Exam 1 (A1)

Class: Bayesian Statistics
Instructor: Luiz Max Carvalho

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?

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*

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

c) (10 marks) Provide a Bayes estimator under (i) quadratic and (ii) 0-1 loss[2];

d) (10 marks) Contrast the estimators obtained in the previous item with the maximum likelihood estimator, in terms of (i) posterior expected loss 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?

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.

Concepts: prior elicitation; point estimation; Bayes risk; integrated risk; posterior predictive.
Level: Easy. Cumbersome, but easy.
Solution: Discussion here will follow mostly from page 11 in Robert (2007). Let us first define the parameter space. Clearly, $p \in \Theta \subseteq [0, 1]$ is our parameter of interest. The experiment with W tells us that it is safe to assume a uniform probability distribution on Θ such that

$$\pi(p) \equiv \text{Beta}(1, 1) = 1 \cdot \mathbb{I}_{[0,1]}(p),$$

where $\mathbb{I}_A(x) = 1$ if $x \in A$ and $\mathbb{I}_A(x) = 0$ otherwise. The dominating measure here is thus the Lebesgue measure on $[0, 1]$. Next, the data-generating process: each roll of the billiard ball produces a Bernoulli random variable Y_i, $i = 1, \ldots, n$ with parameter p: if the ball stops to left of W, $Y_i = 1$, if it stops to the right, $Y_i = 0$. It is well known that $X := \sum_{i=1}^n Y_i \sim \text{Binomial}(n, p)$, i.e.,

$$\Pr(X = x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x},$$

meaning we are justified in writing

$$l(\theta) = p^x (1-p)^{n-x}$$

---

[1] The paper was published posthumously and read to the Royal Society by Bayes's close friend Richard Price (1723-1791).
[2] Recall the discussion in class about how to properly define 0-1 loss as for continuous parameter spaces.

as our likelihood function. Now, let us prove Proposition 1 above, following Edwards (1978) (Proposition 8 therein):

$$f(p, X = x, n) = \Pr(X = x \mid n, p)\pi(p),$$
$$= p^x(1-p)^{n-x}\mathbb{I}_{[0,1]}(t),$$

thus

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

From this, we surmise that

$$\Pr(X = x) = \int_0^1 \binom{n}{x} t^x(1-t)^{n-x}\, dt,$$

and hence we are ready to exhibit the posterior measure:

$$\Pr(p \in (a, b) \mid X = x, n) = \frac{\Pr(p \in (a, b), X = x)}{\Pr(X = x)},$$
$$= \frac{\int_a^b \binom{n}{x} t^x(1-t)^{n-x}\, dt}{\int_0^1 \binom{n}{x} t^x(1-t)^{n-x}\, dt},$$
$$= \frac{\int_a^b \binom{n}{x} t^x(1-t)^{n-x}\, dt}{B(x+1, n-x+1)}, \tag{1}$$

that is $\xi(p \mid n, x) \equiv \text{Beta}(\alpha', \beta')$, with $\alpha' := x + 1$ and $\beta' := n - x + 1$. A sketch of both the prior and the posterior can be found in Figure 1. The expression in (1) immediately informs us that the Bayes
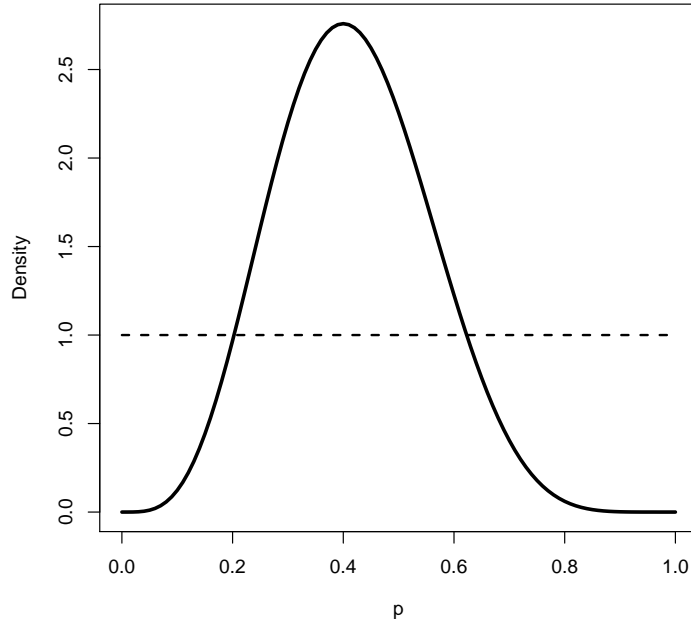

