

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

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

dvs Y har täthet

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

Flerdimensionella stokastiska variabler

Den naturliga generaliseringen till flera dimensioner är att

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

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 .

De simultana fördelningarna bestämmer *marginalfördelningarna*:

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

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

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

Minns oberoende händelser: A och B oberoende om

$$P(A \cap B) = P(A) \cdot P(B)$$

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

Definition: Två stokastiska variabler X och Y är oberoende om

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

för alla mängder (intervall) D_x och D_y .

Speciellt: X och Y är oberoende om

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

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

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

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

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

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 $Y = \max(X_1, \dots, X_n)$ där $X_1, \dots, X_n$ är oberoende med samma fördelning.

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

Väntevärden och andra parametrar

Definition: Väntevärdet för en stokastisk variabel, $E[X]$, definieras som

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

Väntevärdet är ett teoretiskt genomsnitt.

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

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

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

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

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

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

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

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

förutsatt att summan/integralen är absolutkonvergent.

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

$$V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Exempel: Tärningskast

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

(enhet: ”kvadratprickar”)

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

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

Denna har rätt enhet.