Stochastic Processes

Review of Elementary Probability
Lecture I

Hamid R. Rabiee

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Correlation
- Important Theorems

History & Philosophy

- Started by gamblers' dispute
- Probability as a game analyzer !
- Formulated by B. Pascal and P. Fermet
- First Problem (1654) :
  - "Double Six" during 24 throws
- First Book (1657) :
  - *Christian Huygens, "De Ratiociniis in Ludo Aleae", In German, 1657.*

History & Philosophy (Cont'd)

- Rapid development during 18th Century
- Major Contributions:
  - J. Bernoulli (1654-1705)
  - A. De Moivre (1667-1754)

History & Philosophy (Cont'd)

- A renaissance: Generalizing the concepts from mathematical analysis of games to analyzing scientific and practical problems: P. Laplace (1749-1827)

- New approach first book:
  - P. Laplace, "*Théorie Analytique des Probabilités*", In France, 1812.

History & Philosophy (Cont'd)

- 19th century's developments:
  - Theory of errors
  - Actuarial mathematics
  - Statistical mechanics

- Other giants in the field:
  - Chebyshev, Markov and Kolmogorov

History & Philosophy (Cont'd)

- Modern theory of probability (20^{th}) :
  - A. Kolmogorov : Axiomatic approach

- First modern book:
  - *A. Kolmogorov, "Foundations of Probability Theory"*, Chelsea, New York, 1950

- Nowadays, Probability theory as a part of a theory called Measure theory !

History & Philosophy (Cont'd)

- Two major philosophies:
  - Frequentist Philosophy
    - Observation is enough
  - Bayesian Philosophy
    - Observation is NOT enough
    - Prior knowledge is essential

- Both are useful

History & Philosophy (Cont'd)

Frequentist philosophy

- There exist fixed parameters like mean,θ.
- There is an underlying distribution from which samples are drawn
- Likelihood functions(L(θ)) maximize parameter/data
- For Gaussian distribution the L(θ) for the mean happens to be $1/N\sum_i x_i$ or the average.

Bayesian philosophy

- Parameters are variable
- Variation of the parameter defined by the prior probability
- This is combined with sample data p(X/θ) to update the posterior distribution p(θ/X).
- Mean of the posterior, p(θ/X),can be considered a point estimate of θ.

History & Philosophy (Cont'd)

- An Example:
  - A coin is tossed 1000 times, yielding 800 heads and 200 tails. Let p = P(heads) be the bias of the coin. What is p?
- Bayesian Analysis
  - Our prior knowledge (belief) : $\pi(p) = 1 \,(\text{Uniform}(0,1))$
  - Our posterior knowledge : $\pi(p \mid Observation) = p^{800}(1-p)^{200}$
- Frequentist Analysis
  - Answer is an estimator $\hat{p}$ such that
    - Mean : $E[\hat{p}] = 0.8$
    - Confidence Interval : $P(0.774 \le \hat{p} \le 0.826) \ge 0.95$

History & Philosophy (Cont'd)

- Further reading:
  - http://www.leidenuniv.nl/fsw/verduin/stathist/stathist.htm
  - http://www.mrs.umn.edu/~sungurea/introstat/history/indexhistory.shtml
  - www.cs.ucl.ac.uk/staff/D.Wischik/Talks/histprob.pdf

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Correlation
- Important Theorems

