School of Computing
National University of Singapore
CS5240 Theoretical Foundations in Multimedia

Probability Theory

signal: Pr(...)

1 Basics

Random experiment: A repeatable experiment that produces a random outcome. e.g. flipping a coin, rolling a die.

Sample space, S: The set of all possible outcomes.

Event, E: Any subset of S. $E \subset S$

Probability of an event: $\Pr(E)$**, or** $P(E)$**:** the "chance that E will happen".

Example: Flip a coin. $S = \{Head, Tail\}$, $\Pr(Head) = p$, $\Pr(Tail) = 1 - p$.

Notes:

1. $1 \geq \Pr(Event) \geq 0$
2. $\Pr(S) = 1$

3. If $E_1 \cap E_2 = \emptyset$ (called *mutually exclusive*), then $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2)$.

   Intuitively, this means the probability of E_1 *or* E_2 occurring. In general $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$, where $\Pr(E_1 \cap E_2)$ is the probability of E_1 *and* E_2 occurring.

Q: What is $\Pr(\emptyset)$?

Example: Roll a fair (unbiased) die. Let E_1 = "number is even", E_2 = "number is 3".

$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$.

Random Variable: Technically speaking, a random variable X is a function $X : S \to \mathbb{R}$. It associates every member of the Sample Space S with a real number $x \in \mathbb{R}$. Intuitively, we often use X as if it was the member of S.

Example: Roll a die. Let X denote the number that appears. Then possible values of X are {1,2,3,4,5,6}. Strictly speaking, $S = \{one, two, three, four, five, six\}$, and $X : S \to \mathbb{R}$ given by $X(one) = 1, X(two) = 2, X(three) = 3, X(four) = 4, X(five) = 5, X(six) = 6$.

Discrete vs. Continuous Random Variable: Discrete r.v. : X takes on discrete values $x_0, x_1, x_2, \ldots$

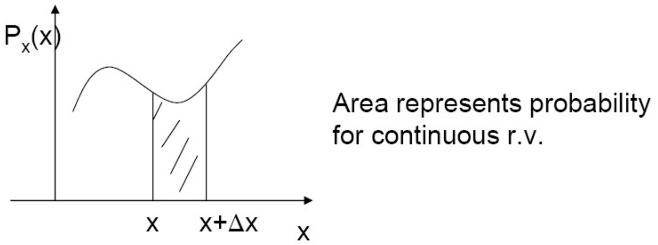
Continuous r.v. : X is continuous. $X \in [a, b]$, e.g. the height of student.

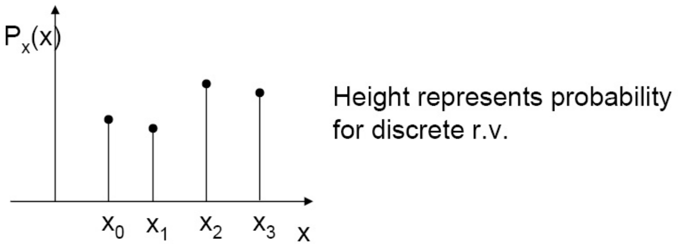
discrete

Probability Mass Function: PMF $P_X(x)$: Probability that the discrete r.v. X takes on value x.

Probability Density Funtion: PDF $P_X(x).\Delta x$: Probability that the continuous r.v. X is between $x \le X \le x + \Delta x$. 概率密度函数.

continuous





Note: $\sum_i P_X(x_i) = 1$ for discrete X, and $\int_{-\infty}^{\infty} P_X(x)dx = 1$, for continuous X.

2 Some common PMF

离散型的分布

1. Bernoulli : $P_X(x_0) = p, \quad P_X(x_1) = 1 - p$. This is for a binary sample space, e.g. flip a coin. p is called the parameter of the Bernoulli PMF. It is the only parameter.
2. Uniform : $P_X(x_i) = \frac{1}{N}, \quad i = 1, \dots, N$, e.g. roll a fair die.
3. Binomial : $P_X(k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad k = 0, 1, \dots, N$. 二项分布.

   e.g. X = number of heads when coin is flipped N times.

