TMA4240 Statistikk
Høst 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalt øving 7
Løsningsskisse

Oppgave 1

a) Regner først ut den kumulative fordelingsfunksjonen til X:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad \text{for} \quad x > 0$$

Skal finne sannsynlighetstettheten til $V = \max(X_1, X_2)$ og regner først ut fordelingsfunksjonen:

$$\begin{aligned} F_V(v) = P(\max(X_1, X_2) \le v) &= P(X_1 \le v \cap X_2 \le v) \\ &\overset{uavh}{=} P(X_1 \le v) P(X_2 \le v) \\ &= F_X(v)^2 = (1 - e^{-\lambda v})^2 = 1 - 2e^{-\lambda v} + e^{-2\lambda v} \end{aligned}$$

Dvs. sannsynlighetstettheten til V blir:

$$f_V(v) = F'(v) = \underline{\underline{2\lambda e^{-\lambda v} - 2\lambda e^{-2\lambda v}}}$$

$$\begin{aligned} \mathrm{E}(V) = \int_0^\infty v(2\lambda e^{-\lambda v} - 2\lambda e^{-2\lambda v}) dv &= 2\int_0^\infty v\lambda e^{-\lambda v} dv - \int_0^\infty v 2\lambda e^{-2\lambda v} dv \\ &= 2\frac{1}{\lambda} - \frac{1}{2\lambda} = \underline{\underline{\frac{3}{2\lambda}}} \end{aligned}$$

Vi har at $\mathrm{E}(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$, dvs. vi har at $\underline{\underline{\mathrm{E}(X) < \mathrm{E}(V) < 2\mathrm{E}(X)}}$ som ventet da V er den største av to X-er. Siden $V = \max(X_1, X_2)$ vil vi forvente at $\mathrm{E}(V) > \mathrm{E}(X)$ og at $\mathrm{E}(V) < \mathrm{E}(X_1 + X_2) = 2\mathrm{E}(X)$.

b) **Merk at tallsvarene i resten av oppgaven kan være ulik de tallsvarene du får selv.**

En kan definere følgende funksjon som simulerer fra fordelding til V:

```
function max_verdi = simulerFraMaks(B, lambda, n)
% Funksjon som simulerer fra maks-fordeling
% B:        antall gjentagelser
% lambda:   modellparameter
```

```
% n:        antall X_i i hver gjentagelse

max_verdi = zeros(B, 1);
for b = 1:B
     utbetaling = exprnd(1/lambda, [n,1]); % trekk fra eksponentialfordelingen
     max_verdi(b) = max(utbetaling); % lagre maksimum
end
end
```

og deretter kalle denne funksjonen:

```
lambda = 0.2; % modellparameter
n = 8; % antall industribranner
B = 500; % antall ganger vi gjentar forsoeket
max_verdi = simulerFraMaks(B, lambda, n); % B realiseringer fra fordelingen til V

figure;
histogram(max_verdi,'Normalization','probability')


% Plot cdf
figure;
[Fhat, v_verdi] = ecdf(max_verdi);
plot(v_verdi, Fhat);


% Korrekt cdf
F = @(v, lambda, n) (1-exp(-lambda * v)).^n;
1-F(v_verdi, lambda, n);
hold on;
plot(v_verdi, F(v_verdi, lambda, n));
legend('Empirisk CDF', 'Teoretisk CDF');
```

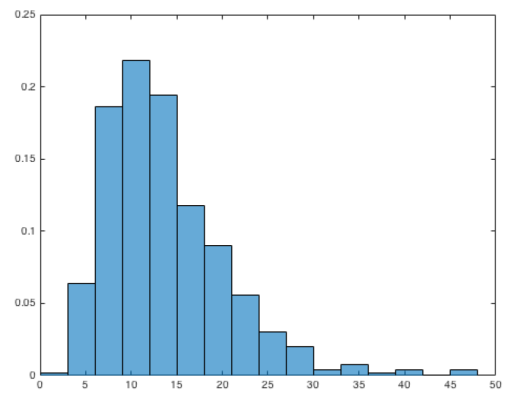
Et histogram av fordelingen til V er gitt i Figur 1. Empirisk og teoretisk kumulative fordelingsfunksjon er plottet i Figur 2, og vi ser at de stemmer bra overens.

c) Vi finner sannsynligheten ved å telle antall ganger $V > 30$ og dele på 500:

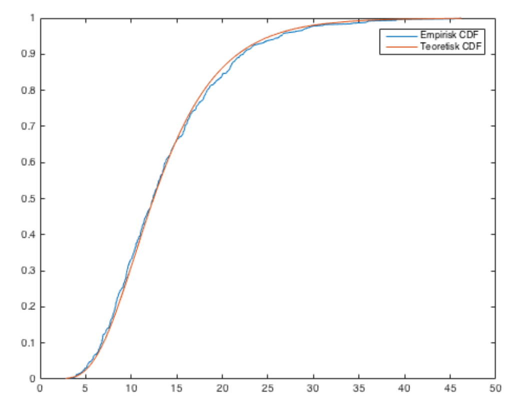
```
% Anslaa sannsynliget V > 30
verdi = 30; %
sannsynliget = sum(max_verdi > verdi) / B

% Sann sannsynlighet
1-F(verdi, lambda, n)
```

Den sanne sannsynligheten er 0.0197, mens den anslåtte sannsynligheten er 0.0220. Dersom vi hadde trukket flere realisasjoner fra V kunne vi ha ventet at den anslåtte sannsynligheten var nærmere den sanne verdien.



Figur 1: Histogram av V.



Figur 2: Empirisk og teoretisk kumulativ fordelingsfunksjon til V.

d) Vi estimerer forventet høyeste utbetaling for 8 branner ved å ta gjennomsnittet av 500 realisasjoner fra V:

```
% Forventet stoerste utbetaling
forventet_utbetaling = mean(max_verdi)
```

Forventet høyeste utbetaling er 13.72 millioner kroner.

e) Kode for å simulere beholdning på konto er gitt under:

```
ta_opp_laan = 30;
ta_ut_fra_konto = 25;
startsum_konto = 30;
laanekostnad = 5;
pris_forsikring = 5;

B = 10000; % antall ganger vi gjentar forsoeket
max_verdi = simulerFraMaks(B, lambda, n);
```

```
beholdning_konto = startsum_konto * ones(B,1);
for b = 1:B
    % Avgjoer om reassurering
    reassurer = randsample(0:1,1,true, [1/3, 2/3]);
    % Ikke reassurer
    if reassurer == 0
        if max_verdi(b) > ta_opp_laan;
            % maa ta opp laan
            beholdning_konto(b) = beholdning_konto(b)-laanekostnad ...
                -ta_ut_fra_konto;
        else
            % kan betale selv
            beholdning_konto(b) = beholdning_konto(b) - max_verdi(b);
        end
    % reassurer
    else
        beholdning_konto(b) = beholdning_konto(b) - pris_forsikring;
    end
end

% Forventet beholdning i fondet etter et aar
forventet_beholdning = mean(beholdning_konto)
```

Forventet beholdning er 22.0804 millioner kroner etter et år.

Oppgave 2

X er kontinuerlig fordelt med sannsynlighetstetthet

$$f_X(x) = \begin{cases} 2x\exp(-x^2) & x \geq 0 \\ 0 & \text{ellers} \end{cases}$$

Formel for transformasjon av variabler finnes i det blå heftet.

a) Vi har

$$U = X - 2 = g(X); x \geq 0$$

slik at

$$X = U + 2 = h(U)$$

med

$$h'(u) = 1.$$

Funksjonen $g(x) = x - 2$ er strengt monoton og deriverbar for alle x. Vi har dermed

$$
\begin{aligned}
f_U(\,u) &= f_X(\,h(u))\,|h'(u)| \\
&= 2\,(u+2)\,\exp(-\,(u+2)^2)\cdot 1 \\
&= 2\,(u+2)\,\exp(-\,(u+2)^2); u \geq -2.
\end{aligned}
$$

Alternativt kan vi ta utgangspunt i kumulativ fordeling. Vi skriver

$$
\begin{aligned}
F_U(u) = \mathrm{P}(U \leq u) &= \mathrm{P}(X - 2 \leq u) \\
&= \mathrm{P}(X \leq u + 2),\ u \geq -2.
\end{aligned}
$$

Dette gir

$$
F_U(u) = F_X(u+2) = 1 - \exp\{-(u+2)^2\},\ u \geq -2.
$$

Derivasjon mhp u gir riktig tetthetsfunksjon.

b) Vi har her

$$
V = -2X = g(X); x \geq 0
$$

der

$$
X = -\tfrac{1}{2}V = h(V)
$$

med

$$
h'(v) = -\tfrac{1}{2}.
$$

Funksjonen $g(x) = -2x$ er strengt monoton og deriverbar for alle x. Dette gir

$$
\begin{aligned}
f_V(\,v) &= f_X(\,h(v))\,|h'(\,v)| \\
&= 2\,(-\tfrac{1}{2}\,v)\,\exp(-\,(-\tfrac{1}{2}\,v)^2)\cdot\tfrac{1}{2} \\
&= -\tfrac{1}{2}\,v\,\exp(-\,(\tfrac{1}{2}\,v)^2); v \leq 0.
\end{aligned}
$$

