Mathematics for Computer Science

Prof.dr.hab. Viorel Bostan

Technical University of Moldova

viorel.bostan@adm.utm.md

Lecture 8

- A probability professor got caught by terroists and and they start to torture him.

- A probability professor got caught by terroists and and they start to torture him.
- They forced the proffesor to put his head in a hot oven and his legs in the refrigerator.

- A probability professor got caught by terroists and and they start to torture him.
- They forced the proffesor to put his head in a hot oven and his legs in the refrigerator.
- "How do you feel?" asked after a while terrorists.

- A probability professor got caught by terroists and and they start to torture him.
- They forced the proffesor to put his head in a hot oven and his legs in the refrigerator.
- "How do you feel?" asked after a while terrorists.
- Professor answered: ...

- A probability professor got caught by terroists and and they start to torture him.
- They forced the proffesor to put his head in a hot oven and his legs in the refrigerator.
- "How do you feel?" asked after a while terrorists.
- Professor answered: ...
- On the average just fine!

- When a large collection of numbers is assembled, we might be interested not in the individual numbers, but in certain descriptive quantities such as the average or the median. In general, the same is true for the probability distribution of a numerically–valued random variable.

- When a large collection of numbers is assembled, we might be interested not in the individual numbers, but in certain descriptive quantities such as the average or the median. In general, the same is true for the probability distribution of a numerically–valued random variable.
- We will discuss two such descriptive quantities:
  the **expected value** and the **variance**.

- When a large collection of numbers is assembled, we might be interested not in the individual numbers, but in certain descriptive quantities such as the average or the median. In general, the same is true for the probability distribution of a numerically–valued random variable.
- We will discuss two such descriptive quantities:
  the **expected value** and the **variance**.
- Both of these quantities apply only to numerically–valued random variables.

- When a large collection of numbers is assembled, we might be interested not in the individual numbers, but in certain descriptive quantities such as the average or the median. In general, the same is true for the probability distribution of a numerically–valued random variable.
- We will discuss two such descriptive quantities:
  the **expected value** and the **variance**.
- Both of these quantities apply only to numerically–valued random variables.
- Consider the following game:
  A die is rolled. If an odd number turns up, we win an amount equal to this number; if an even number turns up, we lose an amount equal to this number.

- When a large collection of numbers is assembled, we might be interested not in the individual numbers, but in certain descriptive quantities such as the average or the median. In general, the same is true for the probability distribution of a numerically–valued random variable.
- We will discuss two such descriptive quantities:
  the **expected value** and the **variance**.
- Both of these quantities apply only to numerically–valued random variables.
- Consider the following game:
  A die is rolled. If an odd number turns up, we win an amount equal to this number; if an even number turns up, we lose an amount equal to this number.
- We want to decide if this is a reasonable game to play.
  You can try a computer simulation.

Winning	n = 100		n = 10000	
	Frequency	Relative Frequency	Frequency	Relative Frequency
1	17	.17	1681	.1681
-2	17	.17	1678	.1678
3	16	.16	1626	.1626
-4	18	.18	1696	.1696
5	16	.16	1686	.1686
-6	16	.16	1633	.1633

In the first run we have played the game 100 times.
In this run our average gain is -0.57.

Winning	n = 100		n = 10000	
	Frequency	Relative Frequency	Frequency	Relative Frequency
1	17	.17	1681	.1681
-2	17	.17	1678	.1678
3	16	.16	1626	.1626
-4	18	.18	1696	.1696
5	16	.16	1686	.1686
-6	16	.16	1633	.1633

In the first run we have played the game 100 times.
In this run our average gain is -0.57.

It looks as if the game is unfavorable.

Discrete Expected Value

Winning	n = 100 Frequency	n = 100 Relative Frequency	n = 10000 Frequency	n = 10000 Relative Frequency
1	17	.17	1681	.1681
-2	17	.17	1678	.1678
3	16	.16	1626	.1626
-4	18	.18	1696	.1696
5	16	.16	1686	.1686
-6	16	.16	1633	.1633

In the first run we have played the game 100 times.
In this run our average gain is -0.57.

