Stat 2911 Lecture Notes

Class 18, 2017

Uri Keich

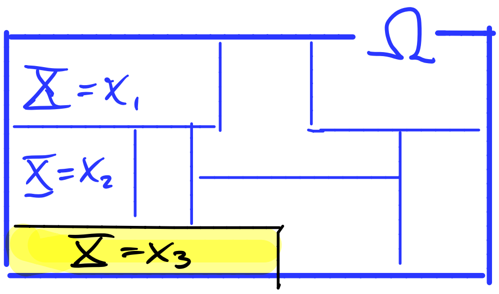
Expected value and variance of the conditional expectation (Rice 4.4)

Conditional Expectation

If X and Y are jointly distributed RVs, then $\forall x$ with $P(X=x)>0$, we define the conditional expectation of Y given $X=x$ as

$$E(Y|X=x)=\sum_y{}' \, y \, P(Y=y|X=x)$$

provided $\sum_y |y| P(Y=y|X=x)<\infty$



$E(Y|X=x_3)$ = the weighted average of Y over $\{\omega : X(\omega)=x_3\}$

Example. Roll a fair die and let Y = outcome

$X = 1_{\{\text{even outcome}\}}$

Ω:

1	3	5
2	4	6

$X=0$

$X=1$

$E(Y|X=0)=3$

$E(Y|X=1)=4$

Claim. If $Y\in L^1$ then $E(Y|X=x)$ is defined $\forall x$ with $P(X=x)>0$.

Random Sum: X_i are iid RVs and N is an ind. $\mathbb{N}_0$-valued RV. $T = \sum_1^N X_i$ is called a random sum.

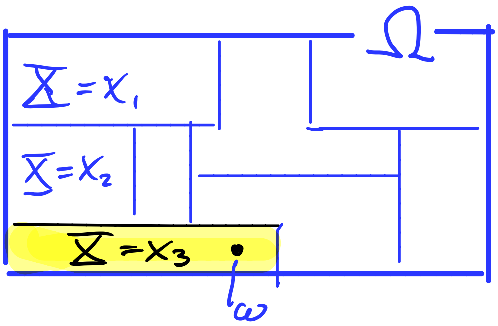
Claim. If $P(N=n)>0$ then the dist. of T given that $N=n$ is the same as the dist. of $S_n = \sum_1^n X_i$.

Cor. If $X_i \in L^1$ then $E(T \mid N=n) = n\,E(X_i)$.

"total expectation law": $E(T) = \sum_n E(T \mid N=n)\,P(N=n)$.

Def. For $Y \in L^1$ we define the RV conditional expectation of Y given X as

$$\underbrace{E(Y \mid X)}_{\text{notation}}(\omega) = E[Y \mid X = X(\omega)]$$



Suppose $\omega \in \Omega$ satisfies $X(\omega) = x_3$.
Then $E(Y \mid X)(\omega) = E(Y \mid X = x_3)$

Example. Fair die, Y = outcome, $X = 1_{\{\text{even outcome}\}}$

Ω:

1	3	5
2	4	6

$X = 0$ $\qquad E(Y \mid X=0) = 3$

$X = 1$ $\qquad E(Y \mid X=1) = 4$

$$E(Y \mid X)(\omega) = \begin{cases} & \omega \in \{1,3,5\} \\ & \omega \in \{2,4,6\} \end{cases}$$

We defined $E(Y|\mathbb{X})$ for all ω for which $P(\mathbb{X}=\mathbb{X}(\omega))>0$.
Since we are dealing with discrete RVs, we can assume here $\forall \omega\in\Omega$, $P(\mathbb{X}=\mathbb{X}(\omega))>0$. $(*)$
That is, we assume $\{x: P(\mathbb{X}=x)>0\}=\mathbb{X}(\Omega)$.

Define $h: \mathbb{X}(\Omega)\longrightarrow\mathbb{R}$ as $h(x)=E(Y|\mathbb{X}=x)$ (well defined if $Y\in\mathcal{L}^1$).

Claim.

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

Proof. For any $\omega\in\Omega$, $P(\mathbb{X}=\mathbb{X}(\omega))>0$ $(*)$ and

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

Example. Fair die: $h(0)=3$
$h(1)=4$

and $E(Y|\mathbb{X})(\omega)=h(\mathbb{X}(\omega))$.

