Phys 362 2020

Lecture 14

Thurs: Read 4.1, 4.2

Mean values, variances, standard deviations

Many statistical situations have numerical variables associated with the possible events and we might want to quantify the typical outcome of these events.

As an example, consider the four cell Galton board game with the illustrated monetary values assigned to each cell. We would want to know a reasonable amount to charge a player per ball. To do this we need to estimate the typical winnings per ball. Suppose that the game is run N times. Let n_A be the number of times A is attained, etc,... Then the earnings per trial are

A	B	C	D
\$4	\$1	\$1	\$4

$$\frac{1}{N}\left(\$4\, n_A + \$1\, n_B + \$1\, n_C + \$4\, n_D\right) = \$4\,\frac{n_A}{N} + \$1\,\frac{n_B}{N} + \$1\,\frac{n_C}{N} + \$4\,\frac{n_D}{N}$$

$$\rightarrow \$4\, p(A) + \$1\, p(B) + \$1\, p(C) + \$4\, p(D)$$

as $N \to \infty$

Then with $p_A = 1/8$, $p_B = 3/8$, $p_C = 3/8$, $p_D = 1/8$ we get:

$$\text{typical earnings per run} = \$4 \times \tfrac{1}{8} + \$1 \times \tfrac{3}{8} + \$1 \times 3/8 + \$4 \times 1/8 = \$1.75$$

This is the notion behind the mean of a variable which depends on random events. The data is generally tabulated as illustrated Then

event, i	prob $p(i)$	value x_i
1	$p(1)$	x_1
2	$p(2)$	x_2
3	$p(3)$	x_3
⋮		

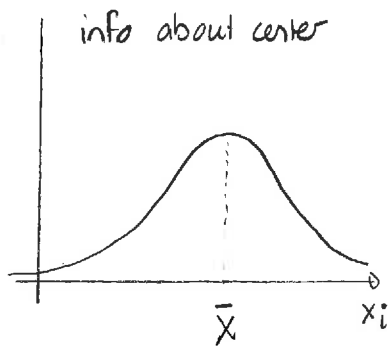
variable

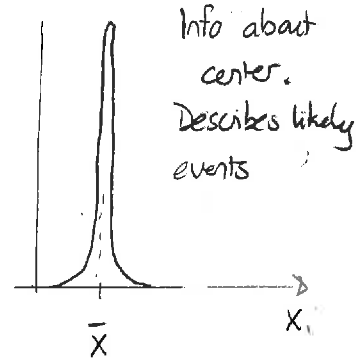
> The mean of x is
>
> $$\bar{x} = \sum_{\text{all } i} x_i \, p(i)$$

This can be extended to any function of the variable, x. Thus if $f = f(x)$ then

> $$\overline{f(x)} = \sum_{\text{all events } i} f(x_i)\, p(x_i)$$

The mean only gives very limited information about the distribution. In some cases it can describe the most likely event or the location of the center of the distribution. We can have two or more distinct distributions with a differing spread. This can be quantified via:





1) variance
2) standard deviation

These are defined as:

> The variance of x is
>
> $$\overline{(\Delta x)^2} = \sum_{\text{all } i} (x_i - \bar{x})^2 \; P(i)$$

Then one can show via algebra that

$$\overline{(\Delta x)^2} = \overline{x^2} - (\bar{x})^2$$

where

$$\overline{x^2} = \sum_{\text{all } i} x_i^2 \, p(i)$$

The <u>standard deviation</u> is defined as:

$$\boxed{\sigma := \sqrt{\overline{(\Delta x)^2}}}$$

and

$$\boxed{\sigma = \sqrt{\overline{x^2} - (\bar{x})^2}}$$

1 Dice rolls: mean and variance

For an unbiased die the probability with which any single one of the six possible outcomes occurs is 1/6. Let x be the face value of the outcome of a die roll.

a) Determine $\overline{x}$.

b) Determine $\overline{x^2}$ and show that $\overline{x^2} \neq \overline{x}^2$.

c) Determine the standard deviation, σ.

