TMA4245 Statistikk
Vår 2015

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer 8, blokk II
Løsningsskisse

Oppgave 1

Betrakt et parallellsystem av to uavhengige komponenter, der levetiden til hver av komponentene er eksponensialfordelt med parameter λ, dvs

$$F_X(x) = 1 - \mathrm{e}^{-\lambda\, x} \text{ for } x \geq 0.$$

Da komponentene danner et paralellsystem, vil systemet fungere dersom minst en av komponentene fungerer. Vi lar dermed levetiden til systemet betegnes ved $V = \max(X_1, X_2)$, og fordelingen til V finnes ved å benytte at komponenten med lengst levetid er mindre eller lik v hvis og bare hvis begge komponentene er mindre eller lik v:

$$F_V(v) = \mathrm{P}(V \leq v) = \mathrm{P}(\max(X_1, X_2) \leq v) = \mathrm{P}(X_1 \leq v \cap X_2 \leq v)$$

$$\overset{Uavh.}{=} \mathrm{P}(X_1 \leq v) \cdot \mathrm{P}(X_2 \leq v) = (F_X(v))^2 = (1 - \mathrm{e}^{-\lambda v})^2 = 1 - 2\mathrm{e}^{-\lambda v} + \mathrm{e}^{-2\lambda v}.$$

Vi har videre

$$f_V(v) = \frac{d}{dv} F_V(v) = 2\,\lambda\,\mathrm{e}^{-\lambda v} - 2\,\lambda\,\mathrm{e}^{-2\,\lambda v}.$$

