Multinomial and Contingency Tables

Multinomial logit model

Introduce data in nes96.pdf

Multinomial logit model

Suppose we have an independent sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ where

$$\begin{aligned} \mathbf{Y}_i &\sim \text{multinomial}(m_i, \mathbf{p}_i) \quad \text{with } \mathbf{p}_i = (p_{i1}, \ldots, p_{iJ}) \\ \mathbb{P}(\mathbf{Y}_i = \mathbf{y}_i) &= \frac{m_i!}{y_{i1}! \cdots y_{iJ}!} p_{i1}^{y_{i1}} \cdots p_{iJ}^{y_{iJ}} \quad \text{for } \mathbf{y}_i \geq 0, \textstyle\sum_j y_{ij} = m_i \end{aligned}$$

A multinomial logit model supposes that

$$p_{ij} = \frac{e^{\eta_{ij}}}{\sum_{k=1}^{J} e^{\eta_{ik}}}$$

where $\eta_{ij} = \mathbf{x}_i^T \boldsymbol{\beta}_j$ for predictor variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and parameter vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J$.

Multinomial logit model

We can interpret $e^{\eta_{ij}}$ as the rate at which an outcome of type j occurs.

Note that for fixed i we can add a constant to each η_{ij} without changing the p_{ij}, so the model is ill-determined. We fix this by setting $\boldsymbol{\beta}_1 = \mathbf{0}$. This is equivalent to dividing top and bottom by $e^{\eta_{i1}}$, or alternatively replacing $\boldsymbol{\beta}_j$ by $\boldsymbol{\beta}_j - \boldsymbol{\beta}_1$.

If $J = 2$ then

$$p_{i1} = \frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}_2}}, \quad p_{i2} = \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}_2}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}_2}}$$

which is just a binomial regression model with responses y_{i2}, parameters $\boldsymbol{\beta}_2$, and a logit link.

Multinomial logit model

We can fit a multinomial logit model by maximum likelihood. The resulting estimator is then consistent and asymptotically efficient, and we can compare nested models by comparing the difference of their deviances to a χ^2 distribution,

Example: See nes96.pdf

For additional reading, see "Faraway on multinomials" posted on LMS.

Contingency tables: two-way tables

Introduce data in wafer.pdf

Contingency tables: two-way tables

Example: semiconductor production (Faraway Ch. 4)

	die contaminated		
wafer quality	no	yes	
good	320	14	334
bad	80	36	116
	400	50	450

Example: Framlingham heart disease study

	heart disease		
cholesterol	yes	no	
low	51	992	1043
high	41	245	286
	92	1237	1329

In each case we ask if the two factors are dependent or not?

Contingency tables: two-way tables

Let y_{ij} be the number of observations with factor 1 at level i and factor 2 at level j:

	factor 2			
factor 1	1	2	3	
1	y_{11}	y_{12}	y_{13}	$y_{1\cdot}$
2	y_{21}	y_{22}	y_{23}	$y_{2\cdot}$
	$y_{\cdot 1}$	$y_{\cdot 2}$	$y_{\cdot 3}$	$y_{\cdot\cdot}$

Let π_{ij} be the probability that an observation has factor 1 at level i and factor 2 at level j. Factor 1 has n_1 levels and factor 2 has n_2 levels.

$$\begin{aligned}
\pi_{i\cdot} &= \sum_j \pi_{ij} \text{ is prob. an obs. has factor 1 at level } i \\
\pi_{\cdot j} &= \sum_i \pi_{ij} \text{ is prob. an obs. has factor 2 at level } j \\
\pi_{\cdot\cdot} &= \sum_i \pi_{i\cdot} = \sum_j \pi_{\cdot j} = 1
\end{aligned}$$

We want to know if $\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$ for all i and j?

Contingency tables

There are four different models for this type of data, depending on how it was collected:

- Multinomial
- Poisson
- Product multinomial
- Hypergeometric

Multinomial model

Suppose that $y_{\cdot\cdot}$ is fixed and observations are independent, then

$$\begin{aligned}
(y_{11}, y_{12}, \ldots, y_{1n_2}, y_{21}, \ldots, y_{n_1 n_2}) &\sim \text{multinomial}\left(y_{\cdot\cdot}, (\pi_{11}, \pi_{12}, \ldots, \pi_{n_1 n_2})\right) \\
\mathbb{P}(Y_{ij} = y_{ij} \text{ for all } i \text{ and } j) &= \frac{y_{\cdot\cdot}!}{\prod_{ij} y_{ij}!} \prod_{ij} \pi_{ij}^{y_{ij}}
\end{aligned}$$

Poisson

Rather than fix $y_{\cdot\cdot}$ we suppose that observations of type (i, j) occur at some rate λ_{ij}, independently of the other types, and we collect observations for a unit period of time:

$$y_{ij} \sim \text{Poisson}(\lambda_{ij}),$$

for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$, and y_{ij} independent.

