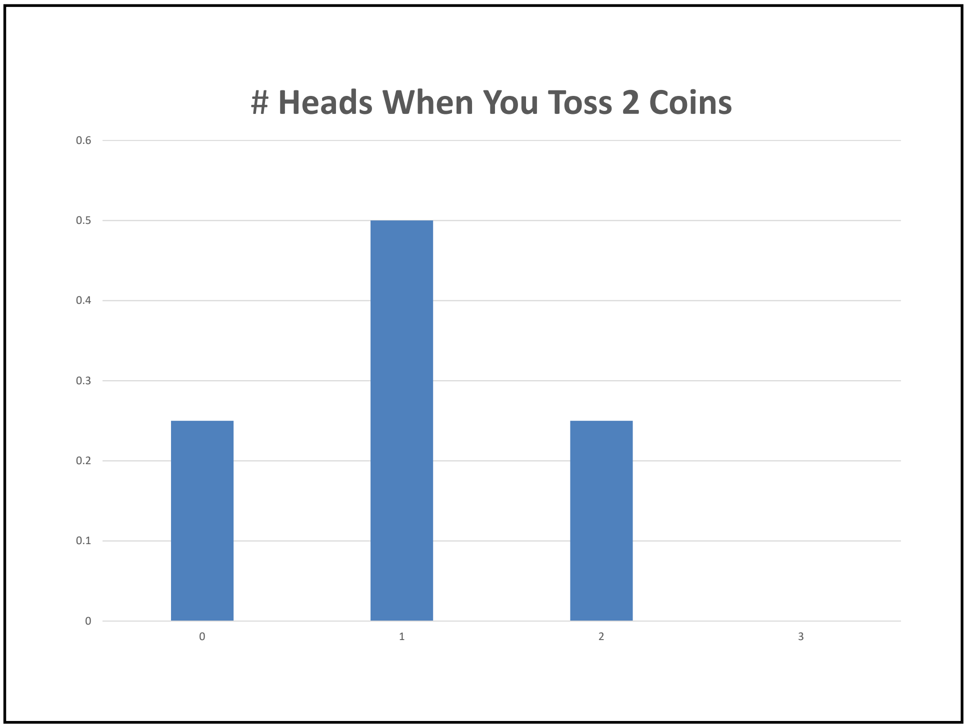
Statistics
Economics 91

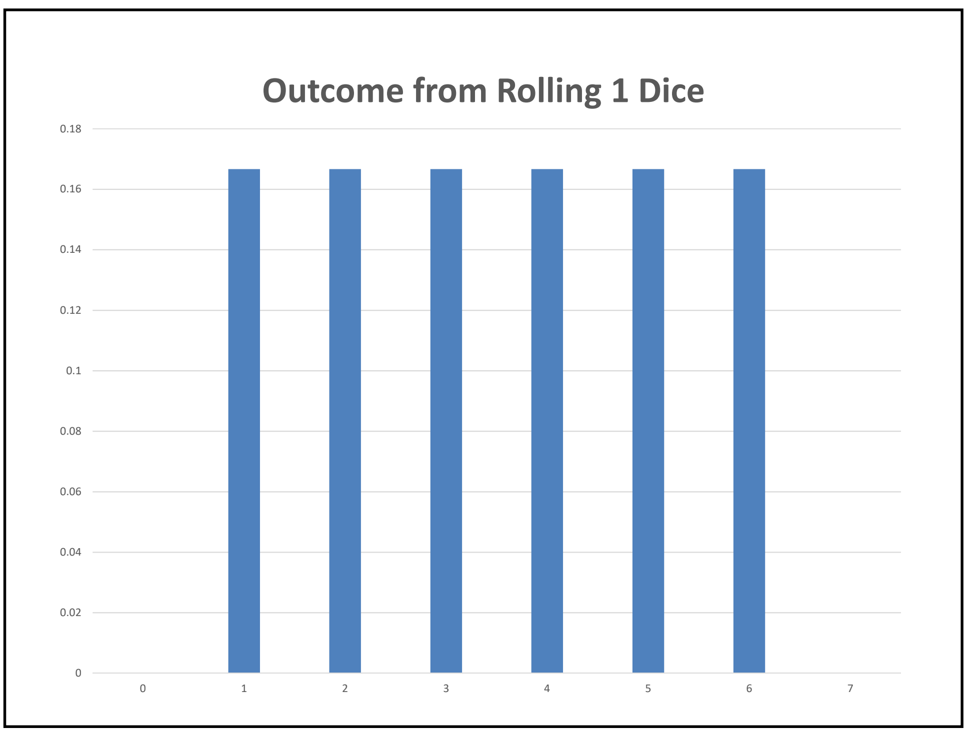
Linus Yamane

Probability Distributions

- Random Variables
  – Discrete
  – Continuous
- Probability Distributions
  – P(x) gives the probability of each possible value x of a random variable X
  – We have discrete probability distributions and continuous probability distributions



Heads When You Toss 2 Coins
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3



Outcome from Rolling 1 Dice
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
1
2
3
4
5
6
7

Probability Distributions

- Mean

$$\mu = E[X] = \sum xP(x)$$

- Variance

$$\sigma^2 = \mathrm{E}[(X-\mu)^2] = \sum (x-\mu)^2\, P(x)$$

- Standard Deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{E[(X-\mu)^2]}$$

Note: we are **not** talking about a sample here

Chuck a Luck

Roulette Wheel

Return on $100 Investment

Investment	Return	E(X)	Var(X)	SD(X)
1	$5 with certainty	$5	0	0
2	$10 with 50% $0 with 50%	$5	25	5
3	$5 with 50% $10 with 25% $0 with 25%	$5	12.50	3.54
4	$5 with 50% $105 with 25% -$95 with 25%	$5	5000	70.71

Risk and Return

Asset	Houses	Treasury Bills 1 year	Treasury Bills 3 year	Dow Jones Industrials	S&P 500 Index
Real Annualized Return	0.2%	1.5%	1.9%	3.3%	3.6%
Risk (SD)	4.9%	2.3%	2.4%	16.3%	16.5%
Time Period	1950-present	1953-present	1953-present	1950-present	1950-present

