The Joint Distribution Function

Given a random pair, \((X,Y)\), the joint distribution function is defined as the probability of the event \(\alpha=\{ (X,Y) \mid X \leq x \text{ and } Y \leq y \}\), for \((x, y)\) in \(\mathbb{R}^2\). That is,

\[ F(x, y) = P(\alpha) \]

This is commonly written as:

\[ F(x, y) = P(X \leq x, Y \leq y) \]

Properties of the Joint Distribution Function

a. \(0 \leq F(x, y) \leq 1\)

This is because \(0 \leq P(\alpha) \leq 1\) for any \(\alpha\) in \(\mathbb{R}=\mathbb{R}\), including \(\alpha=\{ X \leq x \text{ and } Y \leq y \}\).

b. \(F(-\infty, y)=F(x, -\infty)=F(-\infty, -\infty)=0.\)

This is true since the set \(\alpha=\{ X \leq -\infty \text{ and } Y \leq y \}\) is empty.

c. \(F(\infty, \infty)=1.\)

This is true since the set \(\alpha=\{ X > \infty \text{ and } Y > \infty \}\) is empty.
d. \( P(a<X \leq b \text{ and } c<Y \leq d) = F(b,d) - F(a,d) - F(b,c) + F(a,c) \).

Let \( \alpha = \{ a<X \leq b, c<Y \leq d \} \), \( \beta = \{ X \leq b, Y \leq d \} \), \( \delta = \{ X \leq a, Y \leq d \} \), and \( \gamma = \{ X \leq b, Y \leq c \} \). Note that \( \beta \) may be partitioned into the disjoint subsets \( \alpha \) and \( \delta + \gamma \). Consequently,

\[ P(\beta) = P(\alpha) + P(\delta + \gamma). \]

Since \( \delta \) and \( \gamma \) are not disjoint, \( P(\delta + \gamma) = P(\delta) + P(\gamma) - P(\delta \gamma) \). Thus,

\[ P(\beta) = P(\alpha) + P(\delta) + P(\gamma) - P(\delta \gamma), \]

or in terms of the joint distribution function, \( P(\alpha) = F(b,d) - F(a,d) - F(b,c) + F(a,c) \).

e. \( F(\mathcal{X,Y}) \) in non-decreasing in both \( \mathcal{X} \) and \( \mathcal{Y} \).

Using the notation above, \( F(b,d) - F(a,d) = P(\beta - \delta) \geq 0 \) and \( F(b,d) - F(b,c) = P(\beta - \gamma) \geq 0 \).

f. \( P(X=b,Y=d) = F(b,d) - F(b,d-) - F(b-,d+) + F(b-,d-) = \Delta^2 F. \)

This is true since the limit of \( \{ a<X \leq b, c<Y \leq d \} \) as \( a \to b- \) and \( c \to d- \) contains the single point \( (b,d) \). Thus, points of positive probability correspond to points of discontinuity (in both planes) of the distribution function. Note that if the joint distribution function is continuous at \( (b,d) \) in either plane, then \( P(X=b,Y=d) = 0 \).

g. \( F(a+,c+)-F(a+,c)-F(a,c+)+F(a,c)=0. \)

This is true since the limit of \( \{ a<X \leq b, c<Y \leq d \} \) as \( b \to a+ \) and \( d \to c+ \) is empty. Thus, the distribution function is right continuous in all arguments.

h. \( F(\mathcal{Y}) = F(\mathcal{X},\mathcal{Y}). \)

\( F(\mathcal{X},\mathcal{Y}) = P(\alpha) \) where \( \alpha = \lim \{ (X,Y) | X \leq \mathcal{X} \text{ and } Y \leq \mathcal{Y} \} \) as \( \mathcal{X} \to \infty \). Likewise, \( F(\mathcal{Y}) = P(\beta) \), where \( \beta = \{(X,Y) | Y \leq \mathcal{Y} \} \). The result follows immediately, since any element in \( \alpha \) is also an element of \( \beta \), and vice versa. By similar reasoning, \( F(\mathcal{X},\infty)=F(\mathcal{X}) \). The resulting distribution functions are called the marginal distribution functions.
Density Function in the Discrete Case

