COMP170
Discrete Mathematical Tools for Computer Science

Random Variables

Version 2.0: Last updated, May 13, 2007

Discrete Math for Computer Science
K. Bogart, C. Stein and R.L. Drysdale
Section 5.4, pp. 249-262

Random Variables

- What Are Random Variables?
- Binomial Probabilities
- Expected Values
- Expected Values of Sums and Numerical Multiples
- Indicator Random Variables
- The Number of Trials until a First Success

What Are Random Variables?

A **random variable** for an experiment with sample space S is a *function* that *assigns a number* to each element of S.

Example

Flipping a coin n times.

Sample space: set of all sequences of n $\mathtt{H}$'s and $\mathtt{T}$'s.

Random variable "number of heads" takes a sequence and tells us how many heads are in that sequence.

Example:

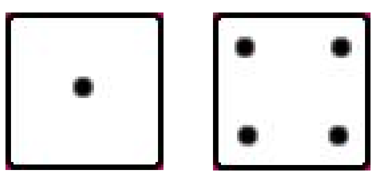
$X(\mathtt{HTHHT}) = 3.$

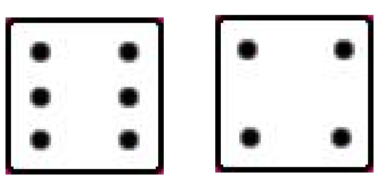
$X(\mathtt{THTHT}) = 2.$

Example 2:

Rolling two dice

Random variable is
"sum of the values showing on top of dice".

$X\Big($  $\Big) = 5$

$X\Big($  $\Big) = 10$

Random Variables

- What Are Random Variables?
- Binomial Probabilities
- Expected Values
- Expected Values of Sums and Numerical Multiples
- Indicator Random Variables
- The Number of Trials until a First Success

Bernoulli Random Variables

A test in which the outcome is either a success or failure.

Examples:	Sucesss
Flipping a coin	A head
Answer to an exam question	A correct answer
A Drug trial	A successful treatment

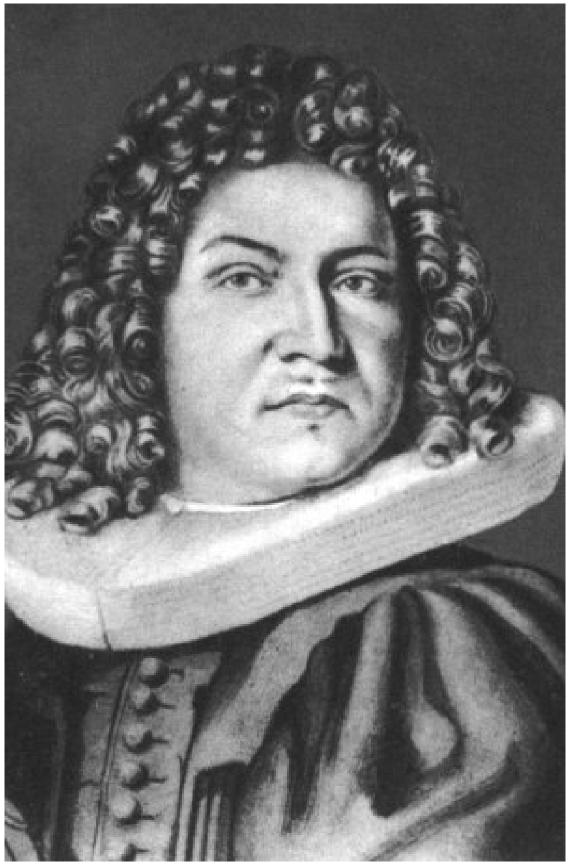
If such a test has

$$P(\text{Success}) = p \quad \text{and} \quad P(\text{Failure}) = q = 1 - p$$

It is called a Bernoulli trial or Benoulli Random Variable with success probability p

Jakob Bernoulli

b. 1654, d. 1705



Swiss Mathematician and Scientist. Famous for his work on probability theory (where *Bernoulli trials* come from) and calculus.

