STAT 200B 2019 Week07

Soyeon Ahn

1 Multinomial

Suppose $Z \in \{1, ..., k\}$ and let $p_j = P(Z = j)$. The parameter $p = (p_1, \dots, p_k)$ is really only $k - 1$ dimensional, since $\sum_{j=1}^{k} p_j = 1$. Suppose we observe an iid sample $Z_1, \dots, Z_n$. Let $X_j = \#\{Z_i : Z_i = j\}$. Then we say $X = (X_1, \dots, X_k)$ has $Multinomial(n, p)$ distribution.

The PDF is

$$f(x_1, \dots, x_k; p) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

Note that the labels $1, \dots, k$ for the Z's are arbitrary, and that $Binomial(n, p)$ distribution is just a special case.

The MLE is $(\hat{p}_1, \dots, \hat{p}_k) = (X_1/n, \dots, X_k/n)$.

Consider testing $H_0 : (p_1, \dots, p_k) = (p_{01}, \dots, p_{ok})$ versus the alternative that they are not equal. The LRT rejects when

$$T(X) = \frac{\mathcal{L}_n(\hat{p})}{\mathcal{L}(p_0)} = \prod_{j=1}^{k} \left(\frac{\hat{p}_j}{p_{0j}} \right)^{X_j}$$

is large. Since I don't know how to calculate the exact probability of this, I'll use the limiting χ^2. That is,

$$\lambda(X) = 2 \log T(X) = 2 \sum_{j=1}^{k} X_j \log \left(\frac{\hat{p}_j}{p_{0j}} \right) \xrightarrow{D} \chi^2_{k-1}$$

The degrees of freedom is $k-1$ because the dimension of Θ is $k-1$ and the dimension of Θ_0 is zero (a point). The approximate size α LRT rejects H_0 when $\lambda(X) \geq \chi^2_{k-1,\alpha}$.

proof sketch Taylor series: $f(x) = log(1 + x) \approx x - \frac{x^2}{2}$ and $O_j = X_j = n\hat{p}_j$ (observed) and $E_j = np_{oj}$ (expected).

$$
\begin{aligned}
2\log T(X) &= 2\sum_{j=1}^{k} X_j \log\left(\frac{\hat{p}_j}{p_{0j}}\right) \\
&= 2\sum O_i \log\left(\frac{O_i}{E_i}\right) \\
&= 2\sum (E_i + O_i - E_i)\log\left(1 + \frac{O_i - E_i}{E_i}\right) \\
&= 2\sum (E_i + O_i - E_i)\left(\frac{O_i - E_i}{E_i} - \frac{(O_i - E_i)^2}{2(O_i - E_i)^2} + O((O_i - E_i)^3)\right) \\
&= 2\sum \left(O_i - E_i + \frac{(O_i - E_i)^2}{E_i} - \frac{(O_i - E_i)^2}{2E_i} + O((O_i - E_i)^3)\right)
\end{aligned}
$$

We use that $\sum O_i - E_i = 0$ since $\sum E_i = \sum O_i$.

$$
\begin{aligned}
&\approx \sum \frac{(O_i - E_i)^2}{E_i} \\
&= \sum \frac{(O_i - E_i)^2}{E_i}.
\end{aligned}
$$

2 Testing independence in contingency tables

A *contingency table* is a table in which observations or individuals are classified according to one or more criteria.

Consider a two-way contingency table with I rows and J columns. For $i = 1, \cdots, I$ And $j = 1, \cdots, J$, let p_{ij} be the probability that an individual selected from the population under consideration is classified in row i and column j. (i.e. in the (i, j) cell of the table).

Let $p_{i.} =$ (cell probability in row i) and $p_{.j} =$ (cell probability in column j). Then we must have $p_{..} = \sum_i \sum_j p_{ij} = 1$.

Suppose a random sample of n subjects is taken, and let n_{ij} be the number of these classified in the (i, j) cell of the table.

Let $n_{i.} = \sum_j n_{ij}$ and $n_{.j} = \sum_i n_{ij}$. So $n_{++} = n$.

We have

$$
(N_{11}, \cdots, N_{1I}, N_{21}, \cdots, N_{IJ}) \sim Multinomial(n; p_{11}, \cdots, p_{1J}, p_{21}, \cdots, p_{IJ}).
$$

We may be interested in testing the null hypothesis that the two classifications are independent. So we test

- H_0: $p_i = p_{i.}p_{.j}$ for all i, j, i.e. independence of columns and rows.
- H_1: p_{ij} are unrestricted.

We have the restrictions like $p_{..} = 1, p_{ij} \geq 0$.

Under H_1, the mles are $\hat{p_{ij}} = \frac{n_{ij}}{n}$.

Under H_0, the mles are $\hat{p}_{i.} = \frac{n_{i.}}{n}$ and $\hat{p}_{.j} = \frac{n_{.j}}{n}$.

