**MAT 3103: Statistics and Probability** **Semester: Summer 2018**

# Chapter 3: Random Variable and Probability Distribution

*Dr Atikur R. Khan (Co-ordinator), Atika Farzana Urmi, Md. Mortuza Ahmmed, Ashfia Tasnim Munia*

## 3.1 Random Variable

**Random variable** is a variable whose values are associated with some probabilities. Let us consider a coin tossing in a cricket match where the sample space is $S = \{H, T\}$. Now, consider that $X$ is a variable represents number of head ($H$) in a coin toss. Thus for this experiment $X$ has values 1 (number of $H$ in $S$) and 0 (second outcome does not have any $H$).

Table 3.1: Number of head in single coin tossing

| Outcome | $X$ = Number of $H$ | $p(X = x)$ = Probability of $X$ when it takes value $x$ |
|---|---|---|
| H | 1 | 0.5 |
| T | 0 | 0.5 |

So, the variable $X$ has values 1 and 0 with probability 0.5 for value 1 and 0.5 for value 0. Thus $X$ is a random variable and $p(x)$ is the probability function of the random variable $X$. Here $X$ takes only discrete values, and so it is a *discrete random variable*.

Now, let us consider sample space for two coins where the sample space is a set of 4 elements shown below.

| | H | T |
|---|---|---|
| H | HH | HT |
| T | TH | TT |

Table 3.2: Number of head in two coin tossing

| Outcome | $X$ = Number of $H$ | $p(X = x)$ = Probability of $X$ when it takes value $x$ |
|---|---|---|
| HH | 2 | 0.25 |
| $\{HT, TH\}$ | 1 | 0.50 |
| TT | 0 | 0.25 |

Thus $X$ takes values 0, 1, 2 with probabilities 0.25, 0.5, 0.25 respectively, and it is a random variable. Here $X$ takes only discrete values, and so it is a *discrete random variable*.

Probability function of discrete random variable must hold two properties:

[1]Prepared by Dr. Atikur R. Khan (atikur@aiub.edu).

(i) $0 \leq p(X = x) \leq 1$

(ii) $\sum_{x=0}^{\infty} p(X = x) = 1$

In the above experiment, $p(X = 0) = 0.25$, $p(X = 1) = 0.5$ and $p(X = 2) = 0.5$, and $\sum_{x=0}^{2} p(X = x) = 1$. *Thus $X$ is a discrete random variable with probability function $p(x)$.* And the pair $\{X, p(X = x)\}$ is known as probability distribution.

The random variable $X$ can be continuous too and its probability function $f(x)$ is known as *probability density function* and it holds similar properties:

(i) $0 \leq f(x) \leq 1$

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

## 3.2 Mean and Variance of Discrete Random Variable

The $r$th mean of a discrete random variable $X$ is

$$\mu_r' = E\left(X^r\right) = \sum_{x=lowest}^{highest} x^r p(x) \tag{3.1}$$

This is also known as $r$th raw moment. Now, when you put $r = 1$ then you get the first raw moment (which is the mean), when you put $r = 2$ you get the second raw moment (not variance, please keep that in mind), for $r = 3$ gives the third raw moment, and so on.

So, for $x$ and $p(x)$ values in Table 3.2, that is for probability distribution $\{x, p(x)\}$ in Table 3.2, we may get

$$\begin{aligned}
\mu_1' &= E\left(X\right) = \sum_{x=0}^{2} xp(x) = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1 \\
\mu_2' &= E\left(X^2\right) = \sum_{x=0}^{2} x^2 p(x) = 0^2 \times 0.25 + 1^2 \times 0.5 + 2^2 \times 0.25 = 1.5 \\
\mu_3' &= E\left(X^3\right) = \sum_{x=0}^{2} x^3 p(x) = 0^3 \times 0.25 + 1^3 \times 0.5 + 2^3 \times 0.25 = 2.5 \\
\mu_4' &= E\left(X^4\right) = \sum_{x=0}^{2} x^4 p(x) = 0^4 \times 0.25 + 1^4 \times 0.5 + 2^4 \times 0.25 = 4.5
\end{aligned}$$

