STK 4020, høsten 2004
Bayesiansk statistikk

Noen ekstraoppgaver fra Nils

1. Vi skal se på det berømte hovedeksemplet i Thomas Bayes' 1763-artikkel, der han utleder en alternativ estimator for en binomisk sannsynlighet. Vi ser også litt på risikofunksjoner.

(a) Anta at Y for gitt p er binomisk (n, p), og at p er uniformt fordelt over $(0, 1)$. Vis at aposteriorifordelingen for p blir

$$\pi(p \mid y) = \frac{(n+1)!}{y!\,(n-y)!}\, p^y (1-p)^{n-y} \quad \text{for } 0 < p < 1.$$

(b) Det naturlige Bayes-estimatet for en parameter er forventningen i aposteriorfordelingen. Vis at denne blir

$$\widehat{p} = \mathrm{E}(p \mid y) = \frac{y+1}{n+2}.$$

Finn også en formel for $\widehat{\sigma}$, standardavviket i aposteriorifordelingen.

(c) For beregningen av $\widehat{p}$ og $\widehat{\sigma}$ er data-resultatet $Y = y$ gitt; det er forsåvidt ikke nødvendig å beregne estimat og usikkerhetstall for andre verdier av data enn for den $Y = y$ man faktisk observerte. Likevel er det interessant og relevant å se på egenskaper som har med «repeterte forsøk» å gjøre. – Beregn risikofunksjonen

$$\mathrm{mse}_1(p) = \mathrm{E}_p\{\widehat{p}(Y) - p\}^2$$

for Bayes-estimatoren $\widehat{p} = \widehat{p}(Y)$, samt risikofunksjonen

$$\mathrm{mse}_2(p) = \mathrm{E}_p\{p^*(Y) - p\}^2$$

for den klassiske estimatoren $p^* = Y/n$.

(d) Sammenlign de to risikofunksjonene, for $n = 5, 10, 100$. Tegn gjerne opp i diagrammer.

(e) Hvilken av de to estimatorene ville du foretrekke å anvende, i en gitt situasjon?

2. Beta-fordelingen med parametre (a, b) for en variabel Z over $(0, 1)$ er den som har tetthet

$$g(z) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, z^{a-1}(1-z)^{b-1} \quad \text{for } z \in (0, 1).$$

Her skal a og b være positive parametre. Merk at med $a = b = 1$ får vi den uniforme fordelingen.

(a) Vis at for en slik variabel gjelder

$$\mathrm{E}Z = \frac{a}{a+b} \quad \text{og} \quad \mathrm{Var}Z = \frac{1}{a+b+1}\mathrm{E}Z(1 - \mathrm{E}Z).$$

(b) La $Y \mid p$ være binomisk (n, p), og anta at p har en $\text{Beta}(a, b)$ som apriorifordeling. Finn aposteriorifordelingen.

(c) Finn Bayes-estimatet $\widehat{p}$ og dets usikkerhet $\widehat{\sigma}$.

3. La $Y \mid \theta$ være normal $(\theta, 1)$. Den kanoniske frekventist-estimator er $\theta^* = Y$ – den er maximum likelihood-estimator, moment-estimator, least squares-estimator, etc.

(a) Finn risikofunksjonen

$$\text{mse}_F(\theta) = \text{E}_\theta(\theta^* - \theta)^2.$$

(b) Mot denne estimator kan man stille opp en Bayes-estimator. Bruk normalfordelingen (θ_0, σ_0^2) som apriorifordeling. Finn aposteriorifordelingen.

(c) Finn risikofunksjonen

$$\text{mse}_B(\theta) = \text{E}_\theta(\widehat{\theta} - \theta)^2$$

for Bayes-estimatoren $\widehat{\theta} = \text{E}(\theta \mid y)$, og sammenlign med den klassiske.

(d) Hvordan vil du tolke θ_0 og σ_0?

4. Jeg nedtegner her noen av de viktigste egenskapene til den multinormale fordeling. Disse er av betydning i mange områder av statistikk og sannsynlighetsteori, også innenfor dette kursets temata. – Vi sier at den stokastiske vektor $X = (X_1, \ldots, X_p)^{\text{t}}$ er multinormalt fordelt med parametre (μ, Σ), der $\mu = (\mu_1, \ldots, \mu_p)^{\text{t}}$ er en p-vektor og $\Sigma = (\sigma_{i,j})_{i,j=1,\ldots,p}$ en symmetrisk, positivt definit $p \times p$-matrise, dersom vektorens sannsynlighetstetthet er

$$f(x) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\{-\tfrac{1}{2}(x-\mu)^{\text{t}}\Sigma^{-1}(x-\mu)\} \quad \text{for } x \in \mathcal{R}^p.$$

Vi skriver da $X \sim \text{N}_p(\mu, \Sigma)$.

