Hypothesis Testing*

Ricardo Dahis

PUC-Rio, Department of Economics

Summer 2023

*Last updated: February 6, 2023, 08:44h

Contents

Contents

Contents

some definitions: null and alternative hypothesis

- definition: a statistical hypothesis is a statement about population parameters
- the goal is to decide which of two complementary hypotheses is true:

$$\text{null hypothesis } \mathbb{H}_0 \quad \text{vs} \quad \text{alternative hypothesis } \mathbb{H}_1$$

- if θ denotes a population parameter, then the general format of the null and alternative hypotheses is $\mathbb{H}_0 : \theta \in \Theta_0$ and $\mathbb{H}_1 : \theta \in \Theta_1$
- examples:
  - if θ represents the effect of a training program, we might be interested in $\mathbb{H}_0 : \theta = 0$ against $\mathbb{H}_1 : \theta \neq 0$
  - if σ^2 is the variance, we might be interested in understanding if volatility is too high defining $\mathbb{H}_0 : \sigma^2 = \sigma_0^2$ against $\mathbb{H}_1 : \sigma^2 > \sigma_0^2$

some definitions: rejection region

- definition: a hypothesis test is a rule that determines for which sample values the decision is to reject or not $\mathbb{H}_0$
  - we define a partition in the sample space $\mathcal{X}$ with two sets: R and R^c
  - if $x \in R$, we elect to reject $\mathbb{H}_0$; if $x \in R^c$, we elect to not reject $\mathbb{H}_0$
  - R is the rejection region and R^c is the acceptance region
  - typically, a hypothesis test is specified in terms of a test statistic $T(\boldsymbol{x})$, but this is not necessary
  - R (and, consequently, R^c) can be defined arbitrarily – but makes little sense to do so if we want a test with good properties

some definitions: power function

- definition: the power function of a hypothesis test with a given rejection region R is the function of $\boldsymbol{\theta}$

$$\beta(\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X} \in R)$$

- be careful: the power function $\neq$ power of the test!

- the terminology is misleading: one should think the power function as the probability of rejecting the null as a function of θ, regardless of whether the null is true or not

some definitions: type-I and type-II errors

- there are two types of error a hypothesis test $\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$ vs $\mathbb{H}_1 : \boldsymbol{\theta} \in \Theta_1$ might make
  - rejecting the null when it is true (false positive): type I error occurs if $\boldsymbol{\theta} \in \Theta_0$ and $x \in R$
  - not rejecting the null when it is false (false negative): type II occurs if $\boldsymbol{\theta} \in \Theta_1$ and $x \notin R$

		decision	
		not reject $\mathbb{H}_0$ $x \notin R$	reject $\mathbb{H}_0$ $x \in R$
truth	$\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$	correct	type I
	$\mathbb{H}_1 : \boldsymbol{\theta} \in \Theta_1$	type II	correct

size and power function

- for each $\boldsymbol{\theta} \in \Theta_0$, $\beta(\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}(X \in R)$ represents the probability that the null hypothesis is rejected while being true.

$$\text{if } \boldsymbol{\theta} \in \Theta_0 : \beta(\boldsymbol{\theta}) \;=\; \mathbb{P}_{\boldsymbol{\theta}}(X \in R) \;=\; \mathbb{P}_{\boldsymbol{\theta}}(\text{type I error}) \;=\; \text{size at } \boldsymbol{\theta}$$

- size varies with $\boldsymbol{\theta}$: we need an aggregate measure for the entire test over the set Θ_0

- example: suppose $X_i \sim N(\mu, 1)$ i.i.d. and that we test $\mathbb{H}_0 : \mu > 0$ against $\mathbb{H}_1 : \mu \leq 0$. We elect to make $R = \{\bar{x}_n \leq 0\}$. The probability of $\bar{x}_n$ being in the rejection region is completely different if $\mu = 0.0001$ or $\mu = 1000$.

- definition: for $0 \leq \alpha \leq 1$, a test with power function $\beta(\boldsymbol{\theta})$ has size α if

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \beta(\boldsymbol{\theta}) = \alpha$$

  whereas it has level α if $\sup_{\boldsymbol{\theta} \in \Theta_0} \beta(\boldsymbol{\theta}) \leq \alpha$.

- ideally, we would have size 0, which is equivalent to $\beta(\boldsymbol{\theta}) = 0$ for all $\theta \in \Theta_0$, but life is never this perfect

power and power function

- for each $\boldsymbol{\theta} \in \Theta_1$, $\beta(\boldsymbol{\theta}) = \mathbb{P}_{\boldsymbol{\theta}}(X \in R)$ represents the probability that the null hypothesis is rejected while being false.

$$\text{if } \boldsymbol{\theta} \in \Theta_1 : \ \beta(\boldsymbol{\theta}) \ = \ \mathbb{P}_{\boldsymbol{\theta}}(X \in R) \ = \ 1 - \mathbb{P}_{\boldsymbol{\theta}}(\text{type II error}) \ = \ \text{power at } \boldsymbol{\theta}$$

- as with size, power varies with $\boldsymbol{\theta}$, but we choose not to define an aggregate measure over $\boldsymbol{\theta} \in \Theta_1$

power function for binomial probability

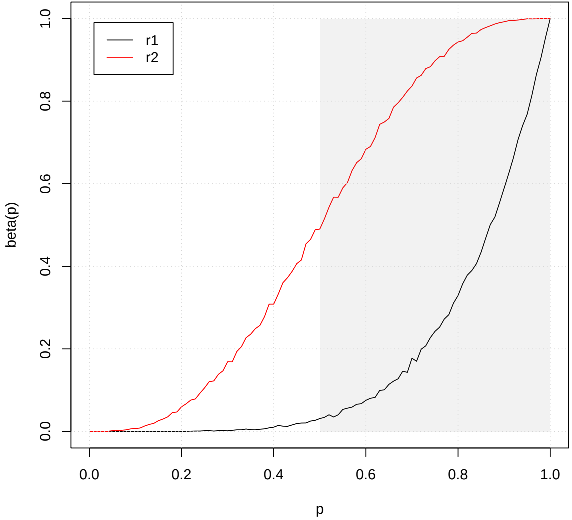
- example 1: let $X \sim \text{Bin}(5, p)$ and consider testing $\mathbb{H}_0 : \Theta_0 = \{p : 0 \leq p \leq 1/2\}$ vs $\mathbb{H}_1 : \Theta_1 = \{p : 1/2 < p \leq 1\}$
- test 1: $x \in R$ if and only if every observation is a success
  - $\beta_1(p) = \mathbb{P}_p(X = 5) = p^5$
  - probability of type I error is pretty low for any $p \leq 1/2$ ($\frac{1}{2^5} = 0.0312$)
  - probability of type II error is less than half only if $p > 0.5^{1/5} = 0.87$
- test 2 $x \in R$ if and only if $X \in \{3, 4, 5\}$
  - $\beta_2(p) = \mathbb{P}_p(X \in \{3, 4, 5\}) = \sum_{x=3}^{5} \binom{5}{x} p^x (1-p)^{5-x}$
  - the price we pay for a much smaller probability of type II error is a larger probability of type I error

R codes

$$\begin{aligned}
\text{test 1} &: \quad x \in R \text{ if and only if every observation is a success} \\
\text{test 2} &: \quad x \in R \text{ if and only if } X \in \{3, 4, 5\}
\end{aligned}$$

```
r1 <- function(p){mean(rbinom(5000,5,p)==5)}
r2 <- function(p){mean(rbinom(5000,5,p)>=3)}
p <- seq(0,1,by=0.01)
plot(p,sapply(p,r1),type='l',ylab='beta(p)',xlab='p')
lines(p,sapply(p,r2),type='l',col='red')
```

R codes

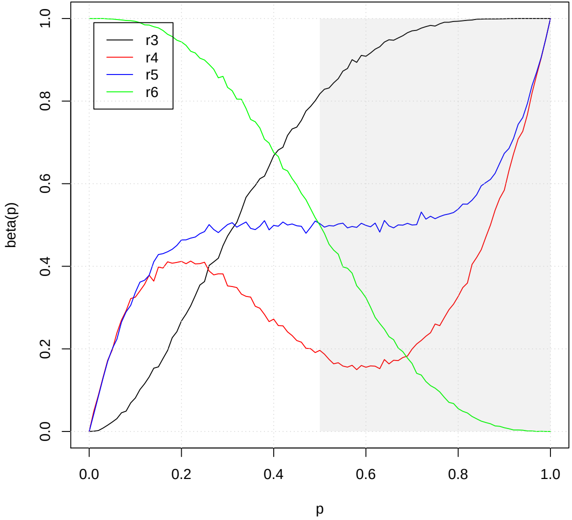


R codes

$$\begin{aligned}
\text{test 3} &: \quad \text{rejects } \mathbb{H}_0 \text{ if and only if } X \in \{2,3,4,5\} \\
\text{test 4} &: \quad \text{rejects } \mathbb{H}_0 \text{ if and only if } X \in \{1,5\} \\
\text{test 5} &: \quad \text{rejects } \mathbb{H}_0 \text{ if and only if } X \in \{1,3,5\} \\
\text{test 6} &: \quad \text{rejects } \mathbb{H}_0 \text{ if and only if } X \in \{1,2\}
\end{aligned}$$

```
r3 <- function(p){mean(rbinom(5000,5,p)>=2)}

r4 <- function(p){

  v <- rbinom(5000,5,p)
  mean((v==1)+(v==5))

}

r5 <- function(p){

  v <- rbinom(5000,5,p)
  mean((v==1)+(v==3)+(v==5))

}

r6 <- function(p){mean(rbinom(5000,5,p)<=2)}
```

R codes



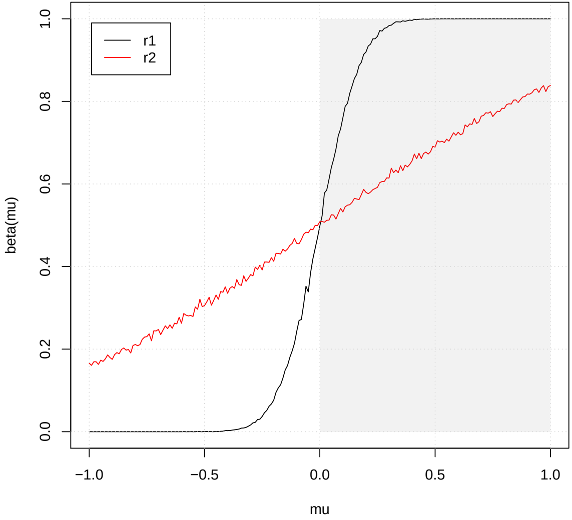
power function for Gaussian mean

- example 2: let $X_1, \ldots, X_n \sim$ iid $N(\mu, 1)$ and consider testing $\mathbb{H}_0 : \mu \leq 0$ versus $\mathbb{H}_0 : \mu > 0$. For that test, we propose two rejection regions

$$\begin{aligned} \text{test 1} &: \quad x \in R \text{ if and only if } \bar{X}_n > 0 \\ \text{test 2} &: \quad x \in R \text{ if and only if } X_1 > 0 \end{aligned}$$

```
n <- 50

rGaussian1 <- function(mu){
  vecTest <- matrix(0,5000,1)
  for (i in 1:5000){vecTest[i,1] <- mean(rnorm(n,mean=mu,sd=1)) > 0}
  mean(vecTest)
}

rGaussian2 <- function(mu){
  vecTest <- matrix(0,5000,1)
  for (i in 1:5000){vecTest[i,1] <-(rnorm(1,mean=mu,sd=1)) > 0}
  mean(vecTest)
}
```

