Homework 09: Due 11/15

Stat061-F23

Prof Amanda Luby

---

1. Suppose we observe $X_1, ..., X_n \sim \text{Bernoulli}(p)$ and we wish to test the hypothesis $H_0 : p = p_0$ against $H_1 : p \neq p_0$.

   (a) What is the sampling distribution of $\hat{p}$?
   (b) What is the sampling distribution of $\hat{p}$ under H_0?
   (c) Show that the acceptance region (the set of test statistics where we fail to reject H_0) corresponds to a $(1 - \alpha)$ confidence interval for $\hat{p}$.

2. Let $Y_1, ..., Y_n \sim N(\mu, 1)$. (*Note:* since σ is known, you should not need the t-distribution)

   (a) Derive the form of the GLRT for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.
   (b) Now, find the GLRT if $H_1 : \mu = \mu_1$. How does the rejection region depend on μ_1?

3. Let $X_1, ..., X_n \sim \text{Exp}(\theta)$. Suppose that we wish to test $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$. Let $X = \sum X_i$ and let δ_c be the test that rejects H_0 if $X \geq c$.

   (a) Find $\pi(\theta|\delta_c)$ and argue that it is a decreasing function of θ.
   (b) Find c such that δ_c has size α_0
   (c) Let $\theta_0 = 2, n = 1$, and $\alpha_0 = 0.1$. Find the precise form of δ_c and sketch its power function.

4. The CLT tells us that, if our sample size is large enough, the sample means should be approximately normally distributed. In Lab05, we also showed that if our sample size was large enough, we could use the t-distribution instead of the normal distribution. However, in Lab05 we were drawing our samples from a normal population model, while the CLT should hold for *any* population model.

   (a) Using the web app at https://gallery.shinyapps.io/CLT_mean/, convince yourself that the CLT holds for normal, uniform, and skewed population data.
   (b) Using skewed population data, try different "levels" of skew and see how that impacts the sampling distribution of $\bar{X}$ for different sample sizes.
   (c) The app above generates an actual population distribution, and so is able to use σ instead of s for the SE of $\bar{X}$. Using the app, find a given level of skew and sample size of at least 5 that seems to *not* work super well for the CLT when the number of samples is large. Then, run your own simulation that generates the sampling distribution using s instead of σ. Create a histogram and overlay the appropriate theoretical t distribution. What do you conclude? *Note:* For the right-skewed data in the app, Low skew comes from a `log-normal(meanlog = 0, sdlog = .25)` distribution, medium skew is from a `log-normal(meanlog = 0, sdlog = .5)` distribution, and high skew comes from a `log-normal(meanlog = 0, sdlog = 1)` distribution.