He often collaborated with his brother Johann Bernoulli, another famous mathematician

For more information, please see
http://en.wikipedia.org/wiki/James_Bernoulli

We are given an *Independent trials process* with two outcomes at each stage: *success* and *failure*.

Examples:	Quantity of Interest
Flipping a coin	# of heads.
Student performance on a test	# of correct answers
Drug trials	# of successful treatments

We analyze:

probability of exactly k successes in n independent trials
with probability p of success on each trial.

Such an independent trials process is called a
Bernoulli trials process

Note that this is the sum of Bernoulli Random Variables

Suppose we have 5 Bernoulli trials,
with probability p success on each trial.
What is the probability of

(a) Success on first 3 trials and failure on last 2?
(b) Failure on the first 2 trials and success on last 3?
(c) Success on Trials 1, 3, and 5, and failure on Trials 2 and 4?
(d) Success on any particular 3 trials and failure on other 2?

By Independence, probability of a sequence of outcomes is product of probabilities of individual outcomes.

So, probability of any sequence of 3 successes and 2 failures is $p^3(1-p)^2$.

More generally, in n Bernoulli trials,
probability of a given sequence of k successes and $n-k$ failures is

$$p^k(1-p)^{n-k}.$$

Probability of a given sequence of k successes and $n - k$ failures

in n Bernoulli trials is $$p^k(1-p)^{n-k}.$$

However, this is **not** the probability of having k successes, because many different sequences could lead to k successes.

How many sequences of n Bernoulli trials have exactly k successes (and $n - k$ failures)?

This is number of ways to choose the k places where success occurs out of n total places which is

$$\binom{n}{k}$$

We have just seen that the

Probability of occurrence of a given sequence of k successes and $n-k$ failures is $p^k(1-p)^{n-k}$

Number of such sequences is $\binom{n}{k}$

Theorem 5.8
The probability of having exactly k successes in a sequence of n independent trials with two outcomes and probability p of success on each trial is given by

$$P(\text{exactly } k \text{ successes}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The Binomial Random Variable X (with parameters n, p) takes on integer values with probability distribution:

$$P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } 0 \le k \le n \\ 0 & \text{Otherwise} \end{cases}$$

Those probabilities are sometimes called **binomial probabilities**, or the **binomial probability distribution**.

Reality Check: This *is* a probability distribution since

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \left(p + [1-p]\right)^n = 1^n = 1$$

Example:

A student takes a ten-question objective test.

Suppose that a student who knows 80% of the course material has probability $.8$ of success on any question, independent of how (s)he did on any other problem.

What is the probability that (s)he earns a grade of 80 or better (out of 100)?

Grade of 80 or better on a ten-question test corresponds to eight, nine, or ten successes in ten trials. So,

$P(80 \text{ or better}) =$

$$\binom{10}{8}(.8)^8(.2)^2 + \binom{10}{9}(.8)^9(.2)^1 + \binom{10}{10}(.8)^{10}(.2)^0 \approx .678.$$

Random Variables

- What Are Random Variables?
- Binomial Probabilities
- Expected Values
- Expected Values of Sums and Numerical Multiples
- Indicator Random Variables
- The Number of Trials until a First Success

Expected Values

Example:

Flipping a fair coin twice, we "expect" to see one head.

Intuition

Four outcomes – one with no heads, two with one head, and one with two heads – giving an average of $$\frac{0+1+1+2}{4} = 1.$$

Expected (average) values might not be possible outcomes

Three flips of a coin: the eight possibilities for the number of heads are $0, 1, 1, 1, 2, 2, 2, 3$, giving an average

$$\frac{0+1+1+1+2+2+2+3}{8} = 1.5.$$

Consider the following game:
You pay me some money, and then you flip 3 coins.
I will pay you \$1.00 for every head that comes up.

Would you play this game if you had to pay me \$2.00? \$1.00?
For this game to be fair, how much do you think it should cost?

