Statistically Speaking: *Measurements and Distributions*

DAVID L. TABB, PH.D.

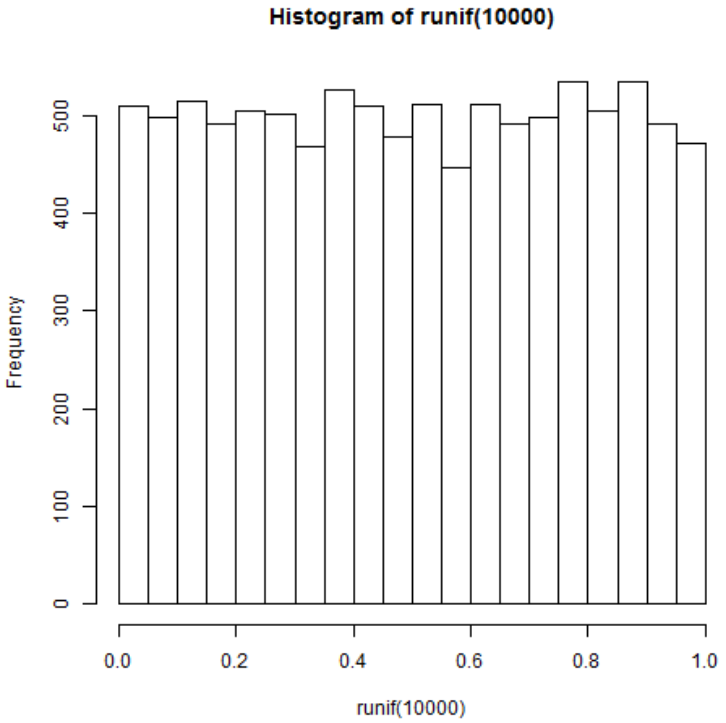
AUGUST 24, 2017

Overview

- Distributions
- Sampling: “Independent, Identically Distributed”
- Log and Linear
- Discrete vs. Continuous

Uniform distribution

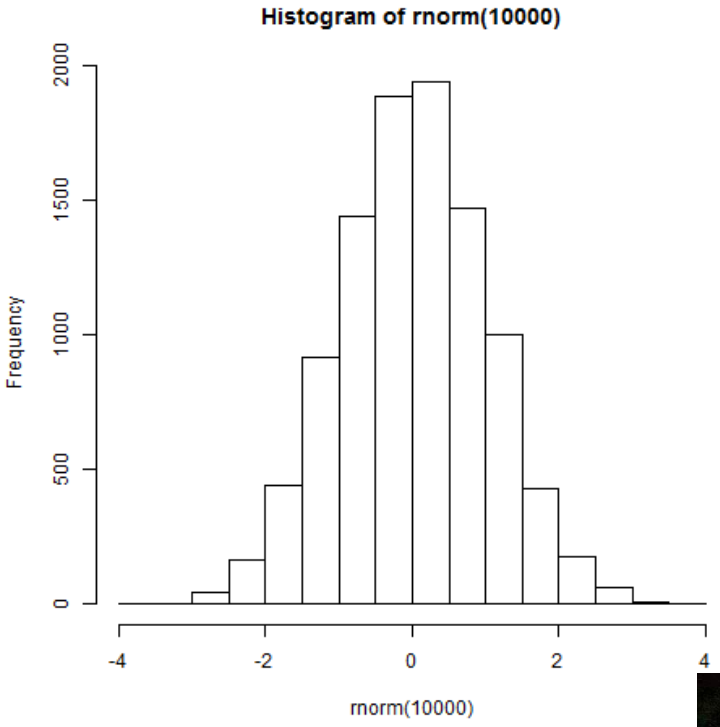
- Any value is just as likely as any other.
- Distribution is *bounded* on both sides.
- Rolling a die is a common example.



The above is a histogram; for the distribution, $\int x = 1$

Gaussian or normal distribution

- Central values are more likely than distant values.
- When many factors contribute to value, a normal dist'n is likely.
- Adult male or female height is a common example.

