Discrete Mathematics and Its Applications

Lecture 5: Discrete Probability: Expected Value and Variance

MING GAO

DaSE@ ECNU
(for course related communications)
mgao@dase.ecnu.edu.cn

Dec. 6, 2018

Outline

1. Expectation
2. Linearity of Expectations
3. Average-Case Computational Complexity
4. Variance
5. Tail Probability
6. Take-aways

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

- After reading the fine print. Is this a good bet?

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

- After reading the fine print. Is this a good bet?
- If 1,000 tickets will be sold. Is this a good bet?

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

- After reading the fine print. Is this a good bet?
- If 1,000 tickets will be sold. Is this a good bet?
- If 10,000 tickets will be sold. Is this a good bet?

Running example Cont'd

Suppose that in order to raise income for a local seniors citizens home, the town council for Pickering decides to hold a charity lottery:

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

- After reading the fine print. Is this a good bet?
- If 1,000 tickets will be sold. Is this a good bet?
- If 10,000 tickets will be sold. Is this a good bet?
- If 100,000 tickets will be sold. Is this a good bet?

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Hence, it is worth to invest the charity lottery.
If there are 10,000 tickets will be sold, how about your answer?

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

Enter the Charity Lottery
One Grand Prize of $20,000
20 additional prizes of $500

Tickets only $10

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Hence, it is worth to invest the charity lottery.

Running example Cont'd

Question: If 1,000 tickets will be sold. Is this a good bet?

We can compute the average win of every investor as follows:

$$avg. = \frac{20000 + 20 \times 500}{1000} = 30 > 10.$$

$$(avg. = \frac{1}{1000} \cdot 20000 + \frac{20}{1000} \times 500 + \frac{979}{1000} \times 0)$$

Hence, it is worth to invest the charity lottery.
If there are 10,000 tickets will be sold, how about your answer?

Expected value

The **expected value** of a r.v. is the sum over all elements in a sample space of the product of the probability of an its element and the value of the r.v. at this element.

Definition

The **expected value**, also called **expectation** or **mean**, of r.v. X on Ω is equal to

$$E(X) = \sum_{\omega \in \Omega} P(\omega) X(\omega).$$

Expected value

The **expected value** of a r.v. is the sum over all elements in a sample space of the product of the probability of an its element and the value of the r.v. at this element.

Definition

The **expected value**, also called **expectation** or **mean**, of r.v. X on Ω is equal to

$$E(X) = \sum_{\omega \in \Omega} P(\omega) X(\omega).$$

The **deviation** of X at $\omega \in \Omega$ is $X(\omega) - E(X)$, the difference between the value of X and the mean of X.

Expected value

The **expected value** of a r.v. is the sum over all elements in a sample space of the product of the probability of an its element and the value of the r.v. at this element.

Definition

The **expected value**, also called **expectation** or **mean**, of r.v. X on Ω is equal to

$$E(X) = \sum_{\omega \in \Omega} P(\omega) X(\omega).$$

The **deviation** of X at $\omega \in \Omega$ is $X(\omega) - E(X)$, the difference between the value of X and the mean of X.

- The expected value is a weighted average of the values of a r.v.;
- The expected value of a r.v. provides a central point for the distribution of values of this r.v..

Expected value

Theorem

If X is a r.v. and $P(X = r)$ is the probability that $X = r$, so that $P(X = r) = \sum_{\omega \in \Omega, X(\omega) = r} P(\omega)$, then

$$E(X) = \sum_{r \in X(\Omega)} P(X = r) \cdot r.$$

Proof.

Suppose that X is a r.v. with range $X(\Omega)$, and let $P(X = r)$ be the probability that r.v. X takes value r.

Expected value

Theorem

If X is a r.v. and $P(X = r)$ is the probability that $X = r$, so that $P(X = r) = \sum_{\omega\in\Omega, X(\omega)=r} P(\omega)$, then

$$E(X) = \sum_{r\in X(\Omega)} P(X = r) \cdot r.$$

Proof.

Suppose that X is a r.v. with range $X(\Omega)$, and let $P(X = r)$ be the probability that r.v. X takes value r.
Consequently, $P(X = r)$ is the sum of the probabilities of the outcomes ω such that $X(\omega) = r$. It follows that

$$E(X) = \sum_{r\in X(\Omega)} P(X = r) \cdot r.$$

Example I

Expected value of tossing coin

Question: A coin is flipped one time. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome $\#$ heads in this outcome. What is the expected value of X if it is a fair coin? What is the expected value of X if it is a biased coin with $P(\{H\}) = p$?

Example I

Expected value of tossing coin

Question: A coin is flipped one time. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome # heads in this outcome. What is the expected value of X if it is a fair coin? What is the expected value of X if it is a biased coin with $P(\{H\}) = p$?

Solution:

$$E(X) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2},$$
$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Note that $E(X^2) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$.

Example II

Expected value of tossing coin

Question: A fair coin is flipped 4 times. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome $\#$ heads in this outcome. What is the expected value of X?

Example II

Expected value of tossing coin

Question: A fair coin is flipped 4 times. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome # heads in this outcome. What is the expected value of X?

Solution:

$$E(X) = 4 \cdot \frac{1}{16} + 3 \cdot \frac{1}{4} + 2 \cdot \frac{3}{8} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{16}$$
$$= \frac{4 + 12 + 12 + 4}{16} = 2$$

Expected value of Binomial r.v.s

Theorem

The expected number of successes when n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Expected value of Binomial r.v.s

Theorem

The expected number of successes when n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Proof.

Let X be the r.v. equal to the number of successes in n trials. We have known that $P(X = k) = C(n,k)p^k q^{n-k}$. Hence, we have

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k \cdot P(X = k) = \sum_{k=1}^{n} k \cdot C(n,k)p^k q^{n-k} \\ &= \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} p^k q^{n-k} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \\ &= np(p+q)^{n-1} = np \end{aligned}$$

$\square$

Expected value of Geometric r.v.s

Theorem

The expected number of successes when a r.v. X follows a Geometric distribution is $\frac{1}{p}$, where p is the probability of success on each trial.

Expected value of Geometric r.v.s

Theorem

The expected number of successes when a r.v. X follows a Geometric distribution is $\frac{1}{p}$, where p is the probability of success on each trial.

Proof.

