Conditional expectation

Saturday, October 7, 2017 11:35 PM

If X, Y are two discrete random variables we can write

$$\mathbb{E}[X \mid Y=y] = \sum_x x\,\mathbb{P}[X=x \mid Y=y]$$

$$= \frac{\sum_x x\,\mathbb{P}[X=x, Y=y]}{\mathbb{P}[Y=y]}$$

$$= \frac{\mathbb{E}[X\, I(Y=y)]}{\mathbb{P}[Y=a]}$$

Expectation of X given Y

$$\mathbb{E}[X \mid Y] := \psi(Y)$$

where $\psi(y) = \mathbb{E}[X \mid Y=y]$

Note $\mathbb{E}[X \mid Y=y]$ is a real number while $\mathbb{E}[X \mid Y]$ is a random variable.

Properties:

i) $\mathbb{E}[aX + Z \mid Y] = a\,\mathbb{E}[X \mid Y] + \mathbb{E}[Z \mid Y]$

ii) $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}X$

iii) $\mathbb{E}[X \cdot f(Y) \mid Y] = f(Y) \cdot \mathbb{E}[X \mid Y]$

iv) If X, Y independent, $\mathbb{E}[X \mid Y] = \mathbb{E}X$.

Pf: $\mathbb{E}(\mathbb{E}[X \mid Y]) = \sum_y \mathbb{P}[Y=y] \cdot \mathbb{E}[X \mid Y=y]$

$$= \sum_y \sum_x x \underbrace{\mathbb{P}[X=x \mid Y=y] \cdot \mathbb{P}[Y=y]}_{= \mathbb{P}[X=x,\, Y=y]}$$

$$= \mathbb{E}X.$$

General Conditional Expectation

If $\mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra and X is a R.V. with $\mathbb{E}|X| < \infty$ then

$\mathbb{E}[X \mid \mathcal{G}]$ is a $\mathcal{G}$ measurable random variable such that for all $B \in \mathcal{G}$,

$$\mathbb{E}\, X\, I(B) = \mathbb{E}\big(\mathbb{E}[X \mid \mathcal{G}] \cdot I(B)\big)$$

– Special Case $\mathcal{G} = \sigma(Y)$ – sets of the form $\{Y \in A\}$ where $A \in \mathcal{B}$.

Note: If X is $\mathcal{G}$-measurable then $\mathbb{E}[X \mid \mathcal{G}] = X$.

Existence: If μ, ν are measures on $\mathcal{G}$ we say that μ is absolutely continuous with respect to ν if $\forall A \in \mathcal{G}$

$$\mu(A) > 0 \implies \nu(A) > 0.$$

—

<u>Theorem</u>: Radon-Nikodym

If μ is absolutely continuous with respect to ν then $\exists\, h \geq 0.$ such that

$$\nu(A) = \int_A h \, d\mu.$$

Sometimes we write $h = \frac{d\nu}{d\mu}$

<u>Existence</u>: Assume $X \geq 0$. Set

$$\nu(A) = \int_A X \, d\mathbb{P} = \mathbb{E}[I(A)X]$$

and $\mu = \mathbb{P}$.

Then $\mathbb{E}[X \mid \mathcal{G}] = \frac{d\nu}{d\mu}$

<u>General X</u>: $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X^+ \mid \mathcal{G}] - \mathbb{E}[X^- \mid \mathcal{G}].$

<u>Uniqueness</u> Suppose that there are two $\mathcal{G}$-measureable R.V. Y_1, Y_2 such that,

$$\mathbb{E}[Y_i \, I(A)] = \mathbb{E}[X \, I(A)] \quad \forall A \in \mathcal{G}.$$

Then $\mathbb{E}[(Y_1 - Y_2) I(A)] = 0 \quad \forall A \in \mathcal{G}.$

If $A_+ = \{ Y_1 - Y_2 > 0 \}$, $A_- = \{ Y_1 - Y_2 < 0 \}$

$$0 = \mathbb{E}[(Y_1 - Y_2) I(A_+)] - \mathbb{E}[(Y_1 - Y_2) I(A_-)]$$
$$= \mathbb{E}[|Y_1 - Y_2|]$$

$\Rightarrow \mathbb{P}[|Y_1 - Y_2| > 0] = 0$

$Y_1 = Y_2$ a.s.

