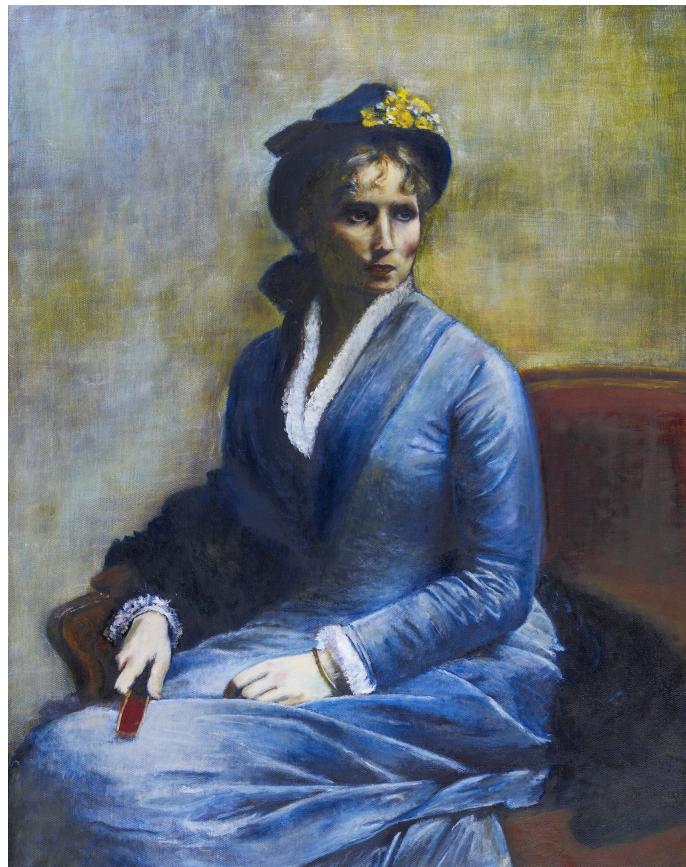


Introduction to Statistics

Anche Kuo

NCTU Computer Science



To my parent, Tiger and Sophie

Who gave me a world to dream and always encourage me to keep dreaming

Contents

1 Statistical Inferences	3
2 Point Estimate	5
2.1 Model of Moment	9
2.2 Maximum Likelihood Estimation	11
2.3 Best Estimator: UMVUE	15
3 Continuous to Point Estimation: sufficiency for UMVUE	20
4 Hypothesis Testing	27
4.1 Confidence Interval	28
4.1.1 How to construct C.I	28
4.1.2 C.I for normal mean	29
4.1.3 C.I for normal variance	30
4.1.4 C.I for difference of means	30
4.2 Best Confident Interval?	31
4.2.1 Shortest C.I	32
4.2.2 Example Shortest C.I	32
4.3 Approximate C.I	33
4.4 Testing Hypothesis	36
5 Most Powerful Test	38
6 UMP test	44
7 Likelihood Ratio Test	49
8 Goodness of Fit Test – chi-square Test	57
9 Appendix	60

1 Statistical Inferences

First recall the two definitions

Definition Suppose we have a random variable with p.d.f $f(x, \theta)$ where θ is an unknown vector, we then call θ a parameter and the set of θ 's possible values, denoted Θ , is called the parameter space

Definition For a random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$, any function $g(\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n)$ independent of parameter θ is called a statistics.

Let $\mathbb{X} \sim f$ be a random variable, if we know p.d.f f , then we know everything we want with that random variable. The problem in statistics is that we have a random point from $\{f(x, \theta) | \theta \in \Theta\}$ where function f is known but parameter θ is unknown. Statistics wants to predict θ . The method for predicting θ is called statistical inference. In the later text of this book, I shall refer "a random point from $\{f(x, \theta) | \theta \in \Theta\}$ " simply by "a random point from $f(x, \theta)$ " for brevity.

Two kinds of statistical inference are called **estimation** and **hypothesis testing**. Estimation means we try to guess the value of θ , hypothesis testing might not care about θ , but test the result of some experiment against the hypothesis. The key in hypothesis testing how sure can we know for certain, that the conclusion draw from the experiment is correct.

More formally, we categorize estimation into

- (1) **Point estimate:** given $\mathbb{X}_1, \dots, \mathbb{X}_n$, what is the function for estimating θ based solely on $\mathbb{X}_1, \dots, \mathbb{X}_n$.
- (2) **Interval estimate:** Given $\alpha \in [0, 1]$, find statistics $T_1(\mathbb{X}_1, \dots, \mathbb{X}_n)$ and $T_2(\mathbb{X}_1, \dots, \mathbb{X}_n)$ with $\alpha = \mathbb{P}_\theta(T_1 \leq \theta \leq T_2)$

Point estimation is meaningless if we don't have interval estimation, because we need to be able to argue how good is the estimation. Put in other words, after point estimation gives us $\hat{\theta}$, we need interval estimation to tell us how close is this $\hat{\theta}$ compared to real θ . That is what interval estimation does. Interval estimation introduces the notion of **confidence**. Confidence is the value α in (2). So interval estimation is, no matter how large is your confidence, we can tell you an interval such that the probability that θ lies within that interval is your confidence. This interval is called **confidence interval**. And of course, a good estimation has a smaller confidence interval for a given confidence.

And more formally, hypothesis testing is that given $\theta_0 \subset \mathbb{R}$ and hypothesis $H : \theta \in \theta_0$, find a rule to decide the acceptance or rejection of H . We will have more to say about interval estimation later.

2 Point Estimate

Definition We call a statistics $\hat{\theta} = \hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ an estimator of θ if it is used to estimate θ . If $X_1 = x_1, \dots, X_n = x_n$ is observed, the value $\hat{\theta}(x_1, \dots, x_n)$ is called estimate of θ .

Two problems for point estimation:

- (a) When many estimators are available, what is the criterion on "good" or "best" estimator.
- (b) Need rules for deriving estimators, depending on the criterion. Two will be introduced.

Before we introduce the first criterion on a good estimator, let's recall the definition of expectation on multivariate random variable:

If $T(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is a statistics (or an estimator), then

$$\mathbb{E}[T(\mathbb{X}_1, \dots, \mathbb{X}_n)] = \begin{cases} \int \dots \int T(x_1, \dots, x_n) f(x_1, \dots, x_n, \theta) dx_1 \dots dx_2 & \text{for continuous case.} \\ \sum \dots \sum T(x_1, \dots, x_n) f(x_1, \dots, x_n, \theta) & \text{for discrete case.} \end{cases}$$

Now the first criterion is based on the expectation:

Definition We call an estimation **unbiased** for parameter θ if it satisfies $\mathbb{E}_{\theta}[\hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)] = \theta$ for $\theta \in \Theta$. Here $\mathbb{E}_{\theta}[\hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)]$ means we are taking expectation of $\hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ after θ is given.

Example $\mathbb{X}_1, \dots, \mathbb{X}_n$ are i.i.d $normal(\mu, \sigma^2)$. Our interest is $\theta = \mu$. $\mathbb{E}[X_1] = \mu$, so $\hat{\theta} = X_1$ is unbiased estimation of μ . In fact, for all $k \in 1, 2, \dots, n$, $\hat{\theta} = \frac{1}{k} \sum_{i_1 < i_2 < \dots < i_k} X_i$ is unbiased estimation for μ .

Example For $X \sim normal(0, \sigma^2)$, X^2 is an unbiased estimator for X 's variance, because $\mathbb{E}[X^2] = (\mathbb{E}[X])^2 + \sigma^2 = \sigma^2$

We want to seek the best estimator in the class of unbiased estimator, but what is best? Depending on the criterion, the best estimator might be different. One intuitive criterion would be, if we have enough points from a random sample, the estimate based on these sample points will be very close to the real point. This sounds like the notion of convergence in analysis, but in probability, how do you define "close to"? As the two point being compared with are functions(random variables).

In analysis, there is the notion of uniform convergence and point-wise convergence for sequence of functions, but both are too strict for probability.

Two p.d.f's can be different on certain set of points but yield the same result under integration over identical set. That is, for p.d.f f and g , it is possible that

$$f \neq g \text{ and } \forall_{A \subset \mathbb{R}} \int_A f = \int_A g$$

and that result of integration is all we want in probability, we want to know the probability of certain event. So in the field of probability, let's look at a weaker version of convergence, called convergence in probability.

Definition We say that X_n converges to X **in probability**, if for every $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$. Denoted $X_n \xrightarrow{P} X$

With the notion of convergence in probability, let's look at the second criterion for a good estimator.

Definition We call an estimation $\hat{\theta}$ **consistent** for θ if $\hat{\theta} \xrightarrow{P} \theta$

Here is a useful theorem to show if an estimator is consistent.

Theorem Suppose that the estimation $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ of parameter θ , satisfies $\mathbb{E}[\hat{\theta}] = \theta$ (unbiased), or $\mathbb{E}[\hat{\theta}] \rightarrow \theta$ w.r.t n (asymptotically unbiased) and $\text{var}(\hat{\theta}) \rightarrow 0$, then $\hat{\theta}$ is consistent for θ , i.e,

$$\hat{\theta} \xrightarrow{P} \theta$$

Before proving this theorem, recall these two inequalities:
 Marcov's inequality: if $X > 0$, then $\mathbb{P}(X \geq u) \leq \frac{\mathbb{E}[X]}{u}$
 Chebyshev's inequality: If $\mathbb{E}[x] = \mu$ and $\text{var}(X) = \sigma^2$, then $\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

Proof

$$\begin{aligned} \mathbb{E}[(\hat{\theta} - \theta)^2] &= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2] \\ &= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + 2\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta)] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \theta)^2] \\ &= (\mathbb{E}[\hat{\theta}] - \theta \text{ is constant with respect to } \hat{\theta}) \text{ var}(\hat{\theta}) + 2\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]](\mathbb{E}[\hat{\theta}] - \theta) + (\hat{\theta} - \theta)^2 \\ &= (\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]] = 0) \text{ var}(\hat{\theta}) + (\mathbb{E}[\hat{\theta}] - \theta)^2 \end{aligned}$$

Finally by Marcov's inequality:

$$0 \leq \mathbb{P}(|\hat{\theta} - \theta| > \epsilon) = \mathbb{P}((\hat{\theta} - \theta)^2 > \epsilon) \leq \frac{\mathbb{E}[(\hat{\theta} - \theta)^2]}{\epsilon^2} \rightarrow 0$$

Example $\{X_n\}$ is a sequence of random variable, and

$$X_n \sim f_n(x) = \begin{cases} 1 - \frac{1}{2^n} & \text{if } x = 0 \\ \frac{1}{2^n} & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Show that $X_n \xrightarrow{P} 0$

Solution

By definition:

Note for $\epsilon > 1$, $\mathbb{P}(|X_n - 0| > \epsilon) = 0$ for all n, so we only look at $\epsilon \in (0, 1]$

$$\mathbb{P}(|X_n - 0| > \epsilon) = \mathbb{P}(X_n = 1) = \frac{1}{2^n} \rightarrow 0$$

By theorem:

$\mathbb{E}[X_n] = \frac{1}{2^n} \rightarrow 0$ and $\mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 = \frac{1}{2^n} - \frac{1}{2^{2n}} \rightarrow 0$, so by the above theorem, $X_n \xrightarrow{P} 0$

The following theorem offers a connection between consistency and unbiasedness.

Theorem Weak Law of Large Number(WLLN).

If $\mathbb{X}_1, \dots, \mathbb{X}_n$ is a random sample with finite mean μ and variance σ^2 exists, then $\bar{\mathbb{X}} \xrightarrow{P} \mu$

Proof Since $\mathbb{E}[\bar{\mathbb{X}}] = \mu$ and $var(\bar{\mathbb{X}}) = \frac{\sigma^2}{n} \rightarrow 0$, by the last theorem, $\bar{\mathbb{X}} \xrightarrow{P} \mu$.

Note how WLLN connects the axiomatic probability to the notion of relative frequency. This law basically tells you relative frequency will converge.

Recall that $var(\mathbb{X}) = \mathbb{E}[(\mathbb{X} - \mu)^2] = \mathbb{E}[\mathbb{X}^2] - \mathbb{E}[\mathbb{X}]^2$, so $\mathbb{E}[\mathbb{X}^2] = var(\mathbb{X}) + \mathbb{E}[\mathbb{X}]^2$

Example Y_1, \dots, Y_n are i.i.d. $f(y) = 3y^2 I_{y \in [0,1]}$, show that there exists a such that $\bar{Y} \xrightarrow{P} a$.

Solution Since $var(Y_i) = \frac{3}{5} - (\frac{3}{4})^2$ exists and $\mu = \frac{3}{4}$, by W.L.L.N, $a = \frac{3}{4}$

Example $Y_1, \dots, Y_n \xrightarrow{i.i.d} normal(\mu, \sigma^2)$, $n = 2k$. Show that estimator for variance $\hat{\sigma}^2 = \frac{1}{2k} \sum_{i=1}^k (Y_{2i} - Y_{2i-1})^2$ is unbiased and consistent for σ^2 .

Solution

$$Y_{2i} - Y_{2i-1} \sim normal(0, 2\sigma^2) \Rightarrow \frac{Y_{2i} - Y_{2i-1}}{\sqrt{2\sigma^2}} \sim normal(0, 1)$$

$$\Rightarrow \frac{\sum_{i=1}^k (Y_{2i} - Y_{2i-1})^2}{2\sigma^2} \sim \chi^2(k)$$

So

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{\sigma^2}{k} \frac{\sum_{i=1}^k (Y_{2i} - Y_{2i-1})^2}{2\sigma^2}\right] = \sigma^2$$

and

$$var(\hat{\sigma}^2) = \frac{\sigma^4}{k^2} var\left(\frac{\sum_{i=1}^k (Y_{2i} - Y_{2i-1})^2}{2\sigma^2}\right) = \frac{2\sigma^4}{k} \rightarrow 0$$

So it is consistent and unbiased.

Example If $U_n \xrightarrow{\mathbb{P}} u$, $g(x)$ is continuous at $x = u$, show that $g(U_n) \xrightarrow{\mathbb{P}} g(u)$

Solution By continuity of g

$$\forall \epsilon > 0 \exists \delta > 0 \forall x |x - u| \leq \delta \rightarrow |g(x) - g(u)| \leq \epsilon$$

$$\Rightarrow \{x : |g(x) - g(u)| > \epsilon\} \subset \{x : |x - u| > \delta\}$$

So

$$\forall \epsilon > 0 \mathbb{P}(|g(x) - g(u)| > \epsilon) \leq \mathbb{P}(|x - u| > \delta) \rightarrow 0$$

Example $Y_1, \dots, Y_n \xrightarrow{i.i.d} f(y) = (\theta + 1)y^\theta I_{\{y \in [0,1]\}}$, is $\frac{2\bar{Y}-1}{1-\bar{Y}} \xrightarrow{\mathbb{P}} \theta$?

Solution Let $g(x) = \frac{2x-1}{1-x}$, g is continuous at $x = \frac{\theta+1}{\theta+2}$, so

$$g(\bar{Y}) = \frac{2\bar{Y}-1}{1-\bar{Y}} \xrightarrow{\mathbb{P}} g\left(\frac{\theta+1}{\theta+2}\right) = \theta$$

Definition Moment: If \mathbb{X} is a random variable having a p.d.f $f(x, \theta)$, its population k th moment is defined as

$$\mathbb{E}_\theta[X^k] = \begin{cases} \sum_{\{x|f(x)>0\}} x^k f(x, \theta) & \text{if } \mathbb{X} \text{ is discrete} \\ \int_{-\infty}^{\infty} x^k f(x, \theta) dx & \text{if } \mathbb{X} \text{ is continuous} \end{cases}$$

We estimate the population moment by sample k th moment as

$$\frac{1}{n} \sum_{i=1}^n X_i^k$$

Why? Suppose that $\mathbb{X}_1, \dots, \mathbb{X}_n$ is a random sample from $f(x, \theta)$. Then, X_1^k, \dots, X_n^k is a random sample, so each X_i^k 's has mean $\mathbb{E}_\theta[X_1^k]$ and variance $\text{var}_\theta(X_1^k)$

By, WLLN, if $\mathbb{X}_1, \dots, \mathbb{X}_n$ is a random sample with mean μ and variance exists, then $\bar{\mathbb{X}} \xrightarrow{\mathbb{P}} \mu$. And since $\mathbb{X}_1, \dots, \mathbb{X}_n$ is a random sample from $f(x, \theta)$, $\frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{\mathbb{P}} \mathbb{E}_\theta[X^k]$

This following example shows why sample variance is called sample variance:

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample with mean μ and variance σ^2 . The sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is an unbiased ($\mathbb{E}[S^2] = \sigma^2$) and consistent ($S^2 \xrightarrow{\mathbb{P}} \sigma^2$) estimation for variance σ^2 .

Proof

$$\begin{aligned} \mathbb{E}[S^2] &= \mathbb{E}\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}[(X_i - \bar{X})^2] \\ &= \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}[X_i^2 - 2X_i\bar{X} + \bar{X}^2] = \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}[X_i^2] - 2\bar{X} \sum_{i=1}^n \mathbb{E}[X_i] + \sum_{i=1}^n \mathbb{E}[\bar{X}^2] \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \mathbb{E}[X_i^2] - n\mathbb{E}[\bar{X}^2] \right) = \frac{1}{n-1} \left(\sum_{i=1}^n (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \right) = \sigma^2 \end{aligned}$$

So S^2 is unbiased.

$$\begin{aligned} S^2 &= \frac{1}{n-1} \left(\sum_{i=1}^n \mathbb{E}[X_i^2] - n\mathbb{E}[\bar{X}^2] \right) = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2] - \mathbb{E}[\bar{X}^2] \right) \\ &\xrightarrow{\mathbb{P}} \frac{n}{n-1} (\mathbb{E}[\mathbb{X}^2] - \mathbb{E}[\bar{\mathbb{X}}^2]) (\because \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i^2] \xrightarrow{\mathbb{P}} \mathbb{E}[\mathbb{X}^2]) = \sigma^2 \end{aligned}$$

So S^2 is consistent.

We now introduce two methods for estimation of θ .

2.1 Model of Moment

Model of moment: Solve estimation from equations for estimating population moments by sample moments. This way it is guaranteed that the estimation

will be unbiased for $\mathbb{E}[\mathbb{X}^k]$.

Definition Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from a distribution with p.d.f $f(x, \theta)$

(a) If θ is univariate, the method of moment estimator $\hat{\theta}$ solves θ for

$$\bar{\mathbb{X}} = \mathbb{E}_\theta[\mathbb{X}]$$

(b) If $\theta = (\theta_1, \theta_2)$ is bivariate, the method of moment estimator $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ solves

$$\begin{cases} \bar{\mathbb{X}} = \mathbb{E}_{\theta_1, \theta_2}(\mathbb{X}) \\ \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbb{X}_i^2] = \mathbb{E}_{\theta_1, \theta_2}(\mathbb{X}^2) \end{cases}$$

(c) So in general, if $\theta = (\theta_1, \dots, \theta_k)$ is k-variate, the method of moment estimator $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ solves

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbb{X}_i] = \mathbb{E}_{\theta_1, \dots, \theta_k}(\mathbb{X}) \\ \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbb{X}_i^2] = \mathbb{E}_{\theta_1, \dots, \theta_k}(\mathbb{X}^2) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbb{X}_i^k] = \mathbb{E}_{\theta_1, \dots, \theta_k}(\mathbb{X}^k) \end{cases}$$

Example Suppose $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} Bernoulli(p)$, we solve for

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbb{X}_i] = p, \text{ so } \hat{p} = \bar{\mathbb{X}}$$

Note it is also consistent since $var(\bar{\mathbb{X}}) = \frac{p(1-p)}{n} \rightarrow 0$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} Poisson(\lambda)$, find the method of moment estimation of λ , and see if it is unbiased and consistent.

