Mathematical Expectation of a Random Variable

Idea: mean value of a random variable

- Definition: Weighted mean of the values of the random variable

$$E(X) = \sum_{x \in X} x P(X = x)$$

- Condition for the existence:

$$E(X) = \sum_{x \in X} |x| P(X = x) < \infty$$

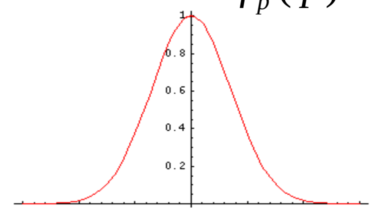
  – Note: Infinite sets can yield paradoxes

Idea: mean value of a random variable

- Definition: Weighted mean of the values of the random variable

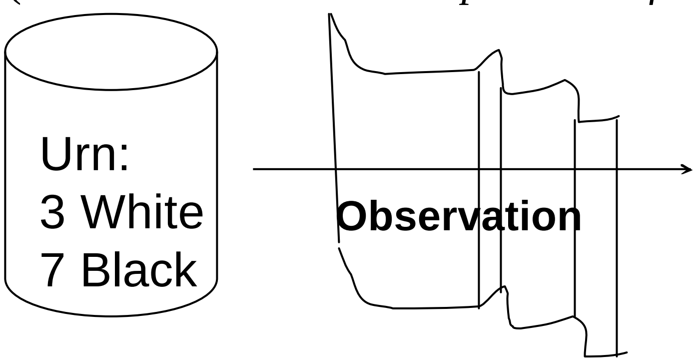
$$E(X) = \sum_{x \in X} x P(X = x)$$

$$f_{\hat{p}}(\hat{p})$$



$$p^*$$

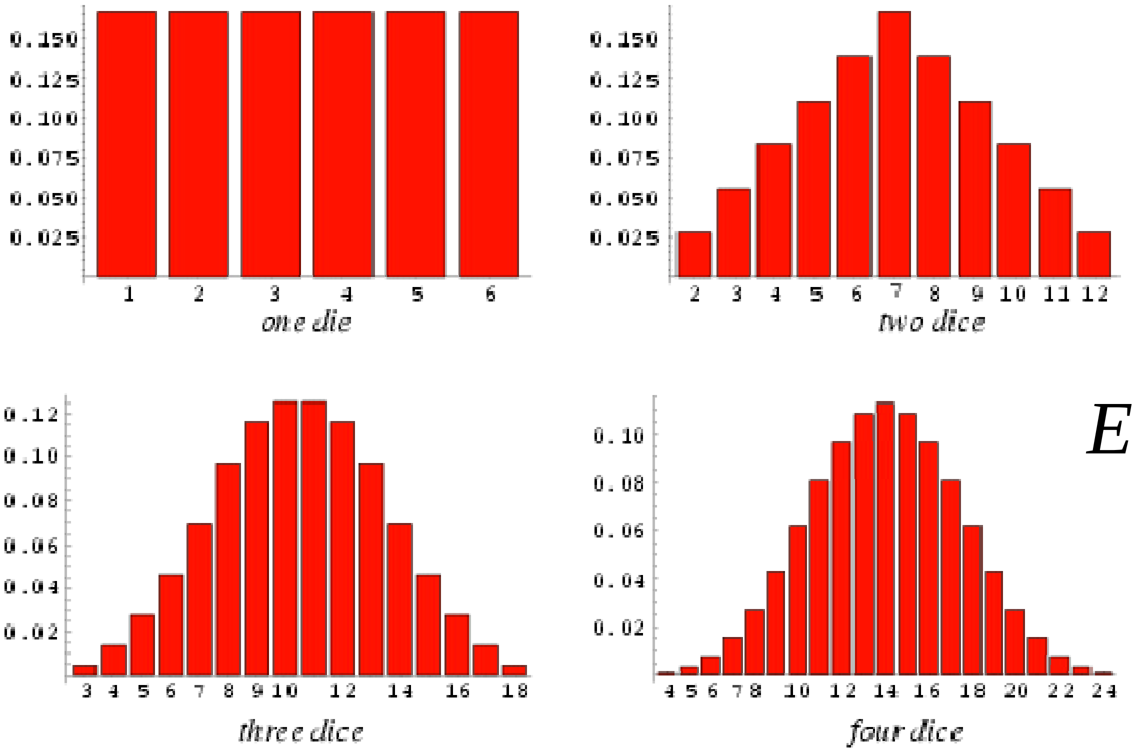
$$p^* = P(white\ observation / composition\ of\ the\ urn)$$



