje tänkbart värde k av den s.v. X multiplicera $g(k)$ med tillhörande sannolikhet, varefter produkterna adderas.



- ▶ **Def:** Antag att en diskret s.v X har väntevärde $E(X) = \mu$.
Då definieras variansen för X som

$$\sigma^2 = E((X - \mu)^2) = \sum_{\text{alla } k} (k - \mu)^2 p_X(k).$$

- ▶ Variansberäkning (räkneregler, mycket användbar, se bevis i Sats 5.6, kap. 5.3).

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

- ▶ Standardavvikelsen, $D(X) = \sqrt{V(X)}$ (mer avnämt spridningsmått), har samma enhet som variabel själv.
- ▶ Ex på tavlan!



KONTINUERLIGA STOKASTISKA VARIABLER

- ▶ En kontinuerlig s.v. kan anta *alla* värden i ett intervall av R^1 eller i flera åtskillda intervall av R^1 , t ex $[0, \infty)$, $[1, 2]$. För en s.v har vi helt kontinuum av tänkbara värden.
- ▶ Utfallen ligger *oändligt tätt* så att **ingen utfall kan antas med positiv sannolikhet**, dvs sannolikhetsfunktion kan inte definieras på samma sätt som vi gjorde för diskreta s.v.
- ▶ Istället: För en kontinuerlig s.v. läggs sannolikhetsmassan 1 ut på R^1 enligt *täthetsfunktion* $f_X(x)$, $x \in R^1$.



KONTINUERLIGA STOKASTISKA VARIABLER (FORTS.)

- ▶ **Def:** En stokastisk variabel X är **kontinuerlig** om det finns icke-negativ funktion $f_X(\cdot)$ sådan att

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

för alla A . $f_X(x)$ kallas för täthetsfunktionen för s.v X .

- ▶ Jämför med diskreta fallen! Summeringen av sannolikhetsfunktionen ersats av integration.
- ▶ Villkor:
 - ▶ $f_X(x) \geq 0$,
 - ▶ $\int_{-\infty}^{\infty} f_X(x) dx = 1$, dvs hela area under täthetsfunktionen är 1.
- ▶ Skilj noga på symbolen X som betecknar en s.v. och x som används som argument i funktionen $f_X(x)$!



VANLIGA KONTINUERLIGA FÖRDELNINGAR

- ▶ *Kontinuerlig likformig fördelning.*
- ▶ **Def:** Om en s.v. X har täthetsfunktion

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

sägs X vara **likformigt fördelad mellan a och b** . Beteckning: $X \in U(a, b)$. U är från engelska *uniform*.

- ▶ **Tolkning:** Eftersom vi betraktar en kontinuerlig s.v. kan vi inte definiera den som att alla värden mellan a och b är lika sannolika (jmf. med det diskreta fallet!) - enskilda värden har alltid sannolikhet noll för kontinuerliga s.v., dvs $P(X = x) = 0$. Det man istället menar är att **sannolikheten att s.v. X ligger i något givet intervall inom (a, b) bara beror på intervallets bredd och inte på *var* intervallet ligger.**



- ▶ *Exponentialfördelning.*
- ▶ **Def:** Om en s.v. X har täthetsfunktion

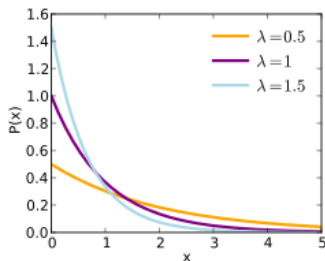
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{om } x > 0, \\ 0 & \text{f.ö.} \end{cases},$$

$\lambda > 0$, sägs X vara **exponentialfördelad**. Bet: $X \in \text{Exp}(\lambda)$.

- ▶ En viktig genskap hos $\text{Exp}(\lambda)$: Exponentialfördelning *saknar minne!*
- ▶ Minneslöshet hos $\text{Exp}(\lambda)$ på tavlan (se anm. 3.2, kap. 3.6)!
- ▶ Förekomst: används för att modellera t ex tiden tills någon får sitt nästa telefonsamtal, tiden tills någon råkar ut för sin nästa bilolycka eller avståndet mellan mutationer på en DNA-sträng.



VANLIGA KONTINUERLIGA FÖRDELNINGAR (FORTS.)



FIGUR: Täthetsfunktion $f_x(s)$ uppritad för tre olika exponentialfördelningar: $\lambda = 0.5$, $\lambda = 1$ respektive $\lambda = 1.5$. Som synes ju mindre λ är, desto mer utbredd är sannolikhetsmassan över intervallet $(0, \infty)$.

- ▶ *Normalfördelning.*
- ▶ **Def:** Om en s.v. X har täthetsfunktion

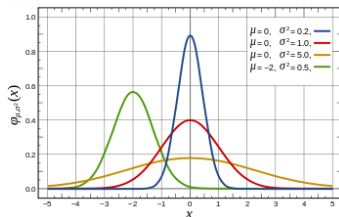
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

där μ och $\sigma > 0$ är givna tal, sägs X vara **normalfördelad**.
Beteckning: $X \in N(\mu, \sigma)$.

