Bayesian Inference

- $X_1, ..., X_n \sim f(x_1, ..., x_n; \theta)$ [really, $f(x_1, ..., x_n|\theta)$ here]
- parameter θ is unknown, so treat it as a random variable
- a random variable must have a distribution, so

$$\theta \sim \pi(\theta),$$

called the prior distribution

- "only" objective is to find

$$\pi(\theta|x_1, ..., x_n) = \frac{f(x_1, ..., x_n|\theta)\pi(\theta)}{\int f(x_1, ..., x_n|\theta)\pi(\theta)d\theta},$$

the posterior distribution, by the Bayes theorem

Example 6.1 (pp. 93–95)

$$\begin{aligned} X_1, X_2, ..., X_n | \lambda &\overset{iid}{\sim} \text{Poisson}(\lambda) \\ \lambda &\sim \text{gamma}(\alpha, \beta) \end{aligned}$$

- want to compute

$$\pi(\lambda | x_1, ..., x_n) = \frac{f(x_1, ..., x_n | \lambda)\pi(\lambda)}{\int f(x_1, ..., x_n | \lambda)\pi(\lambda) d\lambda}$$

- mainly, the denominator

$$m(x_1, ..., x_n) = \int f(x_1, ..., x_n | \lambda)\pi(\lambda) d\lambda$$

$$
\begin{aligned}
m(x_1, ..., x_n) &= \int f(x_1, ..., x_n|\lambda)\pi(\lambda)\ d\lambda \\
&= \int \left[\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] \times \left[\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\right]\ d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)\prod_{i=1}^{n} x_i!} \times \frac{\Gamma(n\bar{x}+\alpha)}{(n+\beta)^{n\bar{x}+\alpha}} \times \\
&\qquad \int \underbrace{\frac{(n+\beta)^{n\bar{x}+\alpha}}{\Gamma(n\bar{x}+\alpha)}\lambda^{n\bar{x}+\alpha-1}e^{-(n+\beta)\lambda}}_{\text{gamma}(n\bar{x}+\alpha,n+\beta)}\ d\lambda \\
&= \frac{\Gamma(n\bar{x}+\alpha)}{\Gamma(\alpha)\prod_{i=i}^{n} x_i!}\left[\frac{\beta}{n+\beta}\right]^\alpha \left[\frac{1}{n+\beta}\right]^{x_1} \cdots \left[\frac{1}{n+\beta}\right]^{x_n}
\end{aligned}
$$

$$
\Rightarrow \qquad \lambda|X_1, ..., X_n \sim \text{gamma}(n\bar{X}+\alpha, n+\beta) \qquad [\text{why}]
$$

Posterior Expectation vs MLE

- a kind of "Bayesian point estimate",

$$\widehat{\lambda}_{bayes} = \mathbb{E}(\lambda|X_1, ..., X_n) = \frac{n\bar{X} + \alpha}{n + \beta}$$

- compare with frequentist point estimate,

$$\widehat{\lambda}_{mle} = \bar{X},$$

as if given an additional β observations with α incidences

Think How important is the choice of (α, β) when n is large?

$$\widehat{\lambda}_{bayes} = \frac{n\bar{X} + \alpha}{n + \beta} = \frac{\bar{X} + (\alpha/n)}{1 + (\beta/n)} \longrightarrow \bar{X} = \widehat{\lambda}_{mle}$$

The Gamma Prior

- choice of gamma mainly so that integral in denominator is tractable

- is a conjugate prior for the Poisson

- but still must choose the parameters (α, β), e.g., gamma$(1/2, 1/2) = \chi^2_{(1)}$, or whatever

- if following the so-called Jeffreys' rule, would choose

$$\text{gamma}(1/2, 0) \quad = \quad \lim_{\varepsilon \downarrow 0} \text{gamma}(1/2, \varepsilon),$$

  but we will skip this topic — not exactly a short story

Simultaneous Estimation

General	Example
$X_i \sim f(\cdot\|\theta_i), \quad i = 1, ..., n$	$X_i \sim \text{Poisson}(\lambda_i)$
$\theta_i \sim \pi(\cdot\|\psi)$	$\lambda_i \sim \text{gamma}(\alpha, \beta)$
$\pi(\theta_i\|x_i, \psi) = \dfrac{f(x_i\|\theta_i)\pi(\theta_i\|\psi)}{\int f(x_i\|\theta_i)\pi(\theta_i\|\psi)d\theta_i}$	$\lambda_i\|X_i, \alpha, \beta \sim \text{gamma}(X_i + \alpha, 1 + \beta)$

Remark Each individual problem (indexed by i) can still have multiple observations $X_{i1}, ..., X_{im_i}$ "as usual", in which case the left side simple leads to $\pi(\theta_i|x_{i1}, ..., x_{im_i}; \psi)$ and the right side to $\text{gamma}(m_i\bar{X}_i + \alpha, m_i + \beta)$. Here, we simplify with $m_i = 1 \ \forall \ i$.

