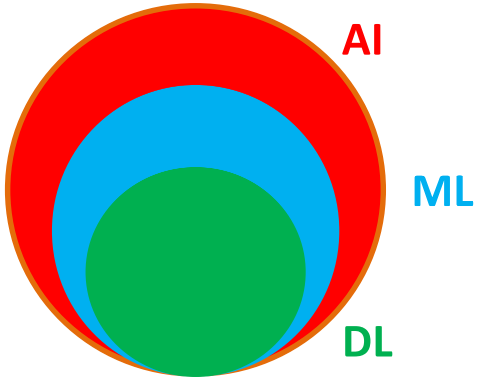
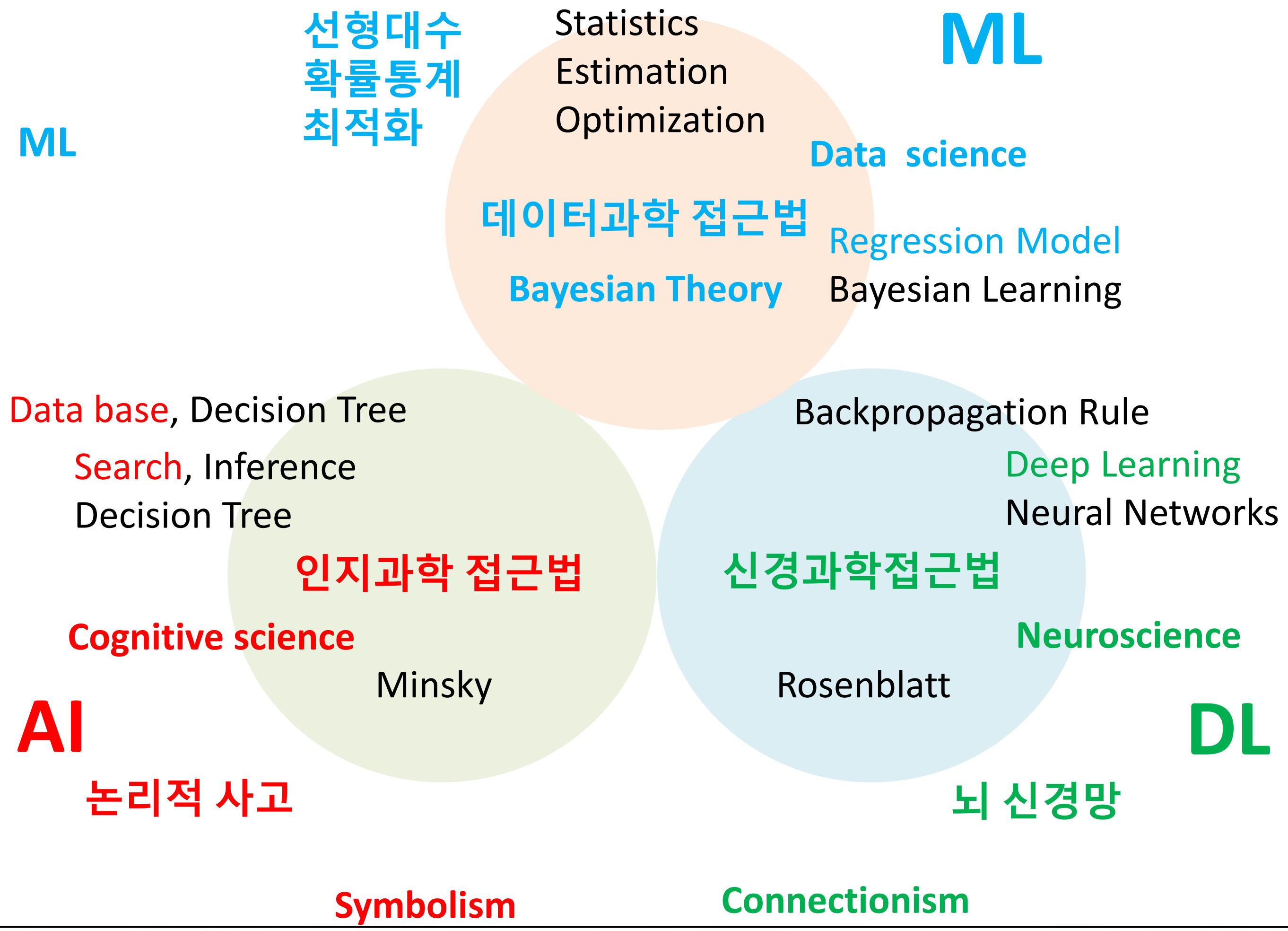
Artificial Intelligence





Model hypothesis
Linear regression
Gaussian: μ, σ

Bayesian learning
Nonlinear model
Density estimation

Deep learning
Neural-net model
Universal mapping

Open access to AI
English terminolog
Math. notations

Probability & Random Variable

Jin Young Choi
Seoul National University

"하나님은 주사위 놀이를 할까요?"

Outline

- **Probability**
- **Conditional Probability**
  - Chain Rule, Total Probability, Independence,
- **Random Variable**
- **Distribution of Random Variable**
- **Joint Probability**
- **Bayes Rule**
- **(Joint) Moment**
  - Mean, (Co)Variance, Expectation, Conditional Expectation
- **Weak Law of Large Numbers**
- **Central Limit Theorem**
- **Random Vector**
- **Random Process**
  - Winner Process, Radom Walk, Markov Process, Ergodicity

Probability

- relative frequency $\neq$ probability
- relative frequency: a measured number, $\hat{p}_i = N_n(O_i)/n$
- probability: a number characterizing an outcome

  (cannot be measured)

$$p_i = \lim_{n \to \infty} N_n\,(O_i)/n, \qquad i = 1\,;\;\; 2, \dots\;, k$$

- probability is a mathematical model, to define probability, we need
  1. a random experiment
  2. outcomes of the experiment
  3. events, each of which is a set of outcomes

Probability

- Finite sample set U with uniform distribution

  $p_A = {}^{|A|}/_{|U|}$

  where $|A|$ is cardinality of the event set A and

  $|U|$ is that of the sample set.

- Example

  \- Fair die (uniform)

  $$P(A) = P(\{2,4,6\}) = \frac{|\{2,4,6\}|}{|\{1,2,3,4,5,6\}|} = 1/2$$

  \- Unfair die (not uniform)

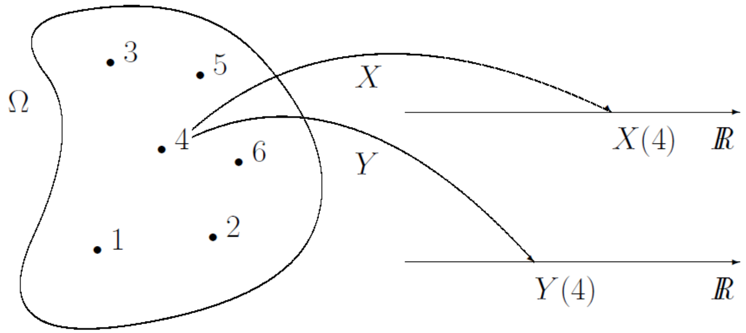
  $$P(A) = P(\{2\}) + P(\{4\}) + P(\{6\})$$
  $$= 1/5 + 1/5 + 1/5 = 3/5 \neq {}^{|A|}/_{|U|}$$

Random variable (r.v.)

- random variable: numerical measurements or observations that have uncertain variability each time they are repeated.
- random variable: real-valued function defined on Ω, given the probability space $(\Omega; A;\ P)$, i.e., $X:\ \Omega\ \rightarrow R$, $Y:\ \Omega\ \rightarrow R$
- example: A coin is tossed 5 times, and if odd number of heads appear, Tom wins \$100, otherwise he loses \$200.

$$\Omega = \{00000, 00001, \cdots, 11111\}$$

Tom's net gain is of interest.



Probability of Random variable

- Probability

$$P(X \in B) := P(X^{-1}(B)) = P\{\omega : X(\omega) \in B\}$$

- example: $\Omega = \{1,2,3,4,5,6\}$, for random variable X,

$\boldsymbol{\omega}$	1	2	3	4	5	6
X	10	-10	20	-10	30	-10

- find $P(X \leq 10) = P(\{1, 2, 4, 6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$

Discrete random variable

- discrete set: having finite number of elements

- example: Ω= {students in this class}, equiprobable selection of a student, $X(\omega) =$ final grade points

- probability mass function (**pmf**)

$$p(x) := P(X = x)$$

  - $0 \leq p(x) \leq 1$
  - $\sum_x p(x) = 1$
  - $P(X \in B) = \sum_{x \in B} p(x)$

Continuous random variable

- continuous set: defined by a range of continuous values
- example: $\Omega = [0, 24]$ represents the time of a specific day;
  $X(\omega) =$ amount of UV light coming into a detector
- probability density function (**pdf**)

$$f(x) := \lim_{\Delta x \to 0} \frac{P(x < X \le x + \Delta x)}{\Delta x}$$

  – $f(x) \ge 0$; $f(x) > 1$ is possible.
  – $\int_{-\infty}^{\infty} f(x)dx = 1$
  – $P(a < X < b) = P(a < X \le b) = P(a \le X < b)$
    $= P(a \le X \le b) = \int_a^b f(x)dx,$
  – $P(X \in B) = \int_B f(x)dx$

Cumulative distribution function

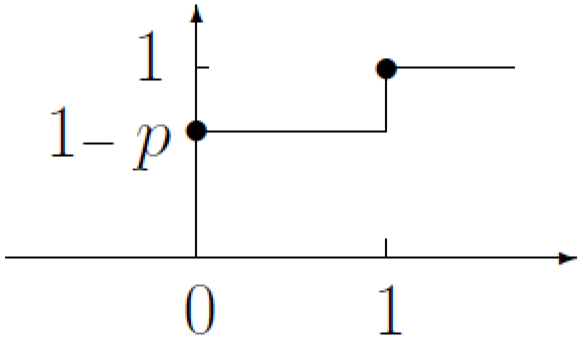
- cumulative distribution function, cdf

$$F(x) := P(X \leq x)$$

for discrete X, $F(x)$ consists of discrete steps.

the step heights are the probability masses.

example: $Ber(p)$ cdf



Cumulative distribution function

- If X is a continuous random variable with density $f(x)$, then

$$\Rightarrow F(x) = P(X \leq x) = \int_{-\infty}^{x} f(v)dv$$

$$= F(a) + \int_{a}^{x} f(v)dv$$

$$\Rightarrow 1 - F(x) = P(X \geq x) = \int_{x}^{\infty} f(v)dv$$

- example: unif(a, b) and $N(m, \sigma^2)$ cdf

Cumulative distribution function

- If X is a *continuous* random variable with density f_X, then

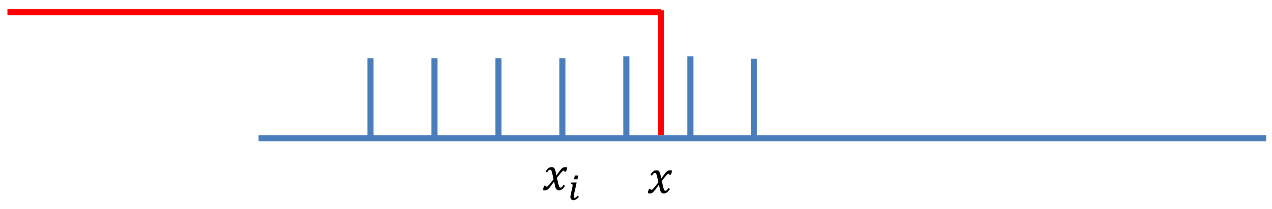
$$F(x) = P(X \le x) = F(a) + \int_a^x f(v)dv$$

$$\Rightarrow f(x) = \frac{d}{dx}F(x)$$

- If X is a *discrete* random variable with density p, then

$$F(x) = \Sigma_i p(x_i)u(x - x_i)$$

$$\Rightarrow f(x) = \frac{d}{dx}F(x) = \Sigma_i p(x_i)\delta(x - x_i)$$: *generalized pdf (↔pmf)*



Cumulative distribution function (cdf)

- properties:

1. $0 \le F(x) \le 1$

2. $P(a < X \le b) = F(b) - F(a)$

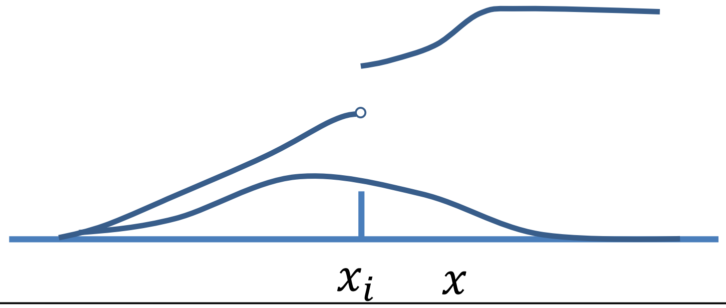
3. $\lim_{x\to-\infty} F(x) = 0$

4. $\lim_{x\to\infty} F(x) = 1$

5. monotone non − decreasing

6. right continuous: $\lim_{\varepsilon\to 0} F(x+\varepsilon) = F(x)$

7. $\lim_{\varepsilon\to 0} F(x-\varepsilon) = F(x) - P(X = x)$



Distribution of discrete r.v.