Solution Solve

$$\bar{\mathbb{X}} = \mathbb{E}_\lambda[\mathbb{X}] \Rightarrow \hat{\lambda} = \bar{\mathbb{X}}$$

it is unbiased. And since

$$var(\hat{\lambda}) = var(\bar{\mathbb{X}}) = \frac{\lambda}{n} \rightarrow 0$$

By WLLN, $\hat{\lambda} \xrightarrow{\mathbb{P}} \lambda$, so it is consistent.

Note if we want to find the method of moment estimation of $\theta = \sqrt{\lambda}$, just solve for $\hat{\lambda}$, then $\hat{\theta} = \sqrt{\hat{\lambda}}$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample, and each X'_i 's has mean μ and variance σ^2 . What is the method of moment estimation of $\theta = \{\mu, \sigma\}$

Solution Solve

$$\begin{cases} \bar{\mathbb{X}} = \mu \\ \frac{1}{n} \sum \mathbb{X}_i^2 = \mathbb{E}[\mathbb{X}^2] = \mu^2 + \sigma^2 \end{cases}$$

So $\hat{\mu} = \bar{\mathbb{X}}$, and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \mathbb{X}_i^2 - \hat{\mu}^2 = \frac{1}{n} (\sum_{i=1}^n \mathbb{X}_i^2 - n\bar{\mathbb{X}}^2)$. To prove that $\hat{\mu}$ is unbiased and consistent is trivial. But if we look at $\hat{\sigma}^2$

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (\mathbb{X}_i^2 - \bar{\mathbb{X}}^2)\right] = \frac{n-1}{n} \mathbb{E}\left[\frac{1}{n-1} \left(\sum_{i=1}^n \mathbb{X}_i^2 - n\bar{\mathbb{X}}^2\right)\right] = \frac{n-1}{n} \sigma^2$$

So $\hat{\sigma}^2$ is biased. But

$$\hat{\sigma}^2 = \frac{1}{n} \sum (\mathbb{X}_i^2 - n\bar{\mathbb{X}}^2) \xrightarrow{\mathbb{P}} \mathbb{E}[\mathbb{X}^2] - \mu^2 = \sigma^2$$

Therefore, $\hat{\sigma}^2$ is consistent.

2.2 Maximum Likelihood Estimation

Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta), \theta \in \Theta$. The joint p.d.f $f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$ for some $\theta \in \Theta$. As p.d.f, it satisfies

$$\forall \theta \in \Theta \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n, \theta) dx_1 \dots dx_n = 1$$

Definition Given x_1, \dots, x_n , the likelihood function of a random sample is defined on its joint p.d.f as a function of θ

$$L(\theta) = f(x_1, \dots, x_n, \theta), \theta \in \Theta$$

For x_1, \dots, x_n fixed, the value $L(\theta)$ is called the likelihood at θ . We know that true θ is in the parameter space Θ , but it is to be estimated from $X_1 = x_1, \dots, X_n = x_n$.

If $L(\theta_1) > L(\theta_2)$, we consider that θ_1 is more reliable than θ_2 to be the true θ when x_1, \dots, x_n are given.

Definition When $X_1 = x_1, \dots, X_n = x_n$ is observed, we have $L(\theta)$ corresponding to these x'_i 's. Let $\hat{\theta} = \hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ be any value of θ that maximizes

$L(\theta)$, we then call $\hat{\theta}$ the maximum likelihood estimator(m.l.e) of θ . When $X_1 = x_1, \dots, X_n = x_n$ is observed, we call $\hat{\theta}(x_1, \dots, x_n)$ the maximum likelihood estimate of θ .

Order statistics Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from a distribution with p.d.f f and c.d.f F . If Y_1, \dots, Y_n is a permutation of $\mathbb{X}_1, \dots, \mathbb{X}_n$ such that $Y_1 \leq Y_2 \leq \dots \leq Y_n$, we call (Y_1, \dots, Y_n) the order statistics of $(\mathbb{X}_1, \dots, \mathbb{X}_n)$.

Some properties:

- (a) $F_{Y_n}(y) = \mathbb{P}(Y_n \leq y) = \mathbb{P}(X_1 \leq y, \dots, X_n \leq y) = \prod_{i=1}^n \mathbb{P}(X_i \leq y) = F(y)^n$
- (b) From (a), $f_{Y_n}(y) = nF(y)^{n-1}f(y)$
- (c) $F_{Y_1}(y) = \mathbb{P}(Y_1 \leq y) = 1 - \mathbb{P}(Y_1 > y) = 1 - \mathbb{P}(X_1 > y, \dots, X_n > y) = 1 - \prod_{i=1}^n \mathbb{P}(X_i > y) = 1 - (1 - F(y))^n$
- (d) From (c), $f_{Y_1} = n(1 - F(y))^{n-1}f(y)$
- (e) $f_{Y_i}(y) = \binom{n}{i-1, 1, n-i} F(y)^{i-1} f(y) (1 - F(y))^{n-i}$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} \text{uniform}(0, \theta)$, what is the m.l.e of θ ? Is this m.l.e unbiased or consistent?

Solution P.d.f of X'_i 's is $f(x) = \frac{1}{\theta} I_{[0, \theta]}$

The likelihood function of $\mathbb{X}_1, \dots, \mathbb{X}_n$, $L(\theta) = \prod_{i=1}^n \frac{1}{\theta} I_{X_i \in [0, \theta]} \neq 0$ if and only if $\forall i \in 1, \dots, n, X_i \in [0, \theta]$. Let $Y_n = \max(\mathbb{X}_1, \dots, \mathbb{X}_n)$, then $L(\theta) = \frac{1}{\theta^n} I_{Y_n \in [0, \theta]}$. And the maximum of $L(\theta)$ is achieved at $\theta = \max(x_1, \dots, x_n)$.

The c.d.f of X is

$$F(x) = \int_0^x \frac{1}{\theta} dt = \frac{x}{\theta}, x \in [0, \theta]$$

The p.d.f of $\hat{\theta} = Y_n$ is

$$g_n(y) = n \frac{y^{n-1}}{\theta^n}, y \in [0, \theta]$$

so

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}[Y_n] = \int_0^\theta y g_n(y) dy = \int_0^\theta n \frac{y^n}{\theta^n} dy = \frac{n}{n+1} \theta$$

\Rightarrow m.l.e is not unbiased.

But $\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}] \rightarrow 0$, so $\hat{\theta}$ is asymptotically unbiased. $\mathbb{E}[\hat{\theta}^2] = \mathbb{E}[y_n^2] = \frac{n}{n+2} \theta^2$, so

$$\text{var}(\hat{\theta}) = \frac{n}{n+2} \theta^2 - (\frac{n}{n+1} \theta)^2 \rightarrow 0 \Rightarrow \hat{\theta} \text{ is consistent for } \theta$$

Note that $\hat{\theta}' = \frac{n+1}{n}\hat{\theta}$ is unbiased and consistent (Think about it, it is actually very reasonable and intuitive that $\hat{\theta}'$ is a better estimator).

Example Let $Y \sim binomial(n, p)$, discuss m.l.e of p .

Solution Likelihood function of Y given $Y = y$ is

$$L(p) = f_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

To find its maximum value,

$$D_p \ln(L(y, p)) = \frac{y}{p} - \frac{n-y}{1-p} = 0 \text{ iff } p = \frac{y}{n}$$

So the m.l.e $\hat{p} = \frac{Y}{n}$. $E[\hat{p}] = \frac{np}{n} = p$, so \hat{p} is unbiased. And $var(\hat{p}) = var(\frac{Y}{n}) = \frac{1}{n^2} var(Y) = \frac{p(1-p)}{n} \rightarrow 0$, so \hat{p} is consistent.

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n \sim normal(\mu, \sigma^2)$, what is m.l.e of μ, σ^2 ?

Solution The likelihood function

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

To find maximum value of μ

$$\begin{aligned} D_\mu \ln(L(\mu, \sigma^2)) &= D_\mu \left[\sum_{i=1}^n \left(-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2} \right) \right] = \frac{(\sum_{i=1}^n x_i - n\mu)}{\sigma^2} \\ &= 0 \text{ iff } \mu = \bar{x}, \text{ so } \hat{\mu} = \bar{X} \end{aligned}$$

To find maximum value of σ^2

$$D_{\sigma^2} L(\bar{x}, \sigma^2) = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^3} = 0 \text{ iff } \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

So $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$.

$\hat{\mu}$ is unbiased and consistent is trivial. So we look only at σ^2 .

$$\mathbb{E}[\hat{\sigma}^2] = \frac{\sigma^2}{n} \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}\right] = \frac{n-1}{n} \sigma^2$$

Here we used the property $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$. So $\hat{\sigma}^2$ is only asymptotically unbiased.

$$var(\hat{\sigma}^2) = \frac{\sigma^4}{n^2} var\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}\right) = \frac{\sigma^4}{n^2} 2(n-1) \rightarrow 0$$

So $\hat{\sigma}^2$ is consistent.

Example $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} uniform(0, 2\theta)$, find m.l.e of θ .

Solution The likelihood function $L(\theta) = \prod_{i=1}^n \frac{1}{2\theta} I_{x_i \in [0, 2\theta]}(x_i) = \frac{1}{(2\theta)^n} I_{\theta \in [\frac{y_n}{2}, \infty]}(\theta)$. So $\hat{\theta} = \frac{Y_n}{2}$. Note that $\mathbb{E}[\frac{Y_n}{2\theta}] = \mathbb{E}[beta(n, 1)] = \frac{n}{n+1}$, so $\mathbb{E}[\frac{Y_n}{2}] = \frac{n\theta}{n+1}$.

Example $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} uniform(\theta_1, \theta_2)$, $\theta_1 < \theta_2$, find m.l.e of θ_1, θ_2 .

Solution The likelihood function $f(x, \theta) = \prod_{i=1}^n \frac{1}{\theta_2 - \theta_1} I_{x_i \in [\theta_1, \theta_2]}(x_i) = \frac{1}{(\theta_2 - \theta_1)^n} I_{\theta_1 \in [-\infty, Y_1]}(\theta_1) I_{\theta_2 \in [Y_n, \theta_2]}(\theta_2)$, so $\hat{\theta}_1 = Y_1, \hat{\theta}_n = Y_n$

$$\begin{aligned} \mathbb{E}[Y_1] &= \int_{\theta_1}^{\theta_2} n\left(1 - \frac{y - \theta_1}{\theta_2 - \theta_1}\right)^{n-1} \frac{1}{\theta_2 - \theta_1} y dy = \frac{n}{(\theta_2 - \theta_1)^n} \int_{\theta_1}^{\theta_2} (\theta_2 - y)^{n-1} y dy \\ &= \frac{n}{(\theta_2 - \theta_1)^n} \left(y\left[-\frac{(\theta_2 - y)^n}{n}\right]_{\theta_1}^{\theta_2} + \int_{\theta_1}^{\theta_2} \frac{(\theta_2 - y)^n}{n} dy\right) = \frac{n\theta_1 + \theta_2}{n+1} \rightarrow \theta_1 \\ \mathbb{E}[Y_n] &= \int_{\theta_1}^{\theta_2} n\left(\frac{y - \theta_1}{\theta_2 - \theta_1}\right)^{n-1} \frac{1}{\theta_2 - \theta_1} y dy = \frac{n}{(\theta_2 - \theta_1)^n} \int_{\theta_1}^{\theta_2} (y - \theta_1)^{n-1} y dy \\ &= \frac{n}{(\theta_2 - \theta_1)^n} \left(y\left[\frac{(y - \theta_1)^n}{n}\right]_{\theta_1}^{\theta_2} - \int_{\theta_1}^{\theta_2} \frac{(y - \theta_1)^n}{n} dy\right) = \frac{\theta_1 + n\theta_2}{n+1} \rightarrow \theta_2 \end{aligned}$$

Example $\mathbb{X}_1, \dots, \mathbb{X}_n$ are independent, and $X_i \sim poisson(\lambda x_i)$ find m.l.e of λ .

Solution The likelihood function L given $X_1 = k_1, \dots, X_n = k_n$ is observed, is $L(\lambda) = \prod_{i=1}^n e^{-\lambda x_i} \frac{(\lambda x_i)^{k_i}}{k_i!}$, so

$$\frac{d}{d\lambda} \ln(L(\lambda)) = \frac{d}{d\lambda} \sum_{i=1}^n (k_i \ln(x_i \lambda) - x_i \lambda - \ln(k_i!)) = \sum_{i=1}^n \left(\frac{k_i}{\lambda} - x_i\right)$$

and $\frac{d}{d\lambda} \ln(L(\lambda)) = 0$ only when $\lambda = \frac{\sum k_i}{\sum x_i}$, and by second derivative test, we know this is when $L(\lambda)$ achieves maximum. So $\hat{\lambda} = \frac{\bar{Y}}{\bar{x}}$

Example $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} Bernoulli(\theta), \Theta \in [0, \frac{1}{2}]$. Find m.l.e of θ .

Solution The maximum likelihood function L given $X_1 = x_1, \dots, X_n = x_n$ is observed, is $L(\lambda) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$, and $\ln(L(\theta)) = \sum x_i \ln(\theta) + (n - \sum x_i) \ln(1 - \theta)$

$$\frac{d}{d\lambda} \ln(L(\lambda)) = \frac{\sum x_i}{\theta} - \frac{(n - \sum x_i)}{1 - \theta} = 0 \text{ if } \theta = \frac{\sum x_i}{n}$$

, and by second derivative test, we know this is when $L(\lambda)$ achieves maximum. Adding the restriction such that $\Theta \in [0, \frac{1}{2}]$, applying some simple calculus, we can show $\hat{\lambda} = \min(\frac{1}{2}, \bar{X})$

Suppose that we have $\hat{\theta} = \hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ for parameter θ , and our intent is m.l.e of $\tau(\theta)$, a function of θ . What is the m.l.e of $\tau(\theta)$?

The space of $\tau(\theta)$ is $\mathcal{T} = \tau(\Theta) = \{z | \exists_{\theta \in \Theta} s.t \tau(\theta) = z\}$.

Theorem If $\hat{\theta} = \hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is the m.l.e of θ and τ is a 1-1 function of θ , then the m.l.e of $\tau(\theta)$ is $\tau(\hat{\theta})$

Proof Because τ is 1-1, for each $y \in \tau(\Theta)$, there exists a unique $x \in \Theta$ such that

$$L_{\tau(\theta)}(y) = L_\theta(x)$$

Where $L_{\tau(\theta)}$ and L_θ are likelihood function for $\tau(\theta)$ and θ respectively. And also

$$L_\theta(x) = L_\theta(\tau^{-1}(\tau(x)))$$

So $L_{\tau(\theta)} = L_\theta(\tau^{-1})$. Now for all $t \in \tau(\Theta)$, we have

$$L_{\tau(\theta)}(y) = L_\theta(\tau^{-1}(y)) = L_\theta(x) \leq L_\theta(\hat{\theta}) = L_{\tau(\theta)}(\tau(\hat{\theta}))$$

Example $Y \sim binomial(n, p)$, m.l.e of p is $\hat{p} = \frac{Y}{n}$.

- (1) $\sqrt{\frac{Y}{n}}$ is the m.l.e for $\tau = \sqrt{p}$
- (2) $(\frac{Y}{n})^2$ is the m.l.e for $\tau = p^2$
- (3) $e^{\frac{Y}{n}}$ is the m.l.e for $\tau = e^p$
- (4) $\ln(\frac{Y}{n})$ is the m.l.e for $\tau = \ln(p)$

2.3 Best Estimator: UMVUE

An unbiased estimation $\hat{\theta} = \hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is called a uniformly minimum variance unbiased estimation (UMVUE), or best estimation if for any unbiased

estimator $\hat{\theta}_i$, we have

$$var_{\theta}(\hat{\theta}) \leq var_{\theta}(\hat{\theta}_i) \text{ for every } \theta \in \Theta$$

Two ways in deriving UMVUE will be introduced.

Cramer-Rao lower bound for UMVUE

Regularity conditions:

- (a) Support of X ($\{x|f(x, \theta) > 0\}$) is an open interval
- (b) Set $\{x|f(x, \theta) = 0\}$ or $\{x|f(x, \theta) > 0\}$ (support of f) is independent of θ
- (c) $\int D_{\theta} f(x, \theta) dx = D_{\theta} \int f(x, \theta) dx = 0$
- (d) If $T = t(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is unbiased, then

$$\int \dots \int t(x_1, \dots, x_n) D_{\theta} f(x_1, \dots, x_n, \theta) dx_1 \dots dx_n = D_{\theta} \int \dots \int t(x_1, \dots, x_n) f(x_1, \dots, x_n, \theta) dx_1 \dots dx_n$$

Some clarification here first, in (b), since p.d.f f is positive, so $\mathbb{R} = \{x|f(x, \theta) = 0\} \cup \{x|f(x, \theta) > 0\}$. This means either one of $\{x|f(x, \theta) = 0\}$ or $\{x|f(x, \theta) > 0\}$ is independent of θ . guarantees the independence of the other. In (c), on what condition is this allowed is given in appendix.

