Topic 19

Extensions on the Likelihood Ratio

Chi-square Tests

Outline

Overview

Chromosomal Recombination

Overview

For the two-sample two-sided sample proportion test, we used the central limit theorem to assert that our test statistic is approximately the square of a standard normal random variable, and hence is a χ^2 random variable with 1 degree of freedom.

These ideas can be extended to the case in which Θ is a d-dimensional parameter space and k of these parameters are, under the null hypothesis, assume to have fixed values. Thus, Θ_0 is $d-k$-dimensional.

As before

- $\hat{\theta}_0$ be the parameter value that maximizes the likelihood for $\theta \in \Theta_0$ and
- $\hat{\theta}$ be the parameter value that maximizes the likelihood for $\theta \in \Theta$.

Procedure

Theorem. Whenever the maximum likelihood estimator has an asymptotically normal distribution, let $\Lambda(\mathbf{x})$ be the likelihood ratio

$$\Lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}.$$

for an d-dimensional parameter space. Consider the hypothesis:

$$H_0 : \theta_i = c_i \text{ for all } i = 1, \ldots, k \quad \text{versus} \quad H_1 : \theta_i \neq c_i \text{ for some } i = 1, \ldots, k.$$

Then under H_0, the distribution of

$$-2 \ln \Lambda_n(X)$$

converge to a χ^2_k distribution as the sample size $n \to \infty$.

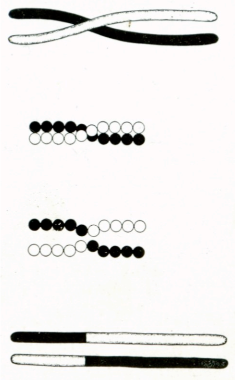
NB. More generally, if Θ_0 is defined has by k linear constraints, this can be changed to the test above by a linear change of variables. Thus,

$$\text{degrees of freedom} = \dim(\Theta) - \dim(\Theta_0).$$

Chromosomal Recombination

During meiosis, paired chromosomes experience crossing over events in the formation of gametes. During prophase I, the four available chromatids (two from each parent) are aligned allowing breaks and reattachments of homologous sites.

Recombination can occur with a small probability at anywhere along chromosome. The number of crossing over events can be modeled as Poisson random variable. The mean number of cross overs for a given chromosomal segment is called its genetic length with Morgans as the unit of measurement.



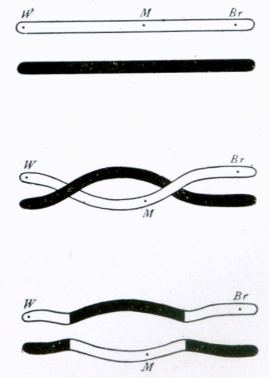
Thomas Morgan's 1916 drawing of describing a chromosomal crossing over event.

Chromosomal Recombination

With whole genome sequences for trios, an individual along with both parents, we can determine the number of crossing over events on both parents and address the question: *Are the number of crossing over events different in sperm and in eggs?* This leads to the hypothesis

$$H_0 : \lambda_m = \lambda_f \quad \text{versus} \quad H_1 : \lambda_m \neq \lambda_f$$

where λ_m and λ_f are the parameters (male and female) in the Poisson random variable that gives the number of crossing over events in the human chromosome across all 22 autosomes.



Thomas Morgan's 1916 drawing of a double crossing over event.

Chromosomal Recombination

The data are n_m and n_f the number of crossing over events for each parent's (male's and female's) chromosome. Thus, assuming that the recombination sites are independent on the two parents, the likelihood function is

$$L(\lambda_m, \lambda_f | n_m, n_f) = \frac{\lambda_m^{n_m}}{n_m!} e^{-\lambda_m} \cdot \frac{\lambda_f^{n_f}}{n_f!} e^{-\lambda_f}.$$

Exercise. Show that the maximum likelihood estimates for the likelihood function above is

$$\hat{\lambda}_m = n_m \quad \text{and} \quad \hat{\lambda}_f = n_f.$$

Thus,

$$L(\hat{\lambda}_m, \hat{\lambda}_f | n_m, n_f) = \frac{n_m^{n_m}}{n_m!} \cdot \frac{n_f^{n_f}}{n_f!} e^{-(n_m+n_f)}.$$

Chromosomal Recombination

Under the null hypothesis, λ_m and λ_f have a common value. Let's denote this by λ_0. Then the likelihood function is

$$L(\lambda_0|n_m, n_f) = \frac{\lambda_0^{n_m}}{n_m!}e^{-\lambda_0} \cdot \frac{\lambda_0^{n_f}}{n_f!}e^{-\lambda_0} = \frac{\lambda_0^{n_m+n_f}}{n_m!n_f!}e^{-2\lambda_0}.$$

Exercise. Show that the maximum likelihood estimate for the likelihood function above is

$$\hat{\lambda}_0 = \frac{n_m + n_f}{2}.$$

Thus,

$$L(\hat{\lambda}_0|n_m, n_f) = \frac{((n_m + n_f)/2)^{n_m+n_f}}{n_m!n_f!}e^{-(n_m+n_f)}.$$

Chromosomal Recombination

The likelihood ratio, after canceling the factorial and exponential factors, is

$$\Lambda(n_m, n_f) = \frac{L(\hat{\lambda}_0 | n_m, n_f)}{L(\hat{\lambda}_m, \hat{\lambda}_f | n_m + n_f)} = \frac{(n_m + n_f)^{n_m + n_f}}{2^{n_m + n_f} n_m^{n_m} n_f^{n_f}}.$$

Our data for two individuals sharing the same parents are $n_m = 56$ and $n_f = 107$. Thus,

$$-2 \ln \Lambda(n_m, n_f) = -2((n_m + n_f)(\ln(n_m + n_f) - \ln 2) - n_f \ln n_f - n_m \ln n_m) = 16.228.$$

To compute the p-value, we first determine the number of degrees of freedom.

Chromosomal Recombination

For this circumstance,

$$\text{degrees of freedom} = \dim(\Theta) - \dim(\Theta_0) = 2 - 1 = 1.$$

Thus we have the R commands,

```
> nm<-56
> nf<-107
> n<-nm+nf
> 1-pchisq(-2*(n*(log(n)-log(2))-nf*log(nf)-nm*log(nm)),1)
[1] 5.615274e-05
```

This very low p-value, 0.0056%, allow us to reject the null hypothesis.