STAT 510: Homework 09

David Dalpiaz

Due: Monday, April 18, 11:59 PM

Exercise 1 (Normal-Normal Model)

Assume:

- Likelihood: $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$
- Prior: $\theta \sim N(a, b^2)$
- σ^2 is a fixed and known quantity

Find the posterior distribution of $\theta \mid X_1, \ldots, X_n$.

Solution

First, for convenience we define

$$\sigma_n = \frac{\sigma}{\sqrt{n}}$$

Next we write the likelihood.

$$\mathcal{L}(\theta) \propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2\right) = \exp\left(-\frac{(\bar{x} - \theta)^2}{2\sigma_n^2}\right)$$

Now we write the prior density.

$$f(\theta) \propto \exp\left(-\frac{(\theta - a)^2}{2b^2}\right)$$

We then put these together to obtain the posterior density.

$$f(\theta \mid x_1, \ldots x_n) \propto \mathcal{L}(\theta) \cdot f(\theta) = \exp\left(-\frac{(\bar{x} - \theta)^2}{2\sigma_n^2} - \frac{(\theta - a)^2}{2b^2}\right) = \text{... magic ...}$$

$$f(\theta \mid x_1, \ldots x_n) \propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{1}{b^2}\right)\left[\theta - \frac{\frac{\bar{x}}{\sigma_n^2} + \frac{a}{b^2}}{\frac{1}{\sigma_n^2} + \frac{1}{b^2}}\right]^2\right)$$

Thus, we can conclude that

$$\theta \mid X_1, \ldots X_n \sim N\left(\bar{\theta}, \tau^2\right)$$

$$\bar{\theta} = w\bar{x} + (1 - w)a$$

$$w = \frac{\frac{1}{\sigma_n^2}}{\frac{1}{\sigma_n^2} + \frac{1}{b^2}}$$

$$\frac{1}{\tau^2} = \frac{1}{\sigma_n^2} + \frac{1}{b^2}$$

Exercise 2 (Gamma-Poisson Model)

Assume:

- Likelihood: $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$
- Prior: $\lambda \sim \text{Gamma}(\alpha, \beta)$

For this and other problems on this homework, use the following parameterization for the Gamma distribution:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Find the posterior distribution of $\lambda \mid X_1, \ldots, X_n$.

Solution

First, we note the likelihood.

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \propto \lambda^{\sum x_i} e^{-n\lambda}$$

We also write the prior's density.

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{\beta\lambda} \propto \lambda^{\alpha-1} e^{\beta\lambda}$$

We then put these together to obtain the posterior density.

$$f(\lambda \mid x_1, \ldots x_n) \propto \mathcal{L}(\lambda) \cdot f(\lambda) = \lambda^{\sum x_i} e^{-n\lambda} \lambda^{\alpha-1} e^{\beta\lambda} = \lambda^{\alpha-1+\sum x_i} e^{-(n+\beta)\lambda}$$

Thus, we can conclude that

$$\lambda \mid X_1, \ldots X_n \sim \text{Gamma}\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n\right)$$

Exercise 3 (Using the Beta-Bernoulli Model)

Assume:

- Likelihood: $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$
- Prior: $p \sim \text{Beta}(\alpha = 5, \beta = 5)$

Use the following data and the posterior mean to arrive at a Bayesian estimate of p. Compare this value of the prior mean.

```
some_data = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1)
```

Solution

For the text, we know the posterior distribution.

$$p \mid X_1, \ldots, X_n \sim \text{Beta}\left(\alpha + \sum_{i=1}^{n} x_i, \quad \beta + n - \sum_{i=1}^{n} x_i\right)$$

Thus the posterior mean is given by

$$\bar{p} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + n}.$$

Note that with this data, we have

$$\sum_{i=1}^{n} x_i = 5$$

with $n = 20$.

