



MVE051/MSG810 - Statistik och diskret matematik (LP4 2022-2023)

Chalmers och Göteborgs universitet

Föreläsning 5 – Flerdimensionella stokastiska variabler

March 29, 2023

Dagens föreläsning

- Flerdimensionella stokastiska variabler
- Marginalfördelning
- Kovarians och korrelation

Tvådimensionella stokastiska variabler

En tvådimensionell stokastisk variabel (X, Y) är ett par av två stokastiska variabler X och Y som är definierade på samma utfallsrum. Ofta är X och Y inte oberoende.

En tvådimensionell stokastisk variabel sägs vara

- diskret om X och Y båda är diskreta, och
- kontinuerlig om både X och Y är kontinuerliga.

Diskreta tvådimensionella stokastiska variabler

Antag att $X, Y: \Omega \rightarrow \mathbb{R}$ är diskreta slumpvariabler.

Gemensam sannolikhetsfunktion

Sannolikhetsfunktionen f_{XY} till den tvådimensionella stokastiska variabeln (X, Y) ges av

$$f_{XY} = P(X = x, Y = y), \quad x, y \in \mathbb{R}$$

Precis som tidigare är $f_{XY}: \mathbb{R}^2 \rightarrow \mathbb{R}$ en sannolikhetsfunktion om och endast om

- $f_{XY}(x, y) \geq 0$ för alla $(x, y) \in \mathbb{R}$, och
- $\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} f_{XY}(x, y) = 1$.

Gemensam fördelningsfunktion

Fördelningsfunktionen F_{XY} till den tvådimensionella stokastiska variabeln (X, Y) ges av

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{i \leq x} \sum_{j \leq y} P(X = i, Y = j) = \sum_{i \leq x} \sum_{j \leq y} f(i, j).$$

Diskreta tvådimensionella stokastiska variabler

Exempel

Låt X vara mängd regn som förväntas under dagen (enligt morgonens väderprognos, mätt i cm), och Y antal studenter (av totalt 4) som tar med ett paraply. Anta att alla sannolikheter $f_{XY}(x, y)$ ges av tabellen:

	0	1	2	3	4
0cm	0.4	0.1	0	0	0
1cm	0.2	0.05	0.04	0.01	0
2cm	0.05	0.05	0.02	0.02	0.01
3cm	0.01	0.01	0.01	0.01	0.01

Då blir

$$\begin{aligned}F(1, 2) &= P(X \leq 1, Y \leq 2) \\&= P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 0, Y = 2) \\&\quad + P(X = 1, Y = 0) + P(X = 1, Y = 1) + P(X = 1, Y = 2) \\&= 0.04 + 0.1 + 0 + 0.2 + 0.05 + 0.04 = 0.43\end{aligned}$$

Diskreta tvådimensionella stokastiska variabler

Antag att $X, Y: \Omega \rightarrow \mathbb{R}$ är diskreta slumpvariabler med gemensam sannolikhetsfunktion f_{XY} .

Marginella sannolikhetsfunktioner

De marginella sannolikhetsfunktionerna för X och Y ges av

$$f_X(x) = P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y) = \sum_{y \in Y(\Omega)} f(x, y)$$

$$f_Y(y) = P(Y = y) = \sum_{x \in X(\Omega)} P(X = x, Y = y) = \sum_{x \in X(\Omega)} f(x, y)$$

Diskreta tvådimensionella stokastiska variabler

Exempel

Låt X vara mängd regn som förväntas under dagen (enligt morgonens väderprognos, mätt i cm), och Y antal studenter (av totalt 4) som tar med ett paraply. Anta att alla sannolikheter $f_{XY}(x, y)$ ges av tabellen:

	0	1	2	3	4	f_X
0cm	0.4	0.1	0	0	0	0.5
1cm	0.2	0.05	0.04	0.01	0	0.3
2cm	0.05	0.05	0.02	0.02	0.01	0.15
3cm	0.01	0.01	0.01	0.01	0.01	0.05
f_Y	0.66	0.21	0.07	0.04	0.02	1

Kontinuerliga tvådimensionella stokastiska variabler

Antag att $X, Y: \Omega \rightarrow \mathbb{R}$ är kontinuerliga slumpvariabler.

