Multiple Dimension Random Variables

Module 5

GPDS
Grupo de Processamento Digital de Sinais
Prof. F. Assis

Continuous Multiple Dimension Joint Distributions - PDF

- The ***d**-Dimension* jointly continuous random variables $X_1, X_2, \cdots, X_d$, can be written as

$$P\Big((X_1, X_2, \cdots, X_d) \in A\Big) = \int \cdots \int_A \cdots \int f_{X_1 X_2 \cdots X_d}(x_1, x_2, \cdots, x_d)\, dx_1\, dx_2 \cdots dx_d$$

The Marginal **PDF**

- The marginal **PDF** of X_i can be obtained by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2 \ldots X_d}(x_1, x_2, \ldots, x_d) dx_2 \cdots dx_d$$

Continuous Multiple Dimension **CDF**

- The joint **CDF** of ***d**-dimension* random variables X_1, X_2,..., X_d is defined as

$$F_{X_1,X_2,\ldots,X_d}(x_1,x_2,\ldots,x_d) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d)$$

Mean and Variance from Multiple Joint Distribution

For continuous random variables, a **joint probability density function** $f_{X_1X_2\ldots X_d}(x_1,x_2,\ldots,x_d)$ is used to determine the probability that:

$(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3,\ldots,\boldsymbol{X}_d) \in R_{\boldsymbol{X}}$ by the multiple integral of $f_{X_1X_2\ldots X_d}(x_1,x_2,\ldots,x_d)$ over the region $R_{\boldsymbol{X}}$.

$$E(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_i\, f_{X_1X_2\ldots X_d}(x_1, x_2, \ldots, x_d)\, dx_1\, dx_2 \ldots dx_d = \int_{-\infty}^{\infty} x_i\, f_{X_i}(x_i)\, dx_i$$

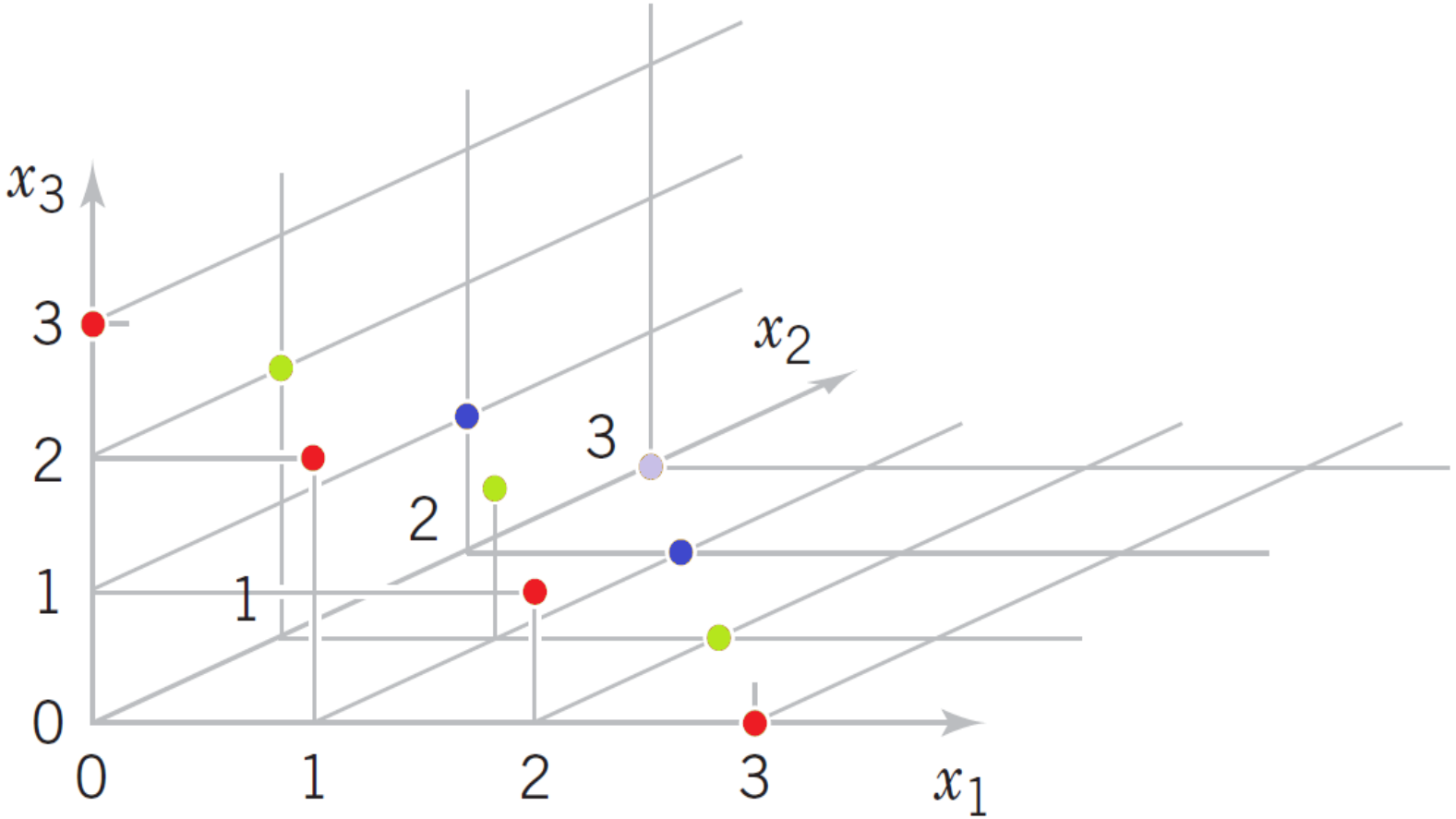
and

$$Var(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} (x_i - \mu_{X_i})^2 f_{X_1X_2\ldots X_d}(x_1, x_2, \ldots, x_d)\, dx_1\, dx_2 \ldots dx_d = \int_{-\infty}^{\infty} (x_i - \mu_{X_i})^2 f_{X_i}(x_i) dx_i$$

Exercise 5.1

- Points that have positive probability in the joint probability distribution of three random variables X_1, X_2, X_3 are shown in figure on the right.

- Suppose the 10 points are equally likely with probability 0.1 each.
- The range is the non-negative integers with $x_1 + x_2 + x_3 = 3$.
- Find marginal probability distribution of $\boldsymbol{X}_2$ and $E[\boldsymbol{X}_2]$.



Exercise 5.1: Solution

- The marginal probability distribution (PMF marginal) of X_2 is found as follows: $x_1 + x_2 + x_3 = 3$

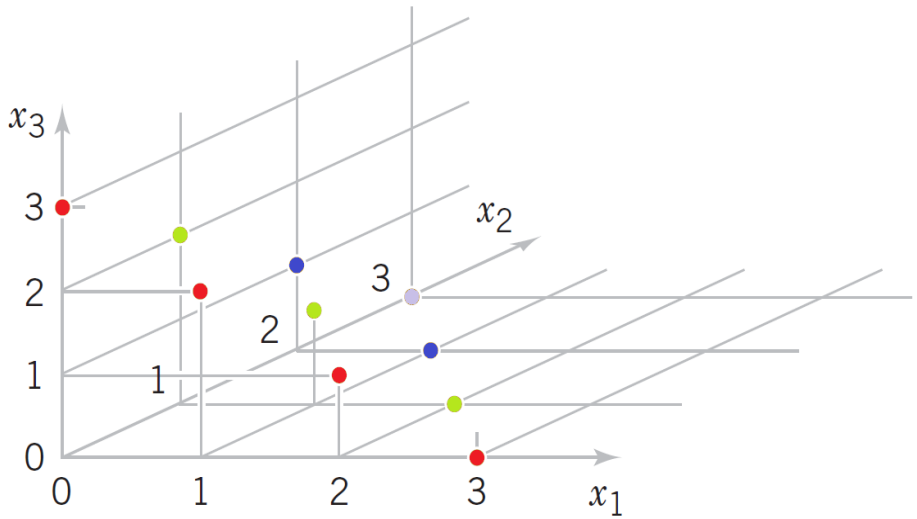
$$P(X_2 = 0) = P_{X_1X_2X_3}(3,0,0) + P_{X_1X_2X_3}(0,0,3) + P_{X_1X_2X_3}(1,0,2) + P_{X_1X_2X_3}(2,0,1) = 0.4$$

$$P(X_2 = 1) = P_{X_1X_2X_3}(2,1,0) + P_{X_1X_2X_3}(0,1,2) + P_{X_1X_2X_3}(1,1,1) = 0.3$$

$$P(X_2 = 2) = P_{X_1X_2X_3}(1,2,0) + P_{X_1X_2X_3}(0,2,1) = 0.2$$

$$P(X_2 = 3) = P_{X_1X_2X_3}(0,3,0) = 0.1$$

Also, $E(X_2) = 0(0.4) + 1(0.3) + 2(0.2) + 3(0.1) = 1$



Sums of Random Variables

$$Y = X_1 + X_2 + \cdots + X_n$$

$$EY = EX_1 + EX_2 + \cdots + EX_n$$

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$$

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

Sums of Independent Random Variables

- If the $\boldsymbol{X}_i$'s are independent, then $Cov(\boldsymbol{X}_i, \boldsymbol{X}_j) = 0$ for $i \neq j$.

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Linear Functions of Random Variables

- Given random variables $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ and constants $c_1, c_2, \ldots c_p$,

$$\boldsymbol{Y} = c_1 \boldsymbol{X}_1 + c_2 \boldsymbol{X}_2 + \ldots + \boldsymbol{c}_p \boldsymbol{X}_p$$

- is a **linear combination** of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$.

