Joint Probability Distributions

So far, we have mostly studied probability models for a single random variable. A named distribution, for example, can be used to compute probabilities, averages, or variances for a *single* X. In most practical applications, we are not only interested in the distribution of separate variables, but in the way in which variables are *related to each other*.

The Discrete Case

Examples:

- Do people who buy tortilla chips at a supermarket also tend to buy salsa?
- Do women generally get better grades in math than men?

DEFINITION: Let X and Y be two discrete random variables. The JOINT PROBABILITY MASS FUNCTION $p(x, y)$ is defined for every pair of possible values (x, y) as

$$p(x, y) = P(X = x \text{ and } Y = y)$$

It can be written down in table form. The MARGINAL PROBABILITY MASS FUNCTIONS of X and Y, denoted by $p_X(x)$ and $p_Y(y)$ are the row, and column sums from the table, respectively.

$$p_X(x) = \sum_y p(x, y), \qquad p_Y(y) = \sum_x p(x, y)$$

Example: Consider the following experiment: We randomly draw one of the numbers 1,2,3 out of a hat and then we toss a fair coin *that* number of times.

Let X = # we draw,
Y = # of heads in the coin tosses.

(a) Are X and Y independent? If not, in what way do they depend on each other?

(b) Write down the joint PMF table for X and Y.

(c) Compute the marginal distributions for X and Y.

Probabilities are computed by adding numbers from the table (Note that all the table entries should always sum to one).

DEFINITION: If X and Y are discrete RV's with joint PMF $p(x,y)$, then

$$P[(X,Y)\in A] = \sum_{(x,y)\in A}\sum p(x,y)$$

Example: (cont.)

(d) Find the probability $P(X = Y)$.

Poll Question 9.1

The Continuous Case

Similarly to the discrete case, we can also define simultaneous probabilities for two continuous random variables. As before, sums will be replaced by integrals in the continuous case.

DEFINITION: Let X and Y be continuous random variables. Then $f(x,y)$ is a JOINT PROBABILITY DENSITY FUNCTION for X and Y if for any two dimensional set A

$$P((X,Y)\in A) = \int\int_A f(x,y)dxdy$$

In particular, if A is a rectangle: $\{(x,y) : a \le x \le b, c \le y \le d\}$, then

$$P[(X,Y)\in A] = P(a \le X \le b, c \le Y \le d) = \int_a^b\int_c^d f(x,y)dydx$$

Similarly to the discrete case, the MARGINAL DENSITY FUNCTIONS are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

Example: Consider the joint pdf given by

$$f(x,y) = \begin{cases} \frac{3}{4}xy^2 & 0 \le x \le 1, 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(X < 0.5)$.

Independence

Recall: Two EVENTS A and B are independent, if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Since we can consider $X = x$ or $Y = y$ as events, we can translate this definition for the case of discrete or continuous random variables.

DEFINITION: Two random variables are said to be INDEPENDENT if for every pair of x and y values

$$p(x, y) = p_X(x) \cdot p_Y(y) \text{ when } X \text{ and } Y \text{ are discrete}$$

or

$$f(x, y) = f_X(x) \cdot f_Y(y) \text{ when } X \text{ and } Y \text{ are continuous}$$

Note: If you want to prove independence, you have to show that the above relationship holds for *every* pair x and y. If you want to show that random variables are not independent, then one counterexample is sufficient.

Example: Refer to the example from page **??**. Are X and Y independent in this case?

Conditional Distributions

Especially, when a problem contains more than one random variable, and the random variables depend on each other, conditional probabilities may become of special interest. Suppose for a randomly chosen San Jose State student X is the number of close family relatives with college education and Y is the number of semesters the student needs to graduate. Then it makes sense to ask questions like: "If you are the first in your family to go to college ($X = 0$), what is the probability that you will graduate in 4 years or less ($Y \leq 8$)?"

