Lecture 7:
Bayesian Estimation Methods

Big Data and Machine Learning for Applied Economics
Econ 4676

Ignacio Sarmiento-Barbieri

Universidad de los Andes

August 31, 2021

Recap

- OLS
- MLE
- Computation

Agenda

The Bayesian Approach

- We've been living in a "frequentist" world
- Observe the data
- Impose some assumptions on the data

$$X_1, X_2, \ldots, X_n \sim_{iid} f(X|\theta) \tag{1}$$

- The parameter θ is thought to be unknown

The Bayesian Approach

- What happens if we have a prior belief of θ?
- How can we incorporate this information?
- Bayesian Statistics gives us a framework to do it a systematic way

A Simple Covid Example

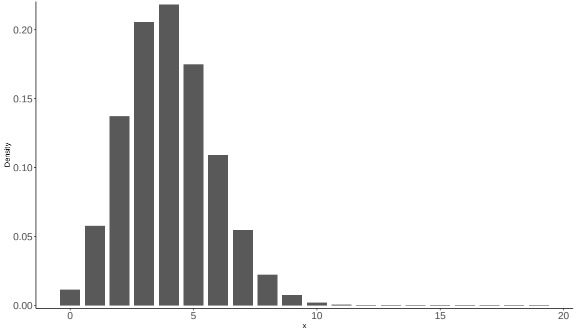
- Suppose we are interested in the prevalence of COVID in a small city.
- The higher the prevalence, the more public health precautions we would recommend be put into place.
- A small random sample of 20 individuals from the city will be checked for the presence of the virus.
- Interest is in θ, the fraction of infected individuals in the city.
- X records the total number of people in the sample who are infected. Before the sample is obtained the number of infected individuals in the sample is unknown.

A Simple Covid Example

- If the value of θ were known, a reasonable sampling model would be

$$X|\theta \sim Binomial(20, \theta) \tag{2}$$

- Suppose we observe the following data, this is consistent with $\theta = .2$



$$Pr(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \tag{3}$$

$$Pr(X = 0) = \binom{20}{0} 0.2^0 (1-0.2)^{20-0} \approx 0.01 \tag{4}$$

A Simple Covid Example
Prior distribution

- Other studies from various parts of the country indicate that the infection rate in comparable cities ranges from about 0.05 to 0.20, with an average prevalence of 0.10.
- How can we incorporate this information?
- Bayes Theorem to the rescue

Bayes Theorem

For this updating we use *Bayes Theorem*

$$\pi(\theta|X) = \frac{f(X|\theta)p(\theta)}{m(X)} \tag{5}$$

with $m(X)$ is the marginal distribution of X, i.e.

$$m(X) = \int f(X|\theta)p(\theta)d\theta \tag{6}$$

It is important to note that Bayes' theorem does not tell us what our beliefs should be, it tells us how they should change after seeing new information.

A Simple Covid Example

Prior distribution

- We can characterize our prior ($p(\theta)$) with a Beta distribution

$$\theta \sim Beta(a, b) \tag{7}$$

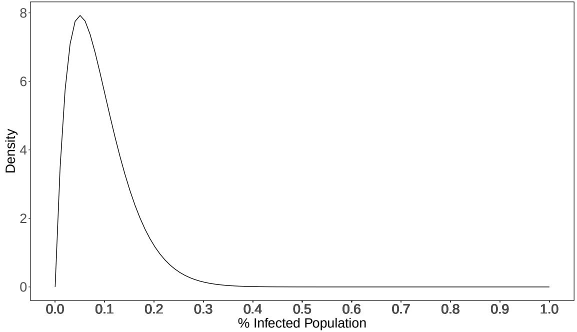
where the density of a Beta takes the form of

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \tag{8}$$

A Simple Covid Example

Prior distribution

For now let, $a = 2$ and $b = 20$.



$$E(\theta) = \frac{a}{a+b} = 0.09 \tag{9}$$

$$Pr(0.05 < \theta < 0.20) = 0.66 \tag{10}$$

A Simple Covid Example

Posterior distribution

$$\pi(\theta|X) = \frac{f(X|\theta)p(\theta)}{m(X)} \tag{11}$$

$$\pi(\theta|X) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\frac{1}{m(x)} \tag{12}$$