- ▶ Viktigaste av alla fördelningar! Kallat även för *Gauss-fördelning*. Benämningen hänsyftar på den tyske matematiker Carl Friedrich Gauss (1777-1855).
- ▶ Att den är så viktig beror på att om man summerar många stokastiska variabler (tänk: något slumpmässighet som beror på många olika faktorer) så är resultatet nästan alltid normalfördelat. Mer om detta under fls. 7, (kap. 6).



VANLIGA KONTINUERLIGA FÖRDELNINGAR (FORTS.)



FIGUR: Täthetsfunktion $f_X(s)$ uppritad för fyra olika normalfördelningar. Man ser att effekten av att ändra μ är att täthetens läge försjuts (med toppen liggandes vid μ), medan fördelningen blir mer koncentrerad när σ är liten, respektive mer utspridd när σ är stor.

VÄNTEVÄRDE OCH VARIANS FÖR KONTINUERLIGA DE S.V.

- ▶ **Def:** Väntevärde av en kontinuerlig s.v X definieras av

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

- ▶ **Sats:** (se Stas 5.1, kap. 5.2) Om $Y = g(X)$ där $g(\cdot)$ är en reelvärd funktion så gäller att

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

- ▶ **Def:** Variansen av en kontinuerlig s.v X definieras av

$$\sigma_X = V(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx.$$

- ▶ På samma sätt som i det diskreta fallet gäller följande (se Sats. 5.6):

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

- ▶ $E(X)$ och $V(X)$ för $X \in \text{Exp}(\lambda)$ på tavlan.



FÖRDELNINGSFUNKTION FÖR EN S.V.

- ▶ Def: Funktionen

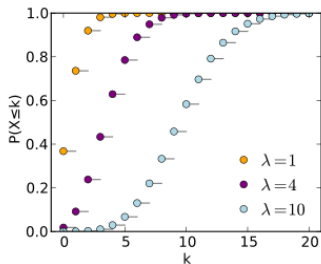
$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}^1$$

kallas för **fördelningsfunktionen** för den s.v. X .

- ▶ Villkor:
 - ▶ $0 \leq F_X(x) \leq 1$, (slh)
 - ▶ $F_X(x)$ är icke-avtagande funktion av x ,
 - ▶ $F_X(x) \rightarrow 0$ då $x \rightarrow -\infty$ och $F_X(x) \rightarrow 1$ då $x \rightarrow \infty$.
 - ▶ $F_X(x)$ är kontinuerlig till höger för varje x .



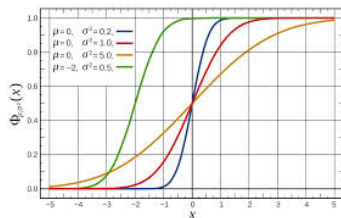
FÖRDELNINGSFUNKTION FÖR EN S.V. (FORTS.)



FIGUR: Fördelningsfunktion $F_X(k)$ upptitad för tre olika Poisson-fördelningar: $\lambda = 1$, $\lambda = 4$ respektive $\lambda = 10$.



FÖRDELNINGSFUNKTION FÖR EN S.V. (FORTS.)



FIGUR: Fördelningsfunktion för $N(\mu, \sigma)$ uppritad för några olika värden på μ och σ .

FÖRDELNINGSFUNKTION FÖR EN S.V. (FORTS.)

- ▶ I det **diskreta** fallet finns det ett nära samband mellan fördelningsfunktionen och sannolikhetsfunktionen:

$$F_X(k) = \sum_{j:j \leq k} p_X(j), \quad p(k) = \begin{cases} F_X(0) & \text{om } k = 0 \\ F_X(k) - F_X(k-1) & \text{f.ö.} \end{cases}$$

- ▶ I det **kontinuerliga** fallet finns ett motsvarande samband mellan fördelningsfunktionen och täthetsfunktionen:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad f_X(x) = \frac{d}{dx} F_X(x),$$

i varje punkt x där $f_X(x)$ är kontinuerlig (se Sats 3.1).

- ▶ Tolkning: bilden på tavlan för båda fallen. Låt $A = (-\infty, x]$,

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



FÖRDELNINGSFUNKTION FÖR EN S.V. (FORTS.)

- ▶ Ofta behöver man beräkna sannolikheter på formen $P(a < X \leq b)$, $P(X \leq a)$, $P(X > a)$, eller liknande. $F_X(x)$ är mycket användbar för sådana beräkningar!
- ▶ En viktig sats (se Sats 3.3): Om $a < b$ så gäller att

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

- ▶ Bevis och exempel på tavlan.
- ▶ **Def:** Lösningen $x = x_\alpha$ till ekvationen

$$F_X(x) = 1 - \alpha$$

kallas för α -quantil för den s.v. X .

