Stat 2911 Lecture Notes

Class 31, 2017

Uri Keich

Conditional expectation and random sums in a more general context, Continuous analog of Law of Total Probability, Prediction (Rice 4.4)

Conditional Expectation

If X and Y have a joint density f_{XY} which is cont. then for any x with $f_X(x) > 0$,

$$E[g(Y) \mid X = x] = \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x)\, dy$$

where $g: \mathbb{R} \to \mathbb{R}$ s.t. $g(Y) \in L^1$.

We define the RV

$$E(Y|X) = h(X),$$

where $h(x) = E(Y | X = x)$.

Claim. If X and Y are RVs defined on Ω and $Y \in L^1$

$$E[E(Y|X)] = E(Y).$$

Mixed dists.

1) $U \sim U(0,1)$ $\qquad X \mid U = p \sim \text{Binom}(m, p)$

2) N is a $\mathbb{N}$-valued RV, which is ind. of the iid

$X_1, X_2, \ldots \sim F_X$ and we define $T = \sum_1^N X_i$.

In all such cases $F_{XY}(x,y) = P(X \le x, Y \le y)$ is well-defined.

Moreover, we can often characterize the conditional dist.

Examples. 1) For $u \in [0,1]$

$$p_{X|U}(x|u) = \binom{m}{x} u^x (1-u)^{m-x} \mathbb{1}_{x \in \{0,1,\dots,m\}}$$

2) For $n \in \mathbb{N}$ with $P(N=n) > 0$ and $t \in \mathbb{R}$,

$$\begin{aligned} F_{T|N}(t|n) &= P(T \leq t \mid N=n) \\ &= \frac{P\left(\sum_{i=1}^{N} X_i \leq t, N=n\right)}{P(N=n)} \\ &= \frac{P\left(\sum_{i=1}^{n} X_i \leq t, N=n\right)}{P(N=n)} \\ &= \frac{P\left(\sum_{i=1}^{n} X_i \leq t\right) P(N=n)}{P(N=n)} \\ &= P\left(\sum_{i=1}^{n} X_i \leq t\right) \\ &= F_{S_n}(t) \qquad \left(S_n = \sum_{1}^{n} X_i\right) \end{aligned}$$

$\Rightarrow T|N=n \sim F_{S_n}$ for any F_X.

Exercise. If the ind. RVs X_i have densities f_i then $\sum_1^n X_i$ has density $((f_1 * f_2) * f_3) \dots * f_n = f_1 * \dots * f_n$.

$\Rightarrow$ If F_X has a density f_X (X_i are cont. RVs), then $F_{T|N=n}$ has a density which is given by $f_{S_n} = \underbrace{f_X * \dots * f_X}_{n\text{-fold convolution}}$.

In both examples we can define $E(Y|\underline{X})$ in the obvious way:

1) $$E(\underline{X}|U=u) =$$
$$\Rightarrow \quad E(\underline{X}|U) = mU.$$

2) If $\underline{X}_i \in L^1$ then $E(T|N=n) =$
$$\Rightarrow \quad E(T|N) = N\,E(\underline{X}_i) \quad \text{(looks familiar?)}$$

In particular, if $N \in L^1$ in addition to $\underline{X}_i \in L^1$,
$$E(T) = E[E(T|N)]$$
$$= E[N\,E(\underline{X}_i)]$$
$$= E(N)\,E(\underline{X}_i). \quad \text{(as before)}$$

The proof of
$$V(Y) = V[E(Y|\underline{X})] + E[V(Y|\underline{X})]$$
relied on $E[E(Y|\underline{X})] = E(Y)$ and $V(Y|\underline{X}) = E(Y^2|\underline{X}) - [E(Y|\underline{X})]^2$, therefore it extends to any jointly distributed $\underline{X}$ and Y with $Y \in L^2$.

In particular, if $\underline{X}_i, N \in L^2$ then
$$V(T) = V[E(T|N)] + E[V(T|N)] \quad (V(T|N=n) = ?)$$
$$= V[N \cdot E(\underline{X}_i)] + E[N \cdot V(\underline{X}_i)]$$
$$= E^2(\underline{X}_i)\,V(N) + V(\underline{X}_i)\,E(N). \quad \text{(as before)}$$

Recall the cont. analog of the law of total prob.:
if $\overline{X}$ is a cont. RV and A is an event then

$$P(A) = \int_{-\infty}^{\infty} P(A \mid \overline{X}=x) f_{\overline{X}}(x)\, dx.$$

Proof. With $Y := 1_A$, $Y \sim$, and therefore

$$E(Y) = E(1_A) =$$

$$h(x) = E(Y \mid \overline{X}=x)$$
$$= E(1_A \mid \overline{X}=x)$$
$$= P(A \mid \overline{X}=x).$$

$$\Rightarrow \quad P(A) = E(Y)$$
$$= E[E(Y \mid \overline{X})]$$
$$= E[h(\overline{X})]$$
$$= \int_{-\infty}^{\infty} h(x) f_{\overline{X}}(x)\, dx$$
$$= \int_{-\infty}^{\infty} P(A \mid \overline{X}=x) f_{\overline{X}}(x)\, dx \quad .$$

Prediction (of unobserved, not necessarily future events)

Suppose X and Y are jointly distributed RVs (two measurements from the same experiment) and suppose we know $X=x$.

