POBABILITY AND STATISTICS, Lessons 11-12.
TESTING STATISTICAL HYPOTHESES

- **Parametric tests**. Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a dominated, identifiable, parametric statistical space, $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. We want to decide between: $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, where $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \Theta$. In case of $|\Theta_0| = 1$ we have a *simple zero-hypothesis*, otherwise it is *composite*. The same for the alternative hypothesis. Our decision is based on the $\mathbf{X} = (X_1, \dots, X_n) \sim \mathbb{P}_\theta$ i.i.d. sample from the sample space $\mathcal{X}$.

  1. We calculate an appropriate statistic $T(\mathbf{X})$ (whose distribution under H_0 is known).
  2. We divide the sample space into *acceptance region* $\mathcal{X}_a$ and *rejection (or critical) region* $\mathcal{X}_c$, where $\mathcal{X}_a \cap \mathcal{X}_c = \emptyset$ and $\mathcal{X}_a \cup \mathcal{X}_c = \mathcal{X}$. Usually, $\mathcal{X}_c = \{\mathbf{x} : T(\mathbf{x}) \geq c_\varepsilon\}$, where c_ε is a quantile (percentile) value of T, and ε is the size of the test (the significance of the test is ε, the confidence is $1 - \varepsilon$).
  3. The decision is: if $\mathbf{X} \in \mathcal{X}_a$, then accept, otherwise reject H_0 with confidence $1 - \varepsilon$.

  - *Definition*: The size of the test defined by $\mathcal{X}_c$ is $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathbf{X} \in \mathcal{X}_c) = \varepsilon$ (supremum of the so-called Type I. errors).
  - *Definition*: The *strength* of the test defined by $\mathcal{X}_c$ is $\beta_n(\theta, \varepsilon) = 1 - \mathbb{P}_\theta(\mathbf{X} \in \mathcal{X}_a) = \mathbb{P}_\theta(\mathbf{X} \in \mathcal{X}_c)$, for $\theta \in \Theta_1$ (1 minus the Type II. error of alternative θ).
  - *Definition*: The test defined by $\mathcal{X}_c$ is *uniformly most powerful test* of size ε, if among the tests with size at most ε, its strength is the largest possible, for any alternative. That is, $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathbf{X} \in \mathcal{X}_c) \leq \varepsilon$, and for any other $\mathcal{X}'_c$ with $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\mathbf{X} \in \mathcal{X}'_c) \leq \varepsilon$, the following also holds: $\mathbb{P}_\theta(\mathbf{X} \in \mathcal{X}_c) \geq \mathbb{P}_\theta(\mathbf{X} \in \mathcal{X}'_c), \quad \forall \theta \in \Theta_1$.
  - *Definition*: The test function $\Psi(\mathbf{X})$ is the probability of rejecting H_0 based on the observation $\mathbf{X}$. It is 1, if $\mathbf{X} \in \mathcal{X}_c$; 0, if $\mathbf{X} \in \mathcal{X}_a$; and may be $p \in (0,1)$, if $\mathbf{X} \in \mathcal{X}_r$ (randomization region), when we cannot decide immediately (discrete cases).
  - **Neyman–Pearson Theorem** (on the existence of a uniformly most powerful test). In a parametric statistical space let $L_\theta(\mathbf{X})$ is the likelihood function based on the i.i.d. sample $\mathbf{X} = (X_1, \dots, X_n)$. Then for the simple alternative

    $$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1$$

    for any $0 < \varepsilon < 1$ there exists a uniformly most powerful test of size ε, and it is defined uniquely (with probability 1) in the following way:

    $$\psi(\mathbf{X}) = 0, \text{ if } \frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})} < c; \quad \psi(\mathbf{X}) = p, \text{ if } \frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})} = c; \quad \psi(\mathbf{X}) = 1, \text{ if } \frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})} > c,$$

    where $p \in [0,1)$ and $c > 0$ are appropriate constants depending on ε.
  - *Remark*: This theorem can be extended to composite hypotheses, when the likelihood function depends monotonously on a sufficient statistic T. In these cases the above inequalities can be reformulated in terms of T (with other constants).
  - See the table enclosed for u (z), t, F, and $Welch$-tests.

- **Non-parametric tests**: H_0 applies not to the parameter.

  - χ^2-test for goodness of fit to a given distribution, and homogenity or independence of two distributions (see the table enclosed).
  - 1- and 2-sample Kolmogorov–Smirnov tests are based on the K–S theorems.