MA201: Probability and Statistics

Presentation by:

Jignesh Patel (jignesh_patel@diu.iiitvadodara.ac.in)

Indian Institute of Information Technology, Vadodara

BOOKS

Text Book:

[1] Probability and Statistics for Computer Scientists, by Michael Baron, CRC Press, second edition (2013).

Reference Books:

[2] Athanasios Papoulis and S. Unnikrishna Pillai, Probability, Random Variables, and Stochastic Processes , Fourth Edition, McGraw-Hill, Europe,2002.

[3] Alberto Leon-Gracia, Probability, Statistics, and Random Processes for Electrical Engineering, Third Edition, Pearson, 2008.

Outline

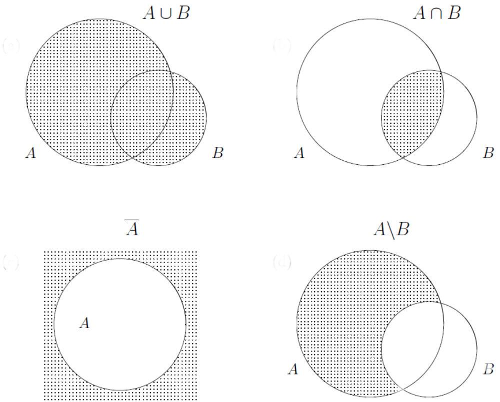
1. Introduction
   - Classical Probability
   - Relative frequency and axiomatic definition of probability
   - Addition rule and conditional probability
   - Multiplication Rule
   - Bayes' theorem
   - Total Probability
2. Random Variables
   - Discrete Random Variable
   - Probability Mass Function
   - Cumulative Density Function
   - Probability Density Function
   - Expectation, Variance
   - Moments
   - Moment Generating Functions
   - Median
   - Quantile
   - Markov inequality
   - Chebyshev inequality
3. Spacial Distribution
   - Discrete uniform
   - Bernoulli
   - Binomial
   - Geometric
   - Negative Binomial
   - Hypergeometric
   - Poisson

Outline

What is Probability?

- Two types of systems
  - Deterministic/Certain
  - Random/ Uncertain
- *Probability* is used to represent/analyse uncertainty.
- Sample Space: All possible outcomes of Experiment.
- Event: Set of outcomes.
- How many events possible for a sample space of n possible outcomes?

Set operations



Classical Probability

$$P(A) = \frac{\mathcal{N}_F}{\mathcal{N}_T} = \frac{\texttt{number of favorable outcomes}}{\texttt{total number of outcomes}}$$

Relative frequency

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

where,

n_A is the number of time A occurs

n is the number of trials

Axioms of Probability

For sample space Ω and event A

$$P(A) \geq 0$$
$$P(\Omega) = 1$$
$$P(E_1 \cup E_2 \cup E_3 ...) = P(E_1) + P(E_2) + P(E_3)...$$

where,

$E_1, E_2, ...$are mutually exclusive events.

Examples

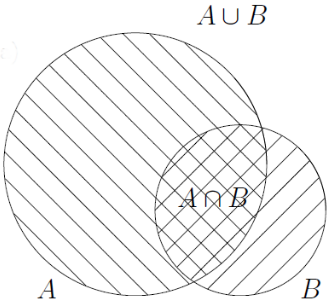
Experiment: Tossing six side fair dice.

1. What is Sample space?
2. What is probability of outcome 6 using classical definition?

Experiment: Tossing six side fair dice twice and note the sum of shown numbers.

1. What is Sample space?
2. What is probability of outcome 7 using classical definition?

Addition rule



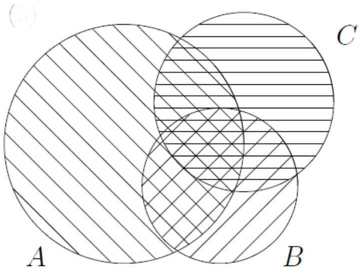
$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}$$

