STAT/MTHE 353: Multiple Random Variables

Queen's University

Review

- S is the *sample space*;
- P is a *probability measure* on S: P is a function from a collection of subsets of S (called the events) to $[0, 1]$. P satisfies the *axioms of probability*;
- A *random variable* is a function $X : S \to \mathbb{R}$. The *distribution* of X is the probability measure associated with X:

$$P(X \in A) = P(\{s : X(s) \in A\}), \quad \text{for any "reasonable" } A \subset \mathbb{R}.$$

Here are the usual ways to describe the distribution of X:

- *Distribution function:* $F : \mathbb{R} \to [0,1]$ defined by

$$F_X(x) = P(X \le x).$$

  It is always well defined.

- *Probability mass function*, or pmf: If X is a *discrete random variable*, then its pmf $p_X : \mathbb{R} \to [0,1]$ is

$$p_X(x) = P(X = x), \quad \text{for all } x \in \mathbb{R}.$$

  Note: : since X is discrete, there is a countable set $\mathcal{X} \subset \mathbb{R}$ such that $p_X(x) = 0$ if $x \notin \mathcal{X}$.

- *Probability density function*, or pdf: If X is a *continuous random variable*, then its pdf $f_X : \mathbb{R} \to [0,\infty)$ is a function such that

$$P(X \in A) = \int_A f(x)\,dx \quad \text{for all } \textit{reasonable } A \subset \mathbb{R}.$$

Joint Distributions

- If $X_1, \ldots, X_n$ are random variables (defined on the same probability space), we can think of

$$\boldsymbol{X} = (X_1, \ldots, X_n)^T$$

  as a random vector. (In this course $(x_1, \ldots, x_n)$ is a row vector and its transpose, $(x_1, \ldots, x_n)^T$, is a column vector.)

- Thus $\boldsymbol{X}$ is a function $\boldsymbol{X} : S \to \mathbb{R}^n$.
- Distribution of $\boldsymbol{X}$: For "reasonable" $A \subset \mathbb{R}^n$, we define

$$P(\boldsymbol{X} \in A) = P(\{s : \boldsymbol{X}(s) \in A\}).$$

- $\boldsymbol{X}$ is called a *random vector* or *vector random variable*.

We usually describe the distribution of $\boldsymbol{X}$ by a function on $\mathbb{R}^n$:

- *Joint cumulative distribution function* (jcdf) is a the function defined for $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ by

$$\begin{aligned} F_{\boldsymbol{X}}(\boldsymbol{x}) &= F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \\ &= P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) \\ &= P(\{X_1 \le x_1\} \cap \{X_2 \le x_2\} \cap \cdots \cap \{X_n \le x_n\}) \\ &= P\Big(\boldsymbol{X} \in \prod_{i=1}^n (-\infty, x_i]\Big) \end{aligned}$$

- If $X_1, \ldots, X_n$ are all discrete random variables, then their *joint probability mass function* (jpmf) is

$$\begin{aligned} p_{\boldsymbol{X}}(\boldsymbol{x}) &= P(\boldsymbol{X} = \boldsymbol{x}) \\ &= P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n), \quad \boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$

  The *finite* or *countable* set of $\boldsymbol{x}$ values such that $p_{\boldsymbol{X}}(\boldsymbol{x}) > 0$ is called the *support* of the distribution of $\boldsymbol{X}$.

Properties of joint pmf:

(1) $0 \le p_{\boldsymbol{X}}(\boldsymbol{x}) \le 1$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

(2) $\displaystyle\sum_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} p_{\boldsymbol{X}}(\boldsymbol{x}) = 1$, where $\boldsymbol{\mathcal{X}}$ is the support of $\boldsymbol{X}$.

If $X_1, \ldots, X_n$ are continuous random variables and there exists $f_{\boldsymbol{X}} : \mathbb{R}^n \to [0,\infty)$ such that for any "reasonable" $A \subset \mathbb{R}^n$,

$$P(X \in A) = \int \cdots \int_A f_{\boldsymbol{X}}(x_1, \ldots, x_n)\,dx_1 \cdots dx_n$$

then

- The $X_1, \ldots, X_n$ are called *jointly continuous*;
- $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ is called the *joint probability density function* (jpdf) of $\boldsymbol{X}$.

