Uppgift 1

En teknolog kastar sex stycken 12-sidiga tärningar samtidigt. Beräkna sannolikheten att hon/han får en stege, dvs att resultatet blir

$\{1+m, 2+m, 3+m, 4+m, 5+m, 6+m\}$

för något $m \in \{0,1,2,3,4,5,6\}$.

Lösning:

Sökt: $P(\text{stege})$

Kom ihåg: $P(\text{stege}) = \dfrac{\#\text{ kombinationer av stege}}{\#\text{ kombinationer}}$

Två synsätt:

1) Med hänsyn till ordning (t.ex. $\{1,3,2,4,5,6\}$ och $\{1,2,3,4,5,6\}$ är två olika kombinationer)

2) Utan hänsyn till ordning (t.ex. $\{1,3,2,4,5,6\}$ och $\{1,2,3,4,5,6\}$ är en kombination)

Vi beräknar $P(\text{stege})$ på båda sätten.

Notera att det finns 7 olika stegar.

1) Med hänsyn till ordning.

komb.: Eftersom vi tar hänsyn till ordning så blir

$$\#\text{ komb.} = 12^6$$

(varje tärning har 12 sidor och vi har 6 tärningar)

komb. med stege: Vi mult. antal olika stegar med antalet olika sätt att få varje stege

$$\#\text{ komb. med stege} = \underbrace{7}_{\text{antal stegar}} \cdot \underbrace{6!}_{\text{antal olika sätt att få varje stege}}$$

Detta ger oss nu

$$P(\text{stege}) = \frac{\#\text{ komb. med stege}}{\#\text{ komb.}} = \boxed{\frac{7\cdot 6!}{12^6}}$$

2) Utan hänsyn till ordning.

komb: Jämfört med 1) så måste vi "exkludera" permutationer av varje kast, så vi dividerar 12^6 med $6!$ (varje kast kan fås på $6!$ olika sätt)

$$\#\text{ komb.} = \frac{12^6}{6!}$$

komb. med stege: Eftersom vi inte tar hänsyn till ordning, så finns det bara ett sätt att få varje stege. Det finns 7 olika stegar, så vi får

$$\#\text{ komb. med stege} = 7\cdot 1 = 7$$

Detta ger oss nu

$$P(\text{stege}) = \frac{\#\text{ komb. med stege}}{\#\text{ komb.}} = \frac{7}{12^6/6!} = \boxed{\frac{7\cdot 6!}{12^6}}$$

Uppgift 2

Låt X och Y vara två oberoende exponentialfördelade slumpvariabler med väntevärde 1. Visa att slumpvariabeln $Z = X + Y$ har täthetsfunktion

$$f_Z(z) = ze^{-z}, \quad z > 0$$

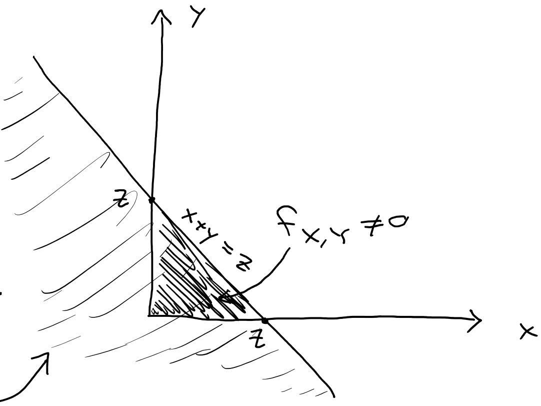
Lösning:

Sökt: $f_Z(z)$.

Kom ihåg: $f_Z(z) = \frac{d}{dz} P(Z \le z)$

Vi betraktar $P(Z \le z)$

$$P(Z \le z) = P(X + Y \le z) \overset{\text{def.}}{=} \iint\limits_{\underbrace{\{(x,y) \in \mathbb{R}^2 : x+y \le z\}}_{\text{se bild}}} f_{X,Y}(x,y)\,dx\,dy$$



X och Y oberoende $\overset{\text{Def. 5.1.5}}{\Longrightarrow} f_{X,Y}(x,y) = f_X(x) f_Y(y)$