For mutually exclusive events,

$$P\{A \cup B\} = P\{A\} + P\{B\}$$

Addition rule

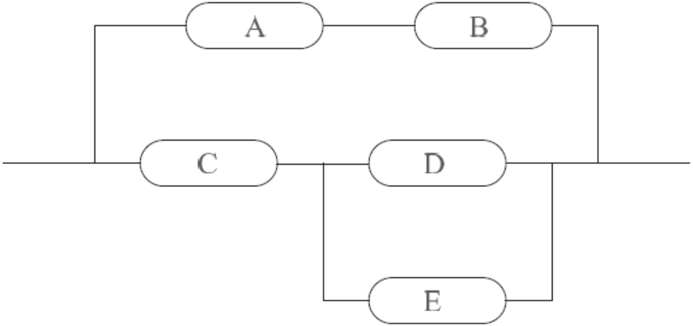
$A \cup B \cup C$



$$\boldsymbol{P}\{A \cup B \cup C\} =$$
$$\boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} + \boldsymbol{P}\{C\}$$
$$- \boldsymbol{P}\{A \cap B\} - \boldsymbol{P}\{A \cap C\} - \boldsymbol{P}\{B \cap C\} +$$
$$\boldsymbol{P}\{A \cap B \cap C\}$$

Reliability

- Probability can be used to check the reliability of the system.
- Calculate reliability of the system in given Figure if each component is operable with probability 0.92 independently of the other components.



Permutations and combinations

Calculation of total and favourable outcomes can be calculated using permutations and combinations.

- Permutations with replacement : k distinguishable objects from set of n with replacements

$$P_r(n,k) = \overbrace{n \cdot n \cdot \ldots \cdot n}^{k \text{ terms}} = n^k$$

- Permutations without replacement : k distinguishable objects from set of n without replacements

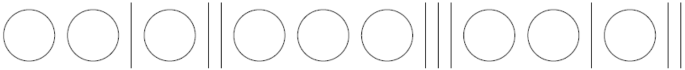
$$P(n,k) = \overbrace{n(n-1)(n-2) \cdot \ldots \cdot (n-k+1)}^{k \text{ terms}} = \frac{n!}{(n-k)!}$$

Permutations and combinations

- Combinations without replacement : k indistinguishable objects from set of n without replacements

$$C(n,k) = \binom{n}{k} = \frac{P(n,k)}{P(k,k)} = \frac{n!}{k!\,(n-k)!}$$

- Combinations with replacement : k indistinguishable objects from set of n with replacements



$$C_r(n,k) = \binom{k+n-1}{k} = \frac{(k+n-1)!}{k!\,(n-1)!}$$

Conditional Probability

Experiment: Tossing 3 coins

1. What is probability the Last coin turns out to be Head?

Conditional Probability

Experiment: Tossing 3 coins

1. What is probability the Last coin turns out to be Head?
2. If first two coins are Heads, then what is probability the Last coin turns out to be Head?

Conditional Probability

Experiment: Tossing 3 coins

1. What is probability the Last coin turns out to be Head?
2. If first two coins are Heads, then what is probability the Last coin turns out to be Head?

- Probability of A given that B already occurred,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional Probability

90% of flights depart on time. 80% of flights arrive on time. 75% of flights depart on time and arrive on time.

1. `You are meeting a flight that departed on time. What is the probability that it will arrive on time?`

Conditional Probability

90% of flights depart on time. 80% of flights arrive on time. 75% of flights depart on time and arrive on time.

1. You are meeting a flight that departed on time. What is the probability that it will arrive on time?
2. You have met a flight, and it arrived on time. What is the probability that it departed on time?

Conditional Probability

90% of flights depart on time. 80% of flights arrive on time. 75% of flights depart on time and arrive on time.

1. You are meeting a flight that departed on time. What is the probability that it will arrive on time?
2. You have met a flight, and it arrived on time. What is the probability that it departed on time?
3. Are the events, departing on time and arriving on time, independent?

Multiplication Rule

- A and B both occurs can be given by,

$$P(A \cap B) = P(A)P(B|A)$$

$$P(A \cap B) = P(B)P(A|B)$$

Multiplication Rule

- A and B both occurs can be given by,

$$P(A \cap B) = P(A)P(B|A)$$

$$P(A \cap B) = P(B)P(A|B)$$

- For A,B and C

$$\begin{aligned} P(A \cap B \cap C) &= P(A \cap B)P(C|(A \cap B)) \\ &= P(A)P(B|A)P(C|(A \cap B)) \end{aligned}$$

Multiplication Rule

- A and B both occurs can be given by,

$$P(A \cap B) = P(A)P(B|A)$$

$$P(A \cap B) = P(B)P(A|B)$$

- For A,B and C

$$\begin{aligned} P(A \cap B \cap C) &= P(A \cap B)P(C|(A \cap B)) \\ &= P(A)P(B|A)P(C|(A \cap B)) \end{aligned}$$