Because you expect to get 1.5 heads,
you expect to make \$1.50.

Therefore, it is reasonable to play this game
as long as the cost is at most \$1.50.

We formalize our intuition by defining:

The **expected value**, or **expectation**, of a random variable X with possible values $\{x_1, x_2, \ldots, x_k\}$ is

$$E(X) = \sum_{i=1}^{k} x_i P(X = x_i).$$

Example:

Suppose a biased coin has probability $\frac{2}{3}$ of coming up Tails. The expected number of tails when flipping the coin 3 times is

$$\sum_{i=0}^{3} i \binom{3}{i} \left(\frac{2}{3}\right)^{i} \left(\frac{1}{3}\right)^{3-i}$$

$$= 0 \cdot 1 \left(\frac{2}{3}\right)^{0}\left(\frac{1}{3}\right)^{3} + 1 \cdot 3 \left(\frac{2}{3}\right)^{1}\left(\frac{1}{3}\right)^{2} + 2 \cdot 3 \left(\frac{2}{3}\right)^{2}\left(\frac{1}{3}\right)^{1} + 3 \cdot 1 \left(\frac{2}{3}\right)^{3}\left(\frac{1}{3}\right)^{0} = 2$$

Another Example

(a) Throwing a fair die: Let X be the number of spots shown. Since each outcome is equally likely

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

(b) Throwing two fair dice. Let Y be number of spots shown. Probabilities are

i	2	3	4	5	6	7	8	9	10	11	12
$Pr(Y=i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$E(Y) = \sum_{i=2}^{12} i Pr(Y=i) = 7$$

Returning to the biased coin tossing

outcomes 's'	TTT,	TTH,	THT,	HTT,	THH,	HTH,	HHT,	HHH
$X(s)$	3	2	2	2	1	1	1	0
$P(s)$	$\frac{8}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{1}{27}$

Notice that if, instead of using the formula $\sum_{i=1}^{k} x_i P(X = x_i)$ on the previous page, we instead summed up $X(s)$ over all outcomes s, weighted by $P(s)$, we get the same answer!

$$\boxed{3 \cdot \frac{8}{27}} + \boxed{2 \cdot \frac{4}{27} + 2 \cdot \frac{4}{27} + 2 \cdot \frac{4}{27}} + \boxed{1 \cdot \frac{2}{27} + 1 \cdot \frac{2}{27} + 1 \cdot \frac{2}{27}} + \boxed{0 \cdot \frac{1}{27}}$$

$$= \boxed{3 \cdot 1 \cdot \frac{8}{27}} + \boxed{2 \cdot 3 \cdot \frac{4}{27}} + \boxed{1 \cdot 3 \cdot \frac{2}{27}} + \boxed{0 \cdot 1 \cdot \frac{1}{27}} = 2$$

What we just saw was a special case of

Lemma 5.9
If a random variable X is defined on a (finite) sample space S, then its expected value is given by

$$E(X) = \sum_{s:s\in S} X(s)P(s).$$

Proof:

Assume that values of the random variable are $x_1, x_2, \ldots, x_k$.

Let F_i stand for "$X = x_i$", so $P(F_i) = P(X = x_i)$.

Take items in sample space, group them together into events F_i, and rework sum into definition of expectation:

$$\sum_{s:s\in S} X(s)P(s) = \sum_{i=1}^{k} \sum_{s:s\in F_i} X(s)P(s)$$

$$= \sum_{i=1}^{k} \sum_{s:s\in F_i} x_i P(s) = \sum_{i=1}^{k} x_i \sum_{s:s\in F_i} P(s)$$

$$= \sum_{i=1}^{k} x_i P(F_i) \qquad = \sum_{i=1}^{k} x_i P(X = x_i) \qquad = E(X).$$

Informal proof:

When we compute the sum in Lemma 5.9, we can group together all elements of the sample space that have X-value x_i and add their probabilities.

