Power and Sample Size for Log-linear Models[1]

STA 312: Fall 2012

[1] See last slide for copyright information.

Overview

Power of a Statistical test

The power of a test is the probability of rejecting H_0 when H_0 is false.

- More power is good, because we want to make correct decisions.
- Power is not just one number. It is a *function* of the parameters.
- Usually,
  - For any n, the more incorrect H_0 is, the greater the power.
  - For any parameter value satisfying the alternative hypothesis, the larger n is, the greater the power.

Statistical power analysis

To select sample size

- Pick an *effect* you'd like to be able to detect – a set of parameter values such that H_0 is false. It should be just over the boundary of interesting and meaningful.
- Pick a desired power, a probability with which you'd like to be able to detect the effect by rejecting the null hypothesis.
- Start with a fairly small n and calculate the power. Increase the sample size until the desired power is reached.
- There may be shortcuts, but this is the idea.

- The main test statistics are

$$G^2 = 2\sum_{j=1}^{c} n_j \log\left(\frac{n_j}{\widehat{\mu}_j}\right) = 2n\sum_{j=1}^{c} p_j \log\left(\frac{p_j}{\widehat{\pi}_j}\right)$$

$$X^2 = \sum_{j=1}^{c} \frac{(n_j - \widehat{\mu}_j)^2}{\widehat{\mu}_j} = n\sum_{j=1}^{c} \frac{(p_j - \widehat{\pi}_j)^2}{\widehat{\pi}_j}$$

- When H_0 is true, their distributions are approximately (central) chi-squared.
- When H_0 is false, their distributions are approximately non-central chi-squared, with non-centrality parameter $\lambda > 0$.

Large-sample target of the restricted MLE as $n \to \infty$

$$\widehat{\boldsymbol{\pi}}_n \to \boldsymbol{\pi}(M)$$

- The notation M means *model* — the "model" given by H_0.
- The non-centrality parameter λ depends on $\boldsymbol{\pi}(M) = \{\pi_1(M), \ldots, \pi_c(M)\}$.

$$X^2 = n\sum_{j=1}^{c} \frac{(p_j - \widehat{\pi}_j)^2}{\widehat{\pi}_j} \qquad \lambda = n\sum_{j=1}^{c} \frac{[\pi_j - \pi_j(M)]^2}{\pi_j(M)}$$

$$G^2 = 2n\sum_{j=1}^{c} p_j \log\left(\frac{p_j}{\widehat{\pi}_j}\right) \quad \lambda = 2n\sum_{j=1}^{c} \pi_j \log\left(\frac{\pi_j}{\pi_j(M)}\right),$$

These formulas for λ are in Agresti's *Categorical data analysis*, p. 241.

Take a closer look

$$X^2 = n\sum_{j=1}^{c}\frac{(p_j - \widehat{\pi}_j)^2}{\widehat{\pi}_j} \qquad \lambda = n\sum_{j=1}^{c}\frac{[\pi_j - \pi_j(M)]^2}{\pi_j(M)}$$

$$G^2 = 2n\sum_{j=1}^{c} p_j \log\left(\frac{p_j}{\widehat{\pi}_j}\right) \qquad \lambda = 2n\sum_{j=1}^{c} \pi_j \log\left(\frac{\pi_j}{\pi_j(M)}\right),$$

- By the Law of Large Numbers, $p_j \to \pi_j$, always.
- When H_0 is correct, $\widehat{\pi}_j \to \pi_j$ for $j = 1, \ldots, c$.
- When H_0 is wrong, may have $\widehat{\pi}_j \to \pi_j$ for some j, but not all.
- So λ is n times a quantity saying how wrong H_0 is.
- Let's call this quantity "effect size."