Theorem (Cramer-Rao lower bound) Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta)$. Suppose that the regularity conditions hold. If $\hat{\tau}(\theta) = t(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is an unbiased estimator for $\tau(\theta)$, then

$$var_{\theta}(\hat{\tau}(\theta)) \geq \frac{(\tau'(\theta))^2}{n \mathbb{E}_{\theta}[(D_{\theta} \ln f(X_1, \theta))^2]} = \frac{(\tau'(\theta))^2}{-n \mathbb{E}_{\theta}[D_{\theta}^2 \ln f(X_1, \theta)]} \text{ for } \theta \in \Theta$$

Proof To simplify notation, we let \mathbf{X} be the random vector $(\mathbb{X}_1, \dots, \mathbb{X}_n)$ and $\mathbf{x} \in \mathbb{R}^n$ be a real vector. And we use $\Delta(\mathbf{x}, \theta)$ to denote $\prod_{i=1}^n f(x_i, \theta)$. Now consider only the continuous distribution. First note that

$$E[D_{\theta} \ln(f(X_1, \theta))] = \int_{-\infty}^{\infty} D_{\theta} \ln f(x, \theta) f(x, \theta) dx = \int_{-\infty}^{\infty} D_{\theta} f(x, \theta) dx = D_{\theta} \int_{-\infty}^{\infty} f(x, \theta) dx = 0$$

and

$$D_{\theta} \int \dots \int \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n = D_{\theta} 1 = 0$$

Since $\hat{\tau}(\theta)$ is unbiased

$$\tau(\theta) = \mathbb{E}_{\theta}[\hat{\tau}(\theta)] = \int \dots \int t(\mathbf{x}) \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n$$

Taking derivative on both sides, we have

$$\begin{aligned}
\tau'(\theta) &= D_\theta \int \dots \int t(\mathbf{x}) \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n - \tau(\theta) D_\theta \int \dots \int \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n \\
&\quad \int \dots \int t(\mathbf{x}) D_\theta \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n - \int \dots \int \tau(\theta) D_\theta \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n \\
&= \int \dots \int (t(\mathbf{x}) - \tau(\theta)) D_\theta \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n
\end{aligned}$$

Note

$$\begin{aligned}
D_\theta \Delta(\mathbf{x}, \theta) &= \sum_{j=1}^n D_\theta f(x_j, \theta) \prod_{i \neq j} f(x_i, \theta) = \sum_{j=1}^n [D_\theta \ln f(x_j, \theta) f(x_j, \theta) \prod_{i \neq j} f(x_i, \theta)] \\
&= \sum_{j=1}^n D_\theta \ln f(x_j, \theta) \Delta(\mathbf{x}, \theta)
\end{aligned}$$

We now have

$$\begin{aligned}
\tau'(\theta) &= \int \dots \int (t(\mathbf{x}) - \tau(\theta)) \sum_{j=1}^n D_\theta \ln(f(x_j, \theta)) \Delta(\mathbf{x}, \theta) dx_1 \dots dx_n \\
&= \mathbb{E}_\theta [(t(\mathbf{X}) - \tau(\theta)) \sum_{j=1}^n D_\theta \ln(f(X_j, \theta))]
\end{aligned}$$

And by Cauchy-Swartz inequality of random variable

$$(E[XY])^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2]$$

We have

$$(\tau'(\theta))^2 \leq \mathbb{E}[(t(\mathbf{X}_1, \dots, \mathbf{X}_n) - \tau(\theta))^2] \mathbb{E}\left[\left(\sum_{j=1}^n D_\theta \ln(f(X_j, \theta))\right)^2\right]$$

and

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{j=1}^n D_\theta \ln(f(X_j, \theta))\right)^2\right] &= \sum_{j=1}^n \mathbb{E}[(D_\theta \ln(f(X_j, \theta)))^2] + \sum_{i \neq j} \mathbb{E}[D_\theta \ln(f(X_i, \theta)) D_\theta \ln(f(X_j, \theta))] \\
&= \sum_{j=1}^n \mathbb{E}[(D_\theta \ln(f(X_j, \theta)))^2] = n \mathbb{E}[(D_\theta \ln(f(X_1, \theta)))^2]
\end{aligned}$$

Hence

$$var(\hat{\tau}) = \mathbb{E}[(t(\mathbb{X}_1, \dots, \mathbb{X}_n) - \tau(\theta))^2] \geq \frac{(\hat{\tau}'(\theta))^2}{n\mathbb{E}_{\theta}[D_{\theta} \ln(f(X_1, \theta))^2]}$$

It remains to prove the final equality, and showing $\mathbb{E}[(D_{\theta} \ln f(\mathbf{x}, \theta))^2] = -\mathbb{E}[D_{\theta}^2 \ln f(\mathbf{x}, \theta)]$ suffice. Because

$$\mathbb{E}[D_{\theta} \ln(f(X_1, \theta))] = \int D_{\theta} \ln f(\mathbf{x}, \theta) f(\mathbf{x}, \theta) dx = 0$$

we can differentiate both sides and obtain

$$\begin{aligned} & D_{\theta} \int D_{\theta} \ln f(\mathbf{x}, \theta) f(\mathbf{x}, \theta) dx \\ &= \int D_{\theta}^2 \ln f(\mathbf{x}, \theta) f(\mathbf{x}, \theta) + D_{\theta} \ln f(\mathbf{x}, \theta) D_{\theta} f(\mathbf{x}, \theta) dx \\ &= \int D_{\theta}^2 \ln f(\mathbf{x}, \theta) f(\mathbf{x}, \theta) + (D_{\theta} \ln f(\mathbf{x}, \theta))^2 f(\mathbf{x}, \theta) dx \\ &= \mathbb{E}[D_{\theta}^2 \ln f(\mathbf{x}, \theta)] + \mathbb{E}[(D_{\theta} \ln f(\mathbf{x}, \theta))^2] = 0 \end{aligned}$$

And the proof is complete.

Theorem If there is an unbiased estimator $\hat{\tau}(\theta)$ with variance achieving Cramer-Rao lower bound, then $\hat{\tau}(\theta)$ is UMVUE of $\tau(\theta)$

Proof By last theorem

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} Poisson(\lambda)$

Solution P.d.f of X'_i s is $f(x, \lambda)$, and

$$D_{\lambda}^2 \ln(f(x, \lambda)) = -\frac{x}{\lambda^2} \Rightarrow \frac{1}{-n\mathbb{E}_{\theta}[D_{\theta}^2 \ln(f(x, \theta))]} = \frac{\lambda}{n}$$

Since $\hat{\lambda} = \bar{X}$ has variance $\frac{\lambda}{n}$, it is UMVUE.

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} Bernoulli(p)$

Solution

$$D_p \ln(f(x, p)) = \frac{x}{p} - \frac{1-x}{1-p}$$

$$D_p^2 \ln(f(x, p)) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$

$$\mathbb{E}[D_p^2 \ln(f(x, p))] = -\frac{1}{p(1-p)}$$

so Cramer-Rao lower bound is $\frac{p(1-p)}{n}$. And again, m.l.e of p , \bar{X} is UMVUE of p .

3 Continuous to Point Estimation: sufficiency for UMVUE

Sufficient statistics Let A and B be two events. The conditional probability of A given B is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

And we showed that $\mathbb{P}(x|B)$ is a probability function too.

In estimation of θ , we have a random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ from distribution $f(x, \theta)$. The information we have about θ is contained in $\mathbb{X}_1, \dots, \mathbb{X}_n$.

Let $U = U(\mathbb{X}_1, \dots, \mathbb{X}_n)$ be a statistic having p.d.f $f_U(u, \theta)$. The conditional p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ given $U = u$ is

$$f(x_1, \dots, x_n, \theta|u) = \begin{cases} \frac{f(x_1, \dots, x_n, \theta)}{f_U(u, \theta)} & \text{if } u(x_1, \dots, x_n) = u \\ \frac{0}{f_U(u, \theta)} & \text{if } u(x_1, \dots, x_n) \neq u \end{cases}$$

If for any $U = u$, conditional p.d.f $f(x_1, \dots, x_n, \theta|u)$ is unrelated to θ , i.e. $\forall_u \forall_{\theta_1, \theta_2} f(x_1, \dots, x_n, \theta_1|u) = f(x_1, \dots, x_n, \theta_2|u)$, then the random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ contains no more information about θ than $U = u$ is observed. This indicates that $U = u$ contains information about θ exactly the same amount as $\mathbb{X}_1, \dots, \mathbb{X}_n$.

Definition Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta), \theta \in \Theta$. We call a statistic $U = u(x_1, \dots, x_n)$ a sufficient statistic if, for any value $U = u$, the conditional p.d.f $f(x_1, \dots, x_n, \theta|u)$ and its domain do not depend on θ .

Example Let $U = (\mathbb{X}_1, \dots, \mathbb{X}_n)$, then

$$f(x_1, \dots, x_n, \theta|u = (\acute{x}_1, \dots, \acute{x}_n)) = \begin{cases} \frac{f(\acute{x}_1, \dots, \acute{x}_n, \theta)}{f(\acute{x}_1, \dots, \acute{x}_n, \theta)} = 1 & \text{if } (x_1, \dots, x_n) = (\acute{x}_1, \dots, \acute{x}_n) \\ \frac{0}{f(\acute{x}_1, \dots, \acute{x}_n, \theta)} & \text{if } (x_1, \dots, x_n) \neq (\acute{x}_1, \dots, \acute{x}_n) \end{cases}$$

which is independent of θ , so random sample $(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is a sufficient statistic.

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta), \theta \in \Theta$. Consider the order statistics Y_1, \dots, Y_n . If $(Y_1, \dots, Y_n) = (y_1, \dots, y_n)$ is observed, sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ has equal chance in the set

$$A = \{(x_1, \dots, x_n) \in \mathbb{R}^n | (y_1, \dots, y_n) \text{ is a permutation of } (x_1, \dots, x_n)\}$$

Then the condition p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ given $(Y_1, \dots, Y_n) = (y_1, \dots, y_n)$ is

$$\begin{cases} \frac{1}{n!} & \text{if } (x_1, \dots, x_n) \in A \\ 0 & \text{otherwise} \end{cases}$$

which is independent of θ . Hence (Y_1, \dots, Y_n) is a sufficient statistic.

Why sufficiency?

We want a statistic with dimension as small as possible and that contains information about θ with the same amount as $(\mathbb{X}_1, \dots, \mathbb{X}_n)$.

Definition If $U = U(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is a sufficient statistic whose range has the smallest dimension, it is called the minimal sufficient statistic.

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $Bernoulli(p)$, $p \in [0, 1]$, the joint p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ is

$$f(x_1, \dots, x_n, p) = p^{\sum x_i} (1-p)^{n-\sum x_i}$$

Consider the statistics $Y = \sum_{i=1}^n X_i$ which has binomial distribution. The conditional p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ given $Y = y$ is

$$f(\mathbf{x}, p|y) = \begin{cases} \frac{p^y (1-p)^{n-y}}{\binom{n}{y} p^y (1-p)^{n-y}} = \frac{y!(n-y)!}{n!} & \text{if } \sum_{i=1}^n x_i = y \\ 0 & \text{otherwise} \end{cases}$$

So Y is sufficient statistics. Furthermore, Y is minimal sufficient statistic, because 1 is the minial possible dimension.

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $U(0, \theta)$. Show that the n th-order statistic $Y_n = \max(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is a minimal sufficient statistic.

Solution The joint p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ is

$$f(x_1, \dots, x_n, \theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } x'_i s \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$

And recall that the p.d.f. of Y_n is $f_{Y_n}(y) = n(\frac{y}{\theta})^{n-1} \frac{1}{\theta}$ for $y \in [0, \theta]$. When $Y_n = y$ is given, all $x'_i s$ satisfies $x_i \leq y$. The conditional p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ given $Y_n = y$ is

$$f(x_1, \dots, x_n, \theta|y) = \begin{cases} \frac{f(x_1, \dots, x_n, \theta, Y_n=y)}{f_{Y_n}(y)} = \frac{\frac{1}{\theta^n}}{n(\frac{y}{\theta})^{n-1} \frac{1}{\theta}} = \frac{1}{ny^{n-1}} & \text{if } x'_i s \leq y \\ 0 & \text{otherwise} \end{cases}$$

So Y_n is a minimal sufficient statistic.

Now we want an easy way to derive sufficient statistics.

Theorem (Factorization theorem) Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta)$. A statistic $U = U(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is sufficient for θ if and only if there exists functions $K_1, K_2 \geq 0$ such that the joint p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ may be formulated as $f(x_1, \dots, x_n, \theta) = K_1(U(x_1, \dots, x_n), \theta)K_2(x_1, \dots, x_n)$

Proof (\Rightarrow) If U is sufficient, then the conditional p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ given $U = \mu$ is

$$f(x_1, \dots, x_n, \theta|u) = \begin{cases} \frac{f(x_1, \dots, x_n, \theta)}{f_U(u, \theta)} & \text{if } u(x_1, \dots, x_n) = \mu \\ 0 & \text{if } u(x_1, \dots, x_n) \neq \mu \end{cases}$$

is unrelated to θ . so $K_1 = f_U(u, \theta)$, $K_2 = f(x_1, \dots, x_n, \theta|u)$.

(\Leftarrow) Suppose that $f(x_1, \dots, x_n, \theta) = K_1(u(x_1, \dots, x_n), \theta)K_2(x_1, \dots, x_n)$. Let $Y_1 = u(\mathbb{X}_1, \dots, \mathbb{X}_n)$, and for $i \in 2, \dots, n$, $Y_i = u_i(\mathbb{X}_1, \dots, \mathbb{X}_n)$ be a 1-1 transformation with inverse function $X_i = w_i(Y_1, \dots, Y_n)$, and Jacobian

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \dots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

The joint p.d.f of Y_1, \dots, Y_n is

$$\begin{aligned} f_{Y_1, \dots, Y_n}(y_1, \dots, y_n, \theta) &= f(w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n), \theta)|J| \\ &= K_1(y_1, \theta)K_2(w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n))|J| \end{aligned}$$

The p.d.f of $U = Y_1$ is

$$\begin{aligned} f_U(y, \theta) &= \int \dots \int K_1(y_1, \theta)K_2(w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n))|J|dy_2 \dots dy_n \\ &= K_1(y_1, \theta) \int \dots \int K_2(w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n))|J|dy_2 \dots dy_n \end{aligned}$$

And note that everything in the integral is unrelated to θ . Then the conditional p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ given $U = u$ is

$$f(x_1, \dots, x_n, \theta|u) = \begin{cases} \frac{f(x_1, \dots, x_n, \theta)}{f_U(u, \theta)} = \frac{K_2(x_1, \dots, x_n)}{\int \dots \int K_2 dy_2 \dots dy_n} & \text{if } u(x_1, \dots, x_n) = u \\ 0 & \text{otherwise} \end{cases}$$

Example $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d}{\sim} Poisson(\lambda)$, so joint p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ is

$$f(x_1, \dots, x_n, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$(\mathbb{X}_1, \dots, \mathbb{X}_n), \sum X_i, \bar{X}$ are all sufficient.

Example $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d}{\sim} normal(\mu, \sigma^2)$, want sufficient statistic from (μ, σ^2) .

Solution Joint p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n, \mu, \sigma^2$ is

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}}$$

$$\begin{aligned} \sum(x_i-\mu)^2 &= \sum(x_i-\bar{x}+\bar{x}-\mu)^2 = \sum(x_1-\bar{x})^2 + n(\bar{x}-\mu)^2 + 2(\bar{x}-\mu)\sum(x_i-\bar{x}) \\ &\quad (n-1)S^2 + n(\bar{x}-\mu)^2 \end{aligned}$$

$$\Rightarrow f(x_1, \dots, x_n, \mu, \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\frac{(n-1)S^2+n(\bar{x}-\mu)^2}{2\sigma^2}} * 1$$

$$\Rightarrow (\bar{X}, S^2) \text{ is minimal sufficient for } (\mu, \sigma^2)$$

Definition (X, Y) are random variables, conditional mean of \mathbb{Y} given $X = x$, $\mathbb{E}[Y|X]$ is defined as

$$\int y f(y|x) dy$$

Conditional variance of \mathbb{Y} given $X = x$, $var(Y|X)$ is defined as

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])^2 | X] = \int (y - \mathbb{E}[Y|X])^2 f(y|x) dy = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2$$

Note both conditional mean and condition variance of Y given $X = x$ are actually function of x . So here we distinguish the constant value of conditional mean (when x is given) and the its general form

Definition random conditional mean is $\mathbb{E}[Y|X]$ and random variance is $var(Y|X)$.

Theorem

- (a) $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$
- (b) $\text{var}(Y) = \mathbb{E}[\text{var}(Y|X)] + \text{var}(\mathbb{E}[Y|X])$

Proof

- (a) trivial
- (b) $\text{var}(Y|X) = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2$, so $\mathbb{E}[\text{var}(Y|X)] = \mathbb{E}[Y^2] - \mathbb{E}[(\mathbb{E}[Y|X])^2]$.
And $\text{var}(\mathbb{E}[Y|X]) = \mathbb{E}[\mathbb{E}[Y|X]^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2 = \mathbb{E}[Y]^2$. Therefore $\mathbb{E}[\text{var}(Y|X)] + \text{var}(\mathbb{E}[Y|X]) = \mathbb{E}[Y^2|X] - \mathbb{E}[Y]^2$

Theorem (Rao-blackwell) If $\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is an unbiased estimation of $\tau(\theta)$ and $U = U(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is sufficient for θ , then

- (a) $\mathbb{E}[\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)|U]$ is a statistic
- (b) $\mathbb{E}[\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)|U]$ is unbiased for $\tau(\theta)$
- (c) $\text{var}_\theta(\mathbb{E}[\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)|U]) \leq \text{var}_\theta(\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n))$ for $\theta \in \Theta$

Proof

- (a) trivial, since the expectation is integrating on terms unrelated to θ
- (b) $\mathbb{E}_\theta[\mathbb{E}[\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)|U]] = \mathbb{E}[\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)] = \tau(\theta)$
- (c) $\text{var}_\theta(\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)) = \mathbb{E}_\theta[\text{var}_\theta(\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)|U)] + \text{var}_\theta(\mathbb{E}[\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)|U])$
 $\Rightarrow \text{var}_\theta(\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)) \geq \text{var}_\theta(\mathbb{E}[\hat{\tau}(\mathbb{X}_1, \dots, \mathbb{X}_n)|U])$

Rao-blackwell theorem gives us a way to improve on unbiased statistic, but we still need other idea to further understand the limitation of this improvement.

Let U be a statistic. If $h(U) = 0$ or $\mathbb{P}(h(U) = 0) = 1$, then p.d.f of $h(U)$ is

$$f_{h(U)}(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mathbb{E}_\theta[h(U)] = 0$$

That is, if $h(U) = 0$ or $\mathbb{P}(h(U) = 0) = 1$, then for all $\theta \in \Theta$, $\mathbb{E}_\theta[h(U)] = 0$.

However, for any U , if for all $\theta \in \Theta$, we have $\mathbb{E}_\theta[h(U)] = 0$, it is not necessary that $\mathbb{P}(h(U) = 0) = 1$.

Example Let $X_1, X_2 \stackrel{i.i.d}{\sim} normal(\mu, \theta)$. $U = X_1 - X_2$, and let $h(U) = U$. And we have $\mathbb{E}_{\mu, \sigma^2}[U] = 0$, but $\mathbb{P}(h(U) = 0) = 0$

Definition Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta)$, a statistic $U = U(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is a complete statistic if for any function h such that $E_\theta[h(U)] = 0$, then

$$\mathbb{P}(h(U) = 0) = 1 \text{ for } \theta \in \Theta$$

To verify $\mathbb{P}(h(U) = 0) = 1$, it is sufficient to prove that $U(\mathbb{R}) - \{h(U) = 0\}$ has measure zero. Notice if $\{h(U) = 0\} \supset U(\mathbb{R})$, then $U(\mathbb{R}) - \{h(U) = 0\} = \emptyset$, and empty set has measure zero.

Example If $X_1, \dots, X_n \stackrel{i.i.d}{\sim} Bernoulli(p)$, we can show $Y = \sum X_i$ is complete. Suppose a function h satisfies $\mathbb{E}(h(Y) = 0) = 1$ for all p , then

$$\sum_{k=0}^n h(k) \binom{n}{k} p^k (1-p)^{n-k} = (1-p)^n \sum_{k=0}^n h(k) \binom{n}{k} \left(\frac{p}{1-p}\right)^k = 0 \text{ for all } p$$

But a polynomial with non-zero coefficient has at most k roots, so it must be the case that $h(k) \binom{n}{k} = 0$ for $k = 0, 1, \dots, n$. And this implies $h(k) = 0$ for $k = 0, 1, \dots, n$. So the range of $Y = \{0, 1, \dots, n\} \subset \{h(k) = 0\}$.

Example If $X_1, \dots, X_n \stackrel{i.i.d}{\sim} uniform(\theta)$, we can show $Y_n = max(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is sufficient and complete.

