Introduction to Probability

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 308

1. Jointly Distributed Random Variables
   - Joint Distribution Functions
   - Independent Random Variables
   - Sums of Independent Random Variables
   - Conditional Distributions: Discrete Case
   - Conditional Distributions: Continuous Case
   - Joint Probability Distributions of Functions of Random Variables

Subsection 1

Joint Distribution Functions

Joint Probability Distribution Functions

- Let X, Y be two random variables.
- We define the **joint cumulative probability distribution function** of X and Y by $F(a, b) = P\{X \leq a, Y \leq b\}$, $-\infty < a, b < \infty$.
- The distribution of X can be obtained from the joint distribution of X and Y as follows:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} = P\{X \leq a, Y < \infty\} \\ &= P(\lim_{b \to \infty} \{X \leq a, Y \leq b\}) \\ &= \lim_{b \to \infty} P\{X \leq a, Y \leq b\} \\ &= \lim_{b \to \infty} F(a, b) \equiv F(a, \infty). \end{aligned}$$

- Similarly, the cumulative distribution function of Y is given by

$$F_Y(b) = P\{Y \leq b\} = \lim_{a \to \infty} F(a, b) \equiv F(\infty, b).$$

- The distribution functions F_X and F_Y are sometimes referred to as the **marginal distributions** of X and Y.

Using the Joint Distribution Function

- All joint probability statements about X and Y can, in theory, be answered in terms of their joint distribution function.
- For instance, suppose we wanted to compute the joint probability that X is greater than a and Y is greater than b.
- This could be done as follows:

$$\begin{aligned} P\{X > a, Y > b\} &= 1 - P(\{X > a, Y > b\}^c) \\ &= 1 - P(\{X > a\}^c \cup \{Y > b\}^c) \\ &= 1 - P(\{X \leq a\} \cup \{Y \leq b\}) \\ &= 1 - [P\{X \leq a\} + P\{Y \leq b\} \\ &\qquad - P\{X \leq a, Y \leq b\}] \\ &= 1 - F_X(a) - F_Y(b) + F(a, b). \end{aligned}$$

A More General Formula

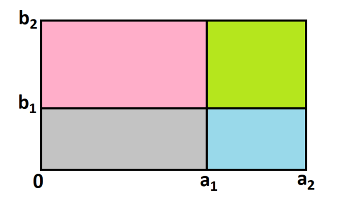
- We found $P\{X > a, Y > b\} = 1 - F_X(a) - F_Y(b) + F(a, b)$.
- This is the same as

$$\begin{aligned}&P\{a < X < \infty, b < X < \infty\}\\&= P(\{X < \infty, Y < \infty\}) - P(\{X < a, Y < \infty\})\\&\qquad - P(\{X < \infty, Y < b\}) + P(\{X < a, Y < b\}).\end{aligned}$$

- This is a special case of, for $a_1 < a_2$, $b_1 < b_2$:

$$\begin{aligned}&P\{a_1 < X \leq a_2, b_1 < Y \leq b_2\}\\&\qquad = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1).\end{aligned}$$

Its verification is based on the "Venn diagram":



Joint Probability Mass Function

- In the case when X and Y are both discrete random variables, it is convenient to define the **joint probability mass function** of X and Y by
$$p(x,y) = P\{X = x, Y = y\}.$$
- The probability mass function of X can be obtained from $p(x,y)$ by
$$p_X(x) = P\{X = x\} = \sum_{y:p(x,y)>0} p(x,y).$$
- Similarly,
$$p_Y(y) = \sum_{x:p(x,y)>0} p(x,y).$$

Example

- Suppose that 3 balls are randomly selected from an urn containing 3 red, 4 white and 5 blue balls.
  - Let X be the number of red balls chosen.
  - Let Y be the number of white balls chosen.
- The joint probability mass $p(i,j) = P\{X = i, Y = j\}$ is given by:

$$\begin{aligned}
p(0,0) &= \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220} & p(0,1) &= \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220} \\
p(0,2) &= \frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{30}{220} & p(0,3) &= \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220} \\
p(1,0) &= \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{30}{220} & p(1,1) &= \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220} \\
p(1,2) &= \frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}} = \frac{18}{220} & p(2,0) &= \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{15}{220} \\
p(2,1) &= \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{12}{220} & p(3,0) &= \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220}.
\end{aligned}$$

Example (Cont'd)

- These probabilities can most easily be expressed in tabular form.

$i \backslash j$	0	1	2	3	$P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$P = \{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

- The probability mass function of X is obtained by computing the row sums.
- The probability mass function of Y is obtained by computing the column sums.

Example

- Suppose that 15 percent of the families in a certain community have no children, 20 percent have 1 child, 35 percent have 2 children and 30 percent have 3.

  Suppose further that in each family each child is equally likely (independently) to be a boy or a girl.

  Suppose a family is chosen at random from this community.
  - Let B be the number of boys in this family;
  - Let G be the number of girls.

  The joint probability mass function is shown in the table:

$i \backslash j$	0	1	2	3	$P\{B=i\}$
0	0.15	0.10	0.0875	0.0375	0.3750
1	0.10	0.175	0.1125	0	0.3875
2	0.0875	0.1125	0	0	0.2000
3	0.0375	0	0	0	0.0375
$P\{G=j\}$	0.3750	0.3875	0.2000	0.0375	

Example (Cont'd)

- Some of the probabilities shown in the table are obtained as follows:

$$
\begin{array}{rcl}
P\{B=0,G=0\} & = & P\{\text{no children}\}=0.15; \\
P\{B=0,G=1\} & = & P\{\text{1 girl and total of 1 child}\} \\
 & = & P\{\text{1 child}\}P\{\text{1 girl}|\text{1 child}\}=(0.20)(\frac{1}{2}); \\
P\{B=0,G=2\} & = & P\{\text{2 girls and total of 2 children}\} \\
 & = & P\{\text{2 children}\}P\{\text{2 girls}|\text{2 children}\} \\
 & = & (0.35)(\frac{1}{2})^2.
\end{array}
$$

- The remaining probabilities in the table follow in a similar way.

Jointly Continuous Random Variables

- We say that X and Y are **jointly continuous** if there exists a function $f(x, y)$, defined for all real x and y, having the property that, for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane),

$$P\{(X, Y) \in C\} = \iint_{(x,y)\in C} f(x, y) dx dy.$$

- The function $f(x, y)$ is called the **joint probability density function** of X and Y.

Interpretation of Joint Probability Density

- If A and B are any sets of real numbers, then, by defining $C = \{(x, y) : x \in A, y \in B\}$, we see from the equation that
$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y) dx dy.$$
- We have
$$\begin{array}{rcl} F(a, b) & = & P\{X \in (-\infty, a], Y \in (-\infty, b]\} \\ & = & \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y) dx dy. \end{array}$$
- It follows, upon differentiation, that
$$f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b)$$
wherever the partial derivatives are defined.

Another Interpretation of Joint Probability Density

- Another interpretation of the joint density function, obtained from

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)dxdy.$$

- We have

$$\begin{aligned} & P\{a < X < a + da, b < Y < b + db\} \\ & \qquad = \int_b^{d+db} \int_a^{a+da} f(x,y)dxdy \\ & \qquad \approx f(a,b)dadb, \end{aligned}$$

  when da and db are small and $f(x,y)$ is continuous at a, b.
- Hence, $f(a,b)$ is a measure of how likely it is that the random vector (X, Y) will be near (a,b).

Probability Densities from Joint Density Functions

- Suppose X and Y are jointly continuous.
  Then, they are individually continuous, and their probability density functions are
  $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$
- This result can be obtained as follows:
  $$\begin{array}{rcl} P\{X \in A\} & = & P\{X \in A, Y \in (-\infty,\infty)\} \\ & = & \int_A \int_{-\infty}^{\infty} f(x,y)dydx \\ & = & \int_A f_X(x)dx. \end{array}$$
  Thus, $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$ is, indeed, the probability density function of X.
- We work similarly to show that $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$ is the probability density function of Y.

