Random Variables

CSGE602013 – Statistics and Probability

Fakultas Ilmu Komputer

Universitas Indonesia

2016

Credits

These course slides were prepared by Alfan F. Wicaksono. **Suggestions**, **comments**, and **criticism** regarding these slides are welcome. Please kindly send your inquiries to alfan@cs.ui.ac.id.

The content was based on previous semester's (odd semester 2013/2014) course slides created by **all previous lecturers.**

References

- Introduction to Probability and Statistics for Engineers & Scientists, 4th ed.,
  - Sheldon M. Ross, Elsevier, 2009.

- A Modern Introduction to Probability and Statistics, Understanding Why and How, Frederik Michel Dekking et al., Springer, 2005.

Outline

- Random Variables
- Discrete Random Variables
- Continuous Random Variables
- Jointly Ditributed Random Variables

Random Variables

We are often NOT interested in all of the details of the experimental result but only in the value of some numerical quantity determined by the result.

In tossing dice…

We are often interested in the sum of the two dice !

We don't care about the values of the individual dice !

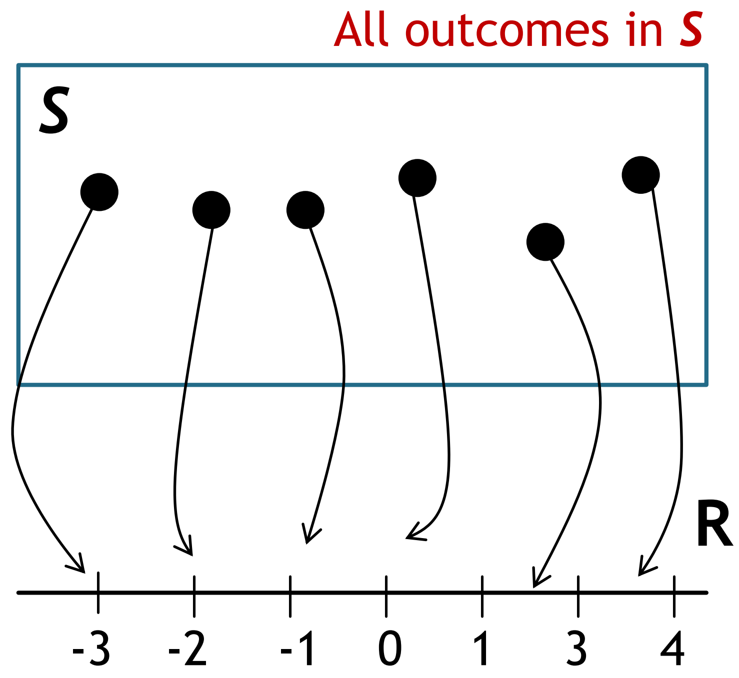
Random Variable

A numerical value to each outcome of a particular experiment.

Quantities of interest that are determined by the result of the experiment.

A Random Variable X, can be thought of as a **function** that assign a numerical value **X(s)** to each possible outcome in an experiment.

$$X : S \to R \qquad (or\ X(s) \in R, \forall s \in S)$$



An RV provides us the power of abstraction and allows us to discard unimportant details of outcomes in an experiment.

Example

Tossing **3** coins → 1:Head, 0:Tail

S = {(0,0,0), (0,0,1), …, (1,1,1)}

If we define random variable **X** as the number of occurrences of Head in a trial.

X(0,0,0) = 0

X(0,0,1) = **X**(0,1,0) = **X**(1,0,0) = 1

X(0,1,1) = **X**(1,0,1) = **X**(1,1,0) = 2

X(1,1,1) = 3

Types of Random Variables

Discrete Random Variable

- Takes only finite or countably infinite number of values
- Arise from discrete event measurement
- Set of possible values is a sequence

Continuous Random Variable

- Takes infinite and uncountable number of values
- Pertinent to continuous measurement
- Ex: an interval

Discrete Random Variable

Letting **X** denote the random variable that is defined as the sum of two fair dice, then

P(**X** = 2) = **P**({(1,1)}) = 1/36

P(**X** = 3) = **P**({(1,2), (2,1)}) = 2/36

P(**X** = 4) = **P**({(1,3), (2,2), (3,1)}) = 3/36

P(**X** = 5) = **P**({(1,4), (2,3), (3,2), (4,1)}) = 4/36

…

P(**X** = 11) = **P**({(5,6), (6,5)}) = 2/36

P(**X** = 12) = **P**({(6,6)}) = 1/36

X is a discrete random variable.

Probability Mass Function (PMF)

Probability mass function $p(x_i)$ of **X** is defined by

$$p(x_i) = P(X = x_i)$$

P(***X*** = x_i) : the probability that the value of **X** is equal to x_i.

PMF of a random var. must follow **THREE** conditions:

First condition

If **X** must assume one of the values x_1, x_2, …, then,

$$\begin{aligned} &0 \le p(x) \le 1 && \textit{for all real } x \\ &p(x_i) > 0, && i = 1,2,\ldots \\ &p(x) = 0, && \textit{all other values of } x \end{aligned}$$

Cont'd->

Second condition

Set of the values x_1, x_2, … can be **finite** or **countably infinite.**

Third condition

and the following property must hold

$$\sum_{x_i} P(x_i) = 1$$

or, if **X** takes value from **finite** set $\{x_1, x_2, \ldots, x_n\}$, then

$$\sum_{i=1}^{n} p(x_i) = 1$$

since

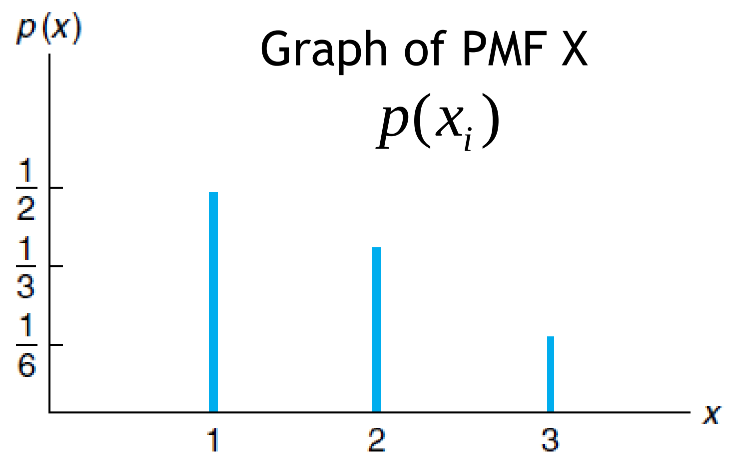
$$1 = P(S) = P\left(\bigcup_{i=1}^{n} X = x_i\right) = \sum_{i=1}^{n} P(X = x_i)$$

Example of PMF

Consider a random variable **X** that is equal to 1, 2, or 3.

