Analysis of Categorical Data
Chapter 1 and 2: Introduction and Contingency Table

Shaobo Jin

Department of Mathematics

Intended Learning Outcome

Through these chapters, you should be able to

1. describe different sampling themes,
2. compute odds ratios and understand their implications,
3. describe confounding and Simpson's paradox,
4. construct partial table and marginal table,
5. evaluate and test associations.

Categorical Variable

A categorical variable has a measurement scale consisting of a set of categories.

- Binary: Yes/No
- Nominal: Volvo/Volkswagen/Toyota/BMW
- Ordinal: Disagree/Neutral/Agree
- Counts: 0, 1, 2, ...

A continuous variable has a measurement scale consisting of all real numbers in an interval.

Distributions of Categorical Data

- Bernoulli distribution $Y \sim \text{Bernoulli}(\pi)$:

$$P(Y = y) = \pi^y (1-\pi)^{n-y}, \quad y = 0, 1,$$

  where π is the success probability.

- Binomial distribution $Y \sim \text{Binomial}(n, \pi)$:

$$P(Y = y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \quad y = 0, 1, ..., n.$$

  where n is the total number of trials and π is the success probability.

- Multinomial distribution $Y \sim \text{Multinonial}(\boldsymbol{n}, \boldsymbol{\pi})$:

$$P(n_1, n_2, ..., n_c) = \frac{n!}{n_1! n_2! \cdots n_c!} \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_c^{n_c},$$

  where $\pi_i = P(\text{outcome } i)$, $\sum_{i=1}^{c} \pi_i = 1$, and $\sum_{i=1}^{c} n_i = n$.

Distributions of Categorical Data

- Poisson distribution $Y \sim \text{Poi}(\mu)$:

$$P(Y = y) = \frac{\mu^y}{y!} \exp\{-\mu\}, \quad y = 0, 1, 2, ...$$

where μ is the mean.

- Negative binomial distribution $Y \sim \text{NegBin}(\mu, \phi)$:

$$P(Y = y) = \frac{\Gamma(y+\phi)}{\Gamma(\phi)\Gamma(y+1)} \left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\mu}{\mu+\phi}\right)^{y}, \quad y = 0, 1, 2, ...$$

where $\Gamma()$ is the gamma function, μ is the mean, and ϕ is the dispersion parameter.

  - $\mathbb{E}(Y) = \mu$ and $\text{var}(Y) = \mu + \mu^2/\phi$.
  - If $\phi \to \infty$, the negative binomial distribution reduces to the Poisson distribution.

Apply Appropriate Methods

A variable's measurement scale determines which statistical methods are appropriate.

- Apply methods appropriate for the actual scale.
- Methods for variables of one type usually can be used with variables at higher levels, but usually not at lower levels.
  - e.g., if we ignore ordering, ordinal data become nominal data. But ordinal data methods cannot be used with nominal data.

In this course, we focus on the case where the response variable is categorical. The covariates/features can be continuous or categorical.

Contingency Table

Let X be a categorical variable with I categories, and Y be a categorical variable with J categories. An $I \times J$ contingency table having I rows for categories of X and J columns for categories of Y displays the frequency counts of outcomes of (X, Y).

	Y			
X	1	2	$\cdots$	J
1	n_{11}	n_{12}	$\cdots$	n_{1J}
2	n_{21}	n_{22}	$\cdots$	n_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}

Joint Distribution

We can also tabulate the joint distribution of (X, Y) as an $I \times J$ table. Let $\pi_{ij} = P(X = i, Y = j)$. Then,

	Y			
X	1	2	$\cdots$	J
1	π_{11}	π_{12}	$\cdots$	π_{1J}
2	π_{21}	π_{22}	$\cdots$	π_{2J}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
I	π_{I1}	π_{I2}	$\cdots$	π_{IJ}

We must have

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \pi_{ij} = 1.$$

Marginal Distribution

	Y				
X	1	2	$\cdots$	J	Total
1	π_{11}	π_{12}	$\cdots$	π_{1J}	π_{1+}
2	π_{21}	π_{22}	$\cdots$	π_{2J}	π_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I	π_{I1}	π_{I2}	$\cdots$	π_{IJ}	π_{I+}
Total	π_{+1}	π_{+2}	$\cdots$	π_{+J}	1

