# HW for PSTAT 207

Haosheng Zhou

Sept, 2022

SECTION 7.2

4.

$$
\begin{aligned}
F_{Y_n}(y) &= \mathbb{P}\left(n[1-F(X_{(n)})] \leq y\right) && (1)\\
&= \mathbb{P}\left(F(X_{(n)}) \geq 1 - \frac{y}{n}\right) && (2)\\
&= \mathbb{P}\left(X_{(n)} \geq F^{-1}\left(1-\frac{y}{n}\right)\right) && (3)\\
&= 1-\mathbb{P}\left(X_{(n)} < F^{-1}\left(1-\frac{y}{n}\right)\right) && (4)\\
&= 1-\left[\mathbb{P}\left(X_1 \leq F^{-1}\left(1-\frac{y}{n}\right)\right)\right]^n && (5)\\
&= 1-\left[F\left(F^{-1}\left(1-\frac{y}{n}\right)\right)\right]^n && (6)\\
&= 1-\left(1-\frac{y}{n}\right)^n \to 1-e^{-y}\ (n\to\infty) && (7)
\end{aligned}
$$

6.

Denote the maximum as $M_n$, so

$$
\begin{aligned}
\forall \varepsilon > 0, \mathbb{P}\left(|M_n - \theta| \geq \varepsilon\right) &= \mathbb{P}\left(\theta - M_n \geq \varepsilon\right) && (8)\\
&= \mathbb{P}\left(M_n \leq \theta - \varepsilon\right) && (9)\\
&= \left[\mathbb{P}\left(X_1 \leq \theta-\varepsilon\right)\right]^n && (10)\\
&= \frac{(\theta-\varepsilon)^n}{\theta^n} \to 0\ (n\to\infty) && (11)
\end{aligned}
$$

for small enough $\varepsilon \in (0,\theta)$, so $M_n \xrightarrow{p} \theta\ (n\to\infty)$.

SECTION 7.3:

1.

$$\log Z_n = \frac{\sum_{i=1}^n \log X_i}{n} \xrightarrow{p} \mathbb{E}\log X_1\ (n\to\infty) \tag{12}$$

by weak law of large numbers. As a result,

$$\mathbb{E}\log X_1 = \int_0^1 \log x\, dx = (x\log x - x)|_{(0,1)} = -1 \tag{13}$$

so

$$Z_n \xrightarrow{p} \frac{1}{e}\ (n\to\infty) \tag{14}$$

SECTION 7.5:

4.

Now $X_n \sim Gamma(n,\beta)$, so the MGF $M_{X_n}(t) = \left(1-\frac{t}{\beta}\right)^{-n}$ and $M_{\frac{X_n}{n}}(t) = M_{X_n}(\frac{t}{n}) = \left(1-\frac{t}{n\beta}\right)^{-n} \to e^{\frac{t}{\beta}}$ $(n\to\infty)$ which is the MGF of the limiting distribution.

SECTION 7.6:

4.

$X \sim B(n,\theta)$ is the sum of $i.i.d.$ Bernoulli random variables, so CLT holds when $n$ is large enough. We see that $\mathbb{E}X = n\theta, Var(X) = n\theta(1-\theta)$ and

$$\frac{X-n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow{d} N(0,1)\ (n\to\infty) \tag{15}$$

now the probability is just

$$\mathbb{P}\left(X > \frac{n}{2}\right) = \mathbb{P}\left(\frac{X-n\theta}{\sqrt{n\theta(1-\theta)}} > \frac{\frac{n}{2}-n\theta}{\sqrt{n\theta(1-\theta)}}\right) = 1-\Phi\left(\frac{\frac{n}{2}-n\theta}{\sqrt{n\theta(1-\theta)}}\right) \geq 0.9 \tag{16}$$

so $\frac{\frac{n}{2}-0.45n}{\sqrt{n\theta(1-\theta)}} = -1.28$ (such $n$ does not exist)

6. Now $\mathbb{E}\overline{X}_n = 75$, so

$$\mathbb{P}\left(|\overline{X}_n - 75| \geq 6\right) \leq \frac{Var(\overline{X}_n)}{36} = \frac{\frac{225}{100}}{36} = \frac{225}{3600} = 0.0625 \tag{17}$$

by CLT

$$\frac{\overline{X}_n - 75}{\sqrt{2.25}} = \frac{\overline{X}_n - 75}{1.5} \xrightarrow{d} N(0,1)\ (n\to\infty) \tag{18}$$

so

$$\begin{aligned} \mathbb{P}\left(|\overline{X}_n - 75| \geq 6\right) &= \mathbb{P}\left(\overline{X}_n \leq 69\right) + \mathbb{P}\left(\overline{X}_n \geq 81\right) && (19)\\ &= \mathbb{P}\left(\frac{\overline{X}_n - 75}{1.5} \leq -4\right) + \mathbb{P}\left(\frac{\overline{X}_n - 75}{1.5} \geq 4\right) && (20)\\ &= \Phi(-4) + 1 - \Phi(4) = 2\Phi(-4) = 0.00006 && (21) \end{aligned}$$

8.3.1

(a):

The joint likelihood is

$$p_{\alpha,\beta}(x) = \prod_{i=1}^{n} \frac{1}{B(\alpha,\beta)} \mathbb{I}_{x_i \in (0,1)} x_i^{\alpha-1}(1-x_i)^{\beta-1} \tag{22}$$

$$= \frac{1}{[B(\alpha,\beta)]^n} (\prod_i x_i)^{\alpha-1} [\prod_i (1-x_i)]^{\beta-1} \mathbb{I}_{x_1,\dots,x_n \in (0,1)} \tag{23}$$

When $\alpha$ is known, $\beta$ unknown,

$$p_\beta(x) = \frac{1}{[B(\alpha,\beta)]^n} [\prod_i (1-x_i)]^{\beta-1} (\prod_i x_i)^{\alpha-1} \mathbb{I}_{x_1,\dots,x_n \in (0,1)} \tag{24}$$

and by factorization theorem, $T(X) = \prod_{i=1}^n (1-X_i)$ is sufficient.

When $\beta$ is known, $\alpha$ is unknown,

$$p_\alpha(x) = \frac{1}{[B(\alpha,\beta)]^n} (\prod_i x_i)^{\alpha-1} [\prod_i (1-x_i)]^{\beta-1} \mathbb{I}_{x_1,\dots,x_n \in (0,1)} \tag{25}$$

and by factorization theorem, $T(X) = \prod_{i=1}^n X_i$ is sufficient.

When both parameters are unknown,

$$p_{\alpha,\beta}(x) = \frac{1}{[B(\alpha,\beta)]^n} (\prod_i x_i)^{\alpha-1} [\prod_i (1-x_i)]^{\beta-1} \mathbb{I}_{x_1,\dots,x_n \in (0,1)} \tag{26}$$

and by factorization theorem, $T(X) = (\prod_{i=1}^n X_i, \prod_{i=1}^n (1-X_i))$ is sufficient.

(b):

The joint likelihood is

$$p_{\alpha,\beta}(x) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} \mathbb{I}_{x_i>0} \tag{27}$$

$$= \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} (\prod_i x_i)^{\alpha-1} e^{-\beta \sum_i x_i} \mathbb{I}_{x_1,\dots,x_n>0} \tag{28}$$

When $\alpha$ unknown, $\beta$ known

$$p_\alpha(x) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} (\prod_i x_i)^{\alpha-1} e^{-\beta \sum_i x_i} \mathbb{I}_{x_1,\dots,x_n>0} \tag{29}$$

and by factorization theorem, $T(X) = \prod_{i=1}^n X_i$ is sufficient.

When $\beta$ unknown, $\alpha$ known

$$p_\beta(x) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} e^{-\beta \sum_i x_i} (\prod_i x_i)^{\alpha-1} \mathbb{I}_{x_1,\ldots,x_n>0} \tag{30}$$

and by factorization theorem, $T(X) = \sum_{i=1}^n X_i$ is sufficient.

When both are unknown,

$$p_{\alpha,\beta}(x) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} e^{-\beta \sum_i x_i} (\prod_i x_i)^{\alpha-1} \mathbb{I}_{x_1,\ldots,x_n>0} \tag{31}$$

and by factorization theorem, $T(X) = (\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$ is sufficient.

(c):

The joint likelihood is

$$p_{N_1,N_2}(x) = \prod_{i=1}^n \frac{1}{N_2 - N_1} \mathbb{I}_{x_i \in \{N_1+1,\ldots,N_2\}} \tag{32}$$

$$= \frac{1}{(N_2 - N_1)^n} \mathbb{I}_{x_1,\ldots,x_n \in \{N_1+1,\ldots,N_2\}} \tag{33}$$

When $N_1$ is known and $N_2$ unknown,

$$p_{N_2}(x) = \frac{1}{(N_2 - N_1)^n} \mathbb{I}_{x_{(n)} \le N_2} \mathbb{I}_{x_{(1)} \ge N_1+1} \mathbb{I}_{x_1,\ldots,x_n \in \mathbb{Z}} \tag{34}$$

so by factorization theorem, $T(X) = X_{(n)}$ is sufficient.

