Lecture 22: Statistical inference

A major approach in statistical analysis does not use any loss-risk.

Three components in statistical inference

- Point estimators (Chapters 3-5)
- Hypothesis tests (Chapter 6)
- Confidence sets (Chapter 7)

Point estimators

Let $T(X)$ be an estimator of $\vartheta \in \mathbb{R}$

Bias: $b_T(P) = E[T(X)] - \vartheta$

Mean squared error (mse):

$$\text{mse}_T(P) = E[(T(X) - \vartheta)^2] = [b_T(P)]^2 + \text{Var}(T(X)).$$

Bias and mse are two common criteria for the performance of point estimators.

Read Example 2.26
Lecture 22: Statistical inference

A major approach in statistical analysis does not use any loss-risk.

Three components in statistical inference

- Point estimators (Chapters 3-5)
- Hypothesis tests (Chapter 6)
- Confidence sets (Chapter 7)

Point estimators

Let \( T(X) \) be an estimator of \( \vartheta \in \mathbb{R} \)

Bias: \( b_T(P) = E[T(X)] - \vartheta \)

Mean squared error (mse):

\[
mse_T(P) = E[T(X) - \vartheta]^2 = [b_T(P)]^2 + \text{Var}(T(X)).
\]

Bias and mse are two common criteria for the performance of point estimators.

Read Example 2.26
Hypothesis tests

To test the hypotheses

\[ H_0 : P \in \mathcal{P}_0 \quad \text{versus} \quad H_1 : P \in \mathcal{P}_1, \]

there are two types of errors we may commit:

- rejecting \( H_0 \) when \( H_0 \) is true (called the \textit{type I error})
- and accepting \( H_0 \) when \( H_0 \) is wrong (called the \textit{type II error}).

A test \( T \): a statistic from \( \mathcal{X} \) to \( \{0, 1\} \).

Probabilities of making two types of errors

Type I error rate:

\[ \alpha_T(P) = P(T(X) = 1) \quad P \in \mathcal{P}_0 \]

Type II error rate:

\[ 1 - \alpha_T(P) = P(T(X) = 0) \quad P \in \mathcal{P}_1, \]

\( \alpha_T(P) \) is also called the power function of \( T \). Power function is \( \alpha_T(\theta) \) if \( P \) is in a parametric family indexed by \( \theta \).
Hypothesis tests

To test the hypotheses

\[ H_0 : P \in \mathcal{P}_0 \quad \text{versus} \quad H_1 : P \in \mathcal{P}_1, \]

there are two types of errors we may commit:

- rejecting \( H_0 \) when \( H_0 \) is true (called the type I error)
- and accepting \( H_0 \) when \( H_0 \) is wrong (called the type II error).

A test \( T \): a statistic from \( \mathcal{X} \) to \( \{0, 1\} \).

Probabilities of making two types of errors

Type I error rate:

\[ \alpha_T(P) = P(T(X) = 1) \quad P \in \mathcal{P}_0 \]

Type II error rate:

\[ 1 - \alpha_T(P) = P(T(X) = 0) \quad P \in \mathcal{P}_1, \]

\( \alpha_T(P) \) is also called the power function of \( T \)

Power function is \( \alpha_T(\theta) \) if \( P \) is in a parametric family indexed by \( \theta \).
Remarks

- Note that these are risks of $T$ under the 0-1 loss in statistical decision theory.
- Type I and type II error probabilities cannot be minimized simultaneously.
- These two error probabilities cannot be bounded simultaneously by a fixed $\alpha \in (0, 1)$ when we have a sample of a fixed size.

