Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

TMA4245 Statistikk
Vår 2013

Øving nummer 5, blokk I
Løsningsskisse

Oppgave 1

X og Y er uavhengige Poisson-fordelte stokastiske variable, X ~p(x;5) og Y ~p(y;10). Fra tabell A.2 finner vi

$P(X \leq 5) = \sum_{x=0}^{5} p(x;5) = 0.616$

og

$P(X \leq 3|X \leq 5) = \frac{P(X \leq 3 \cap X \leq 5)}{P(X \leq 5)} = \frac{P(X \leq 3)}{P(X \leq 5)} = 0.430.$

La $Z = X + Y$. X og Y er Poisson-fordelt og dermed vil variabelen Z være Poisson-fordelt med parameter $\lambda = 5 + 10 = 15$, dvs. $Z \sim$ Poisson(15). Ved bruk av tabell A.2 finner vi

$P(X + Y > 10) = P(Z > 10) = 1 - P(Z \leq 10) = 1 - \sum_{x=0}^{10} = 1 - 0.1185 = 0.882.$

Oppgave 2

Definer

X : antall tankskip som ankommer havnen i løpet av en dag

der

$$X \sim \text{poisson}(\lambda),$$

med

$$\lambda = \text{E}(X) = 2.$$

Havnen kan maksimalt betjene 3 tankskip per dag.

a) Da X er poissonfordelt har vi at

$$P(X=x)=\frac{\lambda^x}{x!}e^{-\lambda}=\frac{2^x}{x!}e^{-2},\quad x=0,1,2,\ldots.$$

Med innsatte verdier for x har vi

x	0	1	2	3	4	5
$P(X=x)$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361

Vi ser dermed at det er størst sannsynlighet for at det ankommer ett eller to tankskip en bestemt dag. Tankskip må dirigeres til andre havner dersom det ankommer mer enn tre tankskip en dag. Vi har dermed

$$P(\text{Ett eller flere tankskip må omdirigeres})=P(X>3)$$
$$=1-P(X\le 3)\overset{\text{tabell}}{=}1-0.8571=0.1429$$

b) Definer

$$Y:\ \text{antall skip som betjenes en dag.}$$

Da $Y\le 3$ har vi at

$$\begin{aligned}P(Y=y)&=P(X=x),\ \ y=0,1,2\\ P(Y=3)&=P(X\ge 3)=1-P(X\le 2).\end{aligned}$$

Vi får dermed

$$\begin{aligned}\mathrm{E}[Y]&=\sum_{y=0}^{3}y\,P(Y=y)\\ &=0\cdot P(X=0)+1\cdot P(X=1)+2\cdot P(X=2)+3\cdot(1-P(X\le 2))\\ &=0.2707+2\cdot 0.2707+3\cdot(1-0.6767)=1.782.\end{aligned}$$

c) Definer

$$k:\ \text{kapasitet.}$$

Vi ønsker å finne k slik at $P(X\le k)\ge 0.90$. Med innsatte verdier for k får vi

k	0	1	2	3	4	5
$P(X\le k)$	0.1353	0.4060	0.6767	0.8571	0.9474	0.9834

Vi har dermed at

$$P(X\le 4)=0.9474>0.90,$$

og vi må dermed bygge ut havnen til kapasitet på fire skip for med minst 90% sannsynlighet å kunne betjene samtlige skip som ankommer en gitt dag.

Oppgave 3

Vi antar X er normalfordelt, $X \sim N(3315, 575^2)$. Ved bruk av tabell A.3 finner vi

a) 1)

$$\begin{aligned}
P(X > 3000) &= 1 - P(X < 3000) \\
&= 1 - P(Z < \frac{3000 - 3315}{575}) \\
&= 1 - P(Z < -0.55) \\
&= 1 - 0.2912 \\
&= 0.709.
\end{aligned}$$

2)

$$\begin{aligned}
P(3000 < X < 3500) &= P(\frac{3000 - 3315}{575} < Z < \frac{3500 - 3315}{575}) \\
&= P(-0.55 < Z < 0.32) \\
&= P(Z < 0.32) - P(Z < -0.55) \\
&= 0.6255 - 0.2912 \\
&= 0.335.
\end{aligned}$$

3)

$$\begin{aligned}
P(X > 3500 | X > 3000) &= \frac{P(X > 3500 \cap X > 3000)}{P(X > 3000)} \\
&= \frac{P(X > 3500)}{P(X > 3000)} \\
&= \frac{P(Z > 0.32)}{P(Z > -0.55)} \\
&= \frac{1 - P(Z < 0.32)}{1 - P(Z < -0.55)} \\
&= \frac{1 - 0.6255}{1 - 0.2912} \\
&= 0.528.
\end{aligned}$$

b) Dersom fødelsvekten er mindre enn c gram, $P(X < c) = 0.01$, vil barnet bli klassifisert som undervektig. La $n = 100$ være antall barn og Y antall av disse barna som var undervektige ved fødselen. Vi antar at fødelsvekten for de ulike barna er uavhengige av hverandre. Fødelsvekten til hvert barn kan da beskrives av en bernoulli variabel med med suksesssannsynlighet (undervektig) $p = P(X < c) = 0.01$. Da vil Y være binomisk fordelt med parametere $n = 100, p = 0.01$, $Y \sim b(y; 100, 0.01)$.

