STAT 24400 Lecture 17
Section 9.1-9.4 Testing Hypotheses

Yibi Huang
Department of Statistics
University of Chicago

Introduction to Hypothesis Testing

Can Dogs Smell Cancer?

Dogs Can Smell Cancer | Secret Life of Dogs | BBC

- https://youtu.be/e0UK6kkS0_M

Case Study: Can Dogs Smell Bladder Cancer?

- A study[1] by M. Willis et al. considered whether dogs could be trained to detect if a person has bladder cancer by smelling his/her urine.
- 6 dogs of varying breeds were trained to discriminate between urine from patients with bladder cancer and urine from control patients without it.
- The dogs were taught to indicate which among several specimens was from the bladder cancer patient by lying beside it.
- Once trained, the dogs' ability to distinguish cancer patients from controls was tested using urine samples from subjects not previously encountered by the dogs.

[1] Olfactory detection of human bladder cancer by dogs: proof of principle study, *British Medical Journal*, vol. 329, September 25, 2004.

Case Study: Can Dogs Smell Bladder Cancer?

- The researchers blinded both dog handlers and experimental observers to the identity of urine samples.
- Each of the 6 dogs was tested with 9 trials. In each trial, one urine sample from a bladder cancer patient was randomly placed among 6 control urine samples.
- Outcome: In the total of 54 trials with the 6 dogs, the dogs made the correct selection 22 times.
  - The dogs were correct for $22/54 \approx 41\%$ of the time,
    - not fabulous
  - If the dogs just guessed at random, they were only expected to be correct for $1/7 \approx 14\%$ of the time
  - Is this difference (41% v.s. 14%) surprising?

Two Competing Hypotheses

Let p be the probability that a dog makes the correct selection on a given trial.

- *Null hypothesis (H_0)*: $p = 1/7$
  "There is nothing going on."
  The dogs just guessed at random.
  - "null" means "nothing surprising is going on".
  - The dogs were just lucky to make more correct selections than expected.

Two Competing Hypotheses

Let p be the probability that a dog makes the correct selection on a given trial.

- *Null hypothesis (H_0)*: $p = 1/7$
  "There is nothing going on."
  The dogs just guessed at random.
  - "null" means "nothing surprising is going on".
  - The dogs were just lucky to make more correct selections than expected.
- *Alternative hypothesis (H_A or H_1)*: $p > 1/7$
  "There is something going on."
  Dogs can do better than random guessing.

Weighing Evidence Using a Test Statistic

The next step of hypothesis testing is to weigh the evidence — how likely the observed data could have occur if H_0 was true?

- If the observed result was very unlikely to have occurred under the H_0, then the evidence raises more than a reasonable doubt in our minds about the H_0.

Weighing Evidence Using a Test Statistic

The next step of hypothesis testing is to weigh the evidence
— how likely the observed data could have occur if H_0 was true?

- If the observed result was very unlikely to have occurred under the H_0, then the evidence raises more than a reasonable doubt in our minds about the H_0.

The *test statistic* is a summary of the data that best reflects the evidence for or against the hypotheses.

- For this study, the test statistics we choose is

$$X = \text{the total number of correct selections in the 54 trials}$$

- A larger X value is a stronger evidence for H_1 and against H_0

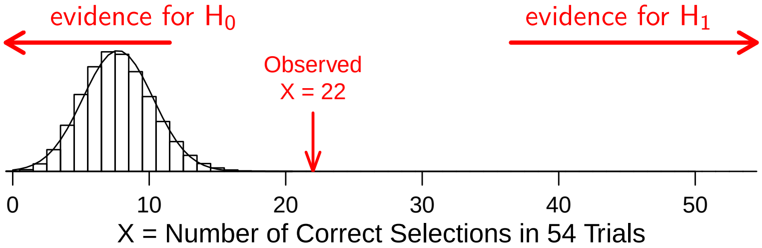
Distribution of the Test Statistic Under H_0

For the "Dogs Smell Cancer" study, if H_0 is true, then

$$X \sim Bin(n = 54,\ p = 1/7) \quad \text{(Why?)}$$

which implies

$$P(X = k) = \binom{54}{k} \left(\frac{1}{7}\right)^k \left(\frac{6}{7}\right)^{54-k}, \quad k = 0, 1, 2, \ldots, 54.$$



Test Procedure & Rejection Region

A test procedure is specified by the following:

1. a *test statistic*
2. a *rejection region*

The null hypothesis H_0 will be **rejected** if and only if **the test statistic falls in the rejection region**.

Test Procedure & Rejection Region

A test procedure is specified by the following:

1. a *test statistic*
2. a *rejection region*

The null hypothesis H_0 will be **rejected** if and only if **the test statistic falls in the rejection region**.

E.g., for the "Dogs Smell Cancer" study, as the strength of evidence for the two hypotheses are reflected by the test statistic

$$X = \# \text{ of correct guesses in the 54 trials.}$$

A sensible **rejection region** is of the form

$$X \geq k \quad \text{for some cutoff } k.$$

and the test procedure is $\boxed{\text{reject } H_0 \text{ if } X \geq k}$.

Test Procedure & Rejection Region

A test procedure is specified by the following:

1. a *test statistic*
2. a *rejection region*

The null hypothesis H_0 will be **rejected** if and only if **the test statistic falls in the rejection region**.

E.g., for the "Dogs Smell Cancer" study, as the strength of evidence for the two hypotheses are reflected by the test statistic

$$X = \# \text{ of correct guesses in the 54 trials.}$$

A sensible **rejection region** is of the form

$$X \geq k \quad \text{for some cutoff } k.$$

and the test procedure is $\boxed{\text{reject } H_0 \text{ if } X \geq k}$.

How to choose the cutoff value k for the rejection region?

Type I and Type II Errors

In a hypothesis test, we make a decision about which of H_0 or H_1 might be true, but our decision might be incorrect.

Type I and Type II Errors

In a hypothesis test, we make a decision about which of H_0 or H_1 might be true, but our decision might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_1 true		

Type I and Type II Errors

In a hypothesis test, we make a decision about which of H_0 or H_1 might be true, but our decision might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_1 true		

Type I and Type II Errors

In a hypothesis test, we make a decision about which of H_0 or H_1 might be true, but our decision might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_1 true		✓

Type I and Type II Errors

In a hypothesis test, we make a decision about which of H_0 or H_1 might be true, but our decision might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type I Error*
	H_1 true		✓

- A *Type I Error* is rejecting the H_0 when it is true.

Type I and Type II Errors

In a hypothesis test, we make a decision about which of H_0 or H_1 might be true, but our decision might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type I Error*
	H_1 true	*Type II Error*	✓

- A *Type I Error* is rejecting the H_0 when it is true.
- A *Type II Error* is failing to reject the H_0 when it is false.