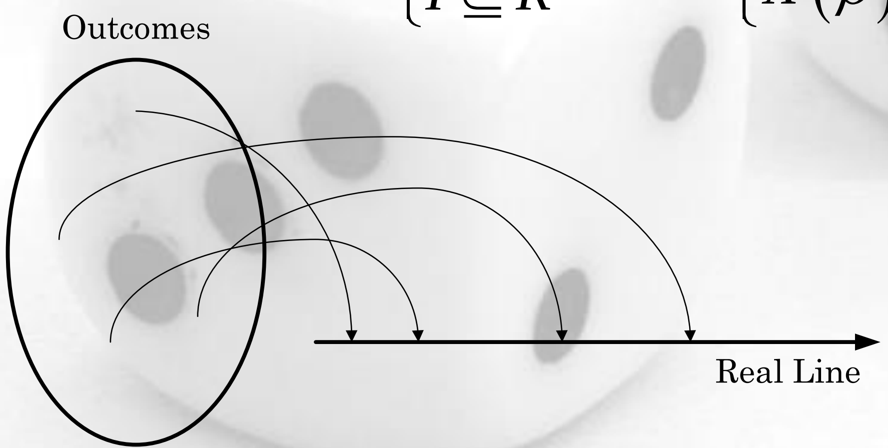
Random Variables

- Probability Space
  - A triple of (Ω, F, P)
    - Ω represents a nonempty set, whose elements are sometimes known as *outcomes* or *states of nature*
    - F represents a set, whose elements are called *events*. The events are subsets of Ω. F should be a “Borel Field”.
    - P represents the probability measure.

- Fact: $P(\Omega)=1$

Random Variables (Cont'd)

- Random variable is a *"function" ("mapping")* from a set of possible outcomes of the experiment to an interval of real (complex) numbers.
- In other words : $\begin{cases} F \subseteq P(\Omega) \\ I \subseteq R \end{cases} : \begin{cases} X : F \to I \\ X(\beta) = r \end{cases}$



Random Variables (Cont'd)

- Example I :
  - Mapping faces of a dice to the first six natural numbers.
- Example II :
  - Mapping height of a man to the real interval (0,3] (meter or something else).
- Example III :
  - Mapping success in an exam to the discrete interval [0,20] by quantum 0.1 .

Random Variables (Cont'd)

- Random Variables
  - Discrete
    - Dice, Coin, Grade of a course, etc.
  - Continuous
    - Temperature, Humidity, Length, etc.

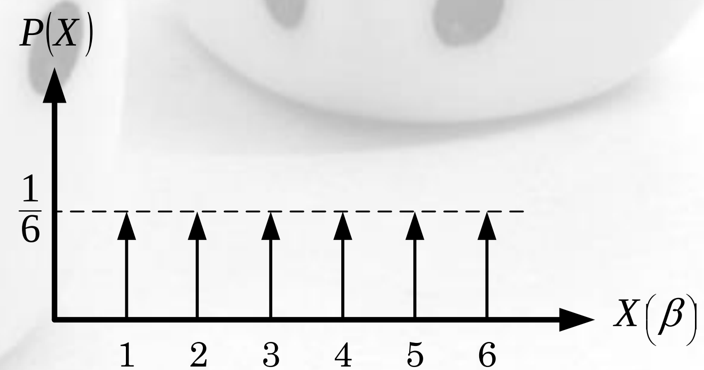
- Random Variables
  - Real
  - Complex

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Correlation
- Important Theorems

Density/Distribution Functions

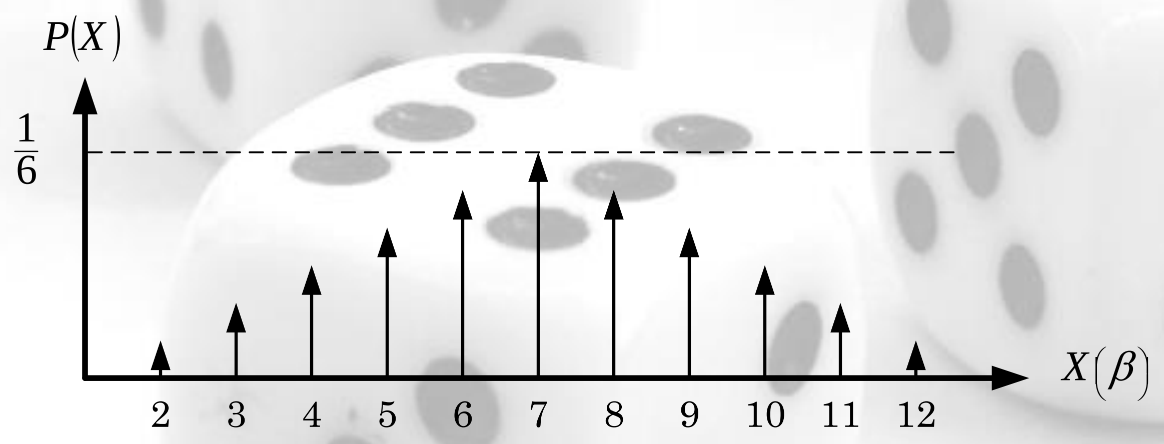
- Probability Mass Function (PMF)
  - Discrete random variables
  - Summation of impulses
  - The magnitude of each impulse represents the probability of occurrence of the outcome
- Example I:
  - Rolling a fair dice