   The Binomial PMF has only two parameters: p and N.
4. Geometric : $P_X(k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots$. This has one parameter p.

   e.g. X = number of coin flips until first head appears. 几何分布
5. Poisson : $P_X(k) = \boxed{\frac{\lambda^k}{k!}} e^{-\lambda}, \quad k = 0, 1, 2, \dots$, and $\lambda > 0$.

   e.g. X = number of cars arriving at a traffic light in a fixed time interval T. λ is the average number of cars arriving in T time.

   The Poisson PMF has one parameter k.

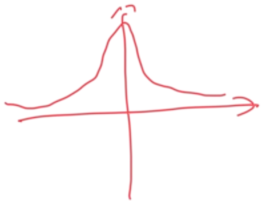
   Check : For Poisson, $\sum_{k=0}^{\infty} P_X(k) = \sum_{k=0}^{\infty} (e^{-\lambda}) \frac{\lambda^k}{k!} = e^{-\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = e^{-\lambda} e^{\lambda} = 1$.

Recall Taylor series: For any z near a, $f(z) \approx f(a) + f'(a)(z-a) + \frac{f''(a)}{2!}(z-a)^2 + \dots + \frac{f^{(k)}}{k!}(z-a)^k + \dots$

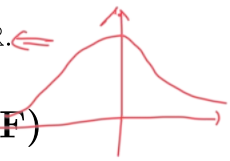
So $e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$
$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} + \dots$
$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} + \dots$

3 Some common PDF

1. Uniform (param a, b): $P_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b; \\ 0, & \text{otherwise.} \end{cases}$
   均匀分布
   back to back!
2. Exponential (param λ): $P_X(x) = \begin{cases} \boxed{\lambda e^{-\lambda x}}, & x \ge 0,\ y > 0; \\ 0, & \text{otherwise.} \end{cases}$
3. $\boxed{\text{Laplace (double exponential, param } \lambda)}$: $P_X(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad \lambda > 0.$ ⇐



4. Gaussian (param μ, σ): $P_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}.$ ⇐
   正态分布 ⇔ 高斯分布.



4 Cumulative Distribution Function (CDF)

Definition: $\boxed{F_X(x) \triangleq \Pr(X \le x).}$

CDF：累积分布函数.

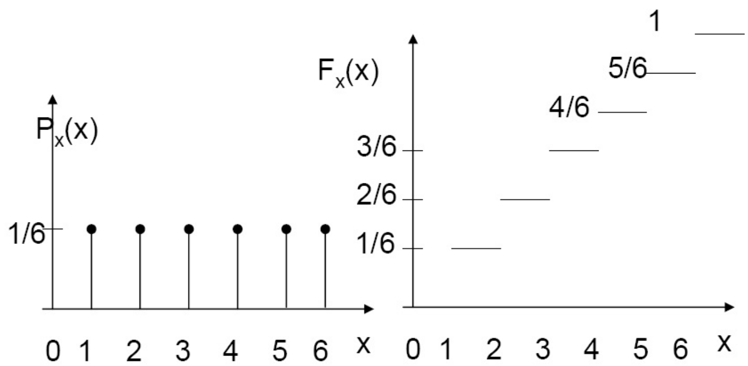
For discrete r.v. $F_X(x) = \sum_{k \le x} P_X(k)$.

For continuous r.v. $F_X(x) = \int_{-\infty}^{x} P_X(u)du$.

Notes:

1. $F_X(-\infty) = 0$
2. $F_X(+\infty) = 1$
3. For discrete r.v. $P_X(k) = F_X(k) - F_X(k-1)$.
   For continuous r.v. $P_X(k) = \frac{dF_X}{dx} |_{x=k} = F'_X(k)$.

Example: Roll a fair die. Let X be the number that appears.



5 Expectation

Let $g(X)$ be a function of r.v. X. Then $g(X)$ is also a r.v.

$E[g(x)] \triangleq \sum_x g(x)P_X(x)$ for discrete X.

$E[g(x)] \triangleq \int_x g(x)P_X(x)dx$ for continuous X ✓

Zeroth moment: $E[X^0] = 1$.

均值 (一阶矩) **Mean (1st moment):** Mean or average or expected value $\mu_X = E[X]$.

二阶矩 **2nd moment:** $E[X^2]$, etc.