Significance Level $\alpha = P(\text{Type I error})$

The *significance level* α of a test procedure is its probability to reject the null hypothesis H_0 when H_0 is true.

$$\alpha = P(\text{Type I error}) = P(\text{reject } \mathrm{H}_0 \mid H_0 \text{ is true})$$

Significance Level $\alpha = P(\text{Type I error})$

The *significance level* α of a test procedure is its probability to reject the null hypothesis H_0 when H_0 is true.

$$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

For the "Dog Smell Cancer" Study, if the test procedure is $\boxed{\text{rejecting } H_0 \text{ if } X \geq 15}$, the significance level would be

$$\begin{aligned}\alpha = P(\text{Type I error}) &= P(H_0 \text{ is rejected when } H_0\ (p = 1/7) \text{ is true}) \\ &= P(X \geq 15 \text{ when } X \sim Bin(n = 54, p = 1/7)) \\ &= \sum_{k=15}^{54} \binom{54}{k} \left(\frac{1}{7}\right)^k \left(\frac{6}{7}\right)^{54-k} \approx 0.0073\end{aligned}$$

Significance Level $\alpha = P(\text{Type I error})$

The *significance level* α of a test procedure is its probability to reject the null hypothesis H_0 when H_0 is true.

$$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

For the "Dog Smell Cancer" Study, if the test procedure is $\boxed{\text{rejecting } H_0 \text{ if } X \geq 15}$, the significance level would be

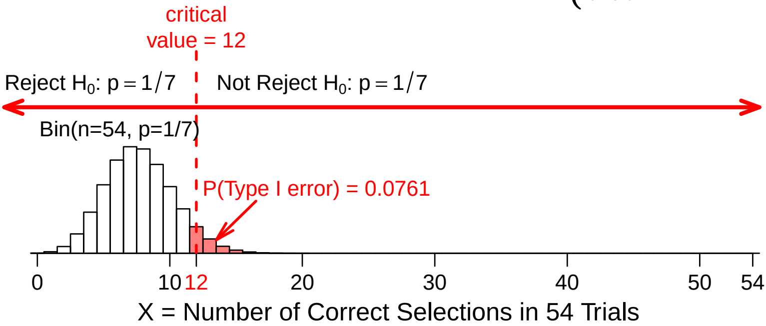
$$\begin{aligned}\alpha = P(\text{Type I error}) &= P(H_0 \text{ is rejected when } H_0\ (p = 1/7) \text{ is true}) \\ &= P(X \geq 15 \text{ when } X \sim Bin(n = 54, p = 1/7)) \\ &= \sum_{k=15}^{54} \binom{54}{k} \left(\frac{1}{7}\right)^k \left(\frac{6}{7}\right)^{54-k} \approx 0.0073\end{aligned}$$

If we reject H_0 when $X \geq 15$, there is a chance of 0.0073 to falsely reject a correct H_0 (Type I error).

P(Type I Error) — Dogs-Smell-Cancer Study

For the test procedure: $\boxed{\text{rejecting } H_0 \text{ when } X \geq k}$,

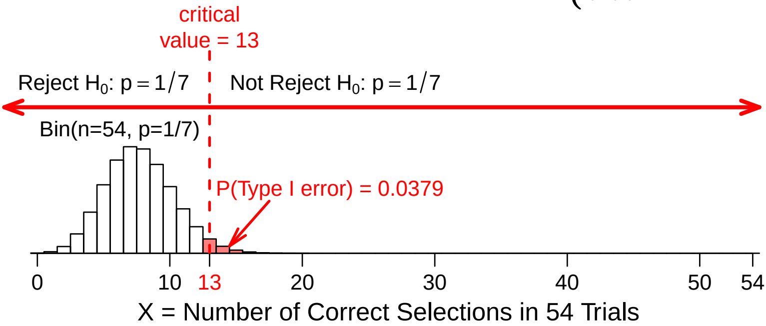
$$P(\text{Type I error}) = P(H_0 \text{ is rejected when } H_0\ (p = 1/7) \text{ is true})$$
$$= P(X \geq k \text{ when } X \sim Bin(n = 54, p = 1/7))$$
$$= \sum_{x=k}^{54} \binom{54}{k} \left(\frac{1}{7}\right)^x \left(\frac{6}{7}\right)^{54-x} \approx \begin{cases} 0.076 & \text{if } k = 12 \\ 0.038 & \text{if } k = 13 \\ 0.017 & \text{if } k = 14 \\ 0.007 & \text{if } k = 15 \end{cases}$$



P(Type I Error) — Dogs-Smell-Cancer Study

For the test procedure: $\boxed{\text{rejecting } H_0 \text{ when } X \geq k}$,

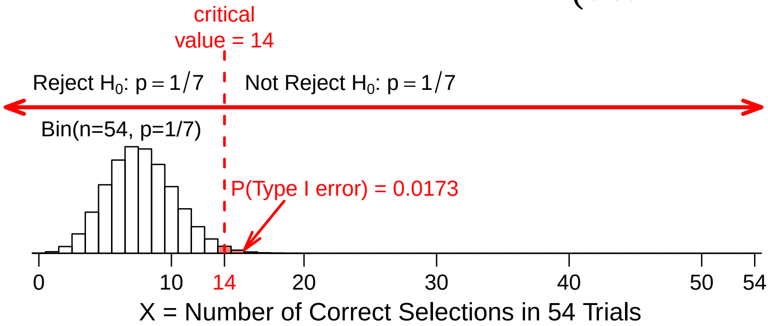
$$
\begin{aligned}
P(\text{Type I error}) &= P(H_0 \text{ is rejected when } H_0\ (p = 1/7) \text{ is true}) \\
&= P(X \geq k \text{ when } X \sim Bin(n = 54, p = 1/7)) \\
&= \sum_{x=k}^{54} \binom{54}{k} \left(\frac{1}{7}\right)^x \left(\frac{6}{7}\right)^{54-x} \approx \begin{cases} 0.076 & \text{if } k = 12 \\ 0.038 & \text{if } k = 13 \\ 0.017 & \text{if } k = 14 \\ 0.007 & \text{if } k = 15 \end{cases}
\end{aligned}
$$



P(Type I Error) — Dogs-Smell-Cancer Study

For the test procedure: rejecting H_0 when $X \geq k$,

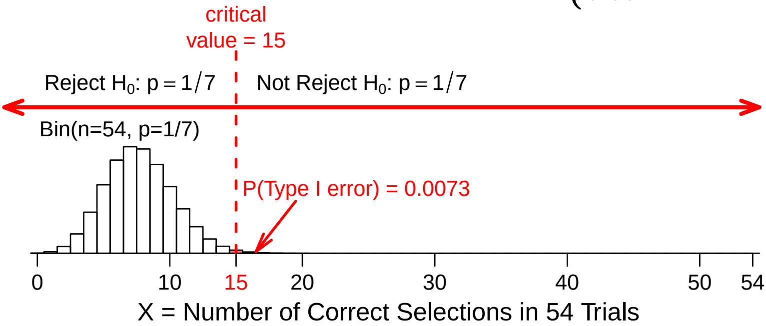
$$P(\text{Type I error}) = P(H_0 \text{ is rejected when } H_0\ (p = 1/7) \text{ is true})$$
$$= P(X \geq k \text{ when } X \sim Bin(n = 54, p = 1/7))$$
$$= \sum_{x=k}^{54} \binom{54}{k} \left(\frac{1}{7}\right)^x \left(\frac{6}{7}\right)^{54-x} \approx \begin{cases} 0.076 & \text{if } k = 12 \\ 0.038 & \text{if } k = 13 \\ 0.017 & \text{if } k = 14 \\ 0.007 & \text{if } k = 15 \end{cases}$$



