Lecture 8 — Two-sample t-test and signed rank test

8.1 Two-sample t-test

Consider the setting of two independent samples $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$, as in Example 7.2 of last lecture. Here μ_X, μ_Y, σ^2 are all unknown; note that we are assuming (for now) a common variance σ^2 for both samples. For the testing problem

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

a natural idea is to reject H_0 for large values of $\bar{X} - \bar{Y}$. Observe that $\bar{X} \sim \mathcal{N}(\mu_X, \frac{\sigma^2}{n})$, $-\bar{Y} \sim \mathcal{N}(-\mu_Y, \frac{\sigma^2}{m})$, and these are independent. Then their sum is distributed[1] as

$$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \tfrac{\sigma^2}{n} + \tfrac{\sigma^2}{m}).$$

Under H_0, $\mu_X - \mu_Y = 0$, so $(\bar{X} - \bar{Y})/\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}} \sim \mathcal{N}(0, 1)$. If σ^2 were known, then a level-α test based on $\bar{X} - \bar{Y}$ would reject when

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} > z(\alpha).$$

Since σ^2 is unknown, we estimate it from the data. We may use both the X_i's and Y_i's to estimate σ^2 by taking the **pooled sample variance**

$$S_p^2 = \frac{1}{m + n - 2} \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{j=1}^{m} (Y_j - \bar{Y})^2 \right),$$

and take as a test statistic

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

To derive the null distribution and rejection threshold for T, we may rewrite this as

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \Bigg/ \sqrt{S_p^2/\sigma^2}.$$

[1] Recall, as a special case of Example 3.3 from Lecture 3, that if $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

By Theorem 7.3 of Lecture 7 (and by independence of the two samples), $\bar{X}, \bar{Y}, \sum_i (X_i - \bar{X})^2, \sum_j (Y_j - \bar{Y})^2$ are all independent, with the last two quantities distributed as $\sigma^2 \chi^2_{n-1}$ and $\sigma^2 \chi^2_{m-1}$. Then under H_0,

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim \mathcal{N}(0,1), \quad \frac{S_p^2}{\sigma^2} \sim \frac{1}{m+n-2} \chi^2_{m+n-2},$$

and these are independent. So the distribution of T is the same for all data distributions $P \in H_0$ and is given by

$$T \sim t_{m+n-2}.$$

The test that rejects H_0 when $T > t_{m+n-2}(\alpha)$ (the upper α point of the t_{m+n-2} distribution) is called the **two-sample *t*-test**.

Remark 8.1. The assumption of common variance σ^2 for the two samples is oftentimes problematic (and violated) in practice. If we assume instead that $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y_1, \ldots, Y_m \overset{IID}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$ for possibly different values of σ_X^2 and σ_Y^2, then $\mathrm{Var}(\bar{X} - \bar{Y}) = \frac{1}{n}\sigma_X^2 + \frac{1}{m}\sigma_Y^2$, and we may estimate this by $\frac{1}{n}S_X^2 + \frac{1}{m}S_Y^2$, where S_X^2 and S_Y^2 are the sample variances of the two samples. Then we may use the test statistic

$$T_{\text{welch}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n}S_X^2 + \frac{1}{m}S_Y^2}}.$$

The distribution of T_{welch} under H_0 is no longer exactly a t distribution, but it was shown by Welch (1947) to be close to the t distribution with

$$\frac{(S_X^2/n + S_Y^2/m)^2}{(S_X^2/n)^2/(n-1) + (S_Y^2/m)^2/(m-1)}$$

degrees of freedom. The test that rejects when T_{welch} exceeds the upper α point of this t distribution is called **Welch's *t*-test** or the **unequal variances *t*-test**.

8.2 Wilcoxon signed rank test

Let's return to the one-sample setting $X_1, \ldots, X_n \overset{IID}{\sim} f$, where we drop the normality assumption and only wish to test

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

Because the shape of f is arbitrary under H_0, the distribution of the t-statistic is no longer the same under every data distribution $P \in H_0$—in particular, it can be very far from t_{n-1} if n is moderately small and f is heavy-tailed. We consider instead the **signed rank statistic** W_+, defined in the following way:

1. Sort $|X_1|, |X_2|, \ldots, |X_n|$ in increasing order. Assign the smallest value (closest to zero) a rank of 1, the next smallest value a rank of 2, etc., and the largest value a rank of n.

