Lecture 8: Generalized Likelihood Ratio Tests

STA261 – Probability & Statistics II

Ofir Harari

Department of Statistical Sciences

University of Toronto

Outline

Generalized Likelihood Ratio Tests   Examples of GLRTs   Comparing Two Population Means
○○○○○○○○○○○   ○○○○○○○○○○○○
○○○○○○○○○   ○○○○○○

Introduction

- Recall that for simple hypotheses, $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$, we defined the Likelihood Ratio statistic

$$\lambda(\underline{X}) = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)}$$

  (to be evaluated at θ_1 and θ_0, based on a sample $X_1, \ldots, X_n \sim f_\theta$).

- Moreover, the beautiful Neyman-Pearson Lemma has taught us that the Likelihood Ratio Test (LRT), based on the rejection region

$$\mathcal{C} = \{\underline{x} : \lambda(\underline{x}) \geq c\} \quad (\text{with } \mathbb{P}\left(\underline{X} \in \mathcal{C} \,\middle|\, \theta = \theta_0\right) = \alpha)$$

  was the Most Powerful (MP) test at level α.

- Sometimes we luck out and can extend these tests to Uniformly Most Powerful (UMP) tests for composite hypotheses – as was the case with the one-tailed tests for the Normal mean with known variance.

this is possible if the rejection region does not dependent on alternative hypothesis

Generalized Likelihood Ratio Tests   Examples of GLRTs   Comparing Two Population Means
○○○○○○○○○○   ○○○○○○○○○○○○
○○○○○○○○○   ○○○○○○

Introduction (cont.)

- Unfortunately, UMP tests – especially practical ones – hardly ever exist

- We have already explained why no two-tailed UMP test for the Normal mean exists

  since one tail test will outperform two tail test

- Even when testing $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu > \mu_0$ for Normal data with *unknown* variance, no test that does not depend on σ^2 can outperform the test that rejects $\mathcal{H}_0$ whenever $\overline{X} \geq \mu_0 + \dfrac{\sigma}{\sqrt{n}} z_{1-\alpha}$ – the UMP test.

  if variance unknown, no test can be more powerful than one in which the variance is known

- In this lecture we will introduce a class of hypotheses tests that are not generally UMP. However –

  - when restricted to simple hypotheses, they are (sometimes) MP tests, and

  - the distribution of the test statistic under $\mathcal{H}_0$ is often known – exactly or approximately.

    asymptotic

UNIVERSITY OF TORONTO

Generalized Likelihood Ratio Tests (GLRT)

Definition

Consider testing $\mathcal{H}_0 : \theta \in \Theta_0$ vs. $\mathcal{H}_1 : \theta \in \Theta_1$, such that $\Theta_0 \uplus \Theta_1 = \Theta$ (the entire parameter space), based on $X_1, \ldots, X_n \sim f_\theta$.

1. The statistic

numerator not limiting to null or alternative hypothesis => MLE

$$\Lambda(\underline{X}) = \frac{\max_{\theta \in \Theta} \mathcal{L}(\theta)}{\max_{\theta \in \Theta_0} \mathcal{L}(\theta)}$$

is called the *generalized likelihood ratio* (GLR).

2. The test based on the rejection region

$$\mathcal{C} = \{\underline{X} \in \mathbb{R}^n \ : \ \Lambda(\underline{X}) \geq c\}$$

for a critical value c satisfying

$$\mathbb{P}\left(\Lambda(\underline{X}) \geq c \middle| \theta \in \Theta_0\right) = \alpha,$$

satisfying probability of type I error bounded above by alpha

is called the *generalized likelihood ratio test* (GLRT) at level α.

UNIVERSITY OF
TORONTO

Generalized Likelihood Ratio Tests (cont.)

$$\Lambda(\underline{X}) = \frac{\max_{\theta \in \Theta} \mathcal{L}(\theta)}{\max_{\theta \in \Theta_0} \mathcal{L}(\theta)}$$

- To find a GLRT at level α one needs to –

    1. Find the <u>unrestricted MLE</u>, $\widehat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$, and calculate the numerator $\mathcal{L}(\hat{\theta})$,

    2. Find the <u>restricted MLE</u>, $\widehat{\theta_0} = \arg\max_{\theta \in \Theta_0} \mathcal{L}(\theta)$, and calculate the denominator $\mathcal{L}(\hat{\theta}_0)$,   restricted to null

    3. If possible, find a simpler statistic $T(\underline{X})$ such that $\Lambda(\underline{X})$ is a monotonically increasing function of $T(\underline{X})$, and whose distribution (assuming $\theta = \widehat{\theta}_0$) is known, and   to calculate alpha

    4. Find a critical value c such that $\mathbb{P}\left(T(\underline{X}) \geq c \middle| \theta = \widehat{\theta}_0\right) = \alpha$.

One Sample t test for the Normal mean

- Consider testing $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu \neq \mu_0$, based on $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where now σ^2 is assumed to be unknown.

- The likelihood is given by

$$\mathcal{L}(\mu, \sigma^2) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(X_i - \mu)^2\right\}.$$

- Plugging in the unrestricted MLEs $\widehat{\mu} = \overline{X}$ and $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, we have

$$\mathcal{L}(\widehat{\mu}, \widehat{\sigma}^2) = \left(2\pi\widehat{\sigma}^2\right)^{-n/2} \exp\left\{-\frac{1}{2\widehat{\sigma}^2} \sum_{i=1}^{n}(X_i - \widehat{\mu})^2\right\}$$

$$= \left(2\pi\widehat{\sigma}^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\frac{2}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}\right\} = \left(2\pi e\widehat{\sigma}^2\right)^{-n/2}.$$

UNIVERSITY OF TORONTO

One Sample t Test (cont.)

mean of one number mu_0 is mu_0, its the only value in the set

- Under $\mathcal{H}_0 : \mu = \mu_0$, the restricted MLE of μ is... $\widehat{\mu}_0 = \mu_0$.