Here

$$\begin{aligned} i\text{th row total: } P\left(X=i\right)=\pi_{i+} &= \sum_{j=1}^{J}\pi_{ij}, \\ j\text{th column total: } P\left(Y=j\right)=\pi_{+j} &= \sum_{i=1}^{I}\pi_{ij}. \end{aligned}$$

Conditional Distribution

	Y				
X	1	2	$\cdots$	J	Total
1	π_{11}	π_{12}	$\cdots$	π_{1J}	π_{1+}
2	π_{21}	π_{22}	$\cdots$	π_{2J}	π_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I	π_{I1}	π_{I2}	$\cdots$	π_{IJ}	π_{I+}
Total	π_{+1}	π_{+2}	$\cdots$	π_{+J}	1

Denote

$$\begin{aligned}\pi_{i|j} = P\left(X = i \mid Y = j\right) &= \frac{P\left(X = i, Y = j\right)}{P\left(Y = j\right)},\\ \pi_{j|i} = P\left(Y = j \mid X = i\right) &= \frac{P\left(X = i, Y = j\right)}{P\left(X = i\right)}.\end{aligned}$$

Independence

Two categorical variables are independent if

$$\pi_{ij} = \pi_{i+}\pi_{+j}, \text{ for all } i \text{ and } j.$$

When X and Y are independent,

$$\begin{aligned}
\pi_{i|j} = P\left(X = i \mid Y = j\right) &= P\left(X = i\right) = \pi_{i+}, \\
\pi_{j|i} = P\left(Y = j \mid X = i\right) &= P\left(Y = j\right) = \pi_{+j}.
\end{aligned}$$

Sampling

When working with a contingency table, sampling theme is important.

1. In Poisson sampling, the cell counts $\{N_{ij}\}$ follow independent Poisson distributions, i.e., $N_{ij} \sim \text{Poisson}(\mu_{ij})$. The joint probability mass function for outcomes $\{n_{ij}\}$ is

$$P(N_{11} = n_{11}, \cdots, N_{IJ} = n_{IJ}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{\mu_{ij}^{n_{ij}}}{n_{ij}!} \exp\{-\mu_{ij}\}.$$

   In Poisson sampling, the total sample size $n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$ is a random variable.

2. In multinomial sampling, the total sample size n is fixed but the row and column totals are not fixed. The cell counts $\{N_{ij}\}$ follow a multinomial distribution

$$P(N_{11} = n_{11}, \cdots, N_{IJ} = n_{IJ}) = \frac{n!}{n_{11}! n_{12}! \cdots n_{IJ}!} \prod_{i=1}^{I} \prod_{j=1}^{J} \pi_{ij}^{n_{ij}}.$$

Independent Multinomial Sampling

Besides Poisson sampling and multinomial sampling, other sampling themes are possible.

In independent multinomial sampling, the row sums are fixed and the rows follow independent multinomial distributions. Then,

$$P\left(N_{11}=n_{11},\cdots,N_{IJ}=n_{IJ}\right) = \prod_{i=1}^{I}\left[\frac{n_{i+}!}{n_{i1}!n_{i2}!\cdots n_{iJ}!}\prod_{j=1}^{J}\pi_{j|i}^{n_{ij}}\right].$$

where $\pi_{j|i}=P\left(\text{column } j \mid \text{row } i\right)$.

Independent Multinomial Sampling

X	Y: 1	$\cdots$	J	Total
1	$P(Y=1 \mid X=1)$	$\cdots$	$P(Y=J \mid X=1)$	1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I	$P(Y=1 \mid X=I)$	$\cdots$	$P(Y=J \mid X=I)$	1

X	Y: 1	$\cdots$	J
1	$P(X=1 \mid Y=1)$	$\cdots$	$P(X=1 \mid Y=J)$
2	$P(X=2 \mid Y=1)$	$\cdots$	$P(X=2 \mid Y=J)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
I	$P(X=I \mid Y=1)$	$\cdots$	$P(X=I \mid Y=J)$
Total	1	$\cdots$	1

Example: Sampling

Helmet use

Suppose that our data can be represented by the following 2×3 table

	Helmet Use		
Gender	No helmet	Traditional helmet	Airbag helmet
Female			
Male			

Which sampling theme is plausible for the following scenarios?

