Chapter 6 Jointly Distributed Random Variables

6.1 Joint distribution functions

- Joint cumulative probability distribution function of X and Y:

  $F(a, b) = P\{X \leq a,\ Y \leq b\} \quad -\infty < a, b < \infty$

- $F_X(a) = P\{X \leq a\} = \lim_{b\to\infty} F(a, b) \equiv F(a, \infty)$

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

- Marginal distribution: $F_X(a), F_Y(b)$

- All joint probability statements about X and Y can be answered in terms of their joint distribution function.

- $P\{X > a,\ Y > b\} = 1 - F_X(a) - F_Y(b) + F(a, b)$

- $P\{a_1 < X \leq a_2,\ b_1 < Y \leq b_2\} = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1)$

- Joint probability mass function of X and Y:

$$p\{x,\ y\} = P\{X = x,\ Y = y\}$$

– $p_X(x) = P\{X = x\} = \sum_y p(x, y)$

– $p_Y(y) = P\{Y = y\} = \sum_x p(x, y)$

Example 6.1a. Suppose that 3 balls are randomly selected from an urn containing 3 red, 4 white, and 5 blue balls.

- If we let X and Y denote, respectively, the number of red and white balls chosen, then the joint probability mass function of X and Y, $p(i, j) = P\{X = i, Y = j\}$, is given by

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

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

- These probabilities can most easily be expressed in tabular form as in Table 6.1
- The reader should note that the probability mass function of X is obtained by computing the row sums, whereas the probability mass function of Y is obtained by computing the column sums.
- As the individual probability mass functions

of X and Y thus appear in the margin of such a table, they are often referred to as being the marginal probability mass function of X and Y, respectively.

Table 6.1 $P\{X = i, Y = j\}$

	0	1	2	3	Row sum $= P\{X = i\}$
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	$\frac{1}{220}$
Column sum $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Example 6.1b. Suppose that 15 percent of the families in a certain community have no children, 20 percent have 1, 35 percent have 2, and 30 percent have 3; and suppose, further, that in each family, each child is equally likely to be a boy or a girl. If a family is chosen at random from this community, the B, the

number of boys, and G, the number of girls, in this family will have the joint probability mass function shown in Table 6.2.

- $$\begin{aligned} P\{B=0,G=0\} &= P\{\text{no children}\} = .15 \\ P\{B=0,G=1\} &= P\{\text{1 girl and total of 1 child}\} \\ &= P\{\text{1 child}\}P\{\text{1 girl} \mid \text{1 child}\} = (.20)\left(\frac{1}{2}\right) \\ P\{B=0,G=2\} &= P\{\text{2 girls and total of 2 children}\} \\ &= P\{\text{2 children}\}P\{\text{2 girls} \mid \text{2 children}\} = (.35)\left(\frac{1}{2}\right)^2 \end{aligned}$$

- Table 6.2 $P\{B=i, G=j\}$

	0	1	2	3	Row sum $=$ $P\{B=i\}$
0	.15	.10	.0875	.0375	.3750
1	.10	.175	.1125	0	.3875
2	.0875	.1125	0	0	.2000
3	.0375	0	0	0	.0375
Column sum $=$ $P\{G=j\}$	.375	.3875	.2000	.0375	

- Joint probability density function of X and Y: $f(x,y)$

- $P\{(X<Y) \in C\} = \int\int_{(x,y)\in C} f(x,y)dxdy$

- $f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)$

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

-

$$\begin{aligned} P\{X \in A\} &= P\{X \in A, Y \in (-\infty, \infty)\} \\ &= \int_A \int_{-\infty}^{\infty} f(x,y)dydx \\ &= \int_A f_X(x)dx \end{aligned}$$

where $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$

- $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$

Example 6.1c. 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)$P\{X < a\}$

(a)

$$P\{X > 1, Y < 1\} = \int_0^1 \int_1^{\infty} 2e^{-x}e^{-2y}dxdy$$

$$
\begin{aligned}
&= \int_0^1 2e^{-2y}(-e^{-x}|_1^{\infty})dy \\
&= e^{-1}\int_0^1 2e^{-2y}dy \\
&= e^{-1}(1-e^{-2})
\end{aligned}
$$

(b)

$$
\begin{aligned}
P\{X<Y\} &= \int\int_{(x,y):x<y} 2e^{-x}e^{-2y}dxdy \\
&= \int_0^{\infty}\int_0^{y} 2e^{-x}e^{-2y}dxdy \\
&= \int_0^{\infty} 2e^{-2y}(1-e^{-y})dy \\
&= \int_0^{\infty} 2e^{-2y}dy - \int_0^{\infty} 2e^{-3y}dy \\
&= 1-\frac{2}{3} \\
&= \frac{1}{3}
\end{aligned}
$$

- (c)

$$
\begin{aligned}
P\{X<a\} &= \int_0^{a}\int_0^{\infty} 2e^{-2y}e^{-x}dydx \\
&= \int_0^{a} e^{-x}dx \\
&= 1-e^{-a}
\end{aligned}
$$

Example 6.1d. 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. (On other words, the point is uniformly distributed within the circle.) If we let the center of the circle denote the origin and define X and Y to be the coordinates of the point chosen (Fig. 6.1), it follows, since (X, Y) is equally likely to be near each point in the circle, that the joint density function of X and Y is given by

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

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, id less than or equal to a.

(d) Find $E[D]$.

(a) Because $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1$

$c \int\int_{x^2+y^2 \le R^2} dxdy = 1$

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

(b)

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

$$\begin{aligned} f_Y(y) &= \frac{2}{\pi R^2}\sqrt{R^2 - y^2} \quad y^2 \le R^2 \\ &= 0 \quad y^2 > R^2 \end{aligned}$$

(c) $D = \sqrt{X^2 + Y^2}$, for $0 \le a \le R$

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

$$= \frac{\pi a^2}{\pi R^2}$$
$$= \frac{a^2}{R^2}$$

(d) From (c) we obtain that the density function of D is $F_D(a) = \frac{2a}{R^2} \qquad 0 \le a \le R$

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

Example 6.1e. 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.

