Probability II: Expectations and Variances

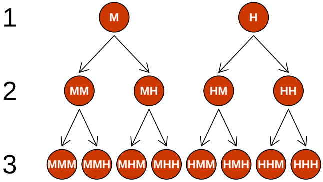
Elie Gurarie

Biol 799 - Lecture 5

January 2, 2017

Some sample spaces are not “numerical”

Example: Sequence of three free throws



- S = {MMM, MMH, MHM, MHH, HMM, HMH, HHM, HHH}

There’s no natural way to order this state space.

Some sample spaces are numerical

Example: Sum of points after three free throws.

- Question II: How many baskets will the basketball player make total?

try 0:	0
try 1:	0 1
try 2:	0 1 2
try 3:	0 1 2 3

- $S = \{0, 1, 2, 3\}$

- This is a naturally "numerical" sample space.
- Every outcome can be assigned a value

Definition

A Random Variable ...

... is a variable whose value is a numerical outcome of a random phenomenon.

Or (more technically) a **random variable** X is a function that takes each element of a sample space S and assigns it to a real number.

Discrete random variable

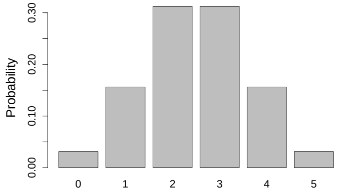
- Every value that a random variable can take is associated with a probability
- The probabilities sum to 1

Value of X	Probability
x_1	p_1
x_2	p_2
x_3	p_3
...	...
x_k	p_k

Example: Number of heads after 5 coin flips

- X is the total number of heads after 5 coin flips
- Possible values of X are: $\{0,1,2,3,4,5\}$
- Probability distribution of X is:
  $P(X = k) = \text{Binomial}(k|n = 5, p = 1/2)$

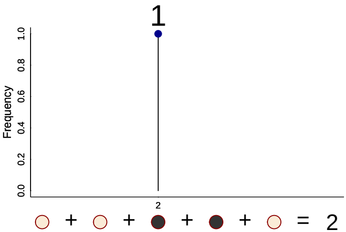
x	$P(X = x)$
0	0.03125
1	0.15625
2	0.31250
3	0.31250
4	0.15625
5	0.03125



Question: How many heads do we *expect* to get?

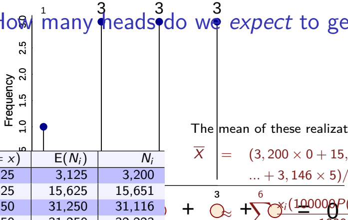
- One way to think of this problem is that if we repeated the experiment many many times - what would the average score be?

n = 1



n = 10

Question: How many heads do we *expect* to get?

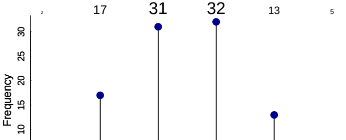


x	$P(X=x)$	E(N_i)	N_i
0	0.03125	3,125	3,200
1	0.15625	15,625	15,651
2	0.31250	31,250	31,116
3	0.31250	31,250	32,223
4	0.15625	15,625	15,664
5	0.03125	3,125	3,146
Sum	**1**	100,000	100,000

The mean of these realizations is:

$$\begin{aligned}\overline{X} &= (3,200 \times 0 + 15,651 \times 1 + \ldots \\ &\quad \ldots + 3,146 \times 5)/100000 = 2.499 \\ &\approx \sum_{i=1}^{6} \frac{x_i(100000 P(X=x_i))}{100000} \\ &= \sum_{i=1}^{6} x_i P(X=x_i)\end{aligned}$$

n = 100



Definition

Expectation

The **expected value** or **expectation** of a discrete random variable X with probability function $f(x)$ is

$$\mathrm{E}(X) = \sum_{i=1}^{n} x_i\, f(x_i)$$

where $\{x_1, x_2, ...x_n\}$ is the set of all values that X can take.
It is essentially the mean of *possible values* weighted by their *probability*.
In statistics, $E(X)$ it is often denoted μ or μ_X.

Question: How many heads do we *expect* to get?

$$\mathrm{E}(X) = \sum_{i=1}^{n} x_i\, p(x_i)$$

x	$f(x)$	$xf(x)$
0	0.03125	0
1	0.15625	0.15625
2	0.31250	0.6250
3	0.31250	0.9375
4	0.15625	0.6250
5	0.03125	0.15625
Sum	1	2.5

So, the **expected value** of X is **2.5**.

