STA 114: Statistics

Notes 11. Central Credible Intervals & Multiparameter Models

Interval summary of a pdf/pmf

We have already seen that a "central range" of a pdf $f(x)$ over a scalar variable x, can be summarized by $[x_{\alpha/2}, x_{1-\alpha/2}]$ for some small $\alpha \in (0,1)$, where for any $u \in (0,1)$, x_u denotes the u-th quantile of $f(x)$. This means, if $F(x)$ denotes the cdf of $f(x)$, then x_u is the smallest number x such that $F(x) \geq u$. If $f(x)$ is a pdf, then $F(x)$ is continuous and so $x_u = F^{-1}(u)$.

For pdf $f(x)$, the interval $[x_{\alpha/2}, x_{1-\alpha/2}]$ comes with the credibility that the area under $f(x)$ within this range equals $F(x_{1-\alpha/2}) - F(x_{\alpha/2}) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha$. For a pmf $f(x)$, the area is approximately $1-\alpha$. Of course, there are many other intervals which would include an area of $1-\alpha$ (e.g., the interval $[x_\beta, x_{1-(\alpha-\beta)}]$ for any $0 < \beta < \alpha$). However, the interval $[x_{\alpha/2}, x_{1-\alpha/2}]$ is "central", because it leaves out exactly $\alpha/2$ area in either tail. We shall call $[x_{\alpha/2}, x_{1-\alpha/2}]$ the $100(1-\alpha)\%$ central credible interval of the pdf/pmf $f(x)$. If $f(x)$ is used to give plausibility scores of a variable X then we shall also identify $[x_{\alpha/2}, x_{1-\alpha/2}]$ as the $100(1-\alpha)\%$ central credible interval for X.

A nice property of this interval is invariance under monotone transformation. Suppose $X \sim f(x)$ and $Y = h(X)$ where $h(x)$ is a monotone (either increasing or decreasing) function. Then the for any $u \in (0,1)$, the u-th quantile y_u of Y is precisely $h(x_u)$, with x_u denoting the u-th quantile of X. Therefore the central $100(1-\alpha)\%$ credible $[y_{\alpha/2}, y_{1-\alpha/2}]$ interval for Y is exactly $[h(x_{\alpha/2}), h(x_{1-\alpha/2})]$, which is obtained by applying the function $h(x)$ to the end points of the $100(1-\alpha)\%$ credible interval of X.

Central credible intervals in scalar parameter Bayesian models

For a Bayesian analysis, we shall talk about prior and posterior credible intervals of a parameter or of a quantity derived from the parameter. For a model $X \sim f(x|\theta)$, with a scalar parameter $\theta \in \Theta$, and prior pdf/pmf $\xi(\theta)$, the central $100(1-\alpha)\%$ prior credible interval for θ is $[\theta_{\alpha/2}, \theta_{1-\alpha/2}]$ and the central $100(1-\alpha)\%$ posterior credible interval for θ is $[\theta_{\alpha/2}(x), \theta_{1-\alpha/2}(x)]$. If we are interested in $\eta = h(\theta)$, where h is a monotone transformation, then the invariance result above helps to get prior and posterior central credible intervals for η.

Example (Opinion poll)**.** For the opinion poll example, with model $X \sim \mathsf{Binomial}(n,p)$, $n = 500$, $p \in [0,1]$ and $\xi(p) = \mathsf{Uniform}(0,1)$, the posterior is $\xi(p|x) = \mathsf{Beta}(x+1, n-x+1)$. The central 95% prior credible interval for p is $[0.025, 0.975]$ (directly from the uniform pdf plot) and the central 95% posterior credible interval, based on data $x = 200$ is $[0.36, 0.44]$ (from `qbeta()`). If we are interested in the log-odds ratio $\eta = \log \frac{p}{1-p}$, then prior and posterior central 95% credible intervals of η are $[-3.7, 3.7]$ and $[-0.54, -0.28]$ (by applying the same transformation to the end points of the above intervals). □

Central credible intervals in multiparameter Bayesian models

If the model parameter is vector valued, and we are interested in some scalar quantity $\eta = h(\theta)$ derived from it (h is now a many-to-one function, so cannot talk of monotonicity), then we must find prior and posterior pdfs of η from those of θ. Once we get these pdfs we can directly apply the central credible interval concept to these derived pdfs.

Example (Lactic acid conentration)**.** Consider n lactic acid measurements modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, $\sigma^2 \in (0, \infty)$, with a $\mathsf{N}\chi^{-2}(1, 1, 1, 1)$ prior $\xi(\mu, \sigma^2)$. For observations (0.86, 1.53, 1.57, 1.81, 0.99, 1.09, 1.29, 1.78, 1.29, 1.58) with $n = 10$, $\bar{x} = 1.38$, $s_x = 0.33$, the posterior pdf is $\xi(\mu, \sigma^2|x) = \mathsf{N}\chi^{-2}(1.34, 11, 11, 0.19)$. We are interested in getting a 95% posterior credible interval for μ.

From the properties of the normal-inverse-chi-square distributions, when $(\mu, \sigma^2) \sim \mathsf{N}\chi^{-2}(m, k, r, s)$, the variable $\eta = \sqrt{k/s}(\mu - m)$ has a $t(r)$ pdf. So a $100(1-\alpha)$% central credible interval for η is $[-z_r(\alpha), z_r(\alpha)]$. Because $\mu = m + \sqrt{s/k} \cdot \eta$ is a monotone transformation of η, we must have $m \mp \sqrt{s/k} \times z_r(\alpha)$ as the central $100(1-\alpha)$% for μ.

