1 True or False

Problem 1.1. Let $Y_n = \max\{X_1, ..., X_n\}$ with $X_i \sim$ Uniform $([0,1])$. Does $Y_n \xrightarrow{p} 1$?

Solution. Yes. This is because

$$\begin{aligned} P(|Y_n - 1| > \epsilon) &= P(1 - Y_n > \epsilon) \\ &= P(X_i < 1 - \epsilon)^n \\ &= (1 - \epsilon)^n \to 0 \end{aligned}$$

which is our definition of convergence in probability. ■

Problem 1.2. Let

$$Z_n = \begin{cases} 0 & \text{if } Y_n < 1 - \frac{1}{n} \\ 1 & \text{otherwise} \end{cases}$$

Does $Z_n \xrightarrow{p} 1$?

Solution. Nope. Suppose by contradiction we do. Then we also should have:

$$P(|Z_n - 1| > 0.1) \to 0 \text{ as } n \to \infty$$

However:

$$\begin{aligned} P(|Z_n - 1| > 0.1) &\geq P(Z_n = 0) \\ &= P\left(Y_n < 1 - \frac{1}{n}\right) \\ &= P\left(X_i < 1 - \frac{1}{n}\right)^n \\ &= \left(1 - \frac{1}{n}\right)^n \to \frac{1}{e} > 0 \end{aligned}$$

which is a contradiction. ■

2 True or False

Problem 2.1. Does a discrete approximation of a uniform distribution always converge in probability to the uniform distribution? Prove or provide a counterexample.

Solution. Nope. The counterexample is the following. Let $X \sim U([0,1])$ and $Y = 1 - X$. Consider the following approximation:

$$X_n = \begin{cases} \frac{1}{n} & \text{if } 0 \leq Y \leq \frac{1}{n} \\ \frac{2}{n} & \text{if } \frac{1}{n} < Y \leq \frac{2}{n} \\ & \vdots \\ 1 & \text{if } \frac{n-1}{n} < Y \leq 1 \end{cases}$$

But in this case,

$$P\{|X - X_n| > 0.1\} = P\{X \geq 0.9\} = 0.1$$

One of the possible sufficient conditions is that

$$X_n = \begin{cases} \frac{1}{n} & \text{if } 0 \leq X \leq \frac{1}{n} \\ \frac{2}{n} & \text{if } \frac{1}{n} < X \leq \frac{2}{n} \\ & \vdots \\ 1 & \text{if } \frac{n-1}{n} < X \leq 1 \end{cases}$$

in which case the convergence in probability goes through. ■

3 Bernoulli Trials

Let $X_1, \ldots, X_n$ denote the number of Bernoulli trials with p until the first success happens, i.e. Geometric(p):

$$P\left(X_i = k\right) = \left(1-p\right)^{k-1} p$$

Problem 3.1. Show that $\mathbb{E}\left[X_i\right] = 1/p$.

Solution. Note that

$$\begin{aligned} \mathbb{E}\left[X_i\right] &= \sum_{k=1}^{\infty} k\left(1-p\right)^{k-1} p \\ \left(1-p\right)\mathbb{E}\left[X_i\right] &= \sum_{k=1}^{\infty} k\left(1-p\right)^{k} p \end{aligned}$$

Subtracting one from the other:

$$\begin{aligned} p\mathbb{E}\left[X_i\right] &= \sum_{k=1}^{\infty}\left[k\left(1-p\right)^{k-1} p - k\left(1-p\right)^{k} p\right] \\ &= p\sum_{k=1}^{\infty}\left[k\left(1-p\right)^{k-1}\right] = p\frac{1}{1-\left(1-p\right)} = 1 \end{aligned}$$

which yields

$$\mathbb{E}\left[X_i\right] = \frac{1}{p}$$

■

Problem 3.2. Construct a consistent estimator $\hat{p}_n$ of p. Is it unbiased?

Solution. Consider

$$\hat{p}_n = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i}$$

Since X_is are independent,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}\left[X_i\right]$$

and the rest follows from the continuous mapping theorem. However, it is not necessarily unbiased. ■

Problem 3.3. Find τ_n, μ, and σ^2 such that $\tau_n\left(\bar{X}_n - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$.

Solution. By CLT, we have:

$$\sqrt{n}\left(\bar{X}_n - \mathbb{E}\left[X_i\right]\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{Var}\left[X_i\right] = \frac{1-p}{p^2}\right)$$

so

$$\tau_n = \sqrt{n}, \quad \mu = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}$$

■

Problem 3.4. Find the limiting distribution of your estimator $\hat{p}_n$.

Solution. Recall that we had:

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{p}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1-p}{p^2}\right)$$

Denote $f(x) = 1/x$ which yields $f'(x) = -1/x^2$. Applying the Delta Method, we have:

$$\begin{aligned}\sqrt{n}\left(f\left(\bar{X}_n\right) - f\left(\frac{1}{p}\right)\right) &\xrightarrow{d} \left[\frac{1}{(1/p)^2}\right]\mathcal{N}\left(0, \frac{1-p}{p^2}\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, p^2(1-p)\right)\end{aligned}$$

■

Problem 3.5. Construct a α test for the null hypothesis $H_0 : 0 \leq p \leq 0.5$ vs. $H_1 : 0.5 < p \leq 1$ that is consistent in level.

