Physical Principles in Biology
Biology 3550
Spring 2024

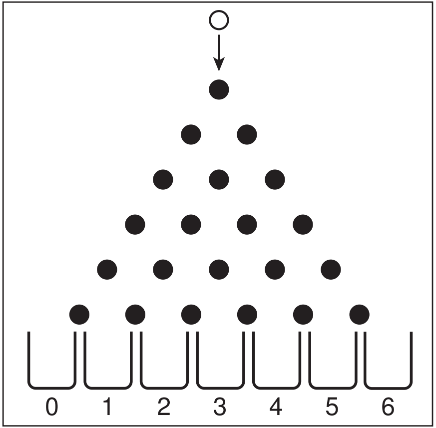
Lecture 9:

Random Variables, the Expected Value, Variance and Standard Deviation

Monday, 29 January 2024

©David P. Goldenberg
University of Utah
goldenberg@biology.utah.edu

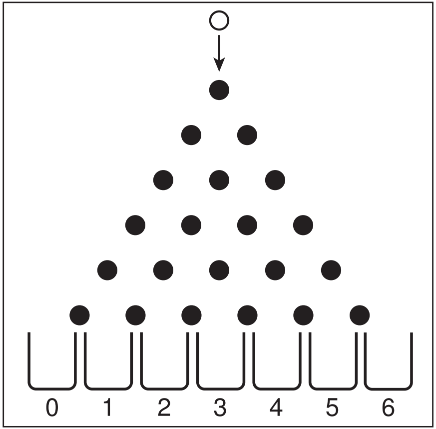
Playing Plinko for Cash



- Suppose that I let you put a ball in the 6-row plinko, and I agree to pay you k dollars if the ball lands in bucket k.
- This is probably going to cost me money!
- How much should I charge you to play?
- How much, on average, am I going to have to pay?

Clicker Question #1

How much should I charge you to play my plinko game (to break even)?
All answers count for now.



A) $1

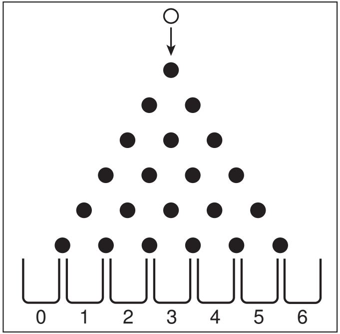
B) $2

C) $3

D) $4

E) $6

A Quick Solution



- Buckets 0 and 6 have equal probabilities. The average payout for these two is $3.
- Buckets 1 and 5 have equal probabilities. The average payout for these two is $3.
- Buckets 2 and 4 have equal probabilities. The average payout for these two is $3.
- The payout for bucket 3 is $3.
- The overall average payout is $3.
- Without knowing any of the actual probabilities!

Random Variables

- Definition: A variable that is assigned a value for each possible outcome or event for a probabilistic process.
- Examples:
  - For a coin toss, we could assign a random variable, x, the value of 1 for heads or 0 for tails.
  - For n successive coin tosses, we could define x to be the number of heads.
  - For the Plinko, we can define the random variable, x, as the number of the bucket that the ball lands in.
  - For any of these, we could define other random variables.

The Expected Value or Expectation

For a random process that has a sample set of n possible outcomes (or a complete set of n non-overlapping events):

- The random variable, x, has values of x_k for $k = 1, 2, 3 \ldots n$
- The n possible outcomes (or events) have probabilities of $p(k)$, for $k = 1, 2, 3 \ldots n$
- The expected value is defined as:

$$E = \sum_{k=1}^{n} p(k) x_k$$

- If the process is repeated a large number of times, the average value of x will approach E.
- For a game of chance, if x_k is the number of dollars paid out for outcome (or event) k, E is the average payout.

The Expected Value is an Average

- The simplest case, where all of the events have equal probabilities:

$$E = \sum_{k=1}^{n} p(k) x_k$$

$$p(k) = \frac{1}{n}$$

$$E = \sum_{k=1}^{n} \frac{1}{n} x_k = \frac{1}{n} \sum_{k=1}^{n} x_k$$

- More generally, *E* is a *weighted* average, in which the values of the more probable outcomes are weighted more heavily than the less probable values.

