

Räkneregler för stokastiska variabler.

För konstanter a och b och en stokastisk variabel X är

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b \quad \mathbb{V}(aX + b) = \mathbb{V}(aX) = a^2\mathbb{V}(X)$$

Bevis:

$$\begin{aligned}\mathbb{E}[aX + b] &= \sum_k (ak + b)\mathbb{P}(X = k) = a \sum_k k\mathbb{P}(X = k) + b \sum_k \mathbb{P}(X = k) = a\mathbb{E}[X] + b \\ \mathbb{V}(aX + b) &= \mathbb{E}[(aX + b) - \mathbb{E}[aX + b]]^2 = \mathbb{E}[(aX + b - a\mathbb{E}[X] - b)^2] = \mathbb{E}[(aX - a\mathbb{E}[X])^2] \\ &= \mathbb{E}[a^2(X - \mathbb{E}[X])^2] = a^2\mathbb{E}[(X - \mathbb{E}[X])^2] = a^2\mathbb{V}(X).\end{aligned}$$

Sats. Låt X och Y vara två stokastiska variabler. Då är

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

Vi bevisar det 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. Om X och Y är två oberoende stokastiska variabler är

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

Bevis: Här bevisas det diskreta fallet:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_k \sum_n kn\mathbb{P}(X = k, Y = n) = \sum_k \sum_n kn\mathbb{P}(X = k)\mathbb{P}(Y = n) \\ &= \underbrace{\sum_k k\mathbb{P}(X = k)}_{=\mathbb{E}[X]} \underbrace{\sum_n n\mathbb{P}(Y = n)}_{=\mathbb{E}[Y]} = \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

Beräkningsformel för varianser:

Sats.

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

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

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

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

Exempel: Area av kvadrat och rektangel: Låt X_1 och X_2 vara två oberoende sidlängder, likformigt fördelade på intervallet $[0, 1]$. Då är

$$E[X] = \frac{1}{2} \quad V(X) = \frac{1}{12}.$$

Vidare så är

$$E[\text{Area rektangel}] = E[X_1 \cdot X_2] = \{\text{oberoende}\} = E[X_1] E[X_2] = \frac{1}{4}.$$

medan

$$E[\text{Area kvadrat}] = E[X^2] = V(X^2) + E[X]^2 = \frac{1}{12} + \frac{1}{4} = \frac{1}{3} \geq m^2.$$

Låt (X, Y) vara en tvådimensionell stokastisk variabel och beteckna $m_X = E[X]$ och $m_Y = E[Y]$. Då är

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

Definition: Kovariansen mellan två stokastiska variabler X och Y definieras och betecknas

$$C(X, Y) = E[(X - m_X)(Y - m_Y)].$$

Variabeln

$$(X - m_X)(Y - m_Y)$$

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

Korrelationen (korrelationskoefficienten) är det dimensionslösa linjära samvariationsmåttet. 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 och endast om $Y = aX + b$.

Med linjäriteten hos väntevärden har vi även följande beräkningsformel (jämför med den för varianser)

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

För två oberoende stokastiska är kovariansen variabler

$$C(X, Y) = E[XY] - m_X m_Y = E[X] E[Y] - m_X m_Y = 0$$

och

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

Två stokastiska variabler X och Y som uppfyller $C(X, Y) = 0$ sägs vara *okorrelerade*.

Sats. Låt $X_1, \dots, X_n$ vara oberoende likafördelade stokastiska variabler med $m = E[X_i]$ och $\sigma = D(X_i)$. Medelvärde

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

har

$$E[\bar{X}] = m \quad V(\bar{X}) = \frac{\sigma^2}{n} \quad D(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

Med Tjebychevs olikhet: För alla $k > 0$ är

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

fås

Sats (Stora talens lag (STL)). Låt $X_1, \dots, X_n$ vara oberoende likafördelade stokastiska variabler med $m = E[X_i]$ och $\sigma = D(X_i)$.

För varje $\delta = kD(\bar{X}) > 0$ fås med Tjebychevs olikhet att

$$P(|\bar{X} - m| > \delta) \leq \frac{1}{k^2} = \frac{V(\bar{X})}{\delta^2} = \frac{\sigma^2}{\delta^2 n} \rightarrow 0$$

då $n \rightarrow \infty$.