Lecture 8

Two Sided Testing in the Bayesian Framework

In the Bayesian context we start with a prior, feed the model data, and end up with a posterior estimate of theta given said data, and we can therefore run inferential operations on the posterior. Say we have the following data, given the Laplacian prior.

$$x = 1, n = 2, \mathcal{F} : \ iid \ Bin(n, \theta)$$

$$P(\theta) = Beta(1, 1) \to P(\theta|x) = Bern(2, 2)$$

We will find that $\hat{\theta}_{MMSE}$, $\hat{\theta}_{MMAE}$, and $\hat{\theta}_{MAP}$ all lie on the same line in this distribution. Simple enough, but how do we find a credible region, as we do with confidence intervals in the frequentist approach. The folloowing is how we produce a % credible region for θ.

$$CR_{\theta,1-\alpha} = [Q[\theta|x, \frac{\alpha_0}{2}], [Q[\theta|x, 1 - \frac{\alpha_0}{2}]$$

Given we want to produce a 95% credible region for θ. We run the following code,

'qbeta(0.025, 2, 2)'
'qbeta(.975, 2, 2)'

Therefore we may say,

$$P(\theta \in [0.094, 0.906]|x) = 95\%$$

This is a real probability statement! The Confidence Interval approach cannot give you such a statement! The Credible Region is highly interpretable. Also, the CR is a proper subset of Θ for $\alpha_0 > 0$. This is not always true with CI's. FOr instance, here is the 95% CI for this data.

$$CI_{\theta,95\%} = [0.5 \pm 1.96\sqrt{\frac{0.5 - 0.5}{2}}] = [-0.21, 1.21] \not\subset \Theta = (0, 1)$$

The above CR is technically a two-sided CR. You can also create a one-sided (i.e. left-sided or right-sided) CR.

$$CR_{L,\theta,1-\alpha} = [smallest \ value \ in \ \Theta \ or \ -\infty, Q[\theta|x, 1 - \alpha_0]]$$

In our case,

$$CR_{L,\theta,1-0.05} = [smallest\ value\ in\ \Theta\ or\ -\infty, Q[\theta|x, 1-0.05]] = 0.865$$

'qbeta(.95, 2, 2)'

Therefore,

$$CR_{L,\theta,1-0.05} = [0, 0.865] \rightarrow P(\theta < 0.865|x) = 95\%$$

The same logic works inversely for the right sided CR test. This is called the credible region approach, but its not the only approach given the posterior. Another approach we will see (but not study further) is called the high density region (HDR) approach. Consider the posterior for θ to imitates the shape of a sin wave. The HDR approach measures the probability that θ exists in high density regions on the PDF. One disadvantage of this approach is that it can be non-continuous, secondly, it is computationally intense. Also, there are no left or right intervals by nature of the HDR.

We have now discussed the second goal of statistcal inference. We will now discuss the third goal, Bayesian hypothesis testing. In the frequentist approach, you either reject or fail to reject the null hypothesis. But we could not answer the probability in which the null hypothesis should fail or not fail. With the Bayesian setup we can immediately compute the following quantities:

$$Bayesian\ p-value = P(H_0, x), P(H_a|x)$$

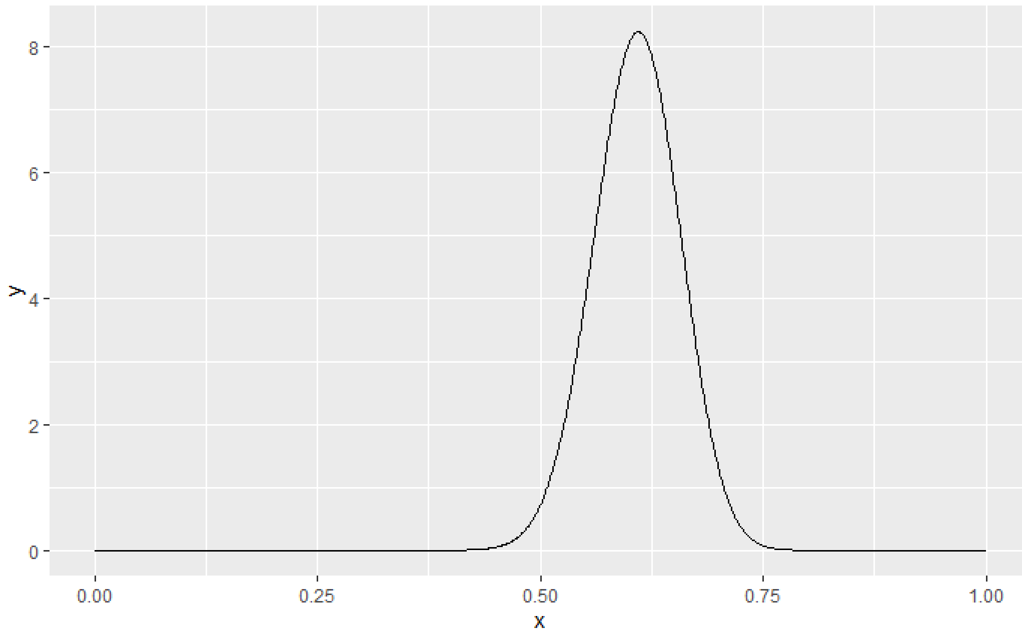
If $P(H_0|x) < \alpha_0$ then you reject H_0. What is so amazing about this process is that we have the value $P(H_0|x)$, whereas we could not explicitly solve for this with the frequentist approach. Let us recreate the example of coin flipping from Lecture 3. Let $n = 3$, where $x = 61$ were heads. Test if the coin is unfairly weighted towards heads at a 5% significance level.

