## Probability Distributions

Let \( P(x, y, \ldots, z) \) represent a joint probability density for the random variables \( X, Y, \ldots, Z \), where values are real numbers, so that \( P(x, y, \ldots, z) \, dx \, dy \, \ldots \, dz \) is interpreted to mean the probability that \( X \) will take a value in the interval \([x, x + dx]\) and \( Y \) will take a value in \([y, y + dy]\) and \( \ldots \) and \( Z \) will take a value in \([z, z + dz]\).

We now define a marginal probability for any subset of the original random variables, say \( X, Y \) to be the integral of the total joint density with respect to the remaining variables, i.e.

\[
P(x, y) = \int P(x, y, z, \ldots, w) \, dz \ldots dw
\]

which represents the density for the joint event \( X \in [x, x + dx] \) and \( Y \in [y, y + dy] \) with no restrictions on the remaining variables. It is the relative density over \( X, Y \) when we have no information about the other variables.

Finally, we define a conditional density for any subset \( X, Y \), conditioned on any remaining variables having prescribed values, say \( Z = z \ldots W = w \), denoted by \( P(x, y | z, \ldots, w) \), to be:

\[
P(x, y | z, \ldots, w) = \frac{P(x, y, z, \ldots, w)}{P(z, \ldots, w)}
\]

which represents the density for \( X, Y \) relative to the knowledge that \( Z, \ldots, W \) have the definite values \( z, \ldots, w \).
we shall say that the variables \( X \) and \( Y \) are independent if and only if \( p(x,y) = p(x)p(y) \) except for a set of measure zero, and more generally we shall say that the groups \( X, Y, U, V, \ldots, W, X \) are mutually independent if random variables are group-wise mutually independent if \( p(x,y,u,v,\ldots,w,x) = p(x)p(y)p(u)p(v)\ldots p(w) \) except for a set of measure zero. Independence then implies that the conditional distribution for a group conditioned on values in another independent group is the same as the marginal for the group, i.e., \( X, Y \) independent \( \Rightarrow p(y|x) = p(y) \) or \( S, T, U, \ldots, W, X \) mutually independent \( \Rightarrow p(w|x) = p(w) \) almost everywhere.

That is, we learn nothing about \( W, \ldots, X \) by being told values of variables in groups independent of \( W, \ldots, X \). 
\[ \text{Include Exp and def Exp (and good exp Exp)} \]

Information

We now define a functional for probability distribution, \( I_{x \ldots y} = I \{ p_{x \ldots y} \} \) to be:

\[
I_{x \ldots y} = \iint \ln p_{x \ldots y} \, dx \ldots dy
\]

\[
= \text{Exp} \{ \ln p_{x \ldots y} \}
\]

i.e., we define the information of a probability distribution to be the Expected logarithm of the probability density.
We can now define Marginal Prob. measures over the product space of sub-classes of \( X \subseteq \ldots \)

\[ P_m(x_i y) = M_p(X_j y) \]

and similarly from this Marginal Measure we derive Marginal Density

\[ p(x, y) = \int p(x, y, z, \ldots) \, dz \ldots \, \hat{m} \]

We further define Conditional measures \( M_p\{x_i y \ldots | z_k \ldots, w_3^2\} \)

to be

\[ = \frac{M_p\{x_i y \ldots | z_k \ldots, w_3^2\}}{M_p\{z_k \ldots, z_3^2\}} \]

which leads to Conditional Density

\[ p^2 \ldots \hat{m} = \frac{P(x_i y \ldots | z \ldots, w)}{P(z \ldots | w)} \]

Finally, we define

\[ I^{3 \ldots \hat{w}}_{x \ldots \hat{y}} = \exp \left\{ \ln p^{3 \ldots \hat{w}}_{x \ldots \hat{y}} \right\} \]

with respect to the conditional measure \( M_p\{x_i y \ldots | z_k \ldots, w_3^2\} \).


9 Probability Distributions:

We shall assume that we have a collection of sets \( X, Y, Z, \ldots \), each with a measure assigned \( M(X), M(Y), \ldots \).

(i.e., a non-negative, countably additive set function over some of the subsets \( X_i, Y_j, \ldots \).)

and furthermore we assign the product measure to the direct product of these sets.

We further assume that we have a probability measure over this direct product, which we shall denote by \( \mu(X, Y, \ldots) \), which we think of as now a function of \( n \) tuples of sets one in \( X, Y, \ldots \), and which we think of as being the probability that a selection of one element from each set will result in the event that \( x \in X, y \in Y, \ldots \).

Now it is known that this always results in the existence of a probability density function

\[ f(x, y, z, \ldots) \]

which is a point (or element) function rather than a set function, such that the integral of \( f(x, y, z, \ldots) \) over any set of the product space with respect to our original product measure is the probability measure of that set.