Analysis of Categorical Data
Chapter 9: Loglinear Model

Shaobo Jin

Department of Mathematics

Intended Learning Outcome

Through this chapter, you should be able to

1. use loglinear models to analyze contingency tables,
2. understand different types of independence,
3. understand the connection between Poisson sampling and multinomial sampling,
4. understand the connection between loglinear model and logistic model,
5. obtain closed form expression of MLE.

Loglinear Model

Suppose that $\begin{pmatrix} N_1 & N_2 & \cdots & N_c \end{pmatrix}$ is a multinomial sample of n subjects. If π_i is the cell probability of cell i, the expected frequency is $\mu_i = n\pi_i$.

Suppose that $\begin{pmatrix} N_1 & N_2 & \cdots & N_c \end{pmatrix}$ is a Poisson sample. The count in each cell i follows a Poisson distribution with expected frequency μ_i.

The loglinear model builds a linear model on μ_i, and is applicable to both multinomial sampling and Poisson sampling. That is, the loglinear model is of the form

$$\log \mu_i = \lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i.$$

Poisson and Multinomial Sampling

Suppose that there are c independent Poisson random variables N_i, $i = 1, 2, ..., c$, with mean μ_i. Their conditional distribution given $\sum_{i=1}^{c} N_i = n$ satisfies

$$
\begin{aligned}
P\left(N_1 = n_1, ..., N_c = n_c \mid \sum_{i=1}^{c} N_i = n\right) &= \frac{P\left(N_1 = n_1, ..., N_c = n_c\right)}{P\left(\sum_{i=1}^{c} N_i = n\right)} \\
&= \frac{n!}{\prod_{i=1}^{c} n_i!} \prod_{i=1}^{c} \left(\frac{\mu_i}{\sum_{j=1}^{c} \mu_j}\right)^{n_i},
\end{aligned}
$$

which is a multinomial distribution with sample size n and outcome probabilities

$$
\left\{\frac{\mu_i}{\sum_{j=1}^{c} \mu_j}\right\}.
$$

Poisson Loglinear Model

Suppose that we have c independent Poisson random variables N_i, each with mean $\mu_i = \exp\left\{\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i\right\}$. The log-likelihood is

$$\ell = \sum_i \left[n_i \left(\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i\right) - \exp\left(\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i\right) - \log\left(n_i!\right)\right].$$

The first-order partial derivatives are

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= \sum_i n_i - \exp\left(\lambda\right) \sum_i \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right), \\ \frac{\partial \ell}{\partial \beta_k} &= \sum_i n_i x_{ik} - \exp\left(\lambda\right) \sum_i \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right) x_{ik}. \end{aligned}$$

Evaluated at the MLEs, we should have

$$\sum_i n_i x_{ik} = \frac{\sum_i n_i}{\sum_i \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right)} \sum_i \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right) x_{ik}.$$

Multinomial Loglinear Model

If we conditional on the sum $\sum_i n_i$, then $(N_1, ..., N_c)$ follows a multinomial distribution with cell probabilities

$$\frac{\mu_i}{\sum_j \mu_j} = \frac{\exp\left(\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i\right)}{\sum_j \exp\left(\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_j\right)} = \frac{\exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right)}{\sum_j \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_j\right)}.$$

The multinomial log-likelihood is

$$\ell = \log\left[\frac{(\sum_i n_i)!}{\prod_i n_i!}\right] + \sum_i n_i \boldsymbol{\beta}^T \boldsymbol{x}_i - \left(\sum_i n_i\right) \log\left[\sum_j \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_j\right)\right].$$

The first-order partial derivative of multinomial log-likelihood is

$$\frac{\partial \ell}{\partial \beta_k} = \sum_i n_i x_{ik} - \left(\sum_i n_i\right) \frac{\sum_i \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right) x_{ik}}{\sum_i \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right)}.$$

Hence, the Poisson loglinear model and the multinomial loglinear model should yield the same $\boldsymbol{\beta}$ estimators.