Write $O_{ij} = n_{ij}$ and $E_{ij} = n\hat{p}_{i.}\hat{p}_{.j} = n_{i.}n_{.j}/n$.

Then

$$2 \log \Lambda = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} O_{ij} \log \left(\frac{O_{ij}}{E_{ij}} \right) \approx \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

the test statistics have a limiting χ^2_ν distribution, where $\nu = (I-1)(J-1)$.

We have $|\Theta_1| = IJ - 1$, because under H_1 the p_{ij}'s sum to one. Also, $|\Theta_0| = (I-1) + (J-1)$ because $p_{1.}, \cdots, p_{I.}$ must satisfy $\sum_i p_{i.} = 1$ and $p_{.1}, \cdots, p_{.I}$ must satisfy $\sum_j p_{.j} = 1$. So

$$|\Theta_1| - |\Theta_0| = IJ - 1 - (I-1) - (J-1) = (I-1)(J-1).$$

3 P-value

More precisely, we proceed as follows: We specify the family of tests by a family of rejection regions $\mathcal{R}$. Assume that $\mathcal{R}$ is linearly ordered by inclusion.

$$R, R^{'} \in \mathcal{R} \text{ implies that } R \subset R^{'} \text{ or } R^{'} \subset R$$

Suppose that for every α, we have a size α test with rejection region R_α. Then,

$$\text{p-value } = \inf\{\alpha : T(X) \in R_\alpha\} = \inf\{\sup_{H_0} P(R) : R \in \mathcal{R}\}$$

That is, the p-value is the smallest level at which we can reject H_0.

Take a test with rejection region R_α having the significance level $\sup_{H_0} P(R)$.

When $p < \alpha$, we reject the null hypothesis at significance level α. There is $R^{'} \in \mathcal{R}$ and $\sup_{H_0} P(R^{'}) < \alpha$. Since $\sup_{H_0} P(R) = \alpha$, we conclude $R^{'} \subset R$ since $R^{'}$ is linearly ordered. the null hypothesis is rejected for the test with rejection region R. Suppose the rejection regions are given by $R = \{T > c\}$ for test statistic T. Assume we have observed the data ω. Then, the p-value is $\inf_{T(\omega)>c} \sup_{H_0} P(\{T > c\}) = \sup_{H_0} P(\{T \geq T(w)\})$

4 Global testing

Suppose we test null hypotheses.

$$\mathcal{H} = \{H_{0,1}, \ldots, H_{0,N}\}$$

We wish to test the intersection null.

$$H_0 : \cap H_{0,i}$$

We wish to test the hypotheses that all the nulls are valid.

We cannot simply test each hypotheses at level α because, if N is large, we are sure to make lots of type I errors just by chance. We need to do some sort of multiplicity adjustment.

The simplest method for testing the global null is Bonferroni's global test.

Bonferroni method Given p-values p_i, reject null hypothesis $H_{0,i}$ if

$$\min p_i \leq \frac{\alpha}{N}$$

Given a prescribed level α, Bonferroni's test tests each $H_{0,i}$ at level $\frac{\alpha}{N}$ and rejects H_0 if $H_{0,i}$'s are rejected. Equivalently, Bonferroni's global test rejects when $\min p_i \leq \frac{\alpha}{N}$.

$$\begin{aligned} P(\min p_i \leq \frac{\alpha}{N}) &= P(\cup\{p_i \leq \frac{\alpha}{N}\}) \\ &\leq \sum P(p_i \leq \frac{\alpha}{N}) \\ &= \sum \frac{\alpha}{N} \text{ since } p_i \sim Unif(0,1) \\ &\leq \alpha \end{aligned}$$

It is possible to show that if the p-values are independent, the level of the test is

$$\begin{aligned} P(\min p_i \leq \frac{\alpha}{N}) &= 1 - P(\cap\{p_i \leq \frac{\alpha}{N}\}) \\ &= 1 - \prod P(p_i \leq \frac{\alpha}{N}) \\ &= 1 - (1 - \frac{\alpha}{N})^N \\ &\to 1 - \exp^{\alpha} \end{aligned}$$

which is approximately α for any α close to zero. We observe that Bonferroni?s test only depends on the smallest p-value. Thus it is suited to situations where we expect at least one of the p-values to be very small.

5 Familywise Error Control

Rather than testing the intersection of many any hypotheses, we want to test hypotheses separately.

	not reject	reject	
$H_{0,i}$ true		V	n_0
$H_{0,i}$ false	U		$N - n_0$
	$N - R$	R	

(The notation is slightly different from the book.)

Let n_0 be the number of null hypotheses that are true and let $n - n_0$ =the number of null hypotheses that are false.

The family-wise error rate (FWER) is the probability of making any false rejections

$$P(\mathrm{V} > 0)$$

Suppose that p_j is a p-value for each null hypothesis. Let $I = \{i : H_{0,i} \text{ is ture } \} \subset \mathcal{H}$. If we reject $H_{0,i}$ for any $i \in I$, we have made an error. Let $R = \{j : \text{we reject } H_{0,j}\} \in \mathcal{H}$ be the set of hypotheses we reject. We say that we have controlled the familywise error rate at level α if

$$P(R \cap I \neq \phi) \leq \alpha$$

One way to control the familywise error rate is the Bonferroni method.