Linear Functions of Random Variables

- $E(\boldsymbol{Y})$ can be found from the joint probability distribution of $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ as follows

$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\left(c_1x_1 + c_2x_2 + \ldots + c_p\mathrm{x}_p\right) f_{X_1X_2\ldots X_p}\left(x_1, x_2, \ldots, x_p\right) dx_1\, dx_2 \ldots dx_p$$

$$= c_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_1\, f_{X_1X_2\ldots X_p}\left(x_1, x_2, \ldots, x_p\right) dx_1\, dx_2 \ldots dx_p$$

$$+ c_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_2\, f_{X_1X_2\ldots X_p}\left(x_1, x_2, \ldots, x_p\right) dx_1\, dx_2 \ldots dx_p +, \ldots,$$

$$+ c_p \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_p\, f_{X_1X_2\ldots X_p}\left(x_1, x_2, \ldots, x_p\right) dx_1\, dx_2 \ldots dx_p$$

Linear Functions of Random Variables

- If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ are random variables, and

$$\boldsymbol{Y} = c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2 + \ldots + c_p\boldsymbol{X}_p$$

- Then in general, the variance of linear function of random variables can be computed in the form

$$Var(Y) = c_1^2 Var(X_1) + c_2^2 Var(X_2) + \cdots + c_p^2 Var(X_p) + \sum_{\substack{i \\ i<j}} \sum_{j} c_i\, c_j Cov(X_i, X_j)$$

- If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ are **independent**

$$Var(\boldsymbol{Y}) = c_1^2 Var(\boldsymbol{X}_1) + c_2^2 Var(\boldsymbol{X}_2) + \cdots + c_p^2 Var(\boldsymbol{X}_p)$$

Exercise 5.2

- A semiconductor product consists of three layers. Supposing that the variances in thickness of the first, second, and third layers are 25, 40, and 30 square nanometers, respectively, and the layer thicknesses are independent.
- What is the variance of the thickness of the final product?
- Consider $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}$ be random variables that denote the thicknesses of the respective layers, and the final product. Then, we can make an approaching

$$\boldsymbol{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3$$

Exercise 5.2:Solution

- The variance of **X** is obtained from the linear combination of random variables

$$Var(X) = \sigma_X^2 = Var(X_1) + Var(X_2) + Var(X_3)$$
$$= 25 + 40 + 30 = 95nm^2$$

- Consequently, the standard deviation of thickness of the final product is

$$\sigma_X = \sqrt{95nm^2} = 9.75nm$$

- Physical interpretation: This shows how the variation in each layer is propagated to the final product.

Reproductive Property of the Gaussian (Normal) Distribution

- If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ are independent, Gaussian random variables with $E[X_i] = \mu_i$ and $\mathrm{Var}[X_i] = \sigma_i^2$, for $i = 1, 2, \ldots, p$,

$$\boldsymbol{Y} = c_1 \boldsymbol{X}_1 + c_2 \boldsymbol{X}_2 + \ldots + c_p \boldsymbol{X}_p$$

is a Gaussian distrubution random variable with

$$E[\boldsymbol{Y}] = c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p$$

and

$$Var[\boldsymbol{Y}] = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_p^2\sigma_p^2$$

Exercise 5.3

- Let the random variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ denote the length and width, respectively, of a manufactured part.
- Assume that $\boldsymbol{X}_1$ is normal with $E[\boldsymbol{X}_1] = 2$ cm and standard deviation 0.1 cm and that $\boldsymbol{X}_2$ is normal with $E[\boldsymbol{X}_2] = 5$ cm and standard deviation 0.2 cm.
- Also assume that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent.
- Determine the probability that the perimeter exceeds 14.56 cm.

Exercise 5.3:Solution

- Then the random variable that describes the perimeter can be express as $\boldsymbol{Y} = 2\boldsymbol{X}_1 + 2\boldsymbol{X}_2$ is a Gaussian (normal) random variable that represents the perimeter of the part.
- We obtain $E[Y]$ and variance of $\boldsymbol{Y}$ as

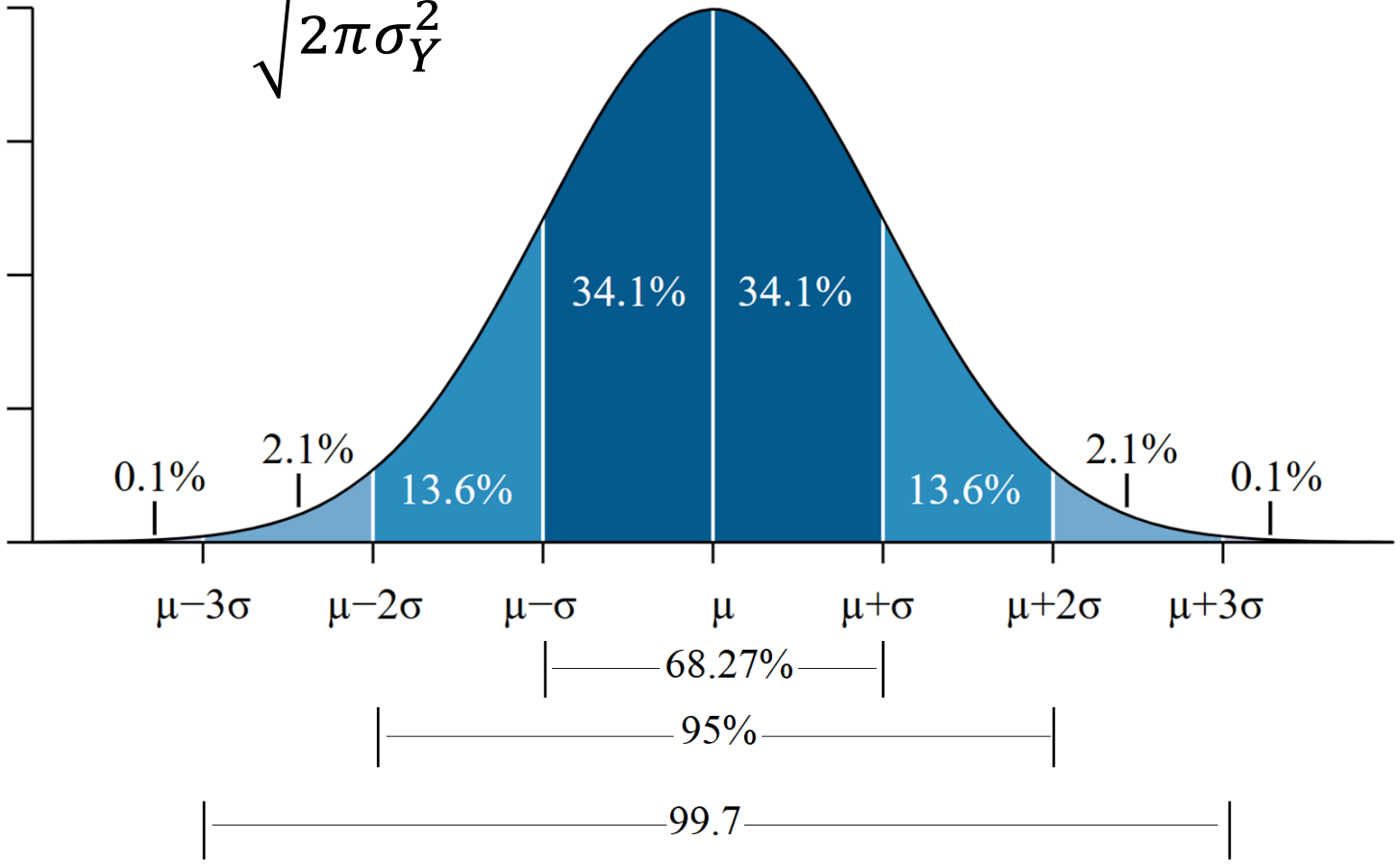
$$\begin{aligned} E[\boldsymbol{Y}] &= 2 \times E[\boldsymbol{X}_1] + 2 \times E[\boldsymbol{X}_2] \\ &= 2 \times 2 + 2 \times 5 = 14 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}[\boldsymbol{Y}] &= 2^2\,\mathrm{Var}[\boldsymbol{X}_1] + 2^2\,\mathrm{Var}[\boldsymbol{X}_2] \\ &= 4 \times 0.1^2 + 4 \times 0.2^2 = 0.2 \end{aligned}$$

Exercise 5.3:Solution

- Remembering the Gaussian distribution

$$f_Y(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{Y} \mid \mu_Y, \sigma_Y^2) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}, -\infty < y < +\infty$$



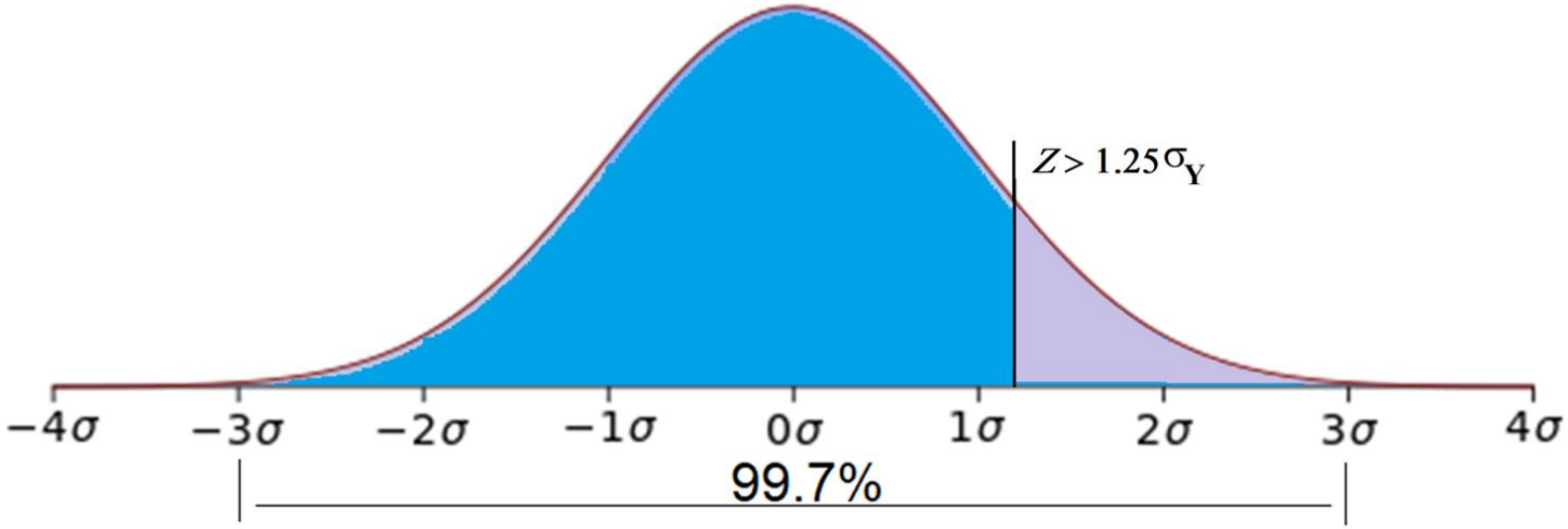
Exercise 5.3:Solution

- $P(Y > 14.56) = \int_{14.5}^{\infty} f_Y(x)\, dx$

$$f_Y(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{Y} \mid \mu_Y, \sigma_Y^2) = \frac{\mathbf{1}}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{\mathbf{1}}{\mathbf{2}}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^{\mathbf{2}}}, -\infty < y < +\infty$$

$$P(Y > 14.56) = P\left(\frac{Y - \mu_Y}{\sigma_Y} > \frac{14.56 - 14}{\sqrt{0.2}}\right)$$

$$P(Z > 1.25) = 0.13$$



General Function of a Discrete Random Variable

- Suppose that $\boldsymbol{X}$ is a **discrete** random variable with probability distribution $P_{\boldsymbol{X}}(x)$. Let $Y = h(X)$ define a one-to-one transformation between the values of $\boldsymbol{X}$ and $\boldsymbol{Y}$ so that the equation $y = h(x)$ can be solved uniquely for x in terms of y.
- Let this solution be $x = u(y)$. Then the probability mass function of the random variable $\boldsymbol{Y}$ is

$$P_Y(y) = P_X[u(y)]$$

Exercise 5.4

- Let $\boldsymbol{X}$ be a geometric random variable with probability distribution

$$\mathbf{P}_X(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots$$

Find the probability distribution of $\boldsymbol{Y} = \boldsymbol{X}^2$.