$$P(composition\ of\ the\ urn / white\ observation)$$

Sum of random variables

- Example: sum of points of *n* dice



http://mathworld.wolfram.com/Dice.html

$$E(X) = \sum_{x \in X} x P(X = x)$$

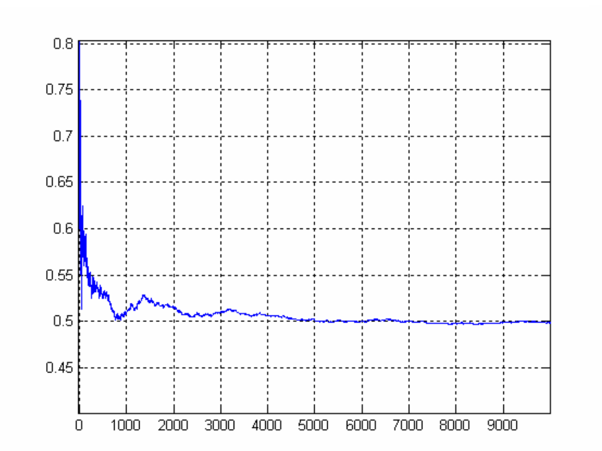
Mistakes of intuition

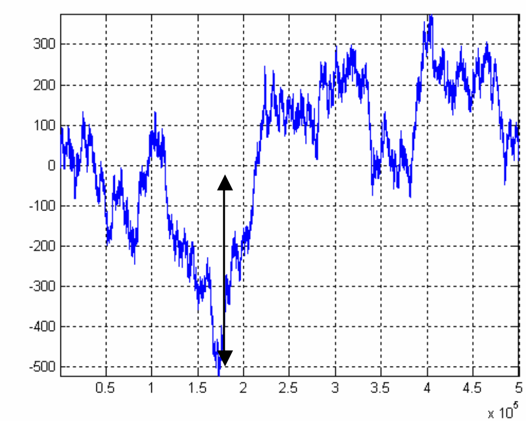
Intuition corresponds to ratio.

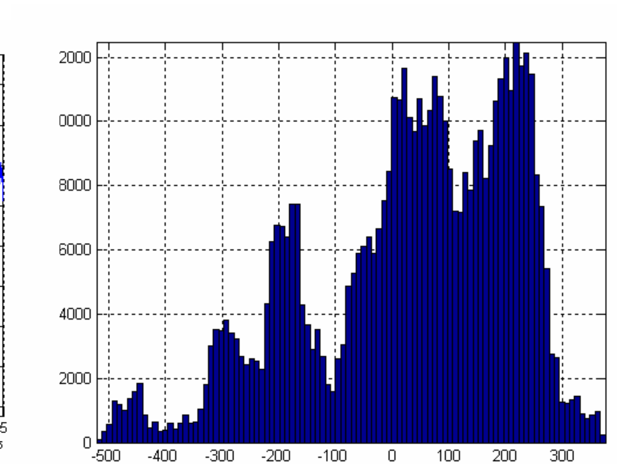
- Convergence on ratio.
- Difference gets as bigger !

$$\frac{favorable}{(favorable+unfavorable)} \rightarrow \frac{1}{2}$$

$$\left| favorable - unfavorable \right| \rightarrow \infty$$



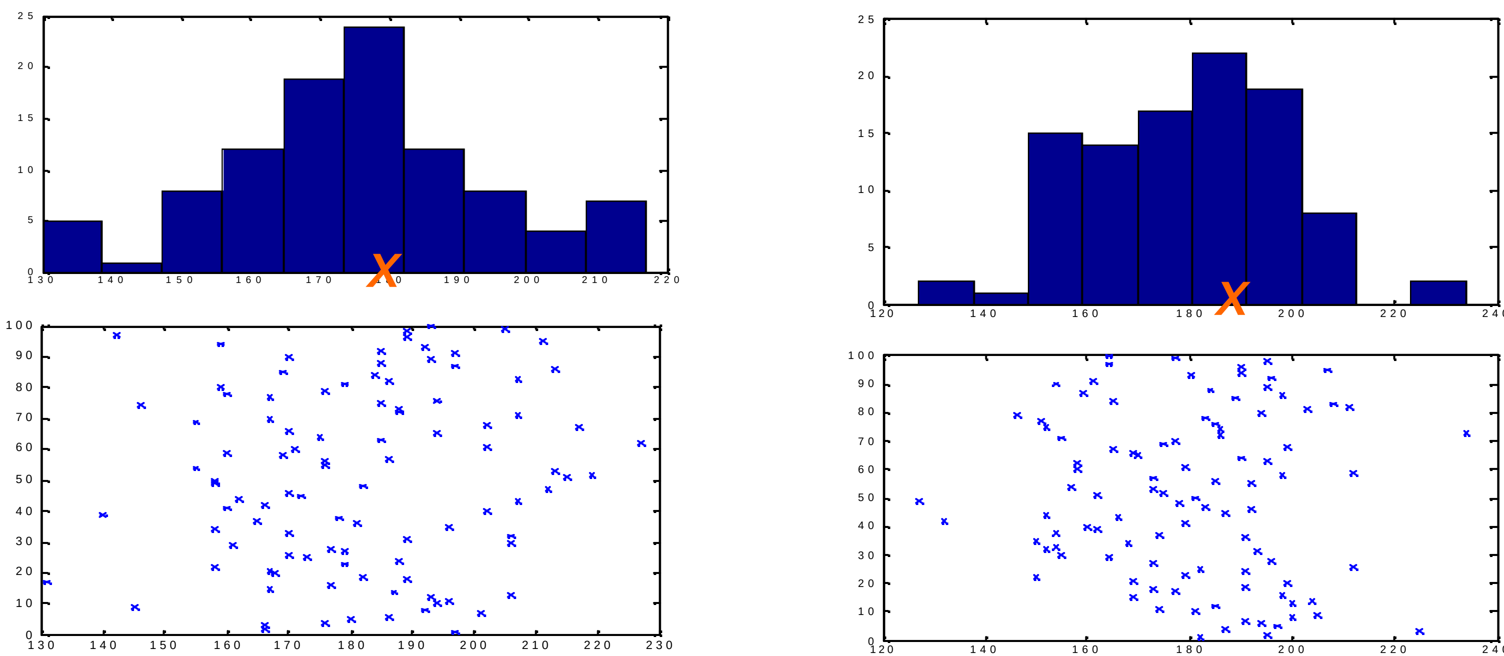




Idea: mean value of a random variable

- Note that it is ***probabilistic interpretation*** of the common sense concept of ***mean***.
- Two samples of the height of a population with a real mean of 180 cm.

