Lecture 7: Multivariate Distributions

MATH 697

Omidali Jazi

September 25, 2018

McGill University

Goals for this Chapter

- Bivariate and multivariate probability distributions.
  - Bivariate normal distribution
  - Multinomial distribution
- Marginal and conditional distributions
- Independence, covariance and correlation
- Expected value & variance of a function of random variables.
  - Linear functions of random variables in particular
- Conditional expectation and variance

Bivariate and Multivariate Distributions

- A bivariate distribution is a probability distribution on two random variables.
  – I.e., it gives the probability on the simultaneous outcome of the random variables
- For example, we may want to know the probability that in the simultaneous roll of two dice each comes up 1
- A multivariate distribution is a probability distribution for more than two random variables.

Joint Probabilities

- Bivariate and multivariate distributions are joint probabilities
  – the probability that two or more events occur
  – It is the probability of the intersection of $n \geq 2$ events: $\{Y_1 = y_1\}, \{Y_2 = y_2\}, ..., \{Y_n = y_n\}$
  – We denote the joint (discrete) probabilities as $P(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n)$

- We sometimes use the shorthand notation $p(y_1, y_2, ..., y_n)$ which is the probability that the event $Y_1 = y_1$ and the event $Y_2 = y_2$, ..., and the event $Y_n = y_n$ all occurred simultaneously.

Bivariate and Multivariate Distributions

- Simple example of a bivariate distribution arises when rolling two dice:
  – Let Y_1 be the number on the first die.
  – Let Y_2 be the number on the second die.

- Then there are 36 possible joint outcomes (remember the mn rule: $6 \times 6 = 36$)
  – The outcomes (y_1, y_2) are $(1,1), (1,2), (1,3), ..., (6,6)$

- Assuming the dice are independent, all the outcomes are equally likely, so

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) = 1/36$$

  for $y_1 = 1, ..., 6$ and $y_2 = 1, ..., 6$.

Bivariate PMF for the Example

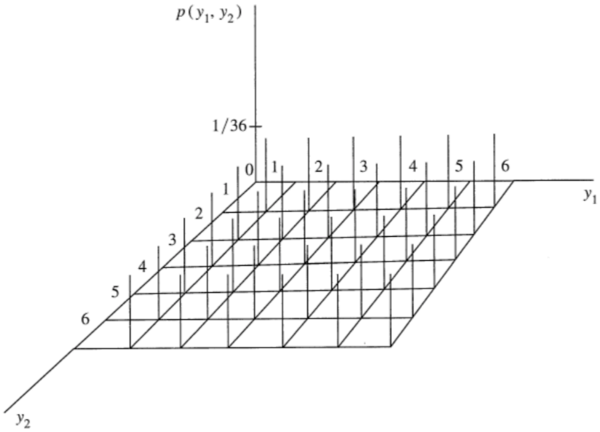


FIGURE **5.1**
Bivariate probability function; y_1 = number of dots on die 1, y_2 = number of dots on die 2

Defining Joint Probability Functions (Discrete Random Variables)

Definition 1

Let Y_1 and Y_2 be discrete random variables. The joint or bivariate pmf for (Y_1, Y_2) is $P(Y_1 = y_1, Y_2 = y_2)$ where $-\infty < y_1 < \infty$ and $-\infty < y_2 < \infty$.

Joint Probability Functions (Discrete Random Variables)

Theorem 2

If Y_1 and Y_2 are joint discrete random variables with joint probability function $p(y_1, y_2)$, then

1. $p(y_1, y_2) \geq 0$ *for all* (y_1, y_2).
2. $\sum_{y_1} \sum_{y_2} p(y_1, y_2) = 1$ *where the sum is over all non-zero* $p(y_1, y_2)$.

