Multivariate Distributions - Basic Facts

Often we are interested in more than one random variable. Examples:

- On a given day, let

$$\begin{aligned} X &= \text{temperature in Toronto at noon} \\ Y &= \text{temperature in Waterloo at noon.} \end{aligned}$$

- $X_1, ..., X_{150}$ = heights of 150 people selected from a given population.
- $S_1, ..., S_{500}$ = closing prices of stocks in the S&P500 index[1].

[1] stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ.

- In statistical applications measurements of many different quantities, or repeated measurements of the same quantity, can be framed as observations of multiple random variables.
- What are the goals of probability theory for more than one random variable?
  - Computing probabilities related to their "joint" behavior.
  - Computing summary quantities describing the random variables.
  - Determining distributional properties of transformations of the r.v's, like their sums or averages.
- To deal with such questions, we will look at extensions of the definitions of
  - probability function/probability density function
  - expected value and variance.

Definition. Suppose that X and Y are *discrete* random variables defined on the same sample space[2].

The **joint probability function** of X and Y is defined as

$$\begin{aligned} f(x,y) &= P(s \in S : X(s) = x \text{ and } Y(s) = y) \\ &= P(X = x, Y = y), \end{aligned}$$

where $x \in range(X), y \in range(Y)$.

Definition. For a collection of n discrete random variables $X_1, ..., X_n$, the joint probability function is defined as

$$f(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n),$$

where $x_1 \in range(X_1), \ldots, x_n \in range(X_n)$.

[2] When we consider two or more random variables, we typically assume they are defined on the same sample space.

MULTIVARIATE DISTRIBUTIONS (CH9)
Basic Facts
Marginal Distributions
Independent r.v.'s
Conditional p.f.
Functions of Random Variables
Multinomial Distribution

Example. Suppose a fair coin is flipped twice. Let X denote the number of heads in the first toss, and let Y denote the total number of heads. Compute the joint probability function of X and Y.

- Any joint p.f. must satisfy:
  1. $f(x, y) \geq 0$
  2. $\sum_{x,y} f(x, y) = 1$.

- Once we know a joint p.f., we can compute probability of any event $A \subseteq range(X, Y)$ using

$$\boxed{P((X, Y) \in A) = \sum_{(x,y) \in A} f(x, y).}$$

For example,

$$P(X > Y) = \sum_{(x,y): x > y} f(x, y).$$

Example. Suppose that X and Y have the following joint p.f.:

	f(x,y)	x = 0	x = 1	x = 2
y	0	.2	.3	.1
	2	.25	.13	.02

Find $P(X + Y \geq 3)$.

Clicker Question(s).

Marginal Distributions

Definition. Suppose that X and Y are *discrete* random variables with joint probability function $f(x,y)$. The **marginal probability function** of X is

$$f_X(x) \equiv f_1(x) := P(X = x) = \sum_{y \in range(Y)} f(x,y).$$

Similarly, the marginal p.f. of Y is

$$f_Y(y) \equiv f_2(y) := P(Y = y) = \sum_{x \in range(X)} f(x,y).$$

Interpretation: the marginal distribution of X is simply the probability function of X when we ignore Y!

Example. Suppose that X and Y have a joint p.f. defined by:

	f(x,y)	x = 0	x = 1	x = 2
y	0	.2	.3	.1
	2	.25	.13	.02

1. Compute the marginal probability functions of X and Y.
2. Compute $E(X)$.

MULTIVARIATE DISTRIBUTIONS (CH9)
Basic Facts
Marginal Distributions
Independent r.v.'s
Conditional p.f.
Functions of Random Variables
Multinomial Distribution

Exercise. Suppose X and Y have joint probability function

$$f(x,y)=\frac{1}{6}\left(\frac{1}{2}\right)^x\left(\frac{2}{3}\right)^y, \quad x,y=0,1,2...$$

1. Compute the marginal probability functions $f_X(x)$ and $f_Y(y)$.
2. Compute $P(X \leq Y)$.

Clicker Question(s).

Independent r.v.'s

Definition. Suppose that X and Y are discrete r.v's with the joint probability function $f(x, y)$ and marginal probability functions $f_X(x)$ and $f_Y(y)$.
X and Y are said to be **independent** random variables iff

$$f(x, y) = f_X(x)f_Y(y) \text{ for all } x \in range(X), y \in range(Y).$$

- The definition of independence simply requires that

  $$P(X = x, Y = y) = P(X = x)P(Y = y)$$

  for all possible x and y, or, equivalently, that the events

  $$\{s : X(s) = x\} \text{ and } \{s : Y(s) = y\}$$

  are independent for all possible x and y.