(a) Vis at enher lineærkombinasjon av X blir normalfordelt:

$$a^{\text{t}}X = \sum_{j=1}^{p} a_j X_j \sim \text{N}(a^{\text{t}}\mu, a^{\text{t}}\Sigma a) = \text{N}\Big(\sum_{j=1}^{p} a_j\mu_j, \sum_{j,k} a_j a_k \sigma_{j,k}\Big).$$

Vis mer generelt at når A er en $k \times p$-matrise og b en k-vektor, så vil

$$Y = AX + b \sim \text{N}_k(A\mu + b, A\Sigma A^{\text{t}}).$$

Den første egenskapen her er også en karakterisering av den multinormale fordeling.

(b) Vis at $\text{E}X = \mu$ og at kovariansmatrisen til X er Σ.

(c) Anta at

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \text{N}_{p+q}\Big(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\Big).$$

Vis at

$$X \mid (Y = y) \sim \text{N}_p(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}).$$

(d) De fem brødre Hjort er henholdsvis 1.97, 1.90, 2.07, 1.94 og 2.00 meter høye. Hvor meget tror du at de forskjellige hjorter veier?

5. Denne oppgaven fyller ut noen detaljer om suffisens, se lærebokens diskusjon side 42. Rammeverket her er en statistisk modell for data $\mathcal{D}$ som avhenger av en parameter θ. Man sier at observatoren $T = T(\mathcal{D})$ er Fisher-suffisient dersom fordelingen for $\mathcal{D}$ gitt $T = t$ ikke avhenger av θ.

(a) La $Y_1, \ldots, Y_n$ være uavhengige og Poisson-fordelte med parameter θ. Vis at $T = \sum_{i=1}^n Y_i$ er suffisient for θ.

(b) Den såkalte faktoriseringssetningen, som ikke skal vises her, sier at T er Fisher-suffisient hvis og bare hvis den simultane sannsynlighetstetther for data, altså likelihood-funksjonen, kan skrives

$$f_{\text{sim}}(\mathcal{D}\,|\,\theta) = L(\theta) = g(T(\mathcal{D})\,|\,\theta)h(\mathcal{D})$$

for visse funksjoner g og h. – Vis ved hjelp av denne setningen at $\sum_{i=1}^n Y_i$ er suffisient i Poisson-situasjonen over.

(c) La $Y_1, \ldots, Y_n$ være uavhengige og normale (μ, σ^2). Vis at radarparet

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i \quad \text{og} \quad S = \Big[\frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2\Big]^{1/2}$$

danner et suffisient par for data.

(d) Prøv å vise at hvis $T = T(\mathcal{D})$ er suffisient, med en faktorisering av simultantettheten for data som over, så blir tettheten for T proporsjonal med $g(t\,|\,\theta)$.

(e) La $\pi(\theta)$ være en vilkårlig prior for θ. Vis at fordelingen for θ gitt data kan skrives

$$\pi(\theta\,|\,\mathcal{D}) = \text{const.}\,\pi(\theta)g(T(\mathcal{D})\,|\,\theta).$$

Den avhenger altså kun av data via $T = T(\mathcal{D})$ alene, og vi sier at $T = T(\mathcal{D})$ er Bayes-suffisient. Dette er altså en konsekvens av Fisher-suffisiens (via faktoriseringssetningen).

(f) Vis i situasjonen over at fordelingen for θ gitt hele datamaterialet $\mathcal{D}$ er det samme som fordelingen for θ gitt kun verdien[e] av $T = T(\mathcal{D})$.

6. La $Y_1, \ldots, Y_n$ være uavhengige og normale (θ, σ^2). I denne oppgaven anses presisjonen σ som et kjent tall, mens θ er ukjent og av primær interesse.

(a) Vis at $\bar{Y}$ er suffisient i denne modellen, og at $\bar{Y} \sim \mathrm{N}(\theta, \sigma^2/n)$.

(b) Anta at θ har prior-fordelingen $\mathrm{N}(\theta_0, \tau_0^2)$. Vis at

$$\begin{pmatrix} \theta \\ \bar{Y} \end{pmatrix} \sim \mathrm{N}_2\Big(\begin{pmatrix} \theta_0 \\ \theta_0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_0^2 \\ \tau_0^2 & \tau_0^2 + \sigma^2/n \end{pmatrix}\Big).$$

(c) Vis ved hjelp av oppgave 4 at

$$\theta\,|\,\text{data} \sim \mathrm{N}(\widehat{\theta}, \tau_n^2),$$

der

$$\widehat{\theta} = \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n}\bar{y} + \frac{\sigma^2/n}{\tau_0^2 + \sigma^2/n}\theta_0$$

og

$$\tau_n^2 = \frac{\sigma^2}{n}\frac{\tau_0^2}{\tau_0^2 + \sigma^2/n}.$$

Den første her er en såkalt kredibilitetsformel, og viser en naturlig metode for å kombinere a priori-gjett med empiriske data.

