Bayesian statistics problem set 1

Florent Leclercq[a)]
CNRS & Sorbonne Université, UMR 7095, Institut d'Astrophysique de Paris, 98 bis boulevard Arago, F-75014 Paris, France

(Dated: 3 April 2025)

I. BAYESIAN INFERENCE WITH VIROLOGY TESTS

A new virology test is developed to detect a particular viral infection. The test has been evaluated and found to have the following characteristics:

- Sensitivity (true positive rate): 98% (if a patient is infected, the test gives a positive result 98% of the time).
- Specificity (true negative rate): 95% (if a patient is not infected, the test gives a negative result 95% of the time).

Assume that in a given population the prevalence (i.e., the prior probability) of the infection is 2%.

1. What is the probability that a patient is actually infected if they obtain a single positive test result?
2. Suppose the patient receives two independent positive test results. Given that the tests are independent and use the same sensitivity and specificity as above, what is now the probability that the patient is infected?
3. Consider now that the patient is tested twice, but the test results are discordant (one positive and one negative, in any order). What is the probability that the patient is infected given one positive and one negative result?
4. Discuss how the prevalence (prior probability) of the infection affects the interpretation of test results. For example, if the prevalence increases to 20%, recalculate the posterior for a positive test result using the same sensitivity and specificity.

II. THE MONTY HALL PROBLEM

Solve the "Monty Hall" problem given in the lectures, using Bayes' theorem.

You are a contestant on a game show, and you are presented with three closed doors. Behind one of the doors is a brand new car, while the other two doors have goats behind them. You have no knowledge of which door has the car and which doors have the goats.

You get to choose one of the three doors, but before it is opened, the game show host (Monty Hall) opens one of the other two doors and shows you that it has a goat behind it. Now, you have the option to stick with your original choice or switch to the remaining unopened door. What is the probability that the car is behind the door you originally chose, and what is the probability that the car is behind the other unopened door?

III. THE DOMINANCE OF THE LIKELIHOOD IN BAYESIAN INFERENCE

Suppose you have a coin with an unknown probability of heads, θ. You decide to model the coin toss outcomes as independent Bernoulli trials and use Bayesian inference to learn θ. Assume a Beta prior for θ:

$$\theta \sim \mathrm{Beta}(a, b). \quad (1)$$

Because the Beta distribution is a conjugate prior for the Bernoulli likelihood, the posterior will also be a Beta distribution. We will demonstrate that for small n the choice of prior matters, but as n grows large, the data "overwhelms" the prior.

1. Derivation of the posterior distribution

[a)] Electronic mail: florent.leclercq@iap.fr; https://www.florent-leclercq.eu/; ORCID: 0000-0002-9339-1404

(a) Write down the likelihood function for k heads in n tosses.

(b) Given the Beta(a, b) prior for θ, derive the expression for the posterior distribution $p(\theta|\text{data})$.

2. Small sample simulation
Assume that the true coin bias is $\theta_{\text{true}} = 0.7$ and that you simulate $n = 10$ coin tosses. Consider three different priors:

   - **Prior A**: Weakly informative prior Beta$(2, 2)$
   - **Prior B**: Beta$(5, 5)$
   - **Prior C**: Informative prior Beta$(20, 20)$

   Perform the following tasks:

   (a) Simulate 10 coin tosses (use a random seed for reproducibility).

   (b) For each prior, compute the posterior parameters.

   (c) Plot the posterior distributions on the same graph.

   (d) Discuss the influence of the prior when the sample size is small.

3. Large sample simulation
Now, simulate $n = 1000$ coin tosses using the same true coin bias $\theta_{\text{true}} = 0.7$ and repeat the analysis using the same three priors:

   - **Prior A**: Weakly informative prior Beta$(2, 2)$
   - **Prior B**: Beta$(5, 5)$
   - **Prior C**: Informative prior Beta$(20, 20)$

   Perform the following:

   (a) Simulate 1000 coin tosses.

   (b) Compute the posterior parameters for each prior.

   (c) Plot the posterior distributions on the same graph.

   (d) Discuss how and why the influence of the prior changes with the larger sample size.