- Similarly to the definition of independent events, the definition of independent r.v's can be used in two "directions":

  "$\Rightarrow$" If we know that X and Y are independent, then we can immediately write their joint p.f. if we know p.f's of X and Y:

  $$f(x, y) = f_X(x)f_Y(y).$$

  "$\Leftarrow$" If we know the joint p.f. $f(x, y)$ of X and Y, then we can verify whether or not these variables are independent by checking if the equation

  $$f(x, y) = f_X(x)f_Y(y).$$

  holds for all possible x and y.

Example. Suppose that X and Y have a joint probability function defined in the table below:

	f(x,y)	x = 0	x = 1	x = 2
y	0	.2	.3	.1
	2	.25	.13	.02

Are X and Y independent random variables? Why?

- It follows from the definition that if X and Y are independent then for all $x \in range(X)$, $y \in range(Y)$ the events

$$\{X = x\} \quad \text{and} \quad \{Y = y\}$$

are independent.

More generally: if X and Y are independent, then for all subsets $A_x \subseteq range(X)$ and $A_y \subseteq range(Y)$ we have

$$P(X \in A_x \text{ and } Y \in A_y) = P(X \in A_x)P(Y \in A_y).$$

Thus, the events

$$\{s : X(s) \in A_x\} \ \text{ and } \ \{s : Y(s) \in A_y\}$$

are independent.

- One way of checking whether or nor variables are independent is to look at their range.

  If the range of X depends on a value Y, then X and Y cannot be independent.

  This holds, for example, when X and Y must satisfy certain constraints, like

$$X + Y = 10$$

or

$$X + Y > 0.$$

Example. Suppose X and Y have joint probability function

$$f(x, y) = \frac{1}{6}\left(\frac{1}{2}\right)^{x}\left(\frac{2}{3}\right)^{y}, \quad x, y = 0, 1, 2...$$

Are X and Y independent?

Clicker Question(s).

- Independence of more than two random variables.

  Definition. If $X_1, ..., X_n$ have joint probability function $f(x_1, ..., x_n)$, and marginal probability functions $f_{X_1}(x_1), ..., f_{X_n}(x_n)$, then $X_1, ..., X_n$ are said to be independent if and only if

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

  for all $x_1, x_2, \ldots, x_n$.

MULTIVARIATE DISTRIBUTIONS (CH9)
Basic Facts
Marginal Distributions
Independent r.v.'s
Conditional p.f.
Functions of Random Variables
Multinomial Distribution

Conditional p.f.

The **conditional probability function** of X given $Y = y$ is defined as

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

provided $f_Y(y) > 0$.

Similarly, $f_Y(y|x)$ is defined as

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

provided $f_X(x) > 0$.

- Note that

> any conditional probability function is a probability function.

- Therefore, conditional probability functions must satisfy the usual conditions p.f's satisfy, namely, for any fixed x:

  (i) $f_Y(y|x) \geq 0$ for any y.

  (ii) $\sum_y f_Y(y|x) = 1$.

- We can work with a conditional p.f. like with any other probability function.

  For example, we can represent the probability that Y belongs to a subset A knowing that $X = x$ as

$$P(Y \in A|X = x) = \sum_{y \in A} f_Y(y|x).$$

- If we have

$$f_X(x|y) = f_X(x), \quad \text{for each } x \text{ and } y,$$

then

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

which implies that X and Y are independent.

Similarly, if

$$f_Y(y|x) = f_Y(y), \quad \text{for each } x \text{ and } y,$$

then X and Y must be independent.

Example. Suppose that X and Y have a joint probability function defined in the table below:

		x		
	f(x,y)	0	1	2
y	0	.2	.3	.1
	2	.25	.13	.02

1. Tabulate the conditional probability function of X given $Y = 0$.
2. Find the probability that $X \neq 0$ knowing that $Y = 0$.

Clicker Question(s).

Exercise. Whenever Nam is a duty Don at Village 1, he is woken up by two types of duty phone calls: emergency calls, and non-emergency calls. Emergency calls arrive according to Poisson distribution with $\lambda = 1$ per 6 hours. Non-emergency also arrive according to Poisson distribution with $\lambda = 3$ per 6 hours, independently of emergency calls.