Relationship between Poisson and Multinomial models

Lemma: if $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, \ldots, k$, X_i independent, then $\sum_i X_i \sim \text{Poisson}(\sum_i \lambda_i)$ and

$$(X_1, \ldots, X_k | \textstyle\sum_i X_i = n) \sim \text{multinomial}(n, (\pi_1, \ldots, \pi_k))$$

where $\pi_i = \lambda_i / \sum_j \lambda_j$.

In our case we have $\pi_{ij} = \lambda_{ij}/\lambda_{\cdot\cdot}$, and the condition $\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$ becomes

$$\lambda_{ij} = \frac{\lambda_{i\cdot}\lambda_{\cdot j}}{\lambda_{\cdot\cdot}}.$$

Testing independence: multinomial model

- H_0 $\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$ (independent factors)
- H_1 π_{ij} unrestricted.

As our test statistic we use the log likelihood ratio for the model under H_0 compared to the model under H_1.
This is just the deviance, since the model under H_1 is the saturated model.

Testing independence: multinomial model

Likelihood under H_0: using the multinomial model we have

$$\begin{aligned}\log \mathcal{L}(\pi_{i\cdot}, \pi_{\cdot j}, \text{ for all } i,j) &= \sum_{i,j} \log(\pi_{i\cdot}\pi_{\cdot j}) y_{ij} + c \\ &= \sum_i y_{i\cdot} \log \pi_{i\cdot} + \sum_j y_{\cdot j} \log \pi_{\cdot j} + c\end{aligned}$$

We need to maximise this subject to the constraints $\sum_i \pi_{i\cdot} = 1$, $\sum_j \pi_{\cdot j} = 1$, $\pi_{i\cdot} \geq 0$, $\pi_{\cdot j} \geq 0$.

We can maximise each term separately using Lagrange multipliers, to get

$$\hat{\pi}_{i\cdot} = \frac{y_{i\cdot}}{y_{\cdot\cdot}}, \quad \hat{\pi}_{\cdot j} = \frac{y_{\cdot j}}{y_{\cdot\cdot}}.$$

Testing independence: multinomial model

Likelihood under H_1: using the multinomial model we have

$$\log \mathcal{L}(\pi_{ij},\ \text{for all } i,j) = \sum_{i,j} \log(\pi_{ij}) y_{ij} + c$$

We need to maximise this subject to the constraints $\sum_{i,j} \pi_{ij} = 1$, $\pi_{ij} \geq 0$.

Using Lagrange multipliers we get

$$\hat{\pi}_{ij} = \frac{y_{ij}}{y_{..}}$$

Testing independence: multinomial model

(scaled) Deviance or LR test statistic:

$$
\begin{aligned}
D &= 2\log\frac{\mathcal{L}(H_1)}{\mathcal{L}(H_0)} \\
&= 2\left(\sum_{ij} y_{ij}\log\frac{y_{ij}}{y_{\cdot\cdot}} - \sum_{ij} y_{ij}\log\frac{y_{i\cdot}y_{\cdot j}}{y_{\cdot\cdot}y_{\cdot\cdot}}\right) \\
&= 2\sum_{ij} y_{ij}\log\frac{y_{ij}}{\hat{\mu}_{ij}} \\
&\approx \chi^2_{(n_1-1)(n_2-1)}
\end{aligned}
$$

where $\hat{\mu}_{ij} = \hat{\pi}_{ij}y_{\cdot\cdot}$ and factor 1 has n_1 levels and factor 2 has n_2 levels. The number of parameters under H_1 is $n_1 n_2 - 1$ and number under H_0 is $n_1 - 1 + n_2 - 1$.

The distribution holds under the null, provided the observed counts are large enough (at least 5 as a rule of thumb). Under the alternative we expect the deviance to be larger than under the null.

Testing independence: Poisson model

- H_0 $\lambda_{ij} = \lambda_{i\cdot}\lambda_{\cdot j}/\lambda_{\cdot\cdot}$ (independent factors)
- H_1 λ_{ij} unrestricted.

As our test statistic we use the log likelihood ratio for the model under H_0 compared to the model under H_1.
This is just the deviance, since the model under H_1 is the saturated model.

