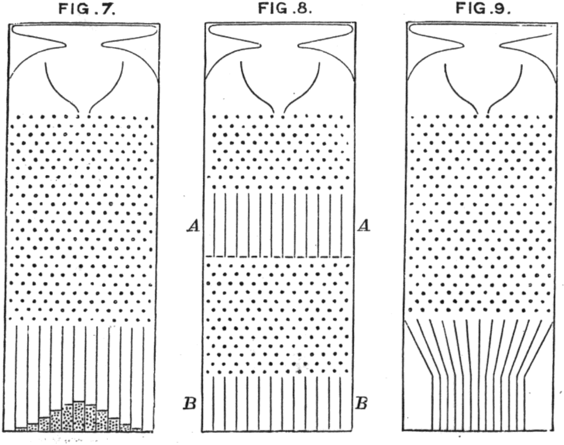
Discrete Random Variables; Expectation

18.05 Spring 2018



http://www.mathsisfun.com/data/quincunx.html

http://www.youtube.com/watch?v=9xUBhhM4vbM

Reading Review

Random variable X assigns a number to each outcome:

$$X : \Omega \to \mathbf{R}$$

"$X = a$" denotes the event $\{\omega \,|\, X(\omega) = a\}$.

Probability mass function (pmf) of X is given by

$$p(a) = P(X = a).$$

Cumulative distribution function (cdf) of X is given by

$$F(a) = P(X \leq a).$$

Example from class

Suppose X is a random variable with the following table.

values of X:	-2	-1	0	4
pmf $p(a)$:	$1/4$	$1/4$	$1/4$	$1/4$
cdf $F(a)$:	$1/4$	$2/4$	$3/4$	$4/4$

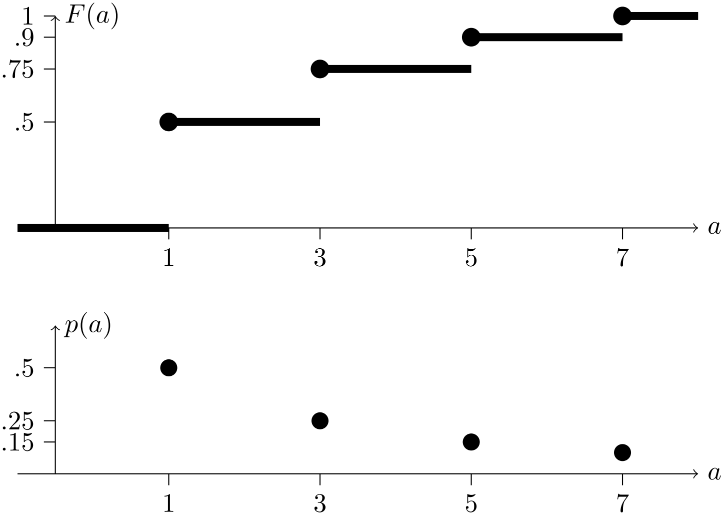
The cdf is the probability 'accumulated' from the left.

Examples. $F(-1) = 2/4$, $F(0) = 3/4$, $F(0.5) = 3/4$, $F(-5) = 0$, $F(5) = 1$.

Properties of $F(a)$:

1. Nondecreasing
2. Way to the left, i.e. as $a \to -\infty$), F is 0
3. Way to the right, i.e. as $a \to \infty$, F is 1.

CDF and PMF



Concept Question: cdf and pmf

X a random variable.

values of X:	1	3	5	7
cdf $F(a)$:	0.5	0.75	0.9	1

1. What is $P(X \leq 3)$?

(a) 0.15 (b) 0.25 (c) 0.5 (d) 0.75

2. What is $P(X = 3)$?

(a) 0.15 (b) 0.25 (c) 0.5 (d) 0.75

1. **answer:** (d) 0.75. $P(X \leq 3) = F(3) = 0.75$.
2. **answer:** (b) $P(X = 3) = F(3) - F(1) = 0.75 - 0.5 = 0.25$.

Expected Value

X is a random variable takes values $x_1, x_2, \ldots, x_n$:

The **expected value** of X is defined by

$$E(X) = p(x_1)x_1 + p(x_2)x_2 + \ldots + p(x_n)x_n = \sum_{i=1}^{n} p(x_i)\, x_i$$

- It is a weighted average.
- It is a measure of central tendency. (Statistics jargon.)

