Chapter 5 Properties of Random Sample

Section 5.5 Convergence Concepts

**Definition 5.5.1:** A sequence of random variables, \( X_1, X_2, \cdots \), converges in probability to a random variable \( X \) if, for every \( \varepsilon > 0 \),
\[
\lim_{n \to \infty} P( | X_n - X | \geq \varepsilon ) = 0, \text{ or equivalently, } \lim_{n \to \infty} P( | X_n - X | < \varepsilon ) = 1.
\]

**Theorem 5.5.2 (Weak Law of Large Numbers):** Let \( X_1, X_2, \cdots \) be iid random variables with \( EX_i = \mu \) and \( VarX_i = \sigma^2 < \infty \). Define \( \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \). Then for every \( \varepsilon > 0 \),
\[
\lim_{n \to \infty} P( | \bar{X}_n - \mu | < \varepsilon ) = 1.
\]

**Proof:** Use Chebychev’s Inequality.

**Note:** WLLN states that the sample mean \( \bar{X} \) is close to \( \mu \) with high probability as \( n \) gets larger. This property of an estimator is known as consistency. In general, to show an estimator is consistent for \( \theta \), one needs to prove that the estimator converges in probability to \( \theta \). Consistency of estimators will be formally defined and discussed in Chapter 10.
Example 5.5.3 (Consistency of $S_n^2$). Let $X_1, X_2, \cdots$, be iid random variables with $EX_i = \mu$ and $VarX_i = \sigma^2 < \infty$ and define $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$, can we prove a WLLN for $S_n^2$? Using Chebychev’s Inequality, we have

$$P(|S_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{E(S_n^2 - \sigma^2)^2}{\varepsilon^2} = \frac{Var(S_n^2)}{\varepsilon^2}.$$ 

And thus, a sufficient condition that $S_n^2$ converges in probability to $\sigma^2$ is that $Var(S_n^2) \to 0$ when $n \to \infty$.

**Note:** If $X_1, \cdots, X_n$ are iid $n(\mu, \sigma^2)$, then $Var(S_n^2) \to 0$ as $n \to \infty$, so that $S_n^2$ is a consistent estimator of $\sigma^2$.

**Theorem 5.5.4:** Suppose that $X_1, X_2, \cdots$ converges in probability to a random variable $X$ and that $h$ is continuous function. Then $h(X_1), h(X_2), \cdots$ converges in probability to $h(X)$.

**Example 5.5.5 (Consistency of $\hat{S}$):** If $S_n^2$ is a consistent estimator of $\sigma^2$, then by Theorem 5.5.4, the sample standard deviation $S_n = \sqrt{S_n^2}$ is a consistent estimator of $\sigma$. Note that $E(S_n)$ is a biased estimator of $\sigma$ (Exercise 5.11).
Definition 5.5.10: A sequence of random variables, $X_1, X_2, \ldots$, converges in distribution to a random variable $X$ if
\[
\lim_{n \to \infty} F_{X_n}(x) = F_X(x) \text{ for all points } x \text{ where } F_X(x) \text{ is continuous.}
\]

Theorem 5.5.12: If the sequence of random variables, $X_1, X_2, \ldots$, converges in probability to $X$, then the sequence converges in distribution to $X$.

Theorem 5.5.13: The sequence of random variable, $X_1, X_2, \ldots$, converges in probability to a constant $\mu$ if and only if the sequence also converges in distribution to $\mu$. That is, the statement
\[
\lim_{n \to \infty} P(|X_n - \mu| > \varepsilon) = 0 \text{ for every } \varepsilon > 0
\]
is equivalent to
\[
\lim_{n \to \infty} P(X_n \leq x) = \begin{cases} 
0, & x < u, \\
1, & x > u.
\end{cases}
\]

Theorem 5.5.14 (Central Limit Theorem): Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exits in neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive $h$). Let $EX_i = \mu$ and $VarX_i = \sigma^2 > 0$. (Both $\mu$ and $\sigma^2$ are finite since the mgf exists) Define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_n$. Let $G_n(x)$ denote the cdf of $\sqrt{n} (\bar{X}_n - \mu) / \sigma$. Then, for any $-\infty < x < \infty$,
\[ \lim_{n \to \infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy. \]

That is, \( \sqrt{n}(\bar{X}_n - \mu) / \sigma \) has a limiting standard normal distribution.