Solution. Recall that we had:

$$\sqrt{n}\left(\hat{p}_n - p\right) \xrightarrow{d} \mathcal{N}\left(0, p^2(1-p)\right)$$

which also implies:

$$\frac{\sqrt{n}\left(\hat{p}_n - p\right)}{\sqrt{\hat{p}_n\left(1-\hat{p}_n\right)}} \xrightarrow{d} \mathcal{N}(0,1)$$

Restricting our attention to tests of the form:

$$\phi_n = I\left\{T_n \geq c_n\right\}$$

we can set:

$$T_n = \frac{\sqrt{n}\left(\hat{p}_n - 0.5\right)}{\sqrt{\hat{p}_n\left(1-\hat{p}_n\right)}}, \quad c_n = z_{1-\alpha}$$

Note that this is consistent in level since for P satisfying the null hypothesis:

$$\begin{aligned}\mathbb{E}_P\left[\phi_n\right] &= P\left(T_n \geq c_n\right) \\ &= P\left(\frac{\sqrt{n}\left(\hat{p}_n - 0.5\right)}{\sqrt{\hat{p}_n\left(1-\hat{p}_n\right)}} \geq z_{1-\alpha}\right) \\ &= P\left(\frac{\sqrt{n}\left(\hat{p}_n - p\right)}{\sqrt{\hat{p}_n\left(1-\hat{p}_n\right)}} + \frac{\sqrt{n}\left(p - 0.5\right)}{\sqrt{\hat{p}_n\left(1-\hat{p}_n\right)}} \geq z_{1-\alpha}\right)\end{aligned}$$

Since under the null,

$$\frac{p - 0.5}{\sqrt{\hat{p}_n (1 - \hat{p}_n)}} \leq 0$$

and so:

$$\mathbb{E}_P [\phi_n] \leq P\left(\frac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{\hat{p}_n (1 - \hat{p}_n)}} \geq z_{1-\alpha}\right) \xrightarrow{d} \alpha$$

■

Problem 3.6. Construct a α test for the null hypothesis $H_0 : 0 \leq p \leq 0.5$ vs. $H_1 : 0.5 < p \leq 1$ that is consistent in level.

Solution. The p-value is defined as the smallest value of α at which we reject the hypothesis. Therefore:

$$\alpha = 1 - \Phi\left(\frac{\sqrt{n}\,(\hat{p}_n - p)}{\sqrt{\hat{p}_n (1 - \hat{p}_n)}}\right)$$

■

4 Confidence Intervals

Suppose $X_1, ..., X_n \sim iid$ where

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p-q \\ -1 & \text{w.p. } q \end{cases}$$

Problem 4.1. Provide a level $1-\alpha$ confidence interval for $p+q$ without using asymptotics.

Solution. We know that $\mathbb{E}\left[|X|\right] = p+q$ and $\mathbb{E}\left[|X|^2\right] = \mathbb{E}\left[X^2\right] = p+q$. This implies:

$$\text{Var}\left[|X|\right] = (p+q) - (p+q)^2 = (p+q)\left(1-(p+q)\right)$$

Therefore:

$$P\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|\right) - \mathbb{E}\left[|X|\right]\right| > \epsilon\right) \leq \frac{\text{Var}\left[|X|\right]}{\epsilon^2} = \frac{(p+q)\left(1-(p+q)\right)}{\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

where equality holds at $p+q = 1/2$. Therefore, let

$$\frac{1}{4n\epsilon^2} = \alpha \Leftrightarrow \epsilon = \frac{1}{2\sqrt{n\alpha}}$$

and construct the confidence interval as the following:

$$c_n = \left[\frac{1}{n}\sum_{i=1}^{n}|X_i| - \frac{1}{2\sqrt{n\alpha}}, \frac{1}{n}\sum_{i=1}^{n}|X_i| + \frac{1}{2\sqrt{n\alpha}}\right]$$

■

Problem 4.2. Provide a level $1-\alpha$ confidence interval for $p-q$ without using asymptotics.

Solution. We know that $\mathbb{E}\left[X\right] = p-q$ and $\mathbb{E}\left[|X|^2\right] = \mathbb{E}\left[X^2\right] = p+q$. This implies:

$$\text{Var}\left[|X|\right] = (p+q) - (p-q)^2$$

Therefore:

$$P\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) - \mathbb{E}\left[X_i\right]\right| > \epsilon\right) \leq \frac{\text{Var}\left[X_i\right]}{\epsilon^2} = \frac{(p+q)-(p-q)^2}{\epsilon^2} \leq \frac{1}{n\epsilon^2}$$

where equality holds at $p = q = 1/2$. Therefore, let

$$\frac{1}{n\epsilon^2} = \alpha \Leftrightarrow \epsilon = \frac{1}{\sqrt{n\alpha}}$$

and construct the confidence interval as the following:

$$c_n = \left[\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{\sqrt{n\alpha}}, \frac{1}{n}\sum_{i=1}^{n}X_i + \frac{1}{\sqrt{n\alpha}}\right]$$

■