Example

- The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Compute (a) $P\{X > 1, Y < 1\}$, (b) $P\{X < Y\}$ and (c) $P\{X < a\}$.

$$\begin{array}{lll} (a) \quad P\{X > 1, Y < 1\} & = & \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}dxdy \\ & = & \int_0^1 2e^{-2y} \int_1^\infty e^{-x}dxdy \\ & = & \int_0^1 2e^{-2y}(-e^{-x}|_1^\infty)dy = e^{-1}\int_0^1 2e^{-2y}dy \\ & = & e^{-1}[-e^{-2y}]_0^1 = e^{-1}(1-e^{-2}). \end{array}$$

Example (Cont'd)

$$
\begin{array}{rrcl}
(b) & P\{X<Y\} & = & \iint_{(x,y):x<y} 2e^{-x}e^{-2y}dxdy \\
 & & = & \int_0^\infty \int_0^y 2e^{-x}e^{-2y}dx\,dy \\
 & & = & \int_0^\infty 2e^{-2y}\int_0^y e^{-x}dx\,dy \\
 & & = & \int_0^\infty 2e^{-2y}(1-e^{-y})dy \\
 & & = & \int_0^\infty 2e^{-2y}dy - \int_0^\infty 2e^{-3y}dy \\
 & & = & (-e^{-2y})|_0^\infty - (-\frac{2}{3}e^{-3y})|_0^\infty \\
 & & = & 1-\frac{2}{3}=\frac{1}{3}. \\
(c) & P\{X<a\} & = & \int_0^a \int_0^\infty 2e^{-2y}e^{-x}dy\,dx \\
 & & = & \int_0^a e^{-x}\int_0^\infty 2e^{-2y}dy\,dx \\
 & & = & \int_0^a e^{-x}(-e^{-2y})|_0^\infty dx \\
 & & = & \int_0^a e^{-x}dx = (-e^{-x})|_0^a = 1-e^{-a}.
\end{array}
$$

Example

- Consider a circle of radius R, and suppose that a point within the circle is randomly chosen in such a manner that all regions within the circle of equal area are equally likely to contain the point.

  Let the center of the circle denote the origin.

  Define X and Y to be the coordinates of the point chosen.

  Since (X, Y) is equally likely to be near each point in the circle, the joint density function of X and Y is given by

$$f(x,y) = \begin{cases} c, & \text{if } x^2 + y^2 \leq R^2 \\ 0, & \text{if } x^2 + y^2 > R^2 \end{cases}, \text{ for some value of } c.$$

  (a) Determine c.
  (b) Find the marginal density functions of X and Y.
  (c) Compute the probability that D, the distance from the origin of the point selected, is less than or equal to a.
  (d) Find $E[D]$.

Example (Part (a))

(a) Recall that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx = 1$.
Thus,

$$c\iint_{x^2+y^2\leq R^2} dydx = 1.$$

We can evaluate $\iint_{x^2+y^2\leq R^2} dydx$ in one of two ways:

- By using polar coordinates

$$\iint_{x^2+y^2\leq R^2} dydx = \int_0^{2\pi}\int_0^R rdrd\theta = 2\pi\int_0^R rdr = 2\pi\frac{R^2}{2} = \pi R^2;$$

- By noting that it represents the area of the circle and is thus equal to πR^2.

Hence,

$$c = \frac{1}{\pi R^2}.$$

Example (Part (b))

(b)

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \frac{1}{\pi R^2}\int_{x^2+y^2\leq R^2} dy \\ &= \frac{1}{\pi R^2}\int_{-c}^{c} dy,\ \text{where } c=\sqrt{R^2-x^2}, \\ &= \frac{2}{\pi R^2}\sqrt{R^2-x^2}, \quad x^2\leq R^2. \end{aligned}$$

It equals 0 when $x^2 > R^2$.

By symmetry, the marginal density of Y is given by

$$f_Y(y) = \begin{cases} \frac{2}{\pi R^2}\sqrt{R^2-y^2}, & y^2 \leq R^2 \\ 0, & y^2 > R^2 \end{cases}$$

Example (Parts (c) & (d))

(c) The distribution function of $D = \sqrt{X^2 + Y^2}$, the distance from the origin, is obtained, for $0 \leq a \leq R$, as follows:

$$\begin{aligned} F_D(a) &= P\{\sqrt{X^2+Y^2} \leq a\} = P\{X^2+Y^2 \leq a^2\} \\ &= \iint_{x^2+y^2 \leq a^2} f(x,y)dydx = \frac{1}{\pi R^2} \iint_{x^2+y^2 \leq a^2} dydx \\ &= \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}, \end{aligned}$$

where we used that $\iint_{x^2+y^2 \leq a^2} dydx$ is the area of a circle of radius a and thus is equal to πa^2.

(d) From part (c), the density function of D is

$$f_D(a) = \frac{2a}{R^2}, \quad 0 \leq a \leq R.$$

Hence,

$$E[D] = \frac{2}{R^2} \int_0^R a^2 da = \frac{2R}{3}.$$

Example

- The joint density of X and Y is given by

$$f(x,y) = \begin{cases} e^{-(x+y)}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Find the density function of the random variable X/Y.

We start by computing the distribution function of X/Y.

For $a > 0$,

$$\begin{aligned} F_{X/Y}(a) &= P\{\tfrac{X}{Y} \le a\} = \iint_{x/y \le a} e^{-(x+y)} dxdy \\ &= \int_0^\infty \int_0^{ay} e^{-(x+y)} dxdy \\ &= \int_0^\infty (1 - e^{-ay}) e^{-y} dy \\ &= \{-e^{-y} + \tfrac{e^{-(a+1)y}}{a+1}\}\big|_0^\infty = 1 - \tfrac{1}{a+1}. \end{aligned}$$

Differentiation shows that the density function of X/Y is given by $f_{X/Y}(a) = \frac{1}{(a+1)^2}, \ 0 < a < \infty$.

Joint Probability Distributions for n Variables

- The **joint cumulative probability distribution function** $F(a_1, a_2, \ldots, a_n)$ of the n random variables $X_1, X_2, \ldots, X_n$ is defined by
$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\}.$$
- The n random variables are said to be **jointly continuous** if there exists a function $f(x_1, x_2, \ldots, x_n)$, called the **joint probability density function**, such that, for any set C in n-space,
$$P\{(X_1, X_2, \ldots, X_n) \in C\} = \iint_{(x_1,\ldots,x_n)\in C} \cdots \int f(x_1, \ldots, x_n) dx_1 dx_2 \cdots dx_n.$$
- For any n sets of real numbers $A_1, A_2, \ldots, A_n$,
$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \int_{A_n} \int_{A_{n-1}} \cdots \int_{A_1} f(x_1, \ldots, x_n) dx_1 dx_2 \cdots dx_n.$$

Example: The Multinomial Distribution

- One of the most important joint distributions is the multinomial distribution, which arises when a sequence of n independent and identical experiments is performed.

  Suppose that each experiment can result in any one of r possible outcomes, with respective probabilities $p_1, p_2, \ldots, p_r$, $\sum_{i=1}^r p_i = 1$.

  Let X_i denote the number of the n experiments that result in outcome number i.

  Then, whenever $\sum_{i=1}^r n_i = n$,

$$P\{X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r\} = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}.$$

  - Any sequence of outcomes for the n experiments that leads to outcome i occurring n_i times for $i = 1, 2, \ldots, r$ will, by the independence of experiments, have probability $p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$ of occurring.
  - The number of such sequences of outcomes is $\frac{n!}{n_1! n_2! \cdots n_r!}$..

Example: The Multinomial Distribution (Remarks)

- The joint distribution whose joint probability mass function is specified by the preceding equation is called the **multinomial distribution**.
- Remarks:
  - When $r = 2$, the multinomial reduces to the binomial distribution.
  - Any sum of a fixed set of the X_is will have a binomial distribution.
    - $\sum_{i \in N} X_i$ represents the number of the n experiments whose outcome is in N;
    - Each experiment will independently have such an outcome with probability $\sum_{i \in N} p_i$.

  That is, if $N \subseteq \{1, 2, \ldots, r\}$, then $\sum_{i \in N} X_i$ will be a binomial random variable with parameters n and $p = \sum_{i \in N} p_i$.

Subsection 2

Independent Random Variables

Independent Random Variables

- The random variables X and Y are said to be **independent** if, for any two sets of real numbers A and B,

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}.$$

- In other words, X and Y are independent if, for all A and B, the events $E_A = \{X \in A\}$ and $F_B = \{Y \in B\}$ are independent.
- It can be shown by using the three axioms of probability that independence will follow if and only if, for all a, b,

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}.$$

- Hence, in terms of the joint distribution function F of X and Y, X and Y are independent if

$$F(a,b) = F_X(a)F_Y(b), \text{ for all } a, b.$$

Independent Discrete Random Variables

- When X and Y are discrete random variables, the condition of independence is equivalent to

$$p(x,y) = p_X(x)p_Y(y), \text{ for all } x, y.$$

- Suppose, first, that the defining equation is satisfied.
  - Let $A = \{x\}$;
  - Let $B = \{y\}$;

  The the defining equation yields the preceding equation.
- If the latter equation is valid, then, for any sets A, B,

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{y \in B} \sum_{x \in A} p(x,y) \\ &= \sum_{y \in B} \sum_{x \in A} p_X(x)p_Y(y) \\ &= \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) \\ &= P\{Y \in B\}P\{X \in A\}. \end{aligned}$$

Independent Jointly Continuous Random Variables

- In the jointly continuous case, the condition of independence is equivalent to
$$f(x,y) = f_X(x)f_Y(y), \text{ for all } x, y.$$
- Loosely speaking, X and Y are independent if knowing the value of one does not change the distribution of the other.
- Random variables that are not independent are said to be **dependent**.