Suppose, we are given the **PMF of X** as follows:

***P*(X = 1) = *p*(1) = 1/2**
***P*(X = 2) = *p*(2) = 1/3**
***P*(X = 3) = *p*(3) = 1/6**



The **PMF of X** can also be written as follows:

x_i	1	2	3
$p(x_i)$	1/2	1/3	1/6

Remember !
The following equation must hold:

$$\sum_{i=1}^{3} P(X = x_i) = \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1$$

Cumulative Distribution Function (CDF)

Cumulative Distribution Function, ***F*** of the random variable ***X*** is defined for any real number **x** by

$$F_X(x) = P(X \leq x) \qquad x \in R$$

How to determine $P(a < X \leq b)$?

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

$$\begin{aligned} P(a < X \leq b) & \\ &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \end{aligned}$$

CDF is also called by "**distribution function**"

Cumulative Distribution Function (CDF)

Basic properties of CDF

1. CDF is a monotone increasing function of X

$$F_X(t_1) \le F_X(t_2), \qquad t_1 \le t_2$$

2.

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to +\infty} F_X(x) = 1$$

3. CDF is right-continuous

$$\lim_{\varepsilon \to 0^+} F_X(x + \varepsilon) = F_X(x)$$

Cumulative Distribution Function for Discrete R.V.

By definition of CDF:

$$F(t) = P(X \le t)$$

In discrete case, it becomes

$$\begin{aligned} F(t) &= P(X \le t) \\ &= \sum_{x_i \le t} P(X = x_i) \\ &= \sum_{x_i \le t} p(x_i) \end{aligned}$$

*we omit subscript **X** on $\mathbf{F_X}$(t) since it's clear here that ***F***(t) is CDF of RV **X**

Example on How to Create CDF of Discrete R.V.

Suppose, we have discrete R.V. **X** and its Probability Mass Function (PMF) as follows:

x_i	1	2	3
$p(x_i)$	1/2	1/3	1/6

Determine the Cumulative Distribution Function (CDF) of random variable **X** !

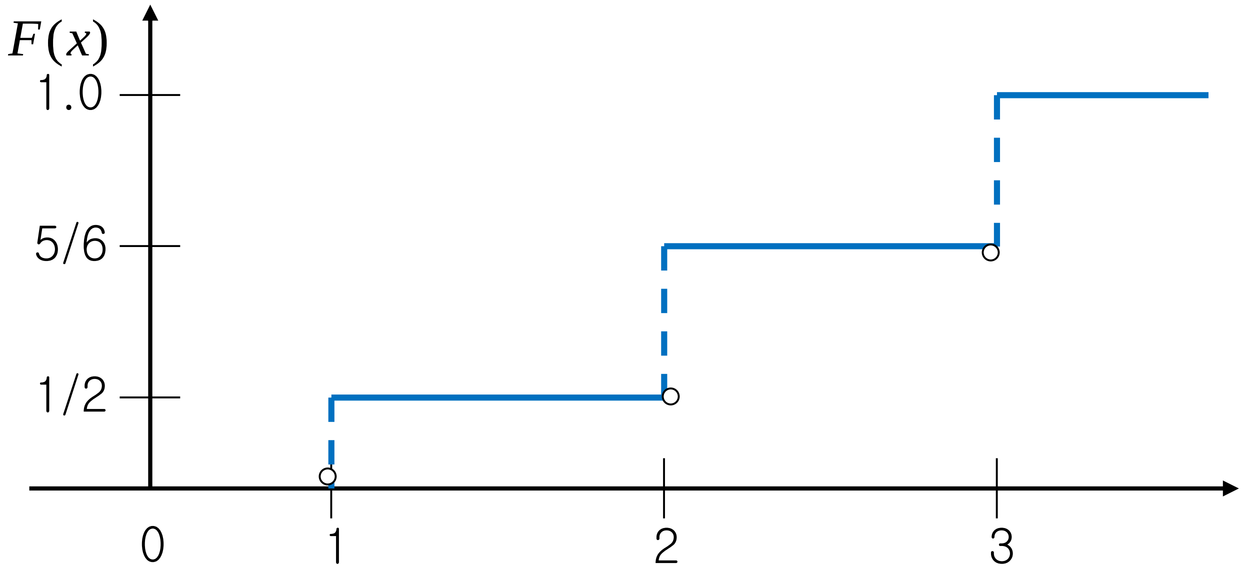
Solution:

$$F_X(a) = \begin{cases} 0 & a < 1 \\ 1/2 & 1 \le a < 2 \\ 1/2 + 1/3 = 5/6 & 2 \le a < 3 \\ 1/2 + 1/3 + 1/6 = 1 & a \ge 3 \end{cases}$$

Although **X** is discrete, CDF must be defined on **real values.**

Example on How to Create CDF of Discrete R.V. (Cont'd)

Draw the **graph** of CDF of X !