**Proof:** Use the theorem 2.3.12.

**Some Notes:**
- assumptions: independence, identical distribution and mgf exists
- finite variance is necessary for convergence to normality (CLT will not apply to rvs from Cauchy distribution)
- how good the approximation is in general depends on the original distribution

**Theorem 2.3.12:** (Convergence of mgfs) Suppose \( \{X_i, i = 1, 2, \ldots\} \) is a sequence of random variables, each with mgf \( M_{X_i}(t) \). Furthermore, suppose that \( \lim_{i \to \infty} M_{X_i}(t) = M_X(t) \) for all \( t \) in a neighborhood of 0, and \( M_X(t) \) is an mgf.

Then there is unique cdf \( F_X \) whose moments are determined by \( M_X(t) \) and for all \( x \) where is continuous, we have \( \lim_{i \to \infty} F_{X_i}(x) = F_X(x) \). That is, convergence, for \( |t| < h \) of mgfs to an mgf implies convergence of cdfs.
**Theorem 5.5.15 (Stronger form of the Central Limit Theorem)** Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $EX_i = \mu$ and $0 < VarX_i = \sigma^2 < \infty$. Define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cdf of $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$. Then, for any $-\infty < x < \infty$,

$$\lim_{n \to \infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy.$$ 

That is, $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ has a limiting standard normal distribution.

**Theorem 5.5.17 (Slutsky’s Theorem)** If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then

a. $Y_n X_n \to aX$ in distribution.

b. $X_n + Y_n \to X + a$ in distribution.

**Example 5.5.18 (Normal approximation with estimated variance)** Suppose that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to n(0,1),$$

but the value of $\sigma$ is unknown. We have seen in example 5.5.3 that if $Var(S_n^2) \to 0$, then $S_n^2 \to \sigma^2$ in probability. By exercise 5.32, we have $\sigma / S_n \to 1$ in probability. By theorem 5.5.17, we have

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sigma}{S_n} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to n(0,1).$$
**Notes (relationship between several convergences)**

1. converges in probability ⇒ converges in distribution
2. converges in probability to a constant ⇔ converges in distribution to a constant
3. Slutsky’s Theorem

**Example 5.5.19 (Estimating the odds)** Suppose that $X_1, X_2, \ldots, X_n$ are iid Bernoulli($p$) random variables. The typical parameter of interest is $p$, which can be estimated by $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. We can obtain the distribution of $n\bar{X}_n$, which is Binomial($n, p$). Sometimes we are interested the odds, $\frac{p}{1-p}$, which may be estimated by $\frac{\bar{X}_n}{1-\bar{X}_n}$. Then what are properties of it? For example, how to calculate the variance of it? The exact calculation may be difficult, but an approximation can be obtained.

**Definition 5.5.20:** If a function $g(x)$ has derivatives of order $r$, that is, $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)$ exists, then for any constant $a$, the Taylor polynomial of order $r$ about $a$ is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!} (x-a)^i.$$
Theorem 5.5.21 (Taylor) If \( g^{(r)}(a) = \frac{d^r}{dx^r} g(x) \bigg|_{x=a} \) exists, then \( \lim_{x \to a} \frac{g(x) - T_r(x)}{(x - a)^r} = 0. \)

For the statistical application of Taylor’s Theorem, we are most concerted with the first-order Taylor series. Let \( T \) be a random variable with mean \( \theta \) and suppose that \( g \) is differentiable function, then
\[
g(t) \approx g(\theta) + g'(\theta)(t - \theta).
\]

