JOINT DENSITIES - RANDOM VECTORS - REVIEW

Joint densities describe probability distributions of a *random vector* $\mathbf{X}$: an n-dimensional vector of random variables, i.e., $\mathbf{X} = (X_1, \ldots, X_n)$, where all $X_i's$ are rvs.

Discrete random vectors are described by the *joint probability density function of* X_i (or joint pdf; also called the *joint probability mass function* (pmf)), $i = \{1, \ldots, n\}$ denoted by

$$P(X = x) = P(s \in S : X_i(s) = x_i \text{ for all } i) = p_X(x_1, x_2, \ldots)$$

Computing probabilities: For any event $A \subseteq \mathbb{N}^n$, $P(X \in A) = \sum_{a \in A} P(X = a)$.

Continuous random vectors are described by the *joint probability density function of* X_i *(or joint pdf)* denoted by $f_X(x_1, \ldots, x_n)$. The pdf has the following properties:

1. $f_X(x_1, \ldots, x_n) \geq 0$ for every $x \in \mathbb{R}^n$.
2. $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1$.
3. For event $A \subset \mathbb{R}^n$, $P(X \in A) = \int_A f_X(x) dx = \int \cdots \int f_X(x_1, ..., x_n) dx_1 \cdots dx_n$.

Marginal distributions. Let X be a continuous/discrete random vector having a joint distribution with pdf/pmf $f(x)$. Then, the one-dimensional distributions of X_i are called *marginal distributions*. We compute the marginal distributions as follows:

If X is a discrete random vector, then the distributions of X_i are given by:

$$p_{X_i}(k_i) = \sum_{\text{all other } X_j = x_j} P(x).$$

If X is a continuous random vector (see p.172), then the marginal distributions of X_i are:

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x) dx_1 \cdots dx_n}_{\text{Integrate over all } x_{j \neq i}}$$

Joint cdf of a vector X. The joint cumulative distribution function of X is defined by

$$F_X(x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

Theorem. Let $F_{X,Y}(u, v)$ be a joint cdf of the vector (X, Y). Then the joint pdf of (X, Y), $f_{X,Y}$, is given by second partial deriveative of the cdf. That is $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$

INDEPENDENT RANDOM VARIABLES

Definition. Two random variables are called independent iff for every intervals A and B on the real line $P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$.

Theorem. The random variables X_i are independent iff their joint pdf is a product of the marginal pdfs, that is

$$f_{X_1,X_2,X_3,\ldots,X_n}(x_1,x_2,x_3,\ldots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n),$$

where $f_{X_1,X_2,X_3,\ldots,X_n}(x_1,x_2,x_3,\ldots,x_n)$ is the joint pdf of the vector $(X_1,X_2,X_3,\ldots,X_n)$, and $f_{X_1}(x_1)$, $f_{X_2}(x_2), \cdots, f_{X_n}(x_n)$ are the marginal pdf's of the variables $X_1,X_2,X_3,\ldots,X_n$.

NOTE: Random variables X and Y are independent iff $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, where $F(x,y)$ is the joint cdf of (X,Y), and $F_X(x)$ and $F_Y(y)$ are the marginal cdf's of the X and Y, respectively.

Random Sample. A random sample of size n from distribution f is a set $X_1,X_2,X_3,\ldots,X_n$ of independent and identically distributed (iid), with distribution f, random variables.

CONDITIONAL DISTRIBUTIONS

Let (X,Y) be a random vector with some joint pdf or pmf. Consider the problem of finding the probability that X=x **AFTER** a value of Y was observed. To do that we develop *conditional distribution* of X given Y=y.

Definition. If (X,Y) is a discrete random vector with pmf $p_{X,Y}(x,y)$, and if $P(Y=y)>0$, then the *conditional distribution* of X given Y=y is given by the *conditional pmf*

$$p_{X|Y=y}(x) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

Similarly, if $P(X=x)>0$, then the *conditional distribution* of Y given X=x is given by the *conditional pmf* $p_{Y|X=x}(y) = \frac{p_{X,Y}(x,y)}{p_X(x)}$.

Definition. If (X,Y) is a continuous random vector with pdf $f_{X,Y}(x,y)$, and if $f_Y(y)>0$, then the *conditional distribution* of X given Y=y is given by the *conditional pdf*

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Similarly, if $f_X(x)>0$, then the *conditional distribution* of Y given X=x is given by the *conditional pdf* $f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$.

Independence and conditional distributions. If random variables X and Y are independent, then their marginal pdf/pmf's are the same as their conditional pdf/pmf's. That is $f_{Y|X=x}(y) = f_Y(y)$ and $f_{X|Y=y}(x) = f_X(x)$, for all y and x where $f_Y(y) > 0$ and $f_X(x) > 0$, respectively.

FUNCTIONS OF RANDOM VARIABLES: PDF OF A SUM, PRODUCT OR QUOTIENT

Let X and Y be **independent** random variables with pdf or pmf's f_X and f_Y or p_X and p_Y, respectively. Then,

If X and Y are discrete random variables, then the pmf of their **sum** W = X+ Y is

$$p_W(w) = \sum_{allx} p_X(x)p_Y(w-x).$$

If X and Y are continuous random variables, then the pdf of their **sum** W = X+ Y is the *convolution* of the individual densities:

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x)f_Y(w-x)dx.$$

If X and Y are independent continuous random variables, then the pdf of their **quotient** $W = Y/X$ is given by:

$$f_W(w) = \int_{-\infty}^{\infty} \mid x \mid f_X(x)f_Y(wx)dx.$$

For a proof of the above result, and the statement below, see pg 181-182 in the text.