Answer: a)

$$\bar{x} = \sum x_i\, p(i)$$

$$= \frac{1+2+3+4+5+6}{6}$$

$$= \frac{21}{6} = \frac{7}{2} = 3.5$$

Outcome	value	prob
1	1	1/6
2	2	1/6
3	3	1/6
4	4	1/6
5	5	1/6
6	6	1/6

b)

$$\overline{x^2} = 1^2\, \tfrac{1}{6} + 2^2\, \tfrac{1}{6} + 3^2\, \tfrac{1}{6} + 4^2\, \tfrac{1}{6} + \cdots + 6^2\, \tfrac{1}{6}$$

$$= \frac{1}{6}\left[1+4+9+16+25+36\right] = \frac{91}{6} = 15.2$$

Then $(\bar{x})^2 = \frac{49}{4} = 12.25 \neq \overline{x^2}$

c)

$$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{15.2 - 12.25} = 1.71$$

Binary random variables

In many cases a single trial results in one of two possible outcomes. For example:

1) coin toss - single coin events $\rightsquigarrow$ heads (H), tails (T)

2) random walk in one dimension $\rightsquigarrow$ in each time step move a fixed distance left or right

   - events: left (L), right (R)

3) measure component of a single spin-½ particle $\rightsquigarrow$ outcomes are either $S_z = +\hbar/2$, $S_z = -\hbar/2$

   - events: spin up ($\uparrow$), spin down ($\downarrow$)

In particular this last example will illustrate many features of statistical physics.

In such cases we often do multiple trials and want to know about the cumulative outcome. For example, in the random walk one might take N steps and ask about the probability of arriving at various distances from the starting point. We assume that the events are independent which means that the outcome of any given single individual trial does not depend on the previous trials. Consider the random walk as an example:

a) walker is originally at the origin $(x=0)$

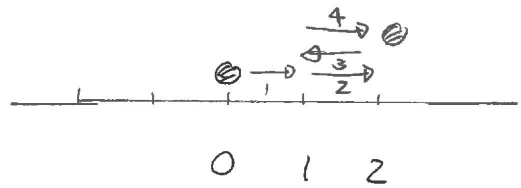
b) each step is 1m

c) for any step the probabilities are:

step	prob
R	p
L	q

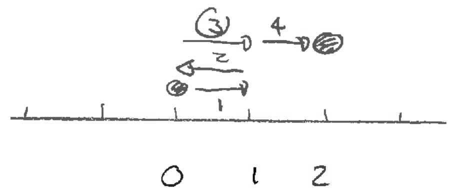
Possible questions to consider are:

1) <u>basic compound events</u>



e.g with 4 steps:

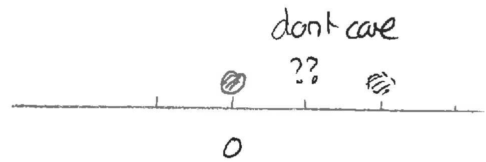
- what is probability of RRLR?
- " " " " RLRR?



2) <u>aggregate/compound events</u>

e.g. with 4 steps:

- what is probability that $x = 2$?



This clear entails multiple ways of reaching the destination and we would have to sum over these

3) <u>average quantities</u>

e.g with N steps:

- what is mean location?

These questions can all be answered using

a) probabilities for the basic events: p, q

b) assigning "joint probabilities" to basic compound events in terms of p, q.

c) using counting to generate probabilities for aggregate compound events

2 Random walk

Consider a one dimensional random walk with 4 steps. Let p be the probability of stepping right in one step and q the probability of stepping left.

a) Provide a relationship between p and q.

b) List all possible sequences of 4 steps and the probability with which each sequence occurs.

c) Let ~~n~~ k be the total number of steps to the right. List: possible values for k, and for each, the associated location at the end of the process x and the probability with which it occurs.

d) Show that the probability that the person takes k steps right, regardless of the order is

$$p(k) = \binom{4}{k} p^k q^{4-k}$$

where

$$\binom{N}{k} := \frac{N!}{k!(N-k)!}.$$

e) Describe how one can determine $\overline{x}$ from $\overline{k}$. Determine $\overline{k}$. *Hint: Note that the binomial expansion gives*

$$(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k}$$

and differentiating with respect to p gives

$$Np(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} k p^k q^{N-k}.$$

f) Determine $\overline{x}$ and verify that the results are correct for $p = 0, 1$, and $1/2$.

Answer: a) $p+q=1$

b) There will be $2^4 = 16$ combinations

event / combination	prob	event	prob
RRRR	p^4	LRRR	qp^3
RRRL	p^3q	LRRL	q^2p^2
RRLR	p^3q	LRLR	q^2p^2
RRLL	p^2q^2	LRLL	q^3p
RLRR	p^3q	LLRR	q^2p^2
RLRL	p^2q^2	LLRL	q^3p
RLLR	p^2q^2	LLLR	q^3p
RLLL	pq^3	LLLL	q^4

c)

k	x	prob
0	-4	q^4
1	-2	$4q^3p$
2	0	$6q^2p^2$
3	2	$4qp^3$
4	4	p^4

d) $p(0) = \binom{4}{0} \cancel{p^0} q^4$ $\quad \binom{4}{0} = \frac{4!}{0!\,4!} = 1$