Thus we get

$$\begin{aligned}
Mean &= \mu_1' = E\left(X\right) = 1 \\
Variance &= \mu_2' - \left(\mu_1'\right)^2 = E\left(X^2\right) - \left[E\left(X\right)\right]^2 = 1.5 - 1 = 0.5
\end{aligned}$$

### 3.2.1 Shape Characteristics of Distribution

By using raw moments we can calculate few other statistical values to learn about the shape of the distribution. At first, we calculate central moments ($\mu_r$) from raw moments ($\mu'_r$) where

$$\mu_r = \sum_{i=0}^{r} \binom{r}{i} (-1)^{r-i} \mu'_i \left(\mu'_1\right)^{r-i}; \text{ where } \mu'_0 = 1 \tag{3.2}$$

To obtain the second, third, and fourth central moments, we may write

$$\mu_2 = \sum_{i=0}^{2} \binom{2}{i} (-1)^{2-i} \mu'_i \left(\mu'_1\right)^{2-i} = \binom{2}{0} (-1)^{2-0} \mu'_0 \left(\mu'_1\right)^{2-0} + \binom{2}{1} (-1)^{2-1} \mu'_1 \left(\mu'_1\right)^{2-1} + \binom{2}{2} (-1)^{2-2} \mu'_2 \left(\mu'_1\right)^{2-2}$$
$$= \mu'_2 - (\mu'_1)^2$$
$$\mu_3 = \sum_{i=0}^{3} \binom{3}{i} (-1)^{3-i} \mu'_i \left(\mu'_1\right)^{3-i} = \mu'_3 - 3\mu'_2(\mu'_1) + 2(\mu'_1)^3$$
$$\mu_4 = \sum_{i=0}^{4} \binom{4}{i} (-1)^{4-i} \mu'_i \left(\mu'_1\right)^{4-i} = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu'_1)^4$$

Shape characteristics of a distribution is summarized by using skewness parameter ($\beta_1$) and curtosis parameter ($\beta_2$):

$$\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} \tag{3.3}$$

and

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \tag{3.4}$$

Thus for our coin tossing experiment, we get

$$\mu'_1 = 1; \text{ which is the mean}$$
$$\mu_2 = 0.5; \text{ which is the variance}$$
$$\mu_3 = 0$$
$$\mu_4 = 0.5$$
$$\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = 0; \text{ comment: symmetric shape based on decision rules described in the next section}$$
$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{0.5}{0.5^2} = 2; \text{ comment: platykurtic based on decision rules described in the next section}$$

### 3.2.2 Decision Rule for Skewness Coefficient

- $\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} > 0$: data are positively skewed or skewed to the right. This means that the right tail of the distribution is longer than the left.
- $\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = 0$: data are perfectly symmetrical (quite unlikely in real world)
- $\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} < 0$: data are negatively skewed (skewed to the left) meaning that the left tail is longer.

Following rules of thumbs (Bulmer, 1979) are also used to interprete statistical results:

- $\beta_1 < -1$ or $\beta_1 > +1$: highly skewed
- $-1 < \beta_1 < -\frac{1}{2}$ or $\frac{1}{2} < \beta_1 < +1$: moderately skewed
- $-\frac{1}{2} < \beta_1 < +\frac{1}{2}$: approximately symmetric

### 3.2.3 Decision Rule for Kurtosis Coefficient

Peter Westfall (2014) noted that higher kurtosis means more of the variance is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations. It is commonly used to say that a higher value indicates a higher and sharper peak, and a lower value indicates a lower and less distinct peak.

Following rules with reference to normal distribution are:

- $\beta_2 < 3$: platykurtic (compared to a normal distribution, its tails are shorter and thinner, and often its central peak is lower and broader).
- $\beta_2 > 3$: leptokurtic (compared to a normal distribution, its tails are longer and fatter, and often its central peak is higher and sharper.).
- $\beta_2 = 3$: mesokurtic.