- Bernoulli pmf : $\Omega = \{0, 1\}$

$$p(k) = \begin{cases} 1-p, & k=0 \\ p, & k=1 \end{cases}$$

- $Ber(p)$, two-valued: Bernoulli trial: success$=1$, failure$=0$

  coin toss: head$=1$, tail$=0$

  Bernoulli trials: independently repeated Bernoulli trial

Distribution of discrete r.v.

- Uniform pmf : $\Omega = \{k | k = l,\ l+1,\ \cdots, m\}$

$$p(k) = \begin{cases} \frac{1}{(m-l+1)} & k = l, l+1, l+2, \cdots, m \\ 0 & \text{else} \end{cases}$$

- $unif\,(l, m)$, equally likely, equiprobable outcomes.

  die toss, random drawing from a deck of 52 cards

Distribution of discrete r.v.

- Geometric pmf : $\Omega = \{k | k = 0,\ 1,\ 2, \cdots\}$

$$p(k) = \begin{cases} (1-p)p^k & k = 0, 1,\ 2, \cdots \\ 0 & \text{else} \end{cases}$$

- $geo(p)$, first failure after k successes in Bernoulli trials

Distribution of discrete r.v.

- Binomial pmf : $\Omega = \{k | k = 0,\ 1,\ 2, \cdots, n\}$

$$p(k) = \begin{cases} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, & k = 0,1,\cdots,n \\ 0, & \text{else} \end{cases}$$

- Multinomial pmf : $\Omega = \{k_1, \cdots, k_m |\ k_i = 0,\ 1,\ 2, \cdots, n\}$

$$p(k_1, \cdots, k_m) = \begin{cases} \frac{n!}{k_1! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m}, & k_i = 0,1,\cdots,n \\ 0, & \text{else} \end{cases}$$

- $bin(p)$, k successes among n Bernoulli trials
- Example: 5 coin tosses; 2 heads; $P\ (head)\ =\ p$

  2 heads cases: 00011, 00101, 00110, 01001, 01010,
  01100, 10001, 10010, 10100, 11000

$$\binom{5}{2} p^2 (1-p)^3 = 10p^2(1-p)^3$$

Distribution of discrete r.v.

- Negative binomial (Pascal) pmf : $\Omega = \{k | k = m,\ m+1,\ \cdots\}$

$$p(k) = \begin{cases} \binom{k-1}{m-1} p^m (1-p)^{k-m}, & k = m, m+1, m+2, \cdots \\ 0, & \text{else} \end{cases}$$

- $Pas(m,p)$, m-th success in k Bernoulli trial

  a salesman, selling an item with prob p, m items to sell
- Example: `How many Bernoulli trials (k) do you need to get $m(=3)$ successes?'

  $k \quad p(k)$

  3 111: $p(3) = \binom{2}{2} p^2(1-p)^0$

  4 0111, 1011, 1101: $p(4) = \binom{3}{2} p^2(1-p)^1$

  5 00111, 01011, 01101, 10011, 10101, 11001: $p(5) = \binom{4}{2} p^2(1-p)^2$

Distribution of discrete r.v.

- Poisson pmf:

$$p(k) = \begin{cases} \dfrac{\lambda^k e^{-\lambda}}{k!}, k = 0,1,2,\cdots \\ 0, \qquad\qquad \text{else} \end{cases}$$

- $Poi(\lambda)$
- How many customers arrive at a store in time t?
- How many neutrinos are detected at a detector in time t?
- How many mosquitos are caught at a trap in time t?
- How many people are born in Seoul in time t?
- How many cars pass a toll gate in time t?
- λ corresponds to the average number.
- $\lambda \sim np\ (large\ n\ \&\ small\ p), bin(n, p) \rightarrow Poi(\lambda)$

Distribution of continuous r.v.

- Uniform **pdf** : $\Omega = \{x | a \leq x \leq b\}$

$$f(x) = \begin{cases} \frac{1}{(b-a)}, & a \leq x \leq b \\ 0, & \text{else} \end{cases}$$

- $unif(a, b)$, equally likely, equiprobable outcomes.

$$dart\ to\ \text{wheel spinning:}\ X \sim unif(0,\ 2\pi)$$

Distribution of continuous r.v.

- Exponential **pdf** : $\Omega = \{x | x \geq 0\}$

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \geq 0 \\ 0, & \text{else} \end{cases}$$

$exp(\lambda)$, time duration, lifetime, interarrival time, cf: $geo(p)$
Shorter time is more likely than longer time.

Distribution of continuous r.v.

- **Laplace pdf** : $\Omega = R$

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}$$

$Lap(\lambda),$ double exponential pdf
difference between two iid exponential rvs.

Distribution of continuous r.v.

- Gaussian **pdf** : $\Omega = R$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$$

denoted by $N(m, \sigma^2)$,

- Multivariate Gaussian **pdf** $\Omega = R^n$

$$f(x) = \frac{1}{\sqrt{(2\pi)^n}\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(x-m)^T\Sigma^{-1}(x-m)\right)$$

Distribution of continuous r.v.

- **Cauchy pdf** : $\Omega = R$

$$f(x) = \frac{\lambda}{\pi(\lambda^2+x^2)}$$

$Cau(\lambda)$, ratio of two iid zero mean Gaussian rvs

Distribution of continuous r.v.

- **Rayleigh pdf** : $\Omega = \{x | x \geq 0\}$

$$f(x) = \begin{cases} \frac{x}{\lambda^2} \exp\left(-\frac{x^2}{2\lambda^2}\right), & x \geq 0 \\ 0, & \text{else} \end{cases}$$

$Ray(\lambda)$, square root of an exponential rv $(X^2 + Y^2)$,

$\sqrt{X^2 + Y^2}$, where $X, Y \sim N(0, \sigma^2)$

Distribution of continuous r.v.

- **Gamma pdf** : $\Omega = \{x | x \geq 0\}$

$$f(x) = \begin{cases} \lambda \dfrac{(\lambda x)^{p-1} e^{-\lambda x}}{\Gamma(p)}, & x \geq 0 \\ 0, & \text{else} \end{cases}$$

$Gam(p, \lambda),$

$p = m$, positive integer $\Rightarrow$ Earlang pdf, sum of iid $\exp(\lambda)$

$p = \frac{k}{2}$, k is a positive integer, $\lambda = \frac{1}{2} \Rightarrow$ chi-squared pdf

Joint Probability

- equal: $P(X = Y) = 1$
- Identical: $\forall x, p_X(x) = p_Y(x)$
- example: three fair-coin tosses;

  X = number of heads; Y = number of tails

  $p_X(0) = p_Y(0) = \frac{1}{8}, p_X(1) = p_Y(1) = \frac{3}{8},$

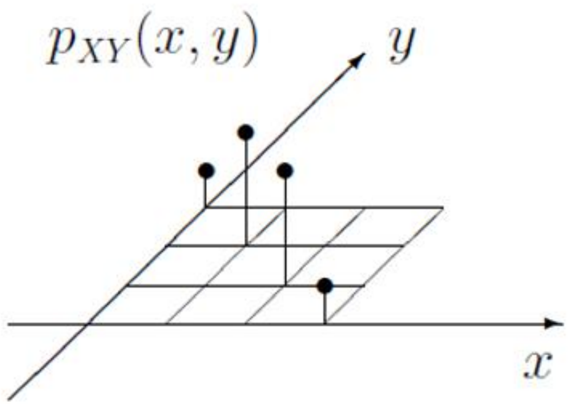
  $p_X(2) = p_Y(2) = \frac{3}{8}, p_X(3) = p_Y(3) = \frac{1}{8}, \rightarrow$ identical

  $P(X = Y) = 0, \rightarrow$ not equal

$p_{XY}(x,y)$	$x=0$	$x=1$	$x=2$	$x=3$
$y=0$	0	0	0	1/8
$y=1$	0	0	3/8	0
$y=2$	0	3/8	0	0
$y=3$	1/8	0	0	0

$\omega =$	hhh	hht	hth	htt	thh	tht	tth	ttt
$X(\omega) =$	3	2	2	1	2	1	1	0

$\omega =$	hhh	hht	hth	htt	thh	tht	tth	ttt
$Y(\omega) =$	0	1	1	2	1	2	2	3



Conditional probability

- Given an event occurred, just as probability changes to conditional probability.

- example: X is the random variable corresponding to the number on a playing card the opponent is putting face down; B is the event that my hand consists of $A(1), 3, 4, 8$, and $Q(12)$.

- Does the event B affect the **pmf** of X?

  $P(X = 1|B)$?

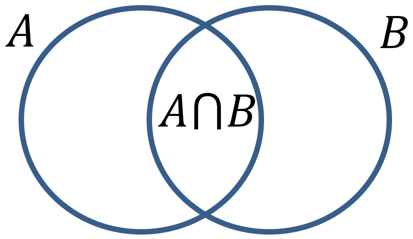
  $P(X = 2|B)$?

Conditional probability

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

- Chain rule.
  1. $P(A \cap B) = P(A|B)P(B)$
  2. $P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)\, P(A_3|A_1 \cap A_2) \cdots P\left(A_n | \cap_{i=1}^{n-1} A_i\right)$

EX) Equally probably (Uniform)



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

Total Probability

- Total Probability.

  1. $P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A \middle| B_i) P(B_i)$

     where $\{B_i, i = 1,2,\cdots,k\}$ is a partition of U.