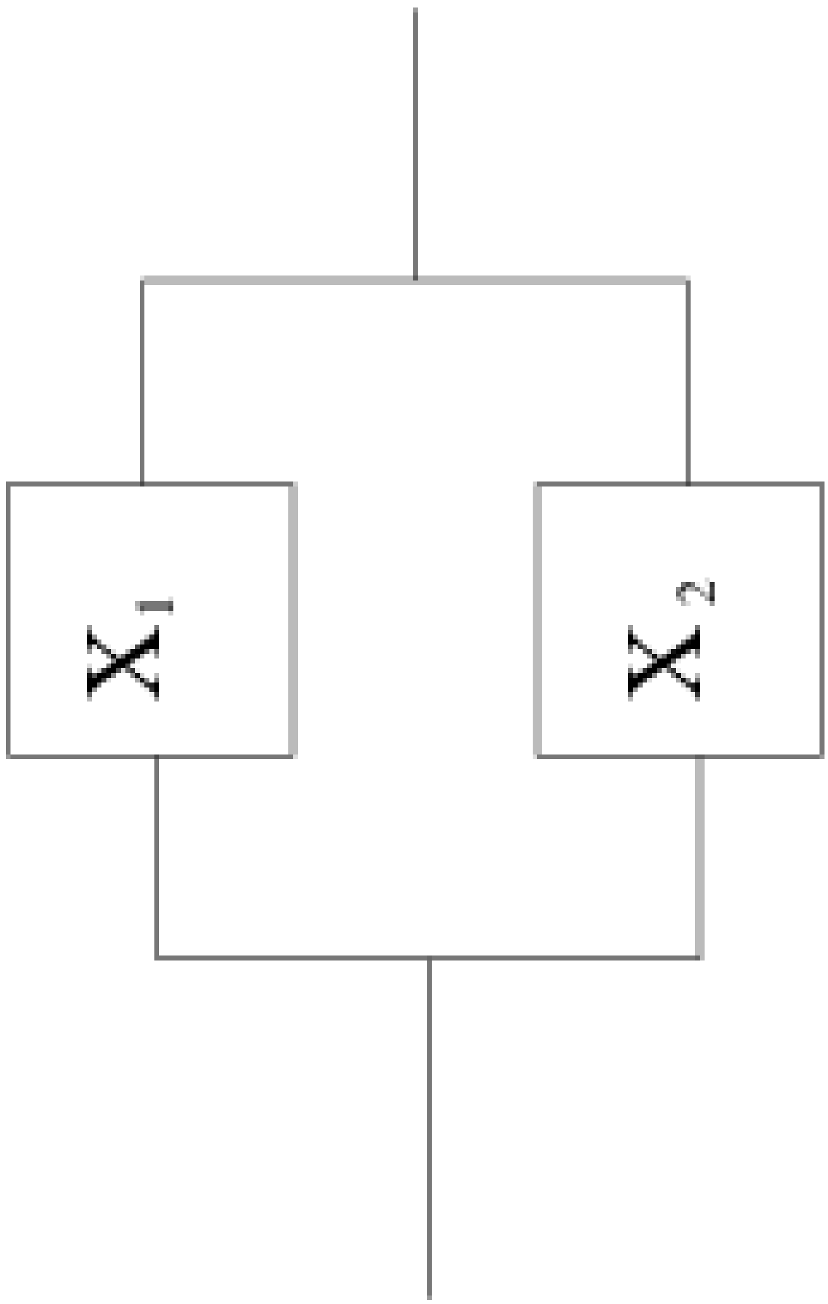
Forventningen til V er gitt ved

$$E(V) = \int_{-\infty}^{\infty} v f_V(v)\, dv.$$

Ved delvis integrasjon får vi dermed

$$E(V) = \int_{-\infty}^{\infty} v f_V(v)\, dv = \int_{0}^{\infty} (2\,\lambda\, v\,\mathrm{e}^{-\lambda v} - 2\,\lambda\, v\,\mathrm{e}^{-2\,\lambda v})\, dv = \frac{3}{2\lambda}.$$

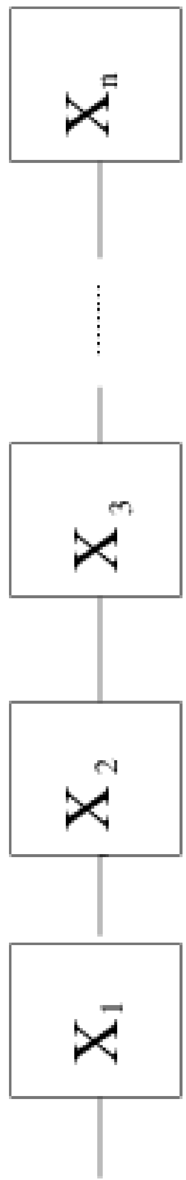
Oppgave 2



Figur 1: Paralellkopling av to komponenter

Betrakt et seriesystem sammensatt av n uavhengige komponenter der levetiden til hver komponent følger en Weibull-fordeling med skalaparameter λ og formparameter α, gitt ved

$$F_X(x) = 1 - \mathrm{e}^{-(\lambda\, x)^\alpha} \text{ for } x \geq 0.$$



Figur 2: Seriekopling av n komponenter

Da vi har et seriesystem, vil systemet fungere frem til første komponent svikter. Vi lar dermed levetiden til systemet betegnes ved $U = \min(X_1, X_2, ..., X_n)$.

$$F_U(u) = \mathrm{P}(U \leq u) = \mathrm{P}(\min(X_1, X_2, ..., X_n) \leq u) = 1 - \mathrm{P}(\min(X_1, X_2, ..., X_n) > u)$$

$$= 1 - \mathrm{P}(X_1 > u \cap X_2 > u \cap \cdots \cap X_n > u) \overset{Uavh.}{=} 1 - \prod_{i=1}^{n} \mathrm{P}(X_i > u)$$

$$= 1 - (1 - F_X(u))^n = 1 - (1 - (1 - \mathrm{e}^{-(\lambda\, x)^\alpha}))^n$$

$$= 1 - (\mathrm{e}^{-(\lambda\, x)^\alpha})^n = 1 - \mathrm{e}^{-(n^{1/\alpha}\, \lambda\, x)^\alpha}.$$

Dette er en Weibull-fordeling med skalaparameter $n^{1/\alpha}\lambda$ og formparameter α.