How can we use this to predict the value of Y?

How shall we measure the quality of our prediction?

The MSE seems reasonable: given X, our predictor of Y is $g(X)$, where $g: \mathbb{R} \to \mathbb{R}$ (measurable).

The prediction MSE is

$$E[Y-g(X)]^2. \qquad (\text{need } Y \in L^2)$$

The prediction problem is to minimize the error, find:

$$\underset{g:\mathbb{R}\to\mathbb{R}}{\operatorname{argmin}} E[Y-g(X)]^2.$$

How do we find g?

Simpler: forget X (say, X is a const. RV) and solve

$$\underset{\alpha\in\mathbb{R}}{\operatorname{argmin}} E[(Y-\alpha)^2] \qquad (Y \in L^2)$$

$$E[(Y-\alpha)^2] = E(Y^2) - 2\alpha E(Y) + \alpha^2 \qquad [\mu = E(Y)]$$

$$= E(Y^2) - \mu^2 + \mu^2 - 2\alpha\mu + \alpha^2$$

$$= V(Y) + (\mu-\alpha)^2$$

$$\geq V(Y) \quad \text{with "=" iff } \alpha = \mu = E(Y)$$

$\Rightarrow$ The best MSE predictor of Y, given no other information, is $E(Y)$.

In particular, if $X \equiv c$ (const. RV), $g(c) := E(Y)$.

Note that g only needs to be defined on $X(\Omega)$.

<u>Claim.</u> If $\varphi: \mathbb{R} \to \mathbb{R}$ (s.t $\varphi(X) \in L^2$ and $Y \in L^2$) then

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

<u>Proof.</u> An intuitive argument that can be readily made rigorous in the discrete case, but generally requires a more general defn of conditional expectation.

$$h_1(x) = E[Y \cdot \varphi(X) \mid X = x]$$

$$= E[Y \cdot \varphi(x) \mid X = x]$$

$$= \varphi(x) \cdot E(Y \mid X = x)$$

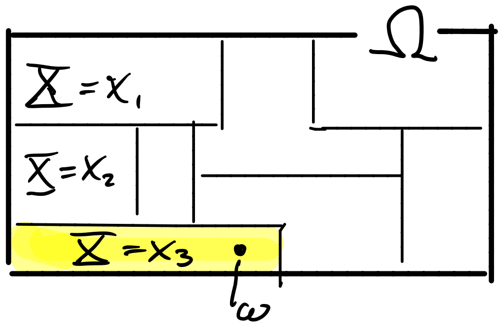
$$= \varphi(x) h(x).$$

$$\Rightarrow \quad E[Y\cdot\varphi(\underline{X})\,|\,\underline{X}] = h_1(\underline{X})$$
$$= \varphi(\underline{X})\,h(\underline{X})$$
$$= \varphi(\underline{X})\,E(Y\,|\,\underline{X}).$$

<u>Cor.</u> $\quad E[\varphi(\underline{X})\,|\,\underline{X}] = \quad .$

<u>Proof.</u> Take $Y \equiv 1$.

Back to the prediction problem.
Say $\underline{X}(\Omega) = \{x_k\}$ (discrete), then
we need to define g on $\underline{X}(\Omega) = \{x_k\}$.



$$E\{\underbrace{[Y-g(\underline{X})]}_{Z=Z_g}{}^2\} = E\{E(\underbrace{[Y-g(\underline{X})]}_{Z}{}^2\,|\,\underline{X})\}$$

$$E([Y-g(\underline{X})]^2\,|\,\underline{X}) = E[Y^2 - 2Y g(\underline{X}) + g(\underline{X})^2\,|\,\underline{X}]$$
$$= E(Y^2\,|\,\underline{X}) - E(Y\cdot 2g(\underline{X})\,|\,\underline{X}) + E[g(\underline{X})^2\,|\,\underline{X}]$$
$$\overset{?}{=} E(Y^2\,|\,\underline{X}) - 2g(\underline{X})E(Y\,|\,\underline{X}) + g(\underline{X})^2$$
$$= E(Y^2\,|\,\underline{X}) - E^2(Y\,|\,\underline{X}) + E^2(Y\,|\,\underline{X}) - 2g(\underline{X})E(Y\,|\,\underline{X}) + g(\underline{X})^2$$
$$= V(Y\,|\,\underline{X}) + [E(Y\,|\,\underline{X}) - g(\underline{X})]^2$$
$$\geq V(Y\,|\,\underline{X}) \quad \text{with "=" iff } g(\underline{X}) = E(Y\,|\,\underline{X}).$$

Claim. If $W, U \in L^1$ and $W \geq U$, then
$E(W) \geq E(U)$ with "=" iff $P(W=U)=1$.