二阶中心矩 **Variance (2nd central moment)** $\sigma_X^2 = Var[X] \triangleq E[(x-\mu_X)^2] = E[X^2] - \mu_X^2$. Standard deviation σ_X = positive square root of variance.

Example: $X \sim$ Poisson(λ).

That is, $P_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$.
The mean $\mu_X = E[X] = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\left(\sum_k k\frac{\lambda^k}{k!}\right)$
How to compute the sum in the parenthesis?

Consider $e^z = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^k}{k!} + \ldots$
Differentiate with respect to z : $e^z = 1 + 2\frac{z}{2!} + \cdots + k\frac{z^{k-1}}{k!} + \ldots$
Multiply by z :

$$ze^z = z + 2\frac{z^2}{2!} + \cdots + k\frac{z^k}{k!} + \ldots \tag{1}$$

which is the sum we want. Therefore $E[X] = e^{-\lambda}\lambda e^{\lambda} = \lambda = \mu_X$.

$Var[X] = E[X^2] - \lambda^2$.
$E[X^2] = \sum_k k^2 e^{-\lambda}\frac{-\lambda^k}{k!} = e^{-\lambda}\left(\sum_k k^2\frac{\lambda^k}{k!}\right)$
Differentiate Equation (1) w.r.t. z : $ze^z + e^z = 1 + 2^2\frac{z}{2!} + \cdots + k^2\frac{z^{k-1}}{k!} + \ldots$
Multiply by z : $z(z+1)e^z = z + 2^2\frac{z^2}{2!} + \cdots + k^2\frac{z^k}{k!} + \ldots$
Therefore, $E[X^2] = e^{-\lambda}\lambda(\lambda+1)e^{\lambda} = \lambda^2 + \lambda$
So that $Var[X] = \lambda^2 + \lambda - \lambda^2 = \lambda$

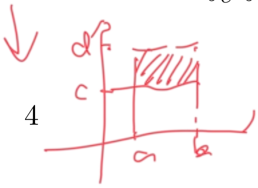
We have just shown that the mean of a Poisson r.v. with parameter λ is λ. Its variance is also λ.

6 Pairs of Random Variables

For a pair of random variables X, Y, we define the joint pdf of $X, Y = P_{X,Y}(x, y)$.

For discrete X, Y, $P_{X,Y}(x, y) = Prob(X = x$ *and* $Y = y)$.

For continuous X, Y, $\Pr(a \leq X \leq b$ and $c \leq Y \leq d) = \int_c^d \int_a^b P_{X,Y}(x, y)dxdy$ = volume under the curved surface.



边缘"分布"

Marginal pdf: $P_X(x) = \sum_y P_{X,Y}(x,y)$ or $\int_{-\infty}^{\infty} P_{X,Y}(x,y)dy$.

$P_Y(y) = \sum_x P_{X,Y}(x,y)$ or $\int_{-\infty}^{\infty} P_{X,Y}(x,y)dx$.

That is, the marginal pdf is obtained from the joint pdf by "integrating (summing) out the other variable".

Conditional pdf: $P_{X|Y}(x \mid y) \triangleq \frac{P_{X,Y}(x,y)}{P_Y(y)}$.

Expectation: $E[g(X,Y)] \triangleq \sum_x \sum_y g(x,y)P_{X,Y}(x,y)$. or $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)P_{X,Y}(x,y)dxdy$

Cross-correlation $R_{XY} \triangleq E[XY]$

Covariance $\sigma_{XY} \triangleq E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y$

Correlation coefficient $\rho_{X,Y} \triangleq \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Note: $-1 \leq \rho_{XY} \leq 1$ because $\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2$ (Cauchy - Schawarz Inequality)

7 Properties

Let X, Y be r.v., and α, β be real numbers.

1. $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$
2. $Var[\alpha X + \beta] = \alpha^2 Var[X]$
3. $Var[\alpha X + \beta Y] = \alpha^2 Var[X] + \beta^2 Var[Y] + 2\alpha\beta\sigma_{XY}$

8 Expectation (2 r.v.)

$$
\begin{aligned}
E[g(X,Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)P_{X,Y}(x,y)dxdy \\
&= \int_Y \int_X g(x,y)P_Y(y)P_{X|Y}(x|y)dxdy \\
&= \int_Y \left[\int_X g(x,y)P_{X|Y}(x,y)dx \right] P_Y(y)dy
\end{aligned}
$$

where $\int_X g(x,y)P_{X|Y}(x,y)dx = E_{X|Y}[g(X,Y)] \triangleq$ conditional expectation

$$\text{So } E[g(X,Y)] = E_Y[E_{X|Y}[g(X,Y)]]$$

Example: Flip a coin, $\Pr(Head) = p, \Pr(Tail) = 1-p$.

