MNFST 102 3.12.2001

Oppgave 1.

a) • Må sjekke at $P(X=x) \geq 0$ for alle x (det er OK her)

• Må sjekke at $\sum_{x=1}^{n} P(X=x) = 1$. (*)

Her er $\sum_{x=1}^{n} \binom{n}{x} p^x ((1-p)^{n-x} = 1-(1-p)^n$

siden $\sum_{x=0}^{n}$ — " — $= 1$.

Dermed holder også (*)

b) I utgangspunktet vil X = ant. barn med A av disse foreldrene være $\text{bin}(n,p)$.
[Forutsetninger: Uavhengighet mellom barn, samme sannsynlighet for alle barn].
Vi observerer X bare hvis $X \geq 1$. Den relevante fordeling er derfor $P(X=x \mid X \geq 1)$ som er den som er gitt ved (1).

c) $$M_X(t) = \sum_{x=1}^{n} e^{tx} \frac{1}{1-(1-p)^n} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \frac{1}{1-(1-p)^n} \left[\sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} - (1-p)^n \right]$$

$$= \frac{(e^t p + 1-p)^n - ((1-p)^n}{1-((1-p)^n}$$

Se at $M_X(t) = \dfrac{M_Y(t) - (1-p)^n}{1-(1-p)^n}$ for $Y \sim bin(n,p)$

Dermed er

$$M_X'(t) = \frac{M_Y'(t)}{1-(1-p)^n}$$

$$M_X''(t) = \frac{M_Y''(t)}{1-(1-p)^n}$$

Det følger at

$$\underline{E(X)} = M_X'(0) = \frac{M_Y'(0)}{1-(1-p)^n} = \underline{\underline{\frac{np}{1-(1-p)^n}}}$$

$$E(X^2) = M_X''(0) = \frac{M_Y''(0)}{1-(1-p)^n} = \frac{np(1-p) + n^2p^2}{1-(1-p)^n}$$

Dermed er

$$Var(X) = \frac{np(1-p) + n^2p^2}{1-(1-p)^n} - \frac{n^2p^2}{1-(1-p)^n}$$

$$= np(1-p)\left[\frac{1}{1-(1-p)^n} - \frac{np(1-p)^{n-1}}{(1-(1-p)^n)^2}\right]$$

etter litt mellomregning.

d) $E(\hat{p}_1) = \dfrac{E(X)}{n} = \dfrac{p}{1-(1-p)^n} \neq p$ for alle $0 < p < 1$.

e) $$\hat{p}_2 = \frac{X-k}{n} = \hat{p}_1 - \frac{k}{n} \qquad (**)$$

$$E(\hat{p}_2) = E(\hat{p}_1) - \frac{k}{n} = \frac{p}{1-(1-p)^n} - \frac{k}{n} \qquad (***)$$

Vi ser at

$$E(\hat{p}_2) = p \quad \text{for alle } 0<p<1$$

$$\Updownarrow$$

$$p \cdot \frac{(1-p)^n}{1-(1-p)^n} = \frac{k}{n} \quad \text{for alle } 0<p<1$$

Men funksjonen på venstre side er ikke konstant i p, så dette kan ikke være tilfelle.

$$E[(\hat{p}-p)^2] = E[(\hat{p} - E(\hat{p}) + E(\hat{p}) - p)^2]$$

$$= E[(\hat{p} - E(\hat{p}))^2 + 2(\hat{p} - E(\hat{p}))(E(\hat{p}) - p) + (E(\hat{p}) - p)^2]$$

$$= Var(\hat{p}) + 0 + (E(\hat{p}) - p)^2$$

Her:

$$E[(\hat{p}_1 - p)^2] = Var(\hat{p}_1) + (E(\hat{p}_1) - p)^2$$

$$= Var(\hat{p}_1) + \left(\frac{p}{1-(1-p)^n} - p\right)^2$$

$$= Var(\hat{p}_1) + \left(\frac{p(1-p)^{n}}{1-(1-p)^{n}} \right)^2$$

Videre er fra (**)

$$E[(\hat{p}_2 - p)^2] = Var(\hat{p}_1) + \left(E(\hat{p}_1) - p - \frac{k}{n}\right)^2$$

$$= Var(\hat{p}_1) + \left(\frac{p(1-p)^{n}}{1-(1-p)^{n}} - \frac{k}{n} \right)^2$$

