Lab 3

Quynh Vu

2023-01-30

Question 1: Consider the happiness example from the lecture, with 118 out of 129 women indicating they are happy. We are interested in estimating θ, which is the (true) proportion of women who are happy. Calculate the MLE estimate $\hat{\theta}$ and 95% confidence interval.

Let $Y_i \sim$ i.i.d. $Bernoulli(\theta)$ and $Y = \sum_{i=1}^{129} Y_i \sim Binom(129, \theta)$

The likelihood function is

$$L(\mathbf{y}; \theta) = \prod_{i=1}^{129} \binom{129}{y_i} \theta^{y_i} (1-\theta)^{129-y_i} = \left[\prod_{i=1}^{129} \binom{129}{y_i} \right] \theta^{\sum_{i=1}^{129} y_i} (1-\theta)^{129 - \sum_{i=1}^{129} y_i}$$

The log-likelihood function is

$$l(\mathbf{y}; \theta) = \sum_{i=1}^{129} log\binom{129}{y_i} + \left(\sum_{i=1}^{129} y_i \right) log(\theta) + \left(129 - \sum_{i=1}^{129} y_i \right) log(1-\theta)$$

Taking derivative of $l(\mathbf{y}; \theta)$ with respect to θ yields

$$\frac{dl(\mathbf{y}; \theta)}{d\theta} = \frac{\sum_{i=1}^{129} y_i}{\theta} - \frac{129 - \sum_{i=1}^{129} y_i}{1-\theta}$$

Setting $\frac{dl(\mathbf{y}; \theta)}{d\theta} = 0$ yields

$$\hat{\theta}_{mle} = \overline{Y} = \frac{\sum_{i=1}^{129} y_i}{129} = \frac{118}{129} \approx 0.915$$

Since $\hat{\theta}_{mle} = \overline{Y}$, the Central Limit Theorem gives $\frac{\hat{\theta}_{mle} - 129\theta}{\sqrt{\frac{\theta(1-\theta)}{129}}} \xrightarrow{D} N(0,1)$ as n $\to \infty$, i.e.

$$\hat{\theta}_{mle} \sim N\left(129\theta, \frac{\theta(1-\theta)}{129} \right)$$

```
qnorm(0.975, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
```

```
## [1] 1.959964
```

The 95% confidence interval of $\hat{\theta}_{mle}$ is

$$\hat{\theta}_{mle} \pm 1.96\sqrt{\frac{\hat{\theta}_{mle}(1-\hat{\theta}_{mle})}{129}} = \left(0.915 - 1.96\sqrt{\frac{0.915(1-0.915)}{129}}, 0.915 + 1.96\sqrt{\frac{0.915(1-0.915)}{129}} \right) \approx (0.87, 0.96)$$

Question 2: Assume a Beta(1,1) prior on θ. Calculate the posterior mean for $\hat{\theta}$ and 95% credible interval.

For any $Beta(\alpha, \beta)$ distribution, the posterior distribution for θ is such that

$$f(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta)\pi(\theta) = \theta^{118}(1-\theta)^{129-118}\theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{(118+\alpha)-1}(1-\theta)^{(11+\beta)-1}$$

so the posterior distribution is $\theta|\mathbf{y} \sim Beta(118 + \alpha, 11 + \beta)$ and the posterior mean for $\hat{\theta}$ is

$$E(\theta|\mathbf{y}) = \tfrac{118+\alpha}{118+\alpha+11+\beta} = \tfrac{118+\alpha}{129+\alpha+\beta} = \left(\tfrac{129}{129+\alpha+\beta}\right)\underbrace{\left(\tfrac{118}{129}\right)}_{\hat{\theta}_{mle}} + \left(\tfrac{\alpha+\beta}{129+\alpha+\beta}\right)\underbrace{\left(\tfrac{\alpha}{\alpha+\beta}\right)}_{E(\theta)} \;(*)$$

For $\theta \sim Beta(1,1)$, the posterior distribution is $Beta(119,12)$ and

- $E(\theta|\mathbf{y}) = \frac{119}{131} \approx 0.91$
- $Var(\theta|\mathbf{y}) = \frac{119\times12}{(119+12)^2(119+12+1)} \approx 0.0006303934$

The 95% credible interval for $\hat{\theta}$ under $Beta(1,1)$ is

$$E(\theta|\mathbf{y}) \pm 1.96\sqrt{Var(\theta|\mathbf{y})} = \big(0.85, 0.95\big)$$

Question 3: Now assume a Beta(10,10) prior on θ. What is the interpretation of this prior? Are we assuming we know more, less or the same amount of information as the prior used in Question 2?

