MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.s077 Spring 2018
Recitation 4 - Hypothesis Testing Friday 03/02

Review.

- **Hypothesis Testing Framework**:
  - A collection of distributions.
    $\mathbf{H_0}$: data $X_1, \cdot, X_n$ distributed according to $P \in \mathcal{C}_0$.
    $\mathbf{H_A}$: data $X_1, \cdot, X_n$ distributed according to $P \in \mathcal{C}_1$.
    where $\mathcal{C}_0$ and $\mathcal{C}_1$ are collections of distributions.
  - Significance Level, α.
    α helps define a family, $\mathcal{R}_\alpha$, which is the rejection region.
    If $\mathcal{X} \in \mathcal{R}_\alpha$ then we reject H_0. Otherwise, we do not reject H_0.
  - Rejection Region.
    The rejection region is of the form:
    $\mathcal{R}_\alpha = \{T(X) \geq t_\alpha\}$
- **Procedure**:
  1. Formulate the two hypotheses as collections of distributions (**critical**). This produces H_0 (null) and H_A (alternate).
  2. Use a statistic whose distribution under the null hypothesis, H_0 is known (or approximately known). This is $T(X)$.
  3. Fix a significance level, α.
  4. Define the rejection region for H_0: $\mathcal{R}_\alpha = \{T(X) \geq t_\alpha\}$.
  5. **Finally**, consider the data to reject or fail to reject H_0.
- z**-test**:
  - One-sample test for mean, i.e.
    Let $\mu = \mathbb{E}[X]$. Then we consider one of:
    1. $\mu = \mu_0$ vs $\mu \neq \mu_0$; (two-sided)
    2. $\mu = \mu_0$ vs $\mu > \mu_0$;
    3. $\mu = \mu_0$ vs $\mu < \mu_0$;
  - **Null Hypothesis**. X_i are iid from from P. P has mean μ_0. Under P, data variance is known, i.e. $Var[X_i] = \sigma_0^2, \forall i$.

- **Statistic**. $Z = \frac{\sum_{i=1}^{n}(X_i - \mu_0)}{\sqrt{n\sigma_0^2}}$
- **Normality**. $Z \sim N(0,1)$ by the CLT. Observe that this holds only asymptotically.
- **Rejection Region**. $\mathcal{R}_\alpha$ is defined using the tail-probabilities of the Normal distribution. For example, for $\alpha = 0.05$, and a two-sided test, we have $\mathcal{R}_\alpha = \{|T(X)| \geq 1.96\}$ which is 2.5% on either tail.
- **Rejection Property**. When $\mu \neq \mu_0$, the statistic is large making rejection "easier" as we get more data.

• t**-test**:

- One-sample test for mean, i.e.
  Let $\mu = \mathbb{E}[X]$. Then we consider one of:
  1. $\mu = \mu_0$ vs $\mu \neq \mu_0$; (two-sided)
  2. $\mu = \mu_0$ vs $\mu > \mu_0$;
  3. $\mu = \mu_0$ vs $\mu < \mu_0$;
- **Null Hypothesis**. X_i are iid from from P. P has mean μ_0. Under P, data variance is **unknown**. So we use $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$.
- **Statistic**. $t = \frac{\sum_{i=1}^{n}(X_i - \mu_0)}{\sqrt{n\hat{\sigma}_0^2}} = (\hat{\mu} - \mu_0)\sqrt{\frac{n}{\hat{\sigma}^2}}$.
- **Magic**. The distribution is independent of μ_0 and $Var[X]$.
- **Normality**. By CLT and LLN. Specifically, by LLN $\hat{\mu} \to \mu_0$ and $\hat{\sigma}^2 \to Var[X]$. Observe that these hold only asymptotically. But the t-distribution is well-defined even for small n.
- **Rejection Region**. $\mathcal{R}_\alpha$ is defined using the tail-probabilities of the Normal distribution. For example, for $\alpha = 0.05$, and a two-sided test, we have $\mathcal{R}_\alpha = \{|T(X)| \geq t_{\alpha,n-1}\}$ which is 2.5% on either tail.
- **Rejection Property**. When $\mu \neq \mu_0$, the statistic is large making rejection "easier" as we get more data. Note that using $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu_0)^2$ would not be a good idea because the statistic would not be small for $\mu \neq \mu_0$, even though $\hat{\sigma}^2$ still converges to $Var[X]$.

• **Exact vs Approximate**:

- For t-test if data were truly Gaussian (with mean μ_0 but unknown variance), the distribution of t-statistic is independent of unknown variance. This is the trick called studentization. The t-distribution is well known. This makes test exact meaning that we can guarantee significance=α by selecting appropriate threshold on t-statistic.

– Classical statisticans were not happy with statements like "asymptotically the test has significance (1 - α)". They really wanted to have exact significance α even for $n = 10$. Such tests were invented (and they are called exact). These exact tests are split into two groups: parametric and non-parametric. The latter ones include things like Kolmogorov-Smirnov, permutation tests and, most notably, sum-rank tests. We will talk about them in recitations two weeks from now and Lectures 8 and 9. Here we discuss parametric exact tests.

  Example: Instead of testing $H : \mu = \mu_0$ vs $K : \mu \neq \mu_0$ one would test $H : X \sim \mathcal{N}(\mu_0, 1)$ vs $K : X \sim \mathcal{N}(\mu, 1), \mu \neq \mu_0$. In this case z-test (with threshold $\Phi^{-1}(\alpha/2)$) is exact. Similarly for t-test and t-distribution. For this reason, z-test and t-tests are sometimes called "parametric". In reality they are applicable (and most often applied) to non-parametric situations, where distributions of samples X_i are very non-Gaussian (but without heavy-tails). The price is of course that significance is now only guaranteed asymptotically in n.

  For heavy-tailed distributions, one uses "fully non-parametric" tests (sign test, Mann-Whitney U-test etc). Parametric vs non-parametric is a matter of debate (i.e. it is a spectrum, rather than binary division). We will consider in upcoming recitations some such non-parametric tests (e.g. permutation test).

Problem 1. [Wasserman, Ex. 10.6] There is a theory that people can postpone their death until after an important event. To test the theory, Phillips and King (1988) collected data on deaths around the Jewish holiday Passover. Of 1919 deaths, 922 died the week before the holiday and 997 died the week after. Think of this as a binomial and test the null hypothesis that $\mu = 1/2$. Construct a 95% confidence interval for μ. Will you reject the null hypothesis at for $\alpha = 0.05$?

Solution. As suggested in the problem, we can model the setting with a Binomial distribution. Specifically, let each person's death be denoted by a random variable X_i. Let $X_i \in \{0, 1\}$ where $X_i = 1$ denotes the event that person i died *before* the holiday, and $X_i = 0$ with denote the event that the person i died *after* the holiday. We assume that each person can die before or after the holiday, independently of others, with probability μ. Therefore, each X_i is an independent Bernoulli random variable with $\mathbb{P}(X_i = 1) = \mu$. Given that we have $n = 1919$ independent and identical random variables, one corresponding to each person in the study, we let

$Y = \sum_{i=1}^{n} X_i$. We know that $Y \sim Binom(n, \mu)$ and it denotes the number of deaths *before* the holiday.

We now proceed to define our null and alternate hypotheses:

- H_0: $\mu = 1/2$.
- H_A: $\mu < 1/2$.

Note that H_A is a one-sided alternative because of the way the problem is stated, i.e. there is a belief that people can postpone their death until after an important event implying $\mu < 1/2$.

Under H_0, $Y \sim Binom(n = 1919, \mu = 1/2)$. For this distribution, we can get exact probabilities of the distribution. For this distribution, we have that $P(Y \leq 222) = 0.05011$ and $P(Y \leq 223) = 0.04557$. Since the Binomial distribution can only take integer values, we know that the 95% confidence interval is represented by $\{Y \geq 223\}$. Since our observations of $y = 222$ falls just outside of this confidence interval, we can reject H_0 at the level of significance $\alpha = 0.05$.

Note. We can also do this problem using the Z-test which is a good candidate because we are testing for the team with allows the CLT (approximate Normality) to kick in. Secondly, we already know that a Binomial distribution can be well-approximated by a Normal distribution centered on the mean of the Binomial when n is moderately large and μ is not too small.

Our H_0 and H_A remain the same as before. We now produce a z-statistic:

$$z = \frac{\hat{\mu} - \mu_0}{\sqrt{\frac{(\mu_0)(1-\mu_0)}{n}}}.$$

Substituting $\hat{\mu} = \frac{922}{1919}, \mu_0 = 0.5, n = 1919$, we get $z = -1.7084$. The one-sided 95% confidence interval for μ under H_0 is represented by a z-score of -1.645. This means that we can reject H_0 because the observed value of the $z - statistic$ is smaller (for one-sided test) than the threshold for level of significance $\alpha = 0.05$.