Föreläsning 6

Simultana fördelningar

Definition Låt X, Y vara två diskreta stokastiska variabler.
(a)

$$p(x,y) = P(X = x, Y = y) \quad \text{för alla möjliga värden } (x,y) \text{ på } (X,Y)$$

kallas *simultan (gemensam) sannolikhetsfunktion* för (X,Y).

(b)

$$p_X(x) = P(X = x) = \sum_{y:p(x,y)>0} p(x,y)$$

för alla möjliga värden x på X och

$$p_Y(y) = P(Y = y) = \sum_{x:p(x,y)>0} p(x,y)$$

för alla möjliga värden y på Y kallas de *marginella sannolikhetsfunktionerna.*

Anmärkning Den marginella sannolikhetsfunktionen $p_X(x) = P(X = x)$ är samtidigt sannolikhetsfunktion av X om man inte tar hänsyn till Y. Vidare är den marginella sannolikhetsfunktionen $p_Y(y) = P(Y = y)$ samtidigt sannolikhetsfunktion av Y om man inte tar hänsyn till X.

Exempel Det finns 3 röda , 4 vita och 5 blåa kulor i en urna. Vi väljer 3 kulor. Låt

$$X : \# \text{ röda kulor i urvalet}$$

$$Y : \# \text{ vita kulor i urvalet}.$$

Bestäm den simultana sannolikhetsfunktionen för (X,Y). Bestäm också de marginella sannolikhetsfunktionerna. Beräkna $P(X + Y = 2)$.

Svar: Klassisk sannolikhet och multiplikationsprincipen ger

$$\begin{aligned}
p(0,0) &= \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220} \\
p(0,1) &= \frac{\binom{4}{1}\cdot\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220} \\
&\vdots \\
p(1,1) &= \frac{\binom{3}{1}\cdot\binom{4}{1}\cdot\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220} \\
&\vdots \quad .
\end{aligned}$$

$$P(X = i, Y = j)$$

i^j	0	1	2	3	$P(X = i)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$P(Y = j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

$$P(X + Y = 2) = p(0, 2) + p(1, 1) + p(2, 0) = \frac{30}{220} + \frac{60}{220} + \frac{15}{220} = \frac{105}{220}\,.$$

Definition Låt X, Y vara två kontinuerliga stokastiska variabler. (a) En funktion $f : \mathbb{R}^2 \longrightarrow [0, \infty)$ kallas *simultan (gemensam) täthetsfunktion* av (X, Y) om

$$P((X, Y) \in C) = \iint\limits_{(x,y)\in C} f(x, y)\, dx\, dy$$

för varje (Borel) mängd $C \subseteq \mathbb{R}^2$.

(b)

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy\,, \quad x \in \mathbb{R},$$

och

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx\,, \quad y \in \mathbb{R},$$

kallas de *marginella täthetsfunktionerna.*

Anmärkningar (1) Den marginella täthetsfunktionen f_X är samtidigt täthetsfunktionen av slumpvariabeln X om man inte tar hänsyn till Y. På samma sätt är den marginella täthetsfunktionen f_Y samtidigt täthetsfunktionen av slumpvariabeln Y om man inte tar hänsyn till X.

(2) Det gäller att

$$\begin{aligned} P(X \in A) &= \int_{x\in A} \int_{y=-\infty}^{\infty} f(x, y)\, dy\, dx \\ &= \int_{x\in A} f_X(x)\, dx\,, \quad A \subseteq \mathbb{R}\,. \end{aligned}$$

Exempel Låt

$$f(x,y) = \begin{cases} 2\,e^{-x}\,e^{-2y} & \text{om} \quad 0 < x < \infty, \quad 0 < y < \infty \\ 0 & \text{annars} \end{cases}$$

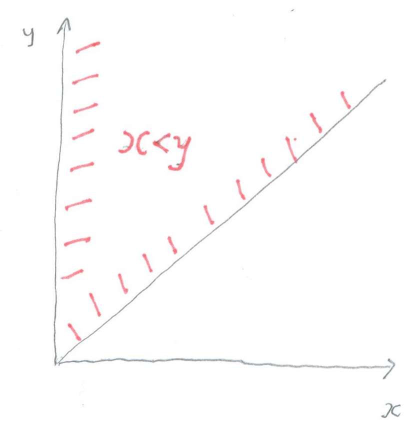
vara den simultana täthetsfunktionen av (X, Y). Beräkna (a) $P(X > 1, Y < 1)$,
(b) $P(X < Y)$ och
(c) $P(X < a)$.