Testing independence: Poisson model

Likelihood under H_0: using the Poisson model we have

$$\begin{aligned}\log \mathcal{L} &= \sum_{ij} (y_{ij} \log \lambda_{ij} - \lambda_{ij}) + c \\ &= \sum_{ij} \left(y_{ij}(\log \lambda_{i\cdot} + \log \lambda_{\cdot j} - \log \lambda_{\cdot\cdot}) - \frac{\lambda_{i\cdot}\lambda_{\cdot j}}{\lambda_{\cdot\cdot}} \right) + c\end{aligned}$$

Maximising this we get

$$\hat{\lambda}_{i\cdot} = y_{i\cdot}, \quad \hat{\lambda}_{\cdot j} = y_{\cdot j}, \quad \hat{\lambda}_{\cdot\cdot} = y_{\cdot\cdot}, \quad \hat{\lambda}_{ij} = \frac{y_{i\cdot} y_{\cdot j}}{y_{\cdot\cdot}}.$$

Likelihood under H_1: the MLE for the full model is just

$$\hat{\lambda}_{ij} = y_{ij}.$$

Testing independence: Poisson model

(scaled) Deviance or LR test statistic:

$$\begin{aligned} D &= 2\sum_{ij}\left(y_{ij}\log\frac{y_{ij}y_{\cdot\cdot}}{y_{i\cdot}y_{\cdot j}} - \left(y_{ij} - \frac{y_{i\cdot}y_{\cdot j}}{y_{\cdot\cdot}}\right)\right) \\ &= 2\sum_{ij} y_{ij}\log\frac{y_{ij}}{\hat{\lambda}_{ij}} \\ &\approx \chi^2_{(n_1-1)(n_2-1)} \text{ (provided our counts are large enough)} \end{aligned}$$

Exactly the same as for the multinomial model!
That is, to test independence of the two factors, we can use the same test statistic with the same distribution under the null, for both the multinomial and the Poisson cases.

Pearson's χ^2 statistic

For the Poisson model, note that under H_0 we have

$$\begin{aligned}\log \lambda_{ij} &= -\log \lambda_{..} + \log \lambda_{i.} + \log \lambda_{.j} \\ &= \gamma + \alpha_i + \beta_j \text{ say.}\end{aligned}$$

That is, estimating $\lambda_{..}$, $\lambda_{i.}$ and $\lambda_{.j}$ under H_0 is just like fitting a Poisson two-way classification model (no interaction) with log-link.

Since $\phi = 1$ for a Poisson regression model, we have that in this case Pearson's χ^2 statistic has a $\chi^2_{n-p} = \chi^2_{(n_1-1)(n_2-1)}$ distribution, and we can alse use this as a goodness-of-fit statistic, and hence use it to test H_0 against H_1.

Pearson's χ^2 statistic

$$
\begin{aligned}
X^2 &= \sum_{ij} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{v(\hat{\mu}_{ij})} \\
&= \sum_{ij} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \\
&= \sum_{ij} \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \\
&\approx \chi^2_{(n_1-1)(n_2-1)}
\end{aligned}
$$

This is a very popular test statistic for a contingency table, though the deviance D generally gives a better test.

As for the deviance we need the counts to be not too small (at least 5 as a rule of thumb), for the χ^2 approximation to hold.

Example

Example: See wafer.pdf.

Contingency tables: three-way classification tables

Three-way classification tables

Example: 4991 Wisconsin male high school seniors were classified according to socio-economic status (low, lower middle, upper middle, and high), the degree of parental encouragement they receive (low and high) and whether or not they have plans to attend college (no, yes). Fienberg (1977, p. 101)

```
encouraged <- gl(2, 1, 16, labels=c("low", "high"))
soc_stratum <- gl(4, 2, 16,
  labels=c("lower", "lower middle", "upper middle", "higher"))
plans <- gl(2, 8, 16, labels=c("no", "yes"))
counts <- c(749, 233, 627, 330, 420, 374, 153, 266,
             35, 133,  38, 303,  37, 467,  26, 800)
wisconsin <- data.frame(counts, encouraged, soc_stratum, plans)
wt <- xtabs(counts ~ soc_stratum + encouraged + plans, wisconsin)
ftable(wt)
```

Three-way classification tables

```
                          plans  no yes
soc_stratum  encouraged
lower        low                749  35
             high               233 133
lower middle low                627  38
             high               330 303
upper middle low                420  37
             high               374 467
higher       low                153  26
             high               266 800
```

What dependencies are there between the three factors?

