STA303 Notes

Ian Zhang

1 Likelihood Function

Let θ be a parameter of some density $f(x;\theta)$ and consider a sample $x_1, \ldots, x_n$. The **likelihood function** is defined as

$$l(\theta|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

and the maximum likelihood estimator is the value $\hat{\theta}$ at which the likelihood function is maximized. Usually, we consider the log-likelihood function $L(\theta) = \log(l(\theta|x_1, \ldots, x_n))$.

- The score function is $\frac{\partial L}{\partial \theta}$
- The observed Fisher information is $-E\left(\frac{\partial^2 L}{\partial \theta^2}\right)$

2 Hypothesis Tests

2.1 Binomial Exact Test

Let Y be the number of successes in n independent Bernoulli trials with success probability π. Estimate π using the proportion of successes:

$$\hat{\pi} = \frac{Y}{n}$$

Some properties of $\hat{\pi}$ include

- $\hat{\pi}$ is an unbiased estimator of π
- $\text{Var}(\hat{\pi}) = \frac{\pi(1-\pi)}{n}$
- $\hat{\pi}$ is a consistent estimator of π (i.e.: $\hat{\pi} \xrightarrow{p} \pi$)
- $\hat{\pi} \dot{\sim} \mathcal{N}\left(\pi, \frac{\pi(1-\pi)}{n}\right)$

The binomial exact test is

$$H_0 : \pi = 0.5 \qquad H_a : \pi > 0.5 \text{ or } \pi < 0.5$$

where we consider $\frac{Y}{n}$ as the test statistic, so The p-value is

$$P(Y \leq y_{obs}|H_0) \text{ or } P(Y \geq y_{obs}|H_0)$$

depending on the test. In other words, the p-value is defined to be the probability that the we observe the current estimate or something more extreme. Usually, we test at level $\alpha = 0.05$, so if the p-value falls below this value, we reject the null. Similarly, if the p-value is greater, we fail to reject H_0.
We can also consider the two-sided test

$$H_0 : \pi = 0.5 \qquad H_a : \pi \neq 0.5$$

In this case, the p-value is $2 \min \{P(Y \geq y_{obs}|H_0), P(Y \leq y_{obs}|H_0)\}$.
In the case that $n \to \infty$, the p-value becomes more complex. Since

$$\hat{\pi} \dot{\sim} \mathcal{N}\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

we can consider the test statistic

$$Z = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim \mathcal{N}(0,1) \text{ by CLT}$$

Then the p-value is $P(Z \geq |z|)$ (for a 1-sided test, get rid of the modulus sign).

2.2 Confidence Intervals

Assuming the test above, consider the test statistic $Z \sim \mathcal{N}(0,1)$, so $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$ where $\Phi(z)$ is the CDF of $\mathcal{N}(0,1)$. Then at $\alpha = 0.05$, since the 0.975 and 0.025 quantiles of $\mathcal{N}(0,1)$ are 1.96 and -1.96 respectively, our confidence interval for π is

$$P\left(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}\right) = 0.95$$

$$P\left(\hat{\pi} - 1.96\sqrt{\frac{\pi(1-\pi)}{n}} \leq \pi \leq \hat{\pi} + 1.96\sqrt{\frac{\pi(1-\pi)}{n}}\right) = 0.95$$

Notice how the bounds of the confidence interval contain the standard error $\sqrt{\frac{\pi(1-\pi)}{n}}$. However, π is what we're estimating, so we replace π with $\hat{\pi}$ in the CI. This produces the **Wald confidence interval**

$$P\left(\hat{\pi} - 1.96\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \leq \pi \leq \hat{\pi} + 1.96\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right) = 0.95$$

Note that this performs poorly when the proportion of successes is very small.

2.3 Score Test

Consider

$$H_0 : \pi = \pi_0 \qquad H_a : \pi \neq \pi_0$$

The score function is

$$S(\pi) = \frac{\partial L(\pi)}{\partial \pi} = \frac{y - n\pi}{\pi(1-\pi)}$$

and the information is

$$I(\pi) = -E\left(\frac{\partial^2 L(\pi)}{\partial \pi^2}\right) = \frac{n}{\pi(1-\pi)}$$

Under regularity conditions, $E(S(\pi)) = 0$ and $\text{Var}(S(\pi)) = I(\pi)$. The test statistic we consider is

$$\frac{S(\pi)}{(I(\pi))^{\frac{1}{2}}}$$

which, under H_0, simplifies to $Z = \frac{\hat{\pi}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$, which is $\mathcal{N}(0,1)$ by the CLT. This is the outline of the **score test**.