It looks as if the game is unfavorable.

To get a better idea, we have played the game $10,000$ times. In this case our average gain is -0.4949.

$$\mu = 1 \cdot \frac{1}{6} - 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} - 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} - 6 \cdot \frac{1}{6} = -0.5.$$

Definition

Let X be a numerically-valued discrete random variable with sample space Ω and distribution function $m(x)$.
The **expected value** $E(X)$ is defined by

$$E(X) = \sum_{x \in \Omega} x \cdot m(x),$$

provided this sum converges absolutely.

Expected value $E(X)$ is reffered as the mean, and denoted also by μ.

If the above sum does not converge absolutely, then we say that X does not have an expected value.

Discrete Expected Value

Example 1.
Let an experiment consist of tossing a fair coin 3 times.
Let X denote the number of heads which appear.

Then, the possible values of X are $0, 1, 2$ and 3.

The corresponding probabilities are $1/8, 3/8, 3/8$, and $1/8$.

Example 1.
Let an experiment consist of tossing a fair coin 3 times.
Let X denote the number of heads which appear.

Then, the possible values of X are $0, 1, 2$ and 3.

The corresponding probabilities are $1/8, 3/8, 3/8$, and $1/8$.

Thus, the expected value of X equals:

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

Later we will see a quicker way to compute this expected value, based on the fact that X can be written as a sum of simpler random variables.

Example 2. Suppose that we toss a fair coin until a head first comes up, and let X represent the number of tosses which were made.

Then the possible values of X are $1, 2, \ldots$, and the distribution function of X is defined by

$$m(i) = \frac{1}{2^i}.$$

Then,

$$E(X) = \sum_{i=1}^{\infty} i \cdot \frac{1}{2^i} = 2.$$

Example 3. Suppose that we flip a coin until a head first appears, and if the number of tosses equals n, then we are paid 2^n dollars.
What is the expected value of the payment?
Solution. Let Y represent the payment. Then,

$$P(Y = 2^n) = \frac{1}{2^n},$$

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1,$$

which is a divergent sum. Thus, random variable Y has no expectation.

Example 3. Suppose that we flip a coin until a head first appears, and if the number of tosses equals n, then we are paid 2^n dollars.
What is the expected value of the payment?
Solution. Let Y represent the payment. Then,

$$P(Y = 2^n) = \frac{1}{2^n},$$

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1,$$

which is a divergent sum. Thus, random variable Y has no expectation.
This example is called the **St. Petersburg Paradox.**

Example 3. Suppose that we flip a coin until a head first appears, and if the number of tosses equals n, then we are paid 2^n dollars.
What is the expected value of the payment?
Solution. Let Y represent the payment. Then,

$$P(Y = 2^n) = \frac{1}{2^n},$$

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1,$$

which is a divergent sum. Thus, random variable Y has no expectation.
This example is called the **St. Petersburg Paradox.**
The fact that the above sum is infinite suggests that a player should be willing to pay any fixed amount per game for the privilege of playing this game.

Example 3. Suppose that we flip a coin until a head first appears, and if the number of tosses equals n, then we are paid 2^n dollars.
What is the expected value of the payment?
Solution. Let Y represent the payment. Then,

$$P(Y = 2^n) = \frac{1}{2^n},$$

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1,$$

which is a divergent sum. Thus, random variable Y has no expectation.
This example is called the **St. Petersburg Paradox.**
The fact that the above sum is infinite suggests that a player should be willing to pay any fixed amount per game for the privilege of playing this game.
How much you would be willing to pay for this privilege?

Example 3. Suppose that we flip a coin until a head first appears, and if the number of tosses equals n, then we are paid 2^n dollars.
What is the expected value of the payment?
Solution. Let Y represent the payment. Then,

$$P(Y = 2^n) = \frac{1}{2^n},$$

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1,$$

which is a divergent sum. Thus, random variable Y has no expectation.
This example is called the **St. Petersburg Paradox.**
The fact that the above sum is infinite suggests that a player should be willing to pay any fixed amount per game for the privilege of playing this game.
How much you would be willing to pay for this privilege?
It is unlikely that your answer is more than 10 dollars. Paradox!