$$PMF = \frac{1}{6}\sum_{i=1}^{6}\delta(X-i)$$

Density/Distribution Functions (Cont'd)

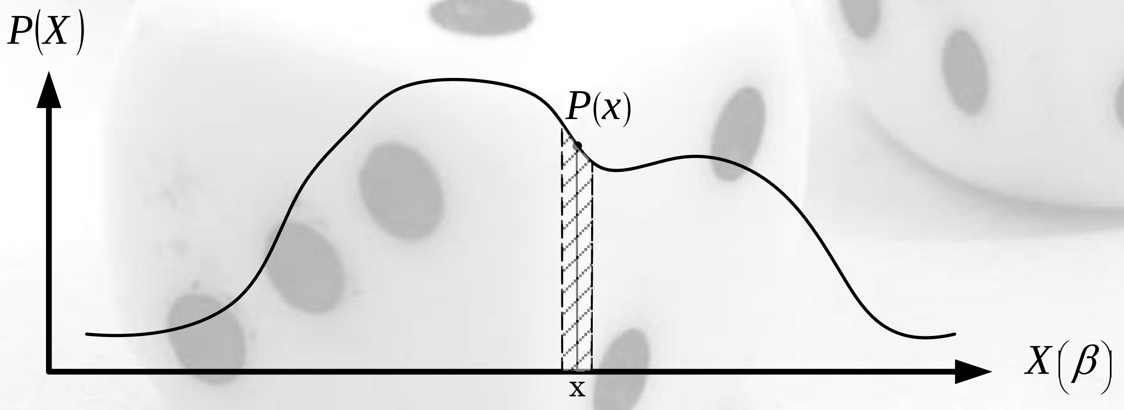
- Example II:
  - Summation of two fair dices



- Note : Summation of all probabilities should be equal to ONE. (Why?)

Density/Distribution Functions (Cont'd)

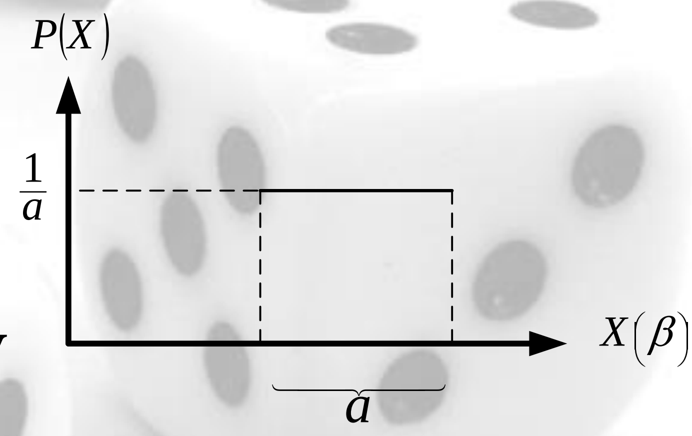
- Probability Density Function (PDF)
  - Continuous random variables
  - The probability of occurrence of $x_0 \in \left(x - \frac{dx}{2}, x + \frac{dx}{2} \right)$ will be $P(x).dx$



