Empirical Distribution, Kolmogorov-Smirnov Test

Fiji earthquakes

Data on the magnitudes of earthquakes near Fiji are on Blackboard under the Class-21 tab. Estimate the cdf $F(x)$. Compute and plot a 95% confidence envelope for F. Find an approximate 95 percent confidence interval for $F(4.9) - F(4.3)$.

Old Faithful

Data on eruption times and waiting times between eruptions of the old faithful geyser (located in Yellowstone National Park) are posted on Blackboard under the Class-21 tab. Estimate the mean waiting time and compute a standard error for the estimate. Also, calculate a 90 percent confidence interval for the mean waiting time. Finally, estimate the median waiting time. We will use this calculation of the median in our discussions next week.

KS problem

Use the Kolmogorov-Smirnov test to test the hypothesis that the 25 values in the table below form a random sample from the uniform distribution on the interval $[0, 1]$.

0.42	0.06	0.88	0.40	0.90
0.38	0.78	0.71	0.57	0.66
0.48	0.35	0.16	0.22	0.08
0.11	0.29	0.79	0.75	0.82
0.30	0.23	0.01	0.41	0.09

Using the table above, test the hypothesis that the 25 values are a random sample from a continuous distribution with pdf:

$$f(x) = \begin{cases} \frac{3}{2} & \text{for } 0 < x \leq \frac{1}{2} \\ \frac{1}{2} & \text{for } \frac{1}{2} < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

KS problem

In class on March 12, we began problem 2 from the Wasserman ecdf chapter. There are four parts to the problem. You've seen parts a and b. Complete problem 2 and then do problem 4.

2. ① $\hat{\delta} = \hat{p} - \hat{q} = \int x \, d\hat{F}_1(x) - \int x \, d\hat{F}_2(x)$

$$= \frac{\sum X_i}{n} - \frac{\sum Y_i}{m} = \hat{p} - \hat{q}$$

$\because V(\hat{p}) = \hat{p}(1-\hat{p}) \quad V(\hat{q}) = \hat{q}(1-\hat{q})$

$\therefore \hat{se}(\hat{\delta}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{q}(1-\hat{q})}{m}}$

$Z_{.05} = 1.64$

$\therefore$ the interval is $(\hat{\delta} - 1.64\,\hat{se},\ \hat{\delta} + 1.64\,\hat{se})$

4. $\because \hat{F}_n(x) = \dfrac{\Sigma I(X_i \le x)}{n}$,

where $I(X_i \le x) = \begin{cases} 1 & \text{if } X_i \le x \\ 0 & \text{if } X_i > x \end{cases}$

let $p = P(X_i \le x) = F'(x)$

then $I(X_i \le x) = \{Y_i\} \sim \text{Bernoulli}(p)$

$\because \hat{F}_n(x) = \dfrac{\Sigma Y_i}{n} = \bar{Y}_n$

$\therefore$ by CLT

$\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

$\therefore \sqrt{n}(\bar{Y}_n - p) \xrightarrow{d} N(0, p(1-p))$