- In general,

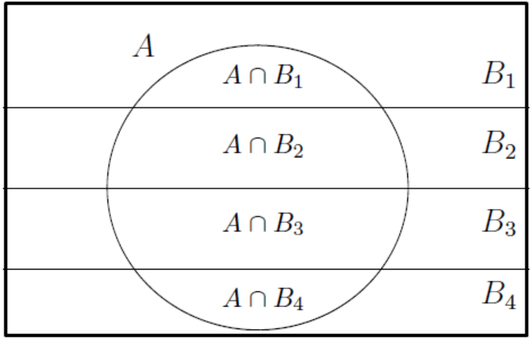
$$P(A_1 \cap A_2 .... \cap A_n) = P(A_1)\Pi_{n=2}^{n} P(A_i | A_1 \cap A_2 ... A_i)$$

Bayes' theorem

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Total Probability



For sample space Ω, Events B_i which is $\forall_i, (B_1 \cup B_2 \cup ....B_i) = \Omega$.

$$P(A) = \Sigma_{\forall_i} P(A|B_i)P(B_i)$$

Some Problems:

We have 4 boxes. B1 contains 2000 items with 5% defective. B2 contains 500 items with 40% defective. B3 contains 1000 items with 10% defective. B4 contains 1000 items with 10% defective. We select one box at random and remove one component at random from the box.

1. What is the probability that the selected item is defective?

Some Problems:

We have 4 boxes. B1 contains 2000 items with 5% defective. B2 contains 500 items with 40% defective. B3 contains 1000 items with 10% defective. B4 contains 1000 items with 10% defective. We select one box at random and remove one component at random from the box.

1. What is the probability that the selected item is defective?
2. We know that the item selected is defective. What is the probability that it comes from B2?

Some Problems:

Among 10 laptop computers, five are good and five have defects. Unaware of this, a customer buys 6 laptops.

1. What is the probability of exactly 2 defective laptops among them?

Some Problems:

Among 10 laptop computers, five are good and five have defects. Unaware of this, a customer buys 6 laptops.

1. `What is the probability of exactly 2 defective laptops among them?`
2. `Given that at least 2 purchased laptops are defective, what is the probability that exactly 2 are defective?`

Some Problems:

A quiz consists of 6 multiple-choice questions. Each question has 4 possible answers. A student is unprepared, and he has no choice but guessing answers completely at random. He passes the quiz if he gets at least 3 questions correctly.

1. What is the probability that he will pass?

Some Problems:

A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9.

1. What is probability in error in module 1 only?

Some Problems:

A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9.

1. What is probability in error in module 1 only?
2. What is probability in error in module 2 only?

Some Problems:

A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9.

1. What is probability in error in module 1 only?
2. What is probability in error in module 2 only?
3. Suppose the program crashed. What is the probability of errors in both modules?

Random Variable

What is Random Variable?

It is transformation/function $\mathcal{X}$ which maps Sample space Ω in the random experiments to a Real Line.

$$\mathcal{X} : \Omega \to R$$

Types:

- Discrete
- Continuos
- Mixed

Discrete Random Variable

A discrete random variable X is defined as a random variable that assumes values from a countable set

Ex: Tossing Coin 3 times

- X_h : Number of heads
- Y : If Number of heads grater than tails 1 else 0.
- X_t : Number of tails
- $Z = \min(X_h, X_t)$

Discrete Random Variable

A discrete random variable X is defined as a random variable that assumes values from a countable set

Ex: Tossing Coin 3 times

- X_h : Number of heads
- Y : If Number of heads grater than tails 1 else 0.
- X_t : Number of tails
- $Z = \min(X_h, X_t)$

Is function of random variable is Random Variable?

- A. Yes.
- B. No.

Discrete Random Variable

Can same out comes from sample space be represented with two different random variable values?

- A. Yes.
- B. No.

Probability Mass Function

Collection of all the probabilities related to X is the distribution of X. The function

$$P(x) = P\{X = x\}$$

Probability Mass Function

Collection of all the probabilities related to X is the distribution of X. The function

$$P(x) = P\{X = x\}$$

$$\Sigma_{\forall_i} P\{X = x_i\} = ?$$

- A. any value between 0 and 1.
- B. 1.