Properties of $E(X)$

- $E(X+Y) = E(X) + E(Y)$ (linearity I)
- $E(aX+b) = aE(X) + b$ (linearity II)
- $E(h(X)) = \sum_i h(x_i)\, p(x_i)$

Three approaches to Expected Value

Roll a six-sided die twice; values X_1 and X_2, $X = X_1 + X_2$. Want to compute $E(X)$.

- sum over outcomes of experiment:

$$\begin{aligned} E(X) &= p(1,1)\cdot(1+1) + p(1,2)\cdot(1+2) + \cdots + p(6,6)\cdot(6+6) \\ &= (1/36)\cdot 2 + (1/36)\cdot 3 + \cdots + (1/36)\cdot 12 \end{aligned}$$

- sum over values of X:

$$\begin{aligned} E(X) &= \mathrm{pmf}(2)\cdot 2 + \mathrm{pmf}(3)\cdot 3 + \cdots + \mathrm{pmf}(12)\cdot 12 \\ &= (1/36)\cdot 2 + (2/36)\cdot 3 + \cdots + (1/36)\cdot 12 \end{aligned}$$

- divide and conquer:

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 7/2 + 7/2 = 7.$$

Meaning of expected value

What is the *expected* average of one roll of a die?

answer: Suppose we roll it 5 times and get (3, 1, 6, 1, 2). To find the average we add up these numbers and divide by 5: ave $= 2.6$. With so few rolls we don't expect this to be representative of what would usually happen. So let's think about what we'd *expect* from a large number of rolls. To be specific, let's (pretend to) roll the die 600 times.

We expect that each number will come up roughly $1/6$ of the time. Let's suppose this is exactly what happens and compute the average.

value:	1	2	3	4	5	6
expected counts:	100	100	100	100	100	100

The average of these 600 values (100 ones, 100 twos, etc.) is then

$$\begin{aligned}\text{average} &= \frac{100\cdot 1 + 100\cdot 2 + 100\cdot 3 + 100\cdot 4 + 100\cdot 5 + 100\cdot 6}{600} \\ &= \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = \boxed{3.5.}\end{aligned}$$

This is the 'expected average'. We will call it the *expected value*

Examples

Example 1. Find $E(X)$

1.	X:	3	4	5	6
2.	pmf:	1/4	1/2	1/8	1/8

3. $E(X) = 3/4 + 4/2 + 5/8 + 6/8 = 33/8$

Example 2. Suppose $X \sim$ Bernoulli(p). Find $E(X)$.

1.	X:	0	1
2.	pmf:	$1-p$	p

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

Example 3. Suppose $X \sim$ Binomial$(12, .25)$. Find $E(X)$.
$X = X_1 + X_2 + \ldots + X_{12}$, where $X_i \sim$ Bernoulli(.25). Therefore

$$E(X) = E(X_1) + E(X_2) + \ldots E(X_{12}) = 12 \cdot (.25) = 3$$

In general if $X \sim$ Binomial(n, p) then $E(X) = np$.

Class example

We looked at the random variable X with the following table top 2 lines.

1.	X:	-2	-1	0	1	2
2.	pmf:	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$
3.	$E(X) =$ -2/5 - 1/5 $+ 0/5 + 1/5 + 2/5 = 0$					
4.	X^2:	4	1	0	1	4
5.	$E(X^2) = 4/5 + 1/5 + 0/5 + 1/5 + 4/5 = 2$					

Line 3 computes $E(X)$ by multiplying the probabilities in line 2 by the values in line 1 and summing.

Line 4 gives the values of X^2.

Line 5 computes $E(X^2)$ by multiplying the probabilities in line 2 by the values in line 4 and summing. This illustrates the use of the formula

$$E(h(X)) = \sum_i h(x_i)\, p(x_i).$$

Continued on the next slide.