P(Type I Error) — Dogs-Smell-Cancer Study

For the test procedure: $\boxed{\text{rejecting } H_0 \text{ when } X \geq k}$,

$$\begin{aligned} P(\text{Type I error}) &= P(H_0 \text{ is rejected when } H_0\ (p = 1/7) \text{ is true}) \\ &= P(X \geq k \text{ when } X \sim Bin(n = 54, p = 1/7)) \\ &= \sum_{x=k}^{54} \binom{54}{k} \left(\frac{1}{7}\right)^x \left(\frac{6}{7}\right)^{54-x} \approx \begin{cases} 0.076 & \text{if } k = 12 \\ 0.038 & \text{if } k = 13 \\ 0.017 & \text{if } k = 14 \\ 0.007 & \text{if } k = 15 \end{cases} \end{aligned}$$



Setting Rejection Region Based on the Significance Level

For the dogs study,

$$P(\text{Type I error}) = \begin{cases} 0.076 & \text{if rejecting H}_0 \text{ when } X \geq 12 \\ 0.038 & \text{if rejecting H}_0 \text{ when } X \geq 13 \\ 0.017 & \text{if rejecting H}_0 \text{ when } X \geq 14 \\ 0.007 & \text{if rejecting H}_0 \text{ when } X \geq 15 \end{cases}$$

Setting Rejection Region Based on the Significance Level

For the dogs study,

$$P(\text{Type I error}) = \begin{cases} 0.076 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 12 \\ 0.038 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 13 \\ 0.017 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 14 \\ 0.007 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 15 \end{cases}$$

To determine the cutoff value k for the rejection region $\{X \geq k\}$, we can first choose a significance level α , which is the maximal P(Type I error) we can tolerate, and then choose the cutoff value so that P(Type I error) does not exceeds the significance level α.

Setting Rejection Region Based on the Significance Level

For the dogs study,

$$P(\text{Type I error}) = \begin{cases} 0.076 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 12 \\ 0.038 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 13 \\ 0.017 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 14 \\ 0.007 & \text{if rejecting } \mathrm{H}_0 \text{ when } X \geq 15 \end{cases}$$

To determine the cutoff value k for the rejection region $\{X \geq k\}$, we can first choose a significance level α , which is the maximal P(Type I error) we can tolerate, and then choose the cutoff value so that P(Type I error) does not exceeds the significance level α.

- If we can tolerate a $\alpha = 5\%$ chance of Type I error, the test procedure can be "rejecting H_0 if $X \geq 13$"
- If we can tolerate a $\alpha = 1\%$ chance of Type I error, the test procedure can be "rejecting H_0 if $X \geq 15$"

Reducing Significance Level Would Increase P(Type II Error)

One might want to avoid a Type I error as much as possible by setting a tiny significance level. However,

smaller significance level $\Rightarrow$ smaller P(Type I error)

Reducing Significance Level Would Increase P(Type II Error)

One might want to avoid a Type I error as much as possible by setting a tiny significance level. However,

$$\begin{aligned} \text{smaller significance level} &\Rightarrow \text{smaller P(Type I error)} \\ &\Rightarrow \text{less likely to reject } H_0 \end{aligned}$$

Reducing Significance Level Would Increase P(Type II Error)

One might want to avoid a Type I error as much as possible by setting a tiny significance level. However,

$$\begin{aligned} \text{smaller significance level} &\Rightarrow \text{smaller P(Type I error)} \\ &\Rightarrow \text{less likely to reject } H_0 \\ &\Rightarrow \text{more likely to make Type II error} \end{aligned}$$

Reducing Significance Level Would Increase P(Type II Error)

One might want to avoid a Type I error as much as possible by setting a tiny significance level. However,

$$\begin{aligned} \text{smaller significance level} &\Rightarrow \text{smaller P(Type I error)} \\ &\Rightarrow \text{less likely to reject } H_0 \\ &\Rightarrow \text{more likely to make Type II error} \\ &\Rightarrow \text{higher P(Type II error)} \end{aligned}$$

Reducing Significance Level Would Increase P(Type II Error)

One might want to avoid a Type I error as much as possible by setting a tiny significance level. However,

$$\begin{aligned}
\text{smaller significance level} &\Rightarrow \text{ smaller P(Type I error)} \\
&\Rightarrow \text{ less likely to reject } H_0 \\
&\Rightarrow \text{ more likely to make Type II error} \\
&\Rightarrow \text{ higher P(Type II error)}
\end{aligned}$$

Suppose the sample size is fixed and a test statistic is chosen, choosing a rejection region with a smaller P(Type I error) would lead to a larger P(Type II error).

P(Type II Error) & Power — Dogs-Smell-Cancer Study

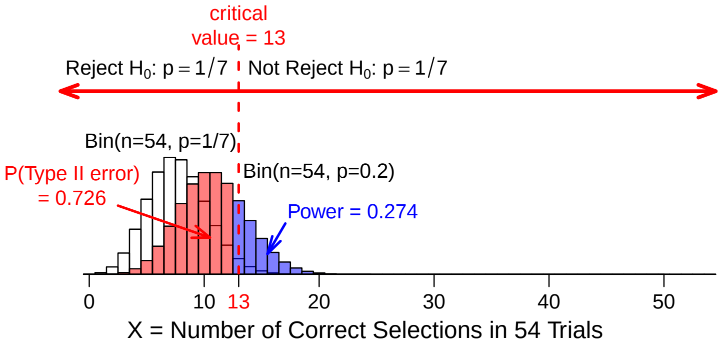
Using the rejection region $X \geq 13$, then

$$\mathrm{P}(\text{Type II error}) = \mathrm{P}(\text{not Reject } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ is FALSE})$$

$$= \mathrm{P}(X < 13 \mid p \neq 1/7) = \sum_{x=0}^{12} \binom{54}{x} p^x (1-p)^{54-x}$$

$$\text{Power} = 1 - \mathrm{P}(\text{Type II error}).$$

Both P(Type II Error) and power are functions of p.



P(Type II Error) & Power — Dogs-Smell-Cancer Study

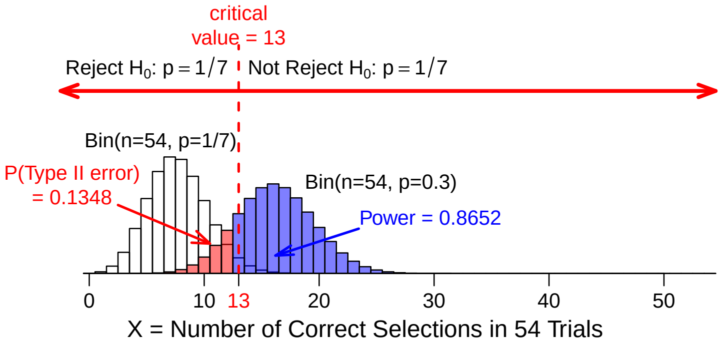
Using the rejection region $X \geq 13$, then

$$\mathrm{P}(\text{Type II error}) = \mathrm{P}(\text{not Reject } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ is FALSE})$$

$$= \mathrm{P}(X < 13 \mid p \neq 1/7) = \sum_{x=0}^{12} \binom{54}{x} p^x (1-p)^{54-x}$$

$$\text{Power} = 1 - \mathrm{P}(\text{Type II error}).$$

Both P(Type II Error) and power are functions of p.