Exercise 5.4:Solution

- Because $\boldsymbol{X} \geq 0$, the transformation is one to one; that is, $y = x^2$ then $x = \sqrt{y}$.

- Therefore, the distribution of the random variable $\boldsymbol{Y}$ is

$$\mathbf{P}_Y(y) = \mathbf{P}\left(\sqrt{y}\right) = p(1-p)^{\sqrt{y}-1}, \qquad y = 1, 4, 9, 16, \ldots$$

General Transformation Function of a Continuous Random Variable

- Suppose that $\boldsymbol{X}$ is a **continuous** random variable with probability distribution $f_X(x)$.
- The function $Y = h(\boldsymbol{X})$ is a one-to-one transformation between the values of $\boldsymbol{Y}$ and $\boldsymbol{X}$, so that the equation $y = h(x)$ can be uniquely solved for x in terms of y.
- Let this solution be $x = u(y)$. The probability distribution of Y is

$$f_Y(y) = f_X[u(y)]|J|$$

- where $\boldsymbol{J} = u'(y)$ is called the **Jacobian** of the transformation and the absolute value of $\boldsymbol{J}$ is used.

General Transformation Function of a Continuous Random Variable

- The equation $f_Y(y) = f_X[u(y)]|J|$ is shown as follows. Let the function $y = h(x)$ be an increasing function of x. Now

$$P(Y \le a) = P[X \le u(a)] = \int_{-\infty}^{u(a)} f_X(x)\,dx$$

- If we change the variable of integration from x to y by using $x = u(y)$, we obtain $dx = u'(y)dy$ and then

$$P(Y \le a) = \int_{-\infty}^{a} f_X[u(y)]u'(y)\,dy$$

General Transformation Function of a Continuous Random Variable

- Because the integral gives the probability that $\boldsymbol{Y} \leq a$, for all values of a contained in the feasible set of values for y, $f_X[u(y)]u'(y)$ must be the probability density of $\boldsymbol{Y}$.
- Therefore, the probability distribution of $\boldsymbol{Y}$ is

$$f_Y(y) = f_X[u(y)]|\mathrm{J}|$$

Exercise 5.5

- Let $\boldsymbol{X}$ be a continuous random variable with probability distribution

$$f_X(x) = \frac{x}{8}, \quad 0 \le x < 4$$

- Find the probability distribution of $\boldsymbol{Y} = h(\boldsymbol{X}) = 2\boldsymbol{X} + 4$.

Exercise 5.5:Solution

- Note that $y = h(x) = 2x + 4$ is an increasing function of x.
- The inverse solution is

$$x = u(y) = (y - 4) / 2,$$

- and from this, we find the Jacobian to be

$$J = u'(y) = dx/dy = 1/2.$$

- Therefore, the probability distribution of $\boldsymbol{Y}$ is

$$f_Y(y) = \frac{(y-4)/2}{8}\left(\frac{1}{2}\right) = \frac{y-4}{32}, \qquad 4 \le y \le 12$$

The Method of Transformations

Theorem

Let X and Y be two jointly continuous random variables. Let $(Z,W) = g(X,Y) = (g_1(X,Y), g_2(X,Y))$, where $g : \mathbb{R}^2 \mapsto \mathbb{R}^2$ is a continuous one-to-one (invertible) function with continuous partial derivatives. Let $h = g^{-1}$, i.e., $(X,Y) = h(Z,W) = (h_1(Z,W), h_2(Z,W))$. Then Z and W are jointly continuous and their joint PDF, $f_{ZW}(z,w)$, for $(z,w) \in R_{ZW}$ is given by

$$f_{ZW}(z,w) = f_{XY}(h_1(z,w), h_2(z,w))|J|,$$

where J is the Jacobian of h defined by

$$J = \det \begin{bmatrix} \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \\ \frac{\partial h_2}{\partial z} & \frac{\partial h_2}{\partial w} \end{bmatrix} = \frac{\partial h_1}{\partial z} \cdot \frac{\partial h_2}{\partial w} - \frac{\partial h_2}{\partial z} \frac{\partial h_1}{\partial w}.$$

The *M*-th Moment of a Random Variable

- The *M*-th **moment random variable** is defined as

$$E[\boldsymbol{X^m}] = \begin{cases} \displaystyle\sum_{\boldsymbol{x} \in \boldsymbol{R_X}} \boldsymbol{x^m p_X(x)}\,, & \boldsymbol{x}\ \textbf{discrete} \\ \displaystyle\int_{-\infty}^{\infty} \boldsymbol{x^m f_X(x) dx}\,, & \boldsymbol{x}\ \textbf{continuous} \end{cases}$$

The *M*-th Central Moment of a Random Variable

- The *M*-th **central moment random variable** is defined as

$$E[(\boldsymbol{X}-\mu_X)^m] = \begin{cases} \displaystyle\sum_{\boldsymbol{x}\in \boldsymbol{R}_X} (\boldsymbol{x}-\boldsymbol{\mu}_X)^{\boldsymbol{m}} \boldsymbol{p}_X(\boldsymbol{x}), & \boldsymbol{x}\ \textbf{discrete} \\ \displaystyle\int_{-\infty}^{\infty} (\boldsymbol{x}-\boldsymbol{\mu}_X)^{\boldsymbol{m}} \boldsymbol{f}_X(\boldsymbol{x})\boldsymbol{dx}, & \boldsymbol{x}\ \textbf{continuous} \end{cases}$$

Definition of a Moment-Generating Function

- The **moment-generating function** of the random variable $\boldsymbol{X}$ is the expected value of $\boldsymbol{e^{sX}}$ and is denoted by $M_X(s)$. That is,

$$M_X(s) = E[\boldsymbol{e^{sX}}] = \begin{cases} \sum_{\boldsymbol{x} \in \boldsymbol{R_X}} \boldsymbol{e^{sx} p_X(x)}, & \boldsymbol{x} \text{ discrete} \\ \int_{-\infty}^{\infty} \boldsymbol{e^{sx} f_X(x) dx}, & \boldsymbol{x} \text{ continuous} \end{cases}$$

Moment-Generating Function

- Assuming that we can differentiate inside the summation and integral signs,

$$\frac{d^m M_X(s)}{ds^m} = \begin{cases} \sum_{\boldsymbol{x} \in \boldsymbol{R_X}} x^m \boldsymbol{e^{sx} p_X(x)}, & \boldsymbol{x} \text{ discrete} \\ \int_{-\infty}^{\infty} x^m \boldsymbol{e^{sx} f_X(x) dx}, & \boldsymbol{x} \text{ continuous} \end{cases}$$

- Now if we set $s = 0$ in this expression, we find that,

$$\left.\frac{d^m M_X(s)}{ds^m}\right|_{s=0} = E[\boldsymbol{X^m}]$$

Moment Generating Functions (MGF)

Remember the Taylor series for e^x: for all $x \in \mathbb{R}$, we have

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Now, we can write

$$e^{sX} = \sum_{k=0}^{\infty} \frac{(sX)^k}{k!} = \sum_{k=0}^{\infty} \frac{X^k s^k}{k!}.$$

Thus, we have

$$M_X(s) = E[e^{sX}] = \sum_{k=0}^{\infty} E[X^k] \frac{s^k}{k!}.$$

Moment Generating Functions (MGF)

- All moments of $\boldsymbol{X}^k$ from its MGF

$$M_X(s) = \sum_{k=0}^{\infty} E[X^k] \frac{s^k}{k!},$$

$$E[X^k] = \frac{d^k}{ds^k} M_X(s)\big|_{s=0}.$$

Exercise 5.6

- Suppose that a random variable $\boldsymbol{X}$ has a binomial distribution, that is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n \qquad \binom{n}{x} \triangleq \frac{n!}{x!(n-x)!}$$

- Determine the moment-generating function and use it to verify that the mean and variance of the binomial random variable are $\mu = np$ and $\sigma^2 = np(1-p)$.

Exercise 5.6:Solution

- From the definition of a moment-generating function, we

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

- This last summation is the binomial expansion of $(x + y)^n = [\, pe^t + (1 - p)]^n$, so

$$M_X(t) = [pe^t + (1-p)]^n$$

Exercise 5.6:Solution

- Taking the first and second derivatives, we obtain

$$M_X(t) = [pe^t + (1-p)]^n$$

$$M_X'(t) = \frac{dM_X(t)}{dt} = npe^t[1 + p(e^t - 1)]^{n-1}$$

- and

$$M_X''(t) = \frac{d^2M_X(t)}{dt^2} = npe^t(1 - p + npe^t)[1 + p(e^t - 1)]^{n-2}$$

Exercise 5.6:Solution

- If we set $t = 0$ in $\frac{dM_X(t)}{dt}$ we obtain

$$\left.\frac{dM_X(t)}{dt}\right|_{t=0} = \mu = np$$

- which is the mean of the binomial random variable $\boldsymbol{X}$.

