

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 (20th) :
 - 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 ($L(\theta)$) maximize parameter/data
- For Gaussian distribution the $L(\theta)$ 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 $p(X/\theta)$ to update the posterior distribution $p(\theta/X)$.
- Mean of the posterior, $p(\theta/X)$, 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 $p = P(\text{heads})$ 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 : $\pi(p|\text{Observation}) = p^{800}(1-p)^{200}$

□ Frequentist Analysis

- Answer is an estimator \hat{p} such that
 - Mean : $E[\hat{p}] = 0.8$
 - Confidence Interval : $P(0.774 \leq \hat{p} \leq 0.826) \geq 0.95$



History & Philosophy (Cont'd)

□ Further reading:

- <http://www.leidenuniv.nl/fsw/verduin/stat/hist/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.

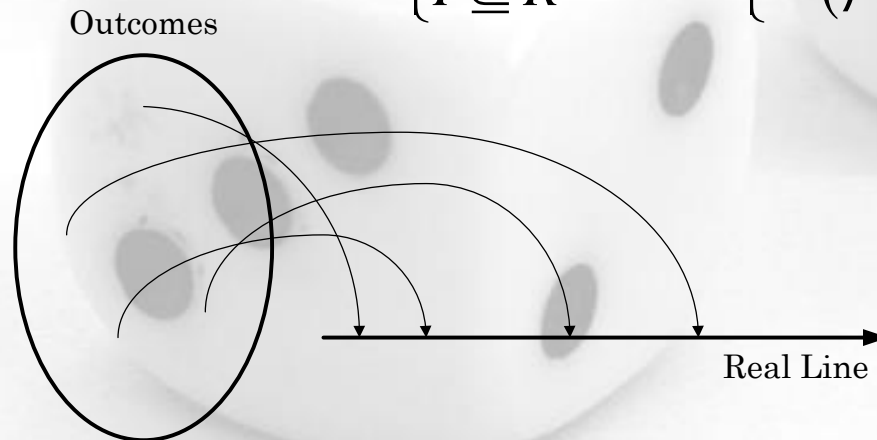
□ Fact: $P(\Omega) = 1$



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 :
$$\begin{cases} F \subseteq P(\Omega) \\ I \subseteq R \end{cases} : \begin{cases} X : F \rightarrow I \\ X(\beta) = r \end{cases}$$



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

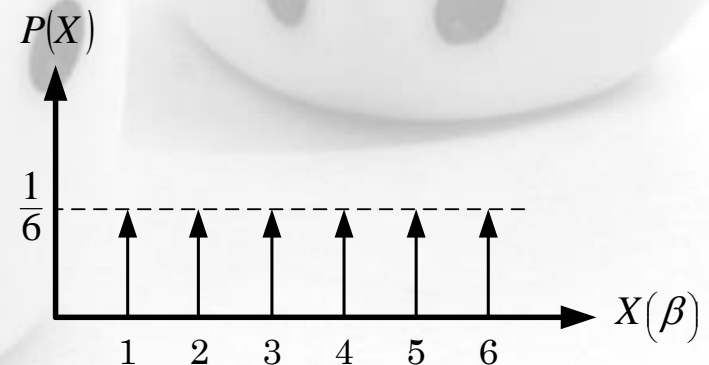
- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Correlation
- Important Theorems



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



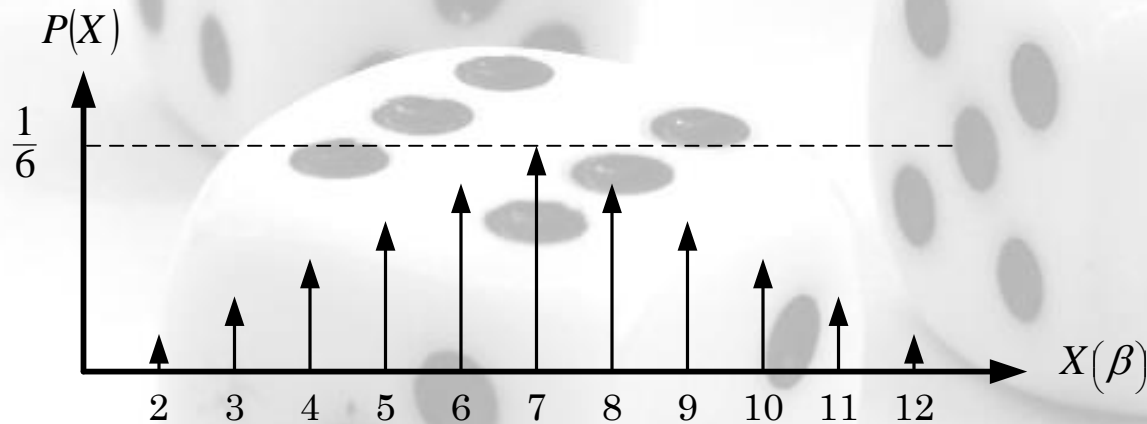
$$PMF = \frac{1}{6} \sum_{i=1}^6 \delta(X - i)$$