Figure 1: **Probability density function of the posterior distribution for p given $x = 4$ and $n = 10$.** The solid line shows $\xi(p \mid n, x)$ and the dashed line sketches $\pi(p)$.

estimator under quadratic loss is the posterior mean,

$$\delta_{\text{S}}(x) = \frac{\alpha'}{\alpha' + \beta'} = \frac{x + 1}{n + 2}$$

3

and the estimator under 0-1 loss is the maximum *a posteriori* (MAP):

$$\delta_{\mathrm{MAP}}(x) = \frac{\alpha' - 1}{\alpha' + \beta' - 2} = \frac{x}{n},$$

i.e. the posterior mode. The maximum likelihood estimator (MLE) is $\delta_{\mathrm{MLE}}(x) = x/n$. Observe that the MAP estimator is thus identical to the MLE and hence achieves the same risk. It remains to compute the risks for the MLE/MAP and posterior mean. In the remainder, it will be convenient to know that if $Z \sim \mathrm{Beta}(a, b)$, the raw moments of a Beta random variable are given by

$$E[Z^r] = \prod_{k=0}^{r-1} \frac{a + r}{a + b + r}. \tag{2}$$

Now, on to the risks. First, recall that the **posterior expected loss** is:

$$\rho(\delta, h(\theta)) = E_\xi \left[(\delta - h(\theta))^2 \right],$$
$$= \delta^2 - 2\delta E_\xi \left[h(\theta) \right] + E_\xi \left[h(\theta)^2 \right],$$

where $\xi(\theta \mid x)$ is our posterior density. We will draw on Example 4.1.5 in Robert (2007) (pg. 171) to save some computations. Specifically,

$$\rho(\delta_{\mathrm{MAP}}, p) = \left(\frac{x - n/2}{n(n+1)} \right)^2 + \frac{(x + 1/2)(n - x + 1/2)}{(n+1)^2(n+2)}.$$

Now, for the posterior mean, we have

$$\rho(\delta_{\mathrm{S}}, p) = \delta_{\mathrm{S}}^2 - 2\delta_{\mathrm{S}} E_\xi \left[p \right] + E_\xi \left[p^2 \right],$$
$$= \left(\frac{\alpha'}{\alpha' + \beta'} \right)^2 - 2 \left(\frac{\alpha'}{\alpha' + \beta'} \right)^2 + \frac{\alpha'}{\alpha' + \beta'} \left(\frac{\alpha' + 1}{\alpha' + \beta' + 1} \right),$$
$$= \frac{x+1}{n+2} \left(\frac{x+2}{n+3} - \frac{x+1}{n+2} \right).$$

A sketch is provided in Figure 2 and, unsuprinsingly, the posterior mean always provides lower conditional risk. Since δ_{S} minimises $\rho(\cdot, p)$, this is exactly what we would expect. Hence, if you are only interested in making **conditional** risk statements, you should always prefer the posterior mean as an estimator. Now, let us compute the integrated risks for both estimators. Recall the definition of **frequentist risk**:

$$R_{\mathrm{F}}(\delta, h) := E_\theta \left[(\delta - h(\theta))^2 \right].$$

The *integrated risk* is then

$$r(\pi, \delta) := E_\pi \left[R_{\mathrm{F}}(\delta, h) \mid x \right].$$

We will again lean on the derivations in Example 4.1.5 of Robert (2007) for the MLE/MAP:

$$R_{\mathrm{F}}(\delta_{\mathrm{MAP}}, p) = \frac{p(1-p)}{n}$$

Hence,

$$r(\pi, \delta_{\mathrm{MAP}}) = \frac{1}{n} \left\{ E_\pi[p] - E_\pi[p^2] \right\}.$$

