Statistics Tutorial (Part 2 of 2)

To understand God's thoughts, we must study statistics, for these are the measure of His purpose.

-- Florence Nightingale (1820-1910)
English nurse, writer and statistician

What we learned last time

- expected value, variance, standard deviation, covariance, correlation
- discrete and continuous probability distributions
- Finance application – mean and variance of a two-asset portfolio

In this lecture. . .

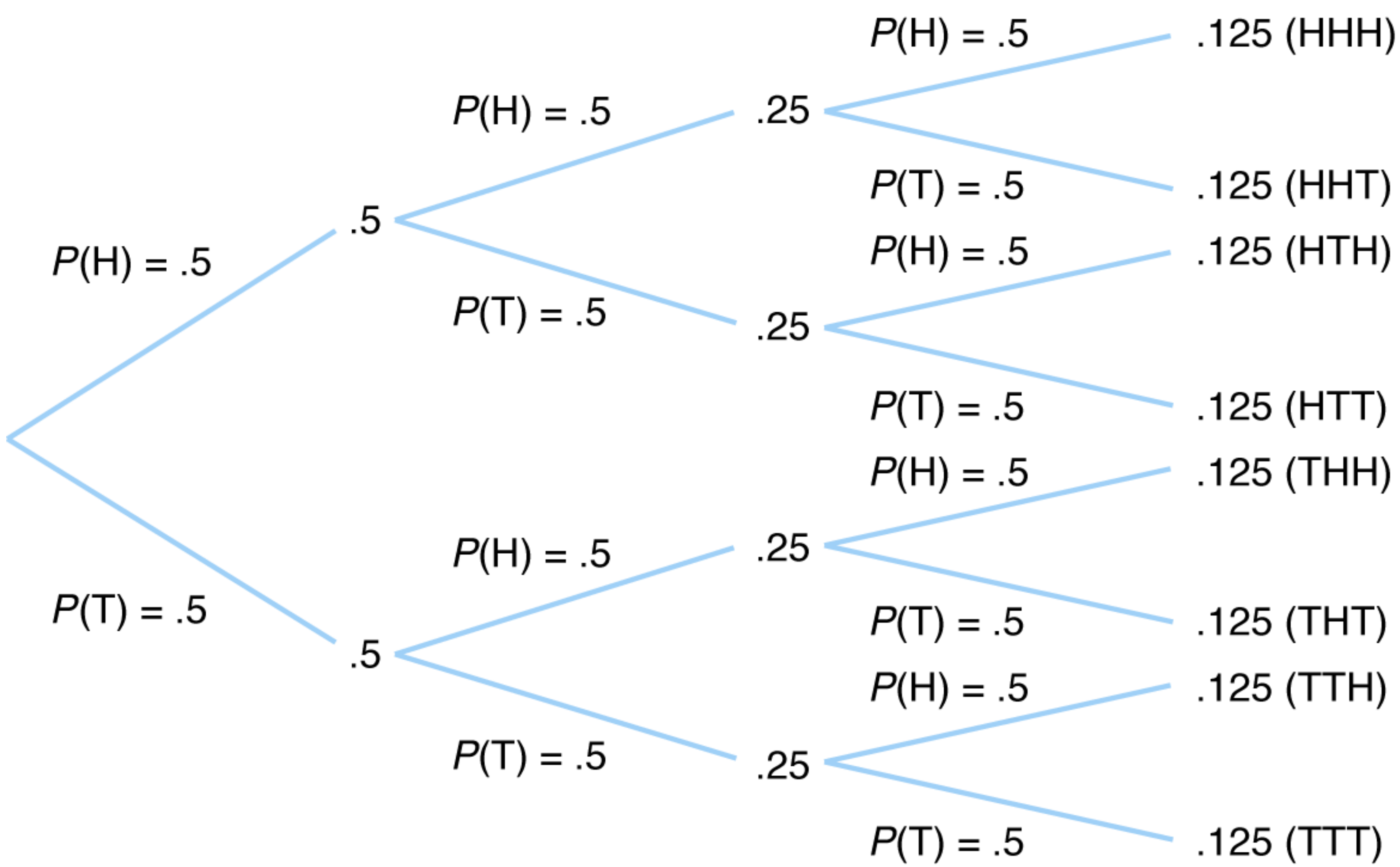
- Bernoulli processes and the binomial distribution
- Central Limit Theorem
- Normal Distribution