You can enjoy a higher rate of return by bearing more risk

Consider Two Investments

Investment A		
Outcome	-$15.04	$43.04
Probability	0.50	0.50

Investment B				
Outcome	$0	$5	$10	$100
Probability	0.49	0.02	0.39	0.10

Do these investment feel equally risky?

μ = \$14 and $\sigma = \$29.04$

Linear Functions of RVs

- Linear functions of random variables
  - Let $Z = a + bX$ (*a* and *b* are constants)
  - where $\mu_x = mean\ of\ x$
  - and $\sigma_x^2 = variance\ of\ x$
- What is the mean and variance of Z?
  - Mean of Z = $E[Z] = \mu_z = a + b\mu_x$
  - Variance of Z = $E(Z - \mu_Z)^2 = \sigma_Z^2 = b^2\sigma_x^2$
  - SD(Z) = $\sigma_z = b\sigma_x$

Joint Probability Distributions

- $P(x, y) = P(X = x\ and\ Y = y)$
  $= P(X = x\ \cap Y = y)$
- $0 \leq P(x, y) \leq 1$
- $\sum_x \sum_y P(x, y) = 1$
- If X and Y are independent
  $P(x, y) = P(x)P(y)$
  for all x and y

Court Room		Innocent	Guilty	
		Y=0	Y=1	
Acquitted	X=0	18	15	33
Convicted	X=1	7	60	67
		25	75	100

		Calvin No Date	Calvin Date	
		Y=0	Y=1	
Hobbes No Date	X=0	42	18	60
Hobbes Date	X=1	28	12	40
		70	30	100

Joint Probability Distribution

- Covariance

$$cov(x,y) = \sigma_{xy}$$

$$= \sum_x \sum_y (x - \mu_x)\left(y - \mu_y\right) P(x,y)$$

- Correlation Coefficient

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{cov(x,y)}{SD(x)SD(y)}$$

- If *X* and *Y* are independent, cov(*x*,*y*)=0
- NOTE: We are not talking about samples here!