Multinomial loglinear model

We can rewrite the Poisson log-likelihood as

$$
\begin{aligned}
\ell &= \sum_i \left[n_i \left(\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i\right) - \exp\left\{\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i\right\} - \log\left(n_i!\right)\right] \\
&= \underbrace{\left\{\sum_i n_i \boldsymbol{\beta}^T \boldsymbol{x}_i - n \log\left[\sum_j \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_j\right)\right]\right\}}_{N_1,\ldots,N_c \mid \sum_i N_i = n} + \underbrace{\{n \log \mu - \mu\}}_{\sum_i N_i \sim \text{Poisson}(\mu)} + C,
\end{aligned}
$$

where $\mu = \sum_i \mu_i$. This means that we can reparametrize the Poisson model with parameters $(\lambda, \boldsymbol{\beta})$ to $(\mu, \boldsymbol{\beta})$.

- The multinomial loglinear model only takes the conditional multinomial distribution part.
- The Poisson loglinear model has one more parameter λ (or μ) than the multinomial logliner model, because of the random sample size.

Independent Model for Two-Way Tables

Consider an $I \times J$ contingency table with multinomial sampling of n subjects. If X and Y are independent, then

$$\pi_{ij} = \pi_{i+}\pi_{+j}.$$

The expected frequency of cell (i, j) is

$$\mu_{ij} = n\pi_{i+}\pi_{+j},$$

which is equivalent to

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y.$$

This is the loglinear model of independence.

Independence under Poisson sampling means independence under multinomial sampling when we conditional on n.

Identification

Consider the model

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y.$$

We can either let $\lambda_I^X = \lambda_J^Y = 0$ or $\sum_i \lambda_i^X = \sum_j \lambda_j^Y = 0$ for identification.

1. If $\lambda_I^X = \lambda_J^Y = 0$,

$$\lambda_i^X = \log \pi_{i+} - \log \pi_{I+}.$$

2. If $\sum_i \lambda_i^X = \sum_j \lambda_j^Y = 0$,

$$\lambda_i^X = \log \pi_{i+} - \frac{1}{I}\sum_{i=1}^{I} \log \pi_{i+}.$$

Saturated Model for Two-Way Table

If X and Y are dependent, the loglinear model becomes

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY},$$

with interaction λ_{ij}^{XY}.

- For identification, we need the constraints $\lambda_I^X = \lambda_J^Y = 0$ and $\lambda_{Ij}^{XY} = \lambda_{iJ}^{XY} = 0$ for all i and j.
- The total number of parameters in this model is

$$\underset{\lambda}{1} + \underset{\lambda_i^X}{I-1} + \underset{\lambda_j^Y}{J-1} + \underset{\lambda_{ij}^{XY}}{(I-1)(J-1)} = IJ.$$

- The number of cells is IJ. Hence, it is the saturated model.

Association in Two-Way Table

The saturated model and the independence model for a two-way table only differ in λ_{ij}^{XY}, which measures deviation from independence. In particular, the log local odds ratio satisfies

$$\begin{aligned}
\log\left(\frac{\mu_{ij}\mu_{i+1,j+1}}{\mu_{i,j+1}\mu_{i+1,j}}\right) &= \left(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}\right) \\
&\quad + \left(\lambda + \lambda_{i+1}^X + \lambda_{j+1}^Y + \lambda_{i+1,j+1}^{XY}\right) \\
&\quad - \left(\lambda + \lambda_i^X + \lambda_{j+1}^Y + \lambda_{i,j+1}^{XY}\right) \\
&\quad - \left(\lambda + \lambda_{i+1}^X + \lambda_j^Y + \lambda_{i+1,j}^{XY}\right) \\
&= \lambda_{ij}^{XY} + \lambda_{i+1,j+1}^{XY} - \lambda_{i,j+1}^{XY} - \lambda_{i+1,j}^{XY}.
\end{aligned}$$

Loglinear Model Under Different Sampling

The connection between the Poisson distribution and the multinomial distribution suggests that the loglinear models

$$\begin{aligned}\log \mu_{ij} &= \lambda + \lambda_i^X + \lambda_j^Y, \\ \log \mu_{ij} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY},\end{aligned}$$

can be used for both multinomial sampling and Poisson sampling.