Let X = number of flips until the first Head appears.

Then $X \sim Geometric(p)$, i.e. $P_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$

What is $E[X]$? By definition, it is $= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{1}{p}$. But this sum is hard to compute.

Another way: let Y = outcome of current flip (trial). Then $E[X] = E_Y\left[E_{X|Y}[X]\right]$

$$E_{X|Y}[X] = \begin{cases} 1 & \text{, when } Y = Head \\ E[X]+1 & \text{, when } Y = Tail \end{cases}$$

so $E[X] = p.1 + (1-p)(E[X]+1) = \frac{1}{p}$

9 Some Definitions

X, Y are *statistically independent* if $P_{X,Y}(x,y) = P_X(x)P_Y(y)$. That is, the joint PDF can be factored into a product of marginal PDFs.

Note: If X, Y are statistically independent, then
$P_{X,Y}(x,y) = \frac{P_{X|Y}(x,y)}{P_Y(y)} = \frac{P_X(x)P_Y(y)}{P_Y(y)} = P_X(x)$
i.e. knowledge of Y provides no knowledge of X.

X, Y are *uncorrelated* means $\rho_{XY} = 0$ or $\sigma_{XY} = 0$ or $E[XY] = E[X]E[Y]$

X, Y are *orthogonal* means $E[XY] = 0$

Note:

1. If $E[X] = 0$ or $E[Y] = 0$, then uncorrelated $\Leftrightarrow$ orthogonal.
2. X, Y independent $\Rightarrow$ X, Y uncorrelated.
   Proof: $E[XY] = \int_y \int_x xyP_{X,Y}(x,y)dxdy$
   $= \int_x \int_y xyP_X(x)P_Y(y)dxdy$
   $= \left(\int_x yP_X(x)dx\right)\left(\int_y yP_Y(y)dy\right) = E[X]E[Y]$
3. However, uncorrelated $\not\Rightarrow$ independent, unless the random variables follow a Gaussian pdf. See Sec. 13.1.

10 Bayes' Rule

$P_{X,Y}(x,y) = P_{X|Y}(x|y)P_Y(y) = P_{Y|X}(y|x)P_X(x)$

$$\begin{aligned} \Rightarrow P_{X|Y}(x|y) &= \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)} \\ &= \frac{P_{Y|X}(y|x)P_X(x)}{\int_{-\infty}^{\infty} P_{Y|X}(y,x)P_X(x)dx} \\ \text{posterior pdf} &= \frac{\text{likelihood} \times \text{prior pdf}}{\text{evidence}} \end{aligned}$$

e.g. Let X denote "have cancer", and Y denote "blood test is positive"

$$\Pr(\text{have cancer} \mid \text{test positive}) = \frac{\Pr(\text{test positive} \mid \text{have cancer}) \Pr(\text{have cancer})}{P_Y(y)}$$

$$P_Y(y) = \Pr(\text{test positive} \mid \text{have cancer}) \Pr(\text{have cancer}) + \Pr(\text{test positive} \mid \text{no cancer}) \Pr(\text{no cancer})$$

So getting a positive blood test does not mean that one is definitely stricken with cancer. The probability depends on:

(a) How prevalent the cancer is in the general population (measured by the prior distribution $\Pr(\text{have cancer})$); and,

(b) How accurate is the blood test when there is cancer (measured by the likelihood $\Pr(\text{test positive} \mid \text{have cancer})$).

Most medical tests are never 100% accurate. They make two kinds of errors: they can be positive when there is no cancer (patient is healthy), and negative when there *is* cancer. In medical terminology, these are called the *sensitivity* and *specificity* of the test.