So $\mathbb{E}[X | \mathcal{G}]$ is unique up to sets of measure 0.

If $\mathcal{G} = \{\emptyset, \Omega\}$ is the trivial σ-algebra then $\mathbb{E}[X | \mathcal{G}]$ is constant R.V. equal to $\mathbb{E}X$.

Check the definitions match for discrete random variables. If $\mathcal{G} = \sigma(Y)$

$$\mathbb{E}[X | Y] := \mathbb{E}[X | \mathcal{G}]$$

If Y is discrete taking values in S then

$$A \in \mathcal{G} \text{ iff } \exists B \subseteq S, \ A = \bigcup_{y \in B} \{Y = y\}$$

In particular if Z is $\mathcal{G}$-measurable then Z is constant on $\{Y = y\}$ so

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

Let $A = \{Y = y\}$, then

$$\mathbb{E}[\, \mathbb{E}[X | Y] \cdot I(A)] = \mathbb{E}[h(y) I(A)]$$
$$= h(y) \cdot \mathbb{P}[Y = y]$$

$$= \mathbb{E}[X I(A)] = \mathbb{E}[X I(Y = y)]$$

$\Rightarrow h(y) = \mathbb{E}[X I(Y=y)]$ ...

$$\Rightarrow h(y) = \frac{\quad}{\mathbb{P}[Y=y]} = \mathbb{E}[X \mid Y=y]$$

matching the definition in the discrete case.

<u>Properties</u>:

Tower property: If $\mathcal{G}_1 \subseteq \mathcal{G}_2$ then

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[X \mid \mathcal{G}_1] \qquad *$$

<u>Special case</u>: $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]]$

First note that $\mathbb{E}[X \mid \mathcal{G}_1]$ is $\mathcal{G}_1$ measurable

so to prove $*$ we need to show that $\forall A \in \mathcal{G}_1$

$$\mathbb{E}\left[I(A)\,\mathbb{E}[X \mid \mathcal{G}_2]\right] = \mathbb{E}[I(A)\,\mathbb{E}[X \mid \mathcal{G}_1]]$$

Since $A \in \mathcal{G}_2$ also

LHS $= \mathbb{E}[I(A)X]$

RHS $= \mathbb{E}[I(A)X]$ ✓

<u>Independence</u>: If A is independent of $\mathcal{G}$ then

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$$

<u>Pf</u>. If $A \in \mathcal{G}$ $\mathbb{E}[I(A)X] = \mathbb{P}[A]\mathbb{E}[X] = \mathbb{E}[I(A)\mathbb{E}[X]]$.

If Y is $\mathcal{G}$ measurable then

$$\mathbb{E}[Y \mid \mathcal{G}] = Y, \qquad \mathbb{E}[XY \mid \mathcal{G}] = Y\,\mathbb{E}[X \mid \mathcal{G}].$$

<u>Examples</u> If $X_1, \dots$ IID $\mathbb{E}X_i = \mu$, $S_n = \sum_{i=1}^{n} X_i$