R codes



power function for Gaussian mean

- example 2 (cont'd): rejection/acceptance region R are generally arbitrary; but it is unlikely that tests with good properties would ensue

- let $X_1, \ldots, X_n \sim$ iid $N(\mu, 1)$ and consider testing $\mathbb{H}_0 : \mu \leq 100$ versus $\mathbb{H}_0 : \mu > 100$. For that test, keep the two previous tests

$$\begin{aligned} \text{test 1} \quad &: \quad x \in R \text{ if and only if } \bar{X}_n > 0 \\ \text{test 2} \quad &: \quad x \in R \text{ if and only if } X_1 > 0 \end{aligned}$$

  this test will have massive size distortions, and power very close to 1.

- in the next example, we conveniently standardize the test statistic.

power function for Gaussian mean

- example 3: let $X_1, \ldots, X_n \sim$ iid $N(\mu, 1)$ and consider testing $\mathbb{H}_0 : \mu \leq \mu_0$ versus $\mathbb{H}_1 : \mu > \mu_0$ using a rejection region $\bar{X}_n > \kappa$.

- we now aim to choose κ such that we know the probability type-I errors, i.e., we aim to devise a test with a defined size
  - in other words, α and n are fixed and we let power roam free

- we know that

$$\beta(\mu) = \mathbb{P}_\mu \left(\bar{X}_n > \kappa \right)$$

  but we can't calculate this probability because μ is not known, so we instead compute

$$\beta(\mu) = \mathbb{P}_\mu \left(\frac{\bar{X}_n - \mu}{1/\sqrt{n}} > \frac{\kappa - \mu}{1/\sqrt{n}} \right) = \mathbb{P} \left(Z > \frac{\kappa - \mu}{1/\sqrt{n}} \right)$$

  with $Z \sim N(0, 1)$.

- important to notice: we've manipulated $\beta(\mu)$ so that it depends on some known distribution (and not on μ). In this way, we may forgo the simulations

power function for Gaussian mean

- we may choose κ to match a test size from

$$\beta(\mu) \quad = \quad \mathbb{P}\left(Z > \frac{\kappa - \mu}{1/\sqrt{n}}\right)$$

- since $\beta(\mu)$ is increasing with μ, maximum $\beta(\mu) = \mathbb{P}\left(Z > \frac{\kappa-\mu}{1/\sqrt{n}}\right)$ subject to $\mathbb{H}_0 : \mu \leq \mu_0$ is achieved at $\mu = \mu_0$
- so we select κ such that

$$\mathbb{P}\left(Z > \frac{\kappa - \mu_0}{1/\sqrt{n}}\right) \quad = \quad \alpha$$

- from the standard normal tables, there is value z_α such that $\mathbb{P}(Z > z_\alpha) = \alpha$. For example, if $\alpha = 0.05$, $z_\alpha \approx 1.64$. Therefore,

$$\frac{\kappa - \mu_0}{1/\sqrt{n}} \quad = \quad z_\alpha \quad \Longrightarrow \quad \kappa \quad = \quad \mu_0 + \frac{z_\alpha}{\sqrt{n}}$$

- the rejection region

$$R \quad = \quad \left\{X : \bar{X}_n > \mu_0 + \frac{z_\alpha}{\sqrt{n}}\right\}$$

  was defined such that the statistical test has size α

power function for Gaussian mean

- this is not necessarily the most convenient formulation: consider testing $\mathbb{H}_0 : \mu \leq \mu_0$ versus $\mathbb{H}_1 : \mu > \mu_0$ using a rejection region $\frac{\bar{X}_n - \mu_0}{1/\sqrt{n}} > c$

$$\begin{aligned}
\beta(\mu) &= \mathbb{P}_\mu\left(\frac{\bar{X}_n - \mu_0}{1/\sqrt{n}} > c\right) = \mathbb{P}_\mu\left(\frac{\bar{X}_n - \mu + \mu - \mu_0}{1/\sqrt{n}} > c\right) \\
&= \mathbb{P}_\mu\left(\frac{\bar{X}_n - \mu}{1/\sqrt{n}} + \frac{\mu - \mu_0}{1/\sqrt{n}} > c\right) = \mathbb{P}_\mu\left(\frac{\bar{X}_n - \mu}{1/\sqrt{n}} > c - \frac{\mu - \mu_0}{1/\sqrt{n}}\right) \\
&= \mathbb{P}\left(Z > c + \frac{\mu_0 - \mu}{1/\sqrt{n}}\right) \text{ with } Z \sim N(0,1)
\end{aligned}$$

- important:
  - $\beta(\mu)$ is increasing in μ, with $\lim_{\mu \to -\infty} \beta(\mu) = 0$, $\lim_{\mu \to \infty} \beta(\mu) = 1$
  - if $\mathbb{P}(Z > c) = \alpha$, then $\beta(\mu_0) = \alpha$, the size of the test
  - to control for size α, we choose $c = z_\alpha$
  - power depends on the distance $\mu_0 - \mu$
  - power increases to 1 as $n \to \infty$

power function for Gaussian mean

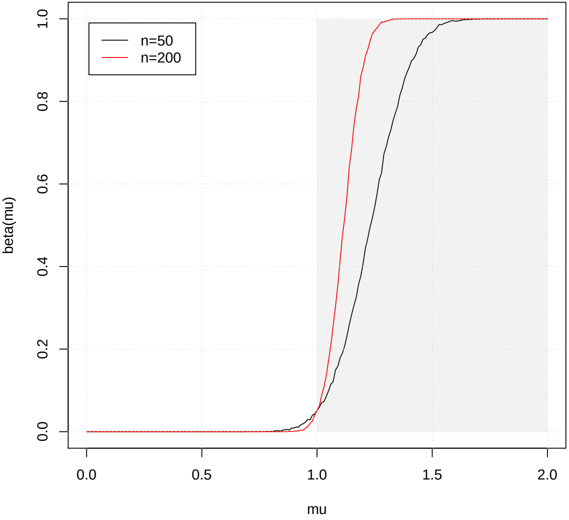
- that is, we have defined the rejection region

$$R = \left\{X : \frac{\bar{X}_n - \mu_0}{1/\sqrt{n}} > z_\alpha\right\} = \left\{X : \bar{X}_n > \mu_0 + \frac{z_\alpha}{\sqrt{n}}\right\}$$

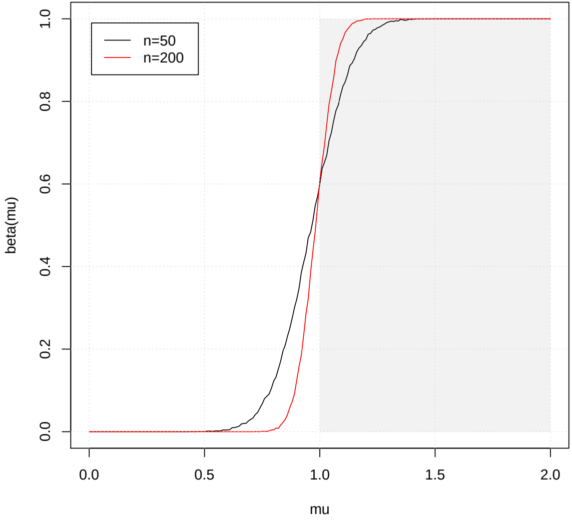
  as we had before.

```
mu0 <- 1

c <- 1.64485

rGaussian2 <- function(mu){

  vecTest <- matrix(0,5000,1)

  for (i in 1:5000){vecTest[i,1] <-
    (sqrt(n)*(mean(rnorm(n,mean=mu,sd=1))-mu0)) > c}

  mean(vecTest)

}
```

$c = 1.64485,\ \alpha = 0.05$



$c = -0.25334, \ \alpha = 0.60$



power function for Gaussian mean

- example 4: suppose now that the probability of type I error must not exceed 0.10 and that of type II error must not exceed 0.20 if $\mu \geq \mu_0 + 1$

- we now aim to choose n such that we <u>know</u> the probability type-I and type-II errors for a given effect size
  - typical application: determination of sample sizes in RCTs.

- using a test that rejects $\mathbb{H}_0 : \mu \leq \mu_0$ if $\sqrt{n}(\bar{X}_n - \mu_0) > c$

$$\beta(\mu) \;=\; \mathbb{P}\left(Z > c + \frac{\mu_0 - \mu}{1/\sqrt{n}}\right) \;=\; \begin{cases} \mathbb{P}(Z > c) = 0.1 & \text{if } \mu = \mu_0 \\ \mathbb{P}(Z > c - \sqrt{n}) = 0.8 & \text{if } \mu = \mu_0 + 1 \end{cases}$$

- from $\mathbb{P}(Z > c) = 0.1$, we get that $c \approx 1.28$

- from $\mathbb{P}(Z > c - \sqrt{n}) = 0.8$, we get that

$$c - \sqrt{n} \;\approx\; -0.84 \;\;\Rightarrow\;\; n \;\approx\; (c + 0.84)^2 \;\approx\; 4.49$$

  or $n \geq 5$

power function for Gaussian mean

- example 5: let $X_1, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$, σ^2 known. A test for $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_1 : \theta \neq \theta_0$ rejects $\mathbb{H}_0$ if $|\bar{X}_n - \theta_0|/(\sigma/\sqrt{n}) > c$.

  the experimenter desires a type-I error of probability 0.05 and a maximum type-II error of 0.25 at $\theta = \theta_0 + \sigma$. What values of n and c achieves this?

