556: MATHEMATICAL STATISTICS I

BASIC PROPERTIES OF MULTIVARIATE DISTRIBUTIONS

A random vector (or vector random variable) $\mathbf{X} = (X_1, \ldots, X_d)$ is a d-dimensional extension of a random variable. We define

- **Joint cdf:** $F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,\ldots,X_d}(x_1, \ldots, x_d)$ defined by

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = P_{X_1,\ldots,X_d}\left[\bigcap_{j=1}^{d}(X_j \in (-\infty, x_j])\right] = P_{X_1,\ldots,X_d}\left[\bigcap_{j=1}^{d}(X_j \leq x_j)\right].$$

  This function has the following properties:

  (i) Limit behaviour:

$$\lim_{\text{Any } j: x_j \longrightarrow -\infty} F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = 0 \qquad \lim_{\text{All } j: x_j \longrightarrow \infty} F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = 1$$

  (ii) Non-decreasing in each dimension: for all j and any $h > 0$

$$F_{X_1,\ldots,X_j,\ldots,X_d}(x_1, \ldots, x_j, \ldots, x_d) \leq F_{X_1,\ldots,X_j,\ldots,X_d}(x_1, \ldots, x_j + h, \ldots, x_d)$$

  (iii) Right-continuous in each dimension: for all j

$$\lim_{h \longrightarrow 0^+} F_{X_1,\ldots,X_j,\ldots,X_d}(x_1, \ldots, x_j + h, \ldots, x_d) = F_{X_1,\ldots,X_j,\ldots,X_d}(x_1, \ldots, x_j, \ldots, x_d)$$

  (iv) Marginalization: without loss of generality, consider $x_1 \longrightarrow \infty$. We have from the definition of the joint cdf that

$$\lim_{x_1 \longrightarrow \infty} F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = F_{X_2,\ldots,X_d}(x_2, \ldots, x_d)$$

  where the right-hand side is the joint cdf of $(X_2, \ldots, X_d)$. This result holds whichever component we allow to increase to infinity. It also holds if we allow more than one component to increase to infinity.

  The joint distribution of $(X_1, \ldots, X_d)$ thus defines the marginal distribution of any subset of the components of $(X_1, \ldots, X_d)$.

- **Joint pmf:** If all the elements of X are discrete, then we can consider the joint pmf denoted $f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,\ldots,X_d}(x_1, \ldots, x_d)$ defined by

$$f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = P_{X_1,\ldots,X_d}\left[\bigcap_{j=1}^{d}(X_j = x_j)\right].$$

  This function has the following properties:

  (i) Boundedness: $0 \leq f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) \leq 1$.

  (ii) Summability: by the probability axioms, if $\mathbb{X}^{(d)}$ denotes the support of the joint pmf

$$\sum_{\mathbf{x} \in \mathbb{X}^{(d)}} f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = 1.$$

- **Joint pdf:** If we can represent

$$F_{X_1,\ldots,X_d}(x_1,\ldots,x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{X_1,\ldots,X_d}(t_1,\ldots,t_d)\, dt_1 \ldots dt_d$$

then X is absolutely continuous with joint pdf $f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,\ldots,X_d}(x_1,\ldots,x_d)$. This function has the following properties:

  (i) Non-negativity: $f_{X_1,\ldots,X_d}(x_1,\ldots,x_d) \geq 0$.

  (ii) Integrability:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_d}(x_1,\ldots,x_d)\, dx_1 \ldots dx_d = 1.$$

- **Conditional pmf/pdf:** for any partition of $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, we may define the conditional pmf/pdf for $\mathbf{X}_2$, given that $\mathbf{X}_1 = \mathbf{x}_1$ as

$$f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}_1}(\mathbf{x}_1)}$$

provided $f_{\mathbf{X}_1}(\mathbf{x}_1) > 0$. This allows us to deduce the chain rule factorization

$$f_{X_1,\ldots,X_d}(x_1,\ldots,x_d) = f_{X_1}(x_1) \prod_{j=2}^{d} f_{X_j|X_1,\ldots,X_{j-1}}(x_j|x_1,\ldots,x_{j-1})$$

provided all the conditional distributions are well-defined. In the factorization, the labelling of the variables is arbitrary.

