STA 250/MTH 342 – Intro to Mathematical Statistics

Lecture 17

Testing composite hypotheses

- Let us move onto the general problem of testing

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

  where Θ_0 and Θ_1 are collections of possible values for θ.
- This includes the previous two scenarios: (1) testing simple hypotheses and (2) testing simple vs composite hypothesis as special cases.

- Most typically there does not exist a single test that performs uniformly the best in the Neyman-Pearson sense against all alternatives in Θ_1. Typically no UMP tests exist.
- So we cannot hope to find a "best" test.
- Our goal now is to find a test that satisfies some restrictions on the Type I error rate α, while perform *reasonably well* in terms of power against all alternatives in Θ_1.

Generalized likelihood ratio test

- The idea is similar to a tournament between two conferences or divisions.
- Imagine Θ_0 and Θ_1 being the Atlantic division and the Coastal division in the ACC.
- Each division chooses its own champion, and then the two champions settle the matter of the ACC championship.
- For example, we may let Θ_0 and Θ_1 each choose a *restricted* maximum likelihood estimate

$$\hat{\theta}_0 = \max_{\theta \in \Theta_0} L(\theta) \quad \text{and} \quad \hat{\theta}_1 = \max_{\theta \in \Theta_1} L(\theta).$$

- Then we settle the competition between Θ_0 and Θ_1 according to a "game" between the two champions—we check whether or not

$$\frac{L(\hat{\theta}_1)}{L(\hat{\theta}_0)} = \frac{\max_{\theta \in \Theta_1} L(\theta)}{\max_{\theta \in \Theta_0} L(\theta)} > K.$$

- This is essentially what we do, but in many applications we may have a problem—$\max_{\theta \in \Theta_1} L(\theta)$ does not exist.
- For this reason let us look at another test statistic which is closely related.
- First let $\Theta = \Theta_0 \cup \Theta_1$. Then we would evaluate the competition in terms of

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}.$$

  This is called the *generalized LR statistic*. Note that $\Lambda \leq 1$.
- Two possibilities:
  1. The unrestricted MLE $\hat{\theta}$ falls on a $\theta \in \Theta_0$: $\Lambda = 1$.
  2. The unrestricted MLE $\hat{\theta}$ falls on a $\theta \in \Theta_1$: $0 < \Lambda < 1$.
- For restrictive Θ_0 (such as a simple null hypothesis, $\theta = 1.0$), no matter whether Θ_0 is true or not, $\hat{\theta}$ will typically lie outside of Θ_0 (why?), and we have $0 < \Lambda < 1$.

- The question now is whether Λ is *small* enough for us to reject H_0.
- This suggests that a *general likelihood ratio test* takes the following form:

$$\text{Reject } H_0 \text{ when } \Lambda < K.$$

- Similar to before, we use pre-specified Type I error rate α to choose the constant K so that

$$P(\Lambda < K|\theta) \leq \alpha \quad \textit{for all } \theta \in \Theta_0.$$

- For simple hypotheses,

$$\Theta_0 = \{\theta_0\} \quad \text{and} \quad \Theta_1 = \{\theta_1\},$$

we have

$$\Lambda = \frac{L(\theta_0)}{\max\{L(\theta_0), L(\theta_1)\}} = \min\left\{1, \frac{L(\theta_0)}{L(\theta_1)}\right\}.$$

- Thus Λ is monotone decreasing in $L(\theta_1)/L(\theta_0)$. In particular

$$\Lambda < K < 1 \quad \text{if and only if} \quad \frac{L(\theta_1)}{L(\theta_0)} > \frac{1}{K} > 1.$$

- So the generalized likelihood ratio test is equivalent to the LR (or Neyman-Pearson) test.
- From now on we shall refer to both the generalized LR test and the Neyman-Pearson test as the *LR test*, and let the context determine which one we mean.

Example: Testing normal mean with known variance

- Our data are i.i.d. observations $X_1, X_2, \ldots, X_n$ from $\mathrm{N}(\mu, \sigma^2)$.
- We want to test

$$H_0 : \theta = (\mu_0, \sigma_0^2) \text{ vs } H_1 : \theta = (\mu_1, \sigma_0^2) \text{ for any } \mu_1 \neq \mu_0.$$

- Sometimes people write this as testing

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu \neq \mu_0$$

  with known $\sigma^2 = \sigma_0^2$.
- This is called a *two-sided* alternative.