Alternativt kan vi ta utgangspunt i kumulativ fordeling. Vi skriver

$$
\begin{aligned}
F_V(v) = \mathrm{P}(V \leq v) &= \mathrm{P}(-2X \leq v) \\
&= \mathrm{P}(X \geq -\frac{v}{2}) \\
&= 1 - \mathrm{P}(X \leq -\frac{v}{2}),\ v \leq 0.
\end{aligned}
$$

Dette gir

$$
F_V(v) = 1 - F_X\left(-\frac{v}{2}\right) = \exp\{-\left(\frac{v}{2}\right)^2\},\ v \leq 0.
$$

Derivasjon mhp v gir riktig tetthetsfunksjon.

c) Vi har

$$W = X^2 = g(X); x \geq 0$$

som gir

$$X = \sqrt{W} = h(W)$$

med

$$h'(w) = \frac{1}{2\sqrt{w}}.$$

Funksjonen $g(x) = x^2$ er strengt monoton og deriverbar for alle $x \geq 0$.

$$\begin{aligned} f_W(\,w) &= f_X(\,h(\,w))\,|h'(\,w)| \\ &= 2\,\sqrt{w}\,\exp(\,-w)\tfrac{1}{2\sqrt{w}} \\ &= \exp(-w); w \geq 0. \end{aligned}$$

Alternativt kan vi ta utgangspunt i kumulativ fordeling. Vi skriver

$$\begin{aligned} F_W(w) = \mathrm{P}(W \leq w) &= \mathrm{P}(X^2 \leq w) \\ &= \mathrm{P}(|X| \leq \sqrt{w}) \\ &= \mathrm{P}(-\sqrt{w} \leq X \leq \sqrt{w}) \\ &= \mathrm{P}(X \leq \sqrt{w}) - \mathrm{P}(X \leq -\sqrt{w}) \\ &= \mathrm{P}(X \leq \sqrt{w}),\ w \geq 0. \end{aligned}$$

Dette gir

$$F_W(w) = F_X\left(\sqrt{w}\right) = 1 - \exp\{-w\},\ w \geq 0.$$

Derivasjon mhp w gir riktig tetthetsfunksjon.

Oppgave 3

Den stokastiske variabelen X er normalfordelt med forventningsverdi μ og varians σ^2,

$$X \sim N(\mu, \sigma^2)$$

og vi skal utlede sannsynlighetstetthetsfunksjonen til den stokastiske variabelen Y, som er gitt ved

$$Y = \frac{X}{\sigma} - \frac{\mu}{\sigma} = \frac{X - \mu}{\sigma}.$$

Den kumulative sannsynlighetsfordelingen til Y er

$$F_Y(y) = P(Y \leq y) = P\left(\frac{X - \mu}{\sigma} \leq y\right) = P(X \leq \mu + \sigma y) = F_X(\mu + \sigma y).$$

Derivasjon med hensyn på y gir sannsynlighetstettheten $f_Y(y)$,

$$\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{d}{dy}F_X(\mu+\sigma y) \\
&= F_X'(\mu+\sigma y)\cdot\frac{d}{dy}(\mu+\sigma y) \\
&= f_X(\mu+\sigma y)\cdot\sigma \\
&= \frac{\sigma}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}((\mu+\sigma y)-\mu)^2} \\
&= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}.
\end{aligned}$$

Dette er tettheten til normalfordelingen med $\mu = 0$ og $\sigma = 1$. Altså har vi at $Y \sim N(0,1)$.

Oppgave 4

a) I dette punktet er $\beta = 10$.

$P(X \leq 4) = 1 - e^{-4/10} = 0.33$

$P(X > 7) = e^{-7/10} = 0.50$

$P(X > 7|X > 4) = \frac{P(X>7)}{P(X>4)} = \frac{0.5}{1-0.33} = 0.74$

b) $P(X > c) = 1 - F(c) = e^{-c/\beta}$

$P(Y > y) = P(X - c > y|X > c) = \frac{P(X>y+c)}{P(X>c)} = e^{-y/\beta}$

Fordelingen til Y gjenkjennes som en eksponensialfordeling med forventningsverdi β.