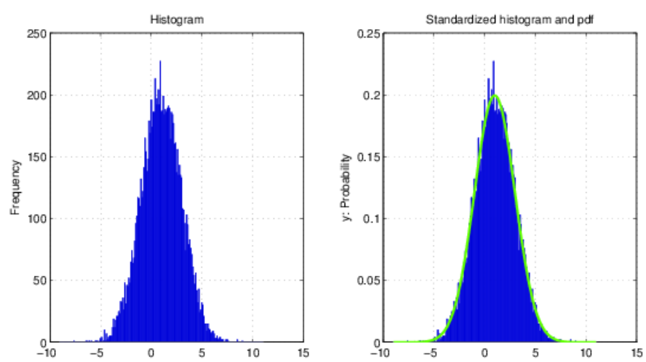
Oppgave 3

Scriptet *run_ confds.m* simulerer n data $x_1, \ldots, x_n$ fra en normalfordeling med forventningsverdi $\mu = 1$ og varians $\sigma^2 = 2^2$ ved å trekke n ganger fra en standard normalfordeling $y_i \sim N(0,1)$ og utføre lineærtransformasjonen

$$x_i = \mu + \sigma \cdot y_i \ \ , \ \ i = 1, \ldots, n$$

Fra uttrykket kan vi greit regne på at da vil $x_i \sim N(\mu, \sigma^2)$. *(I Matlab trekker man fra en standard normalfordeling med funksjonen 'randn').*

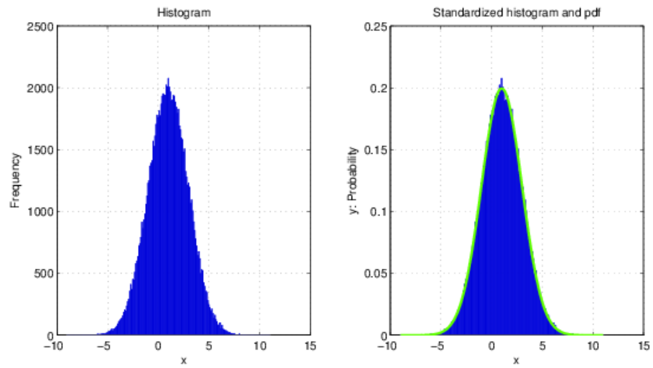
Kjører vi scriptet får vi et histogram av $n = 10000$ simulerte data $x_1, \ldots, x_n$, som f.eks. kan se slik ut



Figur 3: Histogram av $n = 10000$ simulerte data fra $N(1, 2^2)$

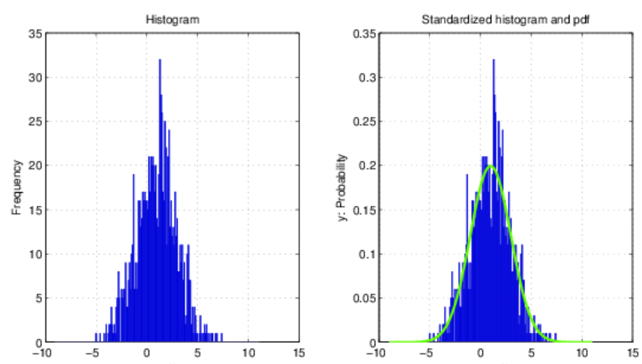
Histogrammet til høyre er standardisert, altså transformert slik at areal under histogramsøylene blir 1. I plottet er det i grønt også tegnet inn kurven for normalfordelingen med forventning 1 og standardavvik 2. Vi ser at de simulerte dataene overlapper normalfordelingen de kommer fra veldig bra. Dette siden vi simulerer såpass mange datapunkter. Det resulterende gjennomsnittet $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = 1.0047$ er veldig nærme den sanne forventningsverdien som også ligger innenfor det estimerte konfidensintervalet [0.96591.0434].

Trekker vi stedet $n = 100000$ data (setter altså parameteren 'n' i scriptet til 100000) kan histogrammet f.eks. se ut som i Fig.2 med estimert forventningsverdi $\hat{\mu} = 0.9983$ og estimert 95% konfidensinterval $[0.9859, 1.0107]$. Igjen er estimatet tilnærmet likt sann forventnigsverdi, som ligger innenfor konfidensintervalet, og overlappen mellom dataene og normalkurven er enda bedre.



Figur 4: Histogram av $n = 100000$ simulerte data fra $N(1, 2^2)$

Trekker vi $n = 1000$ data (setter altså parameteren 'n' i scriptet til 1000) kan histogrammet f.eks. se ut som i Fig.3. med estimert forventningsverdi $\hat{\mu} = 0.9594$ og estimert 95% konfidensinterval $[0.83741.0815]$. Estimatet er fortsatt bra, men ikke like nærme som i tilfellene med høyere n. Vi ser også at estimert konfidensinterval er litt bredere, og at overlappen mellom dataene og normalkurven er dårligere (dette er også fordi vi har så liten oppløsning på histogrammet).



