STAT 346/446 Lecture 9

Hypothesis testing: Methods of finding tests

CB Sections 8.1 and 8.2, DS Section 9.1

Note: We skip last part of Lecture 8

Hypothesis testing

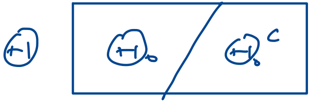
Statistical hypotheses

A statistical **hypothesis** is a statement about a population parameter(s).

Θ: parameter space

There are two complimentary hypothesis in a hypothesis problem:

- **Null hypothesis** H_0
- **Alternative hypothesis** H_1



Usually:

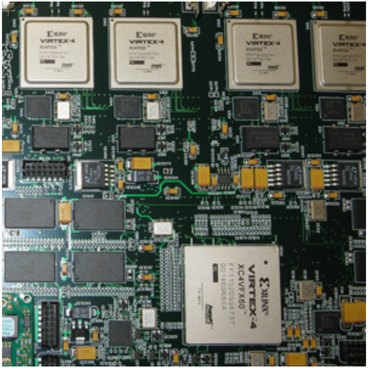
$$H_0 : \theta \in \Theta_0 \qquad \text{and} \qquad H_1 : \theta \in \Theta_0^c$$

For example

$$H_0 : \theta = 0 \qquad \text{and} \qquad H_1 : \theta \neq 0$$

Task: Use data to choose between H_0 and H_1

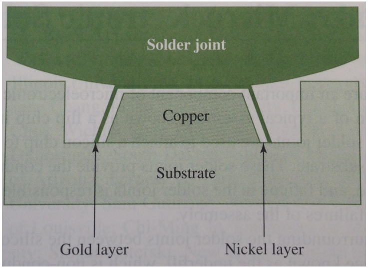
Example: Microelectronic Solder Joints



Picture: AmTECH Microelectronics, Inc.

- **Solder joints** are an important component of microelectronic assembles.
- Solder joints are used to attach a silicon chip to a printed circuit board, called *substrate*.
- Provide the conductive path from the silicon chip to the substrate
- Fatigue in the solder joints cause mechanical and electrical failures of the assembly

Example: Microelectronic Solder Joints – continued



- A critical component of the assembly is the bonding between the solder joint and the substrate
- A bond pad is created in the substrate made of copper, which is coated with thin layers of nickel and gold

- A researcher is investigating a new method for applying the nickel layer
- Thickness of the layer should be 2.775 microns on average
- Model: We will assume a normal model with known variance $\sigma^2 = 0.026^2$

Picture: "Probability and Statistics for Engineers and Scientists" by A Hayter

Microelectronic Solder Joints – data

- An assembly with 16 bond pads is examined and the nickel layer thickness is measured for each pad.
- *Before* collecting the data the thicknesses are *random variables*

$$X_i = \text{ the thickness of bond pad } i, \quad i = 1, 2, \ldots, 16$$

  - We assume that $X_i \overset{\text{ind.}}{\sim} N(\mu, 0.026^2)$, $i = 1, 2, \ldots, n$
  - Want to investigate whether or not $\mu = 2.775$ microns.
- The observed data is (in microns)

$$2.72, 2.79, 2.81, 2.75, 2.77, 2.76, 2.75, 2.75,$$
$$2.81, 2.75, 2.74, 2.77, 2.79, 2.78, 2.80, 2.76$$

That is,

$$x_1 = 2.72,\ x_2 = 2.79,\ x_3 = 2.81,\ \ldots,\ x_{16} = 2.76$$

Microelectronic Solder Joints – data

- Observed sample mean is

2.769

$$\bar{x} = \frac{1}{16}\sum_{i=1}^{16} x_i = 2.76875$$

≠ 2.775 but "very" close
↑ subjective

- Assuming that $X_1, X_2, \ldots, X_n$ are a *random sample* we know that $\overline{X}$ is the best unbiased estimator of the population mean μ
- This new method is supposed to deposit a nickel layer with an average thickness of 2.775 microns
- Based on data, our estimate of the average thickness μ is 2.76875 microns
- *Is there a statistically significant difference between the sample average and the target value?*