Therefore, for $\xi(\mu, \sigma^2|x) = \mathsf{N}\chi^{-2}(1.34, 11, 11, 0.19)$, the central 95% posterior credible interval for μ is $1.34 \mp 0.132 \times 2.2 = [1.05, 1.63]$. □

Some fundamental concepts about multiparameter models

As in the previous example, we might be interested in a single parameter in a multi-parameter Bayesian model. There are various things we can pursue here and all follows from the basic concept of joint, conditional and marginal probability distributions of a collection of variables.

For simplicity, assume we have a two parameter model: $X \sim f(x|\theta_1, \theta_2)$, $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$ with prior pdf/pmf on $\xi(\theta_1, \theta_2)$ that leads to the posterior pdf/pmf $\xi(\theta_1, \theta_2|x) = \text{const} \times f(x|\theta_1, \theta_2)\xi(\theta_1, \theta_2)$ once we observe $X = x$.

So both the prior and the posterior provide plausibility scores on the joint space of θ_1 and θ_2. From either, we can extract the plausibility scores on

- θ_1 alone , or
- θ_2 alone, or
- θ_1 given a specific value of θ_2 or
- θ_2 given a specific value of θ_1.

To further simplify our discussion, let's assume that Θ_1 and Θ_2 are discrete sets, so both $\xi(\theta_1, \theta_2)$ and $\xi(\theta_1, \theta_2|x)$ are pmfs.

When we talk about the plausibility of $\theta_1 = \theta_1'$ (for some fixed number $\theta_1' \in \Theta_1$), without saying anything about θ_2, we are essentially considering all possibilities for θ_2. That is the plausibility of $\theta_1 = \theta_1'$ is the same as the plausibility of the event $\{\theta_1 = \theta_1', \theta_2 \in \Theta_2\}$. The plausibility score of this event, respectively under the prior and the posterior, equals $\xi_1(\theta_1') = \sum_{\theta_2 \in \Theta_2} \xi(\theta_1', \theta_2)$ and $\xi_1(\theta_1'|x) = \sum_{\theta_2 \in \Theta_2} \xi(\theta_1', \theta_2|x)$. The functions $\xi_1(\theta_1)$ and $\xi_1(\theta_1|x)$ defined this way are pmfs over Θ_1 (this is easy to verify), and are called the marginal prior and posterior pmfs of θ_1. We could similarly define the marginal prior and posterior pmfs of

θ_2 as $\xi_2(\theta_2) = \sum_{\theta_1 \in \Theta_1} \xi(\theta_1, \theta_2)$ and $\xi_2(\theta_2|x) = \sum_{\theta_1 \in \Theta_1} \xi(\theta_1, \theta_2|x)$ and use these to talk about θ_2 alone, without any specific mentions of θ_1.

What if we want to talk about plausibility scores of θ_1 when θ_2 has been fixed at θ_2'? The relative (conditional) scores of $\theta_1 = \theta_1'$ against $\theta_1 = \theta_1''$ given $\theta_2 = \theta_2'$ equals $\frac{\xi(\theta_1', \theta_2')}{\xi(\theta_1'', \theta_2')}$ under the prior and $\frac{\xi(\theta_1', \theta_2'|x)}{\xi(\theta_1'', \theta_2'|x)}$ under the posterior. These relative scores are characterized by the conditional pmfs of θ_1 given θ_2 defined as (under prior and posterior)

$$\xi_1(\theta_1|\theta_2) = \frac{\xi(\theta_1, \theta_2)}{\xi_2(\theta_2)}, \quad \xi_1(\theta_2|x, \theta_2) = \frac{\xi(\theta_1, \theta_2|x)}{\xi_2(\theta_2|x)}.$$

For every fixed value of θ_2, the functions $\xi_1(\theta_1|\theta_2)$ and $\xi_1(\theta_1|x, \theta_2)$ are pmfs over $\theta_1 \in \Theta_1$ and will be called the conditional prior and posterior pmfs of θ_1 given θ_2. We could similarly define conditional prior and posterior pmfs of θ_2 given θ_1.

Similar concepts apply when either $\xi(\theta_1, \theta_2)$ or $\xi(\theta_1, \theta_2|x)$ or both are pdfs. We now talk about marginal prior and posterior pdfs of θ_1: $\xi_1(\theta_1) = \int_{\Theta_2} \xi(\theta_1, \theta_2) d\theta_2$ and $\xi_1(\theta_1) = \int_{\Theta_2} \xi(\theta_1, \theta_2|x) d\theta_2$ (and similarly define $\xi_2(\theta_1)$ and $\xi_2(\theta_2|x)$ for θ_2). Conditional prior and posterior pdfs are still given by $\xi_1(\theta_1|\theta_2) = \xi(\theta_1, \theta_2)/\xi_2(\theta_2)$ and $\xi_1(\theta_1|x, \theta_2) = \xi(\theta_1, \theta_2|x)/\xi_2(\theta_2|x)$, but now viewed as pdfs over Θ_1. Similar definition applies to $\theta_2 \in \Theta_1$.

Note that

$$\xi_1(\theta_1|\theta_2) = \text{const} \times \xi(\theta_1, \theta_2), \quad \theta_1 \in \Theta_1$$

where the constant term may involve θ_2, but is constant wrt to the function argument θ_1. So if we can identify $\xi(\theta_1, \theta_2)$ as a constant multiple of some pdf/pmf in θ_1 then $\xi_1(\theta_1|\theta_2)$ must equal that pdf/pmf. The same logic applies to $\xi_1(\theta_1|x, \theta_2)$, $\xi_2(\theta_2|\theta_1)$ and $\xi_2(\theta_2|x, \theta_1)$.