
Chapter 1: Solutions to Exercises

Exercise 1

Assume $x_i > 0$ for all i (otherwise, the likelihood equals 0 and the posterior is undefined). Since the data is independent given θ , the likelihood factors and we get

$$\begin{aligned} p(x_{1:n}|\theta) &= \prod_{i=1}^n p(x_i|\theta) \\ &= \prod_{i=1}^n \theta e^{-\theta x_i} \\ &= \theta^n e^{-\theta \sum x_i}. \end{aligned}$$

Thus, using Bayes' theorem,

$$\begin{aligned} p(\theta|x_{1:n}) &\propto p(x_{1:n}|\theta)p(\theta) \\ &\propto \theta^n e^{-\theta \sum x_i} \theta^{a-1} e^{-b\theta} \mathbb{1}(\theta > 0) \\ &= \theta^{a+n-1} e^{-(b+\sum x_i)\theta} \mathbb{1}(\theta > 0) \\ &\propto \text{Gamma}(\theta \mid a + n, b + \sum x_i). \end{aligned}$$

Therefore, since the posterior density must integrate to 1, we have

$$p(\theta|x_{1:n}) = \text{Gamma}(\theta \mid a + n, b + \sum x_i).$$

Exercise 2

See Figure 1. See Appendix for source code.

Exercise 3

A minimal requirement for using an Exponential model would be that the data is always positive-valued and continuous (not discrete). However, this is not enough to justify using this model. For any parametric model, it is important that the data can be “well-fit” for some choice of parameter values. (For the purposes of this homework, I wouldn’t have expected you to know how to assess this yet—but you should learn it now.) Before seeing the data, this can be judged by knowing roughly how the data is generated (see below), and after seeing the data, by conducting tests such as goodness-of-fit or posterior predictive checks (which we will discuss later).

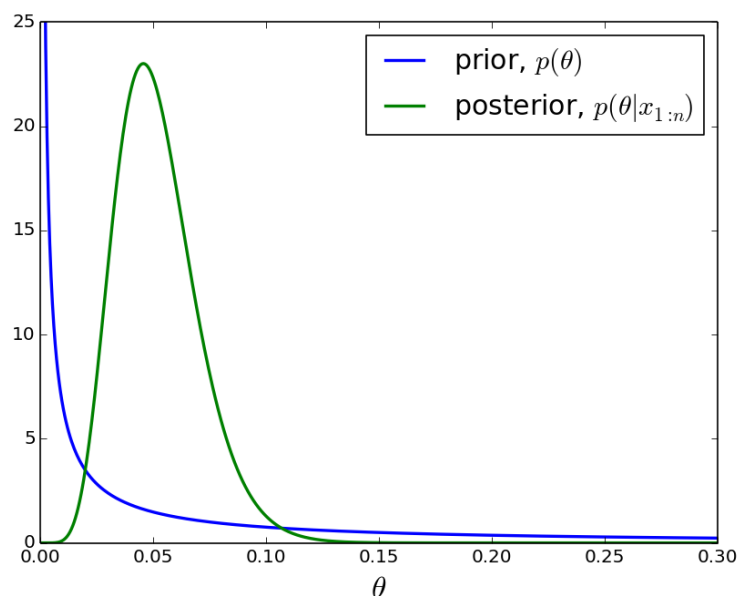


Figure 1: Prior and posterior densities.

The Exponential distribution has certain special properties (e.g., memorylessness, time between events in a Poisson process, characterization in terms of extreme value theory) that make it well-suited to certain applications. For instance, the “memorylessness” property is that if $X \sim \text{Exp}(\theta)$, then for any $s, t \geq 0$,

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t).$$

The Poisson process connection is easier to think about (for me at least). If you can have something that is well-characterized by a Poisson process, then the inter-arrival times are i.i.d. Exponentials. In addition to the examples given in Lecture 1, another example would be the time between occurrences of a very rare disease caused by a genetic mutation, such as acromegaly (which often leads to “gigantism”, see <http://en.wikipedia.org/wiki/Gigantism>), if we assume a constant overall birthrate over the period of time of observation.

