Homework 1 — Bayesian Data Analysis

Kári Hlynsson

kah76@hi.is

Exercise 1: Sample survey

Suppose we are going to sample 100 individuals from a country (with a population size much larger than 100) and ask each sampled person whether they support policy Z or not. Let $Y_i = 1$ if person i in the sample supports the policy, and $Y_i = 0$ otherwise.

(a) Assume $Y_1, \ldots, Y_{100}$ are, conditional on θ, i.i.d. binary random variables with expectation θ. Write down the joint distribution of $\Pr(Y_1 = y_1, \ldots, Y_{100} = y_{100})$ in compact form. Also write down the form $\Pr\left(\sum_{i=1}^{100} Y_i = y\right)$.

(b) For the moment, suppose you believed that $\theta \in \{0.0, 0.1, \ldots, 0.9, 1.0\}$. Given the results of the survey were $\sum_{i=1}^{100} Y_i = 73$, compute $\Pr\left(\sum_{i=1}^{100} Y_i = 73 \,\middle|\, \theta\right)$ for each of these 11 values of θ and plot these probabilites as a function of θ (point mass at each value of θ).

(c) Now suppose you originally had no prior information to believe one of these θ-values over another, and thus $\Pr(\theta = 0.0) = \Pr(\theta = 0.1) = \cdots = \Pr(\theta = 1.0) = \frac{1}{11}$. Use Bayes' rule to compute $p\left(\theta \,\middle|\, \sum_{i=1}^{100} Y_i = 73\right)$ for each θ-value. Make a plot of this posterior distribution as a function of θ (point mass at each value of θ).

(d) Now, suppose you allow θ to be any value in the interval $[0, 1]$. Using the uniform prior density for θ, namely, $\pi(\theta) = 1$, derive and plot the posterior density of θ as a function θ. According to the posterior density, what is the probability of $\theta > 0.8$?

(e) Why are the heights of posterior densities in (c) and (d) not the same?

Solution

Part (a)

Set $\mathbf{Y} := (Y_1, \ldots, Y_n)$ where $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$ and $\mathbf{y} := (y_1, \ldots, y_n) \in \{0, 1\}^n$. Assuming $\sum_{i=1}^n y_i = k$, i.e. that k of the n responses are positive, we wish to derive $\Pr(\mathbf{Y} = \mathbf{y})$. Since all of the Y_i's are i.i.d., we have

$$\Pr(\mathbf{Y} = \mathbf{y}) = \prod_{i=1}^n \Pr(Y_i = y_i) = \theta^k (1-\theta)^{n-k}. \tag{$*$}$$

We also wish to derive $\Pr\left(\sum_{i=1}^n Y_i = k\right)$. The probability of obtaining any arrangement where k of the responses are positive is identical to $(*)$. Therefore, we must account for the number of possible configurations of responses where k are positive. This is equal to the binomial coefficient $\binom{n}{k}$, so we arrive at

$$\Pr\left(\sum_{i=1}^n Y_i = k\right) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \tag{$**$}$$

which is consistent with theory, since summating over all possible values of k yields

$$\sum_{k=0}^n \binom{n}{k} \theta^k (1-\theta)^{n-k} = (\theta + 1 - \theta)^n = 1^n = 1,$$

where we second equation was obtained by the binomial theorem. The probability distribution specified by $(**)$ is called a *binomial distribution*, denoted $\text{Binomial}(n, \theta)$. $\square$

Part (b)

Assign $Y := \sum_{i=1}^{n} Y_i$. As demonstrated in (a), $Y \sim \text{Binomial}(n, \theta)$. The probabilities of observing 73 positive responses from a sample of 100 ranging over $\theta \in \{0.0, 0.1, \ldots, 0.9, 1.0\}$ can be evaluated in R by

```
> theta <- seq(0, 1, by = 0.1)
> probs <- dbinom(x = 73, size = 100, prob = theta)
> print(probs)
 [1] 0.000000e+00 1.114936e-50 4.378461e-30 8.515317e-19 1.750513e-11
 [6] 1.512525e-06 2.204769e-03 7.196692e-02 2.168109e-02 8.758007e-07
[11] 0.000000e+00
```