2.4 Likelihood Ratio Test

The likelihood function of $Y \sim \text{Binomial}(n, \pi)$ is

$$l(\pi) = \pi^y (1-\pi)^{n-y}$$

Consider $H_0 : \pi = \pi_0$, under which we have $l(\pi_0) = \pi_0^y (1-\pi_0)^{n-y}$. From the data itself, we have $l(\hat{\pi}) = \hat{\pi}^y (1-\hat{\pi})^{n-y}$. For the likelihood ratio test, we look at the likelihood ratio

$$\frac{l(\pi_0)}{l(\hat{\pi})} = \left(\frac{\pi_0}{\pi}\right)^y \left(\frac{1-\pi_0}{1-\hat{\pi}}\right)^{n-y}$$

The likelihood ratio test is derived from taking the log of both sides and multiplying by -2. The test statistic we consider is

$$-2\log(\Lambda) = 2\left[y \log\left(\frac{y}{n\pi_0}\right) + (n-y)\log\left(\frac{n-y}{n-n\pi_0}\right)\right] = 2\sum \text{observed}\left[\log\left(\frac{\text{observed}}{\text{expected}}\right)\right]$$

As $n \to \infty$, $-2\log(\Lambda) \xrightarrow{d} \chi_1^2(\alpha)$. Thus, the likelihood ratio test is as follows:
Consider

$$H_0 : \pi = \pi_0 \qquad H_a : \pi \neq \pi_0$$

Reject the null hypothesis at level α if

$$-2\log(\Lambda) > \chi_1^2(\alpha)$$

3 Contingency Tables

A contingency table displays relationships between categorical variables, having I rows for variable X and J rows for variable Y.
For a table 2×2 contingency table, we have

	$I(Y)=1$	$I(Y)=0$	Total
$I(X)=1$	n_{11}	n_{12}	$n_{1+} = n_{11} + n_{12}$
$I(X)=0$	n_{21}	n_{22}	$n_{2+} = n_{21} + n_{22}$
Total	$n_{+1} = n_{11} + n_{21}$	$n_{+2} = n_{21} + n_{22}$	$n = n_{11} + n_{12} + n_{21} + n_{22}$

Here, n_{ij} represents the outcome of the ijth cell. Clearly,

- Each cell has probability $\pi_{ij} = \frac{n_{ij}}{n}$
- Each margin total has probability $\pi_{+i} = \frac{n_{+i}}{n}$ or $\pi_{i+} = \frac{n_{i+}}{n}$

So, we have the three types of probability:

- Marginal: π_{i+} or π_{+i}
- Joint: π_{ij}
- Conditional: $\pi_{i|j}$

In practice, consider a pair of variables where X is whether or not a subject has a disease and Y is the outcome of a subject's diagnostic test (either postive or negative). We introduce 2 definitions:

Definition. The **sensitivity** is the conditional probability that a subject's diagnostic test is positive, given the subject has the disease.

Definition. The **specificity** is the conditional probability that a subject's diagnostic test is negative, given the subject does not have the disease.

3.1 Inference on Contingency Tables

There are 2 types of studies:

- **Experimental:** A study where the researcher introduces interventions and studies the effects. Groups are usually assigned randomly
- **Observational:** Observe existing groups of individuals and assess outcomes, with no attempts at intervention. We can split into 2 subgroups:
  1. **Observational and Propespective:** See what happens to the groups of individuals
  2. **Observational and Retrospective:** Assess to analyze events of interest that have already happened

Suppose we are testing the effectiveness of a new drug. We split the patients randomly into 2 groups: treatment and placebo and observe if their conditions improve or not.

	No Improvement	Improve	Total
Treatment	n_{11}	n_{12}	$n_{1+} = n_{11} + n_{12}$
Placebo	n_{21}	n_{22}	$n_{2+} = n_{21} + n_{22}$
Total	$n_{+1} = n_{11} + n_{21}$	$n_{+2} = n_{21} + n_{22}$	$n = n_{11} + n_{12} + n_{21} + n_{22}$

Suppose we want to measure the effect of interest $\pi_1 - \pi_2$ where $\pi_1 = \frac{\pi_{11}}{\pi_{1+}}$ is the rate at which there is no improvement after treatment and $\pi_2 = \frac{\pi_{21}}{\pi_{2+}}$ is the rate at which there is no improvement after placebo.