Example 4.

Let T be the time for the first success in a Bernoulli trials process.

Take as sample space $\mathbb{N}$ and assign geometric distribution

$$m(j) = P(T = j) = q^{j-1}p.$$

$$\begin{aligned} E(T) &= 1 \cdot p + 2qp + 3q^2p + \dots \\ &= p(1 + 2q + 3q^2 + \dots) \\ &= \frac{p}{(1-q)^2} = \frac{1}{p}. \end{aligned}$$

Example 4.

Let T be the time for the first success in a Bernoulli trials process.

Take as sample space $\mathbb{N}$ and assign geometric distribution

$$m(j) = P(T = j) = q^{j-1}p.$$

$$\begin{aligned} E(T) &= 1 \cdot p + 2qp + 3q^2p + \dots \\ &= p(1 + 2q + 3q^2 + \dots) \\ &= \frac{p}{(1-q)^2} = \frac{1}{p}. \end{aligned}$$

In particular, we see that if we toss a fair coin a sequence of times, the expected time until the first heads is $\frac{1}{1/2} = 2$. If we roll a die a sequence of times, the expected number of rolls until the first six is $\frac{1}{1/6} = 6$.

- The expected value is an average outcome.

- The expected value is an average outcome.
- We know that for any finite experiment the average of the outcomes is not predictable.

Discrete Expected Value

- The expected value is an average outcome.
- We know that for any finite experiment the average of the outcomes is not predictable.
- However, we shall eventually prove that the average will usually be close to $E(X)$ if we repeat the experiment a large number of times.

Discrete Expected Value

- The expected value is an average outcome.
- We know that for any finite experiment the average of the outcomes is not predictable.
- However, we shall eventually prove that the average will usually be close to $E(X)$ if we repeat the experiment a large number of times.
- First need to develop some properties of the expected value.

Discrete Expected Value

- The expected value is an average outcome.
- We know that for any finite experiment the average of the outcomes is not predictable.
- However, we shall eventually prove that the average will usually be close to $E(X)$ if we repeat the experiment a large number of times.
- First need to develop some properties of the expected value.
- Using these properties, and definition of the variance to be introduced later, we shall be able to prove the **Law of Large Numbers**.

Discrete Expected Value

- The expected value is an average outcome.
- We know that for any finite experiment the average of the outcomes is not predictable.
- However, we shall eventually prove that the average will usually be close to $E(X)$ if we repeat the experiment a large number of times.
- First need to develop some properties of the expected value.
- Using these properties, and definition of the variance to be introduced later, we shall be able to prove the **Law of Large Numbers**.
- This theorem will justify mathematically both the frequency concept of probability and the interpretation of expected value as the average value to be expected in a large number of experiments.

Theorem

If X is a discrete random variable with sample space Ω and distribution function $m(x)$, and if ϕ is a function, then

$$E(\phi(X)) = \sum_{x \in \Omega} \phi(x) \cdot m(x),$$

provided the series converges absolutely.

Theorem

Let X and Y be random variables with finite expected values. Then

$$\begin{aligned} E(X+Y) &= E(X) + E(Y), \\ E(cX) &= cE(X), \end{aligned}$$

where c is any constant.

Theorem

Let S_n be the number of successes in n Bernoulli trials with probability p for success on each trial. Then the expected number of successes is np. That is,

$$E(S_n) = np.$$

Proof.

Let X_j be a random variable which has the value 1 if the $j-$th outcome is a success and 0 if it is a failure. Then, for each X_j,

$$\begin{aligned} E(X_j) &= 0 \cdot (1-p) + 1 \cdot p = p, \\ S_n &= X_1 + X_2 + \ldots + X_n. \end{aligned}$$

Theorem

Let S_n be the number of successes in n Bernoulli trials with probability p for success on each trial. Then the expected number of successes is np. That is,

$$E(S_n) = np.$$

Proof.

