Conditional Probabilities and Random Variables

Guy Lebanon

January 6, 2006

We looked previously at conditional probabilities $P(A|B)$ where $A, B \subset \Omega$ are events. In a similar way we can define conditional random variables and their associated cdf,pdf and pmf.

Assume we have a vector random variable $\vec{\boldsymbol{X}} = (X_1, \dots, X_n)$ and its joint cdf $F_{X_1,\dots,X_n}(\vec{\boldsymbol{x}})$, pdf $f_{X_1,\dots,X_n}(\vec{\boldsymbol{x}})$ (for continuous RVs) or pmf $p_{X_1,\dots,X_n}(\vec{\boldsymbol{x}})$ (for discrete RVs). We use the notation $\{X_j : j \neq i\}$ to denote all variables X_j but X_i. We saw previously that the marginal pdf f_{X_i} is given by integrating the joint pdf $f_{X_1,\dots,X_n}(\vec{\boldsymbol{x}})$ over all variables but x_i (similarly - for pmf with summation replacing integration). The same holds for a *vector*-marginal of $\vec{\boldsymbol{X}}$. The marginal pdf or pmf is obtained from the joint by integrating or summing over the unwanted variables e.g.,

$$f_{X_1,X_4}(x_1, x_4) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)\, dx_2 dx_3$$

$$p_{X_1,X_2}(x_1, x_2) = \sum_{x_3\in\mathbb{R}}\sum_{x_4\in\mathbb{R}} p_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4).$$

We can define a one-dimensional random variable $X_i|\{X_j = x_j : j \neq i\}$ e.g., $X_2|X_1 = x_1, X_3 = x_3$ or a vector RV for a subset of $(X_1, \dots, X_n)$ given assigned values to the remaining variables e.g., $X_3, X_1|X_2 = x_2$. Note that the variables that we condition on are assigned values (lower case letters). Thus there is no randomness in these RV with respect to the conditioned variables.

Associated with the conditional RVs $X_i|\{X_j : j \neq i\}$ is the conditional cdf:

$$F_{X_i|\{X_j=x_j:j\neq i\}}(x_i) = \frac{P(X_1 = x_1, \cdots, X_{i-1} = x_{i-1}, X_i \leq x_i, X_{i+1} = x_{i+1}, \cdots, X_n = x_n)}{P(X_1 = x_1, \cdots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \cdots, X_n = x_n)}$$

for example $F_{X_2|X_1=x_1,X_3=x_3}(x_2) = \frac{P(X_1=x_1,X_2\leq x_2,X_3=x_3)}{P(X_1=x_1,X_3=x_3)}$. If $\vec{\boldsymbol{X}}$ is continuous, the denominator in the definition of the conditional cdf is zero. In this case the denominator should be replaced by the pdf $f_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. A more careful definition is available - but it is beyond the scope of this class and requires measure theory.

The conditional pdf is defined as the derivative of the conditional cdf if it exists. It then follows that for continuous RVs the conditional pdf $f_{X_i|\{X_j=x_j:j\neq i\}}(x_i)$ is

$$f_{X_i|\{X_j=x_j:j\neq i\}}(x_i) = \frac{\frac{d}{dx_i}\int_{-\infty}^{x_i} f_{X_1,\dots,X_n}(x_1, \dots, x_n)dx_i'}{f_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)} = \frac{f_{X_1,\dots,X_n}(x_1, \dots, x_n)}{\int_{-\infty}^{\infty} f_{X_1,\dots,X_n}(x_1, \dots, x_n)dx_i}.$$

For example,

$$f_{X_2|X_1=x_1,X_3=x_3}(x_2) = \frac{d}{dx_2}F_{X_2|X_1=x_1,X_3=x_3}(x_2) = \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{\int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(x_1, x_2, x_3)dx_2}.$$

The conditional pmf is defined as

$$p_{X_i|\{X_j=x_j:j\neq i\}}(x_i) = \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_n = x_n)}$$

for example $p_{X_2|X_1=x_1}(x_2) = P(X_1 = x_1, X_2 = x_2)/P(X_2 = x_2)$.