1. We have shown Y_n is sufficient by definition, now we do it by factorization. The joint p.d.f of $\mathbb{X}_1, \dots, \mathbb{X}_n$ is

$$f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n \frac{1}{\theta^n} I_{\{Y_n \leq \theta\}} * 1$$

with $K_1 = f(x_1, \dots, x_n, \theta)$, $K_2 = 1$. 2. To show it is complete, for any function h , if for any $\theta \in \Theta$ we have

$$\mathbb{E}[h(Y_n)] = \int_0^\theta n \frac{y^{n-1}}{\theta^n} h(y) dy = \frac{n}{\theta^n} \int_0^\theta y^{n-1} h(y) dy = 0$$

So for all θ , $\{y | h(y) \neq 0\}$ has measure zero in $[0, \theta]$, and so $\mathbb{P}(h(Y) = 0) = 1$

Theorem (Lehmann-Scheffe) Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta)$, suppose that $U = U(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is a complete and sufficient statistic. If $\hat{\tau}(\theta) = t(U)$ is an unbiased estimator for $\tau(\theta)$, then $\hat{\tau}(\theta)$ is the unique function of U that is unbiased for $\tau(\theta)$ and is UMVUE of $\tau(\theta)$. Note we say

two random random variables X, Y are equal if $\mathbb{P}(X = Y) = 1$ So here we say $\hat{\tau}(\theta)$ is the unique function of U that is unbiased for $\tau(\theta)$, it means

$$\forall_{\text{r.v } Y} \text{ } Y \text{ is function of } U \text{ that is unbiased for } \tau(\theta) \rightarrow \mathbb{P}(Y = \hat{\tau}(\theta)) = 1$$

Proof

First I will prove function of U that's unbiased is unique. Suppose $\hat{\tau}(\theta) = \hat{U}$ is also unbiased for $\tau(\theta)$, then

$$\mathbb{E}[\hat{\tau}(\theta) - \hat{\tau}(\theta)] = \mathbb{E}[\hat{\tau}(\theta)] - \mathbb{E}[\hat{\tau}(\theta)] = \tau(\theta) - \tau(\theta) = 0$$

And since U is complete and $\hat{\tau}(\theta) - \hat{\tau}(\theta)$ is a function of U , we know

$$\mathbb{P}(\hat{\tau}(\theta) - \hat{\tau}(\theta) = 0) = \mathbb{P}(\hat{\tau}(\theta) = \hat{\tau}(\theta)) = 1$$

Second let's prove that it is UMVUE. If T is any unbiased statistic for $\tau(\theta)$, then since U is sufficient, Rao-Blackwell theorem, applies, and we have:

(1). $\mathbb{E}[T|U]$ is an unbiased statistics.

(2). $\text{var}(\mathbb{E}[T|U]) \leq \text{var}(\mathbb{E}[T])$

By first step and (1), $\text{var}(\mathbb{E}[T|U])$ is an unique unbiased statistics, and (2) completes proof.

4 Hypothesis Testing

In this class we will introduce confidence interval (C.I), and hypothesis testing. Before this, we have to introduce a real distribution.

Definition (t-distribution) Suppose that we have

$$\begin{cases} \mathbb{Z} \sim \mathcal{N}(0, 1) \\ \mathbb{V} \sim \chi^2(r) \end{cases}$$

, and the two r.v are independent. We say that $\mathbb{T} = \frac{\mathbb{Z}}{\sqrt{\mathbb{V}/r}}$ has a t-distribution with d.f.r.

We also denote $\mathbb{T} \sim t(r)$.

Theorem If $\mathbb{T} \sim t(r)$, its p.d.f is

$$f_{\mathbb{T}}(t) = \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r} \Gamma(\frac{r}{2}) (1 + \frac{t^2}{r})^{\frac{r+1}{2}}}, t \in \mathbb{R}$$

Proof Since $\mathbb{T} \sim t(r)$, $\mathbb{T} = \frac{\mathbb{Z}}{\sqrt{\mathbb{V}/r}}$. The joint p.d.f of \mathbb{Z} and \mathbb{V} is

$$f_{\mathbb{Z}\mathbb{V}}(\zeta, v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta^2}{2}} \frac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} v^{\frac{r}{2}-1} e^{-\frac{v}{2}}, \zeta \in \mathbb{R}, v > 0$$

We consider transformation $\mathbb{T} = \frac{\mathbb{Z}}{\sqrt{\mathbb{V}/r}}$, $\mathbb{U} = \mathbb{V}$, we have

$$A = \left\{ \begin{pmatrix} \zeta \\ v \end{pmatrix} : \zeta \in \mathbb{R}, v > 0 \right\} \xrightarrow{(\mathbb{T} \quad \mathbb{U})} B \left\{ \begin{pmatrix} t \\ u \end{pmatrix} : t \in \mathbb{R}, u > 0 \right\}$$

For $\begin{pmatrix} t \\ u \end{pmatrix} \in B$, we have inverses $v = u$, $\zeta = \frac{\sqrt{ut}}{\sqrt{r}}$. Unique solution indicates 1-1 transformation. Jacobian is

$$J = \begin{vmatrix} \frac{\partial \zeta}{\partial t} & \frac{\partial \zeta}{\partial u} \\ \frac{\partial v}{\partial t} & \frac{\partial v}{\partial u} \end{vmatrix} = \begin{vmatrix} \frac{\sqrt{u}}{\sqrt{r}} & \frac{u^{-\frac{1}{2}} t}{2\sqrt{r}} \\ 0 & 1 \end{vmatrix} = \frac{\sqrt{u}}{\sqrt{r}}$$

The joint p.d.f of \mathbb{T} and \mathbb{U} is

$$f_{\mathbb{T}\mathbb{U}} = f_{\mathbb{Z}\mathbb{V}}\left(\frac{\sqrt{ut}}{\sqrt{r}}, u\right) \frac{\sqrt{u}}{\sqrt{r}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{u}{r} t^2} \frac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} \frac{\sqrt{u}}{\sqrt{r}}$$

So

$$\mathbb{T} = \frac{1}{\sqrt{2\pi r}} \frac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} \int_0^\infty e^{-\frac{1}{2}(1+\frac{t^2}{r})u} u^{\frac{r+1}{2}-1} du$$

$$\begin{aligned}
&= \frac{1}{\sqrt{\pi r} \Gamma(\frac{r}{2}) 2^{\frac{r+1}{2}}} \frac{1}{(\frac{1}{2}(1 + \frac{t^2}{r}))^{\frac{r+1}{2}}} \int_0^\infty \left(\frac{1}{2}(1 + \frac{t^2}{r})\right)^{\frac{r+1}{2}} e^{-\frac{1}{2}(1 + \frac{t^2}{r})u} u^{\frac{r+1}{2}-1} du \\
&= \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r} \Gamma(\frac{r}{2}) (1 + \frac{t^2}{r})^{\frac{r+1}{2}}}
\end{aligned}$$

Note that $f_T(-t) = f_T(t)$, for $t > 0 \Rightarrow f_T(t)$ is symmetric at $t = 0$

4.1 Confidence Interval

The problem in statistics is that we have a random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ from $f(x, \theta)$, where θ is unknown parameter. The point estimate wants to know what is θ by defining estimation $\hat{\theta} = \hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$. Although we can find unknown θ on consistent estimation, or even UMVUE. However, if $\hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n)$ has a continuous distribution, we have $\mathbb{P}_\theta(\hat{\theta}(\mathbb{X}_1, \dots, \mathbb{X}_n) = \theta) = 0$ for $\theta \in \Theta$.

There is no confidence to claim that $\theta = \hat{\theta}$. We then want one approach that assume some confidence for θ .

Definition Suppose we haee a random sample from $f(x, \theta)$. If there are two statistics $T_1 = t_1(\mathbb{X}_1, \dots, \mathbb{X}_2)$ and $T_2 = t_2(\mathbb{X}_1, \dots, \mathbb{X}_2)$ s.t

$$1 - \alpha = \mathbb{P}_\theta(T_1 \leq \theta \leq T_2)$$

We call the random interval (T_1, T_2) a $100(1 - \alpha)\%$ **confidence interval** for θ . We also call $(t_1(\mathbb{X}_1, \dots, \mathbb{X}_2), t_2(\mathbb{X}_1, \dots, \mathbb{X}_2))$ a $100(1 - \alpha)\%$ C.I interval for θ .

Interpretation of C.I:

(a) A $100(1 - \alpha)\%$ C.I (T_1, T_2) means that it covers the unknown θ with probability $1 - \alpha$. If we sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ many times, then the average

$$\frac{1}{m} \sum_{j=1}^m I(t_1^{(j)} \leq \theta \leq t_2^{(j)}) \sim 1 - \alpha$$

Where m is the number of times we do the experiment, and $(t_1^{(j)}, t_2^{(j)})$ is the interval calculated from statistics (T_1, T_2) based on the j th observation of $\mathbb{X}_1, \dots, \mathbb{X}_n$.

(b) If $T_1 = t_1, T_2 = t_2$ is observed (once), then $\mathbb{P}_\theta(t_1 \leq \theta \leq t_2)$ is 0 or 1.

4.1.1 How to construct C.I

Definition A function of random sample and parameter θ , $\mathbb{Q} = q(\mathbb{X}_1, \dots, \mathbb{X}_n, \theta)$ is called a **pivotal quantity** if its distribution is independent of θ .

So a pivotal quantity may depends on θ , while its distribution is independent of θ . In contrast, a statistics is not dependent on θ , while its distribution might.

For each pivotal quantity $\mathbb{Q} = q(\mathbb{X}_1, \dots, \mathbb{X}_n, \theta)$, there are $a, b \in \mathbb{R}$ such that

$$1 - \alpha = \mathbb{P}_\theta(a \leq q(X_1, \dots, X_n, \theta) \leq b)$$

for $\theta \in \Theta$. Not all pivotal quantity can be used to construct C.I

The interesting pivotal quantity is one which 1-1 transformation as:

$$a \leq q(\mathbb{X}_1, \dots, \mathbb{X}_n, \theta) \leq b \text{ if and only if } T_1 \leq \theta \leq T_2$$

for some statistics $T_1 = t_1(\mathbb{X}_1, \dots, \mathbb{X}_n)$ and $T_2 = t_2(\mathbb{X}_1, \dots, \mathbb{X}_n)$.

Hence, $1 - \alpha = \mathbb{P}_\theta(a \leq q(\mathbb{X}_1, \dots, \mathbb{X}_n, \theta) \leq b) = \mathbb{P}_\theta(T_1 \leq \theta \leq T_2)$ and (T_1, T_2) is a $100(1 - \alpha)\%$ C.I for θ .

In the following examples, Let Z be the r.v. with standard normal distribution $N(0, 1)$. Denote ζ_α with $\alpha = \mathbb{P}(Z \geq \zeta_\alpha)$ or $1 - \alpha = \mathbb{P}(Z \leq \zeta_\alpha)$. We need ζ_α since it satisfies $1 - \alpha = \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq Z \leq \zeta_{\frac{\alpha}{2}})$.

4.1.2 C.I for normal mean

Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be random samples from $N(\mu, \sigma^2)$, we want for $100(1 - \alpha)\%$ C.I for μ . We consider two cases

Case 1 : $\sigma = \sigma_0$ is known.

We have $\mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$ where σ^2 is known constant.

$$\bar{X} \sim N(\mu, \frac{\sigma_0^2}{n}) \Rightarrow \bar{X} - \mu \sim N(0, \frac{\sigma^2}{n}) \Rightarrow \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma_0} \sim N(0, 1)$$

Notice here $\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma_0}$ is our pivoted quantity, because the distribution of $\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma_0}$ is easily obtained and it can be 1 - 1 corresponded to the statistics we seek:

$$\Rightarrow 1 - \alpha = \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq Z \leq \zeta_{\frac{\alpha}{2}}) = \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma_0} \leq \zeta_{\frac{\alpha}{2}})$$

$$= \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \leq \bar{X} - \mu \leq \zeta_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}) = \mathbb{P}(\bar{X} - \zeta_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{X} + \zeta_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}})$$

So C.I for μ is $(\bar{X} - \zeta_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \bar{X} + \zeta_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}})$

Case 2 : σ is unknown.

We have a random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ from $N(\mu, \sigma^2)$ where μ, σ^2 are unknown and we want C.I for μ . Recall that sample mean \bar{X} and sample variance S^2 are independent, We have:

$$\begin{cases} \bar{\mathbb{X}} \sim N(\mu, \frac{\sigma^2}{n}) \\ \frac{(n-1)\mathbb{S}^2}{\sigma^2} \sim \chi^2(n-1) \end{cases} \text{ independent} \Rightarrow \begin{cases} \frac{\bar{\mathbb{X}} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \\ \frac{(n-1)\mathbb{S}^2}{\sigma^2} \sim \chi^2(n-1) \end{cases} \text{ independent}$$

The pivotal quantity is $\frac{N(0,1)}{\chi^2(n-1)} = \frac{\bar{\mathbb{X}} - \mu}{\mathbb{S}/\sqrt{n}} \sim t(n-1)$, and let $t_{\frac{\alpha}{2}}$ be the number s.t $\mathbb{P}(t(n-1) \geq t_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$, we have

$$\begin{aligned} 1-\alpha &= \mathbb{P}(-t_{\frac{\alpha}{2}} \leq \mathbb{T} \leq t_{\frac{\alpha}{2}}) = \mathbb{P}(-t_{\frac{\alpha}{2}} \leq \frac{\bar{\mathbb{X}} - \mu}{\mathbb{S}/\sqrt{n}} \leq t_{\frac{\alpha}{2}}) = \mathbb{P}(-t_{\frac{\alpha}{2}} \frac{\mathbb{S}}{\sqrt{n}} \leq \bar{\mathbb{X}} - \mu \leq t_{\frac{\alpha}{2}} \frac{\mathbb{S}}{\sqrt{n}}) \\ &= \mathbb{P}(\bar{\mathbb{X}} - t_{\frac{\alpha}{2}} \frac{\mathbb{S}}{\sqrt{n}} \leq \mu \leq \bar{\mathbb{X}} + t_{\frac{\alpha}{2}} \frac{\mathbb{S}}{\sqrt{n}}) \end{aligned}$$

So $(\bar{\mathbb{X}} - t_{\frac{\alpha}{2}} \frac{\mathbb{S}}{\sqrt{n}}, \bar{\mathbb{X}} + t_{\frac{\alpha}{2}} \frac{\mathbb{S}}{\sqrt{n}})$ is a $100(1-\alpha)\%$ C.I for μ

4.1.3 C.I for normal variance

We have a random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ from $N(\mu, \sigma^2)$ where μ, σ^2 are unknown and we want $100(1-\alpha)\%$ C.I for σ^2 . We know that $\frac{(n-1)\mathbb{S}^2}{\sigma^2} \sim \chi^2(n-1)$, let $\chi_{\frac{\alpha}{2}}^2$ and $\chi_{1-\frac{\alpha}{2}}^2$ satisfies $\frac{\alpha}{2} = \mathbb{P}(\chi^2(n-1) \leq \chi_{\frac{\alpha}{2}}^2)$ and $\frac{\alpha}{2} = \mathbb{P}(\chi^2(n-1) \geq \chi_{1-\frac{\alpha}{2}}^2)$, so

$$\begin{aligned} 1-\alpha &= \mathbb{P}(\chi_{\frac{\alpha}{2}}^2 \leq \chi^2(n-1) \leq \chi_{1-\frac{\alpha}{2}}^2) = \mathbb{P}(\chi_{\frac{\alpha}{2}}^2 \leq \frac{(n-1)\mathbb{S}^2}{\sigma^2} \leq \chi_{1-\frac{\alpha}{2}}^2) \\ &= \mathbb{P}(\frac{(n-1)\mathbb{S}^2}{\chi_{1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1)\mathbb{S}^2}{\chi_{\frac{\alpha}{2}}^2}) \end{aligned}$$

So $\frac{(n-1)\mathbb{S}^2}{\chi_{1-\frac{\alpha}{2}}^2}, \frac{(n-1)\mathbb{S}^2}{\chi_{\frac{\alpha}{2}}^2}$ is the $100(1-\alpha)\%$ C.I for σ^2 .

4.1.4 C.I for difference of means

Case 1: Suppose we have $\begin{cases} \mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2) \\ \mathbb{Y}_1, \dots, \mathbb{Y}_m \stackrel{i.i.d.}{\sim} N(\mu_y, \sigma_y^2) \end{cases}$ independent. And σ_x^2, σ_y^2 are known. We want $100(1-\alpha)\%$ C.I for $\mu_x - \mu_y$. For simplicity, we look at the case when $n = m$ (the case where $n \neq m$ is reasoned exactly the same). Notice

$$\bar{\mathbb{X}} - \bar{\mathbb{Y}} \sim N(\mu_x - \mu_y, \frac{\sigma_x^2 + \sigma_y^2}{n})$$

and

$$\mathbb{Z} = \frac{\bar{\mathbb{X}} - \bar{\mathbb{Y}} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2 + \sigma_y^2}{n}}}$$

The rest is the same as C.I for normal mean.

Case 2: Suppose we have $\begin{cases} \mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma^2) \\ \mathbb{Y}_1, \dots, \mathbb{Y}_m \stackrel{i.i.d.}{\sim} N(\mu_y, \sigma^2) \end{cases}$ independent.

And σ^2 are unknown. We want $100(1 - \alpha)\%$ C.I for $\mu_x - \mu_y$. Now we have

$$\begin{cases} \bar{\mathbb{X}} \sim N(\mu_x, \frac{\sigma^2}{n}) \\ \frac{(n-1)\mathbb{S}^2}{\sigma^2} \sim \chi^2(n-1) \\ \bar{\mathbb{Y}} \sim N(\mu_y, \frac{\sigma^2}{m}) \\ \frac{(m-1)\mathbb{S}^2}{\sigma^2} \sim \chi^2(m-1) \end{cases} \text{ independent} \Rightarrow \begin{cases} \bar{\mathbb{X}} - \bar{\mathbb{Y}} \sim N(\mu_x - \mu_y, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}) \\ \frac{(n-1)\mathbb{S}_x^2 + (m-1)\mathbb{S}_y^2}{\sigma^2} \sim \chi^2(m+n-2) \end{cases} \text{ indept.}$$

So let

$$T = \frac{\frac{\bar{\mathbb{X}} - \bar{\mathbb{Y}} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}}}{\sqrt{\frac{(n-1)\mathbb{S}_x^2 + (m-1)\mathbb{S}_y^2}{\sigma^2(m+n-2)}}} = \frac{\bar{\mathbb{X}} - \bar{\mathbb{Y}} - (\mu_x - \mu_y)}{\sqrt{\frac{(n-1)\mathbb{S}_x^2 + (m-1)\mathbb{S}_y^2}{(m+n-2)}} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(m+n-2)$$

And we have the pivotal quantity we need.

4.2 Best Confident Interval?

In point estimate, we derive UMVUE as the best estimator for θ . Now can we find the best C.I for some parameter θ ? Is it possible to find the best $(T_1, T_2) = (t_1(\mathbb{X}_1, \dots, \mathbb{X}_n), t_2(\mathbb{X}_1, \dots, \mathbb{X}_n))$ in the class of all $100(1 - \alpha)\%$ C.I for θ ?

Suppose we define "best" C.I as (T_1, T_2) such that

$$E_\theta[(T_2 - T_1)] \leq \theta[(T_2^* - T_1^*)] \text{ for all } (T_1^*, T_2^*) \in 100(\alpha - 1)\% \text{ C.I of } \theta$$

We must call such C.I the uniformly minimum expected length $100(\alpha - 1)\%$ C.I for θ if such C.I exists. However, it has been proven by Wilks that such C.I in general does not exist. So we settle with a less demanding idea of good C.I – shortest C.I

$$\text{Statistical Inference} \left\{ \begin{array}{ll} \text{Point Estimate} & \left\{ \begin{array}{l} \text{Unbiased} \\ \text{Consistent} \\ \text{UMVUE} \end{array} \right. \\ \text{Interval Estimate} & \text{–Shortest C.I} \\ \text{Hypothesis Testing} & \end{array} \right.$$

4.2.1 Shortest C.I

4.2.2 Example Shortest C.I

Example

Let $\bar{X}_1, \dots, \bar{X}_n$ be a random sample from $N(0, 1)$ and , we have seen that $(\bar{X} - \frac{\zeta_{\alpha}}{\sqrt{n}}, \bar{X} + \frac{\zeta_{\alpha}}{\sqrt{n}})$ is a $(1 - \alpha)100\%$ C.I for μ , is it the shortest C.I?