- One of your early practice problems was to show that if $\mu = \mu_0$ was known, the restricted MLE for σ^2 was $\widehat{\sigma}_0^2 = \dfrac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2$. One of your early practice problems was to show that if $\mu = \mu_0$ was known, the restricted MLE for σ^2 was $\widehat{\sigma}_0^2 = \dfrac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2$.

- The denominator of $\Lambda(\underline{X})$ is thus

$$\mathcal{L}(\widehat{\mu}_0, \widehat{\sigma}_0^2) = \left(2\pi\widehat{\sigma}_0^2\right)^{-n/2} \exp\left\{-\frac{1}{2\widehat{\sigma}_0^2} \sum_{i=1}^{n} (X_i - \widehat{\mu}_0)^2\right\}$$

$$= \left(2\pi\widehat{\sigma}_0^2\right)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{\frac{2}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2}\right\} = \left(2\pi e \widehat{\sigma}_0^2\right)^{-n/2}.$$

- Finally

$$\Lambda(\underline{X}) = \frac{\mathcal{L}(\widehat{\mu}, \widehat{\sigma}^2)}{\mathcal{L}(\widehat{\mu}_0, \widehat{\sigma}_0^2)} = \frac{\left(2\pi e \widehat{\sigma}^2\right)^{-n/2}}{\left(2\pi e \widehat{\sigma}_0^2\right)^{-n/2}} = \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{n/2}.$$

One Sample t Test (cont.)

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

- Write

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

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

$$= 0$$

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

then

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

simplification of GLR

One Sample t test (cont.)

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

$$= \left\{ 1 + \frac{1}{n-1} \cdot \frac{\left(\overline{X} - \mu_0\right)^2}{S^2/n} \right\}^{n/2} = \left\{ 1 + \frac{1}{n-1}\left(\frac{\overline{X} - \mu_0}{S/\sqrt{n}}\right)^2 \right\}^{n/2}$$

- Denoting

$$\mathcal{T} = \frac{\overline{X} - \mu_0}{S/\sqrt{n}},$$

  it is clear that $\Lambda(\underline{X})$ is monotonically increasing in $|\mathcal{T}|$, hence the rejection region of the GLRT will be of the form

$$\mathcal{C} = \left\{ \Lambda(\underline{X}) \geq c \right\} = \left\{ |\mathcal{T}| \geq c^* \right\}.$$

- In Lecture 1 we showed that if $\mathcal{H}_0 : \mu = \mu_0$ is true then $\mathcal{T} \sim t_{n-1}$.

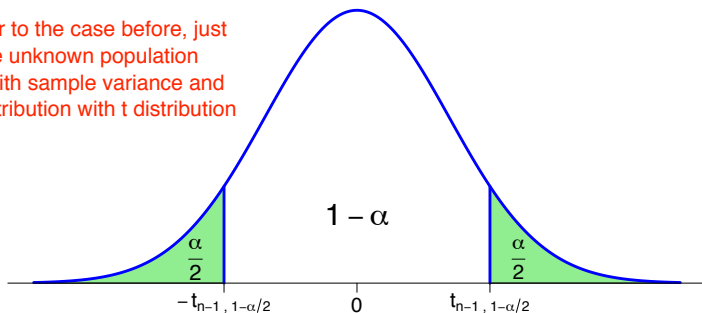# One Sample t Test (cont.)

- Calculating the critical value c^*,

$$\alpha = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| \mu = \mu_0\right) = \mathbb{P}\left(\left|\mathcal{T}\right| \geq c^* \middle| \mu = \mu_0\right)$$

very similar to the case before, just replace the unknown population variance with sample variance and normal distribution with t distribution

- From the figure it is clear that $c^* = t_{n-1,1-\alpha/2}$

UNIVERSITY OF TORONTO

One Sample t Test (cont.)

- To conclude, the rejection region of the GLRT to test $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu \neq \mu_0$ at level α is

$$\mathcal{C} = \left\{ |\mathcal{T}| = \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \right| \geq t_{n-1,1-\alpha/2} \right\}.$$

- This is the test corresponding to the 100(1 − α)% confidence interval we found for μ: $\overline{X} \pm \dfrac{S}{\sqrt{n}} t_{n-1,1-\alpha/2}$

- Similarly, one can show that the rejection region of the right-tailed test $\mathcal{H}_0 : \mu \leq \mu_0$ vs. $\mathcal{H}_1 : \mu > \mu_0$ is

$$\mathcal{C}_{\text{Right}} = \left\{ \mathcal{T} \geq t_{n-1,1-\alpha} \right\}$$

and that of the left-tailed test $\mathcal{H}_0 : \mu \geq \mu_0$ vs. $\mathcal{H}_1 : \mu < \mu_0$ is

$$\mathcal{C}_{\text{Left}} = \left\{ \mathcal{T} \leq -t_{n-1,1-\alpha} \right\}.$$