- CDF is a **step-function** !
- Even though **X** is a **discrete** R.V., CDF must be defined on **real values.**

For Dicrete R.V., $\mathbf{F}_X(t)$ grows only by jumps in discrete steps

$$F_X(t) = F_X(t_i), \qquad for\ t_i \le t < t_{i+1}$$

$$F_X(t_{i+1}) = F_X(t_i) + P(X = t_{i+1})$$

Hence, PMF can be obtained from CDF

$$P(X = x_k) = p_X(x_k) = F_X(x_k) - F_X(x_{k-1})$$

Continuous Random Variable

Example 1

X = life time of a light bulb

X can take any value from positive real number.

Normally, we are not interested in **P(X = k)** for some specific value of **k**; Instead, we are interested in

- Probability within a range $P(a \le X \le b)$, a, b $\in \mathbf{R}^+$
- CDF: $F(x) = P(X \le x)$, x $\in \mathbf{R}^+$

Example 2

Y is the diameter of a randomly chosen cylinder manufactured by the company.

Y can take any value between 49.5 and 50.5.

X and Y are continuous random variable.

Probability Density Function (PDF)

We say that **X** is a continuous random variable if there exists a non-negative function $\boldsymbol{f_x}$**(x)**, defined for all real **x ϵ (-~, ~)**, having property that for any set **B** of real numbers

$$P(X \in B) = \int_B f_x(x)\, dx$$

The function $\boldsymbol{f_x}$**(x)** is called the *probability density function* of random variable **X**.

*subscript x on $\boldsymbol{f_x}$**(x)** can be omitted if it's clear that ***f*****(x)** is PDF of random variable X

***f*(x)** must satisfy

$$1 = P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f_x(x)\, dx$$

Letting **B** = [a, b], we obtain

$$P(a \le X \le b) = \int_{a}^{b} f_x(x)\, dx$$

probability within a range a ≤ x ≤ b

The required probability is the area under the curve f(x) between a and b !

Please remember this !

If we let a = b in the previous equation, then

$$P(X = a) = \int_a^a f_x(x)\, dx = 0$$

In Words, this states that the probability that a continuous random variable will assume any particular value is **zero**.

We can say, it's extremely small ! (considered as zero).

Properties of PDF (Must Hold for a valid PDF)

1. Unlike PMF, since PDF is not a probability, the values of each function can be more than 1.

$$f_x(x) \geq 0, \qquad x \in R$$

2. Since continuous RV maps sample space into an **uncountably** infinite set of real numbers, then its probability has properties:

$$0 \leq \int_a^b f_X(x)dx \leq 1, \qquad \text{for } -\infty \leq a,b \leq \infty$$

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

Problem 1

Let RV **X** denote the diameter of a hole drilled in a sheet metal component. Historical data show that the distribution of **X** can be modeled by a PDF

$$f_X(x) = \begin{cases} 20e^{-20(x-12.5)} & x \geq 12.5 \\ 0 & otherwise \end{cases}$$

(a) Is this a valid PDF ?

(b) What is the probability that a part with a diameter larger than 12.60 mm is scrapped ?

(c) What is the probability that a part with a diameter between 12.5 and 12.6 mm ?

(a) This is a valid PDF since

(1) that exp. function is always positive $f_x(x) \geq 0, \quad x \geq 12.5$

(2)

$$\int_{12.5}^{\infty} f_X(x)\, dx = \lim_{b\to\infty} \int_{12.5}^{b} 20e^{-20(x-12.5)}\, dx = \lim_{b\to\infty}\left(-e^{-20(x-12.5)} \Big|_{12.5}^{b} \right) = 1$$

(b) $P(X > 12.60) = ?$

$$P(X > 12.6) = \int_{12.6}^{\infty} f_X(x)\, dx = \lim_{b\to\infty} \int_{12.6}^{b} 20e^{-20(x-12.5)}\, dx$$

$$= \lim_{b\to\infty}\left(-e^{-20(x-12.5)} \Big|_{12.6}^{b} \right)$$

$$= 0.135$$

(c) $P(12.5 < X < 12.60) = ?$ **Solve the problem (c) !**

Problem 2

Suppose that **X** is a continuous random variable whose PDF is given by

$$f_X(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & otherwise \end{cases}$$

(a) What is the value of C ?

(b) Find ***P*(X** > 1) ?

(a) The value of **C** so that the PDF is valid:

It can be easily shown that **f(x)**

$$f(x) = C\left(4x - 2x^2\right) = C.x.(4 - 2x)$$

is **≥ 0** in the interval 0 < x < 2, **if**

$$C \geq 0$$

So, we need to use the 2nd property of PDF to find the exact value of C:

$$C\int_0^2 (4x - 2x^2)\, dx = C\left(2x^2 - \frac{2x^3}{3}\right)\Bigg|_0^2 = 1$$

$$hence, \quad C = \frac{3}{8}$$

(b) ***P***(**X** > 1) = ? **Try to solve problem (b) !**

Cumulative Distribution Function for Continuous R.V.

- $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(y)\, dy \qquad x \in \Re$

Differentiating both sides yields:

- $f_X(x) = \dfrac{dF_X(x)}{dx}$

Other identities:

- $P(a < X \le b) = P(X \le b) - P(X \le a) = F_X(b) - F_X(a)$
- $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$

The last one is ONLY for continuous case !

Problem 3

From **problem 2**, we have PDF of continuous random variable **X:**

$$f_X(x) = \begin{cases} 20e^{-20(x-12.5)} & x \geq 12.5 \\ 0 & otherwise \end{cases}$$

(a) Determine CDF of **X** ! (that is, Find ***F***(x)) !

(b) Find ***P***(12.5 < X < 12.6) using ***F***(x) !

