

The Bayesian Central Limit Theorem, with some intructions on #1(e) of the second problem set

We have seen Bayes formula:

posterior probability density = constant · prior probability density · likelihood function.

The “Bayesian Central Limit Theorem” says that under certain circumstances, the posterior probability distribution is approximately a normal distribution. Actually more than one proposition is called “the Bayesian Central Limit Theorem,” and we will look at two of them.

First version: Suppose Θ is distributed according to the probability density function $f_{\Theta}(\theta)$, and

$$X_1, X_2, X_3, \dots \mid [\Theta = \theta] \sim \text{i. i. d. } f_{X_1 \mid [\Theta = \theta]}(x)$$

i.e., X_1, X_2, X_3, \dots are conditionally independent given that $\Theta = \theta$, and $f_{X_1 \mid [\Theta = \theta]}(x)$, as a function of x with θ fixed, is the conditional probability density function of X_1 given that $\Theta = \theta$. Suppose the posterior probability density function $f_{\Theta \mid [X_1 = x_1, \dots, X_n = x_n]}$ is everywhere twice-differentiable, for all n .

Let $c =$ the posterior expected value $= \mathbf{E}(\Theta \mid X_1 = x_1, \dots, X_n = x_n)$, and

let $d =$ the posterior standard deviation $= \sqrt{\text{Var}(\Theta \mid X_1 = x_1, \dots, X_n = x_n)}$.

Then if the distribution of U is the posterior distribution, then the distribution of $\frac{U - c}{d}$ is approximately standard normal; more precisely: that distribution approaches the standard normal distribution as n approaches infinity. ■

More tersely: The posterior distribution approaches a normal distribution as the sample-size n grows.

How big a value of n is needed to get a good approximation? For a partial answer to that, consider #1(e) on the second problem set. A 90% posterior probability interval found without using the Bayesian Central Limit Theorem is (0.419705351626, 0.788091845180). These are correct to 12 digits after the decimal point if I can trust my software. The endpoints of this interval differ from the approximations given by the above version of the Bayesian Central Limit Theorem by slightly less than 0.01.

In the case of beta distributions, finding the expected value and the standard deviation is trivial: Just use what is says on page 306 of Schervish & DeGroot. For some other

distributions, finding the expected value and the standard deviation may require laborious integrals, and hence we might use the next version of the Bayesian Central Limit Theorem:

Second version: Start with the same assumptions as in the first version.

Let $c =$ the posterior mode
 $=$ the value of θ that maximizes the posterior density function $f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta)$.

Let $k(\theta) = \log f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta)$.

Let $d = (-k''(c))^{-1/2}$.

Then if the distribution of U is the posterior distribution, then the distribution of $\frac{U - c}{d}$ is approximately standard normal; more precisely: that distribution approaches the standard normal distribution as n approaches infinity. ■

Important: In #1(e) on the second problem set, try both versions of the Bayesian Central Limit Theorem. Compare the result of each with the result alleged to be correct in the second-to-last paragraph on the first page of this handout.

Here is a rough sketch of a proof of the second version. Denote $f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta)$ by $f(\theta)$, and let $k(\theta) = \log f(\theta)$. Then

$$\begin{aligned} f(\theta) &= e^{\log f(\theta)} = e^{k(\theta)} = e^{k(c) + k'(c)(\theta - c) + k''(c)(\theta - c)^2/2 + k'''(c)(\theta - c)^3/6 + \dots} \\ &\cong e^{k(c) + k'(c)(\theta - c) + k''(c)(\theta - c)^2/2} \\ &= e^{k(c) + 0 + k''(c)(\theta - c)^2/2} \\ &= [\text{constant}] \cdot e^{k''(c)(\theta - c)^2/2} \end{aligned}$$

(We must have $k'(c) = 0$ and $k''(c) < 0$
since there is a maximum at c .)

$$= [\text{constant}] \cdot e^{-(\theta - c)^2/(2\sigma^2)}$$

provided $-1/\sigma^2 = k''(c)$, i.e., $\sigma = (-1/k''(c))^{-1/2}$, which is the same thing that followed the words "Let $d =$ " above.

Why do the 3rd- and higher-degree terms evaporate as $n \rightarrow \infty$ above? To answer that you need to know how they depend on n . Those interested in such theoretical questions can work out the details.