Svar: (a)

$$\begin{aligned} P(X > 1, Y < 1) &= \int_{x=1}^{\infty} \int_{y=0}^{1} 2\,e^{-x}\,e^{-2y}\,dy\,dx \\ &= \int_{y=0}^{1} 2\,e^{-2y}\,dy \cdot \int_{x=1}^{\infty} e^{-x}\,dx \\ &= \left(1 - e^{2\cdot 1}\right) \cdot e^{-1} = e^{-1} - e^{-3}\,. \end{aligned}$$

(b)

$$\begin{aligned} P(X < Y) &= \iint\limits_{\{(x,y):x<y\}} 2\,e^{-x}\,e^{-2y}\,dy\,dx \\ &= \int_{y=0}^{\infty} \int_{x=0}^{y} 2\,e^{-x}\,e^{-2y}\,dx\,dy \\ &= \int_{y=0}^{\infty} 2\,e^{-2y} \left(\int_{x=0}^{y} e^{-x}\,dx \right) dy \\ &= \int_{y=0}^{\infty} 2\,e^{-2y}\,\left(1 - e^{-y}\right)\,dy = \cdots = \frac{1}{3}\,. \end{aligned}$$



(c)

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \\ &= \int_{0}^{\infty} 2\,e^{-x}\,e^{-2y}\,dy = e^{-x}\,, \quad x \geq 0, \end{aligned}$$

$$f_X(x) = 0\,, \quad x < 0.$$

Detta medför att $P(X < a) = 0$ om $a \leq 0$ och

$$P(X < a) = \int_0^a f_X(x)\,dx = 1 - e^{-a} \quad \text{om } a > 0.$$

Oberoende stokastiska variabler

Definition Låt X, Y vara två stokastiska variabler. Variablerna X och Y kallas *oberoende* om

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

för vilka som helst två (Borel) mängder $A, B \subseteq \mathbb{R}$.

Sats (a) Låt X, Y vara två diskreta stokastiska variabler. Variablerna X och Y är oberoende om och endast om

$$P(X = i, Y = j) = P(X = i) \cdot P(Y = j)$$

för alla par (i, j) som (X, Y) kan anta. Detta skrivas också som

$$p(i, j) = p_X(i) \cdot p_Y(j)$$

för alla par (i, j).

(b) Låt X, Y vara två kontinuerliga stokastiska variabler. Variablerna X och Y är oberoende om och endast om

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

för (nästan) alla $(x, y) \in \mathbb{R}^2$.

Exempel Låt

$$f(x, y) = \begin{cases} 2\,e^{-x}\,e^{-2y} & \text{om} \quad 0 < x < \infty, \quad 0 < y < \infty \\ 0 & \text{annars} \end{cases}$$

vara den simultana täthetsfunktionen av (X, Y). Är X och Y oberoende?

Svar: Vi har redan beräknat

$$f_X(x) = \begin{cases} e^{-x} & \text{om } x \geq 0 \\ 0 & \text{annars} \end{cases}.$$

Vidare gäller

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\,dx \\ &= \int_0^{\infty} 2\,e^{-x}\,e^{-2y}\,dx = 2e^{-2y}\,, \quad y \geq 0, \\ f_Y(y) &= 0\,, \quad y < 0. \end{aligned}$$

Alltså är X och Y oberoende.

Exempel Två kollegor vill träffas på ett café. Antag att båda kommer oberoende av varandra likformigt fördelad mellan kl.12 och kl.13. Bestäm sannolikheten att den som kommer först behöver vänta längre an 10 minuter.

Svar: Låt

$$X : \text{ankomsttid av den 1:a kollegan, i minuter efter kl. 12}$$

och

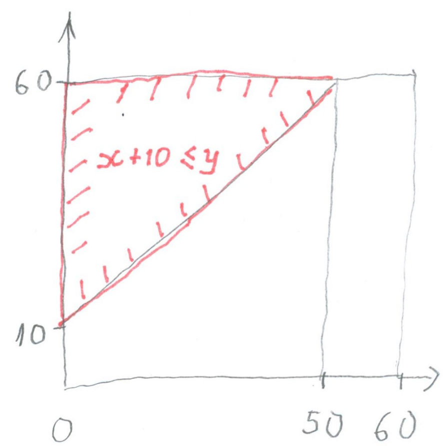
$$Y : \text{ankomsttid av den 2:a kollegan, i minuter efter kl. 12.}$$

Oberoendet av X och Y medför att

$$f(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} \frac{1}{60} \cdot \frac{1}{60} & \text{om} \quad 0 \le x \le 60, \quad 0 \le y \le 60 \\ 0 & \text{annars} \end{cases} .$$