1. We take all cyclists passing Ångström,
2. We only choose 200 cyclists passing Ångström,
3. We take 100 female and 100 male passing Ångström.

Example: A Case-Control Study

Smoking and Lung Cancer

Suppose that 100 patients with lung cancer were admitted last year. For each patient, we record their past smoking behavior. We take another 100 patients without lung cancer and record their past smoking behavior.

	Lung Cancer	
Smoking	Yes	No
Yes		
No		
Total	100	100

Which sampling theme is plausible?

Example: Another Case-Control Study

Smoking and Lung Cancer

Suppose that 100 non-smokers and 100 smokers are recruited in a study. None of them has lung cancer. We will investigate how many will get lung cancer.

	Lung Cancer		
Smoking	Yes	No	Total
Yes			100
No			100

Which sampling theme is plausible?

Effects of Different Sampling

	Cancer	
Smoking	Yes	No
Yes		
No		

Smoking is a binary variable, and Cancer is also a binary variable.

1. Under multinomial sampling, we can obtain $P\left(\text{Smoking} \mid \text{Cancer}\right)$ and $P\left(\text{Cancer} \mid \text{Smoking}\right)$.
2. Under independent multinomial sampling with fixed row sums, we can obtain $P\left(\text{Cancer} \mid \text{Smoking}\right)$.
3. Under independent multinomial sampling with fixed column sums, we can obtain $P\left(\text{Smoking} \mid \text{Cancer}\right)$.

Quantity of Interest

Suppose that we have a 2×2 table. Let $\pi_{1|i}$ and $\pi_{1|j}$ be the success probability in row i and column j, respectively. We are often interested in

1. Difference of proportions: $\pi_{1|1} - \pi_{1|2}$,
2. Relative risk: $\pi_{1|1}/\pi_{1|2}$,
3. Odds:

$$\frac{\pi_{1|i}}{1 - \pi_{1|i}} = \frac{\pi_{i1}}{\pi_{i2}},$$

4. Odds ratio (denoted by θ):

$$\theta = \frac{\mu_{11}/\mu_{12}}{\mu_{21}/\mu_{22}} = \frac{\pi_{1|1}/\left(1 - \pi_{1|1}\right)}{\pi_{1|2}/\left(1 - \pi_{1|2}\right)} = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}},$$

where μ_{ij} be the cell expected frequencies corresponding to $X = i$ and $Y = j$.

Sampling

The above quantities often depend on the sampling themes.

- Under multinomial sampling, we can obtain $P(\text{row } i, \text{column } j)$, $P(\text{column } j \mid \text{row } i)$, and $P(\text{row } i \mid \text{column } j)$.
- Under independent multinomial sampling with fixed row sums, we should work with $P(\text{column } j \mid \text{row } i)$, but not $P(\text{row } i \mid \text{column } j)$.

An interesting property of odds ratio is that different sampling themes lead to the same way of computing the odds ratio. That is, the odds ratio can always be computed.

Estimate Odds Ratio in 2×2 Table

The sample odds ratio is

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}.$$

Estimate Odds Ratio

Table: Effect of planting time on the survival of plum root cuttings

	Survival	
Time	Dead	Alive
at once	217	263
in spring	365	115

Independence: Sufficient Condition

In a 2×2 case let $\pi_{ij} = P(X = i, Y = j)$. Suppose that the odds ratio is 1 as

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = 1.$$

Then, we must have

$$\pi_{12}\pi_{21} = \pi_{11}\pi_{22} = \pi_{11}(1 - \pi_{11} - \pi_{12} - \pi_{21}).$$

Hence,

$$\begin{aligned} \pi_{11} &= \pi_{12}\pi_{21} + \pi_{11}^2 + \pi_{11}\pi_{12} + \pi_{11}\pi_{21} \\ &= (\pi_{12} + \pi_{11})(\pi_{21} + \pi_{11}) \\ &= \pi_{1+}\pi_{+1}. \end{aligned}$$

Likewise, we can show $\pi_{ij} = \pi_{i+}\pi_{+j}$ for all i and j. Hence, $\theta = 1$ implies independence of X and Y.

