Hypothesis Testing

Statistical Experiments

1. A *statistical experiment* consists of the following data:
   (a) A set $X_1, X_2, \dots, X_n$ of *iid* random variables with a common (unknown) distribution $\mathbb{P}$.
   (b) A (parametric, identifiable) statistical model $(E, \{\mathbb{P}_\theta : \theta \in \Theta\})$ for $\mathbb{P}$ which is well-specified (i.e., there exists $\theta^* \in \Theta$ such that $\mathbb{P} = \mathbb{P}_{\theta^*}$).
   (c) A partition of Θ into disjoint sets Θ_0 and Θ_1, which represent the null and alternative hypotheses, respectively.
2. Note that θ^* is fixed (i.e., non-random). The purpose of a statistical experiment is not to determine its location, but rather reject the assertion that it lies in Θ_0.
3. Given an observation $X_1, X_2, \dots, X_n$, we often formulate the null and alternative hypotheses as a function of θ, i.e., we write H_0, H_1 **TODO: Think about this first.**

Tests and Errors

1. A *hypothesis test* is a function $\psi_n : E^n \to \{0, 1\}$.
2. The *Type I Error* associated to ψ_n is

   $$\alpha_{\psi_n}(\theta) = \mathbb{P}_\theta(\psi_n(X_1, X_2, \dots, X_n) = 1 \mid \theta \in \Theta_0).$$

   This represents the probability of rejecting the null hypothesis, given that it is true. Note that in most examples, Θ_0 is a singleton (in these cases, α is a single number), but this is not a guarantee.
3. The *Type II Error* associated to ψ_n is

   $$\beta_{\psi_n}(\theta) = \mathbb{P}_\theta(\psi_n(X_1, X_2, \dots, X_n) = 0 \mid \theta \in \Theta_1).$$

   This likewise represents the probability of failing to reject the null hypothesis, given that it is false.
4. The *Power* of a statistical test ψ_n is 1 minus the type II error. **I think this is wrong. It should be 1 minus the infemum of type II error.**

The Level of a Test

1. An important point about hypothesis testing: While we always want to strike a balance between type I and II errors, we will usually specify a *Level* for our test, which represents a certain amount of type I error we are willing to tolerate. So in this since, we will generally favor minimizing type II error, given a certain level. The level of a test is denoted

   $$\alpha := \sup\{\alpha_\psi(\theta) \mid \theta \in \Theta_0\}.$$

We will often say a test *rejects at level* α.

2. Often, we will not understand the distribution of our test statistic directly, but rather only understand its asymptotic distribution (i.e., considering sample size in the limit). In this case, we will specify an *Asymptotic Level* of a statistical test, also denoted by

$$\alpha := \limsup_{n\to\infty} \sup\{\alpha_{\psi_n}(\theta) \mid \theta \in \Theta_0\}.$$

Relationship between level and power

TODO

The p-value of a Test

1. In general, a test has the form

$$\psi_{n,\alpha} = \mathbf{1}\{T_n > C_\alpha\},$$

where T_n is called the *test statistic*, which in our current examples, will take the form

$$T_n = \sqrt{n}(f(\bar{X}_n) - \theta_0),$$

where $f(\bar{X}_n)$ is an estimator for θ_0, and which converges in distribution to something familiar (so far in our examples, this is usually $N(0, \sigma^2)$). What this means is that by choosing C_α, we can decide the asymptotic level α. Thus, most of the work will be finding the test statistic T_n.
2. In most examples, the null hypothesis is a singleton $\{\theta_0\}$ (for two-sided tests) or a half interval $\{[\theta_0, \pm\infty)\}$ (for one-sided tests). However, this does not generalize well to more abstract null hypotheses.
3. The *Asymptotic p-value* of a test $\psi_{n,\alpha}$ is the smallest asymptotic level α at which $\psi_{n,\alpha}$ rejects H_0, given an observation $X_1, X_2, \ldots, X_n$. Equivalently, this is the probability (under the nearest point of the null hypothesis) of an event at least as extreme as the observation.
4. I should note that I have added some of my own interpretation here. It's not clear from the lectures that p-value makes no sense without an a priori observation, but this is the only way I can make sense of the definition, whereas everything else makes sense as a function of the observation or parameter space.
5. In most straightforward statistical tests, we don't usually use asymptotic levels or p-values. Instead, the exact distribution of T_n will be known, e.g., from a t-distribution or χ^2-distribution.

An example

Recall the kiss example: We observe n people kissing and record whether they turn their heads to the right or left. We model these observations with a Bernoulli distribution, with null hypothesis $\Theta_0 = \{1/2\}$ and alternative hypothesis $\Theta_1 = (0, 1/2) \cup (1/2, 1)$. Let

$$T_n = \sqrt{n} \cdot \frac{|\bar{X_n} - 1/2|}{\sqrt{(1/2)(1-(1/2))}} = 2\sqrt{n}|\bar{X_n} - 1/2|$$

and let

$$\psi_{n,\alpha} = \mathbf{1}(T_n > q_{\alpha/2}).$$

Then by the Central Limit Theorem (assuming the null hypothesis), $\psi_{n,\alpha}$ has asymptotic level α:

$$\text{Level} := \lim_{n\to\infty} \sup_{p\in\Theta_0} (\alpha_{\psi_{n,\alpha}}(p)) = \underbrace{\lim_{n\to\infty} \mathbb{P}_{1/2}(\psi_{n,\alpha} = 1) = \alpha}_{CLT}.$$

Suppose $n = 124$, $\hat{X_n} = 0.645$. Then $T_n = 3.229$. Let α be the value such that $q_{\bar{\alpha}/2} = T_n$, so that $\psi_{n,\alpha}$ has asymptotic level $\bar{\alpha}$. Then $\bar{\alpha}$ is the p-value of this test. Since $T_n \to N(0,1)$ in distribution, we can compute the p-value as $\bar{\alpha} = 10^{-4}$.