Using (2), we have $E_\pi[p] = 1/2$, $E_\pi[p^2] = 1/3$ and $E_\pi[p^3] = 1/4$ and thus

$$r(\pi, \delta_{\mathrm{MAP}}) = \frac{1}{n} \left\{ \frac{1}{2} - \frac{1}{3} \right\},$$
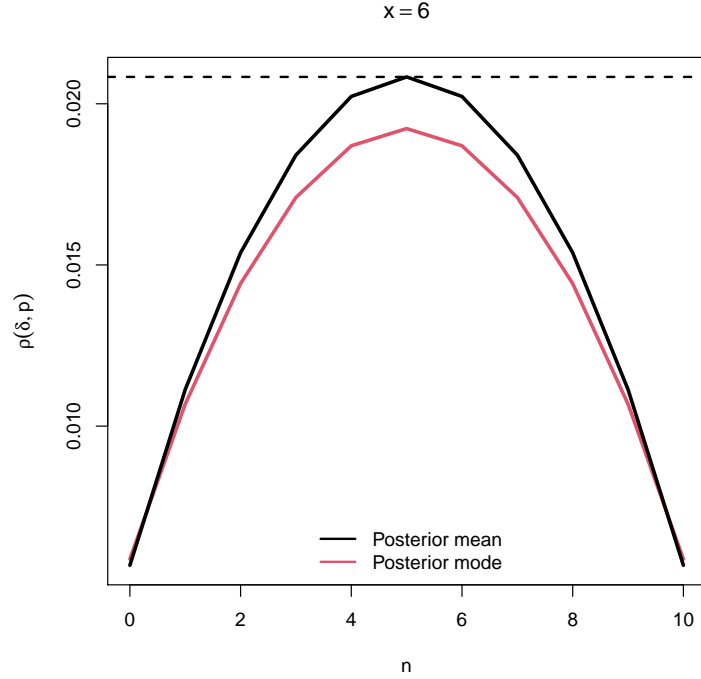$$= \frac{1}{6n}.$$

Figure 2: **Bayes risk for the posterior mean and the posterior mode**. The black line shows $\rho(\delta_{\mathrm{S}}, p)$ whilst the red one depicts $\rho(\delta_{\mathrm{MAP}}, p)$.

Now,

$$R_{\mathrm{F}}(\delta_{\mathrm{S}}(x), p) = E\left[\left(\delta_{\mathrm{S}}(x) - p\right)^2\right],$$

$$= E\left[\delta_{\mathrm{S}}(x)^2\right] - 2pE\left[\delta_{\mathrm{S}}(x)^2\right] + p^2,$$

$$= E\left[\frac{x+1}{n+2}\right] - 2pE\left[\frac{x^2 + 2x + 1}{(n+2)^2}\right] + p^2,$$

Fist, recalling that $E[X^2] = \mathrm{Var}(X) + (E[X])^2$ and, for the data at hand $E[X \mid n, p] = np$ and $\mathrm{Var}(X \mid n, p) = np(1 - p)$. This gives $E[X^2] = np(1 - p) + (np)^2$. Moreover,

$$R_{\mathrm{F}}(\delta_{\mathrm{S}}(x), p) = E\left[\frac{x+1}{n+2}\right] - 2pE\left[\frac{x^2 + 2x + 1}{(n+2)^2}\right] + p^2,$$

$$= \frac{1 + np}{n+2} - \frac{2p\left(np(1-p) + n^2p^2 + 2np + 1\right)}{(n+2)^2} + p^2,$$

$$= \frac{1 + np}{n+2} - \frac{2np^2 + 2n(n-1)p^3 + 4np^2 + 2p}{(n+2)^2} + p^2.$$

We can now compute

$$r(\pi, \delta_{\mathrm{S}}) = E_\pi\left[R_{\mathrm{F}}(\delta_{\mathrm{S}}(x), p)\right],$$

$$= E_\pi\left[\frac{1 + np}{n+2} - \frac{2np^2 + 2n(n-1)p^3 + 4np^2 + 2p}{(n+2)^2} + p^2\right],$$