Joint Distribution Functions

Definition 3

For any random variables (discrete or continuous) Y_1 and Y_2, the joint or bivariate CDF is

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2)$$

for $-\infty < y_1 < \infty$ and $-\infty < y_2 < \infty$

- Note that for two discrete random variables (Y_1, Y_2),

$$F(y_1, y_2) = \sum_{z_1 \leq y_1} \sum_{z_2 \leq y_2} p(z_1, z_2)$$

Example

Back to the Dice Example:

- Find $F(2, 3)$
- Solution:

$$\begin{aligned} P(Y_1 \leq 2, Y_2 \leq 3) &= p(1,1) + p(1,2) + p(1,3) \\ &\quad + p(2,1) + p(2,2) + p(2,3) \\ &= 6/36 \\ &= 1/6 \end{aligned}$$

Properties of Joint CDFs

Theorem 4

If Y_1 and Y_2 are joint random variables with joint CDF $F(y_1, y_2)$, then

1. $F(-\infty, -\infty) = F(y_1, -\infty) = F(-\infty, y_2) = 0$
2. $F(\infty, \infty) = 1$
3. *If $a \leq b$ and $c \leq d$ then*

$$P(a \leq Y_1 \leq b, c \leq Y_2 \leq d) = \\ F(b, d) - F(b, c) - F(a, d) + F(a, c) \geq 0$$

Joint Probability Density Functions

Definition 5

Let Y_1 and Y_2 be continuous random variables with joint distribution function $F(y_1, y_2)$. If there exists a non-negative function $f(y_1, y_2)$ such that

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(u, v) du dv$$

for all $-\infty < y_1 < \infty$, $-\infty < y_2 < \infty$, then (Y_1, Y_2) are said to be jointly continuous random variables. The function $f(y_1, y_2)$ is called the joint probability density function.

Properties of Joint PDFs

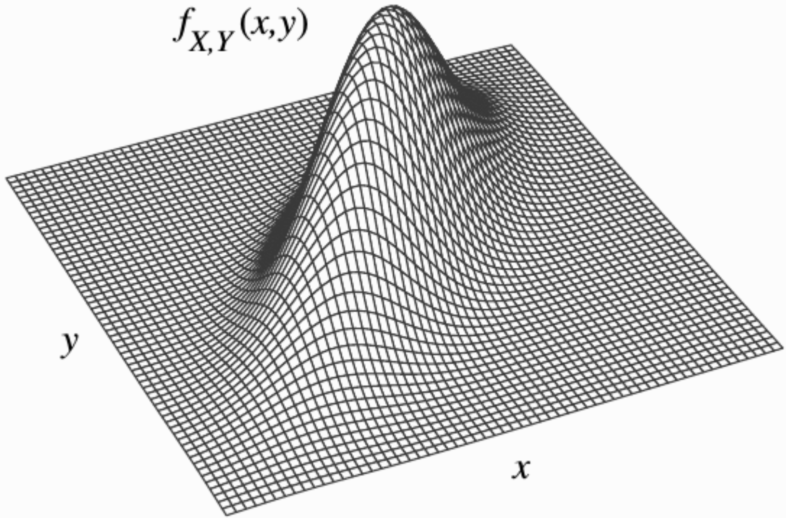
Theorem 6

If Y_1 and Y_2 are random variables with the joint pdf $f(y_1, y_2)$, then

1. $f(y_1, y_2) \geq 0$ *for all* (y_1, y_2).
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1$

Plot

An illustrative joint pdf which is a surface in 3 dimensions



Distribution Functions for Continuous R.V.s

- For jointly continuous random variables, volumes under the pdf surface correspond to probabilities
- For the pdf in the Figure, the probability $P(a_1 \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2)$ corresponds to the volume and it is obtained by