- For $a > 0$,

$$\begin{aligned} F_{X/Y}(a) &= P\left\{\frac{X}{Y} \le a\right\} \\ &= \int\int_{x/y \le a} e^{-(x+y)} dxdy \\ &= \int_0^\infty \int_0^{ay} e^{-(x+y)} dxdy \end{aligned}$$

$$= \int_0^\infty (1 - e^{-ay})e^{-y}dy$$
$$= \left[-e^{-y} + \frac{e^{-(a+1)y}}{a+1}\right]\Bigg|_0^\infty$$
$$= 1 - \frac{1}{a-1}$$

- The density function:

$$f(x,y) = 1/(a+1)^2, \qquad 0 < a < \infty$$

n random variables:

- Joint cumulative probability distribution function:

$$F(a_1, a_2, \ldots, a_n) = P\{X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n\}$$

- Joint probability density function:

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(x_1, x_2, \ldots, x_n)$$

- $P\{(X_1, X_2, \ldots, X_n) \in C\} = \int\int \cdots \int_{(x_1,x_2,\ldots,x_n)\in C} f(x_1, x_2, \ldots, x_n)dx_1dx_2 \cdots dx_n$

- $P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \int_{A_n}\int_{A_{n-1}} \cdots \int_{A_1} f(x_1, x_2, \ldots, x_n)dx_1dx_2 \cdots dx_n$

Example 6.1f.
The multinomial distribution. One of the

most important joint distribution is the multinomial, 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 $p_1, p_2, \ldots, p_r$, $\sum_{i=1}^{r} p_i = 1$. If we denote by X_i, the number of the n experiments that result in outcome number i, then

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

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

- Suppose that a fair die is rolled 9 times.
- The probability that 1 appears three times, 2 and 3 twice each, 4 and 5 once each, and 6 not at all is

$$\frac{9!}{3!2!2!1!1!0!}\left(\frac{1}{6}\right)^9$$

6.2 Independent random variables

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

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

- Equivalent condition of independence:

$$F(a, b) = F_X(a)F_Y(b)$$

  – When X and Y are discrete random variables:

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

  – In the jointly continuous case:

$$f(x, y) = f_X(x)f_Y(y) \quad \text{for all } x, y$$

- Random variables that are not independent are said to be dependent.

Example 6.2a. Suppose that $n + m$ independent trials, having a common success probability p, are performed. If X is the number of successes in the first n trials, and Y is 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,

$$\begin{aligned} P\{X = x, Y = y\} &= \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y} \\ & \quad 0 \le x \le n, 0 \le y \le m \\ &= P\{X = x\} P\{Y = y\} \end{aligned}$$

On the other hand, X and Z will be dependent, where Z is the total number of successes in the $n + m$ trials. (Why is this?)

Example 6.2b. Suppose that the number of people that enter a post office on a given day is a Poisson random variable with parameter λ. Show that if each person that enters the post office is a male with probability p and a female with probability $1-p$, then the number of males and females entering the post office are independent Poisson random variables with respective parameters λp and $\lambda(1-p)$.

- Condition on $X + Y$:

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

- Since $P\{X = i, Y = j | X + Y \neq i + j\} = 0$

$$P\{X = i, Y = j\} = P\{X = i, Y = j | X+Y = i+j\} P\{X+Y = i+j\} \tag{2.3}$$

-

$$P\{X + Y = i + j\} = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \tag{2.4}$$

-

$$P\{X = i, Y = j | X + Y = i + j\} = \binom{i+j}{i} p^i (1-p)^j \tag{2.5}$$

-

$$\begin{aligned} P\{X = i, Y = j\} &= \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\ &= e^{-\lambda} \frac{(\lambda p)^i}{i! j!} [\lambda(1-p)]^j \\ &= \frac{e^{-\lambda p} (\lambda p)^i}{i!} e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^i}{j!} \end{aligned} \tag{2.6}$$

- Hence

$$P\{X = i\} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} \sum_j e^{-\lambda(1-p)} \frac{[\lambda(1-p)^j]}{j!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} \tag{2.7}$$

-

$$P\{Y = j\} = e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!} \tag{2.8}$$

Example 6.2c. A man and a woman decide to meet at a certain location. If each person 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.

- X and Y: The time past 12 that the man and woman arrive.
- X and Y are independent random variables, each of which is uniform$(0, 60)$.
- The desired probability $P\{X + 10 < Y\} +$ $P\{Y + 10 < X\}$.
- By symmetry

$$P\{X + 10 < Y\} = P\{Y + 10 < X\}$$

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

$$= \frac{2}{(60)^2} \int_{10}^{60} (y - 10)\, dy$$
$$= \frac{25}{36}$$

Example 6.2d. **Buffon's needle problem.** A table is ruled with equidistant parallel lines a distance D apart. A needle of length L, where $L \le D$, 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)?

- X: The distance from the middle point of the needle to the nearest parallel line.
- θ: The angle between the needle and the projected line of length X (Fig. 6.2).
- The needle will intersect a line if the hypotenuse of the right triangle in Fig. 6.2 is

less than $L/2$, i.e.

$$\frac{X}{\cos\theta} < \frac{L}{2} \quad \text{or} \quad X < \frac{L}{2}\cos\theta$$

- $X \sim$ uniform$(0, D/2)$; $\theta \sim$ uniform$(0, \pi/2)$
-

$$\begin{aligned} P\left\{X < \frac{L}{2}\cos\theta\right\} &= \int\int_{x<L/2\cos y} f_X(x) f_\theta(y)\, dxdy \\ &= \frac{4}{\pi D}\int_0^{\pi/2} \frac{L}{2}\cos y\, dy \\ &= \frac{2L}{\pi D} \end{aligned}$$

***Example 6.2e. Characterization of the normal distribution.** Let X and Y denote the horizontal and vertical miss distance when a bullet is fired at a target, and 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 $x^2 + y^2$.

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

$$\begin{aligned}
f(x,y) &= f_X(x) f_Y(y) = g(x^2+y^2) \\
f'_X(x) f_Y(y) &= 2x g'(x^2+y^2) \\
\frac{f'_X(x)}{f_X(x)} &= \frac{2x g'(x^2+y^2)}{g(x^2+y^2)} \\
\frac{f'_X(x)}{2x f_X(x)} &= \frac{g'(x^2+y^2)}{g(x^2+y^2)}
\end{aligned}$$

- Consider $x_1^2 + y_1^2 = x_2^2 + y_2^2$, then

$$\begin{aligned}
\frac{f'_X(x)}{x f_X(x)} &= c \\
\frac{d}{dx}(\log f_X(x)) &= cx \\
f_X(x) &= k e^{cx^2/2} \\
f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}
\end{aligned}$$

Proposition 2.1: The continuous (discrete) random variables X and Y are independent if and only if their joint probability density (mass) function can be expressed as

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

Example 6.2f. If the first instance, the joint density function of X and Y is

$$f(x,y) = 6e^{-2x}e^{-3y} \quad 0 < x, y < \infty$$

and is equal to 0 outside this region, are the random variables independent? What if the joint density function is

$$f(x,y) = 24xy \quad 0 < x, y < 1, 0 < x + y < 1$$

and is equal to 0 otherwise?

- $$I(x,y) = \begin{cases} 1 & \text{if } 0 < x, y < 1, 0 < x + y < 1 \\ 0 & \text{otherwise} \end{cases}$$
- $f(x,y) = 24xyI(x,y)$

- They are dependent since the above cannot factor into a part depending only on x and another depending only on y.