Sum of Random Variables

- Let $Z = a + bX + cY$
  - Given the mean and variance of X and Y
  - Constants *a*, *b*, and *c*
- Mean of Z

$$\mu_Z = a + b\mu_X + c\mu_Y$$

- Variance of Z

$$\sigma_Z^2 = b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\sigma_{xy}$$

Special Probability Distributions

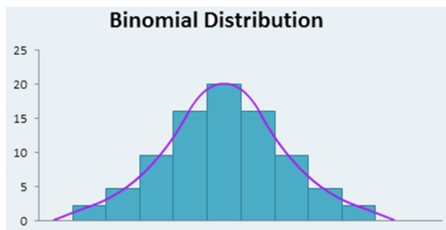
- Bernoulli Trial
- Binomial Distribution
- Poisson Distribution
- Exponential Distribution
- Normal Distribution
- Chi-Square Distribution
- Student *t* Distribution
- Snedecor's *F* distribution

Bernoulli Trial

- 2 possible outcomes (success, failure)
- Repeated trials are independent
- Let X = 1 if success, X = 0 if failure
- P(success) = π and P(failure) = $1 - \pi$
- P(1) = π and P(0) = $1 - \pi$
- Mean = E(X) = π
- Variance = V(X)= $\pi (1 - \pi)$

Binomial Distribution

- Bernoulli trials with probability of success π
- Probability of exactly *n* successes in *n* trials
- $P(x\ successes) = C_x^n \pi^x (1-\pi)^{n-x}$



Binomial Distribution

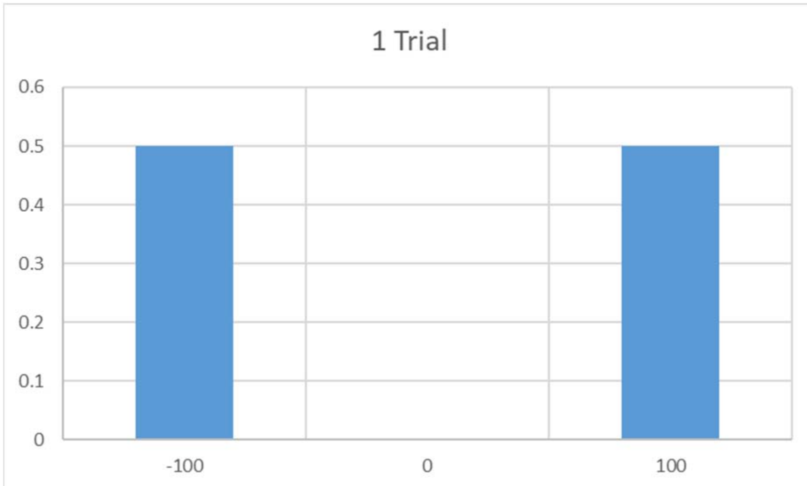
- Mean = $E(X) = n\pi$
- Variance = $V(X) = n\pi(1-\pi)$

Run on a Bank

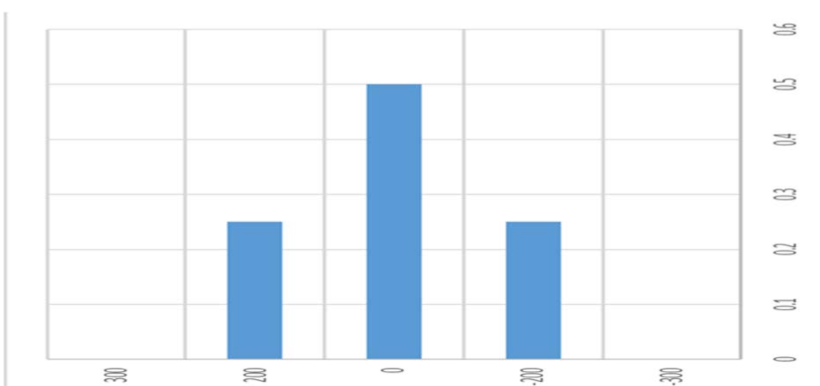
Bank with 2 $1 deposits	
$ withdrawn	probability
0	.81
1	.18
2	.01