Microelectronic Solder Joints – decision rule

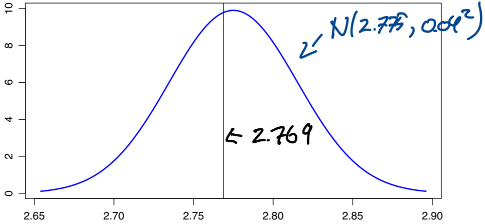
Recall:

- We expect the sample average to vary from sample to sample
- The sampling distribution of $\overline{X}$ is $N(\mu, 0.026^2/16)$
- *If* $\mu = 2.775$ the sampling distribution of $\overline{X}$ is

$$N(2.775, 0.026/16) = N(2.775, 0.04^2)$$

our $\bar{x} = 2.769$ is a sample from this distribution if Ho is true

- Is it plausible that 2.76875 comes from this distribution?



- Yes!
  - There is *no evidence* that the new method does not perform to standards.
  - We **don't reject** $H_0 : \mu = 2.775$

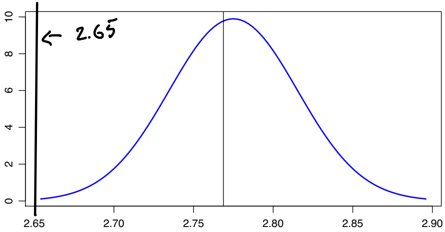
Microelectronic Solder Joints – decision rule

Recall:

- What if we had observed $\overline{x} = 2.65$?
- *If* $\mu = 2.775$ the sampling distribution of $\overline{X}$ is still

$$N(2.775, 0.026/16) = N(2.775, 0.04^2)$$

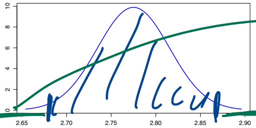
- Is it plausible that 2.65 comes from this distribution?



- No
  - We *have evidence* that the new method does not perform to standards.
  - We **reject** $H_0 : \mu = 2.775$

Microelectronic Solder Joints – decision rule

- Sampling distribution of $\overline{X}$ (if H_0 is true): $N(2.775, 0.026^2/16)$:



reject

don't reject

- Then (if H_0 is true):

$$P(2.709 \leq \overline{X} \leq 2.841) = 0.90 \quad \text{and}$$
$$P(2.696 \leq \overline{X} \leq 2.854) = 0.95 \quad \text{and}$$
$$P(2.671 \leq \overline{X} \leq 2.879) = 0.99$$

e.q to

$$\left| \frac{\bar{X} - 2.775}{\sigma/\sqrt{n}} \right| > z_{\cancel{0.05}\ 0.025}$$

- Suggestion for *decision rule*: Reject H_0 if the observed value of $\overline{X}$ is less than 2.696 or larger than 2.854.
  - The *critical region*: $(\infty, 2.696) \cup (2.854, \infty)$

Microelectronic Solder Joints – summary

- Data model: $X_1, X_2, \ldots, X_n$ is a random sample from $N(\theta, 0.026^2)$
- Null hypothesis $H_0 : \theta = 2.775$
- Alternative hypothesis $H_1 : \theta \neq 2.775$
- Decision rule: Reject H_0 if $\overline{x} < 2.696$ or $\overline{x} > 2.854$

Where we are going:

- How to come up with decision rules – Section 8.2
- How to evaluate hypothesis tests – Section 8.3

Hypothesis testing

Statistical hypotheses

A statistical **hypothesis** is a statement about a population parameter(s).

