

Exam 1 (A1)

Class: Bayesian Statistics
Instructor: Luiz Max Carvalho
Student: Lucas Machado Moschen

26/04/2021

Turn in date: until 28/04/2021 at 23:59h Brasilia Time.

- Please read through the whole exam before starting to answer;
- State and prove all non-trivial mathematical results necessary to substantiate your arguments;
- Do not forget to add appropriate scholarly references *at the end* of the document;
- Mathematical expressions also receive punctuation;
- You can write your answer to a question as a point-by-point response or in “essay” form, your call;
- Please hand in a single, **typeset** (\LaTeX) PDF file as your final main document. Code appendices are welcome, *in addition* to the main PDF document.
- You may consult any sources, provided you cite **ALL** of your sources (books, papers, blog posts, videos);
- You may use symbolic algebra programs such as Sympy or Wolfram Alpha to help you get through the hairier calculations, provided you cite the tools you have used.
- The exam is worth $\min\{\text{your score}, 100\}$ marks.

Background

This exam covers decision theory, prior elicitation and point estimation. You will need a little measure theory here and there and also be sharp with your knowledge of expectations and conditional expectations.

1. Warm up: stretching our Bayesian muscles with a classic.

[Bayes \(1763\)](#)¹: a billiard ball W is rolled on a line of length $L = 1$, with uniform probability of stopping anywhere on L . It stops at $0 < p < 1$. A second ball, O , is then rolled n times under the same assumptions. Let X denote the number of times (out of n) that O stopped to the left of W . Given $X = x$, **what inference(s) can we make on p ?** You may assume $n \geq 2$.

The idea here is to provide a rigorous, principled Bayesian analysis of this 250 years old problem. Here are a few road signs to help you in your analysis:

- a) (10 marks) For the problem at hand, **carefully** define the parameter space, data-generating mechanism, likelihood function and all of crucial elements of a Bayesian analysis. What is the dominating measure?

Solution. We want to make inference about p , the stopping point of the ball W . We know that $0 < p < 1$, so we define the parameter space $\Theta = (0, 1)$. We are considering X a non-negative integer between 0 and n , that is, $\mathcal{X} = \mathbb{Z}_+ \cap [0, n]$.

Each time we roll the ball O , there is a uniform probability of stopping between 0 and $L = 1$. Then,

$$\Pr(\text{position of } O < \text{position of } W) = \Pr(\text{position of } O < p) = p$$

and each trial has a probability p of success and $1 - p$ of fail. Moreover, we do n experiments conditionally independents, what implies that the data-generating mechanism is a Binomial distribution with parameters n and p , that is, $X \sim \text{Bin}(n, p)$.

The density of X is given by the expression

$$f(x|n, p) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

We'll drop the conditionality of n and suppose it's fixed. Then, the likelihood is

$$L(p|x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

with respect to the counting measure. We defined the essential elements for a simple statistical inference. Now we must define the crucial elements of a Bayesian analysis:

- (1) Prior distribution: $\pi(p) = 1$ (over the space Θ). This is because we have prior knowledge that the position of ball W has a uniform probability between 0 and $L = 1$.
- (2) Loss function: The problem didn't express it, however it's important to evaluate the quality of our inference. We can consider the absolute loss, because we are measuring the distance between points in a line.

The dominating measure in Θ is the Lebesgue measure and the dominating measure in \mathcal{X} (which defines the likelihood function [Schervish \(1996\)](#)) is the counting measure. \square

- b) (10 marks) Exhibit your posterior measure and sketch its density.

Hint: it might be convenient to prove the following proposition:

Proposition 1. *The joint distribution of p and X is given by*

$$\Pr(a < p < b, X = x) = \int_a^b \binom{n}{x} t^x (1 - t)^{n-x} dt.$$

¹The paper was published posthumously and read to the Royal Society by Bayes's close friend Richard Price (1723-1791).

