Lecture 4

David Aldous

2 September 2015

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 3.

- Distributions of $\max(X_1, X_2)$ and $\min(X_1, X_2)$
- Properties of Exponential distribution.
- Basics of conditional probability and conditional expectation.
- Global and local interpretations of density function.
- Calculating $\mathbb{E}(X|A)$ from conditional distributions.
- Conditional expectation as a random variable.

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

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

I will give three examples to illustrate this notation.

Example.
First consider

$$S_n = \sum_{i=1}^{n} \xi_i \text{ for i.i.d. } (\xi_i) \text{ with } \mathbb{E}\xi_i = \mu_\xi .$$

Then take N a $\{1, 2, 3, \ldots\}$-valued r.v. with $\mathbb{E}N = \mu_N$, independent of (ξ_i).

What is $\mathbb{E}S_N$*?*

$$S_n = \sum_{i=1}^{n} \xi_i \text{ for i.i.d. } (\xi_i) \text{ with } \mathbb{E}\xi_i = \mu_\xi .$$

N a $\{1, 2, 3, \ldots\}$-valued r.v. with $\mathbb{E}N = \mu_N$, independent of (ξ_i).

- -

$$\mathbb{E}S_n = n\mu_\xi .$$

$$\mathbb{E}(S_N | N = n) = n\mu_\xi .$$

$$\mathbb{E}(S_N | N) = N \ \mu_\xi .$$

$$\mathbb{E}S_N = \mathbb{E}[\mathbb{E}(S_N | N)] = \mathbb{E}[N \ \mu_\xi] = \mu_N \ \mu_\xi .$$

A "geometric probability" example.

Saying that a random point in the plane is **uniform** on a set A is saying that the joint density function of its coordinates (X, Y) is

$$f(x, y) = \frac{1}{\mathrm{area}(A)}, \ (x, y) \in A.$$

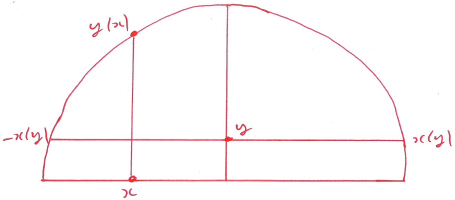
Let's do some calculations in the case where A is the top half of the unit disc. In particular, we will calculate $\mathbb{E}Y$ by first considering $\mathbb{E}(Y|X)$.

$$f(x,y) = \frac{2}{\pi}, \quad (x,y) \in A$$

$$y(x) = \sqrt{1-x^2}$$
$$x(y) = \pm\sqrt{1-y^2}$$

$$f_X(x) = \int_0^{y(x)} f(x,y)\,dy = \frac{2}{\pi}\sqrt{1-x^2}$$

$$f_Y(y) = \int_{-x(y)}^{x(y)} f(x,y)\,dx = \frac{4}{\pi}\sqrt{1-y^2}$$



Conditional distribution of Y given X = x

"obvious" Uniform [0, y(x)]

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2/\pi}{2/\pi\sqrt{1-x^2}} = \frac{1}{\sqrt{1-x^2}}, \quad 0 \le y \le \sqrt{1-x^2}$$

Conditional distribution of Y given $X = x$ is uniform on $[0, \sqrt{1 - x^2}]$. So

$$\mathbb{E}(Y|X = x) = \tfrac{1}{2}\sqrt{1 - x^2}; \quad \text{that is } \mathbb{E}(Y|X) = \tfrac{1}{2}\sqrt{1 - X^2}.$$

We can now calculate

$$\begin{aligned}
\mathbb{E}Y &= \mathbb{E}[\mathbb{E}(Y|X)] \\
&= \tfrac{1}{2}\mathbb{E}\sqrt{1 - X^2} \\
&= \tfrac{1}{2}\int_{-1}^{1} \sqrt{1 - x^2}\ f_X(x)dx \\
&= \pi^{-1}\int_{-1}^{1} (1 - x^2)dx \\
&= \tfrac{4}{3\pi}.
\end{aligned}$$

Example: Likelihood of your team winning, at halftime.

Most team sports are decided by point difference – points scored by home team, minus points scored by visiting team. Write

$$X_1 = \text{ point difference in first half}$$

$$X_2 = \text{ point difference in second half}$$

so

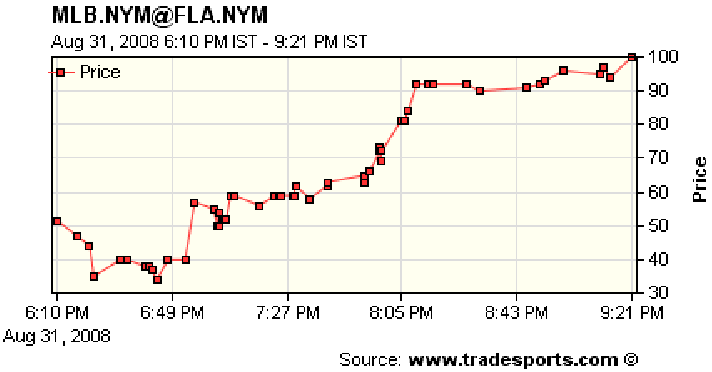
- home team wins if $X_1 + X_2 > 0$
- visiting team wins if $X_1 + X_2 < 0$
- tie if $X_1 + X_2 = 0$.

A reasonable probability model assumes
(i) X_1 and X_2 are i.i.d. random variables.
If the teams are equally talented, then assume
(ii) X_1 has symmetric distribution, that is the same distribution as $-X_1$.
Finally, let me simplify the math by making an unrealistic assumption
(iii) the distribution of X_1 is continuous.