Significance tests

A common approach to finding an “optimal” test is to assign a small bound $\alpha$ to the type I error rate $\alpha_T(P), P \in P_0$, and then to attempt to minimize the type II error rate $1 - \alpha_T(P), P \in P_1$, subject to

$$\sup_{P \in P_0} \alpha_T(P) \leq \alpha. \quad (1)$$

The bound $\alpha$ is called the level of significance. The left-hand side of (1) is called the size of the test $T$. The level of significance should be positive, otherwise no test satisfies (1) except the silly test $T(X) \equiv 0$ a.s. $\mathcal{P}$.
Remarks

- Note that these are risks of $T$ under the 0-1 loss in statistical decision theory.
- Type I and type II error probabilities cannot be minimized simultaneously.
- These two error probabilities cannot be bounded simultaneously by a fixed $\alpha \in (0, 1)$ when we have a sample of a fixed size.

Significance tests

A common approach to finding an “optimal" test is to assign a small bound $\alpha$ to the type I error rate $\alpha_T(P)$, $P \in \mathcal{P}_0$, and then to attempt to minimize the type II error rate $1 - \alpha_T(P)$, $P \in \mathcal{P}_1$, subject to

$$\sup_{P \in \mathcal{P}_0} \alpha_T(P) \leq \alpha.$$  \hspace{1cm} (1)

The bound $\alpha$ is called the level of significance. The left-hand side of (1) is called the size of the test $T$. The level of significance should be positive, otherwise no test satisfies (1) except the silly test $T(X) \equiv 0$ a.s. $\mathcal{P}$. 
Example 2.28

Let $X_1, \ldots, X_n$ be i.i.d. from the $N(\mu, \sigma^2)$ distribution with an unknown $\mu \in \mathbb{R}$ and a known $\sigma^2$.

Consider the hypotheses $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, where $\mu_0$ is a fixed constant.

Since the sample mean $\bar{X}$ is sufficient for $\mu \in \mathbb{R}$, it is reasonable to consider the following class of tests: $T_c(X) = I_{(c, \infty)}(\bar{X})$, i.e., $H_0$ is rejected (accepted) if $\bar{X} > c$ ($\bar{X} \leq c$), where $c \in \mathbb{R}$ is a fixed constant.

Let $\Phi$ be the c.d.f. of $N(0,1)$.

By the property of the normal distributions,

$$\alpha_{T_c}(\mu) = P(T_c(X) = 1) = 1 - \Phi \left( \frac{\sqrt{n}(c - \mu)}{\sigma} \right).$$

Figure 2.2 provides an example of a graph of two types of error probabilities, with $\mu_0 = 0$.

Since $\Phi(t)$ is an increasing function of $t$,

$$\sup_{P \in \mathcal{P}_0} \alpha_{T_c}(\mu) = 1 - \Phi \left( \frac{\sqrt{n}(c - \mu_0)}{\sigma} \right).$$
Example 2.28

Let $X_1, \ldots, X_n$ be i.i.d. from the $N(\mu, \sigma^2)$ distribution with an unknown $\mu \in \mathbb{R}$ and a known $\sigma^2$.

Consider the hypotheses $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, where $\mu_0$ is a fixed constant.

Since the sample mean $\bar{X}$ is sufficient for $\mu \in \mathbb{R}$, it is reasonable to consider the following class of tests: $T_c(X) = l_{(c, \infty)}(\bar{X})$, i.e., $H_0$ is rejected (accepted) if $\bar{X} > c$ ($\bar{X} \leq c$), where $c \in \mathbb{R}$ is a fixed constant.

Let $\Phi$ be the c.d.f. of $N(0, 1)$.

By the property of the normal distributions,

$$\alpha_{T_c}(\mu) = P(T_c(X) = 1) = 1 - \Phi \left( \frac{\sqrt{n}(c - \mu)}{\sigma} \right).$$

Figure 2.2 provides an example of a graph of two types of error probabilities, with $\mu_0 = 0$.