For $n > m$

$$\mathbb{E}[S_n \mid S_m] = \mathbb{E}\Big[S_n + \sum_{i=n+1}^{m} X_i \,\Big|\, S_n\Big]$$

$$= S_n + (m-n)\mu.$$

for $n < m$,

$$\mathbb{E}[S_n \mid S_m] = \sum_{i=1}^{n} \mathbb{E}[X_i \mid S_m]$$

By symmetry $\mathbb{E}[X_1 \mid S_m] = \mathbb{E}[X_2 \mid S_m] = \dots = \mathbb{E}[X_m \mid S_m]$

and $S_m = \mathbb{E}[S_m \mid S_m] = \sum_{i=1}^{m} \mathbb{E}[X_i \mid S_m] = m\,\mathbb{E}[X_i \mid S_m]$

so $\mathbb{E}[X_i \mid S_m] = \frac{S_m}{m}$

$$\Rightarrow \mathbb{E}[S_n \mid S_m] = \frac{n}{m} \cdot S_m.$$

<u>Example:</u> $X_1, X_2, \dots$ $\mathbb{E}X_i = \mu$, N is independent of the X_i, $\mathbb{E}N = \theta$. Find $\mathbb{E}\sum_{i=1}^{N} X_i$.

$$\mathbb{E}\Big[\sum_{i=1}^{N} X_i\Big] = \mathbb{E}\Big(\mathbb{E}\Big[\sum_{i=1}^{N} X_i \,\Big|\, N\Big]\Big)$$

$$= \mathbb{E}[N\mu] = \mu\theta. \quad \text{(why?)}$$

More formally

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i \mid N\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} X_i \, I(N \geq i) \mid N\right]$$

$$= \sum_{i=1}^{\infty} \mathbb{E}\left[X_i I(N \geq i) \mid N\right]$$

$$= \sum_{i=1}^{\infty} I(N \geq i) \, \mathbb{E}\left[X_i \mid N\right]$$

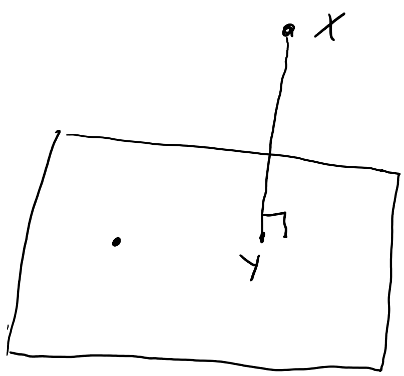
$$= \sum_{i=1}^{\infty} I(N \geq i) \mu = N\mu.$$

Interpretation as a projection

Let $L^2(G)$ be the set of finite variance random variables. A Hilbert space with inner product

$$\langle X, Y \rangle = \mathbb{E} XY, \quad \langle X, X \rangle = \|X\|_{L^2}^2.$$

Then if $\mathbb{E}X^2 < \infty$, and $Y = \mathbb{E}[X \mid G]$ then Y is the projection of X onto $L^2(G)$



For all $Z \in L^2(G)$,

$$\langle X - Y, Z \rangle = 0$$

Since

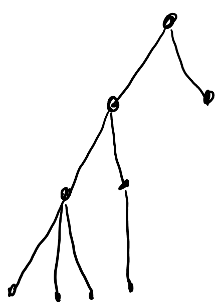
$$\mathbb{E}[(X-Y)Z] = \mathbb{E}\left[\mathbb{E}[(X-Y)Z \mid G]\right]$$

$$= \mathbb{E}\left[Z \, \mathbb{E}[(X-Y) \mid G]\right]$$

$$= \mathbb{E}[Z\, \mathbb{E}[(X-Y) \mid G]]$$

$$= \mathbb{E}[Z\,(Y-Y)] = 0.$$

Branching Processes (Galton-Watson)



$Z_0 = 1$

$Z_1 = 2$

$Z_2 = 2$

$Z_3 = 4$

Each node independently has random number of children with distribution X.

Z_i is #children in generation i.

Q: Does it become extinct, i.e. Z_n is eventually 0?

First moment Let $\mu = \mathbb{E}X$.