Example

df = 8

- Suppose that we have a sample of size 9 from a $\mathcal{N}(\mu, \sigma^2)$ distribution, and we wish to test $\mathcal{H}_0 : \mu = 23$ vs. $\mathcal{H}_1 \neq 23$ at the 5% level. What would your conclusions be, if the sample mean and sample standard deviation were 25.5 and 4, respectively   2 tailed test   quantile = 97.5%

- The rejection region is

$$\mathcal{C} = \left\{ |\mathcal{T}| = \left| \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \right| \geq t_{n-1, 1-\alpha/2} \right\} = \left\{ |\mathcal{T}| \geq \underbrace{t_{8, 0.975}}_{2.306} \right\}$$

less than: cannot reject

- Here

$$\mathcal{T} = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{25.5 - 23}{4/3} = 1.875,$$

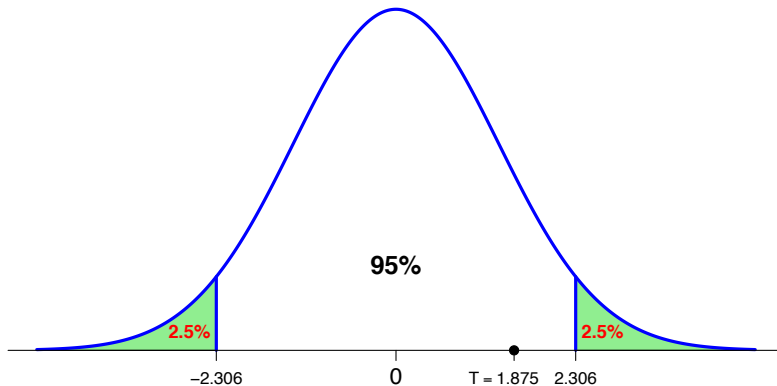hence $|\mathcal{T}| < 2.306$, and we cannot reject $\mathcal{H}_0$ at the 5% level.

- What level could we reject $\mathcal{H}_0$ at?

Example (cont.)

Example (cont.)

- Let us calculate the p-value for this test

- Here, observing a result that is "at least as extreme" as the one we observed means observing $|\mathcal{T}| \geq 1.875$ –
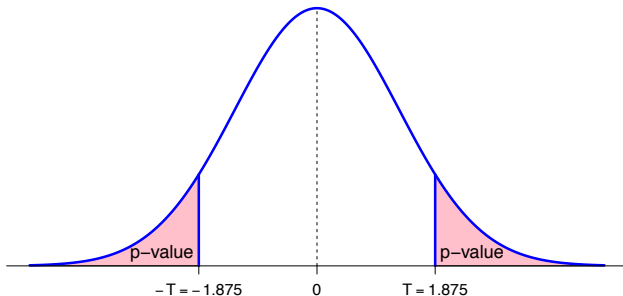
$$\text{p-value} = \mathbb{P}\left(|\mathcal{T}| \geq 1.875 \Big| \mu = 23\right) = 2\mathbb{P}\left(\mathcal{T} \geq 1.875 \Big| \mu = 23\right)$$
$$= 2\Big[1 - \mathbb{P}\left(\mathcal{T} \leq 1.875 \Big| \mu = 23\right)\Big]$$

Example (cont.)

$$\text{p-value} = 2\left[1 - \mathbb{P}\left(\mathcal{T} \leq 1.875 \Big| \mu = 23\right)\right]$$

- Recall that under $\mathcal{H}_0 : \mu = 23$, $\mathcal{T} \sim t_8$

- From the table, the best we can tell is

$$0.95 < \mathbb{P}\left(\mathcal{T} \leq 1.875 \Big| \mu = 23\right) < 0.975$$

$$\implies 0.025 < 1 - \mathbb{P}\left(\mathcal{T} \leq 1.875 \Big| \mu = 23\right) < 0.05$$

$$\implies 0.05 < \text{p-value} = 2\left[1 - \mathbb{P}\left(\mathcal{T} \leq 1.875 \Big| \mu = 23\right)\right] < 0.1$$

  – no wonder we could not reject $\mathcal{H}_0$ at the 5% level – but we would definitely reject it at the 10% level. right tail test at 5% would reject.

- Exact p-value:

  ```
  > 2*(1-pt(1.875, df=8))
  ```

  ```
  [1] 0.09765351
  ```

UNIVERSITY OF TORONTO

F test for the equality of variances

- Suppose now that we have two independent samples,
  $X_1, \ldots, X_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$, and we wish to test $\mathcal{H}_0 : \sigma_X^2 = \sigma_Y^2$ vs. $\mathcal{H}_0 : \sigma_X^2 \neq \sigma_Y^2$

- This time, the likelihood is given by

  mu cancel out in exp{}

  $$\mathcal{L}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2) = \left(2\pi\sigma_X^2\right)^{-m/2} \exp\left\{-\frac{1}{2\sigma_X^2}\sum_{i=1}^{m}(X_i - \mu_X)^2\right\}$$

  L = joint probability of all observations, both X and Y

  $$\times \left(2\pi\sigma_Y^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma_Y^2}\sum_{i=1}^{n}(Y_i - \mu_Y)^2\right\}$$

- It is easy to see that $\widehat{\mu}_X = \overline{X}$, $\widehat{\sigma}_X^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \overline{X})^2 = \frac{(m-1)S_X^2}{m}$,

  restricted MLE

  $\widehat{\mu}_Y = \overline{Y}$ and $\widehat{\sigma}_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{(n-1)S_Y^2}{n}$ are the unrestricted

  MLEs, and

  $$\mathcal{L}(\widehat{\mu}_X, \widehat{\mu}_Y, \widehat{\sigma}_X^2, \widehat{\sigma}_Y^2) \propto \left(\widehat{\sigma}_X^2\right)^{-m/2}\left(\widehat{\sigma}_Y^2\right)^{-n/2} \propto \left(S_X^2\right)^{-m/2}\left(S_Y^2\right)^{-n/2}$$

  what about MLE for mean?