- You can easily derive that $\hat{\mu}_{ij} = n_{i+}n_{+j}/n$ in an independent model with both multinomial sampling and Poisson sampling.
- The saturated model implies $\hat{\mu}_{ij} = n_{ij}$.

Model Fit and Model Comparison

When we fit log-linear models, we often fit a Poisson log-linear model.

- Pearson χ^2 and deviance test whether a model holds by comparing cell fitted values to observed counts.
- If the model fits the data well, they can be approximated by a chi-square distribution with df being the number of cells minus the number of model parameters.

We can also use difference in deviance to compare different models, e.g., compare the model with interaction and the model without interaction. If models are not nested, we can use information criterion.

- Even though we have two models under multinomial sampling, we can still fit Poisson log-linear models and perform model selection via LRT or AIC.

Types of Independence

Consider a three-way $I \times J \times K$ table of variables X, Y, and Z. We can define different types of independence.

1. Mutual independence: the variables are mutually independent, denoted by (X, Y, Z) (generating class).
2. Joint independence: X and Z are jointly independent of Y, denoted by (XZ, Y).
3. Conditional independence: X is independent of Y given Z, denoted by (XZ, YZ) or $X \perp Y \mid Z$.
4. Marginal independence: X and Y are independent when ignoring Z, denoted by (X, Y).
5. Homogeneous association: the conditional odds ratio satisfies $\theta_{ij(1)} = \theta_{ij(2)} = \cdots = \theta_{ij(K)}$.

Marginality/Hierarchical Model

When you include higher-order terms into the model, always include all lower-order terms by the hierarchy principle. A loglinear model that follows this principle is a hierarchical loglinear model.

$$\begin{aligned}
\log \mu_{ij} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.
\end{aligned}$$

Marginal Independence

X is marginally independent of Y, if

$$\pi_{ij+} = \pi_{i++}\pi_{+j+}, \text{ for all } i \text{ and } j.$$

This is denoted by (X, Y) or $X \perp Y$.

Conditional Independence

Categorical variables X and Y are conditionally independent given Z, if independence holds for each partial table within which Z is fixed;

$$\pi_{ij|k} = \pi_{i+|k}\pi_{+j|k}, \text{ for all } i, j, \text{ and } k,$$

where $\pi_{ij|k} = P(X = i, Y = j \mid Z = k)$. This is denoted by (XZ, YZ) or $X \perp Y \mid Z$.

If $X \perp Y \mid Z$, then

$$\begin{aligned}\pi_{ijk} &= P(X = i \mid Y = j, Z = k) P(Y = j, Z = k) \\ &= P(X = i \mid Z = k) P(Y = j, Z = k) \\ &= \pi_{i+k}\pi_{+jk}/\pi_{++k}.\end{aligned}$$

The corresponding loglinear model is

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

Joint Independence

Y is jointly independent of X and Z, if

$$\pi_{ijk} = \pi_{i+k}\pi_{+j+}, \text{ for all } i, j, \text{ and } k.$$

This is denoted by (XZ, Y) or $(X, Z) \perp Y$. The corresponding loglinear model is

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}.$$

- (XZ, Y) means that (X, Y) and (Z, Y).
- Conditional independence is weaker than joint independence, since conditional independence with $\lambda_{jk}^{YZ} = 0$ leads to joint independence.

