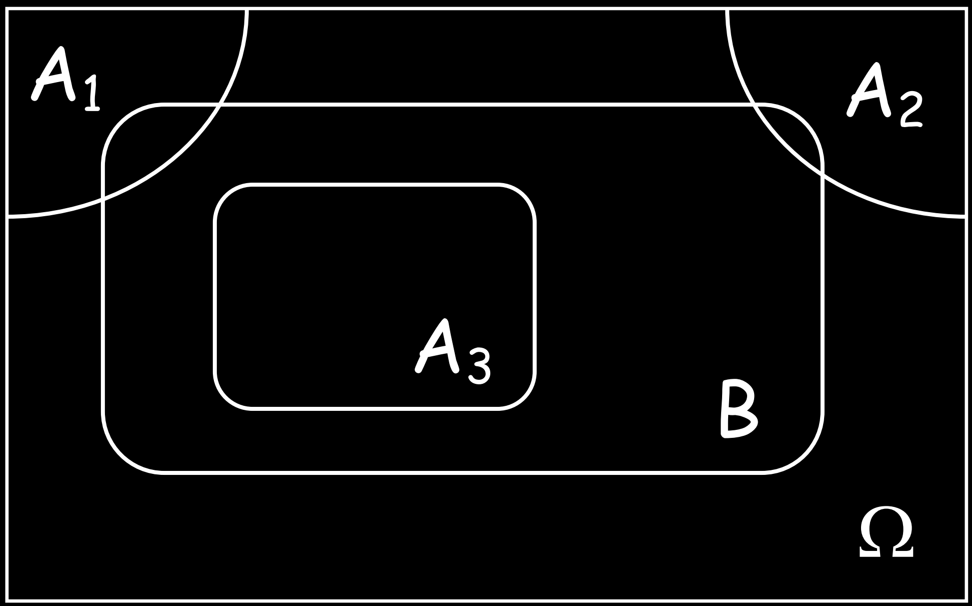
Probability Review

School of Computer and Communications Engineering

Spring 2022

Set Theory



Event: A, B;
Empty set: $\varnothing$;
Sample Space: Ω;
Complement: A^c.

Subset: $A_3 \subseteq B$; Strict Subset: $A_3 \subset B$; Belong: $A_3 \in B$;

Union: $A \cup B$; Intersection: $A \cap B$;

$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$

What is Probability?

Description of possibility.

- Measured relative frequency of occurrence of an event.
- Subjective belief about how “likely” an event is.

Two types of problems concerning probability:

- Specify the probability “model” or learn it (Statistics).
- Use the “model” to compute probability of events (Probability Theory).

We often assume the model and calculate the possibility.

Probabilistic Models

Consists of a Sample Space and a Probability Law

- Sample Space: set of all possible outcomes of an experiment
- Event: Any subset of the sample space, A
- Probability Law: Assigns a probability to every event, P(A)
- Choice of sample space (universe):

  The space should be "collectively exhaustive" (every possible outcome of experiment should be included).

  Every element should be Distinct and Disjoint (mutually exclusive).

Probability Axioms 公理

- Non-negativity: $P(A) \geq 0$ for any event A
- Additivity: Two disjoint events A and B,

  $$P(A \cup B) = P(A) + P(B)$$

- Normalization:

  Probability of the entire sample space, $P(\Omega) = 1$.

  Probability of the empty set, $P(\varnothing) = 0$.

- Discrete probability Law:

  Sample space consists of a finite number of outcomes,

  law specified by probability of single element events.

  - Example: fair coin toss, $\Omega = \{H, T\}$, $P(H) = P(T) = 1/2$
  - Discrete uniform law:

  $P(A)$ = Number of elements in A / Number of elements in Ω

Probability Axioms

- Continuous probability law:

  Ω is a range, probability of any single element event is zero, we have only the probability of event subinterval

- Properties of probability laws

  If $A \subseteq B$, then $P(A) \leq P(B)$

  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

  $P(A \cup B) \leq P(A) + P(B)$

  $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$

Conditional Probability

Given that an event B has occurred, the probability that event A occurred: P(A|B).