Generalized Likelihood Ratio Tests   **Examples of GLRTs**   Comparing Two Population Means

○○○○○○○○○○○
○●○○○○○○○   ○○○○○○○○○○○○
○○○○○○   ○○○○○○○○○○○

when there is more than 1 param, are their respectiveMLE yield MLE as a whole

Test for the equality of variances (cont.)

- Under $\mathcal{H}_0$ we have $\sigma_X^2 = \sigma_Y^2 := \sigma^2$, and the likelihood becomes

$$\mathcal{L}(\mu_X, \mu_Y, \sigma^2) = (2\pi\sigma^2)^{-\frac{m+n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{m}(X_i - \mu_X)^2 + \sum_{i=1}^{n}(Y_i - \mu_Y)^2 \right] \right\}$$

note MLE for mu is same in unrestricted and restricted case

- The restricted MLEs are now $\widehat{\mu}_X^0 = \overline{X}$, $\widehat{\mu}_Y^0 = \overline{Y}$ and

since mean could be different while variance still equates

$$\widehat{\sigma}_0^2 = \frac{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(Y_i - \overline{Y})^2}{m+n} = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n},$$

weighted sum of the two sample variances,

and thus

$$\mathcal{L}(\widehat{\mu}_X^0, \widehat{\mu}_Y^0, \widehat{\sigma}_0^2) \propto (\widehat{\sigma}_0^2)^{-\frac{m+n}{2}} \propto \left[(m-1)S_X^2 + (n-1)S_Y^2 \right]^{-\frac{m+n}{2}}.$$

- The GLR is therefore

$$\Lambda(\underline{X}) \propto \frac{\left[(m-1)S_X^2 + (n-1)S_Y^2 \right]^{\frac{m+n}{2}}}{(S_X^2)^{\frac{m}{2}} (S_Y^2)^{\frac{n}{2}}}$$

Test for the equality of variances (cont.)

$$\Lambda(\underline{X}) \propto \frac{\left[(m-1)S_X^2 + (n-1)S_Y^2\right]^{\frac{m+n}{2}}}{\left(S_X^2\right)^{\frac{m}{2}} \left(S_Y^2\right)^{\frac{n}{2}}}$$

$$= \frac{\left[(m-1)S_X^2 + (n-1)S_Y^2\right]^{\frac{m}{2}}}{\left(S_X^2\right)^{\frac{m}{2}}} \frac{\left[(m-1)S_X^2 + (n-1)S_Y^2\right]^{\frac{n}{2}}}{\left(S_Y^2\right)^{\frac{n}{2}}}$$

$$\propto \left[1 + \frac{n-1}{m-1}\frac{S_Y^2}{S_X^2}\right]^{\frac{m}{2}} \left[1 + \frac{m-1}{n-1}\frac{S_X^2}{S_Y^2}\right]^{\frac{n}{2}} = \left[1 + \mathcal{F}^{*-1}\right]^{\frac{m}{2}} \left[1 + \mathcal{F}^*\right]^{\frac{n}{2}},$$

for $\mathcal{F}^* = \dfrac{(m-1)S_X^2}{(n-1)S_Y^2}$.

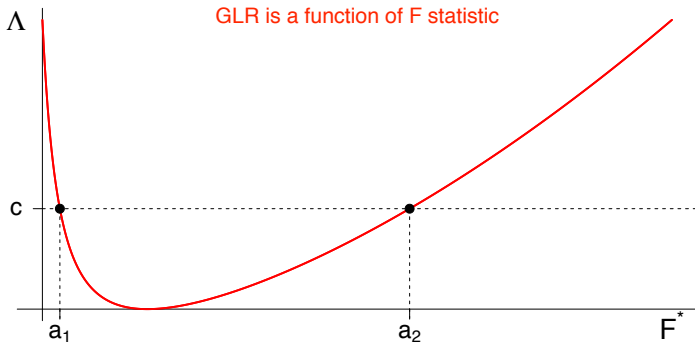## Test for the equality of variances (cont.)

$$\Lambda(\underline{X}) \propto \left[1 + \mathcal{F}^{*-1}\right]^{\frac{m}{2}} \left[1 + \mathcal{F}^{*}\right]^{\frac{n}{2}}$$

GLR is a function of F statistic

$$\mathcal{C} = \left\{\Lambda(\underline{X}) \geq c\right\} = \left\{\mathcal{F}^{*} \leq a_1\right\} \bigcup \left\{\mathcal{F}^{*} \geq a_2\right\}$$

Test for the equality of variances (cont.)

- We have shown that the rejection region of the GLRT can be written as

$$\mathcal{C} = \left\{ \mathcal{F}^* \leq a_1 \right\} \bigcup \left\{ \mathcal{F}^* \geq a_2 \right\}$$

  for $\mathcal{F}^* = \dfrac{(m-1)S_X^2}{(n-1)S_Y^2}$.

