Part II

Lecture: Random Variables, Expectation and Moment Generating Function

Random Variable

Let (Ω, S) be a sample space. A finite single-valued function that maps Ω into $\Re$ is called a random variable (RV) if the inverse images under X of all Borel sets in $\Re$ are events, that is, if
$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in S, \quad \text{for all } B \in \mathcal{B}.$

Random Variable

Let (Ω, S) be a sample space. A finite single-valued function that maps Ω into $\Re$ is called a random variable (RV) if the inverse images under X of all Borel sets in $\Re$ are events, that is, if
$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in S, \quad \text{for all } B \in \mathcal{B}.$

- A RV is a numerical description of the outcome of an experiment.

Random Variable

Let (Ω, S) be a sample space. A finite single-valued function that maps Ω into $\Re$ is called a random variable (RV) if the inverse images under X of all Borel sets in $\Re$ are events, that is, if
$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in S, \quad \text{for all } B \in \mathcal{B}.$

- A RV is a numerical description of the outcome of an experiment.
- A variable that takes on different values as a result of the outcomes of a random experiment.

Random Variable

Let (Ω, S) be a sample space. A finite single-valued function that maps Ω into $\Re$ is called a random variable (RV) if the inverse images under X of all Borel sets in $\Re$ are events, that is, if
$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in S, \quad \text{for all } B \in \mathcal{B}.$

- A RV is a numerical description of the outcome of an experiment.
- A variable that takes on different values as a result of the outcomes of a random experiment.
- The set of all values which X takes up may be discrete or continuous, and accordingly the RV is said to be discrete/continuous.

Random Variable

Let (Ω, S) be a sample space. A finite single-valued function that maps Ω into $\Re$ is called a random variable (RV) if the inverse images under X of all Borel sets in $\Re$ are events, that is, if
$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in S, \quad \text{for all } B \in \mathcal{B}.$

- A RV is a numerical description of the outcome of an experiment.
- A variable that takes on different values as a result of the outcomes of a random experiment.
- The set of all values which X takes up may be discrete or continuous, and accordingly the RV is said to be discrete/continuous.

Random Variable

Let (Ω, S) be a sample space. A finite single-valued function that maps Ω into $\Re$ is called a random variable (RV) if the inverse images under X of all Borel sets in $\Re$ are events, that is, if
$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in S, \quad \text{for all } B \in \mathcal{B}.$

- A RV is a numerical description of the outcome of an experiment.
- A variable that takes on different values as a result of the outcomes of a random experiment.
- The set of all values which X takes up may be discrete or continuous, and accordingly the RV is said to be discrete/continuous.

There do exist subsets of $\Re$ that do not belong to $\mathcal{B}$, and hence there exist real-valued functions defined on Ω that are not RVs.

Let (Ω, S, P) be a probability space, and let X be an RV defined on it.

Probability Function/ Distribution Function

- A function, denoted by $f_i = P\{X = x_i\}$, that provides the probability that a discrete RV X takes on some specific value such that $f_i \geqslant 0$ for all i and $\sum_{i=1}^{\infty} f_i = 1$, is called the probability mass function (pmf) of RV X. Then the distribution function (DF) of X is a function of a real variable x denoted by $F(x)$ and is given by

$$F(x) = P(-\infty < X \leqslant x) = \sum_{j=-\infty}^{i} f_j \ \ (i = 0, \pm 1, \pm 2, \ldots), \ \text{in } x_i \leqslant x < x_{i+1}.$$

Let (Ω, S, P) be a probability space, and let X be an RV defined on it.

Probability Function/ Distribution Function

- A function, denoted by $f_i = P\{X = x_i\}$, that provides the probability that a discrete RV X takes on some specific value such that $f_i \geqslant 0$ for all i and $\sum_{i=1}^{\infty} f_i = 1$, is called the probability mass function (pmf) of RV X. Then the distribution function (DF) of X is a function of a real variable x denoted by $F(x)$ and is given by

$$F(x) = P(-\infty < X \leqslant x) = \sum_{j=-\infty}^{i} f_j \ (i = 0, \pm 1, \pm 2, \ldots), \ \text{in } x_i \leqslant x < x_{i+1}.$$

- A nonnegative real-valued function $f(x)$ that is integrable over $\Re$ and satisfies $\int_{-\infty}^{\infty} f(x)dx = 1$ is called the probability density function (pdf) of some continuous RV X. Then the DF of X is

$$F(x) = P(-\infty < X \leqslant x) = \int_{-\infty}^{x} f(u)du.$$

Let (Ω, S, P) be a probability space, and let X be an RV defined on it.