Exercise 5.6:Solution

- Now if we set $t = 0$ in $\frac{d^2 M_X(t)}{ds^2}$, result:

$$\left.\frac{d^2 M_X(t)}{ds^2}\right|_{t=0} = np(1-p+np) = \mathrm{E}[X^2]$$

$$\sigma^2 = \mathrm{E}[X^2] - (E[X])^2 = np(1-p+np) - (np)^2 = np - np2 = np(1-p)$$

Exercise 5.7

- Find the moment-generating function of the normal random variable and use it to show that the mean and variance of this random variable are μ and σ^2, respectively.
- The moment-generating function is

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-[x^2-2(\mu+t\sigma^2)x+\mu^2]/(2\sigma^2)} dx$$

Exercise 5.7:Solution

- If we complete the square in the exponent, we have

$$x^2 - 2(\mu + t\sigma^2)x + \mu^2 = [x - (\mu + t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4$$

and then

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\{[x-(\mu+t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4\}/(2\sigma^2)} dx$$

$$= e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[x-(\mu+t\sigma^2)]^2/\sigma^2} dx$$

Exercise 5.7:Solution

- Let $u = [x - (\mu + t\sigma^2)]/\sigma$. Then $dx = \sigma du$, and the last expression above becomes

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

- Now the integral is just the total area under a standard normal density, which is 1, so the moment-generating function of a normal random variable is

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

Exercise 5.7:Solution

- Differentiating this function twice with respect to t and setting $t = 0$ in the result, yields

$$\left.\frac{dM_X(t)}{dt}\right|_{t=0} = \mu \quad \text{and} \quad \left.\frac{d^2M_X(t)}{dt^2}\right|_{t=0} = \sigma^2 + \mu^2$$

- To obtain the central moment, it is necessary to subtract the obtained mean

$$\sigma^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Properties of Moment Generating Functions

- If $\boldsymbol{X}$ is a random variable and a is a constant, then

$$M_{X+a}(s) = e^{as} M_X(s)$$

$$M_{aX}(s) = M_X(as)$$

Properties of Moment Generating Functions

- If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are **independent random variables** with moment generating functions $M_{X_1}(s), M_{X_2}(s), \ldots, M_{X_n}(s)$ respectively, and if $\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \ldots + \boldsymbol{X}_n$,then the moment generating function of $\boldsymbol{Y}$ is

$$M_Y(s) = M_{X_1}(s)\, M_{X_2}(s) \ldots M_{X_n}(s)$$

Random Vectors

GPDS
Grupo de Processamento Digital de Sinais

Prof. F. Assis

VECTOR RANDOM VARIABLES

- A **vector random variable X** is a function that assigns a vector of real numbers to each outcome in ***S***, the sample space of the random experiment.
- We use uppercase boldface notation for vector random variables.
- By convention **X** is a column vector (*n* rows by 1 column), so the vector random variable with components corresponds to

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1, X_2, \ldots, X_n]^{\mathrm{T}},$$

Example 5.1 Arrivals at a Packet Switch

- Packets arrive at each of three input ports of a packet switch according to independent Bernoulli trials with p = 1/2.
- Each arriving packet is equally likely to be destined to any of three output ports.
- Let $\mathbf{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ where is the total number of packets arriving for output port i.
- $\mathbf{X}$ is a vector random variable whose values are determined by the pattern of arrivals at the input ports.

Example 5.2 Samples of an Audio Signal

- Let the outcome of a random experiment be an audio signal $X(t)$.

- Let the random variable $X_k = X(kT)$ be the sample of the signal taken at time kT. Where T is the sampling rate.

- An MP3 codec processes the audio in blocks of n samples

$\mathbf{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ is a vector random variable.

Example 5.3 Joint Poisson Counts

- A random experiment consists of finding the number of defects in a semiconductor chip and identifying their locations.
- The outcome of this experiment consists of the vector

$\xi = (n, \boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n)$, where the first component specifies the total number of defects, and the remaining components specify the coordinates of their location.

- Suppose that the chip consists of M regions.
- Let $N_1(\xi), N_2(\xi), \dots, N_M(\xi)$ be the number of defects in each of these regions, that is, $N_i(\xi)$ is the number of $\boldsymbol{y}$'s that fall in region k.
- The vector is $\mathrm{N}(\xi) = (N_1, N_2, \dots, N_M)$ then a vector random variable.

Statistical dependence/independence

- Reviewing exercise 4.3 and exercise 4.6 of module 4. The statements read as follows:
- Exercise 4.3: Consider an experiment carried out in a laboratory, where you have two random variables drawn together. The variable $\boldsymbol{X}$ presents the dynamic range $\boldsymbol{X} = \{0,1\}$ and the variable $\boldsymbol{Y}$ presents the respective dynamic range $\boldsymbol{Y} = \{1,2,3,4,5,6\}$ (results of the measurements are in the table on the right).
- Exercise 4.6: Now consider the random experiment in which you flip a coin and then roll a die. Both the coin and the dice have equiprobable probabilities: $\boldsymbol{X} = \{0,1\}$; $\mathbf{Y} = \{1,2,3,4,5,6\}$.

índice	X	Y
1	1	1
2	1	4
3	0	1
4	1	5
5	0	3
6	1	1
7	0	6
8	0	2
9	1	2
10	0	5
11	0	2
12	1	5
13	1	3
14	0	5
15	1	5
16	1	5
17	0	6
18	0	3
19	0	4
20	1	5
21	1	5
22	0	5
23	1	4
24	1	6

Statistical dependence/independence

- Compare the results of exercises 4.3 and 4.6: for statistical independence we know:

$$P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{P(Y)P(X)}{P(X)} = P(Y)$$

- EX. 4.3
  - PMF Conjunta $P_{XY}(X, Y)$

X/Y	1	2	3	4	5	6
0	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{2}{24}$
1	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{6}{24}$	$\frac{1}{24}$

  - CDF marginal $P_X(x)$

X	$P_X(x)$
0	$\frac{11}{24}$
1	$\frac{13}{24}$

$$P(Y=1|X=0) = \frac{P(X=0,Y=1)}{P(X=0)} = \frac{\frac{1}{24}}{\frac{11}{24}} = \frac{1}{11} \neq$$

$$\neq P(Y=1) = \frac{3}{24} = \frac{1}{8} \text{ (statistically dependent)}$$

- EX. 4.6
  - PMF Conjunta $P_{XY}(X, Y)$

X/Y	1	2	3	4	5	6
0	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$
1	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$

  - CDF marginal $P_X(x)$

X	$P_X(x)$
0	$\frac{12}{24}$
1	$\frac{12}{24}$

$$P(Y=1|X=0) = \frac{P(X=0,Y=1)}{P(X=0)} = \frac{\frac{2}{24}}{\frac{12}{24}} = \frac{2}{12} = \frac{1}{6} =$$

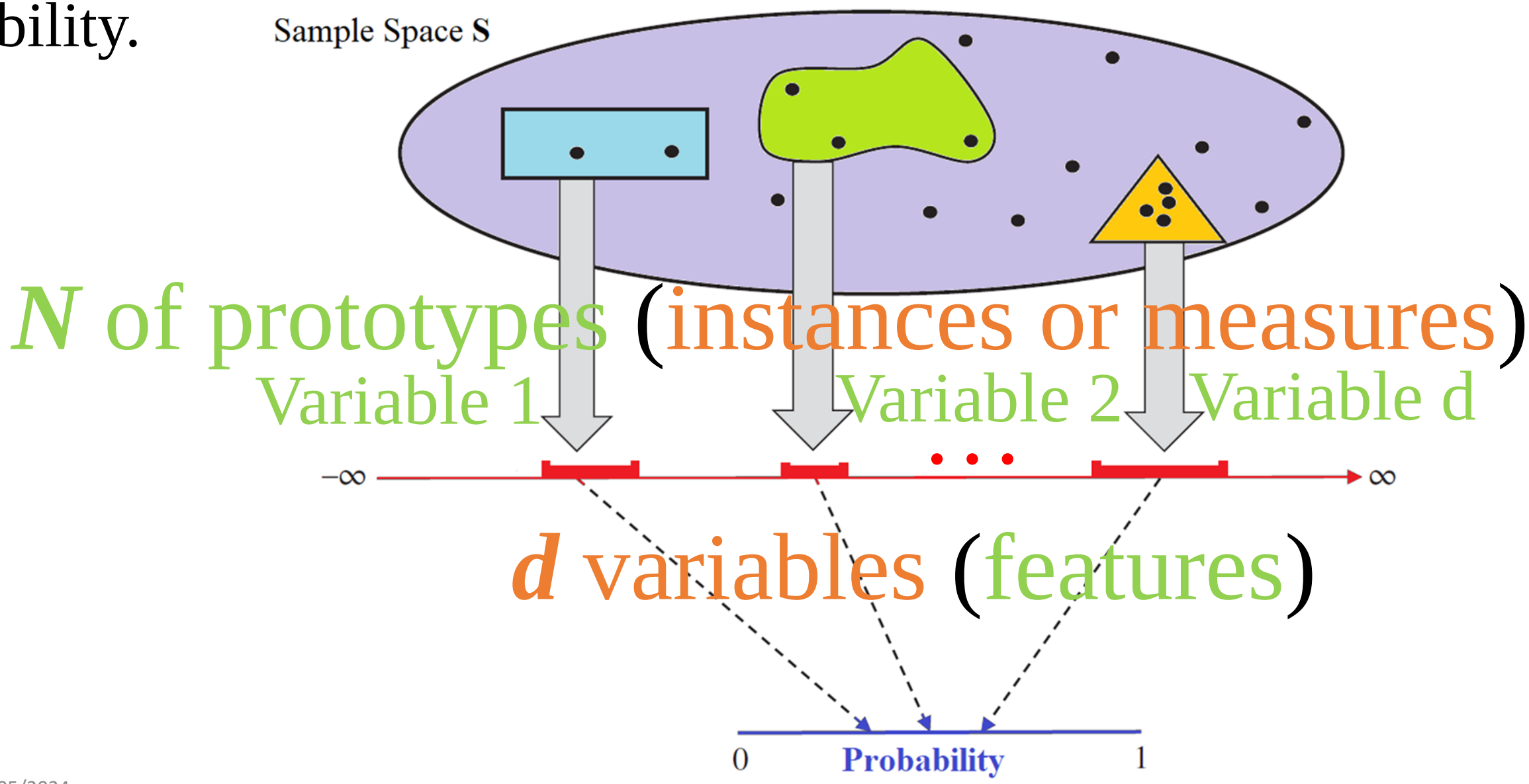
$$= P(Y=1) = \frac{4}{24} = \frac{1}{6} \text{ (statistically independent)}$$

Discrete **d**-dimension data representation

- If a number N of prototypes (instances or measures) are large enough to represent stochastic process **X** in terms of d variables (features).

- Giving the input data vectors X_d and consider a real d-dimensional Hilbert space and a set of **N** points measured in this respective space.

The relationship among sample space, random variable, and probability.



