Problem 2.1

a)

Exponential form for one parameter exponential densities:

$$f_X(x|\theta) = \exp(\eta(\theta)T(x) - A(\theta) + B(x))$$

Applying this to the given density:

$$\begin{aligned} f(x;\theta) &= \theta(1+x)^{-(\theta+1)} \\ &= \exp(\ln(\theta) - (\theta+1)\ln(1+x)) \end{aligned}$$

This leads to:

$$\begin{aligned} A(\theta) &= -\ln(\theta) \\ \eta(\theta) &= -(\theta+1) \\ T(x) &= \ln(1+x) \end{aligned}$$

b)

Using Bayes rule

$$\begin{aligned} f(\theta|X) &\propto \mathcal{L}(X|\theta)g(\theta) \\ &= g(\theta)\prod_{i=1}^{n} f(x_i|\theta) \\ &= \exp(-p\theta)\prod_{i=1}^{n}\theta(1+x_i)^{-(\theta+1)} \\ &= \exp(-p\theta)\theta^n\prod_{i=1}^{n}(1+x_i)^{-(\theta+1)} \\ &= \exp(-p\theta)\theta^n\exp\left(-(\theta+1)\sum_{i=1}^{n}\ln(1+x_i)\right) \\ &\propto \theta^n\exp\left(-\theta\left(p+\sum_{i=1}^{n}\ln(1+x_i)\right)\right) \end{aligned}$$

So it's Gamma with the parameters:

$$\theta|X \sim \mathrm{Gamma}\left(n-1, p+\sum_{i=1}^{n}\ln(1+x_i)\right)$$

c)

Well $g(\theta)$ is a kernel of the exponential distribution and would need some proper scaling:

$$f_g(\theta) = p\exp(-p\theta)$$

So no it's no proper prior because:

$$\int_{\mathbb{R}_+} g(\theta)d\theta \neq 1$$

But because we look at just the kernel for the posterior as well, it does not make a difference, because the scaling drops out anyways.

d)

Because we have quadratic loss we already know that the solution to the Bayes rule is the posterior expected value:

$$\begin{aligned}\arg\min_{\delta\in\mathcal{D}} B(\delta, g) &= \mathbb{E}(\theta|X)\\ &= \frac{n-1}{p+\sum_{i=1}^{n}\ln(1+x_i)}\end{aligned}$$

e)

For this we need to compute the likelihood:

$$\begin{aligned}\mathcal{L}(X|\theta) &= \prod_{i=1}^{n} f(x_i|\theta)\\ &= \prod_{i=1}^{n} \theta(1+x_i)^{-(\theta+1)}\\ &= \theta^n \exp\left(-(\theta+1)\sum_{i=1}^{n}\ln(1+x_i)\right)\end{aligned}$$

log scale the likelihood:

$$\begin{aligned}\ln\mathcal{L}(X|\theta) &= n\ln\theta - (1+\theta)\sum_{i=1}^{n}\ln(1+x_i)\\ \implies \text{FOC}: \ln\mathcal{L}'(X|\theta) &= \frac{n}{\theta} - \sum_{i=1}^{n}\ln(1+x_i) = 0\\ \hat{\theta}_{ML} &= \frac{n}{\sum_{i=1}^{n}\ln(1+x_i)}\end{aligned}$$

This is different from the Bayes-rule because we are not utilizing the information from the prior. The higher p the more smaller values are more common and therefore the estimation is biased downwards.

Problem 2.2

a)

$$
\begin{aligned}
R(\delta, \theta) &= Var(\delta) + \mathrm{Bias}(\delta)^2 \\
&= a^2 Var(X) + (\mathbb{E}(aX + b) - \theta)^2 \\
&= a^2 (n\theta(1-\theta)) + (an\theta + b - \theta)^2 \\
&= a^2 n\theta(1-\theta) + ((an-1)\theta + b)^2 \\
&= a^2 n\theta - a^2 n\theta^2 + (an-1)^2\theta^2 - 2(an-1)\theta b + b^2 \\
&= ((an-1)^2 - a^2 n)\theta^2 + (a^2 n - 2b(an-1))\theta + b^2
\end{aligned}
$$

