Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen desember 2009

Oppgave 1

a) Bestemmer k ved å kreve $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$,

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^1 k(4x+1)\mathrm{d}x = \left[k\left(2x^2+x\right)\right]_0^1 = k(2+1) = 1 \Rightarrow \underline{\underline{k = \frac{1}{3}}}.$$

Den kumulative fordelingsfunksjonen $F(x)$ er gitt ved

$$F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u.$$

Hvis $x < 0$ blir $F(x) = 0$, for $x \in [0,1]$ får vi

$$F(x) = \int_0^x f(u)\mathrm{d}u = \int_0^x \frac{1}{3}(4u+1)\,\mathrm{d}u = \left[\frac{1}{3}\left(2u^2+u\right)\right]_0^x = \frac{2x^2+x}{3},$$

og for $x > 1$ får vi $F(x) = 1$. Dvs.

$$\underline{\underline{F(x) = \begin{cases} 0 & \text{for } x < 0, \\ \frac{2x^2+x}{3} & \text{for } x \in [0,1], \\ 1 & \text{for } x > 1. \end{cases}}}$$

Videre får vi da

$$\mathrm{P}\left(X > \frac{1}{4}\right) = 1 - \mathrm{P}\left(X \leq \frac{1}{4}\right) = 1 - F\left(\frac{1}{4}\right) = 1 - \frac{1}{3}\left(2\cdot\left(\frac{1}{4}\right)^2 + \frac{1}{4}\right) = \underline{\underline{0.875}}$$

og

$$\mathrm{P}\left(X > \frac{1}{2}\,\middle|\, X > \frac{1}{4}\right) = \frac{\mathrm{P}\left(X > \frac{1}{2} \cap X > \frac{1}{4}\right)}{\mathrm{P}\left(X > \frac{1}{4}\right)} = \frac{\mathrm{P}\left(X > \frac{1}{2}\right)}{\mathrm{P}\left(X > \frac{1}{4}\right)} = \frac{1 - \mathrm{P}\left(X \leq \frac{1}{2}\right)}{\mathrm{P}\left(X > \frac{1}{4}\right)}$$

$$= \frac{1 - F\left(\frac{1}{2}\right)}{\mathrm{P}\left(X > \frac{1}{4}\right)} = \frac{1 - \frac{1}{3}\left(2\left(\frac{1}{2}\right)^2 + \frac{1}{2}\right)}{0.875} = \underline{\underline{0.762}}.$$

b) Simultan sannsynlighetstetthet er gitt som

$$f(x,y) = f(x)f(y|x) = \frac{1}{3}(4x+1) \cdot \frac{4x+2y}{4x+1} = \frac{2}{3}(2x+y) \quad \text{for } x,y \in [0,1],$$

og $f(x,y) = 0$ ellers. Marginal sannsynlighetstetthet for Y finnes ved å integrere over x, dvs. for $y \in [0,1]$

$$f(y) = \int_0^1 f(x,y)\mathrm{d}x = \int_0^1 \frac{2}{3}(2x+y)\,\mathrm{d}x = \left[\frac{2}{3}\left(x^2+yx\right)\right]_{x=0}^{x=1}$$

$$= \frac{2}{3}\left(1^2 + y \cdot 1\right) = \underline{\underline{\frac{2}{3}(y+1)}}.$$

For $y \notin [0,1]$ får vi åpenbart $f(y) = 0$. Til slutt regner vi ut

$$\mathrm{P}(Y \leq X) = \int_0^1 \int_0^x f(x,y)\mathrm{d}y\mathrm{d}x = \int_0^1 \int_0^x \frac{2}{3}(2x+y)\,\mathrm{d}y\mathrm{d}x$$

$$\int_0^1 \left[\frac{2}{3}\left(2xy + \frac{1}{2}y^2\right)\right]_{y=0}^{y=x} \mathrm{d}x = \int_0^1 \frac{2}{3}\left(2x^2 + \frac{1}{2}x^2\right)\mathrm{d}x = \int_0^1 \frac{5}{3}x^2\mathrm{d}x = \left[\frac{5}{9}x^3\right]_0^1 = \underline{\underline{\frac{5}{9}}}.$$

Oppgave 2

a) For at X_i'ene skal være negativt binomisk fordelt må det være uavhengighet mellom hver gang en student ringer til firmaet. Dette kan oppnås ved at studentene ringer til firmaet på tilfeldige tidspunkter. Det fungerer for eksempel ikke at en student ringer til firmaet flere ganger like etter hverandre.

