TMA4245 Statistikk
Vår 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalt øving 5
Løsningsskisse

Oppgave 1

a) La X være massen til et tilfeldig valgt egg, målt i gram. Sannsynligheten for at et tilfeldig valgt egg veier mer enn 60 g er

$$\begin{aligned}
P(X > 60) &= P\left(\frac{X - E(X)}{\sqrt{\text{Var}(X)}} > \frac{60 - 70}{\sqrt{16}}\right) \\
&= 1 - P\left(\frac{X - E(X)}{\sqrt{\text{Var}(X)}} \leq \frac{60 - 70}{\sqrt{16}}\right) \\
&= 1 - \Phi\left(\frac{60 - 70}{4}\right) \\
&= 1 - 0.0062 = \underline{\underline{0.9938}}.
\end{aligned}$$

Hvis vi neglisjerer vekten av emballasjen, så vil vekten Y av en tilfeldig valgt pakke med 6 egg, være lik summen av vekten til hvert av de 6 eggene i pakken. Vi har altså at

$$Y = \sum_{i=1}^{6} X_i$$

hvor X_i, $i = 1, \ldots, 6$ er vekten av egg nummer i. Disse er uavhengige og identisk normalfordelte med $E(X_i) = 70$ og $\text{Var}(X_i) = 16$. Siden Y er en lineærkombinasjon av normalfordelte tilfeldige variable, vil den selv være normalfordelt. Videre vil Y ha forventningsverdi

$$E(Y) = E\left(\sum_{i=1}^{6} X_i\right) = \sum_{i=1}^{6} E(X_i) = 6E(X) = 6 \cdot 70 = 420$$

og varians

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^{6} X_i\right) = \sum_{i=1}^{6} \text{Var}(X_i) = 6\text{Var}(X) = 6 \cdot 16 = 96.$$

Pakkevekten Y er altså normalfordelt med forventningsverdi 420 g og varians 96 g.

b) Vi vil finne et tall k, som er slik at $P(Y \geq k) = 0.95$. Denne finner vi ved å skrive opp uttrykket for sannsynligheten, og løse for k,

$$
\begin{aligned}
P(Y \geq k) &= 0.95 \\
P(Y \leq k) &= 1 - 0.95 \\
P\left(\frac{Y - E(Y)}{\sqrt{\mathrm{Var}(Y)}} \leq \frac{k - E(Y)}{\sqrt{\mathrm{Var}(Y)}}\right) &= 0.05 \\
\Phi\left(\frac{k - E(Y)}{\sqrt{\mathrm{Var}(Y)}}\right) &= 0.05 \\
\frac{k - E(Y)}{\sqrt{\mathrm{Var}(Y)}} &= -z_{0.05}.
\end{aligned}
$$

Med normalkvantilen $z_{0.05} = \Phi^{-1}(1 - 0.05) = 1.6449$ får vi fra dette at

$$
k = E(Y) - z_{0.05}\sqrt{\mathrm{Var}(Y)} = 420 - 1.6449 \cdot \sqrt{96} = \underline{\underline{403.8833}},
$$

slik at den garanterte minstevekten blir 403.9 g.

c) Figuren kan f.eks. lages slik:

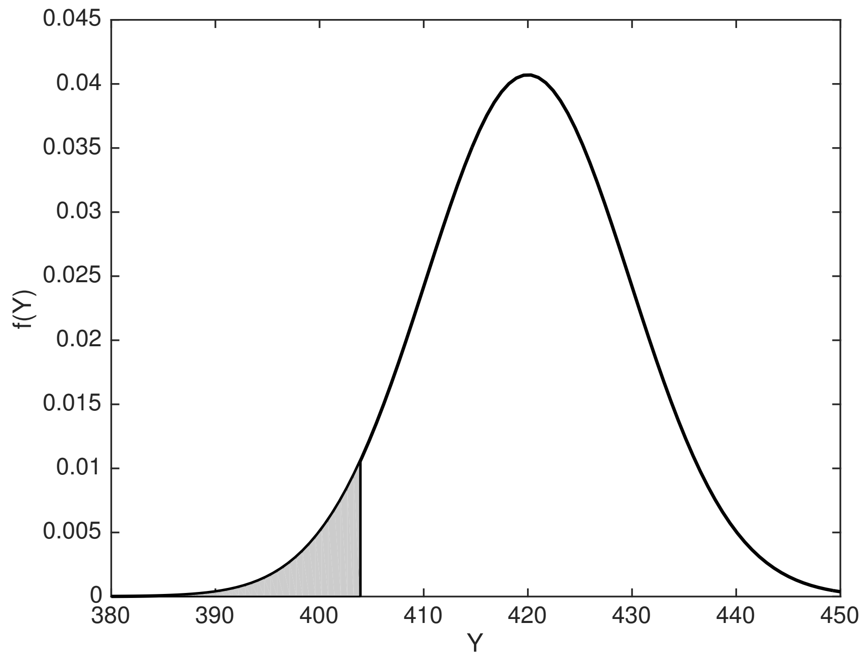
```
mu = 6*70;
sig = sqrt(6*4^2);

ymin = 380;
ymax = 450;
yc = 403.8833;
yy = linspace(ymin,ymax,100);
yy2 = linspace(ymin,yc,40);

figure(1)
hold on
plot(yy,normpdf(yy,mu,sig),'k-','LineWidth',2);
plot([yc,yc],[0,normpdf(yc,mu,sig)],'k-','LineWidth',2);
fill([yy2,fliplr(yy2)],[normpdf(yy2,mu,sig),zeros(size(yy2))],0.8*[1,1,1]);
hold off

xlabel('Y')
ylabel('f(Y)')
set(gca,'FontSize',14)
box on
```