Let's carry out the (generalized) LR test

- Since the unrestricted MLE for θ is $\hat{\theta} = (\hat{\mu}, \sigma_0^2) = (\bar{X}, \sigma_0^2)$, we have

$$
\begin{aligned}
\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} \\
&= e^{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n (X_i - \mu_0)^2 - \sum_{i=1}^n (X_i - \bar{X})^2\right]} \\
&= e^{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n X_i^2 - 2\mu_0 \sum_{i=1}^n X_i + n\mu_0^2 - \sum_{i=1}^n X_i^2 + 2\bar{X}\sum_{i=1}^n X_i - n\bar{X}^2\right]} \\
&= e^{-\frac{1}{2\sigma_0^2}[-2n\mu_0\bar{X} + n\mu_0^2 + 2n\bar{X} - n\bar{X}^2]} \\
&= e^{-\frac{n}{2\sigma_0^2}[\bar{X}^2 - 2\mu_0\bar{X} + \mu_0^2]} \\
&= e^{-\frac{n}{2\sigma_0^2}(\bar{X} - \mu_0)^2}.
\end{aligned}
$$

- So $\Lambda < K$ is equivalent to $(\bar{X} - \mu_0)^2 > C$ or $|\bar{X} - \mu_0| > C'$. The "compromise" test we have guessed earlier!
- We have found that to have a level α test, $C' = \Phi^{-1}(1 - \frac{\alpha}{2}) \cdot \frac{\sigma_0}{\sqrt{n}}$.

- Here no UMP test exists, and the LR test provides a reasonable solution. This is quite generally the case.
- Also, here we used the LR formulation to deduce a simpler form of the test, namely reject when $|\bar{X} - \mu_0| > C'$.
- When the sample size is large, one can often carry out the LR test directly in terms of Λ, due to some nice theoretical properties about the sampling distribution of Λ. (Later!)
- In particular, we can derive an *approximate* sampling distribution of the (generalized) LR statistic under the null distribution when n is large! We will cover this later.

Example: Testing normal mean with unknown variance.

- Again we have i.i.d. data $X_1, X_2, \ldots, X_n$ from $\mathrm{N}(\mu, \sigma^2)$, but now the variance σ^2 is unknown.
- Let us again consider the problem of testing

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

- That is testing

$$\Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, 0 < \sigma^2 < \infty\}$$

vs

$$\Theta_1 = \{(\mu, \sigma^2) : \mu \neq \mu_0, 0 < \sigma^2 < \infty\}.$$

- So

$$\Theta = \Theta_0 \cup \Theta_1 = \{(\mu, \sigma^2) : -\infty < \mu < \infty, 0 < \sigma^2 < \infty\}$$

What is the LR test?

- The LR statistic is

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \frac{L(\mu_0, \hat{\sigma}_0)}{L(\hat{\mu}, \hat{\sigma})}.$$

- Here (exercise!) the restricted MLE is

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{n}$$

and the unrestricted MLE

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}.$$

- Since

$$L(\mu_0, \hat{\sigma}_0) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} e^{-\frac{\sum_{i=1}^n (X_i-\mu_0)^2}{2\hat{\sigma}_0^2}} = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} e^{-n/2},$$

  and

$$L(\hat{\mu}, \hat{\sigma}) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-\frac{\sum_{i=1}^n (X_i-\hat{\mu})^2}{2\hat{\sigma}^2}} = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-n/2},$$

  we have

$$\Lambda = \frac{L(\mu_0, \hat{\sigma}_0)}{L(\hat{\mu}, \hat{\sigma})} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}.$$

- Therefore the LR test rejects when

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} < K \quad \text{or equivalently, when} \quad \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} > K'.$$

- What is the sampling distribution of this test statistic?

- Let us take a closer look at this test

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

- But the numerator

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

- Thus

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = 1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} = 1 + \frac{1}{n-1} \cdot \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2/(n-1)}.$$

- So the LR test rejects when

$$\frac{n(\bar{X}-\mu_0)^2}{\sum_{i=1}^n (X_i-\bar{X})^2/(n-1)} > C.$$

or rejects when

$$\frac{|\sqrt{n}(\bar{X}-\mu_0)|}{s} > C.$$

- This is called the *one-sample (symmetric) t-test*, and

$$T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{s}$$

is called a *t-statistic*.
- An intuitive explanation of this statistic.
- Question: What is the sampling distribution of T under H_0?