Bernoulli Processes

- Properties of a Bernoulli Process:
  - Two possible outcomes for each trial; i.e., one outcome occurs with probability p, whereas the other outcome occurs with probability $1-p$.
  - Outcome probability (p) is constant over time.
  - Trial outcomes are statistically independent.
  - The number of trials is discrete (i.e., equal to n, which is a finite number).

Canonical Example of a Bernoulli Process

- Consider the possible set of outcomes for 3 consecutive coin tosses:



Binomial Distribution

- A *binomial probability distribution function* is used to determine the probability of a number of "successes" in n trials.
- It is a *discrete probability distribution* since the number of successes and trials is discrete.

$$P(r) = \frac{n!}{r!(n-r)!} \cdot p^r \cdot q^{n-r}$$

where: $p =$ probability of a "success"

$q = 1 - p =$ probability of a "failure"

$n =$ number of trials

$r =$ number of "successes" in n trials

Binomial Distribution

- Determine probability of getting exactly two tails in three coin tosses:

$$P(2 \text{ tails}) = P(r = 2) = \frac{n!}{r!(n-r)!} \cdot p^r \cdot q^{n-r}$$

$$= \frac{3 \cdot 2 \cdot 1}{2 \cdot 1(1)} \cdot .25 \cdot .5 = \frac{6}{2} \cdot .125 = .375.$$

Binomial Distribution

- See "Coin Toss Sequences and Probabilities"!

The Central Limit Theorem

- Abraham de Moivre (1667-1754) invented the Central Limit Theorem (CLT).
  - Central Limit Theorem: The distribution of the mean value of a set of "n" *independent* and *identically distributed* random variables, each having mean μ and variance σ^2, approaches a normal distribution with mean μ and variance σ^2/n as n tends toward infinity.
  - In other words, the probability distribution of an average value tends to be normally distributed, even when the distribution from which the average is computed is not normally distributed.

Central Limit Theorem

- Galton Board Demonstration – convergence of binomial to normal: https://youtu.be/6YDHBFVIvIs
- Mathematica demonstration: http://bit.ly/dMdkHX (involving sums of standardized binomial variables)!

Implications of Central Limit Theorem

- What we learned from our numerical example of Central Limit Theorem:
  - As the sample size grows larger, the average is very nearly normally distributed, even though the parent distribution looks anything ***but*** normal.
  - The average will tend to become normally distributed as the sample size increases, regardless of the distribution from which the average is taken.

Implications of Central Limit Theorem

- Why do we care?
  - Since probability distributions typically used in finance and risk management have means and variances, the Central Limit Theorem generally applies.
  - Thus, the normal distribution is widely used in finance and risk management!
    - In portfolio theory, the dominant model for asset allocation decisions is the mean-variance model, which assumes that variance = risk.

The Normal Distribution

- A continuous random variable x has a normal distribution if its probability density function is