Comments:

(a) The joint pdf can be redefined on any set in $\mathbb{R}^n$ that has zero volume. This will not change the distribution of $\boldsymbol{X}$.

(b) The joint pdf may not exists even when each $X_1, \ldots, X_n$ are all (individually) continuous random variables.

Example: ...

Properties of joint pdf:

(1) $f_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$

(2) $\displaystyle\int_{\mathbb{R}^n} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} = \underset{\mathbb{R}^n}{\int \cdots \int} f_{\boldsymbol{X}}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$

- The distributions for the various subsets of $\{X_1, \ldots, X_n\}$ can be recovered from the joint distribution.
- These distributions are called the *joint marginal distributions* (here "marginal" is relative to the full set $\{X_1, \ldots, X_n\}$.

Marginal joint probability mass functions

Assume $X_1, \ldots, X_n$ are discrete. Let $0 < k < n$ and

$$\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$$

Then the marginal joint pmf of $(X_{i_1}, \ldots, X_{i_k})$ can be obtained from $p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ as

$$\begin{aligned} & p_{X_{i_1},\ldots,X_{i_k}}(x_{i_1}, \ldots, x_{i_k}) \\ &= P(X_{i_1} = x_{i_1}, \ldots, X_{i_k} = x_{i_k}) \\ &= P(X_{i_1} = x_{i_1}, \ldots, X_{i_k} = x_{i_k}, X_{j_1} \in \mathbb{R}, \ldots, X_{j_{n-k}} \in \mathbb{R}) \\ &\quad \text{where } \{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\} \\ &= \sum_{x_{j_1}} \cdots \sum_{x_{j_{n-k}}} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \end{aligned}$$

Thus the joint pmf of $X_{i_1}, \ldots, X_{i_k}$ is obtained by summing $p_{X_1,\ldots,X_n}$ over all possible values of the complementary variables $x_{j_1}, \ldots, x_{j_{n-k}}$.

Example: In an urn there are n_i objects of type i for $i = 1, \ldots, r$. The total number of objects is $n_1 + \cdots + n_r = N$. We randomly draw n objects ($n \leq N$) without replacement. Let $X_i = \#$ of objects of type i drawn. Find the joint pmf of $(X_1, \ldots, X_r)$. Also find the marginal distribution of each X_i, $i = 1, \ldots, r$.

Solution: ...

Marginal joint probability density functions

Let $X_1, \ldots, X_n$ be jointly continuous with pdf $f_{\mathbf{X}} = f$. As before, let

$$\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}, \quad \{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$$

Let $B \subset \mathbb{R}^k$. Then

$$\begin{aligned} &P\big((X_{i_1}, \ldots, X_{i_k}) \in B\big) \\ &\quad = P\Big((X_{i_1}, \ldots, X_{i_k}) \in B,\, X_{j_1} \in \mathbb{R}, \ldots, X_{j_{n-k}} \in \mathbb{R}\Big) \\ &\quad = \int_B \cdots \int \left(\int_{\mathbb{R}^{n-k}} \cdots \int f(x_1, \ldots, x_n)\, dx_{j_1} \cdots dx_{j_{n-k}} \right) dx_{i_1} \cdots dx_{i_k} \end{aligned}$$

That is, we "integrate out" the variables complementary to $x_{i_1}, \ldots, x_{i_k}$.

In conclusion, for $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$,

$$\boxed{f_{X_{i_1}, \ldots, X_{i_k}}(x_{i_1}, \ldots, x_{i_k}) = \int_{\mathbb{R}^{n-k}} \cdots \int f(x_1, \ldots, x_n)\, dx_{j_1} \cdots dx_{j_{n-k}}}$$

where $\{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$,

Note: In both the discrete and continuous cases it is important to always know where the joint pmf p and joint pdf f are zero and where they are positive. The latter set is called the *support* of p or f.