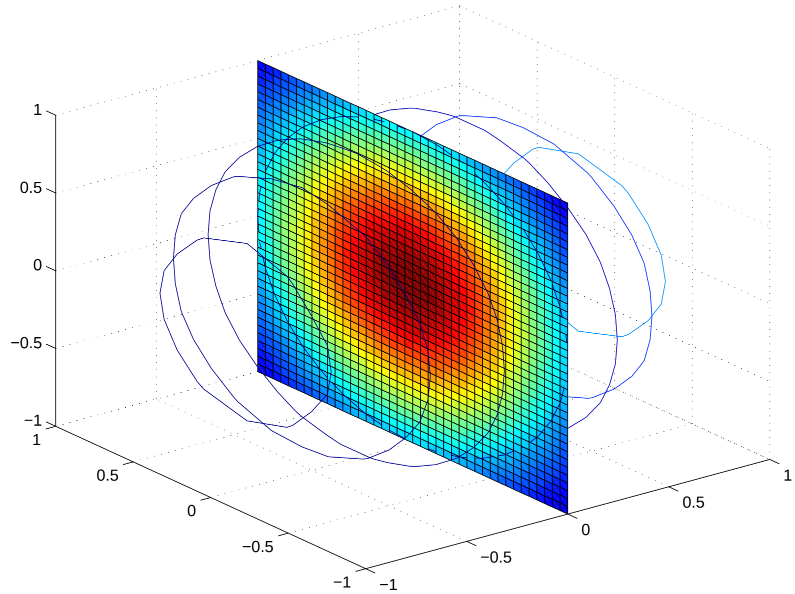

Figure 1: The joint pdf of three independent normal RV X_1, X_2, X_3 appear spherical (the magnitude of f_{X_1,X_2,X_3} decrease exponentially with the Euclidean distance from the center. The conditional distribution $f_{X_2,X_3|X_1=0}(x_2, x_3)$ is displayed as a cross section. It is proportional to the joint pdf $f_{X_1}(0)f_{X_2}(x_2)f_{X_3}(x_3)$ as X_1 is help fixed at 0. Since the three variables are independent, $f_{X_2,X_3|X_1=0}(x_2, x_3) = f_{X_2,X_3}(x_2, x_3) = f_{X_2}(x_3)f_{X_3}(x_3)$ (verify!).

The conditional cdf,pmf and pdf are useful in computing conditional probabilities. This is done in the same way as for the non-conditional case except that all quantities are conditioned on known variables assigned to some of the RVs:

$$P(X \in A|Y = y, Z = z) = \begin{cases} \int_A f_{X|Y=y,Z=z}(x)dx & X \text{ is continuous} \\ \sum_{x\in A} p_{X|Y=y,Z=z}(x) & X \text{ is discrete} \end{cases}.$$

It is interesting to note the duality between the joint, the marginal and the conditional pdf and pmf

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_i|\{X_j=x_j:j\neq i\}}(x_i) f_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = p_{X_i|\{X_j=x_j:j\neq i\}}(x_i) p_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

They lead to pdf and pmf versions of generalizations of Bayes Law $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ for example

$$f_{X_1,X_4|X_2=x_2,X_3=x_3}(x_1, x_4) = \frac{f_{X_2,X_3|X_1=x_1,X_4=x_4}(x_2, x_3) f_{X_1,X_4}(x_1, x_4)}{f_{X_2,X_3}(x_2, x_3)}$$

Example: Suppose that a point X is chosen from a uniform distribution in the interval $[0, 1]$ and that after $X = x$ is observed a point Y is drawn from a uniform distribution on the interval $[x, 1]$. What is the marginal f_Y?

The joint $f_{X,Y}$ is the product of the marginal f_X and the conditional $f_{Y|X=x}$: $f_{X,Y}(x, y) = \frac{1}{1-x} \cdot 1$ for $0 < x < y < 1$ and 0 otherwise. The marginal is therefore $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx = \int_0^y \frac{1}{1-x}dx = -\log(1-y)$ for $0 < y < 1$ and 0 otherwise. We can also find the other conditional $f_{X|Y=y}(x) = f_{X,Y}(x, y)/f_Y(y) = \frac{-1}{(1-x)\log(1-y)}$ for $0 < x < y < 1$ and 0 otherwise.