Section 6.3 Inferences Based on the MLE

Tonglin Zhang
6.3.1 Standard Errors, Bias, and Consistency

Assume one considers an estimator $\tilde{\theta}$ for an unknown parameter $\theta \in \mathbb{R}$. Let $\hat{\theta}$ be the MLE of $\theta$.

- (MSE). The mean-squared error (MSE) of $\tilde{\theta}$ is

$$\text{MSE}_\theta(\tilde{\theta}) = \mathbb{E}_\theta (\tilde{\theta} - \theta)^2,$$

which is a function of $\theta$.

- There is

$$\text{MSE}_\theta(\tilde{\theta}) = \text{Var}_\theta (\tilde{\theta}) + [\mathbb{E}_\theta (\tilde{\theta}) - \theta]^2,$$

where $\mathbb{E}_\theta (\tilde{\theta}) - \theta$ is called the bias.
If \( E(\tilde{\theta}) = \theta \), then \( \tilde{\theta} \) is an unbiased estimator. There is

\[
\text{MSE}_\theta(\tilde{\theta}) = V_\theta(\tilde{\theta}).
\]

The standard error of \( \tilde{\theta} \) is the estimator of the variance of \( \tilde{\theta} \), given by \( \{\text{Var}_\theta(\tilde{\theta})\}^{1/2} \). Comparing the standard deviation, which is \( \{\text{Var}_\theta(\tilde{\theta})\}^{1/2} \), the standard error is an estimator of the standard deviation.
Assume there are two estimators \( \tilde{\theta}_1 \) and \( \tilde{\theta}_2 \). We say \( \tilde{\theta}_1 \) is not worse than \( \tilde{\theta}_2 \) if

\[
\text{MSE}_\theta(\tilde{\theta}_1) \leq \text{MSE}_\theta(\tilde{\theta}_2).
\]

We say \( \tilde{\theta}_1 \) is better than \( \tilde{\theta}_2 \) if

\[
\text{MSE}_\theta(\tilde{\theta}_1) \leq \text{MSE}_\theta(\tilde{\theta}_2)
\]

for all \( \theta \) and

\[
\text{MSE}_\theta(\tilde{\theta}_1) < \text{MSE}_\theta(\tilde{\theta}_2)
\]

for some \( \theta \).
Example 6.3.1. Let $X_1, \cdots, X_n \sim iid N(\mu, \sigma_0^2)$, where $\sigma_0^2$ is known but $\mu$ is not. Find the MLE of $\theta$ as well as its bias, standard deviation, the standard error.

**Solution.** The MLE of $\theta$ is $\bar{X}$. By

$$\bar{X} \sim N(\mu, \frac{\sigma_0^2}{n}),$$

we have

$$\text{Bias}(\bar{X}) = E(\bar{X}) - \mu = 0.$$

Thus, it is unbiased. The standard deviation is

$$\sigma_{\hat{\mu}} = \sigma_{\bar{X}} = \sqrt{\frac{\sigma_0^2}{n}} = \frac{\sigma_0}{\sqrt{n}}.$$

Because $\sigma_0^2$ is known, the standard deviation is known. Thus, the standard error is also $\sigma_0/\sqrt{n}$. 
Examples 6.3.4 and 6.3.5. Let $X_1, \cdots, X_n \sim \text{iid } \mathcal{N}(\mu, \sigma^2)$, where both $\mu$ and $\sigma^2$ are known. Find the MLE of $\mu$ as well as its bias, standard deviation, and standard error. Find the MLE of $\sigma^2$ as well as its bias. If we observe $\bar{x} = 64.517$ and $s = 2.379$ with $n = 30$, what are the standard errors of $\hat{\mu}$ and the bias of $\hat{\sigma}^2$.

Solution: The MLE of $\mu$ is

$$\hat{\mu} = \bar{X}.$$ 

The MLE is $\sigma^2$ is

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

From the previous problem, we have $\text{Bias}(\bar{X}) = 0$ and its standard deviation is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$
6.3.1 Standard Errors, Bias, and Consistency

Because $\sigma^2$ is unknown, we need an estimator of $\sigma^2$ in the computation of its standard error. We have two options. If we choose $\hat{\sigma}^2$, then the standard error is

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

If we choose

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

then the standard error is

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

In many software packages, the second is adopted.
6.3.1 Standard Errors, Bias, and Consistency