Independent events

- Independent events A & B.
  1. $P(A|B) = P(A)$ if $P(B) > 0 \Leftrightarrow P(A \cap B) = P(A|B)P(B) = P(A)P(B)$
  2. $$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)\,P(A_3|A_1 \cap A_2)\cdots P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right)$$
     $$= P(A_1)P(A_2)\,P(A_3)\cdots P(A_n)$$

Joint Probability

- X and Y are independent:
  $\forall B$ and $C \subseteq R$, $P(X \in B, Y \in C) = P(X \in B)P(Y \in C)$

- $X_1, X_2, \cdots, X_n$ are independent:
  $\forall B_1, B_2, \cdots, B_k \subseteq R$,
  $P(X_i \in B_i, i = 1,2,\cdots,n) = \Pi_{i=1}^{n} P(X_i \in B_i)$

Joint Probability

- joint probability mass function, **jpmf**

$$p_{XY}(x,y) := P(X = x, Y = y) = P\big((X,Y) = (x,y)\big) \coloneqq p(x,y)$$

1. $0 \le p(x,y) \le 1$
2. $\Sigma_{(x,y)} p(x,y) = 1$
3. $P((X,Y) \in D) = \Sigma_{(x,y)\in D} p(x,y)$

- extends to multi-variable:

  $p(x_1, \cdots, x_k)$

Joint Probability

- marginal pmf:

$$p(x) = \Sigma_y p(x,y), \quad p(y) = \Sigma_x p(x,y)$$

- X and Y are independent:

$$p(x,y) = p(x)p(y)$$

- $X_1, \cdots, X_k$ are independent:

$$p(x_1, \cdots, x_k) = \Pi_{i=1}^{k} p(x_i)$$

- independent and identically distributed, iid, $X_1, \cdots, X_k$

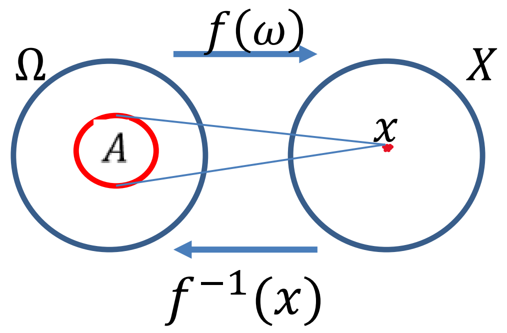
$$p(x_1, \cdots, x_k) = p(x_1) \cdots p(x_k) \text{ and}$$
$$p(x_i) = p\left(x_j\right) \;\; if \; x_i = x_j \;\; for \;\; i \neq j.$$

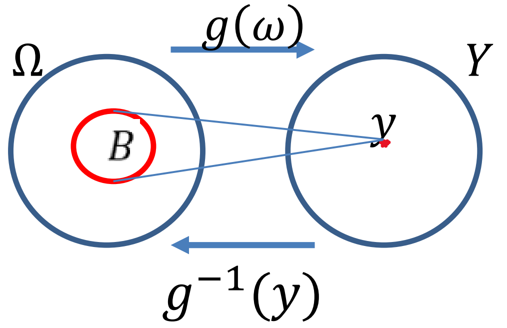
Bayes Rule

- conditional pmf, cpmf:

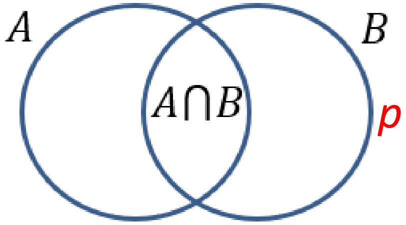
$$p(x|y) := \frac{p(x,y)}{p(y)}$$

$$X = x \in \Omega_X, \; Y = y \in \Omega_Y$$





- $p(x|y)$ is not defined when $p(y) \neq 0$



$$p(x|y) = P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|U|}{|B|/|U|} = \frac{P(A \cap B)}{P(B)} = \frac{p(x,y)}{p(y)}$$

Bayes Rule

- conditional pmf, cpmf:

$$p(x|y,z) = \frac{p(x,y,z)}{p(y,z)}$$

$$p(x,y|z) = \frac{p(x,y,z)}{p(z)}$$

- independence

$$X, Y \ are \ independent \ \rightarrow \ p(x|y) = p(x)$$

Bayes Rule

- chain rule:

$$p(x, y) = p(x)p(y|x)$$

$$p(x_1, \cdots, x_k)$$
$$= p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \cdots p(x_k|x_1, \cdots, x_{k-1})$$

$$p(x, y|w) = p(x|w)p(y|w, x)$$

$$p(x, y, z|w) = p(x, y|w)p(z|w, x, y)$$

Bayes Rule

- total probability law:

$$P(X=x)=p(x)=\Sigma_y p(x,y)=\Sigma_y p(x|y)p(y)$$

- example: X : life expectancy of a 70-year-old person.

– blood condition after 70 years old

H : having high blood pressure, $P(H) = 2/5$

R : having normal blood pressure, $P(R) = 3/5$

– at every year after 70 years old

survival probability of high blood person: $9/10$

survival probability of normal blood person: $19/20$

– what is the probability that a 70 years old person lives until 90 years old ?

Bayes Rule

- example:
  - $p(x) = p(x|H)P(H) + p(x|R)P(R)$
  - $p(x|H) = \begin{cases} \frac{1}{10}\left(\frac{9}{10}\right)^{x-1}, & x = 1,2,\cdots \\ 0, & \text{else} \end{cases} \sim \text{geo}(1/10)$
  - $p(x|R) = \begin{cases} \frac{1}{20}\left(\frac{19}{20}\right)^{x-1}, & x = 1,2,\cdots \\ 0, & \text{else} \end{cases} \sim \text{geo}(1/20)$
  - $p(x) = p(x|H)P(H) + p(x|R)P(R)$

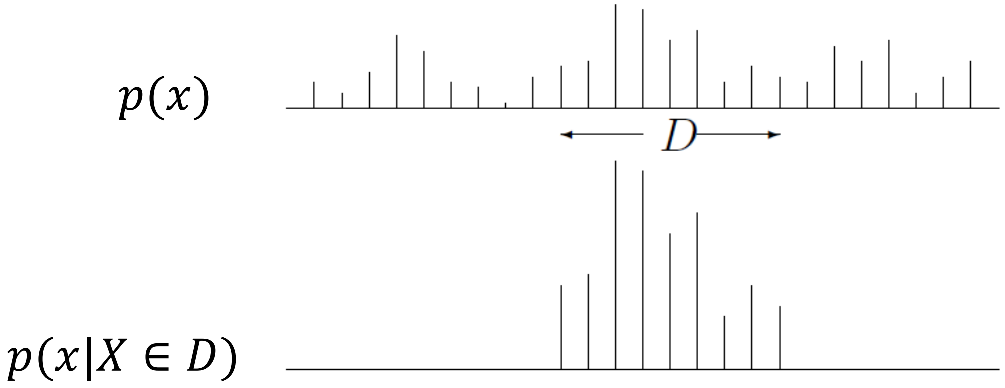
$$= \frac{1}{10}\left(\frac{9}{10}\right)^{x-1} \cdot \frac{2}{5} + \frac{1}{20}\left(\frac{19}{20}\right)^{x-1} \cdot \frac{3}{5}$$

Bayes Rule

- Meaning of conditional probability
  - When the conditioning event is of the form $\{X \in D\}$,

$$p(x|X \in D) = \frac{P(X = x, X \in D)}{P(X \in D)}$$

$$= \begin{cases} \dfrac{P(X = x)}{P(X \in D)}, & X \in D \\ 0, & \text{else} \end{cases}$$



Bayes Rule

- Bayes Rule
  - $p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(x|y)\, p(y)}{\sum_y p(x|y)\, p(y)}$

- Learning and Inference ?

Classification $\quad p(w_i|x) = \frac{p(x, w_i)}{p(x)} = \frac{p(x|w_i)\, p(w_i)}{\sum_{w_i} p(x|w_i)\, p(w_i)}$

Estimation $\quad p(\theta_i|x) = \frac{p(x, \theta_i)}{p(x)} = \frac{p(x|\theta_i)\, p(\theta_i)}{\sum_{\theta_i} p(x|\theta_i)\, p(\theta_i)}$

Exercise

Let X be the exponential random variable.
Find $F(x|X > t)$ and $f(x|X > t)$. How does $F(x|X > t)$ differ from $F(x)$?

- Sol.
  - ✓ $f(x) = \lambda e^{-\lambda x}, x \geq 0.$
  - ✓ $F(x) = \int_0^x \lambda e^{-\lambda v}\, dv = -e^{-\lambda v}\big|_0^x = 1 - e^{-\lambda x}$

Exercise

Let X be the exponential random variable.

Find $F(x|X>t)$ and $f(x|X>t)$. How does $F(x|X>t)$ differ from $F(x)$?

- Sol.
  - ✓ $f(x) = \lambda e^{-\lambda x}, x \geq 0.$
  - ✓ $F(x) = \int_0^x \lambda e^{-\lambda v}\, dv = -e^{-\lambda v}\Big|_0^x = 1 - e^{-\lambda x}$
  - ✓ $F(x|X>t) = \frac{P[\{X\leq x\}\cap\{X>t\}]}{P[\{X>t\}]} = \begin{cases} 0 & x<t \\ \frac{P[\{t<X\leq x\}]}{P[\{X>t\}]} & x \geq t \end{cases}$

Exercise

Let X be the exponential random variable.
Find $F(x|X>t)$ and $f(x|X>t)$. How does $F(x|X>t)$ differ from $F(x)$?

- Sol.
  - ✓ $f(x) = \lambda e^{-\lambda x}, x \geq 0.$
  - ✓ $F(x) = \int_0^x \lambda e^{-\lambda v}\, dv = -e^{-\lambda v}\Big|_0^x = 1 - e^{-\lambda x}$
  - ✓ $F(x|X>t) = \frac{P[\{X\leq x\}\cap\{X>t\}]}{P[\{X>t\}]} = \begin{cases} 0 & x<t \\ \frac{P[\{t<X\leq x\}]}{P[\{X>t\}]} & x \geq t \end{cases}$
  - ✓ $\frac{P[\{t<X\leq x\}]}{P[\{X>t\}]} = \frac{F(x)-F(t)}{1-F(t)} = \frac{e^{-\lambda t}-e^{-\lambda x}}{e^{-\lambda t}} = 1 - e^{-\lambda(x-t)}$

Exercise

Let X be the exponential random variable.

Find $F(x|X > t)$ and $f(x|X > t)$. How does $F(x|X > t)$ differ from $F(x)$?

- Sol.
  - ✓ $f(x) = \lambda e^{-\lambda x}, x \geq 0.$
  - ✓ $F(x) = \int_0^x \lambda e^{-\lambda v}\, dv = -e^{-\lambda v}\Big|_0^x = 1 - e^{-\lambda x}$
  - ✓ $F(x|X > t) = \frac{P[\{X \leq x\} \cap \{X > t\}]}{P[\{X > t\}]} = \begin{cases} 0 & x < t \\ \frac{P[\{t < X \leq x\}]}{P[\{X > t\}]} & x \geq t \end{cases}$
  - ✓ $\frac{P[\{t < X \leq x\}]}{P[\{X > t\}]} = \frac{F(x) - F(t)}{1 - F(t)} = \frac{e^{-\lambda t} - e^{-\lambda x}}{e^{-\lambda t}} = 1 - e^{-\lambda(x-t)}$
  - ✓ $f(x|X > t) = \lambda e^{-\lambda(x-t)}$

Interim Summary

- **Probability**
- **Conditional Probability**
  - Chain Rule, Total Probability, Independence,
- **Random Variable**
- **Distribution of Random Variable**
- **Joint Probability**
- **Bayes Rule**
- **(Joint) Moment**
  - Mean, (Co)Variance, Expectation, Conditional Expectation
- **Weak Law of Large Numbers**
- **Central Limit Theorem**
- **Random Vector**
- **Random Process**
  - Winner Process, Radom Walk, Markov Process, Ergodicity
- **Basic Discrete Structures**
  - Sets, Functions, Sequences, Summations, Recurrence Relations
  - Probability and Counting

Average

- when we cannot have the pmf, we need a statistic, that characterizes the sample, population, or random variable.

- three kinds of averages: mean, mode, median
  - mean: (arithmetic) average:

    $$m := \frac{1}{n} \Sigma_x x$$
  - mode: the most popular value:

    $$x_{\text{mod}} := \max^{-1} P\,(X = x)$$

    mode may not be unique; unimodal, bimodal, multimodal.
  - median: the center value:

    $$P(X < x_{\text{med}}) = P(X > x_{\text{med}})$$

    median may not be unique.

Average

- What do you expect for X?

- expectation, expected value, mean over universe:
  $EX := \Sigma_x x p(x)$ if $\Sigma_x |x| p(x) < \infty$

- expectation is similar to arithmetic average.
- consider n numbers $x_1, \cdots, x_n$.
  let N_x be the number of x_i's with the value x
$$\frac{1}{n}\Sigma_{i=1}^{n} x_i = \frac{1}{n}\Sigma_x x N_x = \Sigma_x x \frac{N_x}{n} \to \Sigma_x x p(x) \text{ as } n \to \infty.$$

Average

- expectation of function of random variables

$$EY = \Sigma_x g(x)p(x)$$

- example: $X \approx \text{unif}(1,10)$,

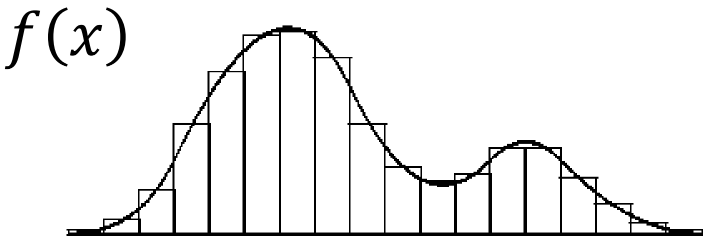
X	1	2	3	4	5	6	7	8	8	10
$Y = g(X)$	5	5	5	5	0	0	0	2	2	3

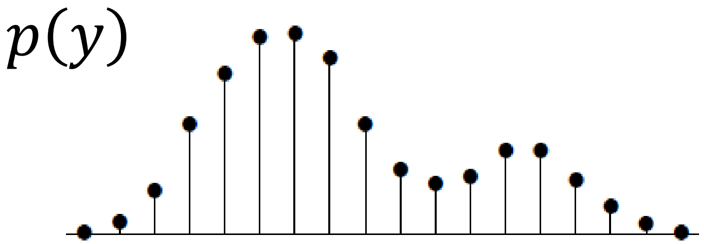
$$\begin{aligned} Eg(X) &= \Sigma_x g(x)p(x) \\ &= 5 \cdot \frac{1}{10} + 5 \cdot \frac{1}{10} + 5 \cdot \frac{1}{10} + 5 \cdot \frac{1}{10} + 0 \cdot \frac{1}{10} + 0 \cdot \frac{1}{10} + 0 \cdot \frac{1}{10} + 2 \cdot \frac{1}{10} + 2 \cdot \frac{1}{10} + 3 \cdot \frac{1}{10} \\ &= 5 \cdot \frac{4}{10} + 0 \cdot \frac{3}{10} + 2 \cdot \frac{2}{10} + 3 \cdot \frac{1}{10} = \Sigma_y y p(y) \end{aligned}$$

Expectation for Continuous r.v.

- expectation: $EX := \int_{-\infty}^{\infty} xf(x)dx$, if $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$

  let Y be a discrete random variable approximating X.





$$EX \approx \Sigma_y yp(y) = \Sigma_y yf(y)\Delta \approx \int xf(x)dx$$

$$Eg(X) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

This can be justified using approximation with a discrete rv.

Moment

- Moment
  - n-th moment: EX^n
  - n-th central moment: $E(X - m_X)^n$
- Mean(average): 1-st moment
- Variance: 2-nd central moment
- n-th central moment is a function of n-th and lower moments.

Moment

- expectation operator is linear.

$$E(aX+b)=aEX+b$$

- variance

$$\begin{aligned} E(X-m_X)^2 &= E(X^2-2m_X X+{m_X}^2) \\ &= EX^2-2m_X EX+{m_X}^2 = EX^2-{m_X}^2 \end{aligned}$$

- standard deviation

$$\sigma_X=\sqrt{\mathrm{var}(X)}$$

- variance is not linear

$$\begin{aligned} \mathrm{var}(aX+b) &= E(aX+b)^2-(E(aX+b))^2 \\ &= E(aX+b)^2-(am_X+b))^2 = a^2E(X-m_X)^2 \\ &= a^2\,\mathrm{var}(X): \ \text{not affected by } b. \end{aligned}$$

Joint moment

- correlation: $\mathrm{cor}(X,Y) = EXY$

- covariance:

$$\mathrm{cov}(X,Y) = E(X - m_X)(Y - m_Y)$$
$$= EXY - m_X m_Y$$

- correlation coefficient:

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$
$$|\rho_{XY}| \leq 1 \text{ [Schwarz ineq]}$$

Joint moment

- uncorrelated X and Y :

$$EXY = EXEY, \mathrm{cov}(X,Y) = 0, \text{ or } \rho_{XY} = 0$$

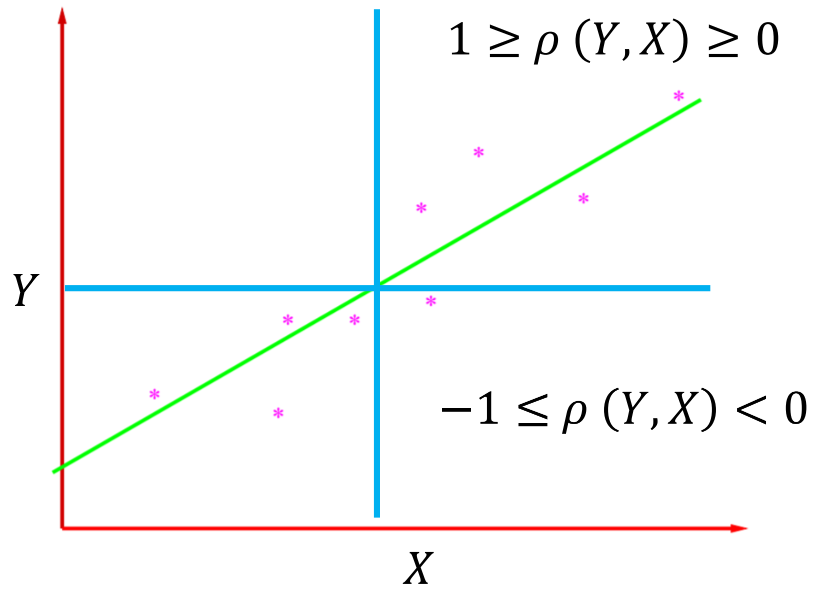
- independent and uncorrelated

independent $\Rightarrow$ uncorrelated
uncorrelated $\not\Rightarrow$ independent

[example]

Let $p(x) = p(-x)$ and $Y = X^2$

$\Rightarrow EX = EX^3 = EX^5 = \cdots = 0$
$\Rightarrow EXY = EX^3 = 0 = EXEY$
$\Rightarrow$ uncorrelated but not independent



Joint moment

- variance of sum of random variables

$$\mathrm{var}(\Sigma_{i=1}^{k} X_i) = \Sigma_{i=1}^{k} \mathrm{var}(X_i) + \Sigma\Sigma_{i\neq j} \mathrm{cov}(X_i, X_j)$$

$$X_i \text{ uncorrelated } \Rightarrow \mathrm{var}(\Sigma_{i=1}^{k} X_i) = \Sigma_{i=1}^{k} \mathrm{var}(X_i)$$

$$X_i \text{ i.i.d. } \Rightarrow \mathrm{var}(\Sigma_{i=1}^{k} X_i) = k\, \mathrm{var}(X_i)$$

- (Cauchy-)Schwarz inequality:

$$|EXY| \leq \sqrt{EX^2}\,\sqrt{EY^2},$$

where equality holds if and only if
$X = aY$ for some a.

$$\Rightarrow |\mathrm{cov}(X,Y)| = |E(X - m_X)(Y - m_Y)| \leq \sigma_X \sigma_Y,$$

$$\Rightarrow |\rho_{XY}| \leq 1$$

cf. $|\sum x_i\, y_i| \leq \sqrt{\sum x_i^2}\, \sqrt{\sum y_i^2}$

Weak law of large numbers

- We often estimate the mean of a random variable by averaging its sample values.

  $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- sample mean:

  : unbiased estimator of m_X if $EX_i = m_X,\ i = 1, \cdots, n$ $\leftarrow EM_n = m_X$

- weak law of large numbers:

  let $X_1, \cdots, X_n$ be i.i.d., $EX_i = m$, $\mathrm{var}(X_i) = \sigma^2$,

  $EM_n = m$, $\mathrm{var}(M_n) = {\sigma^2}/{n} \rightarrow 0\ as\ n \rightarrow \infty$

  $\leftarrow\ X_i$ i.i.d. $\Rightarrow \mathrm{var}(M_n) = \mathrm{var}(\frac{1}{n}\Sigma_{i=1}^{n} X_i) = \frac{1}{n^2}\mathrm{var}(\Sigma_{i=1}^{n} X_i) = \frac{1}{n}\mathrm{var}(X_i)$

Central Limit Theorem

- Let $X_1, \cdots, X_n$ be i.i.d., $EX_i = m$, $\mathrm{var}(X_i) = \sigma^2$,

$$S_n := \Sigma_{i=1}^{n} X_i : \quad \text{mean} = nm, \text{var} = n\sigma^2$$

$$M_n := \frac{1}{n}\Sigma_{i=1}^{n} X_i : \quad \text{mean} = m, \text{var} = \frac{\sigma^2}{n}$$

$$Z_n := \frac{1}{\sqrt{n}}\Sigma_{i=1}^{n} X_i : \quad \text{mean} = \sqrt{n}m, \text{var} = \sigma^2$$

$$Y_n := \frac{1}{\sqrt{n}}\Sigma_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right) : \quad \text{mean} = 0, \text{var} = 1$$

$$Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{X_i - m}{\sigma} \to N(0,1) as\ n \to \infty \qquad \boxed{\to Y_n = \frac{S_n - nm}{\sqrt{n\sigma^2}} \sim N(0,1)}$$

Central Limit Theorem

- Example:

  Consider the sum W of the 100 numbers obtained from repeated tossing of a fair die.

$$P(W \leq 200) = \Sigma_{w=100}^{200} p(w)$$

- Let $X_1, \cdots, X_n$ be i.i.d. repeated tossing of a fair die, $W = \sum_{i=1}^{n} X_i$, $EX_i = m$, $\mathrm{var}(X_i) = \sigma^2$
- $W = \sum_{i=1}^{n} X_i \rightarrow Y = \frac{W-nm}{\sqrt{n\sigma^2}} \sim N(0,1)$

$$\rightarrow P\left(\frac{100-nm}{\sqrt{n\sigma^2}} \leq Y \leq \frac{200-nm}{\sqrt{n\sigma^2}}\right) = \Sigma_{y=y_1}^{y_2} p(y),\ \ y_1 = \frac{100-nm}{\sqrt{n\sigma^2}},\ \ y_2 = \frac{200-nm}{\sqrt{n\sigma^2}}, n = 100$$

Central Limit Theorem

- Example:

  Consider the sum W of the 100 numbers obtained from repeated tossing of a fair die.

$$P(W \le 200) = \sum_{w=100}^{200} p(w),\ W = \sum_{i=1}^{n} X_i$$

$$\rightarrow P\left(\frac{100-nm}{\sqrt{n\sigma^2}} \le Y \le \frac{200-nm}{\sqrt{n\sigma^2}}\right) = \sum_{y=y_1}^{y_2} p(y),\ \ y_1 = \frac{100-nm}{\sqrt{n\sigma^2}},\ \ y_2 = \frac{200-nm}{\sqrt{n\sigma^2}},\ n = 100$$

- Find m and σ^2

Central Limit Theorem

- Example:

  Consider the sum W of the 100 numbers obtained from repeated tossing of a fair die.

$$P(W \le 200) = \Sigma_{w=100}^{200} p(w),\, W = \Sigma_{i=1}^{n} X_i$$

$$\rightarrow P\left(Y \le \frac{200-nm}{\sqrt{n\sigma^2}}\right) = \Sigma_{y=y_1}^{y_2} p(y),\;\; y_1 = \frac{100-nm}{\sqrt{n\sigma^2}},\;\; y_2 = \frac{200-nm}{\sqrt{n\sigma^2}},\, n = 100$$

- Find m and σ^2

$$m = EX_i = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

$$EX_i^2 = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) = \frac{91}{6} = 15.1667$$

$$\sigma^2 = \mathrm{var}(X_i) = EX_i^2 - m^2 = 15.17 - 3.5^2 = 2.9167,$$

$$EW = nm = 100 \times 3.5 = 350$$

$$\mathrm{var}(W) = n\sigma^2 = 100 \times 2.9167 = 291.67$$

Central Limit Theorem

- Example:

  Consider the sum W of the 100 numbers obtained from repeated tossing of a fair die.