## 3.3 Mean and Variance of Continuous Random Variable

Let us consider that $X$ is a random variable with probability function (for continuous random variable we call probability density function)

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \; x > 0, \lambda > 0 \tag{3.5}$$

To verify that $f(x)$ is a probability density function, we verify its properties:

(i) $0 \leq f(x) \leq 1$ for any values of $x > 0$ and $\lambda > 0$

(ii) $\int_0^\infty f(x)dx = \int_0^\infty \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = \left[-e^{-\frac{x}{\lambda}}\right]_0^\infty = 1$

Thus, $f(x)$ is a probability density function of continuous random variable $X$.

Now, the mean of $X$ is the first raw moment $\mu_1' = E(X)$ and can be computed as

$$\begin{aligned} E(X) &= \int_0^\infty x f(x) dx \\ &= \frac{1}{\lambda} \int_0^\infty x e^{-\frac{x}{\lambda}} dx \\ &= \frac{1}{\lambda} \left[ -\lambda x e^{-\frac{x}{\lambda}} \Big|_0^\infty - \int_0^\infty \left(-\lambda e^{-\frac{x}{\lambda}}\right) dx \right]; \text{ using integration by parts } \int uv dx = u \int v dx - \int u' \left( \int v dx \right) dx \\ &= 0 - \lambda e^{-\frac{x}{\lambda}} \Big|_0^\infty \\ &= \lambda \end{aligned}$$

The second raw moment $\mu_2' = E(X^2)$ can be computed as

$$\begin{aligned}
E\left(X^2\right) &= \int_0^\infty x^2 f(x)dx \\
&= \frac{1}{\lambda}\int_0^\infty x^2 e^{-\frac{x}{\lambda}}dx \\
&= \frac{1}{\lambda}\left[-\lambda x^2 e^{-\frac{x}{\lambda}}\Big|_0^\infty - \int_0^\infty 2x\left(-\lambda e^{-\frac{x}{\lambda}}\right)dx\right] \\
&= \frac{1}{\lambda}\left[-\lambda x^2 e^{-\frac{x}{\lambda}}\Big|_0^\infty + 2\lambda\int_0^\infty x\left(e^{-\frac{x}{\lambda}}\right)dx\right] \\
&= \frac{1}{\lambda}\left[-\lambda x^2 e^{-\frac{x}{\lambda}}\Big|_0^\infty + 2\lambda\left(-\lambda x e^{-\frac{x}{\lambda}}\Big|_0^\infty - \int_0^\infty \left(-\lambda e^{-\frac{x}{\lambda}}\right)dx\right)\right] \\
&= \frac{1}{\lambda}\left[-\lambda x^2 e^{-\frac{x}{\lambda}}\Big|_0^\infty + 2\lambda\left(-\lambda x e^{-\frac{x}{\lambda}}\Big|_0^\infty - \lambda^2 e^{-\frac{x}{\lambda}}\Big|_0^\infty\right)\right] \\
&= 2\lambda^2
\end{aligned}$$

Thus the variance of $X$ is

$$V(X) = E(X^2) - [E(X)]^2 = 2\lambda^2 - \lambda^2 = \lambda^2$$

We know $\mu_1' = E(X)$ and $\mu_2' = E(X^2)$. Thus, we need to find other higher ordered moments by integrating as $\mu_r' = E(X^r) = \int_0^\infty x^r f(x)dx$. Alternatively, we can use moment generating functions to calculate $\mu_r'$ to learn about shapes of this distribution.