By definition,

$$F(x) = P(X \le x)$$

$$= \int_{-\infty}^{x} 20e^{-20(y-12.5)}\, dy$$

$$= \int_{12.5}^{x} 20e^{-20(y-12.5)}\, dy$$

$$= -e^{-20(y-12.5)}\Big|_{12.5}^{x} = 1 - e^{-20(x-12.5)}$$

So, the complete answer for **F(x)** is:

$$F(x) = \begin{cases} 0 & x < 12.5 \\ 1 - e^{-20(x-12.5)} & x \ge 12.5 \end{cases}$$

Because CDF must be defined **on real values**

$$P(12.5 < X < 12.6)$$

since **X** is **continuous** RV

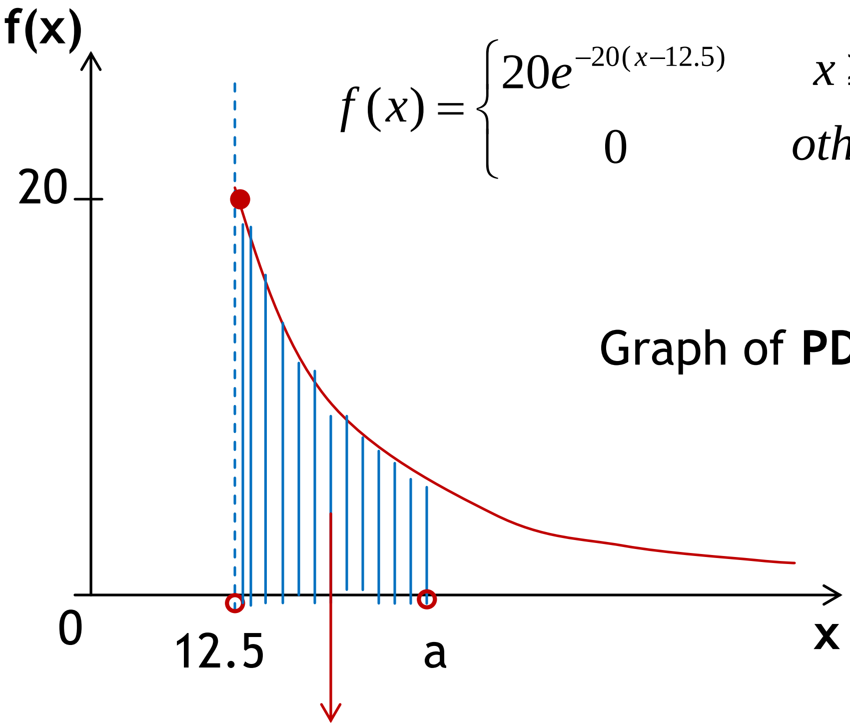
$$= P(12.5 < X \le 12.6)$$

$$= F(12.6) - F(12.5)$$

$$= 1 - e^{-20(12.6-12.5)} - (1 - e^{-20(12.5-12.5)})$$

$$= 1 - 0.135 = 0.865$$

$$f(x)=\begin{cases}20e^{-20(x-12.5)} & x\ge 12.5\\ 0 & otherwise\end{cases}$$



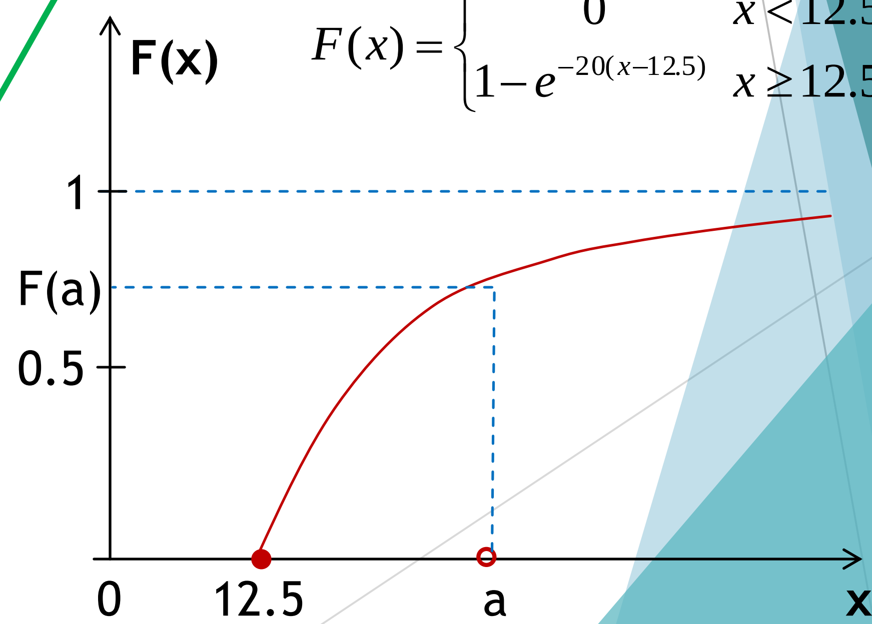
Graph of **PDF**

This area is **F(a)**

$$F(a)=1-e^{-20(a-12.5)}$$

Graph of **CDF**

$$F(x)=\begin{cases}0 & x<12.5\\ 1-e^{-20(x-12.5)} & x\ge 12.5\end{cases}$$



Jointly Distributed Random Variables

We are often interested in the relationships between two or more random variables.

In an experiment into the possible causes of cancer, we might be interested in the relationship between

X: average number of cigarettes smoked daily

Y: age at which an individual contracts cancer

An engineer might be interested in the relationship between

X: the shear strength

Y: diameter of a spot weld in a fabricated sheet steel specimen

Let **X** and **Y** are random variables, **joint cumulative distribution function** of **X** and **Y** is defined as follow

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

If we know ***F***(x, y), we can theoritically compute the probability of any statement concerning the values of **X** and **Y**

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(X \leq x, Y < \infty) = F(x, \infty) \end{aligned}$$

Similarly,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X < \infty, Y \leq y) = F(\infty, y) \end{aligned}$$

For Discrete R.V.

We define the joint probability mass function of **X** and **Y**

$$p(x_i, y_j) = P(X = x_i, Y = y_j) \geq 0$$

satisfying $$\sum_i \sum_j p(x_i, y_j) = 1$$

For Discrete R.V.

Marginal Probability Distribution

Knowing ***P***(**X** = x_i, **Y** = y_j) can determine ***P***(**X** = x_i) & ***P***(**Y** = y_j)

Obtained by summing the joint probability distribution over the values of the other random variable.