- we should first find the power function

$$
\begin{aligned}
\beta(\theta) &= \mathbb{P}_\theta\left(\frac{|\bar{x}_n - \theta_0|}{\sigma/\sqrt{n}} > c\right) \quad = \quad 1 - \mathbb{P}_\theta\left(\frac{|\bar{x}_n - \theta_0|}{\sigma/\sqrt{n}} \leq c\right) \\
&= 1 - \mathbb{P}_\theta\left(-c \leq \frac{\bar{x}_n - \theta + \theta - \theta_0}{\sigma/\sqrt{n}} \leq c\right) \\
&= 1 - \mathbb{P}_\theta\left(-c - \frac{\theta - \theta_0}{\sigma/\sqrt{n}} \leq \frac{\bar{x}_n - \theta}{\sigma/\sqrt{n}} \leq c - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) \\
&= 1 - \mathbb{P}_\theta\left(-c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \leq Z \leq c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) \\
&= 1 - \left[\Phi\left(c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) - \Phi\left(-c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)\right]
\end{aligned}
$$

power function for Gaussian mean

- by hypothesis,

$$
\begin{aligned}
0.05 &= \beta(\theta_0) = 1 - [\Phi(c) - \Phi(-c)] \\
&= 1 - [\Phi(c) - 1 + \Phi(c)] = 2 - 2 \cdot \Phi(c) \\
0.025 &= 1 - \Phi(c)
\end{aligned}
$$

and $c = 1.96$.

- power at $\theta = \theta_0 + \sigma$ is

$$
\begin{aligned}
.75 &\leq \beta(\theta_0 + \sigma) = 1 - \left[\Phi\left(c + \frac{-\sigma}{\sigma/\sqrt{n}}\right) - \Phi\left(-c + \frac{-\sigma}{\sigma/\sqrt{n}}\right)\right] \\
&= 1 + \Phi(-c - \sqrt{n}) - \Phi(c - \sqrt{n}) \\
&= 1 + \Phi(-1.96 - \sqrt{n}) - \Phi(1.96 - \sqrt{n}) \\
&\approx 1 - \Phi(1.96 - \sqrt{n})
\end{aligned}
$$

since $\Phi(-.675) \approx 0.25$, then $1.96 - \sqrt{n} = -.675$, and so $n = 6.943 \approx 7$.

power function for Gaussian mean

- example 6: let $X_1, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$, σ^2 unknown. A test for $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_1 : \theta \neq \theta_0$ rejects $\mathbb{H}_0$ if $|\bar{X}_n - \theta_0|/(s/\sqrt{n}) > c$, where $s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2}$.

  the experimenter desires a type-I error of probability 0.05 and a maximum type-II error of 0.25 at $\theta = \theta_0 + \sigma$. What values of n and c achieves this?

- we should adjust the power function

$$
\begin{aligned}
\beta(\theta) &= \mathbb{P}_\theta\left(\frac{|\bar{x}_n - \theta_0|}{s/\sqrt{n}} > c\right) \;=\; 1 - \mathbb{P}_\theta\left(\frac{|\bar{x}_n - \theta_0|}{s/\sqrt{n}} \leq c\right) \\
&= 1 - \mathbb{P}_\theta\left(-c \leq \frac{\bar{x}_n - \theta + \theta - \theta_0}{s/\sqrt{n}} \leq c\right) \\
&= 1 - \mathbb{P}_\theta\left(-c - \frac{\theta - \theta_0}{s/\sqrt{n}} \leq \frac{\bar{x}_n - \theta}{s/\sqrt{n}} \leq c - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right) \\
&= 1 - \mathbb{P}_\theta\left(-c + \frac{\theta_0 - \theta}{s/\sqrt{n}} \leq t \leq c + \frac{\theta_0 - \theta}{s/\sqrt{n}}\right) \\
&= 1 - \left[F\left(c + \frac{\theta_0 - \theta}{s/\sqrt{n}}\right) - F\left(-c + \frac{\theta_0 - \theta}{s/\sqrt{n}}\right)\right]
\end{aligned}
$$

  where $t \sim t_{n-1}$ with cdf $F(\cdot)$.

power function for Bernoulli with CLT

- example 7: for a random sample $X_1, \ldots, X_n$ of Bernoulli(p) variables, it is desired to test $\mathbb{H}_0 : p = 0.49$ against $\mathbb{H}_1 : p = 0.51$. Use the central limit theorem to determine, approximately, the sample size needed so that the two probabilities of error are both about 0.01. Use a test function that rejects $\mathbb{H}_0$ if $\sum_{i=1}^n X_i$ is large.

- solution: by the CLT,

$$Z = \frac{\sum X_i - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

a test that rejects $\mathbb{H}_0$ if $\sum X_i > c$ has

$$\mathbb{P}\left(Z > \frac{c - n(.49)}{\sqrt{n(.49)(.51)}}\right) = 0.01 \text{ and } \mathbb{P}\left(Z > \frac{c - n(.51)}{\sqrt{n(.49)(.51)}}\right) = 0.01$$

therefore

$$\frac{c - n(.49)}{\sqrt{n(.49)(.51)}} = 2.33 \text{ and } \frac{c - n(.51)}{\sqrt{n(.49)(.51)}} = -2.33$$

solving these equations gives $n = 13.567$ and $c = 6783.5$.

Contents

previous examples

- in most previous examples, we've used rejection regions of the format

$$R = \left\{X : T(X) > \kappa\right\}$$

which is an interval (κ, ∞) for a sufficient statistic $T(X)$.

  - example 2: $R = \left\{X : \bar{X}_n > 0\right\}$
  - example 3: $R = \left\{X : \bar{X}_n > \frac{z_\alpha}{\sqrt{n}+\mu_0}\right\}$
  - example 4: $R = \left\{X : \sqrt{n}(\bar{X}_n - \mu_0) > c\right\}$
  - example 5: $R = \left\{X : |\bar{X}_n - \theta_0|/(\sigma/\sqrt{n}) > c\right\}$
  - example 6: $R = \left\{X : \sum X_i \text{ "large" }\right\}$

- we are going to see that rejection regions of this format are well-grounded by theory

Contents

likelihood ratio test

- it is a very general method of finding acceptance/rejection regions, virtually always applicable and optimal in some sense that we will discuss later

- definition: the LR test for $\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$ against $\mathbb{H}_1 : \boldsymbol{\theta} \in \Theta_1$ is a test with a rejection region of the form $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$, where $0 \leq c \leq 1$ and

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} \ell(\boldsymbol{\theta}|\boldsymbol{x})}{\sup_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}|\boldsymbol{x})} = \frac{\ell(\hat{\boldsymbol{\theta}}_0|\boldsymbol{x})}{\ell(\hat{\boldsymbol{\theta}}|\boldsymbol{x})}$$

- if the restriction is not binding, the constrained maximization $\ell(\hat{\theta}_0|\boldsymbol{x})$ will be the same as the unconstrained maximization $\ell(\hat{\theta}|\boldsymbol{x})$ and $\lambda(\boldsymbol{x}) = 1$

- for now, think c as a fixed constant. We will soon see what that choice entails!

LR test for the Gaussian mean

- example 1: let $(X_1, \ldots, X_n)$ be a random sample from a $N(\mu, 1)$ population and consider testing $\mathbb{H}_0 : \mu = \mu_0$ versus $\mathbb{H}_1 : \mu \neq \mu_0$, then

$$\begin{aligned}
\lambda(\boldsymbol{x}) &= \frac{\ell(\mu_0|\boldsymbol{x})}{\ell(\bar{x}_n|\boldsymbol{x})} = \frac{(2\pi)^{-n/2}\exp\left[-\sum_{i=1}^n (x_i-\mu_0)^2/2\right]}{(2\pi)^{-n/2}\exp\left[-\sum_{i=1}^n (x_i-\bar{x}_n)^2/2\right]} \\
&= \exp\left[-\frac{\sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\bar{x}_n)^2}{2}\right] \\
&= \exp\left[-\frac{n(\bar{x}_n-\mu_0)^2}{2}\right],
\end{aligned}$$

and for $\lambda(\boldsymbol{x}) = c$,

$$\ln c = -\frac{n(\bar{x}_n-\mu_0)^2}{2} \Rightarrow (\bar{x}_n-\mu_0)^2 = -2(\ln c)/n$$

yielding a rejection region

$$\{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\} = \left\{\boldsymbol{x} : |\bar{x}_n - \mu_0| \geq \sqrt{-2(\ln c)/n}\right\}$$

size of a LR test

- in general, to derive a size α LR test that rejects the null $\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$ if $\lambda(\boldsymbol{x}) \leq c$, we choose c such that $\sup_{\boldsymbol{\theta} \in \Theta_0} \mathbb{P}_{\boldsymbol{\theta}}(\lambda(\boldsymbol{x}) \leq c) = \alpha$

- example 1 (cont'd): let $X_1, \ldots, X_n \sim$ iid $N(\mu, 1)$ and consider testing $\mathbb{H}_0 : \mu = \mu_0$ using a LR test that rejects if $|\bar{x}_n - \mu_0| \geq \sqrt{-2(\ln c)/n}$. Then

$$\mathbb{P}\left(|\bar{x}_n - \mu_0| \geq \sqrt{-2(\ln c)/n}\right) \;=\; \mathbb{P}\left(\frac{|\bar{x}_n - \mu_0|}{1/\sqrt{n}} \geq \sqrt{-2(\ln c)}\right) \;=\; \alpha$$

and since $\frac{\bar{x}_n - \mu_0}{1/\sqrt{n}} \sim N(0,1)$ we can choose c such that $\sqrt{-2(\ln c)}$ yields the probability above being equal to α. This will be obtained at $\sqrt{-2(\ln c)} = z_{\alpha/2}$, which implies

$$c \;=\; \exp(-z_{\alpha/2}^2/2)$$

LR test for the exponential distribution

- example 2: let $(X_1, \ldots, X_n)$ be a random sample from an exponential population with pdf

$$f(x_i|\theta) = \begin{cases} e^{-(x_i-\theta)} & x_i \geq \theta \\ 0 & x_i < \theta \end{cases}$$

so the likelihood function is

$$f(\boldsymbol{x}|\theta) = \begin{cases} e^{-(\sum x_i - n\theta)} & x_{(1)} \geq \theta \\ 0 & x_{(1)} < \theta \end{cases}$$

and consider testing $\mathbb{H}_0 : \theta \leq \theta_0$ versus $\mathbb{H}_1 : \theta > \theta_0$

- if $x_{(1)} \geq \theta$, $\ell(\theta|\boldsymbol{x}) = \prod_{i=1}^n f(x_i|\theta)$ is an increasing function of θ. Then unrestricted maximum is obtained at $\hat{\theta} = x_{(1)}$ with maximum

$$\ell(\hat{\theta}|\boldsymbol{x}) \;=\; \ell(x_{(1)}|\boldsymbol{x}) \;=\; e^{-(\sum x_i - n x_{(1)})}$$

LR test for the exponential distribution

- now for the restricted maximum $\ell(\hat{\theta}_0|\boldsymbol{x})$
  - if $x_{(1)} \leq \theta_0$, then restriction is not binding and $\ell(\hat{\theta}_0|\boldsymbol{x}) = \ell(\hat{\theta}|\boldsymbol{x})$
  - if $x_{(1)} > \theta_0$, then $\hat{\theta}_0 = \theta_0$ and $\ell(\theta_0|\boldsymbol{x}) = e^{-(\sum x_i - n\theta_0)}$
- the likelihood test statistic is

$$\lambda(\boldsymbol{x}) = \begin{cases} 1 & x_{(1)} \leq \theta_0 \\ e^{-n(x_{(1)}-\theta_0)} & x_{(1)} > \theta_0 \end{cases}$$

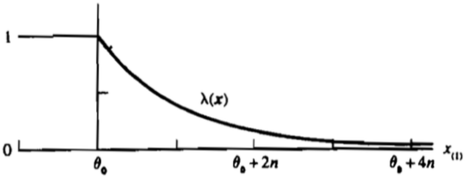


Figure 8.2.1. $\lambda(\mathbf{x})$, *a function only of* $x_{(1)}$.