If X and Y are independent continuous random variables, then the pdf of their **product** $W = XY$ is given by:

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{\mid x \mid} f_X(w/y)f_Y(y)dy = \int_{-\infty}^{\infty} \frac{1}{\mid x \mid} f_X(x)f_Y(w/x)dx.$$

MORE ON EXPECATION, VARIANCE, AND COVARIANCE

Expected value of a function. Let (X, Y) be a discrete random vector with pmf $p(x, y)$ or continuous random vector with pdf $f(x, y)$. Let $g(\cdot)$ be a real values function of X and Y. Then, the expected value of the random variable $g(X, Y)$ is given by:

$$E(g(X,Y)) = \sum_{all x} \sum_{all y} g(x,y)p(x,y), \text{ in discrete case, and}$$

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy, \text{ in continuous case,}$$

provided that the sums and the integrals converge absolutely. This generalizes, as expected, to n-dimensional random vectors X.

Note: Some treat the terms *mean* and *expected value* synonymously. I will try to restrict the use of the term *mean* to refer to the arithmetic mean of a series of numbers.

Expected value of a sum of random variables. Let $X_1, X_2, \ldots, X_n$ be any random variables with finite expected values, and let $a_1, a_2, \ldots, a_n$ be a set of real numbers. Then

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n).$$

Expected value of a product of independent random variables. If X and Y are independent random variables with finite expectations, then $E(XY) = E(X)E(Y)$.

Variance of a sum of independent random variables. Let $X_1, X_2, \ldots, X_n$ be any independent random variables with finite second moments (i.e. $E(X_i^2) < \infty$), and let $a_1, a_2, \ldots, a_n$ be a set of real numbers. Then

$$Var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2Var(X_1) + a_2^2Var(X_2) + \cdots + a_n^2Var(X_n).$$

Covariance. The covariance of random variables X and Y (each with finite variances) is written $Cov(X, Y)$ or $\sigma(X, Y)$ and defined as

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

For random vectors $X = (X_1, \ldots, X_n)^T$ and $Y = (Y_1, \ldots, Y_m)^T$ the *covariance matrix* is

$$Cov(X,Y) = E((X - E(X))(Y - E(Y)^T)) = E(XY^T) - E(X)E(Y)^T.$$

The i, jth element of the covariance matrix is $Cov(X_i, Y_j)$ (sometimes written $\sigma_{X_iY_j}$ or simply σ_{ij}). The ith diagonal term of $Cov(X, X)$ is the variance of X_i ($Var(X_i)$).

Additional properties of covariance: For random vectors X, Y, and Z and constant c,

1. $\sigma(X, c) = \sigma(c, X) = \mathbf{0}$
2. $\sigma(X, Y) = \sigma(Y, X)^T$
3. $\sigma(X + Y, Z) = \sigma(X, Z) + \sigma(Y, Z)$ and $\sigma(X, Y + Z) = \sigma(X, Y) + \sigma(X, Z)$
4. $\sigma(cX, Y) = c\,\sigma(X, Y) = \sigma(X, cY)$

Correlation: If $Cov(X, Y) = 0$ we say X and Y are *uncorrelated* (if $Cov(X, Y) \neq 0$, X and Y are *correlated*).

Note: Independent random variables are always uncorrelated, however the reverse isn't generally true! The one general case where *uncorrelated* implies *independent* is where the joint distribution of X and Y is multivariate-Normal (aka Gaussian).

Note: Correlation is often described using *correlation coefficients* (e.g., see pg 576).

MOMENT GENERATING FUNCTIONS

Definition. The moment generating function $M_X(t)$ of random variable X is given by

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_{\text{all k}} e^{tk} p_X(k) & \text{if X discrete,} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if X continuous,} \end{cases}$$

at all real values t for which the expectation exists.

Use for mgf's. Moment generating functions are very useful in probability and mathematical statistics. They are primarily used for two purposes:

1. Finding moments of random variables, and
2. Identifying distributions of random variables.

Theorem. Let X be a continuous random variable with pdf f(x). Assume that the pdf f(x) is sufficiently smooth for the order of differentiation and integration to be exchanged. Let $M_X(t)$ be the mgf of X. Then

$$M_X^{(r)}(0) = E(X^r),$$

provided that the rth moment of X exists.

Theorem: Identifying distributions. Suppose that X_1 and X_2 are random variables with pdf/pmf's $f_{X_1}(x)$ and $f_{X_2}(x)$, respectively. Suppose that $M_{X_1}(t) = M_{X_2}(t)$ for all t in a neighborhood of 0. Then X_1 and X_2 are equal in distribution, that is $f_{X_1}(x) = f_{X_2}(x)$.

Theorem: Properties of mgf's. This theorem describes the mgf of a linear function and of a sum of independent random variables.

1. **Mgf of a linear function.** Suppose that X is a random variable with mgf $M_X(t)$, and let $Y = aX + b$, where a and b are real numbers. Then, the mgf of Y is given by:

$$M_Y(t) = e^{bt} M_X(at).$$

2. **Mgf of a sum of independent rv's**. Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with mgf's $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$, respectively. Then, the mgf of their sum is the product of the mgf's, that is

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdots M_{X_n}(t).$$