Når Y er en binomisk fordelt variabel med fordeling $b(y; n, p)$, $n \to \infty$ og $p \to 0$, $np \to \mu$ vil vi kunne bruke Poisson-tilnærming til den binomiske fordelingen, dvs $b(y; n, p) \to p(y; \mu)$.

Vi bruker poisson-tilnærmingen $Y \sim p(y;1)$ og finner da ved bruk av tabell A.2

$$\begin{aligned} P(Y>0) &= 1-P(Y<0) \\ &= 1-0.3679 \\ &= 0.63. \end{aligned}$$

$$\begin{aligned} P(Y>1|Y>0) &= \frac{P(Y>1\cap Y>0)}{P(Y>0)} \\ &= \frac{P(Y>1)}{P(Y>0)} \\ &= \frac{1-P(Y<1)}{1-P(Y<0)} \\ &= \frac{1-0.7358}{1-0.3679} \\ &= 0.42. \end{aligned}$$

Oppgave 4

a) The probability that the lamp lights up when the procedure is performed once, is $P(X>25) = P((X-20)/4 > (25-20)/4) = P(Z>1.25) = P(Z<-1.25) = 0.1056$, where Z has the standard normal distribution.

If the procedure is performed three times, let the amounts of by-product be X_1, X_2, X_3. Then the probability that the lamp lights up at least once is $P(X_1 > 25 \cup X_2 > 25 \cup X_3 > 25) = 1 - P(X_1 \le 25,\ X_2 \le 25,\ X_3 \le 25) = 1-(P(X_i \le 25))^3 = 1-(1-P(X_i>25))^3 = 1-(1-0.1056)^3 = 0.285$.

Let Y be the number of times the lamp lights up when the procedure is performed 100 times. Then Y has the binomial distribution with parameters $n=100$ and $p=0.1056$, and

$$\begin{aligned} P(Y\ge 15) = P(Y \ge 14.5) = P\left(\frac{Y-np}{\sqrt{np(1-p)}} \ge \frac{14.5-100\cdot 0.1056}{\sqrt{100\cdot 0.1056\cdot 0.8944}}\right) \\ \approx P(Z>1.28) = P(Z\le -1.28) = 0.100, \end{aligned}$$

using the normal approximation with continuity correction. (The exact binomial probability is 0.104.)

(You are not penalized if you have not applied the continuity correction. If you use 15 instead of 14.5 in the normal approximation, you get 0.074, and if you use 14 (arising from $P(Y\ge 15) = 1-P(Y\le 14)$), you get 0.131 – not very good approximations.)

b) Let n be the number of times the procedure is performed, and let $X_1, X_2, \dots, X_n$ be the amounts of by-product. The total amount of by-product, $Y = \sum_{i=1}^n X_i$, has the normal distribution with mean $20n$ and variance 4^2n. We want to find n such that

$$0.01 \leq P(Y \geq 500) = P\Big(\frac{Y-20n}{4\sqrt{n}} \geq \frac{500-20n}{4\sqrt{n}}\Big) = P\Big(Z \geq \frac{500-20n}{4\sqrt{n}}\Big),$$

that is, $(500-20n)/(4\sqrt{n}) \geq z_{0.01} = 2.326$, or $125-5n \geq z_{0.01}\sqrt{n}$, and we have a quadratic inequality $5n + z_{0.01}\sqrt{n} - 125 \leq 0$ in $\sqrt{n}$. The left hand side is a downward-pointing parabola (as a function of $\sqrt{n}$) having zeros $\big(-z_{0.01} \pm \sqrt{z_{0.01}^2 + 4\cdot 5\cdot 125}\big)/(2\cdot 5)$, that is, -5.24 and 4.77, meaning that $0 \leq \sqrt{n} \leq 4.77$ and $n \leq 22.8$, that is, $n \leq 22$ since n is an integer.