Example

- Suppose that $n+m$ independent trials having a common probability of success p are performed.
  - Let X be the number of successes in the first n trials;
  - Let Y be the number of successes in the final m trials.

  Then X and Y are independent, since knowing the number of successes in the first n trials does not affect the distribution of the number of successes in the final m trials (by the assumption of independent trials).

  In fact, for integral x and y, with $0 \leq x \leq n$, $0 \leq y \leq m$,

$$\begin{aligned} P\{X=x, Y=y\} &= \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y} \\ &= P\{X=x\} P\{Y=y\}. \end{aligned}$$

  In contrast, X and Z will be dependent, where Z is the total number of successes in the $n+m$ trials.

Example

- Suppose that the number of people who enter a post office on a given day is a Poisson random variable with parameter λ.
  Assume each person who enters the post office is:
  - A male with probability p;
  - A female with probability $1-p$.

  Show then that the number of males and females entering the post office are independent Poisson random variables with respective parameters λp and $\lambda(1-p)$.
  - Let X be the number of males;
  - Let Y be the number of females that enter the post office.

  To obtain an expression for $P\{X=i,Y=j\}$, we condition on $X+Y$ as follows:

$$\begin{aligned}&P\{X=i,Y=j\}\\&=P\{X=i,Y=j|X+Y=i+j\}P\{X+Y=i+j\}\\&\ +P\{X=i,Y=j|X+Y\neq i+j\}P\{X+Y\neq i+j\}.\end{aligned}$$

  $P\{X=i,Y=j|X+Y\neq i+j\}$ is clearly 0.

Example (Cont'd)

- So we obtain

  $$P\{X=i,Y=j\}=P\{X=i,Y=j|X+Y=i+j\}P\{X+Y=i+j\}.$$

  Now, because $X+Y$ is the total number of people who enter the post office, we get, by hypothesis,

  $$P\{X+Y=i+j\}=e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}.$$

  Recall that each person entering will be male with probability p. Thus, given that $i+j$ people do enter the post office, the probability that exactly i of them will be male (and thus j of them female) is just $\binom{i+j}{i}p^i(1-p)^j$.
  That is,

  $$P\{X=i,Y=j|X+Y=i+j\}=\binom{i+j}{i}p^i(1-p)^j.$$

Example (Cont'd)

- Now we get

$$\begin{aligned} P\{X=i,Y=j\} &= \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\ &= e^{-\lambda} \frac{(\lambda p)^i}{i!j!} [\lambda(1-p)]^j \\ &= \frac{e^{-\lambda p}(\lambda p)^i}{i!} e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!}. \end{aligned}$$

Hence,

$$P\{X=i\} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} \sum_j e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!}.$$

Similarly,

$$P\{Y=j\} = e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!}.$$

These equations establish the desired result.

Example

- A man and a woman decide to meet at a certain location.
  If each of them independently arrives at a time uniformly distributed between 12 noon and 1 P.M., find the probability that the first to arrive has to wait longer than 10 minutes.
  - Let X be the time past 12 that the man arrives;
  - Let Y be the time past 12 that the woman arrives.

  X and Y are independent random variables, each of which is uniformly distributed over $(0, 60)$.
  The desired probability is $P\{X + 10 < Y\} + P\{Y + 10 < X\}$.
  By symmetry, it equals $2P\{X + 10 < Y\}$. We get:

$$
\begin{aligned}
2P\{X+10<Y\} &= 2\iint_{x+10<y} f(x,y)dxdy \\
&= 2\iint_{x+10<y} f_X(x)f_Y(y)dxdy \\
&= 2\int_{10}^{60}\int_{0}^{y-10}\left(\tfrac{1}{60}\right)^2 dxdy \\
&= \tfrac{2}{(60)^2}\int_{10}^{60}(y-10)dy = \tfrac{25}{36}.
\end{aligned}
$$

Example (Buffon's Needle Problem)

- A table is ruled with equidistant parallel lines a distance D apart.

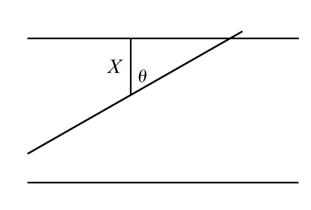
  A needle of length L, where $L \leq D$, is randomly thrown on the table.

  What is the probability that the needle will intersect one of the lines (the other possibility being that the needle will be completely contained in the strip between two lines)?

  We determine the position of the needle by specifying:

  (1) The distance X from the middle point of the needle to the nearest parallel line;

  (2) The angle θ between the needle and the projected line of length X.



  The needle will intersect a line if the hypotenuse of the right triangle in the figure is less than $\frac{L}{2}$, i.e., if $\frac{X}{\cos\theta} < \frac{L}{2}$ or $X < \frac{L}{2}\cos\theta$.

Example (Cont'd)

- X varies between 0 and $\frac{D}{2}$;
  θ varies between 0 and $\frac{\pi}{2}$.
  It is reasonable to assume that they are independent, uniformly distributed random variables over these respective ranges.
  Hence,

$$\begin{aligned} P\{X < \tfrac{L}{2}\cos\theta\} &= \iint_{x<\frac{L}{2}\cos y} f_X(x) f_\theta(y) dx dy \\ &= \frac{4}{\pi D}\int_0^{\pi/2}\int_0^{L/2\cos y} dx dy \\ &= \frac{4}{\pi D}\int_0^{\pi/2}\frac{L}{2}\cos y dy \\ &= \frac{2L}{\pi D}. \end{aligned}$$

Characterization of the Normal Distribution

- Let X and Y denote the horizontal and vertical miss distances when a bullet is fired at a target.

  Assume that:

  1. X and Y are independent continuous random variables having differentiable density functions.
  2. The joint density $f(x, y) = f_X(x)f_Y(y)$ of X and Y depends on (x, y) only through $x^2 + y^2$.

  Loosely put, assumption 2 states that the probability of the bullet landing on any point of the x-y plane depends only on the distance of the point from the target and not on its angle of orientation.

  An equivalent way of phrasing this assumption is to say that the joint density function is *rotation invariant*.

  We show that Assumptions 1 and 2 imply that X and Y are normally distributed random variables.

Characterization of the Normal Distribution (Cont'd)

- Note first that the assumptions yield the relation

$$f(x,y) = f_X(x) f_Y(y) = g(x^2 + y^2).$$

for some function g.

Differentiating with respect to x yields

$$f_X'(x) f_Y(y) = 2xg'(x^2 + y^2).$$

Dividing the latter by the former gives

$$\frac{f_X'(x)}{f_X(x)} = \frac{2xg'(x^2+y^2)}{g(x^2+y^2)} \quad \text{or} \quad \frac{f_X'(x)}{2xf_X(x)} = \frac{g'(x^2+y^2)}{g(x^2+y^2)}.$$

The value of the left depends only on x.

On the other hand, the value of the right depends on $x^2 + y^2$.

Hence, the left-hand side must be the same for all x.

Characterization of the Normal Distribution (Cont'd)

- To see this, consider any x_1, x_2.

  Let y_1, y_2 be such that $x_1^2 + y_1^2 = x_2^2 + y_2^2$.

  Then, from the last equation, we obtain

  $$\frac{f_X'(x_1)}{2x_1 f_X(x_1)} = \frac{g'(x_1^2 + y_1^2)}{g(x_1^2 + y_1^2)} = \frac{g'(x_2^2 + y_2^2)}{g(x_2^2 + y_2^2)} = \frac{f_X'(x_2)}{2x_2 f_X(x_2)}.$$

  Hence, $\frac{f_X'(x)}{x f_X(x)} = c$ or $\frac{d}{dx}(\log f_X(x)) = cx$.

  Upon integration of both sides, we get $\log f_X(x) = a + \frac{cx^2}{2}$ or $f_X(x) = ke^{cx^2/2}$.

  But $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

  Hence, c is necessarily negative, and we may write $c = -\frac{1}{\sigma^2}$.