Example 6.2g. **How can a computer choose a random subset?** Most computers are able to generate the value of, or simulate, a uniform (0,1) random variable by means of a built-in subroutine that (to a high degree of approximation) produces such "random numbers." As a result, it is quite easy for the computer to simulate an indicator (that is, a Bernoulli) random variable.

- Suppose I is an indicator variable such that

$$P\{I = 1\} = p = 1 - P\{I = 0\}$$

- The computer can simulate I by choosing a uniform(0,1) random number U and then letting

$$I = \begin{cases} 1 \text{ if } U < p \\ 0 \text{ if } U \geq p \end{cases}$$

- Suppose that we are interested in having the computer select k of the numbers, $1, 2, \ldots, n$ in such a way that each of the $\binom{n}{k}$ subsets of size k is equally likely to be chosen.
- Simulate n indicator variables $I_1, \ldots, I_n$, of which exactly k will equal 1.
- Those i for which $I_i = 1$ will then constitute the desired subset.
- I_k: If k is selected.
- Simulate n independent uniform$(0, 1)$ random variables $U_1, U_2, \ldots, U_n$.

$$
\begin{aligned}
I_1 &= \begin{cases} 1 \ \text{if } U_1 < k/n \\ 0 \ \text{otherwise} \end{cases} \\
I_{i+1} &= \begin{cases} 1 \ \text{if } U_{i+1} < \frac{k-(I_1+\cdots+I_i)}{n-i} \\ 0 \ \text{otherwise} \end{cases}
\end{aligned}
$$

- At the $i+1$ stage we set I_{i+1} equal to 1 with a probability equal to the remaining number of places in the subset divided by the remaining number of possibilities.

$$P\{I_1 = 1\} = \frac{k}{n}$$

$$P\{I_{i+1} = 1 \mid I_1, \ldots, I_i\} = \frac{k - \Sigma_{j=1}^{i} I_j}{n-i}$$

- Induction on $k+n$.
- It is easy to see that $k+n=2$ is true.
- Suppose that $i_1 < i_2 < \cdots < i_k$ such that $I_{i_1} = \cdots = I_{i_k} = 1$.
- Case $i_1 = 1$:

$$\begin{aligned} & P\{I_1 = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise}\} \\ = \; & P\{I_1 = 1\} P\{I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise} | I_1 = 1\} \\ = \; & \frac{k}{n} \frac{1}{\binom{n-1}{k-1}} = \frac{1}{\binom{n}{k}} \end{aligned}$$

- Case $i_1 \neq 1$:

$$\begin{aligned} & P\{I_1 = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise}\} \\ = \; & P\{I_1 = 0\} P\{I_{i_1} = I_{i_2} = \cdots = I_{i_k} = 1, I_j = 0 \text{ otherwise} | I_1 = 0\} \\ = \; & \left(1 - \frac{k}{n}\right) \frac{1}{\binom{n-1}{k}} = \frac{1}{\binom{n}{k}} \end{aligned}$$

Remark.

- The foregoing method for generating a random subset has a very low memory requirement.
- A faster algorithm that requires somewhat more memory is presented in Sec. 10.1. It uses the last k elements of a random permutation of $(1, 2, \ldots, n)$.

Example 6.2h. Let X, Y, Z be independent and uniformly distributed over (0,1). Compute $P\{X \geq YZ\}$.

- $f_{X,Y,Z}(x, y, z) = f_X(x) f_Y(y) f_Z(z) = 1 \quad 0 \leq x, y, z \leq 1$
-

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

Example 6.2i. Probability interpretation of half-life. Let $N(t)$ denote the number of nuclei contained in a radioactive mass of material at time t. The concept of half-life is often defined in a deterministic fashion by stating that it is an empirical fact that for some value h, called the half-life,

$$N(t) = 2^{-t/h} N(0) \quad t > 0$$

[Note that $N(h) = N(0)/2$.]

- Since the above implies that for any nonnegative s and t,

  $$N(t+s) = 2^{-(s+t)/h} N(0) = 2^{-t/h} N(s)$$

  it follows that no matter how much time s has already elapsed, in an additional time t the number of existing nuclei will decrease by the factor $2^{-t/h}$.

Probabilistic interpretation of the half-life h:

- The deterministic relationship given above results from observations of radioactive masses containing huge numbers of nuclei.

- We can assume that the individual nuclei act independently and with a memoryless life distribution.
- The unique life distribution which is memoryless is the exponential distribution.
- The lifetimes of the individual nuclei are independent random variables having a life distribution that is exponential with median equal to h.
- L: The lifetime of a given necleus.

$$P\{L < t\} = 1 - 2^{-t/h} = 1 - \exp\left\{-t\frac{\log 2}{h}\right\}$$

- Protons decay with a half-life of about $h = 10^{30}$ years.
- The number of decays predicted by the deterministic model:

$$N(0) - N(c) = h(1 - 2^{-c/h})$$

$$
\begin{aligned}
&= \frac{1-2^{-c/h}}{1/h} \\
&\approx \lim_{x\to 0} \frac{1-2^{-cx}}{x} \\
&= c\log 2 \approx .6931c
\end{aligned}
$$

- Since there is a huge number of independent protons, each of which will have a very small probability of decaying within this time period, it follows that the number of protons that decay will have a Poisson distribution with parameter equal to $h(1-2^{-c/h}) \approx c\log 2$.

$$
\begin{aligned}
P\{0 \text{ decays}\} &= e^{-c\log 2} = \frac{1}{2^c} \\
P\{n \text{ decays}\} &= \frac{2^{-c}[c\log 2]^n}{n!}
\end{aligned}
$$

Remark. Independence is a symmetric relation.

Example 6.2j. If the initial throw of the dice in the game of craps results in the sum of the dice equaling 3, then the player will continue to throw the dice until the sum is either 3 or 7. If this sum is 3, then the player wins, and if it is 7, then the player loses.

- Let N denote the number of throws needed until either 3 or 7 appears, and let X denote the value (either 3 or 7) of the final throw.
- Is N independent of X? That is, does knowing which of 3 or 7 occurs first affect the distribution of the number of throws needed until that number appears?
- Most people do not find the answer to this question to be intuitively obvious.
- However, suppose that we turn it around and ask whether X is independent of N. That is, dose knowing how many throws it takes to obtain a sum of either 3 or 7.
- Does this affect the probability that that sum is equal to 3?

- For instance, suppose we know that it takes n throws of the dice to obtain a sum either 3 or 7.
- Does this affect the probability distribution of the final sum?
- Clearly not, since all that is important is that its values is either 3 or 7, and the fact that none of the first $n-1$ throws were either 3 or 7 does not change the probabilities for the nth throw.
- Thus we can conclude that X is independent of N, or equivalently, that N is independent of X.
- Another example: Record value problem
  - $X_1, X_2, \ldots, X_n$ are iid 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*.