Let X_j be a random variable which has the value 1 if the $j-$th outcome is a success and 0 if it is a failure. Then, for each X_j,

$$\begin{aligned} E(X_j) &= 0 \cdot (1-p) + 1 \cdot p = p, \\ S_n &= X_1 + X_2 + \ldots + X_n. \end{aligned}$$

The expected value of the sum is the sum of the expected values:

$$E(S_n) = np.$$

In Poisson distribution parameter λ was obtained as limit of binomial distributions with parameters n and p, such that $np = \lambda$, and $n \to \infty$.

In Poisson distribution parameter λ was obtained as limit of binomial distributions with parameters n and p, such that $np = \lambda$, and $n \to \infty$.

Since for each n, the corresponding binomial distribution has expected value $np = \lambda$, it is reasonable to guess that the expected value of a Poisson distribution with parameter also has expectation equal to λ.

In Poisson distribution parameter λ was obtained as limit of binomial distributions with parameters n and p, such that $np = \lambda$, and $n \to \infty$.

Since for each n, the corresponding binomial distribution has expected value $np = \lambda$, it is reasonable to guess that the expected value of a Poisson distribution with parameter also has expectation equal to λ.

If X and Y are two random variables, then in general

$$E(X \cdot Y) \neq E(X) \cdot E(Y).$$

However, this is true if X and Y are independent.

In Poisson distribution parameter λ was obtained as limit of binomial distributions with parameters n and p, such that $np = \lambda$, and $n \to \infty$.

Since for each n, the corresponding binomial distribution has expected value $np = \lambda$, it is reasonable to guess that the expected value of a Poisson distribution with parameter also has expectation equal to λ.

If X and Y are two random variables, then in general

$$E(X \cdot Y) \neq E(X) \cdot E(Y).$$

However, this is true if X and Y are independent.

Theorem

If X and Y are independent random variables, then

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Example. A coin is tossed twice. $X_i = 1$ if the $i-$th toss is heads and 0 otherwise. We know that X_1 and X_2 are independent. They each have expected value 1/2. Thus

$$E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2) = \frac{1}{4}.$$

Example. Consider a single toss of a coin. We define the random variable X to be 1 if heads turns up and 0 if tails turns up, and we set $Y = 1 - X$. Then

$$E(X) = E(Y) = \frac{1}{2}.$$

But $XY = 0$ for either outcome. Hence, $E(XY) = 0 \neq E(X)E(Y)$.

Snowfall records

We start keeping snowfall records this year and want to find the expected number of records that will occur in the next n years.

The first year is necessarily a record.

The second year will be a record if the snowfall in the second year is greater than that in the first year.

By symmetry, this probability is $1/2$.

More generally, let X_j be 1 if the $j-$th year is a record and 0 otherwise.

To find $E(X_j)$, we need only find the probability that the $j-$th year is a record.

But the record snowfall for the first j years is equally likely to fall in any one of these years, so $E(X_j) = 1/j$.

Therefore, if S_n is the total number of records observed in the first n years,

$$E(S_n) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$$

This is the famous divergent harmonic series. It can be shown that

$$E(S_n) \sim \ln n \quad as \quad n \to \infty$$

Therefore, in ten years the expected number of records is approximately $\ln 10 = 2.3$.

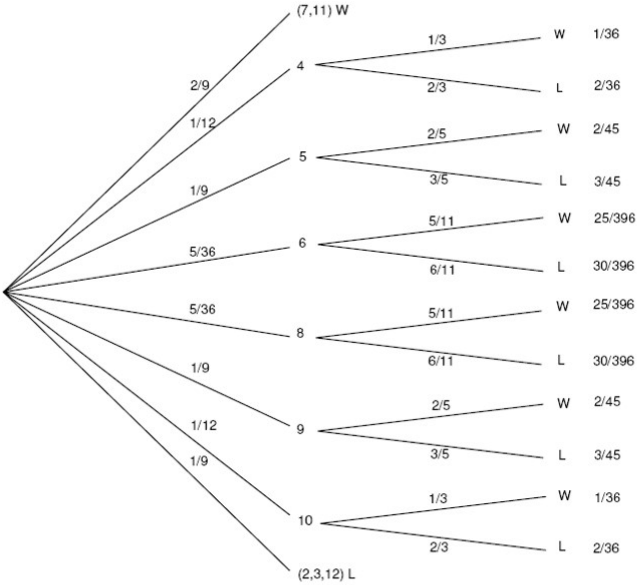
The exact value is the sum of the first ten terms of the harmonic series which is 2.9.


