3F3 Statistical Signal Processing

Howard Mei

October 26, 2020

1 Probability Space

1.1 Notation

- $x \in \mathbf{A}$ x is an element of $\mathbf{A}$ "Set membership"
- $\mathbf{A} \subseteq \boldsymbol{\Omega}$ $\mathbf{A}$ is a subset of $\boldsymbol{\Omega}$
- $\mathbf{A} \subset \boldsymbol{\Omega}$ $\mathbf{A}$ is a proper subset of $\boldsymbol{\Omega}$
- $\mathbf{A} \cup \mathbf{B}$ Union of two sets
- $\mathbf{A} \cap \mathbf{B}$ Intersection of two sets
- $\mathbf{A^c}$ Complementary Set
- $\mathbf{A} \backslash \mathbf{B}$ $\mathbf{A} \cap \mathbf{B^c}$ intersection of A with not B
- $\emptyset$ Empty set

1.2 Probability Space

- **Random experiment** is used to describe any situation which has a set of possible outcomes, each of which occurs with a particular probability.
- **Sample space** Ω is the set of all possible outcomes of the **random experiment**.
- **Event** any subset $\mathbf{A} \subseteq \boldsymbol{\Omega}$
- **Probability** P mapping/function from events to a number in the interval $[0, 1]$. Therefore, specify $\{P(\mathbf{A}), \mathbf{A} \subset \boldsymbol{\Omega}\}$
- **Probability Space** defined as: (Ω, P)
- **Indicator function** for a set or event E defined as:

$$\mathbb{I}_E(t) = \begin{cases} 1 \text{ if } t \in E, \\ 0 \text{ if } t \notin E \end{cases}$$

- Examples:
  - Toss a coin twice. $\Omega = \{HH, HT, TH, TT\}$ - Finite set

– The temperature is a perturbation of seasonal average. $\Omega = (-\infty, \infty)$ - Real line

– Toss a coin n times. One elementary outcome is $\omega = (o_1, o_2, ..., o_n)$

$$\Omega = \{\omega = (o_1, o_2, ..., o_n) : o_i \in \{H, T\}\}.$$

– Toss a coin n times, the event **E** that the first head Occurs on third toss is:

$$\mathbf{E} = \{\omega = (T, T, H, o_4, o_5, ..., o_n) : o_i \in \{H, T\} \text{ for } i > 3\}.$$

$$P(\mathbf{E}) = (1/2)^3$$

1.3 Axioms of probability

A probability P assigns each event **E**, $\mathbf{E} \subset \Omega$, a number in [0,1] and P must satisfy following properties:

- $P(\Omega) = 1$
- For events A,B such that $\mathbf{A} \cap \mathbf{B} = \emptyset$ (i.e. disjoint) then $P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B})$
- if $A_1, A_2...$ are disjoint then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.
- The third one implies the second one.

Examples:

(i)Show that, if event $\mathbf{A} \subset \mathbf{B}$ then $P(A) \leq P(B)$.

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

(ii)Show that, $P(A^c) = 1 - P(A)$

$$\Omega = A \cup A^c$$
$$P(\Omega) = P(A) + P(A^c) = 1$$

(iii)Defining P: Ω is a finite discrete set, i.e. $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$. Let $p_1, p_2, ..., p_n$ be non negative numbers that add to 1. For any event A, set,

$$P(A) = \sum_{i=1}^{n} \mathbb{I}_A(\omega_1) P_i$$

Let $P_i = 1/n$. Then

$$P(\{\omega_i\}) = p_i = 1/n$$

i.e. each outcome is equally likely. This is the *uniform probability distribution.*

1.4 Conditional Probability

- Definition: The conditional probability of event A occurring given that event B has occurred :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ for } P(B) > 0$$

  - Think of $P(A|B)$ as the fraction of times A occurs among those in which B occurs.
  - AB is shorthand for $A \cup B$
  - Example: Verify any set given set G is a probability i.e. $P(\cdot|G)$ is a probability

$$\text{Firstly, } P(\Omega|G) = P(\Omega \cup G)/p(G) = 1$$

$$\begin{aligned}\text{Secondly, for disjoint events A and B } P(A \cap B|G) &= P(AG \cap BG)/p(G) \\ &= (P(AG) + P(BG))/p(G) \\ &= P(A|G) + P(B|G)\end{aligned}$$