- Now set $\mathcal{F} = \dfrac{S_X^2}{S_Y^2} \left(= \dfrac{(n-1)\mathcal{F}^*}{m-1} \right)$ – clearly we can rewrite

$$\mathcal{C} = \left\{ \mathcal{F} \leq b_1 \right\} \bigcup \left\{ \mathcal{F} \geq b_2 \right\}.$$

- To determine b_1 and b_2, recall that they must satisfy

$$\alpha = \mathbb{P}\left(\underline{X} \in \mathcal{C} \Big| \mathcal{H}_0 \right) = \mathbb{P}\left(\mathcal{F} \leq b_1 \Big| \sigma_X^2 = \sigma_Y^2 \right) + \mathbb{P}\left(\mathcal{F} \geq b_2 \Big| \sigma_X^2 = \sigma_Y^2 \right)$$

- Crucially, at the very end of Lecture 1, we showed that if $\mathcal{H}_0 : \sigma_X^2 = \sigma_Y^2$ is true, $\mathcal{F} \sim F_{m-1, n-1}$.
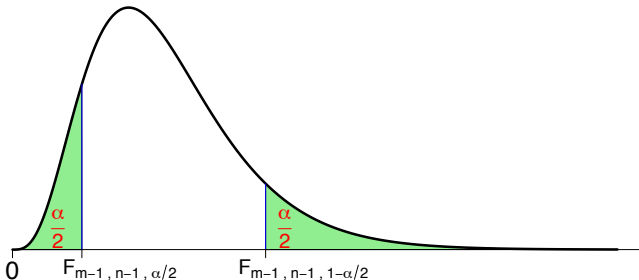
Test for the equality of variances (cont.)

$$\alpha = \mathbb{P}\left(\mathcal{F} \le b_1 \Big| \sigma_X^2 = \sigma_Y^2\right) + \mathbb{P}\left(\mathcal{F} \ge b_2 \Big| \sigma_X^2 = \sigma_Y^2\right)$$

- A natural choice would be $b_1 = F_{m-1,n-1,\alpha/2}$ and $b_2 = F_{m-1,n-1,1-\alpha/2}$ — then the rejection region of the GLRT becomes

$$\mathcal{C} = \left\{\mathcal{F} \le F_{m-1,n-1,\alpha/2}\right\} \bigcup \left\{\mathcal{F} \ge F_{m-1,n-1,1-\alpha/2}\right\}$$

for $\mathcal{F} = \dfrac{S_X^2}{S_Y^2}$.

Example

- Suppose that we observe independent samples of sizes $m = 13$ and $n = 11$ from two Normal populations, whose respective sample variances are $S_X^2 = 0.09$ and $S_Y^2 = 0.14$.

- To test $\mathcal{H}_0 : \sigma_X^2 = \sigma_Y^2$ at the 5% level, the rejection region is

$$\mathcal{C} = \left\{ \mathcal{F} \le F_{m-1,n-1,\alpha/2} \right\} \bigcup \left\{ \mathcal{F} \ge F_{m-1,n-1,1-\alpha/2} \right\}$$

F_{m = 12, n = 10}

$$= \left\{ \mathcal{F} \le \underbrace{F_{12,10,0.025}}_{0.3} \right\} \bigcup \left\{ \mathcal{F} \ge \underbrace{F_{12,10,0.975}}_{3.62} \right\}$$

- However, in this example

$$\mathcal{F} = \frac{S_X^2}{S_Y^2} = \frac{0.09}{0.14} = 0.64,$$

F_{12, 10 0.025} = 1 / F_{10, 12, 0.975} = 1 / 3.37

and we cannot reject $\mathcal{H}_0$ at the 5% level.

⋆ To calculate $F_{12,10,0.025}$ one may use the identity $F_{\nu_1,\nu_2,q} = \dfrac{1}{F_{\nu_2,\nu_1,1-q}}$
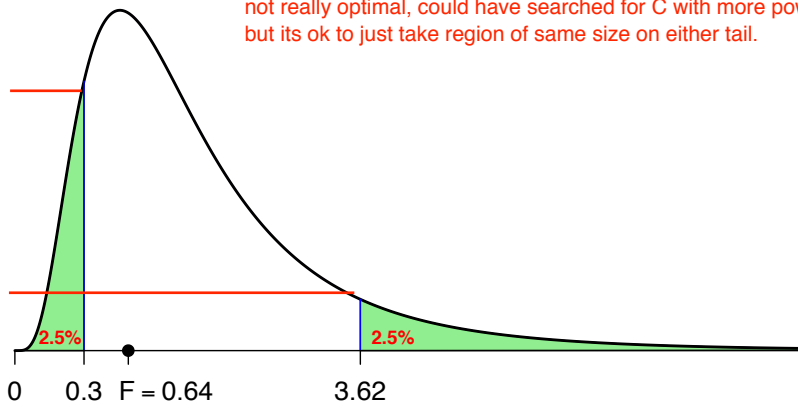
Example (cont.)

not really optimal, could have searched for C with more power
but its ok to just take region of same size on either tail.