Then we have
\[
E(g(T)) \approx E(g(\theta)) + g'(\theta)E(T - \theta) = g(\theta),
\]
and
\[
Var(g(T)) \approx E[g(T) - g(\theta)]^2 = E[g'(\theta)(T - \theta)]^2 = [g'(\theta)]^2 Var(T).
\]

Example 5.5.22 (Continuation of Example 5.5.19) Recall that we are interested in the properties of \( \frac{\bar{X}_n}{1 - \bar{X}_n} \). Let
\[
g(t) = \frac{t}{1 - t}, \quad \theta = E(\bar{X}_n) = p, \quad g'(t) = \frac{1}{(1-t)^2}, \quad \text{thus}
\]
\[
E(g(\bar{X}_n)) \approx g(p) = \frac{p}{1 - p},
\]
and
\[
Var(g(X_n)) = [g'(p)]^2 Var(X_n) = \frac{1}{(1-p)^2} \frac{1}{n} p(1-p) = \frac{p}{n(1-p)^3}.
\]

**Example 5.5.23 (Approximate mean and variance)** Suppose \( X \) is a random variable with \( E(X) = \mu \neq 0 \). If we want to estimate the mean and variance of the random variable \( g(X) \), we have \( E(g(X)) \approx g(\mu) \) and \( Var(g(X)) \approx [g'(\mu)]^2 Var(X) \). Specifically, for \( g(X) = 1/X \), we have \( E(1/X) \approx 1/\mu \) and \( Var(1/X) \approx (1/\mu)^4 Var(X) \).

**Theorem 5.5.24 (Delta Method)** Let \( Y_n \) be a sequence of random variable that satisfies \( \sqrt{n}(Y_n - \theta) \rightarrow n(0, \sigma^2) \) in distribution. For a given \( g \) and a specific value of \( \theta \), suppose that \( g'(\theta) \) exists and is not 0. Then

\[
\sqrt{n}[g(Y_n) - g(\theta)] \rightarrow n(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.}
\]

**Example 5.5.25 (Continuation of Example 5.5.23)** Suppose now that we have the mean of a random sample \( \bar{X}_n \), then we have \( \sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right) \rightarrow n(0, (1/\mu)^4 Var(X)) \).
Section 10.1 - Point Estimation

Section 10.1.1 - Consistency

**Definition 10.1.1** A sequence of estimators $\hat{W}_n = W_n(X_1, \ldots, X_n)$ is a consistent sequence of estimators of the parameter $\theta$ if, for every $\epsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n \to \infty} P_\theta(|W_n - \theta| < \epsilon) = 1 \text{ or } \lim_{n \to \infty} P_\theta(|W_n - \theta| \geq \epsilon) = 0.$$ 

Recall from Chapter 5, we say that $W_n$ converges in probability to $\theta$.

Also, recall an application of Chebychev’s Inequality:

$$P_\theta(|W_n - \theta| \geq \epsilon) \leq \frac{E_\theta[(W_n - \theta)^2]}{\epsilon^2} = \frac{\text{Var}_\theta W_n + (\text{Bias}_\theta W_n)^2}{\epsilon^2}.$$ 

**Theorem 10.1.3** If $W_n$ is a sequence of estimators of a parameter $\theta$ satisfying:

(i) $\lim_{n \to \infty} \text{Var}_\theta W_n = 0$

(ii) $\lim_{n \to \infty} \text{Bias}_\theta W_n = 0$

for every $\theta \in \Theta$, then $W_n$ is a consistent sequence of estimators of $\theta$. 
Theorem 10.1.5 If $W_n$ is a consistent sequence of estimators of a parameter $\theta$.

