TMA4240 Statistikk
Høst 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalt øving 6
Løsningsskisse

Oppgave 1 Togforsinkelsen — Eksamen desember 2003, oppgave 1 av 3

Oppgitt: $\int_0^\infty x^r e^{-ax}\mathrm{d}x = \frac{r!}{a^{r+1}}$ for $a > 0$, $r \geq 0$ heltall.

a) For at $g(x)$ skal være en sannsynlighetstetthet, må vi ha $\int_{-\infty}^\infty g(x)\mathrm{d}x = 1$, dvs at total sannsynlighet er 1. (Bruker formelen med $r = 1$ og $a = 2$.)

$$\int_{-\infty}^{\infty} g(x)\mathrm{d}x = \int_0^\infty kxe^{-2x}\mathrm{d}x = k \cdot \frac{1}{2^2} = \frac{k}{4}.$$

$k/4 = 1$ gir $k = 4$.

For forventet forsinkelse brukes formelen igjen, med $r = 2$ og $a = 2$:

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} xg(x)\mathrm{d}x = \int_0^\infty x \cdot 4xe^{-2x}\mathrm{d}x \\
&= 4\int_0^\infty x^2e^{-2x}\mathrm{d}x = 4 \cdot \frac{2}{2^3} = 1.
\end{aligned}$$

For å vise at det er 0.09 i sannsynlighet for mer enn to minutters forsinkelse, bruker vi delvis integrasjon.

$$\begin{aligned}
\mathrm{Prob}(X > 2) &= \int_2^\infty 4xe^{-2x}\mathrm{dx} \\
&= 4 \cdot \left[-\frac{1}{2}xe^{-2x}\right]_{x=2}^{\infty} + 4 \cdot \int_2^\infty \frac{1}{2}e^{-2x}\mathrm{d}x \\
&= 4 \cdot \frac{1}{2} \cdot 2e^{-4} + \int_2^\infty 2e^{-2x}\mathrm{d}x \\
&= 4e^{-4} + e^{-4} = 5e^{-4} \approx 0.09.
\end{aligned}$$

b) V er binomisk fordelt med $n = 22$ og $p = 0.09$ under forutsetning av at

- Hendelsene ”mer enn 2 minutter forsinket” for to forskjellige dager er uavhengige.
- Sannsynligheten for ”mer enn 2 minutter forsinket” er lik 0.09 hver dag.

Antall forsøk (dager) er bestemt på forhånd, det er to utfall og vi teller antall "suksesser". (Togselskapet ville neppe kalle en dag med mer enn to minutters forsinkelse for en suksess.)

$$\begin{aligned}\text{Prob}(V \geq 2) &= 1 - \text{Prob}(V \leq 1) = 1 - \text{Prob}(V = 0) - \text{Prob}(V = 1) \\ &= 1 - 0.91^{22} - 22 \cdot 0.91^{21} \cdot 0.09^1 = 0.6012.\end{aligned}$$

Ser vi på et år med $n = 220$ virkedager, er $V \sim \text{binomisk}(220, 0.09)$. Da kan vi bruke tilnærmingen til normalfordelingen, dvs

$$\begin{aligned}\text{Prob}(V > 30) &= 1 - \text{Prob}(V \leq 30) \approx 1 - \Phi(\frac{30 + 1/2 - 220 \cdot 0.09}{\sqrt{220 \cdot 0.09 \cdot (1 - 0.09)}} \\ &= 1 - \Phi(2.52) = 0.0059.\end{aligned}$$

c) Setter $x = 2$ i den betingede fordelingen. Da har oppholdstiden Y fordeling $f(y|2) = e^{-y}$ for $y > 0$. Med andre ord er $Y|X = 2$ eksponensialfordelt med parameter (og forventningsverdi) 1.

Simultantetthet finner vi ved å multiplisere;

$$f(x, y) = f(y|x)g(x) = \frac{x}{2}e^{-\frac{xy}{2}} \cdot 4xe^{-2x} = 2x^2 e^{-x(2+\frac{y}{2})} \qquad \text{for } x > 0, y > 0.$$

Marginaltettheten for Y finnes ved å integrere ut x:

$$\begin{aligned}h(y) &= \int_0^\infty f(x, y)\mathrm{d}x = \int_0^\infty 2x^2 e^{-x(2+\frac{y}{2})}\mathrm{d}x \\ &= 2 \cdot \frac{2}{(2 + \frac{y}{2})^3} = \frac{32}{(4 + y)^3}, \qquad \text{for } y > 0.\end{aligned}$$

Her brukte vi enda en gang formelen som var oppgitt, denne gangen med $a = 2 + y/2$ og $r = 2$.