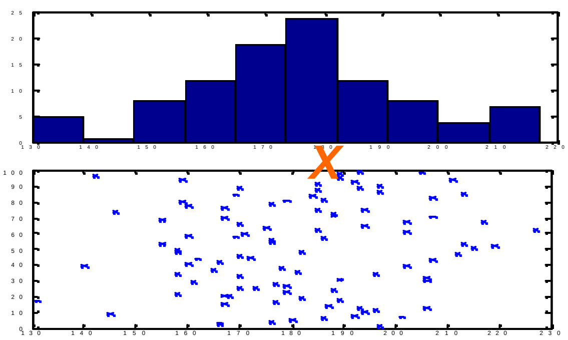
$$Mean(X) = \frac{1}{N}\sum_{x \in X} x$$

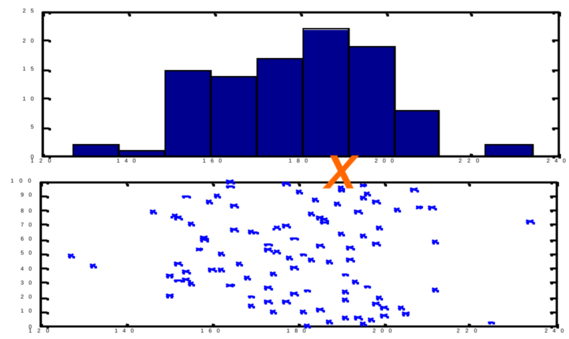


Idea: mean value of a random variable

- Note that it is ***probabilistic interpretation*** of the common sense concept of ***mean***.
- Note that if all possible samples are equally probable, they are equivalent.

$$E(X) = \sum_{x \in X} x P(X = x) \qquad P(X = x) = \frac{1}{N} \qquad Mean(X) = \frac{1}{N} \sum_{x \in X} x$$

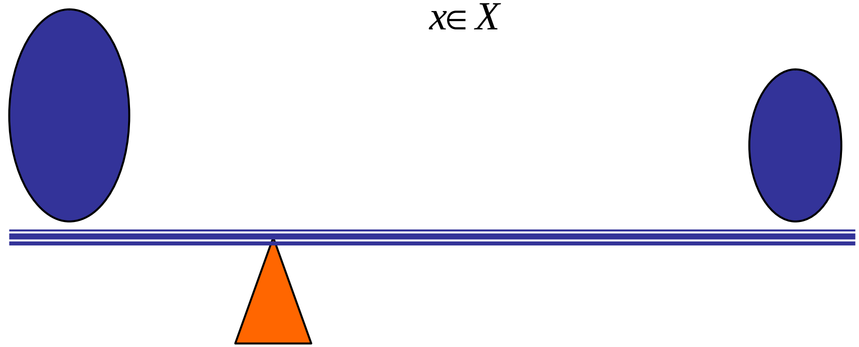




Idea: center of mass/equilibrium

- If the probability P(***X***=x) is interpreted as mass, and the random variable X as distance, the mathematical expectation is the center of mass of the object.

$$E(X) = \sum_{x \in X} xP(X = x)$$



Example: Roulette

- What is the payoff given by the casino?
- Roulette selects at random a number between 1 and 36 (0).
- Players can bet on 18,12,9,6,4,3,2,1 number.
  - A Bet over k numbers has a probability of success of k/36 and of getting a payment x from the casino or lossing a.
  - Expected value is

$$E(X) = \sum_{x \in \{Win, Loss\}} xP(X = x) = \frac{k}{36}x + \left(1 - \frac{k}{36}\right)(-a)$$

  - A fair game would imply

$$E(X) = 0$$

$$x = \left(\frac{36}{k} - 1\right)a$$

Example: Roulette

- What is the payoff given by the casino?
- The roulette has 37 slots (1 to 36+ the 0 slot)
  - In a Real game the casino the odds are
    - 35:1 17:2 11:3

$$E(X) = \sum_{x \in \{Win, Loss\}} xP(X = x) = \left(\frac{36}{k} - 1\right) a \frac{k}{37} + (-a)\left(1 - \frac{k}{37}\right) = -\frac{a}{37}$$