- What is the probability that Nam gets 2 emergency calls and 2 non-emergency calls over 6 hours of sleep.
- Calculate the probability that Nam receives 2 emergency calls given that he received a total of 3 call in 6 hours.

Functions of Random Variables

- Suppose that two random variables X and Y have a joint p.f. $f(x, y)$. Let us define a new random variable U as
$$U = h(X, Y)$$
where h is a given function $h : \mathbb{R}^2 \to \mathbb{R}$.
Question: what is the p.f. of U?

- The answer follows directly from the definition of a joint p.f.:
$$\begin{aligned} f_U(t) = P(U = t) &= P(h(X, Y) = t) \\ &= \sum_{(x,y):\ h(x,y)=t} f(x, y). \end{aligned}$$

Example. Suppose X and Y have joint probability function given by the following table:

		x		
	f(x,y)	0	1	2
y	0	.2	.3	.1
	2	.25	.13	.02

Let $U = X + Y$. Compute the probability function of U.

Exercise. The same as above but for $U = X - Y$ or $U = X \cdot Y$.

Clicker Question(s).

Theorem. Suppose X and Y are independent, and that $X \sim Poi(\mu_1)$ and $Y \sim Poi(\mu_2)$. Then

$$X + Y \sim Poi(\mu_1 + \mu_2).$$

Since X and Y are independent, we have

$$f(x,y) = \frac{\mu_1^x e^{-\mu_1}}{x!} \cdot \frac{\mu_2^y e^{-\mu_2}}{y!}, \quad x, y = 0, 1, 2, \ldots.$$

Thus

$$\begin{aligned} P(T = t) &= P(X + Y = t) = \sum_{x,y:x+y=t} P(X = x, Y = y) \\ &= \sum_{x=0}^{t} P(X = x, Y = t - x) = \sum_{x=0}^{t} f(x, t - x) \\ &= \sum_{x=0}^{t} \frac{\mu_1^x e^{-\mu_1}}{x!} \cdot \frac{\mu_2^{t-x} e^{-\mu_2}}{(t-x)!} \end{aligned}$$

$$
\begin{aligned}
\sum_{x=0}^{t} \frac{\mu_1^x e^{-\mu_1}}{x!} \cdot \frac{\mu_2^{t-x} e^{-\mu_2}}{(t-x)!} &= \mu_2^t e^{-(\mu_1+\mu_2)} \sum_{x=0}^{t} \frac{1}{x!(t-x)!} \left(\frac{\mu_1}{\mu_2}\right)^x \\
&= \frac{\mu_2^t e^{-(\mu_1+\mu_2)}}{t!} \sum_{x=0}^{t} \binom{t}{x} \left(\frac{\mu_1}{\mu_2}\right)^x \\
&= \frac{\mu_2^t e^{-(\mu_1+\mu_2)}}{t!} \left(1+\frac{\mu_1}{\mu_2}\right)^t \\
&= \frac{\mu_2^t e^{-(\mu_1+\mu_2)}}{t!} \left(\frac{\mu_1+\mu_2}{\mu_2}\right)^t \\
&= \frac{(\mu_1+\mu_2)^t}{t!} e^{-(\mu_1+\mu_2)}, \quad t = 0, 1, 2, \ldots.
\end{aligned}
$$

Theorem. If $X \sim Binomial(n, p)$ and $Y \sim Binomial(m, p)$, and X and Y are independent, then[3]

$$X + Y \sim Binomial(n + m, p).$$

Clicker Question(s).

[3] See Problem 9.1.2 for a similar result for the NB distribution.

MULTIVARIATE DISTRIBUTIONS (CH9)
Basic Facts
Marginal Distributions
Independent r.v.'s
Conditional p.f.
Functions of Random Variables
Multinomial Distribution

Multinomial Distribution

Physical Setup. Consider the following experiment:

(1) We have n independent and identical trials, where each trial results in one of k distinct types of outcome. The probabilities of the individual outcomes are

$$p_i, \; 1 \leq i \leq k, \;\; \text{with} \;\; p_1 + p_2 + \cdots + p_k = 1.$$

(2) Let

$$\begin{aligned} X_1 &= \text{the number of times the outcome of type 1 occurs,} \\ X_2 &= \text{the number of times the outcome of type 2 occurs,} \\ &\vdots \\ X_k &= \text{the number of times the outcome of type k occurs.} \end{aligned}$$

Then $(X_1, ..., X_k)$ has a **Multinomial Distribution** with parameters n and $p_1, ..., p_k$.