Proof. By the definition of conditional density, if $f_{p,X}(p, x)$ is the joint density of p and X ,

$$f_{p,X}(p, x) = f_{X|p}(x|p) \cdot \pi(p) = f(x|p).$$

Then

$$\begin{aligned} \Pr(a < p < b, X = x) &= \Pr(p < b, X = x) - \Pr(p \geq a, X = x) \\ &= \Pr(p < b, X \leq x) - \Pr(p < b, \leq x - 1) \\ &\quad - \Pr(p \geq a, X \leq x) + \Pr(p \geq a, X \geq x - 1) \\ &= \int_0^b \sum_{i=0}^x f(i|t) dt - \int_0^b \sum_{i=0}^{x-1} f(i|t) dt \\ &\quad - \int_0^a \sum_{i=0}^x f(i|t) dt + \int_0^a \sum_{i=0}^{x-1} f(i|t) dt \\ &= \int_0^b f(x|t) dt - \int_0^a f(x|t) dt \\ &= \int_a^b \binom{n}{x} t^x (1-t)^{n-x} dt \end{aligned}$$

as we wanted to prove. \square

Solution. The conditional distribution function $F_{p|X}(b|x)$ is defined by the expression

$$F_{p|X}(b|x) = \frac{\Pr(p \leq b, X = x)}{\Pr(X = x)}, \text{ whenever } \Pr(X = x) > 0.$$

By Proposition 1, we can calculate in the following manner:

$$F_{p|X}(b|x) = \frac{\int_0^b \binom{n}{x} t^x (1-t)^{n-x} dt}{\int_0^1 \binom{n}{x} t^x (1-t)^{n-x} dt} = \frac{\int_0^b t^x (1-t)^{n-x} dt}{\int_0^1 t^x (1-t)^{n-x} dt}$$

given that $\Pr(X = x) = \Pr(p \in [0, 1], X = x)$. The posterior measure $\mu_{p|X}$ can be defined through $F_{p|X}(b|x)$. The expressions in the integrals do not have closed form and the ratio can be expressed as a ratio of incomplete beta functions [Wikipedia \(2021b\)](#):

$$F_{p|X}(b|x) = \frac{B(b; x+1, n-x+1)}{B(1; x+1, n-x+1)}.$$

In special, $B(1; x+1, n-x+1) = B(x+1, n-x+1)$ is the beta function.

The density of the posterior measure can be calculated using the Fundamental Theorem of Calculus [Wikipedia \(2021d\)](#)

$$\xi(p|x) = \frac{d}{dp} F_{p|X}(p|x) = \frac{1}{B(x+1, n-x+1)} p^x (1-p)^{n-x}. \quad (1)$$

\square

c) (10 marks) Provide a Bayes estimator under (i) quadratic and (ii) 0-1 loss²;

Proposition 2. *The Bayes estimator δ^π associated with the quadratic loss is the posterior expectation*

$$\delta^\pi(x) = \mathbb{E}^\xi[p]$$

Proof. The quadratic loss is $L(p, \delta) = (p - \delta)^2$. Then,

$$\varrho(\pi, \delta|x) = \mathbb{E}^\xi[(p - \delta)^2] = \mathbb{E}^\xi[p^2] - 2\delta \mathbb{E}^\xi[p] + \delta^2$$

and $\delta^\pi(x) = \min_d \varrho(\pi, d|x)$. Since the quadratic function is convex and

$$\frac{d}{d\delta} \varrho(\pi, \mathbb{E}^\xi[p]) = 2 \mathbb{E}^\xi[p] - 2 \mathbb{E}^\xi[p] = 0,$$

we conclude $\delta^\pi(x) = \mathbb{E}^\xi[p]$ \square

²Recall the discussion in class about how to properly define 0-1 loss as for continuous parameter spaces.

Solution. We divide the solution in quadratic and 0-1 loss.