Example: Suppose X_1, X_2, X_3 are jointly continuous with jpdf

$$f(x_1, x_2, x_3) = \begin{cases} 1 & \text{if } 0 \le x_i \le 1,\ i = 1, 2, 3 \\ 0 & \text{otherwise} \end{cases}$$

Find the marginal pdfs of X_i, $i = 1, 2, 3$, and the marginal jpdfs of (X_i, X_j), $i \neq j$.

Solution: ...

Example: With X_1, X_2, X_3 as in the previous problem, consider the quadratic equation

$$X_1 y^2 + X_2 y + X_3 = 0$$

in the variable y. Find the probability that both roots are real.

Solution: ...

Marginal joint cumulative distribution functions

In all cases (discrete, continuous, or mixed),

$$\begin{aligned} &F_{X_{i_1}, \ldots, X_{i_k}}(x_{i_1}, \ldots, x_{i_k}) \\ &\quad = P(X_{i_1} \le x_{i_1}, \ldots, X_{i_k} \le x_{i_k}) \\ &\quad = P(X_{i_1} \le x_{i_1}, \ldots, X_{i_k} \le x_{i_k}, X_{j_1} < \infty, \ldots, X_{j_{n-k}} < \infty) \\ &\quad = \lim_{x_{j_1} \to \infty} \cdots \lim_{x_{j_{n-k}} \to \infty} F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) \end{aligned}$$

That is, we let the variables complementary to $x_{i_1}, \ldots, x_{i_k}$ converge to ∞

Independence

Definition The random variables $X_1, \ldots, X_n$ are independent if for all "reasonable" $A_1, \ldots, A_n \subset \mathbb{R}$,

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \times \cdots \times P(X_n \in A_n)$$

Remarks:

(i) Independence among $X_1, \ldots, X_n$ usually arises in a probability model *by assumption*. Such an assumption is reasonable if the outcome of X_i "has no effect" on the outcomes of the other X_j's.

(ii) The also definition applies to any n random quantities $X_1, \ldots, X_n$. E.g., each X_i can itself be a vector r.v. In this case the A_i's have to be appropriately modified.

(iii) Suppose $g_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$ are "reasonable" functions. Then if $X_1, \ldots, X_n$ are independent, then so are $g_1(X_1), \ldots, g_n(X_n)$.

Proof For $A_1, \ldots, A_n \subset \mathbb{R}$,

$$\begin{aligned} & P\big(g_1(X_1) \in A_1, \ldots, g_n(X_n) \in A_n\big) \\ &= P\big(X_1 \in g_1^{-1}(A_1), \ldots, X_n \in g^{-1}(A_n)\big) \\ &\quad \text{where } g_i^{-1}(A_i) = \{x_i : g_i(x_i) \in A_i\} \\ &= P\big(X_1 \in g_1^{-1}(A_1)\big) \times \cdots \times P\big(X_n \in g^{-1}(A_n)\big) \\ &= P\big(g_1(X_1) \in A_1\big) \times \cdots \times P\big(g_n(X_n) \in A_n\big) \end{aligned}$$

Since the sets A_i were arbitrary, we obtain that $g_1(X_1), \ldots, g_n(X_n)$ are independent.

Note: If we only know that X_i and X_j are independent for all $i \neq j$, it does not follow that $X_1, \ldots, X_n$ are independent.

Independence and cdf, pmf, pdf

Theorem 1

Let F be the joint cdf of the random variables $X_1, \ldots, X_n$. Then $X_1, \ldots, X_n$ are independent if and only if F is the product of the marginal cdfs of the X_i, i.e., for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

Proof: If $X_1, \ldots, X_n$ are independent, then

$$\begin{aligned} F(x_1, \ldots, x_n) &= P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) \\ &= P(X_1 \le x_1) P(X_2 \le x_2) \cdots P(X_n \le x_n) \\ &= F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n) \end{aligned}$$

The converse that $F(x_1, \ldots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$ for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$ implies independence is out the the scope of this class. □

