Random Experiments and Random Variables

Sorana D. Bolboacă

OBJECTIVES

- Random experiment
- Random variables

Binomial Random Processes

- Two possible outcomes
- Heads or tails
- Make basket or miss basket
- Fatality, no fatality

- With probability p (or 1-p)

Binomial Random Processes

- Predicting a specific versus a general pattern
- Which lotto ticket would you buy?
  - Equally likely (or unlikely) to win:
    - 26 45 8 72 91
    - 26 26 26 26 26 – less likely to be bought
  - Each specific ticket is equally (un)likely to win
  - A ticket that "looks like" ticket A (with alternating values) is more likely than one that "looks like" ticket B (with identical values).

Binomial Random Processes

- Probabilities for specific patterns get smaller as you run more tests
- What is the probability of getting heads on the second test and the tails on all other trials?
  - P(T,H) = 0.25
  - P(T, H, T) = 0.125
  - P(T, H, T, T) = 0.0625

Binomial Random Processes

- What is the probability of getting at least one heads when you toss a coin multiple times?
  - Two tosses: Pr(HT or TH or HH) = 0.75
  - Three tosses: Pr(HTT or THT or TTH or THH or HHT or HHH) = 0.875
  - Four tosses: 0.9375

DEFINITION

- Let X be a quantitative variable measured or observed from an experiment
- The value of X is a random variable

- **Example**:
  - Number of red cells in blood
  - Number of bacteria from the students hands
  - Average of depression score obtained from the application of a test on a sample of patients with malign tumors

Random Variables

- Arithmetic mean
- Standard deviation
- Proportion
- Frequency
  - All are random variables

Types of Random Variables

Discrete:

- Can take a finite number of values
  - The number of peoples with RH- from a sample
  - The number of children with flue from a collectivity
  - The number of anorexic students from university
  - Pulse

Continuous:

- Can take an infinite number of values into a defined range
- Vary continuously in defined range
  - Body temperature
  - Blood sugar concentration
  - Blood pressure

Types of Random Variables

- Generally, means are continuous random variables and frequencies are discrete random variables
- Examples:
  - The mean of lung capacity of a people who work in coal mine
  - The number of patients with chronic B hepatitis hospitalized in Cluj-Napoca between 01/11-05/11/2008.

Random Variable

- A random variable
  - is a quantification of a probability model that allows to model random data
  - Is a quantity that may take any value of a given range that cannot be predicted exactly but can be described in terms of their probability
    - Discrete: generally assessed by counting
    - Continuous: generally assessed by measurements

RANDOM VARIABLE

- When throwing a die with 6 faces, let X be the random variable defined by:

X = the square of the scores shown on the die

What is the expectation of X?

Solution:

- $S = \{1, 2^2, 3^2, 4^2, 5^2, 6^2\} = \{1, 4, 9, 16, 25, 36\}$
- Each face has a probability of 1/6 of occurring, so:

E(X) = 1*1/6+4*1/6+6*1/6+16*1/6+25*1/6+36*1/6 = 91*1/6

RANDOM VARIABLES

- Outcome of an experiment → a quantitative or a qualitative data
- A random variable is a function that assign a unique numerical value to each outcome of an experiment along with an associated probability.
  - Discrete random variable corresponds to an experiment that has a finite or countable number of outcomes.
  - A continuous random variable is a random variable for which the set of possible outcomes is continuous

RANDOM VARIABLES

- Discrete random variable: take a finite discrete value
  - Infinite number of values
    - Number of hospitalization stay: X = {0, 1, 2, …, n, …}
    - Number of bacteria: X = {0, 1, 2, …, n, …}
  - Finite number of values
    - Number of A blood group in a sample of subjects: X = {0, 1, 2, …, n}
- Continuous random variable: random variable where the data can take infinitely many values
  - Infinite number of values

Discrete Probability Distributions

- Probability Distribution: symbols

$$x:\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \Pr(x_1) & \Pr(x_2) & \dots & \Pr(x_n) \end{pmatrix}$$

- **Property:** the probabilities that appear in distribution of a finite random variable verify the formula:

$$\sum_{i=1}^{n} \Pr(x_i) = 1$$

DISCRETE PROBABILITY DISTRIBUTIONS

- The **mean** of discrete probability distribution (called also expected value) is give by the formula:

$$M(x) = \sum_{i=1}^{n} x_i \cdot \Pr(x_i)$$

- Represents the weighted average of possible values, each value being weighted by its probability of occurrence.

- Variance: is a weighted average of the squared deviations in X

$$V(x) = \sum_{i=1}^{n} (x_i - M(x))^2 \cdot \Pr(x_i)$$

- Standard deviation:

$$\sigma(x) = \sqrt{V(x)} = \sqrt{\sum_{i=1}^{n} (x_i - M(x))^2 \cdot \Pr(x_i)}$$

Discrete Probability Distributions

Example: Let *x* be a random variable represented by the number of otitis episodes in the first two years of life in a community. This random variable has the following distribution:

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.129 & 0.264 & 0.271 & 0.185 & 0.095 & 0.039 & 0.017 \end{pmatrix}$$

What is the expected number (average) of episodes of otitis during the first two years of life?

DISCRETE PROBABILITY DISTRIBUTIONS

- What is the **expected number (average)** of episodes of otitis during the first two years of life?