There are two complimentary hypothesis in a hypothesis problem:

- **Null hypothesis** H_0
- **Alternative hypothesis** H_1

Usually:

$$H_0 : \theta \in \Theta_0 \qquad \text{and} \qquad H_1 : \theta \in \Theta_0^c$$

Examples:

$$H_0 : \theta \leq \theta_0 \qquad \text{and} \qquad H_1 : \theta > \theta_0$$

$$H_0 : \theta = \theta_0 \qquad \text{and} \qquad H_1 : \theta \neq \theta_0$$

Hypothesis test

Hypothesis test

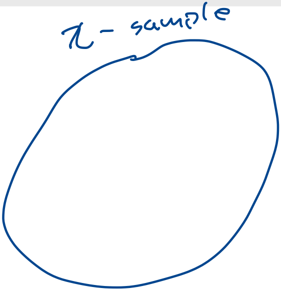
A **hypothesis testing procedure** is a rule that specifies

- For which sample values the decision is made to accept H_0 as true
- For which sample values H_0 is rejected and H_1 is accepted as true

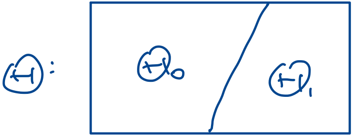
Also called a **decision rule**

Accepting H_0 versus not rejecting H_0

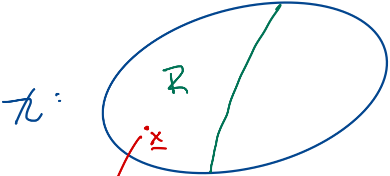
- Intro Stats: We never accept H_0!! – Why?
- Here: decision between H_0 and H_1
  - Accept one, reject the other.



Parameter space



Sample space



Hypotheses live here

would reject H_0 if we observe this $\underline{x}$

Decision rules live here.

R: rejection region

sample space for test statistic

↓

Accepting H_0 versus not rejecting H_0

Recall the argument the hypothesis testing procedure is built upon

- *Assuming that H_0 is true*, we find the sampling distribution of a test statistic.
- If the observed test statistic looks like an observation from the sampling distribution (under the null): i.e. fail to reject H_0

  $$\longrightarrow \text{We Accept } H_0 \text{ (and Reject } H_1\text{)}$$

  We did not find a contradiction/evidence
- If the observed test statistic does **not** look like an observation from the sampling distribution:

  $$\longrightarrow \text{We Reject } H_0 \text{ (and Accept } H_1\text{)}$$

  We found a contradiction/evidence

Accepting H_0 versus not rejecting H_0

- The meaning of "Accept" is subtlety different depending on whether we are accepting H_0 or H_1
  - Accepting H_0 is an inconclusive statement.
  - Accepting H_1 means we have evidence that H_0 is not true
- Therefore we never accept H_0 in introductory statistics courses, only "fail to reject" H_0
- This course: I will assume that you understand this difference and you can use either "Accept" or "fail to reject" H_0

Hypothesis test

- **Rejection region:** The subset of the sample space $\mathcal{X}$ for which H_0 is rejected
  - Also called **critical region**
  - Example: Reject H_0 if $\mathbf{x} \in R$ where

$$R = \{\mathbf{x} = (x_1, \dots, x_n) : \overline{x} < 2.671 \text{ or } \overline{x} > 2.879\} \subset \mathcal{X}$$

- **Acceptance region:** The subset of the sample space $\mathcal{X}$ for which H_0 is accepted
  - Acceptance region is the complement of the rejection region
  - Example: Accept H_0 if $\mathbf{x} \in R^c$

- A decision rule is usually specified in terms of a **test statistic** $W(\mathbf{X}) = T$

Example of a hypothesis test: z-test

- $X_1, \ldots, X_n$ random sample from $N(\mu, \sigma^2)$, σ^2 known
- *Hypotheses:* $H_0 : \mu = 2.7$ vs. $H_1 : \mu \neq 2.7$
- *Decision rule:* Reject H_0 if

$$|z| = \left| \frac{\overline{x} - 2.7}{\sigma/\sqrt{n}} \right| > z_{0.025}$$

where $\overline{x}$ is the observed sample mean

- Critical region:

or $\left\{ t \in \mathbb{R} : \left| \frac{t - 2.7}{\sigma/\sqrt{n}} \right| > z_{0.025} \right\}$

$$\left\{ \mathbf{x} \in \mathbb{R}^n : \left| \frac{\overline{x} - 2.7}{\sigma/\sqrt{n}} \right| > z_{0.025} \right\}$$