If number of outcome is finite & all outcomes are equally likely,

$$P(A|B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of B}}$$

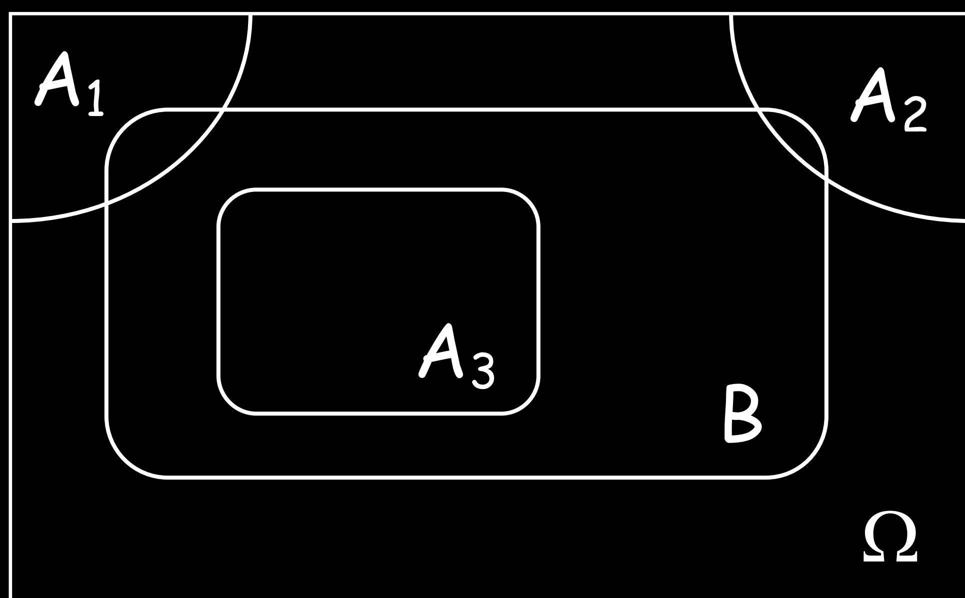
In general,

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

Total Probability Law

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of sample space ($\bigcup_{k=1}^{n} A_k = \Omega$), then for any event B,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots P(A_n \cap B)$$
$$= P(A_1)\, P(B|A_1) + \ldots + P(A_n)P(B|A_n)$$