So ties cannot happen, and by (ii) $\mathbb{P}(\text{Home team wins}) = 1/2$. This is the probability before the game starts. At half time we know the value of X_1, so there is a conditional probability $\mathbb{P}(\text{Home team wins} \mid X_1)$. This is different in different matches, so we can ask

What is the distribution of $\mathbb{P}(\text{Home team wins} \mid X_1)$*?*

First I show some data from baseball.



In this match the initial "price" (perceived probability of home team winning) was close to 50%, and halfway through the game it was about 64%. I have these "halfway through the game" numbers for 30 matches where the initial "price" was close to 50%

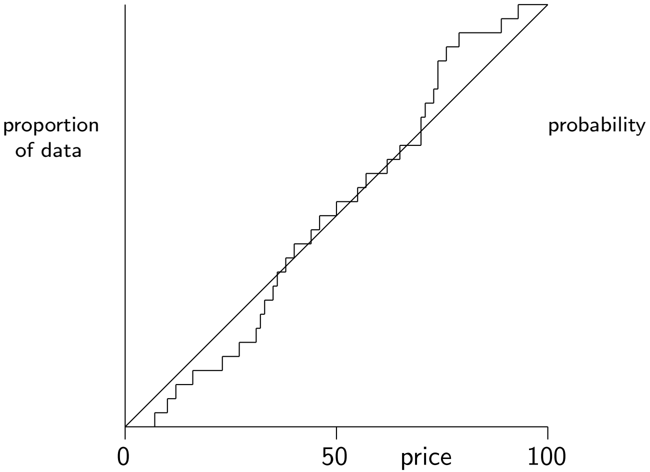


Figure. Empirical distribution function for the baseball data, compared with the uniform distribution.

Write $F(x)$ for the distribution function of each X_i.

$$\begin{aligned}\mathbb{P}(\text{Home team wins} \mid X_1 = x) &= \mathbb{P}(X_1 + X_2 > 0 | X_1 = x) \\ &= \mathbb{P}(X_2 > -x) \\ &= \mathbb{P}(-X_2 < x) \\ &= F(x) \text{ by symmetry assumption}\end{aligned}$$

and this says

$$\mathbb{P}(\text{Home team wins} \mid X_1) = F(X_1).$$

But we know (STAT134) that for X_1 with continuous distribution, $F(X_1)$ always has Uniform$(0, 1)$ distribution.

The baseball data fits this theory prediction quite well. In a low-scoring sport like soccer the "continuous distribution of X_i" assumption is not realistic but we could modify the analysis.

Stochastic Processes

For our purposes a **stochastic process** is just a sequence $(X_0, X_1, X_2, \ldots)$ of random variables. The range space of the X's might be the integers $\mathbb{Z}$ or the reals $\mathbb{R}$ or some more general space S. We envisage observing some physical process changing with time in some random way: X_t is the observed value of the process at time $t = 0, 1, 2, \ldots$. So X_0 is the "initial" value.

Perhaps the simplest example is **simple symmetric random walk**

$$X_t = \sum_{i=1}^{t} \xi_i$$

where (ξ_i) are i.i.d. with $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = \frac{1}{2}$. Here $X_0 = 0$.

There are many ways to study this process, but I want to illustrate the technique "conditioning on the first step" which will be a fundamental technique for studying Markov chains.

Interpret this process as "gambling at fair odds" – bet 1 unit on an event with probability $1/2$, either win or lose the 1 unit. So X_t is your "fortune" after t bets. Suppose

- start with fortune x
- continue until your fortune reach some target amount K or 0.

There is some probability $p(x)$ that you succeed in reaching K; this probability depends on x and on K, but let us take K as fixed. So we know

$$p(K) = 1; \quad p(0) = 0.$$

What can we say about $p(x)$ for $1 \leq x \leq K - 1$?

From the "law of total probability"

$$\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(B^c)\mathbb{P}(A|B^c)$$

we have

$$\begin{aligned} p(x) &= \mathbb{P}(\text{ win first bet}) \times \mathbb{P}(\text{ reach } K|\text{ win first bet}) \\ &+ \mathbb{P}(\text{ lose first bet}) \times \mathbb{P}(\text{ reach } K|\text{ lose first bet}) \end{aligned}$$

That is,

$$(*) \quad p(x) = \tfrac{1}{2}p(x+1) + \tfrac{1}{2}p(x-1), \ 1 \le x \le K-1.$$

This is the simplest case of a **linear difference equation** and the general solution is $p(x) = a + bx$. Because we know the "boundary conditions" $p(0) = 0, \ p(K) = 1$ we can solve to find a, b and find

$$p(x) = x/K, \ 0 \le x \le K.$$

We can use the same "conditioning on the first step" method to study

$$s(x) = \mathbb{E}(\text{number of steps until reach } K \text{ or } 0)$$

starting from x. Here the equation is

$$s(x) = 1 + \tfrac{1}{2}s(x+1) + \tfrac{1}{2}s(x-1), \ 1 \le x \le K-1$$

$$s(0) = 0, \ s(K) = 0.$$

The general form of solution (see textbook for details) is $s(x) = cx(K-x)$; plugging into the equation gives $c = 1$, so

$$s(x) = x(K-x), \ 0 \le x \le K.$$

Conceptual point. The notion of **independence** is used in two conceptually different ways.

- We often use independence as an **assumption** in a model – throwing dice, for instance.
- Given a well-defined math model, events or random variables X, Y either are independent, or are not independent, as a mathematical **conclusion**. For instance, for a uniform random pick of a playing card from a standard deck, the events "card is a King" and "card is a Spade" are independent.

The same point will arise with the Markov property.