Oppgave 2

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,

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

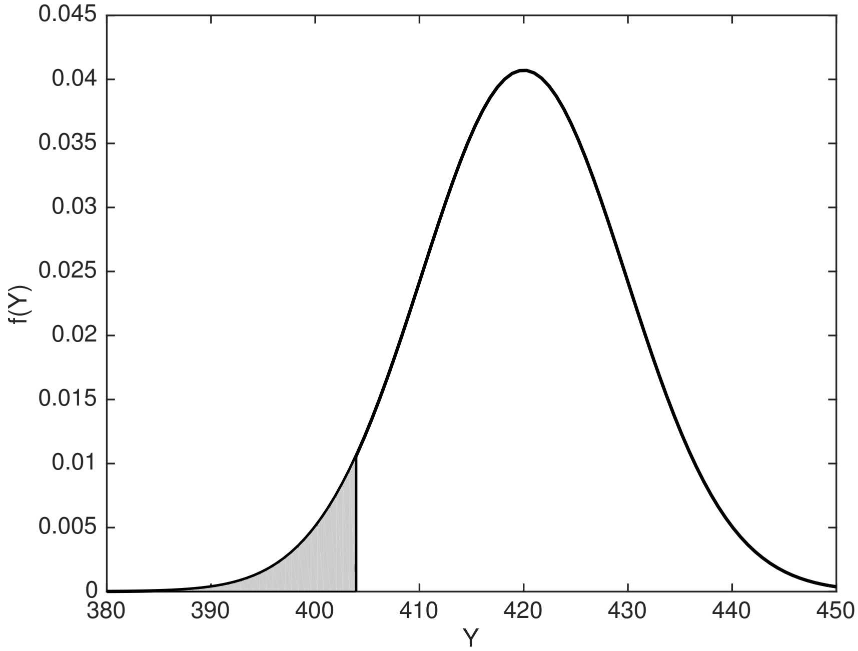
Med normalkvantilen $z_{0.05} = \Phi^{-1}(1 - 0.05) = 1.6449$ får vi fra dette at

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

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);
```



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

```
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 3 Vi lar X være en bestemt pH-måling, og antar at X er normalfordelt med forventningsverdi $\mu = 6.8$ og varians $\sigma^2 = 0.060^2$. Sannsynligheten for at resultatet av målingen er under 6.74 er da

$$\begin{aligned} P(X < 6.74) &= P(\frac{X-6.8}{0.06} < \frac{6.74-6.8}{0.06}) \\ &= \Phi(-1) = 1 - \Phi(1) \\ &= 1 - 0.841 = \underline{\underline{0.159}}. \end{aligned}$$

Videre er sannsynligheten for at resultatet av målingen ligger mellom 6.74 og 6.86 lik

$$\begin{aligned}
P(6.74 < X < 6.86) &= P(X < 6.86) - P(X < 6.74) \\
&= P(\frac{X - 6.8}{0.06} < \frac{6.86 - 6.8}{0.06}) - 0.159 \\
&= \Phi(1) - 0.159 = 0.841 - 0.159 = \underline{\underline{0.682}}.
\end{aligned}$$

Sannsynligheten for at avviket $|X - \mu|$ overstiger 0.06 er

$$\begin{aligned}
P(|X - \mu|) > 0.06) &= P(X - \mu < -0.06) + P(X - \mu > 0.06) \\
&= P(\frac{X - \mu}{0.06} < -1) + P(\frac{X - \mu}{0.06} > 1) \\
&= \Phi(-1) + 1 - \Phi(1) = 2(1 - \Phi(1)) = \underline{\underline{0.318}}.
\end{aligned}$$

Den samme sannsynligheten kan også regnes ut som følger:

$$\begin{aligned}
P(|X - \mu|) > 0.06) &= 1 - P(6.74 < X < 6.86) \\
&= 1 - 0.682 = \underline{\underline{0.318}}.
\end{aligned}$$