In this case we have

$$\bar{p} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + n} = \frac{5 + 5}{5 + 5 + 20} = \boxed{\frac{1}{3}}.$$

Compare this to the prior mean

$$p_0 = \frac{\alpha}{\alpha + \beta} = \frac{5}{5 + 5} = \boxed{\frac{1}{2}}.$$

Also note that the MLE for p is

$$\hat{p} = \frac{\sum x_i}{n} = \frac{5}{20} = 0.25.$$

The posterior is simply a mixture of the MLE and the prior mean.

$$\bar{p} = \left(\frac{n}{\alpha + \beta + n}\right)\hat{p} + \left(\frac{\alpha + \beta}{\alpha + \beta + n}\right)p_0 = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + n}$$

This can be verifyied for this example.

Exercise 4 (Using the Gamma-Poisson Model)

Given:

- Likelihood: $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$
- Prior: $\lambda \sim \text{Gamma}(\alpha = 4, \beta = 2)$

Use the following data and the posterior interval to arrive at a Bayesian interval estimate of λ. Compare this interval to an interval based on the prior distribution.

```
some_data = c(3,3,2,9,1,4,5,4,2,6,7,5,4,4,2,3,6,3,5,5,4,3,5,5,5)
```

Solution

```r
alpha_prior = 4
beta_prior = 2
```

Given our previous results we know the posterior also follows a Gamma distribution. We find the updated parameters in R.

```r
alpha_post = alpha_prior + sum(some_data)
beta_post = beta_prior + length(some_data)
```

The interval based on the posterior distribution is:

```r
qgamma(c(0.025, 0.975), shape = alpha_post, rate = beta_post)
```

```
## [1] 3.314827 4.829360
```

The interval based on the prior distribution is:

```r
qgamma(c(0.025, 0.975), shape = alpha_prior, rate = beta_prior)
```

```
## [1] 0.5449327 4.3836365
```

Exercise 5 (Using the Gamma-Poisson Model, Again)

Given:

- Likelihood: $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$
- Prior: $\lambda \sim \text{Gamma}(\alpha = 7.5, \beta = 1)$

Use the following data and the posterior distribution to calculate the posterior probabilities of the following hypotheses. Compare these probabilities to probabilities based only on the prior distribution.

$$H_0 : \lambda \leq 4 \quad \text{versus} \quad H_1 : \lambda > 4.$$

```r
some_data = c(3, 1, 1, 2, 2, 2, 1, 3, 3, 3, 3, 3, 1, 2, 3)
```

Solution

```r
alpha_prior = 7.5
beta_prior = 1
```

Given our previous results we know the posterior also follows a Gamma distribution. We find the updated parameters in R.

```r
alpha_post = alpha_prior + sum(some_data)
beta_post = beta_prior + length(some_data)
```

```r
# probability of null under posterior
pgamma(4, shape = alpha_post, rate = beta_post)
```

```
## [1] 0.9993185
```

```r
# probability of alternative under posterior
pgamma(4, shape = alpha_post, rate = beta_post, lower.tail = FALSE)
```

```
## [1] 0.000681492
```

```r
# probability of null under prior
pgamma(4, shape = alpha_prior, rate = beta_prior)
```

```
## [1] 0.0762173
```

```r
# probability of alternative under prior
pgamma(4, shape = alpha_prior, rate = beta_prior, lower.tail = FALSE)
```

```
## [1] 0.9237827
```

Notice that after observing the data, our belief has shifted towards the null.

Exercise 6 (Prior vs Data: Effect of Data)

Given:

- Likelihood: $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$
- Prior: $p \sim \text{Beta}(\alpha = 2, \beta = 2)$
- Data: `data_1`, `data_2`, `data_3`

```r
data_1 = c(0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1)
data_2 = c(0,1,0,1,0,0,0,1,0,0,1,0,0,0,1,1,0,1,0,0,1,0,0,0,0)
data_3 = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,1,0,1)
```

Create graphics that show:

