LogLinear Models and Poisson Regression, Stat 701, Spring 2025

We discussed in class on Monday, May 5, the formulation of multinomial-data parametric models in a log-linear parametric regression framework. The goal of this development is to relate the maximization of multinomial likelihood to the Poisson-regression (*quasi-likelihood*) estimating equation, also discussed in that class.

The data are based on categorical outcomes $k = 1, \ldots, K$ for independent individuals indexed $i = 1, \ldots, n$, and consist of predictors $X_k \in \mathbb{R}^d$ (the same for all individuals, varying with outcome-category $k = 1, \ldots, K$, with first component $X_{k,1} = 1$ to enable an intercept-term in the model) together with the outcome $Z_i \in \{1, \ldots, K\}$. The form of the parametric model $p_k(\beta) = P(Z_i = k)$ with $\beta \in \mathbb{R}^d$, is: $p_k(\beta) = \exp(X_k' \beta)$, where the probability-vector property $\sum_{k=1}^K p_k(\beta) = 1$ is enforced in terms of the intercept parameter β_1, through

$$1 \;=\; \sum_{k=1}^{K} p_k(\beta) \;=\; e^{\beta_1} \sum_{k=1}^{K} \exp\,(\sum_{r=2}^{d} \beta_r\, X_{k,r}) \tag{1}$$

In this framework, we have a design matrix $\mathbf{X}$ that is $K \times d$, with rows X_k and entries $X_{k,r}$, and multinomial cell probabilities $p_k = \exp((\mathbf{X}\beta)_k)$. We assume that $\mathbf{X}$ has full rank. After substitution for e^{β_1} using (1), the same model is expressed in terms of a $K \times (d-1)$ design matrix $\mathbf{X}^* = (X_{k,r})_{1 \le k \le K,\, 2 \le r \le d}$ and $d-1$ dimensional parameter $\beta^* = (\beta_2, \ldots, \beta_d)$, as

$$\tilde{p}_k(\beta^*) \;\equiv\; p_k(\beta) \;=\; e^{\beta_1}\, \exp((\mathbf{X}^*\beta^*)_k) \;=\; \frac{\exp(\beta^{*\prime}\, X_k^*)}{\sum_{m=1}^{K} \exp(\beta^{*\prime}\, X_m^*)} \tag{2}$$

where the $d-1$-dimensional rows of $\mathbf{X}^*$ regarded as column-vectors are denoted X_k^*.

We spent much of the May 5 class discussing a couple of examples of this framework.

Example 1. (*Two-way table, predictors a function of tabular indices*) In this and the next Example, the multinomial index k with $K = J \cdot L$ is understood to correspond 1-to-1 with the 2-way table index $k \leftrightarrow (j, l)$ for $j = 1, \ldots, J,\ l = 1, \ldots, L$. An example of a parametric model with $d = 4$, with 3 free coefficients, after accounting for the probability-vector property through equation (1), is a model like that of Homework 6 problem #4, with

$$X_k \;=\; (1,\, j,\, l,\, j \cdot l) \quad , \qquad k \;\leftrightarrow\; (j, l)$$

Example 2. (*Two-way table, general row-column independence*) Example 1 exhibits row-by-column independence $p_{(j,l)} = g(j)\cdot h(l)$ if $\beta_4 = 0$, but as mentioned in an earlier class, in that setting there are only two parameters governing the probabilities $p_k = p_{(j,l)}$, respectively through row-sums $p_{(j,+)}$ and column-sums $p_{(+,l)}$ being governed by proportionality relations

$$p_{(j,+)} \;\equiv\; \sum_{l=1}^{L} p_{(j,l)} \;\propto\; e^{\beta_2\, j} \quad , \qquad p_{(+,l)} \;\equiv\; \sum_{j=1}^{J} p_{(j,l)} \;\propto\; e^{\beta_3\, l}$$

The general row-by-column independence can also be given a loglinear regression form of dimension $d = J + L - 1$ in which the $(J \cdot L) \times (J + L - 1)$ design matrix $\mathbf{X}$ has entries for $k \leftrightarrow (j, l)$:

$$X_{k,r} \;=\; \begin{cases} 1 & \text{if} \quad r = 1 \\ I_{[r-1=j]} & \text{if} \quad r = 2, \ldots, J \\ I_{[r-J=l]} & \text{if} \quad r = J+1, \ldots, L-1 \end{cases} \tag{3}$$

