

Lecture 14: Mixture Distributions in GLMMs

Author: Nick Reich / Transcribed by Hachem Saddiki, edited by Gregory Guranich

Course: Categorical Data Analysis (BIOSTATS 743)

Mixture Distributions

- ▶ A mixture distribution is a probability distribution derived from a convex (weighted) combination of individual probability distributions, called **mixture components**.
- ▶ The weights associated with each mixture component are called **mixture weights**; these are non-negative and sum to 1.
- ▶ **Notation:** Given a set of mixture components $g_k(x)$ and weights w_k indexed by $k = 1, \dots, K$, the mixture distribution can be written as

$$f(x) = \sum_{k=1}^K w_k g_k(x) \tag{1}$$

where $\sum_{k=1}^K w_k = 1$ and $w_k \geq 0 \ \forall k$.

Mixture Distributions (cont'd)

- ▶ **Mixture** distributions are not to be confused with the **convolution** of probability distributions.
- ▶ The former models the **distribution** of a random variable as the weighted sum of two or more mixture distributions; while the latter models the **value** of a random variable as the sum of the **values** of two or more random variables.
- ▶ Mixture distributions are typically used to model a statistical population with subpopulations that are believed to share similar characteristics.
- ▶ Mixture distributions are formulated as **hierarchical models** with hidden (latent) factors, often making exact statistical inference an intractable task.

GLMM as a model for mixture distributions

- ▶ Consider an analysis of children's gender in relation to their mother. Define a binary random variable Y_{ij} which equals 1 if the j -th child born to the i -th mother is a boy.
- ▶ We write a GLMM model for the outcome as follows

$$\begin{aligned}Y_{ij}|\alpha_i &\sim \textit{Bernoulli}(\pi_i) \\ \pi_i &= \textit{logit}^{-1}(\alpha_i) \\ \alpha_i &\sim \textit{Normal}(0, \sigma_\alpha^2)\end{aligned}$$

- ▶ In this example, the population of interest are the children, and the mixture distribution accounts for subpopulations of children that share the same mother. The rationale is that these children, being from the same mother, would share similar (genetic or other) characteristics.

GLMM as a model for mixture distributions (cont'd)

- ▶ Using the law of total expectation, we can compute the expectation of Y_i as:
$$E[Y_i] = E[E[Y_i|\alpha_i]] = E[\pi_i] = E[\text{logit}^{-1}(\alpha_i)]$$
- ▶ Unfortunately, an exact expression for $E[Y_i]$ involves a complicated integral form, and so exact inference is intractable for this model formulation.
- ▶ Nonetheless, in this simple setting, one can appeal to a Monte Carlo simulation to approximate the expectation of interest.
- ▶ Alternatively, one can opt for a different model formulation that is amenable to exact expressions for the expectation and variance of interest.

Beta-Binomial Model

- ▶ Following the same example of children's gender and their mother, we can formulate a Beta-Binomial model for the outcome Y_{ij} as follows:

$$\begin{aligned}Y_i|\pi_i &\sim \textit{Binomial}(n_i, \pi_i) \\ \pi_i &\sim \textit{Beta}(\alpha, \beta)\end{aligned}$$

- ▶ In here, we model π_i as a draw from a Beta distribution.
- ▶ It is easy to see that the Y_i reduces to a Bernoulli distribution when $\alpha = \beta = 1$.

Beta-Binomial Model (cont'd)

- ▶ Another appealing property of the model is that the Beta distribution is conjugate to the Binomial distribution.
- ▶ This leads to a closed form expression for the mean and variance of Y_i :

$$E[Y_i] = \frac{n_i \alpha}{\alpha + \beta} \quad (2)$$

$$Var(Y_i) = \frac{n_i \alpha \beta (\alpha + \beta + n_i)}{(\alpha + \beta)^2 (\alpha + \beta + 1)} \quad (3)$$

Shrinkage Effect for Beta-Binomial

- We can use the conjugacy of the Beta prior and Binomial likelihood to derive a distributional form for the posterior $\pi_i|y_i$.

$$y_i|\pi_i \sim \text{Binomial}(n_i, \pi_i)$$

$$\pi_i|\alpha, \beta \sim \text{Beta}(\alpha, \beta)$$

$$\begin{aligned} p(\pi_i|y_i, \alpha, \beta) &\propto p(y_i|\pi_i)p(\pi_i|\alpha, \beta) \\ &\propto \pi_i^{y_i}(1 - \pi_i)^{n_i - y_i} * \pi_i^{\alpha - 1}(1 - \pi_i)^{\beta - 1} \\ &\propto \pi_i^{y_i + \alpha - 1}(1 - \pi_i)^{n_i - y_i + \beta - 1} \\ &\sim \text{Beta}(y_i + \alpha, n_i - y_i + \beta) \end{aligned}$$

- Having found a closed form expression for the posterior $\pi_i|y_i$, we can compute the posterior mean as

$$E(\pi_i|y_i, \alpha, \beta) = \frac{(y_i + \alpha)}{(n_i + \alpha + \beta)} \quad (4)$$

Shrinkage Effect for Beta-Binomial (cont'd)

- ▶ From Eq(4), we can clearly see the influence of the parameters of the prior (α, β) on the posterior mean $\pi_i|y_i$.
- ▶ As (α, β) increase, $E[\pi|y_i]$ decreases; this is sometimes referred to as a **shrinkage effect**.
- ▶ The effect of the prior on the posterior mean gets weaker as the sample size increases. Intuitively, this reflects the Bayesian perspective of relying more on prior information when few data points are available; and relying on the data when the sample size is large.
- ▶ We provide an illustration of the shape of the posterior $\pi|y_i$ for different settings of the prior $\pi|\alpha, \beta$. We use a fixed sample size of $N = 20$.

Shrinkage Effect: plot of posterior for different prior settings of the Beta-Binomial model