Probability Mass Function

Ex: Tossing 3 Coins

- X : Number of heads

Probability Mass Function

Ex: Tossing 3 Coins

- X : Number of heads
- X can take $0, 1, 2, 3$

Probability Mass Function

Ex: Tossing 3 Coins

- X : Number of heads
- X can take $0, 1, 2, 3$
- $$P\{X = 0\} = 1/8$$
  $$P\{X = 1\} = 3/8$$
  $$P\{X = 2\} = 3/8$$
  $$P\{X = 3\} = 1/8$$

Probability Mass Function

Ex: Tossing 3 Coins twice

- X_1 : Number of heads first time and, X_2 : Number of heads second time
- $Y = X_1 + X_2$
- Y can take $0, 1, 2, 3, 4, 5, 6$

Probability Mass Function

Ex: Tossing 3 Coins twice

- X_1 : Number of heads first time and, X_2 : Number of heads second time
- $Y = X_1 + X_2$
- Y can take $0, 1, 2, 3, 4, 5, 6$

$$P\{Y = 5\} = ?$$

- A. 3/32.
- B. 3/16.
- C. 1/16.
- D. 1/32.

Cumulative Density Function

The cumulative distribution function, or CDF is defined as

$$F_X(x) = P\{X \leq x\} = \sum_{y \leq x} P(y)$$

Some Properties of CDF

- $F_X(\infty) = 1$
- $F_X(-\infty) = 0$
- $P\{X > x\} + P\{X \leq x\} = 1$
- Non decreasing function
- $P\{a < X \leq b\} = F_x(b) - F_x(a)$

Cumulative Density Function

Ex: Tossing 3 Coins

- X : Number of heads
- What is a $F_X(1)$?
  - A. 1/8
  - B. 1/2
  - C. 3/8

Probability Density Function (PDF)

- For all continuous variables, the probability PMF is always equal to zero,

$$P(x) = 0, \forall x.$$

Probability Density Function (PDF)

- For all continuous variables, the probability PMF is always equal to zero,

$$P(x) = 0, \forall x.$$

- CDF, In the continuous case, it equals

$$F_X(x) = P\{X \leq x\} = P\{X < x\}.$$

Probability Density Function (PDF)

- For all continuous variables, the probability PMF is always equal to zero,

$$P(x) = 0, \forall x.$$

- CDF, In the continuous case, it equals

$$F_X(x) = P\{X \leq x\} = P\{X < x\}.$$

- Probability density function (PDF, density) is the derivative of the cdf,

$$f(x) = F_X'(x)$$

.

Probability Density Function (PDF) [1]

Distribution	Discrete	Continuous
Definition	$P(x) = P\{X = x\}(\text{PMF})$	$f(x) = F'(x)(\text{PDF})$
CDF	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$
Total probability	$\sum_{x} P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

Expectation [1]

Expectation or expected value of a random variable X is its mean, the average value.

Example: Tossing Coin. Random Variable X take 1 for heads outcome and 0 for tails.

1. What is Expected value for fair coin?
   A. 0.5
   B. 0.7

Expectation [1]

Expectation or expected value of a random variable X is its mean, the average value.

Example: Tossing Coin. Random Variable X take 1 for heads outcome and 0 for tails.

1. What is Expected value for fair coin?
   A. 0.5
   B. 0.7
2. What is Expected value for coin with tail probability 0.3 coin?
   A. 0.3
   B. 0.7
   C. 0.5

Expectation [1]

- Expectation, discrete case,

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

Expectation [1]

- Expectation, discrete case,

$$\mu = \mathrm{E}(X) = \sum_x xP(x)$$

- Expectation, Continuos case,

$$\mu = \mathrm{E}(X) = \int_x xf(x)dx$$

Example: Expectation [3]

In coin betting game, a person need to pays \$1.50 to toss a coin three times. He get \$1 if the number of heads is 2, \$8 if the number is 3, else nothing. What is the expected amount person win in each game?

1. A. 11/8
2. B. 1/8
3. C. -1/8
4. D. -11/8

Variance [1]

Variance of a random variable is defined as the expected squared deviation from the mean. For discrete random variables, variance is

$$\sigma^2 = \mathrm{Var}(X) = \mathbf{E}(X - \mathbf{E}X)^2 = \sum_x (x-\mu)^2 P(x)$$

Moments [2]

- Moments
  discrete case,

$$m_n = E(X^n) = \sum_x x^n P(x)$$

  continuos case,

$$m_n = E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx$$

Moments [2]

- Moments
  discrete case,
  $$m_n = E(X^n) = \sum_x x^n P(x)$$
  continuos case,
  $$m_n = E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx$$
- Central Moments
  discrete case,
  $$m_n = E((X - E(X))^n) = \sum_x (x - E(X))^n P(x)$$
  continuos case,
  $$m_n = E((X - E(X))^n) = \int_{-\infty}^{\infty} (x - E(X))^n f(x)dx$$

Moments [2]