We have known that $P(X = k) = q^{k-1}p$. Hence, we have

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k \cdot q^{k-1}p = p(\sum_{m=1}^{\infty}\sum_{k=m}^{\infty} q^{k-1}) \\
&= p(\sum_{m=1}^{\infty} \frac{q^{m-1}}{1-q}) = \sum_{m=1}^{\infty} q^{m-1} \\
&= \frac{1}{1-q} = \frac{1}{p}
\end{aligned}
$$

□

Expected value of Geometric r.v.s

Theorem

The expected number of successes when a r.v. X follows a Geometric distribution is $\frac{1}{p}$, where p is the probability of success on each trial.

Proof.

We have known that $P(X=k)=q^{k-1}p$. Hence, we have

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot q^{k-1} p = p(\sum_{m=1}^{\infty}\sum_{k=m}^{\infty} q^{k-1}) \\ &= p(\sum_{m=1}^{\infty} \frac{q^{m-1}}{1-q}) = \sum_{m=1}^{\infty} q^{m-1} \\ &= \frac{1}{1-q} = \frac{1}{p} \end{aligned}$$

$\square$

Example III

Expected value of a dice

Question: Let X be the number that comes up when a fair dice is rolled. What is the expected value of X?

Example III

Expected value of a dice

Question: Let X be the number that comes up when a fair dice is rolled. What is the expected value of X?

Solution:

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$
$$= \frac{21}{6} = \frac{7}{2}$$

Example IV

Expected value of two dices

Question: What is the expected value of the sum of the numbers that appear when a pair of fair dice is rolled?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Example IV

Expected value of two dices

Question: What is the expected value of the sum of the numbers that appear when a pair of fair dice is rolled?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Solution I:

$$E(X) = (2+12)\cdot\frac{1}{36} + (3+11)\cdot\frac{1}{18} + (4+10)\cdot\frac{1}{12}$$
$$+ (5+9)\cdot\frac{1}{9} + (6+8)\cdot\frac{5}{36} + 7\cdot\frac{1}{6} = 7.$$

Example IV Cont'd

Expected value of two dices

Solution II:
Let X_1 and X_2 be the numbers that comes up when the first and the second dices is rolled. How to compute the expected value of $X_1 + X_2$?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$
\begin{aligned}
E(X_1 + X_2) &= E(X_1) + E(X_2) = 2E(X_1) \\
&= 2 \cdot \frac{1}{6} \cdot (1+2+3+4+5+6) = 7.
\end{aligned}
$$

Hence we have

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Linearity of expectations

Theorem

If X_i, $i = 1, 2, \cdots, n$ with n a positive integer, are random variables on Ω, and if a and b are real numbers, then

- $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$;
- $E(aX_i + b) = aE(X_i) + b$.

Proof.

$$E(\sum_{i=1}^{n} X_i) = \sum_{\omega\in\Omega} P(\omega)(\sum_{i=1}^{n} X_i(\omega)) = \sum_{\omega\in\Omega}\sum_{i=1}^{n}(P(\omega)\cdot X_i(\omega)) = \sum_{i=1}^{n} E(X_i)$$

$$\begin{aligned} E(aX_i + b) &= \sum_{\omega\in\Omega} P(\omega)(aX_i(\omega) + b) \\ &= a\sum_{\omega\in\Omega} P(\omega)\cdot X_i(\omega) + b\sum_{\omega\in\Omega} P(\omega) = aE(X_i) + b \end{aligned}$$

$\square$

Expected value of Bernoulli trials

Proof with linearity of expectations

The expected number of successes when n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Expected value of Bernoulli trials

Proof with linearity of expectations

The expected number of successes when n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Proof.

Let X_i be # heads in the $i-$th Bernoulli trial, and X be the number of successes in n mutually independent Bernoulli trials. Hence we have $X = \sum_{i=1}^{n} X_i$, and $E(X_i) = p$.

$$E(X) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = np.$$

□

Expected value in the Hatcheck problem

Question: A new employee checks the hats of n people at a restaurant, forgetting to put claim check numbers on the hats. When customers return for their hats, the checker gives them back hats chosen at random from the remaining hats. What is the expected number of hats that are returned correctly?

Solution:

Expected value in the Hatcheck problem

Question: A new employee checks the hats of n people at a restaurant, forgetting to put claim check numbers on the hats. When customers return for their hats, the checker gives them back hats chosen at random from the remaining hats. What is the expected number of hats that are returned correctly?

Solution:

Let X be the random variable that equals the number of people who receive the correct hat from the checker. Let X_i be the random variable with $X_i = 1$ if the $i-$th person receives the correct hat, otherwise $X_i = 0$. It follows that $X = \sum_{i=1}^{n} X_i$.

Expected value in the Hatcheck problem

Question: A new employee checks the hats of n people at a restaurant, forgetting to put claim check numbers on the hats. When customers return for their hats, the checker gives them back hats chosen at random from the remaining hats. What is the expected number of hats that are returned correctly?

Solution:

Let X be the random variable that equals the number of people who receive the correct hat from the checker. Let X_i be the random variable with $X_i = 1$ if the $i-$th person receives the correct hat, otherwise $X_i = 0$. It follows that $X = \sum_{i=1}^{n} X_i$.

Note that

$$E(X_i) = 1 \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n} = \frac{1}{n}.$$

Expected value in the Hatcheck problem

Question: A new employee checks the hats of n people at a restaurant, forgetting to put claim check numbers on the hats. When customers return for their hats, the checker gives them back hats chosen at random from the remaining hats. What is the expected number of hats that are returned correctly?

Solution:

Let X be the random variable that equals the number of people who receive the correct hat from the checker. Let X_i be the random variable with $X_i = 1$ if the $i-$th person receives the correct hat, otherwise $X_i = 0$. It follows that $X = \sum_{i=1}^{n} X_i$.

Note that

$$E(X_i) = 1 \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n} = \frac{1}{n}.$$

Hence, $E(X) = \sum_{i=1}^{n} E(X_i) = 1$.

Expected number of inversions in a permutation

Question: The ordered pair (i, j) is called an inversion in a permutation of the first n positive integers if $i < j$ but j precedes i in the permutation. For instance, there are six inversions in the permutation $3, 5, 1, 4, 2$; these inversions are $(1, 3), (1, 5), (2, 3), (2, 4), (2, 5), (4, 5)$.