Probability Function/ Distribution Function

- A function, denoted by $f_i = P\{X = x_i\}$, that provides the probability that a discrete RV X takes on some specific value such that $f_i \geqslant 0$ for all i and $\sum_{i=1}^{\infty} f_i = 1$, is called the probability mass function (pmf) of RV X. Then the distribution function (DF) of X is a function of a real variable x denoted by $F(x)$ and is given by

$$F(x) = P\left(-\infty < X \leqslant x\right) = \sum_{j=-\infty}^{i} f_j \;\; (i = 0, \pm 1, \pm 2, \ldots), \;\; \text{in}\; x_i \leqslant x < x_{i+1}.$$

- A nonnegative real-valued function $f(x)$ that is integrable over $\Re$ and satisfies $\int_{-\infty}^{\infty} f(x)dx = 1$ is called the probability density function (pdf) of some continuous RV X. Then the DF of X is

$$F(x) = P\left(-\infty < X \leqslant x\right) = \int_{-\infty}^{x} f(u)du.$$

Let (Ω, S, P) be a probability space, and let X be an RV defined on it.

Probability Function/ Distribution Function

- A function, denoted by $f_i = P\{X = x_i\}$, that provides the probability that a discrete RV X takes on some specific value such that $f_i \geqslant 0$ for all i and $\sum_{i=1}^{\infty} f_i = 1$, is called the probability mass function (pmf) of RV X. Then the distribution function (DF) of X is a function of a real variable x denoted by $F(x)$ and is given by

$$F(x) = P(-\infty < X \leqslant x) = \sum_{j=-\infty}^{i} f_j \;\; (i = 0, \pm 1, \pm 2, \ldots), \;\; \text{in } x_i \leqslant x < x_{i+1}.$$

- A nonnegative real-valued function $f(x)$ that is integrable over $\Re$ and satisfies $\int_{-\infty}^{\infty} f(x)dx = 1$ is called the probability density function (pdf) of some continuous RV X. Then the DF of X is

$$F(x) = P(-\infty < X \leqslant x) = \int_{-\infty}^{x} f(u)du.$$

A real valued function F defined on $(-\infty, \infty)$ that is nondecreasing, right continuous, and satisfies $F(-\infty) = 0$ and $F(+\infty) = 1$ is called a distribution function (DF).

Example

The *Statistical Abstract of the United States* is published annually. It contains a wide variety of information based on the census as well as other sources. The objective is to provide information about a variety of different aspects of the lives of the country's residents. One of the questions asked households to report the number of color televisions in the household. The following table summarizes the data. Develop the probability distribution of the random variable defined as the number of color televisions per household.

Example

.

Number of Color Televisions	Number of Households (1,000s)
0	1,218
1	32,379
2	37,961
3	19,387
4	7,714
5	2,842
Total	101,501

Example

Probability distributions can be estimated from relative frequencies.

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	101,501		**1.000**

$1{,}218 \div 101{,}501 = 0.012$

e.g. $P(\mathbf{X}=4) = P(4) = 0.076 = 7.6\%$

Example

E.g. what is the probability there is at least one television but no more than three in any given household?

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	101,501		**1.000**

"at least one television but no more than three"

$$P(1 \le \mathbf{X} \le 3) = P(1) + P(2) + P(3) = .319 + .374 + .191 = .884$$

The Problem of Medical Representative

New Statement of the Problem

In order to solve the problem we need to to define the following random variable:

$$X = Total\ number\ of\ days\ lost\ out\ of\ the\ remaining\ 27\ days.$$

Here X is a discrete random variable which can take any of the values $0, 1, 2, \ldots, 27$. Now the problem of our interest can be re-stated using this random variable X as

Compute the probability that $X \leqslant 1$, *i.e.*, the Probability that $X = 0\ or\ 1$.

Expectation of a random variable X:

Discrete Case:

Let X be a random variable with *pmf* $f_i = P(X = x_i),\ i = 1, 2, ...$ If $\sum_{i=1}^{\infty} |x_i| f_i < \infty$, then $E(X)$ exists and we write $\mu = E(X) = \sum_{i=1}^{\infty} x_i f_i$. It represents the centre of mass of the probability mass distribution and as such is called a measure of location.

Expectation of a random variable X:

Discrete Case:

Let X be a random variable with *pmf* $f_i = P(X = x_i),\ i = 1, 2, ...$ If $\sum_{i=1}^{\infty} |x_i| f_i < \infty$, then $E(X)$ exists and we write $\mu = E(X) = \sum_{i=1}^{\infty} x_i f_i$. It represents the centre of mass of the probability mass distribution and as such is called a measure of location.