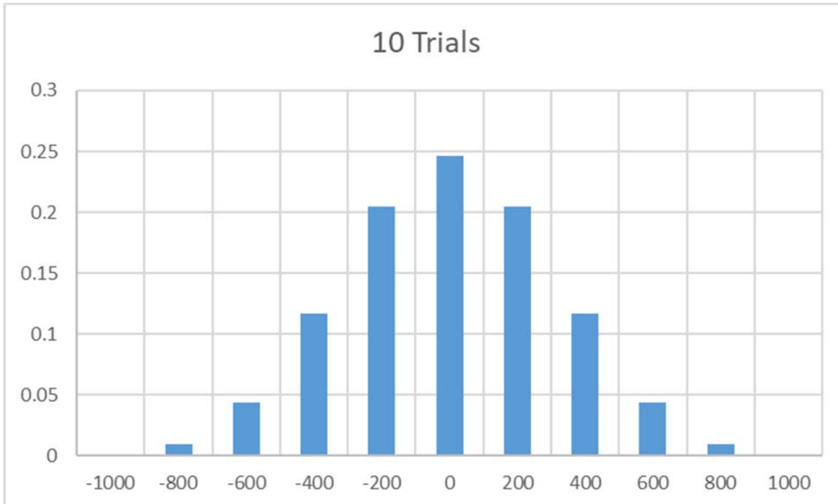
Run on a Bank

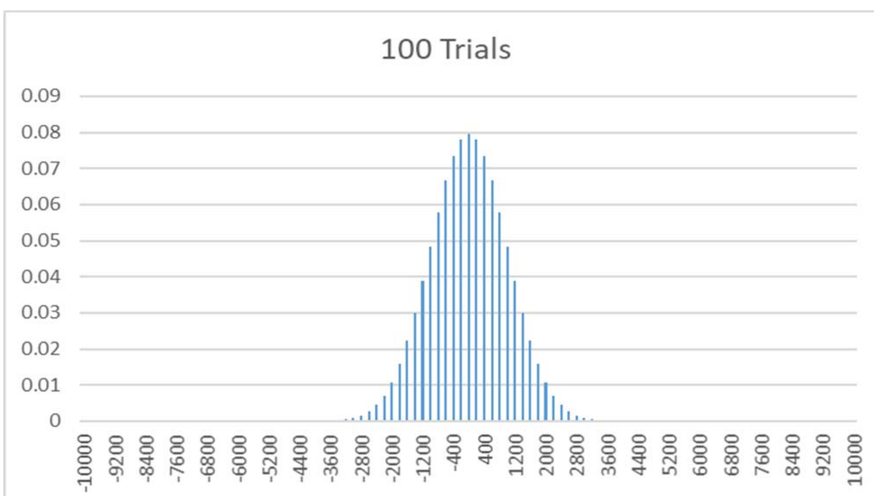
Bank with 10 $1 deposits	
$ withdrawn	probability
0	0.3487
1	0.3874
2	0.1937
3	0.05739
4	0.01116
5	0.00149
6	0.000137
7	0.0000087
8	0.000000364
9	0.000000009
10	0.0000000001

