

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

$$V(X) = E[(X - E[X])^2].$$

Exempel: Tärningskast

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

(enhet: "kvadratprickar")

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

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

Denna har rätt enhet.

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

Sats (Tjebychovs olikhet). För $k > 0$ är

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

Bevis: (Tjebychovs olikhet) Låt X vara en kontinuerlig stokastisk variabel med $\mu = E[X]$ och $\sigma = D(X)$. Då är

$$P(|X - \mu| > \epsilon) = \int_{|x-\mu|>\epsilon} f_X(x) dx = \int_{\frac{(x-\mu)^2}{\epsilon^2} > 1} 1 \cdot f_X(x) dx \leq \int \frac{(x-\mu)^2}{\epsilon^2} \cdot f_X(x) dx = \frac{1}{\epsilon^2} E[(X - \mu)^2] = \frac{\sigma^2}{\epsilon^2}.$$

Med $\epsilon = k\sigma$ fås satsens påstående.

Räknelagar för väntevärden och varianser

Sats. För konstanter a, b är

$$E[aX + b] = aE[X] + b \quad V(aX + b) = a^2V(X).$$

Bevis: (diskreta fallet)

$$\begin{aligned} E[aX + b] &= \sum_{k \in \Omega_X} (ak + b)P(X = k) \\ &= a \sum_{k \in \Omega_X} kP(X = k) + b \sum_{k \in \Omega_X} P(X = k) = aE[X] + b \\ V(aX + b) &= E[((aX + b) - E[aX + b])^2] = E[(aX + b - aE[X] - b)^2] \\ &= E[a^2(X - E[X])^2] = a^2E[(X - E[X])^2] = a^2V(X) \end{aligned}$$

Sats. För två stokastiska variabler X och Y är

$$E[X + Y] = E[X] + E[Y].$$

Bevis: (diskreta fallet)

$$\begin{aligned} \mathbf{E}[X + Y] &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} (x + y) \mathbf{P}(X = x, Y = y) \\ &= \sum_{x \in \Omega_X} x \sum_{y \in \Omega_Y} \mathbf{P}(X = x, Y = y) + \sum_{y \in \Omega_Y} y \sum_{x \in \Omega_X} \mathbf{P}(X = x, Y = y) \\ &= \sum_{x \in \Omega_X} x \mathbf{P}(X = x) + \sum_{y \in \Omega_Y} y \mathbf{P}(Y = y) = \mathbf{E}[X] + \mathbf{E}[Y]. \end{aligned}$$

Sats. För två oberoende stokastiska variabler X och Y gäller

$$\mathbf{E}[XY] = \mathbf{E}[X] \mathbf{E}[Y].$$

Bevis: (kontinuerliga fallet)

$$\begin{aligned} \mathbf{E}[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} y f_Y(y) \underbrace{\int_{-\infty}^{\infty} x f_X(x) dx}_{\mathbf{E}[X]} dy = \mathbf{E}[X] \underbrace{\int_{-\infty}^{\infty} y f_Y(y) dy}_{\mathbf{E}[Y]} = \mathbf{E}[X] \mathbf{E}[Y]. \end{aligned}$$

Exempel: Låt U_1, U_2, U_3 vara oberoende och likaformigt fördelade på intervallet $[0, 1]$. Då är $f_U(x) = 1$ för $0 \leq x \leq 1$ och $\mathbf{E}[U] = 1/2$.

Av U_1 och U_2 konstrueras en rektangel med U_1 och U_2 som sidlängder. Rektangeln har då en genomsnittlig area

$$\mathbf{E}[U_1 \cdot U_2] = \{\text{oberoende}\} = \mathbf{E}[U_1] \mathbf{E}[U_2] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Vidare konstrueras en kvadrat med sidlängd U_3 . Kvadraten har en genomsnittlig area

$$\mathbf{E}[U_3^2] = \int_{-\infty}^{\infty} x^2 f_U(x) dx = \int_0^1 x^2 \cdot 1 dx = \frac{1}{3} > \frac{1}{4}$$

så kvadraten har i genomsnitt större area än rektangeln.

Låt (X, Y) vara en tvådimensionell stokastisk variabel och beteckna $\mu_x = \mathbf{E}[X]$ och $\mu_y = \mathbf{E}[Y]$. Då är

$$\begin{aligned} \mathbf{V}(X + Y) &= \mathbf{E}[(X + Y - \mathbf{E}[X + Y])^2] = \mathbf{E}[(X - \mu_x + Y - \mu_y)^2] \\ &= \mathbf{E}[(X - \mu_x)^2 + (Y - \mu_y)^2 + 2(X - \mu_x)(Y - \mu_y)] \\ &= \mathbf{E}[(X - \mu_x)^2] + \mathbf{E}[(Y - \mu_y)^2] + 2\mathbf{E}[(X - \mu_x)(Y - \mu_y)] \\ &= \mathbf{V}(X) + \mathbf{V}(Y) + 2\mathbf{C}(X, Y). \end{aligned}$$

Definition: Kovariansen mellan två stokastiska variabler X och Y definieras

$$\mathbf{C}(X, Y) = \mathbf{E}[(X - \mu_x)(Y - \mu_y)].$$

Variabeln

$$(X - \mu_x)(Y - \mu_y)$$

mäter om X och Y tenderar att variera åt samma eller motsatt håll. Om X är stor/liten (relativt μ_x) och Y samtidigt är stor/liten (relativt μ_y) så är kovariansen positiv eftersom $+\cdot+=+$ och $-\cdot-=+$. Analogt, om X och Y varierar åt motsatt håll är kovariansen negativ, $+\cdot-= -$ och $-\cdot+= -$. Kovariansen mäter *linjär* samvariation.

Korrelationen (korrelationskoefficienten) är det dimensionslösa linjära samvariationsmättet. Den definieras och betecknas

$$\rho_{X,Y} = \frac{C(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{C(X,Y)}{D(X)D(Y)}.$$

Korrelationskoefficienten uppfyller alltid

$$-1 \leq \rho_{X,Y} \leq 1$$

och $|\rho_{X,Y}| = 1$ om $Y = aX + b$.

Med linjäriteten hos väntevärden har vi även följande beräkningsformel

$$\begin{aligned} C(X,Y) &= E[(X - \mu_x)(Y - \mu_y)] = E[XY - \mu_x Y - \mu_y X + \mu_x \mu_y] \\ &= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y = E[XY] - \mu_x \mu_y. \end{aligned}$$

Speciellt kan man notera att $V(X) = C(X,X)$ så

Sats (Steiners sats).

$$V(X) = E[X^2] - (E[X])^2.$$

För oberoende stokastiska variabler är $E[XY] = E[X]E[Y]$ så

$$C(X,Y) = E[XY] - \mu_x \mu_y = E[X]E[Y] - \mu_x \mu_y = 0$$

och

$$V(X+Y) = V(X) + V(Y) + \underbrace{2C(X,Y)}_{=0} = V(X) + V(Y).$$

Stokastiska variabler (inte nödvändigtvis oberoende) som har $C(X,Y) = 0$ (alt. $\rho_{X,Y} = 0$) sägs vara *okorrelerade*.