Worksheet 20

1. Throughout this worksheet, let $X_1, \dots, X_n$ be a sequence of n i.i.d. continous random variables that have a pdf $f(x)$ and a cdf $F(x)$. Define the random variables $Y_1, \dots, Y_n$ to be the corresponding order statistics, each with a pdf $(g_j(y))$ and cdfs $(G_j(y))$. Write down $G_n(y)$—the cdf of the maximum value—in terms of n and F. Hint: Write out the problem with probabilities before converting to the cdf.

2. In the next few questions, we will work on the density function $g_k(y)$ for an arbitrary k. To start, fix a value y and a positive value Δ. What is the joint probability that $X_1, \dots, X_{k-1}$ are all less than y, that $X_{k+1}, \dots, X_n$ are all greater than $y + \Delta$, and that X_k is in the interval $[y, y + \Delta]$?

3. We are back to another counting question! The probability you have in the previous question counts only one specific configuration of the values X_j that would result in Y_k being in the interval $[y, y+\Delta]$. In general, we could have any set of $k-1$ of the n random variables be less than y, one of the random variables be in the interval $[y, y + \Delta]$, and the rest of the $n - k$ be somewhere greater than $y + \Delta$. (a) How many different configurations are there? (b) What is the probability that Y_k is in the interval $[y, y + \Delta]$?[1]

[1] Technically you are computing the probability that Y_k is in this interval and Y_{k+1} is not. The difference between these will limit to zero in the next question.

4. One way, if it exists, to define the pdf of a random variable Y is:

$$f_Y(y) = \lim_{\Delta \to 0} \left[\frac{1}{\Delta} \times \mathbb{P}[Y \in [y, y + \Delta]] \right] = \lim_{\Delta \to 0} \left[\frac{F_Y(y + \Delta) - F_Y(y)}{\Delta} \right]$$

Where F_Y is the cdf. This comes directly from the fundamental theorem of calculus and the definition of the relationship between the cdf and the pdf. Use this to compute the pdf $g_k(y)$ of the k-th order statistic Y_k. Your answer should be in terms of factorials using only y, k, n, F and f.

5. Let's apply this definition to a special case. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} U(0,1)$. For any $y \in (0,1)$, write down a formula for $F(y)$. Hint: This is easy.

6. Now, write down pdf of the density function $g_k(y)$ for $y \in (0, 1)$ when the X_j's come from a standard uniform distribution? We will simplify this in the next question.

7. Recall Gamma function has the property that $\Gamma(n) = (n - 1)!$ for any integer n. Write your previous question in terms of the Gamma function.

8. Set $\alpha = k$ and $\beta = n - k + 1$ and plug into the solution from the previous question. What is the name for the distribution of the k-th order statistic Y_k from a set of independent random variables from the standard uniform distribution?

9. Let's end with a even more concrete example. Let $X_1, X_2, X_3, X_4 \overset{\text{i.i.d.}}{\sim} U(0,1)$. What are the expected values of the four order statistics Y_1, Y_2, Y_3, Y_4?