This gives us $x_i P(X = x_i)$, which leads us to the definition of the expected value of X.

Random Variables

- What Are Random Variables?
- Binomial Probabilities
- Expected Values
- Expected Values of Sums and Numerical Multiples
- Indicator Random Variables
- The Number of Trials until a First Success

Expected Values of Sums and Numerical Multiples

Example:

Throw two fair dice

X_1 : outcome of first die throw. We know $E(X_1) = \frac{7}{2}$

X_2 : outcome of second die throw. We know $E(X_2) = \frac{7}{2}$

The expected outcome of throwing two dice "should" be the expected outcome of throwing the first plus the expected outcome of throwing the second, i.e.,

$$E(X_1 + X_2) = E(X_1) + E(X_2) = \frac{7}{2} + \frac{7}{2} = 7.$$

We already saw that 7 is the correct answer.

We now see that this formula will *always* be true.

Theorem 5.10
Suppose X and Y are random variables on the (finite) sample space S. Then

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

Proof:
From Lemma 5.9, we may write

$$E(X+Y) = \sum_{s:s\in S} (X(s)+Y(s))P(s)$$
$$= \sum_{s:s\in S} X(s)P(s) + \sum_{s:s\in S} Y(s)P(s)$$
$$= E(X) + E(Y).$$

Another Example

Flip a fair coin and observe whether it comes up H or T.
Define the two random variables X, Y by

$$X = \begin{cases} 1 & \text{if H} \\ 0 & \text{if T} \end{cases} \qquad\qquad Y = \begin{cases} 1 & \text{if T} \\ 0 & \text{if H} \end{cases}$$

Then $E(X) = \frac{1}{2}$ and $E(Y) = \frac{1}{2}$ so $E(X) + E(Y) = 1$

On the other hand, regardless of the value of the coin toss, $X + Y = 1$, so $E(X + Y) = 1$ and the theorem works.

Note, though, that $X \cdot Y = 0$, so

$$E(X) \cdot E(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \neq 0 = E(X \cdot Y).$$

$E(X+Y) = E(X)+E(Y)$ is *always true.* $E(X \cdot Y) = E(X) \cdot E(Y)$ *is* sometimes true and sometimes false *(more later).*

Returning to our "exam" question; If we double the credit we give for each question on the final exam, we would expect students' scores to double.

Let cX denote the random variable we get from X by multiplying all its values by the number c.

Theorem 5.11
Suppose X is a random variable on a sample space S. Then for any number c, we have $E(cX) = cE(X)$.

Theorems 5.10 and 5.11 are typically called **linearity of expectation**.

They can tremendously simplify calculations of expected values

Example

On one flip of a coin, expected number of $\mathtt{H}$ is $.5$.

For n flips, let X_i be number of $\mathtt{H}$ seen on flip i, so that X_i is either 0 or 1. ex : 5 flips: $X_2(\mathtt{HTHHT}) = 0, X_3(\mathtt{HTHHT}) = 1$.

Then X, total number of $\mathtt{H}$ in n flips, is given by
$X = X_1 + X_2 + \ldots + X_n$. (*)

We already saw that X has a binomial distribution so

$$E(X) = \sum_{i=0}^{n} iP(X = i) = \sum_{i=0}^{n} i \binom{n}{i} (0.5)^i (0.5)^{n-i}$$

$$= .5n$$

complicated
want an easier method!

Example An easier method

On one flip of a coin, expected number of H is $.5$.

For n flips, let X_i be number of H seen on flip i, so that X_i is either 0 or 1. ex : 5 flips: $X_2(\mathtt{HTHHT}) = 0, X_3(\mathtt{HTHHT}) = 1.$

Then X, total number of H in n flips, is given by
$X = X_1 + X_2 + \ldots + X_n.$ (*)

Expected value of each X_i is $.5$.