Dessuten har studentene antatt at $X_1, X_2, \ldots, X_n$ er uavhengige. For at dette skal være oppfylt må studentene ringe på tilfeldige tidspunkt uavhengig av hverandre. Flere studenter kan ikke ringe på samme tidspunkt.

For $k = 2$ og $p = 0.1$ får vi

$$\mathrm{P}(X_1 = 2) = \binom{2-1}{2-1} 0.1^2(1-0.1)^{2-2} = 0.1^2 = \underline{\underline{0.01}}$$

og

$$\mathrm{P}(X_1 \leq 4 | X_1 > 2) = \frac{\mathrm{P}(X_1 \leq 4 \cap X_1 > 2)}{\mathrm{P}(X_1 > 2)} = \frac{\mathrm{P}(X_1 = 3 \cup X_1 = 4)}{1 - \mathrm{P}(X_1 = 2)}$$

$$= \frac{\mathrm{P}(X_1 = 3) + \mathrm{P}(X_1 = 4)}{1 - \mathrm{P}(X_1 = 2)} = \frac{\binom{3-1}{2-1} 0.1^2(1-0.1)^{3-2} + \binom{4-1}{2-1} 0.1^2(1-0.1)^{4-2}}{1 - 0.01}$$

$$= \frac{2 \cdot 0.01 \cdot 0.9 + 3 \cdot 0.01 \cdot 0.9^2}{0.99} = \underline{\underline{0.0427}}.$$

b) Rimelighetsfunksjonen blir

$$L(p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left[\binom{x_i - 1}{k - 1} p^k (1-p)^{x_i - k} \right]$$

Tar logaritmen og får

$$\begin{aligned} l(p) = \ln L(p) &= \sum_{i=1}^{n} \left[\ln \left(\binom{x_i - 1}{k - 1} \right) + k \ln(p) + (x_i - k) \ln(1-p) \right] \\ &= \left[\sum_{i=1}^{n} \ln \left(\binom{x_i - 1}{k - 1} \right) \right] + nk \ln(p) + \ln(1-p) \sum_{i=1}^{n} (x_i - k). \end{aligned}$$

Deriverer med hensyn på p og setter lik null for å finne maksimum

$$l'(p) = 0 + \frac{nk}{p} - \frac{1}{1-p} \sum_{i=1}^{n} (x_i - k) = \frac{nk}{p} - \frac{\sum_{i=1}^{n} x_i - nk}{1-p} = 0$$

$$nk(1-p) = \left(\sum_{i=1}^{n} x_i - nk \right) p$$

$$nk - nkp = p \sum_{i=1}^{n} x_i - nkp \Rightarrow p = \frac{nk}{\sum_{i=1}^{n} x_i}.$$

Sannsynlighetsmaksimeringsestimatoren for p blir dermed

$$\underline{\underline{\widehat{p} = \frac{nk}{\sum_{i=1}^{n} X_i}.}}$$

c) Studentene ønsker å teste

$$\mathrm{H}_0 : p = 0.1 \quad \text{mot} \quad \mathrm{H}_1 : p < 0.1.$$

Som testobservator kan man benytte (men andre valg finnes også)

$$V = \sum_{i=1}^{n} X_i = \frac{nk}{\widehat{p}}$$

og forkaste H_0 hvis $V > k^\star$, for en kritisk verdi $k^\star$.

Fra sentralgrenseteorem følger det at $V/n = \bar{X}$ er tilnærmet normalfordelt. Dermed vil også V være tilnærmet normalfordelt. Vi har

$$\mathrm{E}[V] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{k}{p} = \frac{nk}{p}$$

og, siden X_i'ene er uavhengige,

$$\mathrm{Var}[V] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = \sum_{i=1}^{n} \frac{k(1-p)}{p^2} = \frac{nk(1-p)}{p^2}.$$

Når H_0 er riktig har vi dermed at

$$V \sim \mathrm{n}\left(z; \frac{nk}{p_0}, \sqrt{\frac{nk(1-p_0)}{p_0^2}}\right).$$