- Probability Chain Rule

$$P(A_1...A_n) = P(A_1)P(A_2|A_1)...P(A_n|A_{n-1},...,A_1) = P(A_1)\prod_{i=2}^{n} P(A_i|A_{i-1},...,A_1) = \prod_{i=1}^{n} P(A_i|A_{i-1},...,A_1)$$

- Independence: two events A and B are independent if

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

  - if A and B are independent then $P(A|B) = P(A)$

- *Bayes' Theorem*:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

  - Example: A is the event the email is spam and B is the event the email contains "free". We know $P(B|A) = 0.8$ and $P(B|not\ A) = 0.1$ and $P(A) = 0.25$ What is the probability the email is spam given the email contains "Free"?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.8 * 0.25}{0.8 * 0.25 + 0.1 * 0.75} = 0.727$$

  - This is an example of an *expert* system.

1.5 Random Variables

- Definition: Given a probability space(Ω, P), a random variable is a function $X(\omega)$ which maps each element ω of the sample space Ω onto a point on the real line.

  - Example: Flipping a coin twice. Sample Space: $\Omega = \{HH, HT, TH, TT\}$ Define $X(\omega)$ be the number of heads.

ω	$P(\{\omega\})$	$X(\omega)$
TT	0.25	0
TH	0.25	1
HT	0.25	1
HH	0.25	2

x	$\Pr(X = x)$
0	0.25
1	0.5
2	0.25

  - The second table does not mention the sample space. The range of X is listed along with the probability associated.
  - However, there is a sample space lurking behind every definition of a rv.
  - The Probability that $X = x$ is inherited from the definition of (Ω, P) and the mapping $X(\omega)$
  - For any set $A \subset (-\infty, \infty)$, we define

$$Pr(X \in A) = P(\{\omega : X(\omega) \in A\})$$

- Discrete random variable: range is a finite set, say $\{x_1, ..., x_i, ..., x_M\}$ or a countable set, say$\{x_1, x_2, ...\}$.
  - A set E is countable if you can define a one-to-one mapping from E to the set of integers .
  - Examples: all rational number, all even number. The interval$[0, 1]$ is not countable.
  - Definition: Discrete rv X with range $\{x_1, x_2, ...\}$, the pmf is the function $p_x : \{x_1, x_2, ...\} \to [0, 1]$ where

$$p_X(x_i) = Pr(X = x_i)\text{and} \sum_{i=1}^{\infty} p_X(x_i) = 1$$

    The pmf is a complete description: for any set A,

$$Pr(X \in A) = \sum_{i=1}^{\infty} \mathbb{I}_A(x_i)p_X(x_i)$$

- Continuous random variable: defined as having a probability density function(pdf)
  - Definition: A random variable is continuous if there exists a non-negative function $f_X(x) \geq 0$ such that $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ and for any set A

$$Pr(X \in A) = \int_{-\infty}^{\infty} \mathbb{I}_A(x)f_X(x)\,dx$$

  - Example: $A = [a, b]$ then

$$Pr(X \in A) = Pr(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

    * pdf f_X assigns 0 probability to any particular point $x \in \mathbb{R}$ Thus $Pr(X = x) = 0$ for all x.

$$Pr(X \in [a, b]) = Pr(X \in (a, b]) = Pr(X \in (a, b))$$

    * This means a continuous rv has no concentration of probability at points like a discrete rv does

- Cumulative distribution function: Describe both discrete and continuous random variables and is defined to be

$$F_X(x) = Pr(X \leq x)$$

  Properties:

  1. $0 \leq F_X(x) \leq 1$
  2. $F_X(x)$ is non-decreasing as x increases
  3. $\Pr(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$
  4. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$
  5. If X is a continuous r.v. then $F_X(x)$ is continuous
  6. If X is discrete then $F_X(x)$ is right-continuous: $F_X(x) = \lim_{t \downarrow x} F(t)$ for all x