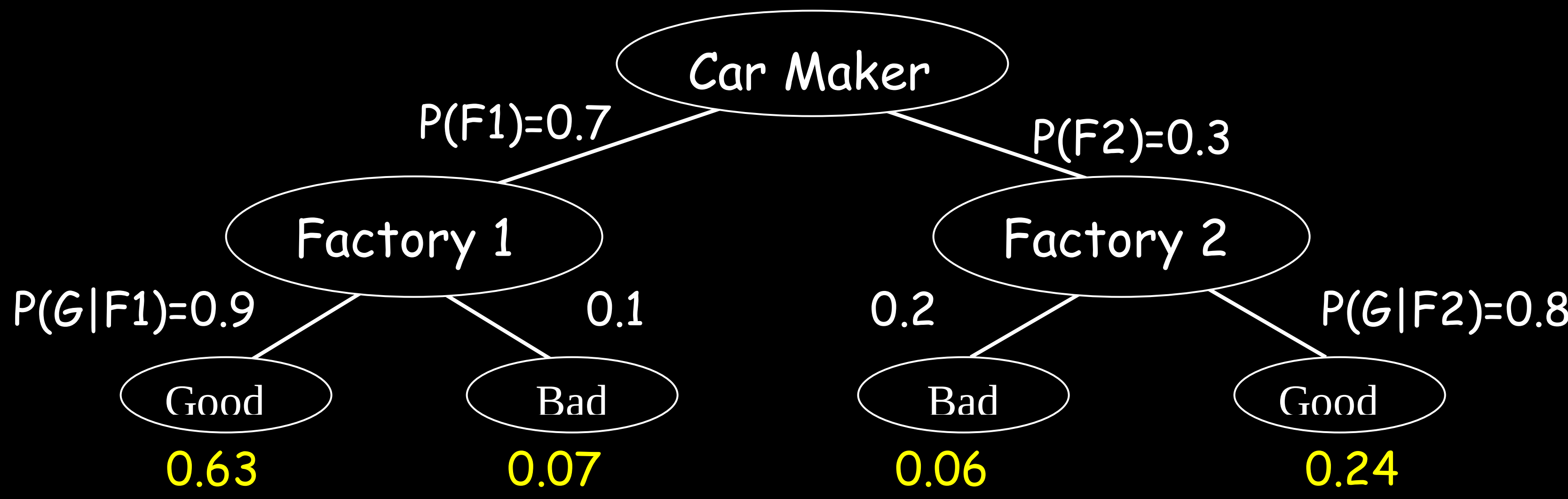
Inference using Bayes Rule

There are multiple "causes" (A_1, A_2, ..., A_n) that result in same "effect" B. Given that we observe the effect B, what is the probability that the cause was A_k ? Answer: use Bayes rule.

Let $A_1, A_2, \ldots, A_n$ be disjoint events which form a partition of the sample space, then for any event B, which has P(B) > 0,

$$P(A_k \mid B) = \frac{P(A_k)P(B \mid A_k)}{P(B)} = \frac{P(A_k)P(B \mid A_k)}{P(A_1)P(B \mid A_1) + \ldots + P(A_n)P(B \mid A_n)}$$

Car Example



What is the probability that the car is OK and was made in Factory 1?

$$P(F1|G) = \frac{P(F1)P(G|F1)}{P(F1)P(G|F1) + P(F2)P(G|F2)} = \frac{0.63}{0.63 + 0.24} = 0.72$$

Independence

If $P(A|B) = P(A)$, so $P(A \cap B) = P(B)P(A)$, A & B are Independent to each other.

Example: A= head in first toss, B = head in second toss.

"Independence" ≠ "mutually exclusive" (disjoint)

- A and B are Disjoint if $P(A \cap B) = 0$. They cannot be independent if $P(A) > 0$ and $P(B) > 0$.
  Example: A = head in a toss, B = tail in a toss
- Independence is on event sequence. Independent events with $P(A) > 0$, $P(B) > 0$ cannot be disjoint.
- For a set of Independent events,

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2) \ldots P(A_n)$$

Monty Hall Game Show

We have a prize in one of 3 sealed envelopes. You pick one. I open one of two remaining, and show you it is empty. Now you got the chance to swap your envelope. Should you swap?

Jobs of Probability

Description, Inference, Predictions, Decision Making, on:

Random Events:

- The waiting time at a toll booth is greater than 5 minutes
- An electronic device fails after 1000 hours

Random Variables:

- The waiting time τ to access a network ($0 < \tau < \infty$)
- The lifetime τ of a hard-disk ($0 < \tau < \infty$)

Random Process:

- Temperature, Voltage, Current, measured as a function of time
- A pixel, $I(x,y)$, at position (x,y) of an Image

Cumulative Distribution Function (cdf)