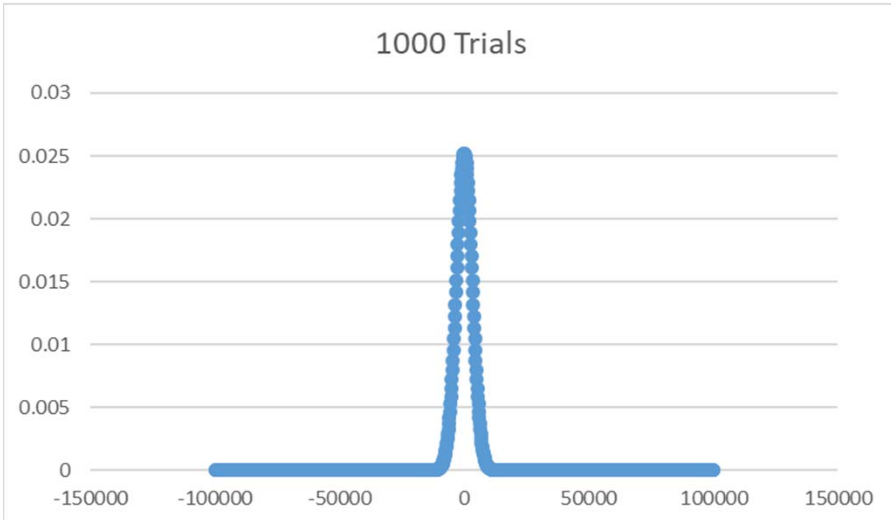

1 Trial
0.6
0.5
0.4
0.3
0.2
0.1
0
-100
0
100

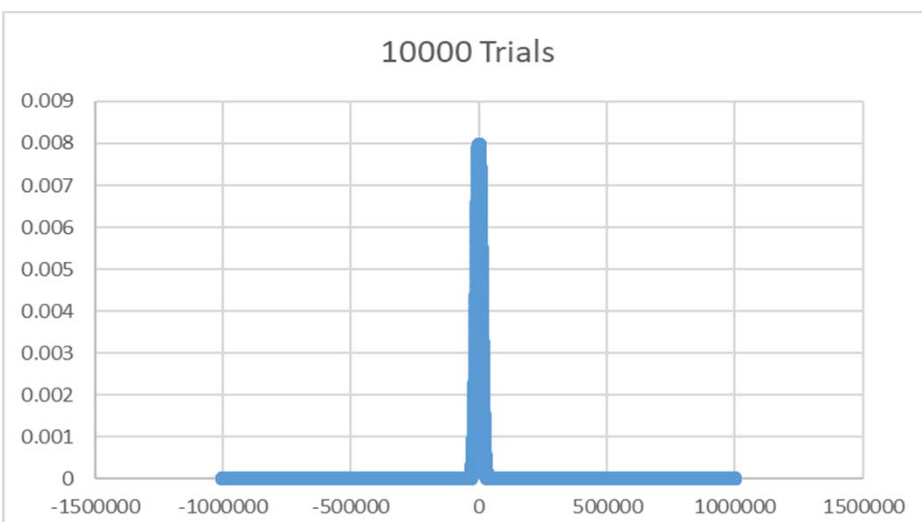
2 Trials

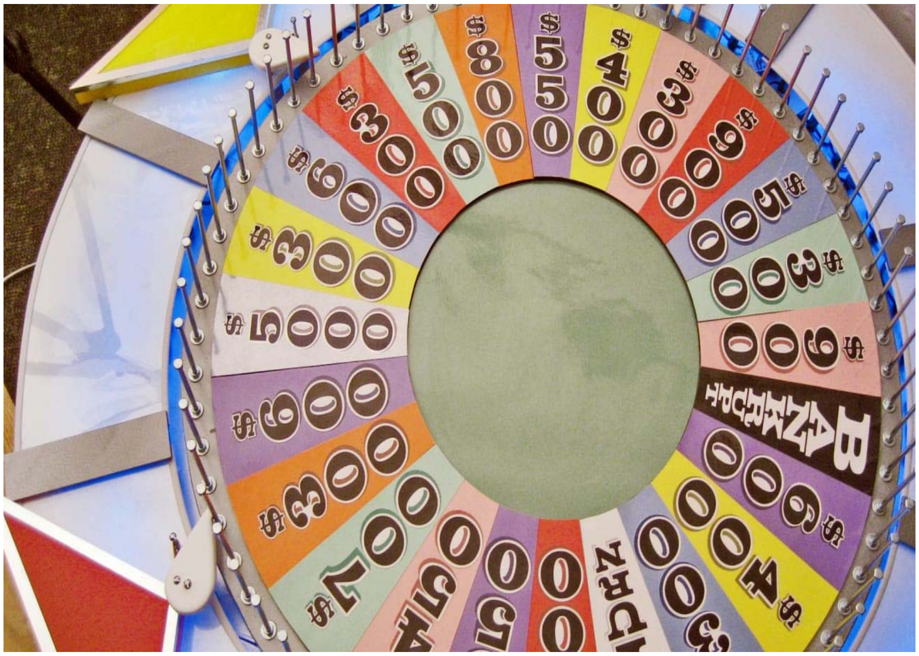

0.6
0.5
0.4
0.3
0.2
0.1
0
300
200
0
-200
-300


10 Trials
0.3
0.25
0.2
0.15
0.1
0.05
0
-1000
-800
-600
-400
-200
0
200
400
600
800
1000


100 Trials
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-10000
-9200
-8400
-7600
-6800
-6000
-5200
-4400
-3600
-2800
-2000
-1200
-400
400
1200
2000
2800
3600
4400
5200
6000
6800
7600
8400
9200
10000


1000 Trials
0.03
0.025
0.02
0.015
0.01
0.005
0
-150000
-100000
-50000
0
50000
100000
150000

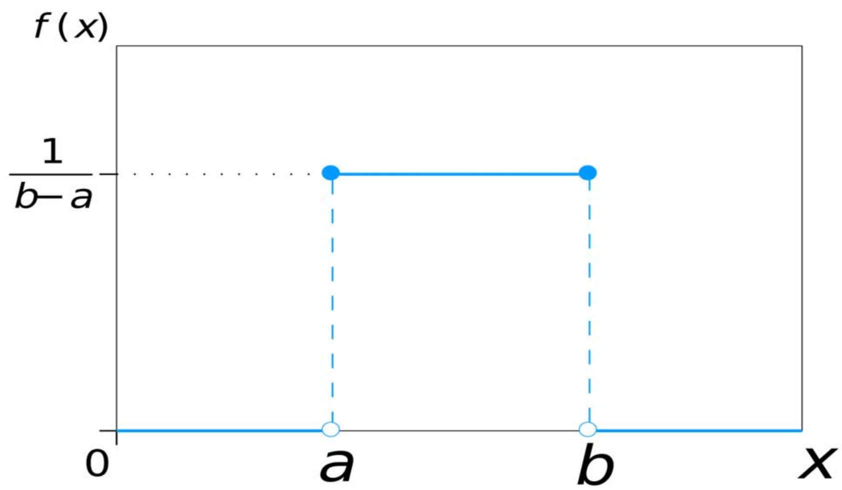

10000 Trials
