Joint Distributions

So far, we have studied probability models for a single random variable. A named distribution, for example, can be used to compute probabilities, means, or variances for a *single* X. In most 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 store also tend to buy salsa?
- Is the number of times students "look up" homework problems on the internet related to their success on a midterm exam?

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

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

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 60. 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 = number we draw,
Y = number 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 PMFs 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

$$P\left((X, Y) \in A\right) = \sum_{(x,y)\in A} p(x, y)$$

Example 60. (cont.)

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

Definition: The joint cumulative distribution function of X and Y is given by

$$F(a, b) = P(X \leq a, Y \leq b), \qquad -\infty < a, b < \infty$$

The marginal distribution functions can be obtained from the joint distribution.

$F_X(a) =$

And similarly, $F_Y(b) = F(\infty, b)$.

Example 61. Express $P(X > a, Y > b)$ in terms of the joint and marginal distribution functions.

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 C

$$P((X,Y) \in C) = \iint_C f(x,y)dx\,dy$$

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

$$P((X,Y) \in C) = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x,y)dydx$$

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

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

Definition: As in the discrete case, the joint distribution function is given by

$$F(a,b) = P(X \leq a, Y \leq b) = P(X \in (-\infty, a], Y \in (-\infty, b]), \qquad -\infty < a, b < \infty$$

and the joint density and joint distribution function are related as usual

$$f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)$$

Example 62. Consider the joint PDF given by

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

Find $P(X < 0.5)$.

Example 63. The joint density of X and Y is given by

$$f(x,y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find the density of the random variable X/Y.

Joint density and distribution functions can, of course, also be defined for more than two random variables. That becomes necessary frequently, when likelihood functions of n observations need to be computed, for example.

$$F(a_1, a_2, \ldots, a_n) = P(X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n)$$

The function $f(x_1, x_2, \ldots, x_n)$ is called the joint density function of the n random variables $X_1, \ldots, X_n$ if for any set $C \in \mathbb{R}^n$,

$$P\left((X_1, \ldots, X_n) \in C\right) = \int \cdots \int_{(x_1,\ldots,x_n)\in C} f(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

Example 64. The multinomial distribution
Suppose a sequence of n identical and independent experiments is performed and suppose that each experiment can result in r different outcomes with respective probabilities $p_1, \ldots, p_r$ $(\sum_{i=1}^{r} p_i = 1)$. Let X_i denote the number of experiments that result in outcome number i, then

$$P(X_1 = n_1, \ldots, X_r = n_r) = \begin{cases} & n_1 + \cdots + n_r = n \\ 0 & \text{else} \end{cases}$$

Example 65. What is the probability that in five rolls of a fair die you see exactly one two, two threes, one four, and one six (in any order)?

Independence

Recall, that 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

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

or

$$F(x, y) = F_X(x)F_Y(y) \text{ for either discrete or continuous RVs}$$

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 66. Refer to the example from page 46. Are X and Y independent in this case?

Fact: The continuous (or discrete) random variables X and Y are independent if and only if their joint probability density (mass) function factors into two parts: one part that depends only on x and one part that depends only on y:

$$f(x, y) = h(x)g(y), \qquad -\infty < x, y < \infty$$

Proof:

Example 67. Suppose the joint density of X and Y is given by

$$f(x,y) = \begin{cases} 24xy & 0 < x, y < 1, 0 < x + y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent?

The concept of independence can be extended to more than two random variables. We say that the random variables $X_1, \ldots, X_n$ are independent, if for all sets of real numbers $A_1, \ldots, A_n$,

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

which is equivalent to

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \text{or} \quad p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

which is also equivalent to

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

An infinite collection of random variables is said to be independent, if every finite subset is independent.

Sums of Independent Random Variables

There are many situations in which the distribution of the sum of two independent random variables X and Y is of interest. This is called a *convolution*.

Example 68. Let X and Y be independent continuous random variables with densities $f_X(x), f_Y(y)$ and distribution functions $F_X(x), F_Y(y)$, respectively. Find the CDF and PDF of $X + Y$.

Definition: The CDF of $X + Y$ obtained above is called the CONVOLUTION of the distributions of X and Y.