Solution:

Expected number of inversions in a permutation

Question: The ordered pair (i, j) is called an inversion in a permutation of the first n positive integers if $i < j$ but j precedes i in the permutation. For instance, there are six inversions in the permutation $3, 5, 1, 4, 2$; these inversions are $(1, 3), (1, 5), (2, 3), (2, 4), (2, 5), (4, 5)$.

Solution:

Let $I_{i,j}$ be the r.v. on the set of all permutations of the first n positive integers with $I_{i,j} = 1$ if (i, j) is an inversion and $I_{i,j} = 0$ otherwise. It follows that if X is the r.v. equal to $\#$ inversions in the permutation, then $X = \sum_{1 \le i < j \le n}^{n} I_{i,j}$.

Expected number of inversions in a permutation

Question: The ordered pair (i,j) is called an inversion in a permutation of the first n positive integers if $i < j$ but j precedes i in the permutation. For instance, there are six inversions in the permutation $3,5,1,4,2$; these inversions are $(1,3),(1,5),(2,3),(2,4),(2,5),(4,5)$.

Solution:

Let $I_{i,j}$ be the r.v. on the set of all permutations of the first n positive integers with $I_{i,j}=1$ if (i,j) is an inversion and $I_{i,j}=0$ otherwise. It follows that if X is the r.v. equal to # inversions in the permutation, then $X = \sum_{1\le i<j\le n}^{n} I_{i,j}$.

Note that

$$E(I_{i,j}) = 1\cdot\frac{1}{2} + 0\cdot\frac{1}{2} = \frac{1}{2}.$$

Expected number of inversions in a permutation

Question: The ordered pair (i,j) is called an inversion in a permutation of the first n positive integers if $i < j$ but j precedes i in the permutation. For instance, there are six inversions in the permutation $3,5,1,4,2$; these inversions are $(1,3),(1,5),(2,3),(2,4),(2,5),(4,5)$.

Solution:

Let $I_{i,j}$ be the r.v. on the set of all permutations of the first n positive integers with $I_{i,j}=1$ if (i,j) is an inversion and $I_{i,j}=0$ otherwise. It follows that if X is the r.v. equal to # inversions in the permutation, then $X=\sum_{1\leq i<j\leq n}^{n} I_{i,j}$.

Note that

$$E(I_{i,j}) = 1\cdot\frac{1}{2}+0\cdot\frac{1}{2}=\frac{1}{2}.$$

Hence, $E(X)=\sum_{1\leq i<j\leq n}^{n} E(I_{i,j})=\frac{n(n-1)}{4}$.

Expectation of independent r.v.s

Theorem

If X and Y are independent r.v.s on a sample space Ω, then

$$E(XY) = E(X)E(Y).$$

Proof: To prove this formula, we use the key observation that event $XY = r$ is the disjoint union of events $X = r_1$ and $Y = r_2$ over all $r_1 \in X(\Omega)$ and $r_2 \in Y(\Omega)$ with $r = r_1 r_2$. We have

$$E(XY) = \sum_{r \in XY(\Omega)} r \cdot P(XY = r) = \sum_{r_1 \in X(\Omega), r_2 \in Y(\Omega)} r_1 r_2 \cdot P(X = r_1 \wedge Y = r_2)$$

$$= \sum_{r_1 \in X(\Omega)} \sum_{r_2 \in Y(\Omega)} (r_1 \cdot P(X = r_1))(r_2 \cdot P(Y = r_2))$$

$$= \sum_{r_1 \in X(\Omega)} (r_1 \cdot P(X = r_1) \sum_{r_2 \in Y(\Omega)} (r_2 \cdot P(Y = r_2))) = \sum_{r_1 \in X(\Omega)} (r_1 \cdot P(X = r_1) E(Y))$$

$$= E(Y) \sum_{r_1 \in X(\Omega)} r_1 \cdot P(X = r_1) = E(X)E(Y).$$

Average-case computational complexity

Computing the average-case computational complexity of an algorithm can be interpreted as computing the expected value of a r.v.. Let the sample space of an experiment be the set of possible inputs a_j, $j = 1, 2, \cdots, n$, and let X be the r.v. that assigns # operations used by the algorithm when given a_j as input. Based on our knowledge of the input, we assign a probability $P(a_j)$ to each possible input value a_j . Then, the **average-case complexity** of the algorithm is

$$E(X) = \sum_{j=1}^{n} P(a_j)X(a_j).$$

Average-case computational complexity

Computing the average-case computational complexity of an algorithm can be interpreted as computing the expected value of a r.v.. Let the sample space of an experiment be the set of possible inputs a_j, $j = 1, 2, \cdots, n$, and let X be the r.v. that assigns # operations used by the algorithm when given a_j as input. Based on our knowledge of the input, we assign a probability $P(a_j)$ to each possible input value a_j . Then, the **average-case complexity** of the algorithm is

$$E(X) = \sum_{j=1}^{n} P(a_j)X(a_j).$$

Average-case computational complexity

Computing the average-case computational complexity of an algorithm can be interpreted as computing the expected value of a r.v.. Let the sample space of an experiment be the set of possible inputs a_j, $j = 1, 2, \cdots, n$, and let X be the r.v. that assigns # operations used by the algorithm when given a_j as input. Based on our knowledge of the input, we assign a probability $P(a_j)$ to each possible input value a_j . Then, the **average-case complexity** of the algorithm is

$$E(X) = \sum_{j=1}^{n} P(a_j)X(a_j).$$

Finding the average-case computational complexity of an algorithm is usually much more difficult than finding its worst-case computational complexity, and often involves the use of sophisticated methods.

Average-case complexity of the linear search algorithm

Question: What is the average-case computational complexity of the linear search algorithm if the probability that x is in the list is p and it is equally likely that x is any of the n elements in the list?

Average-case complexity of the linear search algorithm

Question: What is the average-case computational complexity of the linear search algorithm if the probability that x is in the list is p and it is equally likely that x is any of the n elements in the list?

input
  x: integer
  $a_1, a_2, \ldots, a_n$: distinct integers
procedure *linear search*
$i := 1$
while ($i \leq n$ and $x \neq a_i$)
  $i := i + 1$
if $i \leq n$ **then** *location* $:= i$
else *location* $:= 0$
return *location*

Average-case complexity of the linear search algorithm

Question: What is the average-case computational complexity of the linear search algorithm if the probability that x is in the list is p and it is equally likely that x is any of the n elements in the list?

input
  x: integer
  $a_1, a_2, \ldots, a_n$: distinct integers
procedure *linear search*
$i := 1$
while ($i \leq n$ and $x \neq a_i$)
    $i := i + 1$
if $i \leq n$ **then** *location* $:= i$
else *location* $:= 0$
return *location*

Solution:
We known that $2i + 1$ comparisons are used if x equals the $i-$th element of the list, and $2n + 2$ comparisons are used if x is not in the list.