  For Property 6

  – For a discrete rv with range $x_1, ..., x_i, ..., x_M$

$$F_X(x) = \sum_{j=1}^{M} P(x_j) \mathbb{I}_{[x_j, \infty)}(x) \qquad \text{([touch (not touch)}$$

  is a step function

- CDF and PDF for continuous rv

$$F_X(x) = Pr(X \leq x) = \int_{-\infty}^{x} f_x(t)\, dt$$

$$f_X(t) = \frac{dF_X(t)}{dx}$$

  – CDF is useful when transformation of a random variable

$$\begin{aligned} Y &= r(X) \qquad \text{r is a strcitly increasing fucntion} \\ F_Y(y) &= Pr(Y \leq y) \\ &= Pr(r(X) \leq y) \\ &= Pr(X \leq r^{-1}(y)) \\ &= F_X(r^{-1}(y)) \\ f_Y(y) &= f_X(r^{-1}(y)) * \frac{dr^{-1}(y)}{dy} \end{aligned}$$

2 Multivariates

2.1 Bivariates

2.1.1 Discrete bivariates

- joint pmf: $p_{X,Y}(x_i, y_j) = Pr(X = x_i, Y = y_j)$
- marginal pmf:

$$P_X(x_k) = \sum_{j=1}^{n} P_{X,Y}(x_k, y_j), \qquad P_Y(y_k) = \sum_{i=1}^{m} P_{X,Y}(x_i, y_k)$$

- Independent if:

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \text{for all (x,y)}$$

- Conditional Probability

$$p_{X|Y}(x|y) = \frac{p(X,Y)(x,y)}{P_Y(y)}$$

2.1.2 Continuous bivariates

- For continuous random variables X and Y, we call $f(x,y)$ their **Joint probability density function**:
  - $\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x,y)\,dx\right)\,dy = 1$ and
  - for any sets(events) $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$

$$Pr(X \in A, Y \in B) = \int_{-\infty}^{\infty} \mathbb{I}_B(y)\left(\int_{-\infty}^{\infty} \mathbb{I}_A(x)f(x,y)\,dx\right)\,dy$$

- Independent

$$\text{If and only if: } f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

- Conditional probability density function:

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

  Moreover, for all sets A

$$Pr(X \in A|Y = y) = \int_{-\infty}^{\infty} \mathbb{I}_A(x)f_{X|Y}(x|y)\,dx$$

  Example: Let X_1, X_2 be two independent rvs with $f_1(x_1), f_2(x_2)$ and let $Y = X_1 + X_2$. Find the pdf $f_{X_1,y}$ and f_Y.

  Write the joint pdf using conditional pdf formula:

$$f_{X_1,y}(x_1,y) = f_1(x_1)fY|X_1(y|x_1).$$

  Since $Y = X_2 + x_1,\ fY|X_1(y|x_1) = f_2(y - x_1)$

$$f_Y(y) = \int_{-\infty}^{\infty} f_2(y - x_1)f_1(x_1)\,dx_1$$

  which is the convolution of f_1 and f_2

2.1.3 Expected Value Operations

- Expectation
  - Definition: The *Expected value* or *mean value* or *first moment* of X is

$$\mathbb{E}\{X\} = \begin{cases} \sum_x \quad xp_X(x) & \text{Discrete} \\ \int_{-\infty}^{\infty} \quad xf_X(x)\,dx & \text{Continuous} \end{cases}$$

- Expectation of a function of rv
  - Definition: For any function $r(\cdot)$ compute $\mathbb{E}\{r(X)\}$ by replacing x in the above formulae with $r(x)$ For example, the higher moments are $\mathbb{E}(X^n)$ set $r(X) = X^n$
  - Example: For an event A:

$$\mathbb{E}\{\mathbb{I}_A(X)\} = \begin{cases} \sum_x & \mathbb{I}_A(X)p_X(x) & \text{Discrete} \\ \int_{-\infty}^{\infty} & \mathbb{I}_A(X)f_X(x)\,dx & \text{Continuous} \end{cases}$$