Dermed er differansen i bruttovarians for $\hat{p}_1$ og $\hat{p}_2$ gitt ved

$$\left(\frac{p(1-p)^{n}}{1-(1-p)^{n}} \right)^2 - \left(\frac{p(1-p)^{n}}{1-(1-p)^{n}} - \frac{k}{n} \right)^2$$

$$= \frac{2k\,p(1-p)^{n}}{n\left(1-(1-p)^{n}\right)} - \frac{k^2}{n^2}$$

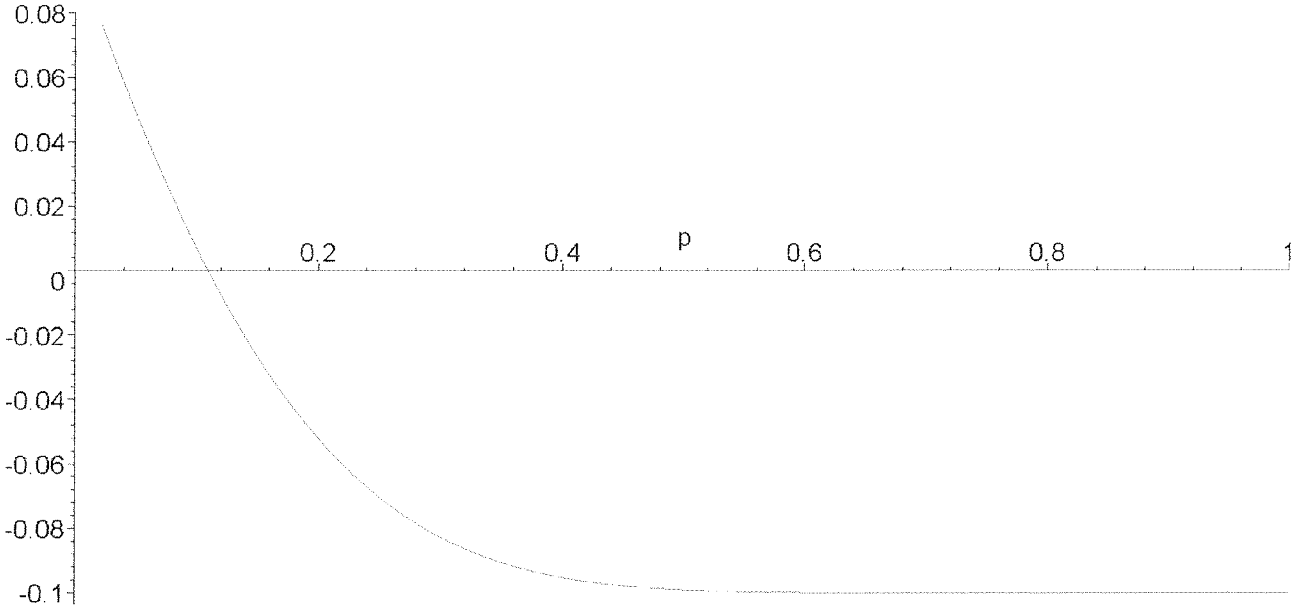
$$= \frac{k}{n} \left[\frac{2p(1-p)^{n}}{1-(1-p)^{n}} - \frac{k}{n} \right]$$

$k=1,\ n=10$:

$$\frac{1}{10} \left[\frac{2p(1-p)^{10}}{1-(1-p)^{10}} - \frac{1}{10} \right]$$

Har plottet denne (se neste side).

Ser at den er positiv for $p < 0.1$ (ca.) (dvs $\hat{p}_2$ best der) og negativ for $p > 0.1$ ($\hat{p}_1$ best der)



Kurven $\dfrac{2p(1-p)^{10}}{1-(1-p)^{10}} - \dfrac{1}{10}$.

<u>Dermed:</u> Hvis p anses liten, vil $\hat{p}_2$ være å foretrekke. Hvis p anses middels stor eller stor, vil $\hat{p}_1$ være å foretrekke.