Theorem 2

Let $X_1, \ldots, X_n$ be discrete r.v.'s with joint pmf p. Then $X_1, \ldots, X_n$ are independent if and only if p is the product of the marginal pmfs of the X_i, i.e., for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n)$$

Proof: If $X_1, \ldots, X_n$ are independent, then

$$\begin{aligned} p(x_1, \ldots, x_n) &= P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\ &= P(X_1 = x_1) P(X_2 = x_2) \cdots P(X_n = x_n) \\ &= p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n) \end{aligned}$$

Proof cont'd: Conversely, suppose that $p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$ for any $x_1, \ldots, x_n$. Then, for any $A_1, A_2, \ldots, A_n \subset \mathbb{R}$,

$$\begin{aligned} P(X_1 \in A_1, \ldots, X_n \in A_n) &= \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} p(x_1, \ldots, x_n) \\ &= \sum_{x_1 \in A_1} \cdots \sum_{x_n \in A_n} p_{X_1}(x_1) \cdots p_{X_n}(x_n) \\ &= \left(\sum_{x_1 \in A_1} p_{X_1}(x_1)\right)\left(\sum_{x_2 \in A_2} p_{X_2}(x_2)\right) \cdots \left(\sum_{x_n \in A_n} p_{X_n}(x_n)\right) \\ &= P(X_1 \in A_1) P(X_2 \in A_2) \cdots P(X_n \in A_n) \end{aligned}$$

Thus $X_1, \ldots, X_n$ are independent. □

Theorem 3

Let $X_1, \ldots, X_n$ be jointly continuous r.v.'s with joint pdf f. Then $X_1, \ldots, X_n$ are independent if and only if f is the product of the marginal pdfs of the X_i, i.e., for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

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

Proof: Assume $f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$ for any $x_1, \ldots, x_n$.Then for any $A_1, A_2, \ldots, A_n \subset \mathbb{R}$,

$$\begin{aligned} P(X_1 \in A_1, \ldots, X_n \in A_n) &= \int_{A_1} \cdots \int_{A_n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\ &= \int_{A_1} \cdots \int_{A_n} f_{X_1}(x_1) \cdots f_{X_n}(x_n)\, dx_1 \cdots dx_n \\ &= \left(\int_{A_1} f_{X_1}(x_1)\, dx_1\right) \cdots \left(\int_{A_n} f_{X_n}(x_n)\, dx_n\right) \\ &= P(X_1 \in A_1) P(X_2 \in A_2) \cdots P(X_n \in A_n) \end{aligned}$$

so $X_1, \ldots, X_n$ are independent.

Proof cont'd: For the converse, note that

$$\begin{aligned} F(x_1, \ldots, x_n) &= P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\, dt_1 \cdots dt_n \end{aligned}$$

By the fundamental theorem of calculus

$$\boxed{\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)}$$

(assuming f is "nice enough").

If $X_1, \ldots, X_n$ are independent, then $F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$. Thus

$$\begin{aligned} f(x_1, \ldots, x_n) &= \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \ldots, x_n) \\ &= \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1}(x_1) \cdots F_{X_n}(x_n) \\ &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \end{aligned}$$ □

Example: ...

Expectations Involving Multiple Random Variables

Recall that the expectation of a random variable X is

$$E(X) = \begin{cases} \sum_x xp(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} xf(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

if the sum or the integral exist in the sense that $\sum_x |x|p(x) < \infty$ or $\int_{-\infty}^{\infty} |x|f(x)\,dx < \infty$.

Example: ...

- If $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is a random vector, we sometimes use the notation

$$E(\boldsymbol{X}) = \big(E(X_1), \ldots, E(X_n)\big)^T$$

- For $X_1, \ldots, X_n$ discrete, we still have $E(\boldsymbol{X}) = \sum_{\boldsymbol{x}} \boldsymbol{x}p(\boldsymbol{x})$ with the understanding that

$$\begin{aligned} \sum_{\boldsymbol{x}} \boldsymbol{x}p(\boldsymbol{x}) &= \sum_{(x_1,\ldots,x_n)} (x_1, \ldots, x_n)^T p(x_1, \ldots, x_n) \\ &= \Big(\sum_{x_1} x_1 p_{X_1}(x_1), \sum_{x_2} x_2 p_{X_2}(x_2), \ldots, \sum_{x_n} x_n p_{X_n}(x_n)\Big)^T \\ &= \big(E(X_1), \ldots, E(X_n)\big)^T \end{aligned}$$

