

Sannolikhet och statistik IX

Johan Jonasson

April 2019

Flerdimensionella fördelningar

Om X och Y är två sv def på samma utfallsrum, kallar vi paret (X, Y) för en tvådimensionell stokastisk variabel. *Den bivariata fördelningsfunktionen* (joint cdf) för (X, Y) ges av

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

Vi har

$$F(x, \infty) = \mathbb{P}(X \leq x, Y < \infty) = \mathbb{P}(X \leq x) = F_X(x)$$

och, förstås, $F(\infty, y) = F_Y(y)$.

Diskreta tvådim sv. Om X och Y är diskreta säger man att (X, Y) är diskret. Obs att

$$V_X \times V_Y = \{(x_j, y_k) : x_j \in V_X, y_k \in V_Y\} \supseteq V_{X,Y}.$$

Vi antar ändå att $V_{X,Y}$ är $V_X \times V_Y$ (det tillkommer ev ett antal värden som (X, Y) antar med sannol 0).

Den bivariata frekvensfunktionen (joint pmf) för (X, Y) ges av

$$p(x_j, y_k) = \mathbb{P}(X = x_j, Y = y_k), (x_j, y_k) \in V_X \times V_Y.$$

Direkt observation:

$$p_X(x_j) = \mathbb{P}(X = x_j, Y \in V_Y) = \sum_{y_k \in V_Y} p(x_j, y_k)$$

och $p_Y(y_k) = \sum_{x_j \in V_X} p(x_j, y_k)$.

Exempel: Antag att (X, Y) antar vart och ett alla par av heltal (x, y) s.a. $1 \leq x \leq y \leq 4$ med samma sannolikhet.

Det finns $4 + 3 + 2 + 1 = 10$ värden, så

$$p(x, y) = \frac{1}{10}, 1 \leq x \leq y \leq 4.$$

$$p_X(x) = \sum_{y=x}^4 \frac{1}{10} = \frac{5-x}{10}, \quad x = 1, 2, 3, 4.$$

$$p_Y(y) = \sum_{x=1}^y \frac{1}{10} = \frac{y}{10}, \quad y = 1, 2, 3, 4.$$

□

Exempel. Slå två tärningar. Låt X vara den första tärningens utfall och Y vara summan. Bestäm $p(x, y)$.

Paret (x, y) är ett möjl värde om $x \in \{1, 2, 3, 4, 5, 6\}$ och $y \in \{x+1, \dots, x+6\}$ och kan isf fås på exakt ett sätt. Alltså

$$p(x, y) = \frac{1}{36}, \quad 1 \leq x \leq 6, \quad x+1 \leq y \leq x+6.$$

□

Exempel: Låt $X \sim Poi(\lambda)$ och sedan givet $X = x$, låt $Y \sim Bin(x, p)$.

Vi får

$$p(x, y) = e^{-\lambda} \frac{\lambda^x}{x!} \binom{x}{y} p^y (1-p)^{x-y}, \quad 0 \leq y \leq x < \infty.$$

Marginalförd för Y blir

$$\begin{aligned} p_Y(y) &= \sum_{x=y}^{\infty} p(x, y) \\ &= e^{-\lambda} \sum_{x=y}^{\infty} \frac{\lambda^x}{x!} \binom{x}{y} p^y (1-p)^{x-y} \\ &= \frac{e^{-\lambda} \lambda^y p^y}{y!} \sum_{x=y}^{\infty} \frac{(\lambda(1-p))^{x-y}}{(x-y)!} \\ &= \frac{e^{-\lambda} (\lambda p)^y}{y!} e^{\lambda(1-p)} \\ &= e^{-\lambda p} \frac{(\lambda p)^y}{y!} \end{aligned}$$

M.a.o. $Y \sim Poi(\lambda p)$.

Konsekvens: Om man i en Poi-process med in λ räknar med impulserna med sannol p , får man en Poi-process med int λp . (S.k. uttunnad Poi-process.)

Sats: X, Y diskreta och oberoende $\Leftrightarrow p(x_j, y_k) = p_X(x_j)p_Y(y_k)$ för alla $x_j \in V_X, y_k \in V_Y$.

Bevis: Om X, Y ober gäller

$$\begin{aligned} p(x_j, y_k) &= \mathbb{P}(X = x_j, Y = y_k) \\ &= \mathbb{P}(X = x_j)\mathbb{P}(Y = y_k) \\ &= p_X(x_j)p_Y(y_k). \end{aligned}$$

Å andra sidan om p kan faktoriseras får vi

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \sum_{x_j \in A} \sum_{y_k \in B} p(x_j, y_k) \\ &= \sum_{x_j \in A} p_X(x_j) \sum_{y_k \in B} p_Y(y_k) \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B).\end{aligned}$$

□

Kont tvådim sv. Paret (X, Y) sägs vara en kontinuerlig tvådim sv om det finns en täthetsfunktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ sådan att

$$\mathbb{P}((X, Y) \in B) = \iint_B f(x, y) dx dy$$

för alla $B \subseteq \mathbb{R}^2$.