Density/Distribution Functions (Cont'd)

□ Example II:

- Summation of two fair dices



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

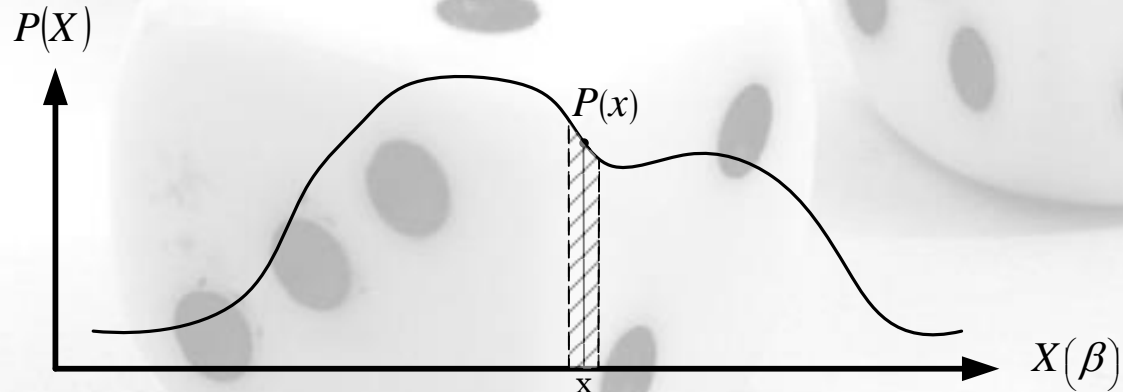


Density/Distribution Functions (Cont'd)

□ 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 $P(x).dx$

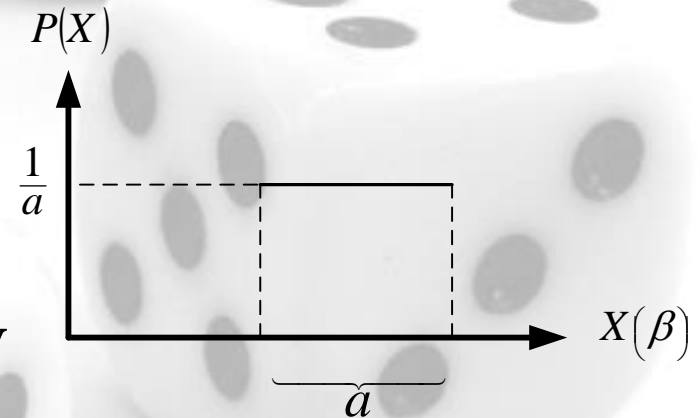


Density/Distribution Functions (Cont'd)