$$= \frac{6(n+2) + 3n(n+2) - (4n + 3n(n-1) + 8n + 6)) + 2(n+2)^2}{6(n+2)^2}.$$

A brief look at a sketch of the risks in Figure 3 shows that the posterior mean offers much higher integrated risk than MAP or MLE. So in this case, if you are concerned with frequentist loss you should use the
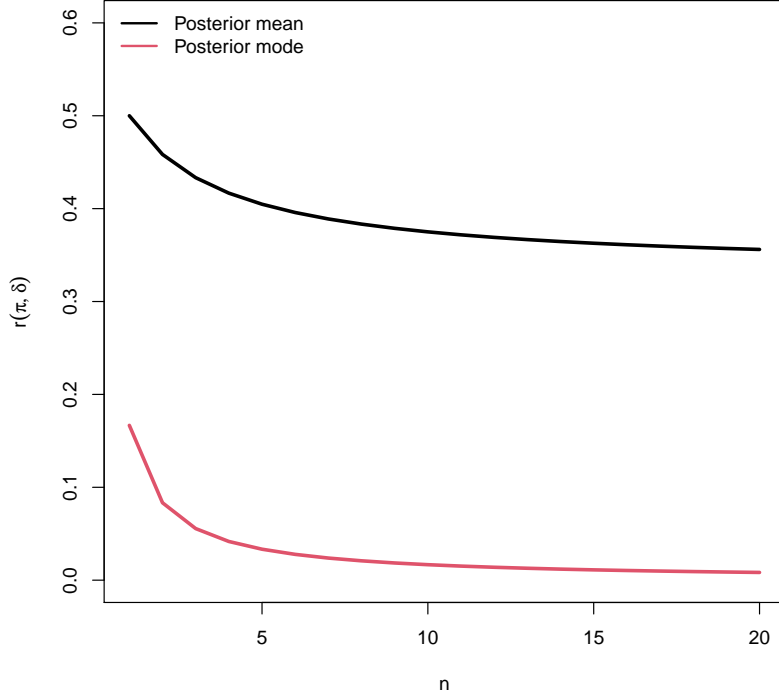
5

Figure 3: **Integrated risk for the posterior mean and the posterior mode**. The black line shows $r(\pi, \delta_{\mathrm{S}})$ whilst the red one depicts $r(\pi, \delta_{\mathrm{MAP}})$.

estimator under 0-1 loss. Finally, we will produce a prediction for the next m rolls, conditional on having observed $x = 6$ balls stop to the left of W out of $n = 9$. This can be accomplished by computing the posterior predictive:

$$
\begin{aligned}
g(\tilde{x} \mid x, n, m) &= \int_0^1 f(\tilde{x} \mid m, z) \xi(z \mid x, n) \, dz, \\
&= \frac{\binom{m}{\tilde{x}}}{B(\alpha', \beta')} \int_0^1 z^{\tilde{x}+\alpha'-1}(1-z)^{m-\tilde{x}+\beta'-1} \, dz, \\
&= \binom{m}{\tilde{x}} \frac{B(\alpha' + \tilde{x}, m - \tilde{x} + \beta')}{B(\alpha', \beta')},
\end{aligned}
$$

i.e. $\Pr(\tilde{X} = \tilde{x} \mid m, n, x)$ is the p.m.f. of a Beta-binomial distribution. This gives a full probability distribution for prediction. Knowing the predictive p.m.f., we can give a point prediction as

$$
E_g[\tilde{X}] = \frac{m(x+1)}{n+2} = m\delta_{\mathrm{S}}.
$$

This result can also be arrived at by the law of total expectation.

Discussion: In this question we have just seen how to conduct a rigorous Bayesian analysis of a problem, using a simple (by modern standards) classical example. We worked through the calculations involved in evaluating estimators under both Bayes and integrated risk and saw explicitly that different choices of risk lead to different estimators being selected. We also produced Bayesian predictions both point and distributional. ∎

6

2. Proper behaviour.

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

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

   is proper almost surely.

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

Now,

c) (10 marks) Take

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

   and

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

   The posterior density is then

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

   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.