Fact: The convolution density of two independent random variables X and Y with respective densities is $f_X(x)$ and $f_Y(y)$ is

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy$$

We will next investigate situations in which the sum of two independent random variables from specific named distributions results in another named random variable.

Example 69. Suppose $X \sim \text{Normal}(\mu = 0, \sigma^2)$ and $Y \sim \text{Normal}(0, 1)$ are independent. Find the density of $X + Y$.

Fact: If $X_1 \sim \text{Normal}(\mu_1, \sigma_1^2)$ and $Y \sim \text{Normal}(\mu_2, \sigma_2^2)$ are independent, then $X+Y$ has a Normal distribution with mean $\mu = \mu_1 + \mu_2$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Fact: Through induction, it follows that the sum of n independent Normal random variables $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$ is also Normal.

$$\sum_{i=1}^{n} X_i \sim \text{Normal}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Example 70. Let $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ be independent. Find the distribution of $X + Y$.

$$P(X+Y) = \sum_{k=0}^{n} P(X=n-k, Y=k)$$
$$= P(X=n-k) \cdot P(Y=k)$$

Use Binomial theorem

Example 71. Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$ be independent. Find the distribution of $X + Y$.

Combinatorial Identity:

$$\binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k}$$

Example 72. Let $X \sim \text{Gamma}(s, \lambda)$ and $Y \sim \text{Gamma}(t, \lambda)$ be independent. Find the distribution of $X + Y$.

Fact: Let $Z_1, \ldots, Z_n$ be independent standard Normal random variables. Then $Y = \sum_{i=1}^{n} Z_i^2$ has χ^2- distribution with n degrees of freedom.

To summarize, we have now shown (or can deduce from what was previously shown) the following relationships between named random variables. That is, if X and Y are independent random variables with the distributions listed below, then their sums also have named distributions.

Distribution of X	Distribution of Y	Distribution of $X + Y$
Bernoulli(p)	Bernoulli(p)	Binomial(n=2,p)
Binomial(n, p)	Binomial(m, p)	Binomial($n + m, p$)
Poisson(λ_1)	Poisson(λ_2)	Poisson($\lambda_1 + \lambda_2$)
Geometric(p)	Geometric(p)	Negative Binomial($r = 2, p$)
Exponential(λ)	Exponential(λ)	Gamma(r=2, λ)
Gamma(s, λ)	Gamma(t, λ)	Gamma($s + t, \lambda$)
Chi-Squared($df = n$)	Chi-squared($df = m$)	Chi-square(df=n+m)
Normal(μ_1, σ_1^2)	Normal(μ_2, σ_2^2)	Normal($\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2$)

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

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

Example 73. Let X and Y be independent Poisson random variables with respective parameters λ_1 and λ_2. Find the conditional distribution of X given that $X + Y = n$.

Example 74. Suppose the joint density of X and Y is

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y} & 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find $P(X > 1|Y = y)$.

The t-Distribution

Definition: Let $Z \sim$ Normal(0,1) and $V \sim \chi^2_n$ be independent. Then

$$t = \frac{Z}{\sqrt{V/n}} \sim t(n)$$

is said to have Student's *t*-distribution with n degrees of freedom.

Historical Note: This distribution is named for William Sealy Gossett. Gossett was a statistician and brewer who worked for the Guinness brewery in the early 1900's in England. He published his statistical papers under the pseudonym "Student" because the Guinness brewery had a policy forbidding employees to publish papers to not divulge any trade secrets. Gossett was a contemporary of Karl Pearson and R.A. Fisher.

Fact: A random variable with a t-distribution with ν degrees of freedom has a continuous distribution on $\mathbb{R}$ with probability density function

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, x \in \mathbb{R}$$

Proof:

Remark: The degree of freedom of the t distribution is defined for any value $n > 0$. But in practice, the degrees of freedom are usually positive integers.

Remark: Similarly to the Normal distribution, probabilities for a t distribution are not computed through a closed-form formula of the CDF but are available in tables and through software.

Example 75. Let $X \sim t(n = 5)$.