$$\int_{b_1}^{b_2} \int_{a_1}^{a_2} f(y_1, y_2) dy_1 dy_2$$

Example

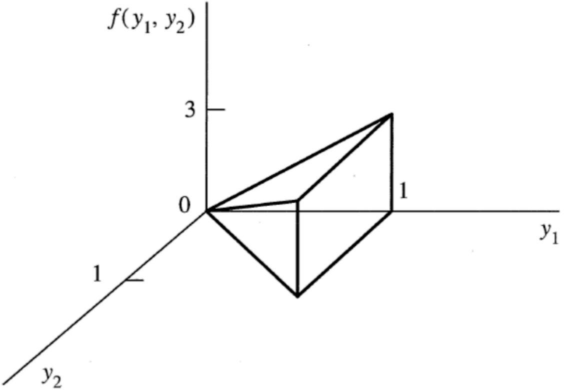
Example 7

Suppose a radioactive particle is located in a square with sides of unit length. Let Y_1 and Y_2 denote the particle's location and assume it is uniformly distributed in square; $f(y_1, y_2) = 1$ for $0 \leq y_1 \leq 1$ and $0 \leq y_2 \leq 1$ and 0 otherwise.

- Find $F(0.2, 0.4)$
- Solution: $\int_0^{0.4} \int_0^{0.2} 1 dy_1 dy_2 = 0.4 \times 0.2 = 0.08$
- Find $P(0.1 \leq Y_1 \leq 0.3, 0 \leq Y_2 \leq 0.5)$
- Solution: $\int_0^{0.5} \int_{0.1}^{0.3} 1 dy_1 dy_2 = 0.5 \times 0.2 = 0.10$

Volumes Correspond to Probabilities

FIGURE 5.4
The joint density function for Example 5.4



Example

Gasoline is stored on a FOB in a bulk tank. Let Y_1 denote the proportion of the tank available at the beginning of the week after restocking. Let Y_2 denote the proportion of the tank that is dispensed over the week. Note that Y_1 and Y_2 must be between 0 and 1 and y_2 must be less than or equal to y_1. Let the joint pdf be $f(y_1, y_2) = 3y_1$ for $0 \leq y_2 \leq y_1 \leq 1$ and 0 otherwise.

- Find $P(0 \leq Y_1 \leq 0.5, 0.25 \leq Y_2)$
- Solution

$$\int_{0.25}^{0.5} \int_{0.25}^{y_1} 3y_1 dy_2 dy_1 = \int_{0.25}^{0.5} 3y_1(y_1 - 0.25) dy_1 = \frac{5}{128}$$

Marginal and Conditional Distributions

- Marginal distributions connect the concept of (bivariate) joint distributions to univariate distributions.
  – As we will see, in the discrete case, the name “marginal distribution” follows from summing across rows or down columns of a table.
- Conditional distributions are what arise when, in a joint distribution, we fix the value of one of the random variables.

Defining Marginal Probability Functions

Definition 8

Let Y_1 and Y_2 be joint discrete random variables with the pmf $p(y_1, y_2)$. The marginal probability functions of Y_1 and Y_2, respectively, are

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2)$$

and

$$p_2(y_2) = \sum_{y_1} p(y_1, y_2)$$

Defining Marginal Density Functions

Definition 9

Let Y_1 and Y_2 be joint continuous random variables with the pdf $f(y_1, y_2)$. The marginal density functions of Y_1 and Y_2, respectively, are

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2$$

and

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1$$

Defining Conditional Probability Functions

Definition 10

If Y_1 and Y_2 are joint discrete random variables with the pmf $p(y_1, y_2)$ and marginal pmfs $p_1(y_1)$ and $p_2(y_2)$, respectively, then the conditional probability function of Y_1 given Y_2 is

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{p(y_1, y_2)}{p(y_2)}$$

- Note that $p(y_1|y_2)$ is not defined if $p_2(y_2) = 0$.
- The conditional probability function of Y_2 given Y_1 is similarly defined.

Defining Conditional Distribution Functions

Definition 11

Let Y_1 and Y_2 be joint continuous random variables with the pdf $f(y_1, y_2)$. The conditional distribution function of Y_1 given $Y_2 = y_2$ is

$$F(y_1|y_2) = \int_{-\infty}^{y_1} \frac{f(u, y_2)}{f_2(y_2)} du$$

Defining Conditional Density Functions

Definition 12

Let Y_1 and Y_2 be joint continuous random variables with pdf $f(y_1, y_2)$ and the marginal density functions $f_1(y_1)$ and $f_2(y_2)$ respectively. For any y_2 such that $f_2(y_2) > 0$, the conditional density function of Y_1 given $Y_2 = y_2$ is

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

The conditional pdf of Y_2 given $Y_1 = y_1$ is similarly defined as

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

Example

A soft drink machine has a random amount Y_2 (in gallons) in supply at the beginning of the day and dispenses a random amount Y_1 during the day. It is not resupplied during the day, so $Y_1 \leq Y_2$, and the joint pdf is $f(y_1, y_2) = 1/2$ for $0 \leq y_1 \leq y_2 \leq 2$ and 0 otherwise.