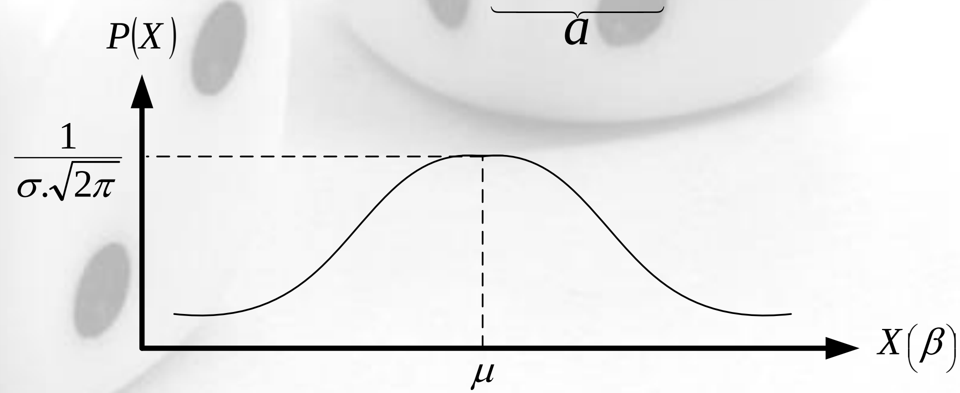
Density/Distribution Functions (Cont'd)

- Some famous masses and densities
  - Uniform Density

$$f(x)=\frac{1}{a}.\left(U(end)-U(begin)\right)$$



  - Gaussian (Normal) Density

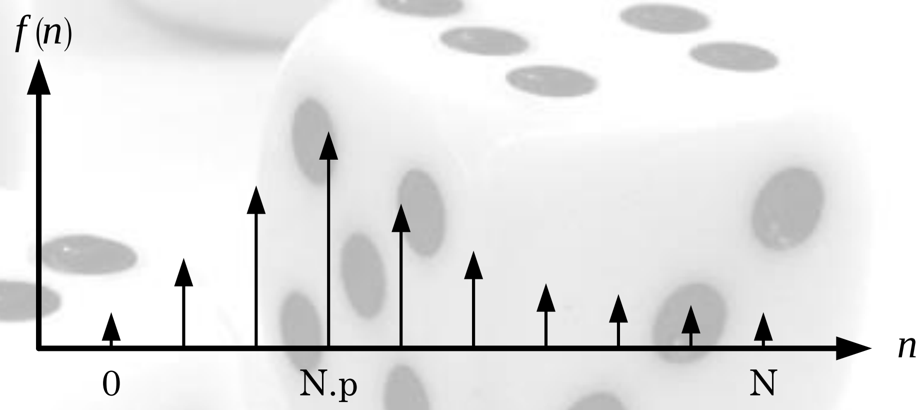


$$f(x)=\frac{1}{\sigma.\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2.\sigma^2}}=N(\mu,\sigma)$$

Density/Distribution Functions (Cont'd)