Take expectation of both sides of (*) and apply Theorem 5.10 repeatedly:

$$\begin{aligned} E(X) &= E(X_1 + X_2 + \ldots + X_n) \\ &= E(X_1) + E(X_2) + \ldots + E(X_n) \\ &= .5 + .5 + \ldots + .5 \\ &= .5n \text{ is expected number of H in } n \text{ flips.} \end{aligned}$$

Example (2)

What is expected number X of correct answers a student will get on an n-question test if he knows 90% of course material and questions on the test are an accurate and uniform sampling of the course material. (Assume student does not guess.)

$P(\text{student gets correct answer on given question}) = .9$.

This is again a binomial probability distribution so

$$E(X) = \sum_{i=0}^{10} iP(X = i) = \sum_{i=0}^{10} i \binom{10}{i} (0.9)^i (0.1)^{n-i}$$

We could evaluate this but, there is an easier way.

Example (2)

What is expected number X of correct answers a student will get on an n-question fiest if he knows 90% of course material and questions on the test are an accurate and uniform sampling of the course material. (Assume student does not guess.)

$P($student gets correct answer on given question$) = .9$.

X_i : number of correct answers on Question i (either 1 or 0).

$E(X_i) = .9$ (why?)

Then $X = X_1 + X_2 + \cdots X_n$ so, by linearity of expectation,

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} .9 = .9n$$

Theorem 5.12
In a Bernoulli trials process with n trials in which each experiment has two outcomes and probability p of success, the expected number of successes is np.

Proof:

X_i : number of successes in ith trial of n independent trials.

Expected number of successes on ith trial is, by definition,

$$E(X_i) = p \cdot 1 + (1 - p) \cdot 0 = p.$$

Number of successes X in all n trials is $X_1 + X_2 + \cdots + X_n$

By Theorem 5.10, expected number of successes in n trials is

$$E(X) = E(X_1) + E(X_2) + \cdots E(X_n) = np$$

Random Variables

- What Are Random Variables?
- Binomial Probabilities
- Expected Values
- Expected Values of Sums and Numerical Multiples
- Indicator Random Variables
- The Number of Trials until a First Success

Indicator Random Variables

A random variable that is 1 if a certain event happens and 0 otherwise is called an **indicator random variable**.

$$X_i = \begin{cases} 1 & \text{if event } i \text{ occurs} \\ 0 & \text{if event } i \text{ does not occur} \end{cases}$$

Property:

$$E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0)$$

$$= P(X_i = 1) = P(\text{event occurs})$$

Sums of indicator random variables
count number of times an event happens.

Because of linearity of expectation,
there is no need for events to be independent.

Example

Recall the problem of the ten-question exam in which the student has probability $.9$ of getting each question correct. We used the random variables

$$X_i = \begin{cases} 1 & \text{if question } i \text{ answered correctly} \\ 0 & \text{if question } i \text{ answered incorrectly} \end{cases}.$$

The fact that $X = X_1 + X_2 + \cdots X_9 + X_{10}$ and linearity of expectation, let us easily calculate

$$E(X) = E(X_1) + E(X_2) + \ldots + E(X_{10}) = 10 \cdot (.9) = 9.$$

These X_i are indicator random variables!

Example; Return to the *Derangement Problem*

Let X be the total number of students who get their own backpacks back after they're all mixed up.

X_i : indicator variable for event E_i
that person i gets correct backpack returned
($X_i = 1$ if person i gets correct backpack; otherwise, $X_i = 0$).

$$X = X_1 + X_2 + \ldots + X_n,$$

so, by linearity of expectation

$$E(X) = E(X_1) + E(X_2) + \ldots + E(X_n),$$

Note that events E_i are not independent.
e.g., when $n = 2$: either both students or neither student get own backpacks returned so $X_1 = X_2$.

What is $E(X_i)$ for a given i?

$$E(X_i) = P(X_i = 1) = P(\text{event occurs}),$$

$$= P(\text{person } i \text{ gets correct backpack})$$

There are $n!$ total permutations of n people.
There are $(n-1)!$ permutations
in which person i's backpack is returned.

$$\Rightarrow \qquad E(X_i) = \frac{(n-1)!}{n!} = 1/n$$

We just showed that $E(X_i) = \frac{1}{n}$.