□ Some famous masses and densities

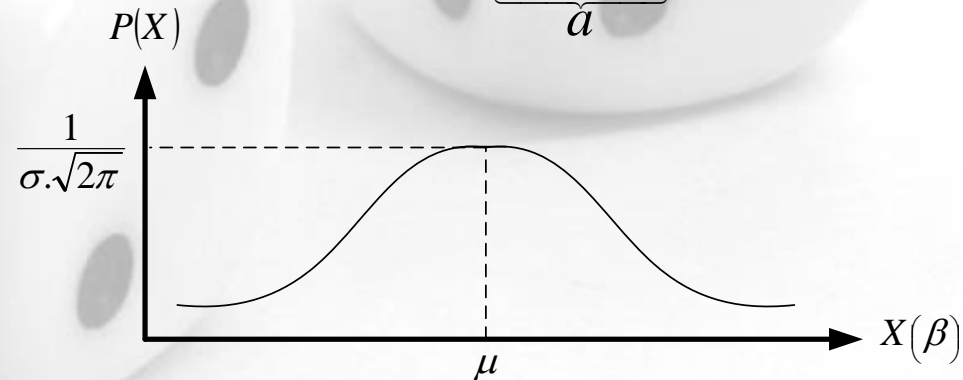
■ Uniform Density

$$f(x) = \frac{1}{a} \cdot (U(\text{end}) - U(\text{begin}))$$



■ Gaussian (Normal) Density

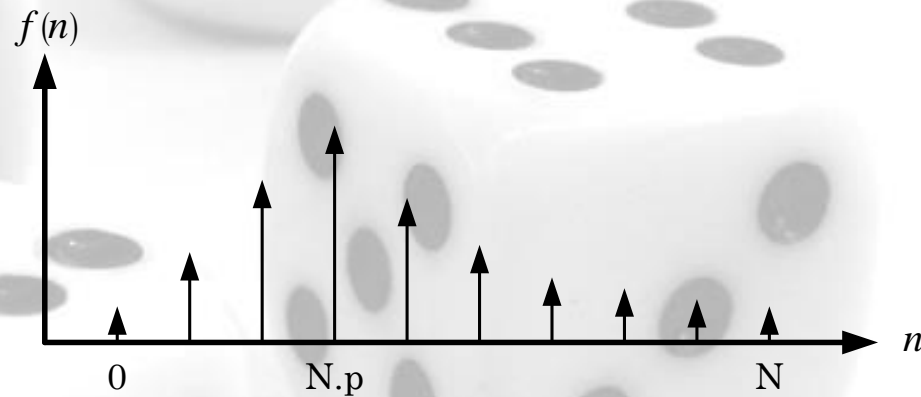
$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = N(\mu, \sigma)$$



Density/Distribution Functions (Cont'd)

■ Binomial Density

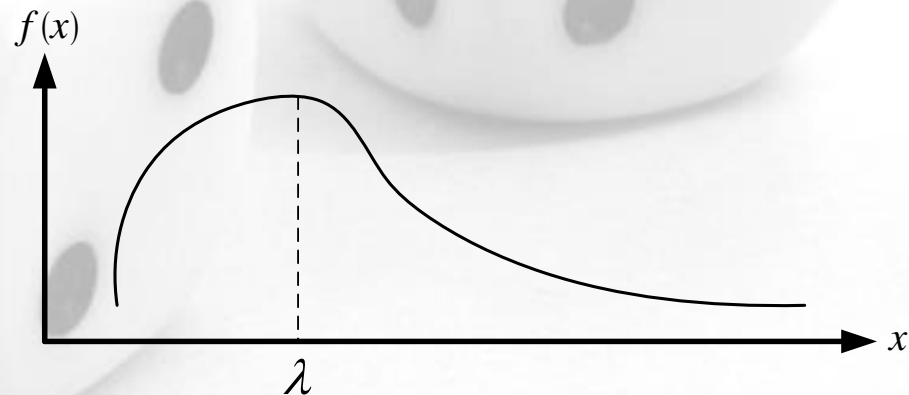
$$f(n) = \binom{N}{n} \cdot (1-p)^n \cdot p^{N-n}$$



■ Poisson Density

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

Note: $x \in \mathbb{N} \Rightarrow \Gamma(x+1) = x!$



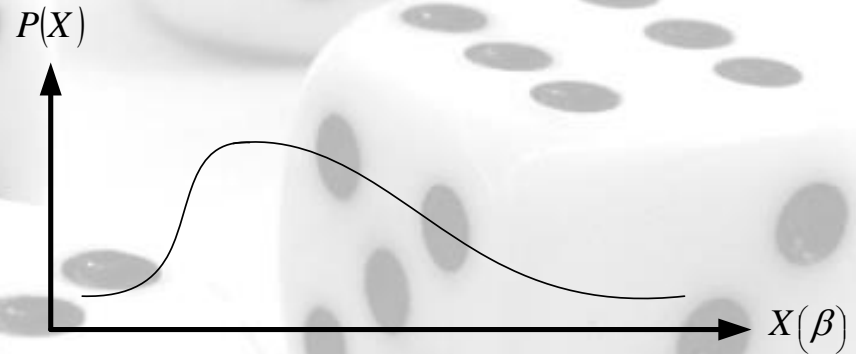
Important Fact: For Sufficiently large N : $\binom{N}{n} \cdot (1-p)^{N-n} \cdot p^n \approx e^{-N.p} \frac{(N.p)^n}{n!}$



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

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

- Rayleigh Density

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



Density/Distribution Functions (Cont'd)