- Binomial Density

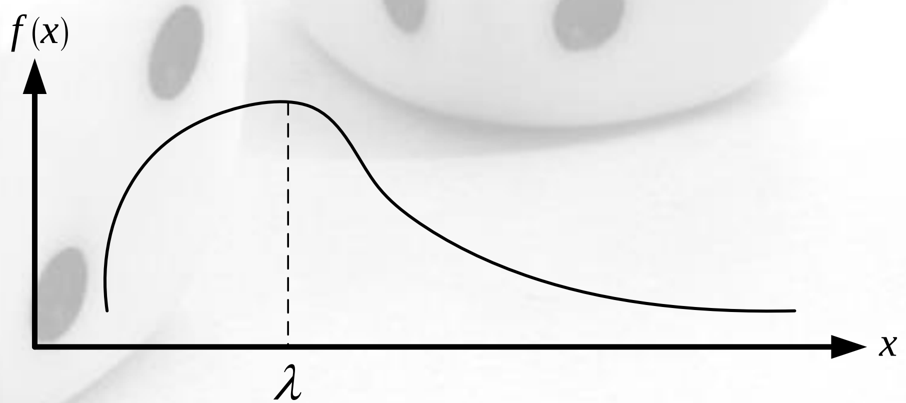
$$f(n)=\binom{N}{n}.(1-p)^n.p^{N-n}$$



- Poisson Density

$$f(x)=e^{-\lambda}\frac{\lambda^x}{\Gamma(x+1)}$$

$$Note: x\in\aleph \Rightarrow \Gamma(x+1)=x!$$



Important Fact: $For\,Sufficiently\,large\,N: \binom{N}{n}.(1-p)^{N-n}.p^n \approx e^{-N.p}\frac{(N.p)^n}{n!}$

Density/Distribution Functions (Cont'd)

- Cauchy Density

$$f(x)=\frac{1}{\pi}\times\frac{\gamma}{(x-\mu)^2+\gamma^2}$$

- Weibull Density

$$f(x)=\frac{k}{\lambda}\times\left(\frac{x}{\lambda}\right)^{k-1}\times e^{-\left(\frac{x}{\lambda}\right)^{k}}$$

Density/Distribution Functions (Cont'd)

- Exponential Density

$$f(x) = \lambda.e^{-\lambda x}.U(x) = \begin{cases} \lambda.e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Rayleigh Density

$$f(x) = \frac{x.e^{-\frac{x^2}{2\sigma^2}}}{\sigma^2}$$

Density/Distribution Functions (Cont'd)

- Expected Value
  - The most likelihood value

$$E[X] = \int_{-\infty}^{\infty} x . f_X(x) dx$$

  - Linear Operator

$$E[a.X + b] = a.E[X] + b$$

- Function of a random variable
  - Expectation

$$E[g(X)] = \int_{-\infty}^{\infty} g(x). f_X(x) dx$$

Density/Distribution Functions (Cont'd)

- PDF of a function of random variables
  - Assume RV "Y" such that $Y = g(X)$
  - The inverse equation $X = g^{-1}(Y)$ may have more than one solution called $X_1, X_2, \ldots, X_n$
  - PDF of "Y" can be obtained from PDF of "X" as follows

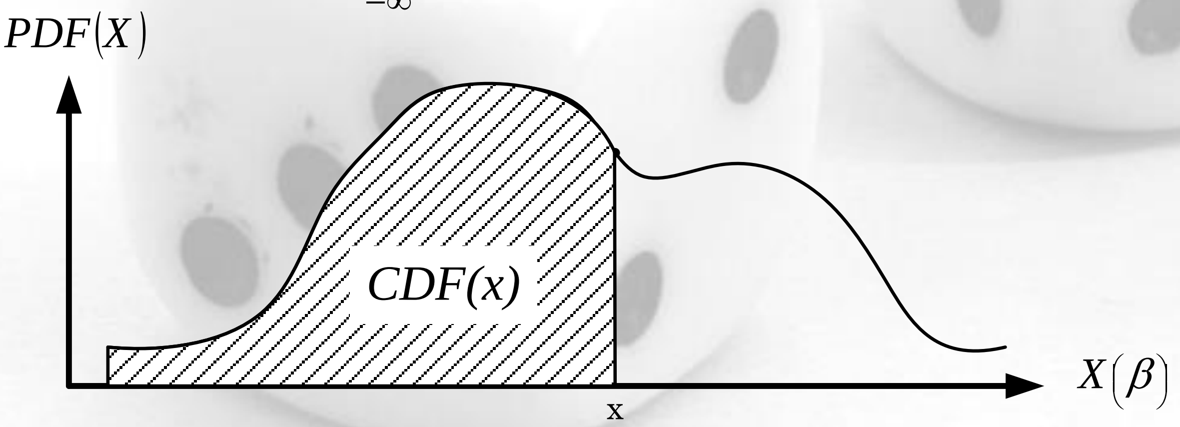
$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{absolute\ value\left(\frac{d}{dx} g(x)\Big|_{x=x_i}\right)}$$

