Bayesian Statistics

Bayes Theorem

Brani Vidakovic

Professor

School of Industrial and Systems Engineering

Ingredients for Bayesian Inference

Before We Begin...

Before We Begin...

Ingrediants for Bayesian Inference

Observations $X_1, \ldots, X_n$ are modeled as

$$X_i \overset{\text{iid}}{\sim} f(x_i|\theta), \qquad i = 1, \ldots, n$$

Joint distribution of sample $X_1, \ldots, X_n$ is

$$f(x_1|\theta) \times f(x_2|\theta) \ldots f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

As a function of θ, the joint distribution $\prod_{i=1}^{n} f(x_i|\theta)$ is called **likelihood**

$$L(\theta|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

Likelihood Principle:

All information about the experiment is contained in the likelihood

Example: Each X_i in a sample is exponential Exp(λ). Let $X_1 = 2$, $X_2 = 3$, and $X_3 = 1$ be the observations. Then the likelihood is:

$$L(\lambda|x_1, x_2, x_3) = \lambda e^{-2\lambda} \times \lambda e^{-3\lambda} \times \lambda e^{-\lambda} = \boxed{\lambda^3 e^{-6\lambda}}$$

If the data are kept unspecified,

$$L(\lambda|x_1, x_2, x_3) = \lambda^3 e^{-\lambda \sum_{i=1}^{3} x_i}$$

Let θ be a parameter in $f(x|\theta)$.

A prior is distribution on θ,

$$\theta \sim \pi(\theta), \qquad \theta \in \Theta,$$

$$\Theta \equiv \text{parameter space.}$$

The joint distribution of (X, θ) is $h(x, \theta)$.

Marginal dsitribution of X is

$$m(x) = \int_{\Theta} h(x, \theta) d\theta$$

Recall Bayes' Rule for events

$$P(AH_i) = \boxed{P(A|H_i)P(H_i) = P(H_i|A)P(A)}$$

$$\Rightarrow P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A)}$$

By Analogy

$$h(x,\theta) = \boxed{f(x|\theta)\pi(\theta) = \pi(\theta|x)m(x)}$$

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

Bayes Theorem

Example: Normal likelihood + Normal prior

$$\begin{cases} x|\theta \sim N(\theta, \sigma^2), \ \ \sigma^2 \text{ known} \\ \theta \sim N(\mu, \tau^2), \ \ \ \mu, \tau^2 \text{ elicited} \end{cases}$$

called hyperparameters

$$h(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\} \times \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\}$$

$$\exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\} \equiv$$

$$\exp\left\{-\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2}\left(\theta - \left(\frac{\tau^2}{\sigma^2+\tau^2}x + \frac{\sigma^2}{\sigma^2+\tau^2}\mu\right)\right)^2 - \frac{1}{2(\sigma^2+\tau^2)}(x-\mu)^2\right\}$$

$$\begin{cases} \theta|X \sim N\left(\frac{\tau^2}{\sigma^2+\tau^2}x + \frac{\sigma^2}{\sigma^2+\tau^2}\mu, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right) \\ X \sim N(\mu, \sigma^2+\tau^2) \end{cases}$$

Posterior mean (mode, median) is

$$\frac{\tau^2}{\sigma^2+\tau^2}x+\frac{\sigma^2}{\sigma^2+\tau^2}\mu = w\times x+(1-w)\times\mu$$

observed **elicited**

$w=\frac{\tau^2}{\sigma^2+\tau^2}$; $\quad 1-w=\frac{\sigma^2}{\sigma^2+\tau^2}$

$\tau^2 \gg \sigma^2 \to w\approx 1,$ the posterior mean close to x

$\sigma^2 \gg \tau^2 \to w\approx 0,$ the posterior mean close to μ

Conjugate Families

- Note that for Normal Likelihood + Normal Prior → Posterior was Normal.

- If for likelihood f and prior π the prior and posterior belong to the same family of distributions, then the pair (f, π) is conjugate.

