Föreläsning 8

Summor av oberoende stokastiska variabler

Faltning

Sats (a) Låt X och Y vara två oberoende diskreta stokastiska variabler som kan anta värdena $\{0, 1, 2, \dots\}$. Det gäller att

$$p_{X+Y}(k) = \sum_{i=0}^{k} p_X(k-i)p_Y(i)\,, \quad k = 0, 1, 2, \dots\,.$$

(b) Låt X och Y vara två oberoende kontinuerliga stokastiska variabler. Det gäller att

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)\,dy\,, \quad a \in \mathbb{R}.$$

Bevis av (a). Det gäller att

$$\begin{aligned} p_{X+Y}(k) &= P(X+Y=k) = P\left(\bigcup_{i=0}^{k}\{X=k-i\}\cap\{Y=i\}\right) \\ (\{X=k-i\} \;\cap\; \{Y=i\} \text{ disjunkta}) &= \sum_{i=0}^{k} P\left(X=k-i, Y=i\right) \\ (X,Y \;\text{ oberoende}) &= \sum_{i=0}^{k} P(X=k-i)\cdot P(Y=i) = \sum_{i=0}^{k} p_X(k-i)p_Y(i)\,. \end{aligned}$$

□

Exempel Låt X, Y vara oberoende $U(0,1)$-fördelade stokastiska variabler. Bestäm f_{X+Y}. Visa på så sätt att en summa av två oberoende rektangelfördelade stokastiska variabler är triangelfördelad. Använd Sats (b).

Lösning. Kom ihåg att

$$f_X(x) = \begin{cases} 1 & \text{om } x \in (0,1) \\ 0 & \text{annars} \end{cases} \quad \text{och} \quad f_Y(y) = \begin{cases} 1 & \text{om } y \in (0,1) \\ 0 & \text{annars} \end{cases}\,.$$

$X+Y$ kan anta värdena mellan 0 och 2. För att bestämma $f_{X+Y}(a)$ skiljer vi mellan två fall, $0 < a \leq 1$ och $1 < a < 2$. Låt $0 < a \leq 1$.

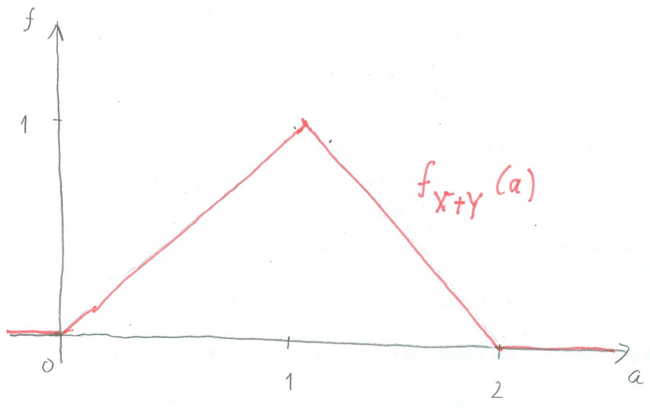
$$\begin{aligned} f_{X+Y}(a) &= \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)\,dy \\ &= \int_{y=0}^{a} f_X(a-y)f_Y(y)\,dy = \int_{y=0}^{a} 1\cdot 1\,dy = a\,, \end{aligned}$$

här är $f_Y(y) = 0$ om $y \notin (0,1)$ och $f_X(a-y) = 0$ om $y \geq a$. Nu låt $1 < a < 2$.

$$\begin{aligned} f_{X+Y}(a) &= \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)\,dy \\ &= \int_{y=a-1}^{1} f_X(a-y)f_Y(y)\,dy = \int_{y=a-1}^{1} 1\cdot 1\,dy = 2-a\,, \end{aligned}$$

här är $f_Y(y) = 0$ om $y \notin (0,1)$ och $f_X(a-y) = 0$ om $y \leq a-1$. Sammanfattningsvist är det

$$f_{X+Y}(a) = \begin{cases} a & \text{om} \quad a \in (0,1] \\ 2-a & \text{om} \quad a \in (1,2) \\ 0 & \text{annars} \end{cases} .$$