P(Type II Error) & Power — Dogs-Smell-Cancer Study

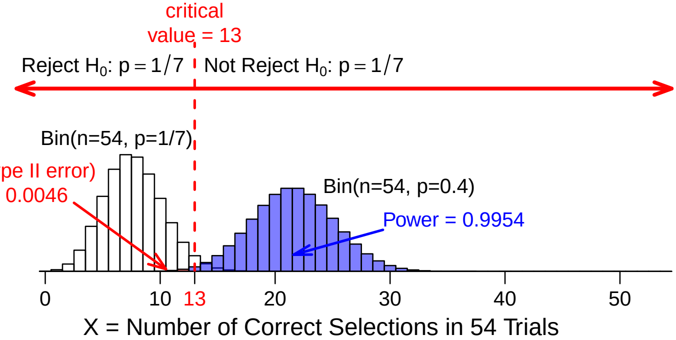
Using the rejection region $X \geq 13$, then

$$\mathrm{P}(\text{Type II error}) = \mathrm{P}(\text{not Reject } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ is FALSE})$$

$$= \mathrm{P}(X < 13 \mid p \neq 1/7) = \sum_{x=0}^{12} \binom{54}{x} p^x (1-p)^{54-x}$$

$$\text{Power} = 1 - \mathrm{P}(\text{Type II error}).$$

Both P(Type II Error) and power are functions of p.



Failing to Reject $H_0 \neq$ Accepting H_0

In the conclusion of a hypothesis test,

- we only say "we reject the H_0" or "we fail to reject the H_0"
- we do NOT say "we accept the H_1" or "we accept the H_0"

- When we fail to reject the H_0, we might have made a Type II error

Failing to Reject $H_0 \neq$ Accepting H_0

In the conclusion of a hypothesis test,

- we only say "we reject the H_0" or "we fail to reject the H_0"
- we do NOT say "we accept the H_1" or "we accept the H_0"

- When we fail to reject the H_0, we might have made a Type II error
- P(Type II error) can be large as it's not controlled.

Failing to Reject $H_0 \neq$ Accepting H_0

In the conclusion of a hypothesis test,

- we only say “we reject the H_0” or “we fail to reject the H_0”
- we do NOT say “we accept the H_1” or “we accept the H_0”

- When we fail to reject the H_0, we might have made a Type II error
- P(Type II error) can be large as it’s not controlled.
- Recall so far we’ve only controlled P(Type I error) by the significance level but haven’t taken any measure to control P(Type II error)

Conclusion of the Dogs Smell Bladder Cancer Study

- There is strong evidence that dogs have some ability to smell bladder cancer,
- However, the dogs were only correct 40% of the time, too low for practical application
- Another study (M. McCulloch et al., Integrative Cancer Therapies, vol 5, p. 30, 2006.) considered whether dogs could be trained to detect whether a person has lung cancer by smelling the subjects' breath. In one test with 83 Stage I lung cancer samples, the dogs correctly identified the cancer sample 81 times.

Summary: Hypothesis Testing

1. We start with a *null hypothesis (H_0)* that represents the status quo.
2. We also have an *alternative hypothesis (H_1)* that represents our research question, i.e. what we're testing for.
3. We then collect data and often summarize the data as a *test statistic*, which is usually a measure gauging whether H_0 or H_A are more plausible
4. We then determine the *sampleing distribution* of the *test statistic* assuming H_0 is true.
   - If the *test statistic* is too far away from what the H_0 predicts, we then reject the H_0 in favor of the H_1.
5. We choose a *significance level* $\alpha =$ maximal P(Type I error) that we can tolerate
6. we set the rejection region based on the significance level
7. we reject H_0 if the test statistic falls in the rejection region, and do not reject otherwise

Likelihood Ratio Tests

Simple & Composite Hypotheses

For $X \sim f(x \mid \theta)$, a hypothesis is called a *simple hypothesis* if it completely specifies the distribution $f(x \mid \theta)$ of X; otherwise it is called a *composite hypothesis*.

Simple & Composite Hypotheses

For $X \sim f(x \mid \theta)$, a hypothesis is called a *simple hypothesis* if it completely specifies the distribution $f(x \mid \theta)$ of X; otherwise it is called a *composite hypothesis*.

Ex. for $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x \mid \theta)$

- H_0: $\theta = 1$ v.s. H_1: $\theta = 2$ $\Rightarrow$ H_0 & H_1 are both simple
- H_0: $\theta = 1$ v.s. H_1: $\theta \neq 1$ $\Rightarrow$ H_0 is simple; H_1 is composite
- H_0: $\theta \leq 1$ v.s. H_1: $\theta \geq 1$ $\Rightarrow$ H_0 & H_1 are both composite

Simple & Composite Hypotheses

For $X \sim f(x \mid \theta)$, a hypothesis is called a *simple hypothesis* if it completely specifies the distribution $f(x \mid \theta)$ of X; otherwise it is called a *composite hypothesis*.

Ex. for $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x \mid \theta)$

- H_0: $\theta = 1$ v.s. H_1: $\theta = 2$ $\Rightarrow$ H_0 & H_1 are both simple
- H_0: $\theta = 1$ v.s. H_1: $\theta \neq 1$ $\Rightarrow$ H_0 is simple; H_1 is composite
- H_0: $\theta \leq 1$ v.s. H_1: $\theta \geq 1$ $\Rightarrow$ H_0 & H_1 are both composite

In some cases, hypotheses might not be about parameters.
e.g., observing i.i.d. pairs (X_i, Y_i) from some joint distribution

- H_0: X & Y are independent
- H_1: X & Y are NOT independent

In this case, both H_0 & H_1 are composite

Likelihood Ratio Tests (LRT)

If H_0: $\theta = \theta_0$ & H_1: $\theta = \theta_1$ are both simple, one can test H_0 v.s. H_1 by comparing their likelihood.

- Higher values of likelihood of $\theta_0 \leftrightarrow \mathsf{H}_0$ seems more plausible
- Higher values of likelihood of $\theta_1 \leftrightarrow \mathsf{H}_1$ seems more plausible

A reasonable test statistic is the ratio of their likelihood

$$\mathsf{LR} = \frac{\text{Likelihood of } \theta_0}{\text{Likelihood of } \theta_1}.$$

We will need to set some threshold c:

- If $\mathsf{LR} < c$ then reject H_0
- If $\mathsf{LR} > c$ then not to reject H_1

(Or use $\leq c$ and $> c$, for discrete cases.)