– A_n: The event that X_n is a record value.
– $P(A_n|A_{n+1}) = P(A_n) = \frac{1}{n}$
– Then A_n and A_{n+1} are independent.

6.3 Sum of independent random variables

- Suppose that X and Y are independent continuous random variables with density functions f_X and f_Y.
- CDF of $X + Y$:

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

- PDF of $X + Y$:

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

Example 6.3a. *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 $X + Y$.

- $$f_X(a) = f_Y(a) = \begin{cases} 1 & 0 < a < 1 \\ 0 & \text{otherwise} \end{cases}$$
- $f_{X+Y}(a) = \int_0^a f_X(a-y)\, dy$
- 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.

Density function of Gamma(t, λ)

$$f(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{t-1}}{\Gamma(t)} \quad 0 < y < \infty$$

Proposition 3.1: If X and Y are independent gamma random variables with respective parameters (s, λ) and (t, λ), then $X + Y$ is a gamma random variable with parameters $(s + t, \lambda)$.

If X_i's are independent gamma(t_i, λ), then

$$\sum_{i=1}^{n} X_i \sim \text{gamma}\left(\sum_{i=1}^{n} t_i, \lambda\right)$$

Example 6.3b. Let $X_1, X_2, \ldots, X_n$ be n independent exponential random variables each having parameter λ. Then, as an exponential random variable with parameter λ is the same as a gamma random variable with parameters (1,λ), we see from Proposition 3.1 that $X_1, X_2, \ldots, X_n$ is a gamma random variable with parameters (n, λ).

Chi-squared distribution:

- If $Z_1, Z_2, \ldots, Z_n$ are independent unit normal random variables, then $Y \equiv \sum_{i=1}^{n} Z_i^2$ is said to have the chi-squared distribution with n degrees of freedom.
- If $n = 1$, then

$$
\begin{aligned}
f_{Z^2}(y) &= \frac{1}{2\sqrt{y}}[f_Z(\sqrt{y}) + f_Z(-\sqrt{y})] \\
&= \frac{1}{2\sqrt{y}}\frac{2}{\sqrt{2\pi}}e^{-y/2} \\
&= \frac{e^{-y/2}(y/2)^{1/2-1}(1/2)}{\sqrt{\pi}}
\end{aligned}
$$

  Y is gamma$(1/2, 1/2)$.

- Thus for any n, Y is gamma$(n/2, 1/2)$ and

$$f_Y(y) = \frac{e^{-y/2}(y/2)^{n/2-1}(1/2)}{\Gamma(n/2)}$$

- When n is an even integer,

$$\Gamma(n/2) = [(n/2) - 1]!$$

- When n is an odd integer,

$$\Gamma(n/2) = [(n/2) - 1] \cdots (1/2)\sqrt{\pi}$$

1. The chi-squared distribution often arises in practice as being the distribution of the square of the error involved when one attempts to hit a target in n-dimensional space when the coordinate errors are taken to be independent unit normal random variables.
2. It is also important in statistical analysis.

Proposition 3.2: If $X_i, i = 1, \ldots, n$ are independent random variables that are normally distributed with respective parameters $\mu_i, \sigma_i^2, 1, \ldots, n$, then $\sum_{i=1}^{n} X_i$ is normally distributed with parameters $\sum_{i=1}^{n} \mu_i$ and $\sum_{i=1}^{n} \sigma_i^2$.

- Assume $X \sim N(0, \sigma^2)$ and $Y \sim N(0, 1)$ are independent.
- Show that $X + Y \sim N(0, 1 + \sigma^2)$.

- $X_i \sim N(\mu_i, \sigma_i^2)$
- $X_1 + X_2 = \sigma_2 \left(\frac{X_1-\mu_1}{\sigma_1} + \frac{X_2-\mu_2}{\sigma_2}\right) + \mu_1 + \mu_2$
- $\frac{X_1-\mu_1}{\sigma_1} \sim N(0, \sigma_1^2/\sigma_2^2)$ and $\frac{X_2-\mu_2}{\sigma_2} \sim N(0,1)$
- Then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Example 6.3c. A club basketball team will play a 44-game season. Twenty-six of these games are against class A teams and 18 are against class B teams. Suppose that the team will win each game against a class A team with probability .4, and will win each game against a class B team with probability .7. Assume also that the results of the different games are independent. Approximate the probability that

(a) the team wins 25 games or more;

(b) the team wins more games against class A teams than it does against class B teams.

- (a)

- X_A, X_B: The number of games the team wins against class A and against class B.
- X_A and X_B: Independent binomial random variables.
- $E[X_A] = 26(.4) = 10.4 \quad \mathrm{Var}(X_A) = 26(.4)(.6) = 6.24$
- $E[X_A] = 18(.7) = 12.6 \quad \mathrm{Var}(X_B) = 18(.7)(.3) = 3.78$
- Normal approximation gives that X_A and X_B are approximately independent normal random variables.
-

$$\begin{aligned} P\{X_A + X_B \geq 25\} &= P\{X_A + X_B \geq 24.5\} \\ &= P\left\{\frac{X_A + X_B - 23}{\sqrt{10.02}} \geq \frac{24.5 - 23}{\sqrt{10.02}}\right\} \\ &\approx P\left\{Z \geq \frac{1.5}{\sqrt{10.02}}\right\} \\ &\approx 1 - P\{Z < .4739\} \\ &\approx .3178 \end{aligned}$$

- (b)

$$\begin{aligned} P\{X_A - X_B \geq 1\} &= P\{X_A - X_B \geq .5\} \\ &= P\left\{\frac{X_A - X_B + 2.2}{\sqrt{10.02}} \geq \frac{.5 + 2.2}{\sqrt{10.02}}\right\} \end{aligned}$$

$$
\begin{aligned}
&\approx P\left\{Z \geq \frac{2.7}{\sqrt{10.02}}\right\} \\
&\approx 1 - P\{Z < .8530\} \\
&\approx .1968
\end{aligned}
$$

Example 6.3d. *Sums of independent Poisson random variables.* If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, compute the distribution of $X + Y$.

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

Example 6.3e. *Sums of independent binomial random variables.* Let X and Y be independent binomial random variables with respective parameters (n, p) and (m, p). Calculate the distribution of $X + Y$.

