Solutions of Exercises on Probability Theory and Bayesian Statistics

Luc Demortier[1]

Problem 1: Eliminating nuisance parameters by conditioning.
In the frequentist paradigm, handling nuisance parameters can be a thorny problem. A method that sometimes works is based on the idea of conditioning. To illustrate this approach, suppose we measure an event count N that is Poisson distributed with mean $\mu\nu$, where μ is the parameter of interest and ν a nuisance parameter. Assume that ν is constrained by the auxiliary measurement of a Poisson variate K with mean $\tau\nu$, where τ is a known constant:

$$N \sim \text{Poisson}(\mu\nu), \tag{1}$$
$$K \sim \text{Poisson}(\tau\nu). \tag{2}$$

In high energy physics one could think of μ as the production cross section for some process of interest and ν as a product of efficiencies, acceptances, and integrated luminosity. One can argue that the *sum* $M \equiv N + K$ provides no information about the *ratio* μ/τ of the above two Poisson means, or about μ itself. It is therefore interesting to seek inferences that condition on M.

1. Compute the conditional distribution of N given M.
   A: By definition of conditional probability, we have:

$$\mathbb{P}(N = n \mid M = m) = \frac{\mathbb{P}(N = n \;\&\; M = m)}{\mathbb{P}(M = m)}, \tag{3}$$

$$= \frac{\mathbb{P}(N = n \;\&\; K = m - n)}{\mathbb{P}(M = m)}, \tag{4}$$

$$= \frac{\mathbb{P}(N = n)\, \mathbb{P}(K = m - n)}{\mathbb{P}(M = m)} \quad \text{[by independence of } N \text{ and } K\text{]}, \tag{5}$$

$$= \frac{[(\mu\nu)^n e^{-\mu\nu}/n!] \times [(\tau\nu)^{m-n} e^{-\tau\nu}/(m-n)!]}{(\mu\nu + \tau\nu)^m e^{-\mu\nu-\tau\nu}/m!}, \tag{6}$$

$$= \binom{m}{n} \frac{(\mu\nu)^n\, (\tau\nu)^{m-n}}{(\mu\nu + \tau\nu)^m}, \tag{7}$$

$$= \binom{m}{n} \frac{\mu^n\, \tau^{m-n}}{(\mu + \tau)^m}, \tag{8}$$

$$= \binom{m}{n} \left(\frac{\mu}{\mu + \tau}\right)^n \left(1 - \frac{\mu}{\mu + \tau}\right)^{m-n}, \tag{9}$$

[1] luc.demortier@mail.rockefeller.edu

which is a binomial distribution for m trials and with probability of success equal to $\mu/(\mu+\tau)$. The ν dependence has been eliminated, allowing this function to be used as likelihood for calculating point and interval estimates for the parameter of interest μ.

2. Next, assume that the expectation value of N is the sum of μ and ν instead of their product, so we have:

$$N \sim \text{Poisson}(\mu+\nu), \tag{10}$$
$$K \sim \text{Poisson}(\tau\nu). \tag{11}$$

What is the conditional distribution of N given M here?

A: By definition of conditional probability, we have:

$$\begin{aligned}
\mathbb{P}(N=n \mid M=m) &= \frac{\mathbb{P}(N=n \;\&\; M=m)}{\mathbb{P}(M=m)}, &(12)\\
&= \frac{\mathbb{P}(N=n \;\&\; K=m-n)}{\mathbb{P}(M=m)}, &(13)\\
&= \frac{\mathbb{P}(N=n)\;\mathbb{P}(K=m-n)}{\mathbb{P}(M=m)}, &(14)\\
&= \frac{[(\mu+\nu)^n e^{-\mu-\nu}/n!]\times[(\tau\nu)^{m-n}e^{-\tau\nu}/(m-n)!]}{(\mu+\nu+\tau\nu)^m e^{-\mu-\nu-\tau\nu}/m!}, &(15)\\
&= \binom{m}{n}\frac{(\mu+\nu)^n\,(\tau\nu)^{m-n}}{(\mu+\nu+\tau\nu)^m}, &(16)\\
&= \binom{m}{n}\left(\frac{\mu+\nu}{\mu+\nu+\tau\nu}\right)^n\left(1-\frac{\mu+\nu}{\mu+\nu+\tau\nu}\right)^{m-n}, &(17)
\end{aligned}$$

which is a binomial distribution for m trials and with probability of success equal to $(\mu+\nu)/(\mu+\nu+\tau\nu)$. The ν dependence can no longer be eliminated in this case, except when $\mu=0$. Although we can't use this conditional probability mass function to compute point and interval estimates for μ, we can still use it to test the hypothesis that $\mu=0$ by the p-value method.

