Quiz 5 occurs November 11, in the discussion section. The quiz will be based upon the problems below.

Quiz 5 Problems

Exercise 1. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with unknown location parameter $\theta > 0$, i.e. X_1 has density

$$g(x) := 1_{x \geq \theta} e^{-(x-\theta)}, \qquad \forall\, x \in \mathbf{R}.$$

Fix $\theta_0 \in \mathbf{R}$. Suppose we want to test that hypothesis H_0 that $\theta \leq \theta_0$ versus the alternative H_1 that $\theta > \theta_0$. That is, $\Theta = \mathbf{R}$, $\Theta_0 = \{\theta \in \mathbf{R} \colon \theta \leq \theta_0\}$ and $\Theta_0^c = \Theta_1 = \{\theta \in \mathbf{R} \colon \theta > \theta_0\}$.

- Explicitly describe the rejection region of the generalized likelihood ratio test for this hypothesis. (Hint: it might be easier to describe the region using $x_{(1)} = \min(x_1, \ldots, x_n)$.)
- (Optional) If H_0 is true, then does
$$2 \log \frac{\sup_{\theta \in \Theta} f_\theta(X_1, \ldots, X_n)}{\sup_{\theta \in \Theta_0} f_\theta(X_1, \ldots, X_n)}$$
converge in distribution to a chi-squared distribution as $n \to \infty$?

Exercise 2. Let $X_1, \ldots, X_n$ be a random sample from a Gaussian random variable with unknown mean $\mu \in \mathbf{R}$ and unknown variance $\sigma^2 > 0$.

Fix $\mu_0 \in \mathbf{R}$. Suppose we want to test that hypothesis H_0 that $\mu = \mu_0$ versus the alternative H_1 that $\mu \neq \mu_0$.

- Explicitly describe the rejection region of the generalized likelihood ratio test for this hypothesis.
- Give an explicit formula for the p-value of this hypothesis test. (Hint: If S^2 denotes the sample variance and $\overline{X}$ denotes the sample mean, you should then be able to use the statistic $\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$. Since we have an explicit formula for Student's t-distribution, you should then be able to write an explicit integral formula for the p-value of this test.)

Exercise 3. Write down the generalized likelihood ratio estimate for the following alpha particle data, as we did in class for a slightly different data set. The corresponding test treats individual counts of alpha particles as independent Poisson random variables, versus the alternative that the probability of a count appearing in each box of data is a sequence of nonnegative numbers that sum to one. (In doing so, you should need to compute a maximum likelihood estimate using a computer.)

Plot the MLE for the Poisson statistic (i.e. plot the denominator of the generalized likelihood ratio test statistic $\frac{\sup_{\theta \in \Theta} f_\theta(X)}{\sup_{\theta \in \Theta_0} f_\theta(X)}$) as a function of λ.

m	$0,1$ or 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	≥ 17
# of Intervals	16	26	58	102	125	146	163	164	120	100	72	54	20	12	10	4

Finally, compute the value s of Pearson's chi-squared statistic S, and compute the probability that $S \geq s$ (assuming H_0 holds). Does the probability $\mathbf{P}(S \geq s)$ give you confidence that the null hypothesis is true?

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

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

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

$$S^2 := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}.$$

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

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

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.