□ Expected Value

- The most likelihood value

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

- Linear Operator

$$E[a \cdot X + b] = a \cdot E[X] + b$$

□ Function of a random variable

- Expectation

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



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, \dots, x_n
 - PDF of “Y” can be obtained from PDF of “X” as follows

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

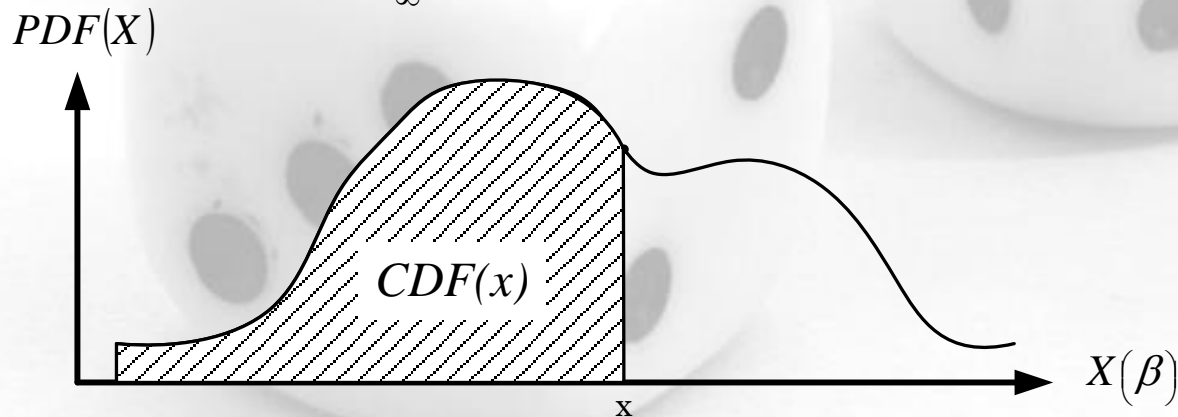


Density/Distribution Functions (Cont'd)

- 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
 - $F(-\infty) = 0$
 - $F(\infty) = 1$



Outline

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



Joint/Conditional Distributions

□ Joint Probability Functions

- Density
- Distribution

$$F_{X,Y}(x, y) = P(X \leq x \text{ and } Y \leq y)$$

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

□ Example I

- In a rolling fair dice experiment represent the outcome as a 3-bit digital number “xyz”.

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

	xyz
1	→ 001
2	→ 010
3	→ 011
4	→ 100
5	→ 101
6	→ 110



Joint/Conditional Distributions (Cont'd)

□ Example II

- Two normal random variables

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

- What is “r” ?

□ Independent Events (Strong Axiom)

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



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

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



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

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



Joint/Conditional Distributions (Cont'd)

□ Conditional Distribution Function

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

$$\begin{aligned} &= \int_{-\infty}^x f_{X|Y}(x|y) dx \\ &= \frac{\int_{-\infty}^x f_{X,Y}(t, y) dt}{\int_{-\infty}^{\infty} f_{X,Y}(t, y) dt} \end{aligned}$$

- Note that “y” is a constant during the integration.
-



Joint/Conditional Distributions (Cont'd)

- Independent Random Variables

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{f_X(x) \cdot f_Y(y)}{f_Y(y)} \\ &= f_X(x) \end{aligned}$$

- 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 $(X_1,Y_1), (X_2,Y_2), \dots, (X_n,Y_n)$

■ Define Jacobean matrix as follows $J =$

$$\begin{bmatrix} \frac{\partial}{\partial X} U & \frac{\partial}{\partial X} V \\ \frac{\partial}{\partial Y} U & \frac{\partial}{\partial Y} V \end{bmatrix}$$

■ The joint PDF will be

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



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: $Crr(X, Y) = E[X \cdot Y]$

□ **Covariance** is normalized correlation

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



Correlation (cont'd)

□ Variance

- Covariance of a random variable with itself

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

□ Relation between correlation and covariance

$$E[X^2] = \sigma_X^2 + \mu_X^2$$

□ 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 ”

$$M_n = E(X^n)$$

- Normalized form

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

- 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 $S_n = \sum_{i=1}^n a_n \cdot X_n$
- PDF of “ S_n ” converges to a **normal distribution** as 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 ”

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



Important Theorems (cont'd)

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

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

- 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

$$E[X.Y]^2 \leq E[X^2].E[Y^2]$$

- Equality holds in case of **linear dependency**.



Acknowledgement

- Thanks to Mr. Jalali for preparing slides



Next Lecture

Elements of Stochastic Processes