Innsatt studentenes observasjoner får vi at

$$v = \sum_{i=1}^{n} x_i = 779.$$

Tilhørende p-verdi blir dermed

$$p = \mathrm{P}\left(V > 779 | \mathrm{H}_0 \text{ riktig}\right) = \mathrm{P}\left(\frac{V - \frac{nk}{p_0}}{\sqrt{\frac{nk(1-p_0)}{p_0^2}}} > \frac{779 - \frac{50\cdot 2}{0.1}}{\sqrt{\frac{50\cdot 2\cdot(1-0.1)}{0.1^2}}} \middle| \mathrm{H}_0 \text{ riktig}\right)$$

$$1 - \mathrm{P}(Z \leq -2.33) = 1 - 0.0099 = \underline{\underline{0.9901}}.$$

Dersom H_0 er riktig er altså sannsynligheten for å observere det studentene har observert eller noe mer ekstremt så stor som 0.9901. Det er dermed intet grunnlag for å hevde at firmaets påstand er feil.

Oppgave 3

a) Forventningsverdien for $\widehat{\mu}$ blir

$$\mathrm{E}\left[\widehat{\mu}\right] = \mathrm{E}\left[\frac{a}{n}\sum_{i=1}^{n} X_i + \frac{b}{m}\sum_{i=1}^{m} Y_i\right] = \frac{a}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] + \frac{b}{m}\mathrm{E}\left[\sum_{i=1}^{m} Y_i\right]$$

$$= \frac{a}{n}\sum_{i=1}^{n} \mathrm{E}\left[X_i\right] + \frac{b}{m}\sum_{i=1}^{m} \mathrm{E}\left[Y_i\right] = \frac{a}{n}\sum_{i=1}^{n} \mu + \frac{b}{m}\sum_{i=1}^{m} \mu = a\mu + b\mu = \underline{\underline{(a+b)\mu}}.$$

Ved å benytte at alle X_i'ene og Y_i'ene er uavhengige av hverandre får vi at variansen til $\widehat{\mu}$ blir

$$\begin{aligned}
\mathrm{Var}\left[\widehat{\mu}\right] &= \mathrm{Var}\left[\frac{a}{n}\sum_{i=1}^{n} X_i + \frac{b}{m}\sum_{i=1}^{m} Y_i\right] = \left(\frac{a}{n}\right)^2 \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] + \left(\frac{b}{m}\right)^2 \mathrm{Var}\left[\sum_{i=1}^{m} Y_i\right] \\
&= \left(\frac{a}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}\left[X_i\right] + \left(\frac{b}{m}\right)^2 \sum_{i=1}^{m} \mathrm{Var}\left[Y_i\right] = \left(\frac{a}{n}\right)^2 \sum_{i=1}^{n} \sigma_A^2 + \left(\frac{b}{m}\right)^2 \sum_{i=1}^{m} \sigma_B^2 \\
&= \underline{\underline{\frac{a^2\sigma_A^2}{n} + \frac{b^2\sigma_B^2}{m}}}.
\end{aligned}$$

b) For at estimatoren skal bli best mulig må den være forventningsrett og variansen må være minst mulig. Kravet om forventningsretthet gir at

$$\mathrm{E}\left[\widehat{\mu}\right] = (a+b)\mu = \mu \Rightarrow a + b = 1 \Rightarrow b = 1 - a.$$

Setter vi dette inn i uttrykket for variansen får vi at

$$\mathrm{Var}\left[\widehat{\mu}\right] = \frac{a^2\sigma_A^2}{n} + \frac{(1-a)^2\sigma_B^2}{m}.$$

Må minimere dette uttrykket med hensyn på a. Gjør dette ved å derivere med hensyn på a og sette lik null:

$$\begin{aligned}
\frac{\partial}{\partial a}\mathrm{Var}\left[\widehat{\mu}\right] &= \frac{2a\sigma_A^2}{n} + \frac{2(1-a)(-1)\sigma_B^2}{m} = 0 \\
am\sigma_A^2 &= (1-a)n\sigma_B^2 \\
a(m\sigma_A^2 + n\sigma_B^2) = n\sigma_B^2 \Rightarrow a &= \frac{n\sigma_B^2}{m\sigma_A^2 + n\sigma_B^2} = \underline{\underline{\frac{\frac{\sigma_B^2}{m}}{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}}}.
\end{aligned}$$