Oppgave 5

En lottorekke kan oppfattes som et ikke-ordnet utvalg på 7 elementer blant tallene 1 til 34, der utvelgingen skjer uten tilbakelegging.

a) Det finnes

$$\binom{34}{7} = 5379616$$

forskjellige lotto-rekker.

Skal finne sannsynligheten for at lottorekka inneholder tallet 34. Denne kan beregnes som

$$P(\text{ lottorekka inneholder tallet 34 }) = 1 - P(\text{ lottorekka inneholder IKKE tallet 34 })$$

der antallet mulige lottorekker er $\binom{34}{7}$ og antallet mulige lottorekker uten tallet 34 er $\binom{33}{7}$ slik at

$$P(\text{ lottorekka inneholder tallet 34 }) = 1 - \frac{\binom{33}{7}}{\binom{34}{7}} = 1 - \frac{27}{34} = \frac{7}{34}.$$

Definer

$$X: \text{ antall rette i lottorekken,}$$

der X følger hypergeometrisk fordeling

$$h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

med $N = 34,\ n = 7$ og $k = 7$.

Sannsynligheten for at en lottorekke vil oppnå 7 rette er gitt ved

$$p = P(\text{ 7 rette }) = h(7; 34, 7, 7) = \frac{\begin{pmatrix} 7 \\ 7 \end{pmatrix}\begin{pmatrix} 27 \\ 0 \end{pmatrix}}{\begin{pmatrix} 34 \\ 7 \end{pmatrix}} = \frac{1}{\begin{pmatrix} 34 \\ 7 \end{pmatrix}} = 1.859 \cdot 10^{-7}.$$

b) Det leveres hver uke inn $n = 19\,000\,000$ rekker i lotto. Definer

$$X: \text{ antall av de } n \text{ innleverte rekkene som oppnår 7 rette i ukens trekning}$$

X er binomisk fordelt, men kan med god tinærmelse regnes å være poissonfordelt da poissonfordelingen er grensefordelingen til binomisk fordeling når $n \to \infty$ og $p \to 0$. Ifølge teorem 5.6 har vi at $n\,p \to \mu$ for $n \to \infty$ slik at parameteren i poissonfordelingen er gitt ved

$$\mu = n\,p = 19\,000\,000 \times 1.859 \cdot 10^{-7} = 3.53.$$

Sannsynlighetstettheten til X er nå gitt ved

$$f_X(x) = \frac{\mu^x}{x!} e^{-\mu}$$

og vi har

$$\begin{aligned} &P(\text{ ingen av de innleverte rekkene har 7 rette }) \\ &= P(X = 0) = \frac{3.53^0}{0!} e^{-3.53} = e^{-3.53} = 0.029. \end{aligned}$$

Legg merke til at vi oppnår samme svar uten poissontilnærmingen, dvs dersom X er binomisk fordelt med parametre $n = 19\,000\,000$ og $p = 1.859 \cdot 10^{-7}$ gitt ved

$$f_X(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x\,(1-p)^{n-x}.$$

Vi har da

$$\begin{aligned} &P(\text{ ingen av de innleverte rekkene har 7 rette }) \\ &= P(X = 0) = \begin{pmatrix} n \\ 0 \end{pmatrix} p^0\,(1-p)^{n-0} = (1-p)^n = (1 - 1.859 \cdot 10^{-7})^{19\,000\,000} = 0.029. \end{aligned}$$

Definer

$$Y: \text{ antall Gull-lotto omganger pr. år}$$

der vi antar at det er 52 uker pr. år. Y følger her en binomisk fordeling med $n = 52$ og $p = 0.029$. Vi har dermed at

$$E(Y) = n\,p = 52 \times 0.029 = 1.5$$

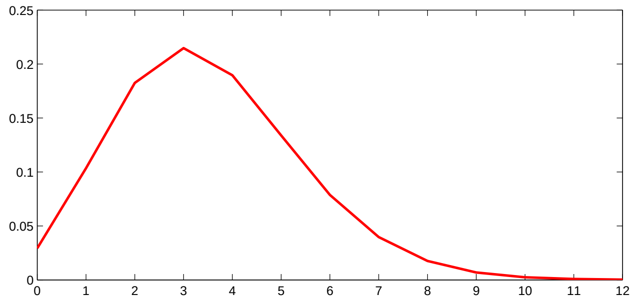
og

$$P(Y = 0) = f_Y(0) = \left(\begin{array}{c} 52 \\ 0 \end{array} \right) p^0\,(1-p)^{52-0} = (1 - 0.029)^{52} = 0.22.$$

c) Eksempel på en mulig kode for Poisson sannsynlighetstetthetsfunksjon:

function funcval = poisson(x,mean)
$funcval = ((mean\hat{\ }x)/(factorial(x)))*(exp(-mean));$

Et plot av verdiene fra Poissontetthetsfunksjonen for x-verdier fra 0 til 12 er vist i Fig 1. Her har vi brukt $\mu = np = 3.53$ som vi fant i Oppgave 5b). Vi bruker heltall ettersom Poisson tetthetsfunksjonen er en diskret sannsynlighetsfordeling som bare er definert for heltallsverdier av x.