- **Independence:** $X_1,\ldots,X_d$ are independent if, **for all** $(x_1,\ldots,x_d)$

$$F_{X_1,\ldots,X_d}(x_1,\ldots,x_d) = \prod_{j=1}^{d} F_{X_j}(x_j)$$

or equivalently

$$f_{X_1,\ldots,X_d}(x_1,\ldots,x_d) = \prod_{j=1}^{d} f_{X_j}(x_j).$$

This definition is equivalent to saying that

$$f_{X_1|X_2,\ldots,X_d}(x_1|x_2\ldots,x_d) = f_{X_1}(x_1)$$

for all possible selections of $x_1,\ldots,x_d$; note that the labelling of the variables is arbitrary, so this definition applies equivalently for any permutation of the labels.

- **Region probabilities:** Let $A \subseteq \mathbb{R}^d$. To compute

$$P_{X_1,\ldots,X_d}[(X_1,\ldots,X_d) \in A]$$

we may write

$$P_{X_1,\ldots,X_d}[(X_1,\ldots,X_d) \in A] = \int \cdots \int_{A} dF_{X_1,\ldots,X_d}(x_1,\ldots,x_d)$$

- **1-1 Transformations:** For continuous variables $(X_1, \ldots, X_d)$ with support $\mathbb{X}^{(d)}$ and joint pdf $f_{X_1,\ldots,X_d}$ we can construct the pdf of a transformed set of variables $(Y_1, \ldots, Y_d)$ using the following steps:

  1. Write down the set of transformation functions $g_1, \ldots, g_d$
  $$\begin{array}{c} Y_1 = g_1 (X_1, \ldots, X_d) \\ \vdots \\ Y_d = g_d (X_1, \ldots, X_d) \end{array} .$$
  2. Write down the set of inverse transformation functions $g_1^{-1}, \ldots, g_d^{-1}$
  $$\begin{array}{c} X_1 = g_1^{-1} (Y_1, \ldots, Y_d) \\ \vdots \\ X_d = g_d^{-1} (Y_1, \ldots, Y_d) \end{array}$$
  3. Consider the joint support of the new variables, $\mathbb{Y}^{(k)}$.
  4. Compute the Jacobian of the transformation: first form the matrix of partial derivatives
  $$D_y = \begin{bmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} & \cdots & \dfrac{\partial x_1}{\partial y_d} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} & \cdots & \dfrac{\partial x_2}{\partial y_d} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial x_d}{\partial y_1} & \dfrac{\partial x_d}{\partial y_2} & \cdots & \dfrac{\partial x_d}{\partial y_d} \end{bmatrix}$$
  where, for each (i, j)
  $$\frac{\partial x_i}{\partial y_j} = \frac{\partial}{\partial y_j} \left\{ g_i^{-1} (y_1, \ldots, y_d) \right\}$$
  and then set $|J (y_1, \ldots, y_d)| = |\det D_y|$

  Note that
  $$\det D_y = \det D_y^\top$$
  so that an alternative but equivalent Jacobian calculation can be carried out by forming $D_y^\top$. *Note also that*
  $$|J (y_1, \ldots, y_d)| = \frac{1}{|J (x_1, \ldots, x_d)|}$$
  where $J (x_1, \ldots, x_d)$ *is the Jacobian of the transformation regarded in the reverse direction (that is, if we start with* $(Y_1, \ldots, Y_d)$ *and transfrom to* $(X_1, \ldots, X_d)$*)*
  5. Write down the joint pdf of $(Y_1, \ldots, Y_d)$ as
  $$f_{Y_1,\ldots,Y_d} (y_1, \ldots, y_d) = f_{X_1,\ldots,X_d} \left(g_1^{-1} (y_1, \ldots, y_d), \ldots, g_d^{-1} (y_1, \ldots, y_d) \right) \times |J (y_1, \ldots, y_d)|$$
  for $(y_1, \ldots, y_d) \in \mathbb{Y}^{(k)}$

- **Expectations:** If $g(.)$ is some k-dimensional function, then
$$\mathbb{E}_{\mathbf{X}}[g(\mathbf{X})] = \mathbb{E}_{X_1,\ldots,X_d}[g(X_1, \ldots, X_d)] = \int \cdots \int g(x_1, \ldots, x_d)\, dF_{X_1,\ldots,X_d}(x_1, \ldots, x_d)$$