$$H_a : \theta > 0.5,\ H_0 : \theta \leq 0.5$$

Assume $p(\theta) = Beta(1, 1)$.

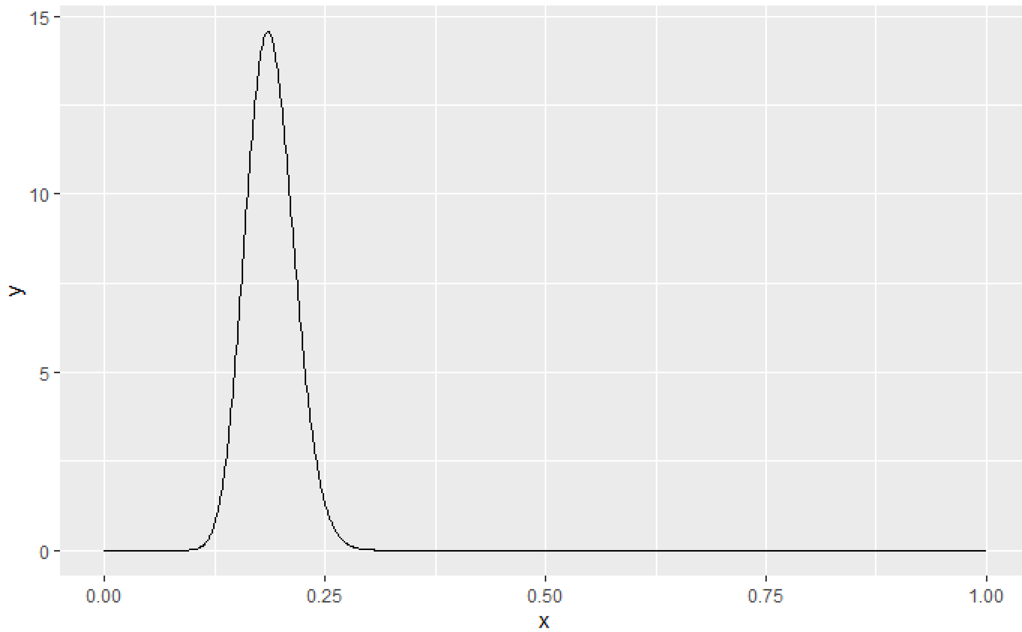
$$P(\theta|x) = Beta(62, 40)$$

$$P(H_0|x) = P(\theta \leq 0.5|x) = \int_0^{0.5} P(\theta|x) d\theta$$



According to the following code, pbeta(0.5, 62, 40)‘ which returns 0.014, or 1.4%, we reject the H_0 such that the coin is weighted unfairly towards heads.

In another example, an Uber driver does 200 rides and gets 37 non-5-star ratings. If his true proportion of non-5-star ratings is more than 25%, then Uber policy is to not fire the driver. Prove that he should be fired (or not) at a 5% significance level. We being by giving him the benefit of a doubt, and say that the null hypothesis is that his true proportion of 5-star ratings is greater than 25%. We will assume the Laplacian prior.



$$H_a : \theta > 25\% \rightarrow H_0 : \theta \leq 25\%$$

$$P(\theta|x) = Beta(38, 164)$$

$$P(\theta \leq 0.25|x) = pbeta(0.25, 38, 164) = 0.983 > 5\%$$

Therefore, we retain H_0 and we do not fire the employee.

In another example, lets test the coin again. Flip the coin 100 times and get 43 heads, $n = 100, x = 43$. Test if the coin is unfair.

$$P(\theta) = Beta(1, 1) \rightarrow P(\theta|x) = Beta(44, 58)$$

$$H_a : \theta \neq 0.5,\ H_0 : \theta = 0.5$$

$$P(H_0|x) = P(\theta = 0.5|x) = 0$$

Therefore do we always reject h_0? Yes. But something here is wrong. Using this approach, two-sided tests are always rejected (if the posterior is continuous). Any infinitely precise theory of θ is wrong in the real world. A coin is never 50.000000...% likely to flip heads. So we need to slightly reframe out hypotheses using a notion of "marginal equivalence" called δ (delta).

$$H_0 :\notin [\theta_0 \pm \delta] \rightarrow H_0 : \theta \in [\theta_0 \pm \delta]$$