Discrete d-dimension data representation – Notation: $\boldsymbol{X} \equiv is\ a\ matrix\ of\ dimension\ (N, d)$

- ***d*** is the dimension number and **N** the total number of measures.

$$\mathbf{X} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ x_1^2 & x_2^2 & \cdots & x_d^2 \\ \vdots & & & \\ x_1^N & x_2^N & \cdots & x_d^N \end{bmatrix}$$

INTRODUCTION TO MACHİNE LEARNİNG

3RD EDİTİON, ETHEM ALPAYDIN

© The MIT Press, 2014

Discrete d-dimension data representation – Notation: $\boldsymbol{X} \equiv (N, d)$

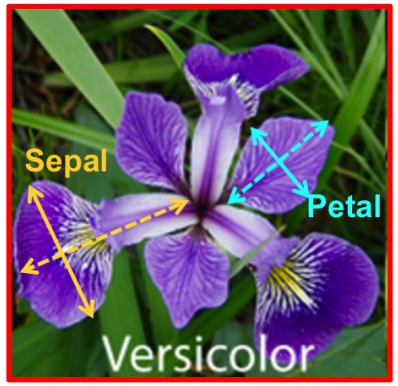
- Another notation for representing data in multiple dimensions

$$\boldsymbol{X_n} = \{x_1[n], x_2[n], \dots, x_d[n]\}_{n=0}^{N-1}, \boldsymbol{X_n} \in R^d, n = 0,1, \dots, N-1$$

$$\boldsymbol{X_n} = \begin{bmatrix} x_1[0] & x_2[0] & x_3[0] & & \dots & x_{d-1}[0] & x_d[0] \\ x_1[1] & x_2[1] & x_3[1] & & \dots & x_{d-1}[1] & x_d[1] \\ \vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\ x_1[N-1] & x_2[N-1] & x_3[N-1] & & \dots & x_{d-1}[N-1] & x_d[N-1] \end{bmatrix}$$

Data Bank example:

- Input data: petal length, petal width, setal length and sepal width.
- Dimension: 4 $(\mathcal{R}^4)$
- Measure (prototype or instance): each line.



{X_1,X_2,X_3,X_4, Class Specification}

5.1, 3.5, 1.4, 0.2, iris-setosa
4.9, 3.0, 1.4, 0.2, iris-setosa
7.0, 3.2, 4.7, 1.4, iris-versicolor
6.3, 3.3, 6.0, 2.5, iris-virginica
5.8, 2.7, 5.1, 1.9, iris-virginica
7.1, 3.0, 5.9, 2.1, iris-virginica
6.3, 2.9, 5.6, 1.8, iris-virginica
4.5, 2.3, 1.3, 0.3, iris-setosa
4.4, 3.2, 1.3, 0.2, iris-setosa
7.0, 3.2, 4.7, 1.4, iris-versicolor
6.4, 3.2, 4.5, 1.5, iris-versicolor
6.9, 3.1, 4.9, 1.5, iris-versicolor
6.5, 3.0, 5.8, 2.2, iris-virginica

…

Expectation value vector

- The **expected value vector** or the **mean vector** of the random vector (matrix) **X** is defined as

$$E[\boldsymbol{X}] = \mu_X = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_d] \end{bmatrix} = \begin{bmatrix} \mu_{\boldsymbol{X}_1} \\ \mu_{\boldsymbol{X}_2} \\ \vdots \\ \mu_{\boldsymbol{X}_d} \end{bmatrix}$$

Correlation matrix

- for a random vector **X**, we define the **correlation matrix**, $\mathbf{R_x}$, as

$$\boldsymbol{R_X} = \mathrm{E}[\boldsymbol{X}^T\boldsymbol{X}] = \begin{bmatrix} \boldsymbol{X}_1^2 & \boldsymbol{X}_1\boldsymbol{X}_2 & \dots & \boldsymbol{X}_1\boldsymbol{X}_d \\ \boldsymbol{X}_1\boldsymbol{X}_2 & \boldsymbol{X}_2^2 & \dots & \boldsymbol{X}_2\boldsymbol{X}_d \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_1\boldsymbol{X}_d & \boldsymbol{X}_2 X_d & \dots & \boldsymbol{X}_d^2 \end{bmatrix} = \begin{bmatrix} E[\boldsymbol{X}_1^2] & \boldsymbol{E}[\boldsymbol{X}_1\boldsymbol{X}_2] & \dots & E[\boldsymbol{X}_1\boldsymbol{X}_d] \\ E[\boldsymbol{X}_1\boldsymbol{X}_2] & E[\boldsymbol{X}_2^2] & \dots & E[\boldsymbol{X}_2\boldsymbol{X}_d] \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{E}[\boldsymbol{X}_1\boldsymbol{X}_d] & E[\boldsymbol{X}_2 X_d] & \dots & E[\boldsymbol{X}_d^2] \end{bmatrix}$$

Covariance matrix

$$\boldsymbol{C}_X = \boldsymbol{\Sigma}_X = \mathrm{E}[(\boldsymbol{X} - E[\boldsymbol{X}])^T(\boldsymbol{X} - E[\boldsymbol{X}])]$$

$$= \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_d) \\ Cov(X_2, X_1) & Var(X_2) & \dots & Cov(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_d, X_1) & Cov(X_d, X_2) & \dots & Var(X_d) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1d} & \sigma_{2d} & \dots & \sigma_d^2 \end{bmatrix}$$

Pearson product-moment correlation coefficient (PPMCC)

$$\boldsymbol{Corr_X} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{12} & 1 & \cdots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{12} & \rho_{2d} & \cdots & 1 \end{bmatrix}$$

$$\rho_{ij} = \frac{E[(X_i - E[X_i])(X_j - E[X_j])]}{\sqrt{\sigma_i{}^2 \sigma_j{}^2}} = \frac{Cov(Xi, Xj)}{\sqrt{Var(X_i)Var(X_j)}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

Multivariate Gaussian (Normal) distribution

- Notation: $\boldsymbol{X} \sim \mathcal{N}(x|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$

$$f_X(\boldsymbol{X}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_X)\}}$$

$\boldsymbol{C}_X = \boldsymbol{\Sigma}_X$ *is the covariance matrix* $(\boldsymbol{d}, \boldsymbol{d})$

$\boldsymbol{\mu}_X = E[\boldsymbol{X}]$ is the expectation $(\boldsymbol{d}, 1)$, $\boldsymbol{d}$ is the dimension.

Multivariate Normal Distribution

$$\boldsymbol{X} \sim \mathcal{N}(x|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X) \Rightarrow f_X(\boldsymbol{X}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_X)\}}$$

$$\boldsymbol{\mu}_X = \begin{bmatrix} \mathrm{E}[\boldsymbol{X}_1] \\ \mathrm{E}[\boldsymbol{X}_2] \\ \vdots \\ \mathrm{E}[\boldsymbol{X}_d] \end{bmatrix} \qquad \boldsymbol{\Sigma}_X = \mathrm{E}[(\boldsymbol{X} - E[\boldsymbol{X}])^T (\boldsymbol{X} - E[\boldsymbol{X}])] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1d} & \sigma_{2d} & \dots & \sigma_d^2 \end{bmatrix}$$

Covariance matrix size: (d, N) x (N, d) = (d, d)

Observe: the dataset has a dimension (N, d), so (N, d)^T = (d, N)

$$Cov(X_i, X_j) = \sigma_{ji} = \sigma_{ij} = (X_i - E[X_i])(X_j - E[X_j]) \quad \sigma_i^2 = E[(X_i - E[X_i])^2]$$

independent and identically distributed (i.i.d.) random variables

Definition Random variables X_1, X_2, ..., X_d are said to be **independent and identically distributed (i.i.d.)** if they are *independent*, and they have the *same marginal distributions*:

$$F_{X_1}(x) = F_{X_2}(x) = \ldots = F_{X_d}(x), \text{ for all } x \in \mathbb{R}.$$

independent and identically distributed (i.i.d.)

- If random variables X_1, X_2, ..., X_d are **i.i.d.**, they will have the same means and variances

$$E[X_1 X_2 \cdots X_{\mathrm{d}}] = E[X_1]E[X_2]\cdots E[X_{\mathrm{d}}] = \text{(statistical Independence)}$$

$$= E[X1]E[X1]\cdots E[X1] = E[X_1]^d \text{ (identically distributed)}$$

Sum of iid Random Variables

- The mean and variance of the sum of ***d*** **independent, identically distributed** (i.i.d.) random variables, each with mean μ and variance σ^2

$$\mathbf{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_d$$

$$E[\boldsymbol{Y}] = E[\boldsymbol{X}_\mathbf{1}] + E[\boldsymbol{X}_\mathbf{2}] + \cdots E[\boldsymbol{X}_\mathbf{d}] = \mathrm{d}\mu$$

$$\mathrm{Var}[\mathbf{Y}] = \mathbf{d}\mathrm{Var}[\mathbf{X}_\mathbf{j}] = \mathrm{d}\sigma_j^2,\ \mathrm{j} = 1,2,\ldots,\mathrm{d}$$

Independent and Identically Distributed (i.i.d.) random variables

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = (F_X(x))^n$$

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = (f_X(x))^n \quad \text{Continuous}$$

$$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = (p_X(x))^n \quad \text{Discrete}$$

Multivariate Gaussian distribution

$$\boldsymbol{X} \sim \mathcal{N}(x|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X) \Rightarrow \mathrm{P}(\boldsymbol{X}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_X)\}}$$

- Example using covarianve matrix: The variables $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_4$ are independents and identically distributed (i.i.d.).

$$\boldsymbol{\Sigma}_X = \begin{bmatrix} 0.68 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.68 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.68 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.68 \end{bmatrix}, \quad \boldsymbol{\mu}_X = \begin{bmatrix} 1.22 \\ 1.22 \\ 1.22 \\ 1.22 \end{bmatrix}$$

Only one marginal PDF is enough to model the joint PDF.

$$PDF: f_X(X) = \left(f_{X_1}(X_1)\right)^4 ; PMF: P_X(X) = P(X_1)^4$$

Multivariate Gaussian distribution

$$\boldsymbol{X} \sim \mathcal{N}(x|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X) \Rightarrow \mathrm{P}(\boldsymbol{X}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_X)\}}$$

- Example using covariance matrix : The variables $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_4$ are independents, but not identically distributed.

$$\boldsymbol{\Sigma}_X = \begin{bmatrix} 0.53 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.32 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.49 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.20 \end{bmatrix}, \quad \boldsymbol{\mu}_X = \begin{bmatrix} 5.93 \\ 2.75 \\ 4.26 \\ 1.32 \end{bmatrix}$$