  Thus, $f_X(x) = ke^{-x^2/2\sigma^2}$.

  That is, X is a normal random variable with parameters $\mu = 0$ and σ^2.

Characterization of the Normal Distribution (Cont'd)

- A similar argument can be applied to $f_Y(y)$ to show that

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\overline{\sigma}} e^{-y^2/2\overline{\sigma}^2}.$$

Furthermore, it follows from Assumption 2 that $\sigma^2 = \overline{\sigma}^2$.

Hence, X and Y are independent, identically distributed normal random variables with parameters $\mu = 0$ and σ^2.

Characterization of Independence

Proposition

The continuous (discrete) random variables X and Y are independent if and only if their joint probability density (mass) function can be expressed as

$$f_{X,Y}(x,y) = h(x)g(y), \quad -\infty < x < \infty, -\infty < y < \infty.$$

- We give the proof in the continuous case.

  First, note that independence implies that the joint density is the product of the marginal densities of X and Y.

  So the preceding factorization will hold when the random variables are independent.

Characterization of Independence (Converse)

- Now, suppose that $f_{X,Y}(x,y) = h(x)g(y)$.

  Then
$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{\infty}^{\infty} f_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{\infty} h(x)dx \int_{-\infty}^{\infty} g(y)dy \\ &= C_1C_2, \end{aligned}$$
  where:
  - $C_1 = \int_{-\infty}^{\infty} h(x)dx$;
  - $C_2 = \int_{-\infty}^{\infty} g(y)dy$.

  Also,
$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = C_2h(x); \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = C_1g(y). \end{aligned}$$
  But $C_1C_2 = 1$. Hence, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Example

- Let X and Y be random variables.

  Tell whether they are independent if their joint density function is:

  (a) $f(x,y) = 6e^{-2x}e^{-3y}$, for $0 < x < \infty$, $0 < y < \infty$, and is equal to 0 outside this region;

  (b) $f(x,y) = 24xy$, if $0 < x < 1$, $0 < y < 1$, $0 < x + y < 1$, and is equal to 0 otherwise.

(a) The joint density function factors.

Thus the random variables are independent (one is exponential with rate 2 and the other exponential with rate 3).

(b) The region in which the joint density is nonzero cannot be expressed in the form $x \in A, y \in B$.

Thus, the joint density does not factor.

Therefore, the random variables are not independent.

Example (Cont'd)

- To explain (b) more explicitly, let

$$I(x, y) = \begin{cases} 1, & \text{if } 0 < x < 1, 0 < y < 1, 0 < x + y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Then

$$f(x, y) = 24xyI(x, y).$$

This clearly does not factor into a part depending only on x and another depending only on y.

Independence of n Random Variables

- The concept of independence may be defined for more than two random variables.
- In general, the n random variables $X_1, X_2, \ldots, X_n$ are said to be **independent** if, for all sets of real numbers $A_1, A_2, \ldots, A_n$,

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}.$$

- As before, it can be shown that this condition is equivalent to asserting that, for all $a_1, a_2, \ldots, a_n$,

$$P\{X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n\} = \prod_{i=1}^{n} P\{X_i \leq a_i\}.$$

- Finally, we say that an infinite collection of random variables is **independent** if every finite subcollection of them is independent.

Example

- Let X, Y, Z be independent and uniformly distributed over $(0, 1)$. Compute $P\{X \geq YZ\}$.

  By independence, for $0 \leq x \leq 1$, $0 \leq y \leq 1$, $0 \leq z \leq 1$,

  $$f_{X,Y,Z}(x, y, z) = f_X(x) f_Y(y) f_Z(z) = 1.$$

  Thus, we get

  $$\begin{aligned} P\{X \geq YZ\} &= \iiint_{x \geq yz} f_{X,Y,Z}(x, y, z) dx dy dz \\ &= \int_0^1 \int_0^1 \int_{yz}^1 dx dy dz \\ &= \int_0^1 \int_0^1 (1 - yz) dy dz \\ &= \int_0^1 (1 - \frac{z}{2}) dz = \frac{3}{4}. \end{aligned}$$

Symmetry of Independence

- The random variables X and Y are independent if their joint density function (or mass function in the discrete case) is the product of their individual density (or mass) functions.
- Therefore, to say that X is independent of Y is equivalent to saying that Y is independent of X - or just that X and Y are independent.
- Sometimes, in considering whether X is independent of Y, it is not at all intuitive that knowing the value of Y will not change the probabilities concerning X.

  Because of this symmetry, it may, then, be beneficial to interchange the roles of X and Y and ask instead whether Y is independent of X.

Example

- Consider the game of craps.

  If the initial throw of the dice results in the sum of the dice equaling 4, then the player will continue to throw the dice until the sum is either 4 or 7.
  - If this sum is 4, then the player wins;
  - If it is 7, then the player loses.

  Let N denote the number of throws needed until either 4 or 7 appears.

  Let X denote the value (either 4 or 7) of the final throw.

  Is N independent of X?

  The question asks whether knowing which of 4 or 7 occurs first affects the distribution of the number of throws needed until that number appears.

  The answer to this question does not seem intuitively obvious.

Example (Cont'd)

- Suppose that we turn the question around and ask whether X is independent of N.

  That is, does knowing how many throws it takes to obtain a sum of either 4 or 7 affect the probability that that sum is equal to 4?

  For instance, suppose we know that it takes n throws of the dice to obtain a sum of either 4 or 7.

  Does this affect the probability distribution of the final sum?

  Clearly not, since all that is important is that its value is either 4 or 7, and the fact that none of the first $n-1$ throws were either 4 or 7 does not change the probabilities for the nth throw.

  Thus, we can conclude that X is independent of N.

  Equivalently (by symmetry), N is independent of X.

Example

- Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed continuous random variables.

  Suppose that we observe these random variables in sequence.

  If $X_n > X_i$ for each $i = 1, \ldots, n-1$, then we say that X_n is a **record value**.

  That is, each random variable that is larger than all those preceding it is called a record value.

  Let A_n denote the event that X_n is a record value.

  Is A_{n+1} independent of A_n?

  That is, does knowing that the nth random variable is the largest of the first n change the probability that the $(n+1)$st random variable is the largest of the first $n+1$?

  While it is true that A_{n+1} is independent of A_n, this may not be intuitively obvious.

Example (Cont'd)

- If we turn the question around and ask whether A_n is independent of A_{n+1}, then the result is more easily understood.

  For knowing that the $(n+1)$st value is larger than $X_1, \ldots, X_n$ clearly gives us no information about the relative size of X_n among the first n random variables.

  By symmetry, it is clear that each of the first n random variables is equally likely to be the largest of this set.

  Therefore,

$$P(A_n | A_{n+1}) = P(A_n) = \frac{1}{n}.$$

  Hence, we can conclude that A_n and A_{n+1} are independent events.

Sequential Verification

- Recall the identity

$$
\begin{aligned}
&P\{X_1 \leq a_1, \ldots, X_n \leq a_n\} \\
&= P\{X_1 \leq a_1\} P\{X_2 \leq a_2 | X_1 \leq a_1\} \cdots \\
&\qquad\qquad P\{X_n \leq a_n | X_1 \leq a_1, \ldots, X_{n-1} \leq a_{n-1}\}
\end{aligned}
$$

- From this, it follows that the independence of $X_1, \ldots, X_n$ can be established sequentially.
- That is, we can show that these random variables are independent by showing that:
  - X_2 is independent of X_1;
  - X_3 is independent of X_1, X_2;
  - X_4 is independent of X_1, X_2, X_3;

    $\vdots$
  - X_n is independent of $X_1, \ldots, X_{n-1}$.

Subsection 3

Sums of Independent Random Variables

Probability Distribution of Sum of Independent Variables

- Suppose that X and Y are independent, continuous random variables having probability density functions f_X and f_Y.
- The cumulative distribution function of $X+Y$ is obtained as follows:

$$\begin{aligned} F_{X+Y}(a) &= P\{X+Y \leq a\} \\ &= \iint_{x+y\leq a} f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)dx f_Y(y)dy \\ &= \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy. \end{aligned}$$

- The cumulative distribution function F_{X+Y} is called the **convolution** of the distributions F_X and F_Y (the cumulative distribution functions of X and Y, respectively).

Probability Density of Sum of Independent Variables

- By differentiating, we find that the probability density function f_{X+Y} of $X+Y$ is given by

$$\begin{aligned} f_{X+Y}(a) &= \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy \\ &= \int_{-\infty}^{\infty} \frac{d}{da}F_X(a-y)f_Y(y)dy \\ &= \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy. \end{aligned}$$

Sum of Two Independent Uniform Random Variables

- If X and Y are independent random variables, both uniformly distributed on $(0,1)$, calculate the probability density of $X+Y$.