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

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

- Hence

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

6.4 Conditional distributions: discrete case

- The conditional probability of E given F:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

- The conditional probability mass function of X given $Y = y$:

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

- The conditional probability distribution function of X given $Y = y$:

$$\begin{aligned} F_{X|Y}(x|y) &= \frac{P\{X \leq x \mid Y = y\}}{P\{Y = y\}} \\ &= \sum_{a \leq x} p_{X|Y}(a|y) \end{aligned}$$

- If X and Y are independent, then

$$p_{X|Y}(x|y) = P\{X = x\}$$

Example 6.4a. Suppose that $p(x,y)$, the joint probability mass function of X and Y, is given by

$$p(0,0) = .4 \quad p(0,1) = .2 \quad p(1,0) = .1 \quad p(1,1) = .3$$

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

- $p_Y(1) = \sum_x p(x,1) = p(0,1) + p(1,1) = .5$
- $P_{X|Y}(0|1) = \frac{p(0,1)}{p_Y(1)} = \frac{2}{5}$
- $P_{X|Y}(1|1) = \frac{p(1,1)}{p_Y(1)} = \frac{3}{5}$

Example 6.4b. 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 $X + Y = n$.

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

- $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$
-
$$P\{X = k | X + Y = n\} = \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1}$$

$$
\begin{aligned}
&= \frac{n!}{(n-k)!k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}
\end{aligned}
$$

6.5 Conditional distributions: continuous case

- Conditional probability density function:
$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$
- $P\{X \in A \mid Y = y\} = \int_A f_{X|Y}(x|y)dx$
- $F_{X|Y}(a|y) = P\{X \le a \mid Y = y\} = \int_{-\infty}^{a} f_{X|Y}(x|y)\,dx$

Example 6.5a. The joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{15}{2}x(2-x-y) & 0 < x, y < 1 \\ 0 & \text{otherwise} \end{cases}$$

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

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

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

Example 6.5b. Suppose that the joint density of X and Y is given by

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

Find $P\{X > 1|Y = y\}$.

- The conditional density of X, given that

$Y = y$

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

- Hence

$$
\begin{aligned}
P\{X > 1|Y = y\} &= \int_1^\infty \frac{1}{y}e^{-x/y}\,dx \\
&= -e^{-x/y}\Big|_1^\infty \\
&= e^{-1/y}
\end{aligned}
$$

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

Suppose that X is a continuous random variable having density function f and N is a discrete random variable.

The conditional density of X given that $N = n$:

$$\frac{P\{x < X < x + dx|N = n\}}{dx} = \frac{P\{N = n|x < X < x + dx\}}{P\{N = n\}} \frac{P\{x < X < x + dx\}}{dx}$$
$$\lim_{dx \to 0} \frac{P\{x < X < x + dx|N = n\}}{dx} = \frac{P\{N = n|X = x\}}{P\{N = n\}} f(x)$$

$$f_{X|N}(x|n) = \frac{P\{N = n|X = x\}}{P\{N = n\}} f(x)$$

Example 6.5c. 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 $n + m$ trails result in n successes?

- $X \sim$ uniform$(0, 1)$: The trial success probability.
- $N \sim$ binomial$(n + m, x)$: The number of success.

- The conditional density of X given that $N = n$: Beta$(n+1, m+1)$

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

- The conditional density is that of a beta random variable with parameters $n+1, m+1$.

*6.6 Order statistics

- $X_1, X_2, \ldots, X_n$ are n independent and identically distributed, continuous random variables having a common density f and distribution function F.

$$
\begin{aligned}
X_{(1)} &= \text{smallest of } X_1, X_2, \ldots, X_n \\
X_{(2)} &= \text{second smallest of } X_1, X_2, \ldots, X_n \\
&\vdots \\
X_{(j)} &= j\text{th smallest of } X_1, X_2, \ldots, X_n
\end{aligned}
$$

$$\vdots$$

$$X_{(n)} = \text{largest of } X_1, X_2, \ldots, X_n$$

- **Order statistics:** $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$
- The order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ will take on the values $x_1 \le x_2 \le \cdots \le x_n$ if and only if for some permutation $(i_1, i_2, \ldots, i_n)$ of $(1, 2, \ldots, n)$

$$X_1 = x_{i_1}, X_2 = x_{i_2}, \ldots, X_n = x_{i_n}$$

$$P\left\{x_{i_1} - \frac{\epsilon}{2} < X_1 < x_{i_1} + \frac{\epsilon}{2}, \ldots, x_{i_n} - \frac{\epsilon}{2} < X_n < x_{i_n} + \frac{\epsilon}{2}\right\}$$
$$\approx \epsilon^n f_{X_1, X_2, \ldots, X_n}(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$$
$$= \epsilon^n f(x_1) \cdots f(x_n)$$
$$P\left\{x_1 - \frac{\epsilon}{2} < X_1 < x_1 + \frac{\epsilon}{2}, \ldots, x_n - \frac{\epsilon}{2} < X_n < x_n + \frac{\epsilon}{2}\right\}$$
$$\approx n!\epsilon^n f(x_1) \cdots f(x_n)$$

- Joint density function of order statistics:

$$f_{X_{(1)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n) \quad x_1 < \cdots < x_n$$

Example 6.6a. Along a road 1 mile long are 3 people "distributed at random." Find the probability that no 2 people are less than a distance of d miles apart, when $d \leq \frac{1}{2}$.

- X_i's are independent uniform$(0,1)$.
- $f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3) = 3! \quad 0 < x_1 < x_2 < x_3 < 1$
- If X_i denotes the position of the ith person, the desired probability is

$$P\{X_{(i)} > X_{(i-1)} + d, i = 2,3\}$$

$$\begin{aligned} P\{X_{(i)} > X_{(i-1)} + d, i = 2,3\} &= \int\int\int_{x_i > x_{i-1}+d \quad i=2,3} f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3)dx_1dx_2dx \\ &= 3! \int_0^{1-2d}\int_{x_1+d}^{1-d}\int_{x_2+d}^{1} dx_3dx_2dx_1 \\ &= 6\int_0^{1-2d}\int_{x_1+d}^{1-d}(1-d-x_2)dx_2dx_1 \\ &= 6\int_0^{1-2d}\int_0^{1-2d-x_1} y_2dy_2dx_1 \end{aligned}$$

where $y_2 = 1-d-x_2$.

- Hence

$$\begin{aligned} &= 3\int_0^{1-2d}(1-2d-x_1)^2dx_1 \\ &= 3\int_0^{1-2d} y_1^2dy_1 \\ &= (1-2d)^3 \end{aligned}$$

- The same method can be used to prove that when there are n people distributed at random over the unit interval the desired probability is

$$[1-(n-1)d]^n \quad \text{when } d \le \frac{1}{n-1}$$

The density function of $X_{(j)}$:

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!(j-1)!}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)$$

$$\binom{n}{j-1, n-j, 1} = \frac{n!}{(n-j)!(j-1)!}$$

Example 6.6b. When a sample of $2n+1$ random variables (that is, when $2n+1$ independent and identically distribute random variables) are observed, the $(n+1)$st smallest is called the **sample median**. If a sample of size 3 from a uniform distribution over (0,1) is observed, find the probability that the sample median is between $\frac{1}{4}$ and $\frac{3}{4}$.