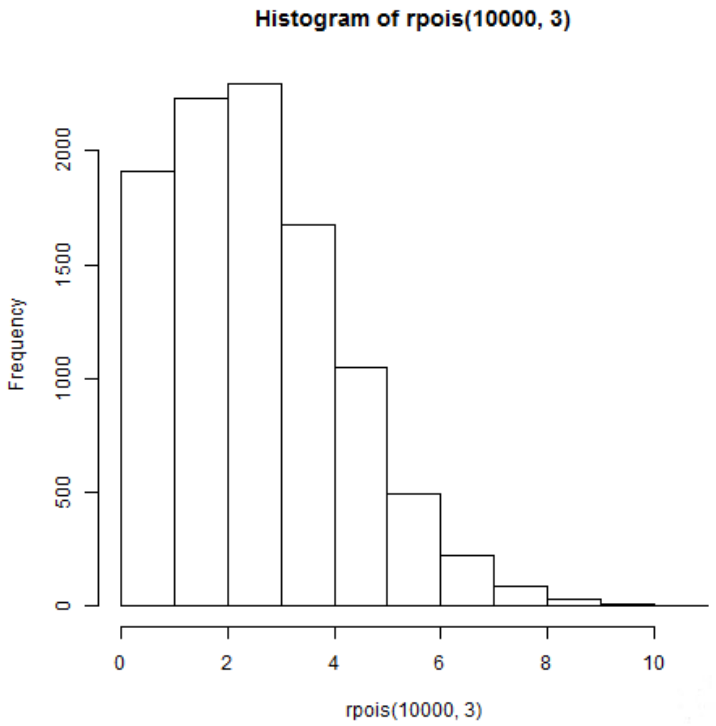


Wikimedia Commons

Carl Gauss (1777-1855)

Poisson distribution

- Given an average rate of occurrence, how many events will take place in an interval?
- How many light bulbs will burn out each year?

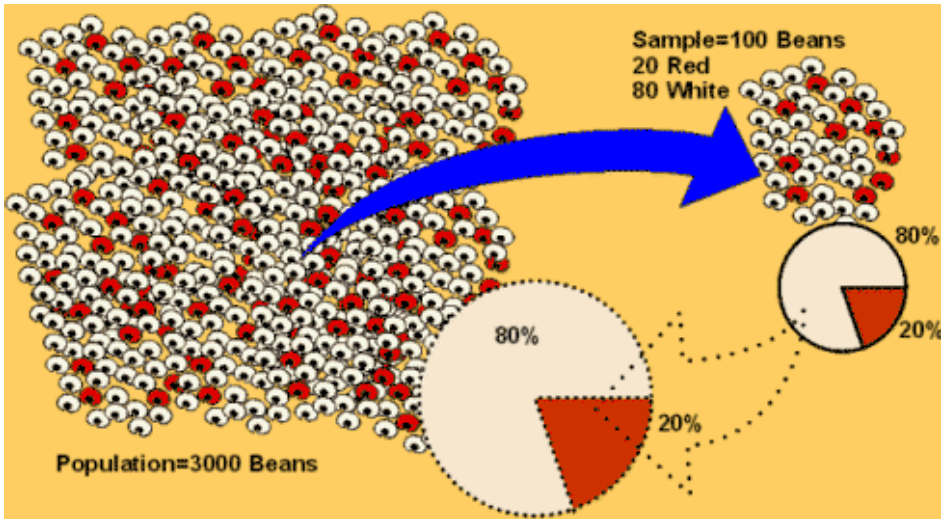


Wikimedia Commons

Siméon Denis Poisson (1781-1840)

Sampling

- We cannot measure all of anything; we must *extrapolate* from a sample.
- We often call the number of specimens or patients the '*n*'.



http://www.strategosinc.com

Bigger '*n*' usually implies better estimation

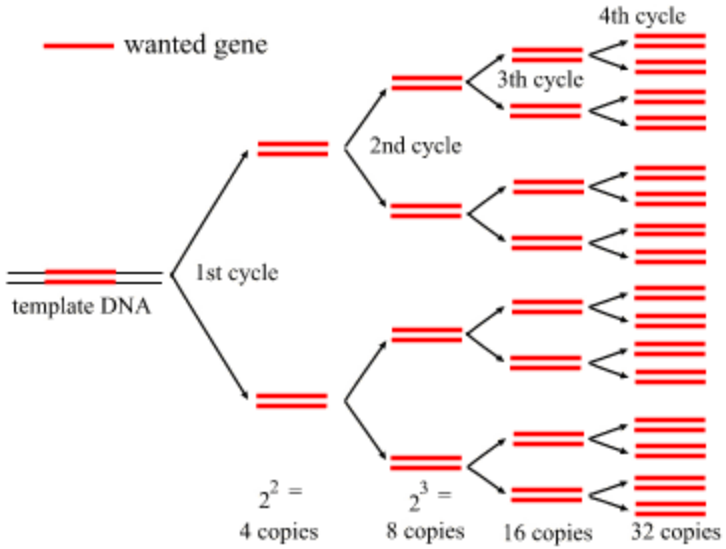
Independent, Identically Distributed

- A jar holds 100 black and 900 red marbles.
- If I have drawn and removed 20 black and 150 red marbles, is the probability that I draw a black marble the same as before?
- I rolled a fair die 10x: 4 4 3 2 6 3 1 3 4 4
- What is the probability the next number is a 4?
- Die rolls are *independent*!