A simple example of something that would not be well-modeled by a Exponential distribution would be anything that could take negative values or is discrete—e.g., the amount of money you have minus what you owe, or the number of children in a family. An example of something positive-valued that would not be well-fit by an Exponential distribution is the distribution of heights of adult humans.

Exercise 4

Since $\ell(s, a) = \mathbb{1}(s \neq a)$ then

$$\begin{aligned}\rho(a, x_{1:n}) &= \mathbb{E}(\ell(S, a) | x_{1:n}) \\ &= \mathbb{E}(\mathbb{1}(S \neq a) | x_{1:n}) \\ &= \mathbb{P}(S \neq a | x_{1:n}) \\ &= 1 - \mathbb{P}(S = a | x_{1:n}).\end{aligned}$$

Thus,

$$\arg \min_a \rho(a, x_{1:n}) = \arg \max_a \mathbb{P}(S = a | x_{1:n}),$$

that is, the action a that minimizes $\rho(a, x_{1:n})$ is the same as the action a that maximizes the posterior probability $\mathbb{P}(S = a | x_{1:n})$.

Exercise 5

Intuitively, the simplest thing that occurs to me is to look at the sample mean $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and see if it is greater than $1/2$. This leads to the decision rule:

$$\delta_1(x_{1:n}) = \mathbb{1}(\bar{x} > \tfrac{1}{2}).$$

Now, for the Bayes rule. Taking $S = X_{n+1}$ and $a = x_{n+1}$, from Exercise 4 we know that the Bayes procedure is to choose x_{n+1} to maximize $\mathbb{P}(X_{n+1} = x_{n+1} | x_{1:n})$, i.e., to maximize the posterior predictive, which we know from Lecture 2 is

$$p(x_{n+1} | x_{1:n}) = \frac{a_n^{x_{n+1}} b_n^{1-x_{n+1}}}{a_n + b_n} \mathbb{1}(x_{n+1} \in \{0, 1\})$$

where $a_n = a + \sum x_i$ and $b_n = b + n - \sum x_i$. This gives us the decision rule

$$\delta_2(x_{1:n}) = \mathbb{1}(a_n > b_n),$$

or, after rearranging terms,

$$\delta_2(x_{1:n}) = \mathbb{1}(\bar{x} > \tfrac{1}{2} + \tfrac{b-a}{2n}).$$

Exercise 6

Yes, for my choice of δ_1 in Exercise 5, they are the same when $a = b$. Qualitatively speaking, the larger a is relative to b , the more the “balance” is tilted towards favoring $x_{n+1} = 1$ when using the Bayes procedure, compared to the intuitive procedure (and vice versa for b larger than a).

When the difference between a and b is sizable (relative to n), the procedure pays less attention to the data and more attention to the prior. On the other hand, as n grows, the influence of the prior becomes less and less since $\frac{b-a}{2n} \rightarrow 0$ as $n \rightarrow \infty$.

(Note: A useful and common way of thinking about a and b is as “pseudo-counts”, that is, as if we had observed a ones and b zeros before the experiment began.)

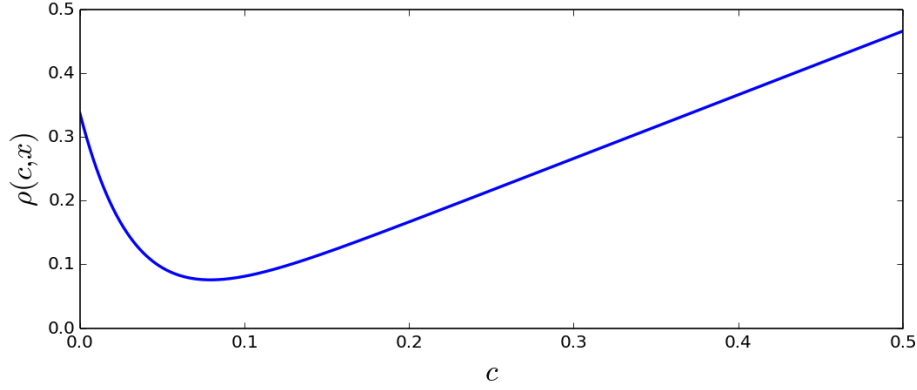


Figure 2: Posterior expected loss for the disease prevalence example.

