Stat 111 Spring 2011 Week 5: Hypothesis testing

1. **Neyman Pearson Tests.**

   a) Describe the Neyman-Pearson parigdym for testing two simple hypotheses.

   b) Explain how a Bayesian would decide between θ_o and θ_1, given prior probabilities on the two values, and how this decision rule is similar to the NP test.

   c) Define the p-value for a test, and demostrate how a posterior probability for H_o may be very different from the p-value. For example, a test for Lyme disease has probability $\alpha = 0.05$ of a false-positive and $\beta = 0.1$ of a false-negative. The CDC claims a deer tick bite has probability $p = 0.01$ of transmitting Lyme. You could generalize this problem by allowing arbitrary values of α, β and p. Give the p-value for a test of H_o: no Lyme vs. H_a: Lyme, and the posterior probability of H_o, given that a person with a deer tick bite tests positive for Lyme.

   d) Imagine testing $H_o : \theta = \theta_o$ vs. $H_a : \theta = \theta_1$ (assume $\theta_1 > \theta_o$) based on a sample of n Uniform$(0, \theta)$ draws. For a single draw ($n = 1$) define a test of level $\alpha = 0.05$. What is the power of this test (as a function of θ_o and θ_1)? If θ_1 were considerably larger than θ_o, you might prefer a test with slightly lower power. Explain.

   e) For an arbitrary n, define a test of the hypotheses in part c with $\alpha = 0.05$. Assuming $\theta_o = 100$, graph the power of this test as a function of n when $\theta_1 = 105$, and as a function of θ_1 when $n = 12$.

   f) For the graph in part d with $n = 12$, add a line for the power of a test that rejects for a large value of $\bar{X}$ (use the Normal approximation to the average of n Uniforms). Show that this power is lower for all values of $\theta_1 > \theta_o$, meaning it is *inadmissible*.

2. **Uniformly Most Powerful Tests.**

   a) Explain why the likelihood ratio test is uniformly most powerful for 1-sided tests, but not for 2-sided tests. Consider a test of $H_o : \theta = \theta_o$ vs. $\theta > \theta_o$ based on a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$ (for example, the X_i's could be the lifetimes of a sample of electrical components). Derive a likelihood ratio test with $\alpha = 0.05$ and find the rejection region when $n = 10$ and $\theta_o = 0.1$. Argue that the Neyman Pearson lemma implies that this test is most powerful.

   b) Setting $\alpha = 0.05$, compare the test in part a to a test that rejects for small values of $X_{(1)}$, the smallest X_i (e.g., the shortest time to failure). With $\theta_o = 0.1$, graph the power for both tests for values of θ ranging from 0.1 to 0.2.

   c) Find the rejection region for the most powerful test of $H_o : \theta = 0.1$ vs. $H_a : \theta < 0.1$. Explain why combining this rejection region with the region in part a does not give the uniformly most powerful test of $H_o : \theta = 0.1$ vs. $H_a : \theta \neq 0.1$ for $\alpha = 0.1$. Give an example of a test that will have more power against some values of θ.

d) Based on the 2-sided test described in part d with $n = 10$, find a 90% confidence interval for θ when the MLE is $\hat{\theta} = 0.16$. What are the 1 and 2-sided p-values for your test?

e) Given independent Normal samples of sizes n_1 and n_2, show how to construct a 95% CI for σ_1/σ_2. Use the method of "pivoting." For example, in my Stat 1 class survey this Spring, the standard deviation for the heights of the $n_1 = 17$ female students was $s_1 = 2.7$ inches, and for the $n = 5$ male students it was $s_2 = 3.3$ inches.

