Expectation

CSGE602013 –STATISTICS AND PROBABILITY

FACULTY OF COMPUTER SCIENCE UNIVERSITAS INDONESIA

References

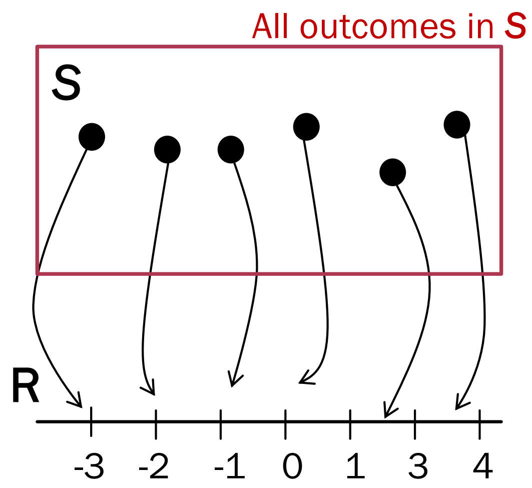
- Introduction to Probability and Statistics for Engineers & Scientists, 4th ed., Sheldon M. Ross, Elsevier, 2009.

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

Random Variable

- Random Variable X, is a function that assigns 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.

PMF and CDF

- The Probability Mass Function (PMF) $p(x_i)$ of X is defined by

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

"the probability that the value of X is *exactly* equal to x_i"

- Cumulative Distribution Function, $F_X(x)$ of the random variable X is defined for any real number x by

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

- "the probability that the value of X is less than or equal to x_i"

Relation of PMF and CDF

- For Discrete R.V., $\boldsymbol{F_X}(t)$ 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

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

Probability Density Function (PDF)

- Probability Density Function (PDF) $f_X(x)$ of RV X for any real number x by

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

- The probability that RV $\boldsymbol{X}$ will be in the interval $\boldsymbol{B}$**,** can be computed by taking the integral of the PDF $f_x(x)$ over $\boldsymbol{B}$
- Letting B = [a, b], the probability within $a \le X \le b$

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

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

Probability Density Function (PDF) (2)

- $f(x)$ must be non-negative

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

- $f(x)$ must satisfy

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

Which represents the probability of every thing that may happen, or the entire sample space (S)

Relation between CDF and PDF

- Given the Cumulative Distribution Function F_x and Probability Density Function f_x
- We know

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\, dx$$

- or

$$\frac{d}{da}F(a) = f(a)$$

Jointly Distributed Random Variables

- Joint PMF of X and Y:

$$p(a_i, b_j) = P(X = a_i, Y = b_j)$$

- Joint CDF of X and Y:

$$F_{XY}(a, b) = F(X \leq a, Y \leq b)$$

- Joint PDF of X and Y:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y)\, dxdy$$

Conditional Distributions

Discrete RV

$$p_{X|Y}(x \mid y) = P(X = x \mid Y = y)$$
$$= \frac{P(X = x, Y = y)}{P(Y = y)}$$
$$= \frac{p(x, y)}{p_Y(y)}$$

Continuous RV

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

EXPECTATION

Expectation

- Random variable X is a numerical value representing each possible outcome in an experiment
- If you have some random variable, X

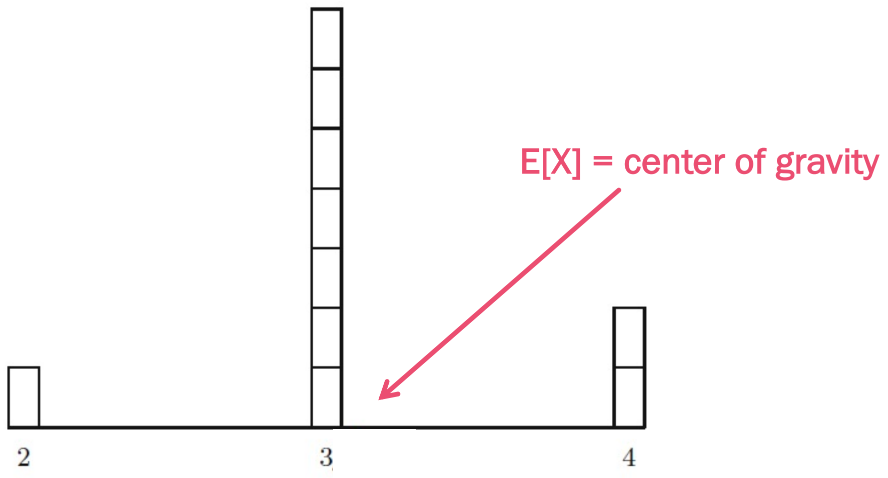
What value do you expect?

- Expectation = Harapan = Mean = Rataan
- $E[X]$ acts as a representative number to describe the random variable X.

Expectation of Discrete RVs

- 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

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.

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**

Indicator Variables

- I is an indicator random variable for an event A as follows:

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

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

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

Properties of Expected Value

1. If a and b are constants, then

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

Prove it!

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

3. $E[b] = b$

Moments

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

- $E[X^n]$ is the **n^th 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 RVs

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

- Consequently,

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

- Generalized

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

Exercise

(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 $\boldsymbol{E}[\boldsymbol{X}^2]$!