The form of the density function is determined by the form of the distribution function. In the discrete case, there are, at most, a countable number of (X,Y) pairs for which $\Delta^2 F \neq 0$. These correspond to points of positive probability. Where $\partial^2 F(x,y)/\partial x \partial y$ or $\partial^2 F(x,y)/\partial y \partial y$ exist, they equal 0. Under these conditions, $F(x,y)$ is a step function in each plane. In this case, the joint density function, $f(x,y)$, is defined as $\Delta^2 F$. (Again, it is conventional to specify a discrete density only at points of positive probability. Any point not explicitly identified as having positive probability is presumed to have zero probability.) Since the joint density function gives the probability of individual points, the probability of an interval is just the sum of the probabilities of the points in that interval. Thus, $F(x,y) = \sum_{\lambda \leq x} \sum_{\eta \leq y} f(\lambda,\eta)$, where $\lambda$ and $\eta$ are summation indices for X and Y respectively.

Density Function in the Continuous Case

$F(x,y)$ is continuous in each plane and at all points. The cross partial $\partial^2 F(x,y)/\partial x \partial y$ exists at all but a countable number of points, and where it does not exist, $\Delta^2 F=0$. Consequently, there are no points of positive probability. Probability is restricted to regions rather than points. In this case, the joint density function, $f(x,y)$, is defined as $\partial^2 F(x,y)/\partial x \partial y$. By the second fundamental theorem of integral calculus, we know that $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\lambda,\eta)d\eta d\lambda$, where $\lambda$ and $\eta$ are variables of integration for X and Y, and where the lower limits are determined by the conditions $F(x,-\infty)=0$ and $F(-\infty,y)=0$. As in the univariate case, the density function does not directly measure probability. It is bivariate integrals of the density function that give probability.
Marginal Density Functions

We have seen that \( F(\mathcal{X}) = F(\mathcal{X}, \infty) \). This result can be used to determine the relationship between the marginal density function of \( X \) and the joint density function of \( (X, Y) \). In the continuous case, the equality above may be rewritten as

\[
\int_{-\infty}^{\infty} f(\lambda) d\lambda = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\lambda, y) dy d\lambda
\]

Differentiating both sides with respect to \( \mathcal{X} \) gives

\[
f(\mathcal{X}) = \int_{-\infty}^{\infty} f(\mathcal{X}, y) dy
\]

Likewise, the marginal density of \( Y \) is

\[
f(Y) = \int_{-\infty}^{\infty} f(\mathcal{X}, y) d\mathcal{X}
\]

In the discrete case, the equality linking the distribution functions may be rewritten as

\[
\sum_{\lambda \in \mathcal{X}} f(\lambda) = \sum_{\lambda \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f(\lambda, y)
\]

Since \( Y \) is being summed over its entire range, \( \sum_{y \in \mathcal{Y}} f(\lambda, y) \) is a function of \( \mathcal{X} \) alone. Thus,

\[
f(\mathcal{X}) = \sum_{y \in \mathcal{Y}} f(\mathcal{X}, y)
\]

since the equality above must hold for all \( \mathcal{X} \).

Conditional Densities and Distributions

Given a bivariate pair \( (X, Y) \) with joint density \( f(\mathcal{X}, \mathcal{Y}) \), we wish to determine the conditional density of \( Y \) given \( X = \mathcal{X} \). We have seen, that for a pair of events \( A \) and \( B \), the conditional probability of \( A \) given \( B \) is \( P(A|B) = P(AB)/P(B) \). In the case of discrete random variables, this rule may be applied directly since the event \( B = \{X = \mathcal{X}\} \) has positive probability. Letting \( A = \{Y = \mathcal{Y}\} \), we have

\[
f_{Y|X=\mathcal{X}}(\mathcal{Y}) = f(\mathcal{Y} | X = \mathcal{X}) = f(\mathcal{X}, \mathcal{Y})/f(\mathcal{X}) = f(\mathcal{X}, \mathcal{Y})/\left[ \sum_{\eta \in \mathcal{Y}} f(\mathcal{X}, \eta) \right]
\]
The corresponding conditional distribution function is

\[ F_{Y|X=x}(y) = \sum_{\eta \leq y} f_{Y|X=x}(\eta) \]

\[ = \sum_{\eta \leq y} f(x, \eta)/f(x) \]