When $N_2$ known and $N_1$ unknown,

$$p_{N_1}(x) = \frac{1}{(N_2 - N_1)^n} \mathbb{I}_{x_{(1)} \ge N_1+1} \mathbb{I}_{x_{(n)} \le N_2} \mathbb{I}_{x_1,\ldots,x_n \in \mathbb{Z}} \tag{35}$$

so by factorization theorem, $T(X) = X_{(1)}$ is sufficient.

When both unknown,

$$p_{N_1,N_2}(x) = \frac{1}{(N_2 - N_1)^n} \mathbb{I}_{x_{(1)} \ge N_1+1} \mathbb{I}_{x_{(n)} \le N_2} \mathbb{I}_{x_1,\ldots,x_n \in \mathbb{Z}} \tag{36}$$

so by factorization theorem, $T(X) = (X_{(1)}, X_{(n)})$ is sufficient.

(d):

The joint likelihood is

$$p_\theta(x) = \prod_{i=1}^n e^{\theta - x_i} \mathbb{I}_{x_i > \theta} \tag{37}$$

$$= e^{n\theta} \mathbb{I}_{x_{(1)} > \theta} e^{-\sum_i x_i} \tag{38}$$

so by factorization theorem, $T(X) = X_{(1)}$ is sufficient.

(e):

The joint likelihood is

$$p_{\mu,\sigma}(x) = \prod_{i=1}^{n} \frac{1}{x_i\sigma\sqrt{2\pi}} e^{-\frac{(\log x_i-\mu)^2}{2\sigma^2}} \mathbb{I}_{x_i>0} \tag{39}$$

$$= \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_i(\log x_i-\mu)^2} (2\pi)^{-\frac{n}{2}} (\prod_{i=1}^{n} x_i)^{-1} \mathbb{I}_{x_1,\ldots,x_n>0} \tag{40}$$

$$= \sigma^{-n} e^{-\frac{n\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum_i \log^2 x_i} e^{\frac{\mu}{\sigma^2}\sum_i \log x_i} (2\pi)^{-\frac{n}{2}} (\prod_{i=1}^{n} x_i)^{-1} \mathbb{I}_{x_1,\ldots,x_n>0} \tag{41}$$

so by factorization theorem, $T(X) = (\sum_{i=1}^n \log^2 X_i, \sum_{i=1}^n \log X_i)$ is sufficient.

(f):

The joint likelihood is

$$p_\theta(x) = \prod_{i=1}^{n} c(\theta) 2^{-\frac{x_i}{\theta}} \mathbb{I}_{x_i\in\{\theta,\theta+1,\ldots\}} \tag{42}$$

$$= c^n(\theta) 2^{-\frac{1}{\theta}\sum_i x_i} \mathbb{I}_{x_1,\ldots,x_n\in\theta+\mathbb{N}} \tag{43}$$

$$= c^n(\theta) 2^{-\frac{1}{\theta}\sum_i x_i} \mathbb{I}_{x_{(1)}\geq\theta} \mathbb{I}_{\forall i,j, x_i-x_j\in\mathbb{Z}} \tag{44}$$

so by factorization theorem, $T(X) = (\sum_{i=1}^n X_i, X_{(1)})$ is sufficient.

(g):

The joint likelihood is

$$p_{\theta,p}(x) = \prod_{i=1}^{n} (1-p) p^{x_i-\theta} \mathbb{I}_{x_i\in\{\theta,\theta+1,\ldots\}} \tag{45}$$

$$= (1-p)^n p^{\sum_i x_i - n\theta} \mathbb{I}_{x_{(1)}\geq\theta} \mathbb{I}_{\forall i,j, x_i-x_j\in\mathbb{Z}} \tag{46}$$

When $p$ known $\theta$ unknown,

$$p_\theta(x) = p^{-n\theta} \mathbb{I}_{x_{(1)}\geq\theta} (1-p)^n p^{\sum_i x_i} \mathbb{I}_{\forall i,j, x_i-x_j\in\mathbb{Z}} \tag{47}$$

so by factorization theorem, $T(X) = X_{(1)}$ is sufficient.

When $\theta$ known $p$ unknown,

$$p_p(x) = (1-p)^n p^{\sum_i x_i - n\theta} \mathbb{I}_{x_{(1)}\geq\theta} \mathbb{I}_{\forall i,j, x_i-x_j\in\mathbb{Z}} \tag{48}$$

so by factorization theorem, $T(X) = \sum_{i=1}^n X_i$ is sufficient.

When both unknown,

$$p_{\theta,p}(x) = (1-p)^n p^{\sum_i x_i - n\theta} \mathbb{I}_{x_{(1)}\geq\theta} \mathbb{I}_{\forall i,j, x_i - x_j \in \mathbb{Z}} \tag{49}$$

so by factorization theorem, $T(X) = (\sum_{i=1}^n X_i, X_{(1)})$ is sufficient.

8.3.2

The joint density is

$$\begin{aligned} p_\sigma(x) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\alpha\sigma)^2}{2\sigma^2}} && (50)\\ &= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\alpha\sigma)^2} && (51)\\ &= \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_i x_i^2} e^{\frac{\alpha}{\sigma}\sum_i x_i} e^{-\frac{\alpha^2}{2}} (2\pi)^{-\frac{n}{2}} && (52)\\ &= g_\sigma(T(X)) \cdot h(x) && (53)\end{aligned}$$

where $g_\sigma(T(X)) = \sigma^{-n} e^{-\frac{1}{2\sigma^2}T_2(x)} e^{\frac{\alpha}{\sigma}T_1(x)}, h(x) = e^{-\frac{\alpha^2}{2}}(2\pi)^{-\frac{n}{2}}$, so by factorization theorem, $T(X)$ is sufficient for $\sigma$.

Now consider $g(x,y) = (1+\alpha^2)x^2 - (1+n\alpha^2)y$, so $\forall\sigma, \mathbb{E}_\sigma g(T(X)) = 0$ but $g(T(X))$ is not almost surely 0 since

$$(1+\alpha^2)\mathbb{E}(\sum_i X_i)^2 = (1+\alpha^2)[n\sigma^2(1+\alpha^2) + n(n-1)\alpha^2\sigma^2] = (1+\alpha^2)n\sigma^2(1+n\alpha^2) \tag{54}$$

and

$$(1+n\alpha^2)\mathbb{E}\sum_i X_i^2 = (1+n\alpha^2)n\sigma^2(1+\alpha^2) \tag{55}$$

but

$$g(T(X)) = (1+\alpha^2)(\sum_i X_i)^2 - (1+n\alpha^2)(\sum_i X_i^2) \tag{56}$$

that's why the family of distribution of $T(X)$ is not complete.

8.3.3

No, if it's complete then since it's sufficient, it's minimal sufficient, it's the function of all sufficient statistic. Since $T(X) = (\sum_i X_i, \sum_i X_i^2)$ is a sufficient statistic for the normal distribution, and notice that $X$ is not a function of $T(X)$ (there exists $x = e_1, y = e_2 \in \mathbb{R}^n$ such that $T(x) = T(y), x \neq y$), so it's not complete.

8.3.11

(a):

Consider likelihood ratio

$$\frac{p_\lambda(x)}{p_\lambda(y)} = \frac{\prod_{i=1}^n \lambda^{x_i} e^{-\lambda}}{\prod_{i=1}^n \lambda^{y_i} e^{-\lambda}} \tag{57}$$

$$= \frac{\lambda^{\sum_i x_i} e^{-n\lambda}}{\lambda^{\sum_i y_i} e^{-n\lambda}} \tag{58}$$

$$= \lambda^{\sum_i x_i - \sum_i y_i} \tag{59}$$

is independent of $\lambda$ iff $\sum_i x_i = \sum_i y_i$, so $T(X) = \sum_{i=1}^n X_i$ is minimal sufficient.

(b):

Consider likelihood ratio

$$\frac{p_\theta(x)}{p_\theta(y)} = \frac{\prod_{i=1}^n \frac{1}{\theta} \mathbb{I}_{x_i \in (0,\theta)}}{\prod_{i=1}^n \frac{1}{\theta} \mathbb{I}_{y_i \in (0,\theta)}} \tag{60}$$

$$= \frac{\mathbb{I}_{x_{(1)}>0} \mathbb{I}_{x_{(n)}<\theta}}{\mathbb{I}_{y_{(1)}>0} \mathbb{I}_{y_{(n)}<\theta}} \tag{61}$$

is independent of $\theta$ iff $x_{(n)} = y_{(n)}$, so $T(X) = X_{(n)}$ is minimal sufficient.

