Signle-Paprameter Models

Gwangsu Kim

JBNU

Second semester of 2024

Binomial Model I

Prior: $\theta \sim \mathcal{B}eta(\alpha, \beta)$

Model: $y \sim \mathcal{B}(n, \theta)$

Posterior:

$$\begin{aligned} p(\theta|y) &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \times \theta^{y}(1-\theta)^{n-y} \\ &= \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1} \end{aligned}$$

Thus, $\theta|y \sim \mathcal{B}eta(\alpha + y, \beta + n - y)$.

Binomial Model II

- A class of prior distributions is called conjugate if the posterior distribution also belongs to the class. The beta distribution is a conjugate family for binomial model.

Posterior Summary I

- After the data were observed, all the information about θ is contained in the posterior. The followings are usual posterior summaries.
- Numerical summary: mean, median, mode, standard deviation, IQR, quantiles
- Interval summary (credible set): equal to the percentile interval, high posterior density (HPD) interval

Posterior Summary II

- For binomial model,

$$
\begin{aligned}
Mod(\theta|y) &= \frac{\alpha+y-1}{\alpha+\beta+n-2}\ (n \geq 2, 1 \geq y \geq n-1), \\
E(\theta|y) &= \frac{\alpha+y}{\alpha+\beta+n} = \frac{\alpha+\beta}{\alpha+\beta+n} \times \frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n} \times \frac{y}{n}.
\end{aligned}
$$

- Note that it is the weighted average of the prior mean and the maximum likelihood estimator (MLE) with weights $\alpha+\beta$ and n. For this reason, sometimes $\alpha+\beta$ is called the prior sample size.

Posterior Summary III

- The $100(1-\alpha)\%$ equal-tail credible set is

$$\left(\mathcal{B}eta(\alpha+y,\beta+n-y)_{1-\alpha/2},\mathcal{B}eta(\alpha+y,\beta+n-y)_{\alpha/2}\right).$$

Example and Uniform prior I

- Example: Probability of Female Birth

  Between years 1745 and 1770, a total of 241, 945 girls and 251, 527 boys were born in Paris. Let θ be the probability of female birth. Assuming no prior knowledge about θ, we use the uniform prior on θ, i.e., $\mathcal{B}eta(1, 1)$. The posterior is

$$\theta|y \sim \mathcal{B}eta\Big(1 + 241, 945, 1 + 251, 527\Big).$$

Example and Uniform prior II

- Bayes computed the marginal probability of y with uniform prior.

$$
\begin{aligned}
p(y) &= \int_0^1 \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta \\
&= \binom{n}{y} \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} \\
&= 1/(n+1), y = 0, \ldots, n.
\end{aligned}
$$

Thus, all possible values of y are equally likely a priori. This is also an appealing reason for using the uniform prior.

Normal Model: unknown mean and known variance I

Prior: $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$

Model: $y_1, \ldots, y_n \sim \mathcal{N}(\theta, \sigma^2)$

Normal Model: unknown mean and known variance II

Posterior:

$$
\begin{aligned}
& p(\theta|y_1,\ldots,y_n) \\
\propto\ & \exp\left(-\frac{1}{2\sigma_0^2}(\theta-\mu_0)^2\right)\prod_{i=1}^{n}\exp\left(-\frac{1}{2\sigma^2}(y_i-\theta)^2\right) \\
\propto\ & \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)\theta^2+\left(\frac{\mu_0}{\sigma_0^2}+\frac{n\bar{y}}{\sigma^2}\right)\theta\right] \\
\propto\ & \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{1/\sigma_0^2}{1/\sigma_0^2+n/\sigma^2}\mu_0+\frac{n/\sigma^2}{1/\sigma_0^2+n/\sigma^2}\bar{y}\right)\right)^2\right] \\
\propto\ & \phi(\theta;\mu_n,\sigma_n^2),
\end{aligned}
$$

Normal Model: unknown mean and known variance III

where $\phi(\cdot;\mu,\sigma^2)$ is the p.d.f. of the normla distribution with mean μ and varaince σ^2, $\mu_n = \frac{1/\sigma_0^2}{1/\sigma_0^2+n/\sigma^2}\mu_0 + \frac{n/\sigma^2}{1/\sigma_0^2+n/\sigma^2}\bar{y}$, and $\sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}$.

- The posterior information (precision or inverse of variance) is the sum of the prior information and sample information:

$$1/\sigma_n^2 = 1/\sigma_0^2 + n/\sigma^2.$$

Normal Model: unknown mean and known variance IV

- If we scale $\sigma_0^2 = \sigma^2/k$, k can be interpreted as the prior sample size. As $n \to \infty$, the sample information dominates the prior information

- The posterior mean is the weighted average of the prior mean and the sample mean with weights of the prior information and the sample information.

Normal Model: unknown mean and known variance V

- As $n \to \infty$, the sample mean dominates the prior mean and the posterior converges to $\mathcal{N}(\bar{y}, \sigma^2/n)$.