Mutual independence

X, Y, and Z are mutually independent, if

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}, \text{ for all } i, j, \text{ and } k.$$

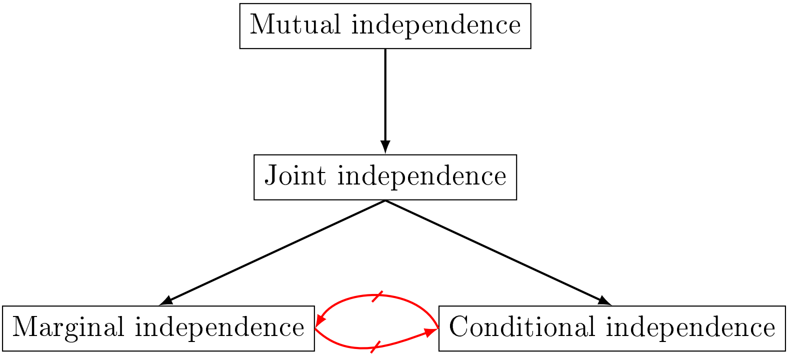
This is denoted by (X, Y, Z). The corresponding loglinear model is

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

- Joint independence is weaker than mutual independence, since joint independence with $\lambda_{jk}^{YZ} = 0$ leads to mutual independence.
- From probability theory perspective,

$$\pi_{ijk} = \underbrace{\pi_{i++}\pi_{++k}}_{\pi_{i+k}}\pi_{+j+}.$$

Relationships in Graph



More General Models

Consider the models

$$
\begin{aligned}
(X, Y, Z):\ \log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z \\
(XZ, Y):\ \log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} \\
(XZ, YZ):\ \log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.
\end{aligned}
$$

They have three, two, and one pair of conditionally independent variables, respectively.

The saturated model is

$$
(XYZ):\ \log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.
$$

Interpretation By Conditional Association

At a fixed level $Z = k$, the local odds ratios are

$$\theta_{ij(k)} = \frac{\pi_{ijk}\pi_{i+1,j+1,k}}{\pi_{i+1,j,k}\pi_{i,j+1,k}}, \quad 1 \leq i \leq I-1,\ 1 \leq j \leq J-1.$$

Suppose that $X \perp Y \mid Z$, where $\pi_{ijk} = \pi_{i+k}\pi_{+jk}/\pi_{++k}$. Then,

$$\theta_{ij(k)} = \frac{\left(\pi_{i+k}\pi_{+jk}/\pi_{++k}\right)\left(\pi_{i+1,+,k}\pi_{+,j+1,k}/\pi_{++k}\right)}{\left(\pi_{i+1,+,k}\pi_{+jk}/\pi_{++k}\right)\left(\pi_{i+k}\pi_{+,j+1,k}/\pi_{++k}\right)} = 1.$$

Hence, $X \perp Y \mid Z$ is equivalent to $\theta_{ij(k)} = 1$ for all possible i, j, and k.

Interpretation By Conditional Association

Consider the model

$$(XY, XZ, YZ): \log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

At a fixed level $Z = k$, the local odds ratios satisfies

$$\begin{aligned} \log \theta_{ij(k)} &= \log \left(\frac{\mu_{ijk}\mu_{i+1,j+1,k}}{\mu_{i+1,j,k}\mu_{i,j+1,k}} \right) \\ &= \lambda_{ij}^{XY} + \lambda_{i+1,j+1}^{XY} - \lambda_{i+1,j}^{XY} - \lambda_{i,j+1}^{XY}, \end{aligned}$$

which implies homogeneous XY association

$$\theta_{ij(1)} = \theta_{ij(2)} = \cdots = \theta_{ij(K)}, \text{ for all } i \text{ and } j.$$

By symmetry, we also have homogeneous XZ association and homogeneous YZ association.