Särskilda fördelningar

Sats (a) Låt $X_1, \ldots, X_n$ vara oberoende normalfördelade stokastiska variabler med parametrarna $(\mu_1, \sigma_1), \ldots, (\mu_n, \sigma_n)$. Då är

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right).$$

(b) Låt $X_1 \sim Bin(m,p)$ och $X_2 \sim Bin(n,p)$ vara oberoende. Då är

$$X_1 + X_2 \sim Bin(m+n, p).$$

(c) Låt $X_1 \sim Po(\lambda_1)$ och $X_2 \sim Po(\lambda_2)$ vara oberoende. Då är

$$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2).$$

Bevis. Del (a) bevisas med hjälp av momentgenererande funktioner, se storseminarium 4.
(b)

$$X_1 = \sum_{i=1}^{m} Y_i \,, \quad X_2 = \sum_{i=m+1}^{m+n} Y_i$$

där $Y_1, \ldots, Y_{m+n}$ är oberoende stokastiska variabler med $P(Y_i = 0) = 1-p$ och $P(Y_i = 1) = p$. Detta medför

$$X_1 + X_2 = \sum_{i=1}^{m+n} Y_i \sim Bin(m+n, p).$$

(c) Vi använder Sats (a). Det gäller att

$$\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} P(X=k-i, Y=i) = \sum_{i=0}^{k} P(X=k-i)\cdot P(Y=i) \\
&= \sum_{i=0}^{k} e^{-\lambda_1}\frac{\lambda_1^{k-i}}{(k-i)!}\cdot e^{-\lambda_2}\frac{\lambda_2^i}{i!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda_1^{k-i}\lambda_2^i \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda_1^{k-i}\lambda_2^i = e^{-(\lambda_1+\lambda_2)}\cdot\frac{(\lambda_1+\lambda_2)^k}{k!}, \quad k=0,1,\dots,
\end{aligned}$$

där vi har använt Binomialsatsen. $\square$

Väntevärde och varians av summor av stokastiska variabler

Sats (a) Låt $X_1,\dots,X_n$ vara stokastiska variabler med $E[|X_i|]<\infty$. Det gäller att

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$$

(b) Låt $X_1,\dots,X_n$ vara *oberoende* stokastiska variabler med $E[X_i^2]<\infty$. Det gäller att

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

Exempel Låt $X_1,\dots,X_n$ vara oberoende och likafördelade stokastiska variabler med $Var(X)=\sigma^2<\infty$. Visa att

$$\lim_{n\to\infty} Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = 0.$$

.

Lösning.

$$\begin{aligned}
\lim_{n\to\infty} Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) &= \lim_{n\to\infty}\left[\frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right)\right] \\
&= \lim_{n\to\infty}\left[\frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right)\right] = \lim_{n\to\infty}\left[\frac{1}{n^2}\cdot n\sigma^2\right] = 0\,.
\end{aligned}$$

Sats (a) Om $X_1\sim N(\mu_1,\sigma_1)$, ..., $X_n\sim N(\mu_n,\sigma_n)$ är oberoende och $a_1,\dots,a_n,b\in\mathbb{R}$ så är

$$\sum_{i=1}^{n} a_iX_i + b \sim N\left(\sum_{i=1}^{n} a_i\mu_i + b, \sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}\right).$$

(b) Om $X_1, \ldots, X_n$ är oberoende $N(\mu, \sigma)$-fördelade stokastiska variabler så är

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

(c) Om $X_1 \sim N(\mu_1, \sigma_1)$, ... , $X_{n_1} \sim N(\mu_1, \sigma_1)$ och $Y_1 \sim N(\mu_2, \sigma_2)$, ... , $Y_{n_2} \sim N(\mu_2, \sigma_2)$ är oberoende så är

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

Exempel Vi betraktar oberoende stokastiska variabler $X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3, Y_4$ som alla är normalfördelade med väntevärde 28 och standardavvikelse 0.25. Låt

$$\overline{X} = \frac{1}{4} \sum_{i=1}^{4} X_i\,, \qquad \overline{Y} = \frac{1}{4} \sum_{i=1}^{4} Y_i\,.$$

(a) Bestäm $P(27.5 < X_1 < 28.5)$.

