7.1 Composite hypotheses and pivotal statistics

Last lecture, we discussed the Neyman-Pearson lemma for testing a simple null hypothesis $H_0$ versus a simple alternative $H_1$. Hypotheses that are not simple are called composite: These do not fully specify the distribution of the observed data, perhaps due to an unknown parameter or a nonparametric description of the distribution. Many testing problems arising in practice have composite hypotheses.

**Example 7.1.** There are $n = 250$ students in S&DS 242. A diagnostic exam is given at the start of the semester, and a comparable exam is given at the end of the semester. Based on the differences in students’ scores, did S&DS 242 improve their knowledge of statistics?

Let $X_i$ be the difference in exam scores for student $i$. There are various ways we might formulate the above question as a hypothesis test: If the observed data $X_1, \ldots, X_n$ look normally distributed, we may choose to do the test assuming a normal model,

\begin{align*}
H_0 & : X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(0, \sigma^2) \text{ for some } \sigma^2 > 0 \quad (7.1) \\
H_1 & : X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2) \text{ for some } \mu > 0 \text{ and } \sigma^2 > 0 \quad (7.2)
\end{align*}

Note that both the null and alternative hypotheses above are composite—neither specifies the value of the variance $\sigma^2$, which is unknown to us. We’ll see that a sensible testing approach will be to estimate its value from the data.

If we are unwilling to model the data as normal, we might instead formulate our hypotheses in a more nonparametric way:

\begin{align*}
H_0 & : X_1, \ldots, X_n \overset{IID}{\sim} f \text{ for some PDF } f \text{ with median 0} \quad (7.3) \\
H_1 & : X_1, \ldots, X_n \overset{IID}{\sim} f \text{ for some PDF } f \text{ with positive median} \quad (7.4)
\end{align*}

We call these hypotheses nonparametric because they do not specify that $f$ follows a particular probability distribution with only a few unknown parameters (normal, exponential, Gamma, ...).

Which formulation we choose may depend on our visual inspection of the data (for skewness, goodness of fit to a normal, etc.) and possibly also on the sample size $n$.

When testing a composite null hypothesis $H_0$, there is a probability of Type I error associated to each data distribution in $H_0$. A test has **significance level** $\alpha$ if

$$\mathbb{P} [\text{reject } H_0] \leq \alpha$$
under every distribution described by $H_0$. In other words, to perform the test at significance level $\alpha$, we must control the maximum probability of Type I error to be $\leq \alpha$, over all possible data distributions in $H_0$.

A difficulty with testing a composite null hypothesis $H_0$ is that we need to reason about the sampling distribution of our chosen test statistic $T$ under every distribution in $H_0$. This can be challenging to do, even by simulation because we may not know which data distribution(s) in $H_0$ we should use to do the simulation. A common simplifying strategy is to find a test statistic $T$ whose sampling distribution is the same for every data distribution in $H_0$. Such a statistic is called pivotal or distribution-free. We will see several examples below and in the next few lectures.

When the alternative $H_1$ is composite, there is also a power (or equivalently, a probability of Type II error) associated to each data distribution in $H_1$. Oftentimes, we cannot simultaneously maximize the power of our test against all possible distributions in $H_1$, so we pick a test that tries to balance power against these possible alternatives.

### 7.2 One-sample $t$-test

Consider testing the hypotheses of Eqs. (7.1) and (7.2) in the normal model. That is, assuming that $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ for unknown parameters $\mu$ and $\sigma^2$, we wish to test

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

If $\sigma^2$ were fixed and known, then by an application of the Neyman-Pearson lemma, the most-powerful level-$\alpha$ test (against any alternative value of the mean $\mu > 0$) would be to reject $H_0$ when $\bar{X} > \frac{\sigma}{\sqrt{n}} z(\alpha)$. This is because the likelihood ratio statistic is decreasing in $\bar{X}$, which has the distribution $\mathcal{N}(0, \sigma^2/n)$ under $H_0$—Example 6.4 worked through this argument in the special case where $\sigma = 1$. Let us write this rejection condition equivalently as

$$
\frac{\sqrt{n} \bar{X}}{\sigma} > z(\alpha).
$$

When $\sigma^2$ is unknown, a natural idea is to estimate $\sigma^2$ 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

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

This statistic $T$ is called the one-sample $t$-statistic. It is a pivotal statistic under the above null hypothesis $H_0$: To see this, suppose that $X_i \sim \mathcal{N}(0, \sigma^2)$. Then we may write $X_i = \sigma Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$. Then $\sqrt{n} \bar{X} = \sigma \sqrt{n} \bar{Z}$. For the sample variance,

$$
S^2 = \sigma^2 \cdot \frac{1}{n-1} \left( (Z_1 - \bar{Z})^2 + \ldots + (Z_n - \bar{Z})^2 \right).
$$
So a factor of $\sigma$ cancels in the numerator and denominator of $T$, yielding

$$T = \frac{\sqrt{n\bar{Z}}}{\sqrt{\frac{1}{n-1} \left((Z_1 - \bar{Z})^2 + \ldots + (Z_n - \bar{Z})^2\right)}}.$$ 

Since the distribution of $Z_1, \ldots, Z_n, \bar{Z}$ is the same for any value of $\sigma^2 > 0$, it follows that the distribution of $T$ is the same for any $\sigma^2 > 0$, i.e. for any distribution specified by $H_0$.