Oppgave 2.

a) Likelihood:

$$L = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}(x_i-\mu_0)^2} \cdot \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\tau} e^{-\frac{1}{2\tau^2}(y_j-\mu_0)^2}$$

$$= \frac{1}{(2\pi)^{\frac{m+n}{2}}} \cdot \frac{1}{\sigma^m} \cdot \frac{1}{\tau^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu_0)^2 - \frac{1}{2\tau^2}\sum_{j=1}^{n}(y_j-\mu_0)^2}$$

Log-likelihood:

$$l = -\frac{m+n}{2}\ln(2\pi) - \frac{m}{2}\ln(\sigma^2) - \frac{n}{2}\ln(\tau^2)$$

$$- \frac{1}{2\sigma^2}\sum(x_i-\mu_0)^2 - \frac{1}{2\tau^2}\sum(y_j-\mu_0)^2$$

$$\frac{\partial l}{\partial(\sigma^2)} = -\frac{m}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum(x_i-\mu_0)^2$$

Settes dette lik 0 fås $\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_0)^2$

Tilsvarende fås $\hat{\tau}^2 = \frac{1}{n}\sum_{j=1}^{n}(Y_j-\mu_0)^2$

b) Vi kan skrive $\frac{m\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{m}\Big(\underbrace{\frac{X_i-\mu_0}{\sigma}}_{Z_i \sim N(0,1)}\Big)^2 = \sum_{i=1}^{m}\underbrace{Z_i^2}_{\chi^2_1} \sim \chi^2_m$

fordi $(N(0,1))^2 \sim \chi^2_1$, og sum av uavh. χ^2-ford. er χ^2-ford med sum frihetsgrader.

Tilsvarende $\frac{n\hat{\tau}^2}{\tau^2} \sim \chi^2_n$

Dermed:

$$E\left(\frac{m\hat{\sigma}^2}{\sigma^2}\right) = m \qquad \text{dvs } \underline{E(\hat{\sigma}^2) = \sigma^2}$$

$$Var\left(\frac{m\hat{\sigma}^2}{\sigma^2}\right) = 2m \quad \Rightarrow \quad \underline{Var(\hat{\sigma}^2) = \frac{2\sigma^4}{m}}$$

Tilsvarende:

$$\underline{E(\hat{\tau}^2) = \tau^2}$$

$$\underline{Var(\hat{\tau}^2) = \frac{2\tau^4}{m}}$$

c) Foreslår testobservatoren

$$F = \frac{\hat{\sigma}^2}{\hat{\tau}^2} = \frac{\frac{m\hat{\sigma}^2}{\sigma^2} / m}{\frac{n\hat{\tau}^2}{\tau^2} / n} \qquad \text{dvs Fisher}(m,n) \text{ hvis } \underline{H_0 \text{ gjelder}}$$

Dermed: Forkast H_0 hvis

$$\underline{F \leq f_{\alpha/2,m,n} \quad \text{eller} \quad F \geq f_{1-\alpha/2,m,n}}$$

Hvis $m = 5$, $n = 4$, $\alpha = 0.05$ er kritiske verdier

$$f_{0.025,5,4} = 0.135, \qquad f_{0.975,5,4} = 9.36$$

Med data: $\hat{\sigma}^2 = \frac{1}{5}[7^2 + 3^2 + 4^2 + 0^2 + 6^2] = 22$

$\hat{\tau}^2 = \frac{1}{4}[11^2 + 9^2 + 1^2 + 13^2] = 93$

Dermed $\frac{\hat{\sigma}^2}{\hat{\tau}^2} = \frac{22}{93} = \underline{0.237}$ dvs IKKE FORKAST H_0.

d) Bruker her at for alle σ^2, τ^2 er

$$\frac{\hat{\sigma}^2/\sigma^2}{\hat{\tau}^2/\tau^2} = \frac{\frac{m\hat{\sigma}^2}{\sigma^2}/m}{\frac{n\hat{\tau}^2}{\tau^2}/n} \sim \text{Fisher}(m,n)$$

Dermed

$$P\left(f_{\alpha/2,m,n} \le \frac{\hat{\sigma}^2/\sigma^2}{\hat{\tau}^2/\tau^2} \le f_{1-\alpha/2,m,n} \right) = 1-\alpha$$