N.B.: If the series $\sum |x_i| f_i$ doesn't converge then $E(X)$ doesn't exist even if the series $\sum x_i f_i$ converge. For example, let $f_i = P\{X = (-1)^{i+1} \frac{3^i}{i}\} = \frac{2}{3^i}\ i = 1, 2, ...,$ then $\sum |x_i| f_i = \sum_{i=1}^{\infty} \frac{2}{i} = \infty$ and so $E(X)$ doesn't exist. Although the series $\sum x_i f_i = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{2}{i}$ is convergent.

Continuous Case:

If X is a continuous random variable with *pdf* $f(x)$ and $\int_{-\infty}^{\infty} |x| f(x)dx < \infty$, then the expectation of X is $E(X) = \int_{-\infty}^{\infty} x f(x)dx$. It provides weighted average of the values of the random variable for which the pdf gives the weights. If an experiment is repeated large number of times, the expected value can be interpreted as the 'long-run average'.

Continuous Case:

If X is a continuous random variable with *pdf* $f(x)$ and $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$, then the expectation of X is $E(X) = \int_{-\infty}^{\infty} x f(x) dx$. It provides weighted average of the values of the random variable for which the pdf gives the weights. If an experiment is repeated large number of times, the expected value can be interpreted as the 'long-run average'.

N.B.: $E(X)$ may not exist for every random variable, for example, for cauchy distribution $f(x) = \frac{1}{\lambda} \frac{1}{1+x^2}, \ x \in (-\infty, \infty)$, the integral $\frac{1}{\pi} \int_{-\infty}^{\infty} |x|/(1+x^2) dx$ diverges. Hence the mean of cauchy distribution doesn't exist.

Continuous Case:

If X is a continuous random variable with *pdf* $f(x)$ and $\int_{-\infty}^{\infty}|x|f(x)dx < \infty$, then the expectation of X is $E(X) = \int_{-\infty}^{\infty} xf(x)dx$. It provides weighted average of the values of the random variable for which the pdf gives the weights. If an experiment is repeated large number of times, the expected value can be interpreted as the 'long-run average'.

N.B.: $E(X)$ may not exist for every random variable, for example, for cauchy distribution $f(x) = \frac{1}{\lambda}\frac{1}{1+x^2},\ x \in (-\infty, \infty)$, the integral $\frac{1}{\pi}\int_{-\infty}^{\infty}|x|/(1+x^2)dx$ diverges. Hence the mean of cauchy distribution doesn't exist.

Exercise

(i) Let

$$f(x) = \begin{cases} 4x^3, & 0 < x < 1 \\ 0, & elsewhere. \end{cases}$$

Then $E(X) = \int_0^1 x.4x^3dx = 4/5$
(ii) Let $f(x) = 2x,\ 0 < x < 1$. Suppose $Y = 1/(1+X)$. Then $E(Y) = \int_0^1 \frac{1}{1+x}.2xdx = 2(1-\ln 2)$.

Higher Order Moment of X:

Let k be a positive integer and c be a constant. The moment of order k or the k^{th} moment of X about the fixed point 'c' is defined to be the mean value $E\{(X-c)^k\}$, if exists.

Raw moment:

The k^{th} moment about the origin i.e., $c = 0$ and we write

$$\alpha_k = E(X^k).$$

Clearly, $\alpha_1 = \mu$, i.e., mean is the 1^{st} order moment of random variable X.

Higher Order Moment of X:

Let k be a positive integer and c be a constant. The moment of order k or the k^{th} moment of X about the fixed point 'c' is defined to be the mean value $E\{(X-c)^k\}$, if exists.

Raw moment:

The k^{th} moment about the origin i.e., $c = 0$ and we write

$$\alpha_k = E(X^k).$$

Clearly, $\alpha_1 = \mu$, i.e., mean is the 1^{st} order moment of random variable X.

Central moment:

If we take $c = E(X) = \mu$, which exists then $E(X-\mu)^k$ is called the k^{th} order central moment and we write

$$\mu_k = E\{(X-\mu)^k\}.$$

We have $\mu_1 = 0$ for all random variable. Note that 0^{th} order moment of any random variable is always unity.

Variance of a Random Variable X:

The second central moment μ_2 is of great importance and is called the variance of X written as $Var(X)$, i.e.,

$$Var(X) = \mu_2 = E\{(X-\mu)^2\} = \sigma^2, \text{ say.}$$

- A measure of the dispersion or variability in the random variable.
- A weighted average of the squared deviations from the mean μ.

