Bayesian Concepts

Chao Cheng

September 19, 2024

Contents

1 Introduction

1. Prior and posterior distribution
2. Predictive probability and application in phase II design
3. Credible interval

From the frequentist perspective, we have the data x and the parameter of the distribution θ and we make estimation/inference about θ. But θ is alwyas treated as a fixed parameter. But from bayesian perspective, θ is also a random variable. First we introduce some notations:

- The prior distribution $\pi(\theta|\alpha)$, where α is fixed parameters for the distribution of θ. This prior distribution of θ represents our previous knowledge of θ before the data x is collected.
- The data distribution $f(x|\theta)$, which is the same as that from frequentist's perspective.
- The posterior distribution $f_{post}(\theta|x)$, which is the distribution of θ based on (conditional on) the observed data. Note that

$$f_{post}(\theta|x) = \frac{f(\theta, x)}{f(x)} = \frac{f(x|\theta)\,\pi(\theta|\alpha)}{f(x)} \propto f(x|\theta)\,\pi(\theta|\alpha),$$

  where the last $\propto$ is taken with respect to θ. So the kernel of posterior distribution of θ given x is determined by $f(x|\theta)\,\pi(\theta|\alpha)$. Sometimes we will write $f_{post}(\theta|x)$ as $f_{post}(\theta|x;\alpha)$ to emphasize that this posterior distribution depends on x and parameter α.

2 Credible Interval

There are mainly two types of credible interval: equal tail interval (ETI) and highest (posterior) density interval (HDI). Denote the CI as $[lower, upper]$, then these two ends satisfies different conditions for ETI and HDI:

- For ETI, the CI has equal probability tails on both sides, that is $F(lower) = (1 - F(upper)) = \alpha/2$ and the CI is $[lower, upper]$.

- For HDI, the CI has equal **density** on both ends, that is $f(lower) = f(upper)$ and $F(upper) - F(lower) = 1 - \alpha$.

  Note: this definition of HDI is not that formal, especially when the distribution is not unimodal.

Note: sometimes we will use SDI (shortest density interval) to mean HDI, since SDI can be seen as a more robust method to compute HDI[Liu et al., 2015].

ETI and HDI can produce same results as long as the distribution is symmetric. But in reality many posterior distributions are skewed, then HDI is a more reasonable CI. A demonstration is shown in Figure 1. Note that HDI can be hard to compute, a starting

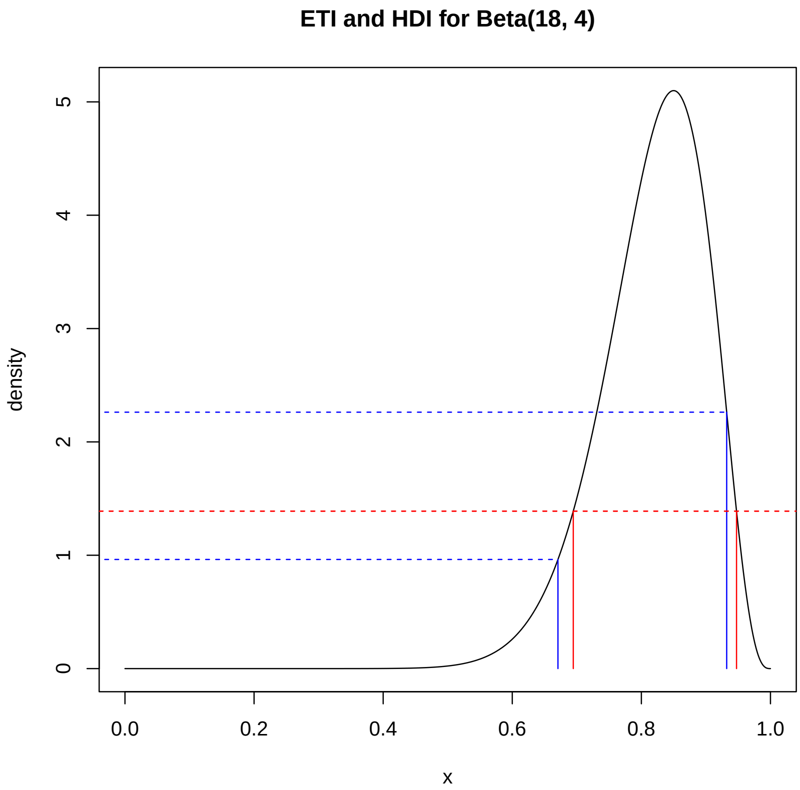

Figure 1: Comparison of 90% ETI and HDI for Beta(18,4)

point of `R` script is shown here, using package `bayestestR`:

```
x <- seq(from = 0, to = 1, by = 0.001)
a <- 18
b <- 4
```

```r
y1 <- dbeta(x, a, b)

# HDI
tmp <- bayestestR::distribution_beta(n = 100000, shape1 = a, shape2 = b)
tmp2 <- bayestestR::ci(tmp, ci = 0.90, method = "HDI")

# Compare HDI and SPI
bayestestR::ci(tmp, ci = 0.90, method = "HDI")
str(bayestestR::hdi(tmp, ci = 0.9))
str(bayestestR::spi(tmp, ci = 0.9))     # SPI is a more robust HDI
```

3 Predictive distribution

Note that if one wants to make some prediction/statement about x, like $x \in A$, then the probability can be computed as

$$P(X \in A) = \int_A f_X(x)\,\mathrm{d}x = \int_A \int f_{X,\theta}(x,\theta)\,\mathrm{d}\theta\mathrm{d}x = \int_A \int f_{X|\theta}(x|\theta)\, f_\theta(\theta)\,\mathrm{d}\theta\mathrm{d}x, \tag{1}$$

where $f_{X|\theta}(x|\theta)$ is the data distribution and $f_\theta(\theta)$ is the distribution of θ, either prior or posterior distribution. (1) is called the **predictive probability** of $X \in A$, which is the probability based on marginal distribution of X.

References

Ying Liu, Andrew Gelman, and Tian Zheng. Simulation-efficient shortest probability intervals. *Statistics and Computing*, 25(4):809–819, jun 2015. doi: 10.1007/s11222-015-9563-8.