Expected Value of x for the Unbiased Six-row Plinko

Bucket (k)	x_k	$p(k)$	$p(k)x_k$
0	0	1/64	0
1	1	6/64	6/64
2	2	15/64	30/64
3	3	20/64	60/64
4	4	15/64	60/64
5	5	6/64	30/64
6	6	1/64	6/64
Total		1	192/64 =3

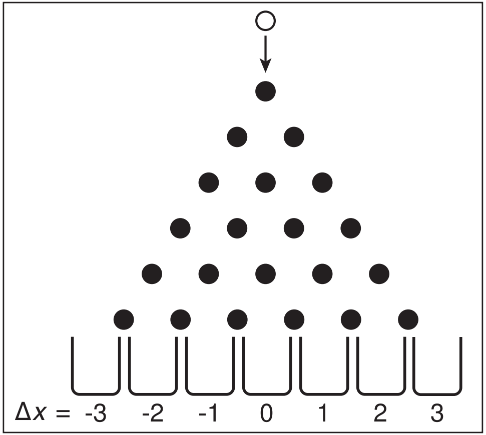
Expected Value of x for a Biased Six-row Plinko: $p(\text{right}) = 0.6$

Bucket (k)	x_k	$p(k)$	$p(k)x_k$
0	0	0.004	0
1	1	0.037	0.037
2	2	0.138	0.276
3	3	0.276	0.829
4	4	0.311	1.244
5	5	0.186	0.933
6	6	0.046	0.280
Total		1	3.6

Expected Value of x for a Biased Six-row Plinko: $p(\text{right}) = 0.4$

Bucket (k)	x_k	$p(k)$	$p(k)x_k$
0	0	0.046	0
1	1	0.187	0.187
2	2	0.311	0.622
3	3	0.276	0.829
4	4	0.138	0.553
5	5	0.037	0.184
6	6	0.004	0.0246
Total		1	2.4

Another Random Variable for the Plinko, Δx



- Δx represents the position of the bucket, relative to the central bucket.
  $\Delta x = x - 3$

Expected Value of Δx for the Unbiased Six-row Plinko

Bucket (k)	Δx_k	$p(k)$	$p(k)\Delta x_k$
0	-3	1/64	-3/64
1	-2	6/64	-12/64
2	-1	15/64	-15/64
3	0	20/64	0
4	1	15/64	15/64
5	2	6/64	12/64
6	3	1/64	3/64
Total		1	0

Expected Value of Δx for a Biased Six-row Plinko: $p(\text{right}) = 0.6$

Bucket (k)	Δx_k	$p(k)$	$p(k)\Delta x_k$
0	-3	0.004	-0.012
1	-2	0.037	-0.074
2	-1	0.138	-0.138
3	0	0.276	0
4	1	0.311	0.311
5	2	0.186	0.373
6	3	0.046	0.139
Total		1	0.6

Notice:

- $\Delta x = x - 3$
- For the unbiased six-row plinko:

$$E(x) = 3$$
$$E(\Delta x) = 0$$

- For the biased six-row plinko:

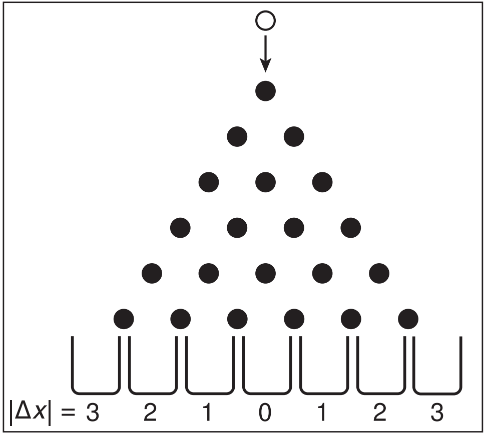
$$E(x) = 3.6$$
$$E(\Delta x) = 0.6$$

- For both, $E(\Delta x) = E(x) - 3$

Two Important Rules

1. If x is a random variable, and a is a constant:
$$E(x + a) = E(x) + a$$
2. And:
$$E(ax) = aE(x)$$

Another Random Variable for the Plinko, $|\Delta x|$



- $|\Delta x|$ represents the distance from the central bucket.

Expected Value of $|\Delta x|$ for the Unbiased Six-row Plinko

Bucket (k)	$\lvert\Delta x\rvert_k$	$p(k)$	$p(k)\,\lvert\Delta x\rvert_k$
0	3	1/64	3/64
1	2	6/64	12/64
2	1	15/64	15/64
3	0	20/64	0
4	1	15/64	15/64
5	2	6/64	12/64
6	3	1/64	3/64
Total		1	60/64 ≈ 0.94

Expected Value of $|\Delta x|$ for a Biased Six-row Plinko: $p(\text{right}) = 0.6$

Bucket (k)	$\lvert\Delta x\rvert_k$	$p(k)$	$p(k)\,\lvert\Delta x\rvert_k$
0	3	0.004	0.012
1	2	0.037	0.074
2	1	0.138	0.138
3	0	0.276	0
4	1	0.311	0.311
5	2	0.186	0.373
6	3	0.046	0.139
Total		1	1.05

Clicker Question #2

What is the expected value of $|\Delta x|$ for a biased six-row plinko with $p(\text{right}) = 0.4$?

