Quant III

Lab 5: Beyesian: Basics

Junlong Aaron Zhou

October 09, 2020

Outline

- Homework Question
- Bayes Basics
- Estimates and Inference

Homework Question:

- Uniform distribution $U(0, \theta)$
- If θ is known:

$$f(y) = \begin{cases} \frac{1}{\theta} & y \in [0, \theta] \\ 0 & \text{otherwuse} \end{cases}$$

Homework Question:

- Uniform distribution $U(0, \theta)$
- If θ is known:

$$f(y) = \begin{cases} \frac{1}{\theta} & y \in [0, \theta] \\ 0 & \text{otherwuse} \end{cases}$$

- Therefore: $f(y) = \frac{1}{\theta}\mathbb{I}(y \leq \theta)$
- $L = \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{I}(y_i \leq \theta)$
- Note $L = 0$ if $\exists y_i > \theta$, and decreasing in θ

Homework Question:

- Uniform distribution $U(0, \theta)$
- If θ is known:

$$f(y) = \begin{cases} \frac{1}{\theta} & y \in [0, \theta] \\ 0 & \text{otherwuse} \end{cases}$$

- Therefore: $f(y) = \frac{1}{\theta}\mathbb{I}(y \leq \theta)$
- $L = \prod_{i=1}^{n} \frac{1}{\theta}\mathbb{I}(y_i \leq \theta)$
- Note $L = 0$ if $\exists y_i > \theta$, and decreasing in θ
- $\hat{\theta}_{MLE}$ is the smallest possible value, i.e. $max\{y_i\}$

Negative Binomial Model

- Remind me later.

Basic Bayes

- **WARNINGS:** Please do ask questions.

Basic Bayes

- **WARNINGS:** Please do ask questions.
- Frequentist v.s.Bayesian
  - Natural, fixed, unknown parameter θ v.s. distribution of θ
  - Infinite sampling (hypothetical)/ Asymptotics v.s. Limited observations
- Take-away: different way of thinking.

Bayesian Estimation: Setup

- Unknown parameter: θ
- Prior: $P(\theta)$
- Posterior $P(\theta|Y)$ given by Bayes Rule:

Bayesian Estimation: Setup

- Unknown parameter: θ
- Prior: $P(\theta)$
- Posterior $P(\theta|Y)$ given by Bayes Rule:

$$\begin{aligned} P(\theta|Y) &= \frac{P(\theta, Y)}{P(Y)} \\ &= \frac{P(Y|\theta)P(\theta)}{\int_\theta P(Y|\theta)P(\theta)} \end{aligned}$$

Setup ctd.

$$P(\theta|Y) = \frac{P(\theta, Y)}{P(Y)}$$
$$= \frac{P(Y|\theta)P(\theta)}{\int_\theta P(Y|\theta)P(\theta)}$$

Setup ctd.

$$P(\theta|Y) = \frac{P(\theta, Y)}{P(Y)}$$
$$= \frac{P(Y|\theta)P(\theta)}{\int_\theta P(Y|\theta)P(\theta)}$$

- Likelihood enters: $L(\theta) = \prod f_\theta(Y) = P(Y|\theta)$

Setup ctd.

$$\begin{aligned} P(\theta|Y) &= \frac{P(\theta, Y)}{P(Y)} \\ &= \frac{P(Y|\theta)P(\theta)}{\int_\theta P(Y|\theta)P(\theta)} \end{aligned}$$

- Likelihood enters: $L(\theta) = \prod f_\theta(Y) = P(Y|\theta)$

$$\begin{aligned} \text{Rewrite posterior: } P(\theta|Y) &= \frac{1}{P(Y)} \times \text{Likelihood} \times \text{Prior} \\ &= \text{Constant} \times \text{Likelihood} \times \text{Prior} \\ &\propto \text{Likelihood} \times \text{Prior} \end{aligned}$$

Setup ctd.

$$P(\theta|Y) = \frac{P(\theta, Y)}{P(Y)} = \frac{P(Y|\theta)P(\theta)}{\int_\theta P(Y|\theta)P(\theta)}$$

- Likelihood enters: $L(\theta) = \prod f_\theta(Y) = P(Y|\theta)$

$$\text{Rewrite posterior: } P(\theta|Y) = \frac{1}{P(Y)} \times \text{Likelihood} \times \text{Prior}$$
$$= \text{Constant} \times \text{Likelihood} \times \text{Prior}$$
$$\propto \text{Likelihood} \times \text{Prior}$$

- Kernel: $P(Y|\theta)P(\theta)$. Once you know this, you know the distribution.

Point Estimation

- MLE (Maximum Likelihood Estimation): $\underset{\theta}{Max} L(\theta)$
- MAP (Maximum A Posteriori): $\underset{\theta}{Max} L(\theta) P(\theta)$

Point Estimation

- MLE (Maximum Likelihood Estimation): $\underset{\theta}{Max} L(\theta)$
- MAP (Maximum A Posteriori): $\underset{\theta}{Max} L(\theta) P(\theta)$
- Prior!
- $MLE = MAP$ when $P(\theta)$ is constant.
- Choose prior is important

Point Estimation

- MLE (Maximum Likelihood Estimation): $\underset{\theta}{Max} L(\theta)$
- MAP (Maximum A Posteriori): $\underset{\theta}{Max} L(\theta) P(\theta)$
- Prior!
- $MLE = MAP$ when $P(\theta)$ is constant.
- Choose prior is important
  - Conjugate prior: get an analytic solution
  - Otherwise: Numerically approximate the posterior (MCMC, VI).
  - Computational convenience, and/or subjective knowledge
  - Sometimes: uninformative prior

Conjugate prior

- Beta is conjugate prior for Bernoulli distribution.
- What does that mean?
- Prior Beta, likelihood derived from Bernoulli $\rightarrow$ Posterior Beta.