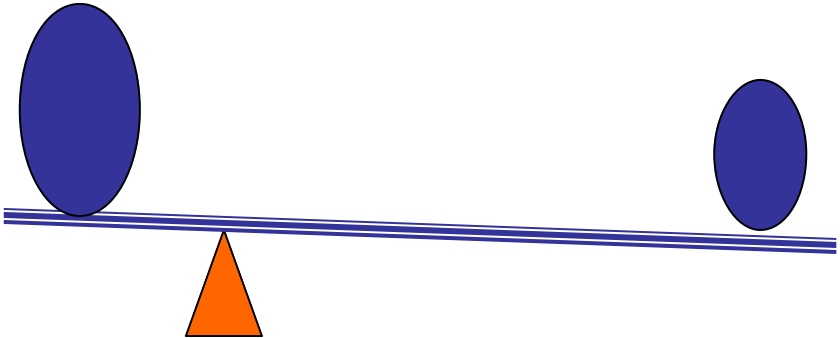
$$E(X) < 0$$

We need to model the variability

Example: Parking ticket

- Which is the expected value of not paying the parking ticket?
  - Parking ticket → 3 euros
  - Parking fine → 50 euros
  - Prob. of getting caught → 0.05

$$E(X) = \sum_{x \in \{Win, Loss\}} xP(X = x) = 3 * 0.95 - 50 * 0.05 = 0.35$$



Example: Parking ticket

- Which is the expected value of not paying the parking ticket?
- Important issues:
  - Subjective value: 3 euros vs. 150 euros

$$E(X) = \sum_{x \in \{Win, Loss\}} xP(X = x) = 3 * 0.95 - 50 * 0.05 = 0.35$$

  - Variability of the savings
  - Possible 'runs' of fines

Expected value of a geometric random variable

- Random Varible X={Number of trials until a success}

$$P(X=i)=p^{i-1}(1-p) \quad \text{for } \mathrm{i}=1,2,3,\cdots$$

$$E(X)=\sum_{n=0}^{\infty} np^{n-1}(1-p)=(1-p)\left(1+2p+3p^{2}+4p^{3}+\cdots\right)$$

- How do we sum this series?

Proof without Words: Differentiated Geometric Series

Roger B. Nelsen

Mathematics Magazine, Vol. 62, No. 5. (Dec., 1989), pp. 332-333.

Expected value of a geometric random variable

- How do we sum this series?

Proof without Words:
Differentiated geometric series

$$1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{4}\right) + 4\left(\frac{1}{8}\right) + \cdots = 4$$

$$1 = 4 - 3(1)$$

$$1 + 2\left(\frac{1}{2}\right) = 4 - 4\left(\frac{1}{2}\right)$$

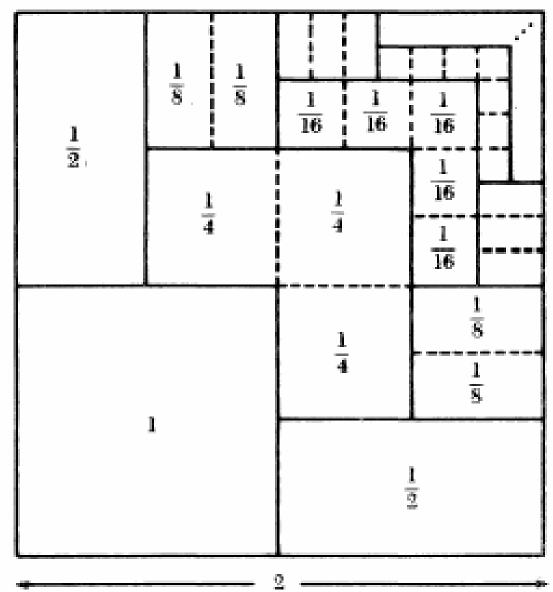
$$1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{4}\right) = 4 - 5\left(\frac{1}{4}\right)$$

$$1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{4}\right) + 4\left(\frac{1}{8}\right) = 4 - 6\left(\frac{1}{8}\right)$$

$$\vdots$$

$$1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{4}\right) + \cdots + n\left(\frac{1}{2}\right)^{n-1} = 4 - (n+2)\left(\frac{1}{2}\right)^{n-1}$$

$$1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{4}\right) + \cdots = 4$$



Expected value of a geometric random variable

- How do we sum this series?

$$1 + 2r + 3r^2 + 4r^3 + \cdots = \left(\frac{1}{1-r}\right)^2, \qquad 0 \le r < 1$$

$$1 = \left(\frac{1}{1-r}\right)^2 - r(2-r)\left(\frac{1}{1-r}\right)^2$$

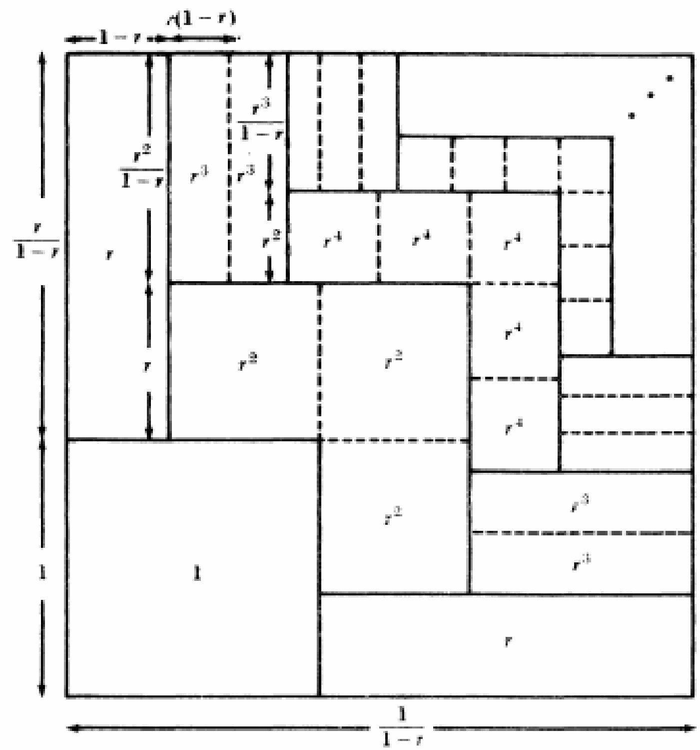
$$1 + 2r = \left(\frac{1}{1-r}\right)^2 - r^2(3-2r)\left(\frac{1}{1-r}\right)^2$$