The joint PDF can be expressed as separable model

$$f_X(X) = f_{X_1}(X_1)\, f_{X_2}(X_2)\, f_{X_3}(X_3)\, f_{X_4}(X_4)$$

$$P_X(X) = P_{X_1}(X_1)\, P_{X_2}(X_2) P_{X_3}(X_3) P_{X_4}(X_4)$$

Multivariate Gaussian distribution

$$\boldsymbol{X} \sim \mathcal{N}(x|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X) \Rightarrow \mathrm{P}(\boldsymbol{X}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_X)\}}$$

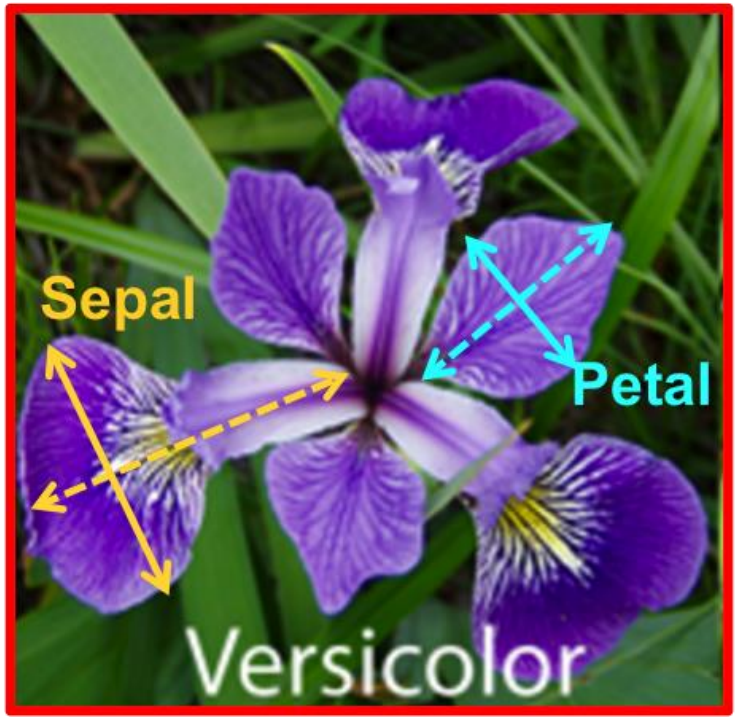
- Example using covariance matrix : Variables $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_4$ are not independent, and neither they are identically distributed.

$$\boldsymbol{\Sigma}_X = \begin{bmatrix} 0.55 & -0.02 & 0.02 & 0.44 \\ -0.02 & 0.72 & -0.21 & -0.07 \\ 0.02 & -0.21 & 2.24 & 0.06 \\ 0.44 & -0.07 & 0.06 & 0.50 \end{bmatrix}, \quad \boldsymbol{\mu}_X = \begin{bmatrix} 4.95 \\ 3.36 \\ 1.45 \\ 0.24 \end{bmatrix}$$

The joint PDF is expressed as non-separable model

Exercise 5.8

- Consider a data bank where a set of measurements of three types of flowers are grouped: iris-setosa, iris-versicolor and iris-virginica. The length and the width of the sepal and petal are measured. The database with 4 columns for each measure is then built.



Exercise 5.8: Solution

- Using Bayes' Law choose randomly 90% of all measures to build a predictor model for each type of type of flower.

- Using the other 10% of the database, test your model and evaluate its accuracy.

- Assume that the four measures taken of each plant are independent random variables.

Exercise 5.8: Naive Bayes Classifier

https://github.com/odubno/gauss-naive-bayes

- **5.1,3.5,1.4,0.2,iris-setosa**
- 4.9,3.0,1.4,0.2,iris-setosa
- 7.0,3.2,4.7,1.4,iris-versicolor
- 6.3,3.3,6.0,2.5,iris-virginica
- 5.8,2.7,5.1,1.9,iris-virginica
- 7.1,3.0,5.9,2.1,iris-virginica
- 6.3,2.9,5.6,1.8,iris-virginica
- 4.5,2.3,1.3,0.3,iris-setosa
- 4.4,3.2,1.3,0.2,iris-setosa
- 7.0,3.2,4.7,1.4,iris-versicolor
- 6.4,3.2,4.5,1.5,iris-versicolor
- 6.9,3.1,4.9,1.5,iris-versicolor
- 6.5,3.0,5.8,2.2,iris-virginica

...

- Three possible classes: iris-setosa, iris-versicolor, iris-virginica.
- N = 150 measurements (or prototypes or instances).
- Dimension = 4.
- Input data (cm): petal length, petal width, setal length and sepal width.

Exercise 5.8: Naive Bayes Classifier

https://github.com/odubno/gauss-naive-bayes

The **Bayes' Rule**

x_j j=1,2,3,...; correspond to one measure (one prototype or one instance) in the **d**-dimension vector space. It is a point in a **d**-dimension vector space.

$$P(C_i|x_j) = \frac{P(x_j|C_i)P(C_i)}{P(x)} = \frac{P(x_j|C_i)P(C_i)}{\sum_k P(x_j|C_k)P(C_k)}$$

$$\sum_{k=1}^{3} P(x_j|C_k)P(C_k) = P(x_j|C_1)P(C_1) + P(x_j|C_2)P(C_2) + P(x_j|C_3)P(C_3)$$

Exercise 5.8: Naive Bayes Classifier

https://github.com/odubno/gauss-naive-bayes

Bayesian Inference

prior *likelihood*

posterior

$$P(C|\mathbf{x}) = \frac{P(C)p(\mathbf{x}|C)}{p(\mathbf{x})}$$

evidence

Exercise 5.8: classes to classify: iris-setosa, iris-versicolor and iris-virginica

- Using **135 instances (90%)** for training and **15 instances (10%)** for testing.

- **The Learning process:**
- **Step1**: Group classes (put learning data together - 90%)

  Observe: This conditions to the class occurrence.

- **Step 2**: Compute the Probability of each class: $P(\boldsymbol{C_k}) = \frac{\boldsymbol{N_k}}{\boldsymbol{N}}$ (relative frequency)

  p(Iris-virginica): $P(\boldsymbol{C_1})$ = 0.33
  p(Iris-setosa): $P(\boldsymbol{C_2})$ = 0.31
  p(Iris-versicolor): $P(\boldsymbol{C_3})$ = 0.36

- **Step 3**: Compute the mean and the standard deviation for **each feature** of **each class**.
- The Likelihood is computed from the Normal Probabilities model.
- The training (learning) is complete.

Exercise 5.8

Grouped Classes

class 0	Class 1	Class 2
5.1,3.5,1.4,0.2,iris-setosa	7.0,3.2,4.7,1.4, versicolor	6.3,3.3,6.0,2.5, virginica
4.9,3.0,1.4,0.2,iris-setosa	6.4,3.2,4.5,1.5, versicolor	5.8,2.7,5.1,1.9, virginica
4.7,3.2,1.3,0.2,iris-setosa	6.9,3.1,4.9,1.5, versicolor	7.1,3.0,5.9,2.1, virginica
4.6,3.1,1.5,0.2,iris-setosa	5.5,2.3,4.0,1.3, versicolor	6.3,2.9,5.6,1.8, virginica
5.0,3.6,1.4,0.2,iris-setosa	6.5,2.8,4.6,1.5, versicolor	6.5,3.0,5.8,2.2, virginica
5.4,3.9,1.7,0.4,iris-setosa	5.7,2.8,4.5,1.3, versicolor	7.6,3.0,6.6,2.1, virginica
4.6,3.4,1.4,0.3,iris-setosa	6.3,3.3,4.7,1.6, versicolor	4.9,2.5,4.5,1.7, virginica
5.0,3.4,1.5,0.2,iris-setosa	4.9,2.4,3.3,1.0, versicolor	7.3,2.9,6.3,1.8, virginica
4.4,2.9,1.4,0.2,iris-setosa	6.6,2.9,4.6,1.3, versicolor	6.7,2.5,5.8,1.8, virginica
4.9,3.1,1.5,0.1,iris-setosa	5.2,2.7,3.9,1.4, versicolor	7.2,3.6,6.1,2.5, virginica
5.4,3.7,1.5,0.2,iris-setosa	5.0,2.0,3.5,1.0, versicolor	6.5,3.2,5.1,2.0, virginica
4.8,3.4,1.6,0.2,iris-setosa	5.9,3.0,4.2,1.5, versicolor	6.4,2.7,5.3,1.9, virginica
4.8,3.0,1.4,0.1,iris-setosa	6.0,2.2,4.0,1.0, versicolor	6.8,3.0,5.5,2.1, virginica
4.3,3.0,1.1,0.1,iris-setosa	6.1,2.9,4.7,1.4, versicolor	5.7,2.5,5.0,2.0, virginica
5.8,4.0,1.2,0.2,iris-setosa	5.6,2.9,3.6,1.3, versicolor	5.8,2.8,5.1,2.4, virginica
5.7,4.4,1.5,0.4,iris-setosa	6.7,3.1,4.4,1.4, versicolor	6.4,3.2,5.3,2.3, virginica
5.4,3.9,1.3,0.4,iris-setosa	5.6,3.0,4.5,1.5, versicolor	6.5,3.0,5.5,1.8, virginica
5.1,3.5,1.4,0.3,iris-setosa	5.8,2.7,4.1,1.0, versicolor	7.7,3.8,6.7,2.2, virginica
5.7,3.8,1.7,0.3,iris-setosa	6.2,2.2,4.5,1.5, versicolor	7.7,2.6,6.9,2.3, virginica
5.1,3.8,1.5,0.3,iris-setosa	5.6,2.5,3.9,1.1, versicolor	6.0,2.2,5.0,1.5, virginica
5.4,3.4,1.7,0.2,iris-setosa	5.9,3.2,4.8,1.8, versicolor	6.9,3.2,5.7,2.3, virginica
5.1,3.7,1.5,0.4,iris-setosa	6.1,2.8,4.0,1.3, versicolor	5.6,2.8,4.9,2.0, virginica
4.6,3.6,1.0,0.2,iris-setosa	6.3,2.5,4.9,1.5, versicolor	7.7,2.8,6.7,2.0, virginica
5.1,3.3,1.7,0.5,iris-setosa	6.1,2.8,4.7,1.2, versicolor	6.3,2.7,4.9,1.8, virginica
4.8,3.4,1.9,0.2,iris-setosa	6.4,2.9,4.3,1.3, versicolor	6.7,3.3,5.7,2.1, virginica
5.0,3.0,1.6,0.2,iris-setosa	6.6,3.0,4.4,1.4, versicolor	7.2,3.2,6.0,1.8, virginica
5.0,3.4,1.6,0.4,iris-setosa	6.8,2.8,4.8,1.4, versicolor	6.2,2.8,4.8,1.8, virginica
5.2,3.5,1.5,0.2,iris-setosa	6.7,3.0,5.0,1.7, versicolor	6.1,3.0,4.9,1.8, virginica
5.2,3.4,1.4,0.2,iris-setosa	6.0,2.9,4.5,1.5, versicolor	6.4,2.8,5.6,2.1, virginica
4.7,3.2,1.6,0.2,iris-setosa	5.7,2.6,3.5,1.0, versicolor	7.2,3.0,5.8,1.6, virginica
4.8,3.1,1.6,0.2,iris-setosa	5.5,2.4,3.8,1.1, versicolor	7.4,2.8,6.1,1.9, virginica
5.4,3.4,1.5,0.4,iris-setosa	5.5,2.4,3.7,1.0, versicolor	7.9,3.8,6.4,2.0, virginica
5.2,4.1,1.5,0.1,iris-setosa	5.8,2.7,3.9,1.2, versicolor	6.4,2.8,5.6,2.2, virginica
5.5,4.2,1.4,0.2,iris-setosa	6.0,2.7,5.1,1.6, versicolor	6.3,2.8,5.1,1.5, virginica
4.9,3.1,1.5,0.1,iris-setosa	5.4,3.0,4.5,1.5, versicolor	6.1,2.6,5.6,1.4, virginica
5.0,3.2,1.2,0.2,iris-setosa	6.0,3.4,4.5,1.6, versicolor	7.7,3.0,6.1,2.3, virginica

