6.S951 Modern Mathematical Statistics Fall 2024

Lecture 13 — October 31, 2024

Prof. Stephen Bates

Scribe: Lorenzo Shaikewitz

1 Outline

Agenda:

1. p-values (continued)
2. Permutation tests
3. Start predictive inference (conformal prediction)

Last time:

1. Confidence intervals $\leftrightarrow$ testing duality
2. p-values

2 p-values (continued)

Recall p-values are a measure of disagreement of data X with the null hypothesis $\{\theta \in \Theta_0\}$. We define a p-value as:

Definition 1 (p-value)**.** *The* p-value *for null* Θ_0 *is a function* $f : \mathcal{X} \to [0, 1]$ *such that*

$$P_{\theta_0}(f(x) \leq t) \leq t \quad \forall\, t \in [0, 1],\, \forall\, \theta_0 \in \Theta_0$$

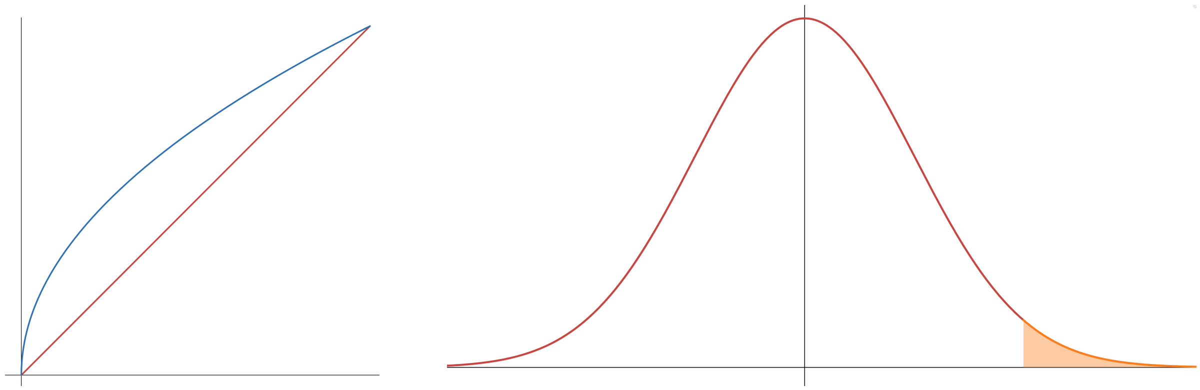
This definition simply means the p-value is a super-uniform random variable (Figure 1a). Intuitively, the p-value is a *tail probability* (Figure 1b). We put the evidence on a standardized $[0, 1]$ scale.

Example 2 (Standard Normal)**.** *Consider* $X \sim \mathcal{N}(\theta, 1)$*, where the null is* $\Theta_0 = \{0\}$*. Then,*

$$f(x) = 1 - \Phi(x)$$

is a p-value. Note that $f(x)$ *is good against alternatives* $\theta > 0$*, meaning it is one-sided. This is visualized in Figure 1b. A two-sided p-value could be:*

$$f(x) = 2(1 - \Phi(|x|))$$



(a) The CDF of a p-value (blue) compared to the CDF of the uniform distribution (red). The p-value CDF is always above uniform.

(b) The PDF of $\mathcal{N}(0,1)$ with a (spooky) tail probability indicated in orange. This tail probability is a p-value.

Figure 1: Intuition for the definition of a p-value.

P-values are typically used for null hypothesis significance testing (p ¡ 0.05 means reject the null). This is formalized in the following:

Proposition 3 (p-values → tests)**.** *If f(x) is a p-value for* Θ_0*, then the following test is level-*α*:*

$$\phi(x) = \begin{cases} 1 & f(x) \leq \alpha \\ 0 & \text{otherwise} \end{cases}$$

Proof. By the definition of the p-value,

$$P(\phi(x) = 1) = P(f(x) \leq \alpha) \leq \alpha$$

□

Example 4 (Universal inference p-value)**.** *Fix* θ_0*,* θ_1*. Let*

$$g(x) = \frac{P_{\theta_1}(x)}{P_{\theta_0}(x)} > 0$$

Notice that $\mathbb{E}_{\theta_0}[g(x)] = \int \frac{P_{\theta_1}(x)}{P_{\theta_0}(x)} P_{\theta_0}(x)\mathrm{d}x = 1$*. Thus, using the Markov inequality,*

$$P(\frac{1}{g(x)} \leq t) = P(g(x) \geq \frac{1}{t}) \leq t$$

This shows that $\frac{1}{g(x)}$ *is a p-value. Note that in general, if you have a quantity with expectation 1 you can use this approach to make a p-value.*

In conclusion, p-values and confidence intervals are complementary:

- p-values summarize a collection of tests $\{\phi_\alpha\}_{\alpha \in [0,1]}$. These are tests of different levels for the same null.

- Confidence intervals summarize a collection of tests $\{\phi_{\tilde{\theta}}\}$ for $\tilde{\theta} \in \Theta$. These are tests of the same level for different nulls.

Example 5. *Let $X_1, \ldots, X_n \sim \mathcal{N}(\theta, 1)$ with null $\Theta_0 = \{0\}$. For a simulation with $n = 1000$, the value $\bar{x} = 0.1$ has a p-value of 0.0008 and a 95% confidence interval $[0.05, 0.15]$.*

3 Permutation Tests

Consider the setting where we sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P_1$ and $Y_1, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} P_2$ with each X_i independent of each Y_i and $X_i, Y_i \in \mathcal{X}$. We want to test if $P_1 = P_2$.

Let $\vec{Z} = (X_1, \ldots, X_n, Y_1, \ldots, Y_m) \in \mathcal{X}^{n+m}$. The key idea of the permutation test is: if $P_1 = P_2$ and $\vec{Z}$ is a vector of i.i.d. entries, then any permutation of the elements of $\vec{Z}$ is equally likely to occur. If the particular permutation $\vec{Z}$ is extreme compared to the distribution of permutations, then $P_1 \neq P_2$ probably.

A permutation test requires a *test function*, which measures the agreement of the first n elements with the last m elements. In general:

Definition 6 (Test Function)**.** *A* test function *is $T : \mathcal{X}^{n+m} \to \mathbb{R}$.*

For example, consider the difference in means:

$$T(x_1, \ldots, x_n, y_1, \ldots, y_n) = \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{m} \sum_{j=1}^{m} Y_j \right|$$

We also need a *permutation*. Let σ denote a permutation of $\{1, \ldots, n+m\}$, where σ maps $(1, \ldots, n+m) \to (\sigma(1), \ldots, \sigma(n+m))$. Let $\vec{Z}_\sigma$ denote $\vec{Z}$ with elements permuted by σ:

$$(\vec{Z}_\sigma)_i = \vec{Z}_{\sigma(i)}$$

The permutation test is the following procedure:

Algorithm 1 Permutation Test

Let $T^* \leftarrow T(Z)$
for $k = 1, \ldots, K$ **do**
  draw random permutation σ_k
  compute $t_k = T(\vec{Z}_{\sigma_k})$
end forreturn $p = \frac{1}{K+1}\left(1 + \sum_{k=1}^{K} \mathbf{1}_{\{T^* \le t_k\}}\right)$

Proposition 7 (Validity of permutation test p-value)**.** *If $P_1 = P_2$, then p in Algorithm 1 is super-uniform.*

Proof. The following is a sketch of the proof in the limiting case. Consider $K \to \infty$. By the law of large numbers,

$$p \to P(T^* \le T(\vec{Z}_\sigma))$$

For a random permutation σ. We can write this in terms of F, the CDF of $T(\vec{Z}_\sigma)$:

$$P(T^* \leq T(\vec{Z}_\sigma)) = 1 - F(T^*) \dot{\sim} \text{Unif}[0,1]$$

where we use the CDF transform to show the distribution is approximately uniform, up to ties in $T(\vec{Z}_\sigma)$ and the finite grid due to the number of permutations. This shows p is super-uniform, which shows it is a p-value. □

We conclude with some remarks about the permutation test.

- (good) The permutation test is valid for all distributions P_1 and test functions T.
- (neutral) The permutation test requires some choice of T.
- (bad) The power may not be optimal.
- (bad) We cannot use the permutation test to get confidence intervals for the mean difference. This is because the permutation test only tests if the distributions are the same, not if their means are the same.