Let $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ be sequences of constants satisfying:

(i) $\lim_{n \to \infty} a_n = a,$

(ii) $\lim_{n \to \infty} b_n = b,$

Then the sequence $U_n = a_n W_n + b_n$ is a consistent sequence of estimators of $a\theta + b$. Specifically, if $a = 1$ and $b = 0$, then $U_n = a_n W_n + b_n$ is a consistent sequence of estimators of $\theta$.

Example 10.1.2 (consistency of $\bar{X}$) Let $X_1, \ldots, X_n$ be a random sample from $n(\mu,1)$, and consider the consistency of $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Example (consistency of $S^2$): Let $X_1, \ldots, X_n$ be a random sample from $n(\mu,\sigma^2)$. Consider the estimators of $\sigma^2: S_n^2$ and $\hat{\sigma}_n^2 = \frac{n-1}{n} S_n^2$ (MLE).

Theorem 10.1.6 (Consistency of MLEs) Let $X_1, \ldots, X_n$ be a random sample from $f(x|\theta)$, and let $L(\theta | x) = \prod_{i=1}^{n} f(x_i | \theta)$ be the likelihood function. Let $\hat{\theta}$ denote the MLE of $\theta$. Let $\tau(\theta)$ be a continuous function
of $\theta$. Under the regularity conditions in Miscellanea 10.6.2 on $f(\theta \mid x)$ and, hence, $L(\theta \mid x)$, for every $\varepsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n \to \infty} P_{\theta} (|\tau(\hat{\theta}) - \tau(\theta)| \geq \varepsilon) = 0.$$ 

That is, $\tau(\hat{\theta})$ is a consistent estimator of $\tau(\theta)$.

**Section 10.1.2 - Efficiency**

**Definition:** Given two unbiased estimators for $\theta$, $W$ and $T$, with variances $Var(W)$ and $Var(T)$, respectively, then the efficiency of $T$ relative to $W$ is defined by the ratio

$$eff(T,W) = \frac{Var(W)}{Var(T)}.$$

**Note:** If $eff(W,T) > 1$, i.e., $Var(W) > Var(T)$, then we say that $T$ is relatively more efficient that $W$.

**Example:** Let $X_1,\ldots,X_n$ be a random sample from $n(\mu,\sigma^2)$. Compute the efficiency of $S_n^2$ relative to $\hat{\sigma}^2_n = \frac{n-1}{n} S_n^2$. (Will not work because $\hat{\sigma}^2_n = \frac{n-1}{n} S_n^2$ is biased)
**Definition 10.1.11** A sequence of estimators $W_n$ is asymptotically efficient for a parameter $\tau(\theta)$ if
\[ \sqrt{n}(W_n - \tau(\theta)) \to n(0, \nu(\theta)) \] in distribution and
\[ \nu(\theta) = \frac{[\tau'(\theta)]^2}{E_\theta(\frac{\partial}{\partial \theta} \log f(X|\theta))^2}. \]
i.e., the asymptotic variance of $W_n$ achieves the Cramer-Rao Lower Bound.

**Theorem 10.1.12** (Consistency and Asymptotic efficiency of the MLEs) Let $X_1, \cdots, X_n$ be iid $f(x|\theta)$, let $\hat{\theta}$ denote the MLE of $\theta$, and let $\tau(\theta)$ be a continuous function of $\theta$. Under the regularity conditions on Miscellanea 10.6.2 (p. 516) on $f(x|\theta)$ and, hence, $L(\theta|x)$,
\[ \sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \to n(0, \nu(\theta)), \]
where $\nu(\theta)$ is the Cramer-Rao Lower Bound. That is, $\tau(\hat{\theta})$ is a consistent and asymptotically efficient estimator of $\tau(\theta)$.

**In other words, under the conditions of Theorem 10.1.12:**
- $\tau(\hat{\theta})$ is a consistent estimator of $\tau(\theta)$.
- $\tau(\hat{\theta})$ has an asymptotic normal distribution and has an asymptotic variance that equals the Cramer-Rao Lower Bound. Therefore,
\[
\frac{\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)]}{\sqrt{\nu(\theta)}} \rightarrow n(0,1) \text{ in distribution.}
\]

**Notes:**

- Most of the common distributions (for instance, the regular exponential family of distributions) satisfy the conditions of Theorem 10.1.12.
- However, when the support depends on the parameter \( \theta \), Theorem 10.1.12 is not applicable.