Några obs

- $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dy dx$.
- $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$.
- $\mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times \mathbb{R}) = \int_A \int_{-\infty}^{\infty} f(x, y) dy dx$, så

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

•

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

Exempel: Låt (X, Y) ha täthet $f(x, y) = c(x + 3y)$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. Vad är c och vad är f_X och f_Y ?

Det måste gälla att $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$, så

$$1 = c \int_0^1 \int_0^1 (x + 3y) dx dy = c \left(\int_0^1 x dx + 3 \int_0^1 y dy \right) = 2c$$

så $c = 1/2$. Alltså

$$f_X(x) = \frac{1}{2} \int_0^1 (x + 3y) dy = \frac{1}{2} \left(x + \frac{3}{2} \right), \quad 0 \leq x \leq 1$$

och

$$f_Y(y) = \frac{1}{2} \int_0^1 (x + 3y) dx = \frac{1}{2} \left(3y + \frac{1}{2} \right), \quad 0 \leq y \leq 1.$$

□

Exempel: Låt $D = \{(x, y) : x^2 + y^2 \leq 1\}$. Låt $(X, Y) \sim \text{likf}(D)$.

Dvs

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x \in D \\ 0, & x \notin D \end{cases}$$

Låt $Z = |(X, Y)| = \sqrt{X^2 + Y^2}$. Vad är förd för Z ?

$$F_Z(z) = \mathbb{P}((X, Y) \in B_z) = \frac{1}{\pi} \text{Area}(B_z)$$

där $B_z = \{(x, y) : x^2 + y^2 \leq z^2\}$. Alltså är $\text{Area}(B_z) = \pi z^2$, så

$$F_Z(z) = z^2, \quad 0 \leq z \leq 1$$

och

$$f_Z(z) = 2z, \quad 0 \leq z \leq 1.$$

Exempel forts: Vad är förd för X ?

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1-x^2}.$$

□

Om $f(x, y) = f_X(x)f_Y(y)$ för alla (x, y) , gäller

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \int_A \int_B f(x, y) dy dx = \int_A f_X(x) dx \int_B f_Y(y) dy \\ &= \mathbb{P}(X \in A) \mathbb{P}(Y \in B) \end{aligned}$$

dvs X och Y är oberoende.

Omvänt: X och Y oberoende $\Rightarrow f(x, y) = f_X(x)f_Y(y)$ för (nästan) alla (x, y) .

Funktioner av två sv. Om (X, Y) är en tvådim sv och $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, vad är förd för $g(X, Y)$? Avgör från fall till fall.

Exempel: $(X, Y) \sim \text{likf}([0, 1]^2)$. Låt $A = XY$. För $0 < a < 1$,

$$\begin{aligned} F_A(a) &= \mathbb{P}(XY \leq a) = \int_0^1 \int_0^{\min(a/x, 1)} dy dx \\ &= \left(a + a \int_a^1 \frac{1}{x} dx \right) \\ &= a - a \ln a. \end{aligned}$$

Derivering ger

$$f_A(a) = -\ln a, \quad 0 < a < 1.$$



OSL i två dim.

- Diskret: $\mathbb{E}[g(X, Y)] = \sum_{(x_j, y_k) \in V_{X, Y}} g(x_j, y_k) p(x_j, y_k)$.
- Kont: $\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$.

Bevis diskret: Låt $V_g = g(V_{X, Y})$.

$$\begin{aligned}
 \mathbb{E}[g(X, Y)] &= \sum_{z \in V_g} z \mathbb{P}(g(X, Y) = z) \\
 &= \sum_{z \in V_g} z \sum_{(x_j, y_k): g(x_j, y_k) = z} p(x_j, y_k) \\
 &= \sum_{z \in V_g} \sum_{(x_j, y_k): g(x_j, y_k) = z} g(x_j, y_k) p(x_j, y_k) \\
 &= \sum_{(x_j, y_k) \in V_{X, Y}} g(x_j, y_k) p(x_j, y_k)
 \end{aligned}$$



Följdsats: för alla sv X och Y gäller $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Bevis: Vi tar det kont fallet. Enl OSL:

$$\begin{aligned}
 \mathbb{E}[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)f(x, y)dx dy \\
 &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x, y)dy dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x, y)dx dy \\
 &= \int_{-\infty}^{\infty} xf_X(x)dx + \int_{-\infty}^{\infty} yf_Y(y)dy \\
 &= \mathbb{E}[X] + \mathbb{E}[Y]
 \end{aligned}$$

□

Exempel: Låt I_1, I_2, \dots, I_n vara sådana att $\mathbb{P}(I_k = 1) = 1 - \mathbb{P}(I_k = 0) = p$. Låt $X = \sum_{k=1}^n I_k$. Då ger följsatsen att

$$\mathbb{E}[X] = \sum_{k=1}^n \mathbb{E}[I_k] = np.$$

Ex.vis visar detta att $X \sim \text{Bin}(n, p) \Rightarrow \mathbb{E}[X] = np$ (fallet då I_k :na är oberoende, vilket följsatsen inte kräver).