og dermed også

$$b = 1 - a = \underline{\underline{\frac{\frac{\sigma_A^2}{n}}{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}}}.$$

c) $\widehat{\mu}$ er en lineærkombinasjon av uavhengige normalfordelte variabler (X_i'ene og Y_i'ene) og er derfor også selv normalfordelt. Dermed har vi at

$$\underline{\underline{\widehat{\mu} \sim \mathrm{n}\left(\widehat{\mu}; (a+b)\mu, \sqrt{\frac{a^2\sigma_A^2}{n} + \frac{b^2\sigma_B^2}{m}}\right)}}.$$

Dermed får vi også

$$\frac{\widehat{\mu}-(a+b)\mu}{\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}} \sim \mathrm{n}(z;0,1)$$

og dermed

$$\mathrm{P}\left(-z_{\frac{\alpha}{2}} \leq \frac{\widehat{\mu}-(a+b)\mu}{\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}} \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Løser vi hver av de to ulikhetene med hensyn på μ og setter de så sammen igjen får vi

$$\mathrm{P}\left(\frac{\widehat{\mu}-z_{\frac{\alpha}{2}}\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}}{a+b} \leq \mu \leq \frac{\widehat{\mu}+z_{\frac{\alpha}{2}}\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}}{a+b}\right).$$

$(1-\alpha)\cdot 100\%$-konfidensintervall for μ er dermed

$$\underline{\underline{\left[\frac{\widehat{\mu}-z_{\frac{\alpha}{2}}\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}}{a+b}, \frac{\widehat{\mu}+z_{\frac{\alpha}{2}}\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}}{a+b}\right]}}.$$

d) Lengden av konfidensintervallet blir

$$L = \frac{\widehat{\mu}+z_{\frac{\alpha}{2}}\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}}{a+b} - \frac{\widehat{\mu}-z_{\frac{\alpha}{2}}\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}}{a+b} = \frac{2z_{\frac{\alpha}{2}}\sqrt{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}}{a+b}$$

og dette uttrykket må følgelig minimeres med hensyn på a og b. For å få enklere regning deriverer vi $\ln(L)$ med hensyn på a og b og setter lik null. Deriverer først med hensyn på a

$$\frac{\partial \ln(L)}{\partial a} = \frac{\partial}{\partial a}\left[\ln(2z_{\frac{\alpha}{2}}) + \frac{1}{2}\ln\left(\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}\right) - \ln(a+b)\right] = \frac{1}{2}\frac{\frac{2a\sigma_A^2}{n}}{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}} - \frac{1}{a+b} = 0$$

Dette gir videre

$$\frac{\frac{a\sigma_A^2}{n}}{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}} = \frac{1}{a+b}$$

$$a+b = a+\frac{b^2}{a}\cdot\frac{\sigma_B^2/m}{\sigma_A^2/n} \Rightarrow \frac{b}{a} = \frac{\sigma_A^2/n}{\sigma_B^2/m}.$$

Tilsvarende fås ved å derivere med hensyn på b at

$$\frac{\partial \ln(L)}{\partial b} = \frac{\partial}{\partial b}\left[\ln(2z_{\frac{\alpha}{2}}) + \frac{1}{2}\ln\left(\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}\right) - \ln(a+b)\right] = \frac{1}{2}\frac{\frac{2b\sigma_B^2}{m}}{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}} - \frac{1}{a+b} = 0$$

som ved tilsvarende regning som over gir

$$\frac{\frac{b\sigma_B^2}{m}}{\frac{a^2\sigma_A^2}{n}+\frac{b^2\sigma_B^2}{m}}=\frac{1}{a+b}$$

$$a+b=\frac{a^2}{b}\cdot\frac{\sigma_A^2/n}{\sigma_B^2/m}+b\Rightarrow\frac{b}{a}=\frac{\sigma_A^2/n}{\sigma_B^2/m}$$

som er det samme uttrykket som vi fant over. Lengden av konfidensintervallet er dermed minimert ved å velge

$$\underline{\underline{\frac{b}{a}=\frac{\sigma_A^2/n}{\sigma_B^2/m}}}.$$

Vi kan legge merke til at de optimale verdiene vi fant for a og b i punkt **b**) også gir samme verdi for forholdet b/a.