$$F_X(\alpha)=P\{X \leq \alpha\}$$

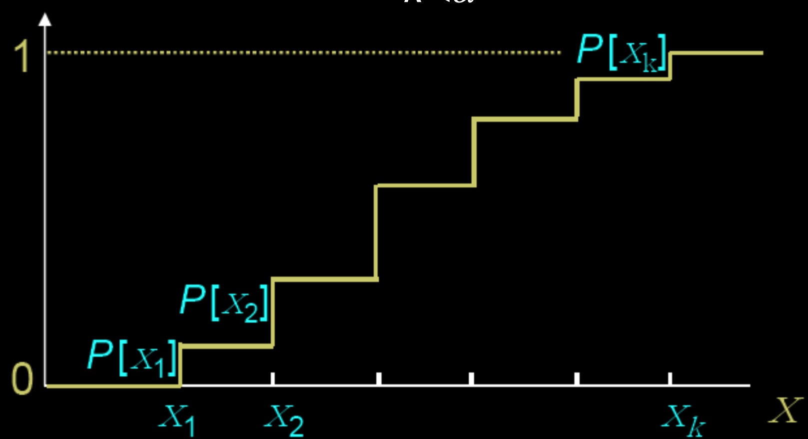
Random Event

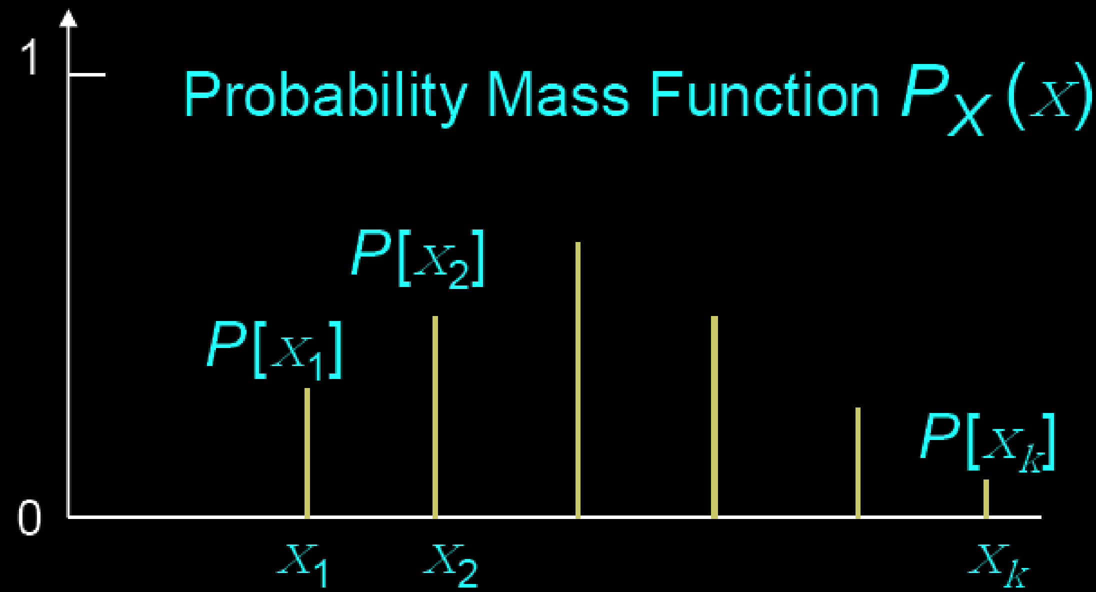
X , takes one of k possible values $\{x_1, x_2, \ldots, x_k\}$, is characterized by the probabilities:

$$P(X = x_1),\ P(X = x_2),\ \ldots,\ P(X = x_k).$$

cdf:

$$F_X(\alpha) = \sum_{k<\alpha} P(X = k)$$





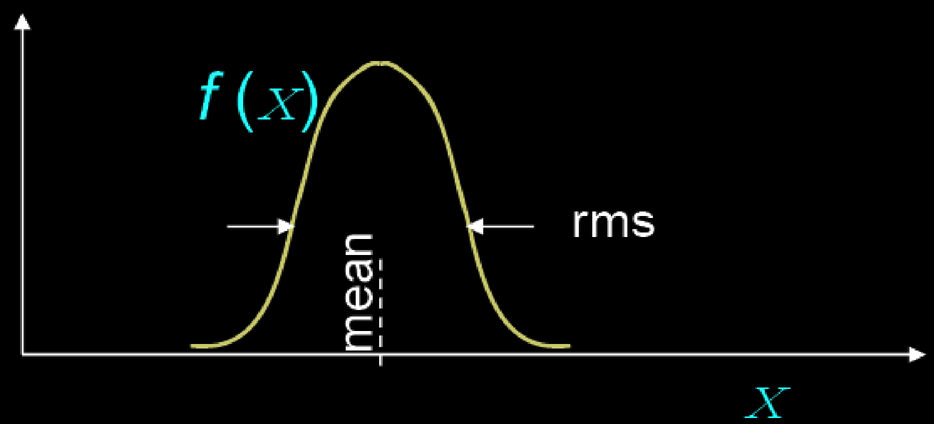
Random Variable

X , takes continuous values, is characterized by a continuous Probability Distribution Function (pdf):

Probability Density Function (pdf)

cdf:

$$F_X(\alpha) = \int_{-\infty}^{\alpha} p_X(t)\,dt$$



Mean: the centroid of the distribution;

RMS: the width of the distribution, a measure of uncertainty degree. (Root Mean Square)

Multiple Discrete Random Variables

Given two discrete r.v.s X and Y,

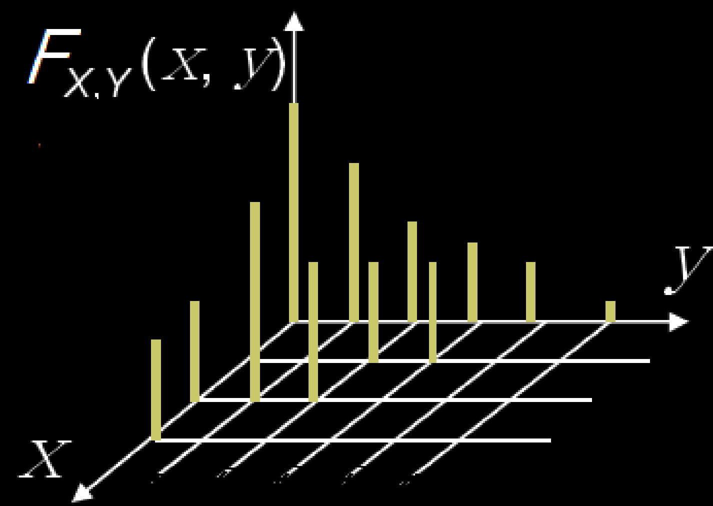
Joint pmf:

$$F(x, y) = P(X = x, Y = y) = P(\{X = x\} \cap \{Y = y\})$$

$$P(\,(x, y) \in A\,) = \Sigma_{(x, y) \in A}\, F(\,x, y)$$

Marginal pmf:

$$F_x(x) = \sum_y F(x, y)$$



Conditioning

pmf of X conditioned on an event A with P(A):

$$F_x(x \mid A) = P(\{X = x\} \mid A) = \frac{P(\{X = x\} \bigcap A)}{P(A)}$$

pmf of X conditioned on $\{Y = y\}$:

$$F_X(x \mid y) = P(\{X = x\} \mid \{Y = y\})$$

$$= \frac{P(\{X = x\} \bigcap \{Y = y\})}{P(\{Y = y\})} = \frac{F_{X,Y}(x, y)}{F_Y(y)} \qquad \sum_x F_X(x \mid y) = 1 \quad , \quad \sum_y F_Y(y \mid x) = 1$$

Independence:

$F_X(x|A) = F_X(x)$, for all x ; or

$F_X(x|y) = F_X(x)$, for all x and y; or

$F_{X,Y}(x, y) = F_X(x)F_Y(y)$; or $E\{X,Y\} = E\{X\}\,E\{Y\}$.