  We have
  $$f_X(a) = f_Y(a) = \begin{cases} 1, & 0 < a < 1 \\ 0, & \text{otherwise} \end{cases}$$

  Hence, we obtain $f_{X+Y}(a) = \int_0^1 f_X(a-y)dy$.

  For $0 \leq a \leq 1$, this yields $f_{X+Y}(a) = \int_0^a dy = a$.

  For $1 < a < 2$, we get $f_{X+Y}(a) = \int_{a-1}^1 dy = 2-a$.

  Hence,
  $$f_{X+Y}(a) = \begin{cases} a, & 0 \leq a \leq 1 \\ 2-a, & 1 < a < 2 \\ 0, & \text{otherwise} \end{cases}$$

  Because of the shape of its density function, the random variable $X+Y$ is said to have a **triangular distribution**.

Sum of n Independent Uniform Random Variables

- Now, suppose that $X_1, X_2, \ldots, X_n$ are independent uniform $(0,1)$ random variables, and let $F_n(x) = P\{X_1 + \cdots + X_n \leq x\}$.
- Whereas a general formula for $F_n(x)$ is messy, it has a particularly nice form when $x \leq 1$.
- We use mathematical induction to prove that

$$F_n(x) = \frac{x^n}{n!}, \quad 0 \leq x \leq 1.$$

- The proceeding equation is true for $n = 1$.
- Assume that $F_{n-1}(x) = \frac{x^{n-1}}{(n-1)!}, \quad 0 \leq x \leq 1$.
- Write $\sum_{i=1}^n X_i = \sum_{i=1}^{n-1} X_i + X_n$ and use the fact that the X_i are all nonnegative, to get, for $0 \leq x \leq 1$,

$$\begin{array}{rcl} F_n(x) & = & \int_0^1 F_{n-1}(x-y) f_{X_n}(y) dy \\ & = & \frac{1}{(n-1)!} \int_0^x (x-y)^{n-1} dy = \frac{x^n}{n!}. \end{array}$$

Number of Variables Needed to Exceed One

- Let $X_1, X_2, \ldots$ be independent uniform $(0,1)$ random variables. Determine $E[N]$, where $N = \min\{n : X_1 + \cdots + X_n > 1\}$.
  Note that N is greater than $n > 0$ if and only if $X_1 + \cdots + X_n \leq 1$.
  Thus, $P\{N > n\} = F_n(1) = \frac{1}{n!}$, $n > 0$.
  Note, also, that $P\{N > 0\} = 1 = \frac{1}{0!}$.
  Hence, for $n > 0$,
  $$P\{N = n\} = P\{N > n-1\} - P\{N > n\} = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}.$$
  Therefore,
  $$E[N] = \sum_{n=1}^{\infty} \frac{n(n-1)}{n!} = \sum_{n=2}^{\infty} \frac{1}{(n-2)!} = e.$$
  That is, the mean number of independent uniform $(0,1)$ random variables that must be summed for the sum to exceed 1 is equal to e.

Sum of Independent Normal Random Variables

Proposition

If X_i, $i = 1, \ldots, n$, are independent random variables that are normally distributed with respective parameters μ_i, σ_i^2, $i = 1, \ldots, n$, then $\sum_{i=1}^n X_i$ is normally distributed with parameters $\sum_{i=1}^n \mu_i$ and $\sum_{i=1}^n \sigma_i^2$.

- Let X and Y be independent normal random variables with:
  - X having mean 0 and variance σ^2;
  - Y having mean 0 and variance 1.

  We determine the density function of $X + Y$.
  Let $c = \frac{1}{2\sigma^2} + \frac{1}{2} = \frac{1+\sigma^2}{2\sigma^2}$.

$$\begin{aligned} f_X(a-y)f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(a-y)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} \\ &= \frac{1}{2\pi\sigma} \exp\left\{-\frac{a^2}{2\sigma^2} + \frac{2ay}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{\sigma^2 y^2}{2\sigma^2}\right\} \\ &= \frac{1}{2\pi\sigma} \exp\left\{-\frac{a^2}{2\sigma^2} - \frac{1+\sigma^2}{2\sigma^2}\left(y^2 - \frac{2ay}{1+\sigma^2}\right)\right\}. \end{aligned}$$

Sum of Independent Normal Random Variables (Cont'd)

- We calculated
$f_X(a-y)f_Y(y) = \frac{1}{2\pi\sigma}\exp\{-\frac{a^2}{2\sigma^2}\}\exp\{-c(y^2 - 2y\frac{a}{1+\sigma^2})\}$.
Hence, we get

$$\begin{array}{rcl} f_{X+Y}(a) & = & \frac{1}{2\pi\sigma}\exp\{-\frac{a^2}{2\sigma^2}\}\exp\{\frac{a^2}{2\sigma^2(1+\sigma^2)}\} \\ & & \times\int_{-\infty}^{\infty}\exp\{-c(y-\frac{a}{1+\sigma^2})^2\}dy \\ & = & \frac{1}{2\pi\sigma}\exp\{-\frac{a^2}{2(1+\sigma^2)}\}\int_{-\infty}^{\infty}\exp\{-cx^2\}dx \\ & = & C\exp\{-\frac{a^2}{2(1+\sigma^2)}\}. \end{array}$$

Note that $C := \frac{1}{2\pi\sigma}\int_{-\infty}^{\infty}\exp\{-cx^2\}dx$ does not depend on a.
This implies that $X+Y$ is normal with mean 0 and variance $1+\sigma^2$.

Sum of Independent Normal Random Variables (Cont'd)

- Suppose that X_1 and X_2 are independent normal random variables with X_i having mean μ_i and variance σ_i^2, $i = 1, 2$.

  Then
  $$X_1 + X_2 = \sigma_2\left(\frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2}\right) + \mu_1 + \mu_2.$$

  Note that:
  - $\frac{X_1 - \mu_1}{\sigma_2}$ is normal with mean 0 and variance $\frac{\sigma_1^2}{\sigma_2^2}$;
  - $\frac{X_2 - \mu_2}{\sigma_2}$ is normal with mean 0 and variance 1.

  From our previous result, we get that $\frac{X_1 - \mu_1}{\sigma_2} + \frac{X_2 - \mu_2}{\sigma_2}$ is normal with mean 0 and variance $1 + \frac{\sigma_1^2}{\sigma_2^2}$.

  This implies that $X_1 + X_2$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_2^2(1 + \frac{\sigma_1^2}{\sigma_2^2}) = \sigma_1^2 + \sigma_2^2$.

  Thus, the proposition is established when $n = 2$.

Sum of Independent Normal Random Variables (Cont'd)

- The general case now follows by induction.

  That is, assume that the proposition is true when there are $n-1$ random variables.

  Now consider the case of n, and write

  $$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n.$$

  By the induction hypothesis, $\sum_{i=1}^{n-1} X_i$ is normal with mean $\sum_{i=1}^{n-1} \mu_i$ and variance $\sum_{i=1}^{n-1} \sigma_i^2$.

  Therefore, by the result for $n=2$, $\sum_{i=1}^{n} X_i$ is normal with mean $\sum_{i=1}^{n} X_i$ and variance $\sum_{i=1}^{n} \sigma_i^2$.

Sum of Independent Poisson Random Variables

- Let X and Y be independent Poisson random variables with respective parameters λ_1 and λ_2.

  We compute the distribution of $X+Y$.

  The event $\{X+Y=n\}$ may be written as the union of the disjoint events $\{X=k,Y=n-k\}$, $0 \leq k \leq n$:

$$\begin{aligned} P\{X+Y=n\} &= \sum_{k=0}^{n} P\{X=k,Y=n-k\} \\ &= \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\} \\ &= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n. \end{aligned}$$

  Thus, X_1+X_2 has a Poisson distribution with parameter $\lambda_1+\lambda_2$.

Sum of Independent Binomial Random Variables

- Let X and Y be independent binomial random variables with respective parameters (n, p) and (m, p).

  By recalling the interpretation of a binomial random variable, we can conclude that $X + Y$ is binomial with parameters $(n + m, p)$.

  This follows because:

  - X represents the number of successes in n independent trials, each of which results in a success with probability p.
  - Y represents the number of successes in m independent trials, each of which results in a success with probability p.

  By hypothesis, X and Y are independent.

  Hence, $X + Y$ represents the number of successes in $n + m$ independent trials when each trial has probability p of success.