11 Central Limit Theorem

Let $X_1, X_2, \ldots, X_N$ be independent, identically distributed (iid) random variables with arbitrary pdf $P_X(x)$. Let the mean and variance be μ, σ^2 respectively. Then the r.v. $\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$ has pdf $Gauss(\mu, \frac{\sigma^2}{N})$ for large N.

Note:

1. $P_X(x)$ is arbitrary, not necessarily Gaussian.
2. $Gauss(\alpha, \beta^2)$ means Gaussian pdf, mean $=\alpha$, variance $=\beta^2$.

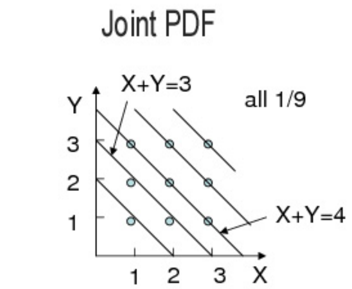
12 Transformations of r.v.

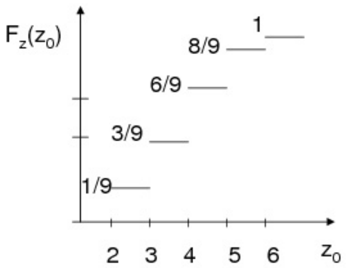
Let r.v. X, Y have joint PDF $P_{X,Y}(X, Y)$
Let $Z = g(X, Y)$ be some function. We know how to compute μ_Z and σ_Z^2 etc. But how to get $P_Z(z)$?

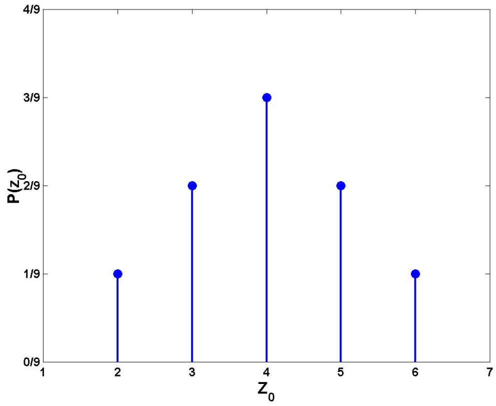
Most general technique is to find CDF first: $F_Z(z)$. Then differentiate to get $P_Z(z)$.

Examples

1. Two 3-sided dice (fair). Let X, Y denote score on a roll of die 1 and die 2 respectively. Let $Z = X + Y$. What is $P_Z(z)$?







$F_Z(z_0) = Prob(Z \leq z_0) = Prob(X + Y \leq z_0)$
$z_0 = 2, F_Z(2) = \frac{1}{9},$
$Z_0 = 3, F_Z(3) = \frac{3}{9},$
$F_Z(4) = \frac{6}{9},$
$F_Z(5) = \frac{8}{9},$
$F_Z(6) = \frac{9}{9} = 1$
The CDF and PDF of Z is are shown in the figure.

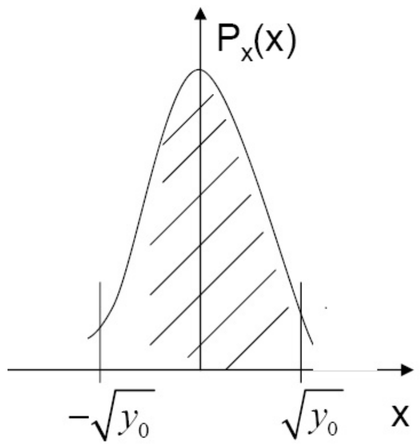
2. Let $X \sim Gauss(0, 1)$ and $Y = X^2$. What is $P_Y(y)$?

$$\begin{aligned}
F_Y(y_0) &= \Pr(Y \leq y_0) \\
&= \Pr(X^2 \leq y_0) \\
&= \Pr(-\sqrt{y_0} \leq X \leq \sqrt{y_0}) \\
&= 2\int_0^{\sqrt{y_0}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\
\frac{dF_Y}{dy_0} &= 2\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y_0})^2} \frac{1}{2\sqrt{y_0}}
\end{aligned}$$

where we have used the Fundamental Theorem of Calculus, and the Product Rule for Differentiation.

$$= \frac{1}{\sqrt{y_0}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_0}$$
$$P_Y(y) = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}}$$

This PDF is known as the χ^2 pdf of degree 1.