Consider the pivotal quantity $Q = \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} \sim N(0, 1)$, let q_1, q_2 satisfy $1 - \alpha = P(q_1 \leq Z \leq q_2)$.

$$\begin{aligned} 1 - \alpha &= \mathbb{P}(q_1 \leq Z \leq q_2) = \mathbb{P}\left(q_1 \leq \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} \leq q_2\right) \\ &= \mathbb{P}\left(\bar{X} - q_2 \frac{1}{\sqrt{n}} \leq \mu \leq \bar{X} - q_1 \frac{1}{\sqrt{n}}\right) \end{aligned}$$

So $(\bar{X} - q_2 \frac{1}{\sqrt{n}}, \bar{X} - q_1 \frac{1}{\sqrt{n}})$ s.t $\int_{q_1}^{q_2} f_Z(t) dt = 1 - \alpha$ is a $100(1 - \alpha)\%$ C.I for μ . The length of this C.I is

$$L = \bar{X} - q_1 \frac{1}{\sqrt{n}} - \bar{X} - q_2 \frac{1}{\sqrt{n}} = (q_2 - q_1) \frac{1}{\sqrt{n}}$$

The shortest C.I is to solve $\min (q_2 - q_1) \frac{1}{\sqrt{n}}$ s.t $\int_{q_1}^{q_2} f_Z(t) dt = 1 - \alpha$. The derivative w.r.t q_1

$$\begin{cases} f_Z(q_2) \frac{\partial q_2}{\partial q_1} - f_Z(q_1) = 0 \Rightarrow \frac{\partial q_2}{\partial q_1} = \frac{f_Z(q_1)}{f_Z(q_2)} \\ 0 = \frac{\partial L}{\partial q_1} = \left(\frac{\partial q_2}{\partial q_1} - 1\right) \frac{1}{\sqrt{n}} \Rightarrow \frac{\partial q_2}{\partial q_1} = 1 \\ \frac{f_Z(q_1)}{f_Z(q_2)} = 1 \Rightarrow f_Z(q_1) = f_Z(q_2) \end{cases}$$

Since $Z = \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} \sim N(0, 1)$ has $f_Z(t)$ symmetric at 0, we have $q_1 = -\zeta_{\frac{\alpha}{2}}, q_2 = \zeta_{\frac{\alpha}{2}}$.

So $(\bar{X} - \frac{\zeta_{\alpha}}{\sqrt{n}}, \bar{X} + \frac{\zeta_{\alpha}}{\sqrt{n}})$ is a shortest C.I.

Example

Let $\bar{X}_1, \dots, \bar{X}_n$ be a random sample from $N(\mu, \sigma^2)$, where μ, σ^2 are unknown. We have seen that $(\bar{X} - t_{\frac{\alpha}{2}} \frac{\$}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}} \frac{\$}{\sqrt{n}})$ is a $100(1 - \alpha)\%$ C.I for μ , is it shortest C.I?

We consider pivotal quantity $T = \frac{\bar{X} - \mu}{\frac{\$}{\sqrt{n}}} \sim t(n - 1)$ Let q_1, q_2 satisfy

$$1 - \alpha = P(q_1 \leq T \leq q_2) = P\left(\bar{X} - q_2 \frac{\$}{\sqrt{N}} \leq \mu \leq \bar{X} - q_1 \frac{\$}{\sqrt{N}}\right)$$

So $(\bar{X} - q_2 \frac{S}{\sqrt{n}}, \bar{X} - q_1 \frac{S}{\sqrt{n}})$ s.t $\int_{q_1}^{q_2} f_Z(t)dt = 1 - \alpha$ is a $100(1 - \alpha)\%$ C.I for μ .
The length of this C.I is

$$L = (q_2 - q_1) \frac{S}{\sqrt{n}}$$

Taking derivative w.r.t q_1 , we have

$$\begin{cases} f_T(q_2) \frac{\partial q_2}{\partial q_1} - f_T(q_1) = 0 \Rightarrow \frac{\partial q_2}{\partial q_1} = \frac{f_T(q_1)}{f_T(q_2)} \\ 0 = \frac{\partial L}{\partial q_1} = \left(\frac{\partial q_2}{\partial q_1} - 1\right) \frac{S}{\sqrt{n}} \Rightarrow \frac{\partial q_2}{\partial q_1} = 1 \end{cases}$$

$$\Rightarrow f_T(q_1) = f_T(q_2), T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

, a symmetric distribution

$$\Rightarrow q_1 = -t_{\frac{\alpha}{2}}, q_2 = t_{\frac{\alpha}{2}}$$

and $(\bar{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}})$ is shortest C.I for μ

4.3 Approximate C.I

So far we have been looking for C.I of continuous random variable. But what about discrete random variable? Finding (t_1, t_2) such that $\mathbb{P}(t_1 \leq \theta \leq t_2)$ isn't always possible, because discrete random variable takes on discontinuous c.d.f. To try to **approximate** C.I for discrete random variable, let's review some probability theorems.

(a) Convergence in distribution (\xrightarrow{d}), we say

$$Y_n \xrightarrow{d} Y$$

if $F_{Y_n}(y) \rightarrow F_Y(y)$ point wise on those y such that F_Y is continuous on y .
Why \xrightarrow{d} ? If $Y_n \xrightarrow{d} Y$, then

$$P(a \leq Y_n \leq b) = F_{Y_n}(b) - F_{Y_n}(a) \rightarrow F_Y(b) - F_Y(a) = P(a \leq Y \leq b)$$

$$\Rightarrow (a \leq Y_n \leq b) \cong P(a \leq Y \leq b) \text{ when } n \text{ is large}$$

(b) Central Limit Theorem (CLT). Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be random sample from a distribution with mean μ and variance $\sigma^2 > 0$, then

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1)$$

Why CLT? For constructing a pivotal quantity $P(-\zeta_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \zeta_{\frac{\alpha}{2}}) \cong P(-\zeta_{\frac{\alpha}{2}} \leq Z \leq \zeta_{\frac{\alpha}{2}}) = 1 - \alpha$

(c) Converge in probability. We say that \mathbb{X}_n converges to \mathbb{X} in **probability**, if for every $\epsilon > 0$, $\mathbb{P}(|\mathbb{X}_n - \mathbb{X}| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$. Denoted $\mathbb{X}_n \xrightarrow{P} \mathbb{X}$

(d) Weak Law of Large Number(WLLN).

If $\mathbb{X}_1, \dots, \mathbb{X}_n$ is a random sample with finite mean μ and variance σ^2 exists, then $\bar{\mathbb{X}} \xrightarrow{P} \mu$.

(e) **Theorem** if $\mathbb{Y}_n \xrightarrow{P} a$, then $g(\mathbb{Y}_n) \xrightarrow{P} g(a)$ for any continuous function g .

(f) (**Slutsky's Theorem**) if $\mathbb{X}_n \xrightarrow{d} a\mathbb{X}$ and $\mathbb{Y}_n \xrightarrow{P} a$, then

$$\begin{cases} \mathbb{X}_n \pm \mathbb{Y}_n & \xrightarrow{d} \mathbb{X} \pm a \\ \mathbb{X}_n \mathbb{Y}_n & \xrightarrow{d} a\mathbb{X} \\ \frac{\mathbb{X}_n}{\mathbb{Y}_n} & \xrightarrow{d} \frac{\mathbb{X}}{a} \end{cases}$$

Note if $\mathbb{Y}_n \xrightarrow{P} a$ then $\mathbb{Y}_n \cong a$ when n is large. If $\mathbb{X}_n \xrightarrow{d} \mathbb{X}$, then $\mathbb{P}(a \leq \mathbb{X}_n \leq b) \cong \mathbb{P}(a \leq \mathbb{X} \leq b)$ when n is large. And convergence in probability is stronger. I.e.

$$\mathbb{X}_n \xrightarrow{P} \mathbb{X} \Rightarrow \mathbb{X}_n \xrightarrow{d} \mathbb{X}$$

Now we are in position to introduce approximate C.I

(1) C.I for μ when distribution is unknown (unknown means μ and σ^2 are unknown). Let

$$\mathbb{X}_1, \dots, \mathbb{X}_n$$

be a random sample from a distribution with mean μ and variance σ^2 , want C.I for μ . First

$$\frac{\bar{\mathbb{X}} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1)$$

by C.L.T. Since the convergence is in distribution, we know we could approximate the distribution by $N(0, 1)$, but σ^2 is unknown, so we need further information about σ^2 . Recall

$$\begin{aligned} \mathbb{S}^2 &= \frac{1}{n-1} \sum \mathbb{X}_i^2 - \frac{n}{n-1} \bar{\mathbb{X}}^2 = \frac{n}{n-1} \frac{1}{n} \sum \mathbb{X}_i^2 - \frac{n}{n-1} \bar{\mathbb{X}}^2 \\ &\xrightarrow{P} \frac{n}{n-1} (\mu^2 + \sigma^2) - \frac{n}{n-1} \mu^2 \quad (\text{by W.L.L.N. and (e)}) \\ &\xrightarrow{P} \sigma^2 \Rightarrow \mathbb{S} \xrightarrow{P} \sigma \end{aligned} \quad (4.4.1)$$

So we combine Z and S

$$\Rightarrow \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \frac{S}{\sigma}} \xrightarrow{d} N(0, 1) \quad (4.4.2)$$

Where

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1) \text{ by C.L.T and } \frac{\sigma}{S} \xrightarrow{P} 1 \text{ by 4.4.1}$$

And by **Slutsky's theorem**, we have 4.4.2. So the analysis now follows similarly as in continuous case:

$$\begin{aligned} 1 - \alpha &= \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq Z \leq \zeta_{\frac{\alpha}{2}}) \cong \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \leq \zeta_{\frac{\alpha}{2}}) \\ &= \mathbb{P}(\bar{X} - \zeta_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + \zeta_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}) \end{aligned}$$

So $(\bar{X} - \zeta_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + \zeta_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}})$ is a $100(1 - \alpha)\%$ C.I for μ

(2) Let $\mathbb{Y} \sim \text{binomial}(n, p)$, want C.I p .

Let $\mathbb{X}_1, \dots, \mathbb{X}_n \xrightarrow{\text{i.i.d.}} \text{Bernoulli}(p)$, and let $\hat{P} = \frac{Y}{n}$.

$$\because \mathbb{Y} \xrightarrow{d} \sum_{i=1}^n \mathbb{X}_i \Rightarrow \begin{cases} \hat{P} = \bar{X} \xrightarrow{P} p \text{ by W.L.L.N} \\ \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0, 1) \text{ by C.L.T} \end{cases}$$

$$\text{Then } \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \sqrt{\frac{p(1-p)}{\hat{P}(1-\hat{P})}} \xrightarrow{d} N(0, 1) \text{ by Slutsky's theorem}$$

So

$$\begin{aligned} 1 - \alpha &= \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq Z \leq \zeta_{\frac{\alpha}{2}}) \cong \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \leq \zeta_{\frac{\alpha}{2}}) \\ &= \mathbb{P}(\hat{P} - \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \zeta_{\frac{\alpha}{2}} \leq p \leq \hat{P} + \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \zeta_{\frac{\alpha}{2}}) \end{aligned}$$

(3) C.I for difference of p' s $\begin{cases} \mathbb{Y}_1 \sim b(n, p_1) \\ \mathbb{Y}_2 \sim b(n, p_2) \end{cases}$ are independent. Want C.I of $p_1 - p_2$.

Let $\hat{P}_i = \frac{Y_i}{n}$, then $\hat{P}_i \xrightarrow{p} p_i$ and $\frac{\hat{P}_i - p_i}{\sqrt{\frac{p_i(1-p_i)}{n}}} \xrightarrow{d} N(0, 1)$

$$\begin{aligned}\Rightarrow \hat{P}_i &\cong N(p_i, \frac{p_i(1-p_i)}{n}) \Rightarrow \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1) + p_2(1-p_2)}{n}}} \cong N(0, 1) \\ &\Rightarrow \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1) + \hat{P}_2(1-\hat{P}_2)}{n}}} \cong N(0, 1)\end{aligned}$$

So $100(1-\alpha)\%$ C.I is $(\hat{P}_1 - \hat{P}_2 - \zeta_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1) + \hat{P}_2(1-\hat{P}_2)}{n}}, \hat{P}_1 - \hat{P}_2 + \zeta_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1(1-\hat{P}_1) + \hat{P}_2(1-\hat{P}_2)}{n}})$

Example Suppose that the exact values of the data x_1, \dots, x_n , are not known, but it is known that 40 of the 50 measurements are larger than t . Find an approximate one-sided lower 95% confidence limit for $\mathbb{P}(x > t)$ based on this information.

Solution Let $\mathbb{X}_1, \dots, \mathbb{X}_{50}$ be 50 i.i.d observations, and let $p = \mathbb{P}(\mathbb{X} > t)$, then $\mathbb{X}_i \stackrel{i.i.d}{\sim} Bernoulli(p)$. So $\hat{p} = \frac{4}{5} = 0.8$, and C.I for p is

$$(0.8 - \sqrt{\frac{0.8 * 0.2}{50}}, 0.8 + \sqrt{\frac{0.8 * 0.2}{50}})$$

4.4 Testing Hypothesis

We have a random sample from $\mathbb{X}_1, \dots, \mathbb{X}_n$. Want hypothesis about θ and conduct hypothesis testing with random sample.

Definition A **statistical hypothesis** is a conjugate about θ . If it is specified with a single value it is called a simple hypothesis. Otherwise it is a composite hypothesis.

Ex: $H : \theta = \theta_0$ is a simple hypothesis. $H : \theta \leq \theta_0, H : \theta = \{\theta_1, \theta_2\}$ are composite hypothesis.

It requires 2 types of hypothesis.

Definition The **null hypothesis** H_0 is a hypothesis that we reject it only if data reveals strongly that it is not true. The **Alternative hypothesis** H_1 , is the hypothesis alternative to the null hypothesis.

To see if some one is guilty, or a new drug is effective, the hypothesis is

- (a) H_0 : Suspect is not guilty, H_1 : Suspect is guilty,
- (b) H_0 : The new drug is not effective, H_1 : The new drug is effective.

We take rejection of H_0 only if there is strong evidence suggesting H_1 is true. So H_0 is a hypothesis we do not reject easily. Or on the other hand, and want to accept H_0 unless evidence suggest strongly otherwise.

Once the hypothesis is determined, we generate a random sample of the new product and define a test for testing hypothesis.

Definition A test is a rule to decide to reject or not to reject a null hypothesis. Usually a test specifies a subset C of the space (range) of random variables $\mathbb{X}_1, \dots, \mathbb{X}_n$, that we reject H_0 if the observations $\mathbb{X}_1, \dots, \mathbb{X}_n \in C$. And not reject H_0 when $\mathbb{X}_1, \dots, \mathbb{X}_n \notin C$. This C is called the critical region.

Two types of error may happen

1. **Type I error:** H_0 is true, but we reject H_0 .
2. **Type II error:** H_1 is true, but we do not reject H_0 .

Definition The **power function** $\pi_C(\theta)$ of a critical region C represents the probability of rejecting H_0 when θ is true.

So let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample, the power function of critical region C is

$$\pi_C(\theta) = \mathbb{P}(\text{rejecting } H_0 | \theta \text{ is true}) = \mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in C | \theta \text{ is true})$$

Notice is $\theta = H_0$ we wish $\pi_C(\theta)$ is small. And if $\theta = H_1$ we wish $\pi_C(\theta)$ is large.

Definition The size of a critical region C is $\sup_{\theta \in H_0} \pi_C(\theta)$. That is, the maximum probability of type I error.

Example If $\mathbb{X}_1, \dots, \mathbb{X}_n$ is a random sample from a distribution with mean and variance are θ and 100, respectively. If the best critical region for testing $H_0 : \theta = 75$ vs. $H_1 : \theta = 78$ is $R = \{x_1, \dots, x_n | c < \bar{x}\}$. Find n and c so that the probability of type I error is 0.05 and probability of type II error is 0.1, approximately.

Solution solve for

$$\begin{aligned} 0.05 &= \mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in R | H_0) = \mathbb{P}\left(\frac{\bar{X} - 75}{100} > \frac{c - 75}{10/\sqrt{n}}\right) \\ &\cong \mathbb{P}(Z > \frac{c - 75}{10/\sqrt{n}}) = \mathbb{P}(Z > \zeta_{0.05}) \end{aligned}$$

So $c = \frac{10}{\sqrt{n}}\zeta_{0.05} + 75$ gives approximately 0.01 probability of type I error. Type II error are found similarly, but instead, we solve for

$$0.1 = \mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \notin R | H_0)$$

5 Most Powerful Test

The rule for choosing a significance level is to fit a significance level α and a test in the class of all tests with size $\leq \alpha$ that minimizes type II error. Usually $\alpha = 0.05, 0.01$ is very small.

Definition Consider a simple hypothesis $H_0 : \theta = \theta_0, H_1 : \theta = \theta_1$. We say that a test with critical region C is the most powerful (MP) test of significance level α if the following holds:

1. $\mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in C | H_0) = \pi_C(\theta_0) = \alpha$ (size of the rejection region is α)
2. $\mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in C | H_1) \geq \mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in A | H_1)$ for every A such that $\mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in A | H_0) = \pi_A(\theta_0) \leq \alpha$ (for every other rejection region of size $\leq \alpha$, type II error is higher)

Recall that if $\mathbb{X}_1, \dots, \mathbb{X}_n$ are random variables, and x_1, \dots, x_n is observed from the random variables. The likelihood function is defined as $L(\theta, x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta)$. The likelihood ratio is defined as $\frac{L(\theta_0, x_1, \dots, x_n)}{L(\theta_1, x_1, \dots, x_n)}$. An intuition for choosing critical region C is if the likelihood ratio is high, then θ_0 is more likely to be true, and if the likelihood ratio is low, then θ_1 is more likely to be true.

Theorem (Neymann-Pearson Theorem) Consider a simple hypothesis $H_0 : \theta = \theta_0, H_1 : \theta = \theta_1$. And let C be the critical region of significance level α . If there exists $k > 0$ such that

1. $\frac{L(\theta_0, x_1, \dots, x_n)}{L(\theta_1, x_1, \dots, x_n)} \leq k$ for all $(x_1, \dots, x_n) \in C$
2. $\frac{L(\theta_0, x_1, \dots, x_n)}{L(\theta_1, x_1, \dots, x_n)} \geq k$ for all $(x_1, \dots, x_n) \notin C$

Then the test with critical region C is the most powerful (MP) test of significance level α .

