Matematisk Statistik och Disktret Matematik, MVE051/MSG810, VT19

Föreläsning 3

Nancy Abdallah

Chalmers - Göteborgs Universitet

April 2, 2019

Kontinuerliga stokastiska variabler

- En **kontinuerlig stokastisk variabel** kan anta alla värden i ett intervall av reella tal (eller alla reella värden).
- För kontinuerliga stokastiska variabler gäller att $P(X = x) = 0$. Däremot $P(a \leq x \leq b) \geq 0$, för alla $a, b \in \mathbb{R}$.
- Det finns en funktion $f(x)$ sådana att

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

  $f(x)$ kallas **täthetsfunktion** *(eng. density function)*.
- **OBS!**

$$P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = P(a < x < b)$$

- En funktion $f(x)$ är en täthetsfunktion om och endast om
  - (i) $f(x) \geq 0$ för alla $x \in \mathbb{R}$, och
  - (ii) $\int_{-\infty}^{+\infty} f(x) = 1$
- Fördelningsfunktionen $F(x)$ är definierad av

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Fördelningsfunktion av en kontinuerlig s.v. är kontinuerlig.

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

$$F'(x) = f(x)$$

Exempel

Visa att funktionen

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

är en täthetsfunktion.

Lösning:

$f(x) \geq 0$ *för alla* $x \in \mathbb{R}$ *och*

$$\int_{-\infty}^{+\infty} f(t)dt = \int_a^b \frac{1}{b-a}dt = \frac{b-a}{b-a} = 1$$

Så är $f(x)$ *en täthetsfunktion. Den stokastiska variabeln som har* $f(x)$ *som täthetsfunktion kallas likformig fördelning.*

Exempel

Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion

$$f(x) = \begin{cases} 12.5x - 1.25 & om 0.1 \leq x \leq 0.5 \\ 0 & annars \end{cases}$$

Beräkna $P(0.3 \leq X \leq 0.6)$ och bestäm fördelningsfunktion $F(x)$.

Lösning:

$$\begin{aligned} P(0.3 \leq X \leq 0.6) &= \int_{0.3}^{0.6} f(x)dx = \int_{0.3}^{0.5} (12.5x - 1.25)dx \\ &= [12.5\frac{x^2}{2} - 1.25x]_{0.3}^{0.5} = 0.75 \end{aligned}$$

Exempel

Fördelningsfunktion:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & om\ x < 0.1 \\ \int_{0.1}^{x}(12.5t - 1.25)dt & om\ 0.1 \leq x < 0.5 \\ 1 & om\ x > 0.5 \end{cases}$$

$$Allts\aa\ F(x) = \begin{cases} 0 & om\ x < 0.1 \\ 6.25x^2 - 1.25x + 0.0625 & om\ 0.1 \leq x < 0.5 \\ 1 & om\ x > 0.5 \end{cases}$$

Vi kan använda $F(x)$ för att beräkna$P(a \leq X \leq b)$.

$P(0.3 \leq X \leq 0.6) = F(0.6) - P(0.3) = 1 - 0.25 = 0.75.$

Väntevärde, varians, standardavvikelse

Låt X vara en kontinuerlig s.v. med täthetsfunktion $f(x)$.

- Väntevärdet av X ges av

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx.$$

- Mer allmänt, Om H är en funktion så kan vi definiera

$$E[H(X)] = \int_{-\infty}^{+\infty} H(x)f(x)dx$$

- Variansen och standardavvikelsen definieras på samma sätt som en diskret s.v., dvs $Var[x] = E(X^2) - E(X)^2$, och $\sigma = \sqrt{Var[x]}$.
- Räkneregler för väntevärdet och variansen är samma som en diskret s.v.

Normalfördelning

- Om en stokastisk variabel X har

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

  som täthetsfunktion, där $\sigma > 0$ och $\mu \in \mathbb{R}$, säges X normalfördelad med parameter μ och σ.
- Notation: $X \in N(\mu, \sigma^2)$.
- $E[X] = \mu$ och $Var[X] = \sigma^2$.

Graf av en normalfördelad s.v.