- Note how $\pi_i = \pi_{\text{no improvement}|i}$

We can estimate π_1 and π_2 through observing proportions. The *risk* of no improvement in treatment is

$$p_1 = \hat{\pi}_1 = \frac{n_{11}}{n_{1+}}$$

while the risk of no improvement in placebo is

$$p_2 = \hat{\pi}_2 = \frac{n_{21}}{n_{1+}}$$

Some properties of p_i are $E(p_i) = \pi_i$, $\text{Var}(p_i) = \frac{\pi_i(1-\pi_i)}{n_i+}$. Assume $p_1 \perp p_2$, thus for a larger sample, we have

$$p_1 - p_2 \dot{\sim} \mathcal{N}\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_{1+}} + \frac{\pi_2(1-\pi_2)}{n_{2+}}\right)$$

by the CLT. A $(1-\alpha)\%$ confidence interval is computed as $p_1 - p_2 \pm z_{\alpha/2}\sqrt{\frac{\pi_1(1-\pi_1)}{n_{1+}} + \frac{\pi_2(1-\pi_2)}{n_{2+}}}$. Consider

$$H_0 : \pi_1 - \pi_2 = 0 \qquad H_a : \pi_1 - \pi_2 \neq 0$$

For this test, we use the test statistic

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_{1+}} + \frac{1}{n_{2+}}\right)}} \sim \mathcal{N}(0,1)$$

where $p = \frac{n_{11}+n_{21}}{n_{1+}+n_{2+}} = \frac{n_{+1}}{n}$ since under H_0, $\pi_1 = \pi_2 = \pi$, which represents a collective rate at which we observe no improvement.

3.1.1 Relative Risk

Sometimes, we're more interested in the ratio of proportions, $\frac{\pi_1}{\pi_2}$, also known as the **Risk Ratio**. Again, we can use p_1 and p_2 to form an estimator $\frac{p_1}{p_2}$. However, this results in the the ratio of a two $\mathcal{N}(0,1)$ random variables, which is Cauchy by nature.

$$\frac{Z_1}{Z_2} \sim \text{Cauchy}$$

Cauchy distribution does not have mean nor variance, thus there does not exist a CLT argument for this. On the other hand, it's still possible to compute $\log\left(\frac{p_1}{p_2}\right) = \log(p_1) - \log(p_2)$, which is normally distributed (by Delta Method). This results in a CI for $\frac{\pi_1}{\pi_2}$ being

$$\log\left(\frac{p_1}{p_2}\right) \pm z_{1-\alpha/2}\sqrt{\frac{1-p_1}{n_{1+}p_1} + \frac{1-p_2}{n_{2+}p_2}}$$

3.2 Odds Ratio

Let the probability of success be π. Then, the odds is defined as

$$\Omega = \frac{\pi}{1-\pi}$$

From a contingency table, we'll always have two categories with probabilities of success π_1 and π_2. Thus, we can define the **Odds Ratio**

$$\frac{\Omega_1}{\Omega_2} = \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_2}{1-\pi_2}} = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}$$

Since it's often the case that π_1 and π_2 are unknown, we can also estimate the odds ratio. Consider the above scenario again, although with simpler variables with placeholders:

	No Improvement	Improve	Total
Treatment	a	b	$a+b$
Placebo	c	d	$c+d$
Total	$a+c$	$b+d$	$n=a+b+c+d$

We estimate the odds ratio

$$\hat{\theta} = \frac{ad}{bc}$$

Notice that

$$\frac{\Omega_1}{\Omega_2} = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} = RR\frac{1-\pi_2}{1-\pi_1}$$

For rare outcomes, we have $\pi_1 \approx 0$ and $\pi_2 \approx 0$, in which case the odds ratio would approximately equal the relative risk.

3.3 Rates and Rates Ratio

Sometimes, it makes more sense to investigate a *rate* rather than a probability. This is because patients could leave the study, or simply the probability of an event can't be calculated realistically, which renders odds and risk as an improper effect measure.
The rate parameter is the parameter of the Poisson distribution, which characterizes the rate of occurence of the events of interest. Suppose we have a Poisson random variable with mean μ. Let T be the total follow up time and λ be the rate of our interest. Then

$$\mu = \lambda T \implies \lambda = \frac{\mu}{T}$$

This means the observed number of events y during the follow up time T has a Poisson distribution, so $Y \sim \text{Poisson}(\lambda T)$.
We use MLE estimation to estimate the rate, since oftentimes μ is unknown. Given $Y \sim \text{Poisson}(\lambda T)$, its log-likelihood function is

$$L(\lambda) = y\log(\lambda T) - \lambda T$$

which has score function $\frac{y}{\lambda} - T$, so its MLE is

$$\hat{\lambda} = \frac{y}{T}$$

The observed Fisher information

$$-\frac{\partial^2 L}{\partial \lambda^2} = \frac{y}{\lambda^2} > 0$$

since Poisson distribution is positive.