Bayes Rule & Conditional Expectation

Bayes rule:

$$F_X(x \mid y) = \frac{F_{X,Y}(x,y)}{F_Y(y)} = \frac{F_Y(y \mid x)F_X(x)}{\sum_\alpha F_Y(y \mid \alpha)F_X(\alpha)}$$

Expectation conditioned on Event A:

$$\mathrm{E}\{X \mid A\} = \sum_x [xF_X(x \mid A)], \quad \mathrm{E}\{g(X) \mid A\} = \sum_x [g(x)F_X(x \mid A)]$$

Expectation conditioned on $\{Y = y\}$:

$$\mathrm{E}\{X \mid Y = y\} = \sum_x [xF_X(x \mid y)]$$

Total Expectation Theorem:

$$\mathrm{E}\{X\} = \sum_y [F_Y(y)E\{X \mid Y = y\}]$$

Discrete Random Function

Given $Z = g(X, Y)$ and pmf of (X, Y), $F_{X,Y}(x,y)$:

pmf of Z: $F_Z(z) = \Sigma_{g(x,y)=z} F_{X,Y}(x,y)$

Expected value of Z:

$E\{Z\} = \Sigma_z [z F_Z(z)] = \Sigma_{(x,y)} g(x,y) F_{X,Y}(x,y)$

Given Joint pmf of n r.v.s:

$$F_{x_1,x_2,\ldots,x_n}(x_1, x_2, \ldots, x_n)$$

Marginal pmf:

$$F_{x_1,x_2,\ldots,x_{n-1}}(x_1, x_2, \ldots, x_{n-1}) = \sum_{x_n} F_{x_1,x_2,\ldots,x_n}(x_1, x_2, \ldots, x_n)$$

$$F_{x_1}(x_1) = \sum_{x_2,\ldots,x_n} F_{x_1,x_2,\ldots,x_n}(x_1, x_2, \ldots, x_n)$$

Multiple Continuous Random Variables

Two r.v.s X and Y are jointly continuous,

Joint pdf: a function $f(x, y) > 0$ that gives

$$P((X, Y) \in B) = \int_B f(x, y)\, dxdy$$

for all subsets B of the 2D plane.

Marginal/Conditional pdf: obtained by integrating the joint pdf over the entire range of one r.v.

$$f(x \mid y) = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx}$$

Conditional

Marginal

Independence

X and Y are independent iff:

$$f(x|y) = f(x) \quad \text{or} \quad f(y|x) = f(y)$$

$$\text{or } f(x,y) = f(x)\, f(y)$$

If X and Y are independent:

- Any two events $\{X \in A\}$ & $\{Y \in B\}$ are independent;
- $E\{g(X)\, h(Y)\} = E\{g(X)\}\, E\{h(Y)\}$;
- $\mathrm{Var}\{X+Y\} = \mathrm{Var}\{X\} + \mathrm{Var}\{Y\}$.
- $\mathrm{Cov}(X,Y) = E\{\, [X - E\{X\}]\, [Y - E\{Y\}]^T \,\} = ?$

Random Function

Two scalar RV: X and Y, 3 constants a, b, c,

$$E\{aX + bY + c\} = a\,E\{X\} + b\,E\{Y\} + c$$

$$\mathrm{Var}\{aX + bY + c\} = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$$

Matrix Version:

$$E\{AX + B\,Y + c\} = A\,E\{X\} + B\,E\{Y\} + c$$

$$\mathrm{Cov}\{AX + B\,Y + c\} = A\,\mathrm{Cov}(X)\,A^T + B\,\mathrm{Cov}(Y)\,B^T + A\,\mathrm{Cov}(X, Y)\,B^T + B\,\mathrm{Cov}(Y, X)\,A^T$$

Distribution of Random Function

Given $Y = g(X)$ and pdf of X, $f_X(x)$:

cdf of Y: $F_Y(y) = P(\, g(X) < y\,) = \int_{-\infty}^{g^{-1}(y)} f_X(x)dx$

let $h(y) = g^{-1}(y)$, pdf of Y:

$$f_Y(y) = \frac{\partial F_Y(y)}{\partial y} = \frac{\partial F_X(h(y))}{\partial y} = \frac{dF_x(h)}{dh}\frac{dh}{dy} = f_X(h)\frac{dh}{dy}$$