- Moments
  discrete case,
  $$m_n = E(X^n) = \sum_x x^n P(x)$$
  continuos case,
  $$m_n = E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx$$
- Central Moments
  discrete case,
  $$m_n = E((X - E(X))^n) = \sum_x (x - E(X))^n P(x)$$
  continuos case,
  $$m_n = E((X - E(X))^n) = \int_{-\infty}^{\infty} (x - E(X))^n f(x)dx$$
- Same way Absolute moments = $E(|X|^n)$ and Generalized moments $E((X - a)^n)$

Moment Generating Functions (MGF) [2]

For discrete random variable, MGF,

$$\Gamma(z) = E(z^X) = \sum_x P(X = x)Z^x$$

with $z = 1$, first derivative

$$\Gamma^{'}(z) = E(X)$$

and second derivative

$$\Gamma^{''}(z) = E(X^2) - (E(X))^2$$

Moment Generating Functions (MGF) [2]

Example: Tossing Coin. Random Variable X take 1 for heads outcome and 0 for tails. Head probability $p = 0.7$ and q tail probability.

1. `MGF`, $\Gamma(z) = \sum_x P(X = x) z^x = pz + q$
2. $\Gamma^{'}(z) = p$

Median [1]

Median M is a number that is exceeded with probability no greater than 0.5 and is preceded with probability no greater than 0.5. That is, M is such that

$$P\{X > M\} \leq 0.5$$

$$P\{X < M\} \leq 0.5$$

Quantile [1]

A p-quantile of a population is such a number x that solves equations

$$P\{X < x\} \leq p$$

$$P\{X > x\} \leq 1 - p$$

Markov inequality [3]

Suppose first that X is a nonnegative random variable with mean $E[X]$. The Markov inequality then states that

$$P[X \geq a] \leq \frac{E[X]}{a}$$

Chebyshev inequality [3]

Now suppose that the mean $E[X] = m$and the variance $VAR[X] = \sigma^2$ of a random variable are known, The Chebyshev inequality states that

$$P[|X - m| \geq a] \leq \frac{\sigma^2}{a^2}$$

Discrete uniform [3]

All random variable value take same probability.

Discrete uniform [3]

All random variable value take same probability.

1. $S_X = \{0, 1, 2, ...n\}$
2. $P(x) = \frac{1}{n}$
3. $E[X] = \frac{n+1}{2}$
4. $VAR[X] = \frac{n^2-1}{12}$

Bernoulli [1]

A random variable with two possible values, 0 and 1, is called a Bernoulli variable, its distribution is Bernoulli distribution, and any experiment with a binary outcome is called a Bernoulli trial.

Bernoulli [1]

A random variable with two possible values, 0 and 1, is called a Bernoulli variable, its distribution is Bernoulli distribution, and any experiment with a binary outcome is called a Bernoulli trial.

1. $S_X = \{0, 1\}$
2. $P(x) \begin{cases} = q = 1 - p \texttt{ if } x = 0 \\ = p \texttt{ if } x = 1 \end{cases}$
3. $E[X] = p$
4. $VAR[X] = p(1 - p)$

Binomial [1]

A variable described as the number of successes in a sequence of independent Bernoulli trials has Binomial distribution. Its parameters are n, the number of trials, and p, the probability of success.

Binomial [1]

A variable described as the number of successes in a sequence of independent Bernoulli trials has Binomial distribution. Its parameters are n, the number of trials, and p, the probability of success.

1. $S_X = \{0, 1, 2, 3, ...n\}$
2. p = Probability of success
3. $P(x) = \binom{n}{x} p^x (1-p)^{n-x}$
4. $E[X] = np$
5. $VAR[X] = np(1-p)$

Geometric [1]

The number of Bernoulli trials needed to get the first success has Geometric distribution.

Geometric [1]

The number of Bernoulli trials needed to get the first success has Geometric distribution.

1. $S_X = \{1, 2, 3, ...n\}$
2. p = Probability of success
3. $P(x) = (1-p)^{(x-1)}p$
4. $E[X] = \frac{1}{p}$
5. $VAR[X] = \frac{1-p}{p^2}$

Negative Binomial [1]

In a sequence of independent Bernoulli trials, the number of trials needed to obtain k successes has Negative Binomial distribution.

Negative Binomial [1]

In a sequence of independent Bernoulli trials, the number of trials needed to obtain k successes has Negative Binomial distribution.