Expectation of the binomial distribution

Binomial distribution:

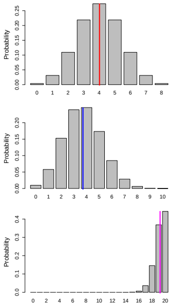
$$f(x|n,p) = \Pr(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Solving for the expectation of the binomial distribution:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{i=0}^{n} f(x|n,p) \\
&= \sum_{i=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{i=0}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{i=0}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{i=0}^{n} \text{Binomial}(x-1|n-1,p) = np
\end{aligned}$$

Example of the binomial expectation

- How many heads to we expect after 8 tosses?
  - $p = 0.5,\ n = 8,\ E(X) = np = 4$
- How many baskets do we expect Shaq to make in 10 attempts?
  - $p = 0.37,\ n = 10,\ E(X) = 3.7$
- How many baskets do we expect Ray Allen to make in 20 attempts?
  - $p = 0.96,\ n = 20,\ E(X) = 19.2$



The **Expectation**

is often denoted μ and called the **mean** of a distribution.

- The expectation tells you the **true mean** of any known, theoretical, distribution
- It is not quite the same as the **sample mean** which we obtain empirically for data.

Some basic arithmetical property of expectations

$$
\begin{aligned}
\mathrm{E}(A+B) &= \mathrm{E}(A)+\mathrm{E}(B) \\
\mathrm{E}(kA) &= k\mathrm{E}(A); \text{ where k is a constant} \\
\mathrm{E}(AB) &= \mathrm{E}(A)\mathrm{E}(B); \text{ only if } A \text{ and } B \text{ are independent}
\end{aligned}
$$

Variance

- Another very important quantity is the **variance** of a distribution.
- It is the **Expected Squared Deviation from the Mean**
  - Convert that to math notation:

$$\mathrm{Var}(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right)$$

- Use the properties of expectation to simplify:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2 - 2X\mathrm{E}(X) + \mathrm{E}(X)^2) \\
&= \mathrm{E}(X^2) - 2\mathrm{E}(X\mathrm{E}(X)) + \mathrm{E}(X)^2 \\
&= \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + \mathrm{E}(X)^2 \\
&= \mathrm{E}(X^2) - \mathrm{E}(X)^2
\end{aligned}$$

Example: Variance of 4 coin flips

- X is the total number of heads after 4 coin flips
- Possible values of X are: $\mathrm{x} = \{0,1,2,3,4\}$
- Probability distribution of X is:
  $P(X = x) = \mathrm{Binomial}(k|n = 4, p = 1/2)$
- Expected value of X is: $\mathrm{E}(X) = \mu = np = 2$

x	$P(X = x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2 P(x)$
0	1/16	-2	4	1/4
1	1/4	-1	1	1/4
2	3/8	0	0	0
3	1/4	1	1	1/4
4	1/16	2	4	1/4
Σ	1			**1**

$$\mathrm{Var}(X) = 1$$

The **Variance**

- The **variance** of a random variable X is defined by the following expressions:

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}((X - \mathrm{E}(X))^2) \\ \mathrm{Var}(X) &= \mathrm{E}(X^2) - \mathrm{E}(X)^2 \end{aligned}$$

- For **discrete random variables**, $X \in \{x_1, x_2, x_3 ... x_n\}$, with known probability function $P(X = x) = f(x)$:

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{i=1}^{n} \left((x_i - \sum_{i=1}^{n} x_i f(x_i))^2 f(x_i) \right) \\ &= \sum_{i=1}^{n} x_i^2 f(x_i) - \left(\sum_{i=1}^{n} x_i f(x_i) \right)^2 \end{aligned}$$

Variance of the binomial distribution

Binomial distribution:

$$f(x|n, p) = \mathrm{Pr}(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Solving for the variance of the binomial distribution:

$$\begin{aligned} \mathrm{Var}(X^2) &= \sum_{i=0}^{n} x^2 f(x|n, p) - \mathrm{E}(X)^2 \\ &= ... \\ &= \text{lots of algebra similar to last time} \\ &= ... \\ &= np(1-p) \end{aligned}$$

The Variance

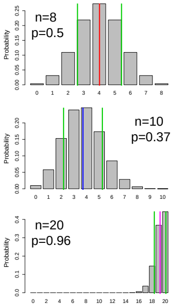
...is often denoted σ^2.

- It tells you something quantitative about the amount of spread in a distribution from the **mean**.
- The square root of the variance, σ is the standard deviation - which is the units of the random variable X.

- This quantity is the **true variance** of any known, theoretical, distribution
- It is not exactly the same as the **sample variance** which we obtain empirically for data.