## 3.4 Moment Generating Function of Random Variable

Moment generating function (MGF) of random variable $X$ is defined as

$$M_X(t) = E(e^{tX}) = \sum e^{tx}p(x); \text{ when } X \text{ is a discrete random variable} \tag{3.6}$$

$$= \int e^{tx}f(x)dx; \text{ when } X \text{ is a continuous random variable} \tag{3.7}$$

Since

$$e^{tX} = 1 + tX + \frac{t^2}{2!}X^2 + \frac{t^3}{3!}X^3 + \frac{t^4}{4!}X^4 + \dots \tag{3.8}$$

$$E\left(e^{tX}\right) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \frac{t^4}{4!}E(X^4) + \dots \tag{3.9}$$

$$M_X(t) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \frac{t^4}{4!}\mu_4' + \dots \tag{3.10}$$

Thus the $r$th raw moment is

$$\mu_r' = \text{Coefficient of } \frac{t^r}{r!} \text{ in } M_X(t) \tag{3.11}$$

$$= \frac{d^r M_X(t)}{dt^r}\Big|_{t=0} \tag{3.12}$$

Please refer to Chapter 4 (Discrete Probability Distribution) and Chapter 5 (Continuous Probability Distribution) for finding moment generating functions and moments of several distributions.

## 3.5 Mean and Variance Under Transformation

Let us assume that $X$ is a random variable with mean $\eta$ and variance $\sigma^2$. Then mean of random variable $Y = a + bX$ is

$$E(Y) = E(a + bX) = a + bE(X) = a + b\eta$$

and the variance of random variable $Y$ is

$$V(Y) = V(a + bX) = V(bX) = b^2 V(X)$$

In the coin tossing experiment, we got the mean $E(X) = 1$ and variance $V(X) = 0.5$. Thus, for that experiment, we may write

$$\begin{aligned}
E(2 + X) &= 2 + E(X) = 2 + 1 = 3, \text{ since } E(constant) = constant \\
E(5X) &= 5E(X) = 5 \times 1 = 5 \\
E(2 + 5X) &= 2 + 5E(X) = 2 + 5 = 7 \\
E(2 - 5X) &= 2 - 5E(X) = 2 - 5 = -3 \\
V(2 + X) &= V(X) = 0.5 \\
V(2 + 5X) &= V(5X) = 5^2 V(X) = 25 \times 0.5 = 12.5 \\
V(2 - 5X) &= V(-5X) = (-5)^2 V(X) = 25 \times 0.5 = 12.5
\end{aligned}$$

## 3.6 Computing Probabilities from Probability Distribution

Probability function for discrete and continuous random variables are known as probability mass function and probability density function, respectively. To compute probability for given range (or values) of random variable $X$:

- add probabilities corresponding to desired values of $X$ when $X$ is a discrete random variable
- integrate the density function over a desired range of values when $X$ is a continuous random variable.

Table 3.3: Computing probabilities

| Probability | Discrete<br>$X$: 0, 1, 2<br>$P(X)$: 0.25, 0.5, 0.25 | Continuous<br>$f(x) = 2e^{-2x},\ x > 0$ |
|---|---|---|
| $P(X > 0)$ | $P(X = 1) + P(X = 2) = 0.5 + 0.25$ | $\int_0^\infty f(x)dx = -e^{-2x}\Big\vert_0^\infty = 1$ |
| $P(X \leq 1)$ | $P(X = 0) + P(X = 1) = 0.25 + 0.5$ | $\int_0^1 f(x)dx = -e^{-2x}\Big\vert_0^1 = 1 - e^{-2}$ |
| $P(1 < X \leq 2)$ | $P(X = 2) = 0.25$ | $\int_1^2 f(x)dx = -e^{-2x}\Big\vert_1^2 = e^{-2} - e^{-4}$ |

## 3.7 Joint and Marginal Probability Density Functions

Joint density function of two random variables $X$ and $Y$ can be written as

$$f(x,y) = \frac{xy}{4}, 0 < x < 4, 0 < y < 1 \tag{3.13}$$

Let us verify following two properties

(i) $0 \leq f(x,y) \leq 1$

(ii) $\int_x \left( \int_y f(x,y) dy \right) dx = 1$

for the above mentioned joint density function. The first property is obvious and even does not require any high school math! Let us show the second property holds true.

