Lecture 3

David Aldous

31 August 2015

This “size-bias” effect occurs in other contexts, such as class size.

If a small Department offers two courses, with enrollments 90 and 10, then

average class (faculty viewpoint) $= (90 + 10)/2 = 50$
average class (student viewpoint) $= (90 \times 90 + 10 \times 10)/100 = 82$.

The specific examples I'm discussing are not so important; the point of these first lectures is to illustrate a few of the 100 ideas from STAT134.

Ideas used in Lecture 2.

- $\mathbb{E}g(X) = \int g(x)\ f_X(x)\ dx$.
- exploit symmetry.
- if X, Y independent then $\mathrm{var}\ (aX + bY) = a^2 \mathrm{var}\ (X) + b^2 \mathrm{var}\ (Y)$
- inventing extra structure
- size-biasing

Different authors use slightly different notation:

$$\mathbb{P}(A) = P(A) = \Pr(A)$$

$$\mathbb{P}(X \le 4) = \mathbb{P}\{X \le 4\}$$

$$\mathbb{E}X = \mathbb{E}[X] = EX$$

$$1_A = \mathbb{1}_A = \mathbb{1}(A) = I(A)$$

Chapter 1 of textbook [PK] provides a review of STAT134 level material.

It's rather boring, but helpful for you to read. Here is material from section 1.5.2: some properties of the Exponential distribution, which plays a prominent role in Poisson processes later.

The Exponential(λ) distribution for a r.v. $X > 0$ is defined to have density function

$$f(x) = \lambda e^{-\lambda x}, 0 < x < \infty$$

and has $\mathbb{E}X = 1/\lambda = \mathrm{s.d.}(X)$. Also $\mathbb{P}(X > x) = e^{-\lambda x}$. Note the parametrization convention; λ is called the *rate*.

There are special properties of independent Exponential r.v.'s – say X_1 with rate λ_1 and X_2 with rate λ_2. Consider

$$U = \min(X_1, X_2) = X_M \text{ for } M = \arg\min(X_1, X_2)$$

$$V = \max(X_1, X_2) = X_N \text{ for } N = \arg\max(X_1, X_2).$$

Let's calculate $\mathbb{P}(M = 1, U > t)$, which is the same as $\mathbb{P}(t < X_1 < X_2)$.

independent Exponentials: X_1 with rate λ_1 and X_2 with rate λ_2.

- -

$$\begin{aligned} \mathbb{P}(t < X_1 < X_2 | X_1 = x_1) &= 0 \text{ if } x_1 < t \\ &= e^{-\lambda_2 x_1} \text{ if } x_1 > t. \end{aligned}$$

Use general formula

$$\mathbb{P}(A) = \int \mathbb{P}(A|X = x) \ f_X(x)dx$$

to get

$$\mathbb{P}(t < X_1 < X_2) = \int_t^\infty e^{-\lambda_2 x_1} \ \lambda_1 e^{-\lambda_1 x_1} \ dx_1$$

$$= \frac{\lambda_1}{\lambda_1 + \lambda_2} \exp(-(\lambda_1 + \lambda_2)t).$$

$$U = \min(X_1, X_2) = X_M \text{ for } M = \arg\min(X_1, X_2)$$

We showed

$$\mathbb{P}(M = 1, U > t) = \tfrac{\lambda_1}{\lambda_1+\lambda_2} \exp(-(\lambda_1 + \lambda_2)t).$$

$$\Rightarrow \mathbb{P}(M = 1) = \tfrac{\lambda_1}{\lambda_1+\lambda_2} \qquad (\text{set } t = 0).$$

So similarly

$$\mathbb{P}(M = 2, U > t) = \tfrac{\lambda_2}{\lambda_1+\lambda_2} \exp(-(\lambda_1 + \lambda_2)t).$$

$$\mathbb{P}(M = 2) = \tfrac{\lambda_2}{\lambda_1+\lambda_2}$$

and by summing

$$\mathbb{P}(U > t) = \exp(-(\lambda_1 + \lambda_2)t).$$

These formulas imply that $U = \min(X_1, X_2)$ has Exponential$(\lambda_1 + \lambda_2)$ distribution and is **independent** of $M = \arg\min(X_1, X_2)$.

At the end we used a conceptual fact. Given two random variables Y, Z – say Y discrete and Z continuous, we know and often use the product formula:

$$\text{if } Y \text{ and } Z \text{ independent then}$$
$$(*) \quad \mathbb{P}(Z < z, Y = y) = \mathbb{P}(Z < z) \times \mathbb{P}(Y = y).$$

The converse is also true: if (*) holds for all z, y then Y and Z are independent.