Concepts: propriety; bias; indicator functions.
Level: Intermediate.
Solution: To answer a), we need to show that $m(x) = \int_{\Theta} f(x \mid t)\pi(t) \, dt < \infty$. We will first address the simpler part of the derivation. Under the assumptions we may take the sampling distribution to be absolutely continuous w.r.t. the Lebesgue measure, and write

$$\int_{\mathcal{X}} m(v) \, dv = \int_{\mathcal{X}} \int_{\Theta} f(v \mid t)\pi(t) \, dv \, dt.$$

Here, the functions involved are (i) non-negative and (ii) continuous and we are dealing with σ-finite probability spaces, so we can apply Fubini's[3] theorem to write:

$$\int_{\mathcal{X}} m(v) \, dv = \int_{\Theta} \int_{\mathcal{X}} f(v \mid t) \, dv \, \pi(t) \, dt, \tag{3}$$

$$= \int_{\Theta} 1 \cdot \pi(t) \, dt = 1.$$

We are *almost*[4] done. All we need to do now is to take care of the case where $m(x)$ is infinite. The only way $m(x)$ can be infinite if (3) is finite is if m is infinite *only* on a set of measure zero. Therefore, we are justified in saying that, if the prior is proper, the posterior is proper *almost surely*. Now, to show b), we will show that unbiasedness of the posterior mean implies something that only happens in a set

---

[3] Guido Fubini (1879-1943) was an Italian mathematician, known for his work in mathematical analysis.
[4] Pun most definitely intended.

of measure zero. First, let E_θ, $E_{X|\theta}$ and E_X denote expectations with respect to the prior, sampling (likelihood) and marginal (data) distributions, respectively. Now let us write the integrated risk as nested expectations:

$$r(\pi, \delta) = E_\theta \left[E_{X|\theta} \left[(\delta(X) - \theta)^2 \right] \right]. \tag{4}$$

Now observe that unbiasedness of $\delta_S(x) := E_{\theta|X}[\theta]$ means that

$$E_{X|\theta} \left[E_{\theta|X}[\theta] \right] = \theta \, \forall \, \theta \in \boldsymbol{\Theta},$$

Assuming unbiasedness of the posterior mean and developing (4), we get

$$r(\pi, \delta(X)) = E_\theta \left[E_{X|\theta} \left[\delta(X)^2 - 2\delta(X)\theta + \theta^2 \right] \right], \tag{5}$$

$$= E_\theta \left[E_{X|\theta} \left[\theta^2 \right] \right] - E_\theta \left[E_{X|\theta} \left[\theta\delta(X) \right] \right] - E_X \left[E_{\theta|X} \left[\theta\delta(X) \right] \right] + E_X \left[E_{\theta|X} \left[\delta(X)^2 \right] \right] \tag{6}$$

$$= E_\theta \left[E_{X|\theta} \left[\theta^2 \right] \right] - E_\theta \left[\theta E_{X|\theta} \left[\delta(X) \right] \right] - E_X \left[\delta(X) E_{\theta|X} \left[\theta \right] \right] + E_X \left[E_{\theta|X} \left[\delta(X)^2 \right] \right], \tag{7}$$

$$= E_\theta \left[E_{X|\theta} \left[\theta^2 \right] \right] - E_\theta \left[\theta^2 \right] - E_X \left[\delta(X)^2 \right] + E_X \left[\delta(X)^2 \right], \tag{8}$$

$$= 0,$$

where we have used Fubini's theorem for expectations plus some rearranging, i.e., we sometimes wrote $E_\theta[E_{X|\theta}[\cdot]]$ and sometimes $E_X[E_{\theta|X}[\cdot]]$ as convenient. These calculations show that assuming unbiasedness leads to an integrated risk of zero, which only happens when $\delta_S(X) = \theta$ with probability 1, which in turn only happens in a set of measure zero. We have therefore obtained our target result. The solution to c) is given in a CrossValidated post by a confused yours truly: https://stats.stackexchange.com/questions/370624/proper-prior-leading-to-improper-posterior/. Reproduced here for convenience. First, consider the correct way to write the likelihood:

$$f(x \mid \theta) = \frac{2x}{\theta} \mathbb{I}_{\leq \sqrt{\theta}}(x),$$

and notice $\mathbb{I}_{\leq \sqrt{\theta}}(x) = \mathbb{I}_{\geq x^2}(\theta)$. Thus, we can write

$$p(\theta \mid x) \propto \frac{4x}{\pi} \frac{1}{\theta(1+\theta^2)} \mathbb{I}_{\geq x^2}(\theta),$$

such that integration now gives

$$m(x) = \int_{\boldsymbol{\Theta}} \frac{4x}{\pi} \frac{1}{t(1+t^2)} \mathbb{I}_{\geq y^2}(t) \, dt = \int_{y^2}^\infty \frac{4x}{\pi} \frac{1}{t(1+t^2)} \, dt,$$

$$= \frac{2x}{\pi} \left(\log(x^4 + 1) - 4\log(x) \right).$$

This last expression is finite except perhaps in a set of measure zero (where x might take infinite values), hence the result is established.