(c):

Consider likelihood ratio

$$\frac{p_p(x)}{p_p(y)} = \frac{\prod_{i=1}^n p(1-p)^{x_i}}{\prod_{i=1}^n p(1-p)^{y_i}} \tag{62}$$

$$= (1-p)^{\sum_i x_i - \sum_i y_i} \tag{63}$$

is independent of $p$ iff $\sum_i x_i = \sum_i y_i$, so $T(X) = \sum_{i=1}^n X_i$ is minimal sufficient.

(d):

Consider likelihood ratio

$$\frac{p_N(x)}{p_N(y)} = \frac{\prod_{i=1}^n \frac{1}{N} \mathbb{I}_{x_i \in \{1,\dots,N\}}}{\prod_{i=1}^n \frac{1}{N} \mathbb{I}_{y_i \in \{1,\dots,N\}}} \tag{64}$$

$$= \frac{\mathbb{I}_{x_1,\dots,x_n \in \{1,\dots,N\}}}{\mathbb{I}_{y_1,\dots,y_n \in \{1,\dots,N\}}} \tag{65}$$

$$= \frac{\mathbb{I}_{x_{(1)}\geq 1} \mathbb{I}_{x_{(n)}\leq N} \mathbb{I}_{x_1,\dots,x_n \in \mathbb{Z}}}{\mathbb{I}_{y_{(1)}\geq 1} \mathbb{I}_{y_{(n)}\leq N} \mathbb{I}_{y_1,\dots,y_n \in \mathbb{Z}}} \tag{66}$$

is independent of $N$ iff $x_{(n)} = y_{(n)}$, so $T(X) = X_{(n)}$ is minimal sufficient.

(e):

Consider joint likelihood

$$\frac{p_{\mu,\sigma}(x)}{p_{\mu,\sigma}(y)} = e^{-\frac{1}{2\sigma^2}[\sum_i (y_i^2 - x_i^2) + 2\mu \sum_i (x_i - y_i)]} \tag{67}$$

to find it's independent of $\mu, \sigma$ iff $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$, so $T(X) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is minimal sufficient statistic.

(f):

Consider joint likelihood

$$\frac{p_{\alpha,\beta}(x)}{p_{\alpha,\beta}(y)} = \frac{\frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} e^{-\beta\sum_i x_i} (\prod_i x_i)^{\alpha-1} \mathbb{I}_{x_1,...,x_n>0}}{\frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} e^{-\beta\sum_i y_i} (\prod_i y_i)^{\alpha-1} \mathbb{I}_{y_1,...,y_n>0}} \tag{68}$$

$$= e^{-\beta(\sum_i x_i - \sum_i y_i)} (\prod_i \frac{x_i}{y_i})^{\alpha-1} \frac{\mathbb{I}_{x_1,...,x_n>0}}{\mathbb{I}_{y_1,...,y_n>0}} \tag{69}$$

to find it's independent of $\alpha, \beta$ iff $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \prod_{i=1}^n \frac{x_i}{y_i} = 1$, so $T(X) = (\sum_{i=1}^n X_i, \prod_{i=1}^n X_i)$ is minimal sufficient statistic.

(g):

Consider joint likelihood

$$\frac{p_{\alpha,\beta}(x)}{p_{\alpha,\beta}(y)} = \frac{\frac{1}{[B(\alpha,\beta)]^n} (\prod_i x_i)^{\alpha-1} [\prod_i (1-x_i)]^{\beta-1} \mathbb{I}_{x_1,...,x_n \in (0,1)}}{\frac{1}{[B(\alpha,\beta)]^n} (\prod_i y_i)^{\alpha-1} [\prod_i (1-y_i)]^{\beta-1} \mathbb{I}_{y_1,...,y_n \in (0,1)}} \tag{70}$$

$$= (\prod_i \frac{x_i}{y_i})^{\alpha-1} (\prod_i \frac{1-x_i}{1-y_i})^{\beta-1} \frac{\mathbb{I}_{x_1,...,x_n \in (0,1)}}{\mathbb{I}_{y_1,...,y_n \in (0,1)}} \tag{71}$$

to find it's independent of $\alpha, \beta$ iff $\prod_{i=1}^n \frac{x_i}{y_i} = 1, \prod_{i=1}^n \frac{1-x_i}{1-y_i} = 1$, so $T(X) = (\prod_{i=1}^n X_i, \prod_{i=1}^n (1-X_i))$ is minimal sufficient statistic.

(h):

Consider joint likelihood

$$\frac{p_\theta(x)}{p_\theta(y)} = \frac{\prod_i \frac{2}{\theta^2}(\theta - x_i) \mathbb{I}_{x_i \in (0,\theta)}}{\prod_i \frac{2}{\theta^2}(\theta - y_i) \mathbb{I}_{y_i \in (0,\theta)}} \tag{72}$$

$$= \frac{\prod_i (\theta - x_i)}{\prod_i (\theta - y_i)} \frac{\mathbb{I}_{x_{(1)}>0, x_{(n)}<\theta}}{\mathbb{I}_{y_{(1)}>0, y_{(n)}<\theta}} \tag{73}$$

$$= \frac{\prod_i (\theta - x_{(i)})}{\prod_i (\theta - y_{(i)})} \frac{\mathbb{I}_{x_{(1)}>0, x_{(n)}<\theta}}{\mathbb{I}_{y_{(1)}>0, y_{(n)}<\theta}} \tag{74}$$

to find it's independent of $\theta$ iff $\forall i, x_{(i)} = y_{(i)}$, so $T(X) = (X_{(1)}, ..., X_{(n)})$ is minimal sufficient statistic (we can also argue by noticing that this is a location family with density $f(x-\theta)$).

8.3.18.

Since $X_1, ..., X_n \sim N(\theta, 1)$, we know that $X$ is multi-variate Gaussian and $(\sum_i a_i X_i, \sum_i X_i)$ is a linear transformation of $X$ so it's also a Gaussian random vector. As a result, being independent is equivalent to being uncorrelated.

$$cov(\sum_i a_i X_i, \sum_j X_j) = \sum_{i,j} a_i cov(X_i, X_j) = \sum_i a_i \tag{75}$$

that's why they are independent iff $\sum_i a_i = 0$.

8.3.20.

WLOG, just show it for $n = 2$.

Consider $V = \frac{X_1}{X_1+X_2}$ and use the Jacobian of the mapping $f : \mathbb{R}^2 \to \mathbb{R}^2, f(x_1, x_2) = \left(x_1, \frac{x_1}{x_1+x_2}\right)$ to see that the joint density

$$
\begin{aligned}
f_{X_1,V}(x_1, v) &= f_{X_1,X_2}(x_1, x_2) \cdot \left|\det \frac{\partial(x_1, x_2)}{\partial(x, v)}\right| && (76)\\
&= \theta^2 e^{-\theta x_1 - \theta(\frac{1}{v}-1)x_1} \cdot \frac{x_1}{v^2} && (77)\\
&= \theta^2 e^{-\theta\frac{x_1}{v}} \cdot \frac{x_1}{v^2} \ (x_1 > 0, 0 < v < 1) && (78)
\end{aligned}
$$

so the marginal density of $V$ is given by

$$
f_V(v) = \int_0^\infty f_{X_1,V}(x_1, v)\, dx_1 = 1 \ (v \in (0, 1)) \tag{79}
$$

so $V \sim U(0, 1)$ is uniform and of course ancillary.

Notice that $\overline{X}$ is complete and sufficient, by Basu's theorem, $\overline{X}$ is independent of $\frac{X_1}{X_1+X_2}$.

5.5.1.

(a):

Binomial is in exponential family only when $n$ is known! Support does not depend on $p$

$$
\begin{aligned}
p_p(x) &= p^{\sum_i x_i}(1-p)^{n-\sum_i x_i} && (80)\\
&= e^{(\sum_i x_i)\log p + (n-\sum_i x_i)\log(1-p)} && (81)\\
&= e^{n\log(1-p) + \sum_i x_i \log\frac{p}{1-p}} && (82)
\end{aligned}
$$

where $c(p) = e^{n\log(1-p)}, Q(p) = \log\frac{p}{1-p}, T(x) = \sum_i x_i$.

(b):

Support does not depend on $\alpha, \beta$

$$
p_{\alpha,\beta}(x) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}(\prod_i x_i)^{\alpha-1} e^{-\beta\sum_i x_i} \tag{83}
$$

If $\alpha$ is known,

$$
p_\beta(x) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}(\prod_i x_i)^{\alpha-1} e^{-\beta\sum_i x_i} \tag{84}
$$

so $c(\beta) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}, h(x) = (\prod_i x_i)^{\alpha-1}, Q(\beta) = -\beta, T(x) = \sum_i x_i$.