(b) Bestäm $P(27.5 < \overline{X} < 28.5)$.

(c) Bestäm $P(-0.3 < \overline{X} - \overline{Y} < 0.3)$.

Lösning.

(a)

$$\begin{aligned} P(27.5 < X_1 < 28.5) &= \Phi\left(\frac{28.5 - 28}{0.25}\right) - \Phi\left(\frac{27.5 - 28}{0.25}\right) \\ &= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0.95\,. \end{aligned}$$

(b) $\overline{X} \sim N(\mu, \sigma/\sqrt{n}) = N(28, 0.125)$

$$\begin{aligned} P(27.5 < \overline{X} < 28.5) &= \Phi\left(\frac{28.5 - 28}{0.125}\right) - \Phi\left(\frac{27.5 - 28}{0.125}\right) \\ &= \Phi(4) - \Phi(-4) = 2\Phi(4) - 1 = 0.99994 \approx 1\,. \end{aligned}$$

beräknad med `http://homepage.stat.uiowa.edu/~mbognar/`.

(c) P.s.s. $\overline{Y} \sim N(\mu, \sigma/\sqrt{n}) = N(28, 0.125)$ som medför att $\overline{X} - \overline{Y}$ har väntevärde 0 och varians

$$Var(\overline{X} - \overline{Y}) = Var(\overline{X}) + (-1)^2 Var(\overline{Y}) = \frac{2}{8^2} = 0.177^2.$$

$$\begin{aligned} P(-0.3 < \overline{X} - \overline{Y} < 0.3) &= \Phi\left(\frac{0.3 - 0}{0.177}\right) - \Phi\left(\frac{-0.3 - 0}{0.177}\right) \\ &= \Phi(1.70) - \Phi(-1.70) = 2\Phi(1.70) - 1 = 0.91\,. \end{aligned}$$

Flerdimensionell normalfördelning

Definition Låt $X_1, \ldots, X_n$ vara stokastiska variabler och $\mathbf{X} = (X_1, \ldots, X_n)^T$.
(a) Vektorn

$$\mu_{\mathbf{X}} \equiv E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$

kallas *väntevärdesvektorn*.
(b) Matrisen

$$Cov_{\mathbf{X}} = \left[\ Cov(X_i, X_j) \ \right]_{i,j=1,\ldots,n}$$

kallas *kovariansmatrisen*.
(c) Stokastiska vektorn $\mathbf{X} = (X_1, \ldots, X_n)^T$ med den simultana täthetsfunktionen

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |det\mathbf{B}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(x-\mu)^T \mathbf{B}^{-1} (x-\mu) \right\}, \quad x \in \mathbb{R}^n,$$

kallas *n-dimensionell normalfördelad* med väntevärdesvektorn μ och kovariansmatrisen $\mathbf{B}$.

Anmärkningar (1) Om $n = 1$ så är $\mathbf{X} = (X_1, \ldots, X_n)^T = X_1$ och $E[\mathbf{X}] = E[X_1]$, $Cov_{\mathbf{X}} = Var(X_1)$. Om $n = 2$ så är

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad E[\mathbf{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix}, \quad Cov_{\mathbf{X}} = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix}.$$

(2) Om $\mathbf{X} = (X_1, \ldots, X_n)^T$ har den simultana täthetsfunktionen i (c) så gäller att $E[\mathbf{X}] = \mu$ och $Cov_{\mathbf{X}} = \mathbf{B}$.