Discussion: This problem helps put the pins in place for the theory of proper priors. Using proper priors has many advantages – specially for model choice, which we will see in a few lectures – and one of the biggest ones is almost sure propriety of the posterior, which authorises its use as a device encoding our beliefs about the quantities of interest in the model. We also saw the importance of judicious use of indicator functions to properly restrict the support of the functions involved and avoid wrong conclusions. Finally, a note on unbiasedness: as we have just shown, the posterior mean can never be unbiased. This does not seem to bother most Bayesians. Why? Quite simply because unbiasedness is not as useful as one might think. In practice, an unbiased estimator with high variance is less useful than a biased estimator with substantially lower variance. Secondly, phenomena such as the *Stein effect*[5] make clear that there is no *the* unbiased estimator in higher dimensions. See the discussion in Section 2.8.2 of Robert (2007), specially page 99. ∎

---

[5]Charles Max Stein (1920-2016) was an American statistician, known for his work on the admissibility of minimax estimators in higher dimensions.

3. Rayleigh dispersion[6].

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

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

This is the Rayleigh[7] 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;

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

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

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

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

a
Concepts: exponential family; Jeffreys's prior; reparametrisation
Level: Intermediate.
Solution: The first item basically asks us to find a reparametrisation of the model. Let us compute, it, then:

$$\eta = 1 - \int_0^t \frac{u}{\sigma^2} \exp\left(-\frac{u^2}{2\sigma^2}\right) du,$$

$$= 1 - \left(1 - \exp\left(-\frac{t^2}{2\sigma^2}\right)\right),$$

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

This immediately gives

$$\sigma^2 = -\frac{t^2}{2\log\eta}.$$

Now, let us derive the joint conditional distribution of $\boldsymbol{x}$:

$$f_n(\boldsymbol{x} \mid \sigma^2) = \sigma^{-2n} \prod_{i=1}^n x_i \exp\left(\frac{1}{2\sigma^2} S^2\right),$$

where $S^2 := \sum_{i=1}^n x_i^2$. So,

$$f_n(\boldsymbol{x} \mid \eta) = \left(-\frac{2\log\eta}{t^2}\right)^n \left(\prod_{i=1}^n x_i\right) \exp\left(\frac{\log\eta}{t^2} S^2\right), \tag{9}$$

answers b). To answer c) we will recall that to say $f(x \mid \theta)$ is in the exponential family is to say that it can be written as:

$$f(x \mid \theta) = h(x) \exp\left\{R(\theta) \cdot T(x) - \psi(\theta)\right\}.$$

Putting $\theta = \sigma^2$ and making $h(x) = \mathbb{I}_{(0,\infty)}(x) \cdot x$, $R(\sigma) = -1/2\sigma^2$, $T(x) = x^2$ and $\psi(\sigma) = 2\log(\sigma)$ shows the Rayleigh distribution to be a single-parameter exponential family. We will now use the special properties of the exponential family's cumulant-generating function to compute the expectation of the sufficient statistic, $T(X) = X^2$. First, consider the natural parametrisation: if $R(\theta) = -1/2\sigma^2$, then the natural parameter is $\alpha = -1/2\sigma^2$ and thus $\psi(\alpha) = -\log(-2\alpha)$. The moment generating function of $T(x)$ is thus

$$M(t) = \exp\left(\psi(\alpha + t) - \psi(\alpha)\right),$$

---

[6]This question's title is pun: https://en.wikipedia.org/wiki/Rayleigh_scattering.
[7]John William Strutt (1842-1919), the Third Baron Rayleigh, was a British physicist, known for his work on Statistical Mechanics.

which gives

$$E[x^2 \mid \sigma] = \left.\frac{d^2}{dt^2} M(t)\right|_{t=0} = -\alpha^{-1} = 2\sigma^2.$$