1. $S_X = \{k, k+1, k+2...\}$
2. p = Probability of success
3. $P(x) = P\{$ the x-th trial results in the k-th success $\}$
   $P(x) = P\{(k-1)$ successes in the first $(x-1)$ trials, and the last trial is a success $\}$
   $P(x) = \binom{x-1}{k-1}(1-p)^{x-k}p^k$
4. $E[X] = \frac{k}{p}$
5. $VAR[X] = \frac{k(1-p)}{p^2}$

Hypergeometric [2]

Describes the probability of k successes in n draws, without replacement, from a finite population of size N that contains exactly K objects with that feature, wherein each draw is either a success or a failure. In contrast, the binomial distribution describes the probability of k successes in n draws with replacement.

Hypergeometric [2]

Describes the probability of k successes in n draws, without replacement, from a finite population of size N that contains exactly K objects with that feature, wherein each draw is either a success or a failure. In contrast, the binomial distribution describes the probability of k successes in n draws with replacement.

1. $P(X = k) = \dfrac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
2. $E[X] = \frac{nk}{N}$
3. $VAR[X] = \frac{nK(n-K)(n-N)}{nN^2(N-1)}$

Poisson [1]

The number of rare events occurring within a fixed period of time has Poisson distribution.

Poisson [1]

The number of rare events occurring within a fixed period of time has Poisson distribution.

1. $\lambda =$ frequency, average number of events
2. $P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$
3. $E[X] = \lambda$
4. $VAR[X] = \lambda$

Uniform Distributions [1]

Uniform distribution is used in any situation when a value is picked "at random" from a given interval.

Uniform Distributions [1]

Uniform distribution is used in any situation when a value is picked "at random" from a given interval.

1. (a, b) = range of values
2. $f(x) = \frac{1}{b-a}, a < x < b$
3. $E[X] = \frac{a+b}{2}$
4. $VAR[X] = \frac{(b-a)^2}{12}$

Exponential distribution [1]

Exponential distribution is often used to model time: waiting time, interarrival time, hardware lifetime, failure time, time between telephone calls, etc.

Exponential distribution [1]

Exponential distribution is often used to model time: waiting time, interarrival time, hardware lifetime, failure time, time between telephone calls, etc.

1. λ = frequency parameter, the number of events per time unit
2. $f(x) = \lambda \exp^{-\lambda x}, x > 0$
3. $E[X] = \frac{1}{\lambda}$
4. $VAR[X] = \frac{1}{\lambda^2}$

Normal distribution [1]

Various fluctuations and measurement errors that consist of accumulated number of small terms appear normally distributed. Normal distribution is often found to be a good model for physical variables like weight, height,e.t.c,

Normal distribution [1]

Various fluctuations and measurement errors that consist of accumulated number of small terms appear normally distributed. Normal distribution is often found to be a good model for physical variables like weight, height,e.t.c,

$$\begin{aligned}
\mu &= \text{ expectation, location parameter} \\
\sigma &= \text{ standard deviation, scale parameter} \\
f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, -\infty < x < \infty \\
\mathbf{E}(X) &= \mu \\
\mathrm{Var}(X) &= \sigma^2
\end{aligned}$$

Gamma distribution [1]

When a certain procedure consists of α independent steps, and each step takes Exponential(λ) amount of time, then the total time has Gamma distribution with parameters α and λ.

Gamma distribution [1]

When a certain procedure consists of α independent steps, and each step takes Exponential(λ) amount of time, then the total time has Gamma distribution with parameters α and λ.

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\ \mathrm{d}x, \qquad \texttt{for } \alpha > 0$$

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, \quad x > 0$$

$$\mathrm{E}(X) = \frac{\alpha}{\lambda}$$

$$\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$$

Joint distribution and marginal distributions

Joint distribution

If X and Y are random variables, then the pair (X,Y) is a random vector. Its distribution is called the joint distribution of X and Y.

Joint distribution

Tossing Coin 3 times

1. X: Number of Heads
2. Y: 0 for Even Number of heads else 1.

Joint distribution

Tossing Dice 2 times. X is min number of dots from both toss and Y is max number of dots in both toss. What is $P\{X = Y = k\}$?

1. A. 1/36
2. B. 2/36
3. C. 6/36
4. D. 0

Marginal distributions

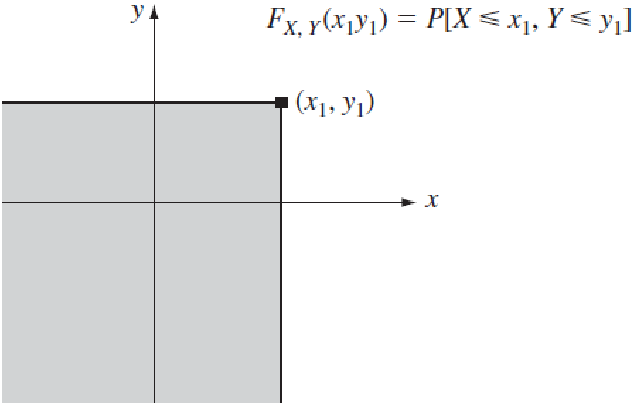
Marginal distribution

Individual distributions of X and Y are then called the marginal distributions.