$\theta = 1$: Sufficient Condition

In a 2×2 table, suppose that X and Y are independent. Then,

$$\pi_{ij} = \pi_{i+}\pi_{+j},$$

for all i and j. Then, the odds ratio satisfies

$$\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{\pi_{1+}\pi_{+1} \times \pi_{2+}\pi_{+2}}{\pi_{1+}\pi_{+2} \times \pi_{2+}\pi_{+1}} = 1.$$

Hence, independence of X and Y implies $\theta = 1$.

Therefore, in a 2×2 table, the odds ratio equals one if and only if X and Y are independent. If $\theta > 1$ (< 1), then the first row is more (less) likely to have a success than the second row, implying dependence.

Confounding

Confounding means that the effect of X on Y depend on the effect of other variables that can influence both X and Y.

$$
\begin{tikzcd}
 & Z \arrow[dl] \arrow[dr] & \\
X \arrow[rr] & & Y
\end{tikzcd}
$$

$$
\begin{tikzcd}
 & \text{Weight in September} \arrow[dl] \arrow[dr] & \\
\text{Diet provided by university} \arrow[rr] & & \text{Weight in December}
\end{tikzcd}
$$

Simpson's Paradox

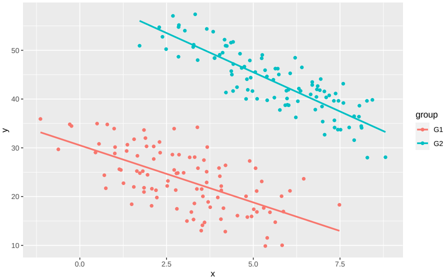
An example that has been analyzed to death is Berkeley university admission rate.

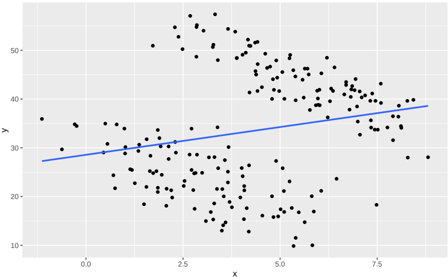
Male		Female	
Applicants	Admitted	Applicants	Admitted
8442	44%	4321	35%

	Male		Female	
Department	Applicants	Admitted	Applicants	Admitted
1	825	62%	108	82%
2	560	63%	25	68%
3	325	37%	493	34%
4	417	33%	375	35%
5	191	28%	393	24%
6	373	6%	341	7%

Women tend to apply to departments with low admission rates, but men tend to apply to departments with high admission rates.

Simpson's Paradox in Regression





Simpson's Paradox In Classification

Classifier	Classification: Correct	Incorrect
1	90	10
2	90	10

Observed	Classifier	Classification: Correct	Incorrect
Success	1	90	0
	2	81	9
Failure	1	0	10
	2	9	1

Partial Table

Suppose that Z is a confounder (or control variable) when studying the XY relationship. We can use the partial table that fixes the levels of Z. That is, for each level of Z, we make a contingency table for X and Y.

		Admission	
Department	Gender	Admitted	Not admitted
1	Male	512	314
	Female	88	19
2	Male	353	207
	Female	17	8
3	Male	120	205
	Female	168	325

Marginal Table

If we make a two-way contingency table by combining the partial tables, then it is a XY marginal table, ignoring Z.