Proof Let's denote the likelihood function $L(\theta, x_1, \dots, x_n)$ as $L(\theta)$. If we have $\int_C L(\theta_0) d\underline{x} = \alpha$ (critical region C has significance level α), and let A be any critical region such that $\int_A L(\theta_0) d\underline{x} \leq \alpha$, we want to show that $\int_C L(\theta_1) d\underline{x} \geq \int_A L(\theta_1) d\underline{x}$.

$$\begin{aligned} \int_C L(\theta_1) d\underline{x} - \int_A L(\theta_1) d\underline{x} &= \int_{C \cap A^c} L(\theta_1) d\underline{x} - \int_{A \cap C^c} L(\theta_1) d\underline{x} \\ &\geq \int_{C \cap A^c} \frac{L(\theta_0)}{k} d\underline{x} - \int_{A \cap C^c} \frac{L(\theta_0)}{k} d\underline{x} = \frac{1}{k} (\int_C L(\theta_0) d\underline{x} - \int_A L(\theta_0) d\underline{x}) \geq 0 \end{aligned}$$

■

Following from Neyman-Pearson theorem, the MP critical region is $C = \{\underline{x} : \frac{L(\theta_0, \underline{x})}{L(\theta_1, \underline{x})} \leq k\}$ where k satisfies

$$\alpha = \mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in C | \theta = \theta_0)$$

$$= \mathbb{P}\left(\frac{L(\theta_0, \mathbb{X}_1, \dots, \mathbb{X}_n)}{L(\theta_1, \mathbb{X}_1, \dots, \mathbb{X}_n)} \leq k | \theta = \theta_0\right)$$

We can also say the test is:

$$\text{Rejecting } H_0 : \theta = \theta_0 \text{ if } \frac{L(\theta_0, x_1, \dots, x_n)}{L(\theta_1, x_1, \dots, x_n)} \leq k$$

However this is not a practical expression as we usually do not know the distribution of $\frac{L(\theta_0, x_1, \dots, x_n)}{L(\theta_1, x_1, \dots, x_n)}$. So given α , how do we solve for k ? We want exact expression for rejection region C so that we can conduct test. Suppose that, for given $k > 0$, there exists constant c such that

$$\frac{L(\theta_0, \underline{x})}{L(\theta_1, \underline{x})} \leq k \iff u(\underline{x}) \leq c$$

and the distribution u is available. The the MP critical region is

$$C = \{u(\underline{\mathbb{X}}) \leq c\} \text{ such that } \alpha = \mathbb{P}(u(\mathbb{X}_1, \dots, \mathbb{X}_n) \leq c | \theta = \theta_0)$$

to simplify notation, we will use:

$$\{u(\underline{\mathbb{X}}) \leq c\} \text{ as the short hand for } \{\underline{x} | u(\underline{x}) \leq c\}$$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $N(\mu, 1)$, want MP critical region of significance level α for $H_0 : \mu = 0$ vs $H_1 : \mu = 1$.

Solution Likelihood function is

$$L(\mu, x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2}} = (2\pi)^{\frac{n}{2}} e^{-\frac{1}{2}[\sum x_i^2 - 2\mu \sum x_i + n\mu^2]}$$

So the likelihood ratio is

$$\frac{L(\mu = 0, \underline{x})}{L(\mu = 1, \underline{x})} = e^{-\sum x_i + \frac{n}{2}} \leq k \iff -\sum x_i \leq \ln k - \frac{n}{2} \iff \bar{x} \geq -\frac{\ln k}{n} + \frac{1}{2} = c$$

Therefore the MP critical region is $C = \{\bar{\mathbb{X}} \geq c\}$. Now we just need to find c . Using the fact that $\frac{\bar{\mathbb{X}} - \mu}{\sigma/\sqrt{n}} \sim \mathbb{Z}$, c is the number such that

$$\alpha = \mathbb{P}(\text{type I error}) = \mathbb{P}(\bar{X} \geq c | \mu = 0) = \mathbb{P}(\sqrt{n}\bar{X} \geq \sqrt{n}c | \mu = 0)$$

$$= \mathbb{P}(\mathbb{Z} \geq \sqrt{n}c) \Rightarrow c = \frac{\zeta_\alpha}{\sqrt{n}}$$

So let the critical region of for $H_0 : \mu = 0$ vs $H_1 : \mu = 1$ be $C = \{\bar{\mathbb{X}} \geq \frac{\zeta_\alpha}{\sqrt{n}}\}$, by Neyman-Pearson theorem, C is the most powerful critical region of significance level α . ■

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $Poisson(\lambda)$. We consider MP test for $H_0 : \lambda = 10$ vs $H_1 : \lambda = 1$

Solution The likelihood function is

$$L(\lambda, x_1, \dots, x_n) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

So the likelihood ratio is

$$\begin{aligned} & \frac{10^{\sum x_i} e^{-10n}}{e^{-n}} = 10^{\sum x_i} e^{-9n} \leq k \\ \iff & \sum x_i \ln(10) \leq \ln(k) + 9n \Rightarrow \sum x_i \leq \frac{\ln(k) + 9n}{\ln(10)} = c \end{aligned}$$

Solving for c , we want

$$\alpha = \mathbb{P}(\text{type I error}) = \mathbb{P}(\sum x_i \leq c | \lambda = 10)$$

$$(\text{Let } \mathbb{Y} \sim Poisson(10n)) = \mathbb{P}(Y \leq c) = \sum_{i=0}^{\lfloor c \rfloor} \frac{(10n)^i e^{-10n}}{i!}$$

Now c can be solved easily. And $C = \{\sum \mathbb{X}_i \leq c\}$ is the MP rejection region of size α . ■

Examples Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be random sample from the following p.d.f

$$f(x|\mu, \tau) = \sqrt{\frac{\tau}{2\pi x^2}} e^{-\frac{\tau}{2x\mu^2}(x-\mu)^2}$$

where $x, \mu, \tau > 0$, show that there exists the most power test for hypothesis:

1. $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$, where τ is known and $\mu_1 > \mu_0$ is known.
2. $H_0 : \tau = \tau_0, H_1 : \tau = \tau_1$, where μ is known and $\tau_1 > \tau_0$ is known.

Solution Firstly the joint distribution

$$f(\underline{x}, \mu, \tau) = \left(\frac{\tau}{2\pi x_i^2} \right)^{\frac{n}{2}} e^{-\sum_{i=1}^n \frac{\tau}{2x_i \mu^2} (x_i - \mu)^2}$$

(1) We want to find k such that the likelihood ratio is smaller than k :

$$\begin{aligned} \frac{f(\underline{x}, \mu_0, \tau)}{f(\underline{x}, \mu_1, \tau)} &= e^{-\sum_{i=1}^n \frac{\tau}{2x_i \mu_0^2} (x_i - \mu_0)^2 + \sum_{i=1}^n \frac{\tau}{2x_i \mu_1^2} (x_i - \mu_1)^2} = e^{-\frac{\tau}{2} [\sum_{i=1}^n x_i (\frac{1}{\mu_0^2} - \frac{1}{\mu_1^2}) - (\frac{n}{\mu_0} - \frac{n}{\mu_1})]} \leq k \\ &\Rightarrow \sum_{i=1}^n x_i \left(\frac{1}{\mu_1^2} - \frac{1}{\mu_0^2} \right) \leq \ln(k) + \left(\frac{n}{\mu_1} - \frac{n}{\mu_0} \right) = k' \end{aligned}$$

And since $\mu_1 > \mu_0$

$$\Rightarrow \sum_{i=1}^n x_i \geq \frac{k'}{\frac{1}{\mu_0^2} - \frac{1}{\mu_1^2}} = c$$

So by Neyman-Pearson, there exists a MP test.

(2) Again, we want to find k such that

$$\begin{aligned} \frac{f(\underline{x}, \mu, \tau_0)}{f(\underline{x}, \mu, \tau_1)} &= \left(\frac{\tau_0}{\tau_1} \right)^{\frac{n}{2}} e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2x_i \mu^2} (\tau_0 - \tau_1)} \leq k \\ &\Rightarrow \sum_{i=1}^n \frac{(x_i - \mu)^2}{2x_i \mu^2} (\tau_1 - \tau_0) \leq \ln(k) \left(\frac{\tau_0}{\tau_1} \right)^{-\frac{n}{2}} = k' \text{ and since } \tau_1 > \tau_0 \\ &\Rightarrow \sum_{i=1}^n \frac{(x_i - \mu)^2}{2x_i \mu^2} \leq \frac{k'}{\tau_1 - \tau_0} = c \end{aligned}$$

So by Neyman-Pearson, there exists a MP test. ■

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $N(0, \sigma)$. Find most powerful test with significance level α for hypothesis $H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 = \sigma_1^2$ where $\sigma_1^2 > \sigma_0^2$

Solution The likelihood ratio is:

$$\begin{aligned} \frac{L(\underline{x}, \sigma_0)}{L(\underline{x}, \sigma_1)} &= \left(\frac{\sigma_1}{\sigma_0} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right)} \leq k \\ &\Rightarrow -\frac{1}{2} \sum_{i=1}^n x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \leq \ln(k) \left(\frac{\sigma_0}{\sigma_1} \right)^{\frac{n}{2}} = k' \end{aligned}$$

And since $\sigma_1^2 > \sigma_0^2$

$$\Rightarrow \sum_{i=1}^n x_i^2 \geq \frac{k'}{\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}} = c$$

So by Neyman-Pearson, there exists a MP test.

To solve for an explicit test, notice $\mathbb{Z}^2 \sim \chi^2(1)$, so:

$$\begin{aligned}\alpha &= \mathbb{P}(\sum X_i^2 \geq c | H_0) = \mathbb{P}\left(\frac{\sum X_i^2}{\sigma^2} \geq \frac{c}{\sigma^2} | H_0\right) = \mathbb{P}\left(\frac{\sum X_i^2}{\sigma_0^2} \geq \frac{c}{\sigma_0^2}\right) \\ &= \mathbb{P}(\chi^2(n) \geq \frac{c}{\sigma_0^2})\end{aligned}$$

So let χ_α^2 be the number such that $\alpha = \mathbb{P}(\chi^2(n) \geq \chi_\alpha^2)$. Then $c = \chi_\alpha^2 \sigma_0^2$. And $C = \{\sum_{i=1}^n \mathbb{X}_i \geq c\}$ is the most powerful critical region of size α . ■

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample with p.d.f $\lambda kx^{k-1}e^{-\lambda x^k}$, with $x, \lambda, k > 0$, and k is known. Find the MP test of significance level α for $H_1 : \lambda_0^2 = \lambda_1^2$, where $\lambda_1 < \lambda_0$.

Solution The likelihood ratio is:

$$\begin{aligned}\frac{f(\underline{x}, \lambda_0)}{f(\underline{x}, \lambda_1)} &= \frac{\lambda_0}{\lambda_1} e^{-(\lambda_0 - \lambda_1) \sum_{i=1}^n x_i^k} \leq k' \\ \Rightarrow \sum_{i=1}^n x_i^k &\geq k' \frac{\lambda_1}{\lambda_0} (\lambda_0 - \lambda_1) = c\end{aligned}$$

So by Neyman-Pearson, there exists a MP test.

To solve for an explicit test, we need to find distribution of \mathbb{X}^k . Let $\mathbb{Y} = \mathbb{X}^k$, then

$$f_{\mathbb{Y}}(y) = f_{\mathbb{X}}(y^{\frac{1}{k}}) \frac{1}{k} y^{\frac{1}{k}-1} = \lambda e^{-\lambda y} \Rightarrow Y \sim \exp(\lambda)$$

So $2\lambda \sum_{i=1}^n \mathbb{X}_i^k \sim \text{Gamma}(\alpha = n, \lambda = \frac{1}{2}) = \chi^2(2n)$

So by Neyman-Pearson, there exists a MP test.

$$\alpha = \mathbb{P}\left(\sum_{i=1}^n \mathbb{X}_i^k \geq c | H_0\right) = \mathbb{P}(2\lambda \sum_{i=1}^n \mathbb{X}_i^k \geq 2\lambda c | H_0) = \mathbb{P}(2\lambda_0 \sum_{i=1}^n \mathbb{X}_i^k \geq 2\lambda_0 c)$$

Let χ_α^2 be the number such that $\alpha = \mathbb{P}(\chi^2(2n) \geq \chi_\alpha^2)$, then $c = \frac{\chi_\alpha^2}{2\lambda_0}$. ■

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $\text{Exp}(1, \theta)$ with p.d.f $f(x) = e^{-(x-\theta)}, x > \theta$. Find the MP test of significance level α for $H_0 : \theta = \theta_0, H_1 : \theta = \theta_1$.

Solution Note $f(x) = e^{-(x-\theta)} I_{\{\theta, \infty\}}(x)$. Let $\mathbb{X}_{(1)}, \dots, \mathbb{X}_{(n)}$ be the ordered statistics of $\mathbb{X}_1, \dots, \mathbb{X}_n$. The joint distribution is

$$f(\underline{x}, \theta) = e^{-\sum_{i=1}^n (x_i - \theta)} \prod_{i=1}^n I_{\{\theta, \infty\}}(x_i)$$

And from the likelihood point of view,

$$f(\underline{x}, \theta) = e^{-\sum_{i=1}^n (x_i - \theta)} \prod_{i=1}^n I_{\{-\infty, x_{(1)}\}}(\theta)$$

So the likelihood function is

$$\frac{f(\underline{x}, \theta_0)}{f(\underline{x}, \theta_1)} = e^{n(\theta_0 - \theta_1)} \frac{I_{\{-\infty, x_{(1)}\}}(\theta_0)}{I_{\{-\infty, x_{(1)}\}}(\theta_1)}$$

Now since $\theta_1 > \theta_0$, if $I_{\{-\infty, x_{(1)}\}}(\theta_0) = 0$, then neither hypothesis is possible given that we observed the data that can not be generated by $\theta = \theta_0$ or $\theta = \theta_1$. If $I_{\{-\infty, x_{(1)}\}}(\theta_0) \neq 0$ and $I_{\{-\infty, x_{(1)}\}}(\theta_1) = 0$ ($\theta_0 \leq x_{(1)} < \theta_1$), then only H_0 is possible. And if $x_{(1)} \geq \theta_1$, then we want k such that $e^{n(\theta_0 - \theta_1)} < k$. As long as $x_{(1)} \geq \theta_1$, k can be anything larger than $e^{n(\theta_0 - \theta_1)} < k$. So by Neyman-Pearson theorem, there exist a most powerful test. To find the explicit test, let

$$\alpha = \mathbb{P}(\mathbb{X}_{(1)} \geq c | H_0) \text{ where } c \geq \theta_1$$

The distribution of $\mathbb{X}_{(1)}$ is given by ordered statistics

$$f_{\mathbb{X}_{(1)}}(x) = n(1 - F_{\mathbb{X}}(x))^{n-1} f_{\mathbb{X}}(x) = n e^{-n(x-\theta)} \sim Exp(n, \theta)$$

So

$$\begin{aligned} \alpha &= \mathbb{P}(\mathbb{X}_{(1)} \geq c | H_0) = \int_c^\infty n e^{-n(x-\theta_0)} \sim Exp(n, \theta_0) dx = e^{-n(c-\theta_0)} \\ &\Rightarrow c = -\frac{\ln(\alpha)}{n} - \theta_0 \end{aligned}$$

Therefore for $c \geq \theta_1$ (α sufficiently small), the set $C = \{\underline{x} | x_{(1)} \geq c\}$ is the most powerful critical region with significance α . ■

6 UMP test

Definition A test with critical region C is called a uniformly most powerful (UMP) test of significance level (size) α for hypothesis

$$H_0 : \theta = \theta_0, H_1 : \theta = \Theta_1$$

If for $\theta_1 \in \Theta_1$, C is the most powerful critical region of significance level α for $H_0 : \theta = \theta_0, H_1 : \theta = \theta_1$

By Neyman-Pearson theorem, MP critical region for $H_0 : \theta = \theta_0, H_1 : \theta = \theta_1$ is

$$C = \{x | \frac{L(\theta_0, x)}{L(\theta_1, x)} \leq k\} = \{u(\bar{x}) \leq c\}$$

Suppose C is independent of $\theta_1 \in \Theta_1$, then C is UMP critical region.

For example, from previous analysis we see if have $\bar{X}_1, \dots, \bar{X}_n$ be a random sample from $N(\mu, \sigma^2)$ where σ^2 is known. MP critical region for $H_0 : \mu = 0, H_1 : \mu = \mu_1 > 0$ is $C = \{\sum \bar{X}_i \geq \sqrt{n}\zeta_\alpha\}$. This is independent of $\mu = \mu_1$ when $\mu_1 > 0$, so C is UMP critical region for $H_0 : \mu = 0, H_1 : \mu > 0$.

Usually UMP test exists for one sided hypothesis:

1. $H_0 : \mu = 0$ vs $H_1 : \mu > 0$
2. $H_0 : \mu = 0$ vs $H_1 : \mu < 0$

And UMP test not exist for two sided hypothesis $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$

Example Let $\bar{X}_1, \dots, \bar{X}_n$ be a random sample from $N(\mu, 1)$, we consider UMP test for $H_0 : \mu = 0$ vs $H_1 : \mu > 0$.

Solution Let $\mu_1 > 0$, we consider simple hypothesis $H_0 : \mu = 0$ vs $H_1 : \mu = \mu_1$. For MP test, the likelihood ratio

$$\begin{aligned} \frac{L(0, x_1, \dots, x_n)}{L(\mu_1, x_1, \dots, x_n)} &\leq k = e^{-\mu_1 \sum x_i + \frac{n}{2} \mu_1^2} \leq k \\ \iff \sum x_i &\geq \frac{\ln k - \frac{n}{2} \mu_1^2}{-\mu_1} = c \end{aligned}$$

And let $R = \{\sum \bar{X}_i > c\}$ be the rejection region

$$\begin{aligned} \Rightarrow \alpha &= \mathbb{P}(\sum \bar{X}_i > c | \mu = 0) = \mathbb{P}(\bar{X} > \frac{c}{n} | \mu = 0) = \mathbb{P}\left(\frac{\bar{X} - 0}{\frac{1}{\sqrt{n}}} \geq \frac{c}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(Z \geq \frac{c}{\sqrt{n}}\right) \Rightarrow c = \sqrt{n}\zeta_\alpha \end{aligned}$$

Notice the above c is independent of μ_1 is because μ_1 is always greater than 0. $\mu_1 > 0$ is the reason why $\sum x_i \geq \frac{\ln k - \frac{n}{2}\mu_1^2}{-\mu_1} = c$ is consistently true. If $\mu_1 \neq 0$ and μ_1 could be positive or negative, then

$$\begin{cases} \sum x_i \geq \frac{\ln k - \frac{n}{2}\mu_1^2}{-\mu_1} = c & \text{for } \mu_1 > 0 \\ \sum x_i \leq \frac{\ln k - \frac{n}{2}\mu_1^2}{-\mu_1} = c & \text{for } \mu_1 < 0 \end{cases} \quad \text{is dependent on } \mu_1$$

■

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $N(0, \sigma^2)$, want UMP critical region for $H_0 : \sigma^2 = 1$ vs $H_1 : \sigma^2 < 1$.

Solution Let $\theta < 1$, we consider first MP critical region for simple hypothesis $H_0 : \sigma^2 = 1$ vs $H_1 : \sigma^2 = \sigma_1^2$, where $\sigma_1^2 < 1$. By Neyman-Pearson theorem, we want the likelihood ratio

$$\begin{aligned} \frac{L(\sigma^2 = 1, x_1, \dots, x_n)}{L(\sigma^2 = \sigma_1^2, x_1, \dots, x_n)} &= \sigma_1^{\frac{n}{2}} e^{-\frac{\sum x_i^2}{2}} = \sigma_1^{\frac{n}{2}} e^{-\frac{\sum x_i^2}{2}(1 - \frac{1}{\sigma_1^2})} < k \\ \iff \sum_i x_i^2 &\leq (\ln(k) - \frac{n}{2} \ln(\sigma_1)) \frac{1 - \frac{1}{\sigma_1^2}}{2} = c \\ \Rightarrow \alpha &= \mathbb{P}\left(\sum_i \mathbb{X}_i^2 < c | H_0\right) = \mathbb{P}(\chi^2(n) < c) \Rightarrow c = \chi_{\alpha}^2 \end{aligned}$$

This critical region is independent of $H_1 = \sigma^2 = \sigma_1^2$, hence $\{\sum_i \mathbb{X}_i^2 < c | H_0\}$ is UMP critical region of significance level α for $H_0 : \sigma^2 = 1$ vs $H_1 : \sigma^2 < 1$.

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $Poisson(\lambda)$. We consider $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$.