- What is the probability that less than half gallon will be sold given that the machine containing 1.5 gallons at the start of the day?

Solution

- Solution: First note that

$$f_2(y_2) = \int_0^{y_2} \frac{1}{2} dy_1 = \frac{y_2}{2} \qquad 0 \leq y_2 \leq 2$$

then

$$\begin{aligned} P(Y_1 \leq 0.5 | Y_2 = 1.5) &= \int_0^{0.5} \frac{f(y_1, y_2 = 1.5)}{f_2(y_2 = 1.5)} dy_1 \\ &= \int_0^{0.5} \frac{1/2}{0.75} dy_1 \\ &= \frac{1}{3} \end{aligned}$$

Expected Value of a Function of R.V.s

Definition 13

Let $g(Y_1, ..., Y_k)$ be a function of the discrete random variables $Y_1, ..., Y_k$, which have joint pmf $p(y_1, ..., y_k)$. Then the expected value of $g(Y_1, ..., Y_k)$ is

$$E\left[g(Y_1, ..., Y_k)\right] = \sum_{y_1} ... \sum_{y_k} g(y_1, ..., y_k) p(y_1, ..., y_k)$$

If $Y_1, ..., Y_k$ are continuous with joint pdf $f(y_1, ..., y_k)$, then

$$E\left[g(Y_1, ..., Y_k)\right] = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(y_1, ..., y_k) f(y_1, ..., y_k) dy_1 ... dy_k$$

Example

An industrial chemical process yields a product with two types of impurities. In a given sample, let Y_1 denote the proportion of impurities in the sample and let Y_2 denote the proportion of "type I" impurities (out of all the impurities). The joint density function of (Y_1, Y_2) is $f(y_1, y_2) = 2(1 - y_1)$ for $0 \leq y_1 \leq 1$ and $0 \leq y_2 \leq 1$ and 0 otherwise.

- Find the expected value of the proportion of "type I" impurities in the sample.
- Solution: $E(Y_2) = \int_0^1 \int_0^1 y_2 2(1 - y_1) dy_2 dy_1 = 0.5$
- Shortcut: We can show from the marginal distributions that Y_1 is $Beta(1,2)$ and Y_2 is $U(0,1)$. Therefore, $E(Y_2) = 0.5$

Independent Random Variables

- We defined the independence of events
  – here we now extend that idea to the independence of random variables.
- For joint random variables (Y_1, Y_2), the probabilities associated with Y_1 are the same regardless of the observed value of Y_2
  – The idea is that learning something about Y_2 does not tell you anything, probabilistically speaking, about the distribution of Y_1.

Defining Independence for R.V.s

Definition 14

Let Y_1 have CDF $F_1(y_1)$, Y_2 have CDF $F_2(y_2)$, and Y_1 and Y_2 have joint CDF $F(y_1, y_2)$. Then Y_1 and Y_2 are independent if and only if

$$F(y_1, y_2) = F_1(y_1)F_2(y_2)$$

for every pair of real numbers (y_1, y_2).

If Y_1 and Y_2 are not independent, then they are said to be dependent.

Determining Independence

Theorem 15

- *Let Y_1 and Y_2 be discrete random variables with the j oint pmf $p(y_1, y_2)$ and marginal pmfs $p_1(y_1)$ and $p_2(y_2)$ respectively. Then Y_1 and Y_2 are independent if and only if*
$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$
for all pairs of real numbers (y_1, y_2).
- *Let (Y_1, Y_2) are continuous random variables with the joint pdf $f(y_1, y_2)$ and the marginal pdfs $f_1(y_1)$ and $f_2(y_2)$ respectively. Then Y_1 and Y_2 are independent if and only if*
$$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$
for all pairs of real numbers (y_1, y_2).