If $\beta$ is known,

$$p_\alpha(x) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}(\prod_i x_i)^{\alpha-1}e^{-\beta\sum_i x_i} \tag{85}$$
$$= \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}e^{-\beta\sum_i x_i}e^{(\alpha-1)\sum_i \log x_i} \tag{86}$$

so $c(\alpha) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}, h(x) = e^{-\beta\sum_i x_i}, Q(\alpha) = \alpha - 1, T(x) = \sum_i \log x_i$.

(c):

Support does not depend on $\alpha, \beta$

$$p_{\alpha,\beta}(x) = \frac{1}{B^n(\alpha,\beta)}(\prod_i x_i)^{\alpha-1}[\prod_i(1-x_i)]^{\beta-1} \tag{87}$$

If $\alpha$ is known,

$$p_\beta(x) = \frac{1}{B^n(\alpha,\beta)}(\prod_i x_i)^{\alpha-1}[\prod_i(1-x_i)]^{\beta-1} \tag{88}$$
$$= \frac{1}{B^n(\alpha,\beta)}(\prod_i x_i)^{\alpha-1}e^{(\beta-1)\sum_i \log(1-x_i)} \tag{89}$$

so $c(\beta) = \frac{1}{B^n(\alpha,\beta)}, h(x) = (\prod_i x_i)^{\alpha-1}, Q(\beta) = \beta - 1, T(x) = \sum_i \log(1-x_i)$

If $\beta$ is known,

$$p_\alpha(x) = \frac{1}{B^n(\alpha,\beta)}(\prod_i x_i)^{\alpha-1}[\prod_i(1-x_i)]^{\beta-1} \tag{90}$$
$$= \frac{1}{B^n(\alpha,\beta)}[\prod_i(1-x_i)]^{\beta-1}e^{(\alpha-1)\sum_i \log x_i} \tag{91}$$

so $c(\alpha) = \frac{1}{B^n(\alpha,\beta)}, h(x) = [\prod_i(1-x_i)]^{\beta-1}, Q(\alpha) = \alpha - 1, T(x) = \sum_i \log x_i$

(d):

Support does not depend on $p$

$$p_p(x) = (1-p)^x p^r = p^r e^{x\log(1-p)} \tag{92}$$

so $c(p) = p^r, Q(p) = \log(1-p), T(x) = x$.

5.5.5.

(a):

Support does not depend on $\alpha, \beta$

$$p_{\alpha,\beta}(x) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}(\prod_i x_i)^{\alpha-1} e^{-\beta\sum_i x_i} \tag{93}$$

$$= \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)} e^{(\alpha-1)\sum_i \log x_i} e^{-\beta\sum_i x_i} \tag{94}$$

so $c(\alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)}, Q_1(\alpha, \beta) = \alpha - 1, T_1(x) = \sum_i \log x_i, Q_2(\alpha, \beta) = -\beta, T_2(x) = \sum_i x_i$.

(b):

Support does not depend on $\alpha, \beta$

$$p_{\alpha,\beta}(x) = \frac{1}{B^n(\alpha, \beta)}(\prod_i x_i)^{\alpha-1}[\prod_i (1 - x_i)]^{\beta-1} \tag{95}$$

$$= \frac{1}{B^n(\alpha, \beta)} e^{(\alpha-1)\sum_i \log x_i} e^{(\beta-1)\sum_i \log(1-x_i)} \tag{96}$$

so $c(\alpha, \beta) = \frac{1}{B^n(\alpha,\beta)}, Q_1(\alpha, \beta) = \alpha - 1, T_1(x) = \sum_i \log x_i, Q_2(\alpha, \beta) = \beta - 1, T_2(x) = \sum_i \log(1 - x_i)$.

8.8.2

(a):

Since loss is SEL, Bayes estimator is just posterior mean.

The likelihood is

$$p_\lambda(x) = \lambda^{\sum_i x_i} e^{-n\lambda} \tag{97}$$

so the posterior is

$$\pi(\lambda|x) \propto \pi(\lambda) \cdot p_\lambda(x) = \lambda^{\sum_i x_i} e^{-(n+1)\lambda} \tag{98}$$

is actually a gamma distribution $\lambda|_x \sim \Gamma(\sum_i x_i + 1, n+1)$, so posterior mean is $\frac{\sum_i x_i + 1}{n+1}$. So $\delta_\pi(X) = \frac{\sum_{i=1}^n X_i + 1}{n+1}$ is the Bayes estimator.

(b):

Just need to get $\mathbb{E}_{\lambda \sim \pi(\cdot|x)} e^{-\lambda}$. Now the posterior distribution is already known, and notice that if $Z \sim \Gamma(\alpha, \beta)$, then

$$\begin{aligned} \mathbb{E}e^{-Z} &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty z^{\alpha-1} e^{-(\beta+1)z}\, dz & (99) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)(\beta+1)^\alpha} \int_0^\infty u^{\alpha-1} e^{-u}\, du \ (u = (\beta+1)z) & (100) \\ &= \frac{\beta^\alpha}{(\beta+1)^\alpha} & (101) \end{aligned}$$

so plugging in numbers to see that $\mathbb{E}_{\lambda \sim \pi(\cdot|x)} e^{-\lambda} = \left(\frac{n+1}{n+2}\right)^{\sum_{i=1}^n x_i + 1}$. So Bayes estimator is $\delta_\pi(X) = \left(\frac{n+1}{n+2}\right)^{\sum_{i=1}^n X_i + 1}$.

8.8.8

Under SEL, the Bayes estimator is the posterior mean. Now the likelihood is

$$p_\theta(x) = \theta^{-n} \ (x_1, ..., x_n \in (0, \theta)) \tag{102}$$

so by the Bayes theorem, the posterior is

$$\begin{aligned} \pi(\theta|x) &\propto p_\theta(x) \cdot \pi(\theta) = \theta^{-n} \frac{\alpha a^\alpha}{\theta^{\alpha+1}} & (103) \\ &\propto \theta^{-n-\alpha-1} \ (\theta \geq a, \theta \geq x_{(n)}) & (104) \end{aligned}$$

let's set $\beta = \max\left\{a, x_{(n)}\right\}$, so $\theta \geq \beta$ is the support of the posterior.

Notice that

$$\int_\beta^\infty \theta^{-n-\alpha-1}\, d\theta = \frac{\beta^{-n-\alpha}}{n+\alpha} \tag{105}$$

so $\pi(\theta|x) = (n+\alpha)\beta^{n+\alpha}\theta^{-n-\alpha-1}$ $(\theta \geq \beta)$ is the posterior density and to calculate the expectation

$$(n+\alpha)\beta^{n+\alpha}\int_\beta^\infty \theta^{-n-\alpha}\,d\theta = (n+\alpha)\beta^{n+\alpha}\frac{\beta^{-n-\alpha+1}}{n+\alpha-1} = \beta\frac{n+\alpha}{n+\alpha-1} \tag{106}$$

so the Bayes estimator is $\delta_\pi(X) = \max\left\{a, X_{(n)}\right\}\cdot\frac{n+\alpha}{n+\alpha-1}$.

8.8.10

Since loss is SEL, Bayes estimator is just posterior mean. Likelihood is

$$p_\theta(x) = e^{n\theta-\sum_i x_i}\ (x_1,...,x_n > \theta) \tag{107}$$

so by the Bayes theorem, the posterior is

$$\begin{aligned}\pi(\theta|x) &\propto p_\theta(x)\cdot\pi(\theta) = e^{n\theta-\sum_i x_i}\cdot e^{-\theta}\ (0<\theta<x_{(1)}) &(108)\\ &\propto e^{(n-1)\theta}\ (0<\theta<x_{(1)}) &(109)\end{aligned}$$

notice that

$$\int_0^{x_{(1)}} e^{(n-1)\theta}\,d\theta = \frac{e^{(n-1)x_{(1)}}-1}{n-1} \tag{110}$$

so $\pi(\theta|x) = \frac{n-1}{e^{(n-1)x_{(1)}}-1}e^{(n-1)\theta}$ $(0<\theta<x_{(1)})$ is the posterior density and to calculate the expectation

$$\begin{aligned}\frac{n-1}{e^{(n-1)x_{(1)}}-1}\int_0^{x_{(1)}}\theta e^{(n-1)\theta}\,d\theta &= \frac{n-1}{e^{(n-1)x_{(1)}}-1}\frac{1}{n-1}\left(x_{(1)}e^{(n-1)x_{(1)}} - \frac{e^{(n-1)x_{(1)}}-1}{n-1}\right) &(111)\\ &= \frac{x_{(1)}e^{(n-1)x_{(1)}}}{e^{(n-1)x_{(1)}}-1} - \frac{1}{n-1} &(112)\end{aligned}$$

so the Bayes estimator is $\delta_\pi(X) = \frac{X_{(1)}e^{(n-1)X_{(1)}}}{e^{(n-1)X_{(1)}}-1} - \frac{1}{n-1}$.