(d) Vis at hvis n er stor, så vil $\widehat{\theta} \approx \bar{y}$ og $\tau_n \approx \sigma/\sqrt{n}$. Vis at det samme holder når n er fast og τ_0 blir stor (en «flat prior»).

(e) Bemerk at Bayesianerens usikkerhet τ_n alltid er mindre enn frekventistens usikkerhet $\sigma/\sqrt{n}$. Dette kan kanskje fremstilles som at «det er alltid en belønning ved å være Bayesianer». Diskuter dette synspunktet.

7. I rammen av forrige oppgave, studer to Bayesianere med prior-parametre (θ_0, τ_0) og (θ_0', τ_0'). Deres resulterende Bayes-estimatorer er $\widehat{\theta}$ og $\widehat{\theta}'$. Hvor uenige kan de klare å bli? Vis ihvertfall at når datamengden øker, så blir de mer og mer enige:

$$\sqrt{n}(\widehat{\theta} - \widehat{\theta}') \to_p 0.$$

Vis også at de blir «mer og mer enige» med den klassiske Fisherianer, som i denne modellen går for $\theta_{\mathrm{ML}} = \bar{Y}$. Kommenter disse funn.

8. Anta den én-dimensjonale parameteren θ skal estimeres med kvadratisk tap, etter å ha observert data X fra en fordeling $f(y \mid \theta)$. Anta at θ har en prior-fordeling $\pi(\theta)$.

(a) Vis at Bayes-estimatoren blir $\widehat{\theta} = \mathrm{E}(\theta \mid Y)$.

(b) Og vis at Bayes-risiko for denne estimatoren, med andre ord minimum Bayes-risiko for vedkommende prior π, blir

$$\mathrm{MBR}(\pi) = \mathrm{E}[\mathrm{Var}(\theta \mid Y)].$$

(c) Anta nå at tapsfunksjonen er vektet kvadratisk, av typen $L(\theta, t) = w(\theta)(t - \theta)^2$. Vis at Bayes-estimatoren tar formen

$$\widehat{\theta} = \frac{\mathrm{E}\{w(\theta)\theta \mid y\}}{\mathrm{E}\{w(\theta) \mid y\}},$$

og at minimum Bayes-risiko kan uttrykkes som

$$\mathrm{MBR}(\pi) = \mathrm{E}w(\theta)\theta^2 - \mathrm{E}w(\theta)\widehat{\theta}^2,$$

under alminnelige antagelser.

9. La $Y_1, \ldots, Y_n$ være uavhengige og normale (θ, σ^2), der σ er kjent. Forventningsparameteren θ skal estimeres med kvadratisk tap.

(a) Vis at risikofunksjonen til estimatoren $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ er konstant og lik σ^2/n.

(b) La θ ha prior-fordelingen $\mathrm{N}(0, \tau_0^2)$. Vis at Bayes-estimatoren blir

$$\widehat{\theta} = \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n} \bar{Y},$$

og beregn dennes risikofunksjon.

(c) Bruk dette, eller andre argumenter, til å finne minimum Bayes risiko for denne prior-fordelingen.

(d) Vis at $\bar{Y}$ er en minimax-estimator for θ.

(e) [Noe vanskeligere.] Vis at $\bar{Y}$ er *den eneste* minimax-estimatoren for θ.

10. La Y være binomisk (n, p), der p er ukjent og skal estimeres med kvadratisk tap.

(a) Finn risikofunksjonen for den vanligste estimatoren $p^* = Y/n$, og vis at dens max-risiko er $1/(4n)$.

(b) La p ha $\mathrm{Beta}(kp_0, kq_0)$-fordelingen som prior, der $q_0 = 1 - p_0$. Merk at $\mathrm{E}p = p_0$ og $\mathrm{Var}p = p_0 q_0/(k+1)$. – Vis at Bayes-estimatoren blir

$$\widehat{p} = \frac{kp_0 + y}{k + n}.$$

(c) Finn risikofunksjonen for $\widehat{p}$, og finn dennes max-risiko. Vis spesielt at risikofunksjonen er konstant når $p_0 = \frac{1}{2}$ og $k = \sqrt{n}$.

(d) Finn et uttrykk for minimum Bayes-risiko for prioren over, ved å beregne forventningen til risikofunksjonen.