Example

- Let $f(y_1, y_2) = 6y_1y_2^2$ for $0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$ and 0 otherwise. Show that Y_1 and Y_2 are independent.
- Solution: We can show that $f_1(y_1) = 2y_1$ for $0 \leq y_1 \leq 1$ and $f_2(y_2) = 3y_2^2$ for $0 \leq y_2 \leq 1$. Since for every y_1 and y_2,

  $$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$

  By Theorem 15, Y_1 and Y_2 are independent.

Determining Independence of Two R.V.s

Theorem 16

Let Y_1 and Y_2 have joint density function $f(y_1, y_2)$ which is positive if and only if $a \leq y_1 \leq b$ and $c \leq y_2 \leq d$, for constants a, b, c, d and $f(y_1, y_2) = 0$ otherwise. Then Y_1 and Y_2 are independent random variables if and only if

$$f(y_1, y_2) = g(y_1)h(y_2)$$

where $g(y_1)$ is a nonnegative function of y_1 alone and $h(y_2)$ is a nonnegative function of y_2 alone.
– Note that $g(y_1)$ and $h(y_2)$ do not have to be density functions.

Example

- Let Y_1 and Y_2 have joint density $f(y_1, y_2) = 2y_1$ for $0 \leq y_1 \leq 1$ and $0 \leq y_2 \leq 1$ and 0 otherwise. Are Y_1 and Y_2 independent?
- Solution: Yes, because of Theorem 16.
- Let Y_1 and Y_2 have joint density $f(y_1, y_2) = 2$ for $0 \leq y_2 \leq y_1 \leq 1$ and 0 otherwise. Are Y_1 and Y_2 independent?
- Solution: No, because of Theorem 16.

Theorem

Theorem 17

Let Y_1 and Y_2 be independent random variables and $g(Y_1)$ and $h(Y_2)$ be functions of only Y_1 and Y_2 respectively. Then

$$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)]$$

provided the expectations exist.

The Covariance of Two Random Variables

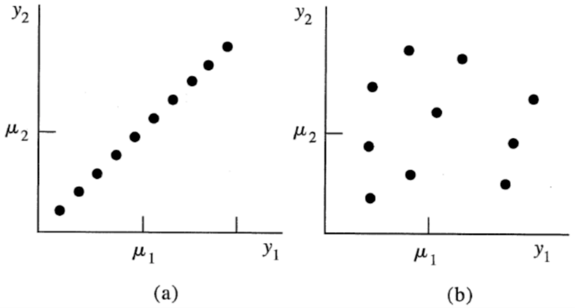
- When looking at two random variables, a common question is whether they are associated with one-another
  – That is, for example, as one tends to increase, does the other do so as well?
  – If so, the colloquial terminology is to say that the two variables are correlated

- Correlation is also a technical term used to describe the association.
  – But it has a precise definition (that is more restrictive than the colloquial use of the term)
  – And there is a second measure from which correlation is derived: covariance

Covariance and Correlation

- Covariance and correlation are measures of dependence between two random variables.
- The figure below shows two extremes: perfect dependence and independence.

FIGURE 5.8 Dependent and independent observations for (y_1, y_2)