10
NINE
SIX
COME
FIELD
Don't Win Bar
WIN LINE
Don't Come Bar
Pays Double
Don't Win Bar

- In the game of craps, the player makes a bet and rolls a pair of dice.
- If the sum of the numbers is 7 or 11 the player wins, if it is 2, 3, or 12 the player loses.
- If any other number results, say r, then r becomes the player's point and he continues to roll until either r or 7 occurs. If r comes up first he wins, and if 7 comes up first he loses.
- Let us calculate the expected winnings on a single play and see if this game is favourable or not.
- We construct a two-stage tree measure as shown on the next page.

- The first stage represents the possible sums for his first roll.
- The second stage represents the possible outcomes for the game if it has not ended on the first roll. In this stage we are representing the possible outcomes of a sequence of rolls required to determine the final outcome. The branch probabilities for the first stage are computed in the usual way assuming all 36 possibilites for outcomes for the pair of dice are equally likely.
- For the second stage we assume that the game will eventually end, and we compute the conditional probabilities for obtaining either the point or a 7.
- For example, assume that the player's point is 6. Then the game will end when one of the eleven pairs, (1, 5), (2, 4), (3, 3), (4, 2), (5, 1), (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1), occurs.
- Assume that each of these possible pairs has the same probability. Then the player wins in the first 5 cases and loses in the last 6.

Craps


(7,11) W
2/9
1/12
1/9
5/36
5/36
1/9
1/12
1/9
4
5
6
8
9
10
(2,3,12) L
1/3
2/3
2/5
3/5
5/11
6/11
5/11
6/11
2/5
3/5
1/3
2/3
W 1/36
L 2/36
W 2/45
L 3/45
W 25/396
L 30/396
W 25/396
L 30/396
W 2/45
L 3/45
W 1/36
L 2/36

- Thus the probability of winning is $5/11$ and the probability of losing is $6/11$.
- From the path probabilities, we can find the probability that the player wins 1 dollar; it is $244/495$.
- The probability of losing is then $251/495$. Thus if X is his winning for a dollar bet,

$$E(X) = 1 \cdot \frac{244}{495} + (-1) \cdot \frac{251}{495} = -0.0141.$$

- The game is unfavorable, but only slightly. The player's expected gain in n plays is $-n(.0141)$. If n is not large, this is a small expected loss for the player.
- The casino makes a large number of plays and so can afford a small average gain per play and still expect a large profit.

In Las Vegas, a roulette wheel has 38 slots numbered $0, 00, 1, 2, \ldots, 36$. The 0 and 00 slots are green, and half of the remaining 36 slots are red and half are black. A croupier spins the wheel and throws an ivory ball. If you bet 1 dollar on red, you win 1 dollar if the ball stops in a red slot, and otherwise you lose a dollar. We wish to calculate the expected value of your winnings, if you bet 1 dollar on red.
Let X be the random variable which denotes your winnings in a 1 dollar bet on red in Las Vegas roulette. Then the distribution of X is given by

$$m_X = \begin{pmatrix} -1 & 1 \\ \frac{20}{38} & \frac{18}{38}. \end{pmatrix}$$

Then,

$$E(X) = (-1) \cdot \frac{20}{38} + 1 \cdot \frac{18}{38} = -0.0526.$$

Variance of Discrete Random Variables

The usefulness of the expected value as a prediction for the outcome of an experiment is increased when the outcome is not likely to deviate too much from the expected value.
Let introduce a measure of this deviation, called the **variance**.