For $\sigma^2$, using $E(S^2) = \sigma^2$, we have

$$\text{Bias}(\hat{\sigma}^2) - \sigma^2 = E(\hat{\sigma}^2) - \sigma^2 = E\left(\frac{n-1}{n}S^2\right) - \sigma^2 = -\frac{\sigma^2}{n}.$$ 

Put the data inside. The standard error of $\hat{\mu}$ is

$$\hat{\sigma_\mu} = \begin{cases} 0.4270 & \text{using } \hat{\sigma}^2 \text{ for } \sigma^2 \\ 0.4344 & \text{using } S^2 \text{ for } \sigma^2 \end{cases}$$
Examples 6.3.2 and 6.3.3. Let $X_1, \cdots, X_n$ be iid $Bernoulli(\theta)$. Find the MLE of $\theta$ as well as its bias, standard deviation, and standard error. If $n = 1000$ and $\sum_{i=1}^{n} x_i = 790$, then what are those answers.

Solution: The MLE of $\theta$ is

$$\hat{\theta} = \bar{X} = \frac{790}{1000} = 0.79.$$ 

The bias is $\text{Bias}(\hat{\theta}) = \mathbb{E}(\bar{X}) - \theta = 0$. Thus, it is unbiased. Its standard deviation is

$$\sqrt{\text{V}(\bar{X})} = \frac{\theta(1 - \theta)}{n}.$$ 

Based on the data, the standard error is

$$\frac{\hat{\theta}(1 - \hat{\theta})}{1000} = 0.0129.$$
Example: Assume $X_1, \cdots, X_n \sim N(\theta, 1)$. Let $\tilde{\theta} = \sum_{i=1}^m X_i / m$, where $m < n$. Justify why $\tilde{\theta}$ is worse than $\hat{\theta} = \sum_{i=1}^n X_i / n$.

Solution: We need to compare their MSE. The one with lower MSE is better.

$$
\text{MSE}(\tilde{\theta}) = \text{MSE} \left( \frac{1}{m} \sum_{i=1}^m X_i \right)
$$

$$
= \text{V} \left( \frac{1}{m} \sum_{i=1}^m X_i \right) + \text{Bias}^2 \left( \frac{1}{m} \sum_{i=1}^m X_i \right)
$$

$$
= \text{V} \left( \frac{1}{m} \sum_{i=1}^m X_i \right) = \frac{\sigma^2}{m}
$$

Similarly, we have $\text{MSE}(\hat{\theta}) = \sigma^2 / n$. By $m < n$, we have

$$
\text{MSE}(\hat{\theta}) < \text{MSE}(\tilde{\theta})
$$

Thus, $\hat{\theta}$ is better.
Consistency of Estimators

An estimator \( \hat{\theta} \) of \( \theta \) is consistent if \( \hat{\theta} \overset{P}{\to} \theta \) for every \( \theta \).

- Consistency is the minimum requirement of estimators: if an estimator is inconsistent, then we say it cannot be used.
- Q: Why \( \hat{\theta} \) equal to a constant cannot be used. A: Because it is not consistent.
Consistency in Previous Examples

**Example 6.3.1.** $E(\bar{X}) = \mu$ and $V(\bar{X}) = \sigma_0^2/n$. By the Chebyshev inequality, we have

$$P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma_0^2}{n\epsilon} \to 0$$

for any $\epsilon > 0$. Thus, $\bar{X} \xrightarrow{P} \mu$. It is consistent.
Example 6.3.4 and 6.3.5. Using the same method, $\bar{X} \xrightarrow{P} \mu$. For $\sigma^2$, we need to use the property

$$(n - 1)S^2 \sim \sigma^2 \chi^2_{n-1} \Rightarrow S^2 \sim \frac{\sigma^2}{n - 1} \chi^2_{n-1},$$

By $E(\chi^2_{n-1}) = n - 1$ and $V(\chi^2_{n-1}) = 2(n - 1)$, we have

$$E(S^2) = E\left(\frac{\sigma^2}{n - 1} \chi^2_{n-1}\right) = \frac{\sigma^2}{n - 1} E(\chi^2_{n-1}) = \sigma^2$$

and

$$V(S^2) = V\left(\frac{\sigma^2}{n - 1} \chi^2_{n-1}\right) = \frac{\sigma^4}{(n - 1)} V(\chi^2_{n-1}) = \frac{2\sigma^4}{n - 1}. $$

Thus, $S^2$ is consistent.
6.3.1 Standard Errors, Bias, and Consistency

For the MLE, we have

$$E(\hat{\sigma}^2) = \frac{n - 1}{n} \sigma^2$$

which goes to $\sigma^2$ as $n \to \infty$. We can also show $V(\hat{\sigma}^2) \to 0$. Thus, it is also consistent.