Example — Normal Likelihood Ratio Tests

Given $X_1, \ldots, X_n$ are i.i.d. $\mathrm{N}(\mu, \sigma^2)$ with known σ^2, recall the likelihood of μ for normal is

$$L(\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2\right)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(\frac{-1}{2\sigma^2} \left[\sum_{i=1}^{n} (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2\right]\right)$$

The LR statistic for testing

$$H_0: \ \mu = \mu_0 \quad \text{v.s.} \quad H_1: \ \mu = \mu_1 \quad \text{where } \mu_0 < \mu_1.$$

is

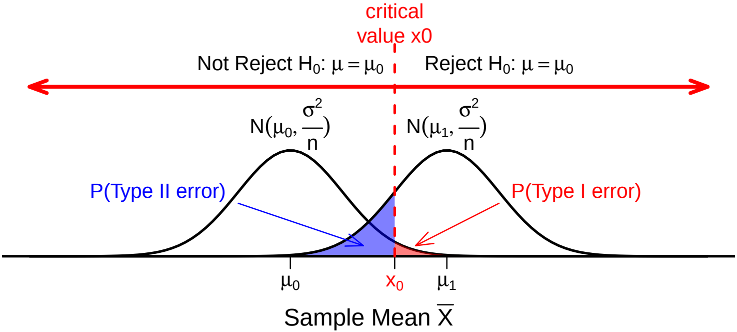
$$\mathrm{LR} = \frac{L(\mu_0)}{L(\mu_1)} = \frac{\exp(\frac{-n}{2\sigma^2}(\overline{X} - \mu_0)^2)}{\exp(\frac{-n}{2\sigma^2}(\overline{X} - \mu_1)^2)} = e^{\frac{n}{2\sigma^2}[(\overline{X}-\mu_1)^2 - (\overline{X}-\mu_0)^2]}$$
$$= e^{\frac{n}{2\sigma^2}(2\overline{X}(\mu_0-\mu_1)+\mu_1^2-\mu_0^2)}$$

As $\mu_0 < \mu_1$, LR $< c$ if and only if $\overline{X} >$ some constant.
Using LR is equivalent to using $\overline{X}$ as the test statistic.
We would reject H_0 if $\overline{X} >$ some critical value x_0.

As $\overline{X} \sim N(\mu_0, \sigma^2/n)$,

$$\mathrm{P}(\text{Type I error}) = \mathrm{P}(\overline{X} > x_0 \mid H_0: \ \mu = \mu_0) = 1 - \Phi\left(\frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right)$$

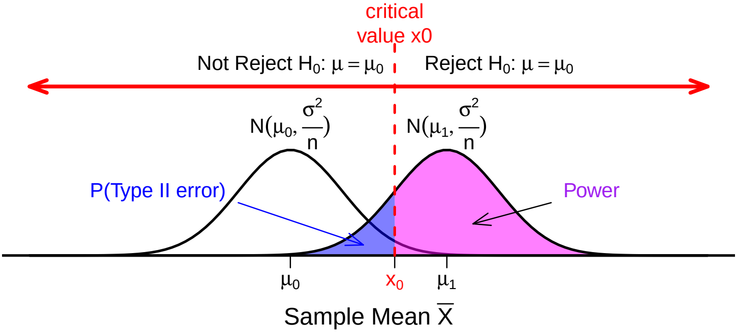
$$\mathrm{P}(\text{Type II error}) = \mathrm{P}(\overline{X} < x_0 \mid H_1: \ \mu = \mu_1) = \Phi\left(\frac{x_0 - \mu_1}{\sigma/\sqrt{n}}\right)$$



As $\overline{X} \sim N(\mu_0, \sigma^2/n)$,

$$\mathrm{P}(\text{Type I error}) = \mathrm{P}(\overline{X} > x_0 \mid H_0: \ \mu = \mu_0) = 1 - \Phi\left(\frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$\mathrm{P}(\text{Type II error}) = \mathrm{P}(\overline{X} < x_0 \mid H_1: \ \mu = \mu_1) = \Phi\left(\frac{x_0 - \mu_1}{\sigma/\sqrt{n}}\right)$$



Generalized Likelihood Ratio Tests

How to perform likelihood ratio test if H_0 or H_1 or both are composite?

General framework: for Data $\sim f(\cdot \mid \theta)$, we test

$$H_0\colon \theta \in \Omega_0, \quad H_1\colon \theta \in \Omega_1$$

where Ω_0, Ω_1 are sets of possible parameter values.

- **Ex1**: for $N(\mu, 1)$, testing $H_0\colon \mu = 0$ vs $H_1\colon \mu \neq 0$
  - $\theta = \mu$, $\Omega_0 = \{0\}$, $\Omega_1 = (-\infty, 0) \cup (0, \infty)$
- **Ex2**: for $N(\mu, \sigma^2)$, testing $H_0\colon \mu = 0$ vs $H_1\colon \mu \neq 0$, σ^2 is unknown
  - $\theta = (\mu, \sigma^2)$, $\Omega_0 = \{0\} \times (0, \infty)$, $\Omega_1 = \big((-\infty, 0) \cup (0, \infty)\big) \times (0, \infty)$

Generalized Likelihood Ratio Tests

How to perform likelihood ratio test if H_0 or H_1 or both are composite?

General framework: for Data $\sim f(\cdot \mid \theta)$, we test

$$\mathrm{H}_0\colon\ \theta \in \Omega_0, \quad \mathrm{H}_1\colon\ \theta \in \Omega_1$$

where Ω_0, Ω_1 are sets of possible parameter values.

- **Ex1**: for $\mathrm{N}(\mu, 1)$, testing $\mathrm{H}_0\colon\ \mu = 0$ vs $\mathrm{H}_1\colon\ \mu \neq 0$
  - $\theta = \mu$, $\Omega_0 = \{0\}$, $\Omega_1 = (-\infty, 0) \cup (0, \infty)$
- **Ex2**: for $\mathrm{N}(\mu, \sigma^2)$, testing $\mathrm{H}_0\colon\ \mu = 0$ vs $\mathrm{H}_1\colon\ \mu \neq 0$, σ^2 is unknown
  - $\theta = (\mu, \sigma^2)$, $\Omega_0 = \{0\} \times (0, \infty)$, $\Omega_1 = \big((-\infty, 0) \cup (0, \infty)\big) \times (0, \infty)$
- **Ex3**: for Exponential(λ), testing $\mathrm{H}_0\colon\ \lambda = 1$ vs $\mathrm{H}_1\colon\ \lambda \neq 1$
  - $\theta = \lambda$, $\Omega_0 = \{1\}$, $\Omega_1 = (0, 1) \cup (1, \infty)$

Generalized Likelihood Ratio (GLR) Tests

One might intend to define the generalized likelihood ratio test statistic to be

$$\Lambda^* = \frac{\max_{\theta \in \Omega_0} \text{Lik}(\theta)}{\max_{\theta \in \Omega_1} \text{Lik}(\theta)} \qquad \begin{matrix} \leftarrow \text{max likelihood under } H_0 \\ \leftarrow \text{max likelihood under } H_1 \end{matrix}$$

However, it's mathematically easier to calculate

$$\Lambda = \frac{\max_{\theta \in \Omega_0} \text{Lik}(\theta)}{\max_{\theta \in (\Omega_0 \cup \Omega_1)} \text{Lik}(\theta)} \qquad \begin{matrix} \leftarrow \text{max likelihood under } H_0 \\ \leftarrow \text{max likelihood under } H_0 \text{ or } H_1 \end{matrix}$$

Using Λ^* or Λ makes no difference:

- Usually we reject H_0 only if Λ^* is small
- Note $\Lambda = \min(\Lambda^*, 1)$. $\Lambda \neq \Lambda^*$ only when $\Lambda^* > 1$, and we won't reject H_0 when $\Lambda^* > 1$

