Statistics 245: Homework 1 due April 11

When solving the problems below as well as future homework problems, give detailed derivations and arguments in order to receive credit. In your solution do not forget to include your name and the homework number. Please staple your pages together.

1. *(Approximate confidence intervals for Poisson distribution)* Let $X_1, \ldots, X_n$ be independent random variables distributed according to a Poisson(λ) distribution. Then the MLE of λ is $\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, and the two r.v.

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \quad \text{and} \quad \frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}/n}} \tag{1}$$

   both have approximately a standard normal distribution $N(0, 1)$ for large n. Using the "pivotal method" derive two approximate confidence intervals for λ. What are the interval midpoints? Are the intervals guaranteed to comprise only nonnegative numbers? Explain.

   The confidence interval using the first pivotal can be found on a lecture note posted on chalk. You can check if my calculation is correct.

   For the two confidence intervals you obtained above, design a simulation study and find which one gives a more accurate coverage probability. You can try, for example, $n = 30$, $\alpha = 0.05$ and $\lambda = 1$. You can use software `R`.

2. *(Sample size determination)* Let X follow a Binomial(n, p) distribution, and let $\hat{p} = X/n$ be the maximum likelihood estimator of the success probability p. Recall that the "Wald" $(1-\alpha)100\%$-confidence interval for p is of the form

$$[\hat{L}, \hat{U}] = \left[\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right].$$

   For $\alpha = 0.05$, find the smallest integer $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the confidence interval has length $\hat{U} - \hat{L} \leq 0.06$ regardless of the value of $\hat{p} \in [0, 1]$.

3. *(Approximate confidence intervals for Binomial distribution)* Let X have a Binomial(n, p) distribution, and let $\hat{p} = X/n$ be the maximum likelihood estimator of the success probability p. In class we discussed two approximate confidence intervals for p, namely,

   - the "Wald"-interval $[\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}]$ (also discussed in problem 2);
   - the interval due to Wilson (1927),

$$\left[\frac{\hat{p} + \frac{z^2}{2n} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}z^2 + \frac{z^4}{4n^2}}}{1 + z^2/n}\right], \qquad z = z_{1-\alpha/2}.$$

(a) For $n = 30$, $p = 0.1$, and confidence level $\alpha = 0.05$, compare the "Wald"- and the "Wilson"-confidence interval in simulations using the software R. More precisely, use the command `rbinom` to generate 100 draws from the Binomial(n, p) distribution. For each one of the 100 draws compute the three approximate confidence intervals (`asin` is the R-command for the arcsine). What proportion of the confidence intervals would we expect to contain $p = 0.1$ if the approximations are good? For each interval type, which proportion of the confidence intervals actually contains $p = 0.1$?

(b) Repeat the simulations for increased sample size $n = 150$. Do your conclusions change?

4. *(Distribution of a ratio)* Show that if X_1 and X_2 are independent exponential random variables with parameter $\lambda = 1$, then X_1/X_2 follows an F distribution. Also, identify the degrees of freedom.

   (Hint: With tools discussed in class you can first find the joint distribution of $(X_1/X_2, X_2)$ and then determine the marginal distribution of X_1/X_2 by integration.)

5. Do problems 16, 17 and 18 in page 241.

6. Let $X_1, \ldots, X_n$ be iid distributed as $N(\mu, \mu^2)$, where $\mu \in \mathbb{R}$ is an unknown parameters.

   (a) Find pivotal(s) for μ.

   (b) Let $\hat{\mu}$ be the MLE of μ. Find a function g such that $\sqrt{n}[g(\hat{\mu}) - g(\mu)] \Rightarrow N(0, 1)$.

   (c) Comment on confidence intervals for μ^2 constructed based on (a) and (b). Which one has smaller length?

7. (Optional) Read the article "Confidence interval" in Wikipedia.