STAT 770 Sep. 14 Lecture Part A

Bayesian Inference for Binomial and Multinomial

Reading for this lecture:

Section 1.6 in Agresti, also Section 17.5 for some history.

We do a 2-slide review of basic Bayesian ideas, in the old-style setting where posteriors can be obtained analytically because the priors have the special *conjugate* form.

Then we apply the theory directly to estimates and "credible intervals" (the Bayesian analogue of CIs) for multinomial data.

General Setup for Bayesian Inference, I

(discrete) data vector $\underline{\mathbf{Y}} \sim p(\underline{\mathbf{y}}, \beta)$ assumed governed by a parametric model, $\beta \in \mathcal{U} \subset \mathbb{R}^d$

Bayesians view the unknown β as a random d-vector and $p(\underline{\mathbf{y}}, \beta)$ as conditional (density or) prob. mass fcn. $P(\underline{\mathbf{Y}} = \underline{\mathbf{y}} \,|\, \beta)$

Before data: assume known prior density $\beta \sim g(b)$ for r.v. β

After data-collection: posterior density governs probabilities for β r.v. given the fixed dataset, $P(\beta \in B \,|\, \underline{\mathbf{Y}}) = \int_B f(b \,|\, \underline{\mathbf{Y}})\, db$

$$f(\beta \,|\, \underline{\mathbf{y}}) = g(\beta) p(\underline{\mathbf{y}} \,|\, \beta) \Big/ \int g(b) p(\underline{y} \,|\, b)\, db\,, \qquad f(\beta \,|\, \underline{\mathbf{Y}}) = g(\beta)\, \frac{L(\beta, \underline{\mathbf{Y}})}{P(\underline{\mathbf{Y}} = \underline{\mathbf{y}})}$$

General Setup for Bayesian Inference, II

Estimation: summary location-statistic for posterior density,

e.g posterior mean $\tilde{\beta}^{Bayes} = E(\beta \,|\, \underline{\mathbf{Y}}) = \int b\, f(\beta \,|\, \underline{\mathbf{Y}})\, db$

Other choices like posterior median OK too; mean minimizes posterior MSE $E\Big((\beta - \tilde{\beta})^2 \,\Big|\, \underline{\mathbf{Y}}\Big)$

Credible Interval: Interval or region $B(\underline{\mathbf{Y}}) \subset \mathbb{R}^d$ such that

$$P(\beta \in B(\underline{\mathbf{Y}}) \,|\, \underline{\mathbf{Y}}) \;=\; \int_{B(\underline{\mathbf{Y}})} f(b \,|\, \underline{\mathbf{Y}})\, db \;=\; 1 - \alpha$$

If $\beta \in \mathbb{R}$, $d = 1$, take $B(\underline{\mathbf{Y}}) = \Big(F^{-1}(\frac{\alpha}{2} \,|\, \underline{\mathbf{Y}}),\, F^{-1}(1 - \frac{\alpha}{2} \,|\, \underline{\mathbf{Y}})\Big)$

interval between quantiles of posterior d.f. $F(b \,|\, \underline{\mathbf{y}}) \;=\; \int_{-\infty}^{b} f(x \,|\, \underline{\mathbf{y}})\, dx$

Conjugate Priors

With respect to data model $p(\underline{\mathbf{y}}, b)$, parametrized family of prior & posterior densities $g(b, \mu)$ is called **conjugate** if the posterior for prior $\beta \sim g(b, \mu)$ is $f(b \,|\, \underline{\mathbf{Y}}) \equiv g(b, \mu^*)$, $\mu^* = \mu^*(\underline{\mathbf{Y}})$, i.e.,

$$g(b)\, p(\underline{\mathbf{y}}, b) \Big/ \int g(x)\, p(\underline{\mathbf{y}}, x)\, dx \equiv g(b, \mu^*(\underline{\mathbf{y}}))$$

We will return to this definition later and recall that there is generally a conjugate prior family when the joint probability mass (or density) function has the exponential family form

$$p(\underline{\mathbf{y}}, \beta) = h(\mathbf{x}) \exp\Big(q(\beta)' T(\underline{\mathbf{y}}) - C(\beta) \Big)$$

(This is a review topic: you should read about it at some point.) We now discuss a special case.