- Thus, $p(\theta|y_1, \ldots, y_n) \approx \mathcal{N}(\bar{y}, \sigma^2/n)$ and

$$E(\theta|y_1, \ldots, y_n) \approx \bar{y}.$$

Normal Model: unknown mean and known variance VI

- The $100(1-\alpha)\%$ credible set of θ is approximately

$$\left(\bar{y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

which is the frequentist confidence interval.

Normal Model: unknown mean and known variance VII

- In summary, the posterior mean and credible set can be approximated by the MLE and the confidence interval. This is true for most parametric model.

Normal Model: known mean and unknown variance I

- For the normal variance, it is easier to deal with the precision $\tau^2 = 1/\sigma^2$.

  Prior:

  $\tau^2 = 1/\sigma^2 \sim \mathcal{G}amma(\alpha, \beta),\ \nu_0\sigma_0^2/\sigma^2 \sim \chi^2_{\nu_0}\ (\nu_0 = 2\alpha, \nu_0\sigma_0^2 = 2\beta)$

  Note that $\chi^2_a \equiv \mathcal{G}amma(a/2, 1/2)$

  Model: $y_1, \ldots, y_n \sim \mathcal{N}(\theta, \sigma^2)$

Normal Model: known mean and unknown variance II

Posterior:

$$
\begin{aligned}
& p(\tau^2|y_1,\ldots,y_n) \\
\propto\ & (\tau^2)^{\alpha-1}\exp(-\beta\tau^2)\prod_{i=1}^{n}(\tau^2)^{1/2}\exp\Big(-\frac{1}{2}\tau^2(y_i-\theta)^2\Big) \\
\propto\ & (\tau^2)^{\alpha+n/2-1}\exp\Big(-\Big\{\beta+\sum_{i=1}^{n}\frac{1}{2}(y_i-\theta)^2\Big\}\tau^2\Big) \\
\propto\ & \mathcal{G}amma\left(\alpha+n/2,\beta+\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2\right).
\end{aligned}
$$

$(\nu_0\sigma_0^2+n\nu)/\sigma^2\sim\chi^2_{\nu_0+n}$ $(n\nu=\sum_{i=1}^{n}(y_i-\theta)^2)$.

Normal Model: known mean and unknown variance III

- $E(\tau^2|y_1, \ldots, y_n)$ is

$$
\begin{aligned}
& \frac{\alpha + n/2}{\beta + \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2} \\
= \quad & \frac{\alpha}{\beta}\frac{\beta}{\beta + \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2} + \frac{n}{\sum_{i=1}(y_i - \theta)^2}\frac{\frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2}{\beta + \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2}.
\end{aligned}
$$

- Thus, in eliciting the prior Gamma parameter, we can set $\frac{\alpha}{\beta}$ = prior precision estimate = 1 / prior variance estimate. Also β effects the prior sample size.

Poisson and Exponential Model I

- Poisson Model

  Prior: $\theta \sim Gamma(\alpha, \beta)$

  Model: $y_1, \ldots, y_n \sim Pois(\theta)$

  Posterior:

$$
\begin{aligned}
p(\theta|y) &\propto \theta^{\alpha-1}\exp(-\beta\theta)\prod_{i=1}^{n}\frac{\theta^{y_i}\exp(-\theta)}{y_i!} \\
&\propto \theta^{\alpha+\sum_{i=1}^{n} y_i - 1}\exp\left(-(\beta+n)\theta\right) \\
&\propto Gamma\left(\alpha+\sum_{i=1}^{n} y_i, \beta+n\right).
\end{aligned}
$$

Poisson and Exponential Model II

- $E(\theta|y_1,\ldots,y_n) = \frac{1}{\beta+n}\alpha + \frac{n}{\beta+n}\frac{\sum_{i=1}^n y_i}{n}$.
- $\text{Var}(\theta|y_1,\ldots,y_n) = \frac{\alpha+\sum_{i=1} y_i}{(\beta+n)^2}$.

- As $n \to \infty$, $E(\theta|y_1,\ldots,y_n) \to \frac{\sum_{i=1}^n y_i}{n}$ and $\text{Var}(\theta|y_1,\ldots,y_n) \to 0$ when $\frac{\sum_{i=1}^n y_i}{n}$ is finite for all n.

Poisson and Exponential Model III

- Exponential Model

  Prior: $\theta \sim \mathcal{G}amma(\alpha, \beta)$

  Model: $y_1, \ldots, y_n \sim Exp(\theta)$

  Posterior:

$$
\begin{aligned}
p(\theta|y) &\propto \theta^{\alpha-1}\exp(-\beta\theta)\prod_{i=1}^{n}\theta\exp(-y_i\theta) \\
&\propto \theta^{\alpha+n-1}\exp\left(-(\beta+\sum_{i=1}^{n}y_i)\theta\right) \\
&\propto \mathcal{G}amma\left(\alpha+n, \beta+\sum_{i=1}^{n}y_i\right).
\end{aligned}
$$

Poisson and Exponential Model IV

- $E(\theta|y_1, \ldots, y_n) = \frac{\alpha+n}{\beta+\sum_{i=1}^n y_i}$.
- Var$(\theta|y_1, \ldots, y_n) = \frac{\alpha+n}{(\beta+\sum_{i=1}^n y_i)^2}$.