Since $\Phi(t)$ is an increasing function of $t$,

$$\sup_{\mathcal{P}_0} \alpha_{T_c}(\mu) = 1 - \Phi \left( \frac{\sqrt{n}(c - \mu_0)}{\sigma} \right).$$
In fact, it is also true that

$$\sup_{P \in \mathcal{P}_1} [1 - \alpha_{T_c}(\mu)] = \Phi \left( \frac{\sqrt{n}(c - \mu_0)}{\sigma} \right).$$

If we would like to use an $\alpha$ as the level of significance, then the most effective way is to choose a $c_\alpha$ (a test $T_{c_{\alpha}}(X)$) such that

$$\alpha = \sup_{P \in \mathcal{P}_0} \alpha_{T_{c_{\alpha}}}(\mu),$$

in which case $c_\alpha$ must satisfy

$$1 - \Phi \left( \frac{\sqrt{n}(c_\alpha - \mu_0)}{\sigma} \right) = \alpha,$$

i.e., $c_\alpha = \sigma z_{1-\alpha}/\sqrt{n} + \mu_0$, where $z_a = \Phi^{-1}(a)$.

In Chapter 6, it is shown that for any test $T(X)$ satisfying $\sup_{P \in \mathcal{P}_0} \alpha_T(P) \leq \alpha$,

$$1 - \alpha_T(\mu) \geq 1 - \alpha_{T_{c_{\alpha}}}(\mu), \quad \mu > \mu_0.$$
Choice of significance level

The choice of a level of significance \( \alpha \) is usually somewhat subjective. In most applications there is no precise limit to the size of \( T \) that can be tolerated. Standard values, 0.10, 0.05, and 0.01, are often used for convenience. For most tests satisfying \( \sup_{P \in \mathcal{P}_0} \alpha_T(P) \leq \alpha \), a small \( \alpha \) leads to a “small” rejection region.

\[ \hat{\alpha} = \inf \{ \alpha \in (0, 1) : T_\alpha(x) = 1 \} \]

Such an \( \hat{\alpha} \), which depends on \( x \) and the chosen test and is a statistic, is called the \textit{p-value} for the test \( T_\alpha \).

\( p \)-value

It is good practice to determine not only whether \( H_0 \) is rejected for a given \( \alpha \) and a chosen test \( T_\alpha \), but also the smallest possible level of significance at which \( H_0 \) would be rejected for the computed \( T_\alpha(x) \), i.e.,
Choice of significance level

The choice of a level of significance $\alpha$ is usually somewhat subjective. In most applications there is no precise limit to the size of $T$ that can be tolerated. Standard values, 0.10, 0.05, and 0.01, are often used for convenience. For most tests satisfying $\sup_{P \in \mathcal{P}_0} \alpha_T(P) \leq \alpha$, a small $\alpha$ leads to a "small" rejection region.

$p$-value

It is good practice to determine not only whether $H_0$ is rejected for a given $\alpha$ and a chosen test $T_\alpha$, but also the smallest possible level of significance at which $H_0$ would be rejected for the computed $T_\alpha(x)$, i.e.,

$$\hat{\alpha} = \inf\{\alpha \in (0, 1) : T_\alpha(x) = 1\}.$$  

Such an $\hat{\alpha}$, which depends on $x$ and the chosen test and is a statistic, is called the $p$-value for the test $T_\alpha$. 

Example 2.29

Let us calculate the $p$-value for $T_{c\alpha}$ in Example 2.28. Note that

$$\alpha = 1 - \Phi \left( \frac{\sqrt{n}(c_\alpha - \mu_0)}{\sigma} \right) > 1 - \Phi \left( \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right)$$

if and only if $\bar{x} > c_\alpha$ (or $T_{c\alpha}(x) = 1$).

Hence

$$1 - \Phi \left( \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right) = \inf\{\alpha \in (0, 1) : T_{c\alpha}(x) = 1\} = \hat{\alpha}(x)$$

is the $p$-value for $T_{c\alpha}$.

It turns out that $T_{c\alpha}(x) = I_{(0,\alpha)}(\hat{\alpha}(x))$.