$$P(W \le 200) = \Sigma_{w=100}^{200} p(w), W = \sum_{i=1}^{n} X_i$$

$$\rightarrow P\left(Y \le \frac{200-nm}{\sqrt{n\sigma^2}}\right) = \Sigma_{y=y_1}^{y_2} p(y),\ \ y_1 = \frac{100-nm}{\sqrt{n\sigma^2}},\ \ y_2 = \frac{200-nm}{\sqrt{n\sigma^2}}$$

$P(100 \le W \le 200) \approx$

Central Limit Theorem

- Example:

  Consider the sum W of the 100 numbers obtained from repeated tossing of a fair die.

$$P(W \le 200) = \Sigma_{w=100}^{200} P_W(w),\, W = \sum_{i=1}^{n} X_i$$

$$\rightarrow P\left(Y \le \frac{200-nm}{\sqrt{n\sigma^2}}\right) = \Sigma_{y=y_1}^{y_2} p_Y(y),\;\; y_1 = \frac{100-nm}{\sqrt{n\sigma^2}},\;\; y_2 = \frac{200-nm}{\sqrt{n\sigma^2}}$$

$$P(100 \le W \le 200) \approx P\left(Y \le \frac{200-nm}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{200-nm}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{200-350}{\sqrt{291.67}}\right) = \Phi(-8.7831) \approx 0$$

Conditional expectation

- conditional expectation:

for an event A, $\quad E(X|A) := \Sigma_x x p(x|A)$

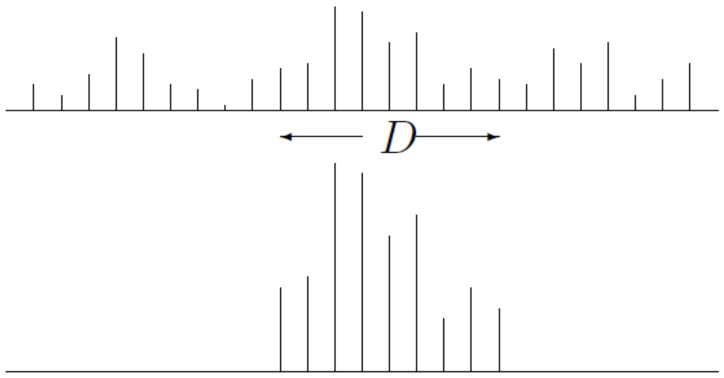
$$E(X|Y=y) := \Sigma_x x p(x|y)$$

this is a function of y (for the given $p(x,y)$)

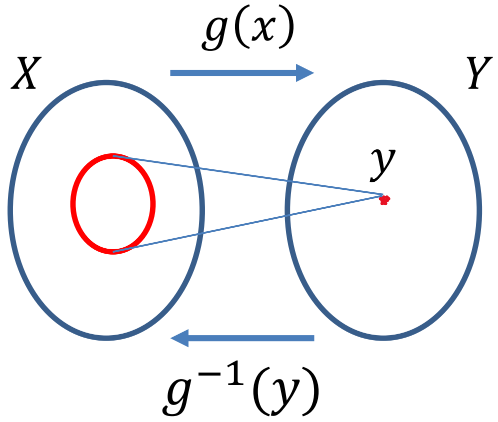
$$E(g(X)|Y=y) = \Sigma_x g(x) p(x|y)$$

$p(x)$

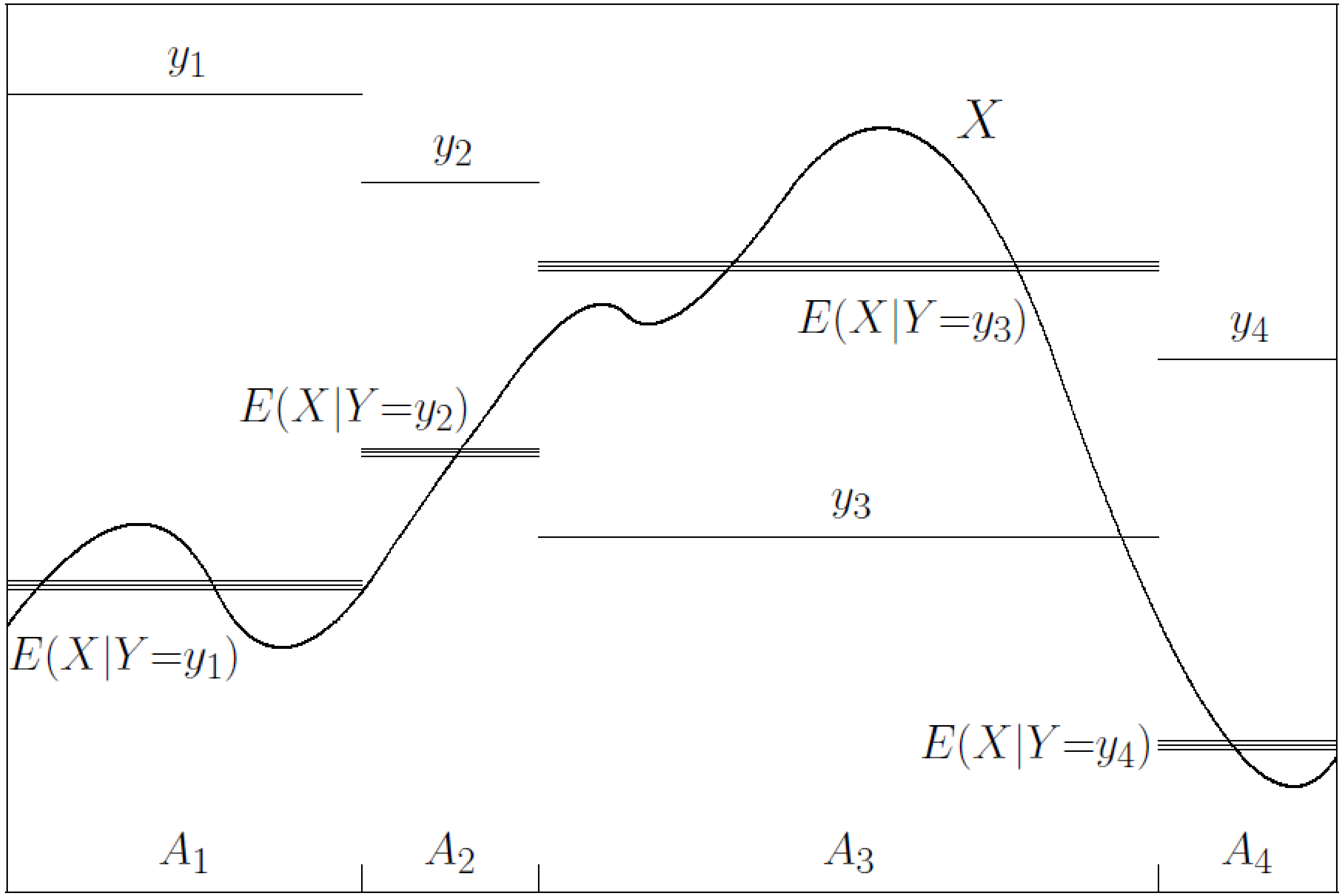
$p(x|y) = p(x|x \in D)$



$D = g^{-1}(y)$



Conditional expectation



Conditional expectation

- properties:

  1. linear : $E(aX + bY|Z) = aE(X|Z) + bE(Y|Z)$
  2. $E(a|X) = a$
  3. $E(X|a) = EX$
  4. $X \geq a \Rightarrow E(X|Y) \geq a$; likewise for $>, \leq,$ and $<$.
  5. $|E(X|Y)| \leq E(|X||Y)$
  6. $E(g(Y)X|Y) = g(Y)E(X|Y)$ [substitution law]
  7. X and Y independent $\Rightarrow E(X|Y) = EX$
  8. $EE(X|Y) = EX$ [total prob law]
  9. g is one-to-one $\Rightarrow E(X|g(Y)) = E(X|Y)$

- $EE(X|Y) = \sum_y \sum_x xp(x|y)p(y) = \sum_y \sum_x xp(x,y) = \sum_x x \sum_y p(x,y) = \sum_x xp(x) = EX$

Interim Summary

- **Probability**
- **Conditional Probability**
  - Chain Rule, Total Probability, Independence,
- **Random Variable**
- **Distribution of Random Variable**
- **Joint Probability**
- **Bayes Rule**
- **(Joint) Moment**
  - Mean, (Co)Variance, Expectation, Conditional Expectation
- **Weak Law of Large Numbers**
- **Central Limit Theorem**
- **Random Vector**
- **Random Process**
  - Winner Process, Radom Walk, Markov Process, Ergodicity
- **Basic Discrete Structures**
  - Sets, Functions, Sequences, Summations, Recurrence Relations
  - Probability and Counting

Random Vectors

- random vector:

$$X = (X_1, \cdots, X_k)^T, EX = (EX_1, \cdots, EX_k)^T$$

- random matrix:

$$EXX^T = \begin{pmatrix} EX_1X_1 & EX_1X_2 & \cdots & EX_1X_k \\ EX_2X_1 & & & EX_2X_k \\ \vdots & & & \vdots \\ EX_kX_1 & EX_kX_2 & \cdots & EX_kX_k \end{pmatrix}$$

$$E(X - \boldsymbol{m}_X)(X - \boldsymbol{m}_X)^T = \begin{pmatrix} E(X_1 - m_1)(X_1 - m_1) & \cdots & E(X_1 - m_1)(X_k - m_k) \\ E(X_2 - m_2)(X_1 - m_1) & \cdots & E(X_2 - m_2)(X_k - m_k) \\ \vdots & \ddots & \cdots \\ E(X_k - m_k)(X_1 - m_1) & \cdots & E(X_k - m_k)(X_k - m_k) \end{pmatrix}$$

Jointly Gaussian random variables

- jointly Gaussian rvs

$$f(x) = \frac{1}{2\pi\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n}\sqrt{|\Sigma|}}\exp\left(-\frac{(\mathbf{x}-\mathbf{m})^T\Sigma^{-1}(\mathbf{x}-\mathbf{m})}{2}\right)$$

$$\Sigma = E[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T]$$

$$= E\begin{bmatrix} (x-m_x)^2 & (x-m_x)(y-m_y) \\ (x-m_x)(y-m_y) & (y-m_y)^2 \end{bmatrix} \quad for\ \ \mathbf{x}^T = [x, y]$$

$$= \begin{bmatrix} {\sigma_x}^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & {\sigma_y}^2 \end{bmatrix}, \quad |\Sigma| = \sigma_x^2\sigma_y^2(1-\rho)$$

Jointly Gaussian random variables

- conditional probability

$$f(x|y) = \frac{1}{(2\pi)^{\frac{k}{2}}\sqrt{|\Sigma_{X|y}|}} \exp\left(-\frac{1}{2}(x - m_{X|y})^t \Sigma_{X|y}^{-1}(x - m_{X|y})\right),$$

where

$$m_{X|y} = A(y - m_Y) + m_X \text{ and}$$

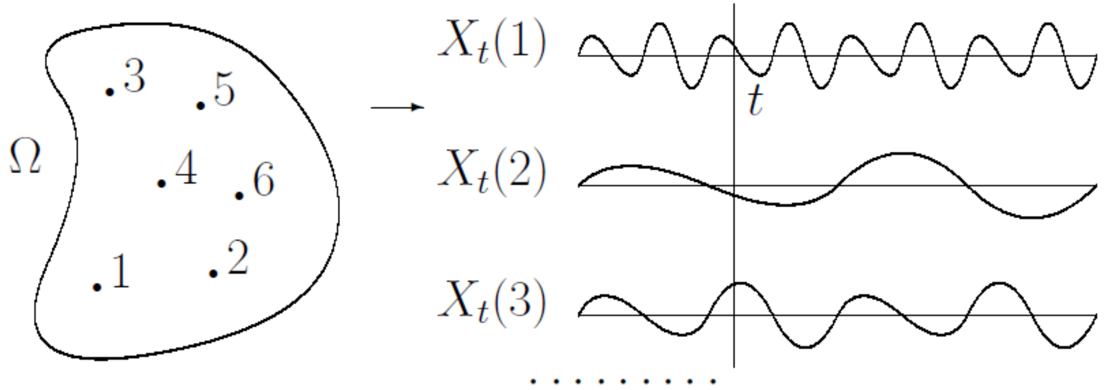
$$\Sigma_{X|y} = \Sigma_X - AC_{YX}, \text{ where } A\Sigma_Y = \Sigma_{XY}.$$