**Example 6.3.2 and 6.3.3.** Do it by yourself.
6.3.2 Confidence Intervals

Let $l = l(data)$ and $u = u(data)$ be two statistics, always satisfying $l < u$. We say $[l, u]$ is a $\gamma$-level confidence interval or $\gamma$-confidence interval for $\theta$ if

$$P_{\theta}(l \leq \theta \leq u) \geq \gamma = 1 - \alpha$$

for every $\theta$. We refer to $\gamma$ as the confidence level of the interval.

Sometimes, we use percentage to expression the confidence level. If we choose $\gamma = 0.95$ ($\alpha = 0.05$), then it is a 95% confidence interval satisfying

$$P_{\theta}(l \leq \theta \leq u) \geq 95\%.$$
Understanding the confidence interval and the confidence level:

(a) Collecting data many times \((n \text{ times})\);
(b) calculate the interval \([l, u]\);
(c) the proportion of \(\theta \in [l, u]\) is approximately greater than or equal to \(\gamma\);
(d) it becomes exactly greater than or equal to \(\gamma\) if \(n \to \infty\).
**Examples**

**Example 6.3.6.** Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2_0)$, where $\sigma^2_0$ is known. We use

$$\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \sim N(0, 1)$$

to find the confidence interval. The result is

$$[\bar{X} - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}],$$

where $z_{alpha/2}$ is the upper quantile (inverse CDF) of $N(0, 1)$. For example, if $\alpha = 0.05$, then we need $z_{0.025}$, leading to the 95% confidence interval for $\mu$ as

$$[\bar{X} - 1.96 \frac{\sigma_0}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma_0}{\sqrt{n}}].$$
6.3.2 Confidence Intervals

**Notations**

- **Definition:** $z_\alpha$ satisfies

$$P(N(0, 1) > z_\alpha) = \alpha \Rightarrow \Phi(z_\alpha) = P(N(0, 1) \leq z_\alpha) = 1 - \alpha.$$

- **Commonly used values:** the following table is generally enough

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
<th>0.001</th>
<th>0.0005</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF</td>
<td>0.9</td>
<td>0.95</td>
<td>0.975</td>
<td>0.99</td>
<td>0.995</td>
<td>0.999</td>
<td>0.9995</td>
</tr>
<tr>
<td>$z_\alpha$</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
<td>3.090</td>
<td>3.291</td>
</tr>
</tbody>
</table>
6.3.2 Confidence Intervals

To understand the concept, we need to do a simulated experiment.

- Assume \( n = 30 \) and \( \sigma_0 = 1 \).
- Collect 30 data points and compute the confidence interval.
- Check whether \( \theta \in [l, u] \).
- Repeat the entire procedure and look at the proportion for the correct confidence intervals.
- Plot it for \( \theta \).
Figure: Coverage Probabilities of the 95% $z$-confidence intervals for $\mu$ in the normal example.
Example 6.3.7. Let $X_1, \cdots, X_n$ be iid $Bernoulli(\theta)$. Then, a $\gamma$-level confidence interval for $\theta$ is

$$[\bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}, \bar{X} + z_{\alpha/2} \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}].$$

- Use $\sum_{i=1}^{1000} x_i = 790$ with $n = 1000$ (Example 6.3.3), we can calculate the confidence interval.
- We also need to do a simulated experiment to understand it.
Example 6.3.7: $z$-confidence interval

Figure: Coverage Probabilities of the 95% $z$-confidence intervals for $\theta$ in the binomial example.
t-Confidence Intervals
Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, where both $\mu$ and $\sigma^2$ are unknown. Then,

- $\bar{X}$ is the MLE of $\mu$;
- $\hat{\sigma}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / n$ is the MLE of $\sigma^2$;
- $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n - 1)$ is the UMVUE (uniform minimum unbiased estimator, not to be taught) of $\sigma^2$;
- $\bar{X}$ and $S^2$ are independent;
- $\bar{X} \sim N(\mu, \sigma^2 / n)$;
- $(n - 1)S^2 \sim \sigma^2 \chi^2_{n-1}$. 
6.3.2 Confidence Intervals

Therefore

\[ T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}, \]

which provides a $\gamma$-level $t$-confidence interval for $\mu$ as

\[ [\bar{X} - t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}], \]

where $t_{\alpha/2,n-1}$ is the quantile values of $t_{n-1}$ distribution. If $\alpha = 0.05$, then the 95%-confidence interval for $\mu$ is

\[ [\bar{X} - t_{0.025,n-1} \frac{S}{\sqrt{n}}, \bar{X} - t_{0.025,n-1} \frac{S}{\sqrt{n}}]. \]
In Example 6.3.5, we have $n = 30$, $\bar{x} = 64.517$, and $s/\sqrt{30} = 0.43434$, then $t_{0.975,29} = 2.0452$, implying that

$$64.517 \pm 2.0452(0.043434) = [63.629, 65.405].$$

To understand the $t$-confidence interval, we also need a simulated example.

