

Matematisk Statistik och Diskret Matematik, MVE051/MSG810, VT22

Timo Vilkas

Chalmers - Göteborgs Universitet



Föreläsning 3
29 mars 2022



Navigation icons: back, forward, search, etc.

Kontinuerliga stokastiska variabler

- En **kontinuerlig stokastisk variabel** X kan anta alla värden i ett delintervall av $\mathbb{R}$ (reella talen).
- För kontinuerliga stokastiska variabler gäller att
 $\rightarrow \mathbb{P}(X = x) = 0$, för alla $x \in \mathbb{R}$. Däremot $\mathbb{P}(a \leq X \leq b) \geq 0$, för alla $a, b \in \mathbb{R}$.
- Om det finns en funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ sådant att

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$$

kallas f **täthetsfunktion** (eng. *density function* eller *probability distribution function, pdf*) till (fördelningen av) X .

- **OBS!** För alla $a < b \in \mathbb{R}$ gäller Δ_a

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b).$$

Navigation icons: back, forward, search, etc.

- En funktion f är en täthetsfunktion om och endast om

① $f(x) \geq 0$ för alla $x \in \mathbb{R}$, och

② $\int_{-\infty}^{+\infty} f(x) dx = 1$

- Fördelningsfunktionen F är då definierad genom

$$F(x) = \underbrace{\mathbb{P}(X \leq x)}_{= \mathbb{P}(X \in (-\infty, x])} = \int_{-\infty}^x f(t) dt$$

OBS! Fördelningsfunktion till en kontinuerlig s.v. är kontinuerlig (i allmänt fall bara högerkontinuerlig).

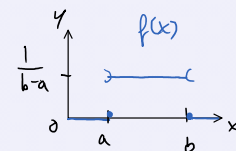
- I varje punkt x där $f(x)$ är kontinuerlig, gäller det att

$$F'(x) = f(x). \quad (\text{bevis: Analysens huvudsats})$$

Exempel

Visa att för $a < b \in \mathbb{R}$, funktionen

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



är en täthetsfunktion

$$\begin{aligned} \text{① } f &\geq 0 \quad \checkmark \quad \text{② } \int_{-\infty}^{\infty} f(x) dx = \int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b 1 dx \\ &= \frac{1}{b-a} \cdot (b-a) = 1 \end{aligned}$$

för t. ex. $c < a$ och $d \in (a, b)$ blir

$$\mathbb{P}(c < X < d) = \int_c^d f(x) dx = \int_a^d \frac{1}{b-a} dx = \frac{d-a}{b-a}.$$

Slumpvariabeln som har f som täthetsfunktion kallas **likformig fördelad** på (a, b) och fördelningen förkortas med **unif** (a, b) .

Exempel

Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion

$$f(x) = \begin{cases} 3x^2 & \text{om } 0 \leq x \leq 1 \\ 0 & \text{annars} \end{cases}$$

Beräkna $\mathbb{P}(0.3 \leq X \leq 0.6)$ och bestäm fördelningsfunktionen F .

- ① $f(x) \geq 0$ för alla x ② $\int_{-\infty}^{\infty} f(x) dx = \int_0^1 3x^2 dx = [x^3]_0^1 = 1 - 0 = 1 \rightarrow$ är en täthetsfunktion
- $\mathbb{P}(0.3 \leq X \leq 0.6) = \int_{0.3}^{0.6} 3x^2 dx = [x^3]_{0.3}^{0.6} = 0.6^3 - 0.3^3 = 0.189$.
- $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & x \leq 0 \\ \int_0^x 3t^2 dt & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$ om $x \in (0,1)$ då
för $x \in (0,1)$: $F(x) = \int_0^x 3t^2 dt = [t^3]_0^x = x^3$.

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Väntevärde, varians, standardavvikelse

Låt X vara en kontinuerlig s.v. med täthetsfunktion f .

- Väntevärdet av X ges då av

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \cdot f(x) dx. \quad \left(\text{så länge } \int_{-\infty}^{+\infty} |x| \cdot f(x) dx < \infty \right)$$

t. ex. $X \sim \text{unif}(a,b)$: $\mathbb{E}[X] = \int_a^b x \cdot \frac{1}{b-a} dx = \left[\frac{x^2}{2(b-a)} \right]_a^b = \frac{a+b}{2}$.

- Mer allmänt, om h är en funktion $\mathbb{R} \rightarrow \mathbb{R}$ så kan vi definiera

$$\mathbb{E}[h(X)] = \int_{-\infty}^{+\infty} h(x) \cdot f(x) dx \quad \left(\text{så länge } \int_{-\infty}^{+\infty} |h(x)| f(x) dx < \infty \right)$$

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

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Normalfördelning

- Om en stokastisk variabel X har täthetsfunktion

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

där $\sigma > 0$ och $\mu \in \mathbb{R}$, kallas X **normalfördelad** med parametrar μ och σ .

- Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$.
- $\mathbb{E}[X] = \mu$ och $\text{Var}[X] = \sigma^2$.

t.ex. $\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$

subst. $t = \frac{x-\mu}{\sigma}$
 $dt = \frac{dx}{\sigma}$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) \exp\left(-\frac{t^2}{2}\right) dt$$

$= \mathbb{P}(Z \in \mathbb{R}) = 1$ (där $Z \sim N(0,1)$)

$$= \underbrace{\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp\left(-\frac{t^2}{2}\right) dt}_{\text{udda}} + \mu \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt = 0 + \mu \cdot 1 = \mu.$$

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Graf till täthetsfunktionen av normalfördelade s.v.

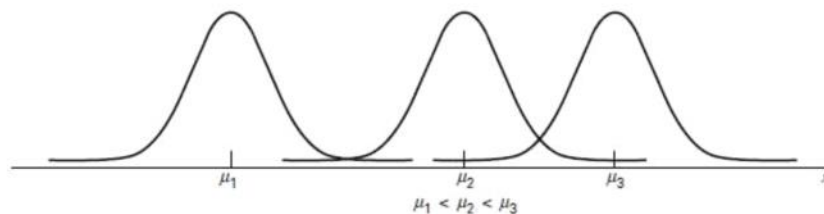


FIGURE 4.6.3 Three normal distributions with different means but the same amount of variability.

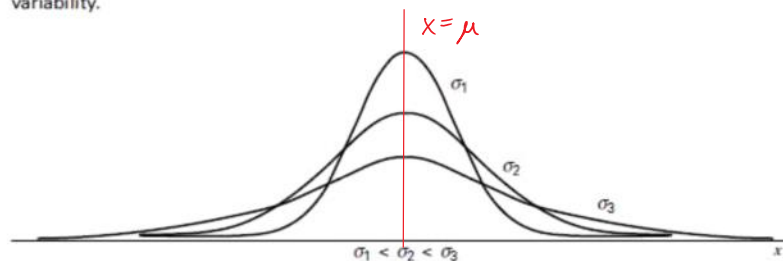


FIGURE 4.6.4 Three normal distributions with different standard deviations but the same mean.

Anmärkning: Täthetsfunktionen till en normalfördelad s.v. är symmetrisk kring linjen $x = \mu$.

T. Vilkas

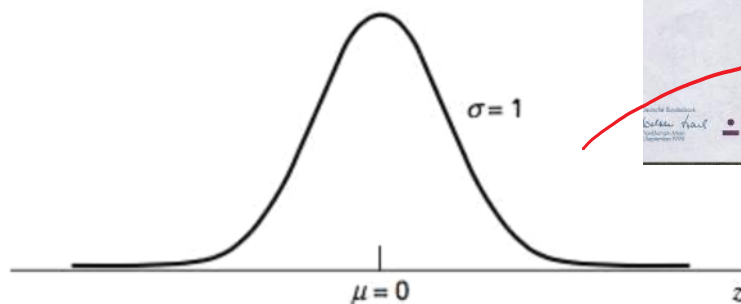
Matematisk statistik & diskret matematik, MVE051/MSG810

Standard Normalfördelning

Om $Z \sim \mathcal{N}(0, 1)$ säges Z att ha **standard normalfördelning**.

Täthetsfunktionen $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ visas i figuren nedan.

Att man använder Z i stället för X ifall av en standard normalfördelning är en gammal konvention.



...fanns med på tyska sedeln!

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Låt $Z \sim \mathcal{N}(0, 1)$. För att beräkna $\mathbb{P}(a < Z < b)$ där a och b är reella tal eller $\pm\infty$, använder man tabell V (s.697-698 i boken) som innehåller värden $F(x)$ för många $x \in \mathbb{R}$.

$$= \mathbb{P}(Z < b) - \mathbb{P}(Z \leq a) = \int_{-\infty}^b f(z) dz - \int_{-\infty}^a f(z) dz,$$

där $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Sats

Om $X \sim \mathcal{N}(\mu, \sigma^2)$ så gäller att $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Bevis: Senare idag.