Remarks

- With the additional information provided by $p$-values, using $p$-values is typically more appropriate than using fixed-level tests in a scientific problem.

- In some cases, a fixed level of significance is unavoidable when acceptance or rejection of $H_0$ is a required decision.
Example 2.29
Let us calculate the p-value for \( T_{c\alpha} \) in Example 2.28. Note that
\[
\alpha = 1 - \Phi \left( \frac{\sqrt{n}(c_{\alpha} - \mu_0)}{\sigma} \right) > 1 - \Phi \left( \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right)
\]
if and only if \( \bar{x} > c_{\alpha} \) (or \( T_{c\alpha}(x) = 1 \)). Hence
\[
1 - \Phi \left( \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right) = \inf\{\alpha \in (0, 1) : T_{c\alpha}(x) = 1\} = \hat{\alpha}(x)
\]
is the p-value for \( T_{c\alpha} \).

It turns out that \( T_{c\alpha}(x) = I_{(0, \alpha)}(\hat{\alpha}(x)) \).

Remarks
- With the additional information provided by p-values, using p-values is typically more appropriate than using fixed-level tests in a scientific problem.
- In some cases, a fixed level of significance is unavoidable when acceptance or rejection of \( H_0 \) is a required decision.
Randomized tests

In Example 2.28, \( \sup_{P \in \mathcal{P}_0} \alpha_T(P) = \alpha \) can always be achieved by a suitable choice of \( c \).

This is, however, not true in general.

We need to consider \textit{randomized tests}.

Recall that a randomized decision rule is a probability measure \( \delta(x, \cdot) \) on the action space for any fixed \( x \).

Since the action space contains only two points, 0 and 1, for a hypothesis testing problem, any randomized test \( \delta(X, A) \) is equivalent to a statistic \( T(X) \in [0, 1] \) with \( T(x) = \delta(x, \{1\}) \) and \( 1 - T(x) = \delta(x, \{0\}) \).

A nonrandomized test is obviously a special case where \( T(x) \) does not take any value in \((0, 1)\).

For any randomized test \( T(X) \), we define the type I error probability to be \( \alpha_T(P) = E[T(X)], \ P \in \mathcal{P}_0 \), and the type II error probability to be \( 1 - \alpha_T(P) = E[1 - T(X)], \ P \in \mathcal{P}_1 \).

For a class of randomized tests, we would like to minimize \( 1 - \alpha_T(P) \) subject to \( \sup_{P \in \mathcal{P}_0} \alpha_T(P) = \alpha \).
Randomized tests

In Example 2.28, \( \sup_{P \in \mathcal{P}_0} \alpha_T(P) = \alpha \) can always be achieved by a suitable choice of \( c \).
This is, however, not true in general.
We need to consider randomized tests.
Recall that a randomized decision rule is a probability measure \( \delta(x, \cdot) \) on the action space for any fixed \( x \).
Since the action space contains only two points, 0 and 1, for a hypothesis testing problem, any randomized test \( \delta(X, A) \) is equivalent to a statistic \( T(X) \in [0, 1] \) with \( T(x) = \delta(x, \{1\}) \) and \( 1 - T(x) = \delta(x, \{0\}) \).
A nonrandomized test is obviously a special case where \( T(x) \) does not take any value in \((0, 1)\).

For any randomized test \( T(X) \), we define the type I error probability to be \( \alpha_T(P) = E[T(X)], P \in \mathcal{P}_0 \), and the type II error probability to be \( 1 - \alpha_T(P) = E[1 - T(X)], P \in \mathcal{P}_1 \).
For a class of randomized tests, we would like to minimize \( 1 - \alpha_T(P) \) subject to \( \sup_{P \in \mathcal{P}_0} \alpha_T(P) = \alpha \).
Randomized tests

In Example 2.28, \( \sup_{P \in \mathcal{P}_0} \alpha_T(P) = \alpha \) can always be achieved by a suitable choice of \( c \).
This is, however, not true in general.
We need to consider randomized tests.
Recall that a randomized decision rule is a probability measure \( \delta(x, \cdot) \) on the action space for any fixed \( x \).
Since the action space contains only two points, 0 and 1, for a hypothesis testing problem, any randomized test \( \delta(X,A) \) is equivalent to a statistic \( T(X) \in [0,1] \) with \( T(x) = \delta(x,\{1\}) \) and \( 1 - T(x) = \delta(x,\{0\}) \).
A nonrandomized test is obviously a special case where \( T(x) \) does not take any value in \((0,1)\).