**Definition 10.1.16** If two estimators \( W_n \) and \( T_n \) satisfy

\[
\sqrt{n}(W_n - \tau(\theta)) \rightarrow n(0,\sigma_W^2),
\]

\[
\sqrt{n}(T_n - \tau(\theta)) \rightarrow n(0,\sigma_T^2),
\]

in distribution, the asymptotic relative efficiency (ARE) of \( T_n \) with respect to \( W_n \) is

\[
ARE(T_n, W_n) = \frac{\sigma_W^2}{\sigma_T^2}.
\]

**Example:** Let \( X_1, \ldots, X_n \) be a random sample from \( n(\mu, \sigma^2) \). Compute the efficiency of \( S_n^2 \) relative to \( \hat{\sigma}_n^2 = \frac{n-1}{n} S_n^2 \). (Will not work because \( \hat{\sigma}_n^2 = \frac{n-1}{n} S_n^2 \) is biased)
Section 10.1.3 Calculations and Comparisons

From the Delta method and asymptotic efficiency of MLEs, the variance of $h(\hat{\theta})$ is:

$$Var(h(\hat{\theta}) | \theta) \approx \frac{[h'(\theta)]^2}{I_n(\theta)}$$

$$= \frac{[h'(\theta)]^2}{E_\theta(-\frac{\partial^2}{\partial \theta^2} \log L(\theta | X))}$$

$$\approx \frac{[h'(\theta)]^2 |_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial \theta^2} \log L(\theta | X) |_{\theta=\hat{\theta}}}.$$ 

Example 10.1.14 (Approximate binomial variance) In example 7.2.7 we saw that $\hat{p} = \overline{X}$ is the MLE of $p$, where $X_1, \ldots, X_n$ are iid from Bernoulli($p$). Then

$$Var_{\hat{p}}(\hat{p}) = \frac{p(1-p)}{n} \text{ (direct calculation).}$$

So

$$Var_{\hat{p}}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}.$$
We can also obtain this calculating:

\[
\frac{\partial^2}{\partial p^2} \log L(p \mid X) \big|_{p=\hat{p}} = \frac{\partial^2}{\partial p^2} \left[ n\hat{p} \log(p) + n(1 - \hat{p}) \log(1 - p) \right] \big|_{p=\hat{p}} \\
= -\frac{n\hat{p}}{p^2} - \frac{n(1 - \hat{p})}{(1 - p)^2} \big|_{p=\hat{p}} = -\frac{n}{\hat{p}(1 - \hat{p})}.
\]

We can calculate the variance of \( \frac{\hat{p}}{1 - \hat{p}} \), the MLE of odds:

\[
\text{Var}_p \left( \frac{\hat{p}}{1 - \hat{p}} \right) = \frac{\left[ \frac{\partial}{\partial p} \left( \frac{p}{1 - p} \right) \right]_{p=\hat{p}}}{\frac{\partial^2}{\partial p^2} \log L(p \mid X) \big|_{p=\hat{p}}} = \frac{\hat{p}}{n(1 - \hat{p})^3}.
\]

We can calculate the variance of \( \frac{\hat{p}}{1 - \hat{p}} \) using the delta method:

\[
\sqrt{n} (\hat{p} - p) \rightarrow n(0, p(1 - p)) \text{ in probability;}
\]

\[
\frac{\partial}{\partial p} \frac{p}{1 - p} = \frac{1}{(1 - p)^2},
\]

So

\[
\sqrt{n} \left( \frac{\hat{p}}{1 - \hat{p}} - \frac{p}{1 - p} \right) \rightarrow n(0, p(1 - p) \left[ \frac{1}{(1 - p)^2} \right]^2) = n(0, \frac{p}{(1 - p)^3}).
\]