- $f_{X_{(2)}}(x) = \frac{3!}{1!1!}x(1-x) \quad 0 < x < 1$
- Hence

$$P\left\{\frac{1}{4} < X_{(2)} < \frac{3}{4}\right\} = 6\int_{1/4}^{3/4} x(1-x)dx$$

$$= 6\left\{\frac{x^2}{2} - \frac{x^3}{3}\right\}\Bigg|_{x=1/4}^{x=3/4} = \frac{11}{16}$$

$$\begin{aligned}
F_{X_{(j)}}(y) &= \int_{-\infty}^{y} \frac{n!}{(n-j)!(j-1)!}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)dx \\
F_{X_{(j)}}(y) &= P\{j \text{ or more of } X_i\text{'s are } \le y\} \\
&= \sum_{k=j}^{n} \binom{n}{k} F^k(y)[1-F(y)]^{n-k} \\
f_{X_{(i)},X_{(j)}}(x_i, x_j) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!(j-1)!} \times \\
&\quad [F(x_i)]^{i-1}[F(x_j)-F(x_i)]^{j-i-1}[1-F(x_j)]^{n-j}f(x_i)f(x_j)
\end{aligned}$$

Example 6.6c. Distribution of the range of a random sample. Suppose that n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ are observed.

- $R = X_{(n)} - X_{(1)}$: The **range** of the observed random variables.

- If the random variables X_i have distribution function F and density function f, then the distribution of R can be obtained from Eq. (6.6) as follows: $a \geq 0$.

$$\begin{aligned} P\{R \leq a\} &= P\{X_{(n)} - X_{(1)} \leq n\} \\ &= \int\int_{x_n - x_1 \leq a} f_{X_{(1)},X_{(n)}}(x_1, x_n) dx_1 dx_n \\ &= \int_{-\infty}^{\infty} \int_{x_1}^{x_1+a} \frac{n!}{(n-2)!} [F(x_n) - F(x_1)]^{n-2} f(x_1) f(x_n) dx_n dx_1 \end{aligned}$$

- Making the change of variable $y = F(x_n) - F(x_1)$,$dy = f(x_n)dx_n$, yields

$$\begin{aligned} \int_{x_1}^{x_1+a} [F(x_n) - F(x_1)]^{n-2} f(x_n) dx_n &= \int_0^{F(x_1+a)-F(x_1)} y^{n-2} dy \\ &= \frac{1}{n-1} [F(x_1 + a) - F(x_1)]^{n-1} \end{aligned}$$

and thus

$$P\{R \leq a\} = n \int_{-\infty}^{\infty} [F(x_1 + a) - F(x_1)]^{n-1} f(x_1) dx_1 \tag{6.7}$$

- When the X_i's are all uniformly distributed on $(0, 1)$:

$$\begin{aligned} P\{R < a\} &= n \int_0^1 [F(x_1 + a) - F(x_1)]^{n-1} f(x_1) dx_1 \\ &= n \int_0^{1-a} a^{n-1} dx_1 + n \int_{1-a}^{1} (1 - x_1)^{n-1} dx_1 \\ &= n(1 - a) a^{n-1} + a^n \end{aligned}$$

- The density function of the range:

$$f_R(a) = \begin{cases} n(n-1)a^{n-2}(1-a) & 0 \le a \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- The range of n independent uniform$(0,1)$ random variables is a Beta$(n-1, 2)$.

6.7 Joint probability distribution of functions of random variables

- The joint probability density function f_{X_1,X_2}.
- $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$.
- Assume that g_1 and g_2 satisfy the following condition:
  1. The equation $y_1 = g_1(x_1, x_2)$ 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 $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
  2. The functions g_1 and g_2 have continuous partial derivatives at all points (x_1, x_2)

and are such that the following 2×2 determinant

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

at all points (x_1, x_2).

The joint density function of Y_1 and Y_2:

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

Example 6.7a. 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}.

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

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

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

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

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

- If X_1 and X_2 are independent $\exp(\lambda_1), \exp(\lambda_2)$, then $f_{Y_1,Y_2}(y_1,y_2)$

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

- If X_1 and X_2 are independent unit normal random variables, then Y_1 and Y_2 are independent $N(0,2)$.

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

Example 6.7b. Let (X,Y) denote a random point in the plane and assume that the rectangular coordinates X and Y are independent unit normal random variables. We are interested in the joint distribution of R, θ, the polar coordinate representation of this point (See Fig. 6.4).

- $r = g_1(x, y) = \sqrt{x^2 + y^2}$ and $\theta = g_2(x, y) = \tan^{-1} y/x$, $0 < r < \infty, 0 < \theta < 2\pi$.
- $\frac{\partial g_1}{\partial x} = \frac{x}{\sqrt{x^2+y^2}}$ and $\frac{\partial g_1}{\partial y} = \frac{y}{\sqrt{x^2+y^2}}$.
- $\frac{\partial g_2}{\partial x} = \frac{-y}{x^2+y^2}$ and $\frac{\partial g_2}{\partial y} = \frac{x}{x^2+y^2}$.
- $J(x, y) = \frac{1}{\sqrt{x^2+y^2}} = \frac{1}{r}$
- $f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$
- $f(r, \theta) = \frac{1}{2\pi} r e^{-r^2/2}$
- R and Θ are independent.
- R is **Rayleigh distribution**. Θ is uniform$(0, 2\pi)$.
- The joint distribution of R^2 and Θ:
  - $d = g_1 = x^2 + y^2$ and $\theta = g_2(x, y) = \tan^{-1} y/x$, $0 < d < \infty, 0 < \theta < 2\pi$.
  - $J = 2$ and $f(d, \theta) = \frac{1}{2} e^{-d/2} \frac{1}{2\pi}$.

- R^2 and Θ are independent with R^2 having an exponential distribution with parameter $1/2$.

- The above result can be used to simulate normal random variables by making a suitable transformation on uniform random variables.
  - U_1 and U_2 are independent uniform$(0, 1)$.
  - $R^2 \equiv -2 \log U_1$ is an exponential distribution with parameter $1/2$.
  - $\Theta \equiv 2\pi U_2$ is a uniform$(0, 2\pi)$.
  - $X_1 = R \cos \Theta = \sqrt{-2 \log U_1} \cos(2\pi U_2)$
  - $X_2 = R \sin \Theta = \sqrt{-2 \log U_1} \sin(2\pi U_2)$

Example 6.7c. If X and Y are independent gamma random variables with parameters (α, λ) and (β, λ), respectively, compute the joint density of $U = X + Y$ and $V = X/(X + Y)$.