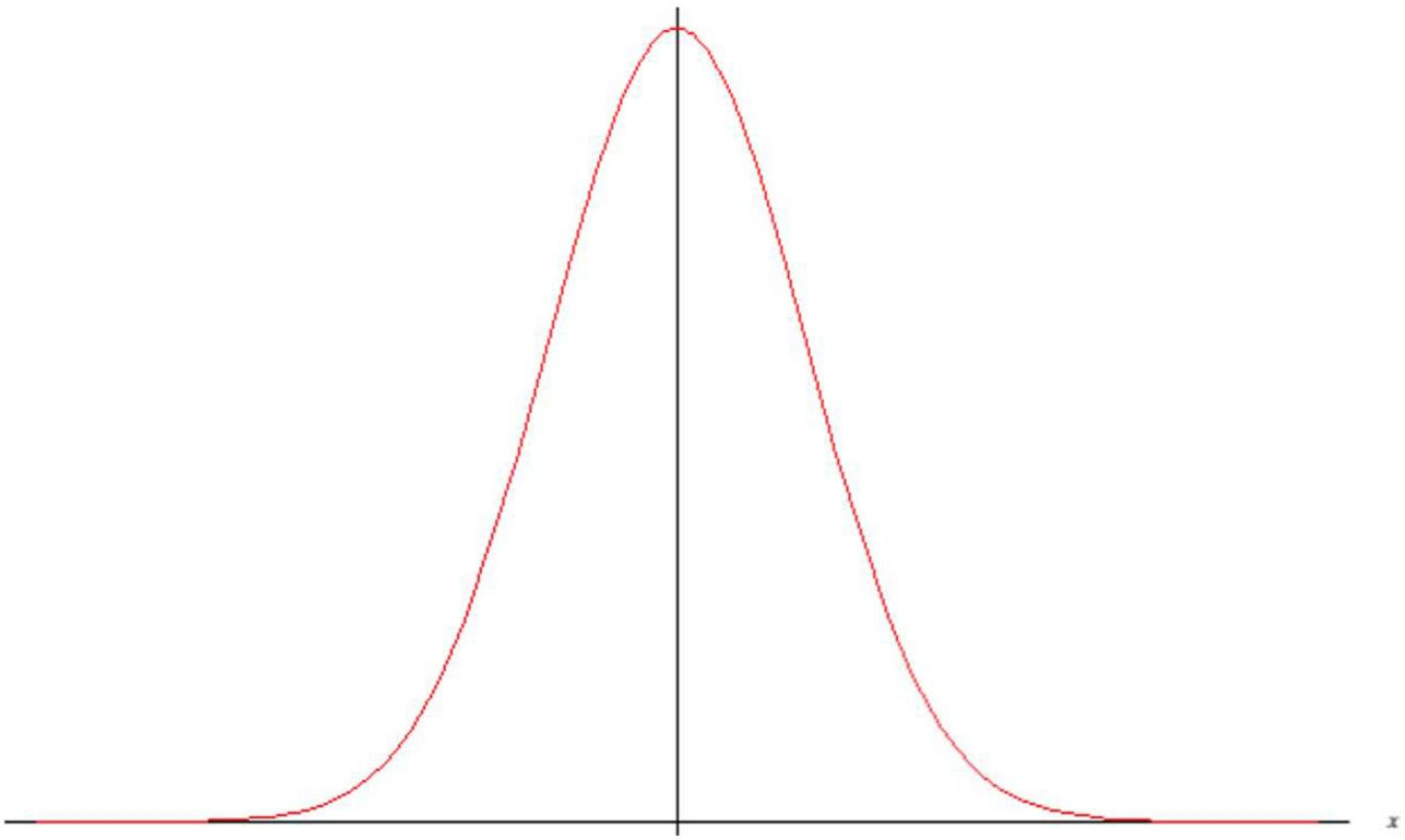
$$f(x) = \frac{e^{-.5\left((x-\mu_x)/\sigma_x\right)^2}}{\sigma_x\sqrt{2\pi}},$$

where $\sigma_x > 0, -\infty < \mu_x < \infty$, and $-\infty < x < \infty$.

- The normal probability density function has two "moments" or parameters: μ (mean, or expected value) and σ (standard deviation).