Density/Distribution Functions (Cont'd)

- Cumulative Distribution Function (CDF)
  - Both Continuous and Discrete
  - Could be defined as the integration of PDF

$$CDF(x) = F_X(x) = P(X \leq x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(x).dx$$



Density/Distribution Functions (Cont'd)

- Some CDF properties
  - Non-decreasing
  - Right Continuous
  - F(-infinity) = 0
  - F(infinity) = 1

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Correlation
- Important Theorems

Joint/Conditional Distributions

- Joint Probability Functions
  - Density
  - Distribution

$$F_{X,Y}(x,y) = P(X \le x \ \ and \ \ Y \le y)$$
$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y)\,dydx$$

- Example I
  - In a rolling fair dice experiment represent the outcome as a 3-bit digital number "xyz".

$$f_{X,Y}(x,y) = \begin{cases} 1/6 & x=0; y=0 \\ 1/3 & x=0; y=1 \\ 1/3 & x=1; y=0 \\ 1/6 & x=1; y=1 \\ 0 & O.W. \end{cases}$$

$$\begin{array}{l} \quad xyz \\ 1 \to 001 \\ 2 \to 010 \\ 3 \to 011 \\ 4 \to 100 \\ 5 \to 101 \\ 6 \to 110 \end{array}$$

Joint/Conditional Distributions (Cont'd)

- Example II
  - Two normal random variables

$$f_{X,Y}(x,y)=\frac{1}{2\pi.\sigma_x.\sigma_y.\sqrt{1-r^2}}e^{-\left(\frac{1}{2\left(1-r^2\right)}\left(\frac{\left(x-\mu_x\right)^2}{\sigma_x{}^2}+\frac{\left(y-\mu_y\right)^2}{\sigma_y{}^2}-\frac{2r\left(x-\mu_x\right)\left(y-\mu_y\right)}{\sigma_x.\sigma_y}\right)\right)}$$

  - What is "r" ?

- Independent Events (Strong Axiom)

$$f_{X,Y}(x,y)=f_X(x).f_Y(y)$$

Joint/Conditional Distributions (Cont'd)

- Obtaining one variable **density** functions

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$$

- **Distribution** functions can be obtained just from the density functions. (How?)

Joint/Conditional Distributions (Cont'd)

- Conditional Density Function
  - Probability of occurrence of an event if another event is observed (we know what "Y" is).

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

  - Bayes' Rule

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x).f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x).f_X(x)dx}$$

Joint/Conditional Distributions (Cont'd)

- Example I
  - Rolling a fair dice
    - X : the outcome is an even number
    - Y : the outcome is a prime number

$$P(X|Y)=\frac{P(X,Y)}{P(Y)}=\frac{1/6}{1/2}=\frac{1}{3}$$

- Example II
  - Joint normal (Gaussian) random variables

$$f_{X|Y}(x|y)=\frac{1}{\sqrt{2\pi}.\sigma_x.\sqrt{1-r^2}}e^{-\left(\frac{1}{2(1-r^2)}\left(\frac{x-\mu_x}{\sigma_x}-r\times\frac{y-\mu_y}{\sigma_y}\right)^2\right)}$$

Joint/Conditional Distributions (Cont'd)

- Conditional Distribution Function

$$F_{X|Y}(x|y) = P(X \le x \ \ while \ \ Y = y)$$

$$= \int_{-\infty}^{x} f_{X|Y}(x|y)dx$$

$$= \frac{\int_{-\infty}^{x} f_{X,Y}(t,y)dt}{\int_{-\infty}^{\infty} f_{X,Y}(t,y)dt}$$

- Note that "y" is a constant during the integration.