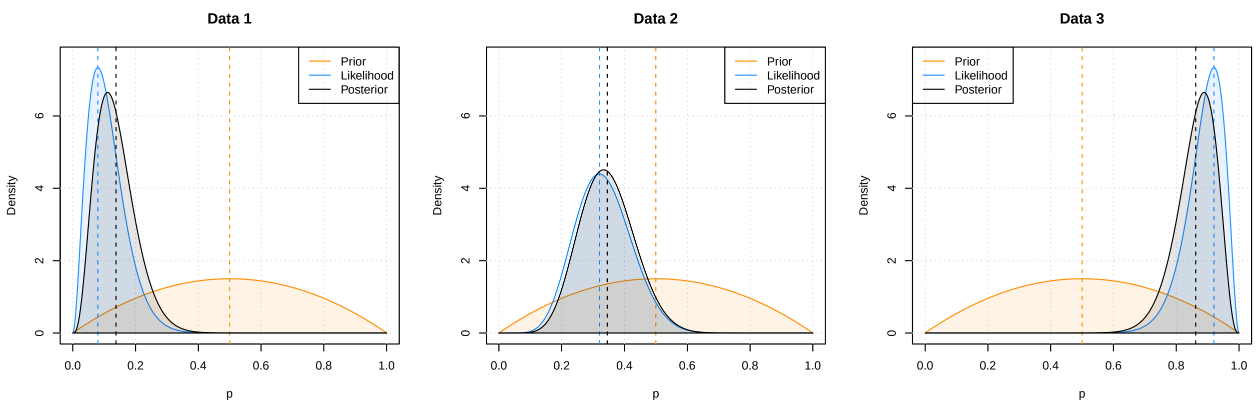
- The prior distribution and an estimate of p based on this distribution
- The likelihood and the MLE for each dataset
- The posterior and an estimate of p based on each of the datasets

Solution

```r
plot_beta_bern = function(alpha, beta, data, l_pos, title) {

  # summary statistics
  n = length(data)
  s = sum(data)

  # grid of values of p to consider
  p = seq(from = 0, to = 1, by = 0.001)

  # prior distribution and resulting estimate
  prior = dbeta(p, alpha, beta)
  prior_est = alpha / (alpha + beta)

  # likelihood and resulting estimate
  # note that we scale the likelihood to "integrate" to 1
  likelihood = dbinom(x = sum(data), size = n, prob = p)
  likelihood = likelihood / sum(0.001 * likelihood)
  mle = mean(data)

  # posterior parameters, distribution, and estimate (posterior mean)
  alpha_post = alpha + s
  beta_post = beta + n - s
  post = dbeta(p, alpha_post, beta_post)
  post_est = alpha_post / (alpha_post + beta_post)

  # setup blank plot
  plot(NA, main = title,
       xlim = c(0, 1), ylim = c(0, 7.6),
```

```r
      xlab = "p", ylab = "Density")

  # add background grid
  grid()

  # add prior distribution and estimate
  polygon(x = p, y = prior,
          col = adjustcolor("darkorange", 0.1), border = "darkorange")
  abline(v = prior_est, col = "darkorange", lty = 2)

  # add likelihood and mle
  polygon(x = p, y = likelihood,
          col = adjustcolor("dodgerblue", 0.1), border = "dodgerblue")
  abline(v = mle, col = "dodgerblue", lty = 2)

  # add posterior and estimate
  polygon(x = p,
            y = post,
            col = rgb(0, 0, 0, 0.1))
  abline(v = post_est, lty = 2)

  # add legend
  legend(l_pos, lwd = 1,
         legend = c("Prior", "Likelihood", "Posterior"),
         col = c("darkorange", "dodgerblue", "black"))

}
```

```r
par(mfrow = c(1, 3))
plot_beta_bern(alpha = 2, beta = 2, data_1, "topright", "Data 1")
plot_beta_bern(alpha = 2, beta = 2, data_2, "topright", "Data 2")
plot_beta_bern(alpha = 2, beta = 2, data_3, "topleft",  "Data 3")
```



Exercise 7 (Prior vs Data: Effect of Prior)

Given:

- Likelihood: $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$
- Prior 1: $p \sim \text{Beta}(\alpha = 2, \beta = 5)$
- Prior 2: $p \sim \text{Beta}(\alpha = 2, \beta = 2)$
- Prior 3: $p \sim \text{Beta}(\alpha = 5, \beta = 2)$