- conditional expectation

$$E(X|Y) = A(Y - m_Y) + m_X$$

Random process

- random process $X_t(\omega), \quad t \in I$
  1. random sequence, random function, or random signal: $X_t : \Omega \rightarrow$the set of all sequences or functions
  2. indexed family of infinite number of random variables: $X_t : I \rightarrow$set of all random variables defined on Ω

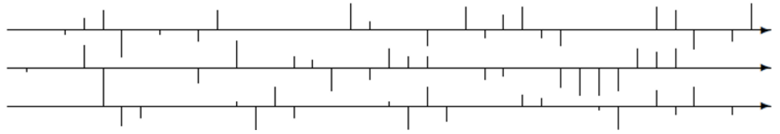


Random process

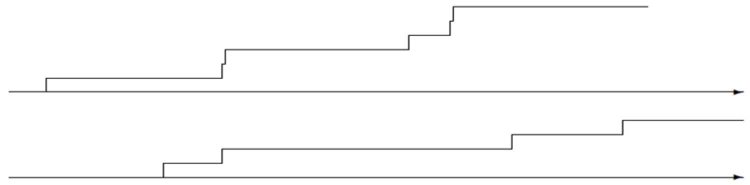
- example:
  - surface temperature of a space shuttle
  - thermal noise of a semiconductor device
  - total number of customers visiting a store up to time t
  - sequence of i.i.d. Bernoulli r. v. : Bernoulli process
  - $X_t = A\cos 2\pi t, Y_t = B\cos 2\pi t$
    where A and B are independent random variables.
  - $X_t = \cos 2\pi(t+\Theta), Y_t = \cos 2\pi(t+\Psi)$
    where $\Theta, \Psi \sim$ uniform and independent.
  - X_n, Y_n, Z_n : i id $N(0,1)$

$$f_{X_i,X_j,Y_k}(x_i,x_j,y_k) = \frac{1}{(2\pi)^{\frac{3}{2}}} e^{-\frac{x_i^2+x_j^2+y_k^2}{2}}$$

discrete-time, continuous-valued - iid Gaussian



continuous-time, discrete-valued-Poisson process



Moment

- mean function:

$$m_X(t) := EX_t = \begin{cases} \Sigma_x x p(x), & \text{disc valued} \\ \int x f(x) dx, & \text{cont valued} \end{cases}$$

- auto-correlation function, acf:

$$R_X(t,s) := EX_t X_s = \begin{cases} \Sigma_{u \in X_t} \Sigma_{v \in X_s} uv\, p(u,v), & \text{disc valued} \\ \int_{X_s} \int_{X_t} uvf(u,v) du dv\,, & \text{cont valued} \end{cases}$$

$$X = (X_{t_1}, \cdots, X_{t_k}) \Rightarrow R_X = \begin{pmatrix} R_X(t_1,t_1) & \cdots & R_X(t_1,t_k) \\ \vdots & & \vdots \\ R_X(t_k,t_1) & \cdots & R_X(t_k,t_k) \end{pmatrix}$$

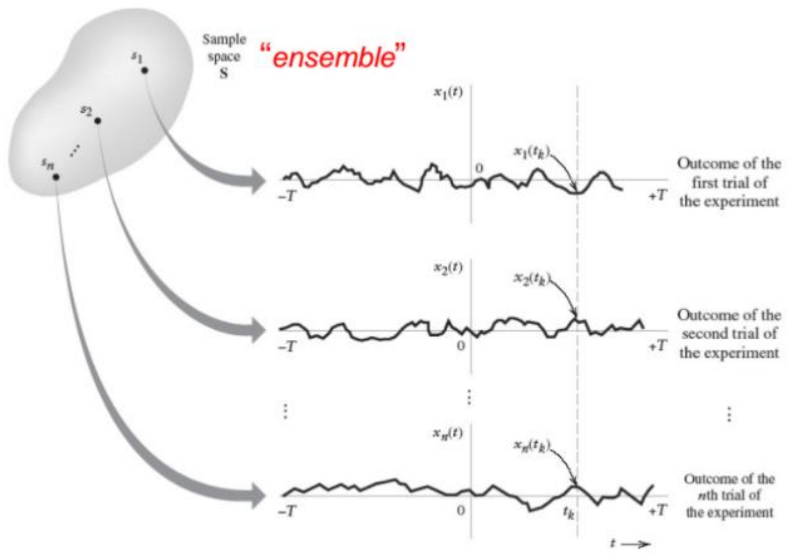


Figure from (Haykin & Moher, 2010)

Moment

- auto-covariance function, acvf:

$$\begin{aligned} C_X(t,s) &:= E(X_t - m_X(t))(X_s - m_X(s)) \\ &= R_X(t,s) - m_X(t)m_X(s) \end{aligned}$$

$$X = (X_{t_1}, \cdots, X_{t_k}) \Rightarrow C_X = \begin{pmatrix} C_X(t_1,t_1) & \cdots & C_X(t_1,t_k) \\ \vdots & & \vdots \\ C_X(t_k,t_1) & \cdots & C_X(t_k,t_k) \end{pmatrix}$$

- cross-correlation function, ccf:

$$R_{XY}(t,s) := EX_tY_s$$

- cross-covariance function, ccvf:

$$\begin{aligned} C_{XY}(t,s) &:= E(X_t - m_X(t))(Y_s - m_Y(s)) \\ &= R_{XY}(t,s) - m_X(t)m_Y(s) \end{aligned}$$

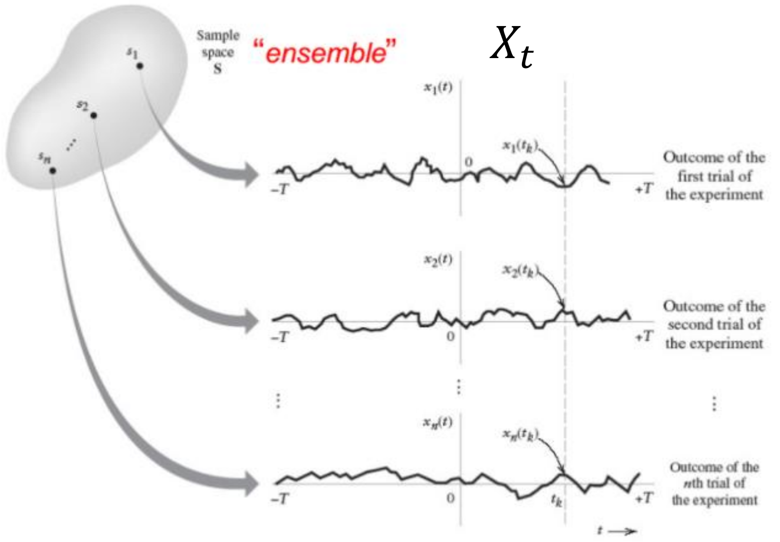


Figure from (Haykin & Moher, 2010)

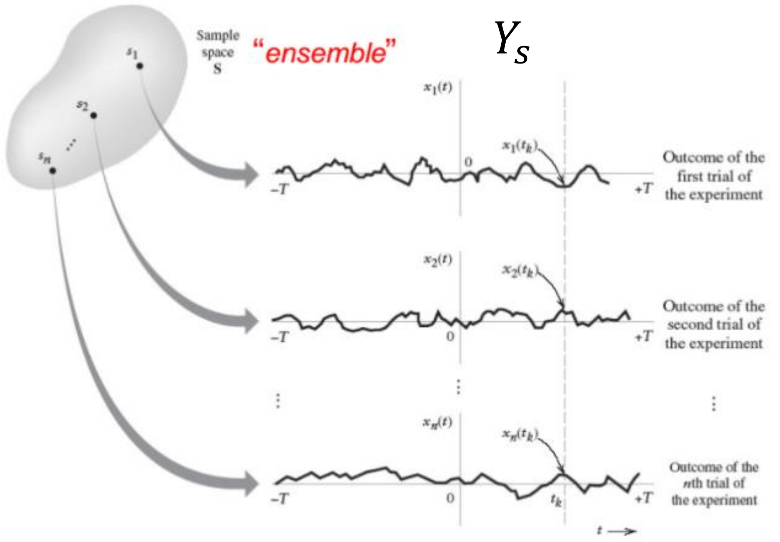


Figure from (Haykin & Moher, 2010)

Moment

- orthogonal processes X_t and Y_t : $E\,X_tY_s = R_{XY}(t,s) = 0$ for any t and s.
- uncorrelated processes X_t and Y_t:

  $EX_tY_s = EX_tEY_s$ for any t and s.

  $R_{XY}(t,s) = m_X(t)m_Y(s)$

  $C_{XY}(t,s) = 0$

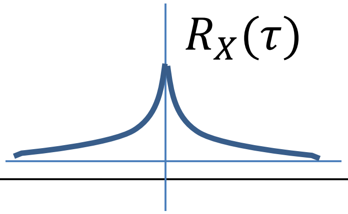
independent $\Rightarrow$ uncorrelated
uncorrelated $\not\Rightarrow$ independent (except jointly Gaussian proc.)

- stationarity: shift invariance for any d, $\tau = s - t$,

$$m_X(t) = m_X(t+d) = m_X(0),$$
$$R_X(t,s) = R_X(t+d, s+d) = R_X(\tau),$$
$$C_X(t,s) = C_X(t+d, s+d) = C_X(\tau),$$
$$R_{XY}(t,s) = R_{XY}(t+d, s+d) = R_{XY}(\tau),$$
$$C_{XY}(t,s) = C_{XY}(t+d, s+d) = C_{XY}(\tau).$$

- strictly stationary

  pdf is shift invariant



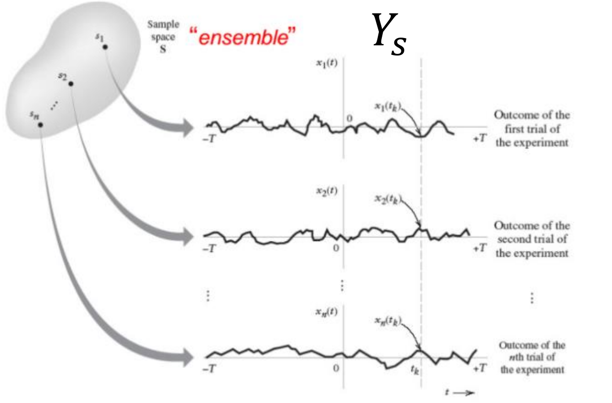


Figure from (Haykin & Moher, 2010)

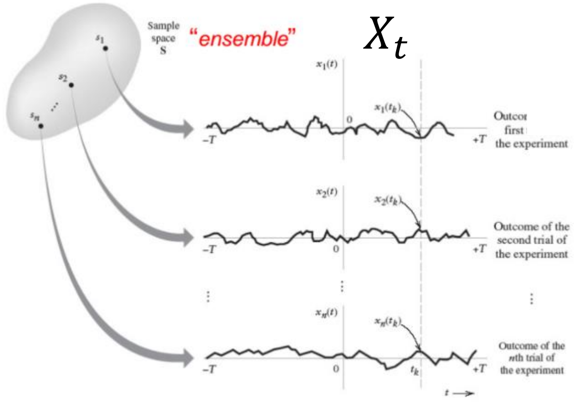


Figure from (Haykin & Moher, 2010)

Power spectral density

- Power spectral density, psd, $S_X(f)$ for wss X_t:

$$S_X(f) := \begin{cases} \sum_{\tau=-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} & \text{disc} - \text{time} \\ \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau & \text{cont} - \text{time} \end{cases}$$

- Fourier inversion:

$$R_x(\tau) = \begin{cases} \int_{-\frac{1}{2}}^{\frac{1}{2}} S_X(f) e^{j2\pi f\tau} df & \text{disc} - \text{time} \\ \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df & \text{cont} - \text{time} \end{cases}$$

Linear time-invariant system

- Linear time-invariant system with impulse response $h(t)$:

$$X_t \longrightarrow \boxed{h(t)} \longrightarrow Y_t$$

$$Y_t = \int_{-\infty}^{\infty} h(t-\tau) X_\tau d\tau$$

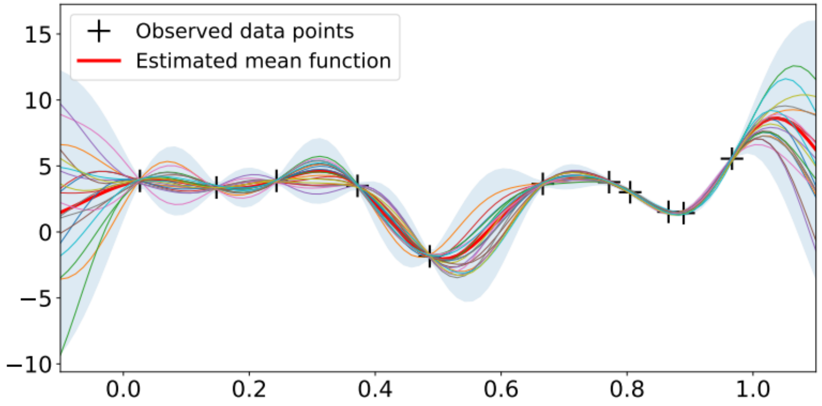
- If X_t is wss, Y_t is also wss.

