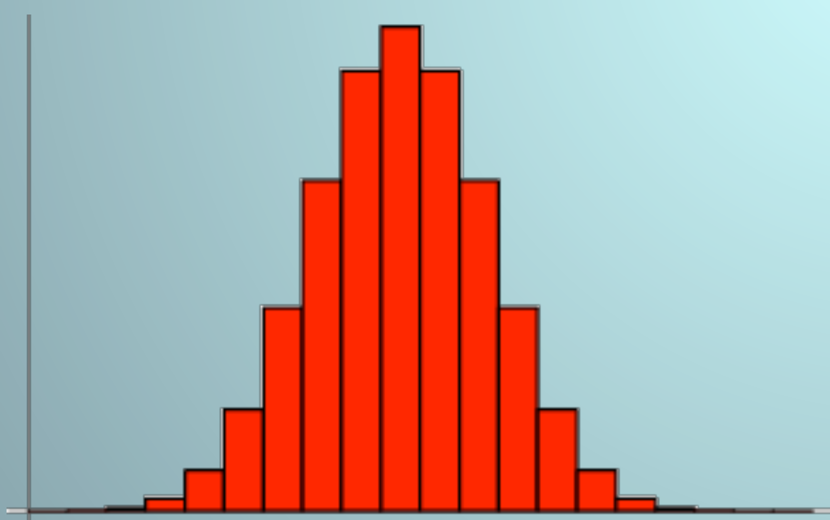
Introduction to Random Variables

Godfried Toussaint

Random Variables and Probability Distributions

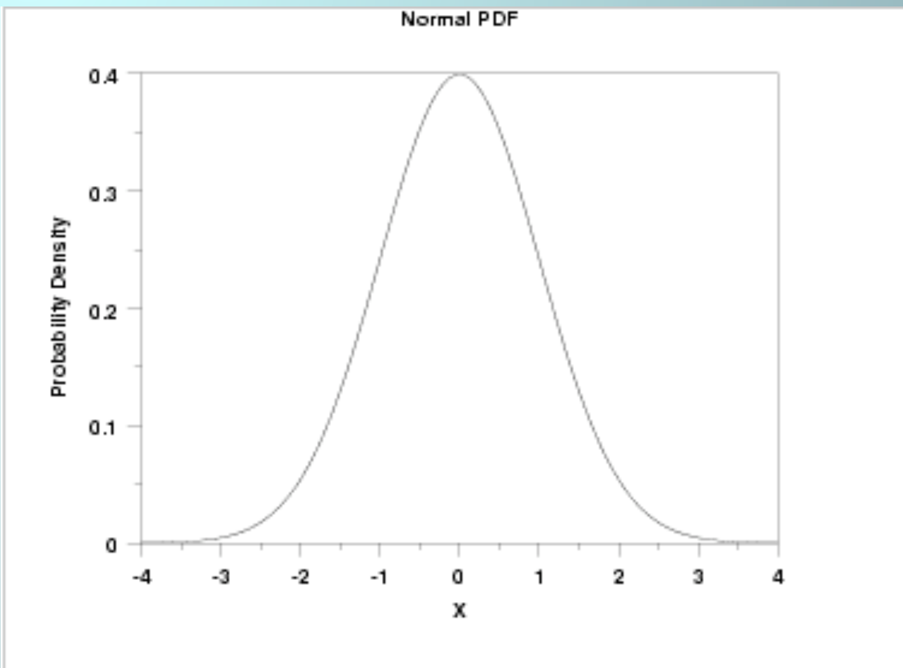
- A *random variable X* is a function of a subset of the outcomes of a set of random experiments. As such a random variable may also be a function of other random variables. Typically it associates a number with each outcome of the sample space (S, P).
- The *probability distribution* of a random variable is the collection of all the possible values that it can take, along with their probabilities:

Discrete case
Binomial distribution



Quincunx demo

Continuous case
Probability density function




Random Variables: Examples

- Let (S, P) be the sample space of rolling a pair of dice. Let X be the sum of the two numbers obtained when the dice are rolled. X is a random variable:

  X[(1, 2)] = 3   X[(5, 3)] = 8   X[(6, 6)] = 12 etc.

- Let (S, P) be the sample space of tossing 10 coins. Let X_H be the random variable #Heads, and X_T be the #Tails. Let X be the random variable $X_H - X_T$.