Sats

Låt $X \sim \mathcal{N}(\mu, \sigma)$ så gäller att

$$\mathbb{P}(\mu - \sigma < X < \mu + \sigma) = \mathbb{P}(-1 < \frac{X-\mu}{\sigma} < 1) = \mathbb{P}(Z < 1) - \mathbb{P}(Z \leq -1) \stackrel{\text{tabell}}{\approx} 0.8413 - 0.1587 = 0.6826$$

$\mathbb{P}(\mu - \sigma < X < \mu + \sigma) = 0.68,$

$\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) = 0.95$ och

$\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997.$

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Exempel

Låt $X \sim \mathcal{N}(14, 8)$ och antag att vi vill beräkna $\mathbb{P}(X \leq 17)$. Då kan vi använda slumpvariabeln $Z = (X - 14)/\sqrt{8}$ som är standard normalfördelad. (enligt satsen ovan)

$$\begin{aligned}\mathbb{P}(X \leq 17) &= \mathbb{P}\left(\frac{X-14}{\sqrt{8}} \leq \frac{17-14}{\sqrt{8}}\right) = \mathbb{P}\left(Z \leq \frac{3}{\sqrt{8}}\right) \\ &\approx \mathbb{P}(Z \leq 1.06) \approx 0.8554.\end{aligned}$$

$\uparrow$
tabell för $\mathcal{N}(0,1)$

För vilket värde av x är $\mathbb{P}(X > x) = 0.6$?

$$\mathbb{P}(X > x) = \mathbb{P}\left(\frac{X-14}{\sqrt{8}} > \frac{x-14}{\sqrt{8}}\right) = 0.6$$

$\underbrace{\frac{X-14}{\sqrt{8}}}_{=: Z \sim \mathcal{N}(0,1)}$

$\Rightarrow \mathbb{P}\left(Z \leq \frac{x-14}{\sqrt{8}}\right) = 1 - 0.6 = 0.4$ enligt tabellen måste då $\frac{x-14}{\sqrt{8}}$ ligga mellan -0.25 och -0.26 , typ $-0.253 \Rightarrow x \approx (-0.253) \cdot \sqrt{8} + 14 \approx 13.28$.

Navigation icons

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Normalfördelning som approximation till Binomialfördelning

Sats

Låt $X \sim \text{Bin}(n, p)$. För fast $p \in (0, 1)$ och stor $n \in \mathbb{N}$ kan X rimligt approximeras med $Y \sim \mathcal{N}(np, np(1-p))$.

Dvs, om n är stor, approximeras X med en normalfördelning med parametrar $\mu = np = \mathbb{E}[X]$ och $\sigma^2 = np(1-p) = \text{Var}[X]$.

I praktiken brukar man betrakta normalapproximationen som rimligt, så länge $\min\{np, n(1-p)\} > 5$.

Anmärkning

Observera att $X \sim \text{Bin}(n, p)$ är diskret medan normalfördelningen är kontinuerlig: Man räknar därför t.ex. för $x \in \mathbb{N}$:

$$P(X \leq x) \approx P(Y \leq x + \frac{1}{2}) \text{ och } P(X < x) \approx P(Y \leq x - \frac{1}{2}).$$

Navigation icons

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Transformation av kontinuerliga slumpvariabler

Sats

Antag att X är en slumpvariabel med täthetsfunktion f_X och $Y = h(X)$, där $h: \mathbb{R} \rightarrow \mathbb{R}$ är strängt monoton och deriverbar. Då är Y en kontinuerlig slumpvariabel med täthetsfunktion

$$f_Y(y) = f_X(h^{-1}(y)) \cdot |(h^{-1})'(y)|.$$

Bevis: Låt h vara strängt växande (fallet strängt avtagande kan du försöka klura ut själv - på exakt samma sätt - och sedan kika i boken).

• Då är $F_Y(y) = P(Y \leq y) = P(h(X) \leq y) \stackrel{h \text{ strängt växande (i symmetri inverterbar): bevarar rikthets riktning}}{=} P(X \leq h^{-1}(y)) = F_X(h^{-1}(y))$

• h är deriverbar, alltså också h^{-1} och det följer

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(h^{-1}(y)) \cdot (h^{-1})'(y) \quad (\text{kedjeregeln}), \text{ där } h \text{ och därmed } h^{-1} \text{ är växande} \Rightarrow (h^{-1})' > 0.$$

Navigation icons

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Exempel

Låt X vara en kontinuerlig slumpvariabel. För $a \neq 0$ och $b \in \mathbb{R}$

samt $Y = aX + b$ gäller $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$.

alltså $Y = h(X)$, där $h(x) = ax + b$ (monoton $\Leftrightarrow a \neq 0$) med invers $h^{-1}(y) = \frac{y-b}{a}$

$[y = ax + b \Leftrightarrow x = \frac{y-b}{a}]$ och $(h^{-1})'(y) = \frac{1}{a}$ ges enligt satsen: $f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \left|\frac{1}{a}\right|$.

Exempel

Om $X \sim \mathcal{N}(\mu, \sigma^2)$ och $Y = \frac{X-\mu}{\sigma}$ så är $Y \sim \mathcal{N}(0, 1)$:

Talen a, b från exemplet ovan motsvarar $\frac{1}{\sigma} > 0$ respektive $-\frac{\mu}{\sigma}$ här, så att $\frac{y-b}{a} = \sigma y + \mu$ och

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

Navigation icons

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810