SF1920/SF1921 Sannolikhetsteori och statistik 6,0 hp
Föreläsning 4
Kontinuerliga stokastiska variabler

Jörgen Säve-Söderbergh

Diskreta och kontinuerliga stokastiska variabler

En diskret stokastisk variabel är en funktion från Ω till en mängd vars element är ändligt många eller uppräkneligt många.

Fördelningen för en diskret stokastisk variabel finns på de ändligt många eller de uppräkneligt många värdena som variabeln antar.

- Binomial-fördelningen lever på värdena $0, 1, \ldots, n$.
- Poisson-fördelningen lever över värdena $0, 1, \ldots$.

Nu önskar vi definiera en stokastisk variabel som antar många fler värden än de diskreta, de antar samtliga värden över ett godtyckligt intervall av reella tal.

Utfallen ligger så tätt att vi inte kan se dem!

Från diskret till kontinuerlig stokastiska variabel

Fördelningsfunktionen om X är en diskret stokastisk variabel ges som

$$F_X(x) = \sum_{j \le x} p_X(j) \, .$$

Alltså kumulerar vi sannolikheterna ända till ett givet x. Kumulerar vi dem allihop har vi ett.

En formel som $\int_{-\infty}^{x} f_X(t)\, dt$ gör samma sak,

Om integralen ska kumulera yta under $f_X(t)$ så måste den vara positiv.
Räknar vi över hela $\mathbb{R}$, d.v.s., måste vi ha $\int_{-\infty}^{\infty} f_X(t)\, dt = 1$.

Kontinuerliga stokastiska variabler

Funktionen $f_X(t)$ kallas **frekvens- eller täthetsfunktion**.

Om det finns en funktion $f_X(t)$ sådan att

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

för alla A, säges X vara en *kontinuerlig variabel*.

Om vi låter $A = (-\infty, x]$ ger detta upphov till fördelningsfunktionen

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

Samband mellan $f_X(x)$ och $F_X(x)$

Theorem

I varje punkt x där $f_X(x)$ är kontinuerlig, gäller att

$$\frac{d}{dx}F_X(x) = f_X(x)$$

Om vi integrerar täthetsfunktionen $f_X(x)$ över intervallet $(-\infty, x]$, så erhåller vi fördelningsfunktionen.

Om vi deriverar fördelningsfunktionen $F_X(x)$, så erhåller vi täthetsfunktionen $f_X(x)$.

Likformig fördelning över intervallet (a, b)

Om den stokastiska variabeln X har en täthetsfunktion som

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

säges X vara *likformigt fördelad på* (a, b). $\qquad X \in U(a, b)$

Låt $a = 0$ och $b = 1$. Då blir intervallet $(0, 1)$.

Formeln för att beräkna sannolikheter är $P(X \in A) = \int_A f_X(x)\,dx$.

Vad är sannolikheten att $0 < X < 0.5$?

$$P(0 < X < 0.5) = \int_0^{0.5} 1dx = [x]_{x=0}^{x=0.5} = 0.5$$

Exponentialfördelning

Definition

Om den stokastiska variabeln X har en täthetsfunktion som

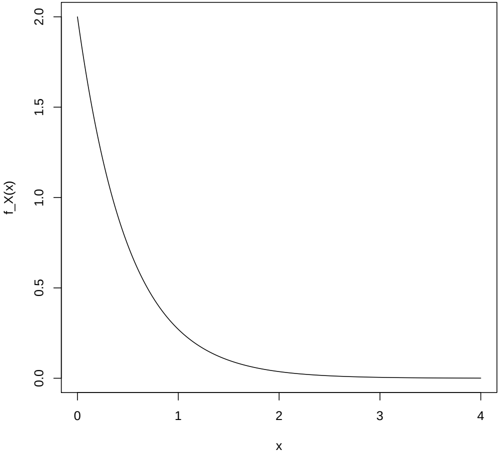
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

där $\lambda > 0$, säges X vara *exponentialfördelad*. $X \in Exp(\lambda)$.

- Väntetidsfördelning. Tiden tills någon händelse inträffar antas ofta följa denna fördelning.
- Exponentialfördelning saknar minne.