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
-1500000
-1000000
-500000
0
500000
1000000
1500000



Continuous Probability Distributions

- Possible outcomes are not countable
- Continuous random variable
  - Time, distance, temperature
- Continuous probability distribution
- Each possible outcome has zero probability
- An interval of possible outcomes has nonzero probability

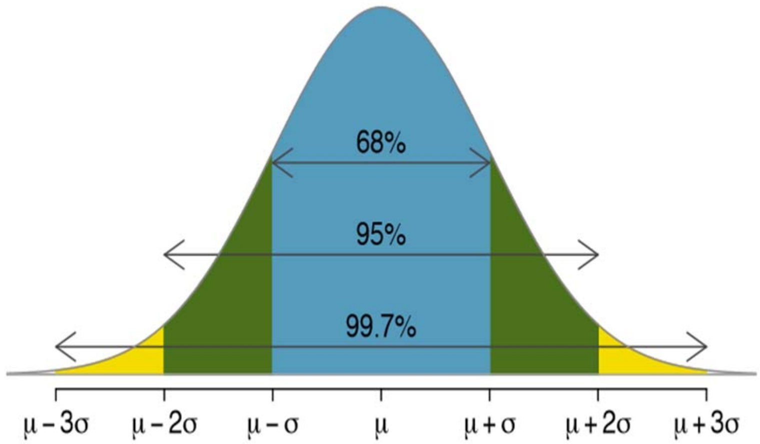
Uniform Probability Distribution



Normal Distribution

- Probability density function
- $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
- Mean = $E(X) = \mu$
- Variance = $V(X) = \sigma^2$
- Mean and variance must be finite
- Gaussian or Laplacian distribution

