Lecture 5: Conditional Expectation

John Sylvester Nicolás Rivera Luca Zanetti Thomas Sauerwald

Lent 2020

UNIVERSITY OF CAMBRIDGE

Outline

Conditional Expectation

Quick-Sort

Coin Flips and Balls-in-Bins

Conditional Probability and Expectation

- Given two events A, B with $\mathbf{P}[A] > 0$ define the *Conditional Probability*

$$\mathbf{P}[B|A] = \mathbf{P}[B \cap A] / \mathbf{P}[A] .$$

  If $\mathbf{P}[A] = 0$, the usual convention is that $\mathbf{P}[B|A] = 0$.
- Given a discrete random variable Y, we define its *Conditional Expectation* with respect to the event A by

$$\mathbf{E}[Y|A] = \sum_{\omega \in \Omega} Y(\omega)\mathbf{P}[\{\omega\}|A] = \sum_{b} b\mathbf{P}[Y = b|A]$$

- When the event $A = \{X = a\}$ where X is another discrete random variable, we define the function $f(a)$ by

$$f(a) = \mathbf{E}[Y|X = a] ,$$

- We define the conditional expectation $\mathbf{E}[Y|X]$, as the **random variable** that takes the value $\mathbf{E}[Y|X = a]$ then $X = a$, i.e. $f(X)$.

Important Remarks

- The conditional expectation $\mathbf{E}[Y|A]$ of Y w.r.t an *event* A is a **deterministic number** .
- The conditional expectation $\mathbf{E}[Y|X]$ of Y w.r.t a *random variable* X is a **random variable** .
- In the definition of $\mathbf{E}[Y|X]$ above X can be a random vector $(X_1, \ldots, X_N)$.

Example: Single Dice

- Let Y be 1 if the dice rolls 1 and 0 otherwise
- Let X_1 be 1 if the dice shows odd number, 0 otherwise
- Let X_2 be 1 is the dice shows a number ≤ 2 , 0 otherwise

$$\mathbf{E}[Y|X_1] = X_1/3 \qquad \mathbf{E}[Y|X_2] = X_2/2 \qquad \mathbf{E}[Y|(X_1, X_2)] = X_1 \cdot X_2$$

Important Remarks (continued)

- The conditional expectation $\mathbf{E}[\,Y|X\,]$ is always a function of X.
- Behind conditional expectation there is the notion of *information*.
- The standard notion of expectation $\mathbf{E}[\,Y\,]$ can be thought of as

  'the best estimate of a random variable Y given no information about it,'

  while the conditional expectation $\mathbf{E}[\,Y|\mathcal{I}\,]$ given $\mathcal{I}$ can be thought of as

  'the best estimate of a random variable Y given information $\mathcal{I}$,'

  where $\mathcal{I}$ above may be an event or a random variable etc.

Conditional Expectation: Two Dice

Suppose we independently roll two standard 6-sided dice. Let X_1 and X_2 the observed number in the first and second dice respectively. Then,

$$\mathbf{E}[\, X_1 + X_2 | X_1 \,] = 3.5 + X_1 \,.$$

Why? Because if $X_1 = a$ then

$$\begin{aligned}
\mathbf{E}[\, X_1 + X_2 | X_1 = a \,] &= \sum_{b=1}^{12} b\mathbf{P}[\, X_1 + X_2 = b | X_1 = a \,] \\
&= \sum_{b=1}^{12} b\mathbf{P}[\, X_1 + X_2 = b, X_1 = a \,] / \mathbf{P}[\, X_1 = a \,] \\
&= \sum_{b=1}^{12} b\mathbf{P}[\, X_2 = b - a, X_1 = a \,] / \mathbf{P}[\, X_1 = a \,] \\
(X_1 \textbf{ independent of } X_2) \quad &= \sum_{b=1}^{12} b\mathbf{P}[\, X_2 = b - a \,] \\
&= \sum_{c=1}^{6} (c + a)\mathbf{P}[\, X_2 = c \,] \\
&= 3.5 + a
\end{aligned}$$

Conditional Expectation: Properties

(1) $\mathbf{E}[\,\mathbf{E}[\,Y|X\,]\,] = \mathbf{E}[\,Y\,]$.

(2) $\mathbf{E}[\,1|X\,] = 1$

(3) *Linearity:*

- For any constant $c \in \mathbb{R}$, $\mathbf{E}[\,cY|X\,] = c\mathbf{E}[\,Y|X\,]$
- $\mathbf{E}[\,Y + Z|X\,] = \mathbf{E}[\,Y|X\,] + \mathbf{E}[\,Z|X\,]$

(4) If X is independent of Y, then $\mathbf{E}[\,Y|X\,] = \mathbf{E}[\,Y\,]$.