Solution Consider simple hypothesis $H_0 : \lambda = 1$ vs $H_1 : \lambda = \lambda_1 \neq 1$. The likelihood ratio is

$$\frac{L(0, x_1, \dots, x_n)}{L(\mu_1, x_1, \dots, x_n)} = e^{-n(1-\lambda_1)\lambda_1 \sum x_i \leq k}$$

$$\Rightarrow \begin{cases} \sum x_i \geq \frac{\ln(\frac{k}{e^{n(1-\lambda_1)}})}{-\ln \lambda_1} & \text{if } \lambda_1 > 1 \\ \sum x_i \leq \frac{\ln(\frac{k}{e^{n(1-\lambda_1)}})}{-\ln \lambda_1} & \text{if } \lambda_1 < 1 \end{cases}$$

Therefore the rejection region depends on λ_1 , and UMP does not exist. ■

The following types of hypothesis will all be considered:

1. $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$

MP test always exist by Neyman-Pearson theorem.

2. $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$

UMP test exists for some distribution that can be derived from Neyman-Pearson theorem.

3. $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

UMP test usually do not exist.

4. $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$

Consider the last type of testing, we need to define the UMP test for it first.

Suppose that we consider hypothesis $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$. For critical region C with power function $\pi_C(\theta) = \mathbb{P}(\underline{\mathbb{X}} \in C | \theta)$. Recall that the size of C is defined as $\sup_{\theta \in \Theta_0} \pi_C(\theta)$.

Definition Consider composite hypothesis $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$. A test with critical region C is called a UMP test of significance level α , if it satisfies

1. $\sup_{\theta \in \Theta_0} \pi_C(\theta) = \alpha$ (size of C is α)

2. For any critical region A such that $\sup_{\theta \in \Theta_0} \pi_A(\theta) \leq \alpha$, we have

$$\pi_C(\theta) \geq \pi_A(\theta) \text{ for all } \theta \in \Theta_1$$

Definition A family of densities $\{f(\theta, x_1, \dots, x_n) | \theta \in \Theta\}$ is said to have a monotone likelihood ratio (MLR) if, for $\theta' < \theta''$, there exists a statistics $T = t(\mathbb{X}_1, \dots, \mathbb{X}_n)$ such that the likelihood ratio

$$\frac{L(\theta', x_1, \dots, x_n)}{L(\theta'', x_1, \dots, x_n)} = h(T)$$

is either nondecreasing or nonincreasing w.r.t T.

There are two types of hypothesis, $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, and $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$. There are also two types of monotonicity. Hence we have to construct UMP critical region.

Note: Suppose that we consider one-sided hypothesis $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ and there is MLR, then the UMP critical region C has monotone power function. This indicates that the size of C is

$$\text{size} = \sup_{\theta \leq \theta_0} \pi_C(\theta) = \pi_C(\theta_0) = \mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in C | \theta_0)$$

Theorem Suppose the family of densities has a MLR in $Y = t(\mathbb{X}_1, \dots, \mathbb{X}_n)$.

(a) Consider hypothesis $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. For $\theta' < \theta''$, if $\text{LR} = \frac{L(\theta')}{L(\theta'')} = h(T)$ is non-decreasing in T, then the UMP critical region of significance level α is $C = \{T \leq t_0\}$ s.t

$$\alpha = \sup_{\theta \leq \theta_0} \mathbb{P}(T \leq t_0 | \theta) = \mathbb{P}(T \leq t_0 | \theta_0)$$

If $\text{LR} = \frac{L(\theta')}{L(\theta'')} = h(T)$ is non-increasing in T, the the UMP critical region of significance level α is $C = \{T \geq t_0\}$ s.t

$$\alpha = \sup_{\theta \leq \theta_0} \mathbb{P}(T \geq t_0 | \theta) = \mathbb{P}(T \geq t_0 | \theta_0)$$

(b) Consider hypothesis $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$. For $\theta' < \theta''$, if $\text{LR} = \frac{L(\theta')}{L(\theta'')} = h(T)$ is non-decreasing in T, then the UMP critical region of significance level α is $C = \{T \geq t_0\}$ s.t

$$\alpha = \sup_{\theta \geq \theta_0} \mathbb{P}(T \geq t_0 | \theta) = \mathbb{P}(T \geq t_0 | \theta_0)$$

If $\text{LR} = \frac{L(\theta')}{L(\theta'')} = h(T)$ is non-increasing in T, the the UMP critical region of significance level α is $C = \{T \leq t_0\}$ s.t

$$\alpha = \sup_{\theta \geq \theta_0} \mathbb{P}(T \leq t_0 | \theta) = \mathbb{P}(T \leq t_0 | \theta_0)$$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $U(0, \theta)$. We consider UMP test for $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$.

Solution The p.d.f of \mathbb{X} is $f(x, \theta) = \frac{1}{\theta} I(0 < x < \theta)$ So the likelihood function is

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} I(0 < x_i < \theta) = \frac{1}{\theta^n} \prod_{i=1}^n I(0 < x_i < \theta) = \frac{1}{\theta^n} I(0 < y_n < \theta)$$

where $y_n = \max\{x_1, \dots, x_n\}$. For $\theta' < \theta''$, the likelihood ratio is

$$\frac{L(\theta')}{L(\theta'')} = \frac{\theta'^n I(0 < y_n < \theta')}{\theta''^n I(0 < y_n < \theta'')} = \begin{cases} \left(\frac{\theta'}{\theta''}\right)^n & \text{if } 0 < y_n < \theta' \\ 0 & \text{if } \theta' \leq y_n < \theta'' \end{cases}$$

This is MLR in $\mathbb{Y}_n = \max\{\mathbb{X}_1, \dots, \mathbb{X}_n\}$. The UMP critical region is $C = \{\mathbb{Y}_n \geq c\}$. The pdf of \mathbb{Y}_n is

$$f_{\mathbb{Y}_n}(y) = n(F(y))^{n-1} f(y) = n\left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} = n \frac{y^{n-1}}{\theta^n}, 0 < y < \theta$$

$$\alpha = P(\mathbb{Y}_n \geq c | \theta = \theta_0) = \int_c^{\theta_0} n \frac{y^{n-1}}{\theta_0^n} dy = 1 - \frac{c^n}{\theta_0^n}$$

$$\Rightarrow c = \theta_0 (1 - \alpha)^{\frac{1}{n}}$$

The UMP critical region of significance level α is $C = \{\mathbb{Y}_n \geq \theta_0 (1 - \alpha)^{\frac{1}{n}}\}$.

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $N(\mu, 1)$, want UMP test for $H_0 : \mu \geq 0$ vs $H_1 : \mu < 0$.

Solution For $\mu' < \mu''$, the likelihood ratio

$$\frac{L(\mu')}{L(\mu'')} = \frac{e^{-\frac{1}{2} \sum (x_i - \mu')^2}}{e^{-\frac{1}{2} \sum (x_i - \mu'')^2}} = e^{-\frac{1}{2} [2(\mu'' - \mu') \sum x_i + n(\mu'^2 - \mu''^2)]}$$

is non-increasing in $\sum x_i$. So the UMP critical region is $C = \{\bar{\mathbb{X}} \leq c\}$

$$\alpha = \mathbb{P}(\bar{\mathbb{X}} \leq c | \mu = 0) = \mathbb{P}(\sqrt{n}\bar{\mathbb{X}} < \sqrt{n}c) \Rightarrow c = \frac{-\zeta_\alpha}{\sqrt{n}}$$

And the UMP critical region of significance level α is $C = \{\mathbb{Y}_n \leq \frac{-\zeta_\alpha}{\sqrt{n}}\}$

7 Likelihood Ratio Test

What can we do if UMP test not exist for composite hypothesis. We consider likelihood ratio test.

Likelihood Ratio Test Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $f(x, \theta), \theta \in \Theta$, the likelihood function is $L(\theta) = \prod_{i=1}^n f(x_i, \theta)$. We consider the hypothesis $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta$. We also consider two maximum likelihood:

$$\max_{\theta \in \Theta_0} L(\theta) \quad \text{and} \quad \max_{\theta \in \Theta} L(\theta) = L(\hat{\theta}_{mle})$$

obviously we have

$$\max_{\theta \in \Theta_0} L(\theta) \leq \max_{\theta \in \Theta} L(\theta)$$

If these two maximum likelihoods are close, we believe H_0 is true.

Definition The generalized likelihood ratio is

$$\lambda = \lambda(x_1, \dots, x_n) = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} = \frac{\max_{\theta \in \Theta_0} L(\theta)}{L(\hat{\theta}_{mle})}$$

The likelihood ration test for $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta$ is:

Rejecting H_0 if $\lambda(x_1, \dots, x_n) \leq \lambda_0$ where λ_0 satisfies $\max_{\theta \in \Theta_0} \mathbb{P}(\lambda(x_1, \dots, x_n) \leq \lambda_0 | \theta) = \alpha$

Suppose that we have a statistics $\mathbb{T} = t(\mathbb{X}_1, \dots, \mathbb{X}_n)$ such that $\lambda(x_1, \dots, x_n) \leq \lambda_0$ if and only if $t(x_1, \dots, x_n) \leq c$. Then the likelihood ration test has critical region $C = \{t(\mathbb{X}_1, \dots, \mathbb{X}_n) \leq c\}$ such that $\alpha = \max_{\theta \in \Theta_0} \mathbb{P}(\lambda(x_1, \dots, x_n) \leq \lambda_0 | \theta)$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from exponential distribution with p.d.f $f(x, \theta) = \theta e^{-\theta x}, x > 0$. We consider LRT for $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$.

Solution The likelihood function is

$$L(\theta) = \theta^n e^{-\theta \sum x_i}, \theta > 0$$

$$\ln L(\theta) = n \ln \theta - \theta \sum x_i \Rightarrow \frac{\partial}{\partial \theta} \ln L(\theta) = \frac{n}{\theta} - \sum x_i = 0$$

$$\Rightarrow \text{mle } \hat{\theta} = \frac{1}{\bar{x}} \Rightarrow \max_{\theta > 0} L(\theta) = L\left(\frac{1}{\bar{x}}\right) = \frac{1}{\bar{x}^n} e^{-n}$$

$$\max_{\theta < \theta_0} L(\theta) = \begin{cases} L(\theta_0) = \theta_0^n e^{-\theta_0 \sum x_i} & \text{if } \theta_0 < \frac{1}{\bar{x}} \\ L\left(\frac{1}{\bar{x}}\right) = \frac{1}{\bar{x}^n} e^{-n} & \text{if } \theta_0 \geq \frac{1}{\bar{x}} \end{cases}$$

LR is

$$\lambda(x_1, \dots, x_n) = \frac{\max_{\theta \leq \theta_0} L(\theta)}{\max_{\theta > 0} L(\theta)} = \begin{cases} (\theta_0 \bar{x})^n e^{-n(\theta_0 \bar{x} - 1)} & \text{if } \bar{x} < \frac{1}{\theta_0} \\ 1 & \text{if } \bar{x} > \frac{1}{\theta_0} \end{cases}$$

For $\bar{x} < \frac{1}{\theta_0}$, $\ln \lambda(x_1, \dots, x_n) = n \ln(\theta_0 \bar{x}) - n(\theta_0 \bar{x} - 1)$

$$\Rightarrow \frac{\partial}{\partial \bar{x}} \ln \lambda = \frac{n}{\bar{x}} - n\theta_0 = n\left(\frac{1}{\bar{x}} - \theta_0\right) > 0$$

Since λ is increasing in \bar{x} , LR critical region is $C = \{\bar{\mathbb{X}} \leq c\}$. Note $X \sim \exp(\theta) = \Gamma(1, \theta)$, so

$$\sum \mathbb{X}_i \sim \Gamma(n, \theta) \Rightarrow 2\theta \sum \mathbb{X}_i \sim \Gamma(n, \frac{1}{2}) = \chi^2(2n)$$

And we want c such that

$$\alpha = \mathbb{P}(\bar{\mathbb{X}} < c | \theta = \theta_0) = \mathbb{P}\left(\sum \mathbb{X}_i < nc | \theta = \theta_0\right) = \mathbb{P}(\theta_0 2 \sum \mathbb{X}_i \leq 2\theta_0 nc) = \mathbb{P}(\chi^2(2n) \leq 2\theta_0 nc)$$

$$\Rightarrow c = \frac{\chi_{\alpha}^2}{2n\theta_0}$$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $N(\mu, 1)$. Want LRT for $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$.

Solution This is a hypothesis without UMP test. The likelihood function id

$$L(\mu) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \sum (x_i - \mu)^2}$$

To find maximum likelihood, let

$$\frac{\partial}{\partial \mu} \ln L(\mu) = -\frac{\partial}{\partial \mu} \frac{1}{2} \sum (x_i - \mu)^2 = 0 \iff \sum (x_i - \mu) = 0$$

So mle $\hat{\mu} = \bar{x}$. $\max_{\mu \neq 0} L(\mu) = L(\hat{\mu}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \sum (x_i - \hat{\mu})^2}$, generalized LR is

$$\frac{\max_{\mu=0} L(\mu)}{\max_{\mu \neq 0} L(\mu)} = e^{-\frac{n}{2} \bar{x}^2} \leq \lambda \iff |\bar{x}| \geq c$$

To find c

$$\alpha = \mathbb{P}(|\bar{\mathbb{X}}| \geq c | \mu = 0) = 1 - \mathbb{P}(-x \leq |\bar{\mathbb{X}}| \leq c | \mu = 0) = 1 - \mathbb{P}(-\sqrt{nc} \leq \mathbb{Z} \leq \sqrt{nc})$$

$$\Rightarrow c = \frac{\zeta_{\frac{\alpha}{2}}}{\sqrt{n}}$$

Example Let $\mathbb{X}_1, \dots, \mathbb{X}_n$ be a random sample from $N(\mu, \sigma^2)$ where μ and σ^2 are unknown. In this case, any one sided hypothesis on μ and σ^2 has no UMP test. We consider LRT for $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$.

Solution It is equivalent to test $H_0 : \mu = 0, \sigma^2 > 0$ and $H_1 : \mu \neq 0, \sigma^2 > 0$. The likelihood function is

$$L(\mu) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum(x_i - \mu)^2}$$

And mle for μ, σ^2 is

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n} \sum (\mathbb{X}_i - \bar{\mathbb{X}})^2 \\ \hat{\mu} &= \bar{\mathbb{X}}\end{aligned}$$

and

$$\max_{\mu=0, \sigma>0} L(\theta) = (2\pi)^{-\frac{n}{2}} \left(\frac{1}{n} \sum x_i^2 \right)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

$$\max_{\mu \neq 0, \sigma>0} L(\theta) = (2\pi)^{-\frac{n}{2}} \left(\frac{1}{n} \sum (x_i - \bar{x})^2 \right)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

The generalized LR is

$$\begin{aligned}\frac{\max_{\mu=0, \sigma^2>0} L(\mu, \sigma)}{\max_{\mu \in \mathbb{R}, \sigma^2>0} L(\mu, \sigma)} &= \left(\frac{\sum (x_i - \bar{x})^2}{\sum x_i^2} \right)^{\frac{n}{2}} \leq \lambda_0 \\ \iff 1 + \frac{n\bar{x}^2}{\sum (x_i - \bar{x})^2} &\geq \frac{1}{\lambda_0^{\frac{n}{2}}} \iff \frac{n\bar{x}^2}{(n-1)\mathbb{S}^2} &\geq \frac{1}{\lambda_0^{\frac{n}{2}}} - 1 \\ \iff \frac{\bar{x}^2}{\mathbb{S}^2/n} &\geq (n-1) \left(\frac{1}{\lambda_0^{\frac{n}{2}}} - 1 \right) = c\end{aligned}$$

To test H_0 , we use $\frac{\bar{\mathbb{X}} - \mu}{S/\sqrt{n}} \sim t(n-1)$

$$\left| \frac{\bar{x}}{S/\sqrt{n}} \right| \geq \sqrt{c} = c_0$$

$$\alpha = \mathbb{P} \left(\left| \frac{\bar{x}}{S/\sqrt{n}} \right| \geq \sqrt{c} \mid \mu = 0 \right) 1 - \mathbb{P}(-c_0 \leq T \leq c_0) \Rightarrow c_0 = t_{\frac{\alpha}{2}}$$

For the problem of testing hypothesis of two means as $H_0 : \mu_x = \mu_y$, we assume that variance of two distribution are equal. How do we know that this is true?

F-distribution A r.v is said to have a F-distribution with degree of freedoms (v_1, v_2) if there are independent $\chi^2(v_1)$ and $\chi^2(v_2)$ such that $F = \frac{\chi^2(v_1)/v_1}{\chi^2(v_2)/v_2}$

We denote $F \sim f(v_1, v_2)$

We denote $f_\alpha(v_1, v_2)$ with $\mathbb{P}(F \geq f_\alpha(v_1, v_2)) = \alpha$. In table, we can find $f_\alpha(v_1, v_2)$ for $\alpha = 0.95, 0.975, 0.995$ only. How can we find $f_\alpha(v_1, v_2)$ for $\alpha = 0.05, 0.025, 0.01$?

Theorem If $F \sim f(v_1, v_2)$ then $\frac{1}{F} \sim f(v_2, v_1)$

Theorem $f_\alpha(v_1, v_2) = \frac{1}{f_{1-\alpha}(v_2, v_1)}$

Proof

$$\begin{aligned}\alpha &= \mathbb{P}(F \leq f_\alpha(v_1, v_2)) = \mathbb{P}\left(\frac{1}{F} \geq \frac{1}{f_\alpha(v_1, v_2)}\right) = 1 - \mathbb{P}\left(\frac{1}{F} \leq \frac{1}{f_\alpha(v_1, v_2)}\right) \\ &= 1 - \mathbb{P}(F(v_2, v_1) \leq \frac{1}{f_\alpha(v_1, v_2)}) \Rightarrow f_\alpha(v_1, v_2) = \frac{1}{f_{1-\alpha}(v_2, v_1)}\end{aligned}$$

We now have a mean to find $f_\alpha(v_1, v_2)$ for $\alpha = 0.05, 0.025, 0.01$ by the inverse of $f_\alpha(v_2, v_1)$ for $\alpha = 0.95, 0.975, 0.995$

Example Inferences for ratio of variances.