  Therefore, $X + Y$ is a binomial random variable with parameters $(n + m, p)$.

Sum of Independent Binomial Random Variables (Cont'd)

- We check this conclusion analytically:

$$\begin{aligned} P\{X+Y=k\} &= \textstyle\sum_{i=0}^{n} P\{X=i, Y=k-i\} \\ &= \textstyle\sum_{i=0}^{n} P\{X=i\}P\{Y=k-i\} \\ &= \textstyle\sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i} \binom{m}{k-i} p^{k-i} q^{m-k+i}, \end{aligned}$$

where $q = 1-p$ and where $\binom{r}{j} = 0$ when $j < 0$.
Thus,

$$\begin{aligned} P\{X+Y=k\} &= p^k q^{n+m-k} \textstyle\sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i} \\ &= \binom{n+m}{k} p^k q^{n+m-k}. \end{aligned}$$

The last equality uses the combinatorial identity

$$\binom{n+m}{k} = \sum_{i=0}^{n} \binom{n}{i}\binom{m}{k-i}.$$

Subsection 4

Conditional Distributions: Discrete Case

Conditional Distributions

- If X and Y are discrete random variables, we define the **conditional probability mass function of X given that $Y = y$**, by

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

  for all values of y such that $p_Y(y) > 0$.
- Similarly, the conditional probability distribution function of X given that $Y = y$ is defined, for all y such that $p_Y(y) > 0$, by

$$F_{X|Y}(x|y) = P\{X \le x|Y = y\} = \sum_{a \le x} p_{X|Y}(a|y).$$

- In other words, the definitions are exactly the same as in the unconditional case, except that everything is now conditional on the event that $Y = y$.

The Case of Independent Variables

- If X is independent of Y, then the conditional mass function and the distribution function are the same as the respective unconditional ones.
- This follows because if X is independent of Y, then

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X=x|Y=y\} \\ &= \frac{P\{X=x,Y=y\}}{P\{Y=y\}} \\ &= \frac{P\{X=x\}P\{Y=y\}}{P\{Y=y\}} \\ &= P\{X=x\}. \end{aligned}$$

Example

- Suppose that $p(x, y)$, the joint probability mass function of X and Y, is given by

$$p(0,0) = 0.4,\ \ p(0,1) = 0.2,\ \ p(1,0) = 0.1,\ \ p(1,1) = 0.3.$$

Calculate the conditional probability mass function of X given that $Y = 1$.

We first note that

$$p_Y(1) = \sum_x p(x,1) = p(0,1) + p(1,1) = 0.5.$$

Hence,

$$p_{X|Y}(0|1) = \frac{p(0,1)}{p_Y(1)} = \frac{2}{5}, \quad p_{X|Y}(1|1) = \frac{p(1,1)}{p_Y(1)} = \frac{3}{5}.$$

Example

- If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, calculate the conditional distribution of X given that $X+Y=n$.

  We calculate the conditional probability mass function of X given that $X+Y=n$ as follows:

$$\begin{aligned} P\{X=k|X+Y=n\} &= \frac{P\{X=k,X+Y=n\}}{P\{X+Y=n\}} \\ &= \frac{P\{X=k,Y=n-k\}}{P\{X+Y=n\}} \\ &= \frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}. \end{aligned}$$

  The last equality follows from the assumed independence of X and Y.

Example (Cont'd)

- Recal that $X + Y$ has a Poisson distribution with parameter $\lambda_1 + \lambda_2$:

$$\begin{aligned} P\{X = k \mid X + Y = n\} &= \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \\ &\quad \times \left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1} \\ &= \frac{n!}{(n-k)!k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} \\ &= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}. \end{aligned}$$

We conclude that the conditional distribution of X given that $X + Y = n$ is the binomial distribution with parameters n and $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Example

- Suppose n independent trials are performed, with each trial resulting in outcome i with probability p_i, $\sum_{i=1}^k p_i = 1$.

  Let the random variables X_i, $i = 1, \ldots, k$, represent, respectively, the number of trials that result in outcome i, $i = 1, \ldots, k$.

  The X_i satisfy the *multinomial distribution* with joint probability mass function

$$P\{X_i = n_i, i = 1, \ldots, k\} = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k},$$

  $n_i \geq 0$, $\sum_{i=1}^k n_i = n$.

  Suppose we are given that n_j of the trials resulted in outcome j, for $j = r+1, \ldots, k$, where $\sum_{j=r+1}^k n_j = m \leq n$.

Example (Cont'd)

- Then each of the other $n-m$ trials must have resulted in one of the outcomes $1,\ldots,r$.

  Thus, it seems that the conditional distribution of $X_1,\ldots,X_r$ is the multinomial distribution on $n-m$ trials with respective trial outcome probabilities, for $i=1,\ldots,r$,

$$P\{\text{outcome } i|\text{outcome is not any of } r+1,\ldots,k\}=\frac{p_i}{F_r},$$

  where $F_r=\sum_{i=1}^r p_i$ is the probability that a trial results in one of the outcomes $1,\ldots,r$.

Example (Cont'd)

- To verify this, let $n_1, \ldots, n_r$, be such that $\sum_{i=1}^r n_i = n - m$. Then

$$P\{X_1 = n_1, \ldots, X_r = n_r | X_{r+1} = n_{r+1}, \ldots, X_k = n_k\}$$
$$= \frac{P\{X_1 = n_1, \ldots, X_k = n_k\}}{P\{X_{r+1} = n_{r+1}, \ldots, X_k = n_k\}}$$
$$= \frac{\frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_r^{n_r} p_{r+1}^{n_{r+1}} \cdots p_k^{n_k}}{\frac{n!}{(n-m)! n_{r+1}! \cdots n_k!} F_r^{n-m} p_{r+1}^{n_{r+1}} \cdots p_k^{n_k}}.$$

  For the probability in the denominator:
  - Regard outcomes $1, \ldots, r$ as a single outcome having probability F_r;
  - Obtain the probability as a multinomial probability on n trials with outcome probabilities $F_r, p_{r+1}, \ldots, p_k$.

  Because $\sum_{i=1}^r n_i = n - m$, we get

$$P\{X_1 = n_1, \ldots, X_r = n_r | X_{r+1} = n_{r+1}, \ldots, X_k = n_k\}$$
$$= \frac{(n-m)!}{n_1! \cdots n_r!} \left(\frac{p_1}{F_r}\right)^{n_1} \cdots \left(\frac{p_r}{F_r}\right)^{n_r}.$$

Example

- Consider n independent trials, with each trial being a success with probability p.

  Given a total of k successes, show that all possible orderings of the k successes and $n-k$ failures are equally likely.

  We want to show that, given a total of k successes, each of the $\binom{n}{k}$ possible orderings of k successes and $n-k$ failures is equally likely.

  Let X denote the number of successes, and consider any ordering of k successes and $n-k$ failures, say, $\boldsymbol{o}=(s,s,\ldots,s,f,f,\ldots,f)$.

  Then

$$\begin{aligned} P(\boldsymbol{o}|X=k) &= \frac{P(\boldsymbol{o},X=k)}{P(X=k)}=\frac{P(\boldsymbol{o})}{P(X=k)} \\ &= \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}}=\frac{1}{\binom{n}{k}}. \end{aligned}$$

Subsection 5

Conditional Distributions: Continuous Case

Conditional Probability Density Function

- If X and Y have a joint probability density function $f(x,y)$, then the **conditional probability density function of X given that $Y = y$** is defined, for all values of y such that $f_Y(y) > 0$, by

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

- To motivate this definition, multiply the left-hand side by dx and the right-hand side by $\frac{dxdy}{dy}$ to obtain

$$\begin{aligned} f_{X|Y}(x|y)dx &= \frac{f(x,y)dxdy}{f_Y(y)dy} \\ &\approx \frac{P\{x \le X \le x+dx, y \le Y \le y+dy\}}{P\{y \le Y \le y+dy\}} \\ &= P\{x \le X \le x+dx | y \le Y \le y+dy\}. \end{aligned}$$

- In other words, for small values of dx and dy, $f_{X|Y}(x|y)dx$ represents the conditional probability that X is between x and $x+dx$ given that Y is between y and $y+dy$.