- Assume $n = 30$.
- Collect 30 data. Compute the $t$-confidence interval.
- Look at the proportion of the interval which contains the true value of $\mu$. 
6.3.2 Confidence Intervals

Example 6.3.5: $t$-confidence interval

Figure: Coverage Probabilities of the 95\% $t$-confidence intervals for $\mu$. 
Testing hypotheses is an important statistical problem. It concerns whether a statement is correct or not. It contains the following items.

- A statement.
- Null hypothesis $H_0$: the statement is true; and the alternative hypothesis $H_1$ ($H_a$, or $H_A$): the statement is false.
- A test statistic $T$.
- Rejection region $C$: if $T \in C$, then conclude $H_1$.
- Q: how to define $T$ and $C$. 

Tonglin Zhang, Department of Statistics, Purdue University 

6.3 Inferences Based on the MLE
Type I error, Type II error, significance level, power function, and P-value:
I decide to move Sections 8.2.1 and 8.2.2 here.

Since the decision can only be made based on data, one cannot guarantee that the decision is always consistent with the truth. Therefore, we propose two types of errors based on the following table.

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>Truth</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>Correct</td>
<td>Type II error</td>
</tr>
<tr>
<td>False</td>
<td>Type I error</td>
<td>Correct</td>
</tr>
</tbody>
</table>
Type I error probability is

\[ P(\text{Conclude } H_1 | H_0). \]

Type II error probability is

\[ P(\text{Conclude } H_0 | H_1). \]

The significance level is

\[ \alpha = \max \{ \text{Type I error probabilities} \}. \]

The power function is

\[ P(\text{Conclude } H_1). \]

The p-value is the largest \( \alpha \) which can reject \( H_0 \). Therefore, if the p-value is larger than \( \alpha \), we conclude \( H_0 \); otherwise, we conclude \( H_1 \).
Examples:

Examples 6.3.9 and 6.3.10. Let $X_1, \cdots, X_n \sim N(\mu, \sigma_0^2)$, where $\sigma_0^2$ is unknown. Assume we want to know whether $\mu = \mu_0$, where $\mu_0$ is a preselected number. Then,

- Statement: $\mu = \mu_0$.
- Null hypothesis $H_0 : \mu = \mu_0$; alternative hypothesis: $H_1 : \mu \neq \mu_0$.
- Test statistic: $\bar{X}$.
- Rejection region (either $\bar{X}$ is too large or $\bar{X}$ is too small):

$$C = \{ | \frac{\bar{X} - \mu_0}{\sigma_0/\sqrt{n}} | > a \} = \{ \bar{X} > \mu_0 + \frac{a\sigma_0}{\sqrt{n}} \text{ or } \bar{X} < \mu_0 - \frac{a\sigma_0}{\sqrt{n}} \}$$

where $a$ is a value to be determined.
Type I error probabilities:

\[ P(\text{Conclude } H_1 | H_0) \]
\[ = P(\bar{X} > \mu_0 + \frac{a\sigma_0}{\sqrt{n}} \text{ or } \bar{X} < \mu_0 - \frac{a\sigma_0}{\sqrt{n}} | \mu = \mu_0) \]
\[ = P(\bar{X} > \mu_0 + \frac{a\sigma_0}{\sqrt{n}} | \mu = \mu_0) + P(\bar{X} < \mu_0 - \frac{a\sigma_0}{\sqrt{n}} | \mu = \mu_0) \]
\[ = [1 - \Phi(a)] + \Phi(-a) \]
\[ = 2\Phi(-a), \]

where \( a > 0 \).