Sampling distribution of T under H_0

- Let

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad \text{and} \quad W = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}.$$

  Then under H_0,

$$Z \sim \mathrm{N}(0,1) \quad \text{and} \quad W \sim \chi^2_{n-1}$$

  and Z and W are *independent*.

- Note that

$$T = \frac{Z}{\sqrt{W/(n-1)}} \sim \frac{\mathrm{N}(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}}.$$

- This sampling distribution is called the *t-distribution* with $n-1$ degrees of freedom.

Choosing the constant C

- Suppose we want to find a level α test, then

$$\begin{aligned}\alpha &= P(|T| > C | H_0) \\ &= 2P(T > C | H_0) \\ &= 2\left(1 - F_{t_{n-1}}(C)\right).\end{aligned}$$

- Thus

$$F_{t_{n-1}}(C) = 1 - \frac{\alpha}{2},$$

  or

$$C = F_{t_{n-1}}^{-1}\left(1 - \frac{\alpha}{2}\right).$$

- So the level α LR test rejects when

$$|T| > F_{t_{n-1}}^{-1}\left(1 - \frac{\alpha}{2}\right).$$

t tests

- In general tests based on the t-statistic

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s}$$

  are called t tests.

- One may use t test for testing one-sided hypotheses as well, e.g.

$$H_0 : \mu \leq \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0.$$

- Intuitively, a reasonable test in this case will be rejecting when

$$T > C.$$

- The proof is similar to the proof for the two-sided case. (See Textbook Example 9.5.12.)

Sampling distribution of T under H_1 (Optional)

- In order to compute the power function of the t-test, we need to know the sampling distribution of T under the alternative when $\mu = \mu_1 \neq \mu_0$.
- Note that when $\mu = \mu_1$,

$$Z' = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}, 1\right).$$

  We still have $W \sim \chi^2_{n-1}$ and is independent of Z'.
- What is the sampling distribution of $T = Z'/\sqrt{W/(n-1)}$?
- Its sampling distribution is called the *non-central t-distribution* with $n-1$ degrees of freedom and non-centrality parameter $\psi = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$.

The non-central t distribution (Optional)

- More generally, the *noncentral t distribution* with k degrees of freedom and noncentrality parameter ψ is defined to be the sampling distribution of

$$\frac{Z+\psi}{\sqrt{W/k}}$$

  where Z and W are independent with $Z \sim \mathrm{N}(0,1)$ and $W \sim \chi^2_k$.
- Note that when $\psi = 0$, this is the (central) t distribution with k degrees of freedom.

The power function of t-test (Optional)

- The power function is given by

$$\begin{aligned}\pi(\mu_1,\sigma) &= P\left(|T| > C|\mu_1,\sigma\right) \\ &= F_{t_{n-1}(\psi)}(-C) + 1 - F_{t_{n-1}(\psi)}(C)\end{aligned}$$

where $\psi = \frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}$ and $C = F^{-1}_{t_{n-1}}\left(1-\frac{\alpha}{2}\right)$.

Comparing the mean of two normal samples

- Now let our data be i.i.d. observations $X_1, X_2, \ldots, X_n$ from $\mathrm{N}(\mu_1, \sigma^2)$, and i.i.d. observations $Y_1, Y_2, \ldots, Y_m$ from $\mathrm{N}(\mu_2, \sigma^2)$.
- The X's and the Y's are *independent* of each other.
- For example, the X's may be samples of SAT scores from High School A, and the Y's are those from High School B.
- Note that we assume the two distributions may have different means μ_1 and μ_2, but with the *same* variance σ^2.
- All three parameters $\theta = (\mu_1, \mu_2, \sigma^2)$ are assumed to be unknown.

- We are interested in testing a null hypothesis that compares the two sample means. For example,

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2.$$

or testing

$$H_0 : \mu_1 \leq \mu_2 \quad \text{vs} \quad H_1 : \mu_1 > \mu_2.$$

- What may be a good test for this purpose?
- This still falls into the general setting of testing two composite hypotheses.