Average-case complexity of the linear search algorithm

Question: What is the average-case computational complexity of the linear search algorithm if the probability that x is in the list is p and it is equally likely that x is any of the n elements in the list?

input
  x: integer
  $a_1, a_2, \ldots, a_n$: distinct integers
procedure *linear search*
$i := 1$
while ($i \leq n$ and $x \neq a_i$)
  $i := i + 1$
if $i \leq n$ **then** *location* $:= i$
else *location* $:= 0$
return *location*

Solution:
We known that $2i + 1$ comparisons are used if x equals the $i-$th element of the list, and $2n + 2$ comparisons are used if x is not in the list.
The probability that x equals a_i , the $i-$th element in the list, is p/n, and the probability that x is not in the list is $q = 1 - p$.

Average-case complexity of the linear search algorithm Cont'd

It follows that the average-case computational complexity of the linear search algorithm is

$$\begin{aligned} E &= \frac{3p}{n} + \frac{5p}{n} + \cdots + \frac{(2n+1)p}{n} + (2n+2)q \\ &= \frac{p}{n}(3 + 5 + \cdots + (2n+1)) + (2n+2)q \\ &= \frac{p}{n}((n+1)^2 - 1) + (2n+2)q \\ &= p(n+2) + (2n+2)q. \end{aligned}$$

Average-case complexity of the linear search algorithm Cont'd

It follows that the average-case computational complexity of the linear search algorithm is

$$\begin{aligned} E &= \frac{3p}{n} + \frac{5p}{n} + \cdots + \frac{(2n+1)p}{n} + (2n+2)q \\ &= \frac{p}{n}(3 + 5 + \cdots + (2n+1)) + (2n+2)q \\ &= \frac{p}{n}((n+1)^2 - 1) + (2n+2)q \\ &= p(n+2) + (2n+2)q. \end{aligned}$$

When $p = 1$, then $q = 0$ and $E = n + 2$.

Average-case complexity of the linear search algorithm Cont'd

It follows that the average-case computational complexity of the linear search algorithm is

$$
\begin{aligned}
E &= \frac{3p}{n} + \frac{5p}{n} + \cdots + \frac{(2n+1)p}{n} + (2n+2)q \\
&= \frac{p}{n}(3 + 5 + \cdots + (2n+1)) + (2n+2)q \\
&= \frac{p}{n}((n+1)^2 - 1) + (2n+2)q \\
&= p(n+2) + (2n+2)q.
\end{aligned}
$$

When $p = 1$, then $q = 0$ and $E = n + 2$.
When $p = 0$, then $q = 1$ and $E = 2(n+1)$.

Average-case complexity of the insertion sort

Question: What is the average number of comparisons used by the insertion sort to sort n distinct elements?

Average-case complexity of the insertion sort

Question: What is the average number of comparisons used by the insertion sort to sort *n* distinct elements?

```
input
  a_1, a_2, ..., a_n: real numbers with n ≥ 2
procedure insertion sort
for j := 2 to n
    i := 1
    while a_j > a_i
        i := i + 1
    m := a_j
    for k := 0 to j − i − 1
        a_{j−k} := a_{j−k−1}
    a_i := m
{a_1, ..., a_n is in increasing order}
```

Average-case complexity of the insertion sort

Question: What is the average number of comparisons used by the insertion sort to sort n distinct elements?

```
input
  a_1, a_2, ..., a_n: real numbers with n ≥ 2
procedure insertion sort
for j := 2 to n
      i := 1
      while a_j > a_i
            i := i + 1
      m := a_j
      for k := 0 to j − i − 1
            a_{j−k} := a_{j−k−1}
      a_i := m
{a_1, ..., a_n is in increasing order}
```

Solution:
We first suppose that X is the r.v. equal to # comparisons used by the insertion sort to sort a list $a_1, a_2, \cdots, a_n$ of n distinct elements.

Average-case complexity of the insertion sort

Question: What is the average number of comparisons used by the insertion sort to sort n distinct elements?

input
  $a_1, a_2, \ldots, a_n$: real numbers with $n \geq 2$
procedure *insertion sort*
for $j := 2$ **to** n
  $i := 1$
  while $a_j > a_i$
    $i := i + 1$
  $m := a_j$
  for $k := 0$ **to** $j - i - 1$
    $a_{j-k} := a_{j-k-1}$
  $a_i := m$
$\{a_1, \ldots, a_n$ is in increasing order$\}$

Solution:
We first suppose that X is the r.v. equal to # comparisons used by the insertion sort to sort a list $a_1, a_2, \cdots, a_n$ of n distinct elements.
Then $E(X)$ is the average number of comparisons used.

Average-case complexity of the insertion sort Cont'd

We let X_i be the r.v. equal to # comparisons used to insert a_i into the proper position after the first $i-1$ elements $a_1, a_2, \cdots, a_{i-1}$ have been sorted. Furthermore, we have $X = \sum_{i=2}^{n} X_i$.

Let $p_j(k)$ denote the probability that the largest of the first j elements in the list occurs at the $k-$th position, that is, that $\max(a_1, a_2, \cdots, a_j) = a_k$, where $1 \le k \le j$.

$$E(X_i) = \sum_{k=1}^{i} p_i(k) \cdot X_i(k) = \sum_{k=1}^{i} \frac{1}{i} \cdot k = \frac{i+1}{2}.$$

$$\begin{aligned} E(X) &= \sum_{i=2}^{n} E(X_i) = \sum_{i=2}^{n} \frac{i+1}{2} = \frac{1}{2} \sum_{i=3}^{n+1} j \\ &= \frac{1}{2} \frac{(n+1)(n+2)}{2} - \frac{1}{2}(1+2) = \frac{n^2+3n-4}{4}. \end{aligned}$$

Thus, the average number of comparisons used by the insertion sort to sort n elements equals $(n^2+3n-4)/4$, which is $\Theta(n^2)$.