(5) if Y is a function of X, i.e. $Y = f(X)$, then $\mathbf{E}[\,YZ|X\,] = Y\mathbf{E}[\,Z|X\,]$. Particularly, $\mathbf{E}[\,X|X\,] = X$

(6) *Tower Property:*

- $\mathbf{E}[\,\mathbf{E}[\,X|(Z,Y)\,]\,|Y\,] = \mathbf{E}[\,X|Y\,]$.

(7) *Jensen Inequality:*

- if f is a convex real function, then $f(\mathbf{E}[\,X|Y\,]) \leq \mathbf{E}[\,f(X)|Y\,]$.

These properties greatly simplify calculations. Example: for our two dice

$$\mathbf{E}[\,X_1 + X_2|X_1\,] \overset{p3}{=} \mathbf{E}[\,X_1|X_1\,] + \mathbf{E}[\,X_2|X_1\,] \overset{p5,p4}{=} X_1 + \mathbf{E}[\,X_2\,] = X_1 + 3.5$$

Exercise: Prove the properties from slide 7

Example: we can prove Property (1), that is $\mathbf{E}[\,\mathbf{E}[\,Y|X\,]\,] = \mathbf{E}[\,Y\,]$, as follows:

$$
\begin{aligned}
\mathbf{E}[\,\mathbf{E}[\,Y|X\,]\,] &= \sum_{\omega\in\Omega} \mathbf{E}[\,Y|X\,](\omega)\cdot\mathbf{P}[\,\{\omega\}\,] \\
&= \sum_{x} \mathbf{E}[\,Y|X=x\,]\,\mathbf{P}[\,X=x\,] \\
&= \sum_{x}\sum_{y} y\mathbf{P}[\,Y=y|X=x\,]\,\mathbf{P}[\,X=x\,] \\
&= \sum_{x}\sum_{y} y\mathbf{P}[\,Y=y,X=x\,] \\
&= \sum_{y} y\mathbf{P}[\,Y=y\,] \\
&= \mathbf{E}[\,Y\,]
\end{aligned}
$$

Bonus Exercise: Prove Property (1) using Properties (4) and (2).

Outline

Conditional Expectation

Quick-Sort

Coin Flips and Balls-in-Bins

Quick-Sort with Random Pivot

Algorithm: QuickSort

Input: Array of different number A.
Output: array A sorted in increasing order

- Pick an element *uniformly* from the array, the so-called pivot .
- If $|A| = 0$ or $|A| = 1$; return A.
- Else
  - Generate two subarrays A_1 and A_2:
    A_1 contains the elements that are smaller than the pivot ;
    A_2 contains the elements that are greater than the pivot ;
  - Recursively sort A_1 and A_2.

Let C_n be the number of comparisons made by Qucik-Sort on n elements.

Recall that Nicolas showed something along the lines of

$$\mathbf{P}[\, C_n \geq 21 n \log n \,] = 1/n.$$

What is $\mathbf{E}[\, C_n \,]$ - the expected number of comparisons?

The expected number of comparisons in Quick-Sort

Let $M_n = \mathbf{E}[\, C_n \,]$ be the expected number of comparisons needed by quick-sort to sort n distinct values. Conditioning on the rank of the pivot gives

$$M_n = \sum_{j=1}^{n} \mathbf{E}\left[\, C_n \,\middle|\, \text{pivot selected is } j\text{th smallest value} \,\right] \cdot \frac{1}{n}$$

If the initial pivot selected is the jth smallest value, then the set of values smaller than it has size $j-1$, and the set of values greater has size $n-j$.

Hence, as $n-1$ comparisons with the pivot must be made, we have

$$M_n = \sum_{j=1}^{n} (n-1+M_{j-1}+M_{n-j})\frac{1}{n} = n-1+\frac{2}{n}\sum_{j=1}^{n-1} M_j.$$

Since $M_0 = 0$.

Thus

$$(n+1)M_{n+1} - nM_n = 2n + 2M_n.$$

Or equivalently

$$(n+1)M_{n+1} = 2n + (n+2)M_n.$$

The expected number of comparisons in Quick-Sort

Rearranging $(n+1)M_{n+1} = 2n + (n+2)M_n$ we have

$$\begin{aligned}
\frac{M_{n+1}}{n+2} &= \frac{2n}{(n+1)(n+2)} + \frac{M_n}{n+1} \\
&= \frac{2n}{(n+1)(n+2)} + \frac{2(n-1)}{n(n+1)} + \frac{M_{n-1}}{n} \\
&= 2\sum_{k=0}^{n-1} \frac{n-k}{(n+1-k)(n+2-k)} \\
&= 2\sum_{i=0}^{n} \frac{i}{(i+1)(i+2)} \\
&= 2\left[\sum_{i=1}^{n} \frac{2}{i+2} - \sum_{i=1}^{n} \frac{1}{i+1}\right] \sim 2\log n.
\end{aligned}$$

Since $M_1 = 0$.