LR test for the exponential distribution

- therefore, a test that rejects $\mathbb{H}_0$ if $\lambda(\boldsymbol{X}) \leq c$ is such that

$$e^{-n(x_{(1)}-\theta_0)} \leq c \Rightarrow -n(x_{(1)}-\theta_0) \leq \ln c \Rightarrow x_{(1)} \geq \theta_0 - \frac{\ln c}{n}$$

rejection region $\{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\} = \{\boldsymbol{x} : x_{(1)} \geq \theta_0 - (\ln c)/n\}$

- now find c that matches a desired size α. General fact:

$$\mathbb{P}(X_i \leq k) = \int_{\theta_0}^{k} e^{-(x-\theta_0)} dx = \left[-e^{-(x-\theta_0)}\right|_{\theta_0}^{k} = 1 - e^{-(x-\theta_0)}$$

therefore the probability that all $X_1, \ldots, X_n$ are greater than k is

$$\mathbb{P}\left(X_{(1)} \geq k\right) = e^{-n(k-\theta_0)}$$

- in the test, $k = \theta_0 - (\ln c)/n$, so we must choose c such that

$$e^{-n(\theta_0-(\ln c)/n-\theta_0)} = \alpha$$

which just implies that $c = \alpha$.

sufficient statistics are sufficient for LR tests

- is it a coincidence that likelihood ratio tests on the normal and exponential depended on sufficient statistics (respectively, $\bar{x}_n$ and $x_{(1)}$)?

- if $T(\boldsymbol{X})$ is a sufficient statistic for θ with pdf/pmf $g(t|\theta)$, then LR tests based on T and its likelihood function $\ell_*(\theta|t) = g(t|\theta)$ should be as good as LR tests based on $\ell(\theta|\boldsymbol{x})$

- theorem (**equivalence**): $\lambda_*\big(T(\boldsymbol{x})\big) = \lambda(\boldsymbol{x})$ for every $\boldsymbol{x}$ in the sample space if $T(\boldsymbol{X})$ is a sufficient statistic for θ

- proof: it follows from the factorization theorem that

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta\in\Theta_0}\ell(\theta|\boldsymbol{x})}{\sup_{\theta\in\Theta}\ell(\theta|\boldsymbol{x})} = \frac{\sup_{\theta\in\Theta_0} g\big(T(\boldsymbol{x})|\theta\big)h(\boldsymbol{x})}{\sup_{\theta\in\Theta} g\big(T(\boldsymbol{x})|\theta\big)h(\boldsymbol{x})} = \frac{\sup_{\theta\in\Theta_0}\ell_*\big(\theta|T(\boldsymbol{x})\big)}{\sup_{\theta\in\Theta}\ell_*\big(\theta|T(\boldsymbol{x})\big)} = \lambda_*\big(T(\boldsymbol{x})\big) \qquad \blacksquare$$

nuisance parameters do not annoy so much

- likelihood tests are also convenient if there are nuisance parameters, that is to say, parameters for which we have no inferential interest

- they do not affect the LR test construction method, though their presence might result in a different test

- example: suppose $X_1, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$ and that we wish to test $\mathbb{H}_0 : \mu \leq \mu_0$ against $\mathbb{H}_1 : \mu > \mu_0$

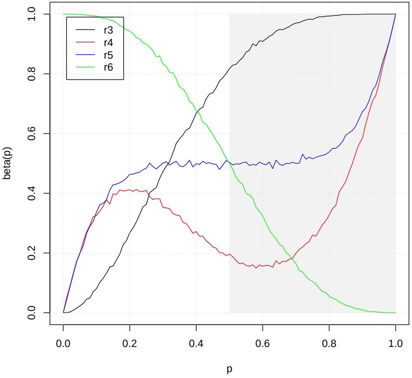
$$\begin{aligned}
\lambda(\boldsymbol{x}) &= \frac{\max_{\mu \leq \mu_0, \sigma^2 \geq 0} \ell(\mu, \sigma^2 | \boldsymbol{x})}{\max_{\mu \in \mathbb{R}, \sigma^2 \geq 0} \ell(\mu, \sigma^2 | \boldsymbol{x})} \\
&= \frac{\max_{\mu \leq \mu_0, \sigma^2 \geq 0} \ell(\mu, \sigma^2 | \boldsymbol{x})}{\ell(\bar{x}_n, \hat{\sigma}^2 | \boldsymbol{x})} \\
&= \begin{cases} 1 & \text{if } \bar{x}_n \leq \mu_0 \\ \frac{\ell(\mu_0, \hat{\sigma}^2 | \boldsymbol{x})}{\ell(\bar{x}_n, \hat{\sigma}^2 | \boldsymbol{x})} & \text{if } \bar{x}_n > \mu_0 \end{cases}
\end{aligned}$$

Contents

most powerful tests

- general principle: a good test should have for a given probability of type-I error the smallest possible probability of type-II error
- definition: unbiased tests are more likely to reject $\mathbb{H}_0$ if the null is false than if it is true, and hence their power functions are such that $\beta(\boldsymbol{\theta}_1) \geq \beta(\boldsymbol{\theta}_0)$ if $\boldsymbol{\theta}_0 \in \Theta_0$ and $\boldsymbol{\theta}_1 \in \Theta_1$

(un)biased tests here?

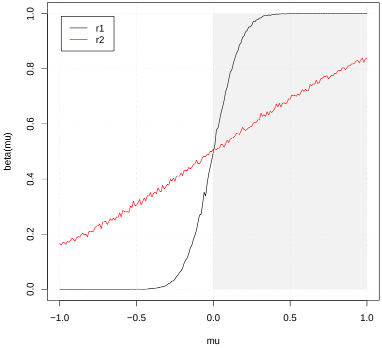


most powerful tests

- definition: let $\mathcal{C}$ be a class of tests for $\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$ versus $\mathbb{H}_1 : \boldsymbol{\theta} \in \Theta_1$, then a test in $\mathcal{C}$ with power function $\beta(\boldsymbol{\theta})$ is a uniformly most powerful class $\mathcal{C}$ test if $\beta(\boldsymbol{\theta}) \geq \tilde{\beta}(\boldsymbol{\theta})$ for every $\boldsymbol{\theta} \in \Theta_1$ and every $\tilde{\beta}(\boldsymbol{\theta})$ that is a power function of a test in class $\mathcal{C}$

- we typically consider the class $\mathcal{C}$ of all level α tests, because we have to control anyway the probability of type I error

which one is most powerful?



Neyman-Pearson lemma

- theorem (**Neyman-Pearson lemma**) (CB 8.3.12): consider testing $\mathbb{H}_0 : \theta = \theta_0$ versus $\mathbb{H}_1 : \theta = \theta_1$, where the pdf/pmf corresponding to θ_i is $f(\boldsymbol{x}|\theta_i)$ for $i = 0, 1$ using a test with rejection region R such that

$$\begin{aligned} \boldsymbol{x} \in R \quad &\text{if} \quad f(\boldsymbol{x}|\theta_1) > kf(\boldsymbol{x}|\theta_0) \\ \boldsymbol{x} \in R^c \quad &\text{if} \quad f(\boldsymbol{x}|\theta_1) < kf(\boldsymbol{x}|\theta_0) \end{aligned}$$

  for some $k \geq 0$, and $\mathbb{P}_{\theta_0}(\boldsymbol{x} \in R) = \alpha$, then

  (i) (Sufficiency) such a test is a UMP level α test

  (ii) (Necessity) if there exists such a test, then every UMP level α test is a size α test

  (iii) (Necessity) every UMP level α test has a rejection region of the above form, except perhaps on a set A of null measure under θ_0 and θ_1: $\mathbb{P}_{\theta_0}(\boldsymbol{X} \in A) = \mathbb{P}_{\theta_1}(\boldsymbol{X} \in A) = 0$

- remember: for $0 \leq \alpha \leq 1$, a test with power function $\beta(\boldsymbol{\theta})$ has size α if

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \beta(\boldsymbol{\theta}) = \alpha$$

  whereas it has level α if $\sup_{\boldsymbol{\theta} \in \Theta_0} \beta(\boldsymbol{\theta}) \leq \alpha$

Neyman-Pearson lemma

- proof (i): let $\phi(\boldsymbol{x})$ denote the test function of the Neyman-Pearson test, taking value 1 if $\boldsymbol{x} \in R$ and zero if $\boldsymbol{x} \in R^c$, and $\tilde{\phi}(\boldsymbol{x})$ any other level α test function $0 \leq \tilde{\phi}(\boldsymbol{x}) \leq 1$

- the Neyman-Pearson rejection region implies that, for every sample point $\boldsymbol{x}$,

$$\left[\phi(\boldsymbol{x}) - \tilde{\phi}(\boldsymbol{x})\right]\left[f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)\right] \quad \geq \quad 0$$

and hence

$$\begin{aligned} 0 \quad &\leq \quad \int \left[\phi(\boldsymbol{x}) - \tilde{\phi}(\boldsymbol{x})\right]\left[f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)\right] \mathrm{d}\boldsymbol{x} \\ &= \quad \beta(\theta_1) - \tilde{\beta}(\theta_1) - k\left[\beta(\theta_0) - \tilde{\beta}(\theta_0)\right] \\ &= \quad \beta(\theta_1) - \tilde{\beta}(\theta_1) - k\left[\alpha - \tilde{\beta}(\theta_0)\right] \\ &\leq \quad \beta(\theta_1) - \tilde{\beta}(\theta_1) \end{aligned}$$

for $k \geq 0$ given that $\alpha - \tilde{\beta}(\theta_0) \geq 0$, hence $\beta(\theta_1) \geq \tilde{\beta}(\theta_1)$. That is, the NP test has greater power than any other test. ■

Neyman-Pearson lemma

- proof (ii): let now $\tilde{\phi}(\boldsymbol{x})$ denote any UMP level α test function and note that, by sufficiency, $\phi(\boldsymbol{x})$ is also UMP level α test. Because ϕ and $\tilde{\phi}$ are both UMP tests, $\beta(\theta_1) = \tilde{\beta}(\theta_1)$, it then follows from

$$\beta(\theta_1) - \tilde{\beta}(\theta_1) - k\left[\beta(\theta_0) - \tilde{\beta}(\theta_0)\right] \quad \geq \quad 0$$

with $k > 0$ that $-k\left[\beta(\theta_0) - \tilde{\beta}(\theta_0)\right] \geq 0 \Rightarrow \beta(\theta_0) - \tilde{\beta}(\theta_0) \leq 0$. Then

$$0 \quad \leq \quad \alpha - \tilde{\beta}(\theta_0) \quad = \quad \beta(\theta_0) - \tilde{\beta}(\theta_0) \quad \leq \quad 0$$

and hence $\tilde{\beta}(\theta_0) = \alpha$ and $\tilde{\phi}$ is in fact a size α test.