The marginal

$$\begin{align}
m(x) &= \int f(X|\theta)p(\theta)d\theta \tag{13}\\
&= \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta \tag{14}\\
&= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{x+a-1}(1-\theta)^{n-x+b-1}d\theta \tag{15}
\end{align}$$

A Simple Covid Example
Posterior distribution

The marginal (cont)

$$m(x) = \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+x)\Gamma(b+n-x)}{\Gamma(a+b+n)}\int_0^1 \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)}\theta^{x+a-1}(1-\theta)^{n-x+b-1}d\theta \quad (16)$$

$$= \binom{n}{x}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+x)\Gamma(b+n-x)}{\Gamma(a+b+n)} \quad (17)$$

The posterior

$$\pi(\theta|X) = \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)}\theta^{x+a-1}(1-\theta)^{n-x+b-1} \quad (18)$$

$$\sim Beta(a+x, b+n-x) \quad (19)$$

A Simple Covid Example

With the posterior we can calculate then any moment of the posterior distribution. For example suppose that for our study none of the sample of individuals is infected (x=0). Then the posterior is

$$\pi(\theta|X=0) \sim Beta(2,40) \tag{20}$$

$a = 2, b = 20, n = 20$. Then

$$E(\theta|X=0) = 0.048 \tag{21}$$

A Simple Covid Example

How did we get there?

$$
\begin{aligned}
E(\theta|X=0) &= \frac{a+x}{a+b+n} && (22)\\
&= \frac{n}{a+b+n}\frac{x}{n} + \frac{a+b}{a+b+n}\frac{a}{a+b} && (23)\\
&= \frac{n}{a+b+n}\bar{x} + \frac{a+b}{a+b+n}\theta_{prior} && (24)\\
&= \frac{n}{a+b+n}0 + \frac{a+b}{a+b+n}\frac{2}{22} && (25)\\
&= 0.048 && (26)
\end{aligned}
$$

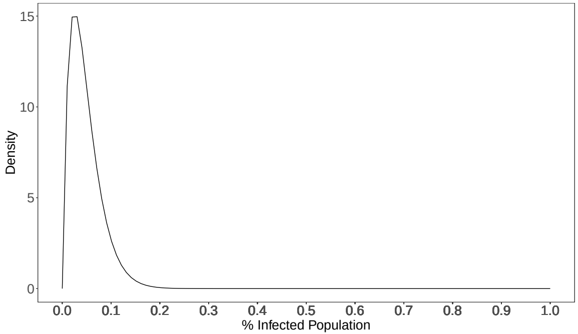
Recall that $a = 2$, $b = 20$, $n = 20$

A Simple Covid Example

Since we have the full distribution we could calculate for example:

$$mode(\theta|X) = 0.025 \tag{27}$$

$$Pr(\theta < 0.20|X = 0) = 0.998 \tag{28}$$



Bayesian Estimation

- The Bayesian approach to stats is fundamentally different from the classical approach we have been taking
- In the classical approach, the parameter θ is thought to be an unknown, but fixed quantity, e.g., $X_i \sim f(\theta)$
- In the Bayesian approach θ is considered to be a quantity whose variation can be described by a probability distribution (*prior distribution*)
- Then a sample is taken from a population indexed by θ and the prior is updated with this sample
- The resulting updated prior is the *posterior distribution*

Conjugate Priors

Definition Let $\mathcal{F}$ denote the class of densities $f(x|\theta)$. A class $\mathcal{C}$ of prior distributions is a conjugate family for $\mathcal{F}$ if the posterior distribution is in the class $\mathcal{C}$ for all $f \in \mathcal{F}$, all priors in $\mathcal{C}$, and all $x \in X$

For example:

- $X \sim D(\theta)$ and $\theta \sim P(\lambda) \rightarrow \theta|X \sim P(\lambda')$
- the normal distribution is a conjugate for the normal family $X \sim N(\mu, \sigma)$ and $\theta \sim N(\mu_0, \sigma_0) \rightarrow \theta|X \sim N(\mu', \sigma')$
- the beta distribution for the binomial family $X \sim Bernoulli(\theta)$ and $\theta \sim Beta(a, b) \rightarrow \theta|X \sim Beta(a', b')$