  X_H(HTHTHTTHTHTT) = 5 H_T(HTHTHTTHTHTT) = 7

  $X = X_H - X_T = -2$

Bernoulli Trials: Tossing an Unfair Coin

- Let (S, P) be the sample space of tossing an unfair coin 5 times. Assume the tosses are independent.
- S = {HEADS, TAILS} and P(HEADS) = p and P(TAILS) = 1-p

- The probability of observing some sequence such as:

HEADS, TAILS, HEADS, TAILS, HEADS is:

$$
\begin{aligned}
P(\text{HTHTH}) &= P(\text{H})P(\text{T})P(\text{H})P(\text{T})P(\text{H}) \\
&= p \cdot (1-p) \cdot p \cdot (1-p) \cdot p \\
&= p^3(1-p)^2
\end{aligned}
$$

Binomial Random Variables

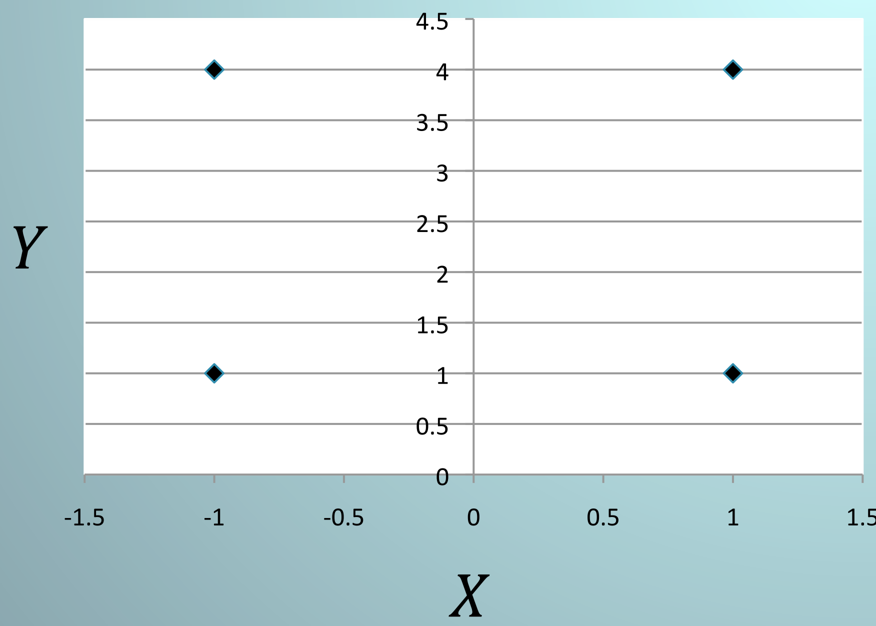
- Let (S, P) be the sample space of tossing an unfair coin n times. Assume the tosses are independent.
- Let X denote the random variable that counts the number of heads, and h be an integer $0 \le h \le n$. What is the probability $P(X = h)$?

- There are exactly $\binom{n}{h}$ sequences of n flips with exactly h heads.
- Therefore $P(X = h) = \binom{n}{h} p^h (1-p)^{n-h}$
- X is called a binomial random variable because it is a term in the expansion of the binomial $(p + q)^n$ where $q = 1 - p$.

Independent Random Variables

- Let (S, P) be a sample space and let X and Y be two random variables defined on (S, P). Then X and Y are *independent* provided that for all values a and b we have:

$$P(X = a \text{ and } Y = b) = P(X = a)P(Y = b)$$

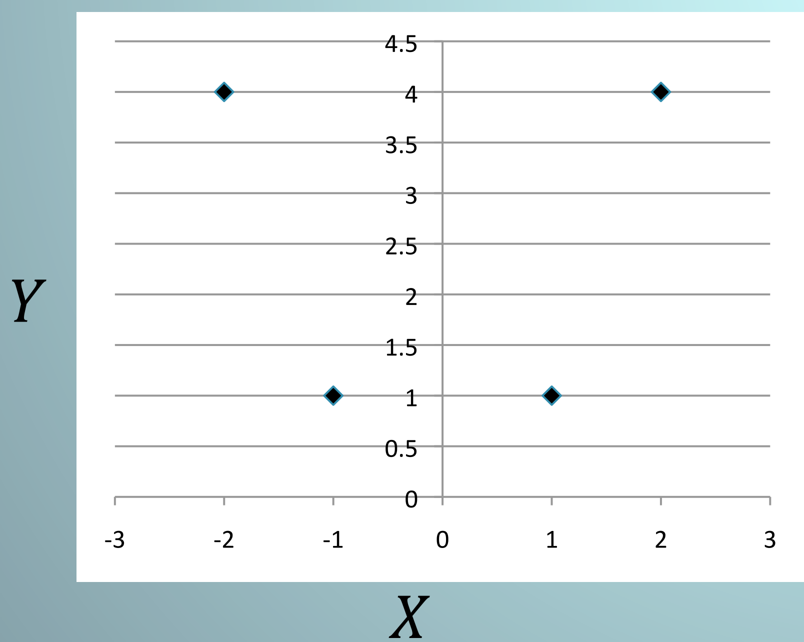


		X		
		−1	1	
Y	1	0.25	0.25	0.50
	4	0.25	0.25	0.50
		0.50	0.50	1.00

Dependent Random Variables

- Let (S, P) be a sample space and let *X* and *Y* be two random variables defined on (S, P). Then *X* and *Y* are *dependent* provided that for some values *a* and *b* we have:

$$P(X = a \text{ and } Y = b) \neq P(X = a)P(Y = b)$$



	X: −2	−1	1	2	
Y: 1	0.00	0.25	0.25	0.00	0.50
4	0.25	0.00	0.00	0.25	0.50
	0.25	0.25	0.25	0.25	1.00

Expectation (mean, average, μ)

- Let X be a real-valued random variable defined on a sample space (S, P). The expectation (or the expected value) of X is:

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

- Example: Let X denote the number rolled by a die. What is the expected value of X?

Expectation (mean, average, μ)

- Let X be a real-valued random variable defined on a sample space (S, P). The expectation (or the expected value) of X is:

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

- Example: Let X denote the number rolled by a die. What is the expected value of

$$E(X) = \sum_{a=1}^{6} X(a)P(a)$$

= X(1)P(1) + X(2)P(2) + X(3)P(3) + X(4)P(4) +X(5)P(5) +X(6)P(6)

= 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6)

= 21/6

= 7/2

= 3.5

Linearity of Expectation

- Let X and Y be real-valued random variables defined on a sample space (S, P). Then

$$E(X+Y) = E(X) + E(Y)$$

Proof: Let Z = X + Y.

$$E(Z) = \sum_{s \in S} Z(s)P(s)$$

$$E(Z) = \sum_{s \in S} [X(s) + Y(s)]P(s)$$

$$E(Z) = \sum_{s \in S} [X(s)P(s) + Y(s)P(s)]$$

$$E(Z) = \sum_{s \in S} X(s)P(s) + \sum_{s \in S} Y(s)P(s)$$

$$E(Z) = E(X) + E(Y)$$

Linearity of Expectation - continued

- Let X and Y be real-valued random variables defined on a sample space (S, P). Let a , b, and c be constants. Then

    1. E(cX) = cE(X)

    2. E(aX+bY) = aE(X) + bE(Y)

    3. and

$$E(c_1X_1 + c_2X_2 + \ldots + c_nX_n) = c_1E(X_1) + c_2E(X_2) + \ldots + c_nE(X_n)$$

Product of Random Variables

- Let (S, P) be the sample space of tossing a fair coin twice. Let X_H be the number of HEADS and X_T the number of TAILS. Let Z = $X_H X_T$.

X_1	X_2	X_H	X_T	Z
T	T	0	2	0
H	T	1	1	1
T	H	1	1	1
H	H	2	0	0

- Then $E(X_H)$ = $E(X_T)$ = 1
- And $E(Z) = \sum_{a \in S} aP(Z = a)$

$$= 0 \bullet P(Z = 0) + 1 \bullet P(Z = 1)$$

$$= 0 \bullet \left(\frac{2}{4}\right) + 1 \bullet \left(\frac{2}{4}\right)$$

$$= \frac{1}{2}$$

- Therefore sometimes $E(X_H X_T) \neq E(X_H)E(X_T)$.

Product of Random Variables - continued

- Let X and Y be independent, real-valued random variables defined on a sample space (S, P).
- Then $E(XY) = E(X)E(Y)$.

Proof: homework.