- (i) Consider the quadratic loss. As proved in Proposition 2, the Bayes estimator δ^π is the posterior mean. By equation 1, we know that the distribution of p given $X = x$ is Beta with parameters $x + 1$ and $n - x + 1$. Then, the posterior mean is

$$\delta^\pi(x) = \int_0^1 p \cdot \xi(p|x) dp = \frac{1}{B(x+1, n-x+1)} \int_0^1 p^{x+1} (1-p)^{n-x} dp = \frac{B(x+2, n-x+1)}{B(x+1, n-x+1)}.$$

By the definition of Beta function,

$$\delta^\pi(x) = \frac{\frac{(x+1)!(n-x)!}{(n+2)!}}{\frac{x!(n-x)!}{(n+1)!}} = \frac{(x+1)!(n-x)!(n+1)!}{x!(n-x)!(n+2)!} = \frac{x+1}{n+2}.$$

- (ii) First we must define a sequence of losses indexed by ϵ ,

$$L_\epsilon(p, d) = 1 - \mathbb{1}_{|p-d| \leq \epsilon}$$

and consider $\epsilon \rightarrow 0$. So

$$\mathbb{E}^\xi[L_\epsilon(p, d)] = 1 - \int_{d-\epsilon}^{d+\epsilon} \xi(p|x) dp.$$

The Bayes estimator is the argument of

$$\max_d \int_{d-\epsilon}^{d+\epsilon} \xi(p|x) dp = \max_d \int_{d-\epsilon}^{d+\epsilon} p^x (1-p)^{n-x} dp.$$

We extend the definition of $\xi(p|x)$ to the real line such that $\xi(p|x) = 0, \forall p \notin [0, 1]$. Fix x and let

$$G(d) = \int_{d-\epsilon}^{d+\epsilon} \xi(p|x) dp.$$

Function G is defined in the decision space $[0, 1]$ and it is continuous. By Weierstrass' theorem Wikipedia (2021c), G has a global maximum. It can not be on the limits, because $G(0) < G(\epsilon)$ and $G(1) < G(1 - \epsilon)$ for every $\epsilon > 0$. If $\bar{d}_\epsilon \in (0, 1)$ is global maximum,

$$G'(\bar{d}_\epsilon) = \xi(\bar{d}_\epsilon + \epsilon|x) - \xi(\bar{d}_\epsilon - \epsilon|x) = 0.$$

Then $\xi(\bar{d}_\epsilon + \epsilon) = \xi(\bar{d}_\epsilon - \epsilon) \implies \log(\xi(\bar{d}_\epsilon + \epsilon)) = \log(\xi(\bar{d}_\epsilon - \epsilon))$. By Rolle's theorem Wikipedia (2021e), there exists $d_\epsilon^* \in (\bar{d}_\epsilon - \epsilon, \bar{d}_\epsilon + \epsilon)$ such that

$$0 = \frac{d}{d\delta} \log \xi(d_\epsilon^*|x) = \frac{x}{d_\epsilon^*} - \frac{n-x}{1-d_\epsilon^*} \implies d_\epsilon^* = \frac{x}{n}.$$

Observe that we can drop the dependence on ϵ and for every $\epsilon > 0$, if \bar{d}_ϵ is global maximum,

$$d^* = \frac{x}{n} \in (\bar{d}_\epsilon - \epsilon, \bar{d}_\epsilon + \epsilon).$$

Therefore, if \bar{d}_ϵ forms a sequence of Bayes estimates associated with the sequence of losses $L_\epsilon(p, d)$, we conclude that \bar{d}_ϵ converges to d^* for each x when $\epsilon \rightarrow 0$.

In summary the Bayes estimator for 0-1 loss is $\delta^\pi(x) = \frac{x}{n}$.

□

- d) (10 marks) Contrast the estimators obtained in the previous item with the maximum likelihood estimator, in terms of (i) posterior expected risk and (ii) integrated risk. Is any of the estimators preferable according to both risks? If not, which estimator should be preferred under each goal risk?

Solution. The maximum likelihood estimator is the argument of

$$\max_p p^x (1-p)^{n-x} = \max_p x \log p + (n-x) \log(1-p).$$

We proved in the last item that there is a maximum and it's attained at

$$\delta_{MLE}(x) = \frac{x}{n}.$$

Denote δ_2 for the Bayes estimator associated to the quadratic loss and δ_1 to the 0-1 loss. We note that $\delta_1 = \delta_{MLE}$, because the prior distribution is uniform.

