9.6 Multinomial

$\mathcal{H}_0 : p = p(\theta) \in \omega_0$
$\mathcal{H}_A : \Omega = \{p_1, p_2, \dots, p_n = 1, p_1 + p_2 + \cdots + p_n = 1\}$
Numerator:

$$max_{p\in\omega_0}(\frac{n!}{x_1!\dots x_m!})p_1^{x_1}((\theta\ )p_2^{x_2}((\theta\ )\cdots p_m^{x_m}(\theta\ )$$

we saw that the maximum will be attained at $(p_1, p_2, \dots, p_n)_i(\hat{\theta}\ )$, and the MLE is $\hat{p}_i = \frac{x_i}{n}$.

$$\Lambda = \prod_{i=1}^{m}(\frac{p_i(\hat{\theta}\ )}{\hat{p}_i})^{x_i}$$

$$-2log\Lambda = 2\sum x_i log(\frac{\hat{p}_i}{p_i(\hat{\theta}\ )}) = 2\sum O_i log(\frac{O_i}{E_i})$$

For large samples, we have $-2log\Lambda \sim \chi^2_{m-1-k}$, where k is the dimension of ω_0.

To see this, use the Taylor expansion around x_0 of the function $f(x) = x\log\frac{x}{x_0}$:

$$\begin{aligned} f'(x) &= \log\frac{x}{x_0} + 1 \qquad f'(x_0) = 1 \\ f''(x) &= \frac{1}{x} \qquad f'(x_0) = \frac{1}{x_0} \\ f(x) &\approx (x - x_0) + \frac{1}{2}\frac{1}{x_0}(x - x_0)^2 \end{aligned}$$

Now taking $x_0 = p_i(\hat{\theta}\ )$, we get the equivalent form of the statistic:

$$-2\log\Lambda \approx n\sum_{i=1}^{m}[\hat{p}_i - p_i(\hat{\theta}\ )] + \frac{(\hat{p}_i - p_i(\hat{\theta}\ ))^2}{p_i(\hat{\theta}\ )} = \sum\frac{[x_i - np_i(\hat{\theta}\ )]^2}{np_i(\hat{\theta}\ )} = \sum\frac{(O_i - E_i)^2}{E_i}$$

This latter is known as Pearson's Chisquare statistic.

9.6.1 Examples:Normal: testing the mean with variance known

Null hypothesis $\mathcal{H}_0 : \mu = \mu_0$.
Alternative may be two-sided ($\mathcal{H}_A : \mu \neq \mu_0$) or one-sided. The same pivotal quantity is used, but the shape of the critical region is different.

Using the maximum likelihood ratio test we get

$$\Lambda = \frac{exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i - \mu_0)^2)}{exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i - \bar{X}\ )^2)} = exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^n(X_i - \mu_0)^2 - \sum_{i=1}^n(X_i - \bar{X}\ )^2)\}$$

$$-2\log\Lambda = \frac{1}{\sigma^2}(\sum_{i=1}^n(X_i - \mu_0)^2 - \sum_{i=1}^n(X_i - \bar{X}\ )^2)$$

and

$$\sum_{i=1}^n(X_i - \mu_0)^2 = \sum_{i=1}^n(X_i - \bar{X}\ )^2 + n(\bar{X}\ - \mu_0)^2$$

the LRT rejects for large values of

$$W = \frac{n(\bar{X} - \mu_0)^2}{\sigma^2}.$$

What is the null distribution of the test statistic W? χ^2_1.

Another "pivotal" quantity is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

which has $\mathcal{N}(0,1)$ distribution if $\mathcal{H}_0$ is true.

9.6.2 Examples:Normal: testing the mean with variance unknown

(And a small sample, otherwise the above applies).
$\mathcal{H}_0$ and $\mathcal{H}_A$ as above. *The pivotal* quantity is

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

with reference distribution t_{n-1} under H0.
In fact it can be shown that this t-test is the likelihood ratio test (try it as a warm-up for the exam).
Example: Effectiveness of a drug in reducing ventricular premature beats(VPB's) a clinician administers the drug to 10 patients, the following reduction after a prescribed, fixed interval is: 0, 7, -2, 14, 15, 14, 6, 16, 19, 26. with $\bar{X} = 11.5$, $s = 8.67$, we set up the test:
$\mathcal{H}_0 : \mu = 0$.
$\mathcal{H}_A : \mu > 0$
Test statistic:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{11.5 - 0}{8.67/\sqrt{10}} = 4.19$$

Take $\alpha = 0.01$, large values of T are extreme.
Under $\mathcal{H}_0$, $T \sim t_9$, and the table gives $Pvalue \approx .001$.

Counterexample:Suppose the averge heart beat for pigs is supposed to be 114 beats/min. The paper gives the heart rates of 4 pigs under anestisia:116,85,116,118. Suppose we want to test $\mu = 114$ against $\mu \neq 114$. Large values of T are extreme, $\bar{X} = 109.25$,$s = 16.19$,

$$T = \frac{109.25 - 114}{16.19/\sqrt{4}} = -0.59$$

$T \sim t_3$ under the null hypothesis, and the observed value is nothing like significant, however, if we didn't put in observation 2 for some reason and redid the test the new mean is 117.33 and the new $s = 1.155$ and the $T = 5$, which for 2 degress of freedom gives a significant p-value, even though the new mean is closer to μ, the standard deviation is so much smaller which makes the ratio larger.