Why we need $\pi_1(M), \ldots, \pi_c(M)$

- For any sample size n, power is an increasing function of λ. To calculate λ, we need not only the true parameter $\boldsymbol{\pi}$, but also where the restricted MLE goes as $n \to \infty$.

$$X^2 = n \sum_{j=1}^{c} \frac{(p_j - \widehat{\pi}_j)^2}{\widehat{\pi}_j} \qquad \lambda = n \sum_{j=1}^{c} \frac{[\pi_j - \pi_j(M)]^2}{\pi_j(M)}$$

$$G^2 = 2n \sum_{j=1}^{c} p_j \log\left(\frac{p_j}{\widehat{\pi}_j}\right) \qquad \lambda = 2n \sum_{j=1}^{c} \pi_j \log\left(\frac{\pi_j}{\pi_j(M)}\right),$$

- We will see that (given a null hypothesis model), $\boldsymbol{\pi}(M)$ is a specific function of the true parameter $\boldsymbol{\pi}$.

The Law of Large Numbers

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be independent with common expected value $\boldsymbol{\mu}$. Then

$$\overline{\mathbf{Y}}_n \rightarrow E(\mathbf{Y}_i) = \boldsymbol{\mu}$$

- For the multinomial, the data values $\mathbf{Y}_i$ are vectors of zeros and ones, with exactly one 1.
- Marginally they are Bernoulli, so $E(\mathbf{Y}_i) = \boldsymbol{\mu} = \boldsymbol{\pi}$.
- The vector of sample proportions is $\overline{\mathbf{Y}}_n = \mathbf{p}$.
- Technical details aside, $\mathbf{p}_n \rightarrow \boldsymbol{\pi}$ as $n \rightarrow \infty$ like an ordinary limit, and the usual rules apply.
- In particular, if $g(\cdot)$ is a continuous function, $g(\mathbf{p}_n) \rightarrow g(\boldsymbol{\pi})$.

Calculating the large-sample target $\boldsymbol{\pi}(M) = \{\pi_1(M), \ldots, \pi_c(M)\}$

- Write the restricted MLE as $\widehat{\boldsymbol{\pi}}_n = g(\mathbf{p}_n)$.
- Let $n \to \infty$, and
- $\widehat{\boldsymbol{\pi}}_n = g(\mathbf{p}_n) \to g(\boldsymbol{\pi}) = \boldsymbol{\pi}(M)$.

A one-dimensional example from earlier

Recall the Jobs example, with $H_0 : \pi_1 = 2\pi_2$

$$
\begin{aligned}
\widehat{\boldsymbol{\pi}} &= \left(\frac{2(n_1+n_2)}{3n}, \frac{n_1+n_2}{3n}, \frac{n_3}{n} \right) \\
&= \left(\frac{2}{3}\left(\frac{n_1}{n}+\frac{n_2}{n}\right), \frac{1}{3}\left(\frac{n_1}{n}+\frac{n_2}{n}\right), \frac{n_3}{n} \right) \\
&= \left(\frac{2}{3}\left(p_1+p_2\right), \frac{1}{3}\left(p_1+p_2\right), p_3 \right) \\
&\to \left(\frac{2}{3}\left(\pi_1+\pi_2\right), \frac{1}{3}\left(\pi_1+\pi_2\right), \pi_3 \right) = \boldsymbol{\pi}(M)
\end{aligned}
$$

Again, we need $\boldsymbol{\pi}(M)$ to calculate λ

$$X^2 = n\sum_{j=1}^{c}\frac{(p_j - \widehat{\pi}_j)^2}{\widehat{\pi}_j} \qquad \lambda = n\sum_{j=1}^{c}\frac{[\pi_j - \pi_j(M)]^2}{\pi_j(M)}$$

$$G^2 = 2n\sum_{j=1}^{c} p_j \log\left(\frac{p_j}{\widehat{\pi}_j}\right) \quad \lambda = 2n\sum_{j=1}^{c}\pi_j \log\left(\frac{\pi_j}{\pi_j(M)}\right),$$

- Divide the test statistic by n and then let $n \to \infty$ to get the effect size.
- Then multiply by a chosen value of n to get λ.
- Increase n until the power is as high as you wish.
- Of course you can (should) do this before seeing any data.