13 Random Vectors

When we have to deal with many random variables, it is easier to put them in a vector $\underline{X}$. Then pdf is $P_{\underline{X}}(\underline{x})$. $P_{\underline{X}}(\underline{x})$ means $P_{x_1,x_2,\ldots,x_n}(x_1, x_2, \ldots, x_n)$

$$\underline{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Joint pdf

Then $E[\underline{X}] = [E[X_1] \quad E[X_2] \ldots E[X_n]]^\top = \underline{\mu}_X$. That is, the expectation of the random vector is the vector of expectations of each random variable.

Let $\underline{g}(\underline{X}) = [g_1(\underline{X}) \quad g_2(\underline{X}) \ldots g_n(\underline{X})]^\top$

$Var[\underline{X}] \triangleq E[(\underline{X} - \underline{\mu}_X)(\underline{X} - \underline{\mu}_X)^\top]$

$= E[\underline{X}\,\underline{X}^\top] - \underline{\mu}_X \underline{\mu}_X^\top$ covariance matrix

Note: $Var[\underline{X}]$ is symmetric and positive semi-definite (p.s.d). The diagonal entries of the covariance matrix are σ_i^2, the variances of each random variable X_i, while the off-diagonal entries are σ_{ij} the covariance between random variables X_i and X_j.

13.1 Multivariate Gaussian

$\underline{X} = [X_1 \quad X_2 \quad \dots \quad X_d]^\top$
Let random vector $\underline{X} \sim Gauss(\underline{\mu}_X, \underline{C}_X)$, i.e. it follows a Gaussian pdf.

Formula: $P_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}(det(\underline{C}_X))^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{X}-\underline{\mu}_X)^\top \underline{C}_X^{-1}(\underline{X}-\underline{\mu}_X)}$
Note:

1. $P_{\underline{X}}(\underline{x})$ is a scalar
2. $\underline{\mu}_X$ mean vector
3. $\underline{C}_X$ covariance matrix
4. d = dimension of $\underline{X}$
5. Let $D^2 = (\underline{X} - \underline{\mu}_X)^\top \underline{C}_X^{-1}(\underline{X} - \underline{\mu}_X)$
   Then D is called the *Mahalanobis distance* of $\underline{X}$ to $\underline{\mu}_X$.
6. Let $\underline{Y} = \underline{A}\ \underline{X} + \underline{b}$, where $\underline{A}$ matrix, $\underline{b}$ vector. Then $\underline{Y} \sim Gauss(\underline{A\mu}_X + b, \underline{AC}_X\underline{A}^\top)$,
   In other words, a linear transformation of a Gaussian random variable produces another Gaussian random variable.
7. $\underline{Z} = \begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix}$, $\underline{X} \in \mathbb{R}^n$, $\underline{Y} \in \mathbb{R}^m$, $\underline{Z} \in \mathbb{R}^{n+m}$ and $\underline{Z} \sim Gauss(\underline{\mu}_z, \underline{C}_z)$ where $\underline{\mu}_z = \begin{bmatrix} \underline{\mu}_x \\ \underline{\mu}_y \end{bmatrix}$, $C_z = \begin{bmatrix} \underline{C}_{xx} & \underline{C}_{xy} \\ \underline{C}_{xy}^\top & \underline{C}_{yy} \end{bmatrix}$.
   Then conditional r.v. $X|Y \sim Gauss(\underline{m}, \underline{S})$, where
   $$\underline{m} = \underline{\mu}_x + \underline{C}_{xy}\underline{C}_{yy}^{-1}(\underline{y} - \underline{\mu}_y)$$
   $$\underline{S} = \underline{C}_{xx} - \underline{C}_{xy}\underline{C}_{yy}^{-1}\underline{C}_{xy}^\top$$
8. $\Phi_x(\underline{\omega}) = e^{-\frac{1}{2}\underline{\omega}^\top \underline{C}_x \underline{\omega} + j\underline{\omega}^\top \underline{\mu}_x}$,
   where $\underline{\omega} = [\omega_1, \omega_2, \cdots \omega_d]^\top$ is the vector of frequencies.
9. If x_i's are uncorrelated, then they are independent.
   <u>Proof:</u>
   $$x_i\text{'s are uncorrelated means } \underline{C}_x = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_d^2 \end{bmatrix}$$
   $$\underline{C}_x^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_d^2} \end{bmatrix}$$

