Statistikgruppen
Handelshøjskolen i København

Statistik for MØK 2. år
Januar 2010
Tue Tjur

Besvarelse af midtvejsprøven

Opgave 1.

Ifølge regnereglen (c), Ssr. side 57, er

$$\mathrm{E}\,(X_1X_2) = \mathrm{E}\,(X_1)\,\mathrm{E}\,(X_2) = 3.5^2 = 12.25.$$

Man kan selvfølgelig også beregne middelværdien af X_1X_2 direkte som $\frac{1}{36}\sum_{x_1=1}^{6}\sum_{x_2=1}^{6} x_1x_2$, hvis man vil ha tiden til at gå.

Opgave 2.

$P(X \geq 2 \text{ og } X \leq 5) = P(X \in \{2,3,4,5\})$

$$= e^{-4}\left(\frac{4^2}{1\times 2} + \frac{4^3}{1\times 2\times 3} + \frac{4^4}{1\times 2\times 3\times 4} + \frac{4^5}{1\times 2\times 3\times 4\times 5}\right)$$

$$= 0.69355.$$

Opgave 3.

Punktsandsynlighederne for den geometriske fordeling med $p = 0.5 = \frac{1}{2}$ er

$$\frac{1}{2}\,,\ \left(\frac{1}{2}\right)^2\,,\ \left(\frac{1}{2}\right)^3\ldots$$

så vi får

$$P\,(X_1 = X_2) = P\,((X_1,X_2) \in \{(0,0),(1,1),(2,2),\ldots\})$$

$$= \left(\frac{1}{2}\right)\times\left(\frac{1}{2}\right) \ + \ \left(\frac{1}{2}\right)^2\times\left(\frac{1}{2}\right)^2 \ + \ \left(\frac{1}{2}\right)^3\times\left(\frac{1}{2}\right)^3 \ + \ldots$$

$$= \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \ldots$$

$$= \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3}.$$

Opgave 4.

(a) Det er klart at $p(x) \geq 0$, og da

$$\int_1^{+\infty} 2x^{-3}\,dx = 2\left[\frac{x^{-2}}{-2}\right]_1^{+\infty} = 2\left(0 - \frac{1}{-2}\right) = 1$$

er p en sandsynlighedstæthed. Middelværdien af X bliver

$$2\int_1^{+\infty} x^{-3}x\,dx = 2\int_1^{+\infty} \frac{1}{x^2}\,dx = 2\left[\frac{-1}{x}\right]_1^{+\infty} = 2\left(0 - \frac{-1}{1}\right) = 2.$$

(b) Vi sætter $Y = t(X) = 1/X$. Transformationen t afbilder intervallet $]1, +\infty[$ aftagende og kontinuert differentiabelt på intervallet $]0, 1[$, så tætheden $q(y)$ for Y bliver for $y \in]0, 1[$

$$q(y) = p(x) \times \mid \frac{dx}{dy} \mid = p(1/y)\frac{1}{y^2} = \frac{2y^3}{y^2} = 2y$$

Tætheden for Y bliver derfor

$$q(y) = \begin{cases} 2y \text{ for } y \in]0, 1[\\ 0 \text{ ellers} \end{cases}$$

Opgave 5.

(a) Ifølge almindelige regneregler for middelværdier og varianser er

$$\mathrm{E}S = 100\,\mathrm{E}X + 10\,\mathrm{E}Y + \mathrm{E}Z = (100 + 10 + 1) \times 4.5 = 499.5.$$

Da

$$\mathrm{var}(X) = \mathrm{var}(Y) = \mathrm{var}(Z) = \mathrm{E}\left(X^2\right) - (\mathrm{E}X)^2 =$$

$$\frac{1}{10}(1 + 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81) - 4.5^2 = 8.25$$

får vi

$$\mathrm{var}(S) = 10000\,\mathrm{var}(X) + 100\,\mathrm{var}(Y) + \mathrm{var}(Z) = 10101 \times 8.25 = 83333.25.$$

(b) S kan åbenbart fortolkes som et tal med tre tilfældige (uafhængige og ligefordelte) cifre (nemlig tallet "XYZ"). Det følger umiddelbart at S er ligefordelt på mængden $\{000, 001, 002, \ldots, 999\} = \{0, 1, 2, \ldots, 999\}$.

Opgave 6.

Tallene opstilles naturligt i en tabel af formen

	død	ikke død	I alt
0–39	53	319	372
40–80	157	218	375
I alt	210	537	747

Den mest naturlige model går ud på at opfatte størrelserne 372 og 375 af de to aldersgrupper som givne tal, og antallene 53 og 157 af døde som observationer af to uafhængige binomialfordelte variable med hver sin sandsynlighedsparameter. Testet for om de to sandsynlighedsparametre er ens, svarende til "ingen forskel mellem dødeligheder i aldersgrupperne", giver

$$-2\log q = 72.95$$

(Pearson: 70.49), som er langt større end 99.99 % fraktilen 15.137 i χ^2–fordelingen med 1 frihedsgrad. Konklusion: Der er ekstremt signifikant forskel mellem dødelighederne i de to aldersgrupper, og denne forskel går helt klart på at sygdommen er farligst for de ældste.