Joint/Conditional Distributions (Cont'd)

- Independent Random Variables

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$= \frac{f_X(x).f_Y(y)}{f_Y(y)}$$

$$= f_X(x)$$

- Remember! Independency is NOT heuristic.

Joint/Conditional Distributions (Cont'd)

- PDF of a functions of joint random variables
  - Assume that $(U,V)=g(X,Y)$
  - The inverse equation set $(X,Y)=g^{-1}(U,V)$ has a set of solutions $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$
  - Define Jacobean matrix as follows $J=\begin{bmatrix} \frac{\partial}{\partial X}U & \frac{\partial}{\partial X}V \\ \frac{\partial}{\partial X}U & \frac{\partial}{\partial Y}V \end{bmatrix}$
  - The joint PDF will be

$$f_{U,V}(u,v)=\sum_{i=1}^{n}\frac{f_{X,Y}(x_i,y_i)}{absolute\ determinant\left(J\big|_{(x,y)=(x_i,y_i)}\right)}$$

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Correlation
- Important Theorems

Correlation

- Knowing about a random variable "X", how much information will we gain about the other random variable "Y" ?
- Shows linear similarity
- More formal: $Crr(X,Y) = E[X.Y]$
- Covariance is normalized correlation

$$Cov(X,Y) = E[(X-\mu_X).(Y-\mu_Y)] = E[X.Y] - \mu_X.\mu_Y$$

Correlation (cont'd)

- Variance
  - Covariance of a random variable with itself

$$Var(X) = \sigma_X{}^2 = E\left[(X - \mu_X)^2\right]$$

- Relation between correlation and covariance

$$E\left[X^2\right] = \sigma_X{}^2 + \mu_X{}^2$$

- Standard Deviation
  - Square root of variance

Correlation (cont'd)

- Moments
  - n^{th} order moment of a random variable "X" is the expected value of "X^n"

$$M_n = E\left(X^n\right)$$

  - Normalized form

$$M_n = E\left(\left(X - \mu_X\right)^n\right)$$

- Mean is first moment
- Variance is second moment added by the square of the mean

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Correlation
- Important Theorems

Important Theorems

- **Central limit theorem**
  - Suppose i.i.d. (Independent Identically Distributed) RVs "X_k" with finite variances
  - Let $S_n = \sum_{i=1}^{n} a_n . X_n$
  - PDF of "S_n" converges to a normal distribution as $\boldsymbol{n}$ increases, regardless to the density of RVs.
- Exception : Cauchy Distribution (Why?)

Important Theorems (cont'd)

- **Law of Large Numbers (Weak)**
  - For i.i.d. RVs "X_k"

$$\forall_{\varepsilon>0} \quad \lim_{n\to\infty} \Pr\left\{\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu_X\right| > \varepsilon\right\} = 0$$

Important Theorems (cont'd)

- **Law of Large Numbers (Strong)**
  - For i.i.d. RVs "X_k"

$$\Pr\left\{ \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu_X \right\} = 1$$

- Why this definition is stronger than before?

Important Theorems (cont'd)

- **Chebyshev's Inequality**
  - Let "X" be a nonnegative RV
  - Let "c" be a positive number

$$\Pr\{X > c\} \leq \frac{1}{c} E[X]$$

- Another form:

$$\Pr\{|X - \mu_X| > \varepsilon\} \leq \frac{\sigma_X{}^2}{\varepsilon^2}$$

- It could be rewritten for negative RVs. (How?)

Important Theorems (cont'd)

- **Schwarz Inequality**
  - For two RVs "X" and "Y" with finite second moments

$$\mathrm{E}[X.Y]^2 \leq \mathrm{E}\left[X^2\right].\mathrm{E}\left[Y^2\right]$$

  - Equality holds in case of linear dependency.

Acknowledgement

- Thanks to Mr. Jalali for preparing slides

Next Lecture

Elements of Stochastic Processes