8.8.11

Consider the risk $R(\theta,\delta_a(X))$ that

$$R(\theta,\delta_a(X)) = \mathbb{E}_\theta\left[\theta - a\left(X_{(1)}-\frac{1}{n}\right)\right]^2 \tag{113}$$

with the density of $X_{(1)}$ as

$$f_1(x) = ne^{n(\theta-x)}\ (x>\theta) \tag{114}$$

so the risk can be calculated as

$$R(\theta,\delta_a(X)) = \int_\theta^\infty\left(\theta+\frac{a}{n}-ax\right)^2 ne^{n(\theta-x)}\,dx = n\int_0^\infty\left[(1-a)\theta+\frac{a}{n}-au\right]^2 e^{-nu}\,du \tag{115}$$

change variables $v = nu$ to see

$$R(\theta, \delta_a(X)) = \int_0^\infty \left[(1-a)\theta + \frac{a}{n} - \frac{a}{n}v\right]^2 e^{-v}\,dv \tag{116}$$
$$= \left[(1-a)\theta + \frac{a}{n}\right]^2 - 2\left[(1-a)\theta + \frac{a}{n}\right]\frac{a}{n} + \frac{2a^2}{n^2} \tag{117}$$
$$= \left[(1-a)\theta + \frac{a}{n}\right]^2 - \frac{2a(1-a)\theta}{n} \tag{118}$$
$$= (1-a)^2\theta^2 + \frac{a^2}{n^2} \tag{119}$$

so $\sup_{\theta>0} R(\theta, \delta_a(X)) = \begin{cases} +\infty & a \neq 1 \\ \frac{1}{n^2} & a = 1 \end{cases}$ so $a^* = 1$ achieves the inf and the minimax decision rule in this class is $\delta(X) = \left(X_{(1)} - \frac{1}{n}\right)$.

In minimax setting, take sup w.r.t. $\theta \in \Theta$ and the parameter space does not depend on sample realization $x$.

8.8.3

Compute the risk

$$R(\theta, \delta_\alpha(X)) = \mathbb{E}_\theta(\theta - \overline{X} - \alpha)^2 \tag{120}$$
$$= (\theta-\alpha)^2 - 2(\theta-\alpha)\mathbb{E}_\theta\overline{X} + \mathbb{E}_\theta\overline{X}^2 \tag{121}$$
$$= (\theta-\alpha)^2 - 2(\theta-\alpha)\theta + \frac{n\theta + n(n-1)\theta^2}{n^2} \tag{122}$$
$$= -\frac{1}{n}\theta^2 + \frac{1}{n}\theta + \alpha^2 \tag{123}$$

so $\sup_{\theta\in(0,1)} R(\theta, \delta_\alpha(X)) = \frac{1}{4n} + \alpha^2$, and $\alpha = 0$ achieves $\inf_{a\in\mathbb{R}} \sup_{\theta\in(0,1)} R(\theta, \delta_\alpha(X))$, so minimax decision rule is $\delta(X) = \overline{X}$.

8.8.5

Compute posterior

$$\pi(\theta|x) \propto p_\theta(x) \cdot \pi(\theta) = \theta^x(1-\theta)^{n-x} \cdot 1\ (0 < \theta < 1) \tag{124}$$

so the posterior is actually $Beta(x+1, n-x+1)$.

Compute expected posterior loss that

$$l(d, x) = \int_\Theta L(\theta, d) \cdot \pi(\theta|x)\,d\theta \tag{125}$$
$$= \frac{1}{B(x+1, n-x+1)} \int_0^1 \frac{(\theta-d)^2}{\theta(1-\theta)} \cdot \theta^x(1-\theta)^{n-x}\,d\theta \tag{126}$$

simplify to get

$$
\begin{aligned}
& \frac{1}{B(x+1,n-x+1)}\int_0^1 (\theta-d)^2\theta^{x-1}(1-\theta)^{n-x-1}\,d\theta && (127)\\
&= \frac{1}{B(x+1,n-x+1)}\left[\int_0^1 \theta^{x+1}(1-\theta)^{n-x-1}\,d\theta - 2d\int_0^1 \theta^{x}(1-\theta)^{n-x-1}\,d\theta + d^2\int_0^1 \theta^{x-1}(1-\theta)^{n-x-1}\,d\theta\right] && (128)\\
&= \frac{B(x+2,n-x)}{B(x+1,n-x+1)} - 2d\frac{B(x+1,n-x)}{B(x+1,n-x+1)} + d^2\frac{B(x,n-x)}{B(x+1,n-x+1)} && (129)\\
&= \frac{x+1}{n-x} - 2d\frac{n+1}{n-x} + d^2\frac{n(n+1)}{x(n-x)} && (130)
\end{aligned}
$$

then minimize EPL w.r.t. $d$ to get

$$d_0(x) = \arg\min_d l(d,x) = \frac{x}{n} \tag{131}$$

so $\delta_\pi(X) = \frac{X}{n}$ is Bayes estimator.

Notice that the Bayes estimator has risk

$$R(\theta,\delta_\pi(X)) = \frac{1}{\theta(1-\theta)}Var\left(\frac{X}{n}\right) = \frac{1}{n} \tag{132}$$

is constant in $\theta$, so it must also be minimax.

SECTION 10.6

2.

(b):

The expected posterior loss is

$$\begin{aligned} l(d_0, x) &= \mathbb{E}_{\theta\sim\pi(\cdot|x)} L(\theta, d_0) && (133) \\ &= 2\pi(\theta_1|x) && (134) \\ l(d_1, x) &= \mathbb{E}_{\theta\sim\pi(\cdot|x)} L(\theta, d_1) && (135) \\ &= \pi(\theta_0|x) && (136) \end{aligned}$$

so now the Bayes test is a test that rejects $H_0$ iff $\pi(\theta_0|x) < 2\pi(\theta_1|x)$, i.e. $\pi(\theta_0)p_{\theta_0}(x) < 2\pi(\theta_1)p_{\theta_1}(x)$ so $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} > \frac{\pi_0}{2\pi_1}$.

This tells us that the family of Bayes test is

$$\delta_k(x) \tag{137}$$

that rejects $H_0$ iff $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = \frac{\theta_1^{\sum_i x_i}(1-\theta_1)^{5-\sum_i x_i}}{\theta_0^{\sum_i x_i}(1-\theta_0)^{5-\sum_i x_i}} = \left(\frac{3}{2}\right)^{\sum_i x_i}\left(\frac{1}{2}\right)^{5-\sum_i x_i} > k$ so $\sum_i x_i > \log_3(32k) = c$. So consider the family of Bayes test

$$\delta_c(x) \tag{138}$$

that rejects $H_0$ iff $\sum_i x_i > c$.

In order to find the minimax test, we want to find $c$ such that the risk function is constant.

$$\begin{aligned} R(\theta_0, \delta_c) &= \mathbb{E}_{X\sim p_{\theta_0}} L(\theta_0, \delta_c(X)) = \mathbb{P}_{\theta_0}\left(\delta_c(X) = d_1\right) && (139) \\ &= \mathbb{P}_{\theta_0}\left(\sum_i X_i > c\right) && (140) \end{aligned}$$

with $\sum_i X_i|_{\theta=\theta_0} \sim B(5, \frac{1}{2})$. So $R(\theta_0, \delta_c) = \sum_{k=c+1}^{5}\binom{5}{k}\frac{1}{32}$.

Similarly,

$$\begin{aligned} R(\theta_1, \delta_c) &= \mathbb{E}_{X\sim p_{\theta_1}} L(\theta_1, \delta_c(X)) = 2\mathbb{P}_{\theta_1}\left(\delta_c(X) = d_0\right) && (141) \\ &= 2\mathbb{P}_{\theta_1}\left(\sum_i X_i \le c\right) && (142) \end{aligned}$$

with $\sum_i X_i|_{\theta=\theta_1} \sim B(5, \frac{3}{4})$. So $R(\theta_1, \delta_c) = 2\sum_{k=0}^{c}\binom{5}{k}(\frac{3}{4})^k(\frac{1}{4})^{5-k}$.