    Then $\mathbb{E}\{\mathbb{I}_A(X)\} = \Pr\{X \in A\}$
  - Example: Take a unit length stick and break it at random. Find the mean of the long piece. Call the longer piece Y and the break point X. Then X is a uniform rv in $[0,1]$, $Y =$ max $\{X, 1-X\}$ and,

$$\begin{aligned} \mathbb{E}Y &= \mathbb{E}(max\{X, 1-X\}) \\ &= \int_{-\infty}^{\infty} max\{x, 1-x\} f_X(x)\,dx \\ &= \int_0^1 max\{x, 1-x\}\,dx \\ &= \int_0^0 .5(1-x)\,dx + \int_0 .5^1 x\,dx = 0.75 \end{aligned}$$

- Expectation of a function of bivariates
  - Definition: The mean of a function $r(X, Y)$ of the bivariate (X, Y) is

$$\mathbb{E}\{r(X,Y)\} = \begin{cases} \sum_y \sum_x & r(x,y)p_{X,Y}(x,y) & \text{Discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & r(x,y)f_{X,Y}(x,y)\,dxdy & \text{Continuous} \end{cases}$$

  - The conditional expectation is

$$\mathbb{E}\{r(X,Y)|Y=y\} = \begin{cases} \sum_x & r(x,y)p_{X|Y}(x|y) & \text{Discrete} \\ \int_{-\infty}^{\infty} & r(x,y)f_{X|Y}(x|y)\,dx & \text{Continuous} \end{cases}$$

  - By using conditional probability we can calculate $\mathbb{E}\{r(X,Y)\}$:

$$\begin{aligned} \mathbb{E}\{r(X,Y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r(x,y)f_{X,Y}(x,y)\,dxdy \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} r(x,y)f_{X|Y}(x|y)\,dx\right)dy \\ &= \int_{-\infty}^{\infty} \mathbb{E}\{r(X,Y)|Y=y\}f_Y(y)\,dy \end{aligned}$$

- Rule of iterated expectation
  Discrete:

$$\mathbb{E}\{r(X,Y)\} = \mathbb{E}(\mathbb{E}\{r(X,Y)|Y\})$$

  Continuous:

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

$$\mathbb{E}\{r(X,Y)\} = \int_{-\infty}^{\infty} \mathbb{E}\{r(X,Y)|Y=y\}f_Y(y)\,dy$$

2.2 Multivariates

2.2.1 Definition

- Let $X_1, X_2, ..., X_n$ be n continuous/discrete random variables. We call $X = (X_1, ..., X_n \in \mathbb{R}^n$ a continuous/discrete random vector.
- Let $X = (X_1, ..., X_n \in \mathbb{R}^n$ be a continuous random vector. Let $f(x_1, ...x_n)$ be a non-negative function that integrates to 1. Then f is called the pdf of the random vector X if

$$Pr(X_1 \in A_1, ..., X_n \in A_n) = \int_{-\infty}^{\infty} \mathbb{I}_{A_n}(x_n) ... \int_{-\infty}^{\infty} \mathbb{I}_{A1}(x_1) f(x_1, ...x_n) d_{x_1} ... d_{x_n}$$

- pdf of X_i is obtained by integrating $f(x_1, .., x_n)$ over the full range except x_i:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1, .., x_n) d_{x_1} ... d_{x_{i-1}} d_{x_{i+1}} d_{x_n}$$

  This is called ith marginal of $f(x_1, .., x_n)$

2.2.2 Independence

- Definition: The n random variables $X_1, ...X_n$ are independent if and only if for every $A_1, ...A_n$

$$Pr(X_1 \in A_1, ..., X_n \in A_n) = Pr(X_1 \in A_1) ... Pr(X_n \in A_n)$$

- joint pdf = product of marginals:

$$f(x_1, ..., x_n) = f_{X_1}(x_1) ... f_{X_n}(x_n)$$

  - Example: The pdf $f(x_1, ..., x_n)$ of a Gaussian random vector $X = (X_1, ..., X_n)$ is

$$\frac{1}{(2\pi)^{n/2} (\det C)^{1/2}} \exp \left\{ -\frac{1}{2} (x - m) C^{-1} (x - m)^T \right\}$$