(i) Here we calculate the posterior expected risk associated with δ_1 and δ_2 .

$$\varrho(\pi, \delta_1|x) = \lim_{\epsilon \rightarrow 0} \mathbb{E}^\xi[L_\epsilon(p, \delta_1)] = 1,$$

$$\text{because } 0 \leq \int_{\delta_1 - \epsilon}^{\delta_1 + \epsilon} \xi(p|x) dp \leq 2\epsilon \cdot \xi(\delta_1|x) \xrightarrow{\epsilon \rightarrow 0} 0.$$

$$\varrho(\pi, \delta_2|x) = \mathbb{E}^\xi[(p - \delta_1)^2] = \mathbb{E}^\xi[p^2] - 2\delta_1 \mathbb{E}^\xi[p] + \delta_1^2 = \mathbb{E}^\xi[p^2] - \mathbb{E}^\xi[p]^2 = \text{Var}^\xi[p].$$

We know that $p|x \sim \text{Beta}(x+1, n-x+1)$. Then, by [Wikipedia \(2021a\)](#),

$$\text{Var}^\xi[p] = \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)} \leq \frac{1}{4} \frac{(n+2)^2}{(n+2)^2(n+3)} = \frac{1}{4(n+3)}.$$

(ii) We calculate the integrated risk

$$r(\pi, \delta_1) = \sum_{x=0}^n \varrho(\pi, \delta_1|x) = \sum_{x=0}^n 1 = n+1$$

and

$$r(\pi, \delta_2) = \sum_{x=0}^n \varrho(\pi, \delta_2|x) = \sum_{x=0}^n \frac{(x+1)(n-x+1)}{(n+2)^2(n+3)} \leq \frac{n+1}{4(n+3)} < \frac{1}{4}.$$

We finish saying that, in terms of (i) and (ii), δ_2 has less risk, then, it would be preferable. However, we shall remember that the 0-1 loss weights too much if you are far from p . When $\epsilon \rightarrow 0$, we are saying all points with distance greater than ϵ to x/n should have error 1, therefore, it was expected δ_2 to have less risk.

The last remark is that MLE is equal to MAP in this case, so the risks are equal. If you are worried with invariance, bias and frequentist properties, MAP is the preferred option, then. \square

- e) (10 marks) Suppose one observes $X = x = 6$ for $n = 9$ rolls. Produce an updated prediction of where the next m balls are going to stop along L . Please provide (i) a point prediction in the form of an expected value and (ii) a full probability distribution.

Solution. Let $Z \sim \text{Bin}(m, p)$. The predictive density of Z is

$$\begin{aligned} g^\pi(z|x) &= \frac{1}{B(x+1, n-x+1)} \int_0^1 \binom{m}{z} p^z (1-p)^{m-z} p^x (1-p)^{n-x} dp \\ &= \frac{1}{B(x+1, n-x+1)} \binom{m}{z} \int_0^1 p^{z+x} (1-p)^{m+n-(z+x)} dp \\ &= \frac{B(z+x+1, m+n-(z+x)+1)}{B(x+1, n-x+1)} \binom{m}{z} \\ &= \frac{B(z+7, m-z+4)}{B(7, 4)} \binom{m}{z}. \end{aligned}$$

Now, we want to make a point prediction through the expected value and we call it $\delta^{pred}(x)$. Hence,

$$\begin{aligned}
\delta^{pred}(6) &= \sum_{z=0}^m z \cdot g^\pi(z|x) \\
&= \sum_{z=1}^m \frac{B(z+7, m-z+4)}{B(7,4)} \frac{m!}{(z-1)!(m-z)!} \\
&= m \sum_{z=1}^m \frac{B((z-1)+8, (m-1)-(z-1)+4)}{B(7,4)} \frac{(m-1)!}{(z-1)!(m-1-(z-1))!} \\
&= \frac{m}{B(7,4)} \sum_{z=0}^{m-1} B(z+8, (m-1)-z+4) \frac{(m-1)!}{z!(m-1-z)!} \\
&= m \frac{B(8,4)}{B(7,4)} = \frac{7}{11}m,
\end{aligned}$$

because $\frac{B(8,4)}{B(7,4)} = \frac{7!3!10!}{6!3!11!} = \frac{7}{11}$. Therefore, we expect 7/11 of the m balls to fall before p and 4/11 after p . \square