Definition

Let X be a numerically valued random variable with expected value $\mu = E(X)$. Then, the **variance** of X, denoted by $V(X)$, is

$$V(X) = E((X - \mu)^2).$$

Note that, $V(X)$ is given by

$$V(X) = \sum (x - \mu)^2 \, m(x),$$

where m is the distribution function of random variable X.

Definition

The **standard deviation** of X, denoted by $D(X)$, is

$$D(X) = \sqrt{V(X)}.$$

We often write σ for $D(X)$ and σ^2 for $V(X)$.

Example. Consider one roll of a die. Let X be the number that turns up. To find $V(X)$, we must first find the expected value of X. This is

$$\mu = E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

To find the variance of X, we form the new random variable $(X - \mu)^2$ and compute its expectation.

x	$m(x)$	$(x-7/2)^2$
1	1/6	25/4
2	1/6	9/4
3	1/6	1/4
4	1/6	1/4
5	1/6	9/4
6	1/6	25/4

From this table we find $E((X-\mu)^2)$ is

$$\begin{aligned} V(X) &= \frac{25}{4}\cdot\frac{1}{6}+\frac{9}{4}\cdot\frac{1}{6}+\frac{1}{4}\cdot\frac{1}{6}+\frac{1}{4}\cdot\frac{1}{6}+\frac{9}{4}\cdot\frac{1}{6}+\frac{25}{2}\cdot\frac{1}{6} \\ &= \frac{35}{12} \end{aligned}$$

and the standard deviation

$$D(X)=\sqrt{35/12}=1.707.$$

Theorem

If X is any random variable with $E(X) = \mu$, then

$$V(X) = E(X^2) - \mu^2.$$

Proof.

We have:

$$\begin{aligned} V(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu\mu + \mu^2. \end{aligned}$$

$\square$

Consider the example of rolling a die.

$$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$$

and,

$$\begin{aligned} V(X) &= E(X^2) - \mu^2 \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 \\ &= \frac{35}{12}, \end{aligned}$$

which is in agreement with the value obtained directly from the definition of $V(X)$.

Properties of Variance

If c is any constant, $E(cX) = cE(X)$ and $E(X + c) = E(X) + c$. These two statements imply that the expectation is a linear function. The variance is not linear.

Theorem

If X is any random variable and c is any constant, then

$$V(cX) = c^2 V(X), \quad V(X + c) = V(X).$$

Proof.

Let $\mu = E(X)$. Then $E(cX) = c\mu$, and

$$\begin{aligned} V(cX) &= E((cX - c\mu)^2) = E(c^2(X - \mu)^2) \\ &= c^2 E((X - \mu)^2) = c^2 V(X). \end{aligned}$$

The second assertion is proved similarly. $\square$

Theorem

Let X and Y be two independent random variables. Then

$$V(X+Y) = V(X) + V(Y).$$

Proof.

Let $E(X) = a$ and $E(Y) = b$. Then

$$\begin{aligned} V(X+Y) &= E((X+Y)^2) - (a+b)^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - a^2 - 2ab - b^2. \end{aligned}$$

Since X and Y are independent, $E(XY) = E(X)E(Y) = ab$. Thus,

$$\begin{aligned} V(X+Y) &= E(X^2) - a^2 + E(Y^2) - b^2 \\ &= V(X) + V(Y). \end{aligned}$$

Theorem

Let $X_1, X_2, \ldots, X_n$ be an independent trials process with $E(X_j) = \mu$ and $V(X_j) = \sigma^2$. Let

$$S_n = X_1 + X_2 + \ldots + X_n$$

be the sum, and

$$A_n = \frac{S_n}{n}$$

be the average. Then

$$\begin{aligned} E(S_n) &= n\mu, \quad V(S_n) = n\sigma^2, \\ E(A_n) &= \mu, \quad V(A_n) = \frac{\sigma^2}{n}. \end{aligned}$$

Proof.

Since all the random variables X_j have the same expected value, we have

$$
\begin{aligned}
E(S_n) &= E(X_1) + \ldots + E(X_n) = n\mu, \\
V(S_n) &= V(X_1) + \ldots + V(X_n) = n\sigma^2.
\end{aligned}
$$

We have seen that, if we multiply a random variable X with mean μ and variance σ^2 by a constant c, the new random variable has expected value $c\mu$ and variance $c^2\sigma^2$. Thus,

$$
\begin{aligned}
E(A_n) &= E\left(\frac{S_n}{n}\right) = \frac{n\mu}{n} = \mu, \\
V(An) &= V\left(\frac{S_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.
\end{aligned}
$$

$\square$

Note that he standard deviation of A_n is given by

$$\sigma(A_n) = \frac{\sigma}{\sqrt{n}}.$$

Note that he standard deviation of A_n is given by

$$\sigma(A_n) = \frac{\sigma}{\sqrt{n}}.$$

This statement implies that in an independent trials process, if the individual terms have finite variance, then the standard deviation of the average goes to 0 as $n \to \infty$.

Since the standard deviation tells us something about the spread of the distribution around the mean, we see that for large values of n, the value of A_n is usually very close to the mean of A_n, which equals μ, as shown above.

This statement is called the **Law of Large Numbers**.

Properties of Variance

Example. Consider n rolls of a die.

We have seen that, if X_j is the outcome if the $j-$th roll, then $E(X_j) = 7/2$ and $V(X_j) = 35/12$.

Thus, if S_n is the sum of the outcomes, and $A_n = S_n/n$ is the average of the outcomes, we have $E(A_n) = 7/2$ and $V(A_n) = (35/12)/n$.

Therefore, as n increases, the expected value of the average remains constant, but the variance tends to 0.

If the variance is a measure of the expected deviation from the mean this would indicate that, for large n, we can expect the average to be very near the expected value.

Consider next the general Bernoulli trials process. As usual, we let $X_j = 1$ if the $j-$th outcome is a success and 0 if it is a failure. If p is the probability of a success, and $q = 1 - p$, then

$$\begin{aligned} E(X_j) &= 0 \cdot q + 1 \cdot p = p, \\ E(X_j^2) &= 0^2 \cdot q + 1^2 \cdot p = p, \\ V(Xj) &= E(X_j^2) - (E(X_j))^2 = p - p^2 = pq. \end{aligned}$$

Thus, for Bernoulli trials, if $S_n = X_1 + X_2 + \ldots + X_n$ is the number of successes, then

$$E(S_n) = np, \quad V(S_n) = npq, \quad D(S_n) = \sqrt{npq}.$$

If $A_n = S_n/n$ is the average number of successes, then

$$E(A_n) = p, \quad V(A_n) = pq/n, \quad D(A_n) = \sqrt{pq/n}.$$

See that expected proportion of successes remains p and the variance tends to 0. This suggests that the frequency interpretation of probability is a correct one.

Bernoulli Trials

Let T denote the number of trials until the first success in a Bernoulli trials process.

Then T is geometrically distributed. What is the variance of T?

$$m_T = \begin{pmatrix} 1 & 2 & 3 & \ldots, \\ p & pq & pq^2 & \ldots \end{pmatrix}$$

We showed that $E(T) = 1/p$. Thus, $V(T) = E(T^2) - (1/p)^2$ so we need only find

$$\begin{aligned} E(T^2) &= 1 \cdot p + 4qp + 9q^2 p + \ldots \\ &= p(1 + 4q + 9q^2 + \ldots) \\ &= \frac{1+q}{p^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} V(T) &= E(T^2) - (E(T))^2 \\ &= \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}. \end{aligned}$$

For example, the variance for the number of tosses of a coin until the first head turns up is $(1/2)/(1/2)2 = 2$.

The variance for the number of rolls of a die until the first six turns up is $(5/6)/(1/6)2 = 30$.

Note that, as p decreases, the variance increases rapidly. This corresponds to the increased spread of the geometric distribution as p decreases.

The End of Lecture