STAT430 Homework #6: Due Friday, March 25, 2022.

Name: KEY

Submit this homework as a pdf file to Canvas.

Question 1

Let $Y_1, ..., Y_n \overset{\text{iid}}{\sim} \text{Exp}(\text{mean} = \theta)$. Note that $E[Y^k] = \theta^k(k!)$ for $k = 1, 2, ...$ Consider estimators

$$\widehat{\theta}_1 = \overline{Y}, \quad \widehat{\theta}_2 = \left(\frac{1}{2n}\sum_{i=1}^{n} Y_i^2\right)^{1/2}, \quad \widehat{\theta}_3 = \left(\frac{n}{n-1}\right)\left(\frac{1}{2n}\sum_{i=1}^{n} Y_i^2\right)^{1/2}.$$

Which of these estimators are consistent for θ?

Short answer: All three estimators are consistent estimators for θ

Shout-out to Kara for suggesting using the properties of convergence in probability. It's easy to think these problems more complicated than they are and harder to take a step back. Convergence in probability leads to the following properties:

1. If $X_n \xrightarrow{p} a$ and $Y_n \xrightarrow{p} b$, then $X_n + Y_n \xrightarrow{p} a + b$.
2. If $X_n \xrightarrow{p} a$ and $Y_n \xrightarrow{p} b$, then $X_n Y_n \xrightarrow{p} ab$.
3. If $X_n \xrightarrow{p} a$ and g is a continuous function at a, then $g(X_n) \xrightarrow{p} g(a)$.

Recall that consistency of an estimator $\hat{\theta}$ means that the estimator converges in probability to its expected value. In particular, if $\hat{\theta}$ is an unbiased estimator for θ, then

$$\lim_{n\to\infty} P\left(|\hat{\theta} - \theta| \geq \epsilon\right) = 1.$$

Okay let's actually start now. It's straightforward to show that $\hat{\theta}_1$ is consistent, using Theorem 9.1 in the book. We just need to show that it is unbiased (it is), and that the limit of the variance goes to zero as n goes to ∞. You can do that.

For $\hat{\theta}_2 = \left(\frac{1}{2n}\sum_{i=1}^{n} Y_i^2\right)^{1/2}$, it is a little less straightforward, but not as hard as I made it seem in office hours.
Consider the estimator $\hat{\theta}_2{}' = \frac{1}{2n}\sum_{i=1}^{\infty} Y_i^2$. Utilizing the provided formula for the moments of X_i, the expectation is

$$E\left[\frac{1}{2n}\sum_{i=1}^{\infty} Y_i^2\right] = \frac{1}{2n}\sum_{i=1}^{\infty} EY_i^2 = \frac{1}{2n}\sum_{i=1}^{\infty} 2\theta^2 = \theta^2.$$

So $\hat{\theta}_2{}'$ is an unbiased estimator of θ^2. Next, the variance of $\hat{\theta}_2{}'$ is

$$V\left[\frac{1}{2n}\sum_{i=1}^{\infty} Y_i^2\right] = \frac{1}{4n^2}\sum_{i=1}^{\infty}(EX_i^4 - (EX_i^2)^2) = \frac{n(4!\theta^4 - 4\theta^4)}{4n^2} = \frac{20\theta^4}{4n}.$$

This goes to zero as $n \to \infty$. Therefore, $\hat{\theta}_2{}'$ is a consistent estimator for θ^2. By definition, this means $\hat{\theta}_2{}' \xrightarrow{p} \theta^2$ and by the third property listed above, since $\sqrt{x}$ is a continuous function for $x > 0$, $\sqrt{\hat{\theta}_2{}'} = \hat{\theta}_2 \xrightarrow{p} \theta$. So $\hat{\theta}_2$ is a consistent estimator for θ.

Using the fact that $\hat{\theta}_2'$ is consistent, it can be seen that $\hat{\theta}_3$ is also a consistent estimator for θ. This is again applying the third property, where $g(x) = \frac{n}{n-1}\sqrt{x}$, which is continuous for $x > 0$ and also using the fact that $\lim_{n\to\infty} \frac{n}{n-1} = 1$.

Question 2

Prove that if $Y_n \sim \text{Binomial}(n, p)$ and $\hat{p}_n = Y_n/n$,

$$\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1-\hat{p}_n)/n}} \to_d N(0,1).$$

Answer: As was stated in class, convergence in probability is stronger than convergence in distribution. We know, by the Weak Law of Large Numbers, that $\hat{p}_n = \frac{1}{n}\sum_{i=1}^n Y_i$ is a consistent estimator for $p = EY_i$, and by the CLT $\hat{p}_n$ is approximately normally distributed with mean p and variance $\frac{p(1-p)}{q}$.

Question 3

The odds of success (for, e.g., a Bernoulli experiment with probability p) are defined as

$$\frac{\text{probability of success}}{\text{probability of failure}} = \frac{p}{1-p}$$

For reasons we won't get into, it is common to want to estimate the log-odds,

$$\log\left(\frac{p}{1-p}\right)$$

(recalling that here, as in almost every case in statistics, "log" refers to natural log, rather than $\log_{10}$).

Suppose that $W_n \sim \text{Binomial}(n, p)$, so that $\hat{p}_n = W_n/n$.

a) We can always express an Binomial random variable as the sum of iid Bernoulli random variables, so that $W_n = \sum_{i=1}^n X_i$, where $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$. Then $\hat{p}_1 = X_1$.

If $Y = \log\left(\frac{X_1}{1-X_1}\right)$, use the first-order Taylor series technique to approximate $\text{E}Y$ and $\text{Var}Y$.

Answer:

---

b) Use the fact that you proved in Question 2 and the delta method to construct an approximate $(1-\alpha)\times 100\%$ confidence interval for the log-odds based on $\hat{p}_n$.

Answer:

c) The *British Medical Journal* reported that, for 114 patients with spondyloarthropathies (a kind of joint disease), 54 of the patients had the ABO secretor state (a genetic feature). Use your computed confidence interval technique from part (b) to report a 95% confidence interval for the log odds of a patient with spondyloarthropathies having the ABO secretor state.

Answer: