Stochastic Processes

Week 01 (version 1.1)

Review of Probability

Introduction to Stochastic Processes

Hamid R. Rabiee

Fall 2025

Outline of Week 01 Lectures

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Linear Correlation
- Important Theorems
- Introduction to Stochastic Processes

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 (On Reasons in the Game of Chance)*”, *In Latin, 1657.*

History & Philosophy (Cont’d)

- Rapid development during 18th Century
- Major Contributions:
  - J. Bernoulli (1654-1705)
  - A. De Moivre (1667-1754)
- 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
- Modern theory of probability (20th Century):
  - A. Kolmogorov : an Axiomatic approach
- First modern book:
  - A. Kolmogorov, "Foundations of Probability Theory", Chelsea, New York, 1950.
- Other giants in the field:
  - Chebyshev, Markov and Kolmogorov

History & Philosophy (Cont'd)

- Two major philosophies:
  - Frequentist Philosophy
    - Observations (dataset X) are sufficient to obtain probability density functions $f(X|\theta)$! (assume θ is fixed but unknown)
  - Bayesian Philosophy:
    - Observations are NOT sufficient but are necessary!
    - Prior knowledge of the parameters (θ) of probability density functions is also essential!

    (assume θ is random but unknown)

History & Philosophy (Cont'd)

Frequentist philosophy

- Parameters (θ) of f(X|θ) are fixed
- There is an underlying distribution from which observations X are drawn
- Likelihood functions f(X|θ) is used for any inference on X
- For Gaussian likelihood function, the most likely for X is happens to be $(1/N)\sum_i x_i$ (sample average)

Bayesian philosophy

- Parameters (θ) of f(X|θ) are random
- Variation of the parameters defined by the prior probability $\pi(\theta)$
- This is combined with f(X|θ) to obtain the posterior f(θ|X)
- Mean of the posterior, f(θ|X) can be considered a point estimate for 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 (believe): $\pi(p) = 1$ (Uniform(0,1))
  - Our posterior knowledge: $\pi(p|Observation) = p^{800}(1-p)^{200}$
- 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)

Nowadays, Probability Theory is considered to be a part of the Measure Theory!

- 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
- Linear Correlation
- Important Theorems
- Introduction to Stochastic Processes

Random Variables

- Borel Field
  - Sets are fundamental in probability theory because they provide the structure for organizing, describing, and calculating the likelihood of different outcomes and events.
  - A **Borel field** (also called a Borel σ-field or Borel σ-algebra) in mathematics, particularly in measure theory and topology, is the smallest collection of subsets of a topological space that includes all open sets and is closed under countable unions, countable intersections, and complements.

Random Variables

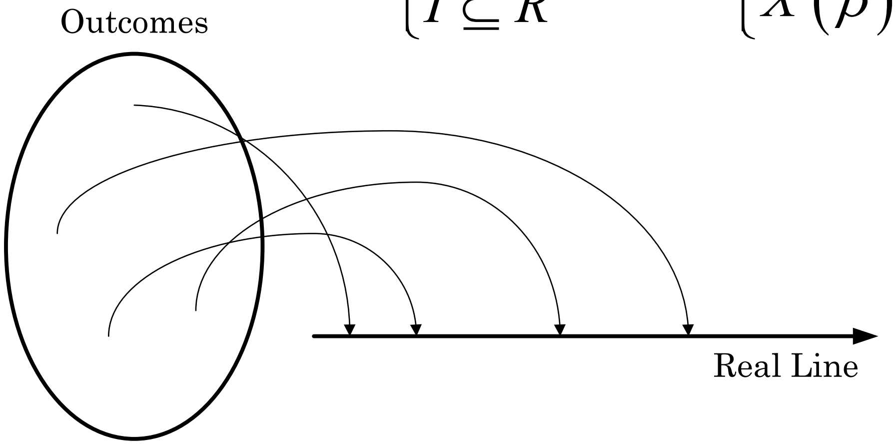
- Probability Space
  - A triple of (Ω, F, P)
    - Ω represents a nonempty set, whose elements are sometimes known as outcomes or states of nature (Sample Space).
    - F represents a set, whose elements are called events. The events are subsets of Ω. F should be a "Borel Field".
    - P represents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

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

Random Variables (Cont'd)

- Random Variable (RV) 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 0.1 increments.

Random Variables (Cont’d)

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

- Random Variables
  - Real
  - Complex

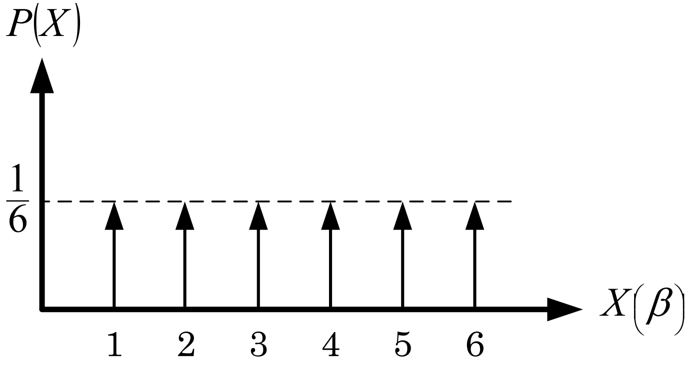
Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Linear Correlation
- Important Theorems
- Introduction to Stochastic Processes

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



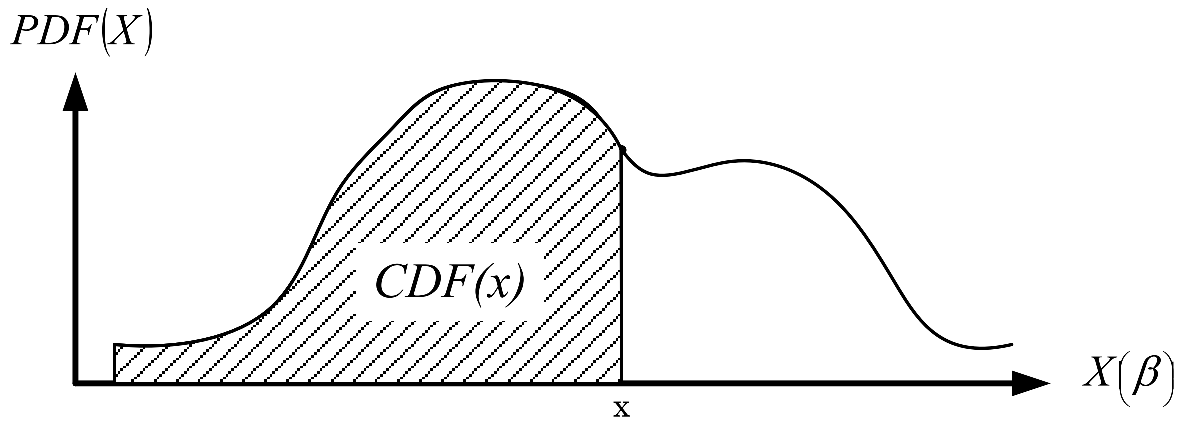
$$PMF = \frac{1}{6}\sum_{i=1}^{6}\delta(X-i)$$

Density/Distribution Functions (Cont'd)

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

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

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

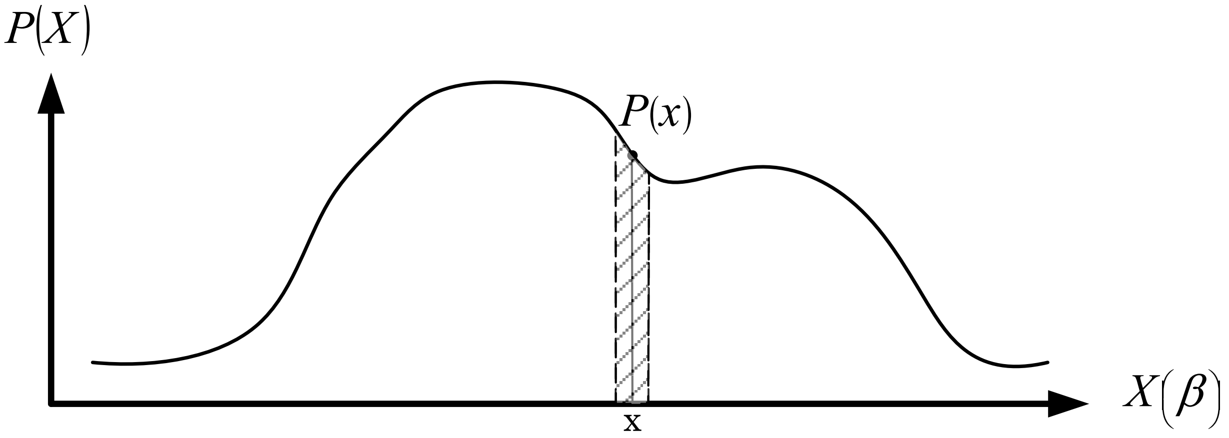


Density/Distribution Functions (Cont’d)

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

Density/Distribution Functions (Cont'd)

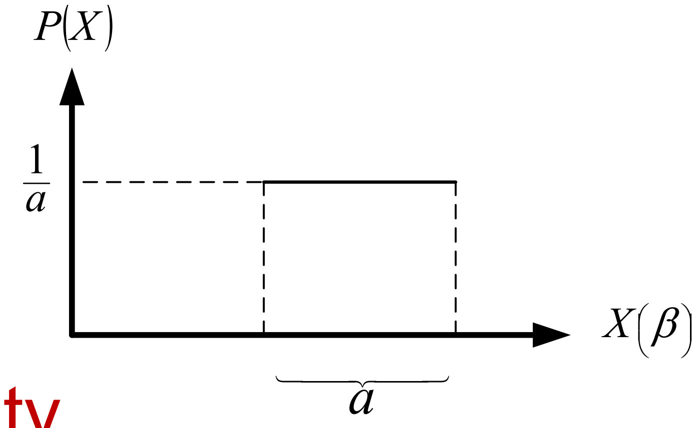
- Probability Density Function (PDF)
  - Continuous random variables
  - The probability of occurrence of $x_0 \in \left(x - \frac{dx}{2}, x + \frac{dx}{2} \right)$ will be $P(x)\,dx$



Density/Distribution Functions (Cont'd)

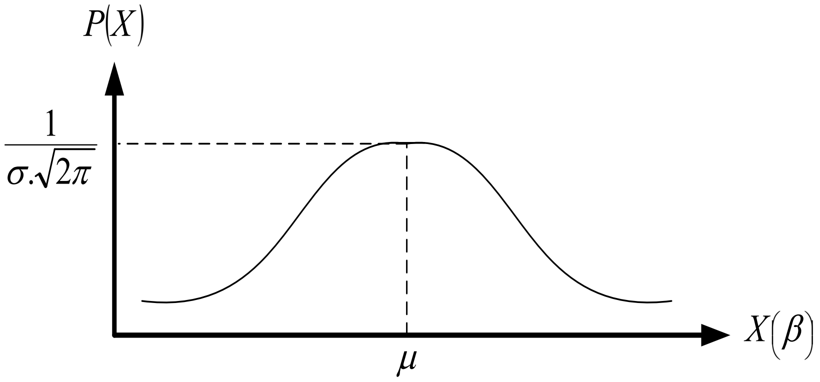
- Some famous masses and densities:
  - Uniform Density

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



  - Gaussian (Normal) Density

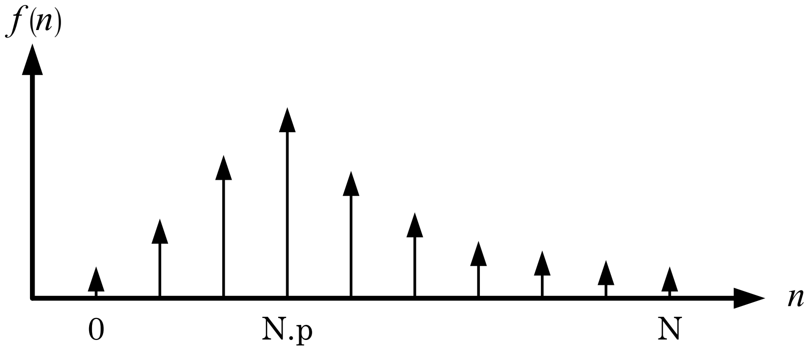
$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2.\sigma^2}} = N(\mu, \sigma)$$



Density/Distribution Functions (Cont'd)

- Binomial Density

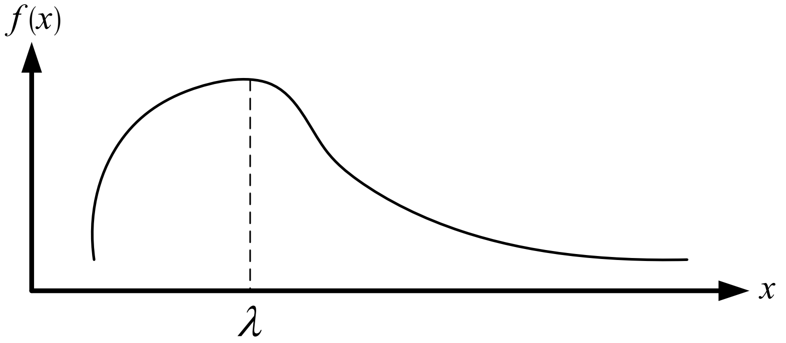
$$f(n) = \binom{N}{n}.(1-p)^{n}.p^{N-n}$$



- Poisson Density

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

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



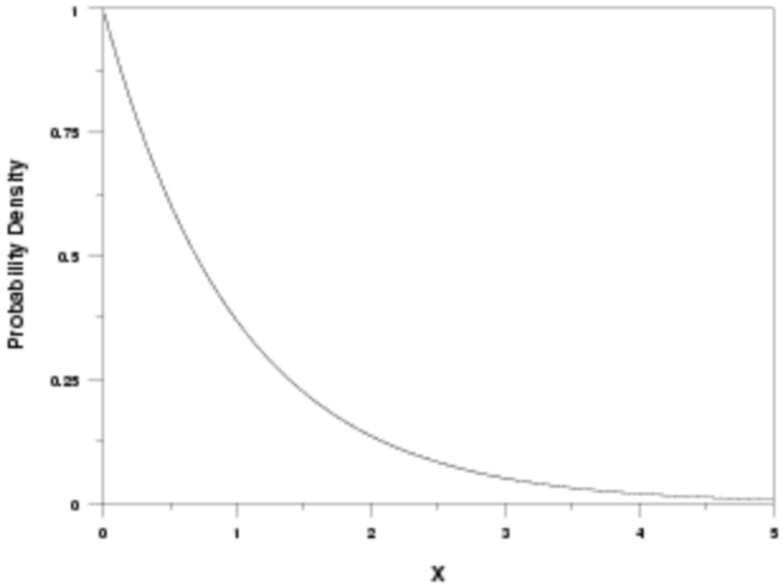
- Important Fact:

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

Density/Distribution Functions (Cont'd)

- Exponential Density

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



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:

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

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Linear Correlation
- Important Theorems
- Introduction to Stochastic Processes

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".

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

xyz

$1 \rightarrow 001$

$2 \rightarrow 010$

$3 \rightarrow 011$

$4 \rightarrow 100$

$5 \rightarrow 101$

$6 \rightarrow 110$

Joint/Conditional Distributions (Cont'd)

- Example II:
  - Two normal (Gaussian) 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).

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

  - Bayes' Rule:

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

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:

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

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:

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

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
- Linear Correlation
- Important Theorems
- Introduction to Stochastic Processes

Correlation

- Knowing about a random variable “X”, how much information will we gain about the other random variable “Y”?
- Covariance: Measures the degree to which two variables change together (linear similarity):

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

- Correlation coefficient (ρ): This is a normalized version of covariance. It is dimensionless and ranges between -1 and 1.

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}$$

Correlation (cont’d)

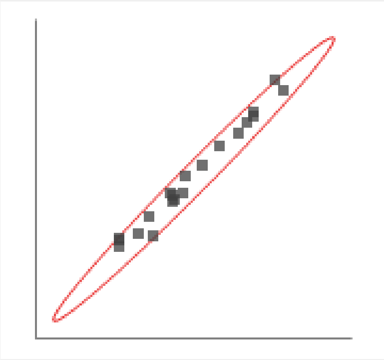
- If the random variables “X” and “Y” are uncorrelated then their covariance and correlation coefficient are zero, and:

$$E\,[XY] = E[X]\,E[Y]$$

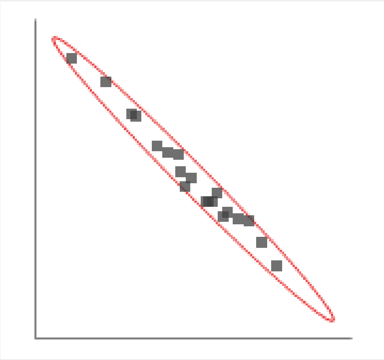
- It is important to note that while uncorrelated variables have zero covariance, this does not necessarily imply that they are independent.
- Independence is a stronger condition that implies no relationship (of any kind), not just linear.
- Consequently, independent random variables are also uncorrelated.

Correlation (cont'd)

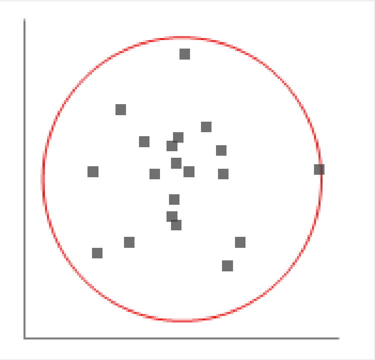
- A correlation scatter plot is a graph to visualize the relationship between two numerical variables.



Positive Correlation



Negative Correlation



No Correlation

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

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Linear Correlation
- Important Theorems
- Introduction to Stochastic Processes

Important Theorems

- **Central limit theorem (CLT)**
  - Consider i.i.d. (Independent Identically Distributed) RVs “X_k” with finite variances
  - Let $S_n = \sum_{i=1}^{n} a_n . X_n$
  - Then PDF of “S_n” converges to a normal distribution as *n* increases, regardless of the initial 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 the weak law of large numbers?

Important Theorems (cont'd)

- **Chebyshev's Inequality**
    - Let "X" be a nonnegative RV
    - Let "c" be a positive number, then: $\Pr\{X > c\} \leq \frac{1}{c} E[X]$
    - The term Chebyshev's inequality may also refer to Markov's inequality, especially in the context of analysis. They are closely related, and some authors refer to Markov's inequality as "Chebyshev's First Inequality,"
- Another form:

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

- This could also be rewritten for negative RVs.

Important Theorems (cont'd)

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

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

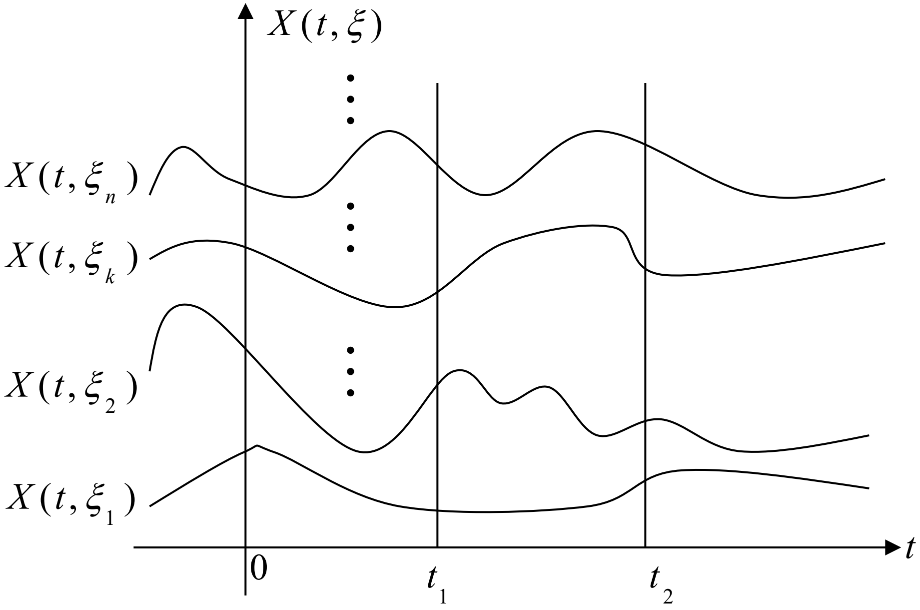
  - Equality holds in case of linear dependency.

Outline

- History/Philosophy
- Random Variables
- Density/Distribution Functions
- Joint/Conditional Distributions
- Linear Correlation
- Important Theorems
- Introduction to Stochastic Processes

Introduction to Stochastic Processes

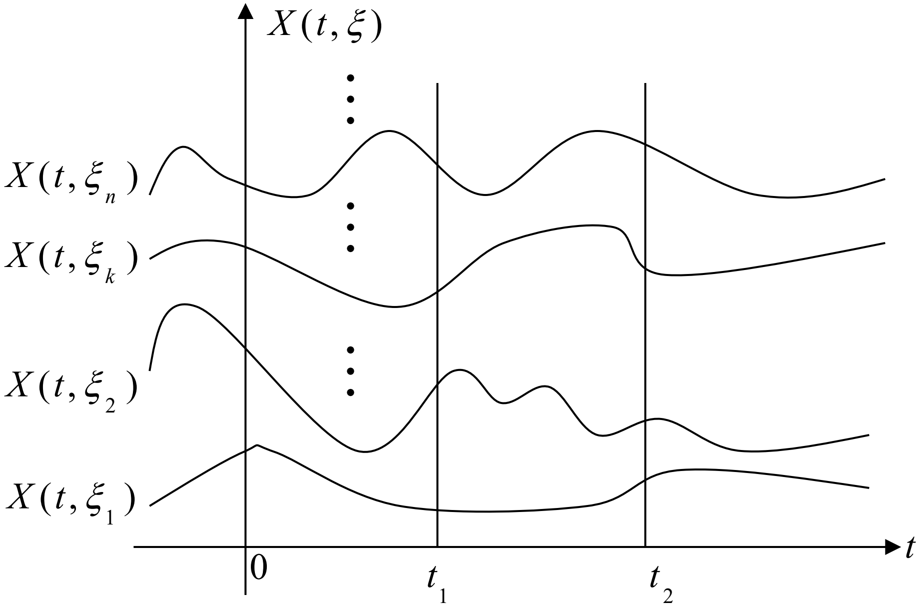
- Let ξ denote the random outcome of an experiment.
- To every such outcome suppose a function $X(t,\xi)$ is assigned.
- The collection of such functions form a stochastic process.
- The set of $\{\xi_k\}$ and the time index *t* can be continuous or discrete (countably infinite or finite).



- For fixed $\xi_i \in S$ (the set of all experimental outcomes), $X(t,\xi)$ is a specific time function.

Introduction to Stochastic Processes

- For fixed *t*, $X_1 = X(t_1, \xi_i)$ is a random variable.
- The ensemble of all such realizations $X(t,\xi)$ over time represents the stochastic process *X*(*t*).



Introduction to Stochastic Processes

- Examples:
- Let $X(t) = a\cos(\omega_0 t + \varphi)$, where φ is a uniformly distributed random variable in $(0, 2\pi)$, represents a stochastic process.
- Stochastic processes are everywhere:
  - stock market fluctuations
  - various queuing systems
  - Earthquake Signals
  - 1-D Audios
  - 2-D Images
  - 3-D Videos

Introduction to Stochastic Processes

- Example 1:

The Random Process (RP) $X(t)$ is defined as:

$X(t) = At + b$, b is a constant, A is a Gaussian rv, $t > 0$

Find $f_X(x,t)$: (use Slide 24)

$$f_A(a) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right) = N(0,1)$$

$$f_X(x,t) = \frac{f_A(a)}{\left|\frac{dx}{dA}\right|}$$

$$A = \frac{X(t) - b}{t} \qquad \left|\frac{dX}{dA}\right| = t, \qquad a = \frac{x - b}{t}$$

$$f_X(x,t) = \frac{1}{t} f_A(a) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{a^2}{2}\right) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-b)^2}{2t^2}\right)$$

Introduction to Stochastic Processes

- Example 1:

The Random Process (RP) $X(t)$ is defined as:
$X(t) = At + b$, b is a constant, A is a Gaussian rv, $t > 0$
What is mean and variance of $X(t)$?

Introduction to Stochastic Processes

- Example 1 continued:

Mean of $X(t)$:

$X(t) = At + b,$ A is $N(0,1)$
$E[X(t)] = E[At + b] = E[A]t + E[b] = 0 \times t + b = b$

Variance of $X(t)$:
$X(t)2 = A^2t^2 + b^2 + 2Abt$

$E[X(t)2] = E[A^2t^2 + b^2 + 2Abt] = E[A^2]t^2 + E[b^2] + E[A]\,2bt = 1 * t^2 + b^2 + 0 * 2bt$

$E[X(t)^2] = t^2 + b^2$
$Var(X[t]) = E[X(t)2] - (E[X(t)])^2 = t^2 + b^2 - b^2 = t^2$

Note: The mean of X(t) is constant, but its variance is a function of time t.

Introduction to Stochastic Processes

Example 2:

$X(t)$: RP

$$X(t) = \underset{\substack{\downarrow \\ \text{constant}}}{A} \cos(\underset{\substack{\downarrow \\ \text{constant}}}{w_0} \underset{\substack{\searrow \\ \text{index} \\ \text{(time)}}}{t} + \underset{\substack{\searrow \\ \text{RV: } Uniform(0, 2\pi)}}{\theta})$$

$a)\ PDF = ?$

$b)\ E[X(t)] = ?$

$c)\ Var[X(t)] = ?$

$$f_\theta(\theta) = \begin{cases} \dfrac{1}{2\pi} & \theta \in (0,2\pi] \\ else & 0 \end{cases}$$

Introduction to Stochastic Processes

Example 2 continued:

$$X(t) = A\cos(w_0 t + \theta) = \mathrm{X_t}(\theta)$$

$$f_X(x,t) = \sum_i \frac{f_\theta(\theta_i)}{\left|\frac{dX_t}{d\theta_i}\right|} = \frac{1}{2\pi} \frac{1[0 < \theta_i \leq 2\pi]}{\left|\frac{dX_t}{d\theta_i}\right|}$$

$A\cos(w_0 t + \theta_i) = x$ → has exactly 2 answers in $(0, 2\pi]$

$$\left|\frac{dX_t}{d\theta_i}\right| = |-A\sin(w_0 t + \theta_i)| = \sqrt{A^2 - X_t^2}$$

$$\rightarrow f_X(x,t) = \frac{2}{2\pi} \frac{1}{\sqrt{A^2 - x^2}} = \frac{1}{\pi\sqrt{A^2 - x^2}} \qquad |X| \leq A$$

Introduction to Stochastic Processes

Example 2 continued:

$X(t) = A\cos(w_0 t + \theta) = \mathrm{X_t}(\theta)$

$$E[X(t)] = E[A\cos(w_0 t + \theta)] = \mathrm{A}\int_0^{2\pi} \cos(w_0 t + \theta)\,\frac{1}{2\pi}\,d\theta = 0$$

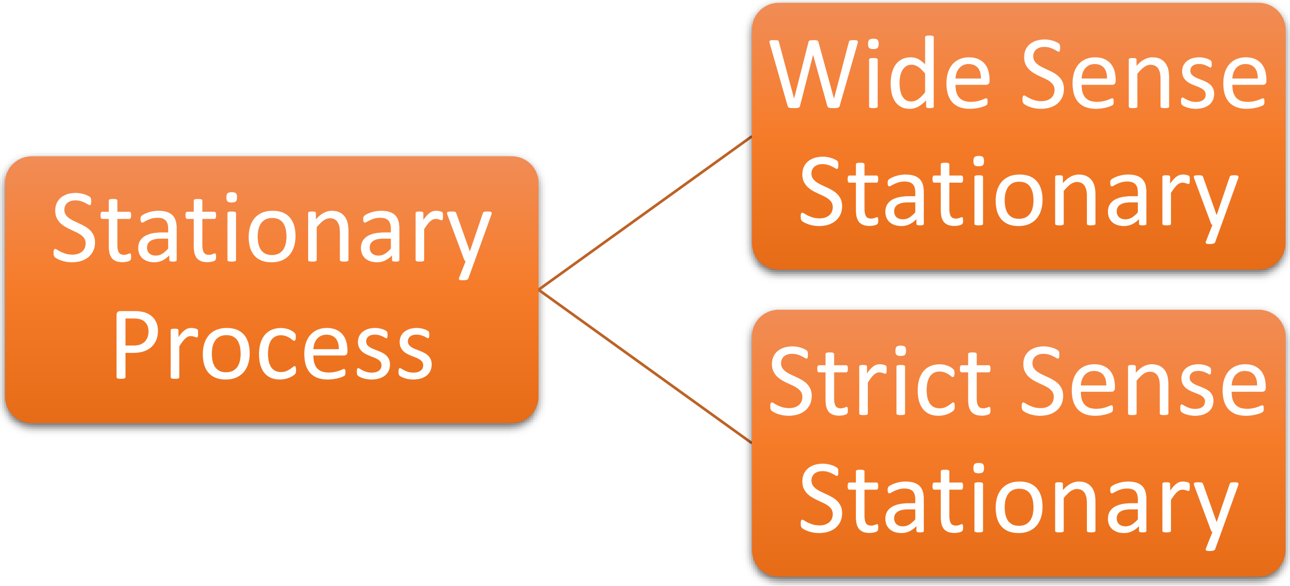
$$V[X(t)] = E[X(t)^2] - (E[X(t)])^2 = E[(A\cos(w_0 t + \theta))^2]$$

$$= \mathrm{A}^2 \int_0^{2\pi} \cos^2(w_0 t + \theta)\,\frac{1}{2\pi}\,d\theta$$

$$= \frac{A^2}{2\pi}\int_0^{2\pi} \frac{1}{2}(1 + \cos(2w_0 t + 2\theta))\,d\theta = \frac{A^2}{2\pi}\int_0^{2\pi} \frac{1}{2}\,d\theta = \frac{A^2}{2}$$

Note: The mean and variance of X(t) is constant in this example.

Introduction to Stochastic Processes
Stationary Processes
(Next Class)



Stationary Process

Wide Sense Stationary

Strict Sense Stationary

Next Week:

Stochastic Processes
Stationary Stochastic Processes

Have a good day!