Now set the values of risk function at two points to be equal

$$16\left(32-\sum_{k=0}^{c}\binom{5}{k}\right)=\sum_{k=0}^{c}\binom{5}{k}3^k \tag{143}$$

$$512=\sum_{k=0}^{c}\binom{5}{k}(3^k+16) \tag{144}$$

to find that such $c$ is between 2 to 3 and does not exist. However, this is telling us that when $\sum_i X_i$ is 0 or 1 or 2 we shall reject $H_1$ and when the sum is 4 or 5 we shall reject $H_0$. So why don't we consider the randomized decision rule that when the sum is 3 we have $p$ probability of rejecting $H_0$. Let's compute the risk function

$$R(\theta_0,\delta_c)=\sum_{k=4,5}^{5}\binom{5}{k}\frac{1}{32}+\binom{5}{3}\frac{1}{32}p \tag{145}$$

$$=\frac{6+10p}{32} \tag{146}$$

and

$$R(\theta_1,\delta_c)=2\sum_{k=0,1,2}\binom{5}{k}\left(\frac{3}{4}\right)^k\left(\frac{1}{4}\right)^{5-k}+2\binom{5}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2(1-p) \tag{147}$$

$$=2\frac{1+15+90+270(1-p)}{4^5} \tag{148}$$

$$=\frac{212+540(1-p)}{4^5}=\frac{752-540p}{1024} \tag{149}$$

so $\sup_\theta R(\theta,\delta_c(X))=\max\left\{\frac{6+10p}{32},\frac{752-540p}{1024}\right\}$, so to minimize it, these two numbers shall be the same and $p=\frac{28}{43}$ so the minimax rule is to reject $H_0$ iff $\sum_{i=1}^5 X_i$ is 4 or 5 and $\frac{28}{43}$ probability to reject when the sum is 3.

(c):

The Bayes test is to reject $H_0$ iff $\left(\frac{3}{2}\right)^{\sum_i x_i}\left(\frac{1}{2}\right)^{5-\sum_i x_i}>k=\frac{\pi_0}{2\pi_1}=\frac{1}{4}$ so it's equivalent to saying $\sum_i x_i\geq 2$.

3.

The expected posterior loss is

$$l(d_0,x)=\mathbb{E}_{\theta\sim\pi(\cdot|x)}L(\theta,d_0) \tag{150}$$

$$=\pi(\theta_1|x) \tag{151}$$

$$l(d_1,x)=\mathbb{E}_{\theta\sim\pi(\cdot|x)}L(\theta,d_1) \tag{152}$$

$$=\pi(\theta_0|x) \tag{153}$$

so now the Bayes test is a test that rejects $H_0$ iff $\pi(\theta_0|x)<\pi(\theta_1|x)$, i.e. $\pi(\theta_0)p_{\theta_0}(x)<\pi(\theta_1)p_{\theta_1}(x)$ so $\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}>\frac{\pi_0}{\pi_1}=2$.

Now let's calculate the likelihood ratio

$$\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = \frac{\theta_1^n e^{-\theta_1 \sum_i x_i}}{\theta_0^n e^{-\theta_0 \sum_i x_i}} = 2^n e^{-\sum_i x_i} > 2 \tag{154}$$

so we shall reject $H_0$ iff $n \log 2 - \sum_i X_i > \log 2$, i.e. $\sum_{i=1}^n X_i < (n-1)\log 2$.

5.

The Bayes solution to classification problem is the MAP estimator, i.e. choose the decision with the largest posterior likelihood. Now

$$\pi(\theta_1|x) = \frac{p_{\theta_1}(x)\pi(\theta_1)}{m(x)} \tag{155}$$

so we just need to compare $p_{\theta_1}(x)\pi(\theta_1) = \frac{2}{5}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x+1)^2}{2}}, p_{\theta_2}(x)\pi(\theta_2) = \frac{2}{5}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, p_{\theta_3}(x)\pi(\theta_3) = \frac{1}{5}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}}$ and find the largest one.

So let's consider $\max\left\{2e^{-X-\frac{1}{2}}, 2, e^{X-\frac{1}{2}}\right\}$ and we see that the Bayes rule is to take the decision $d_i$ if the $i$-th term in the max takes the maximum.

6.

Now in classification with $k$ possible values for parameter $\theta$ as $\theta_1, ..., \theta_k$ with prior $\pi_i = \pi(\theta_i)$ such that $\pi_1 + ... + \pi_k = 1$ with 0-1 loss function. Consider the expected posterior loss

$$\begin{aligned} l(d_i, x) &= \mathbb{E}_{\theta\sim\pi(\cdot|x)} L(\theta, d_i) & (156)\\ &= \sum_{j\neq i} \pi(\theta_j|x) & (157) \end{aligned}$$

so if now we have $\forall j \neq i, \pi_i p_{\theta_i}(x) \geq \pi_j p_{\theta_j}(x)$, then $\forall j \neq i, \pi(\theta_i|x) \geq \pi(\theta_j|x)$ so

$$\forall k \neq i, \sum_{j\neq i} \pi(\theta_j|x) \leq \sum_{j\neq k} \pi(\theta_j|x) \tag{158}$$

and this tells us that in order to minimize the expected posterior loss, we shall set $d = d_i$, which means that the Bayes rule is to accept $\theta_i$.

SECTION 8.6

2.

Set $\frac{\alpha}{\beta} = \overline{X}$, $\frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2} = \frac{\alpha^2+\alpha}{\beta^2} = \frac{1}{n}\sum_{i=1}^n X_i^2$ to solve out moment estimators

$$\hat{\alpha} = \frac{\overline{X}^2}{\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2}, \hat{\beta} = \frac{\overline{X}}{\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2} \tag{159}$$

4.

Set $\frac{\alpha}{\alpha+\beta} = \overline{X}$, $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{1}{n}\sum_{i=1}^n X_i^2$ to solve out moment estimators $\hat{\alpha}, \hat{\beta}$.

5.

Set $e^{\mu+\frac{\sigma^2}{2}} = \overline{X}$, $e^{\sigma^2}e^{2\mu+\sigma^2} = \frac{1}{n}\sum_{i=1}^n X_i^2$ to solve out $\hat{\mu}, \hat{\sigma}$.

SECTION 8.7

2.

(c).

$$p_\theta(x) = (2\pi)^{-\frac{n}{2}}\theta^{-n}e^{-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\theta^2}} \tag{160}$$

$$l(\theta) = -\frac{n}{2}\log(2\pi) - n\log\theta - \frac{\sum_{i=1}^n (x_i-\theta)^2}{2\theta^2} \tag{161}$$

$$= -\frac{n}{2}\log(2\pi) - n\log\theta - \frac{\sum_{i=1}^n x_i^2 - 2\theta\sum_{i=1}^n x_i + n\theta^2}{2\theta^2} \tag{162}$$

$$= -\frac{n}{2}\log(2\pi) - n\log\theta - \frac{\sum_{i=1}^n x_i^2}{2}\theta^{-2} + \sum_{i=1}^n x_i\theta^{-1} - \frac{n}{2} \tag{163}$$

take derivative set to 0

$$-\frac{n}{\theta} - \sum_{i=1}^n x_i\theta^{-2} + \sum_{i=1}^n x_i^2\theta^{-3} = 0 \tag{164}$$

$$-n\theta^2 - \sum_{i=1}^n x_i\theta + \sum_{i=1}^n x_i^2 = 0 \tag{165}$$

solve it to get MLE (and check second derivative).

(e).

$$p_\theta(x) = (2\pi)^{-\frac{n}{2}}\theta^{-\frac{n}{2}}e^{-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\theta}} \tag{166}$$

$$l(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta - \frac{\sum_{i=1}^n (x_i-\theta)^2}{2\theta} \tag{167}$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta - \frac{\sum_{i=1}^n x_i^2 - 2\theta\sum_{i=1}^n x_i + n\theta^2}{2\theta} \tag{168}$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta - \frac{\sum_{i=1}^n x_i^2}{2}\theta^{-1} + \sum_{i=1}^n x_i - \frac{n}{2}\theta \tag{169}$$

take derivative set to 0

$$-\frac{n}{2\theta} - \frac{n}{2} + \frac{\sum_{i=1}^n x_i^2}{2}\theta^{-2} = 0 \tag{170}$$

$$-\frac{n}{2}\theta - \frac{n}{2}\theta^2 + \frac{\sum_{i=1}^n x_i^2}{2} = 0 \tag{171}$$

solve it to get MLE (and check second derivative).

3

We actually just know $Y_1, ..., Y_n$ *i.i.d.* with distribution $B(1, \Phi(-\mu))$, $Y_i = 1$ means $X_i < 0$ and $Y_i = 0$ means $X_i \geq 0$. Now we know that $\sum_{i=1}^n Y_i = m$.