- Data: `some_data`

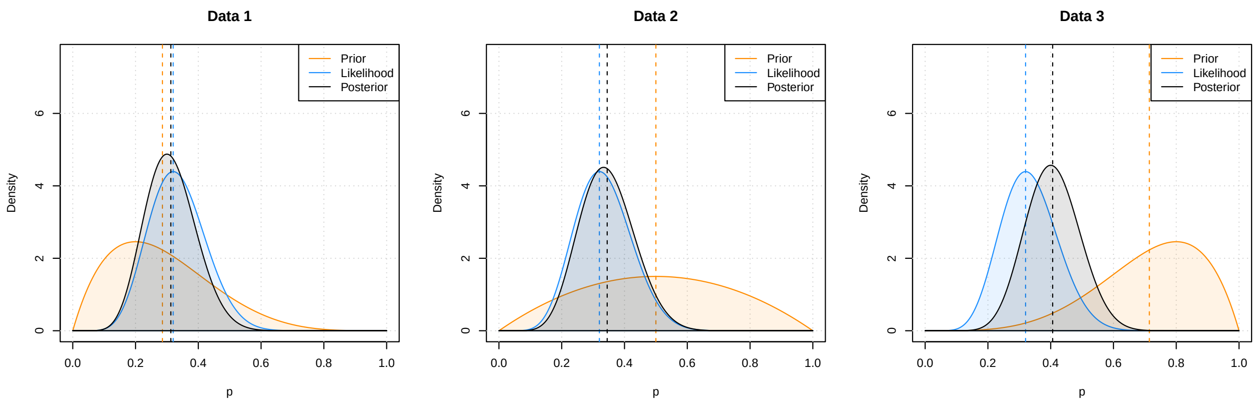
```
some_data = c(0,1,0,1,0,0,0,1,0,0,1,0,0,0,1,1,0,1,0,0,1,0,0,0,0)
```

Create graphics that show:

- The prior distribution and an estimate of p based on each prior
- The likelihood and the MLE given the data
- The posterior and an estimate of p based on each of the priors

Solution

```
par(mfrow = c(1, 3))
plot_beta_bern(alpha = 2, beta = 5, some_data, "topright", "Data 1")
plot_beta_bern(alpha = 2, beta = 2, some_data, "topright", "Data 2")
plot_beta_bern(alpha = 5, beta = 2, some_data, "topright", "Data 3")
```



Exercise 8 (Prior vs Data: Strength of Prior)

Given:

- Likelihood: $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$
- Prior 1: $p \sim \text{Beta}(\alpha = 2, \beta = 2)$
- Prior 2: $p \sim \text{Beta}(\alpha = 5, \beta = 5)$
- Prior 3: $p \sim \text{Beta}(\alpha = 10, \beta = 10)$
- Data: `some_data`

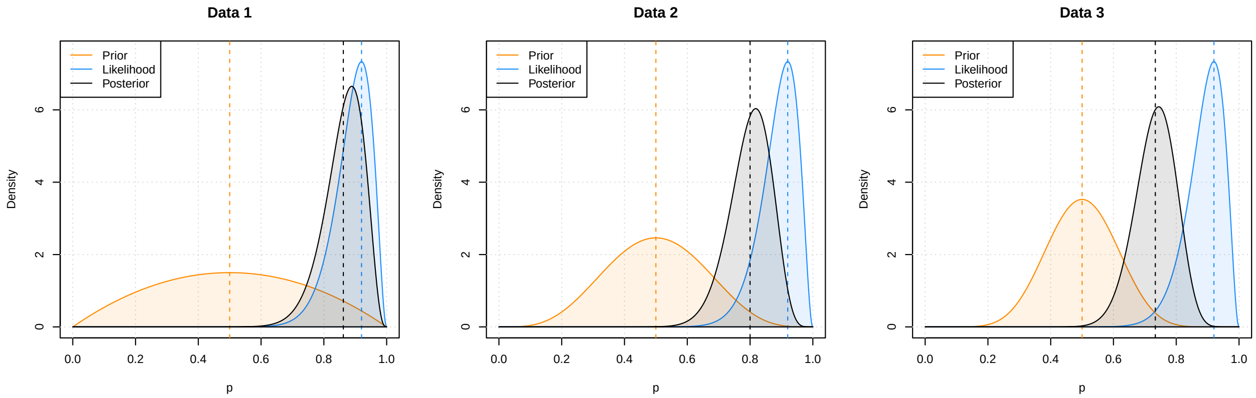
```
some_data = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,0,1,1,1,1,1)
```

