

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 (Tjebychevs olikhet). För $k > 0$ är

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

Lemma [Markovs olikhet]: För en positiv stokastisk variabel X gäller att

$$P(X > c) \leq \frac{1}{c} E[X]$$

för alla $c > 0$.

Bevis: (Kontinuerliga fallet)

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x f_X(x) dx \geq \int_c^{\infty} x f_X(x) dx \geq c \int_c^{\infty} f_X(x) dx = cP(X > c).$$

Bevis: (Tjebychevs olikhet) Med $\mu = E[X]$ är $(X - \mu)^2$ en positiv stokastisk variabel och

$$P(|X - \mu| > c) = P((X - \mu)^2 > c^2) \leq \frac{1}{c^2} E[(X - \mu)^2] = \frac{V(X)}{c^2}.$$

Med $c = kD(X) > 0$ 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)

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

Sats. För två stokastiska variabler X och Y är

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Bevis: (diskreta fallet)

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

Sats. För två oberoende stokastiska variabler X och Y gäller

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

Bevis: (kontinuerliga fallet)

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

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 $\mathbb{E}[U] = 1/2$.

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

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

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

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

så kvadraten har i genomsnitt större area än rektangeln.

Med lineariteten för väntevärden har vi följande beräkningsformel för varianser:

Sats (Steiners sats).

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

(notera att $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$.)

Bevis: Med beteckningen $\mu = \mathbb{E}[X]$ har vi

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu \mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.\end{aligned}$$

Satsen används även ofta för att beräkna $\mathbb{E}[X^2]$ från $\mathbb{V}(X)$ och $\mathbb{E}[X]$ i bevis:

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

Notera att detta medför att

$$\mathbb{E}[\text{Area kvadrat}] = \mathbb{E}[X^2] = \underbrace{V(X)}_{\geq 0} + (\mathbb{E}[X])^2 \geq (\mathbb{E}[X])^2 = \mathbb{E}[\text{Area rektangel}]$$

oavsett sidlängdsfördelningen så länge som rektangelns sidor beskrivs av oberoende stokastiska variabler.

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

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

Definition: Kovariansen mellan två stokastiska variabler X och Y definieras

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

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

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

Korrealtionskoefficienten 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 (jämför med Steiners sats för varianser)

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

För oberoende stokastiska variabler är $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ så

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

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*.