Oppgave 2

Vi lar X være antall tankskip som ankommer havnen i løpet av en dag. Vi har fått oppgitt at $X \sim \text{poisson}(\lambda)$ med

$$
\lambda = E(X) = 2.
$$

Videre vet vi at havnen maksimalt kan betjene 3 tankskip per dag.



Figur 1: Sannsynlighetstettheten til pakkevekten Y, med området under grafen og til venstre for minstevekten fargelagt.

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 følgende punktsannsynligheter:

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. Sannsynligheten for at ett eller flere tankskip må omdirigeres er dermed

$$\begin{aligned}P(\text{omdirigering}) &= P(X>3)=1-P(X\leq 3) \overset{\text{tabell}}{=} 1-0.8571\\ &= \underline{\underline{0.1429}}.\end{aligned}$$

b) Vi lar nå Y være antall skip som betjenes ved havnen en dag. Havnens begrensede kapasitet gjør at $Y \leq 3$, slik at

$$P(Y=y)=\begin{cases}P(X=y) & \text{for } y=0,1,2\\ P(X\geq 3)=1-P(X\leq 2) & \text{for } y=3\\ 0 & \text{for alle andre verdier av } y.\end{cases}$$

Forventet antall skip som betjenes en gitt dag blir dermed

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

c) Vi lar k være havnens kapasitet, altså maksimalt antall skip som kan betjenes på en dag. Vi ønsker å finne k slik at $P(X \leq k) \geq 0.90$. Fra tabellen for poissonfordelingen har vi disse kumulative sannsynlighetene:

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

Dermed ser vi at

$$P(X \leq 3) < 0.90 \quad \text{mens} \quad P(X \leq 4) > 0.90.$$

Havnen trenger altså en kapasitet på $\underline{\underline{k=4}}$ skip per dag for med minst 90% sannsynlighet å kunne betjene samtlige skip som ankommer en gitt dag.

Oppgave 3

a) Vi har at en gjennomlesing av teksten tilsvarer n repeterte forsøk, ett forsøk for hver skrivefeil i teksten. Hvert forsøk resulteterer i suksess (feilen oppdages) eller ikke-suksess (feilen oppdages ikke). Sannsynligheten for suksess er p, og denne er konstant for alle forsøkene. Vi må i tillegg anta at hvert ord leses uavhengig av alle andre ord i teksten, slik at forsøkene er uavhengige.

Vi har $\lambda = 2$ og $s = 8$ og ønsker å finne sannsynligheten for at antall trykkfeil, N, er større enn 10. Vi har $\mu = \lambda s = 2 \cdot 8 = 16$

$$\begin{aligned}
P(N > 10) &= 1 - P(N \leq 10) \\
&= 1 - \sum_{i=1}^{10} P(N=i) \\
&= 1 - 0.0774 \\
&= 0.923.
\end{aligned}$$

Vi har nå gitt $N = 12$ og $p = 0.6$ og œnsker å finne sannsynligheten for at korrekturleseren oppdager alle trykkfeilene.

