Lecture 7 — Composite hypotheses and the t-test

7.1 Composite null and alternative hypotheses

This week we will discuss various hypothesis testing problems involving a composite null hypothesis and a compositive alternative hypothesis. To motivate the discussion, consider the following examples:

Example 7.1. There are 80 students in a STATS 200 class. A diagnostic exam is administered at the start of the quarter, and a comparable exam is administered at the end of the quarter. Did STATS 200 improve students' knowledge of statistics?

Let X_i be the difference in test scores for student i. There are various ways we can formulate the above question as a hypothesis test: If we believe a normal model for the X_i's, $X_1, \ldots, X_{80} \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, then we might formulate our question as the testing problem

$$H_0 : \mu = 0$$
$$H_1 : \mu > 0$$

Note that both the null and alternative hypotheses above are composite, because they do not specify the variance σ^2 (which is unknown). If we are not willing to make a normality assumption, we might assume instead that $X_1, \ldots, X_{80}$ are IID with PDF f, and test

$$H_0 : f \text{ is symmetric around } 0$$
$$H_1 : f \text{ is symmetric around } \mu \text{ for some } \mu > 0$$

or maybe even drop the symmetry assumptions and test

$$H_0 : f \text{ has median } 0$$
$$H_1 : f \text{ has median } \mu \text{ for some } \mu > 0$$

Which formulation we choose and the resulting test statistic we use may depend on our prior knowledge of how test scores are typically distributed and on visual inspection of the data (for departures from normality, symmetry, etc.)

Example 7.2. A friend criticizes the setup of the previous example: It's hard to make two exams that are equally difficult. What if the second exam just happened to be a bit easier?

To address this criticism, we add a control group: We give 100 other students (who are not taking statistics courses this quarter) the same two exams at the start and end of the quarter. Let Y_i be the difference in test scores for student i of this control group. Again, if we believe a normal model $X_1, \ldots, X_{80} \stackrel{IID}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_{100} \stackrel{IID}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$ (with the X's also independent of the Y's), then we might formulate the test as

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X > \mu_Y$$

If we are not willing to assume normality, we might suppose instead that $X_1, \ldots, X_{80}$ are IID with PDF f and $Y_1, \ldots, Y_{100}$ are IID with PDF g, and test

$$H_0 : f = g$$
$$H_1 : f \text{ stochastically dominates } g$$

(This alternative H_1 means that if $X \sim f$ and $Y \sim g$, then $\mathbb{P}[X \geq x] \geq \mathbb{P}[Y \geq x]$ for all $x \in \mathbb{R}$.) Again, how we formulate the testing problem depends on the modeling assumptions we are willing to make.

When testing a composite null hypothesis H_0 against a compositive alternative H_1, there is a probability of type I error associated to each data distribution $P \in H_0$ (the probability of rejecting H_0 if the true distribution were P) and a probability of type II error associated to each data distribution $P \in H_1$ (the probability of accepting H_0 if the true distribution were P). A test has **significance level** α if the maximum probability of type I error for any $P \in H_0$ is α.

This means that to design a level-α test of H_0, we need to control the probability of type I error for every $P \in H_0$, and hence reason about the sampling distribution of our test statistic T under every such data distribution P. In general this can be very difficult, and a common simplifying strategy will be to find a test statistic T that has *the same* sampling distribution under every $P \in H_0$.

7.2 One-sample t-test

Assume $X_1, \ldots, X_n \stackrel{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$ for unknown μ and σ^2, and consider testing

$$H_0 : \mu = 0$$
$$H_1 : \mu > 0$$

If σ^2 were fixed and known, then the uniformly most-powerful level-α test would reject for large values of $\bar{X}$. Specifically, it would reject when $\frac{\sqrt{n}\bar{X}}{\sigma} > z(\alpha)$ (because when $X_1, \ldots, X_n \stackrel{IID}{\sim} \mathcal{N}(0, \sigma^2)$, $\bar{X} \sim \mathcal{N}(0, \frac{\sigma^2}{n})$ so $\frac{\sqrt{n}\bar{X}}{\sigma} \sim \mathcal{N}(0, 1)$).

When σ^2 is unknown, a natural idea is to estimate σ^2 by the sample variance

$$S^2 = \frac{1}{n-1}\left((X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2\right),$$

and to consider the test statistic

$$T = \frac{\sqrt{n}\bar{X}}{S}.$$

To derive the distribution of T under H_0, we first prove the following result:

Theorem 7.3. *Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and let $\bar{X}$ and S^2 be the sample mean and sample variance (where S^2 is defined as above). Then S^2 is independent of $\bar{X}$ and distributed as $S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$.*

Proof. Note that changing the mean μ does not affect the distribution of S^2 and shifts the distribution of $\bar{X}$ by a constant value, which does not affect independence of S^2 and $\bar{X}$. So we may assume without loss of generality $\mu = 0$.