$$1 + 2r + 3r^2 = \left(\frac{1}{1-r}\right)^2 - r^3(4-3r)\left(\frac{1}{1-r}\right)^2$$

$$\vdots$$

$$1 + 2r + 3r^2 + \cdots + nr^{n-1} = \left(\frac{1}{1-r}\right)^2 - r^n(n+1-nr)\left(\frac{1}{1-r}\right)^2$$

$$1 + 2r + 3r^2 + \cdots = \left(\frac{1}{1-r}\right)^2$$



—Roger B. Nelsen
Lewis and Clark College

Expected value of a geometric random variable

- How do we sum this series?

$$E(X)=\sum_{n=0}^{\infty} np^{n-1}(1-p)=(1-p)\left(1+2p+3p^2+4p^3+\cdots\right)$$

$$pE(X)=\sum_{n=1}^{\infty} np^{n}(1-p)=(1-p)\left(p+2p^2+3p^3+4p^4+\cdots\right)$$

$$E(X)-pE(X)=(1-p)\left(1+p+p^2+p^3+p^4+\cdots\right)=$$

$$=(1-p)\sum_{k=0}^{\infty} p^k \underset{\substack{\text{Geometric}\\\text{Series}}}{=} (1-p)\frac{1}{(1-p)}=1$$

$$E(X)=\frac{1}{(1-p)}$$

Expected value of a geometric random variable

- How do we sum this series? *Another way*

$$S(p) = \sum_{k=1}^{\infty} p^k \underset{\substack{\text{Geometric}\\ \text{Series}}}{=} \frac{p}{(1-p)}$$

$$\frac{d}{dp} S(p) = \sum_{k=1}^{\infty} k p^{k-1} = \frac{d}{dp}\left(\frac{p}{(1-p)} \right) = \frac{1}{(1-p)^2}$$

$$E(X) = (1-p)\frac{1}{(1-p)^2}$$

$$\frac{d}{dx}\left(\frac{a(x)}{b(x)} \right) = \frac{b(x)\frac{d}{dx}a(x) - a(x)\frac{d}{dx}b(x)}{b(x)^2}$$

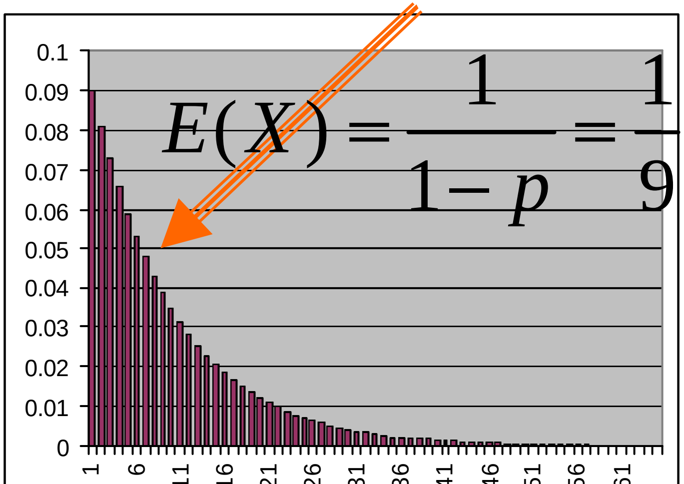
Expected value of a geometric random variable

- Random Varible X={Number of trials until a success}

$$P(X = i) = p^{i-1}(1-p) \quad \text{for } \mathrm{i} = 1,2,3,\cdots$$

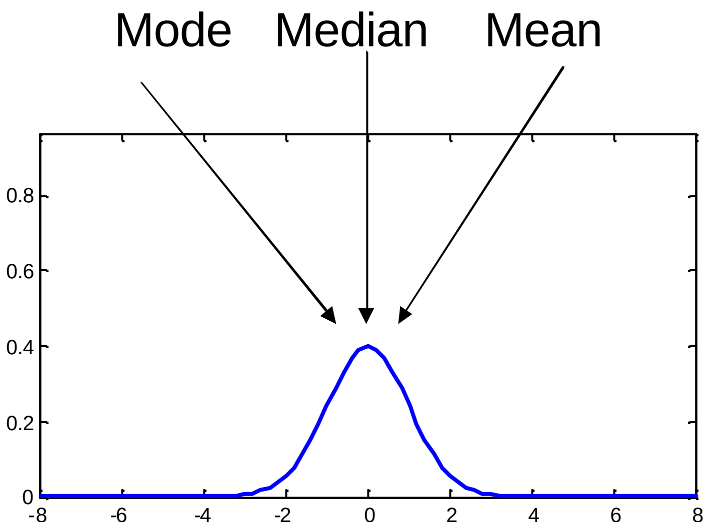
$$E(X) = \sum_{n=0}^{\infty} np^{n-1}(1-p) = (1-p)\left(1 + 2p + 3p^2 + 4p^3 + \cdots\right)$$