Multinomial and Dirichlet Example

Let $\underline{\mathbf{Y}} = \{X_a\}_{a=1}^{n}$ with $P(X_a = c) = \beta_c$ if $1 \le c \le d$, and $P(X_a = d+1) = 1 - \beta_1 - \cdots - \beta_d$

Consider prior $g(b, \underline{\mu}) = c(\underline{\mu}) \left[\prod_{j=1}^{d} b_j^{\mu_j - 1} \right] (1 - b_1 - \cdots - b_d)^{\mu_{d+1} - 1}$

where $b_j > 0$, $j = 1, \ldots, d$, such that $b_1 + \cdots + b_d < 1$, and $c(\underline{\mu})$ is an integration constant, $\underline{\mu} \in (0, \infty)^{d+1}$.

NB. usually, equivalently, defined on **unit (d+1)-dim simplex**

Density is called Dirichlet($\underline{\mu}$): special case $d = 1$ is Beta(μ_1, μ_2)

Verification of Conjugacy Property, Multinomial Example

Denote $\beta_{d+1} = 1 - \beta_1 - \cdots - \beta_d$. Then

$$g(\beta, \underline{\mu}) \cdot P(\underline{\mathbf{Y}} = \underline{\mathbf{y}} \,|\, \beta) \;=\; c(\underline{\mu}) \cdot \left[\prod_{j=1}^{d+1} \beta_j^{\mu_j - 1} \right] \cdot \prod_{a=1}^{n} \beta_{y_a}$$

$$= \; c(\underline{\mu}) \cdot \Big[\prod_{j=1}^{d+1} \beta_j^{\mu_j - 1} \Big] \cdot \prod_{j=1}^{d+1} \beta_j^{N_j} \qquad \text{where} \qquad N_j \equiv \sum_{a=1}^{n} I_{[y_a = j]}$$

Therefore posterior is $f(\beta \,|\, \underline{\mathbf{Y}}) \;=\; \dfrac{c(\mu)}{P(\underline{\mathbf{Y}}{=}\underline{\mathbf{y}})\Big|_{\underline{\mathbf{y}}=\underline{\mathbf{Y}}}} \prod_{j=1}^{d+1} \beta_j^{\mu_j + N_j - 1}$

Since this density as fcn of β is $\propto$ $g(\beta,\; \underline{\mu} + \underline{\mathbf{N}})$ Dirichlet
(where $\underline{\mathbf{N}} = (N_1, \ldots, N_{d+1})$), it is $=$ $g(\beta,\; \underline{\mu} + \underline{\mathbf{N}})$

Beta and Dirichlet: Densities & identities

The Beta(α_1, α_2) density on $(0,1)$ is $\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1}(1-x)^{\alpha_2-1}$

Dirichlet$(\underline{\mu})$ density on $\{(x_1,\ldots,x_d) : \; x_j > 0, \; x_1 + \cdots + x_d < 1\}$

is $\quad = \Gamma(\sum_j \mu_j) \prod_{j=1}^{d+1} \left(x_j^{\mu_j - 1} / \Gamma(\mu_j) \right) \; , \quad x_{d+1} \equiv 1 - x_1 - \cdots - x_d$

Can verify integration constants using Jacobian change-of-variable formula in the identity: for $X_j \sim$ Gamma$(\mu_j, 1)$ **indep.**

$$(X_1, \ldots, X_d) / \sum_{j=1}^{d+1} X_j \sim \textbf{Dirichlet}(\underline{\mu})$$

$(B_1, \ldots, B_d) \sim$ Dirichlet$(\underline{\mu}) \;\Rightarrow\; \sum_{j=1}^{k} B_j \sim$ Beta$(\sum_{j=1}^{k} \mu_j, \sum_{j=k+1}^{d+1} \mu_j)$

Beta and Dirichlet Means & Variances

To find mean of Beta: $\int_0^1 x \cdot \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1}(1-x)^{\alpha_2-1}\,dx$

$$= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \Big/ \left[\frac{\Gamma(\alpha_1+\alpha_2+1)}{\Gamma(\alpha_1+1)\Gamma(\alpha_2)}\right] = \frac{\Gamma(\alpha_1+\alpha_2)\Gamma(\alpha_1+1)}{\Gamma(\alpha_1+\alpha_2+1)\Gamma(\alpha_1)}$$

$$\text{Mean} = \frac{\alpha_1}{\alpha_1+\alpha_2}, \quad \text{similarly} \quad \text{Variance} = \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)}$$

For $\sum_{j=1}^{k} B_j$ based on $\underline{B} \sim \text{Dirichlet}(\underline{\mu})$, put $\alpha_1 = \sum_{j=1}^{k} \mu_j$, and $\alpha_2 = \sum_{j=k+1}^{d+1} \mu_j$ to get mean and variance formulas.