In the continuous case, a direct application of the rule for conditional probability is inappropriate since the event \( B = \{ X = x \} \) has zero probability! Nevertheless, the marginal density is generally non-zero. This fact will allow us to construct a conditional density function in a manner analogous to the discrete case, that satisfies the usual requirements: a) it integrates to one on the restricted sample space, and b) the relative probability of events in the restricted space are unchanged. Specifically, the conditional density function will be defined as

\[ f_{Y|X=x}(y) = f(Y=y \mid X=x) = f(x, y)/f(x) = f(x, y)/[\int \infty \mathbb{I} = f(x, \eta) d\eta] \]

Clearly,

\[ \int \infty f_{Y|X=x}(y) dy = \int \infty f(x, y)/f(x) dy = \int \infty f(x, y) dy / f(x) = f(x)/f(x) = 1 \]

and

\[ f_{Y|X=x}(a)/f_{Y|X=x}(b) = [f(x, a)/f(x)]/[f(x, b)/f(x)] = f(x, a)/f(x, b) \]

Since the ratio of the densities is the same at all points in the restricted sample space, any probability integrals will be the same as well.

The conditional distribution function defined as

\[ F_{Y|X=x}(y) = \int \infty f_{Y|X=x}(\eta) d\eta = \int \infty [f(x, \eta)/f(x)] d\eta \]

Finally, note that a conditional expectation is no different than the usual univariate expectation, apart from the use of the conditional density function. That is, for an arbitrary function \( \psi(y) \),

\[ E_{Y|X=x} [\psi(Y)] = E[\psi(Y) \mid X = x] = \int \infty \psi(y) f_{Y|X=x}(y) dy \]
Joint Mathematical Expectation

Given a random pair, \((X,Y)\), and an arbitrary function, \(\psi(X,Y)\), the mathematical expectation of \(\psi(X,Y)\) is defined as

\[
E[\psi(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x,y)f(x,y)\,dy\,dx
\]

in the continuous case, and as

\[
E[\psi(X,Y)] = \sum_x \sum_y \psi(x,y)f(x,y)
\]

in the discrete case. In some circumstances, the required integral or sum may not exist.

Joint Expectations of Special Interest

There are certain choices for \(\psi(X,Y)\) that are of special interest. We will consider these choices for the continuous case. The reasoning is analogous in the discrete case.

a. Let \(\psi(X,Y)=\gamma(X)\). Then \(E_{X,Y}[\psi(X,Y)]=E_X[\gamma(X)]\). That is, for functions of a single variable, joint expectations reduce to marginal expectations.

a'. Let \(\psi(X,Y)=X\). Then \(E_{X,Y}[X]=\mu_X\), where the subscript refers to the variable \(X\) as opposed to the order of the moment.

a''. Let \(\psi(X,Y) = (X-\mu_X)^2\). Then \(E_{X,Y}[(X-\mu_X)^2] = \sigma_{XX}^2 = \sigma_X^2\).

b. Let \(\psi(X,Y) = (X-\mu_X)(Y-\mu_Y)\). Then \(E[(X-\mu_X)(Y-\mu_Y)]\) defines the "covariance" of \(X\) and \(Y\). The covariance is generally denoted \(\sigma_{XY}\), \(\sigma(X,Y)\), or \(\text{Cov}(X,Y)\). Note that since \(f(x,y)\) is non-negative, \(\sigma_{XY}>0\) implies that, on average, \((X-\mu_X)\) and \((Y-\mu_Y)\) take the same
sign. That is, X tends to be above its mean when Y is above its mean. Likewise, \( \sigma_{XY} < 0 \) implies that, on average, \( (X-\mu_X) \) and \( (Y-\mu_Y) \) take the opposite sign. That is, X tends to be below its mean when Y is above its mean. The covariance may be expressed in terms of the moments of X and Y as \( \sigma_{XY} = E(XY) - E(X)E(Y) \).

c. Let \( \psi(X,Y) = \alpha X + \beta Y \), where \( \alpha \) and \( \beta \) are constants. Then \( E[\psi(X,Y)] = \alpha \mu_X + \beta \mu_Y \). “The mean of a linear transformation is the linear transformation of the means.”

d. Let \( \psi(X,Y) = [(\alpha X + \beta Y) - (\alpha \mu_X + \beta \mu_Y)]^2 \). Then, \( E[\psi(X,Y)] = \alpha^2 \sigma_{XX} + \beta^2 \sigma_{YY} + 2 \alpha \beta \sigma_{XY} \). The variance of a linear combination is a quadratic form in the covariance matrix of the random variables.