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$$

$$S_Y(f) = S_X(f)\, |H(f)|^2$$

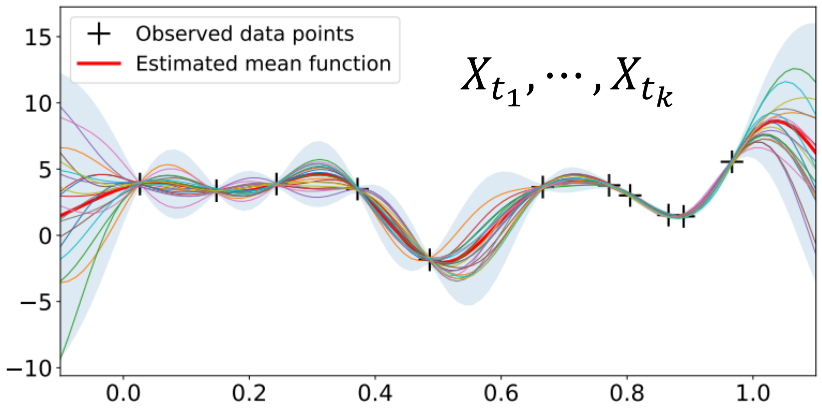
Gaussian process

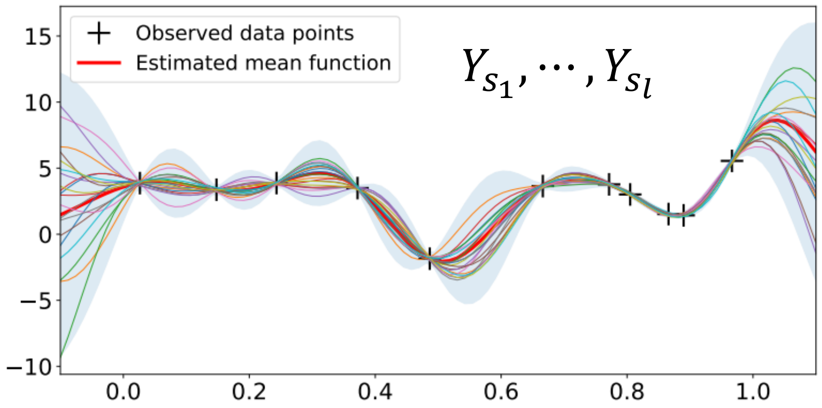
- Gaussian process: for any choice of $t_1, \cdots, t_k$, $(X_{t_1}, \cdots, X_{t_k})^T$ is a Gaussian random vector.
  1. A Gaussian random process is fully characterized by its 1st and 2nd moment, ie, by $m_X(t)$ and $R_X(t, s)$.
  2. Any linear or affine transformation of a Gaussian random process is Gaussian, eg, integral or stable linear filtering.
  3. If samples of a Gaussian random process are uncorrelated, they are independent.
  4. If a Gaussian random process is wss, it is sss.



Gaussian process

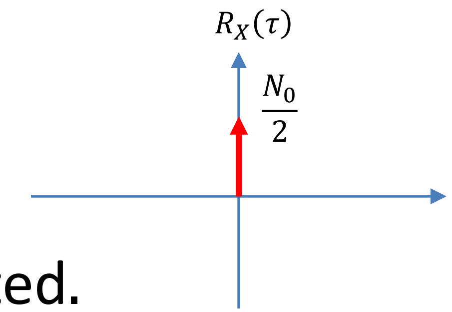
- Jointly Gaussian process: for any choice of $t_1, \cdots, t_k$ and $s_1, \cdots, s_l$, $(X_{t_1}, \cdots, X_{t_k}, Y_{s_1}, \cdots, Y_{s_l})^T$ is a Gaussian random vector.
  1. Jointly Gaussian random process are fully characterized by their 1st and 2nd moments, ie, by $m_X(t), m_Y(t), R_X(t,s), R_Y(t,s)$ and $R_{XY}(t,s)$.
  2. Any linear or affine transformation of a jointly Gaussian random process is Gaussian.
  3. If two jointly Gaussian random process are uncorrelated, they are independent.
  4. If jointly Gaussian random process are jwss, they are jsss.





White noise

- White noise : a wss process X_t with psd $S_X(f) = \frac{N_0}{2}$



$\Rightarrow m_X(t) = 0$
$\Rightarrow R_X(\tau) = \frac{N_0}{2}\delta(\tau)$: $\forall\, \varepsilon > 0, X_t$ and $X_{t+\varepsilon}$ are uncorrelated.
$\Rightarrow P_X = EX_t^2 = \int_{-\infty}^{\infty} \frac{N_0}{2} df = \infty$: infinite average power

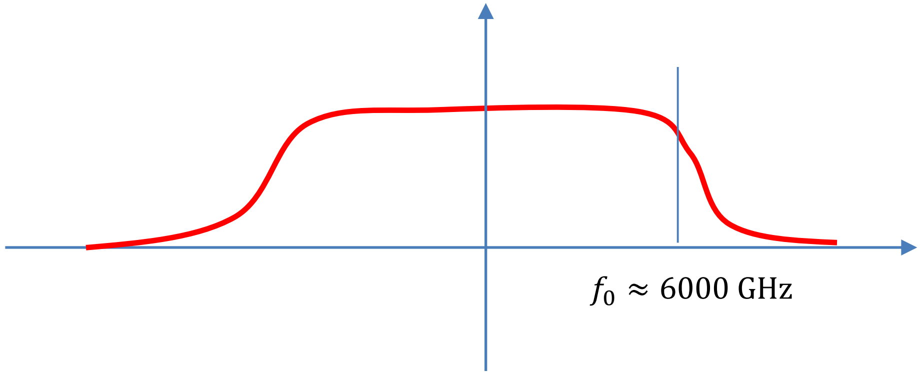
- If X_t is Gaussian, $\forall\varepsilon > 0, X_t$ and $X_{t+\varepsilon}$ are independent.
- In reality what is called a white noise has a psd that is constant up to about 1000 GHz and then gradually tapers off.

White noise

- White noise : a wss process X_t with psd $S_X(f) = \frac{N_0}{2}$
  - Thermal noise model: $S_X(f) = \frac{N_0}{2}\left(\frac{\frac{|f|}{f_0}}{\exp(\frac{|f|}{f_0})-1}\right)$,

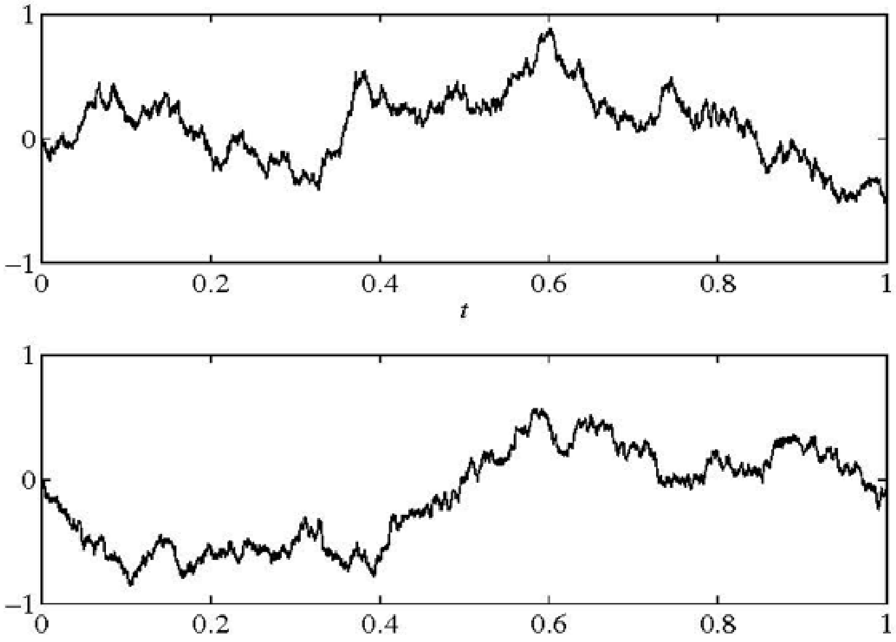
    $f_0 \approx 6000$ GHz, Gaussian $\Rightarrow$ If $\varepsilon > 0.17$ pico-seconds,

    X_t and $X_{t+\varepsilon}$ are considered independent.
  - discrete-time white noise: uncorrelated wss process



Wiener process

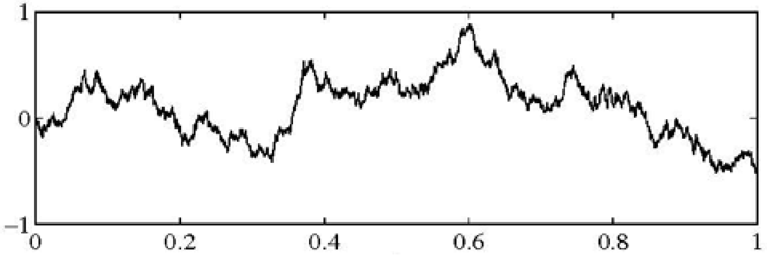
- A Wiener process W_t, also called Brownian motion, describes the motion of a highly excited particle in a fluid, viewed in one coordinate, that does not drift off in one direction.



Wiener process

- Wiener process for $t \geq 0$:
  1. $W_0 = 0$.
  2. For $s < t, W_t - W_s$ is a Gaussian random variable with mean zero and variance $\sigma^2(t-s)$.
  3. For $t_1 < t_2 < \cdots < t_k$, the increments $W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \cdots, W_{t_{k-1}} - W_{t_k}$ are independent: independent increment process.
  4. Each sample path is a continuous function of t.

- $W_t = \int_0^t X_\tau d\tau, t \geq 0$, where X_t is a white noise.