Class example continued

Notice that in the table on the previous slide, some values for X^2 are repeated. For example the value 4 appears twice. Summing all the probabilities where $X^2 = 4$ gives $P(X^2 = 4) = 2/5$. Here's the full table for X^2

1.	X^2:		4		1		0		
2.	pmf:		$2/5$		$2/5$		$1/5$		
3.	$E(X^2)$	$=$	$8/5$	$+$	$2/5$	$+$	$0/5$	$=$	2

Here we used the definition of expected value to compute $E(X^2)$. Of course, we got the same expected value $E(X^2) = 2$ as we did earlier.

Board Question: Interpreting Expectation

(a) Would you accept a gamble that offers a 10% chance to win \$95 and a 90% chance of losing \$5?

(b) Would you pay \$5 to participate in a lottery that offers a 10% percent chance to win \$100 and a 90% chance to win nothing?

- Find the expected value of change in assets in each case.

Discussion on next slide.

Discussion

Framing bias / cost versus loss. The two situations are identical, with an expected value of gaining \$5. In a study, 132 undergrads were given these questions (in different orders) separated by a short filler problem. 55 gave different preferences to the two events. Of these, 42 rejected (a) but accepted (b). One interpretation is that we are far more willing to pay a cost up front than risk a loss. (See *Judgment under uncertainty: heuristics and biases* by Tversky and Kahneman.)

Loss aversion and cost versus loss sustain the insurance industry: people pay more in premiums than they get back in claims on average (otherwise the industry wouldn't be sustainable), but they buy insurance anyway to protect themselves against substantial losses. Think of it as paying \$1 each year to protect yourself against a 1 in 1000 chance of losing \$100 that year. By buying insurance, the expected value of the change in your assets in one year (ignoring other income and spending) goes from negative 10 cents to negative 1 dollar. But whereas without insurance you might lose \$100, with insurance you always lose exactly \$1.

Board Question

Suppose (hypothetically!) that everyone at your table got up, ran around the room, and sat back down randomly (i.e., all seating arrangements are equally likely).

What is the expected value of the number of people sitting in their original seat?

(We explored this with simulations in last Friday's Studio.)

Neat fact: A permutation in which nobody returns to their original seat is called a derangement. The number of derangements turns out to be the nearest integer to $n!/e$. Since there are $n!$ total permutations, we have:

$$P(\text{everyone in a different seat}) \approx \frac{n!/e}{n!} = 1/e \approx 0.3679.$$

It's surprising that the probability is about 37% regardless of n, and that it converges to $1/e$ as n goes to infinity.

`http://en.wikipedia.org/wiki/Derangement`

Solution

Number the people from 1 to n. Let X_i be the Bernoulli random variable with value 1 if person i returns to their original seat and value 0 otherwise. Since person i is equally likely to sit back down in any of the n seats, the probability that person i returns to their original seat is $1/n$. Therefore $X_i \sim Bernoulli(1/n)$ and $E(X_i) = 1/n$. Let X be the number of people sitting in their original seat following the rearrangement. Then

$$X = X_1 + X_2 + \cdots + X_n.$$

By linearity of expected values, we have

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

• It's neat that the expected value is 1 for any n.

• If $n = 2$, then both people either retain their seats or exchange seats. So $P(X = 0) = 1/2$ and $P(X = 2) = 1/2$. In this case, X never equals $E(X)$.

• The X_i are not independent (e.g. for $n = 2$, $X_1 = 1$ implies $X_2 = 1$).

• Expectation behaves linearly even when the variables are dependent.

Deluge of discrete distributions

$$\text{Bernoulli}(p) = \begin{array}{l} 1 \text{ (success) with probability } p, \\ 0 \text{ (failure) with probability } 1-p. \end{array}$$

In more neutral language:

$$\text{Bernoulli}(p) = \begin{array}{l} 1 \text{ (heads) with probability } p, \\ 0 \text{ (tails) with probability } 1-p. \end{array}$$

Binomial(n,p) = # of successes in n independent Bernoulli(p) trials. $\text{pmf}(k) = p^k(1-p)^{n-k}\binom{n}{k}$.

Geometric(p) = # of tails before first heads in a sequence of indep. Bernoulli(p) trials; $\text{pmf}(k) = p(1-p)^k$.

(Neutral language avoids confusing whether we want the number of successes before the first failure or vice versa.)