Suppose that we have

$$\left\{ \begin{array}{l} \mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2) \\ \mathbb{Y}_1, \dots, \mathbb{Y}_m \stackrel{i.i.d.}{\sim} N(\mu_y, \sigma_y^2) \end{array} \right\} \text{independent}$$

Want C.I for $\frac{\sigma_x^2}{\sigma_y^2}$

Recall that we have

$$\begin{aligned}&\left\{ \begin{array}{l} \frac{(n-1)\mathbb{S}_x^2}{\sigma_x^2} \sim \chi^2(n-1) \\ \frac{(m-1)\mathbb{S}_y^2}{\sigma_y^2} \sim \chi^2(m-1) \end{array} \right\} \text{independent} \\ &\Rightarrow F = \frac{\frac{(n-1)\mathbb{S}_x^2}{\sigma_x^2}}{\frac{(m-1)\mathbb{S}_y^2}{\sigma_y^2}} = \frac{\mathbb{S}_x^2}{\mathbb{S}_y^2} \frac{\sigma_y^2}{\sigma_x^2} \sim f(n-1, m-1)\end{aligned}$$

Define $f_{\frac{\alpha}{2}}$ and $f_{1-\frac{\alpha}{2}}$ to be numbers such that:

$$\frac{\alpha}{2} = \mathbb{P}\left(\frac{\mathbb{S}_x^2}{\mathbb{S}_y^2} \frac{\sigma_y^2}{\sigma_x^2} \leq f_{\frac{\alpha}{2}}(n-1, m-1)\right) = \mathbb{P}\left(\frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \frac{\sigma_x^2}{\sigma_y^2} \geq f_{1-\frac{\alpha}{2}}(m-1, n-1)\right)$$

$$\frac{\alpha}{2} = \mathbb{P}\left(\frac{\mathbb{S}_x^2}{\mathbb{S}_y^2} \frac{\sigma_y^2}{\sigma_x^2} \geq f_{1-\frac{\alpha}{2}}(n-1, m-1)\right) = \mathbb{P}\left(\frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \frac{\sigma_x^2}{\sigma_y^2} \leq \frac{1}{f_{1-\frac{\alpha}{2}}(n-1, m-1)}\right)$$

$$\Rightarrow 1 - \alpha = \mathbb{P}\left(\frac{1}{f_{1-\frac{\alpha}{2}}(n-1, m-1)} \leq \frac{\sigma_x^2}{\sigma_y^2} \leq f_{1-\frac{\alpha}{2}}(m-1, n-1)\right)$$

Therefore the C.I for $\frac{\sigma_x^2}{\sigma_y^2}$ with significance level α is

$$\left(\frac{1}{f_{1-\frac{\alpha}{2}}(n-1, m-1)}, f_{1-\frac{\alpha}{2}}(m-1, n-1)\right) \blacksquare$$

When UMP not exist, we may not want to derive LR test because it is too complicated in development. For random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$, if we have a set $A(\mathbb{X}_1, \dots, \mathbb{X}_n)$ such that

$$\alpha = \mathbb{P}((\mathbb{X}_1, \dots, \mathbb{X}_n) \in A(\mathbb{X}_1, \dots, \mathbb{X}_n) | H_0)$$

then $A(\mathbb{X}_1, \dots, \mathbb{X}_n)$ is critical region of significance level α .

Example Hypothesis testing for equality of variance. Suppose that we have

$$\begin{cases} \mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2) \\ \mathbb{Y}_1, \dots, \mathbb{Y}_m \stackrel{i.i.d.}{\sim} N(\mu_y, \sigma_y^2) \end{cases}$$

We want to test hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$ vs $H_1 : \sigma_x^2 \neq \sigma_y^2$.

We have seen that $F = \frac{\sigma_x^2 \mathbb{S}_y^2}{\sigma_y^2 \mathbb{S}_x^2} \sim f(m-1, n-1)$, and

$$\frac{\alpha}{2} = \mathbb{P}\left(\frac{\mathbb{S}_x^2 \sigma_y^2}{\mathbb{S}_y^2 \sigma_x^2} \leq f_{\frac{\alpha}{2}}(n-1, m-1) | H_0\right) = \mathbb{P}\left(\frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \geq f_{1-\frac{\alpha}{2}}(m-1, n-1)\right)$$

$$\frac{\alpha}{2} = \mathbb{P}\left(\frac{\mathbb{S}_x^2 \sigma_y^2}{\mathbb{S}_y^2 \sigma_x^2} \geq f_{1-\frac{\alpha}{2}}(n-1, m-1) | H_0\right) = \mathbb{P}\left(\frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \leq \frac{1}{f_{1-\frac{\alpha}{2}}(n-1, m-1)}\right)$$

$$\Rightarrow \alpha = \mathbb{P}\left(\frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \leq \frac{1}{f_{1-\frac{\alpha}{2}}(n-1, m-1)} \cup \frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \geq f_{1-\frac{\alpha}{2}}(m-1, n-1)\right)$$

A critical region of significance level α for $H_0 : \sigma_x^2 = \sigma_y^2$ vs $H_1 : \sigma_x^2 \neq \sigma_y^2$ is

$$C = \left\{ \frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \leq \frac{1}{f_{1-\frac{\alpha}{2}}(n-1, m-1)} \cup \frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \geq f_{1-\frac{\alpha}{2}}(m-1, n-1) \right\}$$

Actually a test with critical region C is a LRT. But we will not prove it. \blacksquare

Now we come back to the question of C.I and hypothesis testing for the difference of means.

Example Suppose that we have

$$\begin{cases} \mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2) \\ \mathbb{Y}_1, \dots, \mathbb{Y}_m \stackrel{i.i.d.}{\sim} N(\mu_y, \sigma_y^2) \end{cases}$$

We want C.I for $\mu_x - \mu_y$ as a test for $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$

Case 1: suppose it is known that $\sigma_x^2 = \sigma_y^2 = \sigma^2$, we have

$$T = \frac{\bar{\mathbb{X}} - \bar{\mathbb{Y}} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} = \frac{\bar{\mathbb{X}} - \bar{\mathbb{Y}} - (\mu_x - \mu_y)}{\sqrt{\frac{(n-1)\mathbb{S}_x^2 + (m-1)\mathbb{S}_y^2}{\sigma^2(m+n-2)}} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(m+n-2)$$

and $100(1 - \alpha)\%$ C.I is

$$(\bar{\mathbb{X}} - \bar{\mathbb{Y}} - t_{\frac{\alpha}{2}} \sqrt{\dots}, \bar{\mathbb{X}} - \bar{\mathbb{Y}} + t_{\frac{\alpha}{2}} \sqrt{\dots})$$

A test of significance level α is

$$\text{rejecting } H_0 \text{ if } |T| = \frac{|\bar{\mathbb{X}} - \bar{\mathbb{Y}}|}{\sqrt{\dots}} \geq t_{\frac{\alpha}{2}}$$

Hence

$$\mathbb{P}(\text{Type I error}) = \mathbb{P}\left(\frac{|\bar{\mathbb{X}} - \bar{\mathbb{Y}}|}{\sqrt{\dots}} \geq t_{\frac{\alpha}{2}}\right) = \alpha$$

Case II: Suppose that we do not know if $\sigma_x = \sigma_y$, so we have

$$\begin{cases} \mathbb{X}_1, \dots, \mathbb{X}_n \stackrel{i.i.d.}{\sim} N(\mu_x, \sigma_x^2) \\ \mathbb{Y}_1, \dots, \mathbb{Y}_m \stackrel{i.i.d.}{\sim} N(\mu_y, \sigma_y^2) \end{cases}$$

Our interest is C.I for $\mu_x - \mu_y$ as a test for $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$
We consider two situations

1. If sample size n, m are not large ($n, m < 30$), we have two steps to derive C.I and test.
 - (a) step 1: Consider $H_0 : \sigma_x^2 = \sigma_y^2$ vs $H_1 : \sigma_x^2 \neq \sigma_y^2$. We reject H_0 if $F = \frac{\mathbb{S}_y^2}{\mathbb{S}_x^2} \leq f_{\frac{\alpha}{2}}(m-1, n-1)$ or $\geq f_{1-\frac{\alpha}{2}}(m-1, n-1)$. If H_0 is rejected we consider it is too complicated to derive C.I and a test, we would not further do it.
 - (b) step 2: If H_0 is accepted, then we consider $\sigma_x^2 = \sigma_y^2 = \sigma^2$, and test is derived as above.

2. If sample size is large (> 30), then

$$\bar{X} \sim N(\mu_x, \sigma_x^2) \text{ and } \bar{Y} \sim N(\mu_y, \sigma_y^2)$$

Therefore

$$T = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \xrightarrow{d} N(0, 1)$$

So

$$\begin{aligned} 1 - \alpha &= \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq Z \leq \zeta_{\frac{\alpha}{2}}) \cong \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \leq \zeta_{\frac{\alpha}{2}}) \\ &= \mathbb{P}(\bar{X} - \bar{Y} - \zeta_{\frac{\alpha}{2}} \sqrt{\dots} \leq \mu_x - \mu_y \leq \bar{X} - \bar{Y} - \zeta_{\frac{\alpha}{2}} \sqrt{\dots}) \end{aligned}$$

We now consider hypothesis $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$. We have $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \xrightarrow{d} N(0, 1)$, we consider that rejecting H_0 if

$$|T| = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\dots}} \geq \zeta_{\frac{\alpha}{2}}$$

Hence

$$\mathbb{P}(\text{Type I error}) = \mathbb{P}\left(\frac{|\bar{X} - \bar{Y}|}{\sqrt{\dots}} \geq \zeta_{\frac{\alpha}{2}}\right) = 1 - \mathbb{P}\left(\frac{|\bar{X} - \bar{Y}|}{\sqrt{\dots}} \leq \zeta_{\frac{\alpha}{2}}\right) \cong \mathbb{P}(Z \leq \zeta_{\frac{\alpha}{2}}) = \alpha$$

This is a test of significance level α .

Hypothesis Testing for p (probability of success). $\mathbb{Y} \sim \text{Binomial}(n, p)$, want a test for $H_0 : p = p_0$ vs $H_1 : p \neq p_0$. Let $\mathbb{Y}_1, \dots, \mathbb{Y}_n \stackrel{i.i.d.}{\sim} \text{Bernoulli}(p)$, and $\hat{Y} = \frac{\mathbb{Y}}{n} \xrightarrow{\mathbb{P}} p$.

$$\text{by C.L.T } \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\frac{\sum \mathbb{Y}_i}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathbb{P}} N(0, 1)$$

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}} \xrightarrow{\mathbb{P}} N(0, 1) \text{ by Slutsky theorem}$$

Under H_0 ,

$$\frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \rightarrow N(0, 1)$$

So

$$1 - \alpha = \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq \mathbb{Z} \leq \zeta_{\frac{\alpha}{2}})$$

$$\Rightarrow \alpha = \mathbb{P}(|\mathbb{Z}| \geq \zeta_{\frac{\alpha}{2}}) = \mathbb{P}\left(\left|\frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\right| \geq \zeta_{\frac{\alpha}{2}}\right)$$

So one appropriate test for significance level α is rejecting H_0 if $\frac{|\hat{p} - p_0|}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \geq \zeta_{\frac{\alpha}{2}}$

Difference of p 's: Suppose that we have $\begin{cases} \mathbb{X} \sim \text{binomial}(n, p_1) \\ \mathbb{Y} \sim \text{binomial}(m, p_2) \end{cases}$

Want C.I for $p_1 - p_2$ and test for $H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$.

Let $\hat{p}_1 = \frac{\mathbb{X}}{n}$, and $\hat{p}_2 = \frac{\mathbb{Y}}{m}$, by CLT

$$\begin{cases} \frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \rightarrow N(0, 1) \\ \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{m}}} \rightarrow N(0, 1) \end{cases} \Rightarrow \begin{cases} \hat{p}_1 \rightarrow N(p_1, \frac{p_1(1-p_1)}{n}) \\ \hat{p}_2 \rightarrow N(p_2, \frac{p_2(1-p_2)}{m}) \end{cases}$$

$$\Rightarrow \hat{p}_1 - \hat{p}_2 \cong N(p_1 - p_2, \frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m})$$

$$\Rightarrow \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1) + \hat{P}_2(1-\hat{P}_2)}{n}}} \cong N(0, 1)$$

Since

$$1 - \alpha = \mathbb{P}(-\zeta_{\frac{\alpha}{2}} \leq \mathbb{Z} \leq \zeta_{\frac{\alpha}{2}})$$

$$\Rightarrow \alpha = \mathbb{P}(|\mathbb{Z}| \geq \zeta_{\frac{\alpha}{2}}) = \mathbb{P}\left(\frac{|\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)|}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1) + \hat{P}_2(1-\hat{P}_2)}{n}}} \geq \zeta_{\frac{\alpha}{2}}\right)$$

under H_0

$$\alpha = \mathbb{P}\left(\frac{|\hat{P}_1 - \hat{P}_2|}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1) + \hat{P}_2(1-\hat{P}_2)}{n}}} \geq \zeta_{\frac{\alpha}{2}}\right)$$

Therefore an approximate test of significance level α is rejecting H_0 if $\frac{|\hat{P}_1 - \hat{P}_2|}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1) + \hat{P}_2(1-\hat{P}_2)}{n}}} \geq \zeta_{\frac{\alpha}{2}}$.

8 Goodness of Fit Test – chi-square Test

In statistical inference of C.I and hypothesis testing, we need to assume that the random sample is drawn from a fixed distribution (normal or Poisson, etc). How do we know that the assumed distribution is true?

Hypothesis may be:

1. $H_0 : \mathbb{X} \sim N(\mu, \sigma^2)$
2. $H_0 : \mathbb{X} \sim N(5, 100)$
3. $H_0 : \mathbb{X} \sim Poisson(\lambda)$
4. $H_0 : \mathbb{X} \sim Poisson(5)$

Suppose that now we have a random sample $\mathbb{X}_1, \dots, \mathbb{X}_n$ from an unknown distribution $f(x)$, we consider the hypothesis

$$H_0 : \mathbb{X} \sim f_0(x)$$

where f_0 is known without unknown parameters.

Let A_1, \dots, A_k be a partition for space of \mathbb{X} , we define

$$P_j = \mathbb{P}(\mathbb{X} \in A_j | H_0) = \int_{A_j} f_0(x) dx, j = 1, 2, \dots, n$$

$$\Rightarrow \sum_{j=1}^k P_j = 1$$

Let N_j denote the number of x_1, \dots, x_n falling into A_j , then $\sum_{j=1}^k N_j = n$. Note N_j is a statistics of $\mathbb{X}_1, \dots, \mathbb{X}_n$. We have the following theorem

Theorem Let $Q_k = \sum_{j=1}^k \frac{(N_j - nP_j)^2}{nP_j} = \sum_{j=1}^k \frac{(\text{actual } \# - \text{theoretical } \#)^2}{\text{theoretical } \# \text{ in } A_j}$, we have $Q_k \xrightarrow{d} \chi^2(k - 1)$ when H_0 is true.

Proof We consider $k = 2$ only ($N_2 = n - N_1, P_2 = 1 - P_1$)

$$\begin{aligned} Q_2 &= \frac{(N_1 - nP_1)^2}{nP_1} + \frac{(N_2 - nP_2)^2}{nP_2} = \frac{(N_1 - nP_1)^2}{nP_1} + \frac{((n - N_1) - n(1 - P_1))^2}{n(1 - P_1)} \\ &= (N_1 - nP_1)^2 \left(\frac{1}{nP_1} + \frac{1}{n(1 - nP_1)} \right) \\ &= \frac{(N_1 - nP_1)^2}{nP_1(1 - P_1)} \xrightarrow{d} \chi^2(1) \end{aligned}$$

Table 1:

Genotype	Group	Probability.	observations
A_1	Round and Yellow	$P_1 = \mathbb{P}(A_1) = \frac{9}{16}$	$N_1 = 315$
A_2	Round and Green	$P_2 = \mathbb{P}(A_2) = \frac{3}{16}$	$N_2 = 128$
A_3	Angular and Yellow	$P_3 = \mathbb{P}(A_3) = \frac{3}{16}$	$N_3 = 101$
A_4	Angular and Green	$P_4 = \mathbb{P}(A_4) = \frac{3}{16}$	$N_4 = 32$

because

$$N_1 \sim b(n, P_1) \Rightarrow \frac{N_1 - nP_1}{\sqrt{nP_1(1 - P_1)}} \xrightarrow{d} N(0, 1)$$

Definition The Pearson's Chi-square test for hypothesis

$$H_0 : \mathbb{X} \sim f_0(x)$$

is: reject H_0 if $Q_k \geq \chi^2_{1-\alpha}$ Because

$$\begin{aligned} \mathbb{P}(\text{type I error}) &= \mathbb{P}(\text{reject } H_0 | H_0) \cong \mathbb{P}(Q_k \geq \chi^2_{1-\alpha} | H_0) \\ &= \mathbb{P}(\chi^2(k-1) \geq \chi^2_{1-\alpha}) = 1 - \alpha \end{aligned}$$

Example (verification of Mendelian Theory) Shape and color of a pea may be classified into four groups. Their theoretical probabilities and actual observations are listed in table 1:

Is Mendelian theory correct?

Solution We consider hypothesis $H_0 : P_1 = \frac{9}{16}, P_2 = \frac{3}{16}, P_3 = \frac{3}{16}, P_4 = \frac{1}{16}$. Notice we have $k = 4$ and $n = N_1 + N_2 + N_3 + N_4 = 556$, so

$$Q_4 = \sum_{j=1}^4 \frac{(N_j - nP_j)^2}{nP_j} = 0.47 \leq \chi^2_{0.95}(3) = 7.81$$

So we do not reject H_0 and consider Mendelian theory to be correct!

We now consider hypothesis

$$H_0 : \mathbb{X} \sim f(x, \theta_1, \dots, \theta_p)$$

where $\theta_1, \dots, \theta_p$ are unknown parameters. Again, let A_1, \dots, A_k be a partition of space of \mathbb{X} .

Let $\hat{\theta}_1, \dots, \hat{\theta}_k$ be m.l.e's of $\theta_1, \dots, \theta_p$, we define

$$\hat{P}_j = \int_{A_j} f(x, \hat{\theta}_1, \dots, \hat{\theta}_p) dx$$

N_j = # of x_1, \dots, x_n falling into A_j .

Theorem Let

$$Q_k = \sum_{j=1}^k \frac{(N_j - n\hat{P}_j)^2}{n\hat{P}_j}$$

Then $Q_k \xrightarrow{d} \chi^2(k - p - 1)$ if H_0 is true.

9 Appendix

Appendix

Appendix provides proof to harder theorems. Some of which require background in analysis.

A.1 Interchanging Differentiation and Integration If $f(x, t)$ is a function of two variables, satisfying the following 3 conditions:

- (1) $f(x, t)$ exists on a closed cell $[a, b] \times [c, d]$
- (2) $\forall_{t \in [c, d]} \int_a^b f(x, t) dx$ exists
- (3) Given $s \in (c, d)$, for any $\epsilon > 0$ corresponds a δ such that

$$|D_2 f(x, t) - D_2 f(x, s)| \leq \epsilon \text{ for all } x \in [a, b] \text{ and } |t - s| \leq \delta$$

where $D_2 f(x, s)$ means the derivative of f with respect to its second variable at s . Then $D_2 \int_a^b f(x, s) dx = \int_a^b D_2 f(x, s) dx$ (This means, of course, $D_2 \int_a^b f(x, s) dx$ and $\int_a^b D_2 f(x, s) dx$ exist).

Proof If we let

$$g_t(x) = \frac{f(x, t) - f(x, s)}{t - s}$$

then by mean value theorem, there exist u between s and t , such that

$$g_t(x) = D_2 f(x, u)$$

So if $|t - s| \leq \delta$, $|g_t(x) - D_2 f(x, s)| \leq \epsilon$ for all $x \in [a, b]$ by condition (3). Therefore $g_t(x) \rightarrow D_2 f(x, s)$ uniformly on $[a, b]$ as $t \rightarrow s$.

Since $g_t(x) \rightarrow D_2 f(x, s)$ uniformly on $[a, b]$ as $t \rightarrow s$,

$$\begin{aligned} \lim_{t \rightarrow s} \int_a^b g_t(x) dx &= \int_a^b \lim_{t \rightarrow s} g_t(x) dx \\ \lim_{t \rightarrow s} \int_a^b g_t(x) dx &= \lim_{t \rightarrow s} \int_a^b \frac{f(x, t) - f(x, s)}{t - s} dx = \lim_{t \rightarrow s} \frac{\int_a^b f(x, t) dx - \int_a^b f(x, s) dx}{t - s} \\ &= D_2 \int_a^b f(x, t) dx = \int_a^b \lim_{t \rightarrow s} g_t(x) dx = \int_a^b D_2 f(x, s) dx \end{aligned}$$

Remark If $D_2 f$ is continuous on $[a, b] \times [c, d]$, condition (3) holds. Because for every ϵ , corresponds a δ , such that

A.2 Holder's inequality