Lösningar till tentamen i Inferensteori II, 12 januari 2005.

Uppgift 1.

Låt $\hat{q}_n$ vara ML-skattningen av q.

a) Skattningen $\hat{q}_n$ är konsistent om för något $e > 0$, $\lim_{n\to\infty} P\left(|\hat{q}_n - q| > e\right) = 0$ dvs. om $\hat{q}_n$ konvergerar mot q i sannolikhet.

b) Skattningen $\hat{q}_n$ är asymptotisk effektiv om $\lim_{n\to\infty} e(\hat{q}_n) = \lim_{n\to\infty} \dfrac{1/I(q)}{n[Var(\hat{q}_n)]} = 1$, där $I(q)$ är information i en observation och $Var(\hat{q}_n)$ är den asymptotiska variansen (se också Lindgren på sidan 264).

c) Om villkoren A-C från Svenssons kompendium (sidan 6) är uppfyllda och det finns en konsistent lösning $\hat{q}_n$ till likelihoodekvationen så gäller det att

$$\sqrt{n}\left(\hat{q}_n - q\right) \xrightarrow{d} N\left(0, I_X^{-1}(q)\right)$$

när $n \to \infty$.

Uppgift 2.

Se Lindgren på sidor 231-232.

Uppgift 3.

a) $f(\mathbf{x}\,|\,q) = \prod_{i=1}^{n} q\, x_i^{q-1} = q^n \prod_{i=1}^{n} x_i^{q-1} = q^n \exp\left\{q \cdot \sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} \ln x_i\right\}$ visar att fördelningen tillhör till exponentialfamilj så tillräcklig statistika är $\sum_{i=1}^{n} \ln x_i$.

b) $I(q) = -E\left[\dfrac{\partial^2}{\partial^2 q} \ln f(\mathbf{x}\,|\,q)\right]$

$$\ln f(\mathbf{x}\,|\,q) = q \cdot \sum_{i=1}^{n} \ln x_i + n \ln q - \sum_{i=1}^{n} \ln x_i$$

$$\frac{\partial}{\partial q} \ln(\mathbf{x}\,|\,q) = \sum_{i=1}^{n} \ln x_i + \frac{n}{q}, \quad \frac{\partial^2}{\partial^2 q} \ln(\mathbf{x}\,|\,q) = -\frac{n}{q^2} \quad \Rightarrow \quad I(q) = -E\left[-\frac{n}{q^2}\right] = \frac{n}{q^2}.$$

c) $\dfrac{\partial}{\partial q} \ln(\mathbf{x}\,|\,q) = \sum_{i=1}^{n} \ln x_i + \dfrac{n}{q} = 0 \quad \Rightarrow \quad \hat{q} = -\dfrac{1}{\frac{1}{n}\sum_{i=1}^{n} \ln x_i}$

$$\frac{\partial^2}{\partial^2 q} \ln(\mathbf{x}\,|\,q) = -\frac{n}{q^2} < 0 \quad \Rightarrow \quad \textit{likelihooden har maximum i } \hat{q}.$$

d) pga Sats 4.2 från Svenssons kompendium $\sqrt{I(q)}\left(\hat{q}-q\right)\xrightarrow{d} N(0,1)$ dvs.

$\frac{\sqrt{n}}{q}\left(\hat{q}-q\right)\xrightarrow{d} N(0,1)$. Detta ger att $P\left(-z_{a/2}\le \frac{\sqrt{n}}{q}(\hat{q}-q)\le z_{a/2}\right)\approx 1-a$ då n är stort

$$\Leftrightarrow P\left(\frac{\hat{q}}{1+\frac{z_{a/2}}{\sqrt{n}}}\le q\le \frac{\hat{q}}{1-\frac{z_{a/2}}{\sqrt{n}}}\right)\approx 1-a$$

så $(1-a)-$ konfidensintervall för q är $\left(\frac{\hat{q}}{1+\frac{z_{a/2}}{\sqrt{n}}},\frac{\hat{q}}{1-\frac{z_{a/2}}{\sqrt{n}}}\right)$

Uppgift 4.