Wiener process

- mean, auto-correlation, auto-covariance

$$EW_t = 0, \mathrm{var}(W_t) = \sigma^2 t = E[W_t^2] = \int_0^t E[X_\tau^2]d\tau$$

For $t > s, \mathrm{cov}(W_t, W_s) = ?$

$$\begin{aligned}\mathrm{cov}(W_t, W_s) &= EW_tW_s - EW_tEW_s = EW_tW_s \\ &= E(W_t - W_s + W_s)W_s = E(W_t - W_s)W_s + EW_s^2 \\ &= E(W_t - W_s)EW_s + EW_s^2 = \sigma^2 s\end{aligned}$$

$$\Rightarrow R_W(t,s) = EW_tW_s = \sigma^2 \min(t,s)$$

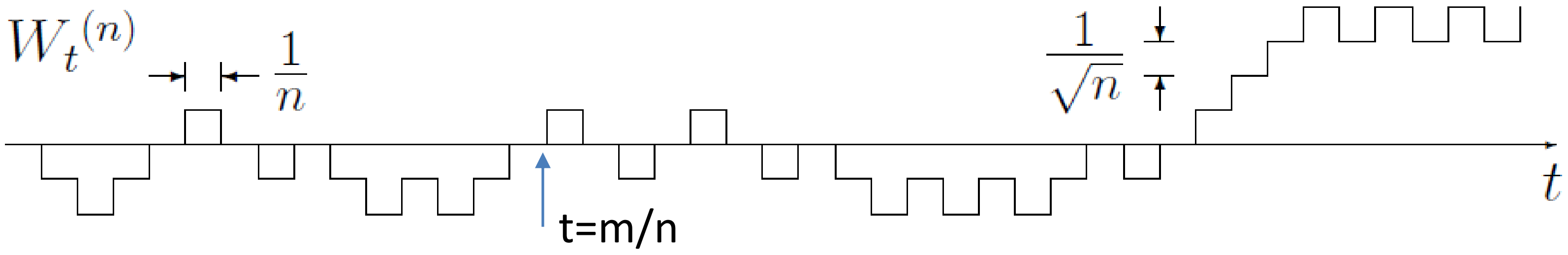
$$\Rightarrow C_W(t,s) = \mathrm{cov}(W_t, W_s) = \sigma^2 \min(t,s)$$

- Wiener processes are nonstationary.

Random walk approx. of Winner process

- random walk approximation:
  $X_1, X_2, \cdots$ is an equiprobable Bernoulli process with values $+1$ and -1,
  $\sigma^2 = p(1-p)$.
  $S_m := \Sigma_{i=1}^{m} X_i$: symmetric random walk

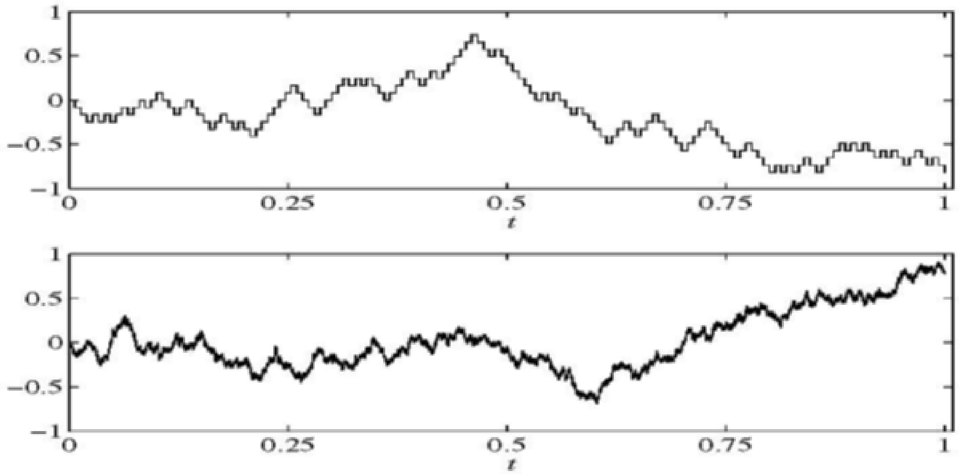
  $W_t^{(n)} := \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor} = \frac{1}{\sqrt{n}} S_m,$
  where $\lfloor \tau \rfloor$ is the greatest integer no greater than τ.



$W_t^{(n)} := \Sigma_{i=1}^{m} \frac{1}{\sqrt{n}} X_i \Rightarrow E\left[{W_t^{(n)}}^2\right] = \Sigma_{i=1}^{m} \frac{1}{n} E[X_i^2] = \frac{1}{n} m\sigma^2 = \sigma^2 t$

Random walk approx. of Winner process

- As $n \to \infty$,
  1. The power of the process is maintained.
  2. By the central limit theorem, $W_t^{(n)}$ becomes Gaussian.
  3. $W_t^{(n)}$ becomes a Wiener process.
  4. As the random walk is an independent increment process, so is the Wiener process.



Markov process

- We discuss jointly discrete cases; jointly continuous cases are similar.

- Markov property:
  For any $t_1 < t_2 \cdots < t_n$ and $x_1, \cdots, x_n$,
  $$p(x_n | x_1, \cdots, x_{n-1}) = p(x_n | x_{n-1})$$

- Markov process: a process with the Markov property
  $$\begin{aligned} &p(x_1, \cdots, x_n) \\ &= p(x_1)p(x_2|x_1) \cdots p(x_n|x_1, \cdots, x_{n-1}) \\ &= p(x_1)p\big(x_2|x_1\big)p(x_3|x_2) \cdots p(x_n|x_{n-1}) \end{aligned}$$

- Examples: binomial counting process, random walk, Poisson process, Wiener process

Markov process

- Chapman-Kolmogorov equation for a Markov process:

$$p(x_3|x_1) = \sum_{x_2} p(x_2|x_1)p(x_3|x_2)$$

proof:

$$\begin{aligned} p(x_3|x_1) &= \sum_{x_2} p(x_2, x_3|x_1) \\ &= \sum_{x_2} p(x_2|x_1)p(x_3|x_1, x_2) \ \ [\mathrm{ch}] \\ &= \sum_{x_2} p(x_2|x_1)p(x_3|x_2) \ [\mathrm{Mp}] \end{aligned}$$

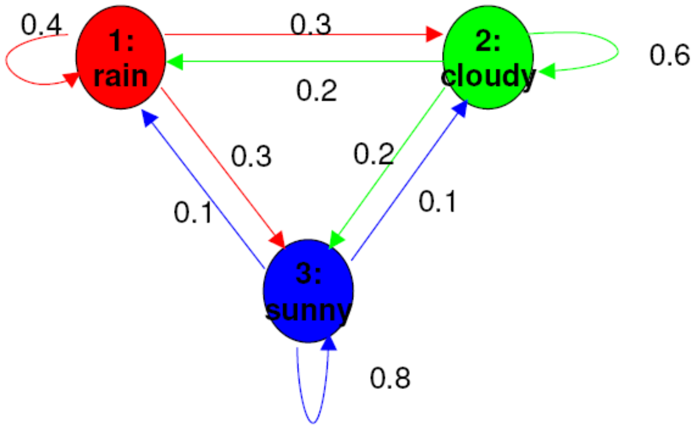
- Generalization

$$p(x_k|x_1) = \sum_{x_2} \cdots \sum_{x_{k-1}} p(x_2|x_1)p(x_3|x_2) \cdots p(x_k|x_{k-1})$$

Markov process

- State transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N1} & p_{NN} & \cdots & p_{NN} \end{bmatrix}$$



  - Where

    $p_{ji} = P(X_n = j | X_{n-1} = i) \qquad 1 \le i, j \le N$

  - With constraints

$$p_{ij} \ge 0, \qquad \sum_{j=1}^{N} p_{ij} = 1$$

- Initial state probability

$$\pi_i = P(X_1 = i), \quad 1 \le i \le N \Longrightarrow \pi^{(n+1)} = \pi^{(n)} P \quad or \;\; \pi^{(n+1)} = P\pi^{(n)}$$

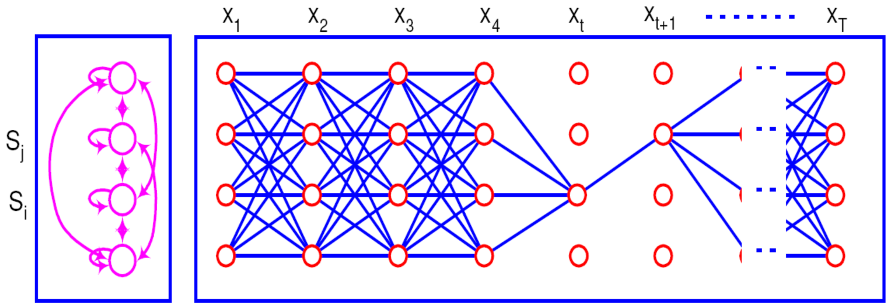
Markov process

- transition matrix P: the matrix whose (i,j)-th element is p_{ij}

  1. $\Sigma_j p_{ij} = 1$
  2. $p_{ij}^{(2)} := [P^2]_{ij} = \Sigma_k p_{ik} p_{kj}$

     $= \Sigma_k P(X_{n+1} = k | X_n = i) P(X_{n+2} = j | X_{n+1} = k)$

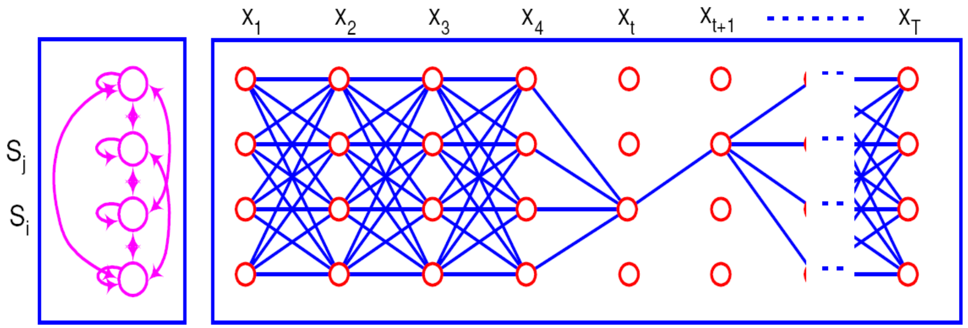
     $= P(X_{n+2} = j | X_n = i)$: Chapman-Kolmogorov equation
  3. $p_{ij}^{(m)} := [P^m]_{ij} = P(X_{n+m} = j | X_n = i)$
  4. $p_{ij}^{(n+m)} = \Sigma_k p_{ik}^{(n)} p_{kj}^{(m)}$: Chapman-Kolmogorov equation



$$\boxed{p(x_k|x_1) = \Sigma_{x_2} \cdots \Sigma_{x_{k-1}} p(x_2|x_1) p(x_3|x_2) \cdots p(x_k|x_{k-1})}$$

Markov process

5. $P(X_{n+m} = j) = \Sigma_i {p_{ij}}^{(m)} P(X_n = i)$
   $\Rightarrow p^{(n+m)} = p^{(n)} P^m$, where $p^{(n)}$ is the row vector
   whose j −th element is $P(X_n = j)$.
   $\Rightarrow p^{(n)} = p^{(0)} P^n$



Markov process

- stationary distribution for $P: \pi^T = (\pi_1, \pi_2, \pi_3, \cdots)$
  that satisfies $\pi = \pi P$ and $\Sigma_i \pi_i = 1$.
  If π exists for P of a Markov chain,

  $$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \cdots \\ \pi_1 & \pi_2 & \pi_3 & \cdots \\ \pi_1 & \pi_2 & \pi_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

  $\lim_{n\to\infty} p^{(n)} = \lim_{n\to\infty} p^{(0)} P^n = \pi$ regardless of $p^{(0)}$.
  The Markov Chain is asymptotically stationary.

Ergodicity

- Ergodicity means equality between time averages and statistical averages for wss processes.
- Assume X_t is wss; consider continuous-time cases.
- Statistical average: $EX_t = m_X$
- Time average:

$$E_T X_t = \frac{1}{T}\Sigma_{t=1}^{T} X_t \text{ or } \frac{1}{T}\int_0^T X_t dt$$
$$E_\infty X_t = \lim_{T\to\infty} E_T\, X_t$$

X_t is ergodic in the mean: $E_\infty X_t = EX_t$

- We can compute the mean by time averaging a sample path.

Ergodicity

- X_t is ergodic in the 2nd moment: $E_\infty X_t^2 = EX_t^2$
  - We can compute the 2nd moment by time averaging the square of a sample path.

- X_t is ergodic in the acf: $E_\infty X_{t+\tau}X_t = EX_{t+\tau}X_t \forall\tau$.
  - We can compute the acf by computing the time-acf of a sample path.

- X_t is ergodic : X_t is ergodic in all moments.
  - We can compute any moment by time averaging the appropriate function of a sample path.