Empirical Bayes

- do not specify ψ, e.g., choose $(\alpha, \beta) = (\frac{1}{2}, \frac{1}{2})$ or whatever
- can integrate out θ_i to obtain the marginal distribution of X_i:

$$m(x_i|\psi) = \int f(x_i|\theta_i)\pi(\theta_i|\psi)d\theta_i$$

- $X_1, ..., X_n \overset{iid}{\sim} m(\cdot|\psi)$ all contain some information about ψ
- can obtain an estimate, say $\widehat{\psi}$, using all of $X_1, ..., X_n$; and use the "plug-in" posterior, $\pi(\theta_i|x_i, \widehat{\psi})$
- e.g., obtain $(\widehat{\alpha}, \widehat{\beta})$ using all $X_1, ..., X_n$, and use Gamma$(X_i + \widehat{\alpha}, 1 + \widehat{\beta})$ for posterior inference

Remark Alternatively, can put another prior on ψ. Called hierarchical Bayes. Can keep on doing it but eventually must stop.

Remarks

Will use $\alpha = 1, \beta = \theta$ in examples to follow, so that there is only one prior parameter (θ) to "worry about" rather than two (α, β).

Will also look at hierarchical Bayes (later).

Example 6.2 (pp. 96–97)

In a data set of $n = 100$ individuals, some have had no accident while various others have had one, two, or three accidents.

None ($X_i = 0$)	One ($X_i = 1$)	Two ($X_i = 2$)	Three ($X_i = 3$)	Total (n)
70	20	8	2	100

Will model $X_i \sim \text{Poisson}(\lambda_i)$, and use prior $\lambda_i \sim \text{gamma}(1, \theta)$ — aka $\text{exponential}(\theta)$. Would like to compare

$$\widehat{\lambda}^{(mle)}_{Amy} \quad \text{vs} \quad \widehat{\lambda}^{(eb)}_{Amy}; \qquad \widehat{\lambda}^{(mle)}_{Bob} \quad \text{vs} \quad \widehat{\lambda}^{(eb)}_{Bob},$$

where Amy is one of those who has had one accident ($X_{Amy} = 1$) and Bob, one of those who has had two ($X_{Bob} = 2$).

The J̲ames-S̲tein Estimator

$X_i \sim \mathrm{N}(\theta_i, \sigma^2), \quad X_1, X_2, ..., X_n$ all independent, $\quad \sigma^2$ known

Fact The celebrated James-Stein estimator,

$$\widehat{\theta}_i^{(js)} = \bar{X} + \left[1 - \frac{(n-3)\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right](X_i - \bar{X}),$$

can be shown to have smaller MSE,

$$\mathbb{E}(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2) = \mathbb{E}\left[\sum_{i=1}^n (\widehat{\theta}_i - \theta_i)^2\right],$$

than $\widehat{\theta}_i^{(mle)} = X_i$ for $n > 3$. But, mathematics aside, what's the intuition? Answer: It is an empirical Bayes estimator of θ_i under a conjugate prior, $\theta_i \sim \mathrm{N}(\mu, \tau^2)$.

Key Steps

(a) posterior of $\theta_i | X_i$ is normal with mean

$$\left[\frac{\tau^2}{\sigma^2+\tau^2}\right] X_i + \left[\frac{\sigma^2}{\sigma^2+\tau^2}\right]\mu$$

$$= \quad \mu + \left[1 - \frac{\sigma^2}{\sigma^2+\tau^2}\right](X_i - \mu)$$

and variance ...

(b) marginal of $X_i \sim \mathrm{N}(\mu, \sigma^2+\tau^2)$

$$\Rightarrow \quad \text{from which to estimate } \mu \text{ and } \frac{1}{\sigma^2+\tau^2}$$

Details for (b)

$$\widehat{\mu} = \bar{X} \quad \text{(of course; what else)}$$

$$\mathbb{E}\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}\right] = \sigma^2 + \tau^2 \quad \text{(related to } \chi^2\text{);}$$

$$\mathbb{E}\left[\frac{n-3}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] = \frac{1}{\sigma^2 + \tau^2} \quad \text{(related to Inv-}\chi^2\text{;}$$

skip details)

Intuition of James-Stein

$$\mathbb{E}(\theta_i|X_i) = \mu + \left[1 - \frac{\sigma^2}{\sigma^2 + \tau^2}\right](X_i - \mu)$$

$$\widehat{\theta}_i^{(js)} = \bar{X} + \left[1 - \frac{(n-3)\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right](X_i - \bar{X})$$