The covariance matrix of the \( (2 \times 1) \) vector X is defined as \( E[(X-\mu)(X-\mu)'] \), where \( \mu \) is the \( (2 \times 1) \) vector of means for the elements of X. Thus, the covariance matrix of X, which is generally denoted \( \Sigma \) or \( \Sigma_X \), is

\[
\Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} \\
\sigma_{12} & \sigma_{22}
\end{bmatrix}
\]

e. Let \( \psi(X,Y) = \exp(tX+sY) \). Then \( E[\exp(tX+sY)] \), when it exists for \( (t,s) \) in an open region about the origin, defines the joint moment generating function of \( (X,Y) \).

\[
E[\exp(tX+sY)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(tX+sY) f(X,Y) dY dX
\]

The joint MGF is generally denoted \( M(t,s) \). If the joint MGF does not exist, the joint characteristic function may be employed. The characteristic function, \( E[\exp(itX+isY)] \), will exist for any random vector.
Properties of the joint Moment Generating Function

a. If the joint MGF exists for \( t \) in an open region about the origin, then moments and cross moments of any orders exist and can be computed from the joint MGF by differentiation.

\[
\frac{\partial^{K+L}}{\partial t^K \partial s^L} \bigg|_{t=s=0} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^K y^L \exp(t x + s y) f(x, y) \, dy \, dx \bigg|_{t=s=0} \\
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^K y^L f(x, y) \, dy \, dx = E(X^K Y^L)
\]

b. The marginal MGFs can be obtained from the joint MGF.

\[
M(t, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(t x) f(x, y) \, dy \, dx \\
= M(t)
\]

c. Let \( Z = \alpha X + \beta Y \). The MGF of a linear combination may be expressed in terms of the joint MGF of \((X, Y)\) as follows.

\[
M_Z(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[r(\alpha x + \beta y)] f(x, y) \, dy \, dx \\
= M_{X,Y}(r\alpha, r\beta)
\]

d. There is a unique relationship between the distribution function of a random variable and its corresponding characteristic function. Knowledge of the distribution function is equivalent to knowledge of the characteristic function, and vice versa.

Statistical Independence

We have seen, that a pair of events \( A \) and \( B \) are statistically independent if and only if \( P(AB) = P(A)P(B) \). Letting \( A = \{X \leq x, Y \leq y\} \) and \( B = \{X \leq x, Y \leq y\} \), this requirement implies

\[
F(x, y) = F(x)F(y)
\]

In the continuous case, this condition may be restated in terms of the density function by differentiating the above.

\[
f(x, y) = f(x)f(y)
\]
The validity of this condition in the discrete case follows directly if we redefine the events as 
\[ A = \{ X = \mathcal{X}, Y \leq \infty \} \text{ and } B = \{ X < \infty, Y = \mathcal{Y} \}. \]

**Implications of Statistical Independence**

The following results are valid under the assumption that X and Y are statistically independent (SI). It must be strongly emphasized, that without statistical independence, these results are generally invalid.

a. Let \( \psi(X, Y) = \gamma(X)\eta(Y) \), where \( \gamma(X) \) and \( \eta(Y) \) are arbitrary functions. Then, under statistical independence, \( E[\psi(X, Y)] = E[\gamma(X)]E[\eta(Y)] \).

b. If X and Y are SI, then W=\( \gamma(X) \) and Z=\( \eta(Y) \) are SI.

Let \( \alpha \) denote a subset of \( \Sigma_W \), \( a = \{ X \mid \gamma(X) \in \alpha \} \), \( \beta \) denote a subset of \( \Sigma_Z \), \( b = \{ Y \mid \eta(Y) \in \beta \} \). Then,

\[
P(W \text{ in } \alpha \text{ and } Z \text{ in } \beta) = P(X \text{ in } a \text{ and } Y \text{ in } b) \]
\[
= P(X \text{ in } a)P(Y \text{ in } b) \]
\[
= P(W \text{ in } \alpha)P(Z \text{ in } \beta) \]

c. We have seen that \( \sigma_{XY} = E(XY) - E(X)E(Y) \). The previous result shows that, under statistical independence, \( E(XY) = E(X)E(Y) \). Hence, if X and Y are SI, then \( \sigma_{XY} = 0 \). It should be emphasized, however, that \( \sigma_{XY} = 0 \) does not imply statistical independence of X and Y.

d. If X and Y are SI, then \( M_{X,Y}(t,s) = M_X(t)M_Y(s) \). The joint MGF of SI random variables factors into the product of the marginal MGFs.