2.5%

2.5%

0   0.3   F = 0.64   3.62

Indep. samples t test for the equality of means

- Suppose that we have two independent samples, $X_1, \ldots, X_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$ again (note that this time we assume that the two populations have the same variance), and we wish to test $\mathcal{H}_0 : \mu_X = \mu_Y$ vs. $\mathcal{H}_0 : \mu_X \neq \mu_Y$. pooled variance

- The likelihood this time is given by

$$\mathcal{L}(\mu_X, \mu_Y, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{m+n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{m}(X_i - \mu_X)^2 + \sum_{i=1}^{n}(Y_i - \mu_Y)^2 \right] \right\}$$

- We won't repeat all the tedious algebra again, but eventually

$$\Lambda \propto \left\{ 1 + \frac{\mathcal{T}^2}{m+n-2} \right\},$$

where

$$\mathcal{T} = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad \text{and} \quad S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}.$$

UNIVERSITY OF TORONTO

Testing for the Equality of Means (cont.)

$$\Lambda \propto \left\{ 1 + \frac{\mathcal{T}^2}{m + n - 2} \right\},$$

where

$$\mathcal{T} = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad \text{and} \quad S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m + n - 2}.$$

- What is the distribution of $\mathcal{T}$ under $\mathcal{H}_0 : \mu_X = \mu_Y$? Because the samples are independent, we have

variance increased here

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_X, \frac{\sigma^2}{m}\right) - \mathcal{N}\left(0, \frac{\sigma^2}{n}\right) = \mathcal{N}\left(\mu_X - \mu_Y, \sigma^2 \left[\frac{1}{m} + \frac{1}{n}\right]\right)$$

$$\overset{\mathcal{H}_0}{=\!=\!=} \mathcal{N}\left(0, \sigma^2 \left[\frac{1}{m} + \frac{1}{n}\right]\right) \quad \text{mean is zero since assume null true}$$

and

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} = \frac{(m-1)S_X^2}{\sigma^2} + \frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2_{m-1} + \chi^2_{n-1} = \chi^2_{m+n-2}.$$

now we determined null distribution for test statistic

Testing for the Equality of Means (cont.)

$$\mathcal{T} = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}$$

$$= \frac{\left(\overline{X} - \overline{Y}\right)\bigg/ \sigma\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}}{\sqrt{\dfrac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2(m+n-2)}}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\dfrac{\chi_{m+n-2}^2}{m+n-2}}} = t_{m+n-2},$$

hence the rejection region for this test is

$$\mathcal{C} = \left\{\, |\mathcal{T}| \geq t_{m+n-2,1-\alpha/2} \right\}.$$

⋆ Similarly, the rejection region for the right-tailed test $\mathcal{H}_0 : \mu_X - \mu_Y \leq 0$ vs.
$\mathcal{H}_0 : \mu_X - \mu_Y > 0$ would be   no need for left tailed test since could just swap null and alternative hypothesis

$$\mathcal{C} = \left\{ \mathcal{T} \geq t_{m+n-2,1-\alpha} \right\}.$$

⋆ S_p^2 is often referred to as the *pooled sample variance* – an unbiased estimator of σ^2 whose variance is smaller than each of S_X^2 and S_Y^2 separately.

Example

test if two population has different mean under normal distribution

We would like to test (at the 5% level, say) whether or not the mean weight of babies born to mothers who smoked during their pregnancy differs from the mean weight of babies whose mothers did not smoke. The following table contains the birth weights (in Kg) of 13 babies born to smoking mothers and 11 babies born to non-smokers –

	Non-smokers	Smokers
	3.99	3.18
	3.79	2.84
	3.60	2.90
	3.73	3.27
	3.21	3.85
	3.60	3.52
	4.08	3.23
	3.61	2.76
	3.83	3.60
	3.31	3.75
	4.13	3.59
	3.26	
	3.54	
Mean	3.67	3.32
Variance	0.09	0.14

Example (cont.)

```
> x <- c(3.99, 3.79, 3.6, 3.73, 3.21, 3.6, 4.08, 3.61, 3.83, 3.31, 4.13, 3.26, 3.54)
> y <- c(3.18, 2.84, 2.9, 3.27, 3.85, 3.52, 3.23, 2.76, 3.6, 3.75, 3.59)
>
> Weight <- c(x,y)
> Smoke <- c(rep("N", length(x)), rep("Y", length(y)))
> boxplot(Weight ~ Smoke, xlab = '', ylab = 'Birth Weight (Kg)')
```
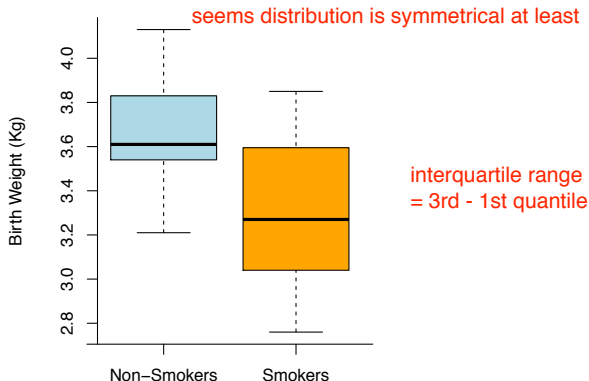
seems distribution is symmetrical at least

so have to confirm
1. normality of population
2. equality of variance
is not entirely un-plausible

interquartile range
= 3rd - 1st quantile

Example (cont.)

$$\begin{cases} \mathcal{H}_0 : \mu_X = \mu_Y \\ \\ \mathcal{H}_1 : \mu_X \neq \mu_Y \end{cases}$$

$$\overline{X} = 3.67, \ S_X^2 = 0.09, \ \overline{Y} = 3.32, \ S_Y^2 = 0.14$$

- First, we need to ask ourselves if the modelling assumptions are reasonable:

  ⋆ The distribution within each group seems symmetrical (i.e. the Normality assumption is not unreasonable)