For the variance of \( \hat{p}(1 - \hat{p}) \), we have
Section 10.3 Hypothesis Testing

Section 10.3.1 Asymptotic Distribution of LRTs

**Theorem 10.3.1 (Asymptotic distribution of the LRT – simple \( H_0 \))** For testing \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \), suppose \( X_1, \ldots, X_n \) are iid \( f(x \mid \theta) \), \( \hat{\theta} \) is the MLE of \( \theta \), and \( f(x \mid \theta) \) satisfies the regularity conditions in Miscellanea 10.6.2. Then under \( H_0 \), as

\[
-2 \log \lambda(x) \rightarrow \chi^2_1 \quad \text{in distribution}
\]

where \( \chi^2_1 \) is a \( \chi^2 \) random variable with 1 degree of freedom. Therefore an approximate \( \alpha \) level test for \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \), Rejects \( H_0 \) when \( -2 \log \lambda(x) > \chi^2_{1, \alpha} \).

**Note:**
- Proof of this Theorem uses Taylor expansion, the asymptotic normality of the MLE and Slutsky’s Theorem.
**Example 10.3.2 (Poisson LRT)** For testing \( H_0: \lambda = \lambda_0 \) versus \( H_1: \lambda \neq \lambda_0 \) based on \( X_1, \ldots, X_n \) iid Poisson(\( \lambda \)), we have

\[
-2\log \lambda(x) = -2\log \left( \frac{\exp(-n\lambda_0)\lambda_0^{n\lambda_0}}{\exp(-n\lambda)\lambda^{n\lambda}} \right) = 2n[\lambda_0 - \hat{\lambda} - \hat{\lambda} \log(\lambda_0 / \hat{\lambda})],
\]

where \( \hat{\lambda} = \bar{x} \) is the MLE of \( \lambda \).

**Theorem 10.3.3** Let \( X_1, \ldots, X_n \) be a random sample from a pdf or pmf \( f(x | \theta) \). Under the regularity conditions in Miscellanea 10.6.2, if \( \theta \in \Theta_0 \) (i.e., under \( H_0 \)), then the distribution of the statistic \(-2\log \lambda(X)\) converges to a chi-squared distribution as \( n \to \infty \). The degrees of freedom, \( v \), of the limiting distribution is the difference between the number of free parameters specified by \( \theta \in \Theta_0 \) and the number of free parameters specified by \( \theta \in \Theta \).

Therefore an approximate \( \alpha \) level test for \( H_0: \theta \in \Theta_0 \) versus \( H_1: \theta \in (\Theta - \Theta_0) \):

Rejects \( H_0 \) when \(-2\log \lambda(x) > \chi^2_{v, \alpha} \).

**Example 10.3.4 (multinomial LRT)** Let \( \theta = (p_1, \ldots, p_5) \) and \( X_1, \ldots, X_n \) iid from this multinomial distribution. The likelihood function is

\[
L(\theta | X) = p_1^{x_1} \cdots p_5^{x_5}.
\]
Consider the testing

\( H_0 : p_1 = p_2 = p_3 \) and \( p_4 = p_5 \) versus \( H_1 : H_0 \) is not true.

The MLE of \( \theta \) is:

\[
\hat{\theta} = \left( \frac{y_1}{n}, \ldots, \frac{y_5}{n} \right) \text{ (under } H_1); \\
\hat{\theta} = \left( \frac{y_1 + y_2 + y_3}{n}, \frac{y_1 + y_2 + y_3}{3n}, \frac{y_1 + y_2 + y_3}{3n}, \frac{y_4 + y_5}{2n}, \frac{y_4 + y_5}{2n} \right) \text{ (under } H_0). 
\]

The limiting distribution of LRT test has 3 degrees of freedom.