Similar to $Beta(1,1)$ prior, $Beta(10,10)$ prior assigns equal weights to the success and failure of the Bernouli random variables Y_i's. As seen above, the Bayesian point estimate of θ, $\hat{\theta}_B = E(\theta|\mathbf{y})$, is a weighted average of the MLE $\hat{\theta}_{mle} = \overline{Y}$ of θ and the prior mean $E(\theta)$ with the respective weights $w_{mle} = \frac{129}{129+\alpha+\beta}$ and $w_\theta = \frac{\alpha+\beta}{129+\alpha+\beta}$.

Since $\hat{\theta}_{mle}$ is a consistent estimator of θ by the Weak Law of Large Number (i.e. $\hat{\theta}_{mle} \xrightarrow{P} \theta$), the Bayes estimate $\hat{\theta}_B$ will be closer to the MLE of θ for larger n. From $(*)$ we see that the sum $\alpha + \beta$ indicates the worth of the prior information relative to a sample of size $n = 129$. In particular, even though $E(\theta) = \frac{1}{2}$ for both $Beta(1,1)$ and $Beta(10,10)$, the latter assigns a bigger weight to the prior opinion.

- Under $Beta(1,1)$ prior, $\hat{\theta}_B = E(\theta|\mathbf{y}) = 0.985\hat{\theta}_{mle} + 0.015E(\theta)$
- Under $Beta(10,10)$ prior, $\hat{\theta}_B = E(\theta|\mathbf{y}) = 0.866\hat{\theta}_{mle} + 0.134E(\theta)$

We are assuming more prior information available under $Beta(10,10)$ and stressing less focus on data.

Question 4: Create a graph in ggplot which illustrates

a) The likelihood (easiest option is probably to use `geom_histogram` to plot the histogram of appropriate random variables)

b) The priors and posteriors in question 2 and 3 (use `stat_function` to plot these distributions). Comment on what you observe.