Variance of a Random Variable X:

The second central moment μ_2 is of great importance and is called the variance of X written as $Var(X)$, i.e.,

$$Var(X) = \mu_2 = E\{(X - \mu)^2\} = \sigma^2, \text{ say.}$$

- A measure of the dispersion or variability in the random variable.
- A weighted average of the squared deviations from the mean μ.

Variance is a characteristic which describes how widely the probability masses are spread about the mean or, in otherwords, an inverse measure of concentration of the probability masses about the mean. It is called a measure of dispersion. It represents the moment of inertia of the probability mass distribution about a line through the mean perpendicular to the line of the distribution.

Variance of a Random Variable X:

The second central moment μ_2 is of great importance and is called the variance of X written as $Var(X)$, i.e.,

$$Var(X) = \mu_2 = E\{(X-\mu)^2\} = \sigma^2, \text{ say.}$$

- A measure of the dispersion or variability in the random variable.
- A weighted average of the squared deviations from the mean μ.

Variance is a characteristic which describes how widely the probability masses are spread about the mean or, in otherwords, an inverse measure of concentration of the probability masses about the mean. It is called a measure of dispersion. It represents the moment of inertia of the probability mass distribution about a line through the mean perpendicular to the line of the distribution.

Formulae for Evaluating Variance

(i) $\sigma^2 = \alpha_2 - \mu^2$ [N.B.: If α_2 exists then only variance exists]
(ii) $\sigma^2 = E\{X(X-1)\} - \mu(\mu-1)$

Example

Find the **mean**, variance, and standard deviation for the population of the number of color televisions per household… (from prev. Example)

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	101,501		**1.000**

$$E(X) = \mu = \sum_{all\ x} xP(x) = 0 \bullet P(0) + 1 \bullet P(1) + \ldots + 5 \bullet P(5)$$

$= 0(.012) + 1(.319) + 2(.374) + 3(.191) + 4(.076) + 5(.028)$

$=$ **2.084**

Example

Find the mean, **variance**, and standard deviation for the population of the number of color televisions per household… (from prev. Example)

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	101,501		**1.000**

$$V(X) = \sigma^2 = \sum_{all\ x} (x-\mu)^2 P(x)$$

$= (0 - 2.084)^2(.012) + (1 - 2.084)^2(.319) + \ldots + (5 - 2.084)^2(.028)$

$=$ **1.107**

Example

Find the mean, variance, and **standard deviation** for the population of the number of color televisions per household… (from prev. Example)

# of Televisions	# of Households	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	101,501		**1.000**

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.107} = \mathbf{1.052}$$

Exercise

The following table represents the number of cars sold per day at DiCarlo Motors, Inc., New York. Construct the probability distribution for the number of cars sold per day. Compute the expected value and the variance for the number of daily sales.

Sales Volume	Number of Days
No sales	54
One car	117
Two cars	72
Three cars	42
Four cars	12
Five cars	3
Total	300

Moment Generating Function (mgf):

Here we consider a function that generate probabilities or moments of a random variables and is quite useful in probability and statistics.

Moment Generating Function (mgf):

Here we consider a function that generate probabilities or moments of a random variables and is quite useful in probability and statistics.

Definition

Let X be a random variable defined on (Ω, S, P). The mgf of X is a function of a real variable t denoted by $M_X(t)$ or $M(t)$ and defined by

$$M(t) = E(e^{tX}) = \begin{cases} \sum_{i=0}^{\infty} p_i e^{ti} & (\text{ for discrete } r.v.) \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & (\text{ for continuous } r.v.) \end{cases}$$

if the expectation exists in some neighbourhood of the origin.

Moment Generating Function (mgf):

Here we consider a function that generate probabilities or moments of a random variables and is quite useful in probability and statistics.

Definition

Let X be a random variable defined on (Ω, S, P). The mgf of X is a function of a real variable t denoted by $M_X(t)$ or $M(t)$ and defined by

$$M(t) = E(e^{tX}) = \begin{cases} \sum_{i=0}^{\infty} p_i e^{ti} & (\text{ for discrete } r.v.) \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & (\text{ for continuous } r.v.) \end{cases}$$

if the expectation exists in some neighbourhood of the origin.

N.B.: For $t = 0$, $M(t)$ certainly exists and $M(0) = 1$. But, for non-zero values of t the series or integral concerned may not convergent for all distributions, for example, Cauchy distribution.