2. Proper behavior.

In this question we will explore propriety and its implications for inference. Consider a parameter space Θ and a sampling model with density $f(x | \theta)$. A density $h : \Theta \rightarrow (0, \infty)$ is said to be proper if

$$\int_{\Theta} h(t) d\mu(t) < \infty,$$

where μ is the dominating measure. Assuming μ to be the Lebesgue measure and that the prior $\pi(\theta)$ is proper, show that

a) (10 marks) The posterior density,

$$\xi(\theta | x) = \frac{f(x | \theta)\pi(\theta)}{m(x)},$$

is proper almost surely.

Proof. We want to prove that $\xi(\theta|x)$ is well defined almost surely, and after

$$\int_{\Theta} \xi(\theta|x) d\mu(t) = \int_{\Theta} \xi(\theta|x) dt < +\infty$$

First, let's prove that

$$m(x) = \int_{\Theta} f(x|\theta)\pi(\theta) d\theta \neq 0 \text{ or } m(x) < +\infty$$

almost surely, that is, $\xi(\theta|x)$ is well defined.

Let $A = \{x \in \mathcal{X} : m(x) = 0\}$ and $B = \{x \in \mathcal{X} : m(x) = +\infty\}$. Then, because the dominating measure is the Lebesgue measure.

$$\Pr(X \in A) = \int_A m(x) dx = 0$$

and

$$\Pr(X \in B) = \int_B m(x) dx = \int_B +\infty dx$$

If $\int_B dx > 0$, we have $\Pr(X \in B) = +\infty$, what is an absurd. So $\int_B dx = 0$ and $\Pr(X \in B) = 0$. We conclude that $\xi(\theta|x)$ is well defined almost surely.

Given the definition of $m(x)$, we know that

$$\int_{\Theta} \xi(\theta|x) dt = 1,$$

what proves our statement \square

- b) (10 marks) The Bayes estimator under quadratic loss, $\delta_S(x) = E_\xi[\theta]$ is biased almost surely. *Hint:* consider what happens to the integrated risk under unbiasedness.

Proof. An estimator δ of θ is unbiased if $E_\theta[\delta(x)] = \theta$, for all $\theta \in \Theta$ (Schervish (1996)). Following the hint, consider the integrated risk

$$r(\pi, \delta_S) = \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta_S(x)) f(x|\theta) dx \pi(\theta) d\theta$$

Applying to the quadratic loss,

$$\begin{aligned} r(\pi, \delta_S) &= \int_{\Theta} \int_{\mathcal{X}} (\delta_S(x) - \theta)^2 f(x|\theta) \pi(\theta) dx d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} (\delta_S(x)^2 - 2\delta_S(x)\theta + \theta^2) f(x|\theta) \pi(\theta) dx d\theta \\ &= \int_{\mathcal{X}} \delta_S(x)^2 m(x) dx + \int_{\Theta} \theta^2 \pi(\theta) d\theta \\ &\quad - \int_{\Theta} \theta \pi(\theta) \int_{\mathcal{X}} \delta_S(x) f(x|\theta) dx d\theta - \int_{\mathcal{X}} \delta_S(x) \int_{\Theta} \theta f(x|\theta) \pi(\theta) d\theta dx. \end{aligned}$$

Observe that

$$\int_{\mathcal{X}} \delta_S(x) f(x|\theta) dx = E_\theta[\delta_S(x)].$$

Suppose $\delta_S(x)$ is unbiased. Then $E_\theta[\delta_S(x)] = \theta$ and

$$\int_{\mathcal{X}} \delta_S(x) f(x|\theta) dx = \theta$$

by definition. The fourth integral is

$$\int_{\Theta} \theta f(x|\theta) \pi(\theta) d\theta = \int_{\Theta} \theta \xi(\theta|x) m(x) = E_\xi[\theta] m(x) = \delta_S(x) m(x).$$

We conclude that $r(\pi, \delta_S) = 0$. Because the integrand is non-negative, we conclude that

$$(\theta - \delta_S(x))^2 f(x|\theta) \pi(\theta) = 0$$

almost surely over $\Theta \times \mathcal{X}$, what is an absurd.