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{i,n} \quad \text{so}$$

$$\mathbb{E}[Z_n \mid Z_{n-1}] = \mathbb{E}\Big[\sum_{i=1}^{Z_{n-1}} X_{i,n} \mid Z_{n-1}\Big] = \mu Z_{n-1}.$$

$$\mathbb{E}Z_n = \mu\, \mathbb{E}Z_{n-1} = \mu^n\, \mathbb{E}Z_0 = \mu^n.$$

If $\mu < 1$ then $\mathbb{E}Z_n \to 0$ so

$\mathbb{P}[Z_n \geq 1] \leq \mathbb{E} Z_n \to 0$. Extinction almost surely.

What about $\mu \geq 1$?

Probability Generating Functions

Define $G_X(s) = \mathbb{E} s^X = \sum_{i=0}^{\infty} s^i \, \mathbb{P}[X=i]$

for X a non-negative integer valued R.V.

Note that

i) $G_X(0) = \mathbb{P}[X=0]$

ii) If X, Y are independent,

$$G_{X+Y}(s) = \mathbb{E} s^{X+Y} = \mathbb{E} s^X \, \mathbb{E} s^Y = G_X(s) \, G_Y(s).$$

iii) $G_X'(s) = \mathbb{E} X s^{X-1}$ and

$$G_X'(1) = \mathbb{E} X.$$

iv) $G_X''(s) = \mathbb{E} X(X-1) s^{X-2} > 0$ if $\mathbb{P}[X \geq 2] > 0$.

$$G_X''(1) = \mathbb{E} X^2 - \mathbb{E} X.$$

Example $X \sim Ber(p)$, $G_X(s) = ps + 1 - p = 1 + p(s-1)$

$Y \sim Bin(n,p)$ $\quad Y = \sum_{i=1}^{n} X_i \quad G_Y(s) = (1 + p(s-1))^n$

$Z \sim Geom(p)$ $\quad \mathbb{P}[Z = k] = (1-p)^k \cdot p$

$$G_Z(s) = \sum_{k=0}^{\infty} (1-p)^k p \cdot s^k = p \sum_{k=0}^{\infty} ((1-p)s)^k$$

$$= \frac{p}{1-(1-p)s}.$$

If $X_1, \dots$ IID with G.F. $G(s)$,

N independent with G.F. $H(s)$, $Z = \sum_{i=1}^{N} X_i$ then

$$G_Z(s) = \mathbb{E}\, s^{\sum_{i=1}^{N} X_i} = \mathbb{E}\left(\mathbb{E}\left[s^{\sum_{i=1}^{N} X_i} \mid N\right]\right)$$

$$= \mathbb{E}\left[\left(\mathbb{E}\left[s^{X_i} \mid N\right]\right)^N\right]$$

$$= \mathbb{E}\left[G(s)^N \right] = H(G(s))$$

Generating function of a branching process.

If $G(s)$ is G.F. of X, $G_n(s) = G_{Z_n}$

$G_{Z_0} = s$, $G_{Z_1} = G_X = G(s)$,

$$G_n(s) = G_{n-1}(G(s)) \quad \text{since } Z_n = \sum_{i=1}^{Z_{n-1}} X_{i,n}.$$

$$= G_{n-2}(G(G(s)))$$

$$= G(\cdots G(s) \cdots)$$

n-fold composition.

Extinction $\iff \mathbb{P}[Z_n = 0] \to 1$

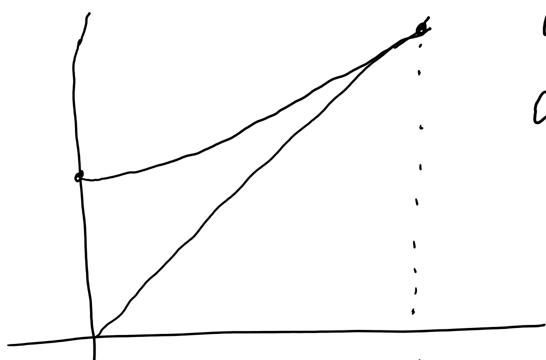
$\iff G_n(0) \to 1$ as $n \to \infty$

Note that $G_n(0)$ is increasing so $G_n(0) \nearrow s_*$. and

$$s_* = \lim G_n(0) = \lim G_{n+1}(0) = \lim G(G_n(0))$$
$$= G(\lim G_n(0))$$
$$= G(s_*)$$

