



Ugeseddel 4 - Sandsynlighedsteori og Statistisk

Levi van Boekel

Opgave 1

Vi betragter fem terningekast.

1.1

Lad den stokastiske variabel X tælle antallet af seksere på en given terning, hvorfor den er binomialfordelt, dvs. $X \sim \text{Bin}(5, \frac{1}{6})$. Ssh. for at få netop 3 seksere er da:

$$p(3) = \binom{5}{3} \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^2 \approx 0.032$$

1.2

Sandsynligheden for at få mindst 3 seksere kan bestemmes som:

$$\begin{aligned} p(x \geq 3) &= 1 - p(0) - p(1) - p(2) = \binom{5}{0} \left(\frac{1}{6}\right)^0 \cdot \left(\frac{5}{6}\right)^5 + \binom{5}{1} \left(\frac{1}{6}\right)^1 \cdot \left(\frac{5}{6}\right)^4 \\ &= + \binom{5}{2} \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^3 = 1 - 0.401 - 0.401 - 0.161 = 0.0355 \end{aligned}$$

1.3

I opgave 1.1 fandt vi ssh. for at få 3 seksere. Da terningen er ligefordelt, er ssh. for at få 3 af de andre øjne det samme. Hvis vi lader Y være en stokastisk variabel, der antallet af gange de samme øjne optræder, har vi da:

$$p(Y = 3) = 6 \cdot 0.032 = 0.192$$

1.4

I opgave 1.2 fandt vi ssh. for at få mindst 3 seksere. Vi kan altså igen bare gange resultatet med 6, idet terningen er ligefordelt og vi dermed opnår sandsynligheden for at få mindst 3 ens.

$$p(Y \geq 3) = 0.0355 \cdot 6 = 0.213$$

Opgave 2

Vi antager, at den stokastiske vektor, (X_1, X_2) har følgende sandsynlighedsfordeling:

Tabel 1: Simultan sandsynlighedsfunktion.

		x_1			
		-1	0	2	6
x_2	3	0	0	1/9	4/27
	1	2/9	0	1/9	1/9
	-2	1/9	1/27	1/27	1/9

Figur 1: Sandsynlighedsfordeling for (X_1, X_2)

2.1

Sandsynligheden for at X_1 er et lige tal er givet ved, hvor den komplementære ssh. bestemmes for at have færre beregninger.

$$p(X_1 = \text{lige}) = 1 - p(X_1 = -1) = 1 - \frac{3}{9} = \frac{6}{9} = \frac{2}{3}$$

2.2

Sandsynligheden for at produktet af de to stokastiske variable er et ulige tal, aflæses af tabellen:

$$p(X_1 \cdot X_2 = \text{ulige}) = p(X_1 = -1, X_2 = 3) + p(X_1 = -1, X_2 = 1) = \frac{2}{9}$$

2.3

Den simultane ssh. findes ved at summere i tabellen

$$p(X_2 > 0, X_1 \geq 0) = \frac{1 + 3 + 3 + 1 + 4 + 3 + 3}{27} = \frac{13}{27}$$

2.4

Den forventede værdi af X_1^2 findes nedenfor:

$$\begin{aligned} \mathbb{E}(X_1^2) &= \sum_{x_1 \in \{-1, 0, 2, 6\}} p(x_1) x_1^2 \\ &= 1 \cdot \frac{9}{27} + 4 \cdot \frac{7}{27} + 36 \cdot \frac{10}{27} \\ &= \frac{9 + 28 + 360}{27} = 14.7 \end{aligned}$$

2.5

Kovariansen beregnes ved:

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \quad (1)$$

Vi starter med at betragte det første led:

$$\mathbb{E}(X_1 X_2) = -\frac{2}{9} + 2\frac{2}{9} + 6\frac{1}{9} + 2\frac{1}{9} - 4\frac{1}{27} + 18\frac{4}{27} + 6\frac{1}{9} - 12\frac{1}{9} \approx 2.74$$

Og bestemmer herefter de to andre:

$$\begin{aligned} \mathbb{E}(X_1) &= -1\frac{3}{9} + 2\frac{7}{27} + 6\frac{10}{27} \approx 2.41 \\ \mathbb{E}(X_2) &= 3\frac{7}{27} + 1\frac{4}{9} - 2\frac{8}{27} \approx 0.62 \end{aligned}$$

Kovariansen er da:

$$\text{Cov}(X_1, X_2) = 2.74 - 2.407 \cdot 0.62 \approx 1.2 \quad (2)$$

2.6

Den betingede middelværdi findes ved, hvor $i = \{-1, 0, 2, 6\}$

$$\mathbb{E}(X_1 \mid X_2 = 1) = \sum_{i=1}^4 x_i \cdot p(x_i \mid X_2 = 1) \quad (3)$$