A) -1.05

B) -0.94

C) 0

D) 0.94

E) 1.05

Expected Value of $|\Delta x|$ for a Biased Six-row Plinko: $p(\text{right}) = 0.4$

Bucket	$\|\Delta x\|$	$p(\|\Delta x\|)$	$p(\|\Delta x\|)\,\|\Delta x\|$
0	3	0.046	0.140
1	2	0.187	0.373
2	1	0.311	0.311
3	0	0.276	0
4	1	0.138	0.138
5	2	0.037	0.074
6	3	0.004	0.012
Total		1	1.05

Two Important Parameters for any Discrete Random Variable

- Expected value, also called the mean (μ)

$$\mu = \sum_{k=1}^{n} p(k) x_k = E(x)$$

- Variance (σ^2)

$$\sigma^2 = \sum_{k=1}^{n} p(k)(x_k - \mu)^2$$

  - Mean of the squares of the differences between x-values and the mean.
  - A measure of the width of the distribution of x values around the mean.
  - Squares are taken so that both positive and negative differences contribute.
  - Square root of the variance, σ, is called the standard deviation.
    σ has the same dimensions as x and μ.

Mean, Variance and Standard Deviation for the Binomial Probability Distribution Function

- The probability function:

$$p(k; n, p_s) = \frac{n!}{k!(n-k)!} p_s^k (1-p_s)^{(n-k)}$$

for $k = 0$ to n.

- The mean of k:

$$\mu = np_s$$

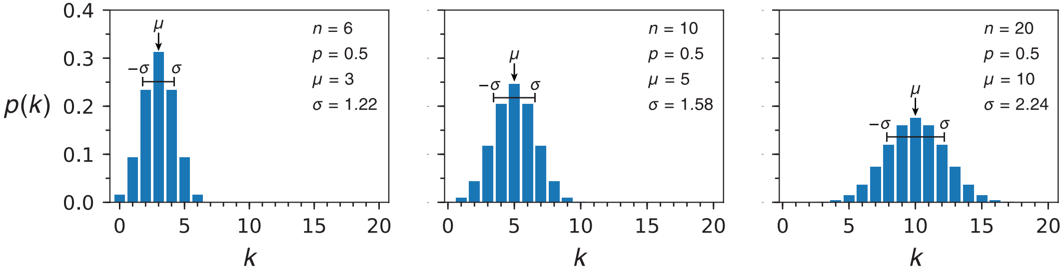
- The variance:

$$\sigma^2 = np_s(1-p_s)$$

- The standard deviation:

$$\sigma = \sqrt{np_s(1-p_s)}$$

Effect of n on the Mean and Standard Deviation for the Binomial Probability Distribution Function

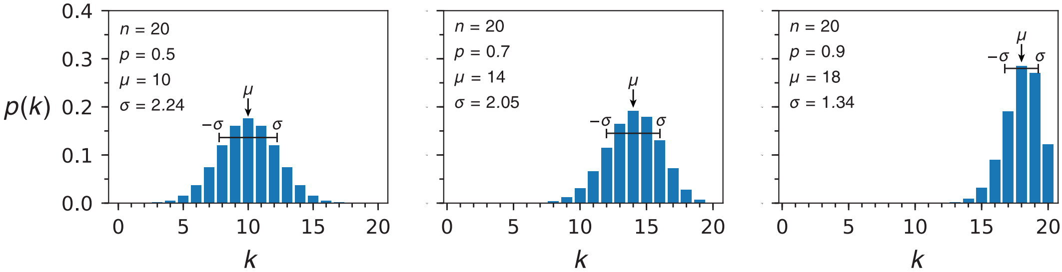


$$\mu = np_s$$

$$\sigma = \sqrt{np(1 - p_s)}$$

- The distribution gets wider as n gets larger, by the factor $\sqrt{n}$.

Effect of p_s on the Mean and Standard Deviation for the Binomial Probability Distribution Function

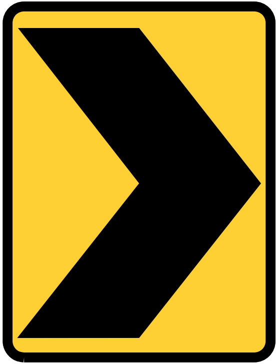


$$\mu = np_s$$

$$\sigma = \sqrt{np_s(1 - p_s)}$$

- For a given value of n, the distribution is widest when $p_s = 0.5$.