Thus Quick-Sort makes $\mathbf{E}[\, C_n \,] \sim 2n \log n$ comparisons in expectation.

Outline

Conditional Expectation

Quick-Sort

Coin Flips and Balls-in-Bins

Example: Expectation of a Geometric Random Variable

Suppose $X_1, X_2, \ldots,$ is an infinite sequence of independent Bernoulli $\mathrm{Ber}(p)$ random variables with parameter p. That is

$$\mathbf{P}[\, X_i = 1 \,] = p, \qquad \mathbf{P}[\, X_i = 0 \,] = 1 - p.$$

We shall think of the $X_i' s$ as *coin flips*.

Let $G = \min\{k \geq 1 : X_k = 1\}$, *the number of coin flips until we get a head.*

G has *geometric distribution* $\mathrm{Geo}(p)$ with parameter p. Indeed,

$$\mathbf{P}[\, G = k \,] = p(1-p)^{k-1}.$$

The expectation of G is given by the formula

$$\mathbf{E}[\, G \,] = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

Let say that we forgot how to compute sums of this type...

We can compute $\mathbf{E}[\, G\,]$ by other means.

- $\mathbf{E}[\, G\,] \stackrel{p1}{=} \mathbf{E}[\,\mathbf{E}[\, G|X_1\,]\,]$
- Conditional on X_1,

$$G = X_1 + (1 - X_1)(1 + G'),$$

  where G' is the number of coins we need to wait to see a head after the first coin.
- $\mathbf{E}[\, X_1 + (1 - X_1)(1 + G')|X_1\,] \stackrel{p3,p5}{=} X_1 + (1 - X_1)\mathbf{E}[\, 1 + G'|X_1\,]$
- G' has geometric distribution of parameter p and it is independent of X_1. Hence

$$\mathbf{E}\left[\, 1 + G'|X_1\,\right] \stackrel{p4}{=} \mathbf{E}\left[\, 1 + G'\,\right] = 1 + \mathbf{E}[\, G\,]$$

- Solve

$$\mathbf{E}[\, G\,] = p + (1 - p)(1 + \mathbf{E}[\, G\,])$$

- To give

$$\mathbf{E}[\, G\,] = 1/p.$$

Example: Balls into Bins

Suppose we have n bins but a random number of balls, say M. Suppose M has finite expectation. What is the expected number of balls in the first bin?

Set Up:

- Balls are assigned to bins uniformly and independently at random
- Let $X_i = 1$ if the ball i falls in bin 1, 0 otherwise
- The total number of balls in bin 1 is $\sum_{i=1}^{M} X_i$
- M is a random variable, and M is independent of all X_i's

Question Rephrased

Let X_i be i.i.d. and M be independent of $\{X_i\}_{i\geq 0}$. If $\mathbf{E}[\,X_i\,] < \infty$ and $\mathbf{E}[\,M\,] < \infty$ then what is $\mathbf{E}\left[\sum_{i=1}^{M} X_i\right]$?

Expectation of the Compound Random Variable $\sum_{i=1}^{M} X_i$

We shall use $\mathbf{E}\left[\sum_{i=1}^{M} X_i\right] \stackrel{P1}{=} \mathbf{E}\left[\mathbf{E}\left[\sum_{i=1}^{M} X_i | M\right]\right]$. Observe that for any $k \in \mathbb{N}$

$$f(k) := \mathbf{E}\left[\sum_{i=1}^{M} X_i \middle| M = k\right] = \mathbf{E}\left[\sum_{i=1}^{k} X_i \middle| M = k\right]$$

$$\stackrel{P3}{=} \sum_{i=1}^{k} \mathbf{E}[\, X_i | M = k\,] \stackrel{P4}{=} \sum_{i=1}^{k} \mathbf{E}[\, X_i\,] = k \cdot \mathbf{E}[\, X_i\,],$$

since X_i are all equidistributed. Thus $\mathbf{E}\left[\sum_{i=1}^{M} X_i | M\right] = f(M) = M \cdot \mathbf{E}[\, X_i\,]$.

To conclude we have

$$\mathbf{E}\left[\sum_{i=1}^{M} X_i\right] = \mathbf{E}[\, M\mathbf{E}[\, X_i\,]\,] \stackrel{P3}{=} \mathbf{E}[\, M\,]\,\mathbf{E}[\, X_i\,].$$