Som gir

$$P\left(\frac{\hat{\tau}^2}{\hat{\sigma}^2} f_{\alpha/2,m,n} \le \frac{\tau^2}{\sigma^2} \le \frac{\hat{\tau}^2}{\hat{\sigma}^2} f_{1-\alpha/2,m,n} \right) = 1-\alpha$$

dvs Konf.int. er

$$\left(\frac{\hat{\tau}^2}{\hat{\sigma}^2} f_{\alpha/2,m,n} \,,\; \frac{\hat{\tau}^2}{\hat{\sigma}^2} f_{1-\alpha/2,m,n} \right)$$

Med observasjonene:

$$[4.23 \cdot 0.135 \,,\; 4.23 \cdot 9.36]$$

$$[0.57 \,,\; 39.39]$$

~~Forventet lengde?~~ $\cancel{E\left[\frac{\hat{\tau}^2}{\hat{\sigma}^2}\right](f_{1-\alpha/2,m,n} - f_{\alpha/2,m,n})}$

$\cancel{= E(F_{m,n}) \cdot (f_{1-\alpha/2,m,n} - f_{\alpha/2,m,n})}$

~~Oppgave 3~~ 2

e) MLE for den felles σ^2 under H_0:

$$\hat{\sigma}_0^2 = \frac{1}{m+n}\left[\sum_{i=1}^{m}(X_i-\mu_0)^2 + \sum_{j=1}^{n}(Y_j-\mu_0)^2\right]$$

Dette følger direkte fra a) siden vi nå har $m+n$ u.i.f. data fra $N(\mu_0, \sigma^2)$

Kall likelihooden i a) for $L(\sigma^2, \tau^2)$.
Da er den generaliserte likelihood ratio:

$$\lambda = \frac{L(\hat{\sigma}_0^2, \hat{\sigma}_0^2)}{L(\hat{\sigma}^2, \hat{\tau}^2)}$$

Her er

$$L(\hat{\sigma}^2, \hat{\tau}^2) = \frac{1}{(2\pi)^{\frac{m+n}{2}}} \cdot \frac{1}{\hat{\sigma}^m} \cdot \frac{1}{\hat{\tau}^n} e^{-\frac{m+n}{2}}$$

$$L(\hat{\sigma}_0^2, \hat{\tau}_0^2) = \frac{1}{(2\pi)^{\frac{m+n}{2}}} \cdot \frac{1}{\hat{\sigma}_0^{m+n}} e^{-\frac{m+n}{2}}$$

Dermed er

$$\lambda = \frac{\hat{\sigma}^m \hat{\tau}^n}{\hat{\sigma}_0^{m+n}} = \frac{\left(\frac{\hat{\sigma}^2}{\hat{\tau}^2}\right)^{\frac{m}{2}}}{\left(\frac{m}{m+n}\left(\frac{\hat{\sigma}^2}{\hat{\tau}^2}\right) + \frac{n}{m+n}\right)^{\frac{m+n}{2}}}$$

dvs. λ er en funksjon av F fra c).

Oppgave 3

a) H_0: sannsynl. for at en bil får en skade avhenger ikke av fargen.

Under H_0 vil derfor, for 200 biler, de forventede antall i hver fargekategori være

Sølv : $0.14 \cdot 200$ =	28
Hvit	28
Blå	60
Rød	58
Annen	26

Dermed blir

$$C = \sum \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(38-28)^2}{28} + \frac{(34-28)^2}{28} + \frac{(54-60)^2}{60} + \frac{(42-58)^2}{58}$$

$$+ \frac{(32-26)^2}{26} = 11.2556$$

Under H_0 er $C \sim \chi^2_{5-1}$ slik at med 5% nivå skal vi forkaste hvis $C \geq 9.488$

Altså: FORKAST H_0 (P-verdi ≈ 0.025)

b) Utsagnet er

$$\frac{P(A|S)}{P(A|R)} = 2$$

Estimering:

$$\frac{P(A|S)}{P(A|R)} = \frac{\frac{P(S|A)\,P(A)}{P(S)}}{\frac{P(R|A)\,P(A)}{P(R)}}$$

$$= \frac{P(S|A)}{P(R|A)} \cdot \frac{P(R)}{P(S)}$$

$$\underset{\text{estimering}}{\uparrow}\!\!= \frac{\frac{38}{200}}{\frac{42}{200}} \cdot \frac{0.29}{0.14} = 1.8741$$

(Stemmer altså ganske bra med påstanden. Men tallet 1.8741 er selvsagt beheftet med usikkerhet.