ENGRD 2700: Basic Engineering Probability and Statistics
Fall 2019

Homework 5

Due **Friday November 1** at 11:59pm. Submit to Gradescope by clicking the name of the assignment. See `https://people.orie.cornell.edu/yudong.chen/engrd2700_2019fa.html#homework` for detailed submission instructions.

The same stipulations from Homework 1 (e.g., independent work, computer code, etc.) still apply.

1. The joint density function of two random variables is given by

$$f_{X,Y}(x,y) = \begin{cases} k(2y+xy) & \text{if } 0 \leq x \leq 1 \text{ and } x \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) What is the constant k?

   (b) What is the marginal pdf of X?

   (c) What is the conditional density of X, given that $Y = 1/4$?

2. The lifetimes X and Y of two batteries (in years) are distributed according to the following joint PDF:

$$f_{X,Y}(x,y) = \begin{cases} 8e^{-4x}e^{-2y} & \text{if } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Compute the probability that both batteries last at least 2 years.

   (b) Find the marginal PDFs $f_X(x)$ and $f_Y(y)$.

   (c) Compute $E[XY]$.

   (d) What is $P(X \leq \frac{Y}{2})$, the probability that the first battery lasts at most half as long as the second?

   (e) Are X and Y independent? Why or why not?

3. For two discrete random variables X and Y, recall the definition of $f_{X|Y}(x|y)$, the *conditional PMF* of X given that $Y = y$. Also recall that conditional PMFs are still PMFs, so they must sum to one; that is, $\sum_x f_{X|Y}(x|y) = 1$.

   A bakery sells two types of cupcakes: red velvet and salted caramel. Let R and S denote the number of each type of cupcake an individual customer buys. Suppose that R and S are distributed according to the following joint PMF:

		s		
$f_{R,S}(r,s)$		0	1	2
	0	0	0.25	0.15
r	1	0.20	0.10	0.10
	2	0.10	0.05	0.05

   (a) Compute $f_{R|S}(1|1)$, the conditional probability that a customer buys 1 red velvet cupcake, given that (s)he buys 1 salted caramel cupcake. How does this probability compare to $f_R(1)$?

   (b) Find $f_{R|S}(r|2)$, the conditional PMF of R, given that the customer buys 2 salted caramel cupcakes. (This involves computing $f_{R|S}(0|2)$, $f_{R|S}(1|2)$, and $f_{R|S}(2|2)$.)

   (c) The *conditional expectation* of X, given that $Y = y$, is defined as

$$E[X \mid Y = y] = \sum_x x \, f_{X|Y}(x|y).$$

   Find $E[R \mid S = 2]$, the expected number of red velvet cupcakes a customer buys, given that (s)he buys 2 salted caramel cupcakes.

4. Given two continuous random variables X and Y, recall the definition of $f_{X|Y}(x|y)$, the *conditional PDF* of X given that $Y = y$. Conditional PDFs are still PDFs, so they must *integrate* to one: $\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = 1$.

   Suppose X and Y are distributed as in the following joint PDF:

   $$f_{X,Y}(x, y) = \begin{cases} 2ye^{-y(2+x)} & x, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

   For $y > 0$, compute $f_{X|Y}(x|y)$.

5. The file `Bwages.csv` contains hourly wages of 1473 randomly selected individuals living in Belgium in 1994. For the questions that follow, attach your plots, as well as any code used to generate them.

   (a) Import this dataset into RStudio, and generate a histogram of the data.

   (b) Construct a normal Q–Q plot, by hypothesizing that the data originate from a Normal($\bar{x}$, s^2) distribution, where $\bar{x}$ and s^2 are the sample mean and sample variance, respectively. Overlay the line $y = x$ onto your plot. (See Recitation 6 for an example of how to do this.)

   (c) Does the normal distribution appear to be a reasonable fit for the data? Why or why not?

   (d) We'll now attempt to fit a *lognormal distribution* to the data. If $X \sim$ Lognormal(μ, σ^2), it has PDF

   $$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/(2\sigma^2)} \qquad x > 0$$

   The lognormal random variable gets its name because $\ln(X)$ is normally distributed (with mean μ and variance σ^2). Construct a lognormal Q–Q plot, by hypothesizing that the data originate from the Lognormal(2.31, 0.41) distribution. Use R's `qlnorm` function to compute theoretical quantiles, and set `meanlog` to 2.31 and `sdlog` to 0.41. Overlay the line $y = x$ onto your plot.

   (e) Does the lognormal distribution appear to be a better fit for the data? Where could the fit be improved? Comment on what you see.