Inference of Conditional Association

The local odds ratios satisfies

$$\log\theta_{ij(k)} = \lambda_{ij}^{XY} + \lambda_{i+1,j+1}^{XY} - \lambda_{i+1,j}^{XY} - \lambda_{i,j+1}^{XY}.$$

When a Poisson log-linear model

$$\log\mu = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}$$

is fitted, we know that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx N\left(\mathbf{0},\ \left(\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X}\right)^{-1}\right).$$

Hence, for any constant vector $\boldsymbol{a}$,

$$\boldsymbol{a}^T \hat{\boldsymbol{\beta}} - \boldsymbol{a}^T \boldsymbol{\beta} \approx N\left(\mathbf{0},\ \boldsymbol{a}^T \left(\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{a}\right).$$

Log-Likelihoood Under Poisson Sampling

Consider the log-linear model with Poisson sampling

$$\log \mu_i \quad = \quad \lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i, \quad i = 1, ..., N.$$

The log-likelihood is

$$\ell \quad = \quad \sum_{i=1}^{N} \left[n_i \left(\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i \right) - \exp \left(\lambda + \boldsymbol{\beta}^T \boldsymbol{x}_i \right) - \log \left(n_i! \right) \right].$$

- The sufficient statistic of β_j is $\sum_i n_i x_{ij}$.
- The first-order partial derivatives are

$$\frac{\partial \ell}{\partial \beta_j} \quad = \quad \sum_{i=1}^{N} \left(n_i - \mu_i \right) x_{ij}.$$

Hence, the MLE must satisfy $\sum_{i=1}^{n} n_i x_{ij} = \sum_{i=1}^{n} \mu_i x_{ij}$.

Sufficient Statistic and Marginal Sum

For a contingency table, N is the number of independence cells, $\boldsymbol{\beta}$ is the vector of effects $\left\{\lambda_i^X, \lambda_j^Y, \cdots\right\}$, and x_{ij} is either 0 or 1.

- The sufficient statistics $\{\sum_i n_i x_{ij}\}$ become some marginal sums.
- $\sum_{i=1}^N n_i x_{ij} = \sum_{i=1}^N \mu_i x_{ij}$ means that the sufficient statistics are the same as the the expected value μ in the marginal tables.

Since minimal sufficient statistic can be expressed as a function of any other sufficient statistics, we can obtain the explicit expression of μ using the minimal sufficient statistics.

Three-Way Table Under Poisson Sampling

Under Poisson sampling, the log-likelihood of a three-way table satisfies

$$\ell = \text{constant} + \sum_{i,j,k} \left[n_{ijk} \log\left(\mu_{ijk}\right) - \mu_{ijk} \right].$$

For the saturated model,

$$\begin{aligned} \ell = \; & \text{constant} + n\lambda + \sum_{i} n_{i++}\lambda_i^X + \sum_j n_{+j+}\lambda_j^Y + \sum_k n_{++k}\lambda_k^Z \\ & + \sum_{i,j} n_{ij+}\lambda_{ij}^{XY} + \sum_{i,k} n_{i+k}\lambda_{ik}^{XZ} + \sum_{j,k} n_{+jk}\lambda_{jk}^{YZ} + \sum_{i,j,k} n_{ijk}\lambda_{ijk}^{XYZ} \\ & - \sum_{i,j,k} \exp\left(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}\right). \end{aligned}$$

The sufficient statistics are $\{n_{i++}\}$, $\{n_{+j+}\}$, $\{n_{++k}\}$, $\{n_{ij+}\}$, $\{n_{i+k}\}$, $\{n_{+jk}\}$, $\{n_{ijk}\}$. Minimal sufficient statistics are $\{n_{ijk}\}$.

Explicit MLE of μ

The minimal sufficient statistics of a given loglinear model are the MLEs for the corresponding marginal totals of μ.