Three-way classification tables

Suppose that factor f has n_f levels, for $f = A, B, C$, and put

$$\begin{aligned}
\pi_{ijk} &= \mathbb{P}(\text{Observe factor } A \text{ at lvl } i,\ B \text{ at lvl } j \text{ and } C \text{ at lvl } k) \\
y_{ijk} &= \text{Observed count of factor } A \text{ at lvl } i,\ B \text{ at lvl } j \text{ and } C \text{ at lvl } k
\end{aligned}$$

As before we use dots $\cdot$ to indicate summation over a particular index.

Let $y_{\cdots} = n$. We suppose that y_{ijk} is the observed value of a Poisson r.v. Y_{ijk} with rate/mean $\lambda_{ijk} = \mu_{ijk} = n\pi_{ijk}$.

We consider various possible dependence structures between the factors.

$A + B + C$: mutually independent factors

Our null hypothesis is $H_0 : \pi_{ijk} = \pi_{i..}\pi_{.j.}\pi_{..k}$. We express this as a Poisson regression model with log link.

$$
\begin{aligned}
\mu_{ijk} &= n\pi_{i..}\pi_{.j.}\pi_{..k} \\
\log \mu_{ijk} &= \log n + \log \pi_{i..} + \log \pi_{.j.} + \log \pi_{..k} \\
&= \delta + \alpha_i + \beta_j + \gamma_k
\end{aligned}
$$

We accept the null hypothesis if an additive model is adequate to explain the observed responses.

We can test the null using the deviance, which is $\chi^2_{n_A n_B n_C - (1 + n_A - 1 + n_B - 1 + n_C - 1)}$, provided the cell counts are reasonable (at least 5)

$A * B + C$: C independent of A and B

$H_0 : \pi_{ijk} = \pi_{ij\cdot}\pi_{\cdot\cdot k}$

$$\log \mu_{ijk} = \log n + \log \pi_{ij\cdot} + \log \pi_{\cdot\cdot k} = \delta + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k$$

Here $\alpha_i + \beta_j + (\alpha\beta)_{ij} = \log \pi_{ij\cdot}$. We use main effects plus an interaction so that we can easily test for the significance of the interaction. In particular, if the interaction is insignificant, then we can remove it to get an additive model $A + B + C$.

That is, we can test for dependence between factor A and factor B by comparing the models $A * B + C$ and $A + B + C$. We test using the difference between the deviances, which will be $\chi^2_{(n_A-1)(n_B-1)}$ if there is no interaction.

Using the difference between two deviances is generally more reliable than testing for model adequacy using a single deviance.

$A * B + A * C$: conditional independence of B and C given A

$H_0 : \mathbb{P}(B = j, C = k|A = i) = \mathbb{P}(B = j|A = i)\mathbb{P}(C = k|A = i)$ that is $\pi_{ijk} = \pi_{ij\cdot}\pi_{i\cdot k}/\pi_{i\cdot\cdot}$.

$$\begin{aligned} \log \mu_{ijk} &= \log n + \log \pi_{ij\cdot} + \log \pi_{i\cdot k} - \log \pi_{i\cdot\cdot} \\ &= \delta + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} \end{aligned}$$

As before, we can test for adequacy of the model using the deviance, or we can compare the model to smaller or larger models (with or without interaction terms) using the difference of the deviances.

$A * B + A * C + B * C$: pairwise dependence

In this case there is no easy formulation of the null hypothesis in terms of the probabilities π_{ijk}. The Poisson regression model has all three pairwise interactions, but not a three-way interaction.

$$\log \mu_{ijk} = \delta + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

The model with three-way interaction is saturated, that is, it gives a perfect fit, so we can not test for this directly. However, if the model with pairwise dependence is inadequate then our only real alternative is the saturated model.

Pearson's χ^2 test

As an alternative to testing model adequacy using the deviance we can use Pearson's χ^2. This again has the form of the sum of (observed $-$ expected)^2/expected over all cells.

For the expected values we use the MLE estimates:

$A + B + C$: $\hat{\mu}_{ijk} = n \frac{y_{i\cdot\cdot}}{n} \frac{y_{\cdot j \cdot}}{n} \frac{y_{\cdot\cdot k}}{n}$

$A * B + C$: $\hat{\mu}_{ijk} = n \frac{y_{ij\cdot}}{n} \frac{y_{\cdot\cdot k}}{n}$

$A * B + A * C$: $\hat{\mu}_{ijk} = \frac{y_{ij\cdot} y_{i\cdot k}}{y_{i\cdot\cdot}}$

We don't have a nice likelihood for the model $A * B + A * C + B * C$, so no nice MLE's either.

Example: a 20-year follow-up study on the effects of smoking, taken from Appleton, French and Vanderpump (1996)
threeway.pdf