The Normal Distribution

$$f(x) = \frac{e^{-.5((x-\mu_x)/\sigma_x)^2}}{\sigma_x \sqrt{2\pi}}$$



The Standard Normal Distribution

- Next, we define the standard normal distribution. This involves transforming the normal random variable x into a *standard normal* random variable z, where $z = (x - \mu_x)/\sigma_x$.
- Note that $E(z) = (E(x) - \mu_x)/\sigma_x = 0$, since $E(x) = \mu_x$.
- Next, calculate σ_z^2;

$$\sigma_z^2 = E[(z - E(z))^2] = E(z^2)$$

$$= E((x - \mu_x)/\sigma_x)^2 = \frac{1}{\sigma_x^2} E((x - \mu_x)^2) = \frac{\sigma_x^2}{\sigma_x^2} = 1.$$

The Standard Normal Distribution

- Since $f(x) = \dfrac{e^{-.5((x-\mu_x)/\sigma_x)^2}}{\sigma_x\sqrt{2\pi}}$, $z = (x\text{-}\mu_z)/\sigma_z$

  $E(z) = \mu_z = 0$, and $\sigma_z = 1$, it follows that the standard normal probability density function for z, $f(z)$, is

$$f(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}},$$

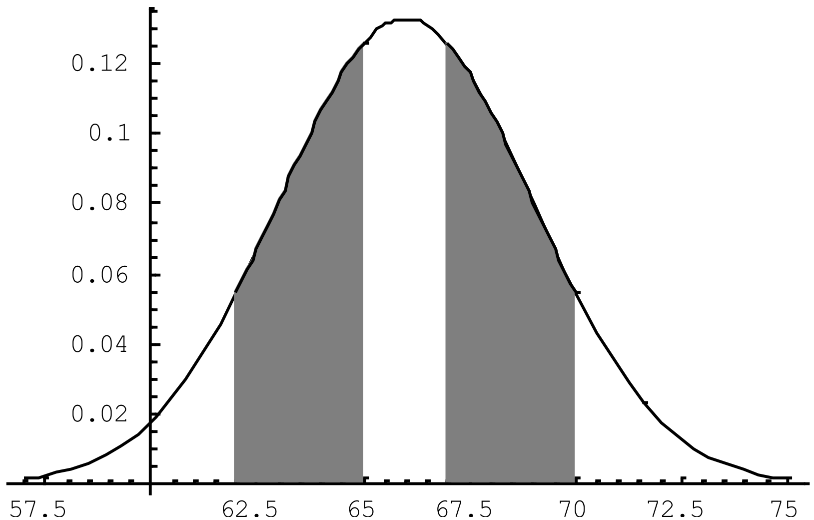
where $-\infty < z < \infty$ and $\int_{-\infty}^{\infty} f(z)dz = 1.0$.

The Standard Normal Distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916

Computing Probabilities using the z Table

- $\mu = 66$ and $\sigma = 3$. Compute $\Pr[62 \le x \le 65]$.
- Note that $\Pr[62 \le x \le 65] = \Pr[67 \le x \le 70]$:



- $\therefore \Pr[62 \le x \le 65] = \Pr[x \le 65] - \Pr[x \le 62]$
  $= \Pr[x \le 70] - \Pr[x \le 67]$.

Computing Probabilities using the z Table

- $\Pr\left[x \leq 70\right] = \Pr\left[z \leq \frac{70-66}{3}\right] = \Pr\left[z \leq 1.33\right]$

  $= .9082.$

- $\Pr\left[x \leq 67\right] = \Pr\left[z \leq \frac{67-66}{3}\right] = \Pr\left[z \leq .33\right]$

  $= .6293.$

- $\therefore \Pr\left[62 \leq x \leq 65\right] = \Pr\left[x \leq 70\right] - \Pr\left[x \leq 67\right]$

  $= .9082 - .6293 = .2789.$

Computing Probabilities using NORMSDIST

- Excel has the standard normal distribution function built in.
- Type "=NORMSDIST(*z*)" in a cell value, and this will return $\Pr\left[\frac{x-\mu_x}{\sigma_x} \le z\right]$:

Probability that z is less than or equal to -.33	0.3707
Probability that z is less than or equal to -1.33	0.0918
Probabillity that x is between 62 and 65	0.2789