which we can visualise by

```
> plot(theta, probs, type = "l", xlab = TeX("\\theta"), ylab = "Density")
```
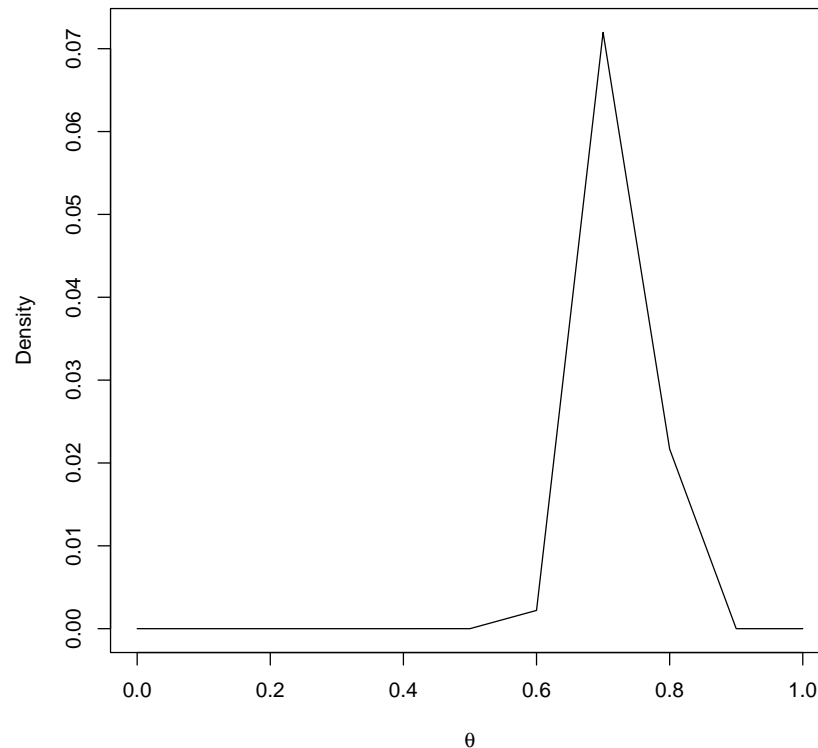

Figure 1: Probabilities of $Y = 73$ for different values of θ, $n = 100$

As shown in the figure, the most probable value is $\theta = 0.7$.

Part (c)

Under the assumption that the prior distribution $\pi(\theta)$ is uniform over its support $\mathcal{S} = \{0.0, 0.1, \ldots, 0.9, 1.0\}$ and that the likelihood function $\mathcal{L}(\theta \mid x)$ for the data is the binomial p.d.f., the posterior distribution becomes

$$p(\theta \mid Y) = \frac{\mathcal{L}(Y \mid \theta)\pi(\theta)}{M(Y)} = \frac{\mathcal{L}(Y \mid \theta)\pi(\theta)}{\sum_{\theta_i \in \mathcal{S}} \mathcal{L}(Y \mid \theta_i)\pi(\theta_i)} = \frac{\binom{n}{k}\theta^k(1-\theta)^{n-k}}{\sum_{\theta_i \in \mathcal{S}} \binom{n}{k}\theta_i^k(1-\theta_i)^{n-k}} = \frac{\theta^k(1-\theta)^{n-k}}{\sum_{\theta_i \in \mathcal{S}} \theta_i^k(1-\theta_i)^{n-k}},$$

so that when $n = 100$, $k = 73$, we have

$$p(\theta \mid Y) = \frac{\theta^{73}(1-\theta)^{27}}{\sum_{\theta_i \in \mathcal{S}} \theta_i^{73}(1-\theta_i)^{27}}.$$