Conditional Cumulative Distribution Function

- The use of conditional densities allows us to define conditional probabilities of events associated with one random variable when we are given the value of a second random variable.
- That is, if X and Y are jointly continuous, then, for any set A,
$$P\{X \in A|Y = y\} = \int_A f_{X|Y}(x|y)dx.$$
- In particular, by letting $A = (-\infty, a]$, we can define the **conditional cumulative distribution function of X given that $Y = y$** by
$$F_{X|Y}(a|y) \equiv P\{X \leq a|Y = y\} = \int_{-\infty}^{a} f_{X|Y}(x|y)dx.$$
- We have been able to give workable expressions for conditional probabilities, even though the event on which we are conditioning (namely, the event $\{Y = y\}$) has probability 0.

Example

- The joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{12}{5}x(2-x-y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Compute the conditional density of X given that $Y = y$, where $0 < y < 1$.

For $0 < x < 1$, $0 < y < 1$, we have

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx} \\ &= \frac{x(2-x-y)}{\int_0^1 x(2-x-y)dx} = \frac{x(2-x-y)}{(x^2 - \frac{1}{3}x^3 - \frac{1}{2}yx^2)|_{x=0}^{x=1}} \\ &= \frac{x(2-x-y)}{\frac{2}{3} - \frac{y}{2}} = \frac{6x(2-x-y)}{4-3y}. \end{aligned}$$

Example

- Suppose that the joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y}, & 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Find $P\{X > 1 | Y = y\}$.
We first obtain the conditional density of X given that $Y = y$.

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{e^{-x/y}e^{-y}/y}{e^{-y}\int_0^\infty (1/y)e^{-x/y}dx} \\ &= \frac{e^{-x/y}e^{-y}/y}{e^{-y}(-e^{-x/y})|_{x=0}^{x=\infty}} = \frac{1}{y}e^{-x/y}. \end{aligned}$$

Hence,

$$P\{X > 1 | Y = y\} = \int_1^\infty \frac{1}{y}e^{-x/y}dx = -e^{-x/y}|_1^\infty = e^{-1/y}.$$

Independent Random Variables

- If X and Y are independent continuous random variables, the conditional density of X given that $Y = y$ is just the unconditional density of X.
- Indeed, in the independent case,

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

Mixed Conditional Distributions

- Suppose that:
  - X is a continuous random variable having probability density function f;
  - N is a discrete random variable.

  Consider the conditional distribution of X given that $N = n$.
- Then
$$\frac{P\{x<X<x+dx|N=n\}}{dx} = \frac{P\{N=n|x<X<x+dx\}}{P\{N=n\}}\frac{P\{x<X<x+dx\}}{dx}$$
- Letting dx approach 0 gives
$$\lim_{dx\to 0}\frac{P\{x<X<x+dx|N=n\}}{dx} = \frac{P\{N=n|X=x\}}{P\{N=n\}}f(x).$$
- Thus, the conditional density of X given that $N = n$ is given by
$$f_{X|N}(x|n) = \frac{P\{N=n|X=x\}}{P\{N=n\}}f(x).$$

Example

- Consider $n + m$ trials having a common probability of success.
  Suppose, however, that this success probability is not fixed in advance but is chosen from a uniform $(0, 1)$ population.
  What is the conditional distribution of the success probability given that the $n + m$ trials result in n successes?
  - Let X be the probability that a given trial is a success.
    By hypothesis, X is a uniform $(0, 1)$ random variable:
    $$f_X(x) = 1, \quad 0 < x < 1;$$
  - Given that $X = x$, the $n + m$ trials are independent with common probability of success x;
    So N, the number of successes, is a binomial random variable with parameters $(n + m, x)$:
    $$P\{N = n|X = x\} = \binom{n+m}{n} x^n (1 - x)^m.$$

Example

- Hence, the conditional density of X given that $N = n$ is

$$\begin{aligned} f_{X|N}(x|n) &= \frac{P\{N = n|X = x\}f_X(x)}{P\{N = n\}} \\ &= \frac{\binom{n+m}{n}x^n(1-x)^m}{P\{N = n\}},\ 0 < x < 1 \\ &= cx^n(1-x)^m, \end{aligned}$$

  where c does not depend on x.
- Hence, if the original or prior (to the collection of data) distribution of a trial success probability is uniformly distributed over $(0,1)$, then the posterior (or conditional) distribution given a total of n successes in $n+m$ trials is $f_{X|N}(x|n) = cx^n(1-x)^m$.

Subsection 6

Joint Probability Distributions of Functions of Random Variables

Functions of Random Variables

- Let X_1 and X_2 be jointly continuous random variables with joint probability density function f_{X_1,X_2}.
- Suppose that $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2.
- We obtain the joint distribution of the random variables Y_1 and Y_2.
- Assume that the functions g_1 and g_2 satisfy the following conditions:
  1. The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, with solutions given by, say, $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$.
  2. The functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2) and are such that, for all (x_1, x_2),

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} \equiv \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0.$$

Joint Density of Functions of Random Variables

- Under these two conditions, it can be shown that the random variables Y_1 and Y_2 are jointly continuous with joint density function given by

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(x_1,x_2)|J(x_1,x_2)|^{-1},$$

where $x_1 = h_1(y_1,y_2)$, $x_2 = h_2(y_1,y_2)$.

- A proof of this would proceed along the following lines:

$$P\{Y_1 \leq y_1, Y_2 \leq y_2\} = \iint_{\substack{(x_1,x_2): \\ g_1(x_1,x_2)\leq y_1 \\ g_2(x_1,x_2)\leq y_2}} f_{X_1,X_2}(x_1,x_2)dx_1dx_2.$$

- The joint density function can now be obtained by differentiating with respect to y_1 and y_2.
- The result of this differentiation will be equal to the right-hand side of the original equation (this is done in advanced calculus).

Example

- Let X_1 and X_2 be jointly continuous random variables with probability density function f_{X_1,X_2}.

  Let $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$.

  Find the joint density function of Y_1 and Y_2 in terms of f_{X_1,X_2}.

  Let $g_1(x_1, x_2) = x_1 + x_2$ and $g_2(x_1, x_2) = x_1 - x_2$.

  Then $J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2.$

  Moreover, the equations $y_1 = x_1 + x_2$ and $y_2 = x_1 - x_2$ have solution $x_1 = \frac{y_1+y_2}{2}$, $x_2 = \frac{y_1-y_2}{2}$.

  Thus, the desired density is

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1,X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right).$$

Example (Cont'd)

- If X_1 and X_2 are independent uniform $(0,1)$ random variables, then

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases} \frac{1}{2}, & 0\leq y_1+y_2\leq 2, 0\leq y_1-y_2\leq 2\\ 0, & \text{otherwise}\end{cases}$$

- If X_1 and X_2 are independent exponential random variables with respective parameters λ_1 and λ_2, then

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases} \frac{\lambda_1\lambda_2}{2}\exp\{-\lambda_1(\frac{y_1+y_2}{2})-\lambda_2(\frac{y_1-y_2}{2})\}, & \\ & y_1+y_2\geq 0, y_1-y_2\geq 0\\ 0, & \text{otherwise}\end{cases}$$

Example (Cont'd)

- If X_1 and X_2 are independent standard normal random variables, then

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= \frac{1}{4\pi}e^{-[(y_1+y_2)^2/8+(y_1-y_2)^2/8]} \\ &= \frac{1}{4\pi}e^{-(y_1^2+y_2^2)/4} \\ &= \frac{1}{\sqrt{4\pi}}e^{-y_1^2/4}\frac{1}{\sqrt{4\pi}}e^{-y_2^2/4}. \end{aligned}$$

  - Both X_1+X_2 and X_1-X_2 are normal with mean 0 and variance 2;
  - We also conclude that these two random variables are independent.
- It turns out that if X_1 and X_2 are independent random variables having a common distribution function F, then X_1+X_2 will be independent of X_1-X_2 if and only if F is a normal distribution function.

Example

- Let (X, Y) denote a random point in the plane, and assume that the rectangular coordinates X and Y are independent standard normal random variables.

  We find the joint distribution of R, Θ, the polar coordinate representation of (x, y).

  Suppose first that X and Y are both positive.