    Where $m = (m_1, ..., m_n)$ is the row vector of means and C is the covariance matrix

$$m_i = \mathbb{E}\{X_i\} \qquad \text{and} \qquad [C]_{i,j} = \mathbb{E}\{(X_i - m_i)(X_j - m_j)\}$$

    Show that if independent, $C_{i,j} = 0$ for $i \neq j$ then

$$f(x_1, ..., x_n) = f_{X_1}(x_1) ... f_{X_n}(x_n)$$

    Proof: Call $C_{i,i} = \sigma_i^2$

$$(x - m) C^{-1} (x - m)^T = \sum_{i=1}^{n} \frac{(x_i - m_i)^2}{\sigma_i^2}$$

    Hence $f(x_1, ..., x_n)$ is

$$\frac{1}{(2\pi)^{n/2} (\det C)^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - m_i)^2}{\sigma_i^2} \right\}$$

$$= \frac{1}{\sqrt{(2\pi)}\sigma_1 ... \sqrt{(2\pi)}\sigma_n} \prod_{i=1}^{n} \exp \left\{ -\frac{1}{2} \frac{(x_i - m_i)^2}{\sigma_i^2} \right\}$$
$$= f_{X_1}(x_1) ... f_{X_n}(x_n)$$

- If $X_1, ..., X_n$ are independent then

$$\mathbb{E}\{\prod_{i=1}^{n} X_i\} = \prod_{i=1}^{n} \mathbb{E}\{X_i\}$$

That is the expectation of the product is the product of expectation

2.2.3 Change of variables

- The change of variable formula can be applied to random vectors. Let

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} g_1(X_1, ..., X_n) \\ \vdots \\ g_n(X_1, ..., X_n) \end{bmatrix}$$

or

$$Y = G(X)$$

- If G is invertible then $X = G^{-1}(Y)$. Let $H(Y) = G^{-1}(Y)$. So

$$\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} h_1(Y_1, ..., Y_n) \\ \vdots \\ h_n(Y_1, ..., Y_n) \end{bmatrix}$$

- The *Jacobian* matrix of partial derivatives of $H(y)$ is formed:

$$J(y) = \begin{bmatrix} \frac{\partial}{\partial y_1} h_1 & \dots & \frac{\partial}{\partial y_n} h_1 \\ & \vdots & \\ \frac{\partial}{\partial y_1} h_n & \dots & \frac{\partial}{\partial y_n} h_n \end{bmatrix}$$

Then

$$f_Y(y) = f_X(H(y)) |\det J(y)|$$

  - Example: Let $X_1, ..., X_n$ be independent Gaussian rv where X_i is $\mathcal{N}(0, 1)$ Let S be an invertible matrix and m a column vector. Let $Y = m + SX$ where $X = (X_1, .., X_n)^T$. Show Y is also a Gaussian random vector.

    Use the Change of variable result:

$$H(Y) = S^{-1}(Y - m)$$

    The Jacobian Matrix$J(y)$:

$$J(y) = S^{-1}$$

    Applying change of variable formula gives

$$f_Y(y) = f_X(S^{-1}(y - m)) |\det S^{-1}|$$

    where $f_X(x_1, ..., x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} x^T x\right\}$

$$f_Y(y) = \frac{|\det S^{-1}|}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(y - m)^T (S^{-1})^T S^{-1} (y - m)\right\}$$

    is the density of a Gaussian vector with mean m and covariance matrix SS^T . Note that $\det S^{-1} = 1/detS$, $\det(SS^T) = \det S \det S^T = (\det S)^2$

– An affine transformation of a Gaussian vector is still a Gaussian vector. This gives a method for generating any Gaussian vector from iid Gaussian random variables.