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p(x_i, x_j)$$

$$P(Y = y_i) = \sum_i P(X = x_i, Y = y_j) = \sum_i p(x_i, x_j)$$

For Discrete R.V.

Joint Cumulative Distribution Function

By definition $$F_{XY}(x, y) = P(X \le x, Y \le y)$$

In discrete case, it becomes

$$\begin{aligned} F_{XY}(a,b) &= P(X \le a, Y \le b) \\ &= \sum_{i:x_i \le a} \sum_{j:y_j \le b} P(X = x_i, Y = y_j) \\ &= \sum_{i:x_i \le a} \sum_{j:y_j \le b} p(x_i, y_j) \end{aligned}$$

For Discrete R.V.

Example

Suppose that 3 batteries are randomly chosen from

- A group of 3 new,
- A group of 4 used but still working, and
- A group of 5 defective batteries.

If we let random variables **X & Y:**

X: the number of new batteries that are chosen

Y: the number of used but still working batteries that are chosen

For Discrete R.V.

The joint PMF of **X** and **Y**, $p(i, j) = P(X = i, Y = j)$ is given by

$i \backslash j$	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 Sums $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

(a) Find $F(1, 1)$!

(b) Find $P(Y = 2)$!

For Continuous R. V.

R.V. **X** and **Y** are jointly continuous if there exists a function ***f***(x, y), having property:

$$f_{XY}(x, y) \geq 0$$

For every set **C** of pairs of real numbers

$$P((X, Y) \in C) = \iint_{(x,y)\in C} f(x, y)\, dx\, dy$$

f(x, y) is called the joint probability density function.

For Continuous R. V.

If **A** and **B** are any sets of real numbers, then by defining

$$C = \{(x, y) \mid x \in A, x \in B\}$$

we have

$$P(X \in A, Y \in B) = \int_B \int_A f(x, y)\, dx\, dy$$

Also, we have

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy$$

For Continuous R. V.

Joint Cumulative Distribution Function

By definition $$F_{XY}(x,y) = P(X \le x, Y \le y)$$

In continuous case, it becomes

$$\begin{aligned} F_{XY}(a,b) &= P(X \le a, Y \le b) \\ &= P(X \in (-\infty, a], Y \in (-\infty, b]) \\ &= \int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)\,dx\,dy \end{aligned}$$

We also have

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

For Continuous R. V.

Marginal Probability Distribution

Obtained by integrating the joint probability distribution over the values of the other random variable.

$$P(X \in A) = P(X \in A, Y \in (-\infty, \infty))$$

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

$$= \int_A f_X(x)\,dx$$

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

Similarly,

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

For Continuous R. V.

Practice Problem

The joint density function of **X** and **Y** is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & otherwise \end{cases}$$

(a) Find P(X > 1, Y < 1)

(b) Find P(X < Y)

(c) F_X (a) = ?

(d) F_{XY} (a, b) = ?

For Continuous R.V.

(a) $$P(X > 1, Y < 1) = \int_0^1 \int_1^\infty 2e^{-x} e^{-2y} \, dx \, dy = e^{-1}(1 - e^{-2})$$

(b) $$P(X < Y) = P(X < Y, Y > 0) = \int_0^\infty \int_0^y 2e^{-x} e^{-2y} \, dx \, dy$$

(c) $$F_X(a) = P(X < a) = P(X < a, Y > 0) = \int_0^a \int_0^\infty 2e^{-x} e^{-2y} \, dy \, dx$$

(d) $$F_{XY}(a,b) = \int_0^a \int_0^b 2e^{-x} e^{-2y} \, dy \, dx$$

Joint probability dist. For *n* random variables

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

In discrete case:

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

In continuous case:

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n)$$

$$= \int_{A_n} \int_{A_{n-1}} \ldots \int_{A_1} f(x_1, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n$$

Independent Random Variables

DEFINITION

The random variables **X** and **Y** are said to be **independent** **if and only if**

In discrete case:

$$p(x, y) = p_X(x)\, p_Y(y) = P(X = x).P(Y = y)$$

In continuous case:

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

Independent Random Variables

Important consequences

Joint cumulative distribution function of **X** and **Y**

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

And, we have (for discrete & continuous case)

$$P(X \in A, Y \in B) = P(X \in A).P(Y \in B)$$

Prove this ☺ !

Independent Random Variables

If there are ***n*** *independent* random variables ...

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

Practice Problem 1

Suppose **X** and **Y** are independent R.V. having joint density func.

$$f(x,y) = \begin{cases} e^{-(x+y)} & x \geq 0, y \geq 0 \\ 0 & otherwise \end{cases}$$

Find PDF of ***X*** and ***Y*** ! Are ***X*** & ***Y*** independent ?

Independent Random Variables

Practice Problem 2

Suppose **X** and **Y** are **independent** R.V. having common density func.

$$f_X(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & otherwise \end{cases} \qquad f_Y(y) = \begin{cases} e^{-y} & y > 0 \\ 0 & otherwise \end{cases}$$

Find density function of random variable **X/Y** !

Conditional Distributions

Discrete Case

If **X** and **Y** are discrete random variables, we define the conditional probability mass function of **X** given that **Y = y**, by $p_{X|Y}(x \mid y) = P(X = x \mid Y = y)$

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

$$= \frac{p(x, y)}{p_Y(y)} \qquad p_Y(y) > 0$$

Conditional Distributions

Continuous Case

If **X** and **Y** have a joint probability density function ***f*(x, y)**, then the conditional probability density function of **X**, given that **Y = y** is

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} \qquad f_Y(y) > 0$$

So,

$$P(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)\, dx$$

Conditional Distributions

Example 1

Suppose that ***p*(x, y)**, the joint probability mass function of **X** and **Y**, is given by

$$p(0,0) = 0.4 \qquad p(0,1) = 0.2 \qquad p(1,0) = 0.1 \qquad p(1,1) = 0.3$$

Calculate the conditional probability mass function of **X** given that **Y = 1** !