- The joint density of X and Y is given by

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\lambda e^{-\lambda x}(\lambda x)^{(\alpha-1)}}{\Gamma(\alpha)}\frac{\lambda e^{-\lambda y}(\lambda y)^{(\beta-1)}}{\Gamma(\beta)} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}e^{-\lambda(x+y)}x^{\alpha-1}y^{\beta-1} \end{aligned}$$

- $u = g_1(x,y) = x+y, v = g_2(x,y) = x/(x+y)$, then

$$\frac{\partial g_1}{\partial x} = \frac{\partial g_1}{\partial y} = 1 \quad \frac{\partial g_2}{\partial x} = \frac{y}{(x+y)^2} \quad \frac{\partial g_2}{\partial y} = -\frac{x}{(x+y)^2}$$

- $$J(x,y) = \begin{vmatrix} 1 & 1 \\ \frac{y}{(x+y)^2} & \frac{-x}{(x+y)^2} \end{vmatrix} = -\frac{1}{x+y}$$

- $x = uv$, and $y = u(1-v)$

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}[uv, u(1-v)]u \\ &= \frac{\lambda e^{-\lambda u}(\lambda u)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)}\frac{v^{\alpha-1}(1-v)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \end{aligned}$$

- U and V are independent gamma$(\alpha+\beta)$ and beta(α,β).
- Suppose that there are $n+m$ jobs to be performed, with each taking an exponential

amount of time with rate λ for performance, and suppose that we have two workers to perform these jobs.

- Worker I will do jobs $1, \ldots, n$, and worker II will do the remaining m jobs.
- If we let X and Y denote the total working times of workers I and II, respectively, then X and Y will be independent gamma(n, λ) and gamma(m, λ).
- Then the above result yields that independently of the working time needed to complete all $n + m$ jobs, the proportion of this work that will be performed by worker I has a beta(n, m).

The joint density function of the n random variables $X_1, X_2, \ldots, X_n$:

- $Y_i = g_i(X_1, X_2, \ldots, X_n), i = 1, 2, \ldots, n$

$$J = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ & & \vdots & \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

$y_i = g_i(x_1, x_2, \ldots, x_n), i = 1, 2, \ldots, n$

- $f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)|J|^{-1}$ where $x_i = h_i(y_1, y_2, \ldots, y_n), i = 1, 2, \ldots, n$

Example 6.7d. Let X_1, X_2 and X_3 be independent unit normal random variables. If $Y_1 = X_1 + X_2 + X_3, Y_2 = X_1 - X_2, Y_3 = X_1 - X_3$, compute the joint density function of Y_1, Y_2, Y_3.

- $$J = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{vmatrix} = 3$$

- $X_1 = \frac{Y_1+Y_2+Y_3}{3}$

- $X_2 = \frac{Y_1 - 2Y_2 + Y_3}{3}$
- $X_3 = \frac{Y_1 + Y_2 - 2Y_3}{3}$
- $$f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) = \frac{1}{3} f_{X_1,X_2,X_3}\left(\frac{y_1 + y_2 + y_3}{3}, \frac{y_1 - 2y_2 + y_3}{3}, \frac{y_1 + y_2 - 2y_3}{3}\right)$$
- $f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{1}{(2\pi)^{3/2}} e^{-\Sigma_{i=1}^{3} x_i^2/2}$
- $f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) = \frac{1}{3(2\pi)^{3/2}} e^{-Q(y_1,y_2,y_3)/2}$

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

Example 6.7e. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed exponential random variables with rate λ. Let

$$Y_i = X_1 + \cdots + X_i \quad i = 1, \ldots, n$$

(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) $Y_1 = X_1, Y_2 = X_1 + X_2, \ldots, Y_n = X_1 + \cdots + X_n$

$$J(x_1, \ldots, x_n) = \begin{vmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & & & & \\ \cdots & \cdots & & & & \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{vmatrix}$$

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

- (b)
$$\begin{aligned} f_{Y_2,\ldots,Y_n}(y_2, \ldots, y_n) &= \int_0^{y_2} \lambda^n e^{-\lambda y_n} dy_1 \\ &= \lambda^n y_2 e^{-\lambda y_n} \quad 0 < y_2 < y_3 < \cdots < y_n \end{aligned}$$

- $$\begin{aligned} f_{Y_3,\ldots,Y_n}(y_3, \ldots, y_n) &= \int_0^{y_3} \lambda^n y_2 e^{-\lambda y_n} dy_2 \\ &= \lambda^n \frac{y_3^2}{2} e^{-\lambda y_n} \quad 0 < y_3 < y_4 < \cdots < y_n \end{aligned}$$

- $$f_{Y_4,\ldots,Y_n}(y_4, \ldots, y_n) = \lambda^n \frac{y_4^2}{3!} e^{-\lambda y_n} \quad 0 < y_4 < \cdots < y_n$$

- Y_n is gamma(n, λ)

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

*6.8 Exchangeable random variables

- The random variables $X_1, \ldots, X_n$ are said to exchangeable if for every permutation $i_1, \ldots, i_n$ of the integers $1, \ldots, n$

$$P\{X_{i_1} \le x_1, \ldots, X_{i_n} \le x_n\} = P\{X_1 \le x_1, \ldots, X_n \le x_n\}$$

- Discrete random variables will be exchangeable if

$$P\{X_{i_1} = x_1, \ldots, X_{i_n} = x_n\} = P\{X_1 = x_1, \ldots, X_n = x_n\}$$

Example 6.8a. Suppose that balls are withdrawn one at a time and without replacement from an urn that initially contains n balls, of which k are considered special, in such a manner that each withdrawal is equally likely to be any of the balls that remain in the urn at the time. Let $X_i = 1$ if the ith ball withdrawn is a special and let it be 0 otherwise. We will show that the random variables $X_1, \ldots, X_n$ are exchangeable.

- Let $(x_1, \ldots, x_n)$ be a vector consisting of k ones and $n-k$ zeros.
- However, before considering the joint mass function evaluated at $(x_1, \ldots, x_n)$, let us try to gain some insight by considering a fixed such vector-for instance, consider the vector $(1, 1, 0, 1, 0, \ldots, 0, 1)$, which is assumed to have k ones and $n-k$ zeros.
- Then
$$p(1, 1, 0, 1, 0, \ldots, 0, 1) = \frac{k}{n}\frac{k-1}{n-1}\frac{n-k}{n-2}\frac{k-2}{n-3}\frac{n-k-1}{n-4}\cdots\frac{1}{2}\frac{1}{1}$$
which follows since the probability that the first ball is special is k/n, the conditional probability that the next one is special is $(k-1)/(n-1)$, the conditional probability that the next one is not special is $(n-k)/(n-2)$, and so on.
- By the same argument, it follows that

$p(x_1, \ldots, x_n)$ can be expressed as the product of n fractions.