What actually is the sampling distribution of $T$ under (any distribution in) $H_0$? To answer this question, let us first prove the following result.

**Theorem 7.2.** Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$, and let $\bar{X}$ and $S^2$ be the above sample mean and sample variance. Then $S^2$ is independent of $\bar{X}$, with distribution $S^2 \sim \frac{\sigma^2}{n-1} \chi^2_{n-1}$.

**Proof.** Changing the mean $\mu$ does not affect $S^2$ and shifts $\bar{X}$ by a constant value, which does not affect independence of $S^2$ and $\bar{X}$. So it is enough to consider the case $\mu = 0$.

**Step 1:** Since $S^2$ is a function of $X_1 - \bar{X}, \ldots, X_n - \bar{X}$, to show that it is independent of $\bar{X}$, it then suffices to show that $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ is independent of $\bar{X}$. The entries of

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

are linear combinations of $(X_1, \ldots, X_n)$, which is multivariate normal. Then by the argument of Problem 1 on Homework 2, $(\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X})$ is also multivariate normal. By Theorem 3.6 from Lecture 3, to show independence of $\bar{X}$ and $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$, it suffices to show that $\text{Cov}[\bar{X}, X_i - \bar{X}] = 0$ for each $i = 1, \ldots, n$. We have

$$\text{Cov}[\bar{X}, X_i - \bar{X}] = \text{Cov}[\bar{X}, X_i] - \text{Cov}[\bar{X}, \bar{X}].$$

For the first term, by bilinearity of covariance and the fact that $\text{Cov}[X_j, X_i] = 0$ for all $j \neq i$,

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

For the second term, since $\bar{X} \sim \mathcal{N}(0, \frac{\sigma^2}{n})$, $\text{Cov}[\bar{X}, \bar{X}] = \text{Var}[\bar{X}] = \frac{\sigma^2}{n}$ also. Then

$$\text{Cov}[\bar{X}, X_i - \bar{X}] = 0$$

as desired. So $\bar{X}$ is indeed independent of $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$, and hence of $S^2$.

**Step 2:** To show the chi-squared distribution for $S^2$, write $X_i = \sigma Z_i$ where $Z_i \sim \mathcal{N}(0, 1)$, and

$$\frac{(n-1)}{\sigma^2} S^2 = (Z_1 - \bar{Z})^2 + \ldots + (Z_n - \bar{Z})^2$$

$$= (Z_1^2 - 2Z_1\bar{Z} + \bar{Z}^2) + \ldots + (Z_n^2 - 2Z_n\bar{Z} + \bar{Z}^2)$$

$$= Z_1^2 + \ldots + Z_n^2 - 2(Z_1 + \ldots + Z_n)\bar{Z} + n\bar{Z}^2$$

$$= (Z_1^2 + \ldots + Z_n^2) - 2n\bar{Z}^2 + n\bar{Z}^2$$

$$= (Z_1^2 + \ldots + Z_n^2) - (\sqrt{n}\bar{Z})^2.$$
Let \( U = \frac{n-1}{\sigma^2} S^2 \), \( W = Z_1^2 + \ldots + Z_n^2 \), and \( V = (\sqrt{n} \bar{Z})^2 \). Then this says

\[
W = U + V.
\]

We showed that \( S^2 \) is independent of \( \bar{X} = \sigma \bar{Z} \), so \( U \) (being a function of \( S^2 \)) is independent of \( V \) (being a function of \( \bar{X} \)). Since \( W \) is the sum-of-squares of \( n \) independent \( \mathcal{N}(0,1) \) variables, and \( V \) is the square of a single \( \mathcal{N}(0,1) \) variable, this implies that \( U \) has the distribution of the sum-of-squares of \( n - 1 \) independent \( \mathcal{N}(0,1) \) variables. That is, \( U = \frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1} \), so \( S^2 \sim \frac{\sigma^2}{n-1} \chi^2_{n-1} \).

**Remark 7.3.** This theorem explains why we often define \( S^2 \) with the normalization \( \frac{1}{n-1} \) rather than \( \frac{1}{n} \): Since \( \chi^2_{n-1} \) is the distribution of the sum-of-squares of \( n - 1 \) independent \( \mathcal{N}(0,1) \) variables, its expectation is \( n - 1 \). Thus we get \( \mathbb{E}[S^2] = \sigma^2 \), so that \( S^2 \) is an *unbiased* estimator for \( \sigma^2 \). Defining \( S^2 \) with \( \frac{1}{n} \) would instead yield a small downward bias (although for moderate to large sample sizes \( n \), this bias is usually inconsequential in practice).