Section 10.3.2 Other Large-Sample Tests

**Definition:** A Wald test is a test based on statistic of the form

\[
Z_n = \frac{W_n - \theta_0}{S_n},
\]

where \( \theta_0 \) is a hypothesized value of the parameter \( \theta \), \( W_n \) is an estimator of \( \theta \), and \( S_n \) is a standard error of \( W_n \), an estimate of the standard deviation of \( W_n \).

**Application of Theorem 10.1.12:** Let \( \hat{\nu}(\theta) \) be consistent estimator for \( \nu(\theta) \), then
\[
\frac{\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)]}{\sqrt{v(\theta)}} \rightarrow n(0,1).
\]

Wald’s statistic is asymptotically standard normal by applying Slutsky’s Theorem.

**Notes**

- We can use Wald’s statistic as a test statistic for constructing **approximate, asymptotic, or large sample** tests for \( \tau(\theta) \). The resulting test is known as the Wald’s test.
- Inverting the Wald’s test will give us an **approximate, asymptotic, or large sample** confidence interval for \( \tau(\theta) \). This is equivalent to treating equation (1) as a pivotal quantity.
- Inference procedures based on the Wald’s statistic do not perform very well in small samples.
- Estimating the variance using the Cramer-Rao Lower Bound will usually result in an underestimation of the true variance.

**Example 10.3.5 (Large-sample binomial tests)** Let \( X_1, \ldots, X_n \) be iid Bernoulli(\( p \)). Consider \( H_0 : p \leq p_0 \) versus \( H_1 : p > p_0 \), where \( 0 < p_0 < 1 \) is specified value. Consider the Wald test:

\[
\frac{\hat{p}_n - p}{\sqrt{(\hat{p}_n(1 - \hat{p}_n) / n)}} \rightarrow n(0,1).
\]
When \( p = p_0 \), \( Z_n = \frac{\hat{p}_n - p_0}{\sqrt{(\hat{p}_n(1 - \hat{p}_n)/n)}} \sim n(0,1) \). We reject \( H_0 \) if \( Z_n > z_\alpha \). The same statistic \( Z_n \) obtains if we use the information number to derive a standard error for \( \hat{p}_n \).

If we are interested in \( H_0 : p = p_0 \) versus \( H_1 : p \neq p_0 \), we know that

\[
\frac{\hat{p}_n - p}{\sqrt{(p(1 - p)/n)}} \rightarrow n(0,1).
\]

So we can use \( Z'_n = \frac{\hat{p}_n - p_0}{\sqrt{(p_0(1 - p_0)/n)}} \rightarrow n(0,1) \) as a test statistic. We reject \( H_0 \) if \(|Z'_n| > z_{\alpha/2}\).

**Definition:** The score statistic is defined as

\[
S(\theta) = \frac{\partial}{\partial \theta} \log f(X \mid \theta) = \frac{\partial}{\partial \theta} \log L(\theta \mid X).
\]

Recall (from p. 336 equation 7.3.8 – proof of the Cramer-Rao inequality) that for all \( \theta \),

\[
E_\theta S(\theta) = E_\theta \left( \frac{\partial}{\partial \theta} \log L(\theta \mid X) \right) = 0,
\]

and hence, \( Var(S(\theta)) = E_\theta S^2(\theta) \). Thus, by Lemma 7.3.11,

\[
Var(S(\theta)) = E_\theta \left( \left[ \frac{\partial}{\partial \theta} \log L(\theta \mid X) \right]^2 \right) = -E_\theta \left( \frac{\partial^2}{\partial \theta^2} \log L(\theta \mid X) \right) = I_n(\theta).
\]

By the asymptotic properties of the MLE,
$$Z_S = \frac{S(\theta_0)}{\sqrt{I_n(\theta_0)}}$$

converges to a standard normal in probability when \( H_0 : \theta = \theta_0 \) is true.

Therefore an approximate \( \alpha \) level test for \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \):

Rejects \( H_0 \) when \( |Z_S| \geq z_{\alpha/2} \).