We can write

$$\int_y f(x,y)dy = \int_0^1 \frac{xy}{4} dy = \frac{x}{4} \frac{y^2}{2} \Big|_{y=0}^{1} = \frac{x}{8}$$

Now, integrating with respect to $x$ we may write

$$\int_x \left( \int_y f(x,y)dy \right) dx = \int_0^4 \frac{x}{8} dx = \frac{x^2}{16} \Big|_0^4 = 1$$

### 3.7.1 Marginal Density Function

Marginal density function of random variable $X$ can obtained from a joint density function by integrating with respect to other random variables. Thus the marginal density function of $X$ in this case is

$$g(x) = \int_y f(x,y)dy = \int_0^1 \frac{xy}{4} dy = \frac{x}{8};\ 0 < x < 4$$

Similarly, the marginal density function for $Y$ is

$$h(y) = \int_x f(x,y)dx = \int_0^4 \frac{xy}{4} dx = 2y;\ 0 < y < 1$$

### 3.7.2 Mean and Variance

Mean and variance of random variable $X$ are

$$\begin{aligned}
E(X) &= \int_x xg(x)dx = \int_0^4 x\frac{x}{8} dx = \int_0^4 \frac{x^2}{8} dx = \frac{x^3}{24} \Big|_0^4 = \frac{64}{24} \\
E(X^2) &= \int_x x^2 g(x)dx = \int_0^4 \frac{x^3}{8} dx = \frac{x^4}{32} \Big|_0^4 = 8 \\
V(X) &= 8 - \left( \frac{64}{24} \right)^2
\end{aligned}$$

Similarly, mean and variance of $Y$ are

$$E(Y) = \int_y yh(y)dy = \int_0^1 2y^2 dy = \frac{2y^3}{3}\Big|_0^1 = \frac{2}{3}$$
$$E(Y^2) = \int_y y^2 h(y)dy = \int_0^1 2y^3 dy = \frac{2y^4}{4}\Big|_0^1 = \frac{1}{2}$$
$$V(Y) = \frac{1}{2} - \left(\frac{2}{3}\right)^2$$

### 3.7.3 Independence of Random Variables

Two random variables $X$ and $Y$ are independent if

$$f(x,y) = g(x)h(y)$$

For the joint density function in (3.13) we can show that

$$f(x,y) = \frac{xy}{4}, 0 < x < 4, 0 < y < 1$$
$$g(x) = \frac{x}{8}, 0 < x < 4 \quad \text{which is the marginal density function of } X$$
$$h(y) = 2y, 0 < y < 1 \quad \text{which is the marginal density function of } Y$$
$$f(x,y) = g(x)h(y)$$

Thus these random variables are independent.

### 3.7.4 Covariance Between Random Variables

Covariance between two random variables $X$ and $Y$ is

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Previously we have found $E(X)$ and $E(Y)$ from the joint density function $f(x,y)$. Now, we find

$$E(XY) = \int_x \int_y xyf(x,y)dydx = \int_x \left(\int_y xyf(x,y)dy\right) dx = \int_0^4 \left(\int_0^1 \frac{x^2y^2}{4} dy\right) dx = \int_0^4 \left(\frac{x^2}{12}\right) dx = \frac{64}{36}$$

Thus the covariance between $X$ and $Y$ under the joint density function in (3.13) is

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{64}{36} - \frac{64}{24} \times \frac{2}{3} = 0$$

Therefore, two random variables $X$ and $Y$ having joint density function as in (3.13) are independent, and their covariance is equal to zero.

## 3.8 Joint and Marginal Probability Mass Functions

Let us consider the two coin tossing experiment discussed earlier. Sample space for this experiment is

| | H | T |
|---|---|---|
| H | HH | HT |
| T | TH | TT |

Let $X$ is a random variable representing the number of head (H) and $Y$ is a random variable representing the number of tail (T). Then both $X$ and $Y$ can take values $0, 1, 2$. From the sample space it is clear that number of head plus number of tail is always two in this experiment, that is, $X + Y = 2$. Thus, values of $X$ depends on values of $Y$ and vice-versa, and so they are not independent. Let us show this by computing marginal probabilities for both random variables.