Conjugate prior ctd.

- Normal is conjugate prior for Normal.
- Consider $\theta \sim N(\mu_0, \sigma_0^2)$
- $Y \sim N(\theta, \sigma^2)$

Conjugate prior ctd.

- Normal is conjugate prior for Normal.
- Consider $\theta \sim N(\mu_0, \sigma_0^2)$
- $Y \sim N(\theta, \sigma^2)$

$$P(\theta|Y) \propto \prod \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\theta)^2}{2\sigma^2}} * \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}}$$

$$\propto e^{-\frac{1}{2}(\frac{1}{\sigma^2}\sum(\theta-y_i)^2+\frac{1}{\sigma_0^2}(\theta-\mu_0)^2)}$$

$$\propto e^{-\frac{1}{2}(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2})(\theta-\frac{\frac{\mu_0}{\sigma_0^2}+\frac{n\bar{y}}{\sigma^2}}{\sigma_0^2+\frac{n}{\sigma^2}})^2}$$

Conjugate prior ctd.

- Normal is conjugate prior for Normal.
- Consider $\theta \sim N(\mu_0, \sigma_0^2)$
- $Y \sim N(\theta, \sigma^2)$

$$P(\theta|Y) \propto \prod \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\theta)^2}{2\sigma^2}} * \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}}$$

$$\propto e^{-\frac{1}{2}(\frac{1}{\sigma^2}\sum(\theta-y_i)^2+\frac{1}{\sigma_0^2}(\theta-\mu_0)^2)}$$

$$\propto e^{-\frac{1}{2}(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2})(\theta-\frac{\frac{\mu_0}{\sigma_0^2}+\frac{n\bar{y}}{\sigma^2}}{\sigma_0^2+\frac{n}{\sigma^2}})^2}$$

$$\theta|Y \sim N(\frac{\frac{\mu_0}{\sigma_0^2}+\frac{n\bar{y}}{\sigma^2}}{\sigma_0^2+\frac{n}{\sigma^2}}, (\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2})^{-1})$$

Point Estimation ctd.

- We can get more statistics, why?
- We know the whole posterior distribution.

Point Estimation ctd.

- We can get more statistics, why?
- We know the whole posterior distribution.
  - Expected value of θ?
  - What is the most likely value of θ? (MAP)
  - The probability of $\theta > 0$?

Inference

- Quick question: what's the difference between point estimation and inference?

Inference

- Quick question: what's the difference between point estimation and inference?
- Frequentist: Confidence Interval (Rely on infinitely sampling)
- Bayesian: We know the whole posterior distribution, and therefore:

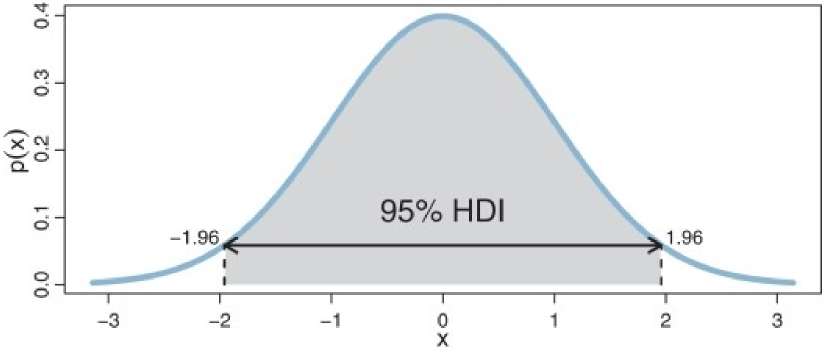
Inference

- Quick question: what's the difference between point estimation and inference?
- Frequentist: Confidence Interval (Rely on infinitely sampling)
- Bayesian: We know the whole posterior distribution, and therefore:
  - Credible Interval
  - We call set A is 95% Credible interval of θ if $Pr(\theta \in A) = \int_A P(\theta|y)d\theta = 0.95$
- Difference?

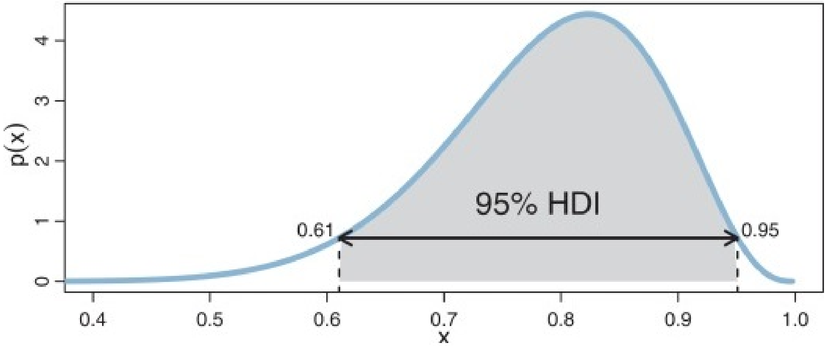
Inference ctd.

- Symmetric credible interval: bounded by α and $1 - \alpha$ quantile.
- Highest posterior density interval (HPD): the density inside the region has to be higher than at any point outside the interval.

HDI



HDI



HDI