We evaluate the posterior density in R by

```
> post0 <- (theta^73 * (1 - theta)^27)/(sum(theta^73 * (1 - theta)^27))
> print(post0)
 [1] 0.000000e+00 1.163146e-49 4.567788e-29 8.883524e-18 1.826206e-10
 [6] 1.577927e-05 2.300105e-02 7.507881e-01 2.261859e-01 9.136709e-06
[11] 0.000000e+00
```

and visualise it with

```
> plot(theta, post0, type = "l", xlab = TeX("\\theta"), ylab = "Posterior Density")
```
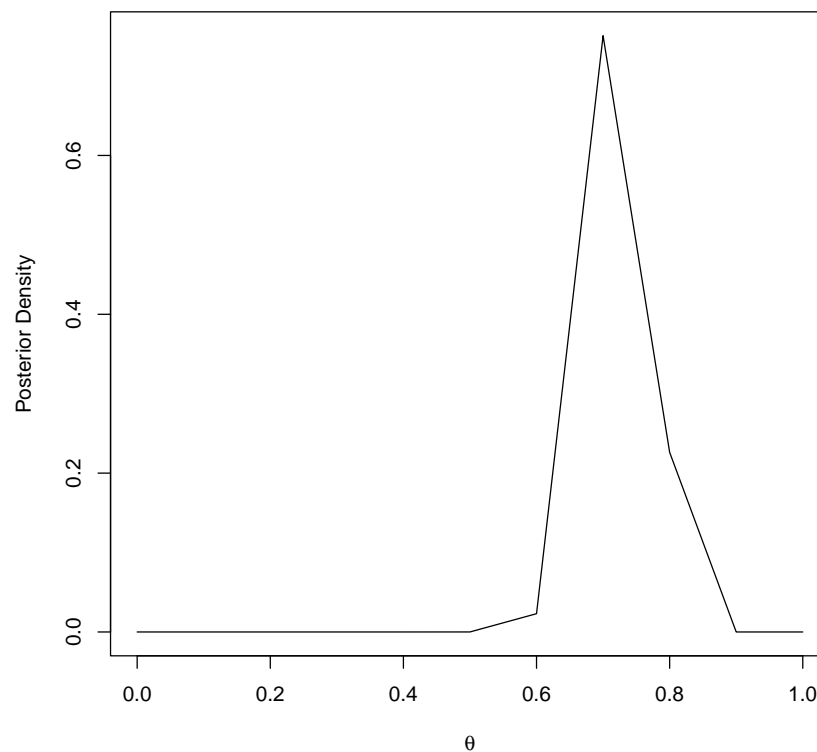

Figure 2: Posterior density of Y assuming $\pi(\theta) \sim \text{Uniform}\{0.0, 0.1, \ldots, 0.9, 1.0\}$

Part (d)

Using the same likelihood function and reassigning $\pi(\theta) = 1$, we obtain the updated posterior allowing for all values of $\theta \in [0, 1]$:

$$p(\theta \mid Y) = \frac{\theta^{73}(1 - \theta)^{27}}{\int_{\mathcal{S}} \theta_i^{73}(1 - \theta_i)^{27}\mathrm{d}\theta_i}$$

We see that the marginal distribution of Y resembles a beta distributed r.v. with $\alpha = 74$, $\beta = 28$. Lengthening by the coefficient of the beta distribution with the specified parameters, we obtain

$$p(\theta \mid Y) = \frac{\frac{\Gamma(102)}{\Gamma(74)\Gamma(28)}\theta^{73}(1 - \theta)^{27}}{\int_{\mathcal{S}} \frac{\Gamma(102)}{\Gamma(74)\Gamma(28)}\theta_i^{73}(1 - \theta_i)^{27}\mathrm{d}\theta_i} = \frac{1}{B(74, 28)}\theta^{73}(1 - \theta)^{27}$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function. Thus we have shown that the posterior follows a beta distribution with parameters $\alpha = 74$ and $\beta = 28$. We evaluate the posterior density with `dbeta()`:

```
> theta <- seq(0, 1, length.out = 1000)
> post1 <- dbeta(x = theta, shape1 = 74, shape2 = 28)
```