$$\begin{aligned} P(X=12|N=12) &= \binom{12}{12} \cdot 0.6^{12} \cdot 0.4^{0} \\ &= 0.6^{12} \\ &= 0.0022. \end{aligned}$$

b) Y_k = antall trykkfeil som gjenstår etter k uavhengige gjennomlesninger. Vi finner først simultanfordelingen til Y_1 og N.
Vi har

$$P(X=x|N=n) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, ..., n$$

Simultanfordelingen til Y_1 og N er da gitt ved

$$\begin{aligned} P(Y_1=u, N=n) &= P(Y_1=u|N=n) \cdot P(N=n) \\ &= P(N-X=u|N=n) \cdot P(N=n) \\ &= P(X=n-u|N=n) \cdot P(N=n) \\ &= \binom{n}{n-u} p^{n-u}(1-p)^u \cdot P(N=n) \end{aligned}$$

for $u = 0, 1, \ldots$ og $n = u, u+1, \ldots$.
Vi finner deretter marginalfordelingen til Y_1.

$$\begin{aligned} P(Y_1=u) &= \sum_{n=u}^{\infty} P(Y_1=u, N=n) \\ &= \sum_{n=u}^{\infty} \binom{n}{n-u} p^{n-u}(1-p)^u e^{-\lambda s} \cdot \frac{(\lambda s)^n}{n!} \\ &= \sum_{n=0}^{\infty} \binom{n+u}{n} p^{n}(1-p)^u e^{-\lambda s} \cdot \frac{(\lambda s)^{n+u}}{(n+u)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!u!} p^{n}(1-p)^u e^{-\lambda s} (\lambda s)^{n+u} \\ &= \frac{(\lambda s)^u e^{-\lambda s}}{u!}(1-p)^u \sum_{n=0}^{\infty} \frac{(\lambda p s)^n}{n!} \\ &= \frac{(\lambda s)^u e^{-\lambda s}}{u!}(1-p)^u e^{\lambda p s} \\ &= \frac{(\lambda s(1-p))^u}{u!} e^{-\lambda s(1-p)}. \end{aligned}$$

Vi ser at marginalfordelingen til Y_1 er $Y_1 \sim \text{Poisson}(\lambda s(1-p))$.

Oppgave 4

a) For hver deltaker har vi følgende situasjon:

- Deltakeren får en serie oppgaver.
- Hver runde har to mulige utfall: Deltakeren klarer ikke oppgaven og går ut av konkurransen (hendelse A), eller han/hun klarer oppgaven og går videre til neste runde (hendelse A').
- Sannsynligheten for ikke å klare oppgaven, $p = P(A)$, er lik i hver runde.
- Resultatene fra hver runde er uavhengige.

Denne situasjonen svarer til en Bernoulli-forsøksrekke, der vi ikke bestemmer antall forsøk på forhånd, men repeterer forsøket (gir nye oppgaver) inntil første gang hendelsen A (klarer ikke oppgaven) inntreffer. Siden X er antall forsøk inntil A inntreffer *første* gang (deltakeren første gang ikke klarer oppgaven), er det rimelig å anta at X er geometrisk fordelt.

Sannsynligheten for at deltakeren går ut i første runde:

$$P(X=1) = f(1) = p(1-p)^{1-1} = p = \underline{\underline{0.10}}$$

Sannsynligheten for at deltakeren fortsatt er med etter fem runder:

$$P(X>5) = 1 - P(X \leq 5) = 1 - F(5) = 1 - (1-(1-p)^5) = (1-p)^5 = 0.90^5 = \underline{\underline{0.59}}.$$

Sannsynligheten for at deltakeren ikke klarer oppgaven i niende runde ($X = 9$), dersom deltakeren klarer oppgavene til og med femte runde ($X > 5$): Her bruker vi betinget sannsynlighet, og resultatet fra forrige spørsmål.

$$\begin{aligned} P(X=9 \mid X>5) &= \frac{P(X=9 \cap X>5)}{P(X>5)} = \frac{P(X=9)}{P(X>5)} = \frac{f(9)}{1-F(5)} \\ &= \frac{p(1-p)^{9-1}}{(1-p)^5} = p(1-p)^3 = 0.10 \cdot 0.90^3 = \underline{\underline{0.073}} \end{aligned}$$

b) Vi har følgende situasjon for hver oppgavelager:

- Resultater for et visst antall (n_1 eller n_2) deltakere blir registrert
- To mulig utfall: Deltakeren klarer færre enn fem oppgaver (hendelse C), eller ikke (dvs. klarer fem eller flere, hendelse C').
- Sannsynligheten for C er lik i for hver deltaker.
- Resultatene for hver deltaker er uavhengige.

Dette svarer til et binomisk forsøk, og Z_1 og Z_2 er dermed binomisk fordelte, med parametre som gitt i oppgaven.