Sökt är $P(\{X + 10 \le Y\} \cup \{Y + 10 \le X\})$. Vi får

$$\begin{aligned} &P(\{X + 10 \le Y\} \cup \{Y + 10 \le X\}) = 2P(X + 10 \le Y) \\ &= 2 \iint_{\substack{\{(x,y): x+10 \le y \\ 0 \le x, y \le 60\}}} \frac{1}{60} \cdot \frac{1}{60} \, dx \, dy \\ &= 2 \int_{y=10}^{60} \int_{x=0}^{y-10} \frac{1}{60} \cdot \frac{1}{60} \, dx \, dy = \cdots = \frac{25}{36} \left(> \frac{2}{3} \right) . \end{aligned}$$



Exempel / Lektionsuppgifter

(1) Antag att

$$f(x,y) = \begin{cases} c\,(x+y)e^{-(x+y)} & 0 < x < \infty \quad 0 < y < \infty \\ 0 & \text{annars} \end{cases} .$$

(a) Bestäm c så att f blir den simultana täthetsfunktionen av ett par (X,Y) stokastiska variabler.

(b) Är X och Y oberoende?

(c) Beräkna sannolikheten $P(X > 2|Y \leq 1)$.

(2) Variablerna X och Y har den simultana täthetsfunktionen

$$f(x,y) = 6x(1-x)$$

för alla (x,y) sådana att $0 \leq x \leq 1$ och $0 \leq y \leq 1$, utanför denna kvadrat är $f(x,y) = 0$.

(a) Bestäm $f_X(x)$ och $f_Y(y)$. Är X och Y oberoende?

(b) Beräkna $E[X-Y]$ och $Var(X-Y)$.

(c) Beräkna sannolikheten att $Y - X > 1/2$.

(3) Variablerna X och Y har den simultana täthetsfunktionen

$$f(x,y) = k(x-y)$$

för alla (x,y) sådana att $0 \leq y \leq x \leq 1$, annars $f(x,y) = 0$.

(a) Bestäm k.

(b) Bestäm $P(X > Y^2)$.

(4) Antag att

$$f(x,y) = \begin{cases} e^{-y} & 0 \leq x \leq y < \infty \\ 0 & \text{otherwise} \end{cases} .$$

Visa att f är den simultana täthetsfunktionen av ett par (X,Y) stokastiska variabler. Beräkna $P(X+Y \geq 1)$. Är X och Y oberoende?

Svar:

(1) (a)

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy &= c\int_0^{\infty}\int_0^{\infty}(x+y)e^{-(x+y)}\,dx\,dy \\
&= c\int_0^{\infty}\int_0^{\infty} x\,e^{-(x+y)}\,dx\,dy + c\int_0^{\infty}\int_0^{\infty} y\,e^{-(x+y)}\,dx\,dy \\
&= c\int_0^{\infty} x\,e^{-x}\int_0^{\infty} e^{-y}\,dy\,dx + c\int_0^{\infty} y\,e^{-y}\int_0^{\infty} e^{-x}\,dx\,dy \\
&= 2c\int_0^{\infty} x\,e^{-x}\,dx \\
&= 2c\,,
\end{aligned}$$

som medför att $c = \frac{1}{2}$.

(b) Det gäller att

$$
\begin{aligned}
f_X(x) &= \frac{1}{2}\int_0^{\infty}(x+y)e^{-(x+y)}\,dy \\
&= \frac{1}{2}\int_0^{\infty} x\,e^{-(x+y)}\,dy + \frac{1}{2}\int_0^{\infty} y\,e^{-(x+y)}\,dy \\
&= \frac{1}{2}\,x\,e^{-x}\int_0^{\infty} e^{-y}\,dy + \frac{1}{2}\,e^{-x}\int_0^{\infty} y\,e^{-y}\,dy \\
&= \frac{1}{2}(x+1)\,e^{-x}\,, \quad x > 0.
\end{aligned}
$$

På samma sätt, $f_Y(y) = \frac{1}{2}(y+1)\,e^{-y}$, $y > 0$. Eftersom $f(x,y) = f_X(x)f_Y(y)$ för alla $x, y > 0$ inte gäller, slumpvariablerna X och Y inte är oberoende.

(c)

$$
\begin{aligned}
P(X > 2|Y \leq 1) &= \frac{P(X > 2, Y \leq 1)}{P(Y \leq 1)} \\
&= \frac{\frac{1}{2}\int_2^{\infty}\int_0^1 (x+y)e^{-(x+y)}\,dy\,dx}{\frac{1}{2}\int_0^1 (y+1)\,e^{-y}\,dy} \\
&= \frac{\frac{1}{2}\int_2^{\infty} x\,e^{-x}\,dx \int_0^1 e^{-y}\,dy + \frac{1}{2}\int_0^1 y\,e^{-y}\,dy \int_2^{\infty} e^{-x}\,dx}{-\frac{1}{2}y\,e^{-y} - e^{-y}|_0^1} \\
&= \frac{\frac{1}{2}(-x\,e^{-x} - e^{-x}|_2^{\infty})(1-e^{-1}) + \frac{1}{2}(-y\,e^{-y} - e^{-y}|_0^1)e^{-2}}{1 - \frac{3}{2}e^{-1}} \\
&= \frac{3e^{-2}(1-e^{-1})}{2-3e^{-1}} + \frac{(1-2e^{-1})e^{-2}}{2-3e^{-1}} \\
&= \frac{4e^{-2} - 5e^{-3}}{2-3e^{-1}}\,.
\end{aligned}
$$