- If the pair (f, π) is conjugate, no need to calculate the normalizing constant in $f(x|\theta)\pi(\theta) \propto \pi(\theta|x)$.

Example: Binomial Likelihood and Beta Prior

$$X|p\sim\mathrm{Bin}(n,p);\quad f(x|p)=\binom{n}{x}p^x(1-p)^{n-x}$$

$$p\sim\mathrm{Be}(\alpha,\beta);\quad \pi(p)=\frac{1}{B(\alpha,\beta)}\,p^{\alpha-1}(1-p)^{\beta-1}$$

$$\pi(p|x)\propto f(x|p)\pi(p)=$$

$$=C\times p^{x+\alpha-1}(1-p)^{n-x+\beta-1}$$

$$\Rightarrow p|X\sim\mathrm{Be}(x+\alpha,n-x+\beta).$$

$$\boxed{X\sim\mathrm{Be}(\alpha,\beta)\Rightarrow EX=\frac{\alpha}{\alpha+\beta}}\qquad E(p|X)=\frac{x+\alpha}{n+\alpha+\beta}$$

$$E(p|X) = \frac{x+\alpha}{n+\alpha+\beta} = \frac{x}{n} \times \frac{n}{n+\alpha+\beta} + \frac{\alpha}{\alpha+\beta} \times \frac{\alpha+\beta}{n+\alpha+\beta}$$

$$\Rightarrow \boxed{\underset{w}{\frac{n}{n+\alpha+\beta}}} \times \boxed{\underset{\hat{p}}{\frac{x}{n}}} + \boxed{\underset{1-w}{\frac{\alpha+\beta}{n+\alpha+\beta}}} \times \boxed{\underset{\text{prior mean}}{\frac{\alpha}{\alpha+\beta}}}$$

Exercise: Show that Poisson Likelihood and Gamma prior form a conjugate pair.

Hint: $x|\lambda \sim \mathrm{Poi}(\lambda)$; $\quad f(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$

$\lambda \sim \mathrm{Ga}(\alpha, \beta)$; $\quad \pi(\lambda) = \frac{\lambda^{\alpha-1} \beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta\lambda}$

$\pi(\lambda|x) \propto \lambda^{x+\alpha-1} e^{-(1+\beta)\lambda}$

$X_1, \ldots, X_n \sim N(\theta, \sigma^2),\ \theta \sim N(\mu, \tau^2)$

$$\bar{X} \sim N\left(\theta, \frac{\sigma^2}{n}\right)$$

$$\theta | X_1, \ldots, X_n \equiv \theta | \bar{X} \sim N\left(\frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2}\bar{X} + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau^2}\mu, \frac{\frac{\sigma^2}{n}\tau^2}{\frac{\sigma^2}{n} + \tau^2}\right)$$

$X_1, \ldots, X_n \sim \mathrm{Poi}(\lambda);\ \lambda \sim \mathrm{Ga}(\alpha, \beta)$

$$\lambda | x \sim \mathrm{Ga}\left(\sum x_i + \alpha, n + \beta\right)$$

Show that posterior mean is weighted average of $\bar{X}$ and $\frac{\alpha}{\beta}$.

Only a handful conjugate pairs.

Limited modeling ability, but computation simple.

Examples

1. Jeremy's IQ
2. 10 flips of a coin; revisited
3. Poisson-Gamma

<u>Example</u>: Jeremy's IQ.

Jeremy models his IQ as $N(\theta, 80)$.

He is a GT student; prior on IQ of a GT student elicited as $N(110{,}120)$.

Jeremy takes IQ test and scores $X = 98$.

Find the posterior for θ, find Bayes' estimator of θ.

$$X|\theta \sim N(\theta, 80), \qquad \theta \sim N(110{,}120)$$

$$\theta|X \sim N\left(\frac{120}{80+120}\times 98 + \frac{80}{80+120}\times 110, \frac{80\times 120}{80+120}\right)$$

$$\sim N(102.8{,}48)$$

The mean of the posterior is a Bayes estimator of a parameter

$$\Rightarrow \ \widehat{\theta_B} = 102.8$$

Classical statistician will estimate θ as $\hat{\theta}_{MLE} = 98$.