- We shall use the abbreviation $(X_1, ..., X_k) \sim \mathit{Multinomial}(n, p_1, ..., p_k)$.
- This model is a generalization of the binomial distribution, where $k = 2$.
- Note that

$$X_1 + X_2 + \cdots X_k = n$$

so one variable is redundant (and we can work with only $k - 1$ variables).
- If $(X_1, ..., X_k) \sim \mathit{Multinomial}(n, p_1, ..., p_k)$, then their joint probability function is

$$f(x_1, ..., x_k) = \frac{n!}{x_1!x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

where $x_1, ..., x_k$ satisfy $x_1 + \cdots + x_k = n$, $x_i \geq 0$. The fact that $\sum f(x_1, ..., x_k) = 1$ follows from the Multinomial Theorem.

- Justification of the form of the joint p.f.:
  - the event

    $$X_1 = x_1,\ X_2 = x_2,\ \ldots, X_k = x_k$$

    occurs if and only if in the n trials we have exactly x_1 outcomes of type 1, x_2 outcomes of type 2, ..., x_k outcomes of type k.

    Probability of any such sequence of outcomes is

    $$p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

  - the number of arrangements of n outcomes with exactly x_1 outcomes of type 1, x_2 outcomes of type 2, ..., x_k outcomes of type k is

    $$\frac{n!}{x_1!x_2! \cdot x_k!}.$$

Example. Consider drawing 5 cards from a standard 52 card deck of playing cards (4 suits, 13 kinds) **with replacement**.

What is the probability that 2 of the drawn cards are hearts, 2 are spades, and 1 is a diamond.

Clicker Question(s).

Example. In Roulette, a small ball is spun around a wheel in such a way so that the probability it lands in a black or red box is 18/38 each, and the probability it lands in a green box is 2/38.

Suppose 10 games are independently played, and let B, R and G denote the number of times the ball landed on black, red, and green, respectively.

(i) Compute the probability that $R = 4$, $B = 4$ and $G = 2$.

(ii) Find the probability that the ball will land in red at least 5 times.

(iii) Find the conditional probability function[4] $f(r|b)$

(iv) Are R and B independent?

(v) Find the probability function of $T = R + B$.

(vi) What is the joint probability function of the number of times the ball lands in red and the number of times the ball lands in black?

[4] A better notation would be $f_{R|B}(r|b)$ or $f_{R|B=b}(r)$, but we will follow the convention in the Notes

Answers:

(i) Since

$$(R, B, G) \sim Multinomial(10, p_1 = \frac{18}{38}, p_2 = \frac{18}{38}, p_3 = \frac{2}{38})$$

the answer is

$$\begin{aligned} P(R = 4, B = 4, G = 2) &= f(4, 4, 2) \\ &= \frac{10!}{4!4!2!}(\frac{18}{38})^4(\frac{18}{38})^4(\frac{2}{38})^2. \end{aligned}$$

where

$$f(r, b, g) = \frac{10!}{r!b!g!} p_1^r p_2^b p_3^g.$$

(ii) We need the marginal p.f. f_R of R.

Consider the following approach ("general reasoning"):

- for each game possible outcomes:

  "s" - if the ball lands in red
  "f" - if the ball lands either in black or in green

- if we define X = number of "s", then we have

$$X \sim Bin(n = 10, p = p_1).$$

Since R counts the number of successes in this experiment,

$$R \sim Bin(n = 10, p = p_1).$$

Thus, the answer is $\sum_{i=5}^{10} \binom{10}{i} \left(\frac{18}{38}\right)^i \left(\frac{20}{38}\right)^{10-i}$.

A similar approach can be used to show

$$B \sim Bin(n = 10, p = p_2) \text{ and } G \sim Bin(n = 10, p = p_3).$$

(iii) To calculate $f(r|b) = \frac{P(R=r,B=b)}{P(B=b)}$, we can use

$$P(R=r, B=b) = P(R=r, B=b, G=10-r-b) = f(r, b, 10-r-b) \quad (1)$$

and

$$P(B=b) = \binom{10}{b} p_2^b (1-p_2)^{10-b}, \quad (2)$$

where the last result follows from $B \sim Bin(10, p_2)$.