□

Now,

- c) (10 marks) Take

$$f(x|\theta) = \frac{2x}{\theta}, x \in [0, \sqrt{\theta}],$$

and

$$\pi(\theta) = \frac{2}{\pi(1+\theta^2)}, \theta > 0.$$

The posterior density is then

$$p(\theta|x) = \frac{1}{m(x)} \frac{4x}{\pi} \frac{1}{\theta(1+\theta^2)},$$

and thus

$$m(x) = \int_0^\infty \frac{4x}{\pi} \frac{1}{t(1+t^2)} dt = \frac{4x}{\pi} \int_0^\infty \frac{1}{t(1+t^2)} dt = \infty,$$

i.e., the posterior is apparently improper. Explain how this “counter-example” is wrong.

Proof. This “counter-example” is wrong because the marginal distribution of x is wrongly defined. Actually, it should be

$$m(x) = \int_0^\infty \frac{4x}{\pi} \frac{1}{t(1+t^2)} \mathbb{1}\{0 \leq x \leq \sqrt{t}\} dt = \int_{x^2}^\infty \frac{4x}{\pi} \frac{1}{t(1+t^2)} dt$$

Then

$$\begin{aligned} m(x) &= \int_{x^2}^\infty \frac{4x}{\pi} \frac{1}{t(1+t^2)} dt \\ &= \frac{4x}{\pi} \int_{x^2}^\infty \frac{1}{t} - \frac{t}{1+t^2} dt \\ &= \frac{2x}{\pi} [2 \log(t) - \log(1+t^2)]_{x^2}^\infty \\ &= \frac{4x}{\pi} \left[\log \left(\frac{t^2}{1+t^2} \right) \right]_{x^2}^\infty \\ &= \frac{4x}{\pi} \log(1+x^{-2}) < +\infty \iff x > 0 \end{aligned}$$

Inasmuch $X = 0$ has null measure (as we expected), the posterior is proper almost surely. \square

3. Rayleigh dispersion³.

Let X_1, \dots, X_n be a random sample from a distribution whose probability density function is

$$f(x | \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \sigma > 0, 0 < x < \infty.$$

This is the Rayleigh⁴ distribution, used in reliability modelling, where the X are failure times (time-until-failure).

- a) (5 marks) For t fixed and known, find $\eta = \Pr(X > t)$ as function of σ^2 ;

Solution. Define $y = x^2$ and $x dx = dy/2$

$$\begin{aligned} \eta = \Pr(X > t) &= \int_t^\infty \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= \int_{t^2}^\infty \frac{1}{2\sigma^2} \exp\left(-\frac{y}{2\sigma^2}\right) dy \\ &= -\exp\left(-\frac{y}{2\sigma^2}\right)_{t^2}^\infty \\ &= \exp\left(-\frac{t^2}{2\sigma^2}\right) \end{aligned}$$

\square

- b) (5 marks) Find the likelihood function for η for a sample $\mathbf{x} = \{x_1, \dots, x_n\}$;

Solution. First, we reparametrize the density function to the parameter η :

$$\eta = \exp\left(-\frac{t^2}{2\sigma^2}\right) \implies \log(1/\eta) = \frac{t^2}{2\sigma^2} \implies \sigma^2 = \frac{t^2}{2\log(1/\eta)}$$

and

$$\exp\left(-\frac{x^2}{2\sigma^2}\right) = \exp\left(-\frac{t^2}{2\sigma^2} \frac{x^2}{t^2}\right) = \eta^{x^2/t^2}.$$

³This question's title is pun: https://en.wikipedia.org/wiki/Rayleigh_scattering.