Exponentialfördelning $Exp(1/2)$

Om vi sätter $\lambda = \frac{1}{2}$, erhåller vi en bestämd exponentialfördelning:



Normalfördelningen

Definition

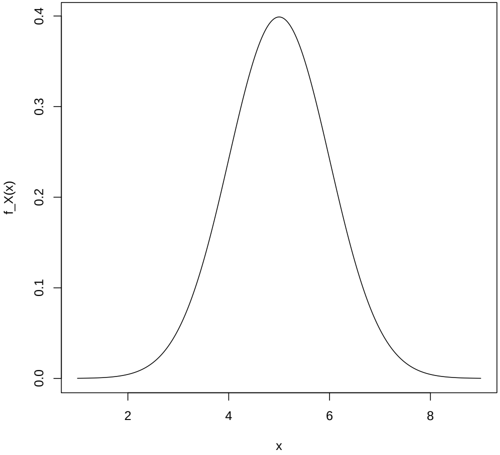
Om den stokastiska variabeln X har en täthetsfunktion som

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

där μ och σ är givna tal ($\sigma > 0$), säges X vara *normalfördelad*. $X \in N\left(\mu, \sigma^2\right)$.

- Statistikens viktigaste fördelning.
- Ungefär 300 år gammal.

Sätt $\mu = 5$ och $\sigma = 1$:



Fördelningsfunktionen för den likformiga fördelningen över $(0,1)$

Vi ska beräkna fördelningsfunktionen för

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{annars} \end{cases}$$

Vi måste skilja på tre olika fall när vi beräknar integralen. För det första, om $-\infty < x < 0$, så har vi

$$F(x) = \int_{-\infty}^{x} f(t)dt = 0,$$

eftersom tätheten är noll överallt över $(-\infty, 0)$. För det andra, om $0 < x < 1$, så följer

$$\begin{aligned} F(x) &= \int_{-\infty}^{0} f(t)dt + \int_{0}^{x} f(t)dt \\ &= \int_{0}^{x} 1dt = [t]_{t=0}^{t=x} = x - 0 = x. \end{aligned}$$

Fördelningsfunktionen för den likformiga fördelningen över $(0, 1)$

Slutligen, om $1 < x < \infty$, har vi att

$$\begin{aligned} F(x) &= \int_{-\infty}^{0} f(t)dt + \int_{0}^{1} f(t)dt + \int_{1}^{x} f(t)dt \\ &= 0 + 1 + \int_{1}^{x} 0dt \\ &= 1 + 0 \\ &= 1. \end{aligned}$$

Alltså blir fördelningsfunktionen

$$F(x) = \begin{cases} 0 & x < 0, \\ x & 0 < x < 1, \\ 1 & x > 1. \end{cases}$$

Skillnader mellan diskreta och kontinuerliga stokastiska variabler

- Sannolikheten att en kontinuerlig stokastisk variabel X antar ett visst tal är noll

$$P(X = x) = \int_x^x f_X(x)dx = 0.$$

Detta är oftast inte fallet med diskreta stokastiska variabler.

- En konsekvens av ovan är att ändpunkter av intervall inte spelar någon roll

$$\begin{aligned} P(a \le X \le b) &= P(a < X < b) = P(a \le X < b) \\ &= P(a < X \le b) = F_X(b) - F_X(a). \end{aligned}$$

Detta gäller inte för diskreta stokastiska variabler.

Lägesmått och spridningsmått för kontinuerliga stokastiska variabler

Om $f_X(x)$ är täthetsfunktionen för en kontinuerlig stokastisk variabel X, så definieras **väntevärdet** för X av

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

Variansen för X är

$$\sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx.$$

Standardavvikelsen för X är

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Likformig fördelning över $[0, 1]$

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Väntevärdet

$$\mu = E(X) = \int_0^1 x(1)dx = \left[\frac{x^2}{2}\right]_{x=0}^{x=1} = \frac{1^2}{2} - \frac{0^1}{2} = \frac{1}{2}.$$

För variansen räknar vi $E(X^2)$ först,

$$E(X^2) = \int_0^1 x^2(1)dx = \left[\frac{x^3}{3}\right]_{x=0}^{x=1} = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}.$$

Då blir variansen

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

α-kvantil

Antag att vi har en viss fördelning som kan framställas med täthetsfunktionen eller fördelningsfunktionen $F_X(x)$.

Definition

Om x är ett tal sådant att arean under tätheten till höger om x är lika med ett tal α, $0 < \alpha < 1$, så bestämmer x den så kallade *α-kvantilen* för fördelningen.

α-kvantilen kan även definieras med hjälp av fördelningsfunktionen :

Lösningen $x = x_\alpha$ till ekvationen

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

kallas α-kvantilen för den stokastiska variabeln X.

α anges ofta i procent (5%-kvantil, 0.1%- kvantil).