If an average for $n = 5$ tests was $\bar{X} = 98$, then

$$\theta|\bar{X} \sim N\left(\frac{120}{\frac{80}{5} + 120} \times 98 + \frac{\frac{80}{5}}{\frac{80}{5} + 120} \times 110, \frac{\frac{80}{5} \times 120}{\frac{80}{5} + 120}\right)$$

$$\sim N(99.4118, 14.1176)$$

<u>Example:</u> Ten flips of a fair coin, revisited

We discussed conjugacy of Binomial likelihood + Beta prior

$$\left.\begin{array}{r} x|p \sim \mathrm{Bin}(n,p) \\ p \sim \mathrm{Be}(\alpha,\beta) \end{array}\right| \qquad p|X \sim \mathrm{Be}(X+\alpha, n-X+\beta)$$

Beta priors – excellent expressive power about the population proportion p.

Next: *betaplots.m*

A realistic prior on p:

$$p \sim \mathrm{Be}(500{,}500)$$

Thus, the posterior for likelihood $X|p \sim \mathrm{Bin}(10, p)$ and observed $X = 0$ is

$$\mathrm{Be}(0 + 500{,}10 - 0 + 500).$$

Thus, the posterior mean is

$$\hat{p}_B = \frac{0+500}{10+500+500} = \frac{500}{1010} = \boxed{0.495}$$

→ More realistic than frequentist's $\hat{p} = 0$.

<u>Example:</u> Poisson-Gamma.

Six plates containing large number of cells each are checked

$X_1 = 2, X_2 = 0, X_3 = 1, X_4 = 5, X_5 = 7, X_6 = 1$ are the number of "marked" cells on the corresponding plates.

Assume $X_i|\lambda \sim \text{Pois}(\lambda)$.

Estimated λ if the prior on λ is Gamma with mean 4 and variance $\frac{1}{4}$.

Likelihood

$$\frac{\lambda^2}{2!}e^{-\lambda}\times\frac{\lambda^0}{0!}e^{-\lambda}\times\frac{\lambda^1}{1!}e^{-\lambda}\times\frac{\lambda^5}{5!}e^{-\lambda}\times\frac{\lambda^7}{7!}e^{-\lambda}\times\frac{\lambda^1}{1!}e^{-\lambda}$$

$$\propto \lambda^{16}e^{-\lambda} \quad (\propto \text{"proportional to"})$$

Prior $\theta\sim\mathrm{Ga}(\alpha,\beta)$, $E\theta = \frac{\alpha}{\beta} = 4$

$$\mathrm{Var}\theta = \frac{\alpha}{\beta^2} = \frac{1}{4}$$

$$\Rightarrow \frac{1}{4} = \frac{\alpha}{\beta}\times\frac{1}{\beta} \Rightarrow \beta = 4\times 4 = 16$$

$$\alpha = 4\times 16 = 64$$

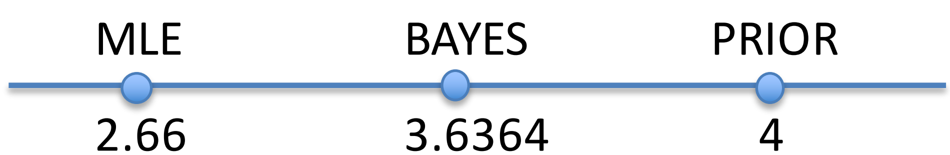
Posterior

$$\lambda | \underset{\sim}{X} \sim \mathrm{Ga}\left(\alpha + \sum x_i \,, n + \beta\right)$$

The mean of posterior is

$$\hat{\lambda}_B = \frac{\alpha + \sum x_i}{n + \beta} = \frac{64 + 16}{6 + 16} = \underline{3.6364}$$

Frequentist estimate: $\hat{\lambda}_{MLE} = \bar{X} = \frac{16}{6} = \underline{2.666}$



Summary