Gemensam sannolikhetsfunktion

En funktion f_{XY} som uppfyller

- $f_{XY}(x, y) \geq 0$ för alla $x, y \in \mathbb{R}$,
- $\iint f_{XY}(x, y) = 1$
- $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dx dy$

kallas för en täthetsfunktion till (X, Y) .

Marginella täthetsfunktioner

$$f_X(x) = \int f(x, y) dy$$

$$f_Y(y) = \int f(x, y) dx$$

Kontinuerliga tvådimensionella stokastiska variabler

Exempel

Låt (X, Y) vara ett par stokastiska variabler med

$$f_{XY}(x, y) = c(1 + x + e^{-y}), \quad x \in [0, 1], y \geq 0.$$

1. Vad är c ?
2. Beräkna $P(X \geq 1/2, Y \geq 3)$
3. Beräkna f_X och f_Y
4. Beräkna $P(X \leq 1/4)$
5. Beräkna $P(Y \geq 1)$

Oberoende stokastiska variabler

Oberoende händelser

Två händelser $A, B \subseteq \Omega$ är oberoende om $P(A \cap B) = P(A)P(B)$.

Oberoende stokastiska variabler

Två stokastiska variabler $X, Y: \Omega \rightarrow \mathbb{R}$ är oberoende om

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

Om X och Y är oberoende så är

$$\begin{aligned}P(X \leq x, Y \leq y) &= \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s, t) \, ds \, dt \\&= \int_{-\infty}^x \int_{-\infty}^y f_X(s) f_Y(t) \, ds \, dt \\&= \int_{-\infty}^x f_X(s) \, ds \int_{-\infty}^y f_Y(t) \, dt = P(X \leq x)P(Y \leq y)\end{aligned}$$

och mer generellt

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

dvs.

$$P(X^{-1}(A) \cap Y^{-1}(B)) = P(X^{-1}(A))P(Y^{-1}(B))$$

Oberoende stokastiska variabler

Exempel (diskret)

Mängden regn X och antalet studenter med paraply Y är inte oberoende eftersom att

$$P(X = 0, Y = 2) = 0 \neq 0.07 \cdot 0.05 = P(X = 0.05)P(Y = 0)$$

Exempel (kontinuerligt)

Låt $f_{XY}(x, y) = c(1 + x + e^{-y})$, där $x, y \in [0, 1]$. Är X och Y oberoende?

Väntevärde

Låt $X, Y: \Omega \rightarrow \mathbb{R}$ ha gemensam täthetsfunktion f_{XY} , och låt $h: \mathbb{R}^2 \rightarrow \mathbb{R}$ vara en funktion.

Om (X, Y) är diskret definieras

$$\mathbb{E}[h(X, Y)] = \sum \sum h(x, y) f_{XY}(x, y).$$

Om (X, Y) är kontinuerlig definieras

$$\mathbb{E}[h(X, Y)] = \iint h(x, y) f_{XY}(x, y) dx dy$$

I båda fallen krävs att $\mathbb{E}[|h(X, Y)|] < \infty$.

Marginella väntevärden

Om vi väljer $h(x, y) = x$ får vi

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} x f_{XY}(x, y) = \sum_{x \in X(\Omega)} x \sum_{y \in Y(\Omega)} f_{XY}(x, y) \\ &= \sum_{x \in X(\Omega)} x f_X(x)\end{aligned}$$

dvs definitionen är konsistent med vår tidigare definition av väntevärden.

Marginella väntevärden

Precis som i det endimensionella fallet finns ett antal räkneregler man kan använda.

Räkneregler för väntevärden

1. $\mathbb{E}[ch(X, Y)] = c\mathbb{E}[h(X, Y)]$
2. $\mathbb{E}[h(X, Y) + g(X, Y)] = \mathbb{E}[h(X, Y)] + \mathbb{E}[g(X, Y)]$

Om X och Y är oberoende gäller (som tidigare) att

3. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Kovarians och korrelation

Låt $X, Y: \Omega \rightarrow \mathbb{R}$ vara två slumpvariabler med väntevärden μ_X och μ_Y .

Kovarians

Covariansen $\text{Cov}(X, Y)$ mellan X och Y är ett mått på hur mycket X och Y samvarierar.

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Specialfall: $\text{Cov}(X, X) = \text{Var}(X)$, och om X och Y är oberoende gäller att $\text{Cov}(X, Y) = 0$.

Räkneregler

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2. $\text{Cov}(cX, Y) = c \text{Cov}(X, Y) = \text{Cov}(X, cY)$
3. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

Kovarians och korrelation

Kovariansen mäter samband mellan X och Y , mer hur stor den är beror även på skalan på X och Y .

Korrelation

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Man kan visa att korrelationen alltid är ett tal mellan -1 och 1
- Om X eller Y är konstant så är motsvarande variation lika med 0 , och då finns inte korrelationen.
- $\rho_{XY} = \pm 1$ bara om $Y = aX + b$ för några tal $a \neq 0$ och b .
- $\rho_{XY} = 0$ används ofta som ett mått på hur starkt linjärt samband det finns mellan X och Y (speciellt inom statistik).
- X och Y oberoende $\rightarrow \text{Cov}(X, Y) = 0$, men motsatsen är inte sann!

Kovarians och korrelation

Exempel

Låt X vara mängd regn som förväntas under dagen (enligt morgonens väderprognos, mätt i cm), och Y antal studenter (av totalt 4) som tar med ett paraply. Anta att alla sannolikheter $f_{XY}(x, y)$ ges av tabellen:

	0	1	2	3	4	f_X
0cm	0.4	0.1	0	0	0	0.5
1cm	0.2	0.05	0.04	0.01	0	0.3
2cm	0.05	0.05	0.02	0.02	0.01	0.15
3cm	0.01	0.01	0.01	0.01	0.01	0.05
f_Y	0.66	0.21	0.07	0.04	0.02	1

Då blir

$$\mathbb{E}[X] = 0 \cdot 0.5 + 1 \cdot 0.5 + 2 \cdot 0.15 + 3 \cdot 0.05 = 0.75$$

$$\mathbb{E}[Y] = 0 \cdot 0.66 + 1 \cdot 0.21 + 2 \cdot 0.07 + 3 \cdot 0.04 + 4 \cdot 0.02 = 0.55$$

$$\text{Var}(X) = 0^2 \cdot 0.5 + 1^2 \cdot 0.3 + 2^2 \cdot 0.15 + 3^2 \cdot 0.05 - 0.75^2 = 0.7875$$

$$\text{Var}(Y) = 0^2 \cdot 0.66 + 1^2 \cdot 0.21 + 2^2 \cdot 0.07 + 3^2 \cdot 0.04 + 4^2 \cdot 0.02 - 0.55^2 = 0.8675$$

$$\mathbb{E}[XY] = 0 \cdot 0 \cdot 0.4 + 0 \cdot 1 \cdot 0.1 + \dots + 3 \cdot 4 \cdot 0.01 = 0.84$$

$$\text{Cov}(X, Y) = 0.84 - 0.75 \cdot 0.55 = 0.4275$$

och

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{0.4275}{\sqrt{0.7875} \cdot \sqrt{0.8675}} = 0.517221$$

Kovarians och korrelation

Exempel

Låt X vara likformigt fördelad på $\{-1, 1\}$, och låt $Y = 1$ om $X = -1$ och likformigt fördelad på $\{-1, 1\}$ om $X = 1$.

	-1	0	1
-1	0	0.5	0
1	0.25	0	0.25

X och Y inte oberoende, men

$$\mathbb{E}[XY] = \mathbb{E}[X] = \mathbb{E}[Y] = 0$$

så $\text{Cov}(X, Y) = \rho_{XY} = 0$.

Betingade fördelningar

Om A och B är händelser och $P(B) > 0$ så definierade vi

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Betingad täthetsfunktion

Den betingade sannolikhetsfunktionen/täthetsfunktionen för den stokastiska variabeln X givet att $Y = y$ ges av

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

På samma sätt ges den betingade sannolikhetsfunktionen/täthetsfunktionen för den stokastiska variabeln Y givet att $X = x$ ges av

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Obs! Om X och Y är oberoende får vi

$$f_{X|y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x).$$

Betingade fördelningar

Exempel

Låt $f(x, y) = c(1 + x + e^{-y})$ där $x, y \in [0, 1]$. Beräkna $f|_{X|y=0.5}$.