Recall how we find the distributions of minima/maxima. The event $\{\min(X_1, X_2) > x\}$ is the event $\{X_1 > x, X_2 > x\}$, so in this example

$$\mathbb{P}(\min(X_1, X_2) > x) = \mathbb{P}(X_1 > x) \times \mathbb{P}(X_2 > x) = \exp(-(\lambda_1 + \lambda_2)x)$$

which repeats the fact that $\min(X_1, X_2)$ has Exponential$(\lambda_1 + \lambda_2)$ distribution. The same argument for maxima tells us

$$\mathbb{P}(\max(X_1, X_2) \le x) = \mathbb{P}(X_1 \le x) \times \mathbb{P}(X_2 \le x) = (1 - e^{-\lambda_1 x})\,(1 - e^{-\lambda_2 x})$$

and so $\max(X_1, X_2)$ has density function

$$\lambda_1 e^{-\lambda_1 x} + \lambda_1 e^{-\lambda_2 x} - (\lambda_1 + \lambda_2) e^{-(\lambda_1+\lambda_2)x}.$$

Review of conditioning.

$$\mathbb{P}(A|B) = \mathbb{P}(A \text{ and } B)/\mathbb{P}(B); \quad \mathbb{P}(A \text{ and } B) = \mathbb{P}(A|B) \times \mathbb{P}(B).$$

In this course we also use conditional expectation, only studied briefly in STAT134. We know

$$\mathbb{E}X = \sum_x x \, \mathbb{P}(X = x) \text{ or } \int x \, f(x) dx.$$

For an event B we calculate conditional expectation given B by using the same formula with the conditional distribution:

$$\mathbb{E}(X|B) = \sum_x x \, \mathbb{P}(X = x|B) \text{ or } \int x \, f_{X|B}(x) dx.$$

If $(B_i, 1 \le i \le k)$ is a **partition** of events then ("law of total probability")

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i) \times \mathbb{P}(B_i).$$

$$\mathbb{E}X = \sum_i \mathbb{E}(X|B_i) \times \mathbb{P}(B_i).$$

If $(B_i, 1 \le i \le k)$ is a **partition** of events then ("law of total probability")

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i) \times \mathbb{P}(B_i).$$

$$\mathbb{E}X = \sum_i \mathbb{E}(X|B_i) \times \mathbb{P}(B_i).$$

- -

So for a discrete r.v. Y

$$\mathbb{P}(A) = \sum_y \mathbb{P}(A|Y = y) \times \mathbb{P}(Y = y).$$

$$\mathbb{E}X = \sum_y \mathbb{E}(X|Y = y) \times \mathbb{P}(Y = y).$$

And analogously for a continuous r.v. Y with density function f_Y

$$\mathbb{P}(A) = \int \mathbb{P}(A|Y = y) f_Y(y) dy.$$

$$\mathbb{E}X = \int \mathbb{E}(X|Y = y) f_Y(y) dy.$$

Let's do two

Elementary examples. (a) X_1, X_2 fair die throws. Calculate $\mathbb{E}(X_1 | X_1 > X_2)$.

(b) X_1, X_2 independent Exponential, rates λ_1 and λ_2. Calculate $\mathbb{E}(X_1 | X_1 > X_2)$.

(a) X_1, X_2 fair die throws. Calculate $\mathbb{E}(X_1 | X_1 > X_2)$.

- -

$$\mathbb{P}(X_1 > X_2) = \tfrac{5}{12}$$

$$\mathbb{P}(X_1 = x | X_1 > X_2) = \frac{\frac{x-1}{36}}{\frac{5}{12}} = \tfrac{x-1}{15},\ 2 \leq x \leq 6$$

$$\mathbb{E}(X_1 | X_1 > X_2) = \sum_{x=2}^{6} \tfrac{x(x-1)}{15} = \tfrac{14}{3}.$$

(b) X_1, X_2 independent Exponential, rates λ_1 and λ_2. Calculate $\mathbb{E}(X_1 | X_1 > X_2)$.

We know

$$\mathbb{P}(X_1 > X_2) = \tfrac{\lambda_2}{\lambda_1+\lambda_2}$$

$$\mathbb{P}(x \le X_1 \le x + dx) = \lambda_1 e^{-\lambda_1 x} dx$$

and so

$$\mathbb{P}(x \le X_1 \le x + dx | X_1 > X_2) = \frac{\lambda_1 e^{-\lambda_1 x} dx \times (1 - e^{-\lambda_2 x})}{\frac{\lambda_2}{\lambda_1+\lambda_2}}$$

which says that the conditional density of X_1 given $\{X_1 > X_2\}$ is

$$f_{X_1|X_1>X_2}(x) = \frac{\lambda_1(\lambda_1 + \lambda_2)}{\lambda_2}(e^{-\lambda_1 x} - e^{-(\lambda_1+\lambda_2)x}).$$

Then

$$\mathbb{E}(X_1|X_1 > X_2) = \int_0^\infty x f_{X_1|X_1>X_2}(x)\ dx = \frac{\lambda_1 + \lambda_2}{\lambda_1\lambda_2} - \frac{\lambda_1}{\lambda_2(\lambda_1 + \lambda_2)}.$$

Conditional expectation as a random variable.

Given r.v.'s (W, Y) consider $\mathbb{E}(W|Y = y)$. This is a number depending on y – in other words it's a **function** of y. Giving this function a name h we have

$$(*) \quad \mathbb{E}(W|Y = y) = h(y) \text{ for all possible values } y \text{ of } Y.$$

We now make a notational convention, to rewrite the assertion (*) as

$$(**) \quad \mathbb{E}(W|Y) = h(Y).$$

The right side is a r.v., so we must regard $\mathbb{E}(W|Y)$ as a r.v.

[PK] page 60 lists properties of (**), but rather hard to understand at first sight. One important property is that the "law of total probability" becomes

$$\mathbb{E}[\ \mathbb{E}(W|Y)\ ] = \mathbb{E}W.$$

Next class will give three examples to illustrate this notation.