The significance level is \( 2\Phi(-a) \).
### 6.3.3 Testing Hypothesis and P-values

- **Type II error probabilities:**

\[
P(\text{Conclude } H_0 | H_1) = P\left( \mu_0 \frac{a\sigma_0}{\sqrt{n}} \leq \bar{X} \leq \mu_0 + \frac{a\sigma_0}{\sqrt{n}} | \mu \neq \mu_0 \right) \\
= \Phi\left( a + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} \right) - \Phi\left( a - \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} \right).
\]
6.3.3 Testing Hypothesis and P-values

The power function is

\[
P(\text{Conclude } H_1) = P(\bar{X} \geq \mu_0 + \frac{a\sigma_0}{\sqrt{n}} \text{ or } \bar{X} < \mu_0 - \frac{a\sigma_0}{\sqrt{n}})
\]

\[
= P(\bar{X} > \mu_0 + \frac{a\sigma_0}{\sqrt{n}}) + P(\bar{X} < \mu_0 - \frac{a\sigma_0}{\sqrt{n}})
\]

\[
= \left[1 - \Phi(a + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0})\right] + \Phi\left(a - \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0}\right).
\]
To understand the $P$-value, we need to change $a$ such that we can just conclude $H_1$. Then, we have

$$a_0 = \left| \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} \right|.$$ 

implying that the $P$-value is

$$p = 2\Phi \left( -\left| \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} \right| \right).$$
Using data of Example 6.3.10, we have $\sigma_0^2 = 4$ and $\mu = 26$. Suppose we want to know whether $\mu = 25$. Then, we choose $\mu_0 = 25$. From the data, we have $\bar{x} = 26.6808$ and $n = 10$. Therefore the P-value is

$$2\Phi \left( -\left| \frac{26.6808 - 25}{2\sqrt{10}} \right| \right) = 2\Phi(-2.6576) = 0.0078.$$  

If we choose $\alpha < 0.0078$, then we conclude $H_0$; otherwise, we conclude $H_1$. Therefore, 0.0078 is the largest significance value for us to conclude $H_1$. 

Tonglin Zhang, Department of Statistics, Purdue University
Example 6.3.11. Let \( X_1, \cdots, X_n \) be an iid sample from \( \text{Bernoulli}(\theta) \). Suppose we want to test \( H_0 : \theta = \theta_0 \). Let \( T = \sum_{i=1}^{n} X_i \). Then, \( T \sim \text{Bin}(n, \theta) \). Then, we reject \( H_0 \) if \( T \leq a \) or \( T \geq b \) for some \( a < b \). Therefore, the rejection region is

\[
\{ T \leq a \text{ or } T \geq b \}. 
\]

Then, the hypotheses are

\( H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0 \).
Type I error probability

\[ P(T \leq a \text{ or } T \geq b | \theta = \theta_0) = P(Bin(n, \theta_0) \leq a) + [1 - P(Bin(n, \theta_0) \geq b)] \]

\[ \approx [1 - \Phi\left(\frac{b - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}}\right)] + \Phi\left(\frac{a - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}}\right). \]

It is also the significance level since \( H_0 \) only contains has one value.
Type II error probability

\[
P(a < T < b|\theta \neq \theta_0) = P(a < Bin(n, \theta) < b) \\
\approx \Phi\left(\frac{b - n\theta}{\sqrt{n\theta(1 - \theta)}}\right) - \Phi\left(\frac{a - n\theta}{\sqrt{n\theta(1 - \theta)}}\right),
\]

where \( \theta \neq \theta_0 \).
We often choose $a$ and $b$ symmetric about $\theta_0$. This is based on the approximation

\[
\frac{T - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}} \sim \text{approx } N(0, 1).
\]

Then, the $p$-value is about

\[
2\Phi\left(-\left|\frac{T - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}}\right|\right).
\]

Suppose we want to test $H_0 : \theta = 1/2$ with $n = 100$. If we observe $T = 54$, then the $P$-value is

\[
2\Phi\left(-\left|\frac{54 - 50}{\sqrt{100(0.5)(0.5)}}\right|\right) = 2\Phi(-0.8) = 0.4238.
\]
Consistency of a test: what is statistical significance practically significant?