Chi-squared test of independence $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$

Passed the Course

Course	Did not pass	Passed	Total
Catch-up	π_{11}	π_{12}	π_{1+}
Mainstream	π_{21}	π_{22}	π_{2+}
Elite	π_{31}	π_{32}	π_{3+}
Total	π_{+1}	π_{+2}	1

MLEs of marginal probabilities are $\widehat{\pi}_{i+} = p_{i+}$ and $\widehat{\pi}_{+j} = p_{+j}$, so

$$\widehat{\pi}_{ij} = p_{i+}p_{+j} \rightarrow \pi_{i+}\pi_{+j} = \pi_{ij}(M)$$

Non-centrality parameters for testing independence

$$X^2 = n\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(p_{ij} - p_{i+}p_{+j})^2}{p_{i+}p_{+j}} \qquad \lambda = n\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(\pi_{ij} - \pi_{i+}\pi_{+j})^2}{\pi_{i+}\pi_{+j}}$$

$$G^2 = 2n\sum_{i=1}^{I}\sum_{j=1}^{J}p_{ij}\log\left(\frac{p_{ij}}{p_{i+}p_{+j}}\right) \qquad \lambda = 2n\sum_{i=1}^{I}\sum_{j=1}^{J}\pi_{ij}\log\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right),$$

With degrees of freedom $df = (I-1)(J-1)$

A cheap way to calculate the non-centrality parameter for any alternative hypothesis

Just for testing independence, so far

$$X^2 = n \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(p_{ij} - p_{i+}p_{+j})^2}{p_{i+}p_{+j}} \qquad \lambda = n \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(\pi_{ij} - \pi_{i+}\pi_{+j})^2}{\pi_{i+}\pi_{+j}}$$

$$G^2 = 2n \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \log\left(\frac{p_{ij}}{p_{i+}p_{+j}}\right) \qquad \lambda = 2n \sum_{i=1}^{I} \sum_{j=1}^{J} \pi_{ij} \log\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right),$$

- Make up some data that represent the alternative hypothesis of interest.
- Sample size does not matter; $n = 100$ would make the cell frequencies percentages.
- Calculate the test statistic on your made-up data.
- Divide by the n you used.
- Now you have an effect size.
- Multiply your effect size by any n, to get a λ.
- Or you can follow the formulas on the right-hand side, but it's the same thing.

Example

- As part of their rehabilitation, equal numbers of convicted criminals are randomly assigned to one of two treatment programmes just prior to their release on parole. How many will be re-arrested within 12 months?
- Suppose the programs differ somewhat in their effectiveness, but not much.
- Say 60% in Programme A will be re-arrested, compared to 70% in Programme B.
- What total sample size is required so that this difference will be detected by a Pearson chi-squared test of independence, with power at least 0.80?
- (Note this test is identical to one implied by the more appropriate product-multinomial model.)

Sample size required for Power of 0.80

Conditional probability of re-arrest 0.60 for Programme A, and 0.70 for B

```
> # R calculation for the recidivism example
> crit = qchisq(0.95,df=1); crit
[1] 3.841459
> dummy = rbind(c(40,60),
+               c(30,70))
> X2 = loglin(dummy,margin=list(1,2))$pearson
2 iterations: deviation 0
> effectsize = X2/200; effectsize
[1] 0.01098901
> wantpow = 0.80; power = 0 ; n = 50; crit = qchisq(0.95,1)
> while(power<wantpow)
+     {
+     n = n+2 # Keeping equal sample sizes
+     lambda = n*effectsize
+     power = 1-pchisq(crit,1,lambda)
+     } # End while power < wantpow
> n; power
[1] 716
[1] 0.8009609
```

What if probability is 0.50 for Programme A?