Methods of finding tests

- Likelihood ratio tests – Section 8.2.1
- Union-Intersection (and Intersection-union) tests – Section 8.2.3
- Bayesian tests – Section 8.2.2

Likelihood ratio tests

- Likelihood function for a random sample $X_1, \ldots, X_n$:

$$L(\theta \mid \mathbf{x}) = f(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

Def: Likelihood ratio tests

The **likelihood ratio test statistic** for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta \mid \mathbf{x})}{\sup_{\Theta} L(\theta \mid \mathbf{x})}$$

A **likelihood ratio test (LRT)** is any test that has a rejection region of the form

$$\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$$ $\Rightarrow$ Reject if $\lambda(x) \leq c$ for some c

where $0 \leq c \leq 1$

The likelihood ratio

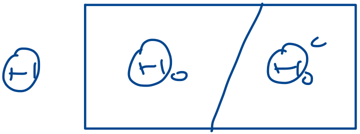
$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta \mid \mathbf{x})}{\sup_{\Theta} L(\theta \mid \mathbf{x})} \qquad \text{Reject } H_0 \text{ if } \lambda(\mathbf{x}) \le c$$

- $0 \le \lambda(\mathbf{x}) \le 1$

$$\rightarrow \; = \frac{\sup_{\Theta_0} L(\theta \mid \underset{\sim}{x})}{L(\hat{\theta} \mid \underset{\sim}{x})}$$

- $\sup_{\Theta} L(\theta \mid \mathbf{x}) = L(\hat{\theta} \mid \mathbf{x})$ where $\hat{\theta}$ is the MLE of θ
- $\sup_{\Theta_0} L(\theta \mid \mathbf{x}) = L(\hat{\theta}_0 \mid \mathbf{x})$ where $\hat{\theta}_0$ maximizes $L(\theta \mid \mathbf{x})$ over Θ_0 only

  → Optimization with constraints
- sup is usually the same as max.
  - **Supremum**: Let S be a set of real numbers. An upper bound for S is a number B such that $x \le B$ for all $x \in S$. The supremum of S is the smallest upper bound for S. A supremum which actually *belongs to the set S* is called a maximum.
  - Example: The open set $(0, 1)$ does not have a maximum, but the supremum is 1



Note $\sup_{\Theta_0} L(\theta \mid \underline{x}) \le \sup_{\Theta} L(\theta \mid \underline{x})$

since $\Theta_0 \subset \Theta$

The likelihood ratio test

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta \mid \mathbf{x})}{\sup_{\Theta} L(\theta \mid \mathbf{x})} \qquad \text{Reject } H_0 \text{ if } \lambda(\mathbf{x}) \leq c$$

- Reject H_0 if $\sup_{\Theta_0} L(\theta \mid \mathbf{x}) \leq c \sup_{\Theta} L(\theta \mid \mathbf{x})$ $= c\, L(\hat{\theta} \mid \underset{\sim}{x})$
- Intuition: If $L(\theta_1 \mid \mathbf{x}) < L(\theta_2 \mid \mathbf{x})$: the observed data $\mathbf{x}$ is more likely when the parameter value is equal to θ_2 than when it is θ_1
- If $\lambda(\mathbf{x})$ is small:
  - There is a parameter value in Θ_0^c for which the observed data are much more likely than for any parameter in Θ_0
  - H_0 should be rejected
- How to select c? - later...

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta \mid \mathbf{x})}{\sup_{\Theta} L(\theta \mid \mathbf{x})} \qquad \text{Reject } H_0 \text{ if } \lambda(\mathbf{x}) \leq c$$

$$= \frac{\sup_{\Theta_0} L(\theta \mid \underline{x})}{L(\hat{\theta} \mid \underline{x})}$$

Know that $\sup_{\Theta_0} L(\theta \mid \underline{x}) \leq L(\hat{\theta} \mid \underline{x})$

If $\lambda(\underline{x})$ is close to 1 ($\sup_{\Theta_0} L(\theta \mid \underline{x})$ is close to $L(\hat{\theta} \mid \underline{x})$)

we accept that $\theta \in \Theta_0$

Let's say $\sup_{\Theta_0} L(\theta \mid \underline{x}) = L(\theta^* \mid \underline{x})$ $\quad \theta^* \in \Theta_0$