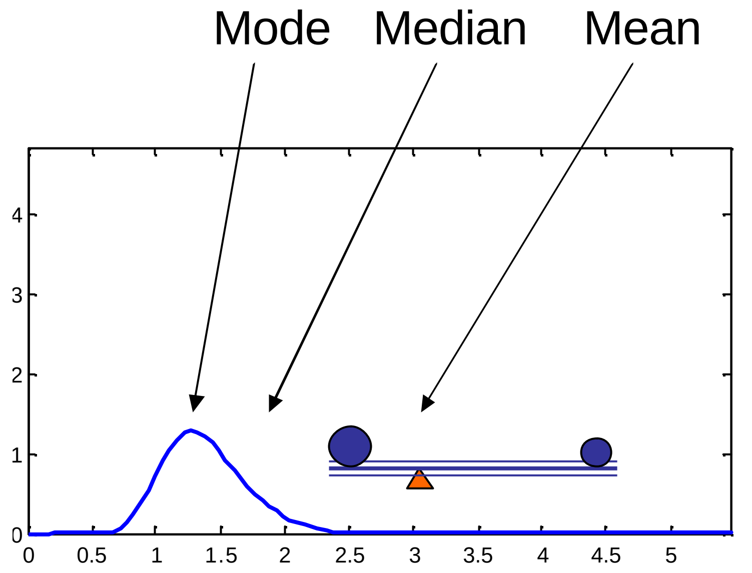
$$E(X) = \frac{1}{1-p}$$



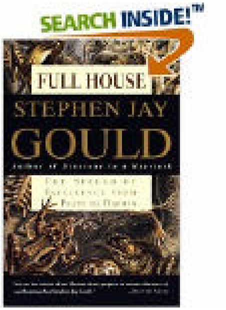
Relation of expectaition and other statistical measures

- Skew distribution vs. symetric distribution





Full House: The Spread of Excellence from Plato to Darwin by Stephen Jay Gould



Property of lineality

- The expectation of a linear combination of random variables is the linear combination of expectations.

$$E(aX + bY) = aE(X) + bE(Y)$$

$$E(aX + bY) = \sum_{w \in \Omega} \left(aX(w) + bY(w)\right)P(w)$$

$$= a\sum_{w \in \Omega} X(w)P(w) + b\sum_{w \in \Omega} Y(w)P(w) =$$

$$= aE(X) + bE(Y)$$

Finding the maximum

- Suppose that n children of differing heights are placed in line at random. You select the first child from the line and walk with her/him along the line until you encounter a child who is taller or until you have reached the end of the line. If you do encounter a taller child, you repeat the process.
- What is the expected value of the number of children selected from the line?

Taken fromTijms, Understanding probability

Finding the maximum

- We define the variable X as the number of children selected from the line.
- We will define the indicator variable:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-}th \text{ child is selected from the line} \\ 0 & \text{otherwise} \end{cases}$$

- Now the number of selected children will be:

$$X = X_1 + X_2 + \cdots + X_n$$

Finding the maximum

- The probability that the ith child is the tallest among the first i children is 1/i
- Therefore:

$$E(X_i) = 0 \cdot \left(1 - \frac{1}{i}\right) + 1 \cdot \frac{1}{i} = \frac{1}{i} \qquad \text{for i} = 1, 2, \cdots \text{n}$$

$$E(X) = E(X_1 + X_2 + \cdots + X_n)$$

$$E(X) = 1 + \frac{1}{2} + \frac{1}{3} \cdots + \frac{1}{n} \approx \ln(n) + \frac{1}{2n} + 0.57722$$

Expected number of distinct birthdays

- What is the expected number of distinct birthdays within a randomly formed group of 100 persons.
  - We define the random variable

$$X_i = \begin{cases} 1 & \text{if the birthday is on day } i \\ 0 & \text{otherwise} \end{cases}$$

  - The number of birthdays is $X = X_1 + X_2 + \cdots + X_3$

Taken fromTijms, Understanding probability

Expected number of distinct birthdays

- What is the expected number of distinct birthdays within a randomly formed group of 100 persons.
  - For each day we have:

$$P(X_i = 0) = \left(\frac{364}{365}\right)^{100}$$

$$P(X_i = 1) = 1 - P(X_i = 0)$$

  - The expected number of distinct birthdays is

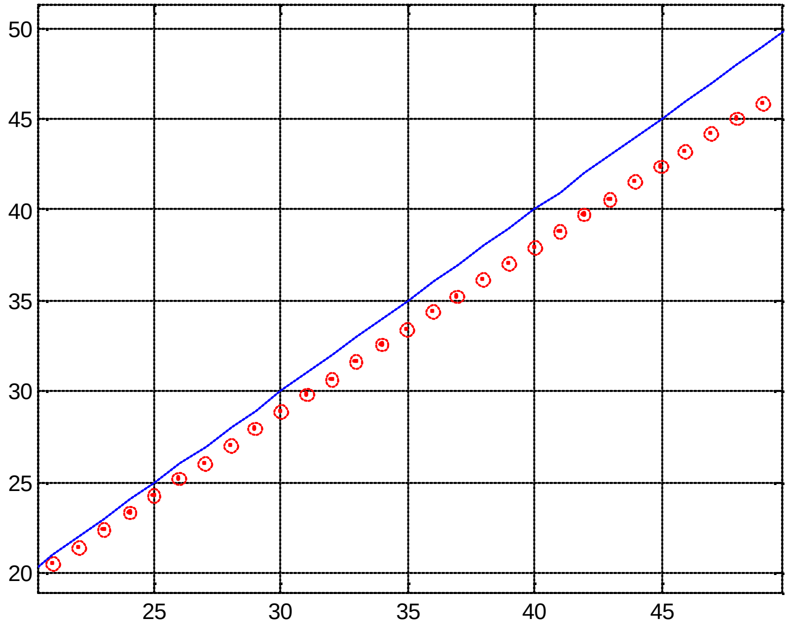
$$E(X) = E(X_1 + X_2 + \cdots + X_{365}) = 365\left(1 - \left(\frac{364}{365}\right)^{100}\right) = 87.6$$