- proof (iii): this implies that

$$\underbrace{\beta(\theta_1) - \tilde{\beta}(\theta_1)}_{=0} - k \underbrace{\left[\beta(\theta_0) - \tilde{\beta}(\theta_0)\right]}_{=0} \quad = \quad \int \left[\phi(\boldsymbol{x}) - \tilde{\phi}(\boldsymbol{x})\right]\left[f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)\right] \mathrm{d}\boldsymbol{x}$$

which implies only if $\tilde{\phi}$ has the same rejection region of the Neyman-Pearson test, except on a set A with $\int_A f(\boldsymbol{x}|\theta_i)\,\mathrm{d}x = 0, \forall i = 1, 2.$ ■

example

- example 1 (CB 8.20): let X be a random variable with distribution under $\mathbb{H}_0$ and $\mathbb{H}_1$ given by

x	1	2	3	4	5	6	7
$f(x\|\mathbb{H}_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$f(x\|\mathbb{H}_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

  use the Neyman-Pearson lemma to find the most powerful test for $\mathbb{H}_0$ against $\mathbb{H}_1$ with size $\alpha = 0.04$. Compute the probability of type-II error.

- solution: by the NP lemma, we should define the rejection region

$$\boldsymbol{x} \in R \quad \text{if} \quad f(\boldsymbol{x}|\theta_1) > kf(\boldsymbol{x}|\theta_0)$$

  that is, $\frac{f(\boldsymbol{x}|\theta_1)}{f(\boldsymbol{x}|\theta_0)} > k$.

x	1	2	3	4	5	6	7
$\frac{f(x\|\mathbb{H}_1)}{f(x\|\mathbb{H}_0)}$	6	5	4	3	2	1	0.84

  so rejecting for large values of k corresponds to small values of x. A test with size $\alpha = 0.04$ is such that $\mathbb{P}(X \leq c|\mathbb{H}_0) = 0.04$, which is achieved at $c = 4$. The type-II error is $\mathbb{P}(X \in \{5, 6, 7\}|\mathbb{H}_1) = .82$.

UMP test for the binomial probability

- example 2: let $X \sim \text{Bin}(2, p)$ and consider testing $\mathbb{H}_0 : p = 1/2$ against $\mathbb{H}_1 : p = 3/4$ using the pmf ratios

$$\frac{f\left(0|p=\frac{3}{4}\right)}{f\left(0|p=\frac{1}{2}\right)} = \frac{\frac{1}{4}\frac{1}{4}}{\frac{1}{2}\frac{1}{2}} = \frac{1}{4} \;;\; \frac{f\left(1|p=\frac{3}{4}\right)}{f\left(1|p=\frac{1}{2}\right)} = \frac{2\frac{1}{4}\frac{3}{4}}{2\frac{1}{2}\frac{1}{2}} = \frac{3}{4} \;;\; \frac{f\left(2|p=\frac{3}{4}\right)}{f\left(2|p=\frac{1}{2}\right)} = \frac{\frac{3}{4}\frac{3}{4}}{\frac{1}{2}\frac{1}{2}} = \frac{9}{4}$$

- if we choose...
  - $k > \frac{9}{4}$ yields the UMP with level $\alpha = 0$
  - $\frac{3}{4} < k < \frac{9}{4}$, the test that rejects $\mathbb{H}_0$ if $X = 2$ is UMP with level

$$\alpha = \mathbb{P}\left(X = 2|\theta = \frac{1}{2}\right) = \frac{1}{4}$$

  - $\frac{1}{4} < k < \frac{3}{4}$, the test that rejects $\mathbb{H}_0$ if $X = \{1, 2\}$ is UMP with level

$$\alpha = \mathbb{P}\left(X = 1 \text{ or } 2|\theta = \frac{1}{2}\right) = \frac{3}{4}$$

  - $k < \frac{1}{4}$ yields the UMP with level $\alpha = 1$

how about sufficiency?

- corollary of NP lemma: suppose $T(\boldsymbol{X})$ is sufficient for θ, with pdf/pmf $g(t|\theta_i)$ corresponding to θ_i ($i = 0, 1$), then any test based on $T(\boldsymbol{X})$ with rejection region S such that

$$\begin{aligned} t \in S \quad &\text{if} \quad g(t|\theta_1) > kg(t|\theta_0) \\ t \in S^c \quad &\text{if} \quad g(t|\theta_1) < kg(t|\theta_0) \end{aligned}$$

  for some $k \geq 0$, where $\mathbb{P}_{\theta_0}\big(T(\boldsymbol{x}) \in S\big) = \alpha$, is a UMP level α test.

- proof: in terms of the original sample $\boldsymbol{X}$, the test based on $T(\boldsymbol{X})$ has rejection region $R = \{\boldsymbol{x} : \ T(\boldsymbol{x}) \in S\}$ such that

$$\begin{aligned} \boldsymbol{x} \in R \quad &\text{if} \quad f(\boldsymbol{x}|\theta_1) = g\big(T(\boldsymbol{x})|\theta_1\big)h(\boldsymbol{x}) > kg\big(T(\boldsymbol{x})|\theta_0\big)h(\boldsymbol{x}) = kf(\boldsymbol{x}|\theta_0) \\ \boldsymbol{x} \in R^c \quad &\text{if} \quad f(\boldsymbol{x}|\theta_1) = g\big(T(\boldsymbol{x})|\theta_1\big)h(\boldsymbol{x}) < kg\big(T(\boldsymbol{x})|\theta_0\big)h(\boldsymbol{x}) = kf(\boldsymbol{x}|\theta_0) \end{aligned}$$

  and $\mathbb{P}_{\theta_0}(\boldsymbol{X} \in R) = \mathbb{P}_{\theta_0}\big(T(\boldsymbol{X}) \in S\big)$, so it is also a UMP level α test by the Neyman-Pearson lemma. ∎

UMP test for the normal mean

- example 3: let $X_1, \ldots, X_n \sim$ iid $N(\mu, 1)$ and consider testing $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu = \mu_1$, with $\mu_0 > \mu_1$. We had that

$$f(\boldsymbol{x}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - \mu)^2}{2\sigma^2}\right\}$$

so, applying the NP lemma,

$$\frac{f(\boldsymbol{x}|\mu_1, 1)}{f(\boldsymbol{x}|\mu_0, 1)} = \exp\left\{\frac{n(\bar{x}_n - \mu_0)^2 - n(\bar{x}_n - \mu_1)^2}{2\sigma^2}\right\} > k$$

so that $(\bar{x}_n - \mu_0)^2 - (\bar{x}_n - \mu_1)^2 > \frac{1}{n} 2\sigma^2 \ln k$. We need to isolate $\bar{x}_n$:

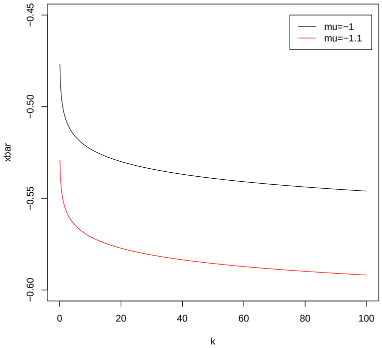
$$\begin{aligned} (\bar{x}_n - \mu_0)^2 - (\bar{x}_n - \mu_1)^2 &= \bar{x}_n^2 - 2\bar{x}_n\mu_0 + \mu_0^2 - \bar{x}_n^2 + 2\bar{x}_n\mu_1 - \mu_1^2 \\ &= -2\bar{x}_n\mu_0 + \mu_0^2 + 2\bar{x}_n\mu_1 - \mu_1^2 \end{aligned}$$

and given that $\mu_1 - \mu_0 < 0$, the rejection region is of the format

$$\bar{x}_n < \frac{\frac{1}{n} 2\sigma^2 \ln k - \mu_0^2 + \mu_1^2}{2(\mu_1 - \mu_0)} \iff \bar{x}_n < c$$

UMP test for the normal mean

- example 3 (cont'd): for $\mu_0 = 0$, $n = 100$ and $\sigma^2 = 1$, this function looks like



equivalent to say that, for any k, there is a c such that $\bar{x}_n < c$. This means that a test with rejection region

$$\bar{x}_n \quad < \quad c \quad = \quad \theta_0 - \frac{\sigma z_\alpha}{\sqrt{n}}$$

is the UMP level α test.

composite hypothesis

- $\mathbb{H}_0$ and $\mathbb{H}_1$ in the Neyman-Pearson lemma are simple hypotheses in that they specify only one possible distribution for sample $\boldsymbol{X}$, i.e., $\mathbb{H}_0$ and $\mathbb{H}_1$ are singletons.