Variance

Definition

Let X be a r.v. on a sample space Ω. The **variance** of X, denoted by $V(X)$, is

$$V(X) = \sum_{\omega \in \Omega} (X(\omega) - E(X))^2 \cdot P(\omega),$$

i.e., $V(X)$ is the weighted average of the square of the deviation of X. The standard deviation of X, denoted $\delta(X)$, is defined as $\sqrt{V(X)}$.

- $V(X) \geq 0$;
- $V(X) = \sum_{x \in X(\Omega)} (x - E(X))^2 \cdot P(X = x)$;
- Informally, it measures how far a set of (random) numbers are spread out from their average value.

Theorem

If X is a r.v. on a sample space Ω, then

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

Proof.

Note that $\sum_{\omega\in\Omega} P(\omega) = 1$, we therefore have

$$\begin{aligned} V(X) &= \sum_{\omega\in\Omega} (X(\omega) - E(X))^2 P(\omega) \\ &= \sum_{\omega\in\Omega} X^2(\omega)P(\omega) - 2E(X)\sum_{\omega\in\Omega} X(\omega)P(\omega) + (E(X))^2 \sum_{\omega\in\Omega} P(\omega) \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2. \end{aligned}$$

$\square$

Corollary

If X is a random variable on a sample space Ω and $E(X) = \mu$, then $V(X) = E((X - \mu)^2)$.

Proof.

Note that $\sum_{\omega \in \Omega} P(\omega) = 1$, we therefore have

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

□

The corollary tells us that the variance of a r.v. X is the expected value of the square of the difference between X and its own expected value.

Example I

Variance of Bernoulli trial

Question: A coin is flipped one time. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome # heads in this outcome. What is the variance of X if it is a biased coin with $P(\{H\}) = p$?

Example I

Variance of Bernoulli trial

Question: A coin is flipped one time. Let Ω be the sample space of the possible outcomes, and let X be r.v. that assigns to an outcome # heads in this outcome. What is the variance of X if it is a biased coin with $P(\{H\}) = p$?

Solution:

$$E(X^2) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$$
$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$
$$V(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p)$$

Example II

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Example II

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?
Solution:

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \cdot P(X=k) = \sum_{k=1}^{n} k(k-1)\cdot P(X=k) + \sum_{k=1}^{n} k \cdot P(X=k) \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{n-k} + np \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j q^{n-2-j} + np \\
&= n(n-1)p^2 (p+q)^{n-2} + np = n(n-1)p^2 + np, \\
V(X) &= E(X^2) - (E(X))^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).
\end{aligned}$$

Example III

Variance of Geometric r.v.s

Question: Let r.v. X be the first occurrence of success requires n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Example III

Variance of Geometric r.v.s

Question: Let r.v. X be the first occurrence of success requires n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Solution:

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \cdot P(X=k) = \sum_{k=1}^{\infty} k(k-1) \cdot P(X=k) + \sum_{k=1}^{\infty} k \cdot P(X=k)$$

$$= \sum_{k=1}^{\infty} k(k-1) \cdot q^{k-1}p + \sum_{k=1}^{\infty} k \cdot q^{k-1}p = \sum_{k=1}^{\infty} k(k-1) \cdot q^{k-1}p + \frac{1}{p}$$

$$\sum_{k=1}^{\infty} k(k-1) \cdot q^{k-1}p = p\sum_{k=2}^{\infty}(2\sum_{j=1}^{k-1} j)q^{k-1} = 2p\sum_{k=2}^{\infty}\sum_{j=1}^{k-1}(jq^{k-1})$$

$$= 2p\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}(jq^{k-1}) = 2p\sum_{j=1}^{\infty}\left[jq^j \sum_{k=j+1}^{\infty}(q^{k-j-1})\right]$$

Example III

Variance of Geometric r.v.s

Question: Let r.v. X be the first occurrence of success requires n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?
Solution:

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \cdot P(X = k) = \sum_{k=1}^{\infty} k(k-1) \cdot P(X = k) + \sum_{k=1}^{\infty} k \cdot P(X = k)$$

$$= \sum_{k=1}^{\infty} k(k-1) \cdot q^{k-1} p + \sum_{k=1}^{\infty} k \cdot q^{k-1} p = \sum_{k=1}^{\infty} k(k-1) \cdot q^{k-1} p + \frac{1}{p}$$

$$\sum_{k=1}^{\infty} k(k-1) \cdot q^{k-1} p = p \sum_{k=2}^{\infty} (2 \sum_{j=1}^{k-1} j) q^{k-1} = 2p \sum_{k=2}^{\infty} \sum_{j=1}^{k-1} (j q^{k-1})$$

$$= 2p \sum_{j=1}^{\infty} \sum_{k=j+1}^{\infty} (j q^{k-1}) = 2p \sum_{j=1}^{\infty} \left[j q^j \sum_{k=j+1}^{\infty} (q^{k-j-1}) \right]$$

Example III Cont'd

Variance of Geometric r.v.s

$$= 2p\sum_{j=1}^{\infty}\left[jq^j\sum_{k=j+1}^{\infty}(q^{k-j-1})\right]$$

$$= 2p\sum_{j=1}^{\infty}\left[jq^j\sum_{k=0}^{\infty}(q^k)\right] = 2p\sum_{j=1}^{\infty}\frac{jq^j}{1-q}$$

$$= 2\sum_{j=1}^{\infty}jq^j = \frac{2q}{p}\sum_{j=1}^{\infty}jq^{j-1}p = \frac{2q}{p}\cdot E(X) = \frac{2q}{p^2}$$

$$E(X^2) = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2}$$

$$V(X) = \frac{2q+p}{p^2} - (\frac{1}{p})^2 = \frac{2q-(1-p)}{p^2} = \frac{q}{p^2}.$$

Nonlinearity of variance

Theorem

If X is a r.v. on Ω, and if a and b are real numbers, then

$$V(aX+b) = a^2 V(X).$$

Proof.

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

$\square$

Bienaymé's formula

Theorem

Question: If X and Y are two independent r.v.s on a sample space Ω, then $V(X+Y) = V(X) + V(Y)$. Furthermore, if X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are pairwise independent r.v.s on Ω, then

$$V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i).$$

Proof:

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

Example II Cont'd

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Example II Cont'd

Variance of Binomial r.v.s

Question: Let r.v. X be the number of successes of n mutually independent Bernoulli trials, where p is the probability of success on each trial. What is the variance of X?