- For example, consider the model (XZ, YZ). Its log-likelihood is

$$\begin{aligned}\ell &= \text{constant} + n\lambda + \sum_i n_{i++}\lambda_i^X + \sum_j n_{+j+}\lambda_j^Y + \sum_k n_{++k}\lambda_k^Z \\ &\quad +0 + \sum_{i,k} n_{i+k}\lambda_{ik}^{XZ} + \sum_{j,k} n_{+jk}\lambda_{jk}^{YZ} + 0 \\ &\quad - \sum_{i,j,k} \exp\left(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + 0 + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + 0\right).\end{aligned}$$

- The minimal sufficient statistics are $\{n_{i+k}\}$, $\{n_{+jk}\}$.
- Hence, we should have $\hat{\mu}_{i+k} = n_{i+k}$, $\hat{\mu}_{+jk} = n_{+jk}$.

MLE of μ: Another Example

Consider the model (X, Y, Z). Its log-likelihood is

$$\begin{aligned}\ell &= \text{constant} + n\lambda + \sum_i n_{i++}\lambda_i^X + \sum_j n_{+j+}\lambda_j^Y + \sum_k n_{++k}\lambda_k^Z \\ &\quad - \sum_{i,j,k} \exp\left(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z\right).\end{aligned}$$

The minimal sufficient statistics are $\{n_{i++}\}$, $\{n_{+j+}\}$, $\{n_{++k}\}$. Hence, we must have $\hat{\mu}_{i++} = n_{i++}$, $\hat{\mu}_{+j+} = n_{+j+}$, and $\hat{\mu}_{++k} = n_{++k}$.

Functions of Minimal Sufficient Statistic

MLE has an invariance property in the sense that a function of MLE is the MLE of the function, i.e., the MLE of $g(\theta)$ is $g\left(\hat{\theta}\right)$, where $\hat{\theta}$ is the MLE of θ. Hence, if we can express a μ as a closed form of expected values of minimal sufficient statistics, then we obtain the closed form expression of such μ.

- Consider the model (XZ, YZ). The MLEs are $\hat{\mu}_{i+k} = n_{i+k}$, $\hat{\mu}_{+jk} = n_{+jk}$. Since
$$\mu_{ijk} = n\pi_{ijk} = \frac{n\pi_{i+k}\pi_{+jk}}{\pi_{++k}} = \frac{\mu_{i+k}\mu_{+jk}}{\mu_{++k}},$$
we should have $\hat{\mu}_{ijk} = n_{i+k}n_{+jk}/n_{++k}$.
- Consider the model (X, Y, Z). The MLEs are $\hat{\mu}_{i++} = n_{i++}$, $\hat{\mu}_{+j+} = n_{+j+}$, and $\hat{\mu}_{++k} = n_{++k}$. Since
$$\mu_{ijk} = n\pi_{i++}\pi_{+j+}\pi_{++k} = \frac{\mu_{i++}\mu_{+j+}\mu_{++k}}{n^2},$$
we should have $\hat{\mu}_{ijk} = n_{i++}n_{+j+}n_{++k}/n^2$.

Decomposable Models

Different from other models in a three-way table, the model (XY, XZ, YZ)

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

has no closed form probabilistic expression in terms of π_{ijk} and its marginals. Hence, we don't have the closed form expression for μ_{ijk}, although we still have closed form expression for μ_{ij+}, μ_{i+k}, and μ_{+jk}.

A loglinear model having no closed form expression for the probabilistic model is nondecomposable. Otherwise, it has a closed form factorization of $\{\pi_{ijk}\}$ and it is called decomposable.

Different Sampling Designs

We said that different sampling designs are connected, e.g., the Poisson loglinear model and the multinomial loglinear model should yield the same $\boldsymbol{\beta}$ estimators.

But they should still be treated differently in model selection.

- If a marginal total is fixed by the sampling design, then the corresponding λ-term must be present in the loglinear model regardless of its lack of statistical significance.
- The reason is that we must keep the MLE of μ to be the same as the observed marginal total.
- For example, if n_{ij+} is fixed by sampling design, then we must pick models that satisfy $\hat{\mu}_{ij+} = n_{ij+}$.