- composite hypotheses: in most realistic problems, the hypotheses of interest specify more than one possible distribution for the sample

one-sided tests:	$\mathbb{H}_0 : \mu \leq \mu_0$ vs $\mathbb{H}_1 : \mu > \mu_0$
two-sided tests:	$\mathbb{H}_0 : \mu = \mu_0$ vs $\mathbb{H}_1 : \mu \neq \mu_0$

- is the Neyman-Pearson lemma applicable? We shall defer this question to when we talk about union-intersection tests.

one-sided tests

- a large class of problems that admit UMP level α tests involve one-sided hypotheses and pdfs/pmfs with the monotone LR property
- definition: a family of pdfs/pmfs $\{g(t|\theta) : \theta \in \Theta\}$ for a univariate random variable T with parameter $\theta \in \mathbb{R}$ has a monotone likelihood ratio if for every $\theta_2 > \theta_1$, $g(t|\theta_2)/g(t|\theta_1)$ is a monotone function of t on $\{t : g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\}$
- interestingly, any exponential family with $g(t|\theta) = h(t)c(\theta)\exp\{w(\theta)t\}$ has an MLR if $w(\theta)$ is nondecreasing
- theorem (**Karlin-Rubin**) (CB 8.3.17): consider testing $\mathbb{H}_0 : \theta \leq \theta_0$ versus $\mathbb{H}_1 : \theta > \theta_0$ using a sufficient statistic T whose pdf/pmf satisfies the MLR property, then the UMP level α test rejects the null if $T > t_0$ with $\mathbb{P}_{\theta_0}(T > t_0) = \alpha$.

one-sided tests

- example: $X_1, \ldots, X_n$ i.i.d. standard normal. Consider testing $\mathbb{H}_0' : \theta \geq \theta_0$ versus $\mathbb{H}_1' : \theta < \theta_0$.

- since $\bar{X}_n$ is sufficient and distribution has a monotone likelihood ratio, we can apply the Karlin-Rubin theorem which states that we should reject the null if

$$\bar{x}_n \quad < \quad \theta_0 - \frac{\sigma z_\alpha}{\sqrt{n}}$$

and the power function is

$$\beta(\theta) \quad = \quad \mathbb{P}_\theta\left(\bar{X}_n < \theta_0 - \frac{\sigma z_\alpha}{\sqrt{n}}\right)$$

which is a decreasing function of θ_0. The value α is given by

$$\sup_{\theta \geq \theta_0} \beta(\theta) \quad = \quad \beta(\theta_0) \quad = \quad \alpha$$

R codes: computations with UMP tests

- example: let $\{X_1, \ldots, X_n\} \sim N(\mu, \sigma^2)$ i.i.d. with σ^2 known, and consider testing $\mathbb{H}_0 : \mu \leq 0$ against $\mathbb{H}_1 : \mu > 0$.
  - test 1: take the test statistic $\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > c$, where $c = z_\alpha$, with rejection region

    $$R_1 = \left\{ X : \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \right\} = \left\{ X : \bar{X}_n > \mu_0 + \sigma \frac{z_\alpha}{\sqrt{n}} \right\}$$

    which is the UMP test of level α.
  - test 2: using only the first 5 observations, also with level α

    $$R_2 = \left\{ X : \frac{\bar{X}_5 - \mu_0}{\sigma/\sqrt{5}} > z_\alpha \right\} = \left\{ X : \bar{X}_5 > \mu_0 + \sigma \frac{z_\alpha}{\sqrt{5}} \right\}$$

R codes: computations with UMP tests

- test 3:

$$R_3 = \left\{ X : \sum_{i=1}^{n} \frac{X_i^2}{\sigma^2} > \kappa \;\; \text{if} \;\; \bar{X}_n > 0 \right\}$$

and we need to find κ such that the probability of rejecting is α.

$$\mathbb{P}(X \in R_3) = \mathbb{P}\left\{ \sum_{i=1}^{n} \frac{X_i^2}{\sigma^2} > \kappa \,\middle|\, \bar{X}_n > 0 \right\} \cdot \mathbb{P}\left(\bar{X}_n > 0\right)$$

while

$$\mathbb{P}\left(\bar{X}_n < 0\right) = \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} < -\sqrt{n}\frac{\mu}{\sigma}\right) = \mathbb{P}\left(Z < \sqrt{n}\frac{\mu}{\sigma}\right)$$

given that $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1)$. Conditional of $\bar{X}_n > 0$, $\sum_{i=1}^{n} \frac{X_i}{\sigma^2} \sim \chi_n^2$ from the χ_n^2 distribution, so we can find a $\kappa = q_{\alpha^*}$ such that $\mathbb{P}\left(\sum_{i=1}^{n} \frac{X_i}{\sigma^2} < q_{\alpha^*}\right) = \alpha^*$.

- taking $\mu = 0$,

$$\mathbb{P}(X \in R_3) = 0.5(1 - \alpha^*) = \alpha \implies \alpha^* = 1 - 2\alpha$$

R codes: computations with UMP tests

```
n <- 500
sigma2 <- 1
alpha <- 0.05
mu <- 0

test1 <- function(x){
  TS <- sqrt(n)*mean(x)/sqrt(sigma2)
  testOutcome <- (TS > qnorm(1-alpha))
}

test2 <- function(x){
  TS <- sqrt(5)*mean(x[1:5])/sqrt(sigma2)
  testOutcome <- (TS > qnorm(1-alpha))
}

test3 <- function(x){
  TS <- sum(x^2/sigma2)
  testOutcome <- (TS > qchisq(1-2*alpha,n))
  if (mean(x) < 0) {testOutcome=0}
  testOutcome
}
```

R codes: computations with UMP tests

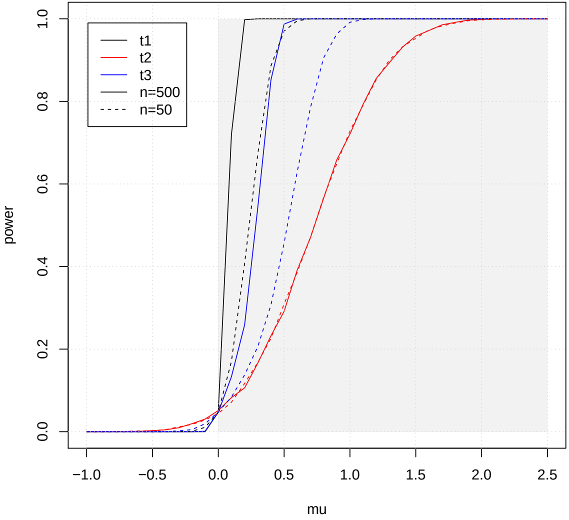
```
testRejFreq <- function(mu){
  testRej <- matrix(0,5000,3)
  for (i in 1:5000){
    x <- rnorm(n,mean=mu,sd=sqrt(sigma2))
    testRej[i,1] <- test1(x)
    testRej[i,2] <- test2(x)
    testRej[i,3] <- test3(x)
  }
  testRejF <- colMeans(testRej)
}
mu <- seq(-1,2.5,by=0.1)
```

R codes: computations with UMP tests

table: rejection frequencies

$n = 50$					
	$\mu = 0$	$\mu = 0.1$	$\mu = 0.2$	$\mu = 0.5$	$\mu = 1$
test 1	0.0472	0.1706	0.4078	0.9702	1.0000
test 2	0.0444	0.0714	0.1168	0.3094	0.7284
test 3	0.0478	0.0840	0.1376	0.4570	0.9916
$n = 500$					
	$\mu = 0$	$\mu = 0.1$	$\mu = 0.2$	$\mu = 0.5$	$\mu = 1$
test 1	0.0534	0.7218	0.9986	1.0000	1.0000
test 2	0.0484	0.0800	0.1214	0.3060	0.6436
test 3	0.0500	0.1376	0.2576	0.9872	1.0000

R codes: computations with UMP tests



Contents

summary so far

- summary of results so far

$\mathbb{H}_0$	$\mathbb{H}_1$	UMP test?	example of R
$\mu = \mu_0$	$\mu = \mu_1$	Neyman-Person lemma	$\bar{x}_n < c$
$\mu = \mu_0$	$\mu > \mu_1$	*(deferred)*	
$\mu \leq \mu_0$	$\mu > \mu_0$	Karlin-Rubin theorem	$\bar{x}_n < c$
$\mu = \mu_0$	$\mu \neq \mu_0$	explore now	

UMPU tests

- if there is no UMP level α test within the class of all tests, we might try to find a UMP level α test within the class of unbiased tests.

- the next example shows that it is not trivial to find an UMP test within the class of α-sized tests.

- example: let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ i.i.d. with σ^2 known, and consider testing $\mathbb{H}_0 : \mu = \mu_0$ versus $\mathbb{H}_1 : \mu \neq \mu_0$.
  - test 1: rejects $\mathbb{H}_0$ if $\bar{X}_n < \mu_0 - \frac{\sigma z_\alpha}{\sqrt{n}}$. The power function is for the test with size α is

$$\begin{aligned}\beta_1(\mu) &= \mathbb{P}_\mu\left(\bar{X}_n < \mu_0 - \frac{\sigma z_\alpha}{\sqrt{n}}\right) = \mathbb{P}_\mu\left(\bar{X}_n - \mu < \mu_0 - \mu - \frac{\sigma z_\alpha}{\sqrt{n}}\right) \\ &= \mathbb{P}_\mu\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < -z_\alpha + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) = \mathbb{P}\left(Z > z_\alpha - \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)\end{aligned}$$

UMPU tests

- example (cont'd): test 2: rejects $\mathbb{H}_0$ if $\bar{X}_n > \mu_0 + \frac{\sigma z_\alpha}{\sqrt{n}}$

  $$\beta_2(\mu) \quad = \quad \mathbb{P}\left(Z > z_\alpha + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$$

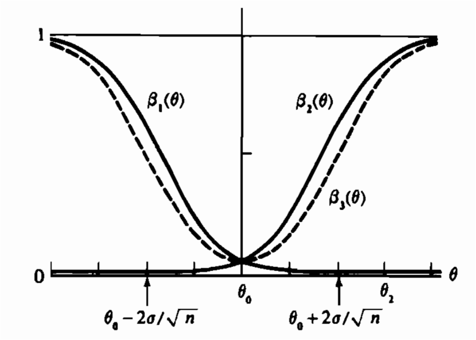
  and take a point $\mu_1 < \mu_0$

  $$\beta_1(\mu_1) \quad = \quad \mathbb{P}\left(Z > z_\alpha - \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right) \quad > \quad \mathbb{P}\left(Z > z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right) \quad = \quad \beta_2(\mu)$$

  because $\mu_0 - \mu_1 > 0$. Now, if $\mu_2 > \mu_0$, we will have that $\mu_0 - \mu_2 < 0$ and the inequality will reverse, that is, $\beta_1(\mu_2) < \beta_2(\mu_2)$.