- Similarly, for jointly continuous $X_1, \ldots, X_n$,

$$\begin{aligned} E(\boldsymbol{X}) &= \int_{\mathbb{R}^n} \boldsymbol{x} f(\boldsymbol{x})\,d\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} (x_1, \ldots, x_n)^T f(x_1, \ldots, x_n)\,dx_1 \cdots dx_n \end{aligned}$$

Theorem 4 ("Law of the unconscious statistician")

Suppose $\boldsymbol{Y} = g(\boldsymbol{X})$ for some function $g : \mathbb{R}^n \to \mathbb{R}^k$. Then

$$E(\boldsymbol{Y}) = \begin{cases} \sum_{\boldsymbol{x}} g(\boldsymbol{x})p(\boldsymbol{x}) & \text{if } X_1, \ldots, X_n \text{ are discrete} \\ \int_{\mathbb{R}^n} g(\boldsymbol{x})f(\boldsymbol{x})\,d\boldsymbol{x} & \text{if } X_1, \ldots, X_n \text{ are jointly continuous} \end{cases}$$

Proof: We only prove the discrete case. Since $\boldsymbol{X} = (X_1, \ldots, X_n)$ can only take a countable number of values with positive probability, the same is true for

$$(Y_1, \ldots, Y_k)^T = \boldsymbol{Y} = g(\boldsymbol{X})$$

so $Y_1, \ldots, Y_k$ are discrete random variables.

Proof cont'd: Thus

$$\begin{aligned} E(\boldsymbol{Y}) &= \sum_{\boldsymbol{y}} \boldsymbol{y}P(\boldsymbol{Y} = \boldsymbol{y}) = \sum_{\boldsymbol{y}} \boldsymbol{y}P(g(\boldsymbol{X}) = \boldsymbol{y}) \\ &= \sum_{\boldsymbol{y}} \boldsymbol{y} \sum_{\boldsymbol{x}:g(\boldsymbol{x})=\boldsymbol{y}} P(\boldsymbol{X} = \boldsymbol{x}) \\ &= \sum_{\boldsymbol{y}} \sum_{\boldsymbol{x}:g(\boldsymbol{x})=\boldsymbol{y}} g(\boldsymbol{x})P(\boldsymbol{X} = \boldsymbol{x}) \\ &= \sum_{\boldsymbol{x}} g(\boldsymbol{x})P(\boldsymbol{X} = \boldsymbol{x}) \\ &= \sum_{\boldsymbol{x}} g(\boldsymbol{x})p(\boldsymbol{x}) \qquad \square \end{aligned}$$

Example: Linearity of expectation...

$$E(a_0 + a_1X_1 + \cdots a_nX_n) = a_0 + a_1E(X_1) + \cdots + a_nE(X_n)$$

Example: Expected value of a binomial random variable...

Example: Suppose we have n bar magnets, each having negative polarity at one end and positive polarity at the other end. Line up the magnets end-to-end in such a way that the orientation of each magnet is random (the two choices are equally likely independently of the others). On the average, how many segments of magnets that stick together do we obtain?

Solution: ...

Transformation of Multiple Random Variables

- Suppose $X_1, \ldots, X_n$ are jointly continuous with joint pdf $f(x_1, \ldots, x_n)$.
- Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable and one-to-one (invertible) function whose inverse g is also continuously differentiable. Thus h is given by $(x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_n)$, where

$$y_1 = h_1(x_1, \ldots, x_n), \ \ y_2 = h_2(x_1, \ldots, x_n), \ \ \ldots, \ \ y_n = h_n(x_1, \ldots, x_n)$$

- We want to find the joint pdf of the vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, where

$$\begin{aligned} Y_1 &= h_1(X_1, \ldots, X_n) \\ Y_2 &= h_2(X_1, \ldots, X_n) \\ &\vdots \\ Y_n &= h_n(X_1, \ldots, X_n) \end{aligned}$$