Using the reparametrisation, we get $E[x^2 \mid \eta] = t^2/\log\eta$. One could also compute $E[X^2]$ directly via integration or by using the information in the Wikipedia article for the Rayleigh distribution[8] to deduce $E[X^2]$ from $\mathrm{Var}(X) = \sigma^2(4-\pi)/2$ and $E[X] = \sigma\sqrt{\pi/2}$. Now we will derive the Jeffreys prior

$$\pi_J(\eta) \propto \sqrt{|I(\eta)|},$$

where $I(\eta)$ is the Fisher information. Dropping constants, differentiating the log of (9) and taking expectations, we get

$$I(\eta) = -E\left[\frac{\partial^2}{\partial\eta^2} \log f_n(\eta)\right] = -\left[\eta^2 \left(\log(\eta)\right)^2\right]^{-1} \text{ for } 0 < \eta < 1,$$

hence $\pi_J(\eta) \propto [\eta \log \eta]^{-1}$. Now, let us explore a different path to this derivation and show how all of this trouble could be avoided. We can exploit the invariance property of the Jeffreys's prior: since σ^2 is a scale parameter, it follows that $\pi_J(\sigma) \propto \sigma^{-1}$. Recall that we have $g(\sigma) = t/\sqrt{-2\log\eta}$. From this we can surmise

$$\pi_J(\eta) \propto \pi_J(g(\sigma))\left|\frac{dg(\sigma)}{d\sigma}\right|,$$

$$\propto \frac{\sqrt{-2\log\eta}}{t} \cdot \frac{t}{\sqrt{(-2\log\eta)^3}},$$

$$\propto \frac{1}{\eta \log \eta}.$$

Good. Now, on to the posterior:

$$\xi(\eta \mid \boldsymbol{x}) \propto \frac{1}{\eta \log \eta} (-\log\eta)^n \eta^{\frac{1}{t^2}S^2},$$

$$\propto (-\log\eta)^{n-1} \eta^{\frac{1}{t^2}S^2 - 1}.$$

It might not look like it, but this is the kernel of a Gamma p.d.f., and therefore

$$\xi(\eta \mid \boldsymbol{x}) = \frac{\left(\frac{1}{t^2}S^2\right)^n}{\Gamma(n)} (-\log\eta)^{n-1} \eta^{\frac{1}{t^2}S^2 - 1}, 0 < \eta < 1.$$

We are now prepared to compute the predictive

$$p(\tilde{x} \mid \boldsymbol{x}) = \int_{\boldsymbol{\Omega}} \xi(t \mid \boldsymbol{x}) f(\tilde{x} \mid t)\, dt,$$

$$= \kappa \int_0^\infty z^n \exp\left(-z\left\{\frac{S^2 + \tilde{x}^2}{t^2}\right\}\right)\, dz,$$

where

$$\kappa := \frac{2\tilde{x}}{\Gamma(n)}\left(\frac{1}{t^2}S^2\right)^n,$$

$$z := -\log\eta.$$

Again making use of the fact that the integrand is the kernel of a Gamma p.d.f., we arrive at our predictive density

$$p(\tilde{x} \mid \boldsymbol{x}) = n\left(\frac{1}{t^2}S^2\right)^n \left(\frac{2\tilde{x}}{t^2}\right)\left(\frac{t^2}{S^2 + \tilde{x}^2}\right)^{n+1} = \frac{2n\tilde{x}\left(S^2\right)^n}{(S^2 + \tilde{x}^2)^{n+1}}.$$

Discussion: In this question we explored a single-parameter exponential distribution, derived a Jeffreys's prior for a transformation of the original parameter and then used this prior to obtain a proper posterior and its corresponding posterior predictive distribution. In our calculations, we made use of special properties of the cumulant-generating function of the exponential family. Additionally, we have used the invariance of the Jeffreys's prior to cut back on calculations and achieve a cleaner solution. ∎

---

[8] https://en.wikipedia.org/wiki/Rayleigh_distribution.

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.

Edwards, A. (1978). Commentary on the arguments of Thomas Bayes. *Scandinavian Journal of Statistics*, pages 116–118.

Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation.* Springer Science & Business Media.