Example of the binomial variance

- What's the variance of heads after 8 coin tosses?
  - $p = 0.5$, $n = 8$, $\mathrm{E}(X) = 4$, $\mathrm{Var}(X) = np(1-p) = 2$
- What's the variance of Shaq's 10 FT attempts?
  - $p = 0.37$, $n = 10$, $\mathrm{E}(X) = 3.7$, $\mathrm{Var}(X) = 2.331$
- What's the variance of Ray Allen's 20 FT attempts?
  - $p = 0.96$, $n = 20$, $\mathrm{E}(X) = 19.2$, $\mathrm{Var}(X) = 0.768$



Basic arithmetical property of variances

- $\text{Var}(kA) = k^2\text{Var}(A)$
- If A and B are independent,

$$\text{Var}(A + B) = \text{Var}(A) + \text{Var}(B)$$

Components of a probability distribution

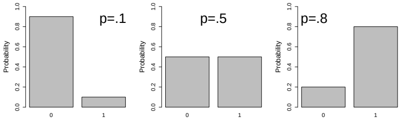
- X: The **random variable** (r.v.)
- x: The possible values (or **support**) of X
  - $X \in \{x_1, x_2, x_3 ... x_n\}$
- $P(X = x|\theta) = f(x, \theta)$: The **probability mass functions**
  - Often contracted to "p.m.f." of "pmf"
  - for continuous r.v.'s, called "**probability density function**" (p.d.f.)
- θ: The **parameters** of the pdf.
- $\text{E}(X) = \mu$: The theoretical **mean** of the pdf.
- $\text{Var}(X) = \sigma^2$: The theoretical **variance** of the pdf.

$X \sim \text{Binomial}(n, p)$

Name:	Binomial Distribution	Total number of successes with probability p after n tries.
Support:	x	0,1,2,3,...,n
Parameters:	$n \in \mathbf{N}$ (positive integers) $p \in [0, 1]$	number of trials (positive integer) probability of success
pmf	$P(X = x \mid n, p)$	$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$
mean	$\text{E}(X)$	np
variance	$\text{Var}(X)$	$np(1-p)$

Bernoulli distribution

- The Bernoulli distribution is the distribution of a **single** event with probability p.
- $P(X = x) = \begin{cases} 1-p & \text{for } x = 0 \\ p & \text{for } x = 1 \end{cases}$
- Examples: a single coin flip, a single FT.



Question: What is the expected value and variance of the Bernoulli distribution?

$X \sim$ Bernoulli(p)

Name:	Bernoulli Distribution	Models number of successes after one trial
Support:	x	0,1
Parameters:	$p \in [0, 1]$	probability of success
pmf	$P(X = x \mid p)$	$\begin{cases} 1-p & \text{for } x = 0 \\ p & \text{for } x = 1 \end{cases}$
mean	$\mathrm{E}(X)$	p
variance	$\mathrm{Var}(X)$	$p(1-p)$

Relationship between Binomial and Bernoulli distribution

- A Binomial(n, p) is the sum of n Bernoulli(p) trials
- Formally, if X_1, X_2, X_3 ... $X_n$are each independent variables with:

$$X_i \sim \text{Bernoulli}(p)$$

then

$$Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$$

Using arithemetic properties of Expectation and Variance

Question: What is the expectation and variance of $Y = \sum_{i=1}^{n} X_i$ where X_i are Bernoulli(p) trials?

- Recall

$$E(A + B) = E(A) + E(B)$$

And (if A and B are independent)

$$\text{Var}(A + B) = \text{Var}(A) + \text{Var}(B)$$

- Then:

$$E(Y) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

- And (very similarly):

$$\text{Var}(Y) = \text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

- These are much easier ways to calculate the mean and variance of the Binomial distribution!

Example Problem: The Lottery

Rules of Washington State Powerball

- Buy a ticket for \$1
- Pick 6 numbers from 1, ..., 53
- Win \$1,000,000 if your numbers are identical to winning numbers (in any order).