DEFINITION: Let X and Y be discrete random variables. Then the CONDITIONAL PMF OF Y GIVEN $X = x$ is given by

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)}$$

Let X and Y be continuous random variables. Then the CONDITIONAL PDF OF Y GIVEN $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \qquad -\infty < y < \infty$$

Expected Values and Variance

Recall: You have learned how to compute expected values and variances for single discrete and continuous random variables in chapters 3 and 4.

$$\text{Discrete:} \quad E(X) = \sum_x xp(x), \quad E(X^2) = \sum_x x^2 p(x)$$

$$\text{Continuous:} \quad E(X) = \int_{-\infty}^{\infty} xf(x)dx, \quad E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$

$$V(X) = E(X^2) - E(X)^2$$

DEFINITION: Let X and Y be jointly distributed random variables with PMF $p(x,y)$ or PDF $f(x,y)$. Then the expected value of any function of X and Y $h(x,y)$ can be computed as

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)p(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f(x,y)dx\,dy & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

Remark: Expected values and variances can be used to describe the behavior of a single random variable. Think back to chapter 3, where we have used $E(X)$ to describe the value of a game or investment strategy and $V(X)$ to describe the risk associated with the game or strategy. If we want to consider the joint behavior of two random variables, then another quantity becomes of interest: How do the random variables influence each other?

DEFINITION: The COVARIANCE of X and Y is denoted by $\text{Cov}(X,Y)$ and can be computed as

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

It describes the JOINT BEHAVIOR of the variables X and Y. However, this quantity depends on the units that X and Y are measured in.

DEFINITION: The CORRELATION COEFFICIENT of X and Y, denoted by $\text{Corr}(X,Y)$ or $\rho(X,Y)$ is defined by

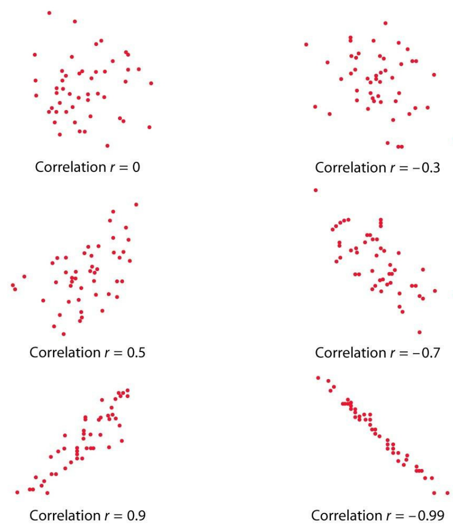
" rho "

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Remarks:

- $-1 \le \rho \le 1$
- ρ does not depend on the units of X and Y.
- If X and Y are independent, then $\rho = 0$. The reverse is not necessarily true!
- $\rho = \pm 1$ iff $Y = aX + b$ for some numbers a, b with $a \neq 0$.

The correlation coefficient measures the strength and direction of a linear relationship between X and Y. The sign of r describes the direction of the association. Positive correlation means that large x-values are associated (on average) with large y values. Negative correlation means that large x-values are associated (on average) with small y-values. The absolute value of r describes the strength of the association. Values close to 0 represent weak association, values close to ± 1 represent strong association.



Poll Question 9.2

Example: Forty students in a statistics class were asked how many siblings (X) and how many sisters (Y) they have. The results are listed in the table below

Siblings, X \ Sisters, Y	0	1	2	3	
0	8	0	0	0	
1	8	9	0	0	
2	1	8	2	0	
3	0	1	2	0	
4	0	0	0	1	
					40

Suppose one student in the class is selected at random. Let X and Y denote the number of siblings and number of sisters of that student, respectively.

(a) Find $p(2, 1)$ and interpret your answer.

$\frac{8}{40}$

Poll Question 9.3

(b) Find the marginal PMFs of X and Y.

(c) What is the average number of siblings (sisters) of students in that class?