Create graphics that show:

- The prior distribution and an estimate of p based on each prior
- The likelihood and the MLE given the data
- The posterior and an estimate of p based on each of the priors

Solution

```
par(mfrow = c(1, 3))
plot_beta_bern(alpha = 2,  beta = 2,  some_data, "topleft", "Data 1")
plot_beta_bern(alpha = 5,  beta = 5,  some_data, "topleft", "Data 2")
plot_beta_bern(alpha = 10, beta = 10, some_data, "topleft", "Data 3")
```



Exercise 9 (Bayes Risk in the Beta-Bernoulli Model)

Suppose $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ and $p \sim \text{Beta}(\alpha, \beta)$. Using squared error loss, find the Bayes estimator and the Bayes risk.

Solution

We have shown previously that

$$p \mid x_1, \ldots x_n \sim \text{Beta}\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i\right)$$

Under squared error loss, the Bayes estimator is the posterior mean.

$$\hat{p} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + n}$$

Under squared error loss the risk is simply the MSE.

$$R\left(p, \hat{p}\right) = \mathbb{E}\left[(p - \hat{p})^2\right]$$

Note that the MSE decomposes into the bias-squared plus variance.

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

Could we simply this quantity? Sure. Are we going to? Nope.

Finally, the Bayes risk is

$$r(f, \hat{p}) = \int R\left(p, \hat{p}\right) f(p) dp = \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)(\alpha + \beta + n)}$$

Note that calculus as hard, and this was solved with a a computer. For the sake of grading, you just need to setup this integral. Here $f(p)$ is a Beta density.

Exercise 10 (The James-Stein Estimator)

Consider $X_1, \ldots, X_k \sim N(\theta_i, 1)$. Define $\theta = (\theta_1, \ldots, \theta_k)$. Consider the loss

$$L\left(\theta, \hat{\theta}\right) = \sum_{j=1}^{k} (\theta_j - \hat{\theta}_j)^2.$$

where $\hat{\theta}$ is some estimator of θ.

Use simulation to compare the risk of the MLE to the James-Stein estimator. Consider at least three simulation setups:

- $k = 2$
- a relative "large" k and a dense θ vector
- a relative "large" k and a sparse θ vector

You are free to further specify k and θ as you wish. You are also free to add additional setups. Summarize your findings.

Solution

```
mse = function(actual, estimated) {
  sum((actual - estimated) ^ 2)
}
```

```
js = function(x) {
  k = length(x)
  max(1 - (k - 2) / sum((x) ^ 2), 0) * x
}
```

First, a quick note about $k = 2$.

```
theta = c(4, 2)
k = length(theta)
x = rnorm(n = k, mean = theta)
js(x)
```

```
## [1] 4.166445 2.619372
```

```
x
```

```
## [1] 4.166445 2.619372
```

When $k = 2$, there is no difference between the James-Stein estimator and the MLE. Thus, we can short circuit this setup and skip the simulation. The results will be the same.

Next, we'll move to a "large" value of k, where "large" will be $k = 200$. First, with a dense θ vector.

```
sim_and_calc_risk = function(theta) {
  x = rnorm(n = k, mean = theta)
  c(mse(theta, js(x)), mse(theta, x))
}
```

```
k = 200
set.seed(42)
theta = runif(n = k, min = 1, max = 10)
set.seed(42)
loss = replicate(n = 10000, sim_and_calc_risk(theta))
rowMeans(loss)
```

```
## [1] 195.2983 200.1476
```

Now, we keep the larger k, but with a sparse θ vector.

```r
k = 200
set.seed(42)
theta = sample(c(0, 1), size = k, replace = TRUE, prob = c(90, 10))
set.seed(42)
loss = replicate(n = 10000, sim_and_calc_risk(theta))
rowMeans(loss)
```

```
## [1]  24.46327 200.14758
```

As expected, in both cases the JS estimator outperforms the MLE. The performance gain is much greater given a sparse θ.

Exercise 11 (Free Points)

The previous two problems were pretty difficult. Draw a smiley face for a free point!

Solution

:)