Solution:

Let X_i be the number of success in the $i-$th Bernoulli trial. Thus, we have $X = \sum_{i=1}^{n} X_i$, X_i and X_j are independent for $i \neq j$.

$$E(X_i^2) = 1^2 p + 0^2(1-p) = p;$$

$$V(X_i) = E(X_i^2) - (E(X_i))^2 = p - p^2 = p(1-p);$$

$$V(X) = \sum_{i=1}^{n} V(X_i) = np(1-p).$$

Example IV

Variance of Binomial r.v.s

Question: Let r.v.s X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are independent and identical distribution r.v.s with $V(X_i) = \sigma^2$. What is the variance of $\frac{1}{n}\sum_{i=1}^{n} X_i$?

Example IV

Variance of Binomial r.v.s

Question: Let r.v.s X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are independent and identical distribution r.v.s with $V(X_i) = \sigma^2$. What is the variance of $\frac{1}{n}\sum_{i=1}^n X_i$?

Solution:

$$V(\frac{1}{n}\sum_{i=1}^n X_i) = (\frac{1}{n})^2 V(\sum_{i=1}^n X_i) = (\frac{1}{n})^2 \sum_{i=1}^n V(X_i)$$

$$= (\frac{1}{n})^2 n\sigma^2 = \frac{\sigma^2}{n}$$

Example IV

Variance of Binomial r.v.s

Question: Let r.v.s X_i, $i = 1, 2, \cdots, n$, with n a positive integer, are independent and identical distribution r.v.s with $V(X_i) = \sigma^2$. What is the variance of $\frac{1}{n}\sum_{i=1}^{n} X_i$?

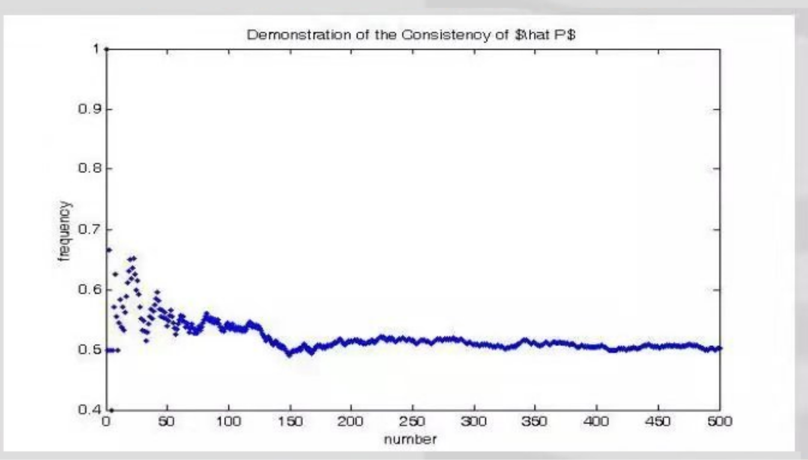
Solution:

$$V(\frac{1}{n}\sum_{i=1}^{n} X_i) = (\frac{1}{n})^2 V(\sum_{i=1}^{n} X_i) = (\frac{1}{n})^2 \sum_{i=1}^{n} V(X_i)$$
$$= (\frac{1}{n})^2 n\sigma^2 = \frac{\sigma^2}{n}$$

That is, the variance of the mean decreases when n increases. It is a good property of variance.

Motivated example

Tossing a fair coin



Markovs inequality

Theorem

Let X be a nonnegative r.v. on a sample space Ω with probability function p. If a is a positive real number, then

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

Proof: Let A be event $A = \{\omega \in \Omega | X \geq a\}$.

Markovs inequality

Theorem

Let X be a nonnegative r.v. on a sample space Ω with probability function p. If a is a positive real number, then

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

Proof: Let A be event $A = \{\omega \in \Omega | X \geq a\}$.

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{\omega \in A} X(\omega)P(\omega) + \sum_{\omega \notin A} X(\omega)P(\omega)$$

$$\geq a \cdot P(A) + \sum_{\omega \notin A} X(\omega)P(\omega) \geq a \cdot P(A).$$

Markovs inequality

Theorem

Let X be a nonnegative r.v. on a sample space Ω with probability function p. If a is a positive real number, then

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

Proof: Let A be event $A = \{\omega \in \Omega | X \geq a\}$.

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{\omega \in A} X(\omega)P(\omega) + \sum_{\omega \notin A} X(\omega)P(\omega)$$
$$\geq a \cdot P(A) + \sum_{\omega \notin A} X(\omega)P(\omega) \geq a \cdot P(A).$$

Hence, we have

$$P(A) = P(X \geq a) \leq \frac{E(X)}{a}.$$

Chebyshevs inequality

Theorem

Let X be a r.v. on a sample space Ω with probability function p. If r is a positive real number, then

$$P(|X(\omega) - E(X)| \geq r) \leq \frac{V(X)}{r^2}.$$

Proof: Let r.v. Y be $Y = |X(\omega) - E(X)|^2$.

Chebyshevs inequality

Theorem

Let X be a r.v. on a sample space Ω with probability function p. If r is a positive real number, then

$$P(|X(\omega) - E(X)| \geq r) \leq \frac{V(X)}{r^2}.$$

Proof: Let r.v. Y be $Y = |X(\omega) - E(X)|^2$.

$$\begin{aligned} P(|X(\omega) - E(X)| \geq r) &= P(|X(\omega) - E(X)|^2 \geq r^2) = P(Y \geq r^2) \\ &\leq \frac{E(Y)}{r^2} = \frac{E(X(\omega) - E(X))^2}{r^2} = \frac{V(X)}{r^2}. \end{aligned}$$

Chebyshevs inequality

Theorem

Let X be a r.v. on a sample space Ω with probability function p. If r is a positive real number, then

$$P(|X(\omega) - E(X)| \geq r) \leq \frac{V(X)}{r^2}.$$

Proof: Let r.v. Y be $Y = |X(\omega) - E(X)|^2$.

$$\begin{aligned} P(|X(\omega) - E(X)| \geq r) &= P(|X(\omega) - E(X)|^2 \geq r^2) = P(Y \geq r^2) \\ &\leq \frac{E(Y)}{r^2} = \frac{E(X(\omega) - E(X))^2}{r^2} = \frac{V(X)}{r^2}. \end{aligned}$$

Example:

Let X be the number of heads in n tosses of a fair coin, then $E(X) = np$, $V(X) = np(1-p)$ and $p = \frac{1}{2}$, we have

$$P(X > \frac{3n}{4}) = P(X - \frac{n}{2} > \frac{n}{4}) < P(|X - \frac{n}{2}| > \frac{n}{4}) < \frac{16np(1-p)}{n^2} = \frac{4}{n}$$

If we toss the coin 1000 times, the probability is less than 0.004.