Compared to 0.70 for B

```
> # What if Programme A reduced re-arrests to 50%?
> dummy = rbind(c(50,50),
+               c(30,70))
> X2 = loglin(dummy,margin=list(1,2))$pearson
2 iterations: deviation 0
> effectsize = X2/200; effectsize
[1] 0.04166667
> wantpow = 0.80; power = 0 ; n = 50; crit = qchisq(0.95,1)
> while(power<wantpow)
+     {
+     n = n+2 # Keeping equal sample sizes
+     lambda = n*effectsize
+     power = 1-pchisq(crit,1,lambda)
+     } # End while power < wantpow
> n; power
[1] 190
[1] 0.8033634
```

Can there be too much power?

What if the sample size is **HUGE**?

```
> # Power of a trivial effect for n = 100,000
> dummy = rbind(c(50,50),
+               c(49,51))
> X2 = loglin(dummy,margin=list(1,2))$pearson
2 iterations: deviation 0
> effectsize = X2/200; effectsize
[1] 0.00010001
> lambda = 100000 * effectsize
> power = 1-pchisq(crit,1,lambda); power
[1] 0.8854098
```

General log-linear models are tougher

Until you think about it

- Explicit formulas for the MLEs under the model are not available in general.
- So can't just re-write them in terms of $p_{ij\ldots}$ and let $n \to \infty$ to get $\pi_{ij\ldots}(M)$.
- However,

We know how the iterative proportional fitting algorithm produces $\widehat{\boldsymbol{\pi}}$

- It's based on certain observed marginal totals.
- These are based on $n_{ij\ldots} = n\, p_{ij\ldots}$,
- So $\widehat{\boldsymbol{\pi}}_n$ is a function of $\mathbf{p}_n$: $\widehat{\boldsymbol{\pi}}_n = g(\mathbf{p}_n)$
- What kind of function?
- It's a sequence of multiplications, all continuous.
- A continuous function of a continuous function is continuous,
- So the entire composite function, though complicated, is continuous, and $\widehat{\boldsymbol{\pi}}_n = g(\mathbf{p}_n) \to g(\boldsymbol{\pi}) = \boldsymbol{\pi}(M)$

$\widehat{\boldsymbol{\pi}}_n = g(\mathbf{p}_n) \rightarrow g(\boldsymbol{\pi}) = \boldsymbol{\pi}(M)$

- The large-sample target of $\widehat{\boldsymbol{\pi}}_n$ is $g(\boldsymbol{\pi})$.
- And we know what the function $g(\cdot)$ is, or anyway we know how to compute it.
- It's the iterative proportional fitting algorithm, applied to sets of marginal probabilities instead of totals.
- Or you could apply it to quantities like $n\,\pi_{i+k}$, and then divide by n at the end.

Note $\widehat{\pi}_j = g_j(\mathbf{p})$ and $\pi_j(M) = g_j(\boldsymbol{\pi})$

It's the same function: Iterative proportional fitting

$$X^2 = n\sum_{j=1}^{c}\frac{(p_j - \widehat{\pi}_j)^2}{\widehat{\pi}_j} \qquad \lambda = n\sum_{j=1}^{c}\frac{[\pi_j - \pi_j(M)]^2}{\pi_j(M)}$$

$$G^2 = 2n\sum_{j=1}^{c} p_j \log\left(\frac{p_j}{\widehat{\pi}_j}\right) \qquad \lambda = 2n\sum_{j=1}^{c}\pi_j \log\left(\frac{\pi_j}{\pi_j(M)}\right)$$

So we can do what we did before with tests of independence.

- Make up a data table that represents the alternative hypothesis of interest.
- Cell frequencies are $n\,\pi_j$, for some convenient n.
- Calculate the test statistic on your made-up data.
- Divide by the n you used; now you have an effect size.
- Multiply your effect size by any n, to get a λ.
- Use that λ to calculate a power value.

Copyright Information