Warning!



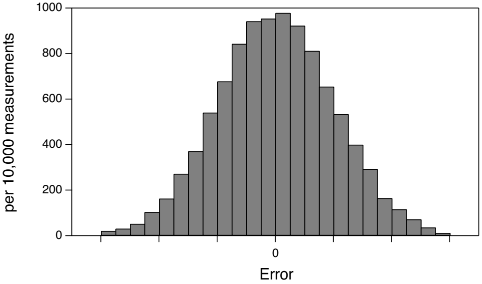
Direction Change

A Slight Diversion: Statistics of Experimental Data

Analyzing Experimental Data: The Working Model

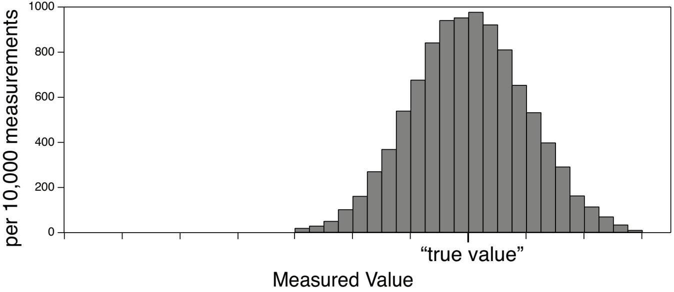
Assumptions:

- The measured values are determined by a “true” value plus random error (positive or negative).
- The random errors are distributed according to a Gaussian function, *i.e.,* a “bell curve”.



- Why is it bell shaped?

Estimating the “True” Value



- The best* estimate of the “true” value is the mean, $\bar{x}$.

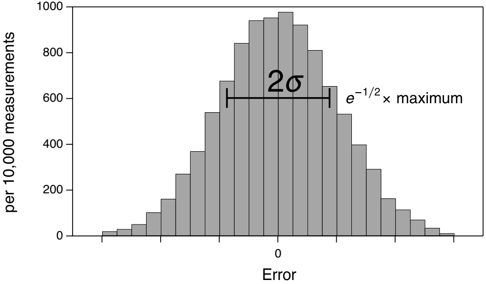
$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

N= number of measurements, x_i is the i^{th} measurement

---

* “Best” means most likely to give the correct value.

Estimating the Distribution Width (σ)

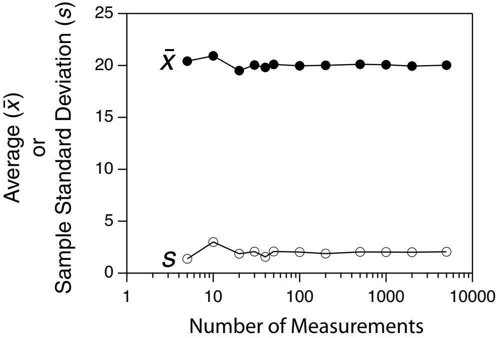


Two ways to estimate σ, the standard deviation:

- From a histogram (takes lots of measurements!)
- The sample standard deviation, s:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{N-1}}$$ an <u>estimate</u> of σ

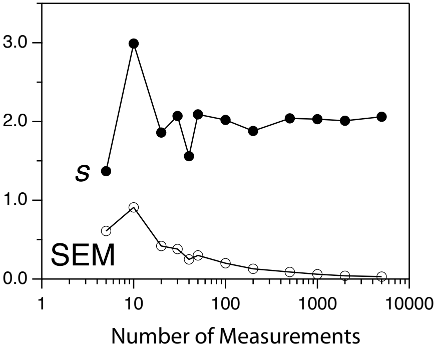
Estimates Improve With More Measurements (A Simulation)



- Estimate of true value ($\bar{x}$) approaches a limiting value (20 mg)
- Estimate of standard deviation (s) approaches a limiting value (2 mg)
- s doesn't approach zero.

Another Useful Statistic: The Standard Error of the Mean (SEM)

$$SEM = \sqrt{\frac{\sum (x - \bar{x})^2}{(N-1)N}} = s/\sqrt{N}$$



- The standard error of the mean represents the uncertainty in the estimate of the mean, $\bar{x}$
- The uncertainty in $\bar{x}$ decreases with more measurements.
- The uncertainty in the mean can be made as small as we like, if we make enough measurements! (Assumes that errors are truly random.)
- Decreasing the uncertainty by half requires four times as many measurements.