For any randomized test \( T(X) \), we define the type I error probability to be \( \alpha_T(P) = E[T(X)], P \in \mathcal{P}_0 \), and the type II error probability to be \( 1 - \alpha_T(P) = E[1 - T(X)], P \in \mathcal{P}_1 \).
For a class of randomized tests, we would like to minimize \( 1 - \alpha_T(P) \) subject to \( \sup_{P \in \mathcal{P}_0} \alpha_T(P) = \alpha \).
Example 2.30
Assume that the sample $X$ has the binomial distribution $Bi(\theta, n)$ with an unknown $\theta \in (0, 1)$ and a fixed integer $n > 1$. Consider the hypotheses $H_0 : \theta \in (0, \theta_0]$ versus $H_1 : \theta \in (\theta_0, 1)$, where $\theta_0 \in (0, 1)$ is a fixed value.
Consider the following class of randomized tests:

$$T_{j, q}(X) = \begin{cases} 1 & X > j \\ q & X = j \\ 0 & X < j, \end{cases}$$

where $j = 0, 1, \ldots, n - 1$ and $q \in [0, 1]$.

$$\alpha_{T_{j, q}}(\theta) = P(X > j) + qP(X = j) \quad 0 < \theta \leq \theta_0$$

$$1 - \alpha_{T_{j, q}}(\theta) = P(X < j) + (1 - q)P(X = j) \quad \theta_0 < \theta < 1.$$ 

It can be shown that for any $\alpha \in (0, 1)$, there exist an integer $j$ and $q \in (0, 1)$ such that the size of $T_{j, q}$ is $\alpha$. 
Confidence sets

\( \vartheta \): a \( k \)-vector of unknown parameters related to the unknown population \( P \in \mathcal{P} \)

\( C(X) \) a Borel set (in the range of \( \vartheta \)) depending only on the sample \( X \)

If

\[
\inf_{P \in \mathcal{P}} P( \vartheta \in C(X)) \geq 1 - \alpha, \tag{2}
\]

where \( \alpha \) is a fixed constant in \((0, 1)\), then \( C(X) \) is called a confidence set for \( \vartheta \) with level of significance \( 1 - \alpha \).

The left-hand side of (2) is called the confidence coefficient of \( C(X) \), which is the highest possible level of significance for \( C(X) \).

A confidence set is a random element that covers the unknown \( \vartheta \) with certain probability.

If (2) holds, then the coverage probability of \( C(X) \) is at least \( 1 - \alpha \), although \( C(x) \) either covers or does not cover \( \vartheta \) whence we observe \( X = x \).
Confidence sets

$\vartheta$: a $k$-vector of unknown parameters related to the unknown population $P \in \mathcal{P}$

$C(X)$ a Borel set (in the range of $\vartheta$) depending only on the sample $X$

If

$$\inf_{P \in \mathcal{P}} P(\vartheta \in C(X)) \geq 1 - \alpha,$$

(2)

where $\alpha$ is a fixed constant in $(0, 1)$, then $C(X)$ is called a confidence set for $\vartheta$ with level of significance $1 - \alpha$.

The left-hand side of (2) is called the confidence coefficient of $C(X)$, which is the highest possible level of significance for $C(X)$.

A confidence set is a random element that covers the unknown $\vartheta$ with certain probability.

