Nikolaj Dybdahl Rathcke (rfq695) og Victor Petren Bach Hansen (grn762)

Statistik og Sandsynlighed - Projekt

Opgave 1

1.1

Lad F være fordelingsfunktionen for Y . Udtryk F ved hjælp af fordelingsfunktionen for standard normalfordelingen Φ, og vis at fordelingen af Y har en tæthed givet ved

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}exp(-\frac{(log(y)-\mu)^2}{2\sigma^2}),\ \ y > 0$$

Vi har at,

$$\begin{aligned} F(y) &= P(Y \leq y) = P(Exp(X) \leq y) \\ &= P(X \leq \ln(y)) = \Phi(\frac{(\ln(y)-\mu)^2}{\sigma}) \end{aligned}$$

for $y > 0$. Derved er $Y' = \frac{1}{y}$
Deraf får vi fra transformationssætning, at

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}exp(-\frac{(ln(y)-\mu)^2}{2\sigma^2}),\ \ y > 0$$

1.2

Vis at Y har middelværdi, og at $EY = e^{\mu+\sigma^2/2}$

For at bestemme en middelværdi for Y, bliver vi først nødt til at se om den overhovedet eksisterer, altså at

$$E|Y| < \infty$$

Det vides dog at $|p(y)| = p(y)$ så hvis $p(y)$ har en middelværdi, bliver $E(Y)$ altså også nødt til at have en.

Vi ved at $Y = e^X$, så kan middelværdien udtrykkes som

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} yp(y)\,\mathrm{d}y \\ E(e^X) &= \int_{-\infty}^{\infty} e^x p(x)\,\mathrm{d}x \end{aligned}$$

Her kan vi indsætte at $p(x)$ er givet ved sætning 4.3.5 i MS,

$$E(e^X) = \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2})\,\mathrm{d}x$$

Her erstattes udtrykket $z = x - \mu \Leftrightarrow x = z + \mu$ og vi skriver det som:

$$\begin{aligned} E(e^X) &= \int_{-\infty}^{\infty} e^{z+\mu} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2}{2\sigma^2})\,\mathrm{d}z \\ &= \int_{-\infty}^{\infty} e^z e^{\mu} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2}{2\sigma^2})\,\mathrm{d}z \end{aligned}$$

Da e^{μ} er en konstant, rykkes den uden for integralet

$$\begin{aligned} E(e^X) &= e^{\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2}{2\sigma^2} + z)\,\mathrm{d}z \\ &= e^{\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2 - 2\sigma^2 z}{2\sigma^2})\,\mathrm{d}z \end{aligned}$$

Her benyttes vinket $x^2 - 2ax = (x-a)^2 - a^2$ og vi skriver udtrykket som

$$\begin{aligned} E(e^X) &= e^{\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(z-\sigma^2)^2 - \sigma^4}{2\sigma^2})\,\mathrm{d}z \\ &= e^{\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(z-\sigma^2)^2}{2\sigma^2} - \frac{\sigma^4}{2\sigma^2})\,\mathrm{d}z \\ &= e^{\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(z-\sigma^2)^2}{2\sigma^2} - \frac{\sigma^2}{2})\,\mathrm{d}z \\ &= e^{\mu+\sigma^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(z-\sigma^2)^2}{2\sigma^2})\,\mathrm{d}z \end{aligned}$$

Her observeres det at udtrykket vi integrerer over, er sandsynlighedstætheden for normalfordelingen med parametrene $N(\mu = \sigma^2, \sigma^2)$, og vi benytter at tæthedsfunktionen for en standard normalfordeling integreres til 1, hvilket giver udtrykket

$$E(Y) = e^{\mu+\sigma^2/2}$$

som også er det vi ønskede at finde.