Problem 2: Eliminating nuisance parameters by Bayesian marginalization.

Here we take the first of the above problems ($N \sim \text{Poisson}(\mu\nu)$) and eliminate the nuisance parameter ν by Bayesian marginalization. One can proceed as follows:

1. Consider the auxiliary measurement of ν, via $K \sim \text{Poisson}(\tau\nu)$.

2. Compute Jeffreys' prior for ν for that auxiliary measurement.

   A: This is

$$\pi_J(\nu) \propto \frac{1}{\sqrt{\nu}}. \tag{18}$$

3. Compute the posterior for ν for that auxiliary measurement.

   A: This is a Gamma density:

$$p(\nu \mid k) = \frac{\tau(\tau\nu)^{k-1/2} e^{-\tau\nu}}{\Gamma(k+1/2)}. \tag{19}$$

4. Use the auxiliary posterior for ν as a prior for ν in the measurement of μ.

   A: Set:

$$\pi(\nu) \equiv p(\nu \mid k). \tag{20}$$

5. Now we still need a prior for μ. In fact, since the problem involves two parameters, μ and ν, what we really need is a conditional prior for μ given ν. Reference analysis provides a method for calculating this conditional prior, and the result is identical to Jeffreys' prior calculated for a fixed value of ν.

   A: Jeffreys' prior is:

$$\pi_J(\mu \mid \nu) \propto \sqrt{\frac{\nu}{\mu}}. \tag{21}$$

   The ν dependence of this prior is important. Strictly speaking, it is not fully specified by the calculation of Jeffreys' prior. Jeffreys' prior only specifies the dependence on μ, up to a proportionality factor, since it is an improper prior. The proportionality factor could have an arbitrary dependence on the nuisance parameter ν. However that dependence is important, since it affects the properties of the posterior. In this problem it turns out that the dependence obtained by simply keeping track of ν during the calculation of the Jeffreys' prior for μ gives a satisfactory posterior.

6. Write out the joint posterior for μ and ν, without trying to normalize it.

   A:

$$p(\mu, \nu \mid n) \propto \frac{(\mu\nu)^n}{n!}\, e^{-\mu\nu}\, \frac{\tau(\tau\nu)^{k-1/2}}{\Gamma(k+1/2)}\, e^{-\tau\nu} \sqrt{\frac{\nu}{\mu}}. \tag{22}$$

7. Integrate out the ν dependence.

   A:

$$p(\mu \mid n) \propto \frac{\mu^{n-1/2}\tau^{k+1/2}}{n!\,\Gamma(k+1/2)}\, \frac{\Gamma(n+k+1)}{(\mu+\tau)^{n+k+1}}. \tag{23}$$

   This can be rewritten to make it look a little more like the result of the conditioning calculation in Problem 1:

$$p(\mu \mid n) \propto \binom{m+1}{n} \left(\frac{\mu}{\mu+\tau}\right)^{n-1/2} \left(1 - \frac{\mu}{\mu+\tau}\right)^{(m+1)-(n-1/2)}, \tag{24}$$

   where $m = n + k$ and the proportionality factor depends on n, k, and τ, but not on μ. After normalizing this posterior over μ we obtain

$$p(\mu \mid n) = \frac{\Gamma(n+k+1)}{\tau\,\Gamma(n+1/2)\,\Gamma(k+1/2)} \left(\frac{\mu}{\mu+\tau}\right)^{n-1/2} \left(1 - \frac{\mu}{\mu+\tau}\right)^{k+3/2}. \tag{25}$$

8. Compare the resulting posterior for μ with the μ dependence of the conditional pmf for N obtained in Problem 1.

   A: The μ dependence of the binomial pmf obtained in Problem 1 has the same functional form as the marginal posterior for μ obtained here. We can consider the binomial pmf as a likelihood function for μ and calculate the corresponding reference prior. This yields:

$$\pi_R(\mu) \propto \frac{1}{\sqrt{\mu}\,(\mu+\tau)}. \tag{26}$$

   The posterior obtained by multiplying the binomial likelihood with this reference prior is identical with the marginal posterior of equation (25).