Joint probabilities $p(x, y)$ for $x = 0, 1, 2$ and $y = 0, 1, 2$ would form a 3 ordered matrix and those probabilities are provided in the following table.

Table 3.4: Joint probability $p(X = x, Y = y)$ where $x$ is number of heads and $y$ is number of tails

| $Y$ \ $X$ | 0 | 1 | 2 | $h(y) = \sum_x p(x, y)$ |
|---|---|---|---|---|
| 0 | 0 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| 1 | 0 | $\frac{2}{4}$ | 0 | $\frac{2}{4}$ |
| 2 | $\frac{1}{4}$ | 0 | 0 | $\frac{1}{4}$ |
| $g(x) = \sum_y p(x, y)$ | $\frac{1}{4}$ | $\frac{2}{4}$ | $\frac{1}{4}$ | $\sum_x \sum_y p(x, y) = \sum_x g(x) = \sum_y h(y) = 1$ |

Marginal probability mass function of $X$ is $g(x) = \sum_y p(x, y)$ and corresponding marginal probabilities are computed in the last row.

Marginal probability mass function of $Y$ is $h(y) = \sum_x p(x, y)$ and corresponding marginal probabilities are computed in the last column.

$$
\begin{aligned}
p(X = 0, Y = 0) = 0, &\text{ but } g(X = 0)h(Y = 0) = \frac{1}{4} \times \frac{1}{4} \\
p(X = 0, Y = 1) = 0, &\text{ but } g(X = 0)h(Y = 1) = \frac{1}{4} \times \frac{2}{4} \\
p(X = 0, Y = 2) = \frac{1}{4}, &\text{ but } g(X = 0)h(Y = 2) = \frac{1}{4} \times \frac{1}{4} \\
&\vdots \\
p(X = 2, Y = 2) = 0, &\text{ but } g(X = 0)h(Y = 2) = \frac{1}{4} \times \frac{1}{4}
\end{aligned}
$$

Thus we can say that for this joint probability mass function $p(X = x, Y = y) \neq g(X = x)h(Y = y)$, and these discrete random variables $X$ and $Y$ are not independent.

## 3.9 Exercises

1. In three coin tossing experiment, show that the number of head $X$ is a random variable.
   (i) Compute mean and variance of random variable $X$
   (ii) Compute skewness and kurtosis coefficients, and comment on the shape of its distribution
   (iii) Calculate $p(X > 2)$, $p(X \geq 2)$, $p(X < 1 \text{ or } X > 2)$, and $p(2 \leq X < 3)$
   (iv) calculate $E(2 + 3X)$, $V(2 + 3X)$ and $V(2 - 3X)$

2. From the probability density function

$$f(x) = \lambda e^{-\lambda x}, x > 0, \lambda > 0$$

   (i) calculate $p(X < 1)$, $p(X > 1)$, $p(1 < X < 1.5)$, and $p(X < 0.5 \text{ or } X > 1.5)$
   (ii) calculate $V(2 - 3X)$

3. Construct sample space for tossing three coins. Assuming $X$ represents number of head and $Y$ represents number of tail in this coin tossing experiment,

   (i) find joint probability mass function $p(x, y)$
   (ii) show that variables are not independent
   (iii) calculate $p(X > 2)$, $p(X \geq 2)$, $p(X < 1 \text{ or } X > 2)$, and $p(2 \leq X < 3)$
   (iv) calculate $E(a + bX)$, $V(a + bX)$ and $V(a - bX)$

4. From the joint probability density function

$$f(x, y) = \frac{x}{2}, 0 < x < 2, 0 < y < 1$$

   (i) find marginal density functions
   (ii) calculate $p(X < 1)$, $p(X > 1)$, $p(1 < X < 1.5)$, and $p(X < 0.5 \text{ or } X > 1.5)$
   (iii) calculate $V(2X - 3Y)$