$\lambda(\underline{x})$ close to 1 : θ^* is almost as plausible value of θ as $\hat{\theta}$ is, so should not be ruled out

$\lambda(\underline{x})$ small : θ^* is much less plausible value of θ than $\hat{\theta}$ so should be rejected

LRT example: Normal model, variance known

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{N}(\theta, 1)$
- Find the form of the likelihood ratio test (LRT) for the following hypotheses
  - (a) $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
  - (b) $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$
  - (c) $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$

A note on notation

- We stated the "one-sided" hypotheses as

$$H_0 : \theta \leq \theta_0 \qquad \text{vs.} \qquad H_1 : \theta > \theta_0$$

- In introductory statistics courses (and often in practice) we state them as

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta > \theta_0$$

- This makes it easier to talk about the "sampling distribution under the null" since $\Theta_0 = \{\theta_0\}$ then only contains one value.
  - "Assuming that the null hypothesis is true" simply means $\theta = \theta_0$
- In actuality we are using the fact that the likelihood function is maximized over $\Theta_0 = (-\infty, \theta_0]$ at the value θ_0

LRT example: Shifted exponential

- Let $X_1, X_2, \ldots, X_n$ be a random sample from

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

- Find the form of the LRT for $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$

LRT and sufficiency

Theorem 8.2.4

Let $T(\mathbf{X})$ be a sufficient statistic for θ and let $\lambda^*(t)$ and $\lambda(\mathbf{x})$ be the LRT statistics based on T and $\mathbf{X}$, respectively. Then

$$\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x}) \qquad \text{for all } \mathbf{x} \text{ in the sample space}$$

proof...

Can simplify things if we know the pdf of T

- Note that

$$\lambda^*(t) = \frac{\sup_{\Theta_0} L^*(\theta \mid t)}{\sup_{\Theta} L^*(\theta \mid t)} = \frac{\sup_{\Theta_0} f_T(t \mid \theta)}{\sup_{\Theta} f_T(t \mid \theta)}$$

- So a simplified expression of the LRT statistic $\lambda(\mathbf{x})$ should only depend on a sufficient statistic

$L(\theta | \underline{X}) = f(\underline{x} | \theta) = h(x)\, g(t | \theta) \quad t = T(\underline{X})$

if T is a sufficient statistic, by fact. thm.

set $c = \int g(t|\theta)\, dt \Rightarrow f(t|\theta) = \frac{1}{c} g(t|\theta)$

$\Rightarrow L(\theta | X) = h(x)\, c\, f(t|\theta)$

$$\Rightarrow \lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta | \underline{X})}{\sup_{\theta \in \Theta} L(\theta | \underline{X})} = \frac{\sup_{\theta \in \Theta_0} \overbrace{h(x)\, c}^{\text{no } \theta} f(t|\theta)}{\sup_{\theta \in \Theta_0} h(x)\, c\, f(t|\theta)}$$

$$= \frac{\sup_{\theta \in \Theta_0} f(t|\theta)}{\sup_{\theta \in \Theta} f(t|\theta)} = \frac{\sup_{\theta \in \Theta_0} L(\theta | t)}{\sup_{\theta \in \Theta} L(\theta | t)}$$

$\rightarrow$ Can work directly with pdf/likelihood of a suff. stat.

LRT example: Normal model, variance known

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{N}(\theta, 1)$
- Know that $\overline{X}$ is a sufficient statistic
- Know that $\overline{X} \sim \mathrm{N}(\theta, 1/n)$
- LRT statistic for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ is:

Union-Intersection and Intersection-Union methods

- A way of combining test procedures
- Potentially useful method of coming up with a test procedure if the null hypothesis can be expressed as an intersection (or a union), i.e.

$$H_0 : \theta \in \bigcap_{k=1}^{K} \Theta_k \qquad \text{or} \qquad H_0 : \theta \in \bigcup_{k=1}^{K} \Theta_k$$