- The successive denominator terms of these fractions will go from n down to 1.
- The numerator term at the location where the vector $(x_1, \ldots, x_n)$ is 1 for the ith time is $k - (i - 1)$, and where it is 0 for the ith time it is $n - k - (i - 1)$.
- Hence, since the vector $(x_1, \ldots, x_n)$ consists of k ones and $n - k$ ones and $n - k$ zeros, we obtain that

$$p(x_1, \ldots, x_n) = \frac{k!(n-k)!}{n!} \quad x_i = 0, 1, \sum_{i=1}^{n} x_i = k$$

- Since this is a symmetric function of $(x_1, \ldots, x_n)$, it follows that the random variables are exchangeable.

If $X_1, X_2, \ldots, X_n$ are exchangeable, it easily follows that each X_i has the same probability distribution. If X and Y are exchangeable discrete random variables, then

$$P\{X = x\} = \sum_y P\{X = x, Y = y\} = \sum_y P\{X = y, Y = x\} = P\{Y = x\}$$

Example 6.8b. In Example 6.8a, let Y_1 denote the selection number of the first special ball withdrawn, let Y_2 denote the additional number that are then withdrawn until the second special ball appears, and in general, let Y_i denote the additional number of balls withdrawn after the $(i-1)$st special ball is selected until the ith is selected, $i = 1, \ldots, k$.

- For instance, if $n = 4, k = 2$ and $X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1$ then $Y_1 = 1, Y_2 = 3$.
- Since $Y_1 = i_1, Y_2 = i_2, \ldots, Y_k = i_k \Leftrightarrow X_{i_1} = X_{i_1+i_2} = \cdots = X_{i_1+\cdots+i_k} = 1, X_j = 0$, otherwise; we obtain from the joint mass function of the X_i that

$$P\{Y_1 = i_1, Y_2 = i_2, \ldots, Y_k = i_k\} = \frac{k!(n-k)!}{n!} \quad i_1 + \cdots + i_k \leq$$

- Hence we see that the random variables $Y_1, \ldots, Y_k$ are exchangeable.
- For instance, it follows from this that the number of cards one must select from a well-shuffled deck until an ace appears has the same distribution as the number of additional cards one must select after the first ace appears until the next one does, and so on.

Example 6.8c. The following is known as Polya's urn model. Suppose that an urn initially contains n red and m blue balls. At each stage a ball is randomly chosen, its color is noted, and it is then replaced along with another ball of the same color. Let $X_i = 1$ if the ith ball selected is red and let it equal 0 of the ith ball is blue, $i \geq 1$. To obtain a feeling for the joint probabilities of these X_i, note the following special cases.

- $P\{X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0\}$
$$= \frac{n}{n+m}\frac{n+1}{n+m+1}\frac{m}{n+m+2}\frac{n+2}{n+m+3}\frac{m+1}{n+m+4}$$
$$= \frac{n(n+1)(n+2)m(m+1)}{(n+m)(n+m+1)(n+m+2)(n+m+3)(n+m+4)}$$

- $P\{X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1\}$
$$= \frac{m}{n+m}\frac{n}{n+m+1}\frac{m+1}{n+m+2}\frac{n+1}{n+m+3}\frac{n+2}{n+m+4}$$
$$= \frac{n(n+1)(n+2)m(m+1)}{(n+m)(n+m+1)(n+m+2)(n+m+3)(n+m+4)}$$

- $P\{X_1 = x_1, \ldots, X_k = x_k\}$
$$\frac{n(n+1)\cdots(n+r-1)m(m+1)\cdots(m+k-r-1)}{(n+m)\cdots(n+m+k-1)}$$

- The random variables $X_1, \ldots, X_k$ are exchangeable.

Example 6.8d. Let $X_1, X_2, \ldots, X_n$ be independent uniform (0,1) random variables, and let $X_{(1)}, \ldots, X_{(n)}$ denote their order statistics. That is, $X_{(j)}$ is the jth smallest of $X_1, X_2, \ldots, X_n$. Also, let

$$\begin{aligned} Y_1 &= X_{(1)}, \\ Y_i &= X_{(i)} - X_{(i-1)}, \quad i = 2, \ldots, n \end{aligned}$$

Show that $Y_1, \ldots, Y_n$ are exchangeable.

- $y_1 = x_1, \ldots, y_i = x_i - x_{i-1} \quad i = 2, \ldots, n$
- $x_i = y_1 + \cdots + y_i \quad i = 1, \ldots, n$
- $f_{Y_1,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = f(y_1, y_1+y_2, \ldots, y_1+\cdots+y_n)$
- $f_{Y_1,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = n!$
  $0 < y_1 < y_1 + y_2 < \cdots < y_1 + \cdots + y_n < 1$
- $f_{Y_1,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = n!$
  $0 < y_i < 1, i = 1, \ldots, n \quad y_1 + \cdots + y_n < 1$

Summary

- Joint cumulative probability distribution function:

$$F(x, y) = P\{X \le x, Y \le y\}$$

  – $F_X(x) = \lim_{y\to\infty} F(x, y)$
  – $F_Y(y) = \lim_{x\to\infty} F(x, y)$

- Joint probability mass function:

$$p(i, j) = P\{X = i, Y = j\}$$

  – $P\{X = i\} = \sum_j p(i, j)$

- $P\{Y = j\} = \sum_i p(i, j)$

- Joint probability density function: $f(x, y)$
  - $P\{(X, Y) \in C\} = \int\int_C f(x, y)\, dxdy$
  - $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$
  - $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$

- Independence

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

$$P\{X_1 \in A_1, \ldots, X_n \in A_n\} = P\{X_1 \in A_1\} \cdots P\{X_n \in A_n\}$$

- The distribution function of $X + Y$:

$$F_{X+Y}(a) = \int_{-\infty}^{\infty} F(a - y) f_Y(y)\, dy$$

- If X_i's are independent $N(\mu_i, \sigma_i^2)$, then

$$\sum_i^n X_i \sim N\left(\sum_i^n \mu_i, \sum_i^n \sigma_i^2\right)$$

- If X_i's are independent Poisson(λ_i), then

$$\sum_i^n X_i \sim \text{Poisson}\left(\sum_i^n \lambda_i\right)$$

- If X_i's are independent gamma(α_i, β), then

$$\sum_i^n X_i \sim \text{gamma}\left(\sum_i^n \alpha_i, \beta\right)$$

- If X_i's are independent binomial(n_i, p), then

$$\sum_i^n X_i \sim \text{binomial}(\sum_i^n n_i, p)$$

- The conditional probability mass function of X given that $Y = y$:

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

- The conditional probability density function of X given that $Y = y$:

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

- The density function of order statistic:

$$f(x_1, \ldots, x_n) = n! f(x_1) \cdots f(x_n) \quad x_1 \leq x_2 \leq \cdots \leq x_n$$

- The random variables $X_1, \ldots, X_n$ are exchangeable if the joint distribution of $X_{i_1}, \ldots, X_{i_n}$ is the same for every permutation $i_1, \ldots, i_n$ of $1, \ldots, n$.