We first show independence of S^2 and $\bar{X}$. The entries of

$$(\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X})$$

are linear combinations of $X_1, \ldots, X_n$, so $(\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X})$ has a multivariate normal distribution by Example 3.4 of Lecture 3. Let's compute

$$\mathrm{Cov}[\bar{X}, X_1 - \bar{X}] = \mathrm{Cov}[\bar{X}, X_1] - \mathrm{Cov}[\bar{X}, \bar{X}].$$

By bilinearity of covariance and the fact that $\mathrm{Cov}[X_j, X_1] = 0$ for all $j \geq 2$,

$$\mathrm{Cov}[\bar{X}, X_1] = \mathrm{Cov}\left[\frac{1}{n}\sum_{j=1}^n X_j, X_1\right] = \frac{1}{n}\sum_{j=1}^n \mathrm{Cov}[X_j, X_1] = \frac{1}{n}\mathrm{Cov}[X_1, X_1] = \frac{1}{n}\mathrm{Var}[X_1] = \frac{\sigma^2}{n}.$$

Since $\bar{X} \sim \mathcal{N}(0, \frac{\sigma^2}{n})$, $\mathrm{Cov}[\bar{X}, \bar{X}] = \mathrm{Var}[\bar{X}] = \frac{\sigma^2}{n}$ also. Then

$$\mathrm{Cov}[\bar{X}, X_1 - \bar{X}] = 0.$$

Similarly $\mathrm{Cov}[\bar{X}, X_i - \bar{X}] = 0$ for every $i = 2, \ldots, n$. By Theorem 3.5 from Lecture 3, this means $\bar{X}$ is independent of $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$, and so $\bar{X}$ is independent of S^2.

To compute the distribution of S^2, we may write

$$\begin{aligned}
(n-1)S^2 &= (X_1 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2 \\
&= (X_1^2 - 2X_1\bar{X} + \bar{X}^2) + \ldots + (X_n^2 - 2X_n\bar{X} + \bar{X}^2) \\
&= X_1^2 + \ldots + X_n^2 - 2(X_1 + \ldots + X_n)\bar{X} + n\bar{X}^2 \\
&= (X_1^2 + \ldots + X_n^2) - 2n\bar{X}^2 + n\bar{X}^2 \\
&= (X_1^2 + \ldots + X_n^2) - n\bar{X}^2.
\end{aligned}$$

Letting $U = (n-1)S^2/\sigma^2$, $W = (X_1^2 + \ldots + X_n^2)/\sigma^2$, and $V = n\bar{X}^2/\sigma^2$, this says $W = U + V$. We showed S^2 is independent of $\bar{X}$, hence U is independent of V. Thus the MGF of W is the product of the MGFs of U and V:

$$M_W(t) = M_U(t)M_V(t).$$

Finally, note that each $X_i/\sigma \sim \mathcal{N}(0,1)$, so $W \sim \chi^2_n$. Also, $\sqrt{n}\bar{X}/\sigma \sim \mathcal{N}(0,1)$, so $V = (\sqrt{n}\bar{X}/\sigma)^2 \sim \chi^2_1$. This means that the MGF of U is, for any $t < \frac{1}{2}$,

$$M_U(t) = \frac{M_W(t)}{M_V(t)} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-(n-1)/2},$$

which is the MGF of the χ^2_{n-1} distribution. So $U \sim \chi^2_{n-1}$, and $S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$. □

Remark 7.4. At the end of Lecture 4, we claimed if $W_1, \ldots, W_6 \overset{IID}{\sim} \mathcal{N}(0,1)$, then $(W_1 - \bar{W})^2 + \ldots + (W_6 - \bar{W})^2 \sim \chi_5^2$. The above theorem verifies this. The theorem also explains why we often define S^2 with the normalization $\frac{1}{n-1}$ rather than $\frac{1}{n}$: As the expectation of a χ_{n-1}^2 random variable is $n-1$, $\mathbb{E}[S^2] = \sigma^2$ so S^2 is an unbiased estimator for σ^2.

Returning to our test statistic

$$T = \frac{\sqrt{n}\bar{X}}{S} = \frac{\sqrt{n}\bar{X}/\sigma}{S/\sigma},$$

we observe that by Theorem 7.3, for $\mu = 0$ and any value of $\sigma^2 > 0$,

$$\frac{\sqrt{n}\bar{X}}{\sigma} \sim \mathcal{N}(0,1), \qquad \frac{S^2}{\sigma^2} \sim \frac{1}{n-1}\chi_{n-1}^2,$$

and these are independent. Hence the distribution of T does not depend on σ, so it is the same under all $P \in H_0$. We give this distribution a name:

Definition 7.5. If $Z \sim \mathcal{N}(0,1)$, $U \sim \chi_n^2$, and Z and U are independent, then the distribution of $Z/\sqrt{\frac{1}{n}U}$ is called the ***t* distribution with *n* degrees of freedom**, denoted t_n.

So under H_0,

$$T \sim t_{n-1}.$$

Letting $t_{n-1}(\alpha)$ denote the upper α point (or $1-\alpha$ quantile) of the distribution t_{n-1}, the test that rejects for $T > t_{n-1}(\alpha)$ is called the **one-sample *t*-test**.

Remark 7.6. The one-sample t-test is often used in paired two-sample settings, such as Example 7.1. There, we actually have two paired samples—the before and after test scores of each student—and we perform the test by first taking the differences of these paired values. In such settings, the test is often called the **paired two-sample *t*-test**, although the statistical procedure is really just a test for one set of IID observations.