Chernoff bound

Theorem

Let X_i be a sequence of independent Bernoulli r.v.s with $P(X_i = 1) = p_i$. Assume that r.v. $X = \sum_{i=1}^n X_i$.

- $P(X < (1-\delta)\mu) < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$, where $\mu = \sum_{i=1}^n p_i$
- $P(X < (1-\delta)\mu) < \exp(-\mu\delta^2/2)$

Proof

For $t > 0$,

$$\begin{aligned} P(X < (1-\delta)\mu) &= P\big(\exp(-tX) > \exp(-t(1-\delta)\mu)\big) \\ &< \frac{\Pi_{i=1}^n E(\exp(-tX_i))}{\exp(-t(1-\delta)\mu)} (Markov\ \ inequality) \end{aligned}$$

Proof of Chernoff bound Cont.

Proof Cont'd

Since $(1 - x < e^{-x})$, we have

$$E(\exp(-tX_i)) = p_i e^{-t} + (1 - p_i) = 1 - p_i(1 - e^{-t}) < \exp(p_i(e^{-t} - 1))$$

$$\Pi_{i=1}^{n} E(\exp(-tX_i)) < \Pi_{i=1}^{n} \exp(p_i(e^{-t} - 1)) = \exp(\mu(e^{-t} - 1))$$

Proof of Chernoff bound Cont.

Proof Cont'd

Since $(1 - x < e^{-x})$, we have

$$E(\exp(-tX_i)) = p_i e^{-t} + (1 - p_i) = 1 - p_i(1 - e^{-t}) < \exp(p_i(e^{-t} - 1))$$
$$\Pi_{i=1}^{n} E(\exp(-tX_i)) < \Pi_{i=1}^{n} \exp(p_i(e^{-t} - 1)) = \exp(\mu(e^{-t} - 1))$$

Hence,

$$\begin{aligned} P(X < (1-\delta)\mu) &< \frac{\exp(\mu(e^{-t} - 1))}{\exp(-t(1-\delta)\mu)} \\ &= \exp(\mu(e^{(-t)} + t - t\delta - 1)) \end{aligned}$$

Proof of Chernoff bound Cont.

Proof Cont'd

Since $(1 - x < e^{-x})$, we have

$$E(\exp(-tX_i)) = p_i e^{-t} + (1 - p_i) = 1 - p_i(1 - e^{-t}) < \exp(p_i(e^{-t} - 1))$$
$$\Pi_{i=1}^{n} E(\exp(-tX_i)) < \Pi_{i=1}^{n} \exp(p_i(e^{-t} - 1)) = \exp(\mu(e^{-t} - 1))$$

Hence,

$$\begin{aligned} P(X < (1-\delta)\mu) &< \frac{\exp(\mu(e^{-t} - 1))}{\exp(-t(1-\delta)\mu)} \\ &= \exp(\mu(e^{(-t)} + t - t\delta - 1)) \end{aligned}$$

Now its time to choose t to make the bound as tight as possible. Taking the derivative of $\mu(e^{(-t)} + t - t\delta - 1)$ and setting $-e^{(-t)} + 1 - \delta = 0$. We have $t = \ln(1/1 - \delta)$.

$$P(X < (1-\delta)\mu) < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}.$$

Proof of Chernoff bound Cont.

Proof of second statement

To get the simpler form of the bound, we need to get rid of the clumsy term $(1-\delta)^{(1-\delta)}$.

Proof of Chernoff bound Cont.

Proof of second statement

To get the simpler form of the bound, we need to get rid of the clumsy term $(1-\delta)^{(1-\delta)}$.

$$(1-\delta)\ln(1-\delta) = (1-\delta)\left(\sum_{i=1}^{\infty} -\frac{\delta^i}{i}\right) > -\delta + \frac{\delta^2}{2}$$

$$(1-\delta)^{(1-\delta)} > \exp\left(-\delta + \frac{\delta^2}{2}\right)$$

Proof of Chernoff bound Cont.

Proof of second statement

To get the simpler form of the bound, we need to get rid of the clumsy term $(1-\delta)^{(1-\delta)}$.

$$(1-\delta)\ln(1-\delta) = (1-\delta)(\sum_{i=1}^{\infty} -\frac{\delta^i}{i}) > -\delta + \frac{\delta^2}{2}$$

$$(1-\delta)^{(1-\delta)} > \exp\left(-\delta + \frac{\delta^2}{2}\right)$$

Furthermore,

$$\begin{aligned} P(X < (1-\delta)\mu) &< \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \\ &< \left(\frac{e^{-\delta}}{\exp\left(-\delta + \frac{\delta^2}{2}\right)}\right)^{\mu} \\ &= \exp\left(-\mu\delta^2/2\right) \end{aligned}$$

Chernoff bound (Upper tail)

Theorem for upper tail

Let X_i be a sequence of independent and Bernoulli r.v.s with $P(X_i = 1) = p_i$. Assume that r.v. $X = \sum_{i=1}^{n} X_i$ and $\mu = \sum_{i=1}^{n} p_i$.

- $P(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$
- $P(X > (1+\delta)\mu) < \exp\left(-\mu\delta^2/4\right)$

Chernoff bound (Upper tail)

Theorem for upper tail

Let X_i be a sequence of independent and Bernoulli r.v.s with $P(X_i = 1) = p_i$. Assume that r.v. $X = \sum_{i=1}^{n} X_i$ and $\mu = \sum_{i=1}^{n} p_i$.