Defining Covariance

Definition 18

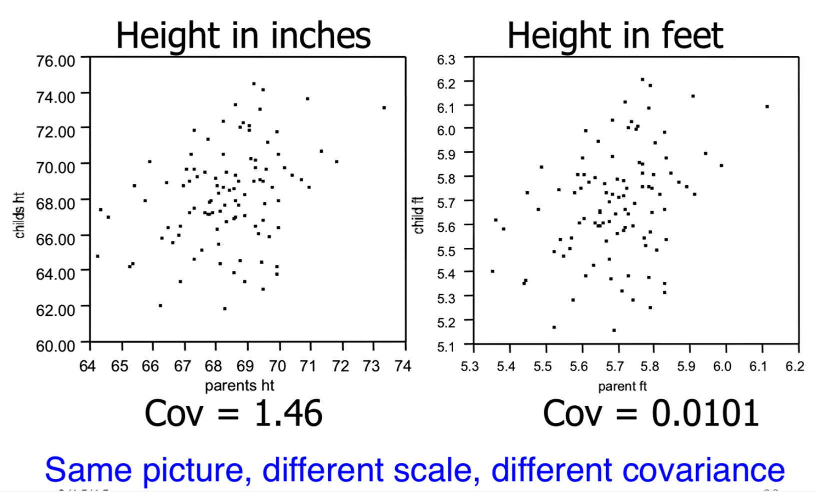
If Y_1 and Y_2 are random variables with means μ_1 and μ_2 respectively. Then the covariance of Y_1 and Y_2 is defined as

$$\begin{aligned} Cov(Y_1, Y_2) &= E\left[(Y_1 - \mu_1)(Y_2 - \mu_2)\right] \\ &= E(Y_1 Y_2) - \mu_1 \mu_2 \end{aligned}$$

Understanding Covariance

- Covariance gives the strength and direction of linear relationship between Y_1 and Y_2:
  - Strong-positive: large positive covariance
  - Strong-negative: large negative covariance
  - Weak positive: small positive covariance
  - Weak negative: small negative covariance
- But what is "large" and "small"?
  - It depends on the measurement units of Y_1 and Y_2

Drawback of Covariance



Correlation of Random Variables

- The correlation (coefficient) of two random variables is defined as:

$$\rho = \frac{Cov(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

As with covariance, correlation is a measure of the dependence of two random variables Y_1 and Y_2

– But note that it's re-scaled to be unit-free and measurement invariant

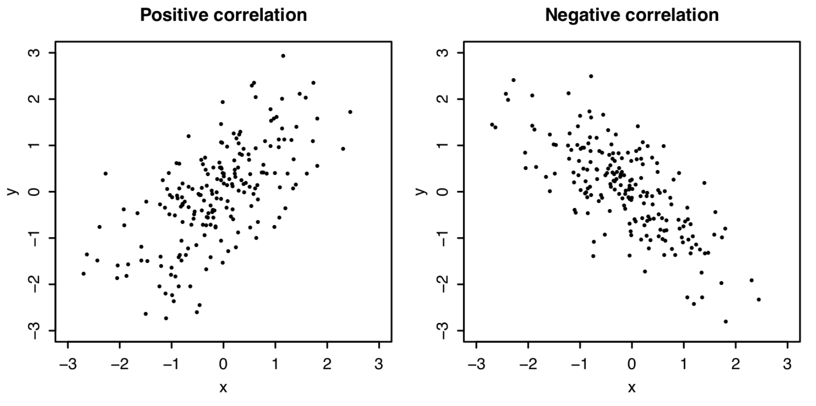
– In particular, for any random variables Y_1 and Y_2

$$-1 \leq \rho \leq 1$$

Interpreting the Correlation Coefficient

- A positive correlation coefficient ($\rho > 0$) means there is a positive association between Y_1 and Y_2.
  - As Y_1 increases Y_2 also tends to increase.
- A negative correlation coefficient ($\rho < 0$) means there is a negative association between Y_1 and Y_2.
  - As Y_1 increases Y_2 tends to decrease.
- A correlation coefficient equal to 0 ($\rho = 0$ means there is no linear association between Y_1 and Y_2.
- Link: http://jose-coto.com/visualizing-bivariate-normal

Visulallization



Example

- Find the covariance between Y_1 and Y_2, where $f(y_1, y_2) = 3y_1$ for $0 \leq y_2 \leq y_1 \leq 1$ and 0 otherwise.
- Solution: By the definition, first

$$E(Y_1Y_2) = \int_0^1 \int_0^{y_1} y_1 y_2 3y_1 dy_2 dy_1 = \int_0^1 \frac{3}{2} y_1^4 dy_1 = \frac{3}{10}$$