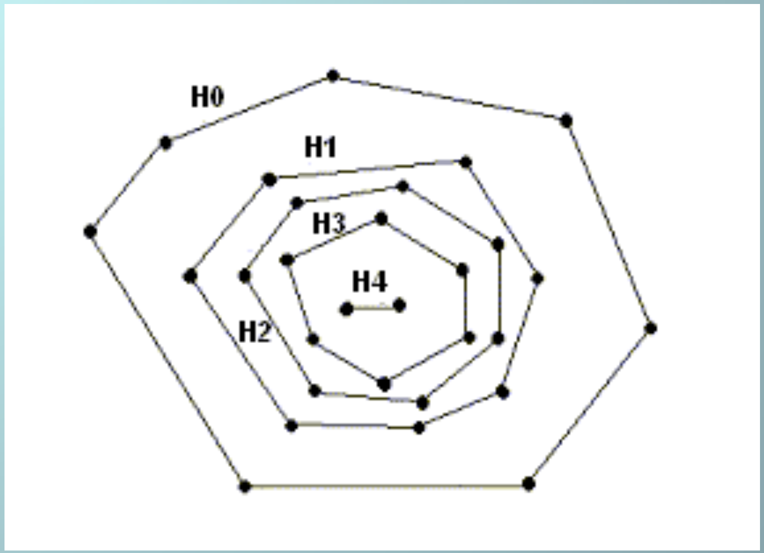
Measures of Central Tendency and Estimators of Location

- The mean – average – expected value
  - Arithmetic mean: $(a^1 + a^2 + \ldots + a^n)/n$
  - Geometric mean: $(a^1 a^2 \cdots a^n)^{1/n}$
- The median – a robust estimator
  - Deleting outliers
- Higher dimensional analogues?

Measures of Central Tendency and Estimators of Location

- The mean – average – expected value
  - Arithmetic mean: $(a^1 + a^2 + \ldots + a^n)/n$
  - Geometric mean: $(a^1 a^2 \cdots a^n)^{1/n}$
- The median – a robust estimator
  - Deleting outliers
- Higher dimensional analogues
  - Mean of each dimension
  - Onion peeling

Onion Peeling



The Arithmetic Mean – Geometric Mean Inequality: Induction Proof

The Arithmetic-Geometric mean inequality: if $a_1, a_2, \ldots, a_n > 0$,

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq \sqrt[n]{a_1 a_2 \cdots a_n}$$

where the equality holds if, and only if, all the a_i's are equal.

Base Case: For $n = 2$ the problem is equivalent to

$(a_1 + a_2)^2 \geq 4a_1a_2$, which is equivalent to $(a_1 - a_2)^2 \geq 0$.

Or alternately expand: $\left(\sqrt{a_1} - \sqrt{a_2}\right)^2$

*Kong-Ming Chong, "*The Arithmetic Mean-Geometric Mean Inequality: A New Proof," *Mathematics Magazine*, Vol. 49, No. 2 (Mar., 1976), pp. 87-88.

The Arithmetic Mean – Geometric Mean Inequality: Induction Proof – *continued…*

Induction Hypothesis: Assume the statement is true for *n*-1.

Proof: Without lost of generality assume that

$$a_1 \leq a_2 \leq \cdots \leq a_n.$$

Let G be the geometric mean $G := \sqrt[n]{a_1 a_2 \cdots a_n}$. Then it follows that

$a_1 \leq G \leq a_n$. Note that since

$$a_1 + a_n \geq \frac{a_1 a_n}{G} + G$$

$$a_1 + a_n - G - \frac{a_1 a_n}{G} = \frac{a_1}{G}(G - a_n) + (a_n - G) = \frac{1}{G}(G - a_1)(a_n - G) \geq 0$$

The AG-Mean Inequality Induction Proof – *continued…*

By the induction hypothesis

$$\frac{a_2 + \cdots + a_{n-1} + \frac{a_1 a_n}{G}}{n-1} \geq \sqrt[n-1]{G^n/G} = G$$

Hence

$$a_2 + \cdots + a_{n-1} + \frac{a_1 a_n}{G} \geq (n-1)G$$

and

$$\frac{a_2 + \cdots + a_{n-1} + \frac{a_1 a_n}{G} + G}{n} \geq G$$

But since

$$a_1 + a_n \geq \frac{a_1 a_n}{G} + G$$

it follows that

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq G$$

Measures of Variability or Spread: Variance

- Let X be a real-valued random variable defined on a sample space (S, P). Let μ = E(X). Then the variance of X is:

$$\begin{aligned} \mathrm{Var}(X) &= E[(X - \mu)^2] \\ &= E[(X - \mu)(X - \mu)] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - E[2\mu X] + E[\mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

- The standard deviation σ is the square root of the variance.
- The range of X: maximum value of X – minimum value of X.