1.3

Bestem varians for Y

Variansen kan udtrykked ved

$$V(Y) = E(Y^2) - (E(Y))^2$$

$E(Y)^2$ kan hurtigt bestemmes som

$$\begin{aligned}(E(Y))^2 &= e^{(\mu+\sigma^2/2)^2}\\ &= e^{2\mu+\sigma^2}\end{aligned}$$

men for at finde $E(Y^2)$ bruges samme metode som opg 1.2

$$\begin{aligned}E(Y^2) &= \int_{-\infty}^{\infty} yp(y)\,\mathrm{d}y = \\ E(e^{2x}) &= \int_{-\infty}^{\infty} e^{2x}p(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} e^{2x}\frac{1}{\sqrt{2\pi\sigma^2}}exp(-\frac{(x-\mu)^2}{2\sigma^2})\,\mathrm{d}x\end{aligned}$$

$(x-\mu)$ substitueres igen med z, altså $z = x - \mu \Leftrightarrow x = z + \mu$ og samme fremgangsmåde benyttes nu endnu en gang:

$$\begin{aligned}
E(e^{2x}) &= \int_{-\infty}^{\infty} e^{2z+2\mu} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2}{2\sigma^2})\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} e^{2z} e^{2\mu} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2}{2\sigma^2})\,\mathrm{d}z \\
&= e^{2\mu} \int_{-\infty}^{\infty} e^{2z} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2}{2\sigma^2})\,\mathrm{d}z \\
&= e^{2\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2}{2\sigma^2} + 2z)\,\mathrm{d}z \\
&= e^{2\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{z^2 - 4\sigma^2 z}{2\sigma^2})\,\mathrm{d}z \\
&= e^{2\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(z-2\sigma^2)^2 - 4\sigma^2}{2\sigma^2})\,\mathrm{d}z \\
&= e^{2\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(z-2\sigma^2)^2}{2\sigma^2} + 2\sigma^2)\,\mathrm{d}z \\
&= e^{2\mu+2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(z-2\sigma^2)^2}{2\sigma^2})\,\mathrm{d}z \\
&= e^{2\mu+2\sigma^2}
\end{aligned}$$

Vi har altså nu at variansen er

$$Var(Y) = E(Y^2) - (E(Y))^2 = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} - = (e^{\sigma^2} - 1)e^{\sigma^2+2\mu}$$

Opgave 2

2.1

Betragt ligefordelingen på $[0,1]$. Gør rede for at den opfylder *Antagelse 1*, og angiv fordelingsfunktionen $F(x)$. Løs ligningen $F(x) = \alpha$ for kendt α og find dermed fraktilfunktionen $q(\alpha)$. Angiv eksplicit 5%-fraktilen, medianen og 95%-fraktilen.

Opfyldelse af antagelse 1:
Da sandsynlighedstætheden for det begrænsede interval $[0,1]$ hvor $a < b$, er givet ved

$$p(x) = \frac{1_{[0,1]}(x)}{1-0}$$

efter 4.1.5 i MS, ses det at denne funktion resulterer i en konstant, hvilket betyder at den er kontinuer. Det ses også at

$$p(x) = \frac{1_{[0,1]}(x)}{1-0} = \begin{cases} 0 & \text{for } x \notin I \\ 1 & \text{for } x \in I \end{cases}$$

hvilket bekræfter vores antagelse 1 fuldt. Denne sandsynlighedstæthed er også vist på fig. 4.1.1, i MS.
Fordelingsfunktionen $F(x)$ for ligefordelingen på intervallet (a,b) er ifl. 4.1.4($fortsat$) på s. 87, givet ved

$$F(x) = \int_{-\infty}^{x} \frac{1_{(a,b)}(y)}{b-a}\,\mathrm{d}y = \begin{cases} 0 & \text{for } x \leq a \\ \frac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x \geq b \end{cases}$$

Som i vores tilfælde altså giver os

$$F(x) = \int_{-\infty}^{x} 1_{[0,1]}(y)\,\mathrm{d}y = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$

For et kendt α bliver løsningen $F(x) = \alpha$ derfor blot som den overstående funktion angiver, idet at vi har at gøre med en ligefordeling.
Kvartilfunktionen fås nu ved at løse ligningen $F(x) = \alpha$, altså

$$\begin{aligned} F(x) &= \alpha \\ q(F(x)) &= q(\alpha) \\ x &= q(\alpha) \end{aligned}$$

Hvilket betyder at vores kvartilfunktion er

$$q(\alpha) = x$$

for $x \in I$
og derfor bliver 5%-fraktilen, medianen og 95%-fraktilen

$$\begin{aligned} q(5\%) &= 0.05 \\ q(50\%) &= 0.50 \\ q(95\%) &= 0.95 \end{aligned}$$

2.2

Angiv fordelingsfunktionen for eksponentialfordelingen med parameter $\lambda = 1$. Find fraktilfunktionen ved at løse ligningen $F(x) = \alpha$. Angiv eksplicit 5%-fraktilen, medianen og 95%-fraktilen.