		Admission	
Department	Gender	Admitted	Not admitted
1	Male	512	314
	Female	88	19
2	Male	353	207
	Female	17	8
3	Male	120	205
	Female	168	325

	Admission	
Gender	Admitted	Not admitted
Male	985	726
Female	273	352

Conditional Association

The associations in partial tables are called conditional associations. Suppose that we have (X, Y, Z) in a $2 \times 2 \times K$ table, where Z is a control variable. Let $\{\mu_{ijk}\}$ be the cell expected frequencies corresponding to $(X = i, Y = j, Z = k)$. Then,

$$\text{conditional odds ratio: } \theta_{XY(k)} = \frac{\mu_{11k}/\mu_{12k}}{\mu_{21k}/\mu_{22k}}, \text{ fixing } Z = k,$$
$$\text{marginal odds ratio: } \theta_{XY} = \frac{\mu_{11+}/\mu_{12+}}{\mu_{21+}/\mu_{22+}},$$

where $\mu_{ij+} = \sum_k \mu_{ijk}$.

Conditional Association

Sample values of $\theta_{XY(k)}$ and θ_{XY} replace μ by n as

$$\text{conditional odds ratio: } \hat{\theta}_{XY(k)} = \frac{n_{11k}/n_{12k}}{n_{21k}/n_{22k}}, \text{ fixing } Z = k,$$

$$\text{marginal odds ratio: } \hat{\theta}_{XY} = \frac{n_{11+}/n_{12+}}{n_{21+}/n_{22+}}.$$

Compute $\hat{\theta}_{XY(k)}$ and $\hat{\theta}_{XY}$

		Admission	
Dept.	Gender	Admitted	Not ad.
1	Male	512	314
	Female	88	19
2	Male	353	207
	Female	17	8
3	Male	120	205
	Female	168	325

	Admission	
Gender	Admitted	Not ad.
Male	985	726
Female	273	352

Different Types of Independence

Suppose that we have (X, Y, Z) in an $I \times J \times K$ table, where Z is a control variable.

- X and Y are conditionally independent at level k of Z if X and Y are independent when $Z = k$:

$$P(Y = j \mid X = i, Z = k) = P(Y = j \mid Z = k), \text{ for all } i, j.$$

- X and Y are conditionally independent given Z if X and Y are independent at every value of Z. It is often denoted by $X \perp Y \mid Z$. In other words, given Z, Y does not depend on X.
- X and Y are (marginally) independent if

$$P(Y = j \mid X = i) = P(Y = j), \text{ for all } i, j.$$

  It is often denoted by $X \perp Y$.

Marginal and Conditional Independence

Suppose that X and Y are conditionally independent given Z. Let $\pi_{ijk} = P(X = i, Y = j, Z = k)$. Then, for all (i, j, k),

$$\begin{aligned} \pi_{ijk} &= P(X = i, Y = j \mid Z = k) P(Z = k) \\ &= P(X = i \mid Z = k) P(Y = j \mid Z = k) P(Z = k) \\ &= \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}}. \end{aligned}$$

But,

$$P(X = i, Y = j) = \sum_{k=1}^{K} \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}} \neq \underbrace{\left(\sum_{k=1}^{K} \pi_{i+k}\right)}_{=P(X=i)} \underbrace{\left(\sum_{k=1}^{K} \pi_{+jk}\right)}_{=P(Y=j)}$$

Hence, conditional independence does not imply marginal independence.

Different Types of Independence

X, Y, and Z are mutually independent if

$$\pi_{ijk} = P(X = i)\, P(Y = j)\, P(Z = k), \text{ for all } i, j, k.$$

1. Mutual independence implies marginal independence.

$$\begin{aligned}
\pi_{ij+} &= \sum_{k=1}^{K} \pi_{ijk} \\
&= \sum_{k=1}^{K} \left(\pi_{i++}\pi_{+j+}\pi_{++k}\right) \\
&= \pi_{i++}\pi_{+j+}.
\end{aligned}$$

2. Mutual independence implies conditional independence.

Back to Odds in $2 \times 2 \times K$ Table

Suppose that X and Y are conditionally independent given Z in a $2 \times 2 \times K$ table. Then, in any partial table with a fixed k, the conditional odds must be

$$\theta_{XY(k)} = \frac{\mu_{11k}/\mu_{12k}}{\mu_{21k}/\mu_{22k}} = 1.$$

Suppose that X and Y are marginally independent ignoring Z in a $2 \times 2 \times K$ table. Then, the marginal odds must be

$$\theta_{XY} = \frac{\mu_{11+}/\mu_{12+}}{\mu_{21+}/\mu_{22+}} = 1.$$

Homogeneous Association

A $2 \times 2 \times K$ table has homogeneous XY association when

$$\theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(K)}.$$

That is, the effect of X on Y is the same at each category of Z. When this occurs, we say there is no interaction between two variables in their effects on the other variable.