  Let $r = g_1(x, y) = \sqrt{x^2 + y^2}$ and $\theta = g_2(x, y) = \tan^{-1} \frac{y}{x}$.

$$\frac{\partial g_1}{\partial x} = \frac{x}{\sqrt{x^2+y^2}}, \qquad \frac{\partial g_1}{\partial y} = \frac{y}{\sqrt{x^2+y^2}},$$
$$\frac{\partial g_2}{\partial x} = \frac{1}{1+(y/x)^2}\left(\frac{-y}{x^2}\right) = \frac{-y}{x^2+y^2}, \quad \frac{\partial g_2}{\partial y} = \frac{1}{x[1+(y/x)^2]} = \frac{x}{x^2+y^2}.$$

  Hence,

$$J(x, y) = \frac{x^2}{(x^2+y^2)^{3/2}} + \frac{y^2}{(x^2+y^2)^{3/2}} = \frac{1}{\sqrt{x^2+y^2}} = \frac{1}{r}.$$

Example (Cont'd)

- By hypothesis, the conditional joint density function of X, Y given that they are both positive is given, for $x > 0, y > 0$, by

$$f(x, y|X > 0, Y > 0) = \frac{f(x, y)}{P(X > 0, Y > 0)} = \frac{2}{\pi} e^{-(x^2+y^2)/2}.$$

  Thus, the conditional joint density function of $R = \sqrt{X^2 + Y^2}$ and $\Theta = \tan^{-1}(\frac{Y}{X})$, given that X and Y are both positive, is

$$f(r, \theta|X > 0, Y > 0) = \frac{2}{\pi} r e^{-r^2/2},\ 0 < \theta < \frac{\pi}{2}, 0 < r < \infty.$$

  Similarly, we can show that

$$\begin{array}{rcl} f(r,\theta|X < 0, Y > 0) & = & \frac{2}{\pi} r e^{-r^2/2},\ \frac{\pi}{2} < \theta < \pi, 0 < r < \infty, \\ f(r,\theta|X < 0, Y < 0) & = & \frac{2}{\pi} r e^{-r^2/2},\ \pi < \theta < \frac{3\pi}{2}, 0 < r < \infty, \\ f(r,\theta|X > 0, Y < 0) & = & \frac{2}{\pi} r e^{-r^2/2},\ \frac{3\pi}{2} < \theta < 2\pi, 0 < r < \infty. \end{array}$$

Example (Cont'd)

- The joint density is an equally weighted average of these 4 conditional joint densities.

  Hence, the joint density of R, Θ is given by

  $$f(r,\theta) = \frac{1}{2\pi} r e^{-r^2/2},\ 0 < \theta < 2\pi, 0 < r < \infty.$$

- This joint density factors into the marginal densities for R and Θ.

  So R and Θ are independent random variables, with:
  - Θ uniformly distributed over $(0, 2\pi)$;
  - R having the Rayleigh distribution with density

    $$f(r) = r e^{-r^2/2},\ 0 < r < \infty.$$

The Case of n Variables

- Suppose the joint density function of the n random variables $X_1, X_2, \ldots, X_n$ is given.
- We want to compute the joint density function of $Y_1, Y_2, \ldots, Y_n$, where

$$Y_1 = g_1(X_1, \ldots, X_n), Y_2 = g_2(X_1, \ldots, X_n), \ldots, Y_n = g_n(X_1, \ldots, X_n).$$

- We assume that:
  - The functions g_i have continuous partial derivatives;
  - The Jacobian determinant

$$J(x_1, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix} \neq 0, \text{ for all } (x_1, \ldots, x_n).$$

The Case of n Variables (Cont'd)

- We also assume that the equations

$$y_1 = g_1(x_1, \ldots, x_n), y_2 = g_2(x_1, \ldots, x_n), \ldots, y_n = g_n(x_1, \ldots, x_n)$$

have a unique solution, say,

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

- Under these assumptions, the joint density function of the random variables Y_i is given by

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)|J(x_1, \ldots, x_n)|^{-1},$$

where $x_i = h_i(y_1, \ldots, y_n)$, $i = 1, 2, \ldots, n$.

Example

- Let X_1, X_2 and X_3 be independent standard normal random variables. If $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_1 - X_2$ and $Y_3 = X_1 - X_3$, compute the joint density function of Y_1, Y_2, Y_3.

  Let $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_1 - X_2$, $Y_3 = X_1 - X_3$.

  The Jacobian of these transformations is given by

$$J = \begin{vmatrix} \frac{\partial g_1}{x_1} & \frac{\partial g_1}{x_2} & \frac{\partial g_1}{x_3} \\ \frac{\partial g_2}{x_1} & \frac{\partial g_2}{x_2} & \frac{\partial g_2}{x_3} \\ \frac{\partial g_3}{x_1} & \frac{\partial g_3}{x_2} & \frac{\partial g_3}{x_3} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{vmatrix} = 3.$$

  The preceding transformations yield that

$$X_1 = \frac{Y_1 + Y_2 + Y_3}{3},\ X_2 = \frac{Y_1 - 2Y_2 + Y_3}{3},\ X_3 = \frac{Y_1 + Y_2 - 2Y_3}{3}.$$

  We conclude that

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = \frac{1}{3} f_{X_1, X_2, X_3}\left(\frac{y_1 + y_2 + y_3}{3}, \frac{y_1 - 2y_2 + y_3}{3}, \frac{y_1 + y_2 - 2y_3}{3}\right).$$

Example (Cont'd)

- Now recall that

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \frac{1}{(2\pi)^{3/2}} e^{-\sum_{i=1}^3 x_i^2/2}.$$

Set

$$\begin{aligned} Q(y_1,y_2,y_3) &= \left(\tfrac{y_1+y_2+y_3}{3}\right)^2 + \left(\tfrac{y_1-2y_2+y_3}{3}\right)^2 + \left(\tfrac{y_1+y_2-2y_3}{3}\right)^2 \\ &= \tfrac{y_1^2}{3} + \tfrac{2}{3}y_2^2 + \tfrac{2}{3}y_3^2 - \tfrac{2}{3}y_2y_3. \end{aligned}$$

Then, we get

$$f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = \frac{1}{3(2\pi)^{3/2}} e^{-Q(y_1,y_2,y_3)/2}.$$

Example

- Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed exponential random variables with rate λ.
  Let $Y_i = X_1 + \cdots + X_i$, $i = 1, \ldots, n$.
  (a) Find the joint density function of $Y_1, \ldots, Y_n$.
  (b) Use the result of Part (a) to find the density of Y_n.

(a) Consider $Y_i = g_i(X_1, \ldots, X_n) = X_1 + \cdots + X_i$, $i = 1, \ldots, n$.
The Jacobian of these transformations is

$$J = \begin{vmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{vmatrix}.$$

Thus, $J = 1$.

Example (Cont'd)

- The joint density function of $X_1, \ldots, X_n$ is given by

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}, \ 0 < x_i < \infty, \ i = 1, \ldots, n.$$

Moreover, the preceding transformations yield

$$X_1 = Y_1, \ X_2 = Y_2 - Y_1, \ \ldots, \ X_i = Y_i - Y_{i-1}, \ \ldots, \ X_n = Y_n - Y_{n-1}.$$

Thus, the joint density function of $Y_1, \ldots, Y_n$ is

$$\begin{aligned} & f_{Y_1,\ldots,Y_n}(y_1, y_2, \ldots, y_n) \\ &= f_{X_1,\ldots,X_n}(y_1, y_2 - y_1, \ldots, y_i - y_{i-1}, \ldots, y_n - y_{n-1}) \\ &= \lambda^n \exp\{-\lambda[y_1 + \textstyle\sum_{i=2}^{n}(y_i - y_{i-1})]\} \\ &= \lambda^n e^{-\lambda y_n} \quad 0 < y_1, 0 < y_i - y_{i-1}, i = 2, \ldots, n \\ &= \lambda^n e^{-\lambda y_n} \quad 0 < y_1 < y_2 < \cdots < y_n. \end{aligned}$$

Example (Cont'd)

(b) To obtain the marginal density of Y_n, let us integrate out the other variables one at a time.
Doing this gives

$$\begin{array}{rcl} f_{Y_2,\ldots,Y_n}(y_2,\ldots,y_n) & = & \int_0^{y_2} \lambda^n e^{-\lambda y_n} dy_1 \\ & = & \lambda^n y_2 e^{-\lambda y_n},\ 0 < y_2 < y_3 < \cdots < y_n; \\ f_{Y_3,\ldots,Y_n}(y_3,\ldots,y_n) & = & \int_0^{y_3} \lambda^n y_2 e^{-\lambda y_n} dy_2 \\ & = & \lambda^n \frac{y_3^2}{2} e^{-\lambda y_n},\ 0 < y_3 < y_4 < \cdots < y_n. \end{array}$$

The next integration yields

$$f_{Y_4,\ldots,Y_n}(y_4,\ldots,y_n) = \lambda^n \frac{y_4^3}{3!} e^{-\lambda y_n},\ 0 < y_4 < \cdots < y_n.$$

Continuing in this fashion gives

$$f_{Y_n}(y_n) = \lambda^n \frac{y_n^{n-1}}{(n-1)!} e^{-\lambda y_n},\ 0 < y_n.$$