Vi räknar

$$f(\mathbf{x}\,|\,q)=2^n q^n\left(\prod_{i=1}^n x_i\right)e^{-q\sum_{i=1}^n x_i^2} \quad \text{och} \quad f(\mathbf{x},q)=f(\mathbf{x}\,|\,q)\cdot g(q)=\frac{2^n q^n}{m}\left(\prod_{i=1}^n x_i e^{-q\left(\sum x_i^2+\frac{1}{m}\right)}\right)$$

Nu beräknar vi den marginella täthetsfunktionen

$$f(\mathbf{x})=\int_0^\infty f(\mathbf{x},q)\,dq=\frac{2^n\left(\prod_{i=1}^n x_i\right)}{m}\int_0^\infty q^n e^{-q\left(\sum x_i^2+\frac{1}{m}\right)}dq=\frac{2^n\left(\prod_{i=1}^n x_i\right)}{m}\cdot\frac{\Gamma(n+1)}{\left(\sum x_i^2+\frac{1}{m}\right)^{n+1}}$$

Så

$$f(q\,|\,\mathbf{x})=\frac{f(\mathbf{x},q)}{f(\mathbf{x})}=\frac{\left(\sum x_i^2+\frac{1}{m}\right)^{n+1} q^n e^{-q\left(\sum x_i^2+\frac{1}{m}\right)}}{\Gamma(n+1)} \quad \text{för} \quad q>0.$$

Vi ses att detta är en gammafördelning $\Gamma\left(n+1,\sum x_i^2+\frac{1}{m}\right)$.

Bayesskattningen är väntevärdet i denna fördelning dvs $\frac{n+1}{\sum x_i^2+\frac{1}{m}}$.

Uppgift 5.

a) Här blir likelihoodkvoten

$$\Lambda=\frac{\left(\frac{1}{2}\right)^n}{\hat{p}^Y(1-\hat{p})^{n-Y}}=\left(2\left(\frac{Y}{n}\right)^{\frac{Y}{n}}\left(1-\frac{Y}{n}\right)^{1-\frac{Y}{n}}\right)^{-n}$$

eftersom ML-skattningen $\hat{p} = \frac{Y}{n}$ som erhålles på vanligt sätt. H_0 ska alltså förkastas då

$$\left(\frac{Y}{n}\right)^{\frac{Y}{n}}\left(1-\frac{Y}{n}\right)^{1-\frac{Y}{n}} > K.$$

Funktionen $g(x) = x^x(1-x)^{1-x}$ är avtagande för $0 < x \le \frac{1}{2}$ och växande för $\frac{1}{2} \le x < 1$. H_0 skall alltså förkastas om Y/n är nära 0 eller 1. Eftersom g är symmetrisk kring $\frac{1}{2}$ kan man också uttrycka det som att H_0 ska förkastas då $\left|\frac{Y}{n}-\frac{1}{2}\right| > K.$

b) Under H_0 är $\left(\frac{Y}{n}-\frac{1}{2}\right)/\sqrt{\frac{1}{4}/100}$ approximativt $N(0,1)$. H_0 ska alltså förkastas på den approximativa nivån $a = 0.05$ om $|Y-50| > 5\cdot 1.96 = 9.8.$

Uppgift 6.

a) Låt $l_1 > 1$ och sätt upp likelihoodkvoten för testet av den enkla hypotesen $H_0 : l = 1$ mot den enkla hypotesen $H_1 : l = l_1$

$$\Lambda^* = \frac{\exp\left\{-\sum_{i=1}^{n} x_i\right\}}{l_1^n \exp\left\{-l_1\sum_{i=1}^{n} x_i\right\}} = l_1^{-n}\exp\left\{(l_1-1)\sum_{i=1}^{n} x_i\right\}$$

Att $\Lambda^* < K'$ är detsamma som att $\sum_{i=1}^{n} x_i < K$ för något K så vi får testet:

Förkasta H_0 om $\sum_{i=1}^{n} x_i < K$. Värdet på K bestäms bara av nivån α och stickprovsstorleken n och inte på l_1 så snart $l_1 > 1$ så testet är likformigt starkast för mothypotesen $H_1 : l > 1$.

b) Med given nivån $a = 0.05$ och styrka för $l = 3, p(3) \ge 0.7,$ får vi sambanden

$$P\left(\sum_{i=1}^{n} X_i < K \mid l = 1\right) = 0.05$$

$$P\left(\sum_{i=1}^{n} X_i < K \mid l = 3\right) = 0.7$$

Det gäller att om λ är det sanna värdet på parametern så är $2l\sum X_i \sim c^2(2n)$.
Vi ska alltså hitta ett K och ett n så att om $Y \sim c^2(2n)$ så ska $P(Y < 2K) = 0.05$ och $P(Y < 6K) \ge 0.7.$ Genom att studera tabellen över $c^2(2n)$ – fördelningen ser vi att det minsta n som uppfyller kravet är (sånär som på avrundningsfel) $n = 5$ dvs. $2n = 10.$