Conditional Distributions

Example 2

The joint density of **X** and **Y** is given by

$$f(x, y) = \begin{cases} \dfrac{12}{5} x(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & otherwise \end{cases}$$

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

Expectation

Expectation of a discrete random variable **X**

$$E[X] = \sum_i x_i \, P(X = x_i) = \sum_i x_i \, p_X(x_i)$$

Expectation of a continuous random variable **X**

$$E[X] = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

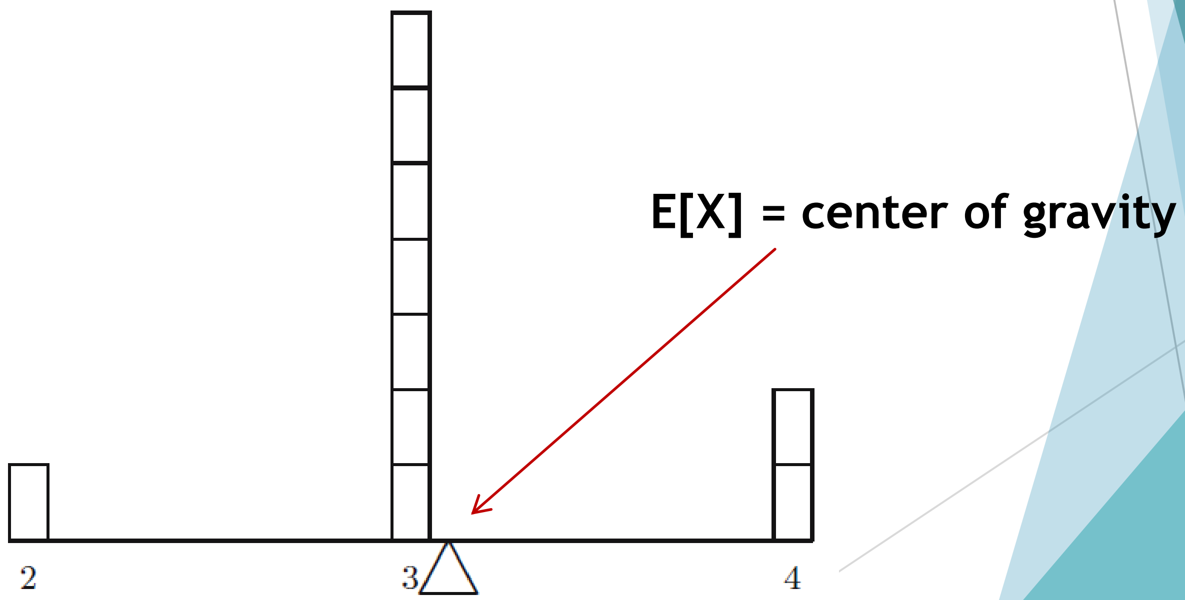
The *expectation* or *expected value* of random variable **X** is also called **mean** of the random variable **X** or **μ**.

Expectation = Mean = *Harapan* = *Rataan*

- **E[X]** acts as a representative number to describe the random variable **X**.
- What do we expect of the random variable **X** ?

In discrete case, you can think **E[X]** as **center of gravity** !

PMF:
P(X = 2) = 0.1
P(X = 3) = 0.7
P(X = 4) = 0.2



Dekking et al., A Modern Introduction to Probability and Statistics : Understanding Why and How, 2005

Example 1

Find **E[X]** where **X** is the outcome when we roll a **fair** die.

$$P(X = x_i) = \frac{1}{6} \qquad x_1 = 1, \ldots, x_6 = 6$$

$$E[X] = \sum_{i=1}^{6} x_i P(X = x_i)$$

$$= 1(\frac{1}{6}) + 2(\frac{1}{6}) + \ldots + 6(\frac{1}{6}) = \frac{7}{2}$$

Example 2

Let **X** be the number of hours delay on the arrival of *commuter line* at UI train station, with the pdf

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & otherwise \end{cases}$$

What is the average time you will have to wait for the train?

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \int_{0}^{\infty} x e^{-x}\, dx = 1$$

E[X] is mean of X

On average, we have to wait for **1 hour**

Expectation of a function of a random variable

If **X** is discrete R.V. with PMF ***p*(x)**, then for any real-valued function **g**

$$E[g(X)] = \sum_{x} g(x) P(X = x) = \sum_{x} g(x) p_X(x)$$

If **X** is continuous R.V. with PDF ***f*(x)**, then for any real-valued function **g**

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx$$

Indicator Variables

Define I an indicator random variable for an event A as follow

$$I = \begin{cases} 1 & if\ \ A\, occurs \\ 0 & if\ \ A\, doesn't\ occur \end{cases}$$

The expectation of *I* is equal to the probability that A occurs.

$$\begin{aligned} E[I] &= 1P(I=1) + 0P(I=0) \\ &= 1P(A) + 0P(A^C) \\ &= P(A) \end{aligned}$$

We will see the use of this type of random variable later ☺

Properties of Expected Value

If **a** and **b** are constants, then

$$E[aX+b]=aE[X]+b$$

Proof:

For discrete case:

$$\begin{aligned} E[aX+b] &= \sum_x (ax+b)p(x) \\ &= a\sum_x xp(x)+b\sum_x p(x) \\ &= aE[X]+b \end{aligned}$$

For continuous case:

Try to show it ! This is practice for you !

Properties of Expected Value

Other properties …

$$E[aX] = aE[X]$$

$$E[b] = b$$

We also call **E[X]** as mean or *first moment* of R.V **X**.

E[X^n] is the *nth moment* of R.V. **X**

$$E[X^n] = \begin{cases} \sum_x x^n p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^n f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

Expected Value of Sums of Random Variables

$$E[g(X,Y)] = \begin{cases} \sum_y \sum_x g(x,y)p(x,y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy & \text{if } X \text{ is continuous} \end{cases}$$

Consequently,

$$E[X+Y] = E[X] + E[Y]$$

In general…

Prove these statements !

$$E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

Problem (easy)

(1) X has three values {0, 1, 2} and the following PMF:

$$p(0) = 0.2 \qquad p(1) = 0.5 \qquad p(2) = 0.3$$

compute **E[X²]** !