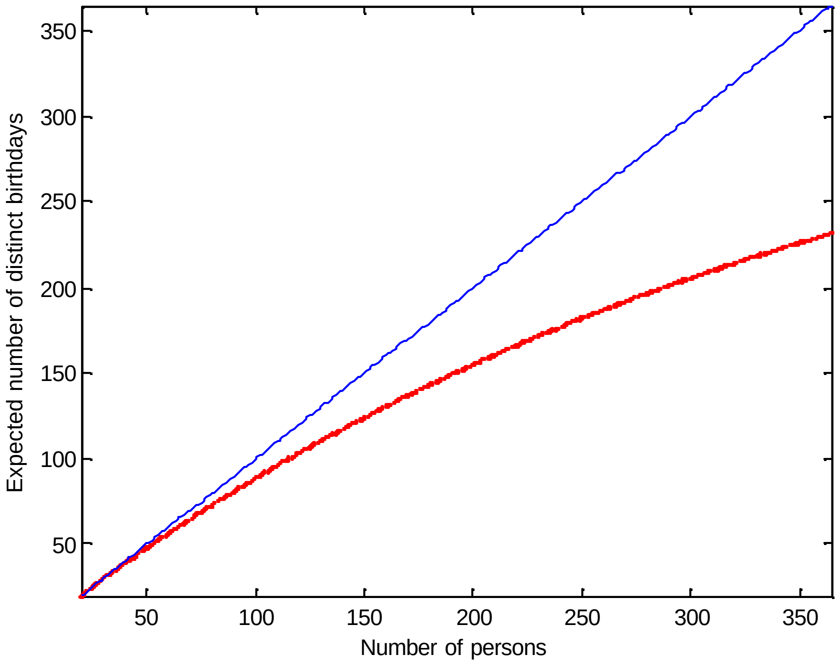
Taken fromTijms, Understanding probability

Expected number of distinct birthdays

- For an arbitrary number of persons.

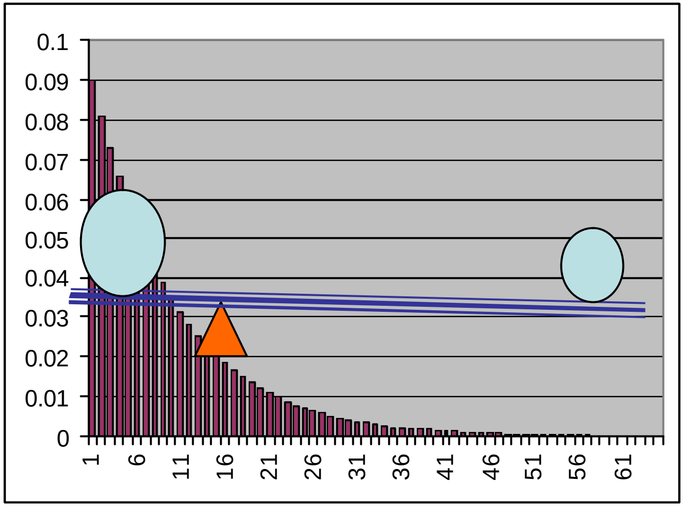
$$E(X) = 365\left(1 - \left(\frac{364}{365}\right)^{n}\right)$$





Other Properties

- If X is a non negative random variable, then $E(X) \geq 0$



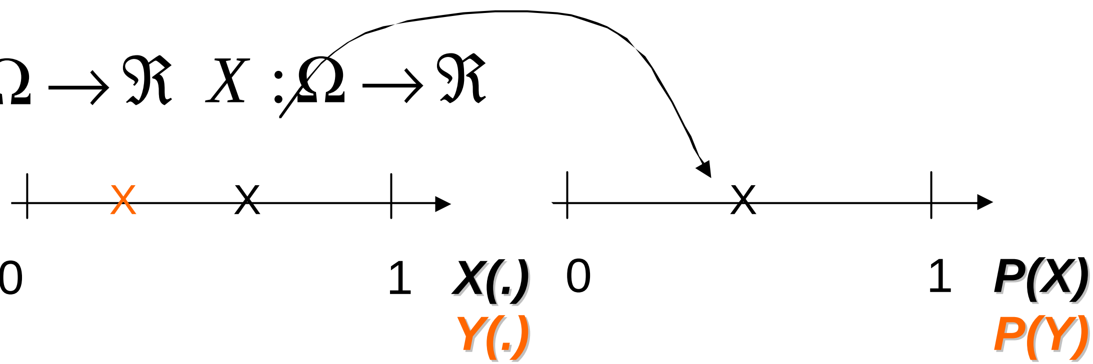
- Also $X \geq Y$ then $E(X) \geq E(Y)$

Other Properties

- What do we mean by

$$X \geq Y \quad \text{then} \quad E(X) \geq E(Y) \quad ?$$

$$Y : \Omega \rightarrow \Re \quad X : \Omega \rightarrow \Re$$



$$E(X) = \sum_{x \in X} x P(X = x)$$

Caveats of intuition

- Does the expectation always exist?