- Can also handle $K = \infty$
- Often useful if we have more than one parameter

Union-Intersection method

- Suppose that H_0 can be expressed as

$$H_0 : \theta \in \bigcap_{k=1}^{K} \Theta_k$$

$$H_1 : \theta \in \left(\bigcap_{k=1}^{K} \Theta_k\right)^c = \bigcup_{k=1}^{K} \Theta_k^c$$

- Suppose that for each k we have a test procedure for

$$H_{0k} : \theta \in \Theta_k \quad \text{vs} \quad H_{1k} : \theta \notin \Theta_k$$

with a rejection region $\{\mathbf{x} : T(\mathbf{x}) \in R_k\}$

- Then the rejection region for the union-intersection test is

$$\bigcup_{k=1}^{K} \{\mathbf{x} : T(\mathbf{x}) \in R_k\}$$

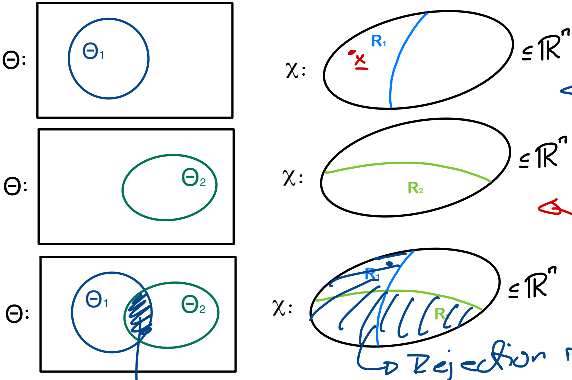
"over all" i.e. intersection

- Note: If one or more H_{0k} is rejected, H_0 must be rejected

Making sense of UIT for $K = 2$

- Θ_1: Null hypothesis for test 1 $\quad$ R_1: Rejection region for test 1
- Θ_2: Null hypothesis for test 2 $\quad$ R_2: Rejection region for test 2

parameter space $\qquad$ sample space



$\Theta_1 \cap \Theta_2$

Rejection region: $R_1 \cup R_2$

$H_0: \theta \in \Theta_1 \cap \Theta_2$

$H_1: \theta \notin \Theta_1 \cap \Theta_2$

i.e. $\theta \in \Theta_1^c \cup \Theta_2^c$

For testing

$H_{01}: \theta \in \Theta_1$

vs. $H_{11}: \theta \notin \Theta_1$

For testing

$H_{02}: \theta \in \Theta_2$

Intersection-Union method

- Suppose that H_0 can be expressed as

$$H_0 : \theta \in \bigcup_{k=1}^{K} \Theta_k$$

- Suppose that for each k we have a test procedure for

$$H_{0k} : \theta \in \Theta_k \quad \text{vs} \quad H_{1k} : \theta \notin \Theta_k$$

  with a rejection region $\{\mathbf{x} : T(\mathbf{x}) \in R_k\}$

- Then the rejection region for the intersection-union test is

$$\bigcap_{k=1}^{K} \{\mathbf{x} : T(\mathbf{x}) \in R_k\}$$

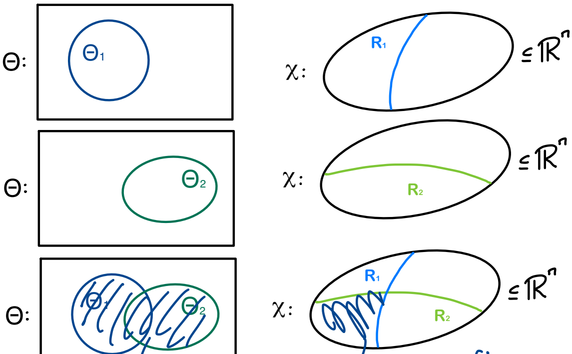
- Note: If all H_{0k} are rejected then H_0 is rejected

Making sense of IUT for $K = 2$

- Θ_1: Null hypothesis for test 1 — R_1: Rejection region for test 1
- Θ_2: Null hypothesis for test 2 — R_2: Rejection region for test 2

$H_0: \theta \in \Theta_1 \cup \Theta_2$

$H_1: \theta \in (\Theta_1 \cup \Theta_2)^c = \Theta_1^c \cap \Theta_2^c$



$\Theta_1 \cup \Theta_2$

Rejection region: $R_1 \cap R_2$

i.e. reject H_0 if both H_{01} and H_{02} are rejected.

