Quiz 5 occurs November 9, in the discussion section. The quiz will be based upon the problems below.

Quiz 5 Problems

Exercise 1. Suppose X is a Gaussian distributed random variable with known variance $\sigma^2 > 0$ but unknown mean. Fix $\mu_0, \mu_1 \in \mathbf{R}$. Assume that $\mu_0 - \mu_1 > 0$. We want to test the hypothesis H_0 that $\mu = \mu_0$ versus the hypothesis H_1 that $\mu = \mu_1$. Fix $\alpha \in (0,1)$. Explicitly describe the UMP test for the class of tests whose significance level is at most α.

Your description of the test should use the function $\Phi(t) := \int_{-\infty}^{t} e^{-x^2/2}dx/\sqrt{2\pi}$, $\Phi\colon \mathbf{R} \to (0,1)$, and/or the function $\Phi^{-1}\colon (0,1) \to \mathbf{R}$. (Recall that $\Phi(\Phi^{-1}(s)) = s$ for all $s \in (0,1)$ and $\Phi^{-1}(\Phi(t)) = t$ for all $t \in \mathbf{R}$.)

Exercise 2. Let $X_1, \ldots, X_n$ be i.i.d. random variables, and denote $X = (X_1, \ldots, X_n)$. Let $\{f_\theta \colon \theta \in \Theta\}$ be a family of multivariable PDFs. Assume that X has distribution f_θ. Suppose Θ consists of two points, i.e. $\Theta = \{\theta_0, \theta_1\}$. Let Z be a sufficient statistic for θ. Consider the likelihood ratio test of the null hypothesis H_0 that $\theta = \theta_0$ versus the alternative H_1 that $\theta = \theta_1$.

Show that the likelihood ratio is a function of Z.

Exercise 3. Suppose X is a binomial distributed random variable with parameters $n = 100$ and $\theta \in [0,1]$ where θ is unknown. Suppose we want to test the hypothesis H_0 that $\theta = 1/2$ versus the hypothesis H_1 that $\theta \neq 1/2$. Consider the hypothesis test that rejects the null hypothesis if and only if $|X - 50| > 10$.

Using e.g. the central limit theorem, do the following:

- Give an approximation to the significance level α of this hypothesis test
- Plot an approximation of the power function $\beta(\theta)$ as a function of θ.
- Estimate p-values for this test when $X = 50$, and also when $X = 70$ or $X = 90$.

Exercise 4. Suppose $X_1, \ldots, X_n$ is a random sample from a Gaussian random variable X with unknown mean $\mu_X \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. Suppose $Y_1, \ldots, Y_m$ is a random sample from a Gaussian random variable Y with unknown mean $\mu_Y \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$.

Assume that $X_1, \ldots, X_n$ is independent of $Y_1, \ldots, Y_m$, i.e. assume that X, Y are independent.

Assume that $n + m > 2$. Define

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \overline{Y} := \frac{1}{m}\sum_{i=1}^{m} Y_i,$$

$$S_X^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad S_Y^2 := \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2,$$
$$S^2 := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}.$$

Show that

$$\frac{\overline{X} - \overline{Y} - \mu_X + \mu_Y}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

has Student's t-distribution with $n+m-2$ degrees of freedom. Deduce the following confidence intervals for the difference of the means

$$\mathbf{P}\Big(\overline{X} - \overline{Y} - tS\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + tS\sqrt{\frac{1}{n} + \frac{1}{m}}\Big)$$
$$= \frac{\Gamma(\frac{p+1}{2})}{\sqrt{p}\sqrt{\pi}\Gamma(p/2)}\int_{-t}^{t}\Big(1 + \frac{s^2}{p}\Big)^{-(p+1)/2} ds,$$

where $p = n+m-2$.

Exercise 5. Suppose you have a random sample of size 6 from a Gaussian random variable with unknown mean $\mu \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$. Suppose this random sample is

$$3,\ 4,\ 5,\ 5,\ 6,\ 7.$$

Explicitly construct a 95% confidence interval for the variance $\sigma^2 > 0$.

Your final answer might depend on the function $\Phi(t) := \int_{-\infty}^{t} e^{-x^2/2}dx/\sqrt{2\pi}$, $\Phi\colon \mathbf{R} \to (0,1)$, and/or $\Phi^{-1}\colon (0,1) \to \mathbf{R}$, and/or the $c_p(t) := \int_0^t \frac{x^{p/2-1}e^{-x/2}}{2^{p/2}\Gamma(p/2)}dx$ and/or c_p^{-1}, and/or the corresponding function for Student's t-distribution.

You should not need to use a central limit theorem.

Exercise 6. Suppose you have a random sample of size 3 from a Gaussian random variable with unknown mean $\mu_X \in \mathbf{R}$ and variance 2. Suppose this random sample is

$$1,\ 2,\ 3.$$

Suppose you have another random sample of size 3 from another Gaussian random variable with unknown mean $\mu_Y \in \mathbf{R}$ and variance 3. Suppose this random sample is

$$3,\ 4,\ 5.$$

Suppose all these random samples are independent of each other.

Explicitly construct a 99% confidence interval for the difference $\mu_X - \mu_Y$.

Your final answer might depend on the function $\Phi(t) := \int_{-\infty}^{t} e^{-x^2/2}dx/\sqrt{2\pi}$, $\Phi\colon \mathbf{R} \to (0,1)$, and/or $\Phi^{-1}\colon (0,1) \to \mathbf{R}$.

You should not need to use a central limit theorem.