Figur 5: Histogram av $n = 1000$ simulerte data fra $N(1, 2^2)$

Det estimerte konfidensintervalet er beregnet som

$$\left[\hat{\mu} - 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \ , \ \hat{\mu} + 1.96 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \right]$$

Når datamengden vokser og estimatet på standardaviket ikke varierer mye ser vi at faktoren $\frac{\hat{\sigma}}{\sqrt{n}}$ går mot 0, altså blir konfidensintervalet smalere jo større datamengden er. Vi merker oss også at vi her har brukt kvantilen $z_{0.025} = 1.96$ fra en normalfordeling selv om vi her bruker estimert varians. Med ukjent varians burde vi egentlig brukt kvantiler fra t-fordeling, men siden datamangden er så stor ($n \geq 1000$) vil t-fordeling med $n-1$ frihetsgrader være tilnærmet lik standard normalfordeling.

Oppgave 4

a) Sannsynligheten for at et batteri virker etter 130 timer er

$$p = P(T > 130) = 1 - P(\leq 130) = 1 - P\left(\frac{T-\mu}{\sigma} \leq \frac{130 - 117.2}{10}\right)$$
$$= 1 - P(Z \leq 1.28) = 1 - 0.8997 = 0.1003$$

Vi har $n = 8$ slike batterier og de er uavhengig av hverandre. Hvis X er antall batterier som virker etter 130 timer, har vi at X består av n uavhengige forsøk, hver med sannsylighet p for 'suksess', X er derfor binomisk fordelt. Radioen virker dersom minst 4 batterier:

$$P(\text{Radioen virker}) = P(X \geq 4) = 1 - P(X \leq 3)$$
$$= 1 - \sum_{x=0}^{3} p(x; n = 8, p = 0.1) = 1 - 0.995 = \underline{\underline{0.005}}$$

b)

$$M_X(t) = \mathrm{E}[e^{tX}] = \sum_{x=0}^{n} e^{tx} p(x; n, p) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (pe^t + 1 - p)^n$$

$$\mathrm{E}[X] = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = n(pe^t + 1 - p)^{n-1}(pe^t)\big|_{t=0}$$
$$= n(p + 1 - p)^{n-1} p = \underline{\underline{np}}$$

Oppgave 5

a) La V være målt vekt, slik at $V \sim N(\mu, \sigma^2) = N(10, 0.2^2)$. Vi får

$$P(V > 10.2) = P\left(\frac{V-\mu}{\sigma} > \frac{10.2 - 10}{0.2}\right) = P(Z > 1)$$
$$= 1 - P(Z \leq 1) = 1 - 0.8413 = \underline{\underline{0.1587}}$$

$$\begin{aligned}
P(|V-\mu|>0.2) &= P(V-\mu>0.2)+P(V-\mu<-0.2) \\
&= P\left(\frac{V-\mu}{\sigma}>\frac{0.2}{0.2}\right)+P\left(\frac{V-\mu}{\sigma}<-\frac{0.2}{0.2}\right) \\
&= P(Z>1)+P(Z\leq -1) = 1-P(Z\leq 1)+P(Z\leq -1) \\
&= 2\cdot P(Z\leq -1) = 2\cdot 0.1587 = \underline{\underline{0.3174}}
\end{aligned}$$

La $\bar{V}=\frac{1}{n}\sum_{i=1}^{n}V_i$, slik at $\bar{V}\sim N(\mu,\sigma^2/n)$. Vi får

$$\begin{aligned}
P(|\bar{V}-\mu|>0.2) &= P(\bar{V}-\mu>0.2)+P(\bar{V}-\mu<-0.2) \\
&= P\left(\frac{\bar{V}-\mu}{\sigma/\sqrt{n}}>\frac{0.2}{0.2/\sqrt{2}}\right)+P\left(\frac{\bar{V}-\mu}{\sigma/\sqrt{n}}<-\frac{0.2}{0.2/\sqrt{2}}\right) \\
&= P(Z>\sqrt{2})+P(Z\leq -\sqrt{2}) \\
&= 1-P(Z\leq \sqrt{2})+P(Z\leq -\sqrt{2}) \\
&= 2\cdot P(Z\leq -1.41) = 2\cdot 0.0793 = \underline{\underline{0.1586}}
\end{aligned}$$