$$Mean(X) = \frac{1}{N}\sum_{x \in X} x$$

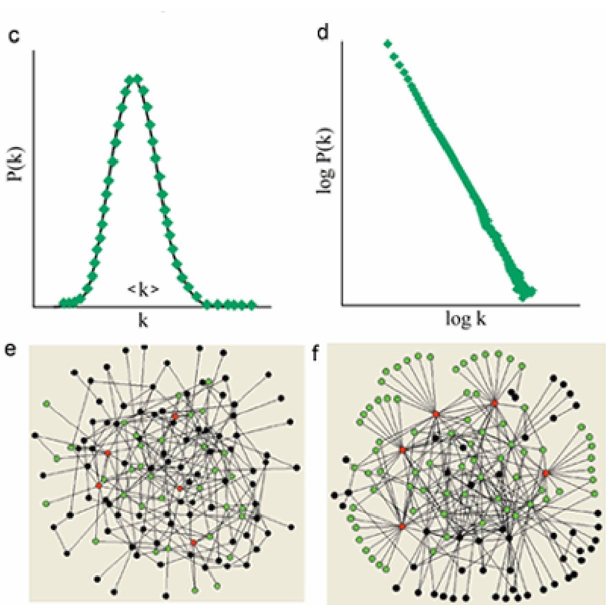
Sample mean always exists

$$E(X) = \sum_{x \in X} xP(X = x)$$

Expectation perhaps gives an infinite value !!!!

$$P(X = x) = \frac{l^{x}}{x!}e^{-l}$$

$$P(X = x) = \frac{1}{x^{a}}$$

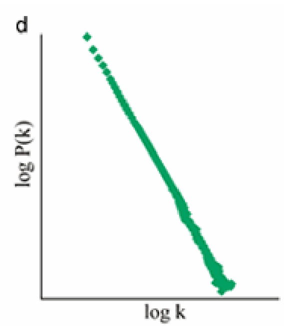


http://physicsweb.org/articles/world/14/7/9/1#pw1407091
The physics of the Web

Caveats of intuition

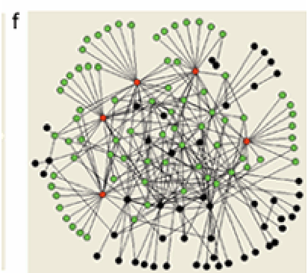
- Does the expectation always exist?
- Example: A Cauchy random variable

$$Mean(X) = \frac{1}{N}\sum_{x\in X} x$$



$$E(X) = \sum_{x\in X} xP(X = x)$$

$$P(X = x) = \frac{1}{p\left(1 + x^2\right)} \qquad \left(-\infty < x < +\infty\right)$$



$$P(X = x) = \frac{1}{x^a} \qquad E(X) \to \infty$$

http://physicsweb.org/articles/world/14/7/9/1#pw1407091
The physics of the Web

Caveats of intuition

- Why is infinite?
- Divergent series $E(X)=\sum_{x\in X} xP(X=x)=\sum_{x=1}^{\infty}\frac{x}{p\left(1+x^2\right)}$
- Note that for high values of x

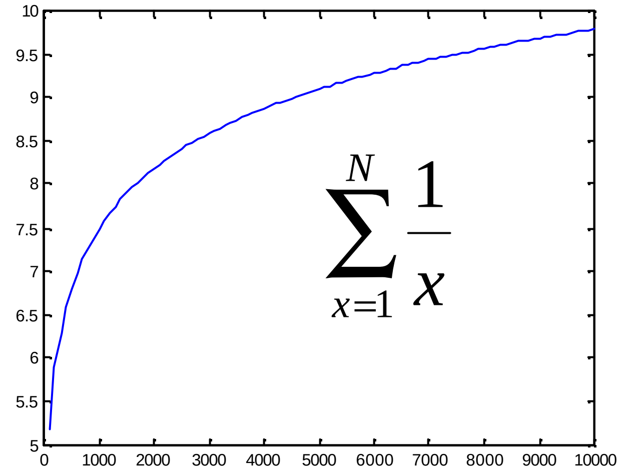
$$E(X)=\sum_{x=1}^{\infty}\frac{x}{p\left(1+x^2\right)}\cong\sum_{x=1}^{x^2>>1}\frac{x}{p\left(1+x^2\right)}+\sum_{x^2>>1}^{\infty}\frac{1}{p}\frac{1}{x}$$

$$E(X)\propto\sum_{x=1}^{\infty}\frac{1}{x}$$

$$P(X=x)=\frac{1}{x^a}$$

Caveats of intuition

- Why is infinite?
- Value of the harmonic series



$$\sum_{x=1}^{\infty} \frac{1}{x} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} \cdots$$

REPASSAR

$$\sum_{x=1}^{\infty} \frac{1}{x} > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} \cdots$$

$$\sum_{x=1}^{\infty} \frac{1}{x} > 1 + 1 + 1 + 1 \cdots = \infty$$

Expected waiting times

- Geometric
- Pareto
- gausian

Expected value of a Binomial