So $(\underline{x} - \underline{\mu}_x)^\top \underline{C}_x^{-1}(\underline{x} - \underline{\mu}_x) = \sum_{k=1}^{d} \left(\frac{x_k - \mu_k}{\sigma_k}\right)^2$
Then

$$\begin{aligned}
\underline{P}_x(\underline{x}) &= \frac{1}{(2\pi)^{d/2}(\det \underline{C}_x)^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_x)^\top \underline{C}_x^{-1}(\underline{x}-\underline{\mu}_x)} \\
&= \frac{1}{\prod_{k=1}^{d}(2\pi)^{1/2}(\sigma_k^2)^{1/2}} e^{-\frac{1}{2}\sum_{k=1}^{d}(\frac{x_k-\mu_k}{\sigma_k})^2} \\
&= \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{1}{2}(\frac{x_k-\mu_k}{\sigma_k})^2} \\
&= P_{x_1} \cdot P_{x_2} \cdot P_{x_3} \cdots P_{x_d}
\end{aligned}$$

In other words, the joint pdf is decomposed into a product of marginal pdfs. This proves that the random variables are statistically independent.

14 Characteristic Function 特征函数

We can view $P_X(x)$ as a signal. Clearly, $\sum_x P_X(x) = 1$ (for discrete X), or $\int_{-\infty}^{\infty} P_X(x)dx = 1$ (for continuous X), so $P_X(x)$ is ablsolutely summable (integrable). Thus we can take the Fourier Transform!

Definition: Characteristic Function: $\phi_X(\omega) \triangleq E[e^{j\omega X}]$.

So $\phi_X(\omega) = \sum_x e^{j\omega x} P_X(x)$; for discrete X.

and $\phi_X(\omega) = \int_{-\infty}^{\infty} e^{j\omega x} P_X(x)dx$; for continuous X.

Notes:

1. Not quite DTFT: $e^{j\omega X}$ instead of $e^{-j\omega X}$
2. But can still be considered Fourier Transform.

Inverse transform:
$P_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(\omega) e^{-j\omega X} d\omega$, for continuous X.
or
$P_X(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_X(\omega) e^{-j\omega X} d\omega$, for discrete X.

Take derivatives: $\frac{d\phi_X}{d\omega} = \frac{d}{d\omega}\left[\int_{-\infty}^{\infty} e^{j\omega x} P_X(x)dx\right]$
$\phi'_X(\omega) = \int_{-\infty}^{\infty} jx e^{j\omega x} P_X(x)dx$

Clearly, $\phi_X^{(n)}(\omega) = \int_{-\infty}^{\infty} (jx)^n e^{j\omega x} P_X(x)dx$

$\Rightarrow \phi_X^{(n)}(0) = j^n E[X^n]$

This gives us a way to compute moments.

Example: Consider two independent r.v. X, Y. Let $Z = X + Y$. What is $P_Z(z)$?

Well, $\phi_Z(\omega) = E[e^{j\omega z}] = E[e^{j\omega(X+Y)}] = E[e^{j\omega X} e^{j\omega Y}]$
So $\phi_Z(\omega) = E[e^{j\omega X}]E[e^{j\omega Y}]$
$\phi_Z(\omega) = \phi_X(\omega)\phi_Y(\omega)$

$\Rightarrow P_Z(z) = P_X(x) * P_Y(y)$; This is convolution!

For continuous Z, $P_Z(z) = \int_{-\infty}^{\infty} P_X(x) P_Y(z-x) dx$

This explains the earlier phenomenon : any pulse-like signal, when convolved with itself many times, produces the *Gaussian*-shape signal.

The Central Limit Theorem says $Y = X_1 + X_2 + \cdots + X_N \sim$ Gaussian for large N.

That is, $\phi_Y(\omega) = (\phi_X(\omega))^N \Rightarrow$ Gaussian

Note : If $X \sim$ Gauss(μ, σ^2), then its Characteristic Function $\phi_X(\omega) = e^{-\frac{1}{2}\omega^2\sigma^2 + j\mu\omega}$