Conditional distributions

$$P\left[Y = y|X = x\right] = \frac{P\left[Y = y, X = x\right]}{P\left[X = x\right]}$$

Joint CDF [3]



Joint CDF

Tossing Dice 2 times. X is min number of dots from both toss and Y is max number of dots in both toss. What is $F_{(X,Y)}(2,3)$?

1. A. 3/36
2. B. 5/36
3. C. 8/36
4. D. 9/36

Properties of expectation

$$\mathbf{E}(aX + bY + c) = a\mathbf{E}(X) + b\mathbf{E}(Y) + c$$

In particular,

$$\mathrm{E}(X + Y) = \mathbf{E}(X) + \mathrm{E}(Y)$$
$$\mathbf{E}(aX) = a\mathbf{E}(X)$$
$$\mathbf{E}(c) = c$$

For independent X and Y,

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$$

Covariance and correlation

$$Cov(X,Y) = E\{(X - E(X))(Y - E(Y)\})$$

Correlation

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Example

A program consists of two modules. The number of errors, X, in the first module and the number of errors, Y , in the second module have the joint distribution, $P(0,0) = P(0,1) = P(1,0) = 0.2, P(1,1) = P(1,2) = P(1,3) = 0.1, P(0,2) = P(0,3) = 0.05.$

Find

1. the marginal distributions of X and Y.
2. the probability of no errors in the first module
3. the distribution of the total number of errors in the program.
4. find out if errors in the two modules occur independently.
5. $Cov(X,Y)$
6. Correlation ρ_{XY}

Conditional Expectation

$$E[Y|X = x] = \sum_y yP\{Y = y|X = x\}$$

Multivariate Gaussian Distribution

A vector-valued random variable $[X = X1...Xn]^T$ is said to have a multivariate normal (or Gaussian) distribution with mean $\mu \in \mathcal{R}^n$ and covariance matrix $\sum$ if its probability density function is given by

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Bivariate Gaussian Distribution

$$f_{X,Y}(x,y) = \frac{\exp\left\{\frac{-1}{2\left(1-\rho_{X,Y}^2\right)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right) + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}}$$

Example

RVs X and Y are jointly Gaussian and their joint distribution is given by,

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{-(x^2-2\rho xy+y^2)}{2(1-\rho^2)}}, -\infty < x, y < \infty.$$

Find the marginal PDFs. Are X and Y independent?

Assignment 4

Show graphically for independently generated random variable X and Y,

$$f_{xy}(x,y) = f_x(x) f_y(y).$$

1. Generate X and Y from Gaussian distribution with 0 mean and variance 1.
2. find $f_{xy}(x,y)$ using $f_x(x)$ and $f_y(y)$.
   HINT: $[f_{xy}(x,y)]_{no\_bins*no\_bins} = [f_x(x)]_{no\_bins\times 1} \times [f_y(y)]_{1\times no\_bins}$
3. Plot joint PDF.

Assignment 5

Three different Gaussian random variables, i.e., $X1, X2$ and $X3$ with 0 mean and 1 variance.,

Compute the covariance matrix of $X1$, $X2$ and $X3$. Covariance matrix (CV)

$$CV = \begin{bmatrix} cov(X1, X1) & cov(X1, X2) & cov(X1, X3) \\ cov(X2, X1) & cov(X2, X2) & cov(X2, X3) \\ cov(X3, X1) & cov(X3, X2) & cov(X3, X3) \end{bmatrix}.$$

Here,

$$cov(X, Y) = E[XY] - E[X]E[Y]$$

Similarly compute correlational matrix.

Assignment 5

```
[[ 0.99203782 -0.01804732  0.03459005]
 [-0.01804732  1.00488834  0.0406933 ]
 [ 0.03459005  0.0406933   0.99985368]]
<matplotlib.colorbar.Colorbar at 0x7f6dd7cf60f0>
```

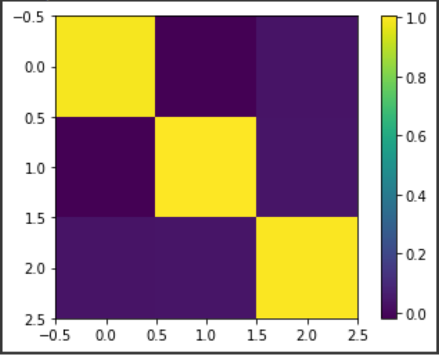


Figure: Output

Assignment 5

Verify the properties of the covariance matrix.