The existence of $M(t)$ in a small neighbourhood of $t = 0$ implies that derivatives of $M(t)$ of all orders exists at $t = 0$. The power series development of $M(t)$ of any distribution will be

$$M(t) = \sum_{k=0}^{\infty} \frac{\alpha_k}{k!} t^k.$$

Thus, the coefficient of t^k is $\frac{\alpha_k}{k!}$, justifying the name *mgf**.

The existence of $M(t)$ in a small neighbourhood of $t = 0$ implies that derivatives of $M(t)$ of all orders exists at $t = 0$. The power series development of $M(t)$ of any distribution will be

$$M(t) = \sum_{k=0}^{\infty} \frac{\alpha_k}{k!} t^k.$$

Thus, the coefficient of t^k is $\frac{\alpha_k}{k!}$, justifying the name *mgf**.
Successive differentiation of $M(t)$ w.r.t. t yields

$$M^{(k)}(0) = \frac{d^k}{dt^k} M(t)]_{t=0} = E(X^k) = \alpha_k$$

($*$ $M(t)$ generates the moments α_k)

विद्यारत्नम् महद्धनम्

The existence of $M(t)$ in a small neighbourhood of $t = 0$ implies that derivatives of $M(t)$ of all orders exists at $t = 0$. The power series development of $M(t)$ of any distribution will be

$$M(t) = \sum_{k=0}^{\infty} \frac{\alpha_k}{k!} t^k.$$

Thus, the coefficient of t^k is $\frac{\alpha_k}{k!}$, justifying the name *mgf**. Successive differentiation of $M(t)$ w.r.t. t yields

$$M^{(k)}(0) = \frac{d^k}{dt^k} M(t)]_{t=0} = E(X^k) = \alpha_k$$

($*$ $M(t)$ generates the moments α_k)
Hence,

$$\mu = E(X) = M'(0),\ E(X^2) = M''(0) \Rightarrow \sigma^2 = M''(0) - \{M'(0)\}^2.$$

Exercise

i. A distribution with *pdf* $f(x) = e^{-x},\ 0 < x < \infty$ and 0 elsewhere has the *mgf* $M(t) = (1-t)^{-1},\ t < 1$. Then $M'(t) = (1-t)^{-2},\ M''(t) = 2(1-t)^{-3}$ giving that $\mu = 1$ and $\sigma = 2 - 1 = 1$.

Exercise

i. A distribution with *pdf* $f(x) = e^{-x},\ 0 < x < \infty$ and 0 elsewhere has the *mgf* $M(t) = (1-t)^{-1},\ t < 1$. Then $M'(t) = (1-t)^{-2},\ M''(t) = 2(1-t)^{-3}$ giving that $\mu = 1$ and $\sigma = 2 - 1 = 1$.

ii. Let $f(x) = \frac{1}{2}e^{-x/2}, x > 0$ and 0 otherwise. Then $M(t) = \frac{1}{1-2t}$ if $t < \frac{1}{2}$ which is an open interval of 0, so *mgf* exists. $M'(t) = \frac{2}{(1-2t)^2}$ and $M''(t) = \frac{4.2}{(1-2t)^3},\ t < \frac{1}{2} \Rightarrow E(X) = 2,\ E(X^2) = 8$ and $Var(X) = 4$.

Exercise

i. A distribution with *pdf* $f(x) = e^{-x},\ 0 < x < \infty$ and 0 elsewhere has the *mgf* $M(t) = (1-t)^{-1},\ t < 1.$ Then $M'(t) = (1-t)^{-2},\ M''(t) = 2(1-t)^{-3}$ giving that $\mu = 1$ and $\sigma = 2 - 1 = 1.$

ii. Let $f(x) = \frac{1}{2}e^{-x/2}, x > 0$ and 0 otherwise. Then $M(t) = \frac{1}{1-2t}$ if $t < \frac{1}{2}$ which is an open interval of 0, so *mgf* exists. $M'(t) = \frac{2}{(1-2t)^2}$ and $M''(t) = \frac{4.2}{(1-2t)^3},\ t < \frac{1}{2} \Rightarrow E(X) = 2,\ E(X^2) = 8$ and $Var(X) = 4.$

iii. Let $f(x) = 1,\ 0 < x < 1.$ Then $M(t) = \frac{e^t - 1}{t}\ \forall t$ and $M'(t) = \frac{(e^t.t - e^t + 1)}{t^2} \Rightarrow M'(0) = \mu = \lim_{t\to 0} M'(t) = \frac{1}{2}.$