Example — Normal LRT (Two-Sided, σ^2 Known)

For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with known σ^2, we want to test

$$H_0\colon\ \mu = \mu_0, \quad \text{against} \quad H_1\colon\ \mu \neq \mu_0.$$

Recall the likelihood of μ for normal is

$$L(\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^n (X_i - \mu)^2)$$

Under H_0, the max $L(\mu)$ is simply $L(\mu_0)$.
Under H_0 or H_1, the max $L(\mu)$ is $L(\overline{X})$. The GLR is thus

$$\Lambda = \frac{L(\mu_0)}{L(\overline{X})} = \frac{\exp(\frac{-1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu_0)^2)}{\exp(\frac{-1}{2\sigma^2} \sum_{i=1}^n (X_i - \overline{X})^2)} = \exp\left(-\frac{n(\overline{X} - \mu_0)^2}{2\sigma^2}\right)$$

Rejecting H_0 if $\Lambda < k$ is equivalent to rejecting H_0 if

$$|Z| = \left|\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}\right| > \text{some constant.}$$

This is the usual two-sided z-test.

Example — Normal LRT (Upper One-Sided, σ^2 Known)

For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with known σ^2, we want to test

$$\mathrm{H}_0\colon\ \mu = \mu_0, \quad \text{against} \quad \mathrm{H}_1\colon\ \mu > \mu_0.$$

Under H_0, the max $L(\mu)$ is again $L(\mu_0)$.
Under H_0 or H_1,

$$\max_{\mu \geq \mu_0} L(\mu) = \begin{cases} L(\overline{X}) & \text{if } \overline{X} \geq \mu_0 \\ L(\mu_0) & \text{if } \overline{X} < \mu_0. \end{cases}$$

The GLR is thus

$$\Lambda = \begin{cases} \exp(-n(\overline{X} - \mu_0)^2 / 2\sigma^2) & \text{if } \overline{X} \geq \mu_0 \\ 1 & \text{if } \overline{X} < \mu_0. \end{cases}$$

Rejecting H_0 if $\Lambda < k$ is equivalent to rejecting H_0 if

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > \text{some constant.}$$

This is the usual upper one-sided z-test.

Example — Normal LRT (Lower One-Sided, σ^2 Known)

Similarly, one can show that the GLR test for

$$H_0\colon \mu = \mu_0, \quad \text{against} \quad H_1\colon \mu < \mu_0.$$

is the usually lower one sided z-test that rejects H_0 if

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < \text{some constant.}$$

Example — Normal LRT (Two-Sided, σ^2 Unknown)

For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with unknown σ^2, we want to test

$$H_0\colon\ \mu = \mu_0, \quad \text{against} \quad H_1\colon\ \mu \neq \mu_0.$$

Recall the likelihood of (μ, σ^2) for normal is

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^{n} (X_i - \mu)^2)$$

Under H_0, the likelihood $L(\mu, \sigma^2)$ is maximized when

$$\mu = \mu_0, \quad \sigma^2 = \widehat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2.$$

and thus

$$\begin{aligned} L(\mu_0, \widehat{\sigma}_0^2) &= (2\pi\widehat{\sigma}_0^2)^{-\frac{n}{2}} \exp(-\tfrac{1}{2\widehat{\sigma}_0^2} \overbrace{\textstyle\sum_{i=1}^{n} (X_i - \mu_0)^2}^{n\widehat{\sigma}_0^2}) \\ &= (2\pi\widehat{\sigma}_0^2)^{-\frac{n}{2}} e^{-n/2}. \end{aligned}$$

Under H_0 or H_1, the likelihood $L(\mu, \sigma^2)$ is maximized when

$$\mu = \overline{X}, \quad \sigma^2 = \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

and thus

$$\begin{aligned} L(\overline{X}, \widehat{\sigma}^2) &= (2\pi\widehat{\sigma}^2)^{-\frac{n}{2}} \exp\Big(-\tfrac{1}{2\widehat{\sigma}^2} \overbrace{\sum_{i=1}^{n} (X_i - \overline{X})^2}^{n\widehat{\sigma}^2}\Big) \\ &= (2\pi\widehat{\sigma}^2)^{-\frac{n}{2}} e^{-n/2}. \end{aligned}$$

The GLR is thus

$$\Lambda = \frac{L(\mu_0, \widehat{\sigma}_0^2)}{L(\overline{X}, \widehat{\sigma}^2)} = \frac{(2\pi\widehat{\sigma}_0^2)^{-\frac{n}{2}} e^{-n/2}}{(2\pi\widehat{\sigma}^2)^{-\frac{n}{2}} e^{-n/2}} = \frac{(\widehat{\sigma}_0^2)^{-\frac{n}{2}}}{(\widehat{\sigma}^2)^{-\frac{n}{2}}} = \left(\frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right)^{-n/2}$$

and consequently

$$\begin{aligned} \Lambda^{-2/n} = \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} &= \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2 + n(\overline{X} - \mu_0)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \\ &= 1 + \frac{n(\overline{X} - \mu_0)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}. \end{aligned}$$

$$\Lambda^{-2/n} = 1 + \frac{n(\overline{X} - \mu_0)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = 1 + \frac{n(\overline{X} - \mu_0)^2}{(n-1)S^2} = 1 + \frac{T^2}{n-1}$$

where

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \text{ is the sample variance, and}$$

$$T = \frac{\overline{X} - \mu_0}{\sqrt{S^2/n}} \text{ is the usual t-statistic}$$

Rejecting H_0 if $\Lambda < k$ is equivalent to rejecting H_0 if

$$|T| = \left| \frac{\overline{X} - \mu_0}{\sqrt{S^2/n}} \right| > \text{some constant.}$$

The GLR test is equivalent to the usual two-sided t-test.

Example — Binomial LRT

For $X \sim \text{Bin}(n, p)$, we want to test

$$\text{H}_0\colon\ p = p_0, \quad \text{against} \quad \text{H}_1\colon\ p \neq p_0.$$

Recall the likelihood of p for Binomial is

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Under H_0, the max $L(p)$ is simply $L(p_0)$.
Under H_0 or H_1, $L(p)$ is maximized when p is the MLE $\widehat{p} = X/n$.
The GLR is thus

$$\Lambda = \frac{L(p_0)}{L(\widehat{p})} = \frac{p_0^X (1-p_0)^{n-X}}{\widehat{p}^X (1-\widehat{p})^{n-X}} = \left(\frac{np_0}{X}\right)^X \left(\frac{n(1-p_0)}{n-X}\right)^{n-X}$$

The GLR statistic is different from the typical one-sample z-stat for proportions:

$$Z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}.$$

Example — Two Sample Problems

Consider two normal random samples of size n_1 and n_2 respectively

$$\left.\begin{array}{ll} X_{11}, X_{12}, \ldots, X_{1n_1} & \overset{\text{iid}}{\sim} N(\mu_1, \sigma^2) \\ X_{21}, X_{22}, \ldots\ldots, X_{2n_2} & \overset{\text{iid}}{\sim} N(\mu_2, \sigma^2) \end{array}\right\} \rightarrow \text{indep., same } \sigma^2.$$

The parameters μ_1, μ_2, and σ^2 are <u>unknown</u>.
We want to test whether the two means are equal

$$\mathrm{H}_0\colon \mu_1 = \mu_2 \quad \text{against} \quad \mathrm{H}_1\colon \mu_1 \neq \mu_2.$$

using GLR as follows.