We want both the Type I error probabilities and the Type II error probabilities small. Usually there is

\[
\text{max}\{\text{Type I error probabilities}\} + \text{max}\{\text{Type II error probabilities}\} = 1.
\]

Because the first term is uniformly controlled by \( \alpha \) (the significance level), the second term is out-of-control. Therefore, type II error probabilities are often considered at individual points.
6.3.3 Testing Hypothesis and P-values

In the normal case, for any $\mu = \mu_1 \neq \mu_0$, the Type II error probability is

$$\Phi\left(a + \frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma_0}\right) - \Phi\left(a - \frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma_0}\right), \ a > 0.$$ 

If $\mu_1 > \mu_0$, then $\sqrt{n}(\mu_0 - \mu_1) \to -\infty$ implying that the about limit is 0 as $n \to 0$. If $\mu_1 < \mu_0$, then $\sqrt{n}(\mu_0 - \mu_1) \to \infty$ also implying that the above limit is 0. Therefore, the Type II probability goes to 0 at the individual level as $n \to \infty$. 
Hypothesis assessment via confidence intervals

Theoretically, the confidence interval problem is equivalent to the (two-sided) testing problem. If want to test $H_0 : \theta = \theta_0$ at 0.05 significance level, then we can compute the 95% confidence interval for $\theta$. We conclude $H_0 : \theta = \theta_0$ if and only if the confidence interval contains $\theta_0$. We can use Example 6.3.12 to understand such an issue.
t-Tests

Let $X_1, \cdots, X_n$ be iid $N(\mu, \sigma^2)$. Then, $\bar{X} \sim N(\mu, \sigma^2 / n)$ and $(n-1)S^2 \sim \sigma^2 \chi^2_{n-1}$ independently. Then,

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

and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$ 

Then,

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$
Therefore, the t-test at $\alpha$ significance level rejects $H_0 : \mu = \mu_0$ if

$$\left| \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \right| \geq t_{\alpha/2}.$$ 

In Example 6.3.10, we obtain $n = 10$, $\bar{x} = 26.6808$ and $s = 2.2050$. For $H_0 : \mu = 25$ against $H_1 : \mu \neq 25$, we obtain the $t$-statistic value as

$$|t| = \left| \frac{26.6808 - 25}{2.2050/\sqrt{10}} \right| = 2.4105 > t_{0.025,9} = 2.2622.$$ 

Thus, we reject $H_0$. 

Tonglin Zhang, Department of Statistics, Purdue University  
6.3 Inferences Based on the MLE
One-Side Tests

Example 6.3.12: Normal distribution with known variances. Let \( X_1, \ldots, X_n \sim iid N(\mu, \sigma^2_0) \). Suppose we want to test

\[ H_0 : \mu \leq \mu_0 \leftrightarrow \mu > \mu_0. \]

- The rejection region is

\[ C = \{ \bar{X} \geq a \} \]

for some \( a > 0 \).
- The type I error probability is

\[ P(\bar{X} \geq a | \mu \leq \mu_0) = 1 - \Phi\left( \frac{a - \mu}{\sigma_0 / \sqrt{n}}, \mu \leq \mu_0, \right) \]

which is increasing in \( \mu \).
The type II error probability is

\[
P(\bar{X} < a | \mu \geq \mu_0) = \Phi\left(\frac{a - \mu}{\sigma_0/\sqrt{n}}\right), \mu > \mu_0,
\]
which is decreasing in \( \mu \).

The power function is

\[
P(\bar{X} \geq a) = 1 - \Phi\left(\frac{a - \mu}{\sigma_0/\sqrt{n}}\right), \mu \in \mathbb{R}.
\]

It is the type I error probability if \( \mu \leq \mu_0 \) (i.e., \( H_0 \) holds) or one minus the type II error probability if \( \mu > \mu_0 \) (i.e., \( H_1 \) holds). The significance level is

\[
\alpha = \max_{\mu \leq \mu_0} [1 - \Phi\left(\frac{a - \mu}{\sigma_0/\sqrt{n}}\right)] = [1 - \Phi\left(\frac{a - \mu_0}{\sigma_0/\sqrt{n}}\right)].
\]
6.3.3 Testing Hypothesis and P-values

▶ If we want to control it by not over 0.05, then we need to select \( a \) such that

\[
[1 - \Phi\left(\frac{a - \mu_0}{\sigma_0/\sqrt{n}}\right)] = 0.05 \Rightarrow a = \mu_0 + 1.645\sigma_0/\sqrt{n}.
\]

▶ If the data set of Example 6.3.10 is used, then the \( p \)-value is

\[
1 - \Phi\left(\frac{26.6808 - 25}{2/\sqrt{10}}\right) = 1 - \Phi(2.6576) = 0.0039.
\]
6.3.3 Testing Hypothesis and P-values

**Figure**: Exact power functions as a function of $\mu$ when $\mu_0 = 2$, $\sigma_0 = 1$, and $n = 10$. 
6.3.3 Testing Hypothesis and P-values

**Figure:** Simulated power functions as a function of $\mu$ when $\mu_0 = 2$, $\sigma_0 = 1$, and $n = 10$. 
Example: Binomial or Bernoulli distribution. Let \( X_1, \cdots, X_n \) be iid \( \text{Bernoulli}(\theta) \). Then \( T = \sum_{i=1}^{n} X_i \sim Bin(n, \theta) \). Suppose we want to test \( H_0 : \theta \leq \theta_0 \) against \( H_1 : \theta > \theta_0 \). We reject \( H_0 \) if \( T \geq a \).