3. **Generalized Likelihood Ratio Tests.**

a) Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$. Show that the Generalized Likelihood Ratio (GLR) test for $H_o : \mu = \mu_o$ vs. $H_a : \mu \neq \mu_o$ rejects for large values of $|t|$, where t is the usual 1-sample t-statistic. Explain why we don't use the null estimate of μ when computing the standard deviation s for our t statistic.

b) With $n = 16$ and $\mu_o = 10$, define rejection regions for your test with $\alpha = 0.05$. Use **pt** in R (the CDF function for the t distribution) to find the power of the test when $\mu = 8$ and $\sigma = 4$. Graph the power as a function of μ.

c) Theorem A on p. 341 states that for large samples, the distribution of $-2 \log \Lambda$ is approximately $\chi^2_{(\eta)}$, where η is the difference in the dimension of the null and alternative parameter spaces. Example A on p. 339 shows that this is exactly true for testing a Normal mean with known variance. Show that it is also true in a more general Normal problem: $H_o : \mu_1 = \mu_{o1}, \mu_2 = \mu_{o2}, \ldots, \mu_k = \mu_{ok}$ vs. H_a : "not H_o", based on $\mathrm{N}(\mu_j, 1)$ samples of size n_j, $j = 1, \ldots, k$.

d) The result from Theorem A applies to all distributions in the *Exponential Family*, but may fail when working with other distributions. For example, consider two samples of Uniform observations, one with upper bound θ_1 and one with upper bound θ_2. Find the null sampling distribution of $-2\log(\Lambda)$ for $H_o : \theta_1 = \theta_{o1}, \theta_2 = \theta_{o2}$ vs. H_a : "not H_o."

e) In the previous setting, find the GLR test of $H_o : \theta_1 = \theta_2$ vs. $H_a : \theta_1 \neq \theta_2$. For example, I took a Uniform samples of size $n = 20$ using R and JMP. For *R*, the largest value was $X_1 = 0.9897$ and for JMP it was $X_2 = 0.9938$ (in theory, $\theta_1 = \theta_2 = 1.0000$). Find a 90% CI for θ_1/θ_2. Note that you will need to derive the conditional distribution of X_1/X_2, given that $X_1 < X_2$.

Problems to turn in:

1. Consider a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \lambda)$.

a) Write out the likelihood function for α and λ. Show that $\bar{X}$ and $\bar{X}_{\text{G}} = (\prod X_i)^{1/n}$ are jointly sufficient for α and λ.

b) If α is known, find the MLE for λ and the asymptotic variance of $\hat{\lambda}$. Compare this to the exact variance of $\hat{\lambda}$. See Section 8.5.2 of Rice.

c) If both α and λ are unknown, find method of moments estimators for these parameters. See Section 8.4 of Rice.

2. Suppose you observe $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bin}(1, \theta)$.

a) Write out the likelihood function and log-likelihood function for θ, and show using the factorization theorem that $S = \sum X_i$ is a sufficient statistic for θ. Graph the likelihood function for $n = 20$ and $s = 5$, and for $n = 100$ and $s = 25$.

b) Find the conditional distribution of $X_1, \ldots, X_n$, given $S = s$, and verify that it does not depend on θ.

c) If you assume $\theta \sim U(0, 1)$, what is the marginal distribution of $S = \sum X_i$? What is the conditional distribution of θ given $S = s$?

d) The Uniform density is also a Beta$(1, 1)$ density. Work out the posterior distribution for θ when the prior distribution is Beta(a, b). Show that the posterior mean is a weighted average of the MLE and the mean of the prior distribution.

e) You want to estimate a player's probability θ of making a free-throw. You are pretty sure θ is larger than 0.4, and that it isn't too close to 1, and settle on $\theta \sim$ Beta$(4, 2)$. After $n = 10$ free-throws, the player made $x = 5$ shots. After $n = 100$ tries, the player made $x = 50$ shots. Find the posterior mean and standard deviation in each case ($n = 10$ and $n = 100$).

f) Graph the prior density, likelihood function, and posterior density on the same axis for each n value. Here is some R code to help you make the graphs. You'll need to add/modify commands to include the posterior density.

```
# create a vector of theta values ranging from 0 to 1 in steps of 0.01.
n=10;  x=5;  theta = seq(0, 1, 0.01)
ptheta = dbeta(theta, 4, 2)
Ltheta = dbeta(theta, x+1, n-x+1)
# note that the likelihood function is proportional to a Beta density.
plot(theta, Ltheta, type="l", xlab = "theta", ylab="")
lines(theta, ptheta, col="red")
# add lines to an existing graph.
legend(0,2, c("prior", "likelihood"), fill=c("red","black"))
# position a legend with the upper-left corner at (0,2).
```