and visualise them by
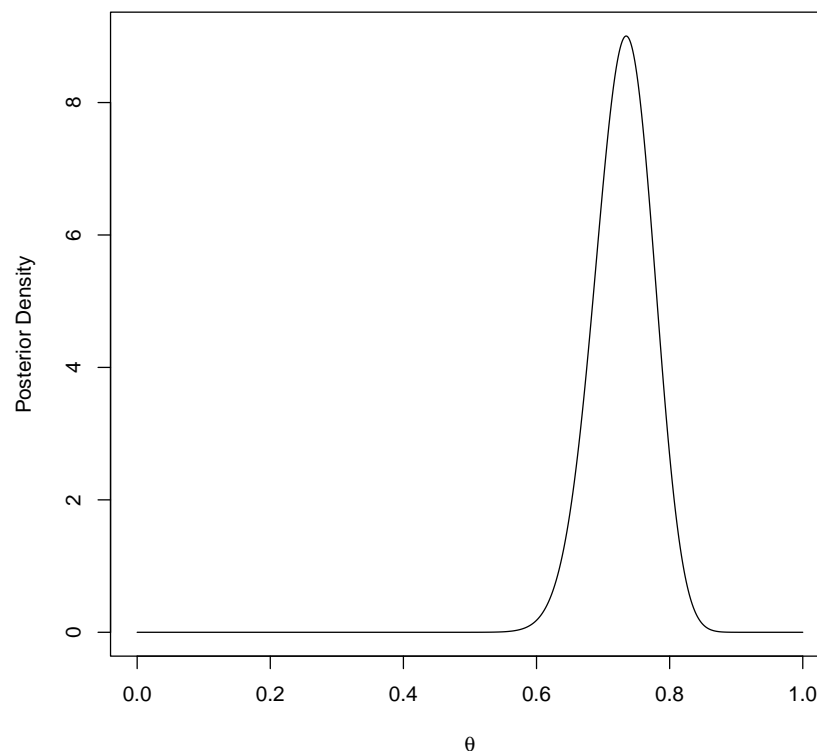
```
> plot(theta, post1, type = "l", xlab = TeX("\\theta"), ylab = "Posterior Density")
```


Figure 3: Posterior density of Y given the prior density $\pi(\theta) = 1$, $\theta \in [0, 1]$

Figure 3 displays a similar distribution to the discrete case where the centre of the distribution has shifted

slightly to the right. The maximum density is attained at $\theta = 0.7297297$ where $p(\theta \mid Y) = 9.044944$, obtained by the R code snippet below:

```
> theta[which(post1 == max(post1))]
[1] 0.7297297
> max(post1)
[1] 9.044944
```

To compute the probability $\Pr(\theta > 0.8)$, we use the `dbeta()` command in R:

```
> pbeta(q = 0.8, shape1 = 74, shape2 = 28, lower.tail = FALSE)
[1] 0.03848785
```

which turns out to be approximately 4%.

Part (e)

Comparing Figures 2 and 3, we see that the distributions are similar although Figure 2 shows a discrete mass function whereas the distribution in Figure 3 is a density. However, the most prominent difference is in the magnitude between these two distributions. This makes sense since we integrate over the distribution in (d) and summate over values of (c). As can be seen below, both turn out to equal 1:

```
> # (c) Sum over the probabilities
> sum(post0)
[1] 1
> # (d) Evaluate the integral of 1/B(74,28) * t^74 * (1 - t)^28 dt
> post1f <- function(t) dbeta(x = t, shape1 = 74, shape = 28)
> integrate(post1f, lower = 0, upper = 1)$value
[1] 1
```

The intuitive explanation is that while we summate over the values in the discrete case (c), such that the probability mass is reflected in one-to-one proportion in the sum, we integrate in (d), which can be viewed as the limit of the infinite series

$$\int_{\mathcal{S}} p(\theta \mid Y)\mathrm{d}\theta = \lim_{\ell(P_i) \to 0} \sum_{P_i \in \mathcal{P}} \ell(P_i)p(\theta = \theta_i \mid Y) \qquad\qquad \text{where } \theta_i \in \overset{\circ}{P_i}$$

where $\mathcal{P} = \{P_i\}_{i=1}^n$ constitutes a partition on the support $\mathcal{S}$, $\ell(P_i)$ yields the length of the interval P_i and θ_i is an arbitrary point on the interior $\overset{\circ}{P_i}$ of P_i. This all goes to say that in the case (d) where we integrate over the posterior density, the densities do not enter the "sum" in one-to-one proportion, so the initial conception that the total density exceeds 1 is misfounded. $\qquad\square$

Exercise 2: Random numbers, probability density functions (pdfs) and cumulative density functions (cdfs)

The goal of this exercise is to generate random numbers, plot the histogram, the empirical pdf and cdf for these numbers, and see how they compare to the theoretical pdf and cdf. The goal is also to compare the sample mean and standard deviation to the theoretical mean and standard deviation.