(2) X has the following PDF:

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & otherwise \end{cases}$$

compute **E[X³]** !

Problem (intermediate)

Suppose there are 20 different types of coupons and suppose that each time one obtains a coupon it is equally likely to be any one of the types.

Compute the expected number of different types that are contained in a set for 10 coupons !

Hint: use indicator random variable ☺

$$X_i = \begin{cases} 1 & \text{If } \textbf{at least one type i} \text{ coupon is in the set of 10} \\ 0 & \text{Otherwise} \end{cases}$$

Mean Square Error (MSE)

Suppose we are to predict the value of a random variable **X**, with say c. The 'square' of the error involved is **(X - c)²**.

The average(mean) squared error (MSE) is minimized when we predict **X = E[X]**.

$$
\begin{aligned}
E[(X-c)^2] &= E[(X-\mu+\mu-c)^2] \\
&= E[(X-\mu)^2 + 2(\mu-c)(X-\mu) + (\mu-c)^2] \\
&= E[(X-\mu)^2] + 2(\mu-c)E[X-\mu] + (\mu-c)^2 \\
&= E[(X-\mu)^2] + (\mu-c)^2 \\
&\geq E[(X-\mu)^2]
\end{aligned}
$$

So, **E[X] itself** is the best predictor for **X** in terms of minimum MSE !

Variance & Covariance

E[X] does **not** tell us anything about the *variation*, or *spread* of random variable **X** !

Random variables **W**, **Y**, and **Z** all have the same expectation = **0**.

$$Z = \begin{cases} -100 \\ -50 \\ 50 \\ 100 \end{cases}$$

Each has probability **0.25**

$$Y = \begin{cases} -1 \\ 1 \end{cases}$$

Each has probability **0.5**

$$W = 0$$

with probability **1**

There is much **greater spread** in the possible values of **Z** than the others

Variance

DEFINITION

If **X** is a random variable with mean **μ**, then the variance of **X**, denoted by **Var(X) or σ²**, is defined by

$$Var(X) = \sigma^2 = E[(X-\mu)^2]$$

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] \\ &= E[(X^2 - 2\mu X - \mu^2)] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

Var(X) measures the **possible variation** of **X**

Variance (σ^2) & Standard Deviation (σ)

Variance (σ^2)

- A value that demonstrates the the spread (around the mean) or a random variable.
- Variance is always positive (or 0)
- As the variance becomes larger, the distribution is more spread out !

Standard Deviation (σ)

- positive square root of the variance

$$SD(X) = \sigma = \sqrt{Var(X)}$$

Example

Compute **Var(X)** when **X** represents the outcome when we roll a fair die.

Previously, we know that **E[X]** = 7/2

$$E[X^2] = \sum_{i=1}^{6} i^2 \, P(X = i)$$

$$= 1^2(\frac{1}{6}) + 2^2(\frac{1}{6}) + \ldots + 6^2(\frac{1}{6}) = \frac{91}{6}$$

$$Var(X) = E[X^2] - (E(X))^2$$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Some Identities of Variance...

For any constants **a** and **b**

$$Var(aX+b) = a^2 Var(X)$$

Proof:

$$\begin{aligned} Var(aX+b) &= E[(aX+b-E[aX+b])^2] \\ &= E[(aX+b-a\mu-b)^2] \\ &= E[(aX-a\mu)^2] \\ &= E[a^2(E-\mu)^2] \\ &= a^2 E[(E-\mu)^2] \\ &= a^2 Var(X) \end{aligned}$$

Some Identities of Variance...

Other consequences ...

$$Var(X+b)=Var(X) \qquad Var(aX)=a^2\,Var(X)$$

$$Var(b)=0$$

Variance of Indicator Random Variable

$$\begin{aligned} Var(I) &= E[I^2]-(E[I])^2 \\ &= E[I]-(E[I])^2 \qquad ,I^2=I \\ &= E[I](1-E[I]) \\ &= P(A)[1-P(A)] \end{aligned}$$

Covariance

DEFINITION

The *Covariance* of two random variables **X** and **Y**, written by **Cov(X, Y)**, is defined by

$$
\begin{aligned}
Cov(X,Y) &= E[(X-\mu_x)(Y-\mu_y)] \\
&= E[XY - \mu_x Y - \mu_y X + \mu_x \mu_y] \\
&= E[XY] - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

Where μ_x and μ_y are the means of **X** and **Y**.

What does "Covariance" actually mean ?

Covariance shows "**linear-relationshipness**" between two variables.

See the next slide regarding **correlation**

Cov(X, Y) = 0 means that random variable **X** and **Y** are **uncorrelated** (no linear relationship).

Uncorrelated ≠ Independent

If **X** and **Y** are **Independent**, then **X** and **Y** are **Uncorrelated**.

BUT !

If **X** and **Y** are **Uncorrelated**, they **can still be Dependent** !!!

Some Identities of Covariance ...

$$Cov(X,X) = Var(X)$$

$$Cov(X,Y) = Cov(Y,X)$$

$$Cov(aX,Y) = a\,Cov(X,Y)$$

$$Cov(X+Z,Y) = Cov(X,Y) + Cov(Z,Y)$$

$$Cov\left(\sum_{i=1}^{n} X_i, Y\right) = \sum_{i=1}^{n} Cov(X_i, Y)$$

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

Just a proof...

$$
\begin{aligned}
Cov(X+Z,Y) &= E[(X+Z)Y]-E[X+Z]E[Y] \\
&= E[XY]+E[ZY]-(E[X]+E[Z])E[Y] \\
&= E[XY]-E[X]E[Y]+E[ZY]-E[Z]E[Y] \\
&= Cov(X,Y)+Cov(Z,Y)
\end{aligned}
$$

Use this to prove the last two identities (previous slide) !