This is due to the fact that these do not seem to come from a Normal distribution to begin with, and the 85 value suggests a long left hand tail.

9.6.3 Examples:Normal: testing the variance with mean unknown

$\mathcal{H}_0 : \sigma^2 = \sigma_0^2$, with $\mathcal{H}_A$ one-sided or two-sided, but in practice frequently $\sigma^2 > \sigma_0^2$.

Pivotal quantity is $(n-1)S^2/\sigma_0^2$, which has a χ^2_{n-1} distribution under $\mathcal{H}_0$.

9.7 Testing Goodness of Fit

There are many ways of testing goodness of fit. We use the simplest, the Pearson's χ^2 statistic test. It is based on the idea that the set of possible values of X can be cut into k distinct pieces; then we compare the number of observations in each piece with what is predicted by the hypothesis. For discrete distributions this is simple; for continuous ones cut points must be chosen. Basic idea: an experiment has k possible outcomes, with probabilities $p_1, p_2, \ldots, p_k$. We perform the experiment n times; let $\{O_i, i = 1, \ldots, k\}$ represent the frequencies of the possible outcomes. Then

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1}$$

where $E_i = np_i$, this is just like the multinomial above.

The condition for knowing the null distribution of the test statistic is that $E_i \geq 5$ for each i. If this fails, group some adjacent classes together: choose them in such a way as to minimise the number of classes lost, or you have to do an exact test.

Examples

Testing a common Poisson rate against a variety of rates: $\mathcal{H}_0$:Poisson(λ)
$\mathcal{H}_A$:Poisson$(\lambda_1, \lambda_2, \ldots, \lambda_n)$
In ω_0: $\hat{\lambda} = \bar{X}$,
Ω:mle of λ_i is $x_i = \tilde{\lambda}_i$.

$$\Lambda = \frac{\prod \hat{\lambda}^{x_i} e^{-\hat{\lambda}}/x_i!}{\prod \tilde{\lambda_i}^{x_i} e^{-\tilde{\lambda_i}}/x_i!} = \prod_{i=1}^{n} (\frac{\bar{x}}{x_i})^{x_i} e^{x_i - \bar{x}}$$

$$-2\log\Lambda = 2\textstyle\sum_{i=1}^{n} x_i \log(\frac{x_i}{\bar{x}}) = 2\sum O \log(\frac{O}{E})$$

Should tend to a χ^2_{n-1}.

This can be seen to be equivalent to a Pearson's χ^2 statistic:

$$\sum_i \frac{(O_i - E_i)^2}{E_i}$$

To see this, use the Taylor expansion around x_0 of the function $f(x) = x \log \frac{x}{x_0}$:

$$f'(x) = \log\frac{x}{x_0} + 1 \qquad f'(x_0) = 1$$

$$f''(x) = \frac{1}{x} \qquad f'(x_0) = \frac{1}{x_0}$$

$$f(x) = (x - x_0) + \frac{1}{2}\frac{1}{x_0}(x - x_0)^2$$

Now taking $x_0 = \bar{x}$, we get the equivalent form of the statistic:

$$\sum O_i \frac{O_i}{E_i} \approx 2\left[\sum (x_i - \bar{x}) + \frac{1}{2}(x_i - \bar{x})^2 \frac{1}{\bar{x}}\right] \approx \frac{n\hat{\sigma}^2}{\bar{x}}$$

Comparing two populations

Distributions may differ in location and/or dispersion, the samples are usually made under different conditions.

When the control distribution is already well known, it is a one sample situation.

To test $\mathcal{H}\,0{:}\mu = \mu_0$, if the sample size is big enough:

1)$\bar{X} \sim \mathcal{N}$

2)$\hat{\sigma}$ is a good approximation of σ.

When sample sizes are small, 2) is worrisome. Instead we use:

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}} = \frac{u}{v}$$

Where $u \sim \mathcal{N}(0,1)$ and $(n-1)v^2 \sim \chi^2_{n-1}$, $(n-1)s^2/\sigma^2$ is χ^2_{n-1} (see theorem 6.3.B, page 181). See page 178, where the t_n distribution is defined as that of the ratio $Z/\sqrt{U/n}$ of $Z \sim \mathcal{N}(0,1)$ variable independent of its denominator $\sqrt{U/n}$, where $U \sim \chi^2_n$. It is a symmetric distribution, with larger tails than the $\mathcal{N}$ ormal.

Comparison of two small Normal samples

If σ^2 is the same for both samples, and is known, we use:

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_x - \mu_Y, \sigma^2(\frac{1}{n} + \frac{1}{m})\right)$$

When σ^2 is the same for both samples, but unkown, we use the pooled variance:

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{m+n-2}$$

It is a theorem, that:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

follows a t_{m+n-2} distribution.

So we have a pivotal quantity even when the variances are unkown.

Of course the power is going to depend on the actual difference between the means: $\Delta = |\mu_x - \mu_y|$.

There are non central t-tables for finding the power.