1. Find the MLE's for μ_1, μ_2, and σ^2 and the max likelihood under $\mathrm{H}_0 \cup \mathrm{H}_1$
2. Find the MLE's for μ_1, μ_2, and σ^2 and the max likelihood under H_0
3. Take the ratio of the two max likelihood

The likelihood and log-likelihood of (μ_1, μ_2, σ^2) based on the two samples are

$$L(\mu_1, \mu_2, \sigma^2) = (2\pi\sigma^2)^{-\frac{n_1+n_2}{2}} e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^{n_1}(X_{1i}-\mu_1)^2+\sum_{j=1}^{n_2}(X_{2j}-\mu_2)^2]}$$

$$\begin{aligned}\ell(\mu_1, \mu_2, \sigma^2) &= -\frac{n_1+n_2}{2}\log(2\pi\sigma^2) \\ &\quad - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n_1}(X_{1i}-\mu_1)^2+\sum_{j=1}^{n_2}(X_{2j}-\mu_2)^2\right)\end{aligned}$$

To solve for the MLE

$$\begin{cases} 0 = \frac{\partial \ell}{\partial \mu_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n_1}(X_{1i}-\mu_1) \\ 0 = \frac{\partial \ell}{\partial \mu_2} = \frac{1}{\sigma^2}\sum_{j=1}^{n_2}(X_{2i}-\mu_2) \\ 0 = \frac{\partial \ell}{\partial \sigma^2} = \frac{-(n_1+n_2)}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}[\sum_{i=1}^{n_1}(X_{1i}-\mu_1)^2+\sum_{j=1}^{n_2}(X_{2j}-\mu_2)^2] \end{cases}$$

The first two equations immediately gives

$$\widehat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i} \stackrel{\text{def}}{=} \overline{X}_1 \quad \text{and} \quad \widehat{\mu}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j} \stackrel{\text{def}}{=} \overline{X}_2.$$

Plugging $\mu_1 = \overline{X}_1$ and $\mu_2 = \overline{X}_2$ into the 3rd equation, we get the MLE for σ^2

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2}{n_1 + n_2}.$$

Plugging the MLEs back to the Likelihood, we get

$$\begin{aligned} L(\overline{X}_1, \overline{X}_2, \widehat{\sigma}^2) &= (2\pi\widehat{\sigma}^2)^{-\frac{n_1+n_2}{2}} \exp\left(\frac{-1}{2\widehat{\sigma}^2}[\overbrace{\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2+\sum_{j=1}^{n_2}(X_{2j}-\overline{X}_2)^2}^{=(n_1+n_2)\widehat{\sigma}^2}]\right) \\ &= (2\pi\widehat{\sigma}^2)^{-\frac{n_1+n_2}{2}} e^{-\frac{n_1+n_2}{2}} \end{aligned}$$

Under H_0: $\mu_1 = \mu_2$, the problem reduces to the MLE and max likelihood with $n_1 + n_2$ observations

$$X_{11}, \ldots, X_{1n_1},\ X_{21}, \ldots, X_{2n_2}.$$

The MLE's for μ and σ^2 are respectively

$$\widehat{\mu} = \frac{\sum_{i=1}^{n_1} X_{1i} + \sum_{j=1}^{n_2} X_{2j}}{n_1 + n_2} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2} \stackrel{\text{def}}{=} \overline{\overline{X}},$$

$$\widehat{\sigma}_0^2 = \frac{\sum_{i=1}^{n_1} (X_{1i} - \overline{\overline{X}})^2 + \sum_{j=1}^{n_2} (X_{2j} - \overline{\overline{X}})^2}{n_1 + n_2}.$$

and the max likelihood under H_0 is

$$L(\overline{\overline{X}}, \overline{\overline{X}}, \widehat{\sigma}^2) = (2\pi\widehat{\sigma}^2)^{-\frac{n_1+n_2}{2}} \exp\left(\frac{-1}{2\widehat{\sigma}_0^2}\Big[\overbrace{\sum_{i=1}^{n_1} (X_{1i} - \overline{\overline{X}})^2 + \sum_{j=1}^{n_2} (X_{2j} - \overline{\overline{X}})^2}^{=(n_1+n_2)\widehat{\sigma}_0^2}\Big]\right)$$

$$= (2\pi\widehat{\sigma}_0^2)^{-\frac{n_1+n_2}{2}} e^{-\frac{n_1+n_2}{2}}$$

The GLR is thus

$$\Lambda = \frac{L(\overline{\overline{X}}, \overline{\overline{X}}, \widehat{\sigma}_0^2)}{L(\overline{X}_1, \overline{X}_2, \widehat{\sigma}^2)} = \frac{(2\pi\widehat{\sigma}_0^2)^{-\frac{n_1+n_2}{2}} e^{-\frac{n_1+n_2}{2}}}{(2\pi\widehat{\sigma}^2)^{-\frac{n_1+n_2}{2}} e^{-\frac{n_1+n_2}{2}}} = \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{-\frac{n_1+n_2}{2}}$$

and consequently

$$\Lambda^{-\frac{2}{n_1+n_2}} = \frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2} = \frac{\sum_{i=1}^{n_1}(X_{1i}-\overline{\overline{X}})^2+\sum_{j=1}^{n_2}(X_{2j}-\overline{\overline{X}})^2}{\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2+\sum_{j=1}^{n_2}(X_{2j}-\overline{X}_2)^2}$$

Using the useful identity

$$\sum_{i=1}^{n_1}(X_{1i}-\overline{\overline{X}})^2 = \sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2 + n_1(\overline{X}_1 - \overline{\overline{X}})^2$$
$$\sum_{j=1}^{n_2}(X_{2j}-\overline{\overline{X}})^2 = \sum_{j=1}^{n_2}(X_{2j}-\overline{X}_2)^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2$$

we get

$$\Lambda^{-\frac{2}{n_1+n_2}} = 1 + \frac{n_1(\overline{X}_1 - \overline{\overline{X}})^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2}{\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2+\sum_{j=1}^{n_2}(X_{2j}-\overline{X}_2)^2}$$

As

$$\overline{X}_1 - \overline{\overline{X}} = \overline{X}_1 - \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2} = \frac{n_2(\overline{X}_1 - \overline{X}_2)}{n_1 + n_2},$$
$$\overline{X}_2 - \overline{\overline{X}} = \frac{n_1(\overline{X}_2 - \overline{X}_1)}{n_1 + n_2}.$$

we get

$$n_1(\overline{X}_1 - \overline{\overline{X}})^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{X}_1 - \overline{X}_2)^2$$

and thus

$$\Lambda^{-\frac{2}{n_1+n_2}} = 1 + \frac{n_1 n_2}{n_1 + n_2}\frac{(\overline{X}_1 - \overline{X}_2)^2}{\widehat{\sigma}^2}$$

Rejecting H_0 when GLR$=\Lambda$ is small is equivalent to rejecting H_0 when $(\overline{X}_1 - \overline{X}_2)^2/\widehat{\sigma}^2$ is large.