- Q1: What is the probability of winning the jackpot?
  - Uniform sample space: Choose 6 from 53
  - $\binom{53}{6} = 22,957,480$
  - $P(Winning) = 1/22957480$

Example Problem: The Lottery

$P(Winning) = 1/22957480$

- Q2: What are your expected winnings?
- Call X the dollar amount of your winnings
  - $X \in \{-1, 999999\}$
  - $P(X = x) \begin{cases} 22957479/22957480 & \text{for } x = 0 \\ 1/22957480 & \text{for } x = 10000000 \end{cases}$
  - $\text{E}(X) = -1 \times \frac{22957479}{22957480} + 999,999 \times \frac{1}{22957480} \approx -0.956$
- So: You will lose 0.96 cents, on average, every time you play.
- **Good luck!**

R code: Generic Powerball

```
> p <- 1/choose(53, 6)
> cost <- 1
> reward <- 1e6
> X <- c(0,reward)-cost
> pmf <- c(1-p,p)
> sum(pmf * X)
[1] -0.9564412
```

Geometric distribution

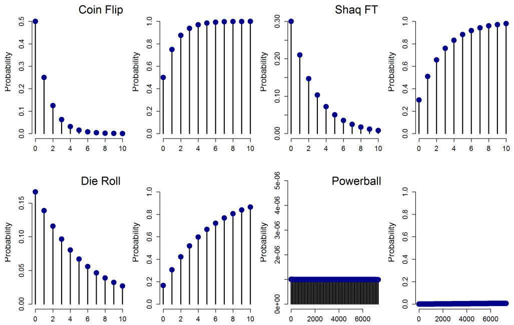
Questions:

- How long (how many rolls) will it take me to roll a 6 on average?
- How long will it take Shaq to make a free throw?
- If I bought a lottery ticket every day, how long would it take me to win?

The Geometric Distribution

- The geometric distribution described the number of Bernoulli trials with probability p before a success - the *waiting time* of a distribution.
- $P(X = k) = p(1-p)^k$ where $k = \{0, 1, 2, 3, ...\}$
- Note: one p is for success, k $(1-p)$'s for failure.

Geometric distribution: examples



Geometric distribution: examples

- What is the probability that Shaq (p=0.3) will miss exactly three times before making it?
  $p(1-p)^3 =$ `dgeom(3, 0.3)`$=$ 10.3%
- What is the probability that Shaq (p=0.3) will miss less than three times in a row?
  $\sum_{i=0}^{2} p(1-p)^i =$ `pgeom(2, 0.3)`$=$ 65%
- What is the probability that Shaq (p=0.3) will miss more than three times in a row?
  $\sum_{i=4}^{\infty} p(1-p)^i =$ `1 - pgeom(4, 0.3)`$=$ 17%
- What is the probability that you will have won SuperLotto at least once after 20 years of playing daily?
  $\sum_{i=0}^{356*20} p(1-p)^i =$ `pgeom(365*20, p = 1/22e6)`$=$ 0.03%

Geometric distribution: Memorylessness

- After 3 misses, what is the probability Shaq will miss 3 more times?
  ALSO 10.3%!
- After 20 years of trying, what is the probability you might win after another 20 years?
  ALSO 0.03%!
- It does not matter how long you have been trying to get a success, the waiting time will always have the same distribution.
- This is called "memorylessness" and is very special.

$$P(X > m + n | X > m) = P(X > n)$$

Geometric distribution

$$\begin{aligned} \mathrm{E}(X) &= \sum_{i=0}^{\infty} k\, p\, (1-p)^k = \frac{1-p}{p} \\ \mathrm{Var}(X) &= \sum_{i=0}^{\infty} k^2\, p\, (1-p)^k - \mathrm{E}(X)^2 = \frac{1-p}{p^2} \end{aligned}$$

- How long (how many flips) will it take me before I get a head from a fair coin on average?
  Answer: $\mu_x = (1 - 1/2)/1/2 = 1$ flip, $\sigma_x = \sqrt{2}$
- How long will it take me to roll a 6 on average?
  Answer: $\mu_x = (1 - 1/6)/1/6 = 5$ rolls, $sigma_x = \sqrt{30}$
- How long will it take Shaq to make a free throw?
  Answer: $\mu_x = (1 - 0.3)/0.3 = 2.333$ attempts, $\sigma_x = 2.789$
- How long would it take me to win Powerball?
  Answer: $\mu_x = (1 - 1/23e6)/23e6 = 22e6$ days $= 63{,}013$ years,
  $\sigma_x = 22e6$ days

$X \sim$ Geometric(p)

Name:	Geometric Distribution	Waiting time of success for Bernoulli trials
Support:	x	0,1,2,3 ...
Parameters:	$p \in [0, 1]$	probability of success
pmf	$P(X = x\|p)$	$p(1-p)^x$
mean	$\mathrm{E}(X)$	$\frac{1-p}{p}$
variance	$\mathrm{Var}(X)$	$\frac{1-p}{p^2}$
Special Feature:	**Memorylessness!**	