2. Define W_+ as the sum of the ranks corresponding to only the positive values of $X_1, \ldots, X_n$.

As an example, suppose we have four observations $X_1 = 2, X_2 = -4, X_3 = -1, X_4 = 10$. Then the ranks of these four observations would be $2, 3, 1, 4$. Observations X_1 and X_4 are positive, so $W_+ = 2 + 4 = 6$.

We expect W_+ to be larger under H_1 than under H_0, because high-rank observations are more likely to be positive under H_1. The test that rejects for large W_+ is called **Wilcoxon's signed rank test**. The following theorem states that W_+ has the same distribution under every $P \in H_0$, and provides a method for determining the null distribution and rejection threshold for W_+ when n is large. (When n is small, we can determine the exact null distribution of W_+ by computing W_+ for all 2^n possible combinations of $+$ and $-$ signs for the ranked data.)

Theorem 8.2. *The distribution of W_+ is the same for every PDF f that is symmetric about 0. For large n, this distribution is approximately $\mathcal{N}(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24})$. (More formally, $\sqrt{\frac{24}{n(n+1)(2n+1)}}\left(W_+ - \frac{n(n+1)}{4}\right) \to \mathcal{N}(0,1)$ in distribution as $n \to \infty$.)*

Proof sketch. We'll show that the distribution of W_+ is the same for every f, and that $\mathbb{E}[W_+] = \frac{n(n+1)}{4}$ and $\mathrm{Var}[W_+] = \frac{n(n+1)(2n+1)}{24}$. We'll provide only a heuristic explanation of why W_+ is asymptotically normal.

Let $f_0(x_1, \ldots, x_n) = \prod_{i=1}^n f(x_i)$ be the joint PDF of the data. By symmetry of f about 0, $f_0(\pm x_1, \ldots, \pm x_n)$ is the same for each of the 2^n combinations of $+/-$ signs. This implies, conditional on $|X_1|, \ldots, |X_n|$, the signs of $X_1, \ldots, X_n$ are independent and each equal to $+$ or $-$ with probability $\frac{1}{2}$. Then, letting $I_k = 1$ if the value with rank k is positive and $I_k = 0$ if it is negative, $I_1, \ldots, I_n \overset{IID}{\sim} \text{Bernoulli}(\frac{1}{2})$ for any PDF f that is symmetric about 0.

The signed rank statistic is

$$W_+ = \sum_{k=1}^n k I_k.$$

Since $I_1, \ldots, I_n$ have the same distribution under any symmetric PDF f about 0, the distribution of W_+ is the same for all such PDFs f. We compute

$$\mathbb{E}[W_+] = \sum_{k=1}^n k\mathbb{E}[I_k] = \frac{1}{2}\sum_{k=1}^n k = \frac{n(n+1)}{4},$$
$$\mathrm{Var}[W_+] = \sum_{k=1}^n \mathrm{Var}[kI_k] = \sum_{k=1}^n k^2 \mathrm{Var}[I_k] = \frac{1}{4}\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{24},$$

where the computation for variance uses that $I_1, \ldots, I_n$ are independent.

To explain why W_+ is approximately normally distributed, define the **empirical CDF** of $|X_1|, \ldots, |X_n|$ by

$$F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{|X_i| \le t\}.$$

($F_n(t)$ is the fraction of values of $|X_i|$ that are at most t.) Then the rank associated with X_i is exactly $nF_n(|X_i|)$, so

$$W_+ = \sum_{i=1}^{n} nF_n(|X_i|)\mathbb{1}\{X_i > 0\}.$$

When n is large, one may show that $F_n(t)$ is, with high probabiliy, close to the true CDF $F(t)$ of $|X_i|$ for every $t \in \mathbb{R}$, and hence that the difference between W_+ and

$$\tilde{W}_+ = \sum_{i=1}^{n} nF(|X_i|)\mathbb{1}\{X_i > 0\}$$

is negligible. But $\tilde{W}$ is just the sum of IID random variables $Y_i := nF(|X_i|)\mathbb{1}\{X_i > 0\}$, and hence asymptotically normally distributed by the CLT. □