Concept Question

1. Let $X \sim \text{binom}(n, p)$ and $Y \sim \text{binom}(m, p)$ be independent. Then $X + Y$ follows:

(a) $\text{binom}(n + m, p)$ (b) $\text{binom}(nm, p)$
(c) $\text{binom}(n + m, 2p)$ (d) other

2. Let $X \sim \text{binom}(n, p)$ and $Z \sim \text{binom}(n, q)$ be independent. Then $X + Z$ follows:

(a) $\text{binom}(n, p + q)$ (b) $\text{binom}(n, pq)$
(c) $\text{binom}(2n, p + q)$ (d) other

1. **answer:** (a). Each binomial random variable is a sum of independent Bernoulli(p random variables, so their sum is also a sum of Bernoulli(p) r.v.'s.

2. **answer:** (d) This is different from problem 1 because we are combining Bernoulli(p) r.v.'s with Bernoulli(q) r.v.'s. This is not one of the named random variables we know about.

Board Question: Find the pmf

$X = \#$ of successes before the *second* failure of a sequence of independent Bernoulli(p) trials.

Describe the pmf of X.

Hint: this requires some counting.

Answer is on the next slide.

Solution

X takes values 0, 1, 2, ... The pmf is $p(n) = (n+1)p^n(1-p)^2$.

For concreteness, we'll derive this formula for $n = 3$. Let's list the outcomes with three successes before the second failure. Each must have the form

$$\_\_ \ \_\_ \ \_\_ \ \_\_ \ F$$

with three S and one F in the first four slots. So we just have to choose which of these four slots contains the F:

$$\{FSSSF, SFSSF, SSFSF, SSSFF\}$$

In other words, there are $\binom{4}{1} = 4 = 3 + 1$ such outcomes. Each of these outcomes has three S and two F, so probability $p^3(1-p)^2$. Therefore

$$p(3) = P(X = 3) = (3+1)p^3(1-p)^2.$$

The same reasoning works for general n.

Dice simulation: geometric(1/4)

Roll the 4-sided die repeatedly until you roll a 1.
Record $X = \#$ of rolls *BEFORE* the 1.

Example: If you roll (3, 4, 2, 3, 1) then record 4.
Example: If you roll (1) then record 0.

Fiction

Gambler's fallacy: [roulette] if black comes up several times in a row then the next spin is more likely to be red.

Hot hand: NBA players get 'hot'.

Fact

P(red) remains the same.

The roulette wheel has no memory. (Monte Carlo, 1913).

The data show that player who has made 5 shots in a row is no more likely than usual to make the next shot.

(Currently, there seems to be some disagreement about this.)

Gambler's fallacy

"On August 18, 1913, at the casino in Monte Carlo, black came up a record twenty-six times in succession [in roulette]. [There] was a near-panicky rush to bet on red, beginning about the time black had come up a phenomenal fifteen times. In application of the maturity [of the chances] doctrine, players doubled and tripled their stakes, this doctrine leading them to believe after black came up the twentieth time that there was not a chance in a million of another repeat. In the end the unusual run enriched the Casino by some millions of francs."

Hot hand fallacy

An NBA player who made his last few shots is more likely than his usual shooting percentage to make the next one?

https://doi.org/10.1016/0010-0285(85)90010-6
(There is still some controversy about this. Some statisticians feel that the authors of the above paper erred in their analysis of the data and the data do support the notion of a hot hand in basketball.)

Amnesia

Show that *Geometric*(p) is memoryless, i.e.

$$P(X = n + k \mid X \geq n) = P(X = k)$$

Explain why we call this memoryless.

Explanation given on next slide.

Proof that geometric(p) is memoryless

One method is to look at the tree for this distribution. Here we'll just use the formula that defines conditional probability. To do this we need to find probabilities for the events used in the formula.

Let A be '$X = n + k$' and let B be '$X \geq n$'.

We have the following:

- $A \cap B = A$. This is because $X = n + k$ guarantees $X \geq n$. Thus, $P(A \cap B) = P(A) = p^(n + k)(1 - p)$
- $P(B) = p^n$. This is because B consists of all sequences that start with n successes.

We can now compute the conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{p^{n+k}(1-p)}{p^n} = p^k(1-p) = P(X = k).$$

This is what we wanted to show!