- Further, $E(Y_1) = \int_0^1 \int_0^{y_1} y_1 3y_1 dy_2 dy_1 = \frac{3}{4}$
- likewise, $E(Y_2) = \int_0^1 \int_0^{y_1} y_2 3y_1 dy_2 dy_1 = \frac{3}{8}$
- Hence, $Cov(Y_1, Y_2) = \frac{3}{10} - \frac{3}{4} \times \frac{3}{8} = \frac{3}{160}$

Correlation and Independence

Theorem 19

If Y_1 and Y_2 are independent random variables. Then $Cov(Y_1, Y_2) = 0$. That is, independent random variables are uncorrelated.

Notes

- High correlation merely means that there is a strong linear relationship, not that Y_1 causes Y_2 or Y_2 causes Y_1.
- E.g., there could be a third factor that causes both Y_1 and Y_2 to move in the same direction
  - No causal connection between Y_1 and Y_2.
- High correlation does not imply causation.
  - Zero correlation does not mean there is no relationship between Y_1 and Y_2
- Link: http://www.abs.gov.au/websitedbs/a3121120.nsf/home/

Mean, Variance, Covariance of Linear Combinations of R.V.s

Theorem 20

Let $Y_1, Y_2, ..., Y_n$ and $X_1, X_2, ..., X_m$ be random variables with $E(Y_i) = \mu_i$ and $E(X_i) = \xi_i$. Define $U_1 = \sum_{i=1}^{n} a_i Y_i$ and $U_2 = \sum_{j=1}^{m} b_j X_j$ for constants $a_1, ..., a_n$ and $b_1, ..., b_n$. Then:

- $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$ *and* $E(U_2) = \sum_{j=1}^{m} b_i \xi_j$
- $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum \sum_{1 \leq i < j \leq 1} a_i b_j Cov(Y_i, Y_j)$
- *Similar expression for* $V(U_2)$.
- $Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(Y_i, X_j)$

Example (Exercise)

- Let Y_1, Y_2, and Y_3 be random variables, where
  - $E(Y_1) = 1$, $E(Y_2) = 2$, $E(Y_3) = -1$
  - $V(Y_1) = 1$, $V(Y_2) = 3$, and $V(Y_3) = 5$
  - $Cov(Y_1, Y_2) = -0.4$, $Cov(Y_1, Y_3) = 0.5$, and $Cov(Y_2, Y_3) = -0.2$
- Find the expected value and variance of $U = Y_1 - 2Y_2 - Y_3$.
- If $W = 3Y_1 + Y_2$, find $Cov(U, W)$.

Defining a Multinomial Experiment

Definition 21

A multinomial experiment has the following properties:

1. The experiment consists of n identical trials.
2. The outcome of each trial falls into one of k classes or cells.
3. The probability that the outcome of a single trial falls into cell i, p_i, $i = 1, ..., k$ remains the same from trial totrialand $p_1 + p_2 + \cdots + p_k = 1$.
4. The trials are independent.
5. The random variables of interest are $Y_1, Y_2, ..., Y_k$, the number of outcomes that fall in each of the cells, where $Y_1 + Y_2 + \cdots + Y_k = n$.

Defining the Multinomial Distribution

Definition 22

Assume that $p_i > 0$, $i = 1, ..., k$ are such that $p_1 + p_2 + \cdots + p_k = 1$. The random variables $(Y_1, Y_2, ..., Y_k)$ have a multinomial distribution with parameters. $p_1, p_2, ..., p_k$ if their joint distribution is

$$p(y_1, ..., y_k) = \frac{n!}{y_1! ... y_{k-1}! y_k!} p_1^{y_1} ... p_{k-1}^{y_{k-1}} p_k^{y_k}$$

where, $y_k = (n - \sum_{i=1}^{k-1} y_i)$ and $p_k = (1 - \sum_{i=1}^{k-1} p_i)$.

Further, for each i, $y_i = 0, 1, 2, ..., n$ and $\sum_{i=1}^{k} y_i = n$.