E(X)=1*0.425+2*0.275+3*0.075+4*0.025
E(Y)=1*0.45+2*0.1+3*0.025

(d) What is the probability that a student in this class has no brothers?

Poll Question 9.4

(e) Given that a student in this class has no sisters, find the probability that the student has at least one brother.

P(X-Y>=1|Y=0)=0.225/0.425

(f) Are the random variables X and Y independent in this example?

(g) Find $E(XY)$.

E(XY)=x*y*p(x,y)(for all x,y)

(h) Compute the correlation coefficient of X and Y (using the facts that $\sigma_X = 0.954$ and $\sigma_Y = 0.741$). Interpret this value.

Distribution of Linear Combinations

Previously, we have discussed the joint distributions of two discrete or continuous random variables. Furthermore, we have learned how to compute expected values and variances of jointly distributed random variables and to compute covariances and correlations to describe linear association. These same quantities can be used to investigate the behavior of linear combinations of random variables.

DEFINITION: Given a collection of random variables $X_1, X_2, \ldots, X_n$ and numerical constants $a_1, a_2, \ldots, a_n$, the RV

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

is called a LINEAR COMBINATION of the X_i's.

PROPOSITION: Suppose $X_1, \ldots, X_n$ are random variables with means $\mu_1, \ldots, \mu_n$ and variances $\sigma_1^2, \ldots, \sigma_n^2$, respectively.

- Then, whether or not the X_i's are independent

$$E(a_1X_1 + \cdots + a_nX_n) = a_1\mu_1 + \cdots + a_n\mu_n$$

  Recall that the expectation of a sum is *always* the sum of the expectations.

- If the $X_1, \ldots, X_n$ are independent

$$V(a_1X_1 + \cdots + a_nX_n) = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$$

- If the $X_1, \ldots, X_n$ are not independent

$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jCov(X_i, X_j)$$

Example: Let X_1 and X_2 be random variables with means $\mu_1 = 1$ and $\mu_2 = 5$ and variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 4$ and covariance $Cov(X_1, X_2) = \cancel{3}$ 0.3 Compute

(a) $E(X_1 - 2X_2)$ =E(X1)-2E(X2)=-9

Poll Question 9.5

(b) $V(X_1 + 2X_2)$ X1,X2 are not independent cause Cov(X1,X2)!=0
V(X1)+4V(X2)+2Cov(x1,x2)+2Cov(x2,x1)

(c) $V(X_1 - 2X_2)$ V(X1)+4V(X2)-2Cov(x1,x2)-2Cov(x2,x1)

Facts: You have seen some of the facts below mentioned in class (or proved them on your homework).

- The sum of n independent Bernoulli(p) random variables has a Binomial(n, p) distribution.
- The sum of two independent Binomials (say, Binomial(n_1, p) and Binomial(n_2, p)) with the same success probability p is Binomial($n_1 + n_2, p$).
- The sum of r independent Geometric(p) random variables has a Negative Binomial(r, p) distribution.
- The sum of two independent Negative Binomial random variables (say NegBin(r_1, p) and NegBin(r_2, p) with the same success probability p) is NegativeBinomial($r_1 + r_2, p$).
- The sum of n independent Poisson random variables, with parameters $\lambda_1, \ldots, \lambda_n$, respectively has a Poisson distribution with parameter $\lambda_1 + \cdots + \lambda_n$.
- The sum of r independent Exponential(λ) random variables has a Gamma(r, λ) distribution.
- The sum of two independent Gamma (say Gamma(r_1, λ) and Gamma(r_2, λ) with the same rate λ) is Gamma($r_1 + r_2, \lambda$).
- The sum of n independent Normal random variables with means $\mu_1, \ldots, \mu_n$ and variances $\sigma_1^2, \ldots, \sigma_n^2$ has a Normal distribution with mean $\mu_1 + \cdots + \mu_n$ and variance $\sigma_1^2 + \cdots + \sigma_n^2$.