– To Generate a $\mathcal{N}(m, \Sigma)$ vector:

  * Decompose the symmetric matrix $\Sigma = SS^T$.
  * Output $m + SX$ where $X = (X_1, ..., X_n)^T$ where $X_1, ..., X_n$ are independent $\mathcal{N}(0, 1)$

2.2.4 Characteristic function

- Definition: The characteristic function of a discrete or continuous random variable X is:

$$\varphi_X(t) = \mathbb{E}\{\exp(itX)\}, \qquad t \in \mathbb{R}$$

For a random vector $X = (X_1, X_2, ..., X_n)$,

$$\varphi_X(t) = \mathbb{E}\{\exp(it^T X)\}, \qquad t \in \mathbb{R}^n$$

Similarly to Fourier Transform, the characteristic function uniquely describes a pdf.

– Example: Show $\varphi_X(t) = \exp(itX)\exp(-\frac{1}{2}\sigma^2 t^2)$ when X is a Gaussian random variable with mean μ and variance σ^2.

$$\begin{aligned}
&\mathbb{E}\{\exp(itX)\} \\
&= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\
&= e^{it\mu} \int_{-\infty}^{\infty} e^{its} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}s^2\right) ds, \qquad \text{let } s = x - \mu \\
&= e^{it\mu} e^{-\frac{1}{2}\sigma^2 t^2} \qquad \text{Fourier transform table}
\end{aligned}$$

– Example: Compute the characteristic function $\varphi_Y(t)$ of $Y = \sum_{i=1}^{n} X_i$ where X_i are independent random variables.

$$\begin{aligned}
&\mathbb{E}\{\exp(itY)\} \\
&= \mathbb{E}\{\exp(itX_1)\exp(itX_2)... \exp(itX_n)\} \\
&= \mathbb{E}\{\exp(itX_1)\}\mathbb{E}\{\exp(itX_2)\}...\mathbb{E}\{\exp(itX_n)\} \\
&= \varphi_{X_1}(t)...\varphi_{X_n}(t)
\end{aligned}$$

– The characteristic function of the sum of independent random variables is the product of their individual characteristic functions.

– Example: (Moments) Using $\varphi_X(t)$, compute $\mathbb{E}\{X^n\}$

$$\frac{d^n}{dt^n}\varphi_X(t) = \mathbb{E}\left\{\frac{d^n}{dt^n}\exp(itX)\right\} = \mathbb{E}\{i^n X^n \exp(itX)\}$$

Thus $i^n \mathbb{E}\{X^n\} = \frac{d^n}{dt^n}\varphi_X(t = 0)$ (Putting t=0 for the above equation and make the exponential go to 1)

- Equality of characteristic functions: Suppose that X and Y are random vectors with same characteristic functions: $\varphi_X(t) = \varphi_Y(t)$ for all $t \in \mathbb{R}^n$. Then X and Y have the same probability distribution

- Example using characteristic function: Let $X_1, X_2, ..., X_n$ be independent Gaussian random variables where X_i is $\mathcal{N}(0,1)$. Then $Y = m + SX$, where $m \in \mathbb{R}^d$ where $d < n$, is the multivariate Gaussian with mean m and covariance SS^T.

  Verify the result using characteristic function, that is let $t \in \mathbb{R}^d$ and compute $\mathbb{E}\{\exp(it^T Y)\}$

$$\begin{aligned}\exp(it^T Y) &= \exp(it^T m)\exp(it^T SX) \\ &= \exp(it^T m)\exp(ir_1 X_1)...\exp(ir_n X_n)\end{aligned}$$

  Where vector $r = t^T S$

$$\begin{aligned}&\mathbb{E}\{\exp(it^T Y)\} \\ &= \exp(it^T m)\mathbb{E}\{\exp(ir_1 X_1)...\exp(ir_n X_n)\} \\ &= \exp(it^T m)\exp(-\frac{1}{2}r_1^2)...\exp(-\frac{1}{2}r_n^2) \\ &= \exp(it^T m)\exp(-\frac{1}{2}t^T SS^T t)\end{aligned}$$

3 Random process

3.1 Definition of random process

- Definition:A discrete random (or stochastic) process is one of the following infinite collection of random variables

$$\{..., X_{-1}, X_0, X_1, ...\} \quad or \quad \{X_0, X_1, ...\} \quad or \quad \{X_1, X_2, ...\}$$