(a) Find $P(X > 0.5)$

(b) Find x, such that $P(X \leq x) = 0.2$

Fact: For $n \to \infty$, the t-distribution converges in distribution to the standard normal distribution. That means that for any fixed x

$$F_{t_\nu}(x) \to \Phi(x), \quad \text{as } \nu \to \infty$$

Proof:

In statistics, a common procedure for comparing the mean of a population to a fixed constant or for comparing the means of two independent populations to each other are one-sample and two-sample t-tests. In these tests, the sample mean $\bar{x}$ and the sample variance s^2 of an independent sample of observations $x_1, \dots, x_n$ are computed as follows:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Fact: If $x_1, \dots, x_n$ are an independent sample from a Normal population with mean μ and variance σ^2, then $\bar{x}$ and s^2 are independent and

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1)$$

and

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Proof: Will be provided when we discuss Section 7.8 of the text in a few weeks.

Note: Let $x_1, \ldots, x_n$ be an independent random sample from a Normal distribution with mean μ and variance σ^2. The t-test statistic is defined as

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Then the t-test statistic has a t-distribution with $\nu = n - 1$ degrees of freedom.

Proof:

The F-Distribution

Definition: Let X be a continuous random variable with density:

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2} x^{(m/2)-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(n+mx)^{(m+n)/2}}, \quad x > 0$$

Then X is said to have an F-distribution with degrees of freedom m and n. We will write $X \sim F_{m,n}$.

Remark: The F-distribution arises frequently in quotients of sums of squares (for instance in ANOVA or regression). m is also referred to as the numerator degree of freedom and n is also referred to as the denominator degree of freedom.

Fact: If $X \sim F_{m,n}$ then

$$E[X] = \frac{n}{n-2}, \quad \text{for } n > 2, \qquad Var(X) = \frac{2n^2(n+m-2)}{m(n-2)^2(n-4)}, \quad \text{for } n > 4$$

Theorem: Let $U \sim \chi^2_m$ and $V \sim \chi^2_n$ be independent. Then

$$X = \frac{U/m}{V/n} \sim F_{m,n}$$

That is, the quotient of two independent χ^2-random variables, each scaled by their respective degree of freedom, has an F distribution.

Joint Distributions of Function of Random Variables

Recall, that we have previously studied how the density of a random variable $Y = g(X)$ can be found (under certain circumstances) from the density of X. Now, we will extend this same concept to functions of jointly distributed random variables.

That is, assume that X_1 and X_2 are jointly distributed random variables with joint density f_{X_1,X_2}. Suppose that the random variables Y_1 and Y_2 can be computed as functions of X_1 and X_2, say $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$. Assume that the functions g_1 and g_2 satisfy the following conditions:

- The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, say $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
- The functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that the 2×2 determinant

$$J(x_1, x_2) = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0$$

  at all points (x_1, x_2). (This matrix of partial derivatives is called the Jacobian of the function $(x_1, x_2) \mapsto (y_1, y_2)$. Since it's usually the determinant of this matrix that is needed for computations, some people also refer to the determinant of the matrix of partial derivatives as "the Jacobian".)

Fact: If the two conditions above are satisfied, then the random variables Y_1 and Y_2 have the joint distribution

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)|J(x_1, x_2)|^{-1}$$

Example 76. Let $X_1 \sim$ Normal(0,1) and $X_2 \sim$ Normal(0,1) be independent. Find the joint distribution of $X_1 + X_2$ and $X_1 - X_2$.

Multivariate Normal Distribution

One of the most commonly used joint distributions in practice is the multivariate Normal distribution. It can be defined for any number of variables but for now, we will focus on the two-dimensional case.

Definition: The random variables X_1 and X_2 are said to have a bivariate Normal distribution, if for constants μ_1, μ_2 and $\sigma_1^2 > 0, \sigma_2^2 > 0, -1 < \rho < 1$, their joint density function is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right)$$

or, in vector form

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

$\boldsymbol{\mu}$ is called the mean vector of the distribution and Σ is called the covariance matrix. ρ is the correlation between X_1 and X_2. Shown below is the graph of a two-dimensional Normal density function with mean vector $\mu = (1,1)'$ and covariance matrix

$$\Sigma = \begin{pmatrix} 0.9 & 0.4 \\ 0.4 & 0.3 \end{pmatrix}$$

In general, the mean vector will determine the center (peak) of a two-dimensional Normal distribution and the eigenvectors of the covariance matrix will determine the axes of orientation.