Given p-values p_i, reject null hypothesis $H_{0,i}$ if

$$p_i \leq \frac{\alpha}{N}$$

$$\begin{aligned}
P(\text{making a false rejection}) &= P(p_i \leq \frac{\alpha}{N} \text{ for some } i \in I) \\
&\leq \sum_{i \in I} P(p_i \leq \frac{\alpha}{N}) \\
&= \sum_{i \in I} \frac{\alpha}{N} \text{ since } p_i \sim Unif(0,1) \text{ for } i \in I \\
&= \frac{\alpha |I|}{N} \leq \alpha
\end{aligned}$$

The Bonferroni method is very conservative because it is trying to make it unlikely that you would make even one false rejection.

False Discovery Control

As the number of hypotheses being tested grows, controlling the FWER becomes a more and more stringent criterion. Intuitively, the chance of a false rejection grows as more hypotheses are tested. In modern applications, practitioners routinely test thousands or tens of thousands of hypotheses, and controlling the FWER is unacceptably conservative.

The false discovery proportion is the ratio of false rejections to rejections:

$$\begin{aligned}
\text{FDP} &= \frac{V}{R} \text{ if } R > 0 \\
&= 0 \text{ if } R = 0
\end{aligned}$$

False discovery rate is the mean of FDP, i.e, the mean of the number of false rejections divided by the number of rejections (the proportion of rejections that are incorrect).

The observations are

- FDP $\leq 1\{V > 0\}$. Thus the FDR is at most the FWER.
- Under the global null, all rejections are false. Thus the FDR is equal to the FWER under the global null.

Since FWER is a more stringent criterion than FDR, any procedure that controls the FWER also controls the FDR. FDR is a weaker type of control .

Benjamini and Hochberg suggested the following procedure, which guarantees $FDR \leq \alpha$:

1. For each test, compute the $p-value$. Let $P_{(1)} < \cdots < P_{(N)}$ denote the ordered p-values.
2. Select $j = \max\{i : P_{(i)} < \frac{i\alpha}{N}\}$. Let $T = P_j$.
3. Reject all null hypotheses for which the p-value $\leq P_{(j)}$. In other words, R = $\{i : P_i \leq T\}$

Consider, in general, rejecting all hypothesis for which $P_i < t$. Let $W_i = 1$ if $H_{0,i}$ is true and $W_i = 0$ otherwise. Let $\hat{G}$ be the empirical distribution of the p-values and let $G(t) = E(\hat{G})$. In this case,

$$\text{FDP} = \frac{\sum_i W_i I(P_i < t)}{\sum_i I(P_i < T)} = \frac{\frac{1}{N}\sum_i W_i I(P_i < t)}{\frac{1}{N}\sum_i I(P_i < T)}$$

Hence,

$$\begin{aligned} E(\text{ FDP }) &\approx \frac{E(\frac{1}{N}\sum_i W_i I(P_i < t))}{E(\frac{1}{N}\sum_i I(P_i < T))} \\ &= \frac{t|I|}{G(t)} \leq \frac{t}{G(t)} \approx \frac{t}{\hat{G}(t)} \end{aligned}$$

Let $t = P_{(i)}$ for some i; then $\hat{G}(t) = \frac{i}{N}$. Thus, $FDR \leq P_{(i)}\frac{N}{i}$. Setting this equal to α we get $P_{(i)} < \frac{i\alpha}{N}$,the Benjamini-Hochberg rule

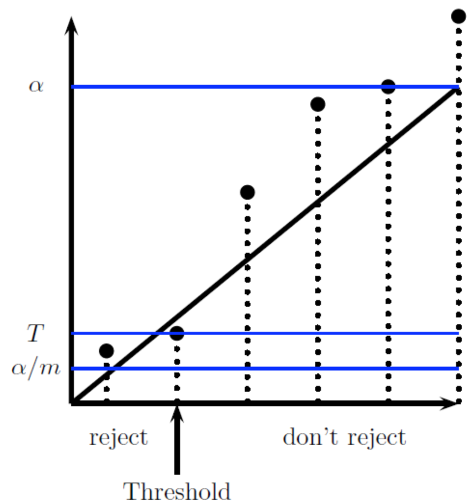

FIGURE 10.6. The Benjamini-Hochberg (BH) procedure. For uncorrected testing we reject when $P_i < \alpha$. For Bonferroni testing we reject when $P_i < \alpha/m$. The BH procedure rejects when $P_i \leq T$. The BH threshold T corresponds to the rightmost undercrossing of the upward sloping line.

Please refer to Larry Wasserman's course at CMU and Emmanuel Candes' course at Stanford University.