Sats Låt $\mathbf{X} = (X_1, \ldots, X_n)^T$ vara n-dimensionell normalfördelad med väntevärdesvektorn $\mu_{\mathbf{X}}$ och kovariansmatrisen $Cov_{\mathbf{X}}$. Låt $b \in \mathbb{R}^k$ och $\mathbf{A}$ vara en $k \times n$-matris. Då är linjärkombinationen

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + b$$

k-dimensionell normalfördelad med väntevärdesvektorn

$$\mu_{\mathbf{Y}} = \mathbf{A}\mu_{\mathbf{X}} + b$$

och kovariansmatrisen

$$Cov_{\mathbf{Y}} = \mathbf{A} Cov_{\mathbf{X}} \mathbf{A}^T.$$

Exempel / Lektionsuppgift Låt $\mathbf{X} = (X_1, X_2)^T$ vara 2-dimensionell normalfördelad med väntevärdesvektorn $(1, 0)^T$ och kovariansmatrisen

$$Cov_{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Bestäm $P(X_1 > X_2)$. *Ledning:* Låt $Y = (1, -1)\mathbf{X}$. Då är $P(X_1 > X_2) = P(Y > 0)$. Använd senaste satsen.

Lösning.

$$P(X_1 > X_2) = P(X_1 - X_2 > 0) = P\left(\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} > 0 \right) = P(Y > 0)\,.$$

Med

$$\mathbf{A} = [\,1 - 1\,], \quad b = 0 \quad \text{och} \quad \mu_{\mathbf{X}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

får vi

$$\mu_{\mathbf{Y}} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 \quad \text{och} \quad Cov_{\mathbf{Y}} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2$$

och således $Y \sim N(1, \sqrt{2})$. Detta medför att

$$\begin{aligned} P(X_1 > X_2) &= P(Y > 0) = P\left(\frac{Y-1}{\sqrt{2}} > \frac{0-1}{\sqrt{2}} \right) \\ &= P\left(Z > -\frac{1}{\sqrt{2}} \right) = \Phi(0.707) = 0.76 \end{aligned}$$

där $Z \sim N(0, 1)$.

Exempel / Lektionsuppgift Låt X_1, X_2 vara oberoende $N(0, 1)$-fördelade stokastiska variabler och låt $\alpha \in [0, 2\pi)$. Låt den stokastiska vektorn $\mathbf{Y}$ defineras genom

$$\mathbf{Y} \equiv \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{bmatrix} \sin\alpha & \cos\alpha \\ -\cos\alpha & \sin\alpha \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

(a) Bestäm väntevärdesvektorn och kovariansmatrisen för $\mathbf{Y}$.

Ledning: Använd $\sin^2\alpha + \cos^2\alpha = 1$.

(b) Ange fördelningen av $Y_1 + Y_2$ (namnet av fördelningen och parametrarna).

Lösning.

(a) Låt

$$\mathbf{A} = \begin{bmatrix} \sin\alpha & \cos\alpha \\ -\cos\alpha & \sin\alpha \end{bmatrix} \quad \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Det gäller att $\mathbf{Y} = \mathbf{A}\mathbf{X}$. Det medför att

$$\mu_{\mathbf{Y}} = \mathbf{A} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

och

$$\begin{aligned} \mathbf{C}_{\mathbf{Y}} &= \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T = \begin{bmatrix} \sin\alpha & \cos\alpha \\ -\cos\alpha & \sin\alpha \end{bmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} \sin\alpha & -\cos\alpha \\ \cos\alpha & \sin\alpha \end{bmatrix} \\ &= \begin{bmatrix} \sin\alpha & \cos\alpha \\ -\cos\alpha & \sin\alpha \end{bmatrix} \begin{bmatrix} \sin\alpha & -\cos\alpha \\ \cos\alpha & \sin\alpha \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

(b) Som linjärkombination av normalfördelade stokastiska variabler är $Y_1 + Y_2$ också normalfördelad.

$$\begin{aligned} &E[Y_1 + Y_2] = E[Y_1] + E[Y_2] = 0 \\ &V(Y_1 + Y_2) = V(Y_1) + V(Y_2) + 2Cov(Y_1, Y_2) = 1 + 1 + 2 \cdot 0 = 2\,, \end{aligned}$$

dvs.

$$Y_1 + Y_2 \sim N(0, \sqrt{2})\,.$$

Jörg-Uwe Löbus, jorg-uwe.lobus@liu.se