Flerdimensionella stokastiska variabler

Den naturliga generaliseringen till flera dimensioner är att

$$\mathsf{P}\left((X,Y)\in D\right) = \begin{cases} \displaystyle\sum\sum_{(x,y)\in D} p_{X,Y}(x,y) & \text{om } (X,Y) \text{ diskret} \\[2ex] \displaystyle\int\int_{(x,y)\in D} f_{X,Y}(x,y)\,dx\,dy & \text{om } (X,Y) \text{ kontinuerlig.} \end{cases}$$

Funktionerna $p_{X,Y}(x,y)$ och $f_{X,Y}(x,y)$ kallas den *simultana* sannolikhetsfunktionen och den *simultana* täthetsfunktionen för de stokastiska variablerna X och Y.

Minns oberoende händelser: A och B oberoende om $\mathsf{P}(A\cap B) = \mathsf{P}(A)\cdot\mathsf{P}(B)$. Två stokastiska variabler borde vara oberoende om

$$\mathsf{P}\left(\{X=x\}\cap\{Y=y\}\right) = \mathsf{P}(X=x)\,\mathsf{P}(Y=y)$$

Definition: Två stokastiska variabler är oberoende om

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \qquad \text{för alla } x \text{ och } y$$

om (X,Y) är diskret, och

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \qquad \text{för alla } x \text{ och } y$$

om (X,Y) är kontinuerlig.

Funktioner av stokastiska variabler

Funktioner av stokastiska variabler är nya stokastiska variabler. Låt X vara en stokastisk variabel och $g(x)$ en reellvärd funktion. Då är $Z = g(X)$ en stokastisk variabel.

Exempel: X och Y oberoende diskreta stokastiska variabler. Låt $Z = X+Y$. Oberoendet ger att $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ och

$$\mathsf{P}(X+Y=n) = \sum_k \mathsf{P}(X=k)\,\mathsf{P}(Y=n-k)$$

I det kontinuerliga fallet får man motsvarande

$$f_{X+Y}(z) = \int f_X(t)\,f_Y(z-t)\,dt$$

Formeln kallas *faltningsformeln.*

Exempel: Låt $X_1,\ldots,X_n$ vara oberoende, likafördelade stokastiska variabler med fördelningsfunktion $F_X(t)$. Med $Z = \max(X_1,\ldots,X_n)$ så

$$\begin{aligned} F_Z(t) &= \mathsf{P}(Z\le t) = \mathsf{P}(\max(X_1,\ldots,X_n)\le t) = \mathsf{P}(X_1\le t,\ldots,X_n\le t) \\ &= \mathsf{P}(X_1\le t)\cdots\mathsf{P}(X_n\le t) = (F_X(t))^n \end{aligned}$$

Väntevärden och andra parametrar

Beskrivande parametrar för en stokastisk variabel:

- Genomsnittligt värde, väntevärde $\mathsf{E}[X]$
- Spridningsmått, varians $\mathrm{V}(X)$ och standardavvikelse $\mathrm{D}(X)$

Definition: *Väntevärdet* för en stokastisk variabel, $\mathsf{E}[X]$, definieras som

$$\mathsf{E}[X] = \begin{cases} \displaystyle\sum_{k \in S_X} k \cdot \mathsf{P}(X = k) & \text{om } X \text{ är diskret} \\ \displaystyle\int_{-\infty}^{\infty} x \cdot f_X(x)\,dx & \text{om } X \text{ är kontinuerlig} \end{cases}$$

Väntevärdet är ett teoretiskt genomsnittligt värde.

Exempel: X är resultatet av ett tärningskast. $\mathsf{P}(X = i) = 1/6$ för $i = 1, 2, \ldots, 6$.

$$\mathsf{E}[X] = \sum_{k \in S_X} k \mathsf{P}(X = k) = \sum_{k=1}^{6} k \frac{1}{6} = \frac{1}{6} + \frac{2}{6} + \cdots \frac{6}{6} = 3.5.$$

Exempel: X är exponentialfördelad, dvs $f_X(x) = \frac{1}{m} \mathrm{e}^{-x/m}$ för $x \geq 0$.

$$\mathsf{E}[X] = \int_0^{\infty} x \cdot \frac{1}{m} \mathrm{e}^{-x/m}\,dx = \cdots = m.$$

Sats (Den aningslöse statistikerns sats). För en reellvärd funktion $g(x)$ och en stokastisk variabel X är

$$\mathsf{E}[g(X)] = \begin{cases} \displaystyle\sum_{k \in S_X} g(k) \cdot \mathsf{P}(X = k) & \text{om } X \text{ är diskret} \\ \displaystyle\int_{-\infty}^{\infty} g(x) \cdot f_X(x)\,dx & \text{om } X \text{ är kontinuerlig} \end{cases}$$

förutsatt att summan/integralen är absolutkonvergent.

Spridningsmått?

Definition: *Variansen* för en stokastisk variabel definieras och betecknas

$$\mathrm{V}(X) = \mathsf{E}\left[(X - \mathsf{E}[X])^2\right].$$

Definition: *Standardavvikelsen* för en stokastisk variabel definieras och betecknas

$$\mathrm{D}(X) = \sqrt{\mathrm{V}(X)}.$$

Denna har rätt enhet.

Exempel: Tärningskast

$$\mathrm{V}(X) = \mathsf{E}\left[(X - 3.5)^2\right] = \sum_{k=1}^{6} (k - 3.5)^2 \frac{1}{6} = \frac{35}{12} \approx 2.917$$

(enhet: ”kvadratprickar”)

Exempel: Tärningskast

$$\mathrm{D}(X) = \sqrt{\mathrm{V}(X)} = \sqrt{\frac{35}{12}} \approx 1.71$$

(enhet: ”prickar”)

En liten insyn i vad standardavvikelsen betyder kan fås genom följande.

Sats (Tjebychevs olikhet). För alla $k > 0$ är

$$\mathsf{P}(|X - \mathsf{E}[X]| > k\mathrm{D}(X)) \leq \frac{1}{k^2}.$$