Will Monroe
CS 109

Lecture Notes #12
July 21, 2017

Joint Distributions

Based on a chapter by Chris Piech

Often you will work on problems where there are several random variables (often interacting with one another). We are going to start to formally look at how those interactions play out.

For now we will consider joint probabilities with two random variables X and Y.

Discrete Case

In the discrete case, a **joint probability mass function** tells you the probability of any combination of values for the two random variables, $X = a$ and $Y = b$:

$$p_{X,Y}(a, b) = P(X = a, Y = b)$$

(The “,” means “and”.) If you want to calculate the probability of an event only for one variable, you can calculate a “marginal" from the joint probability mass function:

$$p_X(a) = P(X = a) = \sum_y P(X = a, Y = y) = \sum_y P_{X,Y}(a, y)$$
$$p_Y(b) = P(Y = b) = \sum_x P(X = x, Y = b) = \sum_x P_{X,Y}(x, b)$$

This is just the first version of the general law of total probability ($P(A) = \sum_i P(E_i A)$), rewritten with probability mass functions.

In the discrete case, we can define the function $p_{X,Y}$ non-parametrically. Instead of using a formula for p we simply state the probability of each possible outcome.

Continuous Case

In the continuous case, a **joint probability density function** tells you the relative likelihood of values for the two random variables, $X = a$ and $Y = b$.

Random variables X and Y are *jointly continuous* if there exists a probability density function (PDF) $f_{X,Y}$ such that:

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \int_{a_1}^{a_2} dx \int_{b_1}^{b_2} dy\, f_{X,Y}(x, y)$$

Using the PDF we can compute marginal probability densities:

$$f_X(a) = \int_{-\infty}^{\infty} dy\, f_{X,Y}(a, y)$$
$$f_Y(b) = \int_{-\infty}^{\infty} dx\, f_{X,Y}(x, b)$$

Joint Cumulative Distribution Functions

If X and Y are discrete random variables with joint probability mass function $p_{X,Y}(x, y)$, then the **joint cumulative distribution function** $F_{X,Y}(x, y)$ is defined as

$$F_{X,Y}(a, b) = P(X \leq a, Y \leq b) = \sum_{x=-\infty}^{a} \sum_{y=-\infty}^{b} p_{X,Y}(x, y)$$

For jointly continuous random variables with joint probability density function $f_{X,Y}(x, y)$:

$$F_{X,Y}(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{a} dx \int_{-\infty}^{b} dy\, f_{X,Y}(x, y)$$

The joint CDF is especially useful for computing probabilities that jointly continuous random variables lie in certain ranges. Let $F(a, b)$ be the joint cumulative distribution function (CDF) of X and Y. Then

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F(a_2, b_2) - F(a_1, b_2) + F(a_1, b_1) - F(a_2, b_1)$$

Lemma for Computing Expectations

If Y is a non-negative random variable the following hold (for discrete and continuous random variables respectively):

$$E[Y] = \sum_{i=0}^{\infty} P(Y > i) = \sum_{i=1}^{\infty} P(Y \geq i)$$

$$E[Y] = \int_{0}^{\infty} di\, P(Y > i) \tag{1}$$

Expectation with Multiple RVs

Expectation over a joint isn't nicely defined because it is not clear how to compose the multiple variables. However, expectations over functions of random variables (for example sums or multiplications) are nicely defined: $E[g(X, Y)] = \sum_{x,y} g(x, y)p(x, y)$ for any function $g(X, Y)$. When you expand that result for the function $g(X, Y) = X + Y$ you get a beautiful result:

$$\begin{aligned}
E[X + Y] = E[g(X, Y)] &= \sum_{x,y} g(x, y)p(x, y) = \sum_{x,y} [x + y]p(x, y) \\
&= \sum_{x,y} xp(x, y) + \sum_{x,y} yp(x, y) \\
&= \sum_{x} x \sum_{y} p(x, y) + \sum_{y} y \sum_{x} p(x, y) \\
&= \sum_{x} xp(x) + \sum_{y} yp(y) \\
&= E[X] + E[Y]
\end{aligned}$$

This can be generalized to multiple variables:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Example 1

A disk surface is a circle of radius R. A single point imperfection is uniformly distributed on the disk with joint PDF:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{else} \end{cases}$$

Problem: What is $f_X(y)$?

Solution:
$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} dy\, f_{X,Y}(x, y) \\ &= \frac{1}{\pi R^2} \int_{x^2+y^2 \leq R^2} dy \\ &= \frac{1}{\pi R^2} \int_{y=-\sqrt{R^2-x^2}}^{+\sqrt{R^2-x^2}} dy \\ &= \frac{2\sqrt{R^2 - x^2}}{\pi R^2} \end{aligned}$$

Problem: What is $E[D]$?

Solution:
$$\begin{aligned} E[D] &= \int_0^R da\, P(D > a) \qquad \text{by lemma (1)} \\ &= \int_0^R da \left(1 - \frac{a^2}{R^2}\right) \\ &= \left[a - \frac{a^3}{3R^2}\right]_{a=0}^{R} = \frac{2}{3}R \end{aligned}$$