9.07 Introduction to Statistics for Brain and Cognitive Sciences

Emery N. Brown

Homework Assignment 2
September 14, 2016
Due September 21, 2016 at 5:00pm

1. Suppose that X has the probability density function $f(x) = c(1-x^2)$ for $0 \le X \le 1$ and $f(x) = 0$ otherwise.

   A. Find c.
   B. Find the cumulative density function (cdf).
   C. Find $\Pr(0.1 \le X \le 0.9)$. Draw this area under the cdf.

   D. Find $x_{0.95}$ such that $\Pr(X \le x_{0.95}) = 0.95$. The value $x_{0.95}$ is a 95% upper confidence bound for X.

2. Dustin Pedroia of the Boston Red Sox has a batting average of 300. Interpret this as meaning that his probability of a hit is 0.3. A data analyst wishes to analyze his performance using a binomial model.

   A. In a typical game he bats 4 times. What is the probability that he gets 2 hits or more?

   B. In 4 at bats, how many hits is he expected to get?

   C. Recently, Pedroia got 11 hits in a row. What is the probability of this event? (The most number of hits ever in a row by a major league baseball player is $12!$)

3. The average number of purchases per second on the Amazon book website is 10. Assume that the number of purchases pr second is a Poisson random variable X.
   A. Compute the probability of there being 9, 10 or 11 purchases per second.
   B. Compute the probability of there being fewer than 2 purchases per second.
   C. Suppose that instead of knowing that the average number of purchases per second is 10, you are told that $\Pr(X = 0) = 0.082$. Find λ and $Var(X)$.

4. The cdf of the Rayleigh probability model is

$$F(x) = 1 - \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

for $x \in [0, \infty)$ and $\sigma > 0$.

A. Find the probability density of $F(x)$.
B. Find the upper 15th and lower 5th percentiles of the Rayleigh probability model.
C. Compute $\Pr(0.25 < X < 0.75)$.
D. Take $\sigma = 1$ and plot the pdf and the cdf.

5. 50% of the US adult population has hypertension (high blood pressure). Suppose that a new non-invasive test for diagnosing hypertension has been designed based on using heart rate variability along with blood pressure measurements. The new test will classify 25% of people with hypertension as not having hypertension and 15% of people without hypertension as having hypertension.

A. What is the prevalence of hypertension in the US adult population?
B. What is the false negative probability of the test?
C. What is the false positive probability of the test?
D. What is the sensitivity of the test?
E. What is the specificity of the test?
F. What is the predictive value negative of the test?
G. What is the predictive value positive of the test?

6. For the probability density in **Problem 1** find the inverse cdf F^{-1}. Using F^{-1} and **Algorithm 3.1**, simulate a random sample of 200 draws from the cdf $F(x)$. Construct the empirical cdf of the random sample. To do so, if the random sample is $x_1, x_2, ..., x_{200}$ arrange the sample from smallest to largest to obtain the ordered sample $x_{(1)}, x_{(2)}, ..., x_{(200)}$. The empirical cdf $F_n(x)$ is defined as

$$F_n(x) = \frac{\text{number of } x_i \le x}{n}$$

or the terms of the ordered sample as

$$F_n(x) = \begin{cases} 0 & x < x_{(1)} \\ \dfrac{k}{n} & x_{(k)} \le x < x_{(k+1)} \\ 1 & x_{(n)} < x \end{cases}$$

where $n = 200$ At every observation, the empirical cdf has a jump of $\frac{1}{n}$.

Plot the empirical cdf of the random sample on the same graph with the cdf superimposed. Does the empirical cdf agree qualitatively with $F(x)$?

7. A fertility specialist is following a cohort of 5,000 families. The probability of triplets in human births is 0.001.

   A. What is an appropriate probability model for this problem? Explain.
   What is an efficient way to compute the probabilities for this model? Explain.

   B. What is the probability that there will be exactly one set of triplets among the members of the cohort?

   C. How many births would you expect from the cohort?

   D. The Canadian colleague of the fertility specialist is following a similar 5,000 family cohort. She has observed 10 triplet births in her cohort. Do you think the triplet birth probability is the same in the two countries? In other words, how likely would the 10 triplet births be assuming that the probability in the Canadian cohort is the same as in the U.S. cohort?

9.07 Statistics for Brain and Cognitive Science

Fall 2016