170S Week 6 Discussion Notes

Colin Ni

February 13, 2025

Bayesian effects that we have seen

As usual, let X be a random variable depending on a parameter θ, and let $X_1, \dots, X_n$ be a sample drawn from X.

Recall that the core idea of Bayesian statistics is to make a guess, called the prior, on the distribution of θ and not just rely solely on the sample $X_1, \dots, X_n$. Thus we have two forces at play, our prior distribution and the sample, and Bayes' theorem gives us a posterior distribution on θ. We make our estimate by finding the conditional mean (or more generally by minimizing a chosen loss function) of this posterior.

On the following page, we tabulate the examples we have seen / will see on the homework. Let us discuss three common themes in these examples.

(1) Our prior and the example pull on our Bayes estimate in a predictable way. For example, in the binomial with beta-prior example, the Bayes estimate is a weighted average of the MLE y/m (which only involves the data) of the probability θ of success and the mean $\alpha/(\alpha+\beta)$ of our prior (which does not see the data):

$$\frac{\alpha + y}{m + \alpha + \beta} = \frac{m}{m + \alpha + \beta}\frac{y}{m} + \frac{\alpha + \beta}{m + \alpha + \beta}\frac{\alpha}{\alpha + \beta}.$$

You will see similar things on the homework.

(2) As $n \to \infty$, the effect of the prior weakens, and the Bayes estimate becomes a purely-data estimate (*e.g.* the MLE). For example, in the binomial with beta-prior example, as $n \to \infty$ we recover the MLE y/m of the probability θ of success (*e.g.* see the weighted average above), and in the normal with normal-prior example, as $n \to \infty$ we recover the MLE $\bar{X}$ of the mean μ.

(3) We have control over the strength of the prior. For example, in the normal with normal-prior example, taking σ_0^2 to be very small is tantamount to the prior saying it is extremely confident the mean should be close to θ_0, so the Bayes estimate is closer to the mean of the prior. For another example, in the binomial with beta-prior example, the bigger $\alpha + \beta$ is, the stronger our prior, because in our decomposition in (1) the first coefficient becomes 0 and the second coefficient becomes 1.

distr and parameter	prior	posterior	conditional mean of posterior
$Y \sim \mathrm{Binom}(m, \theta)$	$\theta \sim \mathrm{Beta}(\alpha, \beta)$	$\mathrm{Beta}(\alpha + y, m - y + \beta)$	$\dfrac{\alpha + y}{m + \alpha + \beta}$
$X_i \sim \mathrm{Norm}(\theta, \sigma^2)$	$\theta \sim \mathrm{Norm}(\theta_0, \sigma_0^2)$	$\theta \sim \mathrm{Norm}\left(\dfrac{\bar{X}\sigma_0^2 + \theta_0\sigma^2/n}{\sigma_0^2 + \sigma^2/n}, \dfrac{(\sigma^2/n)\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\right)$	$\dfrac{\bar{X}\sigma_0^2 + \theta_0\sigma^2/n}{\sigma_0^2 + \sigma^2/n}$
$Y_i \sim \mathrm{Pois}(\theta)$	$\theta \sim \mathrm{Gamma}(\alpha, \beta)$	$\theta \sim \mathrm{Gamma}\left(\alpha + n\bar{Y}, \dfrac{1}{n + \frac{1}{\beta}}\right)$	$\dfrac{\alpha\beta + n\beta\bar{Y}}{n\beta + 1}$
$X_i \sim \mathrm{Gamma}\,(\alpha, \theta)$	$\dfrac{1}{\theta} \sim \mathrm{Gamma}(\alpha_0, \theta_0)$	$\theta \sim \mathrm{Gamma}\left(\alpha_0 + n\alpha, \dfrac{1}{\frac{1}{\theta_0} + n\bar{X}}\right)$	$\dfrac{\alpha_0\theta_0 + n\theta_0\alpha}{1 + n\theta_0\bar{X}}$
$f(x \mid \theta) = 3\theta x^2 e^{-\theta x^3}$	$\theta \sim \mathrm{Gamma}\left(4, \dfrac{1}{4}\right)$	$\theta \sim \mathrm{Gamma}\left(n + 4, \dfrac{1}{4 + \sum_i x_i^3}\right)$	$\dfrac{n + 4}{4 + \sum_i x_i^3}$

Confidence intervals for Norm$(0, 1)$

Confidence intervals for the distribution $Z \sim \mathrm{Norm}(0, 1)$ is the common underlying mechanic for confidence intervals for other statistics, such as means, differences of two means, and proportions. The key definition to remember is as follows; it gives the threshold z_α for which only a proportion α of sampled Z values is above z_α, namely $z_\alpha = F_Z^{-1}(1-\alpha)$.