b) Vi har $X_1\sim N(\mu_A,\sigma^2)$ og $X_2\sim N(\mu_B,\sigma^2)$ som er uavhengig av hverandre. Vi får ved fremgangsmåte 1:

$$\begin{aligned}
\mathrm{E}[\hat{\mu}_A] &= \mathrm{E}[X_1] = \mu_A \\
\mathrm{Var}[\hat{\mu}_A] &= \mathrm{Var}[X_1] = \sigma^2 \\
\mathrm{E}[\hat{\mu}_B] &= \mathrm{E}[X_2] = \mu_B \\
\mathrm{Var}[\hat{\mu}_B] &= \mathrm{Var}[X_2] = \sigma^2
\end{aligned}$$

Vi har $Y_1\sim N(\mu_A+\mu_B,\sigma^2)$ og $Y_2\sim N(\mu_A-\mu_B,\sigma^2)$ som er uavhengig av hverandre. Vi får ved fremgangsmåte 2:

$$\begin{aligned}
\mathrm{E}[\tilde{\mu}_A] &= \mathrm{E}[(Y_1+Y_2)/2] = \frac{1}{2}(\mathrm{E}[Y_1]+\mathrm{E}[Y_2]) = \frac{1}{2}(\mu_A+\mu_B+\mu_A-\mu_B) = \mu_A \\
\mathrm{Var}[\tilde{\mu}_A] &= \mathrm{Var}[(Y_1+Y_2)/2] = \frac{1}{4}(\mathrm{Var}[Y_1]+\mathrm{Var}[Y_2]) = \frac{1}{4}(\sigma^2+\sigma^2) = \sigma^2/2 \\
\mathrm{E}[\tilde{\mu}_B] &= \mathrm{E}[(Y_1-Y_2)/2] = \frac{1}{2}(\mathrm{E}[Y_1]-\mathrm{E}[Y_2]) = \frac{1}{2}(\mu_A+\mu_B-\mu_A+\mu_B) = \mu_B \\
\mathrm{Var}[\tilde{\mu}_B] &= \mathrm{Var}[(Y_1-Y_2)/2] = \frac{1}{4}(\mathrm{Var}[Y_1]+\mathrm{Var}[Y_2]) = \frac{1}{4}(\sigma^2+\sigma^2) = \sigma^2/2
\end{aligned}$$

Begge fremgangsmåtene gir forventningsrette estimatorer, så vi velger den med minst varians, dvs. fremgangsmåte 2: $\tilde{\mu}_A$ og $\tilde{\mu}_B$.

c) Vi har $\tilde{\mu}_A=u_1(Y_1,Y_2)=(Y_1+Y_2)/2$ og $\tilde{\mu}_B=u_2(Y_1,Y_2)=(Y_1-Y_2)/2$, som gir oss at $Y_1=w_1(\tilde{\mu}_A,\tilde{\mu}_B)=\tilde{\mu}_A+\tilde{\mu}_B$ og $Y_2=w_2(\tilde{\mu}_A,\tilde{\mu}_B)=\tilde{\mu}_A-\tilde{\mu}_B$. Fra transformasjonsformelen for to variabler har vi da at

$$g_{\tilde{\mu}_A,\tilde{\mu}_B}(\tilde{\mu}_A,\tilde{\mu}_B) = f_{Y_1,Y_2}(w_1(\tilde{\mu}_A,\tilde{\mu}_B),w_2(\tilde{\mu}_A,\tilde{\mu}_B))\cdot|J|$$

hvor

$$J=\begin{vmatrix}\delta w_1/\delta\tilde{\mu}_A & \delta w_1/\delta\tilde{\mu}_B \\ \delta w_2/\delta\tilde{\mu}_A & \delta w_2/\delta\tilde{\mu}_B\end{vmatrix} = \begin{vmatrix}1 & 1 \\ 1 & -1\end{vmatrix} = -2.$$