$= q^4$ ✓

$p(1) = \binom{4}{1} p^1 q^3$ $\quad \binom{4}{1} = \frac{4!}{1!\,3!} = 4$

$= 4pq^3$ ✓

$p(2) = \binom{4}{2} p^2 q^2$ $\quad \binom{4}{2} = \frac{4!}{2!\,2!} = \frac{4.3}{2} = 6$

$= 6p^2q^2$ ✓

$p(3) = \binom{4}{3} p^3 q$ $\quad \binom{4}{3} = \frac{4!}{3!\,1!} = 4$

$= 4p^3q$ ✓

$p(4) = \binom{4}{4} p^4 q^0$ $\quad \binom{4}{4} = \frac{4!}{0!\,4!} = 1$

$= p^4$ ✓

e) $x = n_{\text{steps right}} - n_{\text{steps left}}$ $\quad \underbrace{n_{\text{steps right}}}_{k} + n_{\text{steps left}} = N$

$= k - (N-k)$

$= 2k - N$

So $\bar{x} = 2\bar{k} - N$

We need

$$\bar{k} = \sum k\, p(k)$$

$$= \sum_{k=0}^{4} k \binom{4}{k} p^{4} q^{4-k}$$

Now

$$\sum_{k=0}^{N} k \binom{N}{k} p^{N} q^{N-k} = \underset{4}{N}\, p\, \underbrace{(p+q)}_{1}^{N}$$

$$\Rightarrow \quad \bar{k} = 4p\, 1^{N} \quad \Rightarrow \quad \bar{k} = 4p.$$

Thus $\bar{x} = 8p - 4$

f) For $p = 0$ always left $\Rightarrow$ $\bar{x} = -4$ & always occurs.

$p = 1$ always right $\Rightarrow$ $\bar{x} = 4$

$p = 1/2$ equally likely left vs right $\Rightarrow$ $\bar{x} = 0$

Processes involving multiple trials of a process with a binary outcome are called Bernoulli processes. They follow the general framework.

Single trial

Outcome	Probability
$s = +1$	$p = p(+1)$
$s = -1$	$q = p(-1)$

with $p + q = 1$

Multiple trials

Construct basic compound events

For N trials list sequence of outcomes:

$$s_1, s_2, s_3, \ldots, s_N$$

± 1 ± 1

Construct probability for this

$$p(s_1, s_2, \ldots, s_N) = p(s_1)p(s_2) \ldots p(s_N)$$

if events are independent.

Aggregate / compound events

Let k = # occurences of $s = +1$

$N - k$ = # " " $s = -1$

List aggregate events + probs

outcome, k	prob $p(k)$
$k=0$	$p(k=0)$
$k=1$	$p(k=1)$
$k=2$	$p(2)$
$k=3$	$p(3)$
$\vdots$	$\vdots$
$k=N$	$p(N)$

If events are independent

$$p(k) = \binom{N}{k} p^k q^{N-k}$$

The following general result applies:

Consider a Bernoulli process with N trials each yielding $+1$ or -1. Then the probability of attaining k outcomes of $+1$ and $N-k$ of -1 is

$$p(k) = \binom{N}{k} p^k q^{N-k}$$

Binomial (Bernoulli distribution

where $$\binom{N}{k} = \frac{N!}{k!\,(N-k)!}$$

and p = prob single trial gives $+1$

q = " " " " -1

Proof: List the ways in which one can attain k outcomes of $+1$ and $N-k$ of -1.:

$$\underbrace{+++\cdots+}_{k}\ \underbrace{---\cdots-}_{N-k}$$

$$\underbrace{+++\cdots+}_{k-1}-+\underbrace{-\cdots-}_{N-k-1}$$

$$\vdots$$

$$\underbrace{--\cdots-}_{N-k}\ \underbrace{+\cdots+}_{k}$$

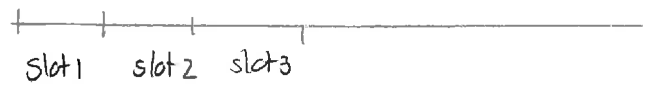
The probability with which each occurs is

$$p^k q^{N-k}$$

However, there are multiple arrangements of k "+" and $N-k$ "−" signs. Let $W(k)$ be the number of ways. Then

$$p(k) = W(k)\, p^k q^{N-k}.$$

We need to compute $W(k)$. This is the number of ways in which + can be placed in k out of N "slots." We can label the signs by order of insertion. Let



$+_1$ = first plus sign $\rightsquigarrow$ N choices

$+_2$ = second " " $\rightsquigarrow$ exclude slot for $+_1$ $\rightsquigarrow$ $N-1$ choices.