(a) Generate $B = 3000$ numbers from the gamma distribution with parameters $\alpha = 2$, $\beta = 0.1$ (R:rgamma). Compute the sample mean and the sample standard deviation and compare to the theoretical mean and standard deviation (R:mean,sd).

(b) Plot the theoretical density (pdf) of the gamma distribution (R:dgamma). Plot empirical density based on the data on the same graph. In R one can use

```
library(lattice)
densityplot(y)
```

given that y are the data. Plot the histogram on another graph (R:histogram).

(c) Plot the theoretical cumulative density function (cdf) of the gamma distribution (R:pgamma). Plot empirical cumulative density based on the data on the same graph. In R one can use

```
n <- length(y)
y_sort <- sort(y)
plot(y_sort, (1:n)/n, type = "s", ylim = c(0,1))
```

where again y are the data.

Solution

Part (a)

Start with the random number generation, with the RNG seed set to 42 for reproducibility:

```
> y <- rgamma(n = 3000, shape = 2, rate = 0.1)
```

Compare theoretical and sample moments:

```
> # Theoretical mean
> 2 / 0.1
[1] 20
> # Sample mean
> mean(y)
[1] 19.87275
> # Theoretical variance
> 2 / 0.1^2
[1] 200
> # Sample variance
> var(y)
[1] 202.2949
```

We see that the theoretical and sample moments are very similar, with a difference of 0.12725 between the theoretical and sample mean, and a difference of 2.29486 between the theoretical and sample mean. In fact, as n increases, we would expect these estimates to become even better by the law of large numbers.

Part (b)

The following code snippet was used to generate the plot below in Figure 4:

```
> t <- seq(min(y), max(y), length.out = 3000)
> ft <- dgamma(x = t, shape = 2, rate = 0.1)
>
> par(mfrow = c(1, 2))
> plot(density(y), main = NA, xlab = "Y", col = "blue", ylim = range(ft, density(y)$y))
> lines(t, ft, col = "red")
> legend("topright", inset = c(0.05, 0.05), legend = c("Theoretical", "Sample"),
+ col = c("blue", "red"), lty = 1)
> hist(y, main = NA, xlab = "Y")
```
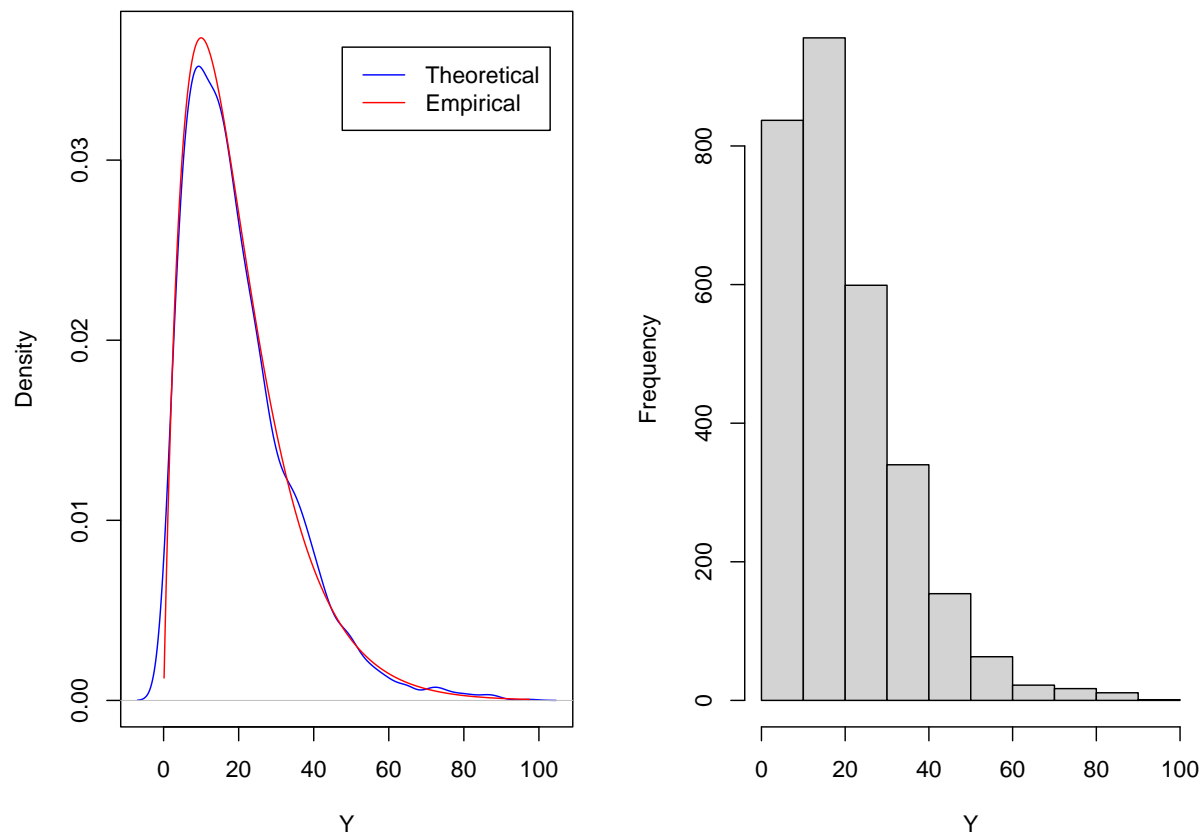