- For example let's consider testing

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2.$$

  Let's find the LR test. This is very similar to the problem of testing the mean for a single normal sample.
- The joint likelihood is

$$L(\mu_1, \mu_2, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{(n+m)/2}} e^{-\frac{\sum_{i=1}^n (X_i - \mu_1)^2 + \sum_{j=1}^m (Y_j - \mu_2)^2}{2\sigma^2}}.$$

- We can solve for the global (i.e. unrestricted) MLE $(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2)$

$$\hat{\mu}_1 = \bar{X} \qquad \hat{\mu}_2 = \bar{Y}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{n+m} = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m}.$$

- Under the null, we have $\mu_1 = \mu_2 = \mu$, and we can again solve for the restricted MLE $(\hat{\mu}, \hat{\sigma}_0^2)$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j}{n+m}$$

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{n} (X_i - \hat{\mu})^2 + \sum_{j=1}^{m} (Y_j - \hat{\mu})^2}{n+m}.$$

- So the generalized LR is

$$\Lambda = \frac{L(\hat{\mu}, \hat{\mu}, \hat{\sigma}_0^2)}{L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2)} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n+m}{2}},$$

and we reject when $\hat{\sigma}^2 / \hat{\sigma}_0^2 < K$ or equivalently when $\hat{\sigma}_0^2 / \hat{\sigma}^2 > K'$.

- We have

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum_{i=1}^n (X_i - \hat{\mu})^2 + \sum_{j=1}^m (Y_j - \hat{\mu})^2}{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}$$

$$= 1 + \frac{n(\bar{X} - \hat{\mu})^2 + m(\bar{Y} - \hat{\mu})^2}{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}$$

  because $\sum_{i=1}^n (X_i - \hat{\mu})^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \hat{\mu})^2$.
- Check that

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = 1 + \frac{1}{n+m-2} \cdot T^2$$

  where

$$T = \frac{\bar{X} - \bar{Y}}{s_{pooled} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

  with

$$s_{pooled}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{m+n-2} = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}.$$

- An intuitive explanation of this statistic.

- So we can reject when $T^2 > C$ or equivalently when $|T| > C'$.
- Again, to find C', we need to know the sampling distribution of T under H_0.
- Since

$$W = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{\sigma^2}$$

is a χ^2_{n+m-2} distribution (*why?*) and it is independent (*why?*) of

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}},$$

which is N(0,1) under the null.

- Therefore
$$T = \frac{Z}{\sqrt{W/(n+m-2)}}$$
has a t_{n+m-2} distribution under H_0.
- So the cutoff constant C' to make the test have level α is
$$C' = F^{-1}_{t_{n+m-2}}(1-\alpha/2)$$
- This test is called a (two-sided) *two-sample* t-test. (Draw a figure.)

- For example, suppose $n = 10$, $m = 12$, and we get $\bar{X} = 2.2$, $\bar{Y} = 3.1$, and $s^2_{pooled} = .82$, then

$$T = \frac{2.2 - 3.1}{\sqrt{.82 \cdot \left(\frac{1}{10} + \frac{1}{12}\right)}} = -2.32.$$

  So we will reject H_0 at the .05 level but not at the .01 level.

- Similarly, we can derive the corresponding one-sided two-sample t-test for testing

$$H_0 : \mu_1 \leq \mu_2 \quad \text{vs} \quad H_1 : \mu_1 > \mu_2.$$

  which rejects when $T > C''$.

- The above example assumes that the two samples come from normal distributions with *the same* variance.
- What if the two distributions are $\mathrm{N}(\mu_1, \sigma_1^2)$ and $\mathrm{N}(\mu_2, \sigma_2^2)$, where the two variances are unknown and can be different?
- This seemingly simple extension makes finding the *exact* sampling distribution for the LR test extremely difficult.
- In fact it's still an *open* problem, called the ***Behrens-Fisher*** problem.

- Welch proposes to use the following test statistic that imitates (but is not!) a t-statistic

$$T_w = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

- Note that the denominator is a natural estimator for the standard devaition of the numerator

$$\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}.$$

- What is the sampling distribution of T_w under H_0?

- The *exact* sampling distribution of T_w under H_0 (that is $\mu_1 = \mu_2$) is *unknown* but can be approximated by a t distribution with the following degrees of freedom:

$$\mathrm{df} = \frac{\left(s_X^2/n + s_Y^2/m\right)^2}{\frac{(s_X^2/n)^2}{n-1} + \frac{(s_Y^2/m)^2}{m-1}}.$$

- The test based on T_w is called *Welch's approximate t-test*.