- The rejection region should be
  \[
  C = \{ T \geq a \}.
  \]

- The type I error probability is
  \[
  P(T \geq a | \theta \leq \theta_0) = P(Bin(n, \theta) \geq a) = 1 - P(Bin(n, \theta) \leq a - 1),
  \]
  for \( \theta \leq \theta_0 \).

- The type II error probability is
  \[
  P(T < a | \theta > \theta_0) = P(Bin(n, \theta) \leq a - 1), \theta \geq \theta_0.
  \]
6.3.3 Testing Hypothesis and P-values

- The power function is

\[ P(T \geq a) = 1 - P(Bin(n, \theta) \leq a - 1), \theta \in (0, 1). \]

- The significance level is

\[ \alpha = \max_{\theta \leq \theta_0} [1 - P(Bin(n, \theta) \leq a - 1)] = 1 - P(Bin(n, \theta_0) \leq a - 1). \]

If \( \alpha = 0.05 \) and \( \theta_0 = 0.05 \) are chosen, then we can find \( a \) by choosing the minimum \( a \) satisfying \( \alpha \leq 0.05 \).

- We have the following table.

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>9</td>
<td>12</td>
<td>15</td>
<td>18</td>
<td>20</td>
<td>26</td>
<td>32</td>
<td>59</td>
</tr>
</tbody>
</table>

If \( n = 100 \) and \( T = 60 \), the \( p \)-value is

\[ P(Bin(100, 0.5) \geq 60) = 0.02844. \]
6.3.3 Testing Hypothesis and P-values

Figure: Simulated power functions as a function of $\theta$ when $\theta_0 = 0.5$. 
6.3.4 Inferences for the Variance

Let \( X_1, \ldots, X_n \sim N(\mu, \sigma^2) \). Then,

\[
(n - 1)S^2/\sigma^2 \sim \chi^2_{n-1}.
\]

Then, we can find \( a \) and \( b \) such that

\[
P(\chi^2_{n-1} \leq a) = P(\chi^2_{n-1} \leq b) = \alpha/2.
\]

This provides the \( \gamma \)-level confidence interval as

\[
\left[ \frac{(n - 1)S^2}{\chi^2_{\alpha/2}, n - 1}, \frac{(n - 1)S^2}{\chi^2_{\alpha/2, n-1}} \right].
\]

In example 6.3.10, we have \( s^2 = 4.8620 \) and \( n = 10 \). Then, the 95\% confidence interval for \( \sigma^2 \) is

\[
\left[ \frac{9(4.8620)}{19.023}, \frac{9(4.8620)}{2.700} \right] = [2.3002, 16.207].
\]
6.3.5 Sample-Size Calculations: Confidence Intervals

- We need to determine the number of observations before data collection.
- We cannot use data information in sample-size calculation problems.
- To reduce the cost, we want to make sample size as small as possible.
- We also want to have enough data to draw conclusions.
- Therefore, there is a trade-off.
Example 6.3.16. Note that the length of confidence interval is \(2z_{\alpha/2}\sigma_0/\sqrt{n}\). We can obtain the minimum \(n\) such that the length is less than \(2\delta\), a preselected value. Then, we have

\[
\frac{2z_{\alpha/2}\sigma_0}{\sqrt{n}} \leq 2\delta \Rightarrow n \geq \sigma_0^2 \left(\frac{z_{\alpha/2}}{\delta}\right)^2.
\]