Let $g : \mathbb{R}^n \to \mathbb{R}^n$ denote the inverse of h. Let $B \subset \mathbb{R}^n$ be a "nice" set. We have

$$P\big((Y_1, \ldots, Y_n) \in B\big) = P\big(h(\boldsymbol{X}) \in B\big) = P\big((X_1, \ldots, X_n) \in A\big)$$

where $A = \{\boldsymbol{x} \in \mathbb{R}^n : h(\boldsymbol{x}) \in B\} = h^{-1}(B) = g(B)$.

The multivariate change of variables formula for $\boldsymbol{x} = g(\boldsymbol{y})$ implies that

$$\begin{aligned} P\big((X_1, \ldots, X_n) \in A\big) &= \int \cdots \int_{g(B)} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\ &= \int \cdots \int_{B} f\big(g_1(y_1, \ldots, y_n), \ldots, g_n(y_1, \ldots, y_n)\big) |J_g(y_1, \ldots, y_n)|\, dy_1 \cdots dy_n \\ &= \int \cdots \int_{B} f(g(\boldsymbol{y})) |J_g(\boldsymbol{y})|\, d\boldsymbol{y} \end{aligned}$$

where J_g is the Jacobian of the transformation g.

We have shown that for "nice" $B \subset \mathbb{R}^n$

$$P\big((Y_1, \ldots, Y_n) \in B\big) = \int_B f(g(\boldsymbol{y})) |J_g(\boldsymbol{y})|\, d\boldsymbol{y}$$

This implies the following:

Theorem 5 (Transformation of Multiple Random Variables)

Suppose $X_1, \ldots, X_n$ are jointly continuous with joint pdf $f(x_1, \ldots, x_n)$. Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable and one-to-one function with continuously differentiable inverse g. Then the joint pdf of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T = h(\boldsymbol{X})$ is

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_n) = f\big(g_1(y_1, \ldots, y_n), \ldots, g_n(y_1, \ldots, y_n)\big) |J_g(y_1, \ldots, y_n)|.$$

Example: Suppose $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ has joint pdf f and let $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$, where $\boldsymbol{A}$ is an invertible $n \times n$ (real) matrix. Find $f_{\boldsymbol{Y}}$.

Solution: ...

Often we are interested in the pdf of just a single function of $X_1, \ldots, X_n$, say $Y_1 = h_1(X_1, \ldots, X_n)$.

(1) Define $Y_i = h_i(X_1, \ldots, X_n)$, $i = 2, \ldots, n$ in such a way that the mapping $h = (h_1, \ldots, h_n)$ satisfies the conditions of the theorem (h has an inverse g which is continuously differentiable).
Then the theorem gives the joint pdf $f_{\boldsymbol{Y}}(y_1, \ldots, y_n)$ and we obtain $f_{Y_1}(y_1)$ by "integrating out" $y_2, \ldots, y_n$:

$$f_{Y_1}(y_1) = \int_{\mathbb{R}^{n-1}} \cdots \int f_{\boldsymbol{Y}}(y_1, \ldots, y_n)\, dy_2 \ldots dy_n$$

A common choice is $Y_i = X_i$, $i = 2, \ldots, n$.

(2) Often it is easier to directly compute the cdf of Y_1:

$$\begin{aligned} F_{Y_1}(y) &= P(Y_1 \le y) = P\big(h_1(X_1, \ldots, X_n) \le y\big) \\ &= P\big((X_1, \ldots, X_n) \in A_y\big) \\ &\quad \text{where } A_y = \{(x_1, \ldots, x_n) : h(x_1, \ldots, x_n) \le y\} \\ &= \int_{A_y} \cdots \int f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \end{aligned}$$

Differentiating F_{Y_1} we obtain the pdf of Y_1.

Example: Let $X_1, \ldots, X_n$ be *independent* with common distribution Uniform$(0, 1)$. Determine the pdf of $Y = \min(X_1, \ldots, X_n)$.

Solution: ...