Föreläsning 8 – integraler och statistik

Studenterna värmer upp genom att hitta på funktioner $f(x)$ vars integral över hela definitionsmängden har värdet 1 och vars primitiva funktion $F(x)$ växer från 0 till 1 över hela definitionsmängden…

Inledning – några nya begrepp

- Ett *utfallsrum* är *mängden* av alla möjliga utfall (möjliga resultat) vid ett slumpmässigt försök.

  Ex) Kön hos slumpmässigt vald person i klassen: $\Omega = \{kvinna, man, intersex\}$

  Ex) Ålder hos slumpmässigt vald person i klassen: $\Omega = [0, \infty[$

- En *stokastisk variabel* X är en *funktion* – ja just det en funktion – som är definierad i utfallsrummet Ω och ger de olika utfallen tillhörande värden, vanligen reella tal.

  Ex) Kön hos slumpmässigt vald person, *diskret* variabel X: $X: \{man, kvinna\} \to \{0, 1\}$

  Ex) Ålder hos slumpmässigt vald person, *kontinuerlig* variabel X: $X \in [0, \infty[$

- Notera att X (stora bokstaven X) är den stokastiska variabeln, medan x (lilla bokstaven x) har sparats för att kunna användas som ingående variabel i våra beskrivande matematiska modeller – t.ex. i en Riemann-integrerbar täthetsfunktion $f_X(x)$ som beskriver tätheten för olika värden x utifrån verkligheten via den stokastiska variabeln X.

Integraler och statistik – beräkningar för kontinuerliga stokastiska variabler

- Defitinion 1

  En Riemann-integrerbar funktion (från reella tal till reella tal) som uppfyller både

  1. $f_X(x) \geq 0$ *(sannolikheter är aldrig negativa)*
  2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$ *(totala sannolikheten är 100 %)*

  kallas täthetsfunktion (eng. density-function) $f_X(x)$ för den stokastiska variabeln X.

- Täthetsfunktionen $f_X(x)$ beskriver hur olika värden av X är fördelade – ju tätare runt vissa x-värden, desto högre funktionsvärde och integralen över hela utfallsrummet har alltid värdet 1 (= 100 %).

- Defitinion 5

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

kallas en fördelningsfunktion till den stokastiska variabeln X.

- Fördelningsfunktion $F_X(b)$ summerar sannolikheten fram till ett visst värde, i detta fall $x = b$ och därmed gäller att sannolikheten $P(a < X \leq b) = F_X(b) - F_X(a)$

- Sats 6

$$F_X{}'(x) = f_X(x)$$

Därmed är fördelningsfunktionen $F_X(x)$ en primitiv funktion till täthetsfunktionen $f_X(x)$

Därmed är sannolikheten

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

och den åskådliggörs som en yta under funktionskurvan.

- Sats 7

För varje fördelningsfunktion $F_X(x)$ gäller att:

  - $F_X(x) \to 1$ då $x \to \infty$
  - $F_X(x) \to 0$ då $x \to -\infty$
  - $F_X(x)$ är växande
  - $F_X(x)$ är åtminstone högerkontinuerlig för varje x (räcker så)

Notera att av oändligt många möjliga primitiva funktioner $F_X(x)$ till täthetsfunktionen $f_X(x)$ så är fördelningsfunktionen just den $F_X(x)$ som har en anpassad konstant C sådan att $F_X(x)$ stämmer enligt ovan.

- Definition 10

Låt $\alpha \in [0, 1]$ – i detta fall en sannolikhet.

Lösningen till ekvationen

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

alltså lösningen till

$$F(x) = \alpha$$

kallas då α-kvantilen för den stokastiska variabeln X och betecknas x_α.

- x_α (så kallade kvantiler) anger det x-värde som från vänster summerar just sannolikheten α och motsvarar i en integrationsgräns.

  ”Nollprocentskvatilen till vänster och hundraprocentskvantilen till höger”

- Definition 11

  Särskilt intressanta kvantiler är de så kallade kvartilerna är $x_{0.25}$ (kallad nedre kvartilen), $x_{0.50}$ (kallad medianen eller mellersta kvartilen) och $x_{0.75}$ (kallad övre kvartilen)

Exempeluppgift

Nedre kvartilen $x_{0.25}$ för täthetsfunktionen $f_X(x) = \begin{cases} 0 & \text{då } x < 0 \\ \frac{x}{2} & \text{då } 0 \leq x \leq 2 \\ 0 & \text{då } x > 2 \end{cases}$ ges av

$$\int_{-\infty}^{b} f_X(x)dx = \int_0^b \frac{x}{2}\, dx = 0.25 \quad \Leftrightarrow \quad \left[\frac{x^2}{4}\right]_0^b = 0.25 \quad \Leftrightarrow \quad b = x_{0.25} = 1$$

- Definition 14

  Väntevärdet – även kallat det förväntade värdet – är ett annat *lägesmått* och ges av

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

  I vardagen uppskattas vanligen väntevärdet med ett stickprovsmedelvärde.

  För en kontinuerlig sannolikhetsfördelning sammanfaller väntevärdet med *tyngdpunkten* av arean under täthetsfunktionen $f_X(x)$.

Exempeluppgift

Väntevärde för $f_X(x) = \begin{cases} 0 & \text{då } x < 0 \\ \frac{x}{2} & \text{då } 0 \leq x \leq 2 \\ 0 & \text{då } x > 2 \end{cases}$

$$\int_{-\infty}^{\infty} x\, f_X(x)dx = \int_0^2 x \cdot \frac{x}{2} dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{8}{6} = \frac{4}{3}$$

- Definition 18

  Låt $\mu = E(X)$ enligt ovan.

  Spridningsmåttet *variansen* $V(X)$ för den stokastiska variabeln X är talet som ges av:

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$$

- Väntevärdet av den stokastiska variabeln X^2 (se Appendix) ges av:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$$

  För variansen fås därmed:

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - 2\mu \underbrace{\int_{-\infty}^{\infty} x f_X(x)\,dx}_{=E(X)=\mu} + \mu^2 \underbrace{\int_{-\infty}^{\infty} f_X(x)\,dx}_{=1}$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - 2\mu^2 + \mu^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \mu^2$$

$$= E(X^2) - \big(E(X)\big)^2$$

  Sammanfattningsvis har vi:

- Sats 20

$$V(X) = E(X^2) - \big(E(X)\big)^2$$