$\vdots$

$+_k$ = "k" plus sign $\rightsquigarrow$ $N-(k-1)$ choices.

So the number of these arrangements is

$$N(N-1)\ldots(N-(k-1)) = \frac{N(N-1)\ldots 2.1}{(N-k)(N-k-1)\ldots 1}$$

$$= \frac{N!}{(N-k)!}$$

But this overcounts because, for example

$+_1 +_2 +_3 \ldots +_k - \cdots -$

$+_2 +_1 +_3 \cdots +_k - \cdots -$

are counted as distinct. How many ways can one arrange k objects? There are k choices for $+_1$, followed by $(k-1)$ for $+_2$, ... Thus the total number of choices is $k(k-1)\ldots 2.1 = k!$ Thus

$$W(k) = \frac{N!}{(N-k)!\,k!} = \binom{N}{k}$$

3 Binomial coefficients

Show that

$$\binom{N}{n} = \binom{N}{N-n}.$$

Ans:

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

$$\binom{N}{N-n} = \frac{N!}{(N-n)!\,(N-(N-n))!}$$

$$= \frac{N!}{(N-n)!\,n!}$$

4 Bernoulli process probabilities

For a Bernoulli process with $N = 4$ evaluate $p(n)$ for all n if:

a) $p = \frac{1}{2}$, and

b) $p = \frac{1}{4}$.

Answer

$$p(n) = \binom{N}{n} p^n q^{N-n}$$

a) $$p(n) = \binom{N}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^{N-n} = \binom{4}{n}\left(\frac{1}{2}\right)^4$$

so

n	$p(n)$
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16

b) $$p(n) = \binom{4}{n}\left(\frac{1}{4}\right)^n\left(\frac{3}{4}\right)^{4-n} = \frac{1}{4^4}\binom{4}{n} = \frac{1}{256}\binom{4}{n} 3^{4-n}$$

n	$p(n)$
0	81/256
1	108/256
2	54/256
3	12/256
4	1/256

Means and standard deviations for Bernoulli processes

The binomial theorem gives:

$$(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k}.$$

and this enables us to compute $\bar{k}, \overline{k^2}, \overline{k^3}, \ldots$ We can then compute any function of k such as $x = 2k - N$, since $\bar{x} = 2\bar{k} - N$.

We can show:

$$\boxed{\begin{array}{l} \text{For a Bernoulli distribution:} \\ \bar{k} = pN \\ \overline{k^2} = \bar{k}^2 + pqN \end{array}}$$

<u>Proof.</u>

$$\bar{k} = \sum_{k=0}^{N} k \binom{N}{k} p^k q^{N-k}$$

$$= \sum_{k=0}^{N} \binom{N}{k} p \frac{\partial}{\partial p}(p^k) q^{N-k}$$

$$= p \frac{\partial}{\partial p} \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k}$$

$$= p \frac{\partial}{\partial p} (p+q)^N$$

$$= pN \underbrace{(p+q)}_{=1}{}^{N-1} = pN$$

$$\Rightarrow \bar{k} = pN$$

For $\overline{k^2}$

$$\frac{\partial}{\partial p}(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} k p^{k-1} q^{N-k}$$

$$\frac{\partial^2}{\partial p^2}(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} k(k-1) p^{k-2} q^{N-k}$$

$$= \frac{1}{p^2} \sum_{k=0}^{N} \binom{N}{k} (k^2-k) p^k q^{N-k}$$

$$= \frac{1}{p^2} \left[\overline{k^2} - (\overline{k}) \right]$$

But $\frac{\partial}{\partial p}(p+q)^N = N(p+q)^{N-1}$

$$\frac{\partial^2}{\partial p^2}(p+q)^N = N(N-1)(p+q)^{N-2}$$

and $p+q = 1 \Rightarrow$

$$\frac{\partial^2}{\partial p^2}(p+q)^N = N(N-1)$$

Thus

$$N(N-1) = \frac{1}{p^2} \left[\overline{k^2} - \overline{k} \right]$$

$$\Rightarrow p^2(N^2 - N) = \overline{k^2} - pN$$

$$\Rightarrow \overline{k^2} = p^2 N^2 + N p\underbrace{(1-p)}_{q}$$

$$\Rightarrow \overline{k^2} = (\overline{k})^2 + Npq$$