Siden Y_1 og Y_2 er uavhengige, har vi $f_{Y_1,Y_2}(y_1, y_1) = f_{Y_1}(y_1)f_{Y_2}(y_2)$ og vi får følgende:

$$\begin{aligned}
g_{\tilde{\mu}_A,\tilde{\mu}_B}(\tilde{\mu}_A, \tilde{\mu}_B) &= f_{Y_1,Y_2}(w_1(\tilde{\mu}_A, \tilde{\mu}_B), w_2(\tilde{\mu}_A, \tilde{\mu}_B)) \cdot |J| \\
&= f_{Y_1}(w_1(\tilde{\mu}_A, \tilde{\mu}_B))f_{Y_2}(w_2(\tilde{\mu}_A, \tilde{\mu}_B)) \cdot |-2| \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}(\tilde{\mu}_A + \tilde{\mu}_B - (\mu_A + \mu_B))^2\right\} \\
&\quad \cdot \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}(\tilde{\mu}_A - \tilde{\mu}_B - (\mu_A - \mu_B))^2\right\} \cdot 2 \\
&= \left(\frac{1}{\sqrt{2\pi}}\frac{\sqrt{2}}{\sigma}\right)^2 \exp\left\{-\frac{1}{2\sigma^2}\left[(\tilde{\mu}_A + \tilde{\mu}_B)^2 - 2((\tilde{\mu}_A + \tilde{\mu}_B)\right.\right. \\
&\quad \cdot(\mu_A + \mu_B) + (\mu_A + \mu_B)^2 + (\tilde{\mu}_A - \tilde{\mu}_B)^2 - 2(\tilde{\mu}_A - \tilde{\mu}_B) \\
&\quad \left.\left.\cdot(\mu_A - \mu_B) + (\mu_A - \mu_B)^2\right]\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\frac{\sqrt{2}}{\sigma}\right)^2 \exp\left\{-\frac{1}{2\sigma^2}\left[\tilde{\mu}_A^2 + 2\tilde{\mu}_A\tilde{\mu}_B + \tilde{\mu}_B^2 - 2\tilde{\mu}_A\mu_A\right.\right. \\
&\quad -2\tilde{\mu}_A\mu_B - 2\tilde{\mu}_B\mu_B + \mu_A^2 + 2\mu_A\mu_B + \mu_B^2 + \tilde{\mu}_A^2 - 2\tilde{\mu}_A\tilde{\mu}_B \\
&\quad +\tilde{\mu}_B^2 - 2\tilde{\mu}_A\mu_A + 2\tilde{\mu}_A\mu_B + 2\tilde{\mu}_B\mu_A - 2\tilde{\mu}_B\mu_B + \tilde{\mu}_A^2 \\
&\quad \left.\left.-2\mu_A\mu_B + \mu_B^2\right]\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\frac{\sqrt{2}}{\sigma}\right)^2 \exp\left\{-\frac{1}{2\sigma^2}\left[2\tilde{\mu}_A^2 + 2\tilde{\mu}_B^2 - 4\tilde{\mu}_A\mu_A - 4\tilde{\mu}_B\mu_B\right.\right. \\
&\quad \left.\left.+2\tilde{\mu}_A^2 + 2\tilde{\mu}_B^2\right]\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}}\frac{\sqrt{2}}{\sigma}\right)^2 \exp\left\{-\frac{2}{2\sigma^2}\left[(\tilde{\mu}_A - \mu_A)^2 + (\tilde{\mu}_B - \mu_B)^2\right]\right\} \\
&= \frac{1}{\sqrt{2\pi}}\frac{\sqrt{2}}{\sigma}\exp\left\{-\frac{2}{2\sigma^2}(\tilde{\mu}_A - \mu_A)^2\right\} \\
&\quad \cdot \frac{1}{\sqrt{2\pi}}\frac{\sqrt{2}}{\sigma}\exp\left\{-\frac{2}{2\sigma^2}(\tilde{\mu}_B - \mu_B)^2\right\} \\
&= g_{\tilde{\mu}_A}(\tilde{\mu}_A)g_{\tilde{\mu}_B}(\tilde{\mu}_B)
\end{aligned}$$

og dermed er $\tilde{\mu}_A$ og $\tilde{\mu}_B$ uavhengige ($\tilde{\mu}_A \sim N(\mu_A, \sigma^2/2)$ og $\tilde{\mu}_B \sim N(\mu_B, \sigma^2/2)$).