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.129 & 0.264 & 0.271 & 0.185 & 0.095 & 0.039 & 0.017 \end{pmatrix}$$

- $M(x) = 0 \cdot 0.129 + 1 \cdot 0.264 + 2 \cdot 0.271 + 3 \cdot 0.185 + 4 \cdot 0.095 + 5 \cdot 0.039 + 6 \cdot 0.017$
- $M(x) = 0 + 0.264 + 0.542 + 0.555 + 0.38 + 0.195 + 0.102$
- $M(x) = 2.038$

Discrete Probability Distributions

x_i	$Pr(x_i)$	$x_i*Pr(x_i)$	$x_i-M(x)$	$(x_i-M(x))^2$	$(x_i-M(x))^2*Pr(x_i)$
0	0.129	0	-2.038	4.153	0.536
1	0.264	0.264	-1.038	1.077	0.284
2	0.271	0.542	-0.038	0.001	0.000
3	0.185	0.555	0.962	0.925	0.171
4	0.095	0.38	1.962	3.849	0.366
5	0.039	0.195	2.962	8.773	0.342
6	0.017	0.102	3.962	15.697	0.267
		M(x)=**2.038**			V(x)=**1.967**
					σ(x)=**1.402**

Discrete Probability Distributions By Examples

- **Bernoulli:** head versus tail (two possible outcomes)
- **Binomial:** number of 'head' obtained by throwing a coin of n times
- **Poisson:** number of patients consulted in a emergency office in one day

Bernoulli Distribution

- A random variable whose response are dichotomous is called a Bernoulli random variable
- The experiment has just two possible outcomes (success and failure – dichotomial variable):
  - Gender: boy or girl
  - Results of a test: positive or negative
- Probability of success = p
- Probability of failure = 1-p

X	1	0
Pr(X=x)	p	1-p

- Mean of X: $M(x) = 1 \cdot p + 0 \cdot (1-p)$
- Variance of X: $V(x) = p \cdot (1-p)$

Binomial Distribution

- An experiment is given by repeating a test of *n* times (*n* = known natural number).
- The possible outcomes of each attempt are two events called success and failure
- Let noted with *p* the probability of success and with *q* the probability of failure ($q = 1 - p$)
- The *n* repeated tests are independent
- In a binomial experiment:
  - It consists of a fixed number n of identical experiments
  - There are only two possible outcomes in each experiments, denoted by S (success) and F (failure)
  - The experiments are independent with the same probability of S (denoted p)
- ${}_4C_2$ = 4 choose 2 (combination choosing 2 from 4)

BINOMIAL DISTRIBUTION

Mean	**M(x) = n·p**
Variance	V(x) = n·p·q
Standard deviation	σ(x) = √(n·p·q)

- The number of successes X obtained by performing the test n times is a random variable of n and p parameters and is noted as Bi(n,p)
- The random variable X can take the following values: 0, 1, 2,...n
- Probability that X to be equal with a value k is given by the formula:

$$\Pr(X = k) = C_n^k p^k q^{n-k}$$

where:

$$C_n^k = \frac{n!}{k! \cdot (n-k)!}$$

Binomial Distribution

- Suppose that 90% of adults with joint pain report symptomatic relief with a specific medication. If the medication is given to 10 new patients with joint pain, what is the probability that it is effective in exactly 7 subjects?
- Assumptions of the binomial distribution model?
  - The outcome is pain relief (yes or no), and we will consider that the *pain relief* is a success
  - The probability of success for each subject is 0.9 (p=0.9)
  - The final assumption is that the subjects are independent, and it is reasonable to assume that this is true.
- $Pr(x=7) = {}_{10}C_7 \cdot 0.9^7 \cdot (1-0.9)^{10-7} = 0.0574$ → there is a 5.74% chance that exactly 7 of 10 patients will report pain relief when the probability that any one reports relief is 90%.

BINOMIAL DISTRIBUTION

$$Pr(x=k) = C_n^k p^k q^{n-k}$$

- What is the probability to have 2 boys in a family with 5 children if the probability to have a baby-boy for each birth is equal to 0.47 and if the sex of children successive born in the family is considered an independent random variable?

- p=0.47
- q=1-0.47=0.53
- n=5
- k=2
- **Pr(x=2)=10·0.47^2·0.53^3**
- **Pr(x=2) = 0.33**

$$C_5^2 = \frac{5!}{2!\cdot(5-2)!} = \frac{5\cdot 4\cdot 3\cdot 2\cdot 1}{2\cdot 1\cdot (3\cdot 2\cdot 1)} = \frac{120}{12} = 10$$

POISSON DISTRIBUTION

- Random Poisson variable take a countable infinity of values (0,1,2,...,k,...) that is the number of achievements of an event within a given range of time or place
  - number of entries per year in a given hospital
  - white blood cells on smear
  - number of decays of a radioactive substance in a given time T

POISSON DISTRIBUTION

- POISSON random variable:
  - Is characterized by theoretical parameter θ (expected average number of achievement for a given event in a given range)
- Symbol: Po(θ)
- Poisson Distribution:

$$X:\begin{pmatrix} k \\ e^{-\theta}\cdot\dfrac{\theta^k}{k!} \end{pmatrix}$$

$$\Pr(x=k)=\frac{e^{-\theta}\cdot\theta^k}{k!}$$

- Mean of expected values: M(x) = θ
- Variance: V(x) = θ

POISSON DISTRIBUTION

- The mortality rate for specific viral pathology is 7 per 1000 cases. What is the probability that in a group of 400 people this pathology to cause 5 deaths?
- n=400
- p=7/1000=0.007
- θ=n·p=400·0.007=2.8
- e=2.718281828=2.72

Pr(x=5) = $(2.72^{-2.8} \cdot 2.8^{5})/(5\cdot4\cdot3\cdot2\cdot1)$ =10.45/120
Pr(X=5) = 0.09

RECALL

- Random variables could be discrete or continuous.
- For random variables we have:
  - Discrete probability distributions
  - Continuous probability distribution
- It is possible to compute the expected values (means), variations and standard deviation for discrete and random variables.

Thank you!