Consider joint likelihood

$$p_\mu(y) = [\Phi(-\mu)]^{\sum_{i=1}^n y_i}[1-\Phi(-\mu)]^{n-\sum_{i=1}^n y_i} \tag{172}$$

$$= [\Phi(-\mu)]^{\sum_{i=1}^n y_i}[\Phi(\mu)]^{n-\sum_{i=1}^n y_i} \tag{173}$$

$$l(\mu) = \sum_{i=1}^n y_i \cdot \log\Phi(-\mu) + (n - \sum_{i=1}^n y_i)\cdot\log\Phi(\mu) \tag{174}$$

$$l'(\mu) = -\sum_{i=1}^n y_i \cdot \frac{\varphi(-\mu)}{\Phi(-\mu)} + (n - \sum_{i=1}^n y_i)\cdot\frac{\varphi(\mu)}{\Phi(\mu)} = 0 \tag{175}$$

to solve out $\hat{\mu} = \Phi^{-1}(\frac{n-m}{n})$ as the MLE.

4

(a)

$$p_{\alpha,\beta}(x) = \beta^{-n}e^{-\beta^{-1}(\sum_{i=1}^n x_i - n\alpha)} \tag{176}$$

$$l(\alpha,\beta) = -n\log\beta - \beta^{-1}(\sum_{i=1}^n x_i - n\alpha) \tag{177}$$

$$\frac{\partial l}{\partial\alpha} = \frac{n}{\beta} > 0, \frac{\partial l}{\partial\beta} = -\frac{n}{\beta} + \frac{1}{\beta^2}(\sum_{i=1}^n x_i - n\alpha) \tag{178}$$

so when $\beta$ is fixed and $\alpha$ gets larger, likelihood is larger. However, $\alpha$ has upper bound $X_{(1)}$ so $\hat{\alpha} = X_{(1)}$. Now maximize w.r.t. $\beta$ so set partial derivative to 0 to get $\hat{\beta} = \overline{X} - X_{(1)}$ to be MLE.

(b)

$$\mathbb{P}(X_1 \geq 1) = \begin{cases} 1 & \alpha \geq 1 \\ e^{-\frac{1}{\beta}(1-\alpha)} & \alpha < 1 \end{cases} \tag{179}$$

so the MLE should be

$$\mathbb{P}(X_1 \hat{\geq} 1) = \begin{cases} 1 & X_{(1)} \geq 1 \\ e^{-\frac{1}{\overline{X}-X_{(1)}}(1-X_{(1)})} & X_{(1)} < 1 \end{cases} \tag{180}$$

by functional invariance.

15

$$p_\mu(x) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_i (x_i-\mu)^2} \tag{181}$$

$$l(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_i (x_i-\mu)^2 \ (\mu \geq 0) \tag{182}$$

$$l'(\mu) = \sum_i (x_i - \mu) = -n\mu + n\overline{X} \tag{183}$$

so if $\overline{X} \geq 0$, then $\hat{\mu} = \overline{X}$ is the MLE and otherwise $\hat{\mu} = 0$ is the MLE.

SECTION 8.4

2

We want to find unbiased estimator for $\sigma^p$ so naturally we consider $S^p$ where $V = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1} = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$. So now

$$(n-1)^{\frac{p}{2}} \frac{\mathbb{E}S^p}{\sigma^p} = \mathbb{E}V^{\frac{p}{2}} = \frac{(\frac{1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_0^\infty v^{\frac{p}{2}} v^{\frac{n-3}{2}} e^{-\frac{v}{2}}\, dv \tag{184}$$

$$= \frac{(\frac{1}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} 2^{\frac{n+p-1}{2}} \int_0^\infty u^{\frac{n+p-3}{2}} e^{-u}\, du \tag{185}$$

$$= 2^{\frac{p}{2}} \frac{\Gamma(\frac{n+p-1}{2})}{\Gamma(\frac{n-1}{2})} = C \tag{186}$$

and we find that $\hat{\sigma^p} = \frac{(n-1)^{\frac{p}{2}}}{C} S^p$ is an unbiased estimator for $\sigma^p$.

It's obvious that the minimum MSE estimator does not exist (since parameter space contains more than 1 elements). Now if we consider the minimum MSE estimator among all unbiased estimator, it must be UMVUE and notice that the estimator we have already got is a function of $(\overline{X}, S^2)$, the complete sufficient statistic and that it's unbiased, it must be UMVUE.

5

If there exists $T(X)$ unbiased then

$$\forall\lambda > 0, \mathbb{E}T = \sum_{k=0}^{\infty} T(k)\frac{\lambda^k}{k!}e^{-\lambda} = \frac{1}{\lambda} \tag{187}$$

so

$$\forall\lambda > 0, \sum_{k=0}^{\infty} T(k)\frac{\lambda^k}{k!} = \frac{e^{\lambda}}{\lambda} = \sum_{k=0}^{\infty}\frac{1}{\lambda}\frac{\lambda^k}{k!} \tag{188}$$

and $\forall\lambda > 0, \sum_{k=0}^{\infty}[T(k)-\frac{1}{\lambda}]\frac{\lambda^k}{k!} = 0$ since $\mathbb{E}|T| < \infty$ the series is absolute convergent. This tells us that $\forall k, T(k) = \frac{1}{\lambda}$ which is impossible since the estimator cannot contain the unknown parameter, contradiction!

9

Let's notice that $d(\theta) = \mathbb{E}\mathbb{I}_{X_1=0}$ so take $U = \mathbb{I}_{X_1=0}$ as an unbiased estimator and notice that $X_1, ..., X_n \sim NB(1,p)$ so

$$p_p(x) = p^n(1-p)^{\sum_i x_i} \tag{189}$$

and likelihood ratio $\frac{p_p(x)}{p_p(y)} = (1-p)^{\sum_i x_i - \sum_i y_i}$ so $T = \sum_{i=1}^n X_i$ is minimal sufficient. Now notice that

$$p_p(x) = e^{\log p + x\log(1-p)} \tag{190}$$

so $T$ is also complete by verifying OSC. Now $V = \mathbb{E}(U|T)$ must be UMVUE so

$$\begin{aligned}\mathbb{E}(U|T=t) &= \mathbb{P}\left(X_1 = 0|T = t\right) &(191)\\ &= \frac{\mathbb{P}\left(X_1 = 0, X_2 + ... + X_n = t\right)}{\mathbb{P}\left(T = t\right)} &(192)\\ &= \frac{\mathbb{P}\left(X_1 = 0\right)\mathbb{P}\left(X_2 + ... + X_n = t\right)}{\mathbb{P}\left(T = t\right)} &(193)\end{aligned}$$

with $T \sim NB(n,p), X_2 + ... + X_n \sim NB(n-1,p)$ so

$$\begin{aligned}\mathbb{E}(U|T=t) &= \frac{p\binom{t+n-2}{t}(1-p)^t p^{n-1}}{\binom{t+n-1}{t}(1-p)^t p^n} &(194)\\ &= \frac{n-1}{t+n-1} &(195)\end{aligned}$$

so UMVUE is $\frac{n-1}{\sum_{i=1}^n X_i + n - 1}$.

15

$X_1, ..., X_n \sim P(\lambda)$, now take unbiased estimator $U = \mathbb{I}_{X_1=k}$ and we have complete sufficient statistic $T =$

$\sum_{i=1}^n X_i$. By Rao-Blackwell, we just have to compute

$$\begin{aligned}
\mathbb{E}(U|T=t) &= \mathbb{P}\left(X_1=k|T=t\right) &(196)\\
&= \frac{\mathbb{P}\left(X_1=k, X_2+...+X_n=t-k\right)}{\mathbb{P}\left(T=t\right)} &(197)\\
&= \frac{\mathbb{P}\left(X_1=k\right)\mathbb{P}\left(X_2+...+X_n=t-k\right)}{\mathbb{P}\left(T=t\right)} &(198)\\
&= \frac{\frac{\lambda^k}{k!}e^{-\lambda}\frac{[(n-1)\lambda]^{t-k}}{(t-k)!}e^{-(n-1)\lambda}}{\frac{(n\lambda)^t}{t!}e^{-n\lambda}} &(199)\\
&= \binom{t}{k}\frac{1}{n^k}\left(1-\frac{1}{n}\right)^{t-k} &(200)
\end{aligned}$$

so the UMVUE is $\binom{\sum_{i=1}^n X_i}{k}\frac{1}{n^k}\left(1-\frac{1}{n}\right)^{\sum_{i=1}^n X_i - k}$.

READING

SECTION 8.7

7

$X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned}
p_{\mu,\sigma}(x) &= (2\pi)^{-\frac{1}{2}}\sigma^{-1}e^{-\frac{(x-\mu)^2}{2\sigma^2}} &(201)\\
l(\mu,\sigma) &= -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2} &(202)\\
l'_\mu &= \frac{x-\mu}{\sigma^2} &(203)\\
l'_\sigma &= -\frac{1}{\sigma} + (x-\mu)^2\sigma^{-3} &(204)
\end{aligned}$$

so we can always set $\mu = x$ and $\sigma \to 0$ so the likelihood explodes to infinity, that's why there's no MLE.