Good and bad news:

- Nice because gives us a closed form for the posterior. However, whether a conjugate family is a reasonable choice is left to you!
- Downside, if we choose another families, then these results are no longer available. Then we have to use sampling-based methods (Gibbs Sampler, MCMC, etc)

Bayesian Linear Regression

Consider

$$y_i = \beta x_i + u_i \;\; u_i \sim_{iid} N(0, \sigma^2 I) \tag{29}$$

The likelihood function is

$$f(y|\beta,\sigma,x) = \Pi_{i=1}^{n} \frac{1}{(\sqrt{2\pi\sigma^2})} e^{-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2} \tag{30}$$

Now consider that the prior for β is $N(\beta_0, \tau^2)$

$$p(\beta) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\beta - \beta_0)^2} \tag{31}$$

Bayesian Linear Regression

The Posterior distribution then

$$\pi(\beta|y,x) = \frac{f(y,x|\beta)p(\beta)}{m(y,x)} \tag{32}$$

$$= \frac{f(y|x,\beta)f(x|\beta)p(\beta)}{m(y,x)} \tag{33}$$

by assumption $f(x|\beta) = f(x)$

$$= f(y|x,\beta)p(\beta)\frac{f(x)}{m(y,x)} \tag{34}$$

$$\propto f(y|x,\beta)p(\beta) \tag{35}$$

Bayesian Linear Regression (Detour)

Useful Result:

Suppose a density of a random variable θ is proportional to

$$exp\left(\frac{-1}{2}(A\theta^2 + B\theta)\right) \tag{36}$$

Then $\theta \sim N(m, V)$ where

$$m = \frac{-1B}{2A} \quad V = \frac{1}{A} \tag{37}$$

Bayesian Linear Regression (we are back)

$$P(\beta|y,X) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n exp\left(\frac{-1}{2\sigma^2}\sum(y_i-\beta x_i)^2\right) exp\left(\frac{-1}{2\tau^2}(\beta-\beta_0)^2\right) \tag{38}$$

$$\propto exp\left[\frac{-1}{2}\left(\frac{1}{\sigma^2}\sum(y_i-\beta x_i)^2+\frac{-1}{\tau^2}(\beta-\beta_0)^2\right)\right] \tag{39}$$

Bayesian Linear Regression (we are back)

Using the previous detour

$$A = \frac{1}{\sigma^2}\sum x_i^2 + \frac{1}{\tau^2} \tag{40}$$

$$B = -2\frac{1}{\sigma^2}\sum y_i x_i + \frac{1}{\tau^2}\beta_0 \tag{41}$$

Then $\beta \sim N(m, V)$ with

$$m = \frac{\frac{1}{\sigma^2}\sum y_i x_i + \frac{1}{\tau^2}\beta_0}{\left(\frac{1}{\sigma^2}\sum x_i^2 + \frac{1}{\tau^2}\right)} \tag{42}$$

$$V = \frac{1}{A} \tag{43}$$

Bayesian Linear Regression (we are back)

$$m = \left(\frac{\frac{\sum x_i^2}{\sigma^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \right) \frac{\sum x_i y_i}{\sum x_i^2} + \left(\frac{\frac{1}{\tau^2}}{\frac{\sum x_i^2}{\sigma^2} + \frac{1}{\tau^2}} \right) \beta_0 \tag{44}$$

$$m = \omega \hat{\beta}_{MLE} + (1 - \omega)\beta_0 \tag{45}$$

Remarks

- If prior belief is strong $\tau \downarrow 0 \rightarrow \omega \downarrow 0 \implies m = \beta_0$
- If prior belief is weak $\tau \uparrow \infty \rightarrow \omega \uparrow 1 \implies m = \beta_{MLE}$

Review & Next Steps

- Bayesian Estimation
- **Next Class:** Cont. Bayesian Stats.

Further Readings

- Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2, pp. 337-472). Pacific Grove, CA: Duxbury. Chapter 7
- Hoff, P. D. (2009). A first course in Bayesian statistical methods (Vol. 580). New York: Springer.