An Example of Fixed Marginal Total

		Mortality		
Seedlings	Depth	Dead	Alive	Total
Longleaf	High	41	59	100
	Low	11	89	100
Slash	High	12	88	100
	Low	5	95	100

We treat Mortality as our response variable. By design, the row totals are fixed to 100. Our loglinear model should ensure $\hat{\mu}_{i+k} = 100$. Hence, λ_{ik}^{DS} must be included in our model.

Loglinear Model and Logistic Model

Consider an $I \times 2 \times K$ table. Suppose that we have fitted a (XY, XZ, YZ) loglinear model. Then,

$$\begin{aligned}
& \log \frac{P(Y=1 \mid X=i, Z=k)}{P(Y=2 \mid X=i, Z=k)} \\
= \; & \log \frac{\pi_{i1k}}{\pi_{i2k}} = \log \frac{\mu_{i1k}}{\mu_{i2k}} \\
= \; & \left(\lambda + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ}\right) \\
& - \left(\lambda + \lambda_i^X + \lambda_2^Y + \lambda_k^Z + \lambda_{i2}^{XY} + \lambda_{ik}^{XZ} + \lambda_{2k}^{YZ}\right) \\
= \; & \underbrace{\left(\lambda_1^Y - \lambda_2^Y\right)}_{\alpha} + \underbrace{\left(\lambda_{i1}^{XY} - \lambda_{i2}^{XY}\right)}_{\beta_i^X} + \underbrace{\left(\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ}\right)}_{\beta_k^Z},
\end{aligned}$$

which is an additive logistic model.

In the log-linear model, we model the association λ_{ik}^{XZ}. But in the logistic model, we do not model λ_{ik}^{XZ}, as they cancel out.

Loglinear Model and Logistic Model

Many other loglinear models for an $I \times 2 \times K$ table is equivalent to a logistic model.

- For (Y, XZ),

$$\log \frac{P\left(Y=1 \mid X=i, Z=k\right)}{P\left(Y=2 \mid X=i, Z=k\right)} = \underbrace{\left(\lambda_1^Y - \lambda_2^Y\right)}_{\alpha}.$$

- For (XY, XZ),

$$\log \frac{P\left(Y=1 \mid X=i, Z=k\right)}{P\left(Y=2 \mid X=i, Z=k\right)} = \underbrace{\left(\lambda_1^Y - \lambda_2^Y\right)}_{\alpha} + \underbrace{\left(\lambda_{i1}^{XY} - \lambda_{i2}^{XY}\right)}_{\beta_i^X}.$$

- For (XZ, YZ),

$$\log \frac{P\left(Y=1 \mid X=i, Z=k\right)}{P\left(Y=2 \mid X=i, Z=k\right)} = \underbrace{\left(\lambda_1^Y - \lambda_2^Y\right)}_{\alpha} + \underbrace{\left(\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ}\right)}_{\beta_k^Z}.$$

Higher Dimensions

We have introduced the loglinear models for three-way tables $(I \times J \times K)$. It can be generalized to contingency tables of higher dimensions.

$$
\begin{aligned}
\log \mu_{hijk} &= \lambda + \lambda_h^W + \lambda_i^X + \lambda_j^Y + \lambda_k^Z, \\
\log \mu_{hijk} &= \lambda + \lambda_h^W + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{hi}^{WX} + \lambda_{hj}^{WY} + \lambda_{hk}^{WZ}, \\
\log \mu_{hijk} &= \lambda + \lambda_h^W + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{hi}^{WX} + \lambda_{hj}^{WY} + \lambda_{hk}^{WZ} + \lambda_{ik}^{XZ} \\
&\quad + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}, \\
\log \mu_{hijk} &= \lambda + \lambda_h^W + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{hi}^{WX} + \lambda_{hj}^{WY} + \lambda_{hk}^{WZ} + \lambda_{ik}^{XZ} \\
&\quad + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{hij}^{WXY}.
\end{aligned}
$$