Example: Two-sided t-test

- Let $X_1, X_2, \ldots, X_n$ be iid. $\mathrm{N}(\mu, \sigma^2)$, both μ and σ^2 unknown.

  $\Theta_0 = \{\mu_0\}$
- Want to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$

  $(-\infty, \mu_0] \cap [\mu_0, \infty) = \{\mu_0\}$
- We can write H_0 as $H_0 : \underbrace{\{\mu : \mu \leq \mu_0\}}_{\Theta_1} \cap \underbrace{\{\mu : \mu \geq \mu_0\}}_{\Theta_2}$
- The LRT for $H_{01} : \mu \leq \mu_0$ vs. $H_{11} : \mu > \mu_0$ is

$$\text{reject } H_{01} \text{ if} \qquad \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \geq t_L \leftarrow \text{some number}$$

- The LRT for $H_{02} : \mu \geq \mu_0$ vs. $H_{12} : \mu < \mu_0$ is

$$\text{reject } H_{02} \text{ if} \qquad \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \leq t_U$$

Example: Two-sided t-test – continued

- The Union-intersection test for H_0 is therefore: Reject H_0 if

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \geq t_L \qquad \text{or} \qquad \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \leq t_U$$

- If $t_L = -t_U$ we get: Reject H_0 if

$$\left| \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \right| \geq t_L$$

- This is called the **two-sided t-test** → same as the IUT
- Same as the likelihood ratio test

Example: Two parameters

- Two parameters that are important for assessing quality of upholstery fabric:
  - θ_1 : mean breaking strength
  - θ_2 : probability of passing a flammability test

- Standards: $\theta_1 > 50$ and $\theta_2 > 0.95$

- Suppose we will collect data:
  - $X_i =$ breaking strength of unit i, $i = 1, \ldots, n$
  - $Y_j = 1$ if unit j does not catch fire (0 otherwise), $j = 1, \ldots, m$

- Assume that $X_1, \ldots, X_n$ are i.i.d $N(\theta, \sigma^2)$ and $Y_1, \ldots, Y_m$ are iid $\mathrm{Bernoulli}(\theta_2)$

Example: Two parameters - continued

- Assume that $X_1, \ldots, X_n$ are i.i. $N(\theta, \sigma^2)$ and $Y_1, \ldots, Y_m$ are iid $\mathrm{Bernoulli}(\theta_2)$
- Want to test the hypothesis

$$H_0 : \theta_1 \leq 50 \text{ or } \theta_2 \leq 0.95 \qquad \text{vs.} \qquad H_1 : \theta_1 > 50 \text{ and } \theta_2 > 0.95$$

- Determine the rejection region for this test

Bayesian tests

- Want to test $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \notin \Theta_0$
- All Bayesian inference is based on the posterior distribution

$$p(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta)\, p(\theta)}{\int f(\mathbf{x} \mid \theta)\, p(\theta) d\theta}$$

- Since θ is treated as a random variable, we can get the probability (prior or posterior) that the null hypothesis is true:

$$\overset{\text{prior}}{P(\theta \in \Theta_0)} \qquad \text{or} \qquad \overset{\text{posterior}}{P(\theta \in \Theta_0 \mid \mathbf{X})}$$

- One way to do Bayesian testing: Pick the hypothesis with higher posterior probability:

$$\text{accept } H_0 \text{ if} \quad P(\theta \in \Theta_0 \mid \mathbf{X}) \geq P(\theta \in \Theta_0^c \mid \mathbf{X})$$

Bayesian tests

Since $P(\theta \in \Theta_0 \mid \underline{X}) + P(\theta \in \Theta_0^c \mid \underline{X}) = 1$

- Bayesian test for $H_0 : \theta \in \Theta$ versus $H_1 : \theta \notin \Theta_0$:

$$\text{reject } H_0 \text{ if} \quad P(\theta \in \Theta_0^c \mid \mathbf{X}) > \frac{1}{2}$$