- $P(X > (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$
- $P(X > (1+\delta)\mu) < \exp\left(-\mu\delta^2/4\right)$

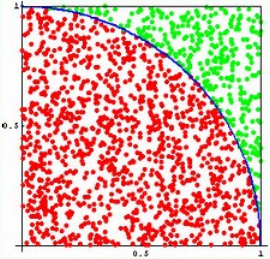
Example

Let X be the number of heads in n tosses of a fair coin, then $\mu = \frac{n}{2}$ and $\delta = \frac{1}{2}$, we have

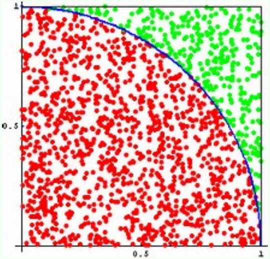
$$P(X > \frac{3n}{4}) = P(X > (1+\frac{1}{2})\frac{n}{2}) < \exp\left(-\frac{n}{2}\delta^2/4\right) = \exp\left(-n/32\right)$$

If we toss the coin 1000 times, the probability is less than $\exp\left(-125/4\right)$.

Why is this algorithm accurate?



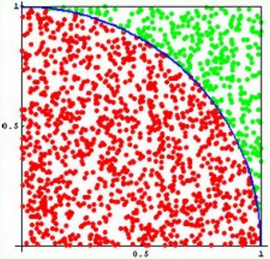
Why is this algorithm accurate?



For this case, sample space $\Omega = \{(x, y) | 0 \leq x, y \leq 1\}$, and $C = \{(x, y) | x^2 + y^2 \leq 1 \wedge x, y \geq 0\}$.
Let E be an event that the point locates in the circle area C. Then we have

$$P(E) = \frac{S(C)}{S(\Omega)} = \frac{\pi}{4}.$$

Why is this algorithm accurate?



For this case, sample space $\Omega = \{(x, y) | 0 \leq x, y \leq 1\}$, and $C = \{(x, y) | x^2 + y^2 \leq 1 \wedge x, y \geq 0\}$. Let E be an event that the point locates in the circle area C. Then we have

$$P(E) = \frac{S(C)}{S(\Omega)} = \frac{\pi}{4}.$$

Let X_i be a r.v., where $X_i = 1$ means a generated point p_i inside in the circle, otherwise 0, i.e., $X_i = I_C(P_i)$. Hence,

$$E(X_i) = \frac{\pi}{4}, E(\sum_{i=1}^{n} X_i) = \frac{n\pi}{4}, \text{ and } V(\sum_{i=1}^{n} X_i) = \frac{n\pi(4-\pi)}{16}.$$

Why is this algorithm accurate? Cont'd

Chebyshev bound

Hence, we have

$$Y = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{n} I_C(P_i)}{n} = \frac{\sum_{i=1}^{n} I_C(P_i)}{\sum_{i=1}^{n} I_C(P_i) + \sum_{i=1}^{n} I_{\Omega - C}(P_i)}.$$

In terms of the Chebyshev bound, we have

Why is this algorithm accurate? Cont'd

Chebyshev bound

Hence, we have

$$Y = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{n} I_C(P_i)}{n} = \frac{\sum_{i=1}^{n} I_C(P_i)}{\sum_{i=1}^{n} I_C(P_i) + \sum_{i=1}^{n} I_{\Omega - C}(P_i)}.$$

In terms of the Chebyshev bound, we have

$$\begin{aligned} & P(Y - \frac{\pi}{4} > \frac{\pi}{4}) < P(|X - \frac{n\pi}{4}| > \frac{n\pi}{4}) \\ & \leq \frac{V(X)}{(\frac{n\pi}{4})^2} \\ & = \frac{n\pi(4-\pi)}{16} \frac{16}{n^2\pi^2} \approx \frac{1}{4n} \end{aligned}$$

Why is this algorithm accurate? Cont'd

Chebyshev bound

Hence, we have

$$Y = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{n} I_C(P_i)}{n} = \frac{\sum_{i=1}^{n} I_C(P_i)}{\sum_{i=1}^{n} I_C(P_i) + \sum_{i=1}^{n} I_{\Omega-C}(P_i)}.$$

In terms of the Chebyshev bound, we have

$$\begin{aligned} P(Y - \frac{\pi}{4} > \frac{\pi}{4}) &< P(|X - \frac{n\pi}{4}| > \frac{n\pi}{4}) \\ &\leq \frac{V(X)}{(\frac{n\pi}{4})^2} \\ &= \frac{n\pi(4-\pi)}{16}\frac{16}{n^2\pi^2} \approx \frac{1}{4n} \end{aligned}$$

When $n = 100$, the probability of large deviation is less than 0.0025.

Why is this algorithm accurate? Cont'd

Chernoff bound

In terms of the Chernoff bound, we have

Why is this algorithm accurate? Cont'd

Chernoff bound

In terms of the Chernoff bound, we have

$$\begin{aligned} P(Y - \frac{\pi}{4} > \frac{\pi}{4}) &= P(X - \frac{n\pi}{4} > \frac{n\pi}{4}) \\ &= P(X > (1+1)\frac{n\pi}{4}) \\ &< \exp\left(-\frac{n\pi}{4} 1^2/4\right) \\ &= \exp\left(-n\pi/16\right). \end{aligned}$$

Why is this algorithm accurate? Cont'd

Chernoff bound

In terms of the Chernoff bound, we have

$$\begin{aligned} P(Y - \frac{\pi}{4} > \frac{\pi}{4}) &= P(X - \frac{n\pi}{4} > \frac{n\pi}{4}) \\ &= P(X > (1+1)\frac{n\pi}{4}) \\ &< \exp\left(-\frac{n\pi}{4}1^2/4\right) \\ &= \exp\left(-n\pi/16\right). \end{aligned}$$

When $n = 100$, the probability of large deviation is less than $\exp\left(-75/4\right)$.

Why is this algorithm accurate? Cont'd

Chernoff bound

In terms of the Chernoff bound, we have

$$\begin{aligned} P(Y - \frac{\pi}{4} > \frac{\pi}{4}) &= P(X - \frac{n\pi}{4} > \frac{n\pi}{4}) \\ &= P(X > (1+1)\frac{n\pi}{4}) \\ &< \exp\left(-\frac{n\pi}{4} 1^2/4\right) \\ &= \exp\left(-n\pi/16\right). \end{aligned}$$

When $n = 100$, the probability of large deviation is less than $\exp\left(-75/4\right)$.

Please explain which inequalities give better tail bounds? **Why?**

Take-aways

Conclusions

- Random variable
- Bernoulli Trials and the Binomial Distribution
- Average-Case Computational Complexity
- Variance
- Tail Probability