Statistics for Astronomers
Solutions to Homework #7

Prof. Sundar Srinivasan

January 10, 2021

1. If $X \sim \mathscr{N}(\mu, \sigma^2)$, the likelihood of observing a value $X = x$ is given by

$$\mathscr{L}(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \Longrightarrow \ln\mathscr{L} = \text{constant} - \ln\sigma - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

The first-order partial derivatives w.r.t. the parameters are

$$\frac{\partial}{\partial\mu}\mathscr{L} = \left(\frac{x-\mu}{\sigma^2}\right) \qquad \frac{\partial}{\partial\sigma}\mathscr{L} = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}$$

The first derivative w.r.t. μ is already in a convenient form (numerator is the deviation of the data from the population mean, denominator is a function only of the variance), so we can compute the Fisher Information as

$$\mathcal{I}(\mu) = \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\mathscr{L}\right)^2\right] = \frac{1}{\sigma^2}\mathbb{E}\left[\left(\frac{x-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} = \text{constant}$$

Therefore, the Jeffreys prior for the population mean has the Uniform distribution.

For the standard deviation, we compute the second-order partial derivative to compute the Fisher Information and the Jeffreys prior:

$$\frac{\partial^2}{\partial\sigma^2}\mathscr{L} = \frac{1}{\sigma^2} - 3\frac{(x-\mu)^2}{\sigma^4}$$

$$\mathcal{I}(\sigma) = -\mathbb{E}\left[\frac{\partial^2}{\partial\sigma^2}\mathscr{L}\right] = \frac{1}{\sigma^4}\,\mathbb{E}\left[3\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right] \propto \frac{1}{\sigma^2}$$

$$\Longrightarrow \pi_{_J}(\sigma) \propto \sqrt{\mathcal{I}(\sigma)} \propto \frac{1}{\sigma}$$

That is, the Jeffreys prior for the standard deviation is the logarithmic prior (since $\ln\sigma$ is uniformly distributed).

2. (a) The likelihood for x successes in N trials is

$$\mathscr{L}(\theta) = \binom{N}{x}\,\theta^x\,(1-\theta)^{N-x} \Longrightarrow \ln\mathscr{L}(\theta) = \text{constant} + x\ln\theta + (N-x)\ln(1-\theta), \qquad (1)$$

and its derivative can be written in terms of the mean $\mathbb{E}[x] = N\theta$ and $\mathrm{Var}[x] = N\theta(1-\theta)$:

$$\frac{\partial}{\partial\theta}\mathscr{L}(\theta) = \frac{x}{\theta} - \frac{N-x}{1-\theta} = N\,\frac{x-\mathbb{E}[x]}{\mathrm{Var}[x]}$$

The square root of the Fisher Information is then

$$\sqrt{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\mathscr{L}(\theta)\right)^2\right]} = N\sqrt{\mathbb{E}\left[\frac{x-\mathbb{E}[x]}{\mathrm{Var}[x]}\right]} = \sqrt{\frac{N}{\theta(1-\theta)}} \propto \mathrm{Beta}\left(\frac{1}{2},\frac{1}{2}\right),$$

so that the Jeffreys prior is $\pi_J(\theta) = \mathrm{Beta}\left(\frac{1}{2},\frac{1}{2}\right)$.

(b) The mean and variance of Uniform$(0,1)$ are $\frac{1}{2}$ and $\frac{1}{12}$. The mean and variance of $\mathrm{Beta}\left(\alpha,\beta\right)$ are $\frac{\alpha}{\alpha+\beta}$ and $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Setting $\alpha=\beta=\frac{1}{2}$, the mean and variance of the Jeffreys prior are $\frac{1}{2}$ and $\frac{1}{8}$ respectively. Both prior distributions have the same mean, and the Jeffreys prior has a higher variance around this mean. A higher variance means less information. This is consistent with the Jeffreys prior being non-informative.