Vi har at sandsynlighedstætheden for eksponentialfordelingen, fra 4.1.6 med parameter $\lambda = 1$ er givet ved

$$p(x) = \lambda e^{-\lambda x} = e^{-x},\ x > 0$$

hvor fordelingsfunktionen for eksponentialfordelingen så er givet ved (tager højde for at $x > 0$

$$F(x) = \int_0^x e^{-y}\,\mathrm{d}y = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-x} & \text{for } x > 0 \end{cases}$$

Vores fraktilfunktion er den inverse fordelingsfunktion

$$q(\alpha) = \frac{-ln(1-\alpha)}{\lambda} = -ln(1-\alpha),\ 0 \leq \alpha < 1$$

Hvoraf vi kan finde 5%-fraktilen, medianen og 95%-fraktilen.

$$\begin{aligned} q(0.05) &\approx 0.051 \\ q(0.5) &\approx 0.069 \\ q(0.95) &\approx 2.996 \end{aligned}$$

2.3

Angiv fordelingsfunktionen for eksponentialfordelingen med parameter $\lambda = 2$. Find fraktilfunktionen ved at løse ligningen $F(x) = \alpha$. Angiv eksplicit 5%-fraktilen, medianen og 95%-fraktilen.

Vi har at sandsynlighedstætheden for eksponentialfordelingen, fra 4.1.6 med parameter $\lambda = 2$ er givet ved

$$p(x) = \lambda e^{-\lambda x} = 2e^{-2x},\ x > 0$$

hvor fordelingsfunktionen for eksponentialfordelingen så er givet ved (tager højde for at $x > 0$

$$F(x) = \int_0^x 2e^{-2y}\,\mathrm{d}y = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-2x} & \text{for } x > 0 \end{cases}$$

Vores fraktilfunktion er

$$q(\alpha) = \frac{-ln(1-\alpha)}{\lambda} = \frac{-ln(1-\alpha)}{2},\ 0 \leq \alpha < 1$$

Hvoraf vi kan finde 5%-fraktilen, medianen og 95%-fraktilen.

$$\begin{aligned} q(0.05) &\approx 0.026 \\ q(0.5) &\approx 0.347 \\ q(0.95) &\approx 1.498 \end{aligned}$$

2.4

Ladd X være en reel stokastisk variabel hvis fordeling har en tæthed p, der opfylder $Antagelse1$. Lad X have fordelingsfunktion F og lad den tilhørende fraktilfunktion være q. Betragt den transformerede stokastiske variabel

$$Y = a + bX$$

for $a, b \in \mathbb{R},\ b > 0$. Find fordelingsfunktionen $\tilde{F}$ for Y og vis at Y har fraktilfunktionen

$$\tilde{q}(\alpha) = a + bq(\alpha),\ \alpha \in (0,1)$$

Vi skriver vores fordelingsfunktion som

$$
\begin{aligned}
F_Y(x) &= P(Y \leq x) \\
&= P(a + bX \leq x) \\
&= P(X \leq \frac{x-a}{b}) \\
&= F_X(\frac{x-a}{b})
\end{aligned}
$$

Nu sætter vi $F_X(\frac{x-a}{b}) = \alpha$ og tager kvartilfunktionen på begge sider og isolerer vores x

$$
\begin{aligned}
q(F_X(\frac{x-a}{b})) &= q(\alpha) \\
\frac{x-a}{b} &= q(\alpha) \\
x &= a + bq(\alpha)
\end{aligned}
$$

og da $x = q(\alpha)$, eller $\tilde{q}(\alpha)$ i vores tilfælde, får vi altså det ønskede udtryk

$$
\tilde{q}(\alpha) = a + bq(\alpha),\ \alpha \in (0, 1)
$$

2.5

Forklar hvorfor standard normalfordelingen har 0 som median. Hvad er medianen af normalfordelingen med middelværdi μ og varians σ^2?