Definition (Risk)**.** Risk is the probability of an event occuring within a specific time period.

It's imperative to note that the risk and rate parameter are not the same. Unlike risk, the rate parameter *does not* correspond to a follow-up period of a fixed length.

- Instead characterizes the instantaneous occurence of the outcome event at any given time

Additionally, we can also compute the **rate ratio**: $\frac{\lambda_1}{\lambda_0}$.

3.4 Delta Method

Let $\hat{\theta}$ be the MLE of some distribution. 2 important properties to note are

- Invariance: If g is some injective function of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$
- Asymptotic normality: Let $I(\theta)$ be the observed Fisher information. Then $\hat{\theta} \xrightarrow{d} \mathcal{N}(\theta, I(\theta)^{-1})$

Theorem. Suppose $a_n(X_n - \theta) \xrightarrow{d} Z$ where a_n converges increasingly monotonically to ∞. If g is differentiable at θ with derivative $g'(\theta)$, then

$$a_n(g(X_n) - g(\theta)) \xrightarrow{d} g'(\theta)Z$$

By the Delta Method, if $\hat{\theta}$ is the MLE, then for any differentiable function g, we have

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] \xrightarrow{d} \mathcal{N}(0, [g'(\theta)]^2 I(\theta)^{-1})$$

We can use the Delta Method to compute the variance of $\log(OR)$. Recall that

$$OR = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}$$

Let $\hat{\pi}_i$ be the MLE of π_i respectively. Then since $g(\pi_i) = \log(\Omega_i)$ is injective, $g(\hat{\pi}_i)$ is the MLE of π_i. By the Delta Method,

$$g(\hat{\pi}_i) \dot{\sim} \mathcal{N}\left(g(\pi_i), \frac{1}{n_i \pi_i (1-\pi_i)}\right)$$

By computation, we can see that $\text{Cov}(\hat{\pi}_1, \hat{\pi}_2) \to 0$, thus $\text{Cov}(g(\hat{\pi}_1), g(\hat{\pi}_2)) = 0$. So,

$$\text{Var}(\log(\hat{\theta})) = \sum_{i=1}^{2} \frac{1}{n_i \pi_i (1-\pi_i)}$$

where $\hat{\theta}$ is the MLE of the odds ratio. This means

$$\text{Var}(\log(\hat{\theta})) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

and a 95% approximate CI for $\log(\theta)$ is

$$\log(\hat{\theta}) \pm 1.96\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

3.5 Independence

Definition. Random variables X and Y are related if the conditional distribution of Y given $X = x$ changes as x changes.

Definition. Random variables X and Y are statistically independent if the conditional distribution of Y given $X = x$ is identical for all x.

- When the rows and columns of a contingency table are independent, then $\pi_{ij} = \pi_{i+}\pi_{j+}$, which implies $\pi_{1+} = \pi_{i|j}$

3.5.1 Tests of Independence

- An odds ratio or rate ratio of 1 indicates independent variables

Consider the table

	$Y = 0$	$Y = 1$	Total
$X = 0$	a	b	$a + b$
$X = 1$	c	d	$c + d$
Total	$a + c$	$b + d$	$n = a + b + c + d$

The joint probabilities are π_{ij}, marginals π_{i+} or π_{+j}. To test independence, consider the test

$$H_0 : X \perp Y (OR = 1) \qquad H_a : X \not\perp Y (OR \neq 1)$$

If the null hypothesis is true, then the expected value of cell (i, j) is $\mu_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j}$. There are three tests we consider:

1. Pearson's χ^2 test
2. Likelihood ratio test
3. Fisher's exact test

Which test we use depends on what we fixed.
If we assume only the row margins are fixed then each n_{ij} is Binomial(n_{i+}, π_i) distributed. We use Pearson's χ^2 test. The test statistic is

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}$$

which is χ^2_ν, where $\nu = (IJ - 1) - (I - 1) - (J - 1)$, where I is the number of rows, J is the number of columns. In the 2×2 case, $\nu = 1$, so the test statistic above is χ^2_1 distributed.