UMPU tests

- the problem is that the class of tests is too wide: we may restrict the class of tests to search among α-level unbiased tests.
- test 3: reject $\mathbb{H}_0$ if $\bar{X}_n > \theta_0 + \frac{\sigma z_{\alpha/2}}{\sqrt{n}}$ or $\bar{X}_n > \theta_0 - \frac{\sigma z_{\alpha/2}}{\sqrt{n}}$



- it happens that this test is the UMP test
- note that there is a loss of power compared to tests 1 and 2 at some parameter points

Contents

union-intersection tests

- in some situations, tests for complicated null hypotheses can be developed from tests for simpler null hypotheses

- suppose that the null hypothesis can be conveniently expressed as

$$\mathbb{H}_0 : \boldsymbol{\theta} \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma$$

  and there are tests available for each testing problem $\mathbb{H}_0^{(\gamma)} : \boldsymbol{\theta} \in \Theta_0^\gamma$ versus $\mathbb{H}_1^{(\gamma)} : \boldsymbol{\theta} \in \Theta_1^\gamma$, with rejection regions $\{\boldsymbol{x} : \ T_\gamma(\boldsymbol{x}) \in R_\gamma\}$

- if any hypothesis $\mathbb{H}_0^{(\gamma)}$ is rejected, then $\mathbb{H}_0$ must also be rejected. Then the rejection region for the UI test is $\bigcup_{\gamma \in \Gamma} \{\boldsymbol{x} : \ T_\gamma(\boldsymbol{x}) \in R_\gamma\}$

- in some situations, it is possible to simplify the expression for the rejection region of a union-intersection test

$$\bigcup_{\gamma \in \Gamma} \{\boldsymbol{x} : \ T_\gamma(\boldsymbol{x}) \in R_\gamma\} \ = \ \{\boldsymbol{x} : \sup_{\gamma \in \Gamma} T_\gamma(\boldsymbol{x}) > c\}$$

  and hence $T(\boldsymbol{x}) = \sup_{\gamma \in \Gamma} T_\gamma(\boldsymbol{x})$

Gaussian union-intersection tests

- example: let $X_1, \dots, X_n \sim$ iid $N(\mu, \sigma^2)$ and consider testing $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu \neq \mu_0$
- we may write the null hypothesis as the intersection of $\mathbb{H}_0^L : \{\mu : \mu \leq \mu_0\}$ and $\mathbb{H}_0^U : \{\mu : \mu \geq \mu_0\}$

$$\text{LR tests} \quad \begin{cases} \text{reject } \mathbb{H}_0^L : \quad \mu \leq \mu_0 \quad \text{if} \quad \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n} \geq t_L \\ \text{reject } \mathbb{H}_0^U : \quad \mu \geq \mu_0 \quad \text{if} \quad \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n} \leq t_U \end{cases}$$

- union-intersection test

$$\text{reject } \mathbb{H}_0 : \quad \mu = \mu_0 \quad \text{if} \quad t_L \leq \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n} \quad \text{or} \quad \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n} \leq t_U,$$

which coincides with the two-sided LR t-test if $t_L = -t_U \geq 0$ and then we can write

$$\text{reject } \mathbb{H}_0 : \quad \mu = \mu_0 \quad \text{if} \quad \sqrt{n}\,\frac{|\bar{X}_n - \mu_0|}{S_n} \geq t_L$$

which is also called the two-sided t-test

union-intersection test and Neyman-Pearson lemma

- let $X_1, \ldots, X_n \sim$ iid $N(\mu, 1)$. From the NP lemma, the α-level uniformly most powerful test for $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu = \mu_1$, $\mu_1 < \mu_0$, has rejection region

$$R = \left\{ x : \bar{x}_n < \mu_0 - \frac{\sigma z_\alpha}{\sqrt{n}} \right\}$$

- now consider testing $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu < \mu_0$. We can write

$$\begin{aligned} \mathbb{H}_0^{(\gamma)} &: \quad \mu = \mu_0 \\ \mathbb{H}_1^{(\gamma)} &: \quad \mu = \gamma \end{aligned}$$

  with $\gamma \in \Gamma = \{\gamma : \gamma < \mu_0, \gamma \in \mathbb{R}\}$, which is a union-intersection test.

- notice that, for each of these tests, the rejection region R is unchanged. It follows that the rejection region for the UI test is

$$\bigcup_{\gamma \in \Gamma} \{\boldsymbol{x} : \ T_\gamma(\boldsymbol{x}) \in R_\gamma\} \quad = \quad R$$

  and also $\sup_{\gamma \in \Gamma} T_\gamma(\boldsymbol{x}) = T(x)$.

- note that each of those tests are the UMP test individually.. it follows that rejection region R also constitutes the UMP for the composite hypothesis!

intersection-union tests

- suppose that we may conveniently express the null as a union

$$\mathbb{H}_0 : \boldsymbol{\theta} \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma$$

  and there are tests available for each testing problem $\mathbb{H}_0^{(\gamma)} : \boldsymbol{\theta} \in \Theta_0^\gamma$ versus $\mathbb{H}_1^{(\gamma)} : \boldsymbol{\theta} \in \Theta_1^\gamma$, with rejection regions $\{\boldsymbol{x} : T_\gamma(\boldsymbol{x}) \in R_\gamma\}$

- if all hypotheses $\mathbb{H}_0^{(\gamma)}$ is rejected, then $\mathbb{H}_0$ must be rejected. The rejection region for the IU test is $\bigcap_{\gamma \in \Gamma} \{\boldsymbol{x} : T_\gamma(\boldsymbol{x}) \in R_\gamma\}$

- in some situations, it is possible to simplify the expression for the rejection region of a intersection-union test

$$\bigcap_{\gamma \in \Gamma} \{\boldsymbol{x} : T_\gamma(\boldsymbol{x}) \in R_\gamma\} = \{\boldsymbol{x} : \inf_{\gamma \in \Gamma} T_\gamma(\boldsymbol{x}) \geq c\}$$

  and hence $T(\boldsymbol{x}) = \inf_{\gamma \in \Gamma} T_\gamma(\boldsymbol{x})$

Contents

p-values

- so far, a statistical test would report only whether $\mathbb{H}_0$ got accepted or rejected at a certain α-level, but not by how much

- p-values are another way of conveying information about the outcome of the statistical test: what is the minimum α such that $\mathbb{H}_0$ is rejected?

$$\mathbb{H}_0 \text{ rejected } \alpha = 0.10$$
$$\mathbb{H}_0 \text{ rejected at } \alpha = 0.05$$
$$\mathbb{H}_0 \text{ not rejected at } \alpha = 0.01$$

  so lower values are indicative of "more convincing" rejections

- definition: the p-value is the smallest significance level such that x is in the rejection region

$$p(\boldsymbol{x}) \quad = \quad \inf\{\alpha : \boldsymbol{x} \in R_\alpha\}$$

  where R_α is the rejection region at significance level α

p-values

- example: take our well-known rejection region

$$R_\alpha = \left\{ x : \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \right\}$$

for the test of $\mathbb{H}_0 : \mu \leq \mu_0$ against $\mathbb{H}_1 : \mu > \mu_0$. Note that

$$\left\{ x : \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \right\} = \left\{ x : \Phi\left(\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right) > \alpha \right\}$$

for a given x, the p-value is the infimum α such that $\Phi\left(\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}\right) > \alpha$ holds,

$$p = \Phi\left(\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right)$$

Contents

Contents

inference and set estimation

- we would like to make statements of the form $\boldsymbol{\theta} \in C(\boldsymbol{x})$, where the set estimate $C(\boldsymbol{x}) \subset \Theta$ depends only on the realization of the sample

- if θ is a scalar, $C(\boldsymbol{x})$ will typically be an interval

- our goal is to build intervals in which the true parameter lies with a certain probability

$$\begin{aligned} \mathbb{P}\left(\mu = \bar{X}_n\right) &= 0 \qquad \text{point estimation} \\ \mathbb{P}\left(\mu \in C(\boldsymbol{x})\right) &\geq 0 \qquad \text{interval estimation} \end{aligned}$$

- definition: an interval estimate of a parameter $\theta \in \Theta \subset \mathbb{R}$ is any pair of statistics $L(\boldsymbol{x})$ and $U(\boldsymbol{x})$ that satisfy $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$ for all $\boldsymbol{x} \in S_{\boldsymbol{X}}$, whereas the random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ corresponds to the interval estimator

- it is possible that $L(\boldsymbol{X}) = -\infty$ or $U(\boldsymbol{X}) = \infty$

- we will see soon that this topic is very much connected to hypothesis testing

interval coverage

- example: if $X_1, \ldots, X_4 \sim$ iid $N(\mu, 1)$, $[\bar{X}_4 - 1, \bar{X}_4 + 1]$ is a interval estimator of μ. The probability that $\mu \in C(\boldsymbol{x})$ is

$$\begin{aligned} \mathbb{P}\big(\mu \in [\bar{X}_4 - 1, \bar{X}_4 + 1]\big) &= \mathbb{P}(\bar{X}_4 - 1 \leq \mu \leq \bar{X}_4 + 1) = \mathbb{P}(|\bar{X}_4 - \mu| \leq 1) \\ = \mathbb{P}\left(\frac{|\bar{X}_4 - \mu|}{1/\sqrt{4}} \leq \frac{1}{1/\sqrt{4}}\right) &= \mathbb{P}(|Z| \leq 2) = 0.9544 \end{aligned}$$

- definition: the probability that the interval estimator $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ of θ includes the true parameter value θ is the coverage probability
- definition: the confidence coefficient of $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ is the infimum of the coverage probabilities, namely, $\inf_{\theta \in \Theta} \mathbb{P}_\theta(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})])$
- since θ is unknown, the best we can offer is the infimum coverage probability, that is to say, the confidence coefficient
- keep in mind that the random quantity is the interval $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$, but not θ, which is unknown but a fixed quantity
  - in the example above, the bounds depended on $\bar{X}_n$, which is a random quantity

scale uniform interval estimator

- example: let $X_1, \ldots, X_n \sim \text{iid}\, \mathrm{U}(0, \theta)$ and consider $[aX_{(n)}, bX_{(n)}]$ with $1 \leq a < b$. The coverage probability is

$$\mathbb{P}_\theta \left(aX_{(n)} \leq \theta \leq bX_{(n)} \right) \quad = \quad \mathbb{P} \left(\theta / b \leq X_{(n)} \leq \theta / a \right)$$

and cdf of $X_{(n)}$ is

$$\begin{aligned} \mathbb{P}\left(X_{(n)} \leq k \right) &= \prod_{i=1}^{n} \mathbb{P}\left(X_i \leq k \right) = \prod_{i=1}^{n} \int_0^k \frac{1}{\theta} \, \mathrm{d}x \\ &= \prod_{i=1}^{n} \frac{k}{\theta} = \left[\frac{k}{\theta} \right]^n \\ \mathbb{P}\left(\theta / b \leq X_{(n)} \leq \theta / a \right) &= \left[\frac{\theta / a}{\theta} \right]^n - \left[\frac{\theta / b}{\theta} \right]^n = a^{-n} - b^{-n} \end{aligned}$$