Given $Z = g(X, Y)$ and pdf of (X, Y), $f_{X,Y}(x,y)$:

cdf of Z: $F_Z(z) = P(\, g(X,Y) < z\,)$

$$= \iint_{g(x,y)<z} f_{X,Y}(x,y)dxdy$$

Characteristic Function & Moment

CF: Fourier transform of pdf (the 1st CF)

$$\Phi_X(\omega) = E\{e^{-j\omega X}\} = \int_{-\infty}^{\infty} e^{-j\omega t} p_X(t)dt$$

$$\Rightarrow \quad p_X(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\omega t}\Phi_X(\omega)d\omega$$

Moment: defined with the *k*th-order differential of CF.

$$m_k = j^{-k} \left. \frac{d^k \Phi_X(\omega)}{d\omega^k} \right|_{\omega=0}$$

Cumulant

For Vector Random Process:

$$\Phi_X(\omega_1,\omega_2,\cdots,\omega_n) = E\{\exp[-j(\omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n)]\}$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-j(\omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n)} p(X_1, X_2, \cdots, X_n) dX_1 dX_2 \cdots dX_n$$

Logarithm CF (LCF / the 2nd CF):

$$\Psi_X(\omega_1,\omega_2,\cdots,\omega_k) = \ln[\Phi_X(\omega_1,\omega_2,\cdots,\omega_k)]$$

Cumulant: by the *r*th-order partial differential of LCF,

$$c_r \triangleq (-j)^r \left.\frac{\partial^r \Psi(\omega_1,\omega_2,\cdots,\omega_n)}{\partial\omega_1^{k_1}\partial\omega_2^{k_2}\cdots\partial\omega_n^{k_n}}\right|_{\omega_1=\omega_2=\cdots\omega_n=0}$$

where $r = k_1 + k_2 + \ldots + k_n$

Higher Order Spectrum (HOS)

HOS: multi-dimensional FT of Cumulant of Real VR

$$S_r(\omega_1,\omega_2,\cdots,\omega_{r-1})$$
$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-j(\omega_1\tau_1+\omega_2\tau_2+\cdots+\omega_{r-1}\tau_{r-1})} c_r(\tau_1,\tau_2,\cdots,\tau_{r-1})\,d\tau_1 d\tau_2\cdots d\tau_{r-1}$$

Bispectrum (3rd order HOS)

$$S_{3,X}(\omega_1,\omega_2)$$

Trispectrum (4th order HOS)

$$S_{4,X}(\omega_1,\omega_2,\omega_3)$$

Some useful Theorems

For two functions of Independent r.v.s:

$$E\{ g(X)\, h(Y) \,|_{Y=\beta} \} = h(\alpha)\, E\{ g(X) \,|_{Y=\beta} \};$$

$$E\{ g(X)\, h(Y) \} = E\{ h(Y)\, E\{ g(X) \,|\, Y\} \};$$

$$Var(X) = E\{ Var(X|Y) \} + Var(E\{X|Y\});$$

$$Cov(X) = E\{ Cov(X|Y) \} + Cov(E\{X|Y\});$$ Matrix version

For 3 Independent random variables:

$$Cov(X,Y) = E\{ Cov(X,Y|Z) \} + Cov(E\{X|Z\}, E\{Y|Z\}).$$

For a Independent random process $\{X\}$:

$$y_k = g_k(X_1, \ldots, X_n), \quad k=1,2, \ldots, n;$$

$$p(y_1, \ldots, y_n) = p(X_1, \ldots, X_n) \quad \mathrm{J}^{-1}(X_1, \ldots, X_n)$$

$$\mathrm{J}(X_1, \ldots, X_n) = \begin{vmatrix} \frac{\partial g_1}{\partial X_1} & \cdots & \frac{\partial g_1}{\partial X_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial X_1} & \cdots & \frac{\partial g_n}{\partial X_n} \end{vmatrix}$$

Central Limit Theorem

If many independent random variables are summed, the pdf of the sum tends toward the Gaussian density, no matter what their individual densities are.