Normal Distribution



Normal Distribution and Risk

Deviations Relative to Mean		
Range	Probability of Being in Range	P(outside of Range) 1 in n
+/- 1σ	68.8%	3
+/- 2σ	95.5%	22
+/- 3σ	99.7%	370
+/- 4σ	99.994%	15,787
+/- 5σ	99.99994%	1,744,278
+/- 6σ	99.9999998%	506,797,346

Normal Distribution and Financial Assets

- Expected rate of occurrence of a negative shock (Normal Distribution)
- Assume 250 trading days a year

Shock	Interval		Shock	Interval
1σ	6 Days		5σ	14,000 years
2σ	44 days		6σ	4 million years
3σ	3 years		7σ	3 billion years
4σ	126 years		8σ	6 trillion years

Equivalent probabilities of some larger shocks

20σ : Winning the cosmological lottery (drawing one atomic particle vs all particles in the universe)

25σ : Winning the CA lottery 17-18 times in a row

Normal Distribution

- Defined by the mean μ and variance σ^2
- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- Made famous by the Central Limit Theorem
- Sample means are normally distributed even if the original variables themselves are not normally distributed as long as we have finite means and finite variances.

Chi-Squared Distribution

- The chi-squared distribution (also chi-square or χ_k^2-distribution) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables
- Mean = $E\left(\chi_k^2\right) = k$
- Variance = $V\left(\chi_k^2\right) = 2k$
- Mean and variance must be finite.

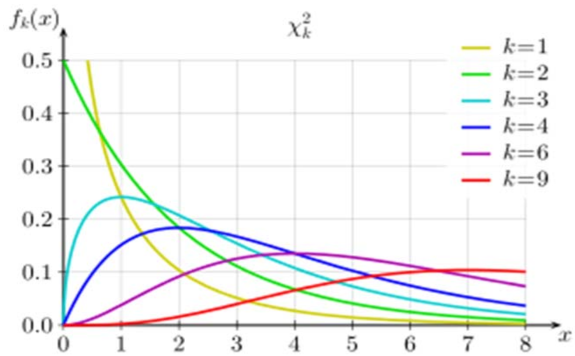
Chi-Squared Distribution

- Friedrich Robert Helmert (1843 – 1917) discovered in 1876

- Rediscovered by Karl Pearson (1857-1936) in 1900

Chi-Squared Distribution

- Tests of variance
- Tests of independence
- Tests of goodness of fit

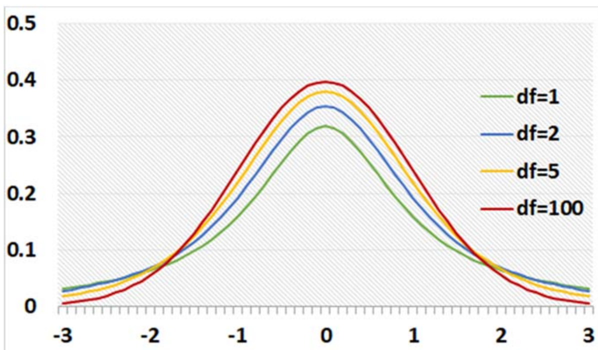


Student's *t* Distribution

- Similar to the normal distribution
- Symmetrical, bell-shaped, mean of zero
- Has fatter tails than a normal
- If $Z \sim N(0,1)$ and $W \sim \chi_k^2$ then $\frac{Z}{\sqrt{W/k}} \sim Student's\ t$ with *k* degrees of freedom
- Mean = $E(t_k) = 0$
- Variance = $V(t_k) = \frac{k}{k-2}$ for *k*>2

Student’s *t* Distribution

- Developed by William S. Gossett in 1908 while working for Guinness in Dublin
- Used the pseudonym Student



Student’s *t* Distribution

- Estimating means of samples when σ is unknown
- Estimating differences between two means
- Analyzing regression results

F-Distribution

- Fisher-Snedecor F-Distribution
- Ronald Fisher (1890-1962)
- George Snedecor (1881-1974)