17

(a):

$$\begin{aligned}
M(t) = \mathbb{E}e^{tT(X)} &= \int e^{tT(x)}e^{\eta T(x)+D(\eta)+S(x)}\,dx &(205)\\
&= e^{D(\eta)}\int e^{(t+\eta)T(x)+S(x)+D(\eta+t)-D(\eta+t)}\,dx &(206)\\
&= e^{D(\eta)-D(\eta+t)} &(207)
\end{aligned}$$

since the density integrates to 1 when $\eta$ is replaced with $\eta+t$. So $\mathbb{E}T(X) = M'(0) = -e^{D(\eta)-D(\eta)}D'(\eta) = -D'(\eta)$ and $\mathbb{E}[T(X)]^2 = M''(0) = e^{D(\eta)-D(\eta)}([D'(\eta)]^2 - D''(\eta)) = [D'(\eta)]^2 - D''(\eta)$ so $Var(T(X)) = -D''(\eta)$.

(b):

Now if $\mathbb{E}_\eta T(X) = T(x)$ has solution $\eta_0(x)$, then $-D'(\eta_0) = T(x)$. The likelihood is

$$p_\eta(x) = e^{\eta T(x) + D(\eta) + S(x)} \tag{208}$$
$$l(\eta) = \eta T(x) + D(\eta) + S(x) \tag{209}$$
$$l'(\eta) = T(x) + D'(\eta) \tag{210}$$

so $l'(\eta_0) = 0$ takes the largest likelihood (since variance is positive, $D''(\eta_0) < 0$ so it's a maximum), so $\eta_0(X)$ is MLE.

Prove uniqueness, if there is another statistic $U(X)$ which is also MLE, then $\forall x \in \mathscr{X}, D'(\eta_0(x)) = D'(U(x))$ so $\forall x \in \mathscr{X}, \eta_0(x) = U(x)$ and contradiction!

SECTION 8.4

1

Sample mean $\overline{X}$ is unbiased for $p$ and sample variance $S^2$ is unbiased for $p(1-p)$. So $\mathbb{E}[\overline{X} - S^2] = p - p(1-p) = p^2$ and $\overline{X} - S^2$ is unbiased estimator for $p^2$.

3

$$MSE(\alpha S^2) = \mathbb{E}(\alpha S^2 - \sigma^2)^2 \tag{211}$$
$$= [\mathbb{E}\alpha S^2 - \sigma^2]^2 + Var(\alpha S^2) \tag{212}$$
$$= (1-\alpha)^2 \sigma^4 + \alpha^2 Var(S^2) \tag{213}$$

since $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$ has variance $2(n-1)$, so $Var(S^2) = \frac{2}{n-1}\sigma^4$. So $MSE(\alpha S^2) = \sigma^4 \left(\frac{2}{n-1}\alpha^2 + (1-\alpha)^2\right)$ and minimize w.r.t. $\alpha$ to get $\alpha^* = \frac{n-1}{n+1}$ so the minimum MSE estimator is $\frac{n-1}{n+1}S^2$.

Now $Var(\frac{n-1}{n+1}S^2) = \frac{2(n-1)}{(n+1)^2}\sigma^4$ and $Var(S^2) = \frac{2}{n-1}\sigma^4$ so $\frac{n-1}{n+1}S^2$ has less variance.

6

(a):

The complete sufficient statistic is $T = \sum_{i=1}^n X_i \sim B(n,p)$ so let's consider the factorial moment

$$\mathbb{E}[T(T-1)...(T-s+1)] = \sum_{k=s}^n k(k-1)...(k-s+1)\binom{n}{k} p^k (1-p)^{n-k} \tag{214}$$
$$= \frac{n!}{(n-s)!}p^s \tag{215}$$

this is telling us that UMVUE for $p^s$ is just $\frac{(n-s)!}{n!}\sum_{i=1}^n X_i(\sum_{i=1}^n X_i - 1)...(\sum_{i=1}^n X_i - s + 1)$.

(b):

Now we can already find a function of $T$ to have expectation $p^s$ and we just have to find a function of $T$ with expectation $(1-p)^{n-s}$. Now $Y_i = 1 - X_i \sim B(1,q)$ with $p+q=1$ so if $Q = \sum_{i=1}^n Y_i \sim B(n,q)$

$$\mathbb{E}[Q(Q-1)...(Q-n+s+1)] = \sum_{k=n-s}^{n} k(k-1)...(k-n+s+1)\binom{n}{k}q^k(1-q)^{n-k} \tag{216}$$

$$= \frac{n!}{s!}q^{n-s} \tag{217}$$

so now this is telling us that $\frac{s!}{n!}Q(Q-1)...(Q-n+s+1)$ has expectation $(1-p)^{n-s}$ and since $Q = n-T$, we see that the UMVUE is

$$\frac{(n-s)!}{n!}\sum_{i=1}^n X_i(\sum_{i=1}^n X_i - 1)...(\sum_{i=1}^n X_i - s + 1) + \frac{s!}{n!}(n - \sum_{i=1}^n X_i)(n-1-\sum_{i=1}^n X_i)...(s+1-\sum_{i=1}^n X_i) \tag{218}$$

12

$p(\lambda) = \mathbb{E}U$ where $U = \mathbb{I}_{X_1 \le t_0}$, now $X_1, ..., X_n \sim \Gamma(1,\lambda)$ so $T = \sum_{i=1}^n X_i \sim \Gamma(n,\lambda)$ is complete sufficient statistic, by Rao-Blackwell, $V = \mathbb{E}(U|T)$ is UMVUE so

$$\mathbb{E}(U|T=t) = \mathbb{P}(X_1 \le t_0 | T = t) \tag{219}$$

and we want to see the distribution of $X_1|_T$. Note that

$$F_{T|X_1}(t|x) = \mathbb{P}(T \le t | X_1 = x) \tag{220}$$

$$= \mathbb{P}(X_2 + ... + X_n \le t - x | X_1 = x) \tag{221}$$

$$= \mathbb{P}(X_2 + ... + X_n \le t - x) \tag{222}$$

$$= \frac{\lambda^{n-1}}{\Gamma(n-1)}\int_0^{t-x} u^{n-2}e^{-\lambda u}\,du \tag{223}$$

$$f_{T|X_1}(t|x) = \frac{\lambda^{n-1}}{\Gamma(n-1)}(t-x)^{n-2}e^{-\lambda(t-x)} \tag{224}$$

so

$$f_{(T,X_1)}(t,x) = \frac{\lambda^{n-1}}{\Gamma(n-1)}(t-x)^{n-2}e^{-\lambda(t-x)} \cdot \lambda e^{-\lambda x} \tag{225}$$

$$= \frac{\lambda^n}{\Gamma(n-1)}(t-x)^{n-2}e^{-\lambda t} \tag{226}$$

and the other conditional density is

$$f_{X_1|T}(x|t) = \frac{\frac{\lambda^n}{\Gamma(n-1)}(t-x)^{n-2}e^{-\lambda t}}{\frac{\lambda^n}{\Gamma(n)}t^{n-1}e^{-\lambda t}} \tag{227}$$

$$= (n-1)\frac{1}{t}\left(1 - \frac{x}{t}\right)^{n-2} \tag{228}$$

and now we see

$$V = \int_0^{t_0} (n-1)\frac{1}{t}\left(1-\frac{x}{t}\right)^{n-2} dx \tag{229}$$

$$= 1 - \left(1-\frac{t_0}{t}\right)^{n-1} \tag{230}$$

if $t > t_0$ and 1 otherwise.

17

$X_1, ..., X_n \sim N(\theta, 1), p = \Phi(x-\theta) = \mathbb{E}\mathbb{I}_{X_1 \leq x}$ now $T = \sum_{i=1}^n X_i \sim N(n\theta, n)$ is complete sufficient so $V = \mathbb{P}(X_1 \leq x|T)$ is UMVUE. Now calculate $\mathbb{P}(X_1 \leq x|T=t)$ and it's obvious that $X_1|_{T=t}$ is still Gaussian with $\mathbb{E}X_1|_{T=t} = \frac{t}{n}$ by symmetricity and by complicated calculations $Var(X_1|_{T=t}) = \frac{n-1}{n}$ so $X_1|_{T=t} \sim N(\frac{t}{n}, \frac{n-1}{n})$ and

$$V = \Phi([x-\frac{t}{n}]\sqrt{\frac{n}{n-1}}) = \Phi\left((x-\overline{X})\sqrt{\frac{n}{n-1}}\right) \tag{231}$$

is the UMVUE.