Figure 4: Comparisonal plot of the theoretical and empirical densities

Part (c)

The code below was used to create the plot shown in Figure 5:

```
> n <- length(y)
> ecdf <- sort(y)
> tcdf <- pgamma(q = t, shape = 2, rate = 0.1)
>
> plot(ecdf, 1:n/n, type = "s", ylim = c(0, 1), col = "blue", xlab = "Y",
+ ylab = "Cumulative Density")
> lines(t, tcdf, col = "red")
> legend("topright", inset = c(0.05, 0.05), legend c("Theoretical", "Empirical"),
+ col = c("blue", "red"), lty = 1)
```
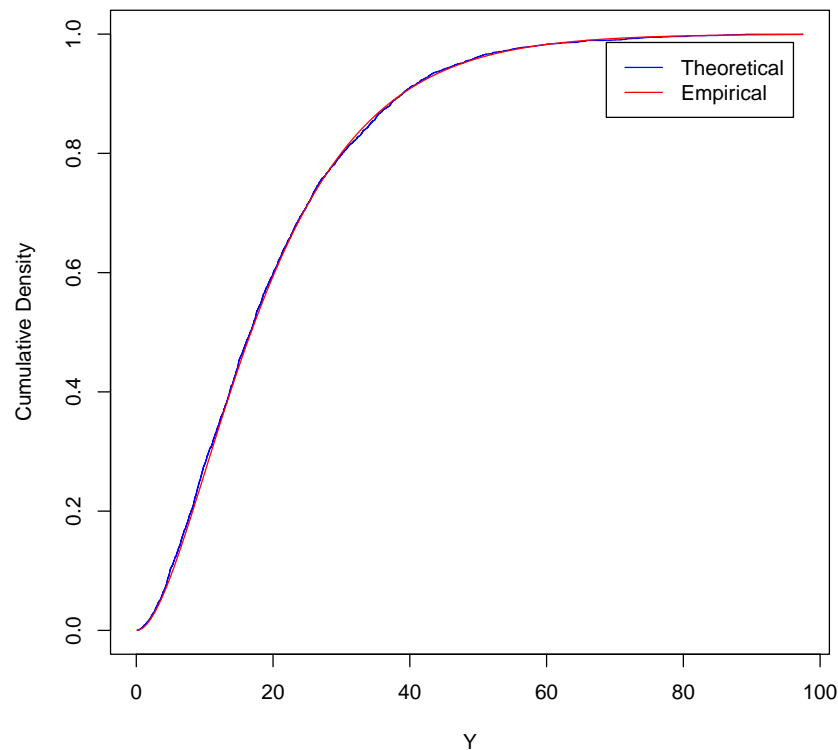

Figure 5: Comparison of the theoretical and empirical cumulative density functions.

We see that in both (b) and (c), the distributions match closely, which is consistent since the sample vector $\mathbf{y} = \{y_1, \ldots, y_{3000}\}$ was drawn from the theoretical distribution and should therefore resemble it with increasing precision as B grows larger.