Variance and Covariance

$$
\begin{aligned}
Var(X+Y) &= Var(X) + Var(Y) + Cov(X,Y) + Cov(Y,X) \\
&= Var(X) + Var(Y) + 2Cov(X,Y)
\end{aligned}
$$

This can be used to show

$$
Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} Cov(X_i, X_j)
$$

If X and Y are **independent** random variables, then

$$Cov(X,Y) = 0$$

Tidak berlaku sebaliknya ! ☺

Prove it using the definition of independent R.V. ! (see next slide)

As a consequence,

$$Var(X+Y) = Var(X) + Var(Y)$$

In general, for independent R.V. $X_1, X_2, \ldots, X_n$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

If **X** and **Y** are **independent** then

$$E[XY] = E[X]E[Y]$$

See for discrete case:

$$
\begin{aligned}
E[XY] &= \sum_i \sum_j x_i y_j \; P(X = x_i, Y = y_j) \\
&= \sum_i \sum_j x_i y_j \; P(X = x_i)P(Y = y_j) \qquad \text{Why ?} \\
&= \sum_i x_i P(X = x_i) \sum_j y_j \; P(Y = y_j) \\
&= E[Y] \sum_i x_i P(X = x_i) \\
&= E[Y]E[X]
\end{aligned}
$$

Prove it for continuous case !

Example

Compute the *variance* of the sum obtained when 10 **independent** rolls of a **fair die** are made.

Let $\mathbf{X_i}$ denote the outcome of the i^{th} roll,

$$Var\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} Var(X_i)$$

$$= 10.\frac{35}{12}$$

$$= \frac{175}{6}$$

Previously, we determined

$$Var(X_i) = \frac{35}{12}$$

Correlation & Covariance

- Positive value of **Covariance** indicates Y tends to increase as X does.
- Negative value of **Covariance** indicates Y tends to decrease as X increase.
- The **Correlation** between two random variables **X** & **Y** is defined as

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X).Var(Y)}}$$

As before, it can easily be seen that **Correlation** lies between – 1 and 1, and independent R.V.s have correlation **0**.

Correlation & Covariance

Practice Problem

What if random variable **X** and **Y** have linear relationship ?

$$Y = aX + b$$

where **a ≠ 0**

(1) compute $Cov(X, Y)$

(2) compute $Corr(X, Y)$

Markov & Chebyshev’s Inequality

Markov's Inequality

If **X** is a random variable that takes only **non-negative values**, then for any value **a > 0**

$$P(X \geq a) \leq \frac{E[X]}{a}$$

$$\begin{aligned}
E[X] &= \int_0^\infty x f(x)\, dx \\
&= \int_0^a x f(x)\, dx + \int_a^\infty x f(x)\, dx \\
&\geq \int_a^\infty x f(x)\, dx \geq \int_a^\infty a f(x)\, dx \\
&= a \int_a^\infty f(x)\, dx = aP(X \geq a)
\end{aligned}$$

Chebyshev's Inequality

If **X** is a random variable with mean **μ** and variance **σ²**, then for any value **k > 0**

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{or} \quad P(|X-\mu| < k\sigma) \geq 1-\frac{1}{k^2}$$

The formulas can also be written as follow:

$$P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2} \quad \text{or} \quad P(|X-\mu| < k) \geq 1-\frac{\sigma^2}{k^2}$$

Prove this using Markov's inequality !

Weak Law of Large Numbers

If we have an event **A**, what does **P(A)** actually mean ?

We have feeling that, if we perform the random experiment **n** times and $\mathbf{S_n}$ be the number of times that event **A** occurs, then $\mathbf{S_n/n}$ is approximately **P(A)**.

At least, in the sense that

$$\lim_{n\to\infty} \frac{S_n}{n} = P(A)$$

Weak law of large number makes this **intuitive notion** more precise

$$P\left(\left|\frac{S_n}{n} - P(A)\right| \geq \varepsilon\right) \to 0 \qquad as\ n \to \infty$$

Prove it using **chebyshev's Inequality**

Arnold O. Allen, Probability, Statistics, and Queueing Theory (with computer science application), 1978

Weak Law of Large Numbers
(another version, the same notion)

Let $\mathbf{X_1}$, $\mathbf{X_2}$, …, $\mathbf{X_n}$ be a sequence of independent and identically distributed (**i.i.d**) random variables, each having mean $\mathbf{E[X_i] = \mu}$.

Then, for any $\boldsymbol{\varepsilon > 0}$

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) \to 0 \qquad as\ n \to \infty$$

Prove it using **chebyshev's Inequality** by the fact that:

$$E\left[\frac{X_1 + \ldots + X_n}{n}\right] = \mu \qquad Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Example

Suppose the number of cars produced by a factory in a week follow a random variable with mean 50. The variance of production during a week is 25

a. What is the probability that the factory will produce this week will exceed 75?

By Markov's Ineq. $P(X > 75) \le \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$

b. What is the probability that the production this week will be between 40 and 60 cars? $P(50 - k\sigma < X < 50 + k\sigma) = ?$

By Chebyshev's Ineq. $P(|X - 50| \le (2.5)) > 1 - \frac{1}{2^2} = \frac{3}{4}$

c. What is the probability that the average cars produced each week in the next 100 weeks are not within 10 of 50? How about a large number of weeks?

By Chebyshev's Ineq for n. $P(|\frac{\sum_{i=1}^{100} X_i}{n} - 50| > 10) \le \frac{25}{100(10^2)} = \frac{1}{400}$

If n = ∞, by Weak Law of Large Number, the probabality=0

Proof: $P(|\frac{\sum_{i=1}^{\infty} X_i}{n} - 50| > 10) \le \frac{25}{\infty(10^2)} = 0$

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the **mean**, or both the **mean** and the **variance**, of the probability distribution are known.

Of course, if the actual distribution were known, then the desired probabilities could be exactly computed and we would not need to resort to bounds.