Similar to the likelihood curve, the posterior density under $Beta(1,1)$ also peaks at the MLE $\hat{\theta} = \overline{Y}$. When the prior is not flat as in the case of $Beta(10,10)$, the peak of the posterior density curve is between that of the corresponding prior and the likelihood curves. This is because the posterior mean is the weighted average of the prior mean and the data. Also, since $w_{mle} > w_{theta}$, the peak of the posterior density curve under $Beta(10,10)$ prior is closer to that of the likelihood curve. For the two priors, the posterior has smaller variance than the prior (i.e. the spread of prior curves is larger), implying data give us more information about where θ lies in its range.

```
n = 129; y = 118; theta = seq(0, 1, by = 0.001)
BinomLikelihood <- function(y,n){
    L = function(theta) dbinom(y,n,theta)
    L(theta)
```

```r
}
L_theta <- BinomLikelihood(y, n) ## Likelihood
prior_beta11 <- function(theta) {
    dbeta(theta, 1, 1, ncp = 0)
}
posterior_beta11 <- function(theta) {
    dbeta(theta, 119, 12, ncp = 0)
}
prior_beta1010 <- function(theta) {
    dbeta(theta, 10, 10, ncp = 0)
}
posterior_beta1010 <- function(theta) {
    dbeta(theta, 128, 21, ncp = 0)
}
data <- data.frame(theta, L_theta)
theme_set(theme_minimal())
ggplot(data, aes(x=theta)) +
    geom_line(aes(y = L_theta), color = "darkred", ) +
    stat_function(fun = prior_beta11, geom = "line", color = "steelblue") +
    stat_function(fun = posterior_beta11, geom = "line", color = "darkgoldenrod") +
    stat_function(fun = prior_beta1010, geom = "line", color = "blueviolet") +
    stat_function(fun = posterior_beta1010, geom = "line", color = "darkolivegreen3") +
    geom_vline(xintercept = y/n, linetype="dashed", color = "cyan", size=1)+
    ylab(expression(L(theta) ~ and ~ pi(theta) ~ and ~ f(theta|y))) +
    xlab(expression(theta)) +
    labs(title = "Beta prior and Binomial distribution")
```
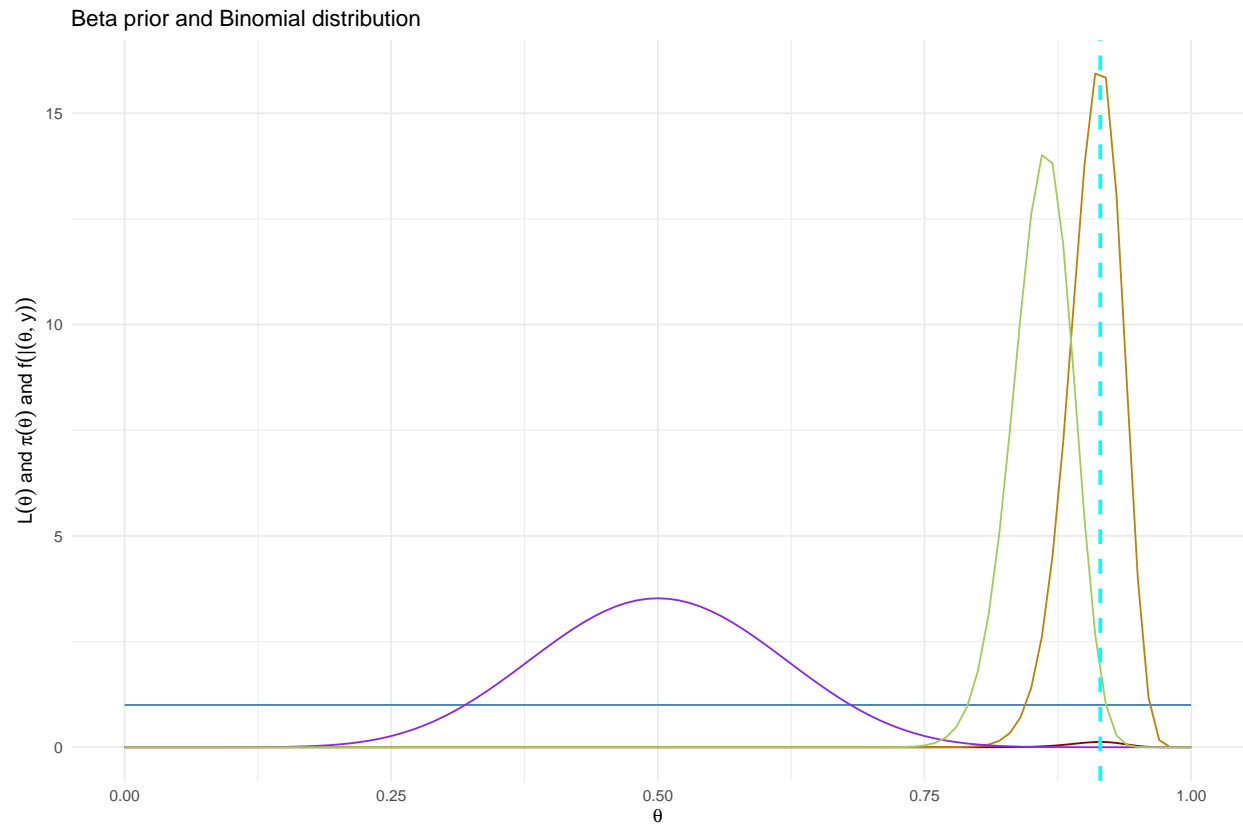

Beta prior and Binomial distribution

Question 5 (No R code required): A study is performed to estimate the effect of a simple training program on basketball free-throw shooting. A random sample of 100 college students is recruited into the study. Each student first shoots 100 free-throws to establish a baseline success probability. Each student then takes 50 practice shots each day for a month. At the end of that time, each student takes 100 shots for a final measurement. Let θ be the average improvement in success probability. θ is measured as the final proportion of shots made minus the initial proportion of shots made. Given two prior distributions for θ (explaining each in a sentence) based on your best knowledge:

The non-informative prior distribution is chosen to provide very limited information relative to the experiment (e.g. assuming the improvement in success probability of each recruited student is "flat" or equally likely over time), so the data can "speak for themselves" and take the primary role in the posterior distribution and justify whether the sample is a good representative of the whole population.

The subjective (or informative) prior, on the other hand, is chosen based on belief or available information from literature review or previous experiments on (relatively) similar sampled populations in order to provide as much information as possible to the estimation of the increase in success probability between the final and the initial shooting attempts by recruited students, allowing high accuracy degree and limiting bias.