Exercise 7

The mean of the Beta distribution $\text{Beta}(a, b)$ is $a/(a + b)$. Thus, the prior mean is $\mathbb{E}(\boldsymbol{\theta}) = a/(a + b)$. From Lecture 1, the posterior is $\text{Beta}(a_n, b_n)$ where $a_n = a + \sum x_i$ and $b_n = b + n - \sum x_i$. Hence, the posterior mean is

$$\begin{aligned} \mathbb{E}(\boldsymbol{\theta}|x_{1:n}) &= \frac{a_n}{a_n + b_n} \\ &= \frac{a + \sum x_i}{a + b + n} \\ &= \frac{a + b}{a + b + n} \frac{a}{a + b} + \frac{n}{a + b + n} \frac{1}{n} \sum x_i \\ &= (1 - t)\mathbb{E}(\boldsymbol{\theta}) + t\bar{x} \end{aligned}$$

with $t = n/(a + b + n)$.

Exercise 8

See Figure 2. See Appendix for source code.

Exercise 9

A canonical example of this is in medical diagnosis. Suppose S indicates whether a certain patient has a certain type of cancer ($S = 1$) or not ($S = 0$). A screening test is performed, yielding data x , and from this a decision is made as to whether the patient tests positive or not. If a positive diagnosis is made but the patient does not have cancer (“false positive”), the patient will unnecessarily suffer emotional distress, physical discomfort from further testing, and monetary costs. On the other hand, if a negative diagnosis is made but the patient actually does have cancer (“false negative”), she or he could die prematurely. Due to the asymmetry in these consequences, the 0 – 1 loss is typically not appropriate. A better loss function would be as follows:

		Diagnosis	
		0	1
Truth	0	0	α
	1	β	0

where $\alpha, \beta > 0$ but $\alpha \neq \beta$. (Note: This is purely hypothetical, and should not be construed as recommended practice.)

A Source code in Julia language

Exercise 2

```
using PyPlot
draw_now() = (pause(0.001); get_current_fig_manager()[:window][:raise_]())

# Data
x = [20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0]
n = length(x)

# Prior: Gamma(a,b)
a = 0.1
b = 1.0

# Gamma density
f(t,a,b) = exp((a-1)*log(t) - b*t + a*log(b) - lgamma(a))

# Make plots
figure(1); clf(); hold(true)
t = linspace(eps(),0.3,500)
# Plot prior
plot(t,f(t,a,b),label="prior, \p(\theta)",linewidth=2)
# Plot posterior
A = a + n
B = b + sum(x)
plot(t,f(t,A,B),label="posterior, \p(\theta|x_{1:n})",linewidth=2)
# Adjust plot settings
ylim(0,25)
xlabel("\theta",fontsize=20)
legend(loc=1,fontsize=18)
draw_now()
# Save figure
savefig("homework.png",dpi=120)
```

Exercise 8

```
using PyPlot
draw_now() = (pause(0.001); get_current_fig_manager()[:window][:raise_]())
```

```

# Beta density
f(t,a,b) = exp((a-1)*log(t) + (b-1)*log(1-t) - lbeta(a,b))
dt = 1/5000
t = [dt/2:dt:1-dt/2]

# Data
n = 30
z = 1

# Prior: Beta(a,b)
a = 0.05
b = 1.0

# Posterior: Beta(A,B)
A = a + z
B = b + n - z
p_t = f(t,A,B)

# Loss function
loss(theta,theta_hat) = abs(theta-theta_hat).*(1.0 + 9.0*(theta_hat.<theta))

# Posterior expected loss
# Compute using a Riemann sum approximation.
# Note: When z=0 this is not a good approximation since the density blows up
#       at 0.
rho(theta_hat) = sum(loss(t,theta_hat).*p_t*dt)

# Plot posterior expected loss
figure(1); clf(); hold(true)
rhos = [rho(th)::Float64 for th in t]
plot(t,rhos,linewidth=2)
xlim(0,0.5)
ylim(0,0.5)
xlabel("\$c\$",fontsize=20)
ylabel("\$\\rho(c,x)\$",fontsize=20)
draw_now()

# Find minimizer
println(t[indmin(rhos)])

# Save figure
savefig("rho.png",dpi=120)

```