⁴John William Strutt (1842-1919), the Third Baron Rayleigh, was a British physicist, known for his work on Statistical Mechanics.

Then, the reparametrized density is

$$f(x|\eta) = \frac{2}{t^2} \log(1/\eta) x \eta^{x^2/t^2}$$

and the likelihood function is

$$L(\eta|\mathbf{x}) = \prod_{i=1}^n \frac{2}{t^2} \log(1/\eta) x_i \eta^{x_i^2/t^2} = \frac{2^n}{t^{2n}} \log(1/\eta)^n \left(\prod_{i=1}^n x_i \right) \eta^{t^{-2} \sum_{i=1}^n x_i^2}$$

□

c) (10 marks) Is the Rayleigh distribution a member of the exponential family? Justify your answer;

Proof. Let $\Theta = \mathbb{R}_{++}$ be the parameter space (in σ) and $\mathcal{X} = \mathbb{R}_{++}$ be the space state. Define

$$C(\sigma) = \frac{1}{\sigma^2} \geq 0, \quad h(x) = x \geq 0,$$

$$R(\sigma) = -\frac{1}{2\sigma^2} \in \mathbb{R} \quad \text{and} \quad T(x) = x^2 \in \mathbb{R}$$

By definition, $f(x|\sigma) = C(\sigma)h(x) \exp\{R(\sigma)T(x)\} = \frac{x}{\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$ is from the exponential family of dimension 1, that is, the Rayleigh distribution is a member of the exponential family. □

d) (10 marks) Find the Jeffreys' prior for η , $\pi_J(\eta)$;

Solution. We use $\log(1/\eta) = -\log(\eta)$. Let's divide the calculation in the following steps:

(1) The log-likelihood is

$$l(\eta|x) = \log L(\eta|x) = \log(2/t^2) + \log(x) + \log(\log(1/\eta)) + \frac{x^2}{t^2} \log(\eta);$$

(2) The first derivative of l is

$$\frac{d}{d\eta} l(\eta|x) = \frac{\frac{d}{d\eta} \log(1/\eta)}{\log(1/\eta)} + \frac{x^2}{t^2} \frac{1}{\eta} = -\frac{1}{\eta \log(1/\eta)} + \frac{x^2}{t^2} \frac{1}{\eta} = \frac{1}{\eta \log(\eta)} + \frac{x^2}{t^2} \frac{1}{\eta};$$

(3) The second derivative of l is

$$\frac{d^2}{d\eta^2} l(\eta|x) = -\frac{1 + \log(\eta)}{(\eta \log(\eta))^2} - \frac{x^2}{t^2} \frac{1}{\eta^2};$$

(4) The second moment of Rayleigh distribution is

$$\mathbb{E}_\eta[X^2] = \frac{2}{t^2} \log(1/\eta) \int_0^\infty x^3 \eta^{x^2/t^2} dx = \frac{1}{t^2} \log(1/\eta) \int_0^\infty y \eta^{y/t^2} dy$$

using that $x^2 = y$ and $2x dx = dy$. Integrating by parts the last integral, we obtain

$$\int_0^\infty y \eta^{y/t^2} dy = y \frac{t^2}{\log(\eta)} \eta^{y/t^2} \Big|_0^\infty - \frac{t^2}{\log(\eta)} \int_0^\infty \eta^{y/t^2} dy = -\frac{t^4}{\log(\eta)^2} \eta^{y/t^2} \Big|_0^\infty = \frac{t^4}{\log(\eta)^2}$$

and

$$\mathbb{E}_\eta[X^2] = \frac{\log(1/\eta)}{t^2} \frac{t^4}{\log(\eta)^2} = -\frac{t^2}{\log(\eta)}$$