Definition. For any probability $\alpha \in (0, 1)$, the number $z_\alpha \in \mathbb{R}$ is defined to satisfy $F_Z(z_\alpha) = 1 - \alpha$.

Warmup. Carefully draw a big picture of the pdf of Z, and draw $\alpha = 0.16$ as the area under the curve over a right-tail region. Recall the 68–95–99.7 rule, and use this to label z_α.

Suppose now that we want intervals (a, b) such that $\mathbb{P}(a < Z < b) = \alpha$, or in other words confidence intervals for the values of Z. We can use the z_α values to do this in many ways, but here are the usual ones:

- The centered-at-zero interval $(-z_{\alpha/2}, z_{\alpha/2})$.
- The left-tail interval $(-\infty, z_\alpha)$.
- The right-tail interval $(-z_\alpha, \infty)$.

Confidence intervals for other statistics

To find confidence intervals for some statistic, the strategy is often to reduce the statistic to $Z \sim \mathrm{Norm}(0, 1)$ and then use these z_α values.

Example. If $X \sim \mathrm{Norm}(\mu, \sigma^2)$, then we can construct an α-confidence interval $[\mu - z, \mu + z]$ (that is centered at μ) for the values of X: since $X \overset{\mathrm{d}}{=} \sigma Z + \mu$, we have

$$\mathbb{P}(\mu - z < X < \mu + z) = \mathbb{P}\left(-\frac{z}{\sigma} < Z < \frac{z}{\sigma}\right) = \alpha$$

if and only if $z = \sigma z_{\alpha/2}$.

Example. Let $X_1, \ldots, X_n$ be a sample from $\mathrm{Norm}(\mu, \sigma^2)$, and consider the sample mean $\bar{X}$. Since the X_i are normal, $\bar{X}$ is normal, and recalling (*e.g.* from the week 2 discussion notes) that $\mathbb{E}(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n$, it follows that

$$\bar{X} \sim \mathrm{Norm}\left(\mu, \frac{\sigma^2}{n}\right).$$

Thus we can construct an α-confidence interval $[\mu - z, \mu + z]$ for the values of $\bar{X}$: since $\bar{X} \overset{\mathrm{d}}{=} \sigma Z/\sqrt{n} + \mu$, we have

$$\mathbb{P}(\mu - z < \bar{X} < \mu + z) = \mathbb{P}\left(-\frac{\sqrt{n}z}{\sigma} < Z < \frac{\sqrt{n}z}{\sigma}\right) = \alpha$$

if and only if $z = \sigma z_{\alpha/2}/\sqrt{n}$.

Example. Let $X_1, \ldots, X_n$ be a large sample from a random variable X with mean μ and known variance σ^2. By the central limit theorem

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathrm{d}} \mathrm{Norm}(0,1) \qquad \text{as } n \to \infty.$$

Thus since our sample is large, we can approximate

$$\bar{X} \sim \mathrm{Norm}\left(\mu, \frac{\sigma^2}{n}\right),$$

which is our previous example.

Example. Let $X_1, \ldots, X_{n_X}$ be a sample from $\mathrm{Norm}(\mu_X, \sigma_X^2)$, and let $Y_1, \ldots, Y_{n_Y}$ be a sample from $\mathrm{Norm}(\mu_Y, \sigma_Y^2)$, where we assume σ_X^2, σ_Y^2 are known. Then

$$\bar{X} - \bar{Y} \sim \mathrm{Norm}(\mu_X - \mu_Y, \sigma_X^2/n_X + \sigma_Y^2/n_Y).$$

Thus we can construct an α-confidence interval $[\mu_X - \mu_Y - z, \mu_X - \mu_Y + z]$ for the values of $\bar{X} - \bar{Y}$: since

$$\bar{X} - \bar{Y} \stackrel{\mathrm{d}}{=} \left(\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right) Z + (\mu_X - \mu_Y),$$

we have

$$\mathbb{P}((\mu_X - \mu_Y) - z < \bar{X} - \bar{Y} < (\mu_X - \mu_Y) + z) = \mathbb{P}\left(\frac{-z}{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} < Z < \frac{z}{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right) = \alpha$$

if and only if $z = (\sigma_X^2/n_X + \sigma_Y^2/n_Y) z_{\alpha/2}$.