- As $n \to \infty$, $E(\theta|y_1, \ldots, y_n) \to \frac{n}{\sum_{i=1}^n y_i}$ and Var$(\theta|y_1, \ldots, y_n) \to 0$ when $\frac{\sum_{i=1}^n y_i}{n}$ is finite for all n.

Nonconjugate Analysis I

- Often the prior distribution elicited does not belong to a conjugate family. In this case, the posterior distribution may not be a well known distribution and the posterior summaries, mean, median and quantiles, are not available in closed form.

- In this case, a simple way around is to generate a random sample (posterior sample) from the posterior and approximate the posterior with the empirical distribution of the posterior sample.

Nonconjugate Analysis II

- For example, let $\theta_1, \ldots, \theta_m$ be a posterior sample from the posterior $p(\theta|y)$. Then

$$\begin{aligned} \text{posterior mean} &\approx \frac{1}{m}\sum_{i=1}^{m}\theta_i. \\ \text{posterior } \alpha\text{-quantile} &= \theta\left([\alpha m]\right). \end{aligned}$$

Nonconjugate Analysis III

- Inverse-Probability Transformation

1. Let $U \sim (0, 1)$ and F is a continuous cdf. Then

$$X = F^{-1}(U) \sim F$$

Exmaple: Let $F(x) = 1 - e^{-\lambda x}, x \geq 0$, and $U \sim U(0, 1)$. Then

$$X = -\frac{1}{\lambda} \log(1 - U) \sim Exp(\lambda).$$

Remark: The above is still true even for general cdf F with

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}.$$

Nonconjugate Analysis IV

- Accpetance-Rejection Method

  1. Suppose we wish to generate a random number following density f, from which it is difficult to generate a random variable directly. Suppose g is another density from which we can generate random sample and

  $$f(x) \le Mg(x), x \in \text{supp}(f).$$

  2. Algorithm

     1. Generate $X \sim g$ and $U \sim U(0, 1)$ $(X \perp U)$.
     2. Accept $Y = X$ if $U \le f(X)/Mg(X)$.
     3. Otherwise, return to 1.

Nonconjugate Analysis V

3 Proof:

$$
\begin{aligned}
P(Y \leq y) &= P\left(X \leq y \middle| U \leq \frac{f(X)}{Mg(X)}\right) = \frac{P\left(X \leq y, U \leq \frac{f(X)}{Mg(X)}\right)}{P\left(U \leq \frac{f(X)}{Mg(X)}\right)} \\
&= \frac{EP\left(X \leq y, U \leq \frac{f(X)}{Mg(X)} \middle| X\right)}{EP\left(U \leq \frac{f(X)}{Mg(X)} \middle| X\right)} = \frac{E\left[I(X \leq y)\frac{f(X)}{Mg(X)}\right]}{E\left[\frac{f(X)}{Mg(X)}\right]} \\
&= \frac{\int_{x \leq y} \frac{f(x)}{Mg(x)} g(x)dx}{\int \frac{f(x)}{Mg(x)} g(x)dx} = \frac{\int_{x \leq y} f(x)dx}{\int f(x)dx} = \int_{x \leq y} f(x)dx.
\end{aligned}
$$

Non-informative Prior I

- Criticism on subjective prior: The main criticism of the Bayesian statistics has been use of subjective prior in the data analysis. Although it is essential to use the prior information in some analyses, many classical statisticians feel uncomfortable when different analysts have different results. Bayesians answer this criticism as follows

Non-informative Prior II

- When use of a statistical model is subjective, how can use of prior be criticized?
- In almost all problems, we do have prior knowledge.
- With a large sample size, objectivity will be attained.

Non-informative Prior III

- To respond to the criticism on prior, some Bayesians (object Bayesians) made effort to produce noninformative prior or prior with no information. In normal model with unknown mean and known variance, use of the prior $\pi(\theta) = 1$, yields

$$
\begin{aligned}
p(\theta|y) &\propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(y_i - \theta)^2\right) \\
&\propto \exp\left(-\frac{n}{2\sigma^2}(\theta - \bar{y})^2\right) \\
&\propto \mathcal{N}(\bar{y}, \sigma^2/n).
\end{aligned}
$$

Non-informative Prior IV

- This posterior corresponds to the posterior with the prior variance = ∞. Note that although the prior is improper, i.e., it does not integrate to 1, the posterior is proper.

- In the normal model with known mean and unknown variance, the prior $\pi(\log\sigma) = 1$ or $\pi(\sigma) = 1/\sigma$ or $\pi(\sigma^2) = 1/\sigma^2$ or $\pi(\tau^2) = 1/\tau^2$ can be use to produce similar results, yielding

$$\tau^2|y \sim \mathcal{G}amma\left(n/2, \sum_{i=1}^{n}(y_i-\theta)^2/2\right).$$

Non-informative Prior V

- Jeffreys' prior: $\pi(\theta) \propto \left[\det\left(I\left(\theta\right)\right)\right]^{1/2}$.