(2) X has the following PDF:

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

compute $\boldsymbol{E}[\boldsymbol{X}^3]$!

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 !

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

Hint: use indicator random variable ☺

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 $X = E[X]$.

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

- $E[X]$ itself is the best predictor for X in terms of minimum MSE !

Latihan

- Seorang mahasiswa yang jarang menghadiri kuliah baru sempat mempelajari 5 bab dari 8 bab yang akan diujikan. Ke-lima bab tersebut baru mulai dibaca dan dipelajarinya semalam sebelum ujian secara tergesa-gesa, sehingga ia hanya memahami 60% dari tiap bab yang dipelajarinya, dan tidak mengerti sama sekali terhadap 3 bab lainnya. Ujian ada 5 soal dengan nilai tiap soal 20 poin. Dalam kondisi ini, mahasiswa tersebut merasa hanya akan mendapat nilai ujian 30. Dapatkah anda meyakinkan mahasiswa tersebut bahwa ia bisa saja mendapatkan nilai lebih?
- Asumsi yang bisa anda gunakan antara lain:
  - Pemahaman tiap bab independent dengan bab lain.
  - Untuk bab yang telah dipelajari dan dipahami sebesar 60%, maka nilai pada soal yang mengevaluasi bab tersebut adalah 60% dari poin soal tersebut.
  - Tiap materi pada tiap bab hanya akan dievaluasi oleh satu soal. Dengan kata lain tidak akan ada dua soal yang mengevaluasi pemahaman pada satu bab yang sama.
  - Untuk soal yang mengevaluasi bab yang tidak dimengerti sama sekali, asumsikan kondisi terburuk: mendapat nilainya nol pada soal tersebut.

What does E[X] tell you?

- E[X] tells you the *mean*.. Measure of central tendency!
- 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**

What does E[X] tell you?

- But there is much greater spread in the possible values of Z than the others

Measure of variance!

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

Each has probability **0.25**

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

Each has probability **0.5**

$$W = 0$$

with probability **1**

VARIANCE AND COVARIANCE

Variance

- If X is a random variable with mean μ, then the variance of X, denoted by Var(X) or σ^2 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 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 knew that $E[X] = 7/2$

- Recall:

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

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

Example

- Compute $Var(X)$ when X represents the outcome when we roll a fair die.
- Previously, we knew that $E[X] = 7/2$
- Then:

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

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

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

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

Identities of Variance

1. For any constants a and b $Var(aX+b)=a^2Var(X)$

2. $Var(X+b)=Var(X)$ $Var(aX)=a^2\ Var(X)$
   $Var(b)=0$

3. 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

- 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 $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ are the means of **X** and **Y**.

What does “Covariance” actually mean?

- Covariance shows “linear-relationshipness” between two variables
- 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 incorrelated, they can still be dependent

Latihan

Diberikan *X* dan *Y* adalah variabel acak diskret dengan *joint pmf* sebagai berikut:

	Y=0	Y=1	Y=2
X=0	1/6	1/4	1/8
X=1	1/8	1/6	1/6

a. Hitunglah $P(Y=1|X=0)$

b. Hitunglah $Var(Y)$

c. Hitunglah $Cov(X,Y)$

Identities of Covariance

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

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

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

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

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

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

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

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

Dependency and Covariance

- If X and Y are independent random variables, then

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

  - It's not valid conversely!

- As a consequence,

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

- In general, for independent R.V. X_1, X_2, ..., X_n

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

Independent RVs

- If X and Y are independent then

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

- See for discrete case:

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

Example

- Compute the variance of the sum obtained when 10 independent rolls of a fair die are made.
- Let X_i denote the outcome of the i^{th} roll,

Example

- Compute the variance of the sum obtained when 10 independent rolls of a fair die are made.
- Let 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 with X.
- Negative value of Covariance indicates Y tends to decrease as X increases.
- 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.

Exercise

If random variable X and Y have linear relationship $\boldsymbol{Y = ax + b}$ where $\boldsymbol{a \neq 0}$

(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**

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

$$E[X] = \int_0^\infty xf(x)\,dx$$

$$= \int_0^a xf(x)\,dx + \int_a^\infty xf(x)\,dx$$

$$\geq \int_a^\infty xf(x)\,dx \geq \int_a^\infty af(x)\,dx$$

$$= a\int_a^\infty f(x)\,dx = aP(X \geq a)$$

Chebyshev's Inequality

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

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

or

The formulas can also be written as follow:

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

or

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 S_n be the number of times that event **A** occurs, then 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 X_1, X_2, …, X_n be a sequence of independent and identically distributed (**i.i.d**) random variables, each having mean **$E[X_i] = \mu$.**

Then, for any **$\varepsilon > 0$**

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

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. $$P(|X - 50| \le 10) > 1 - \frac{25}{10^2} = \frac{3}{4}$$

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. $$P\left(\left|\frac{\sum_{i=1}^{100} X_i}{n} - 50\right| > 10\right) \le \frac{25}{100(10^2)} = \frac{1}{400}$$

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

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

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.