By dividing (1) by (2), we get for $r = 0, 1, \ldots, 10-b$:

$$\begin{aligned}
\frac{\frac{10!}{r!b!(10-r-b)!} p_1^r p_2^b p_3^{10-b-r}}{\frac{10!}{b!(10-b)!} p_2^b (1-p_2)^{10-b}} &= \frac{(10-b)!}{r!(10-b-r)!} \frac{p_1^r p_3^{10-b-r}}{(1-p_2)^{10-b}} \\
&= \binom{10-b}{r} \frac{p_1^r p_3^{10-b-r}}{(1-p_2)^r (1-p_2)^{10-b-r}} \\
&= \binom{10-b}{r} \left(\frac{p_1}{1-p_2}\right)^r \cdot \left(\frac{p_3}{1-p_2}\right)^{10-b-r}.
\end{aligned}$$

This shows that if we know that $B = b$, then

$$R \sim Bin(10 - b, \frac{p_1}{1 - p_2}).$$

A similar method can be used to show that conditionally on $B = b$ we have

$$G \sim Bin(10 - b, \frac{p_3}{1 - p_2}).$$

We can also find the p.f. of R given $B = b$ by using "general reasoning":

- if we know that $B = b$, then the outcomes of only $10 - b$ rolls are uncertain, as they can be either red or green.
- since the rolls are independent, we can describe the number of "red" outcomes using the binomial distribution. For this we need the probability of "success".

$$\begin{aligned} &P(\text{ball lands in red}|\text{ball lands in red or in green}) \\ &\qquad = \frac{P(\text{ball lands in red AND ball lands in red or in green})}{P(\text{ball lands in red or in green})} \\ &\qquad = \frac{P(\text{ball lands in red})}{P(\text{ball lands in red or in green})} \\ &\qquad = \frac{p_1}{p_1 + p_3} = \frac{p_1}{1 - p_2}. \end{aligned}$$

Thus, if we know that $B = b$, then

$$R \sim Bin(10 - b, \frac{p_1}{1 - p_2}).$$

In the above reasoning we use the information that in 10 rolls we get red r times.
However we do not know on which rolls exactly the ball lands in the red.

Can you explain why we do not need this information?

(iv) R and B are not independent since

$$R + B + G = 10.$$

Another approach is to verify that

$$f(r|B = b) \neq f_R(r) \quad (\text{or } f(r, b) \neq f_R(r) f_B(b)).$$

(v) To find the p.f. of $T = R + B$ we can use "general reasoning"[5]:

- in 10 independent rolls define "success" and "failure" as

  "s" - if the ball lands in red **or** black

  "f" - if the ball lands in green

  Note that $T = R + B$ counts the number of successes in the above 10 experiments.

- since the probability of success is $p_1 + p_2$, we have

$$T \sim Bin(10, p_1 + p_2).$$

[5] see the Notes for an approach where we find the p.f. of T

(vi) Since

$$P(R = r, B = b) = P(R = r, B = b, G = 10 - r - b),$$

the joint pdf of R and B is

$$\begin{aligned} f_{(R,B)}(r, b) &= f(r, b, 10 - r - b) \\ &= \frac{10!}{r!b!(10 - r - b)!} \left(\frac{18}{38}\right)^r \left(\frac{18}{38}\right)^b \left(\frac{2}{38}\right)^{10-r-b}. \end{aligned}$$

Properties of the multinomial distribution

Suppose $(X_1, ..., X_k) \sim Mult(n, p_1, ..., p_k)$. Then

(i) For each i, $i = 1, 2, \ldots, k$, we have

$$X_i \sim Bin(n, p_i).$$

(ii) $X_i = n - \sum_{j \neq i} X_j$, so $X_1, X_2, \ldots, X_k$ are dependent.

(iii) For each $i, j = 1, 2, \ldots, k$, $i \neq j$, we have

$$X_i | X_j = m \sim Bin\left(n - m, \frac{p_i}{1 - p_j}\right)$$

(iv) For any i, j, $i \neq j$, we have[6]

$$(X_i, X_j, n - X_i - X_j) \sim Mult(n, p_i, p_j, 1 - p_i - p_j).$$

[6] see Problem 9.2.1

Clicker Question(s).