If \(\sigma_0^2 = 10\), \(\gamma = 0.95\), \(\delta = 0.5\), then we want the 95% confidence interval not over 1, leading

\[
n \geq 10(1.96/0.5)^2 = 153.6 \Rightarrow n = 164.
\]
However, if $\sigma^2$ is unknown, then it becomes

$$n \geq s^2 \left( \frac{t_{\alpha/2,n-1}}{\delta} \right)^2,$$

which depends on the sample. Therefore, the method cannot be used. This is still a research problem today.
Example 6.3.17. Let $X_1, \cdots, X_n$ be iid $Bernoulli(\theta)$. Then $T = \sum_{i=1}^{n} \sim Bin(n, \theta)$. The approximate confidence interval is

$$\bar{x} \pm z_{\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}.$$ 

If the length not over $2\delta$, then we need

$$n \geq \bar{x}(1-\bar{x})\left(\frac{z_{\alpha/2}}{\delta}\right)^2.$$ 

Using

$$\bar{x}(1-\bar{x}) \leq 1/4,$$

we can choose

$$n \geq \frac{1}{4}\left(\frac{z_{\alpha/2}}{\delta}\right)^2.$$
This choice guarantees the length of the confidence interval not over $2\delta$. If $\gamma = 0.95$ ($\alpha = 0.05$) and $\delta = 0.1$ (the length not over 0.2), then we need

$$n \geq \frac{1}{4} \left( \frac{1.96}{0.1} \right)^2 = 96.04$$

and we choose $n = 97$. If $\delta = 0.01$ (the length not over 0.02), then

$$n \geq \frac{1}{4} \left( \frac{1.96}{0.01} \right)^2 = 9604.$$
We can only control type II error probabilities at the individual level.

The type II error probability equals one minus the power function.

We need to make the power function as large as possible. This is called the more powerful way.
Example 6.3.18. Let $X_1, \cdots, X_n \sim N(\mu, \sigma_0^2)$ with a known $\sigma_0^2$. If we test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ at $\alpha$ significance level, then the power function

$$\beta(\mu) = 1 - \Phi\left(\frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} + z_{\alpha/2}\right) + \Phi\left(\frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} - z_{\alpha/2}\right)$$

$$= \Phi\left(-\frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} - z_{\alpha/2}\right) + \Phi\left(\frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} - z_{\alpha/2}\right)$$
Rather than the method introduce by the book, I decide to introduce another method. If $\mu > \mu_0$ (same for the case when $\mu < \mu_0$), then the first term is larger. There is

$$
\beta(\mu) \leq \Phi\left(-\frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} - z_{\alpha/2}\right).
$$

If a preselected $\beta$ is chosen and we solve $n$ by

$$
\beta_0 = \Phi\left(-\frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} - z_{\alpha/2}\right) \Rightarrow n \geq \sigma_0^2 \left(\frac{z_{\beta_0} + z_{1-\alpha/2}}{\mu_0 - \mu}\right)^2.
$$

This guarantees that $\beta(\mu) \leq \beta_0$ at $\mu$. 
Example. Let $X_1, \ldots, X_n \sim N(\mu, \sigma_0^2)$ with a known $\sigma_0^2$. Consider the test for

$$H_0 : \mu \leq \mu_0 \leftrightarrow \mu > \mu_0.$$ 

Assume we reject $H_0$ if

$$\bar{X} \geq \mu_0 + \frac{z_\alpha \sigma_0}{\sqrt{n}}.$$ 

For any $\mu > \mu_0$, the type II error probability is

$$\beta(\mu) = P(\bar{X} \leq \mu_0 + z_\alpha \frac{\sigma_0}{\sqrt{n}} | \mu > \mu_0)$$

$$= \Phi\left( \frac{\mu_0 + z_\alpha \sigma_0 / \sqrt{n} - \mu}{\sigma_0 / \sqrt{n}} \right)$$

$$= \Phi\left( z_\alpha + \frac{\sqrt{n} (\mu_0 - \mu)}{\sigma_0} \right).$$
If we want $\beta(\mu) \leq \beta_0$, then we need

$$\Phi(z_\alpha + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0}) \leq \beta_0$$

$$\Rightarrow z_\alpha + \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} \leq -z_{\beta_0}$$

$$\Rightarrow n \geq \sigma^2_0 \left(\frac{z_{\beta_0} + z_\alpha}{\mu_0 - \mu}\right)^2.$$ 

For instance, assume $\mu_0 = 0$, $\sigma_0 = 2$, and $\mu = 0.1$. If $\alpha = 0.05$, then $z_\alpha = z_{0.05} = 1.645$. If we want $\beta(0.1) \geq 0.01$, then $z_{\beta_0} = z_{0.01} = 2.33$. Thus,

$$n \geq \sigma^2_0 \left(\frac{1.645 + 2.33}{0.1}\right)^2 = 6320.25 \Rightarrow n = 6321.$$
Example 6.3.19. Binomial case, not analytically solvable, but we can numerically derive the result.

Example 6.3.20. If $\sigma^2$ is unknown, there is not a clear way to find $n$. This is a research problem.