  Notation: $\{X_n\}_{n_i}^{j} = \{X_i, X_{i+1}, ..., X_j\}$

  - Example: Random phase cosine. Let $X_n = \cos(2\pi fn + \phi)$ where ϕ is a Uniform random variable drawn from $[0, 2\pi)$ To generate

$$\{X_n\}_{n=0}^{\infty} = \{X_0, X_1, ...\}$$

    first sample ϕ and then set

$$X_n = \cos(2\pi fn + \phi)$$

    for $n = 0, 1, ...$

  - Example: infinite collection of independent random variables
    Let $0 < q < 1$ and $U_1, U_2, ...$ be iid discrete random variables such that

$$Pr(U_n = 1) = q, \qquad Pr(U_n = -1) = 1 - q$$

  - Example: Random walk
    Generate the sequence $U_1, U_2, ...$ as in the previous example and define a new random process $X_0, X_1 ...$ as follows: set $X_0 = 0$ and

$$X_n = X_{n-1} + U_n$$

for $n > 0$
We could equivalently write

$$X_n = \begin{cases} X_{n-1} + 1 & w.p.q \\ X_{n-1} - 1 & w.p.1 - q \end{cases}$$

and $X_0 = 0$.

- Definition (Finite dimensional distributions)
  - To completely specify a discrete time random process $X_0, X_1, ...,$ we must specify their joint probability density function
  $$f_{X_0,X_1,...,X_n}(x_0, x_1, ..., x_n)$$
  for all integers $n \geq 0$ when $X_0, X_1, ...$ is a collection of continuous random variables
  - For discrete time random process $X_0, X_1, ...,$ we must specify their joint probability mass function
  $$p_{X_0,X_1,...,X_n}(x_0, x_1, ..., x_n)$$
  for all integers $n \geq 0$
  - For any fixed n, you can treat $(X_0, X_1, ..., X_n)$ as a random vector and just as in the case of random vectors, we use their joint pdf or joint pmf to describe how the random vector should be generated.
  - For many interesting random processes, specifying $p_{X_0,X_1,...,X_n}(x_0, x_1, ..., x_n$ is not too arduous. One such process which underpins many real world statistical models is a **Markov chain**.

3.2 Markov Chain

- Example
A gambler has initial wealth r bets and keep playing until wealth is R or zero. Amount bet is b at every bet. The random process now is:

$$X_{n+1} = \begin{cases} X_n & \text{if} X_n \in \{0, R\} = 0 \\ X_n + b & w.p.q \\ X_n - b & w.p.1 - q \end{cases}$$

The generate X_{n+1}, only the value of X_n is needed and not its past values. Any discrete time random process with this property is called a Markov process.
It can be shown that the probability of wealth doubling when $q \leq 0.5$is

$$\left[1 = \left(\frac{1-q}{q}\right)^{r/b}\right]^{-} 1$$

	b=1pound	b=10pence	b=1pence
$q = 0.5$	0.5	0.5	0.5
$q = 0.49$	0.40	0.02	4.3×10^{-18}

Table: Probability of doubling wealth with an initial fortune of 10 pounds.

- Definition of Markov chain Let $\{X_n\}_{n\geq 0}$ be discrete random variables taking values in $S = \{1, ..., L\}$.

– The transition probability matrix Q is a non-negative matrix

$$\begin{bmatrix} Q_{1,1} & Q_{1,2} & \cdots & Q_{1,L} \\ Q_{1,1} & Q_{1,2} & \cdots & Q_{1,L} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{L,1} & Q_{L,2} & \cdots & Q_{L,L} \end{bmatrix}$$

and each row sums to one.

$Q_{1,L} = Pr(X_{n+1} = L|X_n = 1)$ from state 1 jump to state L is the probability of L given current state

– The conditional pmf of X_n given $X_0 = i_0, ..., X_{n-1} = i_{n-1}$

$$Pr(X_n = i_n|X_0 = i_0, ..., X_{n-1} = i_{n-1}) = Q_{i_{n-1},i_n} = Pr(X_n = i_n|X_{n-1} = i_{n-1})$$