Note that in this example

$$\begin{aligned} E[E(Y|\mathbb{X})] &= E[h(\mathbb{X})] \\ &= h(0)P(\mathbb{X}=0)+h(1)P(\mathbb{X}=1) \\ &= 3\cdot\tfrac{1}{2}+4\cdot\tfrac{1}{2}=3\tfrac{1}{2} \\ &= E(Y). \end{aligned}$$

Example. Random Sum $T = \sum_{1}^{N} X_i$, $X_i \in L^1$.

For $n \in N(\Omega)$,

$$h(n) = E(T \mid N = n)$$
$$= n E(X_i).$$

(numbers)

Therefore,

$$E(T \mid N) = h(N)$$
$$= N E(X_i).$$

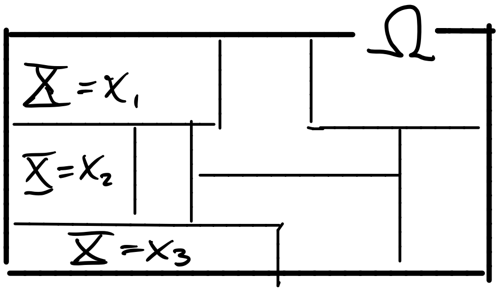
(RVs)

Note: $E[E(T \mid N)] = E(N) E(X_i)$.
$= E(T)$ for the faulty monitor.

This is the second such example; is it the rule?

Theorem. If $Y \in L^1$ then $E(Y \mid X) \in L^1$ and

$$E[E(Y \mid X)] = E(Y).$$



$E(Y \mid X) \equiv E(Y \mid X = x_n)$ on any event $X = x_n$.

Averaging the averages gives you the overall average.

__Theorem.__ If $Y \in \mathcal{L}^1$ then $E(Y|\mathbb{X}) \in \mathcal{L}^1$ and

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

__Proof.__

Recall that $E(Y|\mathbb{X}) = h(\mathbb{X})$, where $h(x) = E(Y|\mathbb{X}=x)$

Therefore, $E(Y|\mathbb{X}) \in \mathcal{L}^1$ iff $h(\mathbb{X}) \in \mathcal{L}^1$:

$$\sum_{x \in \mathbb{X}(\Omega)}{}' |h(x)|\, P(\mathbb{X}=x) = \sum_{x}{}' \underbrace{|E(Y|\mathbb{X}=x)|}_{\text{why well-def'd?}} \cdot P(\mathbb{X}=x)$$

$$= \sum_{x}{}' \Big| \sum_{y}{}' y \cdot P(Y=y|\mathbb{X}=x) \Big| \cdot P(\mathbb{X}=x)$$

$$\leq \sum_{x}{}' \sum_{y}{}' |y| \cdot P(Y=y|\mathbb{X}=x) \cdot P(\mathbb{X}=x)$$

$$\overset{(*)}{=} \sum_{y}{}' |y| \sum_{x}{}' P(Y=y|\mathbb{X}=x) \cdot P(\mathbb{X}=x)$$

$$= \sum_{y}{}' |y|\, P(Y=y) < \infty$$

$\Rightarrow E(Y|\mathbb{X}) \in \mathcal{L}^1$, and moreover, repeating w/o $|\cdot|$:

$$E[E(Y|\mathbb{X})] = \sum_{x}{}' h(x)\, P(\mathbb{X}=x)$$

$$= \underbrace{\sum_{x}{}' E(Y|\mathbb{X}=x)\, P(\mathbb{X}=x) = E(Y)}_{\text{"Total expectation law"}}$$

(*) is justified by the absolute convergence. $\square$

Cor. If $T = \sum_1^N X_i$ is a random sum with $X_i \in \mathcal{L}^1$

then
$$E(T) = E(N)\,E(X_i)$$

Proof.

We saw that $E(T|N) = N \cdot E(X_i)$.

Exercise: $T \in \mathcal{L}^1$.

It follows that

$$E(T) = E[E(T|N)]$$
$$= E[N \cdot E(X_i)]$$
$$= E(X_i) \cdot E(N). \qquad \square$$

What about $V[E(Y|X)]$? Is it $V(Y)$?

Example. $N \sim \text{Poisson}(\lambda)$ $\quad X_i \sim \text{Bernoulli}(p)$ $\quad T = \sum_1^N X_i$

$\Rightarrow T \sim \text{Poisson}(\lambda p)$ $\quad \Rightarrow V(T) = \lambda p$

$$V[E(T|N)] = V[N\,E(X_i)]$$
$$= p^2\,V(N)$$
$$= p^2 \lambda$$
$$\underset{p \in (0,1)}{<} p\lambda = V(T)$$

does it make sense?