(5) The Fisher Information is $I(\eta) =$

$$-\mathbb{E}_\eta \left[\frac{d^2}{d\eta^2} l(\eta|x) \right] = \frac{1 + \log(\eta)}{(\eta \log(\eta))^2} + \frac{1}{t^2 \eta^2} \mathbb{E}_\eta[X^2] = \frac{1 + \log(\eta)}{\eta^2 \log(\eta)^2} - \frac{1}{\log(\eta) \eta^2} = \frac{1}{\eta^2 \log(\eta)^2}$$

(6) The Jeffreys' prior is, then,

$$\pi_J(\eta) \propto \frac{1}{\eta |\log(\eta)|} = -\frac{1}{\eta \log(\eta)} = \frac{1}{\eta \log(1/\eta)}$$

a improper prior in $(0, 1)$. We emphasize that $\log(\eta) \leq 0$, because $\eta \leq 1$.

□

e) (10 marks) Using $\pi_J(\eta)$, derive the predictive distribution $p(\tilde{x} | \mathbf{x})$.

Solution. Let $S^2 = \sum_{i=1}^n x_i^2$ and we derive the posterior distribution

$$\xi(\eta | \mathbf{x}) \propto \log(1/\eta)^n \eta^{t^{-2}S^2} \cdot \frac{1}{\eta \log(1/\eta)} = \log(1/\eta)^{n-1} \eta^{t^{-2}S^2-1}.$$

Denote $k = t^{-2}S^2$ and

$$R(n) = \int_0^1 \log(1/\eta)^{n-1} \eta^{k-1} d\eta.$$

The marginal distribution of \mathbf{x} is

$$m(\mathbf{x}) = R(n) = \log(1/\eta)^{n-1} \frac{\eta^k}{k} \Big|_0^1 + \int_0^1 (n-1) \frac{\eta^k}{k} \frac{\log(1/\eta)^{n-2}}{\eta} d\eta = \frac{n-1}{k} R(n-1)$$

with $R(1) = \frac{1}{k}$. With this recurrence equation, we obtain

$$m(\mathbf{x}) = R(n) = \frac{(n-1)!}{k^{n-1}} R(1) = k^{-n} (n-1)!.$$

We conclude that

$$\xi(\eta | \mathbf{x}) = \frac{(t^{-2}S^2)^n}{(n-1)!} \log(1/\eta)^{n-1} \eta^{t^{-2}S^2-1}.$$

Now we are ready to calculate the predictive distribution

$$\begin{aligned} p(\bar{x} | \mathbf{x}) &= \int_0^1 \frac{2}{t^2} \bar{x} \log(1/\eta) \eta^{\bar{x}^2/t^2} \frac{(t^{-2}S^2)^n}{(n-1)!} \log(1/\eta)^{n-1} \eta^{t^{-2}S^2-1} d\eta \\ &= \frac{2(t^{-2}S^2)^n}{t^2(n-1)!} \bar{x} \int_0^1 \log(1/\eta)^n \eta^{t^{-2}(S^2+\bar{x}^2)-1} d\eta \\ &= \frac{2(t^{-2}S^2)^n}{t^2(n-1)!} \bar{x} (t^{-2}(S^2+\bar{x}^2))^{-n-1} n!, \text{ [by our calculation of } R(n)\text{]} \\ &= 2n \frac{S^{2n}}{(S^2+\bar{x}^2)^{n+1}} \bar{x} \end{aligned}$$

□

Bibliography

- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, 53:370–418.
- Schervish, M. (1996). *Theory of Statistics*. Springer Series in Statistics. Springer New York.
- Wikipedia, t. f. e. (2021a). Beta distribution. https://en.wikipedia.org/wiki/Beta_distribution.
- Wikipedia, t. f. e. (2021b). Beta function. https://en.wikipedia.org/wiki/Beta_function.
- Wikipedia, t. f. e. (2021c). Extreme value theorem. https://en.wikipedia.org/wiki/Extreme_value_theorem.
- Wikipedia, t. f. e. (2021d). Fundamental theorem of calculus. https://en.wikipedia.org/wiki/Fundamental_theorem_of_calculus.
- Wikipedia, t. f. e. (2021e). Rolle's theorem. https://en.wikipedia.org/wiki/Rolle%27s_theorem.