Exercise 5.8

$$\boldsymbol{X} \sim \mathcal{N}(x|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X) \Rightarrow f_X(X|C_k) \cong P_X(X|C_k)$$

$$= \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}\{(\boldsymbol{x}-\boldsymbol{\mu}_X)^T \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_X)\}}$$

- $\Sigma_{versicolor}$

$$= \begin{bmatrix} 0.53 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.32 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.49 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.20 \end{bmatrix}$$

- Σ_{setosa}

$$= \begin{bmatrix} 0.35 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.37 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.17 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.11 \end{bmatrix}$$

- $\mu_{iris-versicolor} = \begin{bmatrix} 5.93 \\ 2.75 \\ 4.26 \\ 1.32 \end{bmatrix}$:

- $\mu_{setosa} = \begin{bmatrix} 5.01 \\ 3.43 \\ 1.47 \\ 0.25 \end{bmatrix}$

As the variables are **independent** the model is separable, $f_X(X|Class_i)$ can also be written as

$$f_X(X|Class_i) = f_{X_1}(x_1|Class_i) f_{X_2}(x_2|Class_i) f_{X_3}(x_3|Class_i) f_{X_4}(x_4|Class_i)$$

Exercise 5.8

Posterior Probabilities: Predictor model test

- Iris-versicolor: 5.68e-16, **Iris-setosa: 0.99**, Iris-virginica:1.22e-25
- **Iris-versicolor: 0.99**, Iris-setosa: 3.92e-67, Iris-virginica: 1.83e-4
- Iris-versicolor: 5.82e-17, **Iris-setosa: 1.0**, Iris-virginica:1.387e-26
- **Iris-versicolor: 0.65**, Iris-setosa: 3.02e-141, **Iris-virginica: 0.34**

...

- **Accuracy**: **93.33**% (1 Misclassification)

Probability Bounds: Bonferroni Inequalities

For any events $A_1, A_2, \ldots, A_n$, we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i);$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j);$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k).$$

$$\vdots$$

Markov's Inequality

- Consider ***X*** as any positive continuous random variable.
- If ***X*** is any nonnegative random variable, then, for a > 0

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Markov's Inequality: proof

Let X be any positive continuous random variable, we can write

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{0}^{\infty} x f_X(x) dx \\ &\geq \int_{a}^{\infty} x f_X(x) dx \\ &\geq \int_{a}^{\infty} a f_X(x) dx \\ &= a \int_{a}^{\infty} f_X(x) dx \\ &= a P(X \geq a). \end{aligned}$$

Thus, we conclude

$$P(X \geq a) \leq \frac{EX}{a}, \qquad \text{for any } a > 0.$$

Chebyshev's inequality

- If **X** is any random variable, then for any $b > 0$ we have

$$P(|X - E[X]| \geq b) \leq \frac{Var(X)}{b^2}$$

Chebyshev's inequality: demonstration

Let X be any random variable. If you define $Y = (X - EX)^2$, then Y is a nonnegative random variable, so we can apply Markov's inequality to Y. In particular, for any positive real number b, we have

$$P(Y \geq b^2) \leq \frac{EY}{b^2}.$$

But note that

$$\begin{aligned} EY &= E(X - EX)^2 = Var(X), \\ P(Y \geq b^2) &= P\big((X - EX)^2 \geq b^2\big) = P\big(|X - EX| \geq b\big). \end{aligned}$$

Thus, we conclude that

$$P\big(|X - EX| \geq b\big) \leq \frac{Var(X)}{b^2}.$$

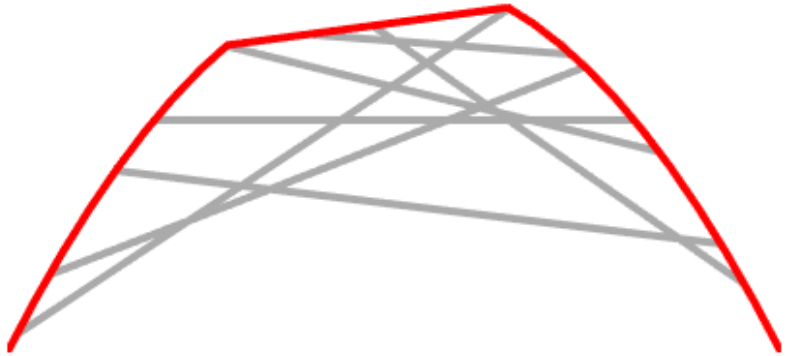
Cauchy-Schwarz Inequality

For any two random variables X and Y, we have

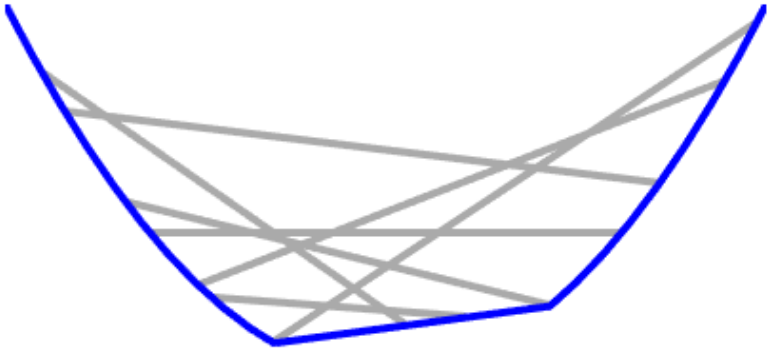
$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]},$$

where equality holds if and only if $X = \alpha Y$, for some constant $\alpha \in \mathbb{R}$.

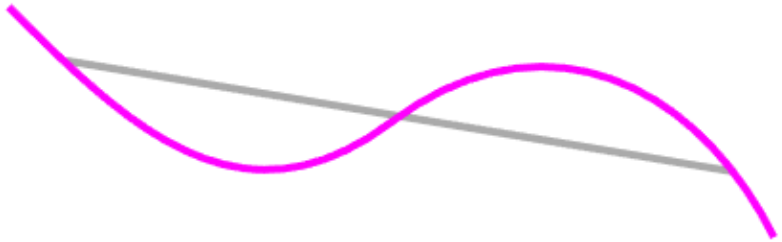
A convex function and a concave function.



A concave function:
no line segment joining
two points on the graph
lies above the graph
at any point



A convex function:
no line segment joining
two points on the graph
lies below the graph
at any point

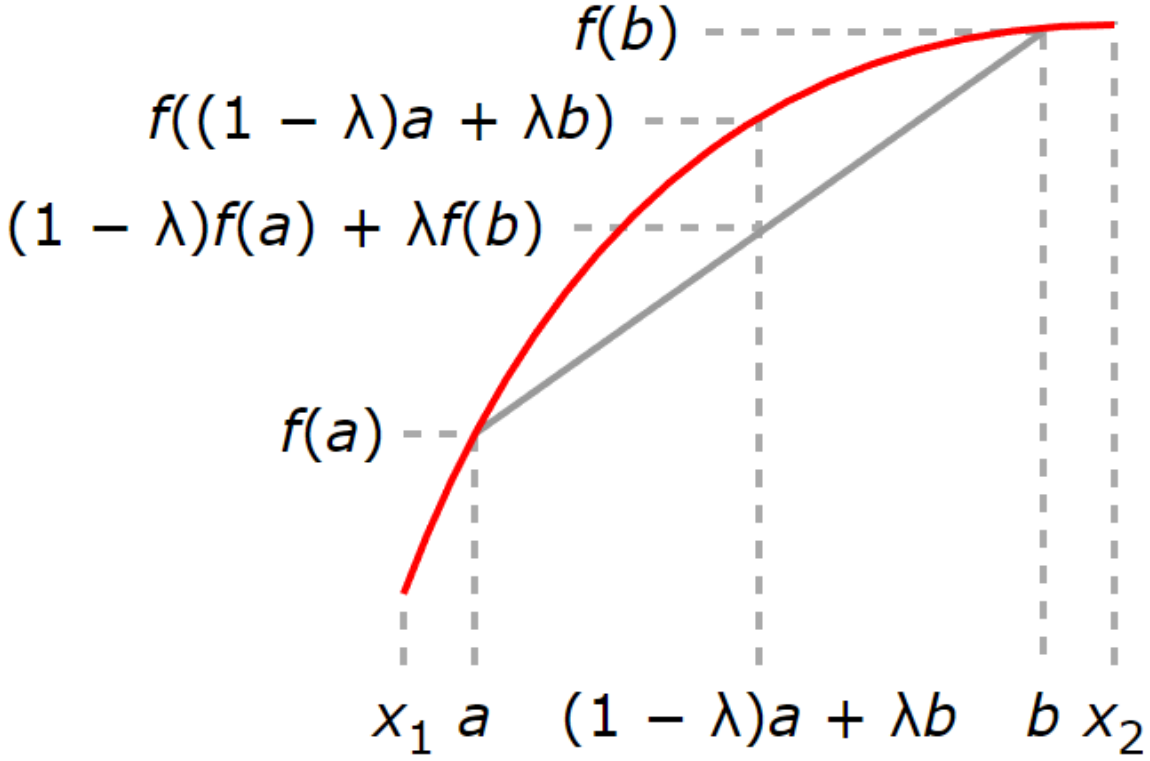


A function that is neither
concave nor convex:
the line segment shown lies
above the graph at some
points and below it at others

A concave function.