(or equiv: if $P(\theta \in \Theta_0 \mid X) \leq \frac{1}{2}$)

- $P(\theta \in \Theta_0^c \mid \mathbf{X})$ is the test statistic

- Rejection region is

$$\{\mathbf{x} : P(\theta \in \Theta_0^c \mid \mathbf{X} = \mathbf{x}) > 0.5\}$$

- Or: If we want to guard against falsely rejecting H_0 we could pick a larger percentage, e.g.

$$\text{reject } H_0 \text{ if} \quad P(\theta \in \Theta_0^c \mid \mathbf{X}) > 0.9$$

Example: Normal-normal model

- Let $X_1, X_2, \ldots, X_n$ be iid. $\mathrm{N}(\theta, \sigma^2)$, where σ^2 is known.
- Suppose the prior on θ is $\mathrm{N}(\mu, \tau^2)$
- Want to test $\theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$

  $\Theta_0 = (-\infty, \theta_0]$ $\quad \Theta_0^c = (\theta_0, \infty)$
- We know that the posterior distribution of $\theta \mid \mathbf{X}$ is $\mathrm{N}(\tilde{\mu}, \tilde{\sigma}^2)$ where

$$\tilde{\mu} = \frac{n\tau^2}{n\tau^2 + \sigma^2}\overline{x} + \frac{\sigma^2}{n\tau^2 + \sigma^2}\mu \quad \text{and} \quad \tilde{\sigma}^2 = \frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2}$$

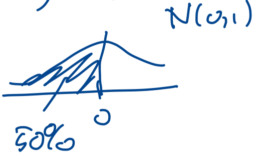
- Find the Bayesian decision rule : Reject if $P(\theta > \theta_0 \mid \underline{X}) > \frac{1}{2}$

$$\Rightarrow 1 - P(\theta \leq \theta_0) = 1 - P\left(\frac{\theta - \tilde{\mu}}{\tilde{\sigma}} \leq \frac{\theta_0 - \tilde{\mu}}{\tilde{\sigma}}\right)$$

$$= 1 - \Phi\left(\frac{\theta_0 - \tilde{\mu}}{\tilde{\sigma}}\right) > \frac{1}{2} \Rightarrow \Phi\left(\frac{\theta_0 - \tilde{\mu}}{\tilde{\sigma}}\right) < \frac{1}{2}$$

$$\Rightarrow 1 - P(\theta \le \theta_0) = 1 - P\left(\frac{\theta - \tilde{\mu}}{\hat{\tau}} \le \frac{\theta_0 - \tilde{\mu}}{\hat{\tau}} \right)$$

$$= 1 - \Phi\left(\frac{\theta_0 - \tilde{\mu}}{\hat{\tau}} \right) > \frac{1}{2} \, {}^{*} \Rightarrow \Phi\left(\frac{\theta_0 - \tilde{\mu}}{\hat{\tau}} \right) < \frac{1}{2}$$



$$\Rightarrow \frac{\theta_0 - \tilde{\mu}}{\hat{\tau}} < \Phi^{-1}(0.5) = 0$$

$$\Rightarrow \theta_0 - \tilde{\mu} < 0 \Rightarrow \theta_0 < \tilde{\mu}$$

$$\Rightarrow \frac{n\tau^2}{n\tau^2 + \sigma^2} \bar{x} + \frac{\sigma^2}{n\tau^2 + \sigma^2} \mu > \theta_0$$

no undetermined constant c, properties of the test determined by the choice $\frac{1}{2}$ *

$$\Rightarrow n\tau^2 \bar{x} > \theta_0 (n\tau^2 + \sigma^2) - \sigma^2 \mu$$

$$\Rightarrow \bar{x} > \theta_0 + \frac{\sigma^2(\theta_0 - \mu)}{n\tau^2}$$

Recall LRT: Reject if $\bar{x} > \theta_0 + \sqrt{\frac{2}{n} \log c}$

$\Rightarrow$ Get the same form