SF1901 Sannolikhetsteori och statistik I

Jimmy Olsson

Föreläsning 5
10 november 2016

Idag

Väntevärde och varians (Kap. 5.1–5.3)

Flerdimensionella stokastiska variabler (Kap. 4.1–4.7)

Oberoende stokastiska variabler

Idag

Väntevärde och varians (Kap. 5.1–5.3)

Flerdimensionella stokastiska variabler (Kap. 4.1–4.7)

Oberoende stokastiska variabler

Väntevärde för stokastiska variabler

- *Väntevärdet* för en stokastisk variabel X definieras som

$$\mathbb{E}(X) = \begin{cases} \sum_k k\, p_X(k) & \text{om } X \text{ är diskret,} \\ \int_{-\infty}^{\infty} x\, f_X(x)\, dx & \text{om } X \text{ är kontinuerlig.} \end{cases}$$

- Väntevärdet kan tolkas som
  - ”masscentrum” för fördelningen, dvs. ett mått på fördelningens *läge*.
  - medelvärdet för variabelns utfall om försöket utförs ett stort antal gånger.
- Ibland betecknas väntevärdet även som μ_X.

Väntevärde för funktioner av stokastiska variabler

- Låt nu X vara en stokastisk variabel, g en reellvärd funktion och sätt $Y = g(X)$.
- Vi diskuterade hur täthetsfunktionen/sannolikhetsfunktionen för Y kan bestämmas i vissa fall.
- Dessutom gäller följande eleganta resultat för $\mathbb{E}(Y)$.

Sats

$$\mathbb{E}(Y) = \begin{cases} \displaystyle\sum_k g(k)\, p_X(k) & \textit{om } X \textit{ är diskret,} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & \textit{om } X \textit{ är kontinuerlig.} \end{cases}$$

Exempel: exponentialfördelning

- Låt $X \in \text{Exp}(\lambda)$. Beräkna
  (a) $\mathbb{E}(X)$,

  svar: $\dfrac{1}{\lambda}$

  (b) $\mathbb{E}(X^2)$.

  svar: $\dfrac{2}{\lambda^2}$

Varians och standardavvikelse för stokastiska variabler

- *Variansen* för en stokastisk variabel X definieras som

$$\mathbb{V}(X) = \mathbb{E}[(X - \mu_X)^2],$$

där $\mu_X = \mathbb{E}(X)$, dvs.

$$\mathbb{V}(X) = \begin{cases} \displaystyle\sum_k (k - \mu_X)^2 \, p_X(k) & \text{om } X \text{ är diskret,} \\ \displaystyle\int_{-\infty}^{\infty} (x - \mu_X)^2 \, f_X(x) \, dx & \text{om } X \text{ är kontinuerlig.} \end{cases}$$

- Variansen mäter hur mycket X avviker från $\mathbb{E}(X)$ i medeltal och ger sålunda ett mått på fördelningens *spridning*.
- Ibland betecknas variansen även som σ_X^2.
- Man visar enkelt att $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Standardavvikelse för stokastiska variabler

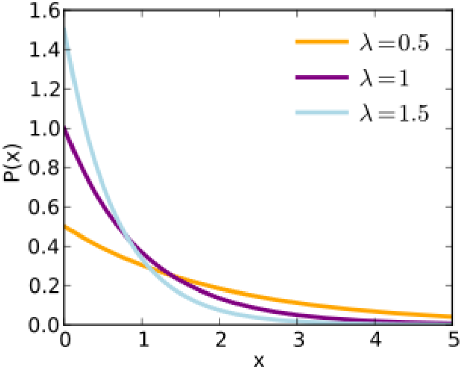
- *Standardavvikelsen* är nära besläktad med variansen och definieras som
$$\mathbb{D}(X) = \sqrt{\mathbb{V}(X)} = \sigma_X.$$
- En fördel med standardavvikelsen är att den har samma enhet som X själv.

Exempel: Exponentialfördelning

- Låt $X \in \text{Exp}(\lambda)$. Beräkna $\mathbb{V}(X)$ och $\mathbb{D}(X)$.

$$\text{svar: } \mathbb{V}(X) = \frac{1}{\lambda^2}, \quad \mathbb{D}(X) = \frac{1}{\lambda}\,(= \mathbb{E}(X)!)\,.$$

Exempel: exponentialfördelning (forts.)