(e) Vis at estimatoren

$$\widetilde{p} = \frac{\frac{1}{2}\sqrt{n} + Y}{\sqrt{n} + n} = \frac{1}{\sqrt{n}+1} \frac{1}{2} + \frac{\sqrt{n}}{\sqrt{n}+1} \frac{Y}{n}$$

er en minimax-estimator for p.

(f) Finn det intervall av p-verdier der minimax-estimatoren $\widetilde{p}$ har bedre risiko enn standardestimatoren Y/n. Uttal deg om situasjonen der n vokser.

(g) Vis at $\widetilde{p}$ også er den eneste minimax-estimatoren for p.

11. La $Y_1, \ldots, Y_n$ være uavhengige og uniformt fordelte over $(0, \theta)$.

(a) Vis at $W = \max\{Y_1, \ldots, Y_n\}$ blir en F-suffisient og dermed B-suffisient observator. Forklar hvorfor enhver aposteriori-fordeling for θ bare vil avhenge av data via W.

(b) Finn bruttovariansen $\mathrm{E}(cW - \theta)^2$ for estimatoren cW, og finn den beste estimatoren av denne type. [Svar: $c_0 = \frac{n+2}{n+1}$.] Sammenlign med Markov-estimatoren $\frac{n+1}{n} W$ (som er den med minst varians blant alle forventningsrette estimatorer).

(c) Den naturlige tapsfunksjon her er

$$L(\theta, t) = (t - \theta)^2/\theta^2 = (t/\theta - 1)^2.$$

Vis at estimatoren

$$\widehat{\theta} = \frac{n+2}{n+1} W$$

har konstant risiko $1/(n+1)^2$.

(d) La $\theta \sim \mathrm{Par}(\theta_0, a)$ a priori, der a og θ_0 er positive parametre. Dette er Pareto-fordelingen med tetthet $(a/\theta_0)(\theta_0/\theta)^{a+1}$ på $[\theta_0, \infty)$. Vis at

$$\mathrm{E}\theta^{-1} = \frac{a}{a+1}\theta_0^{-1} \quad \text{og} \quad \mathrm{E}\theta^{-2} = \frac{a}{a+2}\theta_0^{-2},$$

med påfølgende formel

$$\frac{\mathrm{E}\theta^{-1}}{\mathrm{E}\theta^{-2}} = \frac{a+2}{a+1}\theta_0$$

for minimatoren av $\mathrm{E}(t/\theta - 1)^2$ over alle t.

(e) Vis at θ gitt data får en $\mathrm{Par}(\max(\theta_0, W), n+a)$-fordeling. Bruk dette til å vise at Bayes-estimatoren blir

$$\widetilde{\theta} = \frac{\mathrm{E}(\theta^{-1} \,|\, \mathrm{data})}{\mathrm{E}(\theta^{-2} \,|\, \mathrm{data})} = \frac{n+a+2}{n+a+1}\max(\theta_0, W).$$

(f) Hvilken grense-prior vil ha estimatoren $\widehat{\theta}$ som sin Bayes-løsning? Kommenter svaret.

(g) Finn risikofunksjonen til $\widetilde{\theta}$, og sammenlign denne med risikofunksjonen til $\widehat{\theta}$.

(h) Vis at minimum Bayes-risiko for $\mathrm{Pareto}(\theta_0, a)$-prioren blir $\mathrm{MBR}(\theta_0, a) = 1/(n+a+1)^2$.

(i) Vis at $\widehat{\theta}$ er en minimax-estimator for θ.

12. La Y være Poisson-fordelt med parameter θ. Denne skal estimeres med den naturlig vektede tapsfunksjonen $L(\theta, t) = (t-\theta)^2/\theta$.

(a) Vis at estimatoren Y har konstant risiko lik 1. – Resten av oppgaven har blant annet som formål å vise at Y er minimax.

(b) La θ ha $\mathrm{Gamma}(a, b)$ som prior. Vis at aposteriori-fordelingen blir $\mathrm{Gamma}(a+y, b+1)$.

(c) Vis at Bayes-estimatoren blir

$$\widehat{\theta} = \frac{a+y-1}{b+1}$$

dersom $a + y \geq 1$.

(d) Finn dennes risikofunksjon, og sammenlign med risikofunksjonen til Y.

(e) Vis at minimum Bayes-risiko for Gamma-prioren blir $1/(b+1)$.

(f) Vis at Y er minimax.

(g) [Noe vanskeligere.] Vis at Y også er den eneste minimax-estimatoren for θ.

(h) Generaliser til situasjonen der $Y_1, \ldots, Y_n$ er uavhengige Poisson-variable med parameter θ. Vis at $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$ er minimax med konstant risiko $1/n$.