Problem 3: Sampling to a foregone conclusion.

This is an exercise to illustrate the Law of the Iterated Logarithm. Write a little Monte Carlo program to do the following:

- Generate random numbers X_i from a Gaussian distribution with zero mean and unit standard deviation.

- As you generate them, compute a "running significance" Z_n:

$$Z_n \equiv \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} X_i \right|, \tag{27}$$

  where n is the number of Gaussian variates generated so far.

- Make a plot of Z_n versus n. At what value of n does the first crossing of $Z_n = 2$ occur? What about $Z_n = 3$? Compare with the curve $\sqrt{2\ln\ln n}$. What happens if, instead of checking the significance after each new data point, we only check it after every 100 new data points?

Problem 4: Bayesian intervals for an exponential lifetime.

Consider an exponential decay time t with probability density $f(t \mid \tau) = e^{-t/\tau}/\tau$. Derive Jeffreys' prior for this problem and compute the corresponding posterior. Construct equal-tailed intervals from this posterior and compute their frequentist coverage. Repeat with a flat prior.

A: The likelihood function is

$$\mathcal{L}(\tau) = \frac{1}{\tau} e^{-t/\tau}. \tag{28}$$

Jeffreys' prior is obtained by taking the square root of the expectation value of the second derivative of the negative logarithm of this expression:

$$\pi_J(\tau) = \sqrt{\mathbb{E}\left[-\frac{d^2}{d\tau^2}\ln\mathcal{L}\right]}, \tag{29}$$

$$= \sqrt{\mathbb{E}\left[\frac{-1}{\tau^2} + \frac{2t}{\tau^3}\right]}, \tag{30}$$

$$= \sqrt{\frac{-1}{\tau^2} + \frac{2}{\tau^2}} \tag{31}$$

$$= \frac{1}{\tau}. \tag{32}$$

The posterior is therefore

$$p(\tau \mid t) = \frac{e^{-t/\tau}/\tau^2}{\int_0^\infty e^{-t/\tau}/\tau^2\, d\tau} = \frac{t}{\tau^2}\, e^{-t/\tau}. \tag{33}$$

To construct a central interval $[\tau_1, \tau_2]$ with γ credibility we solve the following equations for τ_1 and τ_2:

$$\int_0^{\tau_1} p(\tau \mid t)\, d\tau = \frac{1-\gamma}{2} \quad \text{and} \quad \int_{\tau_2}^{\infty} p(\tau \mid t)\, d\tau = \frac{1-\gamma}{2}. \tag{34}$$

This yields

$$\tau_1 = \frac{-t}{\ln\left(\frac{1-\gamma}{2}\right)} \quad \text{and} \quad \tau_2 = \frac{-t}{\ln\left(\frac{1+\gamma}{2}\right)}. \tag{35}$$

The frequentist coverage of this interval is

$$\mathbb{P}_\tau[\tau_1(T) \leq \tau \leq \tau_2(T)] = \mathbb{P}_\tau\left[-\tau\ln\frac{1+\gamma}{2} \leq T \leq -\tau\ln\frac{1-\gamma}{2}\right] \tag{36}$$

$$= \int_{-\tau\ln\frac{1+\gamma}{2}}^{-\tau\ln\frac{1-\gamma}{2}} \frac{1}{\tau} e^{-t/\tau}\, dt \tag{37}$$

$$= \gamma. \tag{38}$$

Hence the frequentist coverage is exactly equal to the credibility.

If we had used a flat prior instead of Jeffreys' prior, the posterior would be proportional to the likelihood:

$$p(\tau \mid t) \propto \frac{1}{\tau} e^{-t/\tau}. \tag{39}$$

Unfortunately this posterior is not normalizable and is therefore improper:

$$\int_0^\infty \frac{1}{\tau} e^{-t/\tau}\, d\tau = \int_0^\infty \frac{e^{-u}}{u}\, du = \infty. \tag{40}$$

It can not be used to compute intervals.