We always have $G(1) = \mathbb{E}\, 1^X = 1$ is a fixed point

Case 1 $\mu < 1$



$G''(s) = \mathbb{E}(X(X-1)s^X) \geq 0$

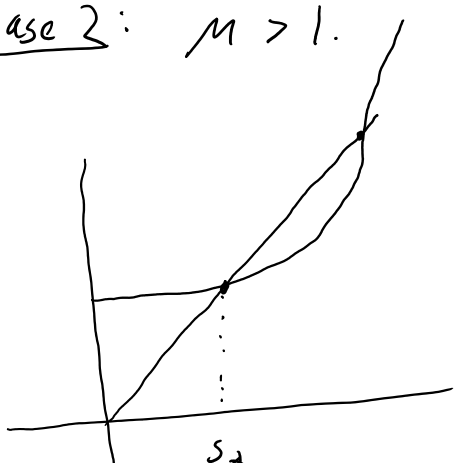
$G'(1) = \mathbb{E}X = \mu < 1$ so

$G'(s) \leq \mu < 1$ for $0 \leq s \leq 1$

So $s=1$ is the only solution to $G(s)=s$ in $[0,1]$.

Of course we already knew this!

Case 2: $\mu > 1$.

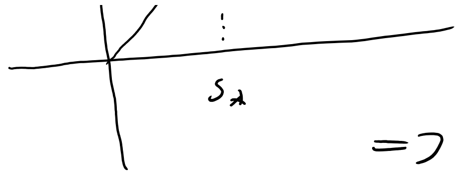


$G'(1) = \mu > 1$ so

$G(s) < s$ for $s \in (1-\varepsilon, 1)$.

$G''(s) = \mathbb{E}\, X(X-1)s^X > 0$ so

G is strictly convex.



u is strictly convex.

$\Rightarrow$ Exactly 1 solution $G(s_*) = s_*$ in $[0,1)$.

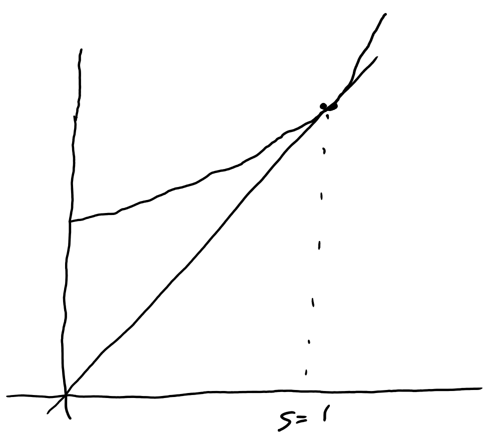
Since for $s < s_*$, $G(s) \le G(s_*) = s_*$

so $\forall n$, $G_n(0) \le s_*$. Therefore $G_n(0) \uparrow s_* < 1$.

$\mathbb{P}[\text{Extinction}] = s_* < 1$.

<u>Case 3</u> $\mu = 1$.

Trivial case $\mathbb{P}[X = 1] = 1$. Otherwise $\mathbb{P}[X \geq 2] > 0$.



Then if $H(s) = G(s) - s$,

$H(1) = 0$,

$H'(1) = G'(1) - 1 = 0$.

$H''(s) = \mathbb{E}\, X(X-1)s^X > 0$

for all $s \geq 0$.

So H is convex and so has a unique minima at $s = 1$.

$\Rightarrow$ 1 is the unique solution to $G(s) = s$

so $\mathbb{P}[\text{Extinction}] = 1$.

In this case $\mathbb{E}[Z_n \mid Z_{n-1}] = Z_{n-1}$ which defines a martingale.