Figur: Täthetsfunktioner för Exp(λ)-fördelningar med olika λ.

Idag

Väntevärde och varians (Kap. 5.1–5.3)

Flerdimensionella stokastiska variabler (Kap. 4.1–4.7)

Oberoende stokastiska variabler

Tvådimensionella s.v.

- Till att börja med betraktar vi det tvådimensionella fallet.

Definition

En *tvådimensionell stokastisk variabel* är en $\mathbb{R}^2$-*värd* funktion (X, Y) definierad på ett utfallsrum Ω.

- Med andra ord, (X, Y) associerar varje utfall i Ω med ett *talpar* i $\mathbb{R}^2$. Koordinaterna X och Y är endimensionella s.v.

Diskret tvådimensionell s.v.

- Om X och Y båda antar ett ändligt eller uppräkneligt oändligt antal värden $\{k_0, k_1, k_2, \ldots\}$ sägs den tvådimensionella s.v. (X, Y) vara *diskret*. Vi låter $\{k_0, k_1, k_2, \ldots\} = \{0, 1, 2, \ldots\}$ i det följande.

Definition

Funktionen

$$p_{X,Y}(j, k) = \mathbb{P}(X = j, Y = k), \quad j \in \{0, 1, 2, \ldots\}, k \in \{0, 1, 2, \ldots\},$$

kallas *sannolikhetsfunktionen* för den diskreta tvådimensionella s.v. (X, Y).

- I analogi med det endimensionella fallet gäller att $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} p_{X,Y}(j, k) = 1$.

Kontinuerlig tvådimensionell s.v.

- Analog med det endimensionella fallet är även följande definition.

Definition

Om det finns en icke-negativ funktion $f_{X,Y}(x,y)$ sådan att

$$\mathbb{P}((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$$

för alla A sägs den tvådimensionella s.v. (X,Y) vara *kontinuerlig*. Funktionen $f_{X,Y}(x,y)$ kallas *täthetsfunktionen för* (X,Y).

- I definitionen ovan betecknar $\iint_A$ en s.k. *dubbelintegral*, vilken beskriver *volymen* under grafen $A \ni (x,y) \mapsto f_{X,Y}(x,y)$. Jfr. det endimensionella fallet.

Exempel: likformig fördelning

- Antag att (X, Y) har *likformig fördelning* över $0 < x, y < 1$, dvs.
$$f_{X,Y}(x, y) = 1, \quad 0 < x, y < 1.$$
Bestäm $\mathbb{P}(X > Y)$.
$$\text{svar: } \frac{1}{2}$$

Kontinuerlig tvådimensionell s.v.

- Vi noterar att

$$1 = \mathbb{P}((X, Y) \in \mathbb{R}^2) = \iint_{\mathbb{R}^2} f_{X,Y}(x, y)\, dx\, dy,$$

  dvs. att totala volymen under grafen (ytan) utgör den totala sannolikhetsmassan $(= 1)$.

- Den *marginella täthetsfunktionen* f_X *för* X erhålls genom att man integrerar bort y-argumentet från $f_{X,Y}$ enligt

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy.$$

  Analogt,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx.$$

- Visa bild!

Tvådimensionella s.v.

- Vi är i allmänhet intresserade av att beräkna sannolikheten $\mathbb{P}((X, Y) \in A)$ att talparen hamnar i någon mängd $A \subset \mathbb{R}^2$, och här kommer följande till användning.

Definition

Funktionen

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y), \quad (x, y) \in \mathbb{R}^2,$$

kallas för *den simultana fördelningsfunktionen* för den tvådimensionella s.v. (X, Y).

Fler än två dimensioner

- Ovan definitioner utvidgas på ett naturligt sätt till fler än två dimensioner.
- En *n-dimensionell s.v.* $(X_1, \dots X_n)$ är således en $\mathbb{R}^n$-värd funktion definierad på ett utfallsrum Ω.
- Den simultana fördelningsfunktionen definieras i detta fall som

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n), \qquad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

och även den simultana sannolikhets- och täthetsfunktionen $p_{X_1,\dots,X_n}$ resp. $f_{X_1,\dots,X_n}$ definieras analogt med det tvådimensionella fallet.

Idag

Väntevärde och varians (Kap. 5.1–5.3)

Flerdimensionella stokastiska variabler (Kap. 4.1–4.7)

Oberoende stokastiska variabler

Oberoende stokastiska variabler

Definition

De s.v. X och Y kallas *oberoende* om

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \qquad (**)$$

för alla mängder A och B.