  ⋆ We have already tested for the equality of the variances of the two groups

- The rejection region for this example is

$$\text{m+n-2 = 13 + 11 -2 = 22}$$

$$\mathcal{C} = \left\{ |\mathcal{T}| \geq t_{m+n-2, 1-\alpha/2} \right\} = \left\{ |\mathcal{T}| \geq \underbrace{t_{22, 0.975}}_{2.074} \right\}$$

and the pooled variance is

$$S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} = \frac{12 \times 0.09 + 10 \times 0.14}{13 + 11 - 2} = 0.113.$$

Example (cont.)

$$\mathcal{C} = \big\{\, |\mathcal{T}| \geq 2.074 \big\}, \;\; S_p^2 = 0.113$$

- The test statistic here takes the value

$$\frac{\overline{X} - \overline{Y}}{S_p\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} = \frac{3.67 - 3.32}{\sqrt{0.013} \times \sqrt{\dfrac{1}{13} + \dfrac{1}{11}}} = 2.54,$$

  well inside the rejection region! Therefore at the 5% level we reject $\mathcal{H}_0$ (and conclude that the mean weights differ).

- The p-value should tell us how significant exactly the result is:

$$\text{p-value} = \mathbb{P}\left(\,|\mathcal{T}| \geq 2.54 \,\Big|\, \mu_X = \mu_Y\right) = 2\mathbb{P}\left(\mathcal{T} \geq 2.54 \,\Big|\, \mu_X = \mu_Y\right)$$

  t distribution is symmetric hence

- Table (22 degrees of freedom):

$$0.99 < \mathbb{P}\left(\mathcal{T} \leq 2.54 \,\Big|\, \mu_X = \mu_Y\right) < 0.995 \implies 0.01 < \text{p-value} < 0.02$$

- Exact:

```
> 2*(1-pt(2.54, df=22))
[1] 0.01865869
```

Interpreting the results

- The difference between the mean birth weights in the two populations has been shown to be *statistically significant* (with p-value < 0.02)

- If we performed a right-tailed test ($\mathcal{H}_1 : \mu_X - \mu_Y > 0$) instead, the results would be even more overwhelming (p-value < 0.01 – why?)

  easier to reject one-tailed test... since p-value is less...

- Can we now conclude that smoking $\implies$ lower birth weight?

- Careful...

- Unless we carry out a designed experiment in a controlled environment, allocating subjects randomly to the treatment (smokers) and the control groups to avoid research bias, we cannot infer *causation*.

- But obviously it is unethical to ask pregnant women to smoke for the sake of science... so do not expect such an experiment to take place.

Interpreting the results (cont.)

- As it is, the data is purely *observational – the* allocation to treatment groups (smokers/non-smokers) in not under the researcher's control.

- The risk in drawing far-reaching conclusions in this kind of setting is ignoring the possible existence of a *confounder*: a factor other than the 'treatment factor' (smoking), that is associated with both the treatment and the *response* (birth weight). smoking -> birth weight

- For example, smoking and malnutrition tend to go hand-in-hand.

- In a controlled environment, we could have dictated a dietary regime that would rule out nutrition as a possible cause of birth weight differences.

- Without controlling nutrition (for example), we cannot prove that the true diagram is not as follows –

Interpreting the results (cont.)
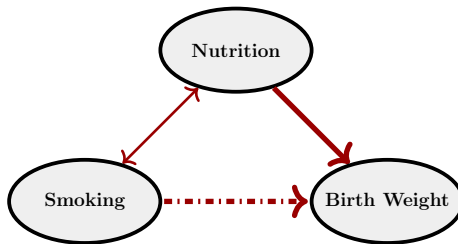
- In light of the above, the best we can say at this point is that smoking is *associated* with lower birth weights.

Confidence interval for the mean difference

- In general,

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} \sim t_{m+n-2}$$

- Doing the usual thing, we gather that

$$\mathbb{P}\Big(\overline{X} - \overline{Y} - S_p\, t_{m+n-2,1-\alpha/2} \sqrt{\frac{1}{m} + \frac{1}{n}}$$

$$\leq \mu_X - \mu_Y$$

$$\leq \overline{X} - \overline{Y} + S_p\, t_{m+n-2,1-\alpha/2} \sqrt{\frac{1}{m} + \frac{1}{n}}\Big) = 1 - \alpha,$$

hence

$$\overline{X} - \overline{Y} \pm S_p\, t_{m+n-2,1-\alpha/2} \sqrt{\frac{1}{m} + \frac{1}{n}}$$

forms a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$.

UNIVERSITY OF
TORONTO

Example

- In the birth weight example we had

$$\overline{X} = 3.67, \;\; \overline{Y} = 3.32, \;\; S_p^2 = 0.113, \;\; m = 13, \;\; n = 11$$

larger sample size hence restricts CI

- A 95% confidence interval for the mean weight difference would then be

$$\overline{X} - \overline{Y} \pm S_p \, t_{m+n-2,1-\alpha/2} \sqrt{\frac{1}{m} + \frac{1}{n}}$$

$$= 3.67 - 3.32 \pm \sqrt{0.113} \times \underbrace{t_{22,0.975}}_{2.074} \sqrt{\frac{1}{13} + \frac{1}{11}}$$

$$= 0.35 \pm 0.29 = [60 \text{ grams}, \, 640 \text{ grams}]$$

CI is based on alpha = 0.019; higher alpha => smaller interval

- Recall that we tested $\mathcal{H}_0 : \mu_X - \mu_Y = 0$ vs. $\mathcal{H}_1 : \mu_X - \mu_Y \neq 0$ and had p-value $= 0.019$

- Did you expect 0 to be contained in the 95% interval for the mean difference? Would you expect 0 to be contained in a 99% interval?