- If one of the cells is zero, then the test statistic blows up towards infinity
  - Use Yate's Continuity Correction, which has the form

$$\chi^2_c = \sum_i \sum_j \frac{(|n_{ij} - \mu_{ij}| - 0.5)^2}{\mu_{ij}}$$

For fixed n, the cell counts follow a multinomial distribution (since n is the sum of the cell counts). So, we have the likelihood ratio statistic

$$\Lambda = 2\sum_i \sum_j n_{ij} \log\left(\frac{n_{ij}}{\mu_{ij}}\right)$$

which is χ^2_1 under H_0, so asymptotically, the two test statistics presented are equal.
For Fisher's exact test, suppose the row and column margins are fixed (i.e.: n_{i+} and n_{+j} are fixed for each i, j), thus each cell count is Hypergeometric(n, n_{i+}, n_{+j}). For example, we have

$$P(n_{11} = t) = \frac{\binom{n_{1+}}{t}\binom{n_{2+}}{n_{+1}-t}}{\binom{n}{n_{+1}}}$$

In the a, b, c, d notation:

$$P(n_{11} = a) = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}}$$

So, since n_{11} is hypergeometric, then the one sided p-value is $P(n_{11} \geq a)$. For the two sided, define the set $S = \{t : P(n_{11} = t) \leq P(n_{11} = a)\}$. Then the two sided p-value is

$$\sum_{t \in S} P(n_{11} = t)$$

3.6 Confounding and Interaction Variables

3.6.1 Confounding

Suppose there are 3 categories, X, Y, Z, where Z has K categories and we want to know the association between X and Y. For each fixed value of Z, we have the tables

	$Y = 0$	$Y = 1$	Total
$X = 0$	a_i	b_i	$a_i + b_i$
$X = 1$	c_i	d_i	$c_i + d_i$
Total	$a_i + c_i$	$b_i + d_i$	n_i

The marginal table cells are $a = a_1 + a_2 + \cdots + a_K$.
A confounding variable is a third variable when marginally associated causes different conclusions from the conditionals. The presence of such a variable presents bias in the marginal association (the association with a instead of each a_i). To combat this, we use the Cochran-Mantel-Haenszel Test to test conditional independent of $2 \times 2 \times K$ tables.

$$H_0: \text{ Conditional independence in the } 2 \times 2 \times K \text{ tables}$$

Let the count in the first cell of the kth table be n_{11k}, so

$$\mu_{11k} = \frac{(a_k + b_k)(a_b + c_k)}{n_k} \qquad \text{Var}(n_{11k}) = \frac{(a_k + b_k)(a_k + c_k)(b_k + d_k)(c_k + d_k)}{n_k^2(n_k - 1)}$$

since n_{11k} is hypergeometrically distributed. Consider the test statistic

$$\chi^2_{CMH} = \frac{(\sum_k (n_{11k} - \mu_{11k}))^2}{\sum_k \text{Var}(n_{11k})}$$

which is χ^2_1 distributed.
For the kth conditional table, the conditional odds ratio can be estimated as

$$\hat{\theta}_k = \frac{a_k d_k}{b_k c_k}$$

Define weights

$$w_k = \frac{b_k c_k}{n_k}$$

The adjusted odds ratio can be obtained by weighting the conditional odds ratios.

$$\hat{\theta}_{CMH} = \frac{\sum_k \frac{a_k d_k}{n_k}}{\sum_k \frac{b_k c_k}{n}}$$

which has variance

$$\text{Var}(\log(\hat{\theta}_{CMH})) = \frac{\sum_k w_k^2 \left(\frac{1}{a_k} + \frac{1}{b_k} + \frac{1}{c_k} + \frac{1}{d_k}\right)}{(\sum_k w_k)^2}$$

3.6.2 Interaction

Sometimes, the conditional odds ratios at each level of Z are very different. If so, need to first check for statistical significance between the odds ratios, so we use the Test of Homogeneity.

$$H_0: \text{ the odds ratios are homogeneous}$$

Consider the test statistic

$$\chi^2_H = \sum_k \frac{(\log(\hat{\theta}_k) - \log(\hat{\theta}_{CMH}))^2}{\text{Var}(\log(\hat{\theta}_k))}$$

which is χ^2_1 distributed.
If we conclude there is a statistical difference, then we say there is an interaction between Z and X on Y (Z is an interaction term).

3.6.3 Confounding vs Interaction

For a confounding variable, we observe that the conditional effect measures are not statistically different, and that the marginal effect is not the same from the conditional, so we estimate the adjusted effect through weighting. For interactions, the conditional effect measures are statistically significantly different, and the marginal effect is not the same from the conditional. We don't want to adjust the effect and need to produce case specific effect measures since we want to observe the influence Z has on X for the outcome.