X och Y exp.-fördelade med väntevärde 1 $\Rightarrow f_X(x) = e^{-x}, \ x > 0, \quad f_Y(y) = e^{-y}, \ y > 0$

$$\Rightarrow f_{X,Y}(x,y) = e^{-x}e^{-y}, \quad x > 0, \ y > 0$$

Så detta ger oss

$$\iint\limits_{\{(x,y) \in \mathbb{R}^2 : x+y \le z\}} f_{X,Y}(x,y)\,dx\,dy \overset{\text{se bild}}{=} \int_0^z \left(\int_0^{z-y} e^{-x}\,dx \right) \cdot e^{-y}\,dy = \int_0^z \left[\frac{e^{-x}}{-1} \right]_{x=0}^{x=z-y} \cdot e^{-y}\,dy =$$

$$= \int_0^z \Big(\underbrace{-e^{-(z-y)}}_{= e^{-z}e^{y}} + \underbrace{e^{0}}_{=1} \Big) e^{-y}\,dy = \int_0^z \left(e^{-y} - e^{-z}\cancel{e^{y}}\cancel{e^{-y}} \right) dy =$$

$$= \left[\frac{e^{-y}}{-1} \right]_{y=0}^{y=z} - \left[ye^{-z} \right]_{y=0}^{y=z} = -e^{-z} + e^{0} - ze^{-z} + 0 =$$

$$= 1 - e^{-z} - ze^{-z}$$

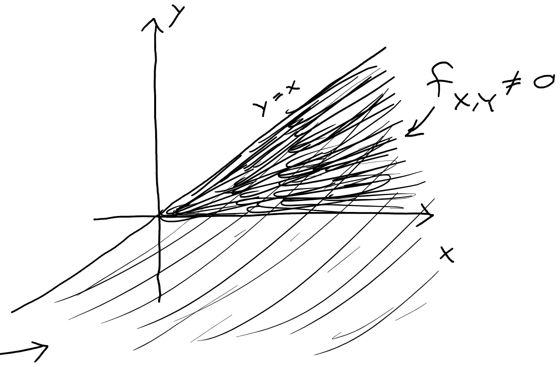
Vi deriverar m.a.p. z

$$f_Z(z) = \frac{d}{dz} P(Z \le z) = \frac{d}{dz}\left(1 - e^{-z} - ze^{-z}\right) = \cancel{e^{-z}} - \cancel{e^{-z}} + ze^{-z} = \boxed{ze^{-z}}$$

Uppgift 3

Beräkna $P(X>Y)$ för en tvådimensionell kontinuerlig slumpvektor (X,Y) med täthetsfunktion

$$f_{X,Y}(x,y) = \frac{1}{2}(x+y)e^{-x-y}, \quad x>0,\ y>0.$$



Lösning:

Sökt: $P(X>Y)$.

$$P(X>Y) \overset{\text{def.}}{=} \iint\limits_{\underbrace{\{(x,y)\in\mathbb{R}^2:\ x>y\}}} f_{X,Y}(x,y)\,dx\,dy =$$

se bild

$$= \int_0^\infty \left(\int_y^\infty \frac{1}{2}(x+y)e^{-x}e^{-y}\,dx \right) dy = \frac{1}{2}\int_0^\infty \left(\int_y^\infty xe^{-x} + ye^{-x}\,dx \right) e^{-y}\,dy =$$

$$= \frac{1}{2}\int_0^\infty e^{-y}\cdot \Big[\underbrace{-e^{-x}(x+1)}\Big]_{x=y}^{x=\infty} + y\cdot e^{-y}\cdot \left[\frac{e^{-x}}{-1}\right]_{x=y}^{x=\infty} dy =$$

$$\frac{d}{dx}(-e^{-x}(x+1)) =$$
$$= -(-e^{-x}\cdot x + \cancel{e^{-x}}\cdot 1 - \cancel{e^{-x}})$$
$$= xe^{-x}$$

$$= \frac{1}{2}\int_0^\infty e^{-y}\cdot\Big(-\underbrace{\lim_{x\to\infty} xe^{-x}}_{=0} - \underbrace{\lim_{x\to\infty} e^{-x}}_{=0} + e^{-y}(y+1)\Big) + ye^{-y}\cdot\Big(-\underbrace{\lim_{x\to\infty} e^{-x}}_{=0} + e^{-y}\Big)\,dy =$$

$$= \frac{1}{2}\int_0^\infty e^{-2y}(y+1) + ye^{-2y}\,dy = \frac{1}{2}\int_0^\infty 2ye^{-2y} + e^{-2y}\,dy =$$

$$= \frac{1}{2}\Big[\underbrace{-\frac{1}{2}e^{-2y}(2y+1)}\Big]_{y=0}^{y=\infty} \underbrace{+ \frac{1}{2}\left[\frac{e^{-2y}}{-2}\right]_{y=0}^{\infty}} =$$

$$= +\Big(\frac{1}{\cancel{2}}\cdot(\cancel{+2})e^{-2y}\cdot 2y + \underbrace{\frac{1}{2}\cancel{e^{-2y}}\cdot 2}_{=e^{-2y}} + \underbrace{\frac{1}{2}\cdot(\cancel{-2})e^{-2y}}_{=-e^{-2y}}\Big) =$$
$$= 2ye^{-2y}$$