4. Comparing posterior means and variances
For both the small sample ($n = 10$) and large sample ($n = 1000$) cases, compute the posterior mean and variance for each prior. Recall that for a Beta(α, β) distribution:

   - The mean is: $\mu = \dfrac{\alpha}{\alpha + \beta}$
   - The variance is: $\sigma^2 = \dfrac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

   Write Python code to compute these quantities and observe how they converge with more data.

This exercise shows that while the prior can have a strong influence when data are scarce (as seen with $n = 10$), its effect diminishes as the amount of data increases (as seen with $n = 1000$). In the large-sample limit, the posterior is dominated by the likelihood, ensuring that the inference about θ is driven primarily by the observed data rather than the prior beliefs. This is a key idea in Bayesian statistics and is sometimes referred to as "Bayesian consistency."

Feel free to experiment further by modifying the priors or the true value of θ to see how robust this phenomenon is under different settings.

IV. IGNORANCE PRIORS FOR AN URN PROBLEM

Suppose we have an urn containing red and white balls. The probability of drawing a red ball is an unknown parameter θ (with $\theta \in [0, 1]$). In the absence of any prior information about θ, we might wish to assign a prior probability density $\pi(\theta)$ that represents "complete ignorance." However, as Jaynes famously argued, the idea of "ignorance" depends on how one parameterises the problem. In what follows we will explore different invariance requirements for an "uninformative" prior and see that they lead to different answers.

1. Invariance under label exchange
Because the labels "red" and "white" are arbitrary, one might argue that our state of ignorance should be invariant under exchanging the two colors.

(a) Explain why this invariance condition seems natural for representing ignorance, and write the functional equation that $\pi(\theta)$ must satisfy.

(b) Verify that the uniform prior

$$\pi(\theta) = 1, \quad 0 < \theta < 1, \tag{2}$$

satisfies this invariance.

(c) Show that the uniform prior on $[0, 1]$ is also the maximum entropy prior.

2. Scale invariance of the odds
An alternative idea is to require that our state of ignorance be invariant under reparameterisation. A common reparameterisation is in terms of the *odds*:

$$\phi = \frac{\theta}{1-\theta}, \quad \text{with inverse } \theta = \frac{\phi}{1+\phi}. \tag{3}$$

Since $\phi \in [0, \infty)$, one might require that ignorance about ϕ be expressed by a prior that is *scale invariant*. In other words, if we "rescale" the odds by a positive constant, our ignorance should remain the same.

(a) Functional equation for scale invariance
Assume that the prior density for ϕ, called $\pi_\phi(\phi)$, satisfies the following invariance property for any scaling factor $a > 0$:

$$\pi_\phi(a\phi) = \frac{1}{a}\,\pi_\phi(\phi). \tag{4}$$

Show that (up to a multiplicative constant) the unique solution of this functional equation is

$$\pi_\phi(\phi) \propto \frac{1}{\phi}. \tag{5}$$

(b) Transforming back to the θ parameter
The densities $\pi(\theta)$ and $\pi_\phi(\phi)$ are related by the usual change-of-variables formula:

$$\pi(\theta) = \pi_\phi(\phi) \left|\frac{\mathrm{d}\phi}{\mathrm{d}\theta}\right|. \tag{6}$$

i. Compute $\dfrac{\mathrm{d}\phi}{\mathrm{d}\theta}$.
ii. Express $\pi(\theta)$ in terms of θ by using the result from question (a).

3. Jeffreys prior
The Jeffreys prior is defined by

$$\pi_J(\theta) \propto \sqrt{|I(\theta)|}, \tag{7}$$

i.e. its density is proportional to the square root of the determinant of the Fisher information matrix $I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log p(r|\theta)\right)^2\right]$.

(a) Compute the Fisher information for the binomial likelihood of our problem,

$$p(r|\theta) = \binom{N}{r}\,\theta^r\,(1-\theta)^{N-r}. \tag{8}$$

Show that

$$\pi_J(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}}. \tag{9}$$

(b) Compare the three candidate priors for θ (the uniform prior, the odds-invariance prior, and the Jeffreys prior). Discuss what these differences imply about the notion of a "noninformative" or "ignorance" prior.

ACKNOWLEDGEMENTS

I thank Benjamin Wandelt, Alan Heavens and colleagues of the SOC of the ICIC Data Analysis workshop 2021 for their own lectures, from which this document is inspired.