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 7.2, for \( \mu = 0 \) and any value of \( \sigma^2 > 0 \),

\[
\frac{\sqrt{n} \bar{X}}{\sigma} \sim \mathcal{N}(0,1), \quad \frac{S^2}{\sigma^2} \sim \frac{1}{n-1} \chi^2_{n-1},
\]

and these are independent. We give this distribution for \( T \) a name:

**Definition 7.4.** If \( Z \sim \mathcal{N}(0,1) \), \( U \sim \chi^2_n \), 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 \( t_n \).

This is a continuous distribution whose PDF is unimodal and symmetric around 0. (It is possible to derive a formula for this PDF, but the formula is usually less useful than the above definition.)

To summarize, under the normal null hypothesis of Eq. (7.1), the one-sample *t*-statistic has the distribution

\[
T \sim t_{n-1}.
\]

Letting \( t_{n-1}(\alpha) \) denote the upper-\( \alpha \) point of the distribution \( t_{n-1} \), the test that rejects \( H_0 \) when \( T > t_{n-1}(\alpha) \) is called the (one-sided) **one-sample t-test**. For testing against alternatives \( H_1 : \mu \neq 0 \) instead of \( H_1 : \mu > 0 \), to balance the power of the test against both positive and negative alternatives, it is common to use the two-sided test which rejects \( H_0 \) when \( |T| > t_{n-1}(\alpha/2) \).
7.3 Sign test

Consider now testing the nonparametric hypotheses of Eqs. (7.3) and (7.4): Assuming $X_1, \ldots, X_n$ are IID and distributed according to a PDF $f$, we wish to test

- $H_0 : f$ has median 0
- $H_1 : f$ has median $\mu$ for some $\mu > 0$

The above $t$-statistic is no longer pivotal and does necessarily have the distribution $t_{n-1}$ under this more general null hypothesis $H_0$.

Let’s consider instead the sign statistic

$$S = \sum_{i=1}^{n} 1\{X_i > 0\}$$

where

$$1\{X_i > 0\} = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i \leq 0 \end{cases}$$

Under any distribution belonging to the null hypothesis $H_0$, since the PDF $f$ has median 0, we have $P[X_i > 0] = \frac{1}{2}$. Then each variable $1\{X_i > 0\}$ has distribution Bernoulli($\frac{1}{2}$), and these are independent, so

$$S \sim \text{Binomial}(n, \frac{1}{2}).$$

This holds under any distribution in $H_0$, so $S$ is pivotal. Letting $b_n(\alpha)$ be the upper-$\alpha$ point of Binomial($n, \frac{1}{2}$), the test which rejects $H_0$ when $S > b_n(\alpha)$ is called the sign test.

**Remark 7.5.** When $n$ is large, we may approximate the threshold value $b_n(\alpha)$ using a normal approximation: Note that Binomial($n, \frac{1}{2}$) is the sum of $n$ IID Bernoulli random variables with mean $\frac{1}{2}$ and variance $\frac{1}{4}$. By the Central Limit Theorem, if $S \sim \text{Binomial}(n, \frac{1}{2})$, then

$$\sqrt{4n} \left( \frac{S}{n} - \frac{1}{2} \right) \to \mathcal{N}(0, 1)$$

in distribution as $n \to \infty$. In other words, multiplying by $\sqrt{n/4}$ and adding $n/2$, the Binomial($n, \frac{1}{2}$) distribution for $S$ is approximately the distribution $\mathcal{N}(\frac{n}{2}, \frac{n}{4})$—here the standard deviation of the normal is $\sqrt{n/4}$. Then the upper-$\alpha$ point $b_n(\alpha)$ is also approximately the upper-$\alpha$ point of this normal distribution, which is $\frac{n}{2} + \frac{\sqrt{n}}{2} z(\alpha)$. An approximate sign test may reject $H_0$ when

$$S > \frac{n}{2} + \frac{n}{4} z(\alpha)$$

instead of $S > b_n(\alpha)$.

Note that this is equivalent to rejecting $H_0$ when

$$\sqrt{4n} \left( \frac{S}{n} - \frac{1}{2} \right) > z(\alpha).$$
For any finite sample size \( n \), the Type I error of this test is not guaranteed to be \( \leq \alpha \). However, under \( H_0 \), we have

\[
\mathbb{P} \left[ \sqrt{\frac{4n}{n}} \left( \frac{S}{n} - \frac{1}{2} \right) > z(\alpha) \right] \to \alpha
\]

as \( n \to \infty \) because of the convergence in distribution from Eq. (7.5). So the Type I error will be approximately \( \alpha \) when the sample size \( n \) is large. Such a test is sometimes called an asymptotic level-\( \alpha \) test.