1. Symmetric,i.e.,$C_X = C_X^T$.
2. Its eigenvalues are greater than equal to zero
3. It is positive semi-definite, i.e., for any real valued vector a,

$$a^T C_X a \geq 0$$

Assignment 5

Generate covariance matrix of correlated data. Take face images as the data.

`Show that data and noise are uncorrelated. Take Image files as your data and standard gaussian noise.`

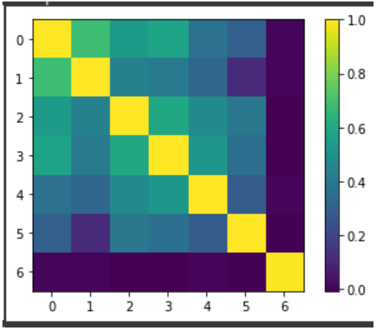


Figure: Output

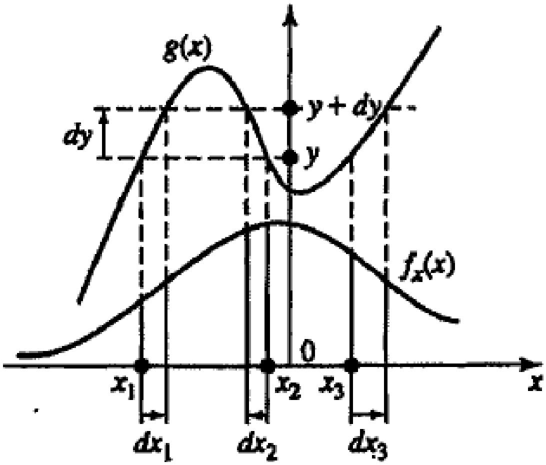
Function of Random variable: discrete

- Example: 3 digit binary number generator. X is random variable showing the decimal value. Y is sum of all the digits.
- Example: $Y = aX + b$ for some RVs X and Y

Function of Random variable: continuous

- Example: $Y = aX + b$ for some RVs X and Y
- Let $X \sim \mathcal{N}(\mu, \sigma^2)$. PDF of $Y = aX + b$.
- For any $Y = g(X)$, to calculate PDF
  1. Calculate $F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$, Here $f_X(x)$ is known.
  2. $f_Y(y) = \frac{d}{dy} F_Y(y)$.
- Example: $Y = X^2$

General Formula Function of Random variable [2]



$$f_Y(y) = \frac{f_X(x_1)}{g'(x_1)} + \frac{f_X(x_2)}{g'(x_2)} + \frac{f_X(x_3)}{g'(x_3)}$$

Sum of random variables [2]

- Example: $Z = X + Y$. X and Y are independent.
- For Discrete case:
$$P\{Z = z\} = \sum_{x} P_X(x) P_Y(z - x)$$
.
- For continuos case:
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$
.
- This is also called as convolution.

Function of random vector [2]

- Example: $X = \frac{X}{Y}$.
- Formula: For random vector $X = [X_1, X_2 ... X_n]$ and $Y = [Y_1, Y_2 ... Y_n]$,

$$f_Y(Y_1, Y_2 ...) = \frac{f_{X_1 X_2 ..}({x_1}^{(1)} {x_2}^{(1)} ...)}{|J({x_1}^{(1)} {x_2}^{(1)} ..)|} + \frac{f_{X_1 X_2 ..}({x_1}^{(2)} {x_2}^{(2)} ...)}{|J({x_1}^{(2)} {x_2}^{(1)} ..)|} ...,$$

$$J(x_1 x_2 ...) = \begin{vmatrix} \partial y_1/\partial x_1 & \partial y_1/x_2 & .. \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 & .. \\ . & . & . \\ . & . & . \end{vmatrix}$$

Order statistics [2]

- Let $X_1, X_2, ..., X_n$ are independent identical random variable.
- $X_{(1)} \leq X_{(2)} .... \leq X_{(k)} .... \leq X_{(n)}$, here $X_{(k)}$ is k^{th} order statistics.

References

[1] M. Baron, *Probability and statistics for computer scientists*, 2nd ed. Chapman and Hall/CRC, 2013.

[2] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. McGraw-Hill, Europe, 2002.

[3] A. L. Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering,*, 3rd ed. Pearson, 2008.

Thank You