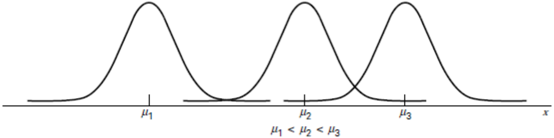


FIGURE 4.6.3 Three normal distributions with different means but the same amount of variability.

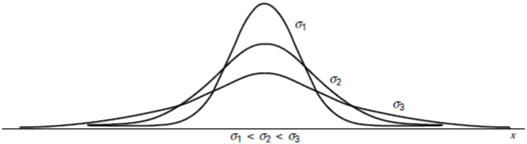
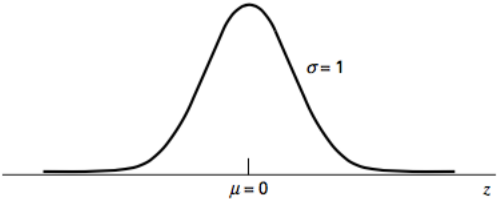


FIGURE 4.6.4 Three normal distributions with different standard deviations but the same mean.

Standard Normalfördelning

Om $X \in N(0, 1)$, säges X att ha en standard normalfördelning. Täthetsfunktionen ser ut som figuren nedan. I så fall använder vi Z i stället av X.



Anmärkning: Grafen av en normalfördelad s.v. är symetrisk kring linjen $z = \mu$.

Låt $Z \in N(0, 1)$. För att beräkna $P(a < Z < b)$ där a och b är reella tal eller $\pm\infty$, använder man tabell V s.697-698 som ger flera värden för $F(x)$.

Sats

Om $X \in N(\mu, \sigma)$ så gäller att $\frac{X-\mu}{\sigma} \in N(0, 1)$.

Sats

Låt $X \in N(\mu, \sigma)$ så gäller att

$$P(\mu - \sigma < X < \mu + \sigma) = 0.68$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.95$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$$

Exempel

Låt $X \in N(17, 5)$ och antag att vi vill beräkna $P(X \leq 20)$. Då utnyttjas variableln $Z = (X - 17)/\sqrt{5}$ för vilken gäller att $Z \in N(0, 1)$.

$$\begin{aligned} P(X \leq 20) &= P\left(\frac{X-17}{\sqrt{5}} \leq \frac{20-17}{\sqrt{5}}\right) = P(Z \leq 1.34) \\ &= F(1.34) = 0.9099 \end{aligned}$$

För vilken värde av x är $P(X > x) = 0.6$?

$P\left(\frac{X-17}{\sqrt{5}} > \frac{x-17}{\sqrt{5}}\right) = P\left(z > \frac{x-17}{\sqrt{5}}\right) = 0.6$

$\Rightarrow P\left(z < \frac{x-17}{\sqrt{5}}\right) = 1 - 0.6 = 0.4 \Rightarrow \frac{x-17}{\sqrt{5}} \approx -2.55$

Alltså $x \approx -2.55\sqrt{5} + 17 = 11.3$

Normalfördelning som approximation till Binomialfördelning

Sats

Låt $X \in Bin(n, p)$. Om [$p \leq 0.5$ och $np > 5$] eller [$p > 0.5$ och $n(1-p) > 5$] så approximeras X med $N(np, np(1-p))$.

Dvs, om n är stor, approximeras X med en normalfördelning med parameter $\mu = np$ och $\sigma^2 = np(1-p)$.

Anmärkning

Märk att Binomialfördelningen är diskret medan normalfördelningen är kontinuerlig: Vi beräknar $P(X \leq x) \approx P(Y \leq x + \frac{1}{2})$ eller $P(X < x) \approx P(Y \leq x - \frac{1}{2})$.

Transformation av kontinuerliga s.v.

- Antag att X har täthetsfunktion f_X och antag variabeln Y är relaterad till X med $h(Y) = X$ där h är strängt monoton och deriverbar. Då gäller för täthetsfunktionen f_Y till Y att

$$f_Y(y) = f_X(h(y))|h'(y)|$$

- Exempel: Om $X = aY + b$ så äar $f_Y(y) = f_X(ay + b)|a|$.
- Exempel: Om $X \in N(\mu, \sigma^2)$ och $Y = \frac{X-\mu}{\sigma}$ så blir $X = \sigma Y + \mu$ och

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}((\sigma y + \mu) - \mu)^2\right)\sigma \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \end{aligned}$$