- Let $\boldsymbol{f}$ be a function defined on the interval $[x_1, x_2]$ e $\lambda \in [0,1]$.

- This function is **concave** according to the definition if, for every pair of numbers a and b with $x_1 \leq \boldsymbol{a} \leq x_2$ and $x_1 \leq \boldsymbol{b} \leq x_2$, the line segment from $(\boldsymbol{a}, f(\boldsymbol{a}))$ to $(\boldsymbol{b}, f(\boldsymbol{b}))$ lies on or below the graph of the function, as illustrated in the following figure



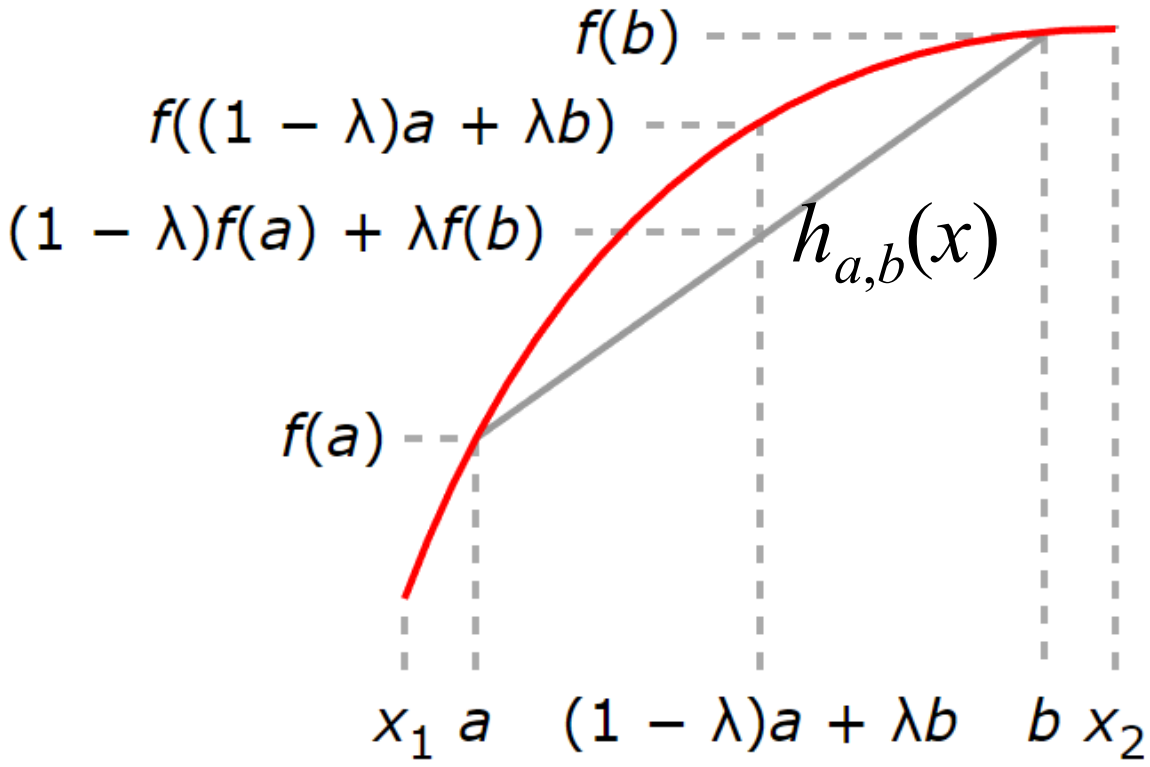
A concave function.

- Denote the height of the line segment from $(a, f(a))$ to $(b, f(b))$ at the point x by $h_{a,b}(x)$.

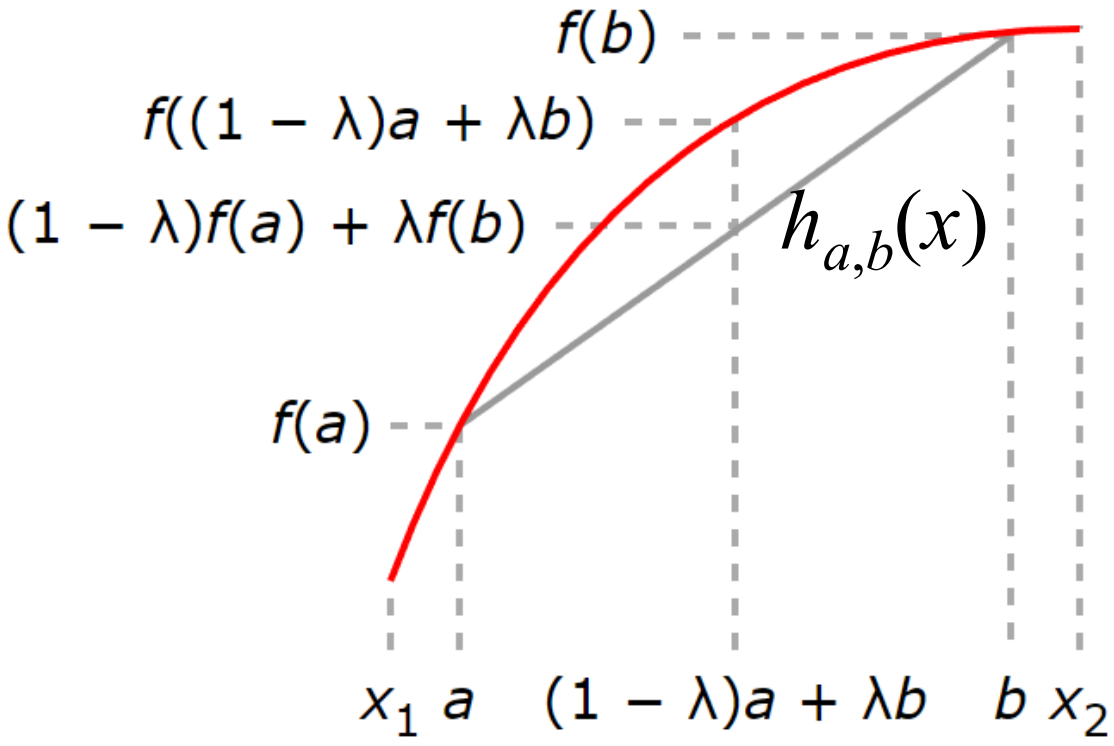
$$h_{a,b}(x) = (1 - \lambda)f(a) + \lambda f(b)$$

- Then according to the definition, the function f is concave if and only if for every pair of numbers, $\boldsymbol{a}$ and $\boldsymbol{b}$ with $x_1 \leq \boldsymbol{a} \leq x_2$ and $x_1 \leq \boldsymbol{b} \leq x_2$ we have

$$f(x) \geq h_{a,b}(x) \text{ for all } x \text{ with } a \leq x \leq b$$



A concave function.



- Every point x with a ≤ x ≤ b may be written as $x = (1 - \lambda)\boldsymbol{a} + \lambda b$, where λ is a real number from 0 to 1(When λ = 0, we have x = a; when λ = 1 we have x = b).

$$h_{a,b}(x) = (1 - \lambda)\mathrm{f}(a) + \lambda\mathrm{f}(b)$$

- Thus, the condition above is equivalent to

$$f((1 - \lambda)a + \lambda b) \geq (1 - \lambda)\mathrm{f}(a) + \lambda\mathrm{f}(b)$$

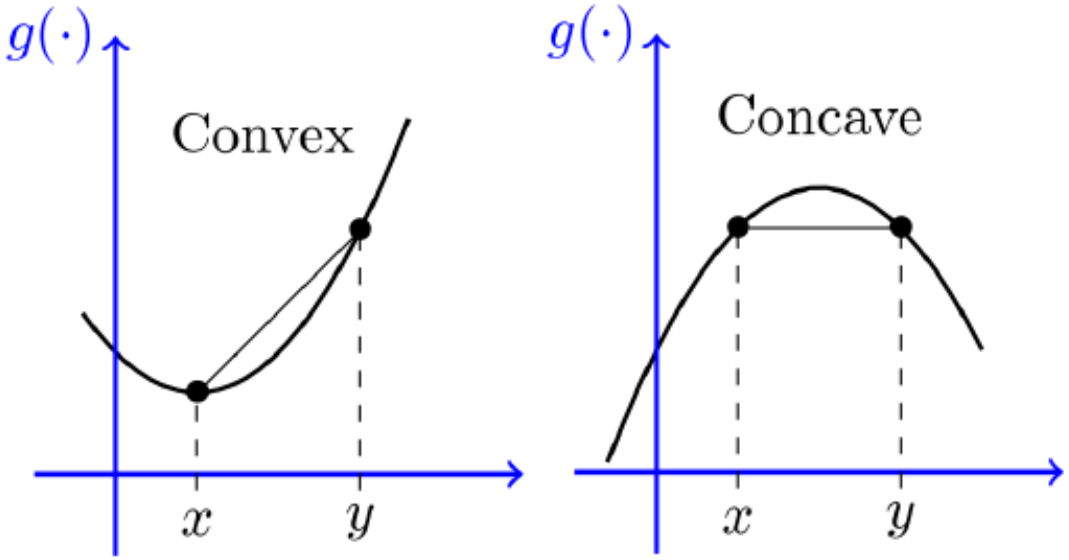
A convex function and a concave function.

Consider a function $g : I \to \mathbb{R}$, where I is an interval in $\mathbb{R}$. We say that g is a **convex** function if, for any two points x and y in I and any $\alpha \in [0, 1]$, we have

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y).$$

We say that g is **concave** if

$$g(\alpha x + (1-\alpha)y) \geq \alpha g(x) + (1-\alpha)g(y).$$



Jensen's Inequality

- Observe that variance of every random variable $\boldsymbol{X}$ is a positive value

$$Var(X) = \sigma_X^2 = E[(X - \mu_{\boldsymbol{X}})^2] = E[\boldsymbol{X}^2] - (E[\boldsymbol{X}])^2 \geq 0$$

- Thus (defining $g(\boldsymbol{X}) = \boldsymbol{x}^2$):

$$E[\boldsymbol{X}^2] \geq (E[\boldsymbol{X}])^2 \Longrightarrow E[g(\boldsymbol{X})] \geq g(E[\boldