18.650. Statistics for Applications
Fall 2016. Problem Set 6

Due Friday, Oct. 21 at 12 noon

NOTE: there was a typo in Problem 2, Question 3

Problem 1 Hypotheses testing - Basics

Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{E}xp(\lambda)$, for some unknown parameter $\lambda > 0$ and let λ_0 be a (known) fixed positive number.

1. Consider the following hypotheses:

$$H_0 : \text{"}\lambda = \lambda_0\text{"} \text{ vs. } H_1 : \text{"}\lambda \neq \lambda_0\text{"}.$$

Give a test with asymptotic level α, for any $\alpha \in (0, 1)$.

2. Consider the following hypotheses:

$$H_0 : \text{"}\lambda \leq \lambda_0\text{"} \text{ vs. } H_1 : \text{"}\lambda > \lambda_0\text{"}.$$

Give a test with asymptotic level (at most) α, for any $\alpha \in (0, 1)$.

3. At a call center, the times between two consecutive calls are modeled as i.i.d. exponential random variables, with some unknown parameter λ. The employer of this center wants to determine whether more employees should be hired. In order to do so, he/she wants to get some information about λ. The value itself of λ is irrelevant, but if $\lambda \geq 1$, meaning that the average time between two calls is no larger than 1 (minute), the employer's decision will be to hire more employees. To make his/her decision, the employer observes the times between 50 consecutive calls, and finds out that the average these observations is 0.98.
In statistical words, the employer needs to test the following hypotheses:

$$H_0 : \text{"}\lambda \leq 1\text{"} \text{ vs. } H_1 : \text{"}\lambda > 1\text{"}.$$

a) Based on the observations and using the same test as in the previous question, should the employer reject the null hypothesis ?
b) What is the p-value of the test ?

Problem 2 A Student one-sided test

Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, for some unknown parameter $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$. We want to test the following hypotheses at non asymptotic level α (for some fixed $\alpha \in (0, 1)$:

$$H_0 : \text{"}\mu > 0\text{"} \text{ vs. } H_1 : \text{"}\mu \leq 0\text{"}.$$

1. Recall the maximum likelihood estimator $(\hat{\mu}, \hat{\sigma}^2)$ of (μ, σ^2).
2. Let $S = \sqrt{n-1}\frac{\hat{\mu}-\mu}{\sqrt{\hat{\sigma}^2}}$. Prove that S is a Student random variable with $n-1$ degrees of freedom.
3. Propose a test with non asymptotic level α. Prove your answer.

Problem 3 Test of independence for Bernoulli random variables

Let X, Y be two Bernoulli random variables and denote by $p = \mathbb{P}[X=1]$, $q = \mathbb{P}[Y=1]$ and $r = \mathbb{P}[X=1, Y=1]$.

1. Prove that X and Y are independent if and only if $r = pq$.
2. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of n i.i.d. copies of (X, Y). Based on this sample, we want to test whether X and Y are independent, i.e., whether $r = pq$.
   a) Define $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\hat{q} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{r} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$. Prove that these are, respectively, consistent estimators of p, q and r.
   b) Show that the vector $(\hat{p}, \hat{q}, \hat{r})$ is asymptotically normal and find the asymptotic covariance matrix.
   c) Using the previous question combined with the Delta-method, prove that

   $$\sqrt{n}\left(\hat{r} - \hat{p}\hat{q} - (r - pq)\right) \xrightarrow[(d)]{n\to\infty} \mathcal{N}(0, V),$$

   where V depends on p, q and r.
   d) Consider the following hypotheses:

   $$H_0 : \text{"}X \text{ and } Y \text{ are independent"} \text{ vs. } H_1 : \text{"}X \text{ and } Y \text{ are not independent"}.$$

   Assuming that H_0 is true, show that $V = pq(1-p)(1-q)$ and propose a consistent estimator of V.
   e) Using the last two questions, propose a test with asymptotic level α, for any $\alpha \in (0, 1)$.

   We would like to know whether the facts of being happy and being in a relationship are independent of each other. In a given population, 1000 people (aged at least 21 years old) are sampled and asked two questions: "Do you consider yourself as happy ?" and "Are you involved in a relationship ?". The answers are summarized in the following table:

	Happy	Not happy
In a relationship	205	301
Not in a relationship	179	315

Would you reject independence of being happy and being in a relationship, with asymptotic level 5% ? Compute the p-value of your test.