Oppgave 4

a) Fra oppgaveteksten har vi den betingede sannsynlighetsfordelingen til T gitt λ,

$$f_{T|\lambda}(t|\lambda) = \lambda e^{-\lambda t}, \quad t \geq 0,$$

og sannsynlighetsfordelingen til λ,

$$f_\lambda(\lambda) = \theta e^{-\lambda\theta}, \quad \lambda \geq 0,$$

hvor θ inngår som en parameter. Simultanfordelingen til T og λ er da gitt ved

$$f_{T,\lambda}(t, \lambda) = f_{T|\lambda}(t|\lambda) \cdot f_\lambda(\lambda) = \lambda\theta e^{-\lambda(t+\theta)},$$

og marginalfordelingen til T kan finnes ved å integrere ut "mellomleddet" λ,

$$\begin{aligned}
f_T(t) &= \int_0^\infty f_{T,\lambda}(t, \lambda) d\lambda = \int_0^\infty \lambda\theta e^{-\lambda(t+\theta)} d\lambda \\
&= \theta \left(\left[\lambda \left(-\frac{1}{t + \theta} \right) e^{-\lambda(t+\theta)} \right]_0^\infty + \int_0^\infty \frac{1}{t + \theta} e^{-\lambda(t+\theta)} d\lambda \right) \\
&= \theta \left(0 - 0 + \left[\frac{1}{(t + \theta)^2} e^{-\lambda(t+\theta)} \right]_0^\infty \right) \\
&= \theta \left(0 + \frac{1}{(t + \theta)^2} \right) = \frac{\theta}{(t + \theta)^2}.
\end{aligned}$$

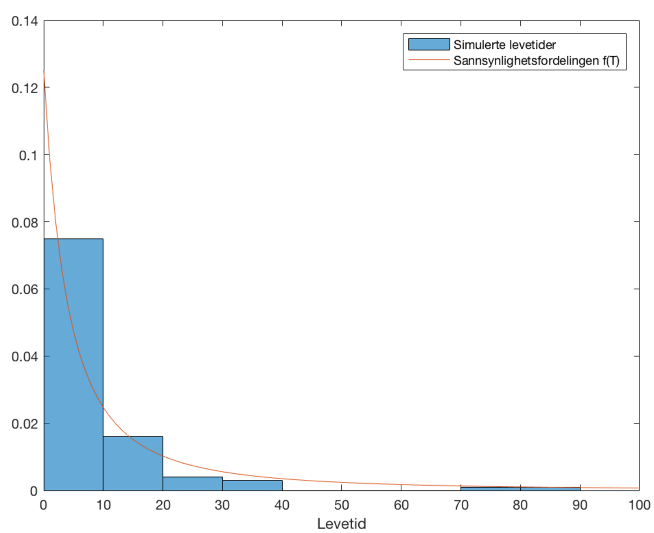
b) Simuleringen kan gjennomføres i Matlab på følgende måte:

```
theta=8 #Konstant verdi.

N=100 #Vi er interessert i levetiden til 100 komponenter.

for i=1:N

    #Simuler parameterverdien lambda
    #fra en eksponensialfordeling med parameterverdi theta:
    lambda=exprnd(theta);

    #Simuler levetiden til en
    #komponent med parameterverdi lambda.
    #Denne er eksponensialfordelt gitt lambda.
    t(i)=exprnd(lambda);

end
```

c) Histogram og plott av sannsynlighetstettheten lages i Matlab:

```
figure
histogram(t,'Normalization','pdf')
hold on

T=linspace(0,100,100)
plot(T,theta./(T+theta).^2)
```

Resultatet er vist i figuren nedenfor. Vi ser at den teoretiske levetiden stemmer ganske bra med de simulerte verdiene. Merk at histogrammet du oppnår kan se noe annerledes ut enn det du ser her. Dette skyldes at man vil ende opp med forskjellige levetider fra en simulering til en annen.



Oppgave 5 For eksponentialfordelingen har vi

$$\begin{aligned}
P(X \geq t+s | X > s) &= \frac{P(X \geq t+s \cap X > s)}{P(X > s)} \\
&= \frac{P(X \geq t+s)}{P(X > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} = \underline{\underline{P(X \geq t)}}.
\end{aligned}$$

For den geometriske fordelingen har vi, for en vilkårlig $a \in \{0, 1, 2, \ldots\}$,

$$P(X \geq a) = P(\text{Ingen suksesser på de første } a-1 \text{ forsøkene}) = (1-p)^{a-1},$$

og vi får dermed

$$\begin{aligned}
P(X \geq t+s | X > s) &= \frac{P(X \geq t+s \cap X > s)}{P(X > s)} \\
&= \frac{P(X \geq t+s)}{P(X > s)} \\
&= \frac{P(X \geq t+s)}{P(X \geq s+1)} \\
&= \frac{(1-p)^{(t+s)-1}}{(1-p)^{(s+1)-1}} \\
&= \frac{(1-p)^{t+s-1}}{(1-p)^{s}} \\
&= (1-p)^{t-1} = \underline{\underline{P(X \geq t)}}.
\end{aligned}$$