So this leads to the definitions:

$$
\begin{aligned}
\alpha &= (an-1)^2 - a^2 n \\
\beta &= a^2 n - 2b(an-1) \\
\gamma &= b^2
\end{aligned}
$$

b)

So using the hint we get the condition:

$$
R'(\delta_{a,b}, \theta) = 2\alpha\theta + \beta = 0
$$

This means that (because $\theta \neq 0$ is possible):

$$
\begin{aligned}
\alpha &= 0 \\
\beta &= 0
\end{aligned}
$$

So inserting this into the definitions of α and β:

$$
\begin{aligned}
(an-1)^2 - a^2 n &= 0 \\
a^2 n - 2b(an-1) &= 0 \\
\implies an - 1 &= a\sqrt{n} \\
a(n-\sqrt{n}) &= 1 \\
a_M &= \frac{1}{n-\sqrt{n}} \\
\implies a^2 n &= 2b(an-1) \\
\frac{n}{(n-\sqrt{n})^2} &= 2b\left(\frac{\sqrt{n}}{n-\sqrt{n}}\right) \\
\frac{\sqrt{n}}{2(n-\sqrt{n})} &= b_M
\end{aligned}
$$

And this means, that the risk is constant at:

$$
R(\delta_{\alpha_M,\beta_M},\theta) = \frac{n}{4(n-\sqrt{n})^2}
$$

Problem 2.3

a)

To find the cdf we need to find the antiderivative of the pdf:

$$
F(x;\theta) = \begin{cases} 0 & x \leq 0 \\ x^\theta & x \in (0,1) \\ 1 & \text{else} \end{cases}
$$

The cdf for the sample maximum is therefore, following the formula:

$$
F_{X_{[n]}}(x) = F_X(x)^n
$$

And therefore we get:

$$
F_{X_{[n]}}(x;\theta) = \begin{cases} 0 & x \leq 0 \\ x^{n\theta} & x \in (0,1) \\ 1 & \text{else} \end{cases}
$$

b)

$$
\begin{aligned}
\alpha &= P_{H_0}(\vec{x} \in C_n) \\
&= 1 - P(x_{[n]} < 0.90^{3/n}) \\
&= 1 - (0.90^{3/n})^{n/3} \\
&= 1 - 0.9 = 0.1
\end{aligned}
$$

So the size is 10%.

c)

It holds trivially that:

$$X_{[n]} \xrightarrow{d} Y$$

Where:

$$P(Y=1)=1$$

And it holds that:

$$0.9^{3/n} \to 1$$

So asymptotically we get:

$$C_n \to C = \{\vec{x} : x_{[n]} \geq 1\}$$

And the size:

$$\begin{aligned}\alpha &= P_{H_0}(Y \in C) \\ &= 1 - P(Y < 1) = 1 - 0 = 1\end{aligned}$$

But it should converge to 0 so no, it's not consistent

d)

Proposition (Without randomization)

Let $\vec{\mathbf{X}}$ be a (random) sample from $f(\vec{\boldsymbol{x}};\theta)$. Furthermore, let $\{k(\theta_1) > 0\}$ be positive constants depending on $\theta_1 \in \Theta_1$ and C_r a critical region s.t.

(1') $\quad \mathrm{P}(\vec{\mathbf{X}} \in C_r; \theta_0) = \alpha, \qquad 0 < \alpha < 1;$

(2') $\quad \dfrac{f(\vec{\boldsymbol{x}};\theta_1)}{f(\vec{\boldsymbol{x}};\theta_0)} \geq k(\theta_1) \qquad \forall\, \vec{\boldsymbol{x}} \in C_r$ *and* $\forall\, \theta_1 \in \Theta_1;$

(3') $\quad \dfrac{f(\vec{\boldsymbol{x}};\theta_1)}{f(\vec{\boldsymbol{x}};\theta_0)} < k(\theta_1) \qquad \forall\, \vec{\boldsymbol{x}} \notin C_r$ *and* $\forall\, \theta_1 \in \Theta_1.$

Then C_r is the uniformly most powerful critical region of size α for testing the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \in \Theta_1$.

10 / 31