Standard normalfordelingen har medianen 0, da den er symmetrisk om y-aksen. Dette medfører at der er et lige stort areal (sandsynlighed) under kurven på hver side af y-aksen og derved at medianen er på y-aksen eller ved $x = 0$.
Da en normal fordeling er en version af standard normalfordelingen, hvor variansen gør den højere eller bredere og middelværdien rykker hele grafen venstre eller højere, bibeholder den altså sin symmetri, og derfor vil medianen stadig ligge hvor kurven har sit toppunkt. Da det er middelværdien der rykker kurven på x-aksen, betyder dette at medianen er lig middelværdien μ.

2.6

Lad Z være logaritmisk normalfordelt med parametre (0, 1). Vis at medianen for Z er lig 1.
Vink: Brug at Z = exp(X) hvor X er standard normalfordelt.

Vi har fra opgave 1.1, at følgende er en sandsynlighedstæthed

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}exp(-\frac{(log(y)-\mu)^2}{2\sigma^2})$$

Ved indsættelse af $(\mu = 0, \sigma^2 = 1)$ fås

$$p(y) = \frac{1}{\sqrt{2\pi}}\frac{1}{y}exp(-\frac{(log(y))^2}{2})$$

Dette kan ikke integreres for negative tal på grund af leddet $log(y)$, men da det er en sandsynlighedstæthed medfører det at det hele integrerer til 1. Da vi ønsker at vise medianen er 1, kan vi blot integrere fra 1 til ∞

$$F(x) = \int_1^\infty \frac{1}{\sqrt{2\pi}}\frac{1}{y}exp(-\frac{(log(y))^2}{2}) = \frac{1}{2}$$

Som altså betyder at arealet fra $-\infty$ til 1 også er $\frac{1}{2}$ og medianen er derfor 1.

Alternativt kan man bruge forklaringen fra opgave 2.7 til at konkludere at medianen må være $Z < e^0 = 1$

2.7

Lad Z være logaritmisk normalfordelt med parametre (μ, σ^2). Find medianen for Z.

Siden $Z = exp(X)$ hvor X er normalfordelt betyder det, som vi konkluderede i opgave 2.5, at X har median lig middelværdi. Dette gør, at sandsynligheden for at få et $x < \mu$ er $\frac{1}{2}$. Derved er sandsynligheden for at Z er mindre end $\frac{1}{2}$ det samme som $Z < e^\mu$. Altså er medianen e^μ.

2.8

Lad Xvære en reel stokastisk variabel hvis fordeling har en tæthed p, der opfylder $Antagelse 1$. Lad X have fordelingsfunktion F og lad den tilhørende fraktilfunktion være q. Betragt den transformerede stokastiske variabel

$$Y = FX$$

Vis at Y er ligefordelt på $[0, 1]$

Idet at vores fordeling for X er kontinuert, er X en stokastisk variabel. Da, Xs tæthed p opfylder vores $Antagel 1$, kan vi benytte os af Sætning 4.4.2 i MS, hvilket siger

$$q(y) = \begin{cases} p(F^{-1}(y))/|F'(F^{-1}(y))| & \text{for } y \in [v, h] \\ 0 & \text{for } y \notin [v, h] \end{cases}$$

Da vores $F' = p$, bliver $q(y)$ altså

$$q(Y) = p(F^{-1}(y))/p(F^{-1}(y)) = 1$$

for $y \in [0, 1]$, hvilket betyder at Y er ligefordelt på intervallet.

2.9

Lad p være en sandsynlighedstæthed på et interval $I \subset \mathbb{R}$. Antag at p opfylder $Antagelse 1$. Lad F være den tilhørende fordelingsfunktion, og lad q være fraktilfunktionen. Lad U være en reel stokastisk variabel, der er ligefordelt på $[0, 1]$.
Vis at den stokastiske variabel

$$X = q(U)$$

har fordelingsfunktionen F og dermed en tæthed p

$$\tilde{F}(x) = P(q(U) \leq x)$$

Fordelingsfunktionen tages på begge sider af uligheden. Uligheden vendes ikke da $F'(x) = p(x) > 0$ ifølge *Antagelse 1.*

$$\tilde{F}(x) = P(U \leq F(x))$$

Højre side af lighedstegnet kan beskrives som $F(x)$ da X er ligefordelt, og $F(x)$ er kontinuert på intervallet 0 til 1

$$\tilde{F}(x) = F(x)$$

Da er F fundet og det ses at denne er kontinuert da den opfylder *Antagelse 1*, og dermed er den differentiabel og har en tæthed p