Recall that X is the total number of students who get their own backpacks back after they're all mixed up, and, by linearity of expectation,

$$E(X) = E(X_1) + E(X_2) + \ldots + E(X_n).$$

$$\Rightarrow \qquad E(X) = n \cdot \frac{1}{n} = 1$$

This means that
$E($number of students who get their own backpack back $) = 1$

Note that this is independent of n.

Random Variables

- What Are Random Variables?
- Binomial Probabilities
- Expected Values
- Expected Values of Sums and Numerical Multiples
- Indicator Random Variables
- The Number of Trials until a First Success

The Number of Trials until a First Success

How many times should we expect to have to flip a coin until we first see a head? Why?
How many times should we expect to have to roll two dice until we see a sum of 7? Why?

Intuitively:

We should have to flip a coin twice to see a head.

However, we could conceivably flip a coin forever without seeing a head, so should we really expect to see a head in two flips?

Probability of getting 7 on two dice is $1/6$.

Does that mean we should expect to have to roll the dice six times before we see 7?

Analysis

Not finite sample spaces.

Instead, consider process of repeating independent trials with probability p of success until success occurs and then stopping.

Possible outcomes are the infinite set $\{S, FS, FFS, \ldots, F^iS, \ldots\}$, where F^iS stands for sequence of i failures followed by a success.

The natural probability weight we would assign to F^iS would be $(1-p)^i p$.
Does this make sense?

$$P(S) = p, \quad P(FS) = (1-p)p, \; \ldots, \; P(F^iS) = (1-p)^i p, \ldots.$$

Their sum is

$$\sum_{i=0}^{\infty}(1-p)^i p = p\sum_{i=0}^{\infty}(1-p)^i = p\frac{1}{1-(1-p)} = \frac{p}{p} = 1.$$

so this is a good probability distribution

> Probability distribution $P(F^iS) = (1-p)^i p$ is called a **geometric distribution** because of the geometric series we used in proving that probabilities sum to 1.

Theorem 5.13
Suppose we have a sequence of trials in which each trial has two outcomes, success and failure, and in which the probability of success at each step is some $p > 0$. Then the expected number of trials until the first success is $1/p$.

Proof:

Consider random variable X, which is i if first success is on Trial i. That is, $X(F^{i-1}S) = i$.

Probability that first success is on Trial i is $(1-p)^{i-1}p$, because for this to happen, there must be $i-1$ failures followed by 1 success.

Expected number of trials is expected value of X, which is, by definition of expected value and previous two sentences,

$$
\begin{aligned}
E(\text{number of trials}) &= \sum_{i=1}^{\infty} p(1-p)^{i-1} i \\
&= p \sum_{i=1}^{\infty} (1-p)^{i-1} i \\
&= \frac{p}{1-p} \sum_{i=1}^{\infty} (1-p)^{i} i \\
&= \frac{p}{1-p} \frac{1-p}{p^2} \quad = \frac{1}{p}.
\end{aligned}
$$

Example

For a fair coin, $P(\text{getting a head}) = \frac{1}{2}$.
Applying Theorem 5.13, we see that expected number of times we need to flip a fair coin until we see a head is

$$\frac{1}{\frac{1}{2}} = 2.$$

Example

When throwing two fair dice, the probability of seeing a 7 is $\frac{1}{6}$. So, applying Theorem 5.13, we see that expected number of times we need to roll two dice until we get a 7 is

$$\frac{1}{\frac{1}{6}} = 6$$

When throwing two fair dice, the probability of seeing a 6 is $\frac{5}{36}$. So, applying Theorem 5.13, we see that expected number of times we need to roll two dice until we get a 6 is

$$\frac{1}{\frac{5}{36}} = \frac{36}{5} = 7.2$$