If (2) holds, then the coverage probability of $C(X)$ is at least $1 - \alpha$, although $C(x)$ either covers or does not cover $\vartheta$ whence we observe $X = x$. 
Example 2.32

Let $X_1, \ldots, X_n$ be i.i.d. from the $N(\mu, \sigma^2)$ distribution with both $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ unknown.

Let $\theta = (\mu, \sigma^2)$ and $\alpha \in (0, 1)$ be given.

Let $\bar{X}$ be the sample mean and $S^2$ be the sample variance.

Since $(\bar{X}, S^2)$ is sufficient (Example 2.15), we focus on $C(X)$ that is a function of $(\bar{X}, S^2)$.

From Example 2.18, $\bar{X}$ and $S^2$ are independent and $(n - 1)S^2/\sigma^2$ has the chi-square distribution $\chi^2_{n-1}$.

Since $\sqrt{n}(\bar{X} - \mu)/\sigma$ has the $N(0, 1)$ distribution,

$$P \left( -\tilde{c}_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \tilde{c}_\alpha \right) = \sqrt{1 - \alpha},$$

where $\tilde{c}_\alpha = \Phi^{-1} \left( \frac{1 + \sqrt{1 - \alpha}}{2} \right)$ (verify).

Since the chi-square distribution $\chi^2_{n-1}$ is a known distribution, we can always find two constants $c_{1\alpha}$ and $c_{2\alpha}$ such that
Example 2.32

Let $X_1, \ldots, X_n$ be i.i.d. from the $N(\mu, \sigma^2)$ distribution with both $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ unknown.

Let $\theta = (\mu, \sigma^2)$ and $\alpha \in (0, 1)$ be given.

Let $\bar{X}$ be the sample mean and $S^2$ be the sample variance.

Since $(\bar{X}, S^2)$ is sufficient (Example 2.15), we focus on $C(X)$ that is a function of $(\bar{X}, S^2)$.

From Example 2.18, $\bar{X}$ and $S^2$ are independent and $(n - 1)S^2 / \sigma^2$ has the chi-square distribution $\chi^2_{n-1}$.

Since $\sqrt{n}(\bar{X} - \mu) / \sigma$ has the $N(0, 1)$ distribution,

$$P \left( -\tilde{c}_\alpha \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq \tilde{c}_\alpha \right) = \sqrt{1 - \alpha},$$

where $\tilde{c}_\alpha = \Phi^{-1} \left( \frac{1 + \sqrt{1 - \alpha}}{2} \right)$ (verify).

Since the chi-square distribution $\chi^2_{n-1}$ is a known distribution, we can always find two constants $c_{1\alpha}$ and $c_{2\alpha}$ such that
Example 2.32 (continued)

\[ P \left( c_1 \alpha \leq \frac{(n-1)S^2}{\sigma^2} \leq c_2 \alpha \right) = \sqrt{1 - \alpha}. \]

Then

\[ P \left( \tilde{c}_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \tilde{c}_\alpha, c_1 \alpha \leq \frac{(n-1)S^2}{\sigma^2} \leq c_2 \alpha \right) = 1 - \alpha, \]

or

\[ P \left( \frac{n(\bar{X} - \mu)^2}{\tilde{c}_\alpha^2} \leq \sigma^2, \frac{(n-1)S^2}{c_2 \alpha} \leq \sigma^2 \leq \frac{(n-1)S^2}{c_1 \alpha} \right) = 1 - \alpha. \] (3)

The left-hand side of (3) defines a set in the range of \( \theta = (\mu, \sigma^2) \) bounded by two straight lines, \( \sigma^2 = (n-1)S^2/c_{i\alpha}, \ i = 1, 2, \) and a curve \( \sigma^2 = n(\bar{X} - \mu)^2/\tilde{c}_\alpha^2 \) (see the shadowed part of Figure 2.3). This set is a confidence set for \( \theta \) with confidence coefficient \( 1 - \alpha \), since (3) holds for any \( \theta \).