This design matrix achieves the result that for $k \leftrightarrow (j, l)$, $1 \le j \le J$, $1 \le l \le L$,

$$p_k(\beta) \;=\; p_{(j,l)} \;=\; \exp\Big(\sum_{r=1}^{J+L-1} \beta_r\, X_{k,r} \Big) \;=\; e^{\beta_1} \cdot e^{\beta_{j+1}\, I_{[j<J]}} \cdot e^{\beta_{J+l}\, I_{[l<L]}}$$

which is evidently a product of a function of j by a function of l (which is the defining feature of row-by-column independence), and

$$p_{(j,+)} \;=\; \sum_{l=1}^{L} p_{(j,l)} \;\propto\; \exp(\beta_{j+1}\, I_{[j<J]}) \quad , \qquad p_{(+,l)} \;=\; \sum_{j=1}^{J} p_{(j,l)} \;\propto\; \exp(\beta_{J+l}\, I_{[l<L]})$$

Remark. Allowing predictors or covariates that differ by individual in these models is more difficult when $K > 2$, and we do not discuss that here. Such models are multi-outcome logistic regression models. The loglinear case $K = 2$ with general individual-level predictors is the logistic-regression model that we have already discussed.

Relation between Poisson-regression and Multinomial MLEs

Now we are considering a general loglinear multinomial model, with *iid* Z_i's and $P(Z_i = k) \;=\; \tilde{p}_k(\beta^*) \;=\; p_k(\beta) \;=\; \exp(X_k'\beta)$ for $k = 1, \ldots, K$. Letting $N_k \;=\; \sum_{i=1}^{n} I_{[Z_i=k]}$, it is these counts $\underline{N} = (N_1, \ldots, N_K)$ that are jointly distributed Multinom$(n,\, (p_1, \ldots, p_K))$, and that [in a general multinomial model, whether or not parameterized by a lower-dimensional parameter β or β^*], are sufficient statistics for $(p_1, \ldots, p_K)$. In the multinomial model parameterized here by β^* in equation (2), the count-statistics $\underline{N}$ are sufficient for β^*, and the log-likelihood and likelihood equation determining MLEs for β^* are

$$\log L_1(\beta^*) = \sum_{k=1}^{K} N_k \log \, \tilde{p}_k(\beta^*) \quad \Rightarrow \quad \nabla_{\beta^*} \log L_1(\beta^*) = \sum_{k=1}^{K} X_k^* \, \{N_k \, - \, n \, \tilde{p}_k(\beta^*)\} = \underline{0} \tag{4}$$

The last step is to establish the connection with the MLEs in a Poisson Regression model. Suppose we treat the counts N_k for $k = 1, \ldots, K$ as independent $\text{Poisson}(\exp(X_k'\gamma)$ variables, where $\gamma \in \mathbb{R}^d$ is a parameter just like β. The log-likelihood $\log L_2$ and likelihood equation, obtained by setting $\nabla_\gamma \log L_2$ to $\underline{0}$, are respectively

$$\log L_2(\gamma) \;=\; \sum_{k=1} \{ -e^{X_k'\gamma} \,+\, N_k \, X_k\} \quad \Rightarrow \quad \sum_{k=1}^{K} X_k \, (N_k \,-\, e^{X_k'\hat{\gamma}}) \;=\; \underline{0} \tag{5}$$

The first entry of the likelihood equation in (5) says that $n \;=\; \sum_{k=1}^{K} N_k \;=\; \sum_{k=1}^{K} \exp(X_k' \, \hat{\gamma})$, which shows that $\hat{\gamma}_1$ is uniquely determined from $(\hat{\gamma}_2, \ldots, \hat{\gamma}_d)$ from the formula

$$\exp(\hat{\gamma}_1) \;=\; n / \sum_{k=1}^{K} \exp\,(\sum_{r=2}^{d} X_{k,r} \, \hat{\gamma}_r) \tag{6}$$

which is just like formula (1). Substituting for $\exp(\hat{\gamma}_1)$ in the likelihood equation in (5), omitting the first entry of that d-dimensional vector equation, and denoting $\gamma^* = (\gamma_2, \ldots, \gamma_d)$, yields

$$\sum_{k=1}^{K} X_k^* \left\{ N_k \,-\, \frac{n \, \exp(\hat{\gamma}^{*\prime} \, X_k^*)}{\sum_{m=1}^{K} \exp(\hat{\gamma}^{*\prime} \, X_m^*)} \;=\; \underline{0} \right\} \tag{7}$$

This proves that $\hat{\gamma}^*$ in (7) satisfies exactly the same likelihood equation as the MLE of β^* in (4). Therefore the MLEs are the same, and we can rely on Poisson regression software to fit loglinear models.