- Det visar sig vara tillräckligt att $(**)$ gäller för alla rektanglar:

Sats

De s.v. X och Y är oberoende om och endast om

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

för alla $(x,y) \in \mathbb{R}^2$.

Oberoende stokastiska variabler

- Det gäller vidare att

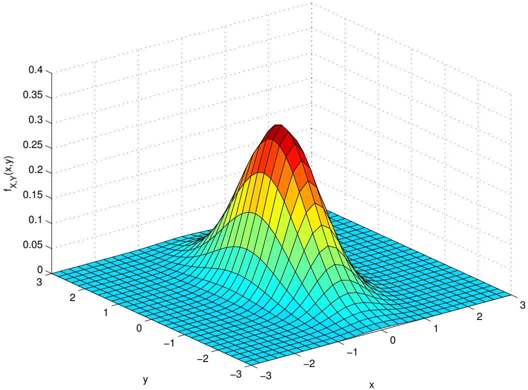
$$
\begin{aligned}
&F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \text{för alla } x, y \\
\Leftrightarrow &\begin{cases} p_{X,Y}(j,k) = p_X(j)p_Y(k) & \text{för alla } j, k \quad \text{(diskreta fallet)}, \\ f_{X,Y}(x,y) = f_X(x)f_Y(y) & \text{för alla } x, y \quad \text{(kontinuerliga fallet)}. \end{cases}
\end{aligned}
$$

- Ovan definition kan utan vidare utvidgas till fler än två s.v. enligt samma mönster som i Föreläsning 2 (oberoende mängder). T.ex. är de s.v. X, Y och Z oberoende om de är parvis oberoende och dessutom

$$F_{X,Y,Z}(x,y,z) = F_X(x)F_Y(y)F_Z(z)$$

för alla $(x,y,z) \in \mathbb{R}^3$.

Exempel: normalfördelade s.v.



Figur: Simultan täthetsfunktion $f_{X,Y}(x,y) = \exp\{-x^2/(2\sigma_x^2) - y^2/(2\sigma_y^2)\}/(2\pi\sigma_x\sigma_y)$ för två oberoende normalfördelade s.v. med $\sigma_x = 0.5$ och $\sigma_y = 1$.

Maximum av n oberoende s.v.

- Låt $X_1, X_2, \ldots, X_n$ vara oberoende s.v. med gemensam fördelnings- och täthetsfunktion F resp. f.
- Vi sätter $Z = \max(X_1, \ldots, X_n)$.
- Då gäller

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(X_1 \leq z, \ldots, X_n \leq z)$$
$$= \prod_{i=1}^{n} \mathbb{P}(X_i \leq z) = (F(z))^n.$$

- Kedjeregeln ger vidare

$$f_Z(z) = \frac{dF_Z}{dz}(z) = n(F(z))^{n-1} f(z).$$

Minimum av n oberoende s.v.

- Låt $X_1, X_2, \ldots, X_n$ vara som ovan, men sätt istället $Z = \min(X_1, \ldots, X_n)$.
- Genom att ta komplementet kan vi räkna som tidigare enligt

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) = 1 - \mathbb{P}(Z > z) \\ &= 1 - \mathbb{P}(X_1 > z, \ldots, X_n > z) = 1 - \prod_{i=1}^{n} \mathbb{P}(X_i > z) \\ &= 1 - \prod_{i=1}^{n} (1 - \mathbb{P}(X_i \leq z)) = 1 - (1 - F(z))^n. \end{aligned}$$

- Kedjeregeln ger så

$$f_Z(z) = \frac{dF_Z}{dz}(z) = n(1 - F(z))^{n-1} f(z).$$

Exempel: ”en kedja är inte starkare än sin svagaste länk”

- Påfrestningen (enhet kN) som en länk tål är en stokastisk variabel som är Rayleighfördelad med täthetsfunktionen

$$f_X(x) = \frac{1}{4} x e^{-x^2/8}, \quad x > 0.$$

Vad är sannolikheten att en kedja som består av 4 länkar brister om den utsätts för belastningen 0.5 kN?

$$\text{svar: } 1 - e^{-1/8} \approx 0.12$$

Nästa föreläsning

- Kovarians och korrelation,
- Summor av oberoende s.v.,
- Stora talens lag.