(c) The prior predictive distribution for data x is $p(x) = \int_0^1 d\theta\; p(x|\theta)\;\pi(\theta)$, where $p(x|\theta)$ is numerically equal to the likelihood given by Equation (1). For the Uniform prior, we get

$$p(x) = \int_0^1 d\theta\,\binom{N}{x}\theta^x\,(1-\theta)^{N-x} = \binom{N}{x}\,\frac{\Gamma(x+1)\,\Gamma(N-x+1)}{\Gamma(N+2)} = \frac{1}{N+1}$$

That is, each of the $N+1$ values of x $(0 \le x \le N)$ is equally likely. The prior predictive distribution for the data is therefore a discrete uniform distribution.
For the Jeffreys prior,

$$p(x) = \int_0^1 d\theta\,\binom{N}{x}\theta^x\,(1-\theta)^{N-x}\,\frac{1}{\pi}\,\theta^{-\frac{1}{2}}\,(1-\theta)^{-\frac{1}{2}} = \frac{1}{\pi}\,\binom{N}{x}\,\frac{\Gamma(x+\frac{1}{2})\,\Gamma(N-x+\frac{1}{2})}{\Gamma(N+1)}$$

This function has maxima at $x=0$ and $x=N$ of value $\dfrac{\Gamma(N+\frac{1}{2})}{\Gamma(N+1)} \approx 0.032$, and its shape reflects that of the Jeffreys prior, which places a large weight at both extremes. Figure 2 compares the predictive distributions corresponding to the two priors.

(d) We will use the following results for this part of the problem:

$$\mathbb{E}[\mathrm{Beta}(\alpha,\beta)] = \frac{\alpha}{\alpha+\beta};\qquad \mathrm{Var}[\mathrm{Beta}(\alpha,\beta)] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The posterior distribution for θ is obtained from Bayes' Theorem: $p(\theta|x) = \dfrac{p(x|\theta)\;\pi(\theta)}{p(x)}$.

Using the likelihood from Equation (1), for the Uniform prior, the posterior is distributed according to $\mathrm{Beta}(x+1, N-x+1)$, the mean and variance for which are $\dfrac{x+1}{N+2}$ and $\dfrac{(x+1)(N-x+1)}{(N+2)^2(N+3)}$

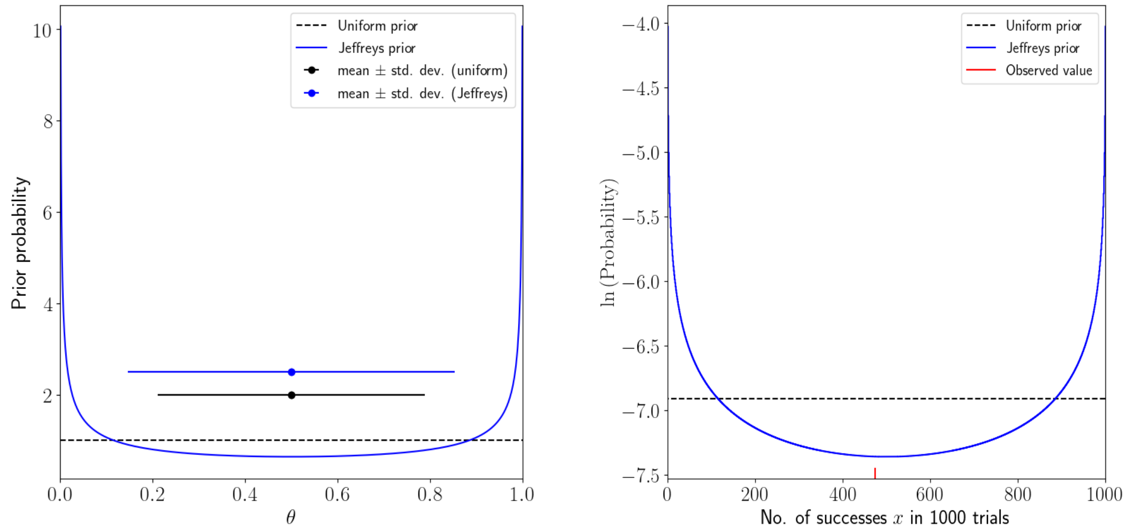

Figure 1: *Left*: distributions for θ according to the Uniform (black dashed) and Jeffreys (blue) priors, with the corresponding means and standard deviations shown. *Right*: prior predictive distributions for the Uniform (black dashed) and Jeffreys (blue) priors.