$$= \frac{1}{4}\cdot\Big(-\underbrace{\lim_{y\to\infty} 2ye^{-2y}}_{=0} - \underbrace{\lim_{y\to\infty} e^{-2y}}_{=0} + 1 - \underbrace{\lim_{y\to\infty} e^{-2y}}_{=0} + 1\Big) =$$

$$= \frac{1}{4}\cdot 2 = \boxed{\frac{1}{2}}$$

Uppgift 4

Redogör för den/de viktiga skillnaden/skillnaderna mellan parametriska och icke-parametriska statistiska testmetoder.

Lösning:

Parametriska testmetod: Antar någon fördelning för datan. T.ex. normal, T, χ^2, F (alternativt använda approximation; t.ex. CLT)

I praktiken, fungerar bra, t.ex., för stora stickprov.

Icke-parametriska testmetoder: Antar inte någon specifik fördelning.

I praktiken, fungerar bra för både stora och små stickprov.

Fungerar nästan lika bra som para. testmetoder när antagandena för para. är uppfyllda.

Fungerar betydligt bättre än para. när antagandena för para. inte är uppfyllda.

Se Kap. 8.7 och kap. 10.6 för exempel på icke-parametriska testmetoder.

Se Kap 8.7 och Kap. 10.6 i boken för fler detaljer.

Uppgift 5

En teknolog har hittat en tiokrona och vill undersöka om myntet är äkta genom att statistiskt testa om sannolikheten att få krona är lika stor som sannolikheten att få klave vid slantsingling. Beskriv i detalj hur en korrekt sådant statistiskt test kan utföras.

Lösning:

Låt p = sannolikhet för klave.

Vi använder ett två-sidigt test

$$H_0: p = p_0 := 0.5$$

$$H_1: p \neq p_0 = 0.5$$

Vi vill göra följande steg

1. Bestämma signifikansnivå α, stickprovsstorlek n.

2. Antag H_0 är sann: $p = p_0$.

3. Om n är litet, då använder vi binomialfördelning för testet
   Om n är stort, då använder vi normalfördelning (som approx.) för testet

4. Låt
$$X_i = \begin{cases} 1, & \text{Sannolikhet } p \\ 0, & \text{Sannolikhet } 1-p \end{cases}$$
   Vi låter
$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

5. Använd fördelningen från 3 för att hitta gränser L_1 och L_2 sådant att
$$P(L_1 \leq \hat{p} \leq L_2) = 1 - \alpha \qquad (\Leftrightarrow \quad P(\hat{p} \notin [L_1, L_2]) = \alpha)$$
   under antagande av H_0 (oftast här man behöver använda tabell).

6. Om vi observerar realisering av $\hat{p}$ som inte ligger i intervallet $[L_1, L_2]$ så förkastar vi H_0.

Se 9.2 för fler detaljer.

Uppgift 6

Redogör för antaganden och härledningar som ligger bakom linjär regression. Härledningarna behöver ej utföras i detalj utan behöver endast beskrivas principiellt.

Lösning:

Se Kap. 11.1 - 11.3. Några viktiga aspekter

Data: (y_i, x_i), $i = 1, \ldots, n$

Enkel linjär reg.: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, man brukar anta att $E(\varepsilon_i) = 0$

Mål: Skatta β_0 och β_1.

Ett sätt att skatta β_0 och β_1 är Minstakvadratmetoden

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Idé för härledning: Minimera kvadraten av felen; dvs, derivera summan av kvadraterna m.a.p. b_1 och b_0 och sätt bägge lika med noll. Lös för b_1 och b_0.

Om vi dessutom antar att $\underbrace{\varepsilon_i \sim \text{Normal}(0, \sigma^2)}_{\Leftrightarrow\, Y_i \sim \text{Normal}(\beta_0 + \beta_1 x_i,\, \sigma^2)}$ då kan vi skriva ned fördelningarna för B_1 och B_0 (b_1 och b_0 är realiseringar av B_1 och B_0)

$$B_1 \sim \text{Normal}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

$$B_0 \sim \text{Normal}\left(\beta_0, \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})}\sigma^2\right)$$

Idé för härledning: Normalfördelning följer från att man kan skriva

$B_1 = c_1 Y_1 + \ldots + c_n Y_n$

$\Rightarrow B_1 \sim \text{Normal}$

$B_0 = \bar{Y} - B_1 \bar{x}$

$\Rightarrow B_0 \sim \text{Normal}$

Detta ger oss möjligheten att bestämma konfidensintervall och testa hypoteser på parametrarna.