ceiling(6*(runif(10)))

PCR: Exponentials in Biotechnology

- Each cycle of PCR doubles the amount of the wanted gene (ideally).
- If you count cycles until detection of gene is possible, you have a *logarithmic* measurement!

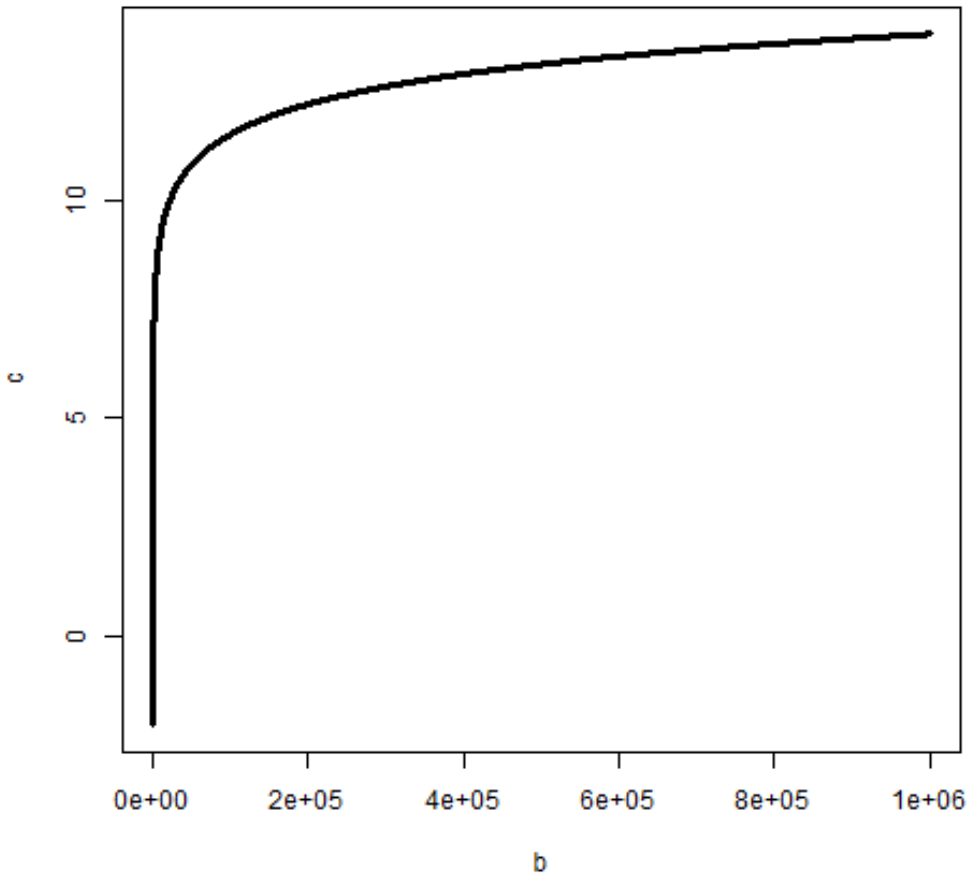


Andy Vierstraete
https://users.ugent.be/~avierstr

What is a logarithm?

- $\log_A B = C$
- $A^C = B$

- For this "natural log" graph, A is set to *e*= 2.718282



e is natural because curve slope = 1 when e^0 (Euler)

Measures divide by type

CONTINUOUS

- Height, Width, Depth
- Time
- pH, Concentration
- Temperature, Pressure

DISCRETE

- Counts / Integers
- Categories / Types
- Genders
- Dice rolls

Reducing a continuous variable to a discrete one loses information.

Closing thoughts

- Our data come from distributions, and knowing a few key types is highly valuable.
- Dependencies among our data should alter our expectations of them.
- Logarithmic expressions are common in biotechnology. Know how to interpret them.
- When designing an experiment, limit the times you record continuous data in discrete forms.