- For example, suppose $n = 10$, $m = 12$, and we get $\bar{X} = 2.2$, $\bar{Y} = 3.1$, and $s_X^2 = .75$, $s_Y^2 = .94$, then

$$T_w = \frac{2.2 - 3.1}{\sqrt{\frac{.75}{10} + \frac{.94}{12}}} = -2.30.$$

- The d.f. for the approximate t distribution is

$$\text{df} = \frac{(.75/10 + .94/12)^2}{\frac{(.75/10)^2}{10-1} + \frac{(.94/12)^2}{12-1}} = 19.9$$

- Do we reject the null hypothesis? What is the 0.05 level quantile of $t_{19.9}$?

Comparing paired samples

- Up to this point we have been considering two *independent* samples from normal distributions.
- Sometimes the two samples that we want to compare are not independent of each other.
- For example, $X_1, X_2, \ldots, X_n$ may be the number of hours of sleep that n individuals get on Day 1.
- After applying some treatment for insomnia to these *same* n individuals, we measure the hours of sleep they get on Day 2—$Y_1, Y_2, \ldots, Y_n$.
- So (X_i, Y_i) are two measurements taken on the same individual i.
- Two samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ collected this way are called *paired* samples.
- A better way of writing the data may be $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

- We are still interested in comparing the mean of the two samples. For example, we want to test

$$H_0 : \mu_1 = \mu_2 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2.$$

- Note that now the observations in the two samples are *dependent*, so the earlier two-sample t-test cannot be directly applied.

- Question: Can we still apply the previous two-sample t-test?

How about a one-sample t-test?

- A useful strategy to handle this problem is to take the difference between the paired observations

$$U_i = X_i - Y_i.$$

- Now $U_1, U_2, \ldots, U_n$ form a single sample of *i.i.d.* observations, and we basically want to test

$$H_0 : \mu_U = 0 \quad \text{vs} \quad H_1 : \mu_U \neq 0.$$

Example: Bivariate normal data

- Suppose the paired data (X_i, Y_i) are i.i.d. bivariate normal random vectors with mean

$$E(X_i) = \mu_1 \quad E(Y_i) = \mu_2,$$

variance

$$\mathrm{Var}(X_i) = \sigma_1^2 \quad \mathrm{Var}(Y_i) = \sigma_2^2$$

and correlation

$$\mathrm{corr}(X_i, Y_i) = \rho.$$

- So the joint pdf of (X_i, Y_i) is

$$f(x,y) = \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_1\sigma_2} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]}.$$

(We don't need this pdf here.)

- Question: Which test should we use?

- All we need to know about the bivariate normal distribution here is that the different

$$U_i = X_i - Y_i$$

are i.i.d. normal random variables with some mean μ_U and some variance σ_U^2.
- Now testing $H_0 : \mu_U = 0$ reduces to the simple case of testing the mean of a normal distribution with unknown variance.
- So we can apply the *one-sample* t-test treating the U_i's as our data—reject when $|T_U| > C$ where

$$T = \frac{\sqrt{n}(\bar{U} - \mu_{U,0})}{s_U}.$$

For the current example, $\mu_{U,0} = 0$ because under H_0, $\mu_1 - \mu_2 = 0$.
- We can of course choose our null hypothesis to be

$$H_0 : \mu_1 = \mu_2 + 5$$

for example, in which case $\mu_{U,0} = 5$.

- Which design is better—paired or independent samples?
- What are the factors that affect the power of the test?

What if the one sample or two sample data are not Gaussian?

- Can we still apply the *t*-tests to compare their means?

A quick sum up

- We have learned the (generalized) LR test for testing composite null and alternative hypotheses.
- We have seen an example on testing the mean of a normal distribution with *unknown* variance, in which case the (generalized) LR test is equivalent to the t-test.
- In all of the examples we have seen, we use the general form of the LR test to derive specific forms in terms of test statistics (e.g. the t-statistic). Then we choose the "constant" C in these tests based on the *exact* sampling distribution of the test statistic under the null.

Next . . .

- For many problems, one cannot easily find the exact sampling distribution of the test statistics.
- Fortunately, when the sample size is large, one can find the approximate sampling distribution of the generalized likelihood ratio statistic Λ under the null.
- This allows us to specify the test for a given level α in terms of Λ directly.

$$\mathscr{R}(\alpha) = \{x : \Lambda < C\}$$

  such that $P(\Lambda < C | H_0) \approx \alpha$.