Funktioner av stokastiska variabler

Funktioner av stokastiska variabler $Y = g(X)$ är nya stokastiska variabler.

Exempel: X är en kontinuerlig stokastisk variabel med täthet $f_X(x)$. och $S_X = (-\infty, \infty)$. Med $Y = X^2$ är Y en kontinuerlig stokastisk variabel med $S_Y = [0, \infty)$ och fördelningsfunktion

$$F_Y(t) = \mathsf{P}(Y \leq t) = \mathsf{P}\left(X^2 \leq t\right) = \mathsf{P}\left(-\sqrt{t} \leq X \leq \sqrt{t}\right) = F_X(\sqrt{t}) - F_X(-\sqrt{t}),$$

dvs Y har täthet

$$f_Y(t) = \frac{d}{dt}F_Y(t) = f_X(\sqrt{t})\frac{1}{2\sqrt{t}} - f_X(-\sqrt{t})\frac{-1}{2\sqrt{t}} = \frac{1}{2\sqrt{t}}\left(f_X(\sqrt{t}) + f_X(-\sqrt{t})\right).$$

Flerdimensionella stokastiska variabler

Den naturliga generaliseringen till flera dimensioner är att

$$\mathsf{P}((X,Y) \in D) = \begin{cases} \displaystyle\sum\sum_{(x,y)\in D} p_{X,Y}(x,y) & \text{om } (X,Y) \text{ diskret} \\ \displaystyle\iint_{(x,y)\in D} f_{X,Y}(x,y)\,dx\,dy & \text{om } (X,Y) \text{ kontinuerlig.} \end{cases}$$

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.

De simultana fördelningarna bestämmer *marginalfördelningarna*:

$$\int_A f_X(x)\,dx = \mathsf{P}(X \in A) = \mathsf{P}((X,Y) \in A \times (-\infty,\infty)) = \int_A \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx$$

vilket ger $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$. På samma sätt för diskreta fördelningar så är

$$p_X(x) = \mathsf{P}(X = x) = \mathsf{P}\left(\bigcup_y \{X = x, Y = y\}\right) = \sum_y \mathsf{P}(X = x, Y = y) = \sum_y p_{X,Y}(x,y).$$

Minns oberoende händelser: A och B oberoende om

$$\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B)$$

Om vi översätter det till en utsaga om stokastiska variabler så får vi

Definition: Två stokastiska variabler X och Y är oberoende om

$$\mathsf{P}(\{X \in D_x\} \cap \{Y \in D_y\}) = \mathsf{P}(X \in D_x) \cdot \mathsf{P}(Y \in D_y)$$

för alla mängder (intervall) D_x och D_y.

Speciellt: X och Y är oberoende om

$$F_{X,Y}(x,y) := \mathsf{P}(X \leq x, Y \leq y) = \mathsf{P}(X \leq x)\,\mathsf{P}(Y \leq y) = F_X(x)F_Y(y)$$

för alla x och y, eller i det diskreta fallet:

$$p_{X,Y}(x,y) = \mathsf{P}(\{X = x\} \cap \{Y = y\}) = \mathsf{P}(X = x)\,\mathsf{P}(Y = y) = p_X(x)p_Y(y)$$

för alla x och y, eller i det kontinuerliga fallet

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

för alla x och y.

Exempel: Summa av två oberoende diskreta stokastiska variabler X och Y. Då har Z sannolikhetsfunktionen

$$\begin{aligned}
p_Z(n) &= \mathsf{P}(Z=n) = \mathsf{P}(X+Y=n) = \mathsf{P}\left(\bigcup_k \{X=k, Y=n-k\}\right) \\
&= \sum_k \mathsf{P}(X=k, Y=n-k) = \sum_k \mathsf{P}(X=k)\,\mathsf{P}(Y=n-k) = \sum_k p_X(k) p_Y(n-k)
\end{aligned}$$

Motsvarande för summa av oberoende kontinuerliga stokastiska variabler är

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

Formeln kallas för faltningsformeln.

Exempel: Rekord $Y = \max(X_1, \ldots, X_n)$ där $X_1, \ldots, X_n$ är oberoende med samma fördelning.

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

Väntevärden och andra parametrar

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 genomsnitt.

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 antalet kast till första 5:a. X är ffg(p), $p = 1/6$. Då är

$$\begin{aligned}
\mathsf{E}[X] &= \sum_k k\mathsf{P}(X=k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p = p \sum_{k=0}^{\infty} k q^{k-1} = p\frac{d}{dq}\sum_{k=0}^{\infty} q^k = p\frac{d}{dq}\frac{1}{1-q} \\
&= p\frac{1}{(1-q)^2} = p\frac{1}{p^2} = \frac{1}{p}.
\end{aligned}$$

Exempel: X är exponentialfördelad med parameter λ, dvs $f_X(x) = \lambda e^{-\lambda x}$ för $x \ge 0$.

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

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

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

förutsatt att summan/integralen är absolutkonvergent.

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

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

Exempel: Tärningskast

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

(enhet: ”kvadratprickar”)

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

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

Denna har rätt enhet.