Figur 1: Poissontetthetsfunksjonen, $f_X(x) = \frac{\mu^x}{x!}e^{-\mu}$ for heltallsverdier av x fra 0 til 12 med $\mu = 3.53$.

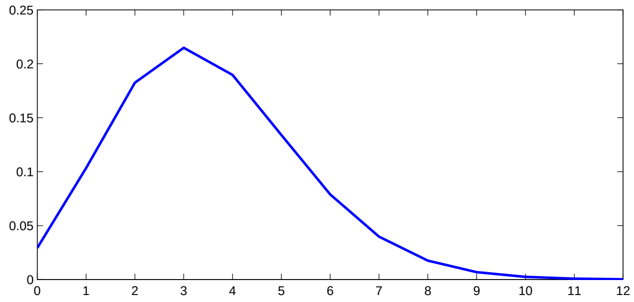
d) Eksempel på en mulig kode for den binomiske sannsynlighetstetthetsfunksjonen:

function funcval = binomial(n,x,p)
$funcval = nchoosek(n,x) * (p\hat{\ }x) * ((1-p)\hat{\ }(n-x));$

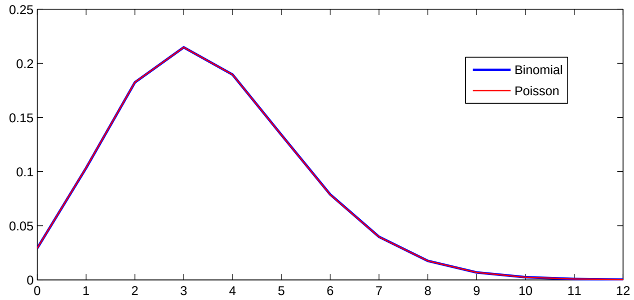
Et plot av verdiene fra den binomiske tetthetsfunksjonen for x-verdier fra 0 til 12 er vist i Fig 2. Her har vi brukt $n = 19\,000\,000$ og $p = 1.859 \cdot 10^{-7}$ som vi fant i Oppgave 5b).

I Fig 3 har vi plottet de to tetthetsfunksjonene, Poisson og binomisk, mot hverandre. Vi ser at tetthetsfunksjonene er tilnærmet identiske ettersom de overlapper hverandre nesten perfekt.

e) Vi skal nå øke antall valgbare tall fra 34 til m, der $m > 34$ slik at det med sannsynlighet minst 0.1 ikke finnes rekker med 7 riktige i en uke der det innleveres $n = 19\,000\,000$ rekker. Vi må i denne oppgaven bruke de tidligere definerte ligningene. Vi har at



Figur 2: Binomisk tetthetsfunksjon, $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for heltallsverdier av x fra 0 til 12 med $n = 19\,000\,000$ og $p = 1.859 \cdot 10^{-7}$.



Figur 3: Poissontetthetsfunksjonen av b) plottet mot den binomiske tetthetsfunksjonen.

$$
\begin{aligned}
&P(\text{ ingen av de innleverte rekkene har 7 rette }) \\
&= P(X = 0) = (1-p)^{19\,000\,000} \geq 0.1
\end{aligned}
$$

Vi har videre at p er sannsynligheten for at en lottorekke vil oppnå 7 rette, og denne er gitt ved

$$
p = P(7 \text{ rette }) = h(7; m, 7, 7) = \frac{\begin{pmatrix} 7 \\ 7 \end{pmatrix} \begin{pmatrix} m-k \\ 0 \end{pmatrix}}{\begin{pmatrix} m \\ 7 \end{pmatrix}} = \frac{1}{\begin{pmatrix} m \\ 7 \end{pmatrix}}
$$

slik at

$$
\left(1 - \frac{1}{\begin{pmatrix} m \\ 7 \end{pmatrix}} \right)^{19\,000\,000} \geq 0.1
$$

Prøver med innsatte verdier av m:

m	35	36
$\left(1 - \frac{1}{\binom{m}{7}}\right)^{19\,000\,000}$	0.059	0.103

Vi ser av tabellen at ulikheten er oppfylt for $m = 36$, og har dermed at $m \geq 36$ dersom det med sannsynlighet minst 0.1 ikke skal finnes rekker med 7 riktige i en uke der det innleveres $n = 19\,000\,000$ rekker.