- example: (cont'd) consider alternatively $[X_{(n)} + c, X_{(n)} + d]$

$$\begin{aligned}\mathbb{P}_\theta(X_{(n)} + c \le \theta \le X_{(n)} + d) &= \mathbb{P}_\theta\left(\theta - d \le X_{(n)} \le \theta - c\right) \\ &= \left[\frac{\theta - c}{\theta}\right]^n - \left[\frac{\theta - d}{\theta}\right]^n \\ &= (1 - c/\theta)^n - (1 - d/\theta)^n\end{aligned}$$

which depends on θ, with confidence coefficient zero $(\theta \to \infty)$

interval estimator for a Gaussian sample mean

- example: if $X_1, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$, σ^2 known. Consider testing $\mathbb{H}_0 : \mu = \mu_0$ against $\mathbb{H}_1 : \mu \neq \mu_0$. We would then typically use the rejection region

$$R \quad = \quad \left\{ X : |\bar{X}_n - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

since test has size α, $\mathbb{P}(x \in R^c | \mu = \mu_0) = 1 - \alpha$. But

$$\begin{aligned} R^c &= \left\{ X : |\bar{X}_n - \mu_0| < z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\} = \left\{ X : -z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu_0 < z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\} \\ &= \left\{ X : -\bar{X}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < -\mu_0 < -\bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\} \\ &= \left\{ X : \bar{X}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\} \end{aligned}$$

i.e., there is a probability $1 - \alpha$ that μ_0 is in the interval above.

interval estimator for a Gaussian sample mean

- there is a clear correspondence between confidence sets and tests
  - the acceptance region is a set in the sample space such that $\mathbb{H}_0 : \mu = \mu_0$ is not rejected. It is a function of μ_0, but not data

$$A(\mu_0) \quad = \quad \left\{ \boldsymbol{x} : \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{x}_n \leq \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right\}$$

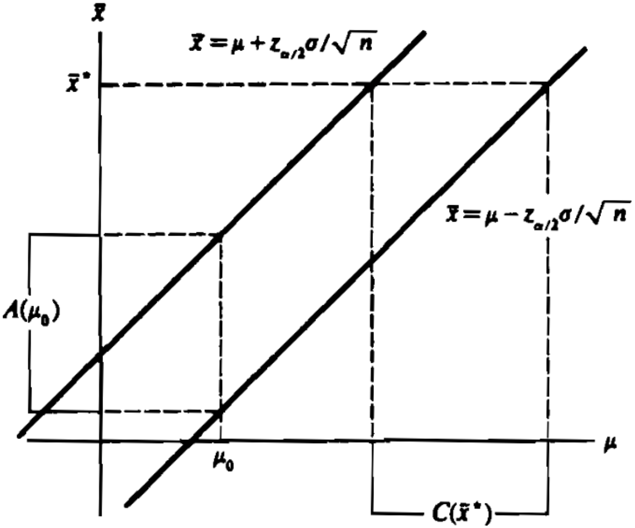
  - the confidence interval is set with plausible values of the parameters. It is a function of data, but not parameters

$$C(\boldsymbol{x}) \quad = \quad \left\{ \mu : \bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right\}$$

  - therefore

$$\boldsymbol{x} \in A(\mu_0) \quad \Longleftrightarrow \quad \mu_0 \in C(\boldsymbol{x})$$

interval estimator for a Gaussian sample mean



rejection regions and confidence intervals

- this notion can be made formal

- theorem (CB 9.2.2): for each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of level α of $\mathbb{H}_0 : \theta = \theta_0$. For each $\boldsymbol{x} \in \mathcal{X}$, define

$$C(\boldsymbol{x}) \quad = \quad \{\theta_0 : \boldsymbol{x} \in A(\theta_0)\}$$

then the random set $C(\boldsymbol{X})$ is a $1-\alpha$ confidence set. Conversely, let $C(\boldsymbol{X})$ be a $1-\alpha$ confidence set. Define

$$A(\theta_0) \quad = \quad \{\boldsymbol{x} : \theta_0 \in C(\boldsymbol{x})\}$$

then $A(\theta_0)$ is the acceptance region of a level-α test with $\mathbb{H}_0 : \theta = \theta_0$.

rejection regions and confidence intervals

- proof: $A(\theta_0)$ is acceptance region of a level-α test so $\mathbb{P}_{\theta_0}(\boldsymbol{X} \notin A(\theta_0)) \leq \alpha$ and $\mathbb{P}_{\theta_0}(\boldsymbol{X} \in A(\theta_0)) \geq 1 - \alpha$. Then

$$\mathbb{P}_\theta(\theta \in C(\boldsymbol{X})) \;=\; \mathbb{P}_\theta(\boldsymbol{X} \in A(\theta)) \;\geq\; 1 - \alpha$$

  so $C(\boldsymbol{X})$ is a $1 - \alpha$ confidence set.

- the type-I error probability for $\mathbb{H}_0 : \theta = \theta_0$ with acceptance region $A(\theta_0)$ is

$$\mathbb{P}_{\theta_0}(\boldsymbol{X} \notin A(\theta_0)) \;=\; \mathbb{P}_{\theta_0}(\theta_0 \notin C(\boldsymbol{X})) \;\leq\; \alpha$$

  so this is a α-level test. $\blacksquare$

Contents

how to gauge performance

- two relevant quantities:
  - size of the interval: length or volume
  - coverage probability: probability that true parameter is in the set
- the latter is generally a function of the parameter, so we usually take the infimum over the parameter space.
  - this is the confidence coefficient
- we will soon see that performances of tests and set estimates are closely connected

how to gauge performance

- question: we can optimize the length of an interval while keeping coverage probability constant at $1-\alpha$?

- example: take $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$, σ known. Then

$$\mathbb{P}\left(a \le \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le b\right) \quad = \quad \mathbb{P}\,(a \le Z \le b) \quad = \quad 1-\alpha$$

  gives the confidence interval

$$\left\{\mu : \bar{x}_n - b\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x}_n - a\frac{\sigma}{\sqrt{n}}\right\}$$

- what choice of a and b minimizes length while keeping $1-\alpha$ coverage?
  - minimize $b-a$ with $\mathbb{P}(a \le Z \le b) = 1-\alpha$

how to gauge performance

a	b	$P(Z < a)$	$P(Z > b)$	$b - a$
-1.34	2.33	.09	.01	3.67
-1.44	1.96	.075	.025	3.40
-1.65	1.65	.05	.05	3.30

- table suggests that $a = -b = 1.65$ is the optimum
- it is not a requirement that the interval should symmetric: this is a consequence of the symmetry of the normal distribution

how to gauge performance

- theorem (CB 9.3.2): let $f(x)$ be a unimodal pdf. If an interval $[a, b]$ satisfies

  (i) $\int_a^b f(x)dx = 1 - \alpha$

  (ii) $f(a) = f(b) > 0$

  (iii) $a \leq x^* \leq b$, where x^* is the mode of $f(x)$

  then $[a, b]$ is the shortest interval among all intervals such that $\int_a^b f(x)dx = 1 - \alpha$.

proof

optimality

- since there is a correspondence between confidence sets and hypothesis tests, there must be some correspondence between their optimalities

- consider a situation where $\boldsymbol{X} \sim f(\boldsymbol{x}|\theta)$ and construct a confidence set $C(\theta)$ for θ by inverting an acceptance region $A(\theta)$

- definition: the probability of true coverage is $\mathbb{P}_\theta(\theta \in C(\boldsymbol{X}))$

- definition: the probability of false coverage is the probability that θ' is covered when θ is the true parameter

$$\mathbb{P}_\theta(\theta' \in C(\boldsymbol{X})) \quad \text{if} \quad \theta' \neq \theta$$

- definition: the $1-\alpha$ confidence set that minimizes the probability of false coverage is called the uniformly most accurate confidence set (UMA)

optimality

- theorem (CB 9.3.5): let $\boldsymbol{X} \sim f(\boldsymbol{x}|\theta)$ where θ is real-valued. For each $\theta_0 \in \Theta$, let $A^*(\theta_0)$ be the UMP level-α acceptance region of a test of $\mathbb{H}_0 : \theta = \theta_0$ versus $\mathbb{H}_1 : \theta > \theta_0$. Let $C^*(\boldsymbol{x})$ be the $1 - \alpha$ confidence set formed by inverting the UMP acceptance regions. Then, for any other confidence region $C^*(\boldsymbol{X})$,

$$\mathbb{P}_\theta(\theta' \in C^*(\boldsymbol{X})) \;\leq\; \mathbb{P}_\theta(\theta' \in C(\boldsymbol{X}))$$

  that is, $C^*(\boldsymbol{X})$ is a UMA lower confidence bound.

- proof: let $\theta' < \theta$. Then

$$\begin{aligned} \mathbb{P}_\theta\left(\theta' \in C^*(\boldsymbol{X})\right) &= \mathbb{P}_\theta\left(\boldsymbol{X} \in A^*(\theta')\right) \\ \overset{UMP}{\leq} \mathbb{P}_\theta\left(\boldsymbol{X} \in A(\theta')\right) &= \mathbb{P}_\theta\left(\theta' \in C(\boldsymbol{X})\right) \end{aligned}$$

■

Contents

Reference:

- Casella and Berger, Ch. 8 and 9

Exercises:

- 8.1–8.3, 8.5–8.8, 8.12–8.19, 8.22(a), 8.27, 8.28, 8.32, 8.37, 8.51
- 9.1-9.14, 9.16-9.17, 9.23, 9.34-9.42, 9.47-9.52

how to gauge performance

- proof: let $[a', b']$ be any interval with $b' - a' < b - a$. There are two cases: $b' \leq a$ and $b' > a$. If $b' \leq a$, then $a' \leq b' \leq a \leq x^*$ and

$$\int_{a'}^{b'} f(x)dx \quad \leq \quad f(b')(b' - a')$$

since $x \leq b' \leq x^* \Rightarrow f(x) \leq f(b')$. Now,

$$f(b')(b' - a') \quad \leq \quad f(a)(b' - a')$$

since $f(x)$ is nondecreasing for $b' \leq a \leq x^*$ and

$$f(a)(b' - a') \quad < \quad f(a)(b - a) \quad \leq \quad \int_a^b f(x)dx \quad = \quad 1 - \alpha$$

since, using (ii) and (iii), $f(x) \geq f(a)$ for $a \leq x \leq b$. So $[a', b']$ cannot have the same coverage probability. Complete argument for $b' \leq a$ case. ■

back