**Example 10.3.6 (Binomial score test)** Consider the example 10.3.5 and the test: \( H_0 : p = p_0 \) versus \( H_1 : p \neq p_0 \),

We have:

$$S(p) = \frac{\hat{p}_n - p}{p(1 - p)/n} \quad \text{and} \quad I_n(p) = \frac{n}{p(1 - p)}$$

So the score statistic is

$$Z_S = \frac{S(p_0)}{\sqrt{I_n(p_0)}} = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1 - p_0)/n}}.$$ 

So we rejects \( H_0 \) when \( |Z_S| \geq z_{\alpha/2} \).

**Notes:**

- Asymptotic tests implies that the Type I error probability will be approximately \( \alpha \) if \( \theta \in \Theta_0 \) and for large \( n \), i.e., \( \lim_{n \to \infty} P(\text{Reject } H_0 \mid \theta \in \Theta_0) = \alpha \). It does not imply \( \lim_{n \to \infty} \sup_{\theta \in \Theta_0} P(\text{Reject } H_0 \mid \theta \in \Theta_0) = \alpha \).
Section 10.4 Interval Estimation

Section 10.4.1 MLE Based Method

If \( X_1, \ldots, X_n \) from \( f(x \mid \theta) \) and \( \hat{\theta} \) is the MLE of \( \theta \), then

\[
\text{var}_\theta(h(\hat{\theta}) \mid \theta) \approx \frac{[h'(\theta)]^2}{\frac{\partial^2}{\partial \theta^2} \log L(\theta \mid X) \big|_{\theta=\hat{\theta}}} \quad \text{and} \quad \frac{h(\hat{\theta}) - h(\theta)}{\sqrt{\text{var}_\theta(h(\hat{\theta}) \mid \theta)}} \rightarrow n(0,1),
\]

So the approximate confidence interval is:

\[
h(\hat{\theta}) - z_{\alpha/2} \sqrt{\text{var}_\theta(h(\hat{\theta}) \mid \theta)} \leq h(\theta) \leq h(\hat{\theta}) + z_{\alpha/2} \sqrt{\text{var}_\theta(h(\hat{\theta}) \mid \theta)}.
\]

Example 10.4.1 (Confidence Interval for Odds) We know that the MLE of odds \( p / (1 - p) \) is \( \hat{p} / (1 - \hat{p}) \) and its approximate variance is \( \text{var}_\hat{p} \left( \frac{\hat{p}}{1 - \hat{p}} \right) \approx \frac{\hat{p}}{n(1 - \hat{p})^3} \).

Section 10.4.1 Score Test Based Method
\[ Q(X | \theta) = \frac{\partial}{\partial \theta} \log L(\theta | X) \]
\[ \sqrt{-E_\theta \left( \frac{\partial^2}{\partial \theta^2} \log L(\theta | X) \right)} \rightarrow n(0,1). \]

So we can construct the set:
\[ \{ \theta : |Q(x | \theta)| \leq z_{\alpha/2} \} . \]

**Example 10.4.2 (Binomial score interval)** Consider the sufficient statistic \( Y = \sum_{i=1}^{n} X_i \), we have

\[ Q(Y | p) = \frac{\partial}{\partial \theta} \log L(p | Y) \]
\[ \sqrt{-E_\theta \left( \frac{\partial^2}{\partial \theta^2} \log L(p | Y) \right)} = \frac{\hat{p} - p}{\sqrt{p(1-p) / n}} (\hat{p} = y / n) . \]

So we have
\[ \{ p : \frac{\hat{p} - p}{\sqrt{p(1-p) / n}} \leq z_{\alpha/2} \} . \]

In addition, consider the Wald Test, we have:
\[ \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p}) / n}} \rightarrow n(0,1). \]

So we can get:
\{ p : \left| \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p}) / n}} \right| \leq z_{a/2} \}.