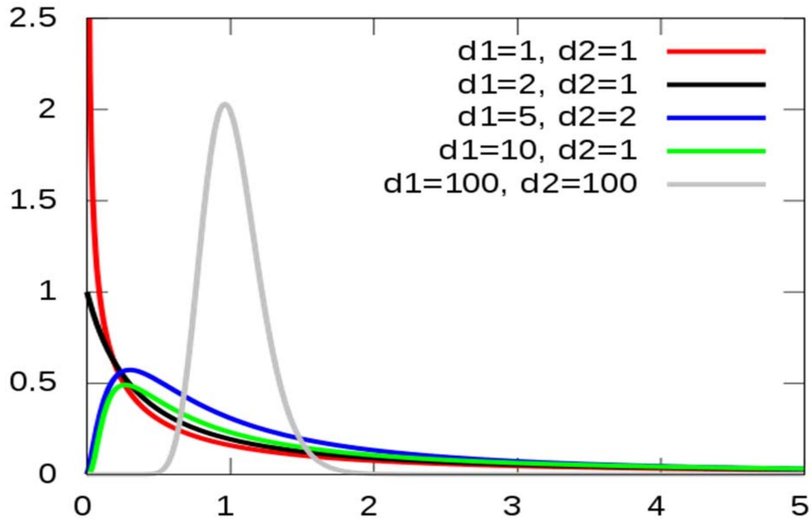
F-Distribution

- Basically the ratio of two Chi-squared random variables
- If $W \sim \chi_m^2$ and $Y \sim \chi_n^2$, both independent, then

$$\frac{W/_m}{Y/_n} \sim F(m, n)$$

- Used in comparing the variances of two populations
- Used in analysis of variance (ANOVA)

F-Distribution



Poisson Distribution

- Number of car accidents every day
- Number of arrivals of customers at the Pit Stop per hour
- Number of calls to the Dean of Students office every hour
- Pieces of mail per day you receive
- Siméon Denis Poisson (1781-1840)

Poisson Distribution

- Assume the probability of an event occurring is proportional to the length of the time interval
- The probability of two or more occurrences is negligible compared to the probability of one occurrence
- And the number of occurrences in any non-overlapping time intervals are independent

Poisson Distribution

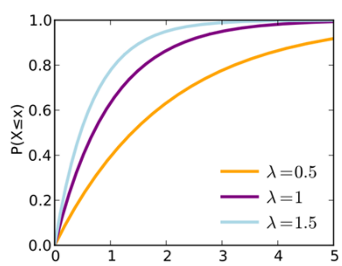
- Discrete probability distribution where
- $P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$
  - Where $x = 0, 1, 2, 3, \ldots$
  - e = 2.71828 (base of natural log)
  - λ = average number of success in the relevant time period
- Mean $E[X] = \lambda$
- Variance $V[X] = \lambda$

Exponential Distribution

- The amount of time it will take before the next accident on the 10 freeway
- The amount of time before the Uber driver shows up
- The amount of time it takes before the next customer comes in the door
- A continuous probability distribution describing the time between events in a Poisson point process

Exponential Distribution

- $P(X \leq x) = 0$ for $x \leq 0$
- $P(X \leq x) = 1 - e^{\frac{-x}{\mu}}$ for $x \geq 0$
- $P(X \geq x) = 1$ for $x \leq 0$
- $P(X \geq x) = e^{\frac{-x}{\mu}}$ for $x \geq 0$



- μ = mean time
- e = 2.71828 (base of natural log)

$\lambda = \frac{1}{\mu}$

Exponential Distribution

- Mean = $E(X) = \mu$
- Variance = $V(X) = \mu^2$
- Probability distribution only takes on positive values
- Probability distribution is continuous
- Probability distribution is NOT symmetric
- Probability distribution has NO memory

Memorylessness

- $P(T > s + t \mid T > s) = P(T > t)$
  for all $s, t \geq 0$

- The distribution of this "waiting time" until a certain event does not depend on how much time has already elapsed
- Not the perfect description of light bulbs, but an approximation