Da:

$$\begin{aligned} \mathbb{E}(X_1 \mid X_2 = 1) &= -1 \cdot \frac{p(X_2 = -1, X_2 = 1)}{\frac{4}{9}} + 2 \cdot \frac{p(X_1 = 2, X_2 = 1)}{\frac{4}{9}} + 6 \cdot \frac{p(X=6, X_2 = 1)}{\frac{4}{9}} \\ &= \frac{(-1 \cdot \frac{2}{9} + 2 \cdot \frac{1}{9} + 6 \cdot \frac{1}{9})}{\frac{4}{9}} \\ &= 1.5 \end{aligned}$$

Sandsynligheden for at x_1 antager den betingede middelværdi er 0, da den ikke er en del af udfaldsrummet.

2.7

Vi har følgende hvor $i = \{-2, 1, 3\}$

$$\mathbb{E}(X_2 \mid X_1 = 2) = \sum_{i=1}^4 x_i \cdot p(x_i \mid X_1 = 2) \quad (4)$$

Altså:

$$\mathbb{E}(X_2 \mid X_1 = 2) = \frac{-2 \cdot \frac{1}{27} + \frac{1}{9} + \frac{3}{9}}{\frac{7}{27}} \approx 1.42$$

Opgave 3

Agent A står ved en trafikeret vej og håber på at fange en ledig taxa. Der kommer i gennemsnit en hver halve time. Vi antager derfor, at antal taxaer pr. minut er Poissonfordelt med parameter $\frac{1}{30}$, og at disse antal er uafhængige fra minut til minut.

3.1

Vi ved, at taxaerne pr. minut er Poissonfordelt med parameteren $\lambda = \frac{1}{30}$. Hvis vi nu vil generalisere fordelingen til at tage højde for t -minutter, kan vi blot skrive: $\lambda = t \frac{1}{30}$. Det gælder, idet vi kan opfatte taxaerne i første minut som en stokastisk variabel X_1 med $\lambda = \frac{1}{30}$, taxaerne i andet minut som en anden (uafhængig) stokastisk variabel, $X_2 = X_1 + \frac{1}{30}$. Poisson fordelingsgens foldningsegenskab sikrer nu, at summen af de to blot er 2λ , hvorfor vi generelt kan skrive, λt . I vores tilfælde ønsker vi at bestemme ssh. for at, at der ikke kommer nogen taxa i 30 minutter. Vi indsætter da $t = 30$ og får:

$$p(x_{30} = 0) = \frac{\left(\frac{1}{30}30\right)^0}{0!} e^{-\frac{1}{30}30} = e^{-1} \approx 0.36$$

3.2

Sandsynligheden for at A må vente mere end 90 minutter beregnes tilsvarende:

$$p(x_{90} = 0) = \frac{\left(\frac{1}{30}90\right)^0}{0!} e^{-\frac{1}{30}90} = e^{-3} \approx 0.04$$

3.3

Sandsynligheden for at A får en taxa allerede inden, der er gået 10 minutter, kan bestemmes som den komplementære ssh. til, at A ikke får en Taxa, inden der er gået 10 minutter. Da:

$$p(x_{10} > 0) = 1 - \frac{\left(\frac{1}{30}10\right)^0}{0!} e^{-\frac{1}{30}10} = 1 - e^{-\frac{1}{3}} \approx 0.283$$

Opgave 4

4.1

Vi lader Y være en stokastisk variabel, der angiver antallet af cykler, der bliver stjålet blandt de 10 forsikringstagere. Det gælder altså, at: $Y \sim \text{Bin}(10, \frac{1}{20})$, fordi vi netop har antaget, at udfaldene er indbyrdes uafhængige og har samme ssh. Særligt antagelsen om, at alle har samme ssh. virker ikke plausibel. Fordelingen er simuleret i STATA:

k	pmf	cdf
0	.5987369	.5987369
1	.3151247	.9138616
2	.0746348	.9884964
3	.0104751	.9989715
4	.0009648	.9999363
5	.0000609	.9999973
6	2.67e-06	.9999999
7	8.04e-08	1
8	1.59e-09	1
9	1.86e-11	1
10	9.77e-14	1

Figur 2: Sandsynlighedsfordeling for Y

4.2

For en binomialfordelt stokastisk variabel er middelværdien blot produktet af sandsynlighedsparameteren og antalsparameteren. Altså:

$$\mathbb{E}(Y) = 10 \cdot 0.05 = 0.5 \quad (5)$$

Den forventede nettoprofit er da: $10 \cdot 400 - 0.5 \cdot 4000 = 2000$.