Yes! averaging reduces the variance.

Recall that we defined the conditional expectation of Y given $X=x$ for $x\in X(\Omega)$ as:

$$E(Y|X=x)=\sum_y{}' y\, P(Y=y|X=x)$$

$$=\sum_y y\, q_x(y),$$

where $q_x(y)=p_{Y|X}(y|x):=P(Y=y|X=x)$.

It is easy to verify that q_x is a pmf and the expected value of the corresponding distribution is given by

$$\sum_y{}' y\cdot q_x(y)=E(Y|X=x).$$

The conditional expectation of Y given $X=x$ is the same as the expected value of the conditional distribution.

The variance of $Y\in L^2$ is given by EY^2-E^2Y.

We can define the conditional variance of Y given $X=x$ as

$$V(Y|X=x)=E(Y^2|X=x)-E^2(Y|X=x).$$

Note that if $Y\in L^2$ then $Y^2\in L^1$ so the RHS is well-defined for $x\in X(\Omega)$.

Analogously to the unconditional expectation, we have:

$$E[\varphi(Y)|X=x] = \sum_y{}' \varphi(y)\, q_x(y) . \qquad (q_x(y) = P_{Y|X}(y|x))$$

Therefore, for $Y \in L^2$ and $x \in X(\Omega)$,

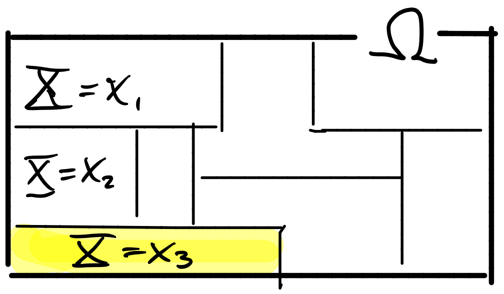
$$g(x) \stackrel{d}{=} E(Y^2|X=x) = \sum_y{}' y^2 q_x(y) \quad (<\infty)$$

$$h(x) \stackrel{d}{=} E(Y|X=x) = \sum_y{}' y\, q_x(y) \quad (\in \mathbb{R})$$

$$\Rightarrow \sigma(x) \stackrel{d}{=} V(Y|X=x) = g(x) - h^2(x) \in \mathbb{R}$$

Easy exercise: $\sigma(x) = \sum_y{}' \Big[y - \sum_{y'}{}' y' q_x(y')\Big]^2 \cdot q_x(y)$

That is the conditional variance of Y given $X=x$ is, again, the variance of the conditional distribution q_x.



$V(Y|X=x_3)$ = the variance of Y over $\{\omega : X(\omega)=3\}$

Example. $T = \sum_1^N X_i$ is a random sum, N is ind. of the iid $X_i \in L^2$, $n \in N(\Omega)$.

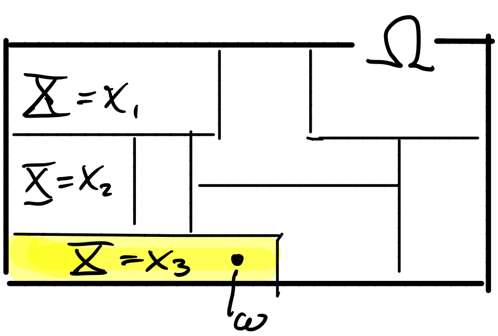
The conditional dist. of T given $N=n$ is the same as that of $S_n = \sum_1^n X_i$.

$$\Rightarrow V(T|N=n) = V(S_n) =$$

Again, for $Y \in L^2$ we can define the RV

$$V(Y|X)(\omega) := V(Y|X = X(\omega))$$
$$= v(X(\omega))$$
$$= g(X(\omega)) - h^2(X(\omega)),$$

where $g(x) = E(Y^2|X=x)$ and $h(x) = E(Y|X=x)$.



Suppose $\omega \in \Omega$ satisfies $X(\omega) = x_3$.
Then $V(Y|X)(\omega) = V(Y|X = x_3)$

$E(Y|X)$ and $V(Y|X)$ are const. on the events $\{X = x_n\}$, giving the mean and variance on each such set.

Example. $T = \sum_1^N X_i$ as before.

We found $V(T|N=n) = n \cdot V(X_i) = v(n)$

$$\Rightarrow V(T|N) = v(N) =$$