1. No, because we reject null at 5% level already
2. Note alpha smaller than p-value., hence   we cannot reject null <=> 0 is contained in the CI 39 / 45

Paired sample t test

- Under the right circumstances, researchers may opt for a *randomized paired design*: randomly selecting n paired experimental units from the reference population, within each pair assigning one of the pair to one treatment group at random (and the other to the other group), and measure the responses at the end of the assigned time.

restrict confounding variables

- "Experimental pairs" must consist of two highly correlated units –

  - the same person (blood pressure before and after treatment)

  - identical twins (with and without treatment)

  - two students with identical admission grades (one attended a public school and the other a private school)

    - etc.

Generalized Likelihood Ratio Tests   Examples of GLRTs   **Comparing Two Population Means**
○○○○○○○○○○   ○○○○○○○○○○○○
○○○○○○○○   ○●○○○○

Paired sample t test (cont.)

- The rationale behind paired samples:

  - If the two samples are independent (as before), for each pair (X_i, Y_i) we have

    variance increased when X and Y independent.

    $$\text{Var}(X_i - Y_i) = \text{Var}(X_i) + \text{Var}(Y_i)$$

  - If, on the other hand, the two samples are paired, we have

    $$\text{Var}(X_i - Y_i) = \text{Var}(X_i) + \text{Var}(Y_i) - 2\underbrace{\text{Cov}(X_i, Y_i)}_{>>0} << \text{Var}(X_i) + \text{Var}(Y_i),$$

    variance of difference decreases when X and Y dependent

- The pair-to-pair variability can be reduced significantly relative to the independent samples setup   reduction of noise

- We should thus expect to detect differences between the group means more easily $\implies$ increase in power!

Paired sample t test (cont.)

- Our aim is to test hypotheses on the mean-difference between the treatment groups' responses based on a paired-sample design –

$$\begin{cases} \mathcal{H}_0 : \mu_D = 0 \\ \\ \mathcal{H}_1 : \mu_D > 0, \neq 0, < 0 \end{cases}$$

  where $D := X - Y$ and $\mu_D := \mu_X - \mu_Y$.

- We make the following assumptions:
  1. Different pairs are independent

  2. The difference $X - Y$ follows a Normal distribution

- then the differences sample $\{D_i = X_i - Y_i\}$ can be calculated, and the problem reduces to the one sample t test, based on the test statistic

this form of hypothesis is same as one sample t test

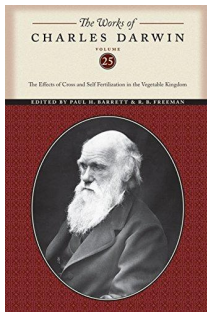$$\mathcal{T} = \frac{D}{S_D / \sqrt{n}},$$

  whose distribution under $\mathcal{H}_0 : \mu_D = 0$ is t_{n-1}.

Example

Darwin (1809-1882) performed investigations of the effects of cross- and self-fertilization on plants. His hypothesis was that cross-fertilization produces more robust and vigorous individuals. In one of his experiments, Darwin paired 15 cross-fertilized and 15 self-fertilized seedlings (to control different soil, watering and light conditions), and let the pairs grow in different pots. After a number of weeks, the stalk height of each plant was measured.

Example (cont.)

D = Cross - Self

1. independence: plant each pair far away from each other

2. normality: use q-q plot for example, hard to determine for small sample size of 15

this is sample mean and sample variance (1/n-1 remember)

Pair #	Cross (cm)	Self (cm)	D (cm)
1	59.7	44.2	15.5
2	30.5	51.8	−21.3
3	53.3	50.8	2.5
4	55.9	50.8	5.1
5	48.5	46.7	1.8
6	54.6	47.2	7.4
7	56.1	47.2	8.9
8	51.8	38.9	12.9
9	46.5	41.9	4.6
10	54.9	45.7	9.2
11	59.2	41.4	17.8
12	53.3	45.7	7.6
13	56.1	32.5	23.6
14	58.4	39.4	19.0
15	30.5	45.7	−15.2
Mean		51.3	44.7	6.6
Variance		84.2	26.7	142.9

note variance in D is even higher than from Cross and Self. Didnt do a job in reducing variance.

Example (cont.)

- Darwin had a clear research question in mind: does cross-fertilized lead to increased stalk heights? Hence one-tailed test

$$\mathcal{H}_0 : \mu_D \leq 0 \ \text{ vs. } \ \underline{\mathcal{H}_1 : \mu_D > 0} \ \ (D = \text{Cross} - \text{Self})$$

- Here

$$\mathcal{T} = \frac{\overline{D}}{S_D/\sqrt{n}} = \frac{6.6}{\sqrt{142.9/15}} = 2.15,$$

df = 15 - 1 = 14

and recalling that under $\mathcal{H}_0 : \mu_D = 0$ the distribution of $\mathcal{T}$ is $\underline{t_{14}}$, the table tells us that

$$0.01 < \text{p-value} = \mathbb{P}\left(\mathcal{T} \geq 2.15 \Big| \mu_D = 0\right) < 0.025,$$

or exactly–

```
> 1-pt(2.15, df=14)
```
so reject at significance level of 0.025 but not at 0.01
```
[1] 0.0247585
```

★ Note that the the p-value for a two-tailed test would be 0.0495 – right inside the 5% mark.