Note that the binomial distribution is just a special case of the multinomial with $k = 2$.

Expected Value, Variance, and Covariance of the Multinomial Distribution

Theorem 23

If $(Y_1, Y_2, ..., Y_k)$ have a multinomial distribution with parameters n and $p_1, p_2, ..., p_k$. Then

- $E(Y_i) = np_i$
- $V(Y_i) = np_i(1 - p_i)$
- $Cov(Y_i, Y_j) = -np_ip_j$ *for* $i \neq j$.

The Multivariate Normal Distribution

- The multivariate normal distributions is a joint distribution for $k > 1$ random variables
$$f(y) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right\}$$

- The pdf for a multivariate normal is where
  - Y and μ are k dimensional vectors for the location at which to evaluate the pdf and the means, respectively.
  - Σ is the variance-covariance matrix.
  - For the bivariate normal: (Y_1, Y_2) with $\mu = (\mu_1, \mu_2)$ and
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Defining the Conditional Expectation

Definition 24

If Y_1 and Y_2 are any random variables, the conditional expectation of $g(Y_1)$ given that $Y_2 = y_2$ is defined as

- if Y_1 and Y_2 are jointly continuous

$$E(g(Y_1)|Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1)f(y_1|y_2)dy_1$$

- if Y_1 and Y_2 are jointly discrete.

$$E(g(Y_1)|Y_2 = y_2) = \sum_{y_1} g(y_1)p(y_1|y_2)$$

Example

For random variables Y_1 and Y_2 with the joint pdf $f(y_1, y_2) = 1/2$ for $0 \leq y_1 \leq y_2 \leq 2$ and 0 otherwise.

- Find the conditional expectation of Y_1 given $Y_2 = 1.5$.
- Solution: First, $f_2(y_2) = \int_0^{y_2} \frac{1}{2} dy_1 = \frac{y_2}{2}$ for $0 \leq y_2 \leq 2$
- Therefore, $f(y_1|y_2) = \frac{1/2}{y_2/2} = \frac{1}{y_2}$. That is, Y_1 given Y_2 is a uniform distribution in $(0, y_2)$. Hence, $E(Y_1|Y_2 = 1.5) = \frac{1.5}{2} = 0.75$
- Alternatively, using the definition

$$E(Y_1|Y_2) = \int_0^{y_2} y_1 f(y_1|y_2) dy_1 = \frac{y_2}{2}$$

So, $E(Y_1|Y_2 = 1.5) = 0.75$

Repeated Conditional Expectation Formula

Theorem 25

Let Y_1 and Y_2 denote random variables. Then

$$E(Y_1) = E[E(Y_1|Y_2)]$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Conditional Variance Formula

Theorem 26

Let Y_1 and Y_2 denote random variables. Then

$$V(Y_1) = E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)]$$

Example

Let Y_1 have an exponential distribution with mean λ and the conditional density of Y_2 given $Y_1 = y_1$ be $f(y_2|y_1) = 1/y_1$ for $0 \leq y_2 \leq y_1$ and 0 otherwise.

- Find $E(Y_2)$ and $V(Y_2)$, the unconditional mean and variance of Y_2.
- Solution: First note that Y_2 given Y_1 is a uniform distribution in $(0, y_1)$. Hence, $E(Y_2|Y_1) = \frac{Y_1}{2}$ and $V(Y_2|Y_1) = \frac{Y_1^2}{12}$
- Therefore from the pervious theorems, $E[Y_2] = E[E(Y_2|Y_1)] = E(\frac{Y_1}{2}) = \frac{\lambda}{2}$ and

$$V(Y_2) = E\left[\frac{Y_1^2}{12}\right] + V\left[\frac{Y_1}{2}\right] = \frac{5\lambda^2}{12}$$

What We Have Just Learned

- Bivariate and multivariate probability distributions
  - Bivariate normal distribution
  - Multinomial distribution
- Marginal and conditional distributions
- Independence, covariance and correlation
- Expected value and variance of a function of random variables
  - Linear functions of random variables in particular
- Conditional expectation and variance