4.2

Sandsynligheden for at mere end en cykel bliver stjålet kan vi bestemme som den komplementære hændelse til, at netop 0 eller en cykel bliver stjålet. Da:

$$P(Y > 1) = 1 - \binom{10}{0} \left(\frac{1}{20}\right)^0 \left(\frac{19}{20}\right)^{10} + \binom{10}{1} \left(\frac{1}{20}\right)^1 \left(\frac{19}{20}\right)^9 \approx 0.0861 \quad (6)$$

4.3

Hvis der i stedet er 100, der har tegnet en forsikring, er de forventede udgifter: $\mathbb{E}(U) = 100 \cdot 0,05 \cdot 4000 = 20.000$, mens indtægterne er: $\mathbb{E}(I) = 100 \cdot 400 = 40.000$, nettooverskuddet er da 20.000.

4.4

Vi skal nu approksimere ssh. for at $Y > 10$. Vi kan opskrive vores stokastiske variabel som tilnærmelsesvis Poisson fordelt $\lambda = 100 \cdot 0,05 = 5$, idet den forventede værdi netop er lig λ .

$$p(Y > 10) = 1 - \sum_{x=0}^{10} \frac{5^x}{x!} \exp(-5) = 1 - 0,98631 \approx 0,01369$$

Hvis vi i stedet bruger Binomialfordelingen, skal vi bestemme:

$$P(Y > 10) = 1 - \sum_{i=1}^{10} \binom{100}{x_i} \left(\frac{1}{20}\right)^{x_i} \left(\frac{1}{20}\right)^{100-x_i} = 1 - 0,98853 = 0,01147$$

4.5

Hvis antallet af forsikringstagere er 200, er indtægterne givet ved 80.000, hvis udgifterne skal overstige dette, skal der være flere end 20 skader. Vi skal altså bestemme, hvor $\lambda = 200 \cdot 0,05 = 10$. Da:

$$p(Y > 20) = 1 - \sum_{x=0}^{20} \frac{10^x}{x!} \exp(-10) = 1 - 0,99842 \approx 0,00158$$

Hvis vi i stedet bruger Binomialfordelingen, skal vi bestemme:

$$P(Y > 20) = 1 - \sum_{i=1}^{20} \binom{200}{x_i} \left(\frac{1}{20}\right)^{x_i} \left(\frac{1}{20}\right)^{200-x_i} \approx 1 - 0,99884 = 0,00116$$

Opgave 5**5.1**

Vi får oplyst, at $P(Y_{30} = 0) = 0,1$. Altså:

$$\begin{aligned} 0,1 &= \frac{\lambda^0}{0!} e^{-30\lambda} \\ \ln(0,1) &= -30\lambda \\ \lambda &= -\frac{\ln(0,1)}{30} \approx 0,077 \end{aligned}$$

Den daglige ankomstintensitet er da lig 0,077, som svarer til $\mathbb{E}(Y_1)$, altså det forventede antal jobtilbud om dagen.

5.2

De to andre typer har en daglig ankomstintensitet på:

$$\begin{aligned} \lambda_2 &= \frac{-\ln(0,2)}{30} \approx 0,054 \\ \lambda_3 &= \frac{-\ln(0,05)}{30} \approx 0,099 \end{aligned}$$

Hvis vi antager, at der er uafhængighed mellem jobtilbudene, kan vi opstille en model for det samlede antal af jobtilbud som de tre arbejdsløse vil modtage ved: $Z \sim \text{Pois}((0.077 + 0.054 + 0.099)t)$. Sandsynligheden for, at der ikke kommer noget jobtilbud indenfor 10 dage til de tre arbejdsløse bliver da:

$$P(Z_{10} = 0) = e^{-(0.077+0.054+0.099) \cdot 10} \approx 0.1$$

5.3

Sandsynligheden for, at de forskellige typer (nu kaldt A, B, C) ikke får et job efter 60 dage er:

$$P(A_{60} = 0) = e^{-0.077 \cdot 60} \approx 0.0099$$

$$P(B_{60} = 0) = e^{-0.054 \cdot 60} \approx 0.0391$$

$$P(C_{60} = 0) = e^{-0.099 \cdot 60} \approx 0.0026$$

Andelen er kan da blot findes ved at dividere de respektive sandsynligheder med summen af sandsynlighederne:

$$\text{Andel A} = \frac{0.0099}{0.0516} \approx 0.192$$

$$\text{Andel B} = \frac{0.0391}{0.0516} \approx 0.758$$

$$\text{Andel C} = \frac{0.0026}{0.0516} \approx 0.05$$