- Suppose that $X \perp Y \mid Z$. Then the table has homogeneous XY association since $\theta_{XY(k)} = 1$ for all k.
- If there is interaction, the effect of X on Y depends on Z.

Test Homogeneous Association

For a $2 \times 2 \times K$ table, we can test homogeneous association using the Breslow-Day test.

$$
\begin{aligned}
H_0: &\quad \theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(K)}. \\
H_1: &\quad H_0 \text{ is not true.}
\end{aligned}
$$

Keep in mind that homogeneous association does not mean that the marginal odds ratio is the same as the conditional odds ratio.

	$Z=1$		$Z=2$	
X	$Y=1$	$Y=2$	$Y=1$	$Y=2$
1	100	20	100	100
2	200	20	60	30

Odds Ratio to $I \times J$ Table

Suppose that we have an $I \times J$ table.

1. There are $\binom{I}{2}$ pairs of rows and $\binom{J}{2}$ pairs of columns. For rows a and b and columns c and d, there are $\binom{I}{2}\binom{J}{2}$ odds ratios of the form
$$\frac{\mu_{ac}/\mu_{ad}}{\mu_{bc}/\mu_{bd}} = \frac{\mu_{ac}\mu_{bd}}{\mu_{bc}\mu_{ad}}.$$
2. The local odds ratios are
$$\theta_{ij} = \frac{\pi_{ij}/\pi_{i+1,j}}{\pi_{i,j+1}/\pi_{i+1,j+1}} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i+1,j}\pi_{i,j+1}},$$
for $i = 1, ..., I-1$ and $j = 1, ..., J-1$. There are $(I-1)(J-1)$ local odds ratios. They determine all odds ratios formed from pairs of rows and pairs of columns.

Maximum Likelihood Estimator

We often estimate a parameter θ (not necessarily odds ratio) by maximum likelihood estimator. Under some regularity conditions, the distribution of the maximum likelihood estimator can be approximated by

$$N\left(\theta, \mathcal{I}^{-1}(\theta)\right),$$

where

$$\mathcal{I}(\theta) = \operatorname{var}\left[\frac{\partial \ell(\theta)}{\partial \theta}\right] = -\mathrm{E}\left[\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^T}\right]$$

is the Fisher information matrix and $\ell(\theta)$ is the log-likelihood function.

Wald Statistics and Delta Method

The Wald test statistic for a unidimensional parameter θ is

$$Z = \frac{\hat{\theta} - \theta_0}{\text{standard error of } \hat{\theta}},$$

where θ_0 is some hypothesized value of θ. If the true value of θ is θ_0 and $\hat{\theta}$ is asymptotically normal, then Z is approximately $N(0,1)$. That is,

$$\hat{\theta} - \theta_0 \approx N\left(0, \operatorname{var}\left[\hat{\theta}\right]\right).$$

For a continuously differentiable function $g(\theta)$, the delta method implies that

$$g\left(\hat{\theta}\right) - g(\theta_0) \approx N\left(0, \left[\frac{\partial g(\theta_0)}{\partial \theta^T}\right] \operatorname{var}\left[\hat{\theta}\right] \left[\frac{\partial g(\theta_0)}{\partial \theta^T}\right]^T\right).$$

Likelihood Ratio Test

Suppose that we want to test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$. The likelihood ratio test statistic is

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta)}.$$

Under some regularity conditions,

$$-2 \log \lambda(x) \approx \chi_v^2,$$

when sample size increases, where the degrees of freedom v is the number of free parameters when $\theta \in \Theta_0 \cup \Theta_1$ minus the number of free parameters when $\theta \in \Theta_0$.