Dr. Z.'s Probability Lecture 15 Handout: Independent random variables and their sum

By Doron Zeilberger

Version of Nov. 10, 2017 (Thanks to JINHUA XU)

Important Definition: Two random variables X and Y are **independent** if for any two sets A and B of real numbers

$$P[\,X \in A\,,\, Y \in B\,] = P[\,X \in A\,] \cdot P[\,Y \in B\,]$$

Obvious but important properties of independent random variables

1. If X and Y are independent discrete random variable, with prob. mass functions $p_X(x)$ and $p_Y(y)$ respectively, then the joint density function $p(x,y)$ is their product:

$$p(x,y) \;=\; p_X(x)p_Y(y) \quad .$$

2. If X and Y are independent continuous random variable, with density functions $f_X(x)$ and $f_Y(y)$ respectively, then the joint density function $f(x,y)$ is given by

$$f(x,y) = f_X(x)f_Y(y) \quad .$$

In other words, the joint density function is the **product** of the *marginal* distributions.

Problem 15.1: Two friends decide to meet at a certain restaurant. If each of them independently arrives at a time uniformly distributed between 1pm and $1:30$pm. Find the probability that the first to arrive has to wait longer than 7 minutes.

Sol. to problem 15.1: Since the arrival time (in minutes) after 1pm for Friend 1 and Friend 2 are *independent*, measured in minutes, and the density functions are

$$f_X(x) = \begin{cases} \frac{1}{30} \quad , & if \quad 0 \leq x \leq 30 \quad ; \\ 0, & otherwise \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{30} \quad , & if \quad 0 \leq y \leq 30 \quad ; \\ 0, & otherwise \end{cases}$$

The **joint** density function is their product, namely

$$f(x,y) = \begin{cases} \frac{1}{900} \quad , & if \quad 0 \leq x,y \leq 30 \quad ; \\ 0, & otherwise \end{cases} \quad .$$

The probability that Friend 1 has to wait longer than 7 minutes for Friend 2 is

$$P[\,X + 7 < Y\,] \quad .$$

This is given by the double integral

$$\int\int_{\substack{y-x>7\\0\le x,y\le 30}} f(x,y)\,dx\,dy$$

$$= \int\int_{\substack{y-x>7\\0\le x,y\le 30}} \frac{1}{900}\,dx\,dy$$

$$= \frac{1}{900}\int_7^{30}\int_0^{y-7} dx\,dy$$

$$= \frac{1}{900}\int_7^{30}(y-7)\,dy = \frac{1}{900}\frac{(y-7)^2}{2}\Big|_7^{30} = \frac{1}{900}\frac{23^2}{2} = \frac{529}{1800} \quad .$$

By symmetry the probability that Friend 2 has to wait longer than 7 minutes for Friend 1 is the same, so the final answer is twice that: $\frac{529}{900} = 0.5877777778\ldots$.

Ans. to 15.1: The probability that the first to arrive has to wait longer than 7 minutes is $\frac{529}{900} = 0.5877777778\ldots$.

Note: Since in this case we have the **uniform distribution**, we don't need calculus; A quicker way would be to draw the region

$$\{0 \le x \le 30 \quad , \quad 0 \le y \le 30 \quad , \quad y > x+7\} \quad ,$$

and note that it is a right-angled triangle with vertices $(0,7)$, $(0,30)$ and $(23,30)$ whose base has length 23 and height is 23, and from middle-school we know that the area of the triangle is $\frac{1}{2}\cdot 23^2$.

Warning: This shortcut only applies when the integrand is a constant, i.e. for uniform distributions.

Problem 15.2 Two friends decide to meet at a certain restaurant at 1:30pm. One of them, Mr. A, has a propensity to arrive earlier than the agreed time. The density function of his arrival time is

$$a(x) = \begin{cases} \frac{1}{20}, & if \quad -15 < x < 5\,; \\ 0, & otherwise \end{cases} \quad .$$

(x is measured in minutes)

The other friend, Mr. B, has a propensity to arrive later than the agreed time. The density function of his arrival time is

$$b(x) = \begin{cases} \frac{1}{20}, & if \quad -5 < x < 15\,; \\ 0, & otherwise \end{cases} \quad .$$

Find the probability that, surprisingly Mr. B arrived before Mr. A.

Comments: Some people got confused, since $b(x)$ is defined using the variable-name x. It would have been more straightforward to define the pdf b by

$$b(y) = \begin{cases} \frac{1}{20}, & if \quad -5 < y < 15\,; \\ 0, & otherwise \end{cases} \quad .$$

but you can use, as arguments any **variable name**, for exapmle, the following is equally valid

$$b(DonaldTrump) \;=\; \begin{cases} \frac{1}{20}, & if \quad -5 < (DonaldTrump < 15\,; \\ 0, & otherwise \end{cases} \quad .$$

Now $f(x,y)$ is $a(x)b(y)$ since this is a joint densitiy function of independent random variables. Once again x, y are place-holders, and you can use any names for them. But having decided on the names, you have to use x for the function a and y for the function b.