(0.4731 and 0.0158^2) respectively.

The product of the likelihood with the Jeffreys prior results in a posterior distributed according to Beta$(x+\frac{1}{2}, N-x+\frac{1}{2})$. Its mean and variance are $\frac{x+\frac{1}{2}}{N+1}$ and $\frac{(x+\frac{1}{2})(N-x+\frac{1}{2})}{(N+1)^2(N+2)}$ (0.4730 and 0.0158^2) respectively.

Given that both x and N are large, the posterior shapes and the associated means and variances are almost identical for the two priors (Figure **??**). This demonstrates the insensitivity to the prior for large datasets (the posterior is data-dominated rather than prior-dominated).

(e) The required probability can be computed using the cumulative distributions of the posteriors:

```
 N, x = 1000, 403
print(1 - scipy.stats.beta.cdf(0.5, x+1, N-x+1)) #for the Uniform prior
print(1 - scipy.stats.beta.cdf(0.5, x+1/2, N-x+1/2)) #for the Jeffreys prior
```

The required probabilities are 3.86×10^{-10} and 3.78×10^{-10} respectively.

(f) The posterior predictive distribution for a new observation $\widetilde{x}$ is given by $p(\widetilde{x}|x) = \int_0^1 d\theta\, p(\widetilde{x}|\theta)\, p(\theta|x)$.

The observation $\widetilde{x}$ takes values 0 and 1. We compute the above integral for both these values for both priors below.

For the Uniform prior, $p(\widetilde{x} = 0|x) = \frac{\Gamma(N+2-x)}{\Gamma(N+1-x)} \frac{1}{N+2} = \frac{N+1-x}{N+2} \approx 0.527$, so that $p(\widetilde{x} = 1|x) = 1 - p(\widetilde{x} = 0|x) = \frac{x+1}{N+2} \approx 0.473$.

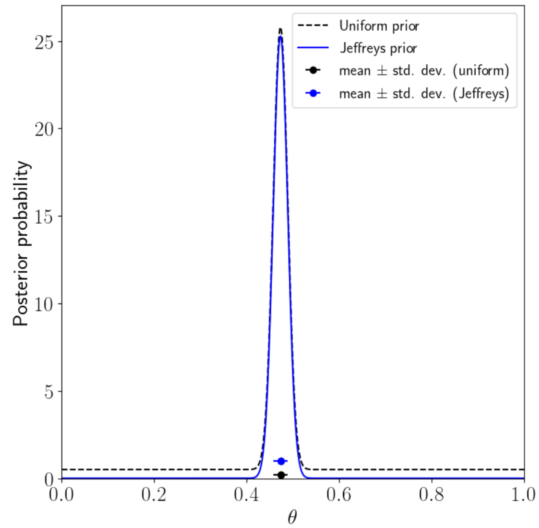

Figure 2: Posterior distributions for θ according to the Uniform (black dashed, shifted up for clarity) and Jeffreys (blue) priors, with the corresponding means and standard deviations shown.

Similarly, for the Jeffreys prior, $p(\widetilde{x} = 0|x) \approx 0.527$ and $p(\widetilde{x} = 1|x) \approx 0.473$.

Due to the large amount of data that has already been obtained for the coin, and because the outcome of each coin flip is independent of the previous outcomes, the probability of obtaining a head or tail on the next coin flip is very tightly constrained.