Distribution of the Two-Sample T-Statistic (Equal σ^2)

The *two-sample T-statistic* is defined to be

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})S^2}},$$

where

$$S^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2}{n_1 + n_2 - 2} = \frac{n_1 + n_2}{n_1 + n_2 - 2}\widehat{\sigma}^2,$$

which is proportional to $(\overline{X}_1 - \overline{X}_2)^2/\widehat{\sigma}^2$.

Distribution of the Two-Sample T-Statistic (Equal σ^2)

The *two-sample T-statistic* is defined to be

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})S^2}},$$

where

$$S^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2}{n_1 + n_2 - 2} = \frac{n_1 + n_2}{n_1 + n_2 - 2}\widehat{\sigma}^2,$$

which is proportional to $(\overline{X}_1 - \overline{X}_2)^2/\widehat{\sigma}^2$.

$$\left.\begin{array}{l} \frac{1}{\sigma^2}\sum_{i=1}^{n_1}(X_{1i} - \overline{X}_1)^2 \sim \chi^2_{n_1-1} \\ \frac{1}{\sigma^2}\sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2 \sim \chi^2_{n_2-1} \end{array}\right\}\text{indep.} \Rightarrow V = \frac{(n_1+n_2-2)S^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}.$$

Moreover, under H_0: $\mu_1 = \mu_2$,

$$\overline{X}_1 - \overline{X}_2 \sim N\left(0, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right) \Rightarrow Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})\sigma^2}} \sim N(0,1).$$

Putting everything together, we have

$$T = \frac{Z}{\sqrt{V/(n_1 + n_2 - 2)}} \sim t_{n_1+n_2-2}.$$

Example — Exponential

Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$.
Testing $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$.

- likelihood:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-X_i \lambda} = \lambda^n \exp(-\lambda \sum_{i=1}^{n} X_i) = \lambda^n e^{-n\lambda \overline{X}}$$

- Under H_0, max $L(\lambda) = L(\lambda_0)$
- Under H_0 or H_1, $L(\lambda)$ is maximized when λ is the MLE $\widehat{\lambda} = 1/\overline{X}$.
- The GLR is thus

$$\Lambda = \frac{L(\lambda_0)}{L(1/\overline{X})} = \frac{\lambda_0^n e^{-n\lambda_0 \overline{X}}}{\overline{X}^{-n} e^{-n}} = e^n (\lambda_0 \overline{X})^n e^{-n\lambda_0 \overline{X}}.$$

Null Distribution of GLR

To use the GLR as a test statistic for testing H_0 vs H_1...

- $\Lambda \leq 1$ always
- If $\Lambda \approx 1$, data are consistent with H_0
  - no reason to reject H_0
- $\Lambda \ll 1$ is evidence for H_1

How small does Λ need to be to reject H_0? Our goal:

$$\mathrm{P}(\Lambda < \text{(the threshold we choose)} \mid H_0 \text{ is true}) \approx \alpha$$

We need to know the (approximate) null distribution of Λ

Null Distribution of GLR

Under some regularity conditions, the large sample distribution of GLR is

$$-2\log(\Lambda) \approx \chi^2_{d\,-\,d_0}, \text{ where } \begin{cases} d = \text{ dimension of } \Omega_0 \cup \Omega_1 \\ d_0 = \text{ dimension of } \Omega_0 \end{cases}$$

Part of the conditions:
Ω_0 is interior to $\Omega_0 \cup \Omega_1$, not at the boundary

- Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, 1)$
  testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ . . . . valid
- Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ where σ^2 is unknown
  testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$ . . . . valid
- Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ Exponential(λ)
  testing $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$ . . . . valid
- Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ where σ^2 is known
  testing $H_0 : \mu = 0$ vs $H_1 : \mu > 0$ . . . . NOT valid

How to Determine d & d_0?

- Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, with unknown μ & known σ^2
  test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\Rightarrow d_0 = 0$, $d = 1$
- Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$with μ & σ^2 unknown,
  test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\Rightarrow d_0 = 1$, $d = 2$

How to Determine d & d_0?

- Data: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, with unknown μ & known σ^2
  test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\Rightarrow d_0 = 0$, $d = 1$
- Data: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$with μ & σ^2 unknown,
  test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\Rightarrow d_0 = 1$, $d = 2$
- Data: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ with μ & σ^2 unknown,
  test $H_0 : (\mu, \sigma^2) = (0, 1)$ vs $H_1 : (\mu, \sigma^2) \neq (0, 1)$
  $\Rightarrow d_0 = 0$, $d = 2$
- Data: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim}$ Exponential(λ),
  test $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$
  $\Rightarrow d_0 = 0$, $d = 1$

Back to the GLR for Exponential

Data: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$,
testing $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$.

Recall the GLR is

$$\Lambda = e^n (\lambda_0 \overline{X})^n e^{-n\lambda_0 \overline{X}}$$

Then

$$-2\log(\Lambda) = 2n\log(\lambda_0 \overline{X}) - 2n(\lambda_0 \overline{X} - 1) \sim \chi^2_1.$$

At the $\alpha = 0.05$ significance level, we reject H_0: $\lambda = \lambda_0$ if

$$-2\log(\Lambda) > 3.84.$$

Back to Normal LRT w/ Known σ^2

For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with known σ^2, we want to test

$$\text{H}_0\text{: } \mu = \mu_0, \quad \text{against} \quad \text{H}_1\text{: } \mu \neq \mu_0.$$

Recall the GLR is

$$\Lambda = \frac{L(\mu_0)}{L(\overline{X})} = \frac{\exp(\frac{-1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu_0)^2)}{\exp(\frac{-1}{2\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2)} = \exp\left(-\frac{n(\overline{X} - \mu_0)^2}{2\sigma^2}\right)$$

Observe

$$-2\log(\Lambda) = \frac{n(\overline{X} - \mu_0)^2}{\sigma^2}$$

Under H_0,

$$\overline{X} \sim \text{N}(\mu_0, \sigma^2/n) \Rightarrow \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} \sim \text{N}(0,1) \Rightarrow \frac{n(\overline{X} - \mu_0)^2}{\sigma^2} \sim \chi_1^2$$

In this special case, the asymptotic approximation is the exact distrib.

Back to Normal LRT w/ Unknown σ^2

For $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ w/ unknown σ^2,
testing H_0: $\mu = \mu_0$, against H_1: $\mu \neq \mu_0$. Recall the GLR is

$$\Lambda = \left(1 + \frac{n(\overline{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right)^{-n/2}$$

and consequently

$$-2\log\Lambda = n\log\left(1 + \frac{n(\overline{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right) = n\log\left(1 + \frac{(\overline{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2/n}\right)$$

Under H_0, $\overline{X} \to \mu_0$ and $\sum_{i=1}^n (X_i - \overline{X})^2/n \to \sigma^2$ as $n \to \infty$, we know

$$\frac{(\overline{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2/n} \to 0 \text{ in prob. as } n \to \infty.$$

and $\log(1 + x) \approx x$ when $x \approx 0$, we have

$$-2\log\Lambda \approx n \cdot \frac{(\overline{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2/n} \to \frac{n(\overline{X} - \mu_0)^2}{\sigma^2} \sim \chi_1^2.$$