Sol. to 15.2: The **joint** density function is their product, namely

$$f(x,y) = \begin{cases} \frac{1}{400} \quad , & if \quad -15 < x < 5\,, -5 < y < 15 \quad ; \\ 0, & otherwise \end{cases} \quad .$$

We need $P[X > Y]$. By drawing a diagram, the region inside $\{-15 < x < 5\} \times \{-5 < y < 15\}$ for which $x > y$ is the right-angled triangle whose vertices are $(-5,-5)$,$(5,-5)$ and $(5,5)$. Its base has length 10 and its height is 10, hence its area is $10^2/2 = 50$. Multiplying by $\frac{1}{400}$, we get the desired probability is $\frac{50}{400} = \frac{1}{8}$.

Ans. to Problem 15.2: The probability that Mr. B will arrive before Mr. A. is $\frac{1}{8}$.

Important Fact: If X and Y are **independent** continuous random variables with density functions $f_X(x)$ and $f_Y(y)$ respectively, then the density function of their **sum**, $X+Y$, is given by the **convolution**

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy \quad .$$

Note: In practice this is very complicated to do by hand (except for the uniform distribution).

Problem 15.3: Let X and Y be independent continuous random variables with densities

$$f_X(x) = \begin{cases} e^{-x}\,, & if \quad x \geq 0\,; \\ 0 & otherwise. \end{cases} \qquad , \qquad f_Y(y) = \begin{cases} 3e^{-3y}\,, & if \quad y \geq 0\,; \\ 0 & otherwise. \end{cases} \quad .$$

Find the density function $f_Z(z)$ for $Z = X + Y$.

Sol. to 15.3:

$$f_Z(z) = \int_0^{\infty} f_X(z-y)\, f_Y(y)\, dy \quad .$$

But $f_X(z-y) = 0$ when $z - y < 0$, in other words, when $y > z$, so the integral can be written as

$$f_Z(z) = \int_0^{z} f_X(z-y)\, f_Y(y)\, dy \quad .$$

Hence

$$f_Z(z) \;=\; \int_0^z e^{-(z-y)}\, 3e^{-3y}\, dy = \int_0^z 3\, e^{-z+y-3y}\, dy = 3\, e^{-z} \int_0^z e^{-2y}\, dy = 3\, e^{-z} \frac{e^{-2y}}{-2}\Big|_0^z$$

$$= \frac{3\,e^{-z}}{2}(-e^{-2z} - -e^{0}) = \frac{3}{2}(e^{-z} - e^{-3z}) \quad .$$

Of course it is 0 when $z < 0$, so

Ans. to 15.3:

$$f_Z(z) = \begin{cases} \frac{3}{2}(e^{-z} - e^{-3z})\,, & if \quad z \geq 0\,; \\ 0 \quad otherwise. \end{cases} \quad .$$

Important Fact: If X and Y are **independent** Poisson distributions of parameters μ_1 and μ_2, then $X + Y$ is also a Poisson distribution, and its parameter is $\mu_1 + \mu_2$.

Note: This fact was already mentioned in Lecture 9.

Problem 15.4: The number of people who go to a certain diner for breakfast, lunch, and dinner are all Poisson distribution with paramaters 10, 20, and 5 respectively. Assuming that they are independent of each other. What is the probability that the total number of customers during a given day is 36.

Sol. of 15.4: The parameter of the total number of customers is $10 + 20 + 5 = 35$. Hence

$$P[NumberOfCustomers = 36] = e^{-35} \cdot \frac{35^{36}}{36!} = 0.06540449260\ldots\ldots \quad .$$

Ans. to 15.4: The probability that the total number of customers during a given day is exactly 36 equals $0.06540449260\ldots$.

Another Important Fact: If X and Y are **independent** Normal distributions with means μ_1, μ_2 respectively, and variances σ_1^2, σ_2^2 respectively, then $X + Y$ is yet another Normal distribution with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

More generally, if you have more than two independent random variables. The mean of the sum is the sum of the means and the variance of the sum is the sum of the variances.

Also recall that for any constant,c (and any distribution X, not just normal) $E[cX] = cE[X]$ and $Var(cX) = c^2 Var(X)$.

Problem 15.5: Let X_1, X_2, X_3 be **independent** Normal distributions with means $1, 3, 5$ respectively and standard-deviations $2, 3, 4$. What is the mean of their average $(X_1 + X_2 + X_3)/3$. What is its standard deviation?

Sol. of 15.5: Let's first handle $X_1 + X_2 + X_3$ (and later divide by 3). The mean of $X_1 + X_2 + X_3$ is $1 + 3 + 5 = 9$. The **variances** of X_1, X_2, X_3 are $4, 9, 16$, so the variance of $X_1 + X_2 + X_3$ is $4 + 9 + 16 = 29$. Hence the mean of $(X_1 + X_2 + X_3)/3$ is $\frac{9}{3} = 3$ and the its variance is $\frac{29}{9}$, and so its standard deviation is $\frac{\sqrt{29}}{3} = 1.795054935\ldots$.

Ans. to 15.5: The mean and standard deviation of $(X_1 + X_2 + X_3)/3$ are 3 and $\sqrt{\frac{29}{3}} = 1.795054935\ldots$ respectively.