(2) (a) $f_X(x) = 6x(1-x)$, $0 \leq x \leq 1$, och $f_Y(y) = 1$, $0 \leq y \leq 1$. X och Y är oberoende.

(b)

$$
E[X] = \int_0^1 x \cdot 6x(1-x)\,dx = \left[2x^3 - \frac{3}{2}x^4\right]_0^1 = \frac{1}{2}
$$

$$
E[Y] = \int_0^1 y \cdot 1\,dy = \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2}
$$

$$
E[X - Y] = E[X] - E[Y] = 0\,.
$$

$$
E[X^2] = \int_0^1 x^2 \cdot 6x(1-x)\,dx = \frac{3}{10} \qquad E[Y^2] = \int_0^1 y^2 \cdot 1\,dy = \frac{1}{3}
$$

$$
Var(X-Y) = Var(X) + Var(-Y) = Var(X) + Var(Y)
$$

$$
= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 = \frac{2}{15}\,.
$$

(c)

$$\begin{aligned} P\left(Y - X > \frac{1}{2}\right) &= P\left(Y > \frac{1}{2} + X\right) \\ &= \int_0^{\frac{1}{2}} \left(\int_{\frac{1}{2}+x}^{1} f(x,y)\, dy \right) dx \\ &= \int_0^{\frac{1}{2}} \left(\int_{\frac{1}{2}+x}^{1} 6x(1-x)\, dy \right) dx \\ &= \int_0^{\frac{1}{2}} \left(\frac{1}{2} - x \right) 6x(1-x)\, dx \\ &= \frac{3}{32}. \end{aligned}$$

(3) (a) Vi har att

$$\int_0^1 \int_0^x k(x-y)\, dy\, dx = \int_0^1 k \left[-\frac{(x-y)^2}{2} \right]_0^x dx = k \int_0^1 \frac{x^2}{2}\, dx = \frac{k}{6}.$$

Så $k = 6$.

(b) $0 \le Y \le 1$ medför att $Y > Y^2$ och

$$P(X > Y^2) \ge P(X > Y).$$

$$P(X > Y^2) = 1$$

fås från $P(X > Y) = 1$.

(4)

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dy dx = \int_0^{\infty} \int_x^{\infty} e^{-y}\, dy\, dx = \int_0^{\infty} e^{-x}\, dx = 1$$

Dessutom, $f \ge 0$. Således är f en täthetsfunktion. Det gäller att

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_x^{\infty} e^{-y}\, dy = e^{-x}, \quad x > 0.$$

$$f_Y(y) = \int_0^{\infty} f(x,y)\, dx = \int_0^{y} e^{-y}\, dx = ye^{-y}, \quad y > 0.$$

Eftersom $f(x,y) = f_X(x) f_Y(y)$ inte gäller för alla $x, y > 0$, de stokastiska variablerna X och Y inte är oberoende. För $P(X + Y \ge 1)$ rita integrationsområdet.

$$\begin{aligned} P(X + Y \ge 1) &= \int_0^{\frac{1}{2}} \int_{1-x}^{\infty} e^{-y}\, dy dx + \int_{\frac{1}{2}}^{\infty} \int_x^{\infty} e^{-y}\, dy dx \\ &= \int_0^{\frac{1}{2}} e^{-(1-x)}\, dx + \int_{\frac{1}{2}}^{\infty} e^{-x}\, dx \\ &= e^{-\frac{1}{2}} - e^{-1} + e^{-\frac{1}{2}} = 2e^{-\frac{1}{2}} - e^{-1}. \end{aligned}$$

Alternativt visar man

$$\begin{aligned} P(X+Y \geq 1) &= 1 - P(X+Y \leq 1) \\ &= 1 - \int_0^{\frac{1}{2}} \int_x^{1-x} e^{-y} \, dy dx \\ &= 1 - \int_0^{\frac{1}{2}} \left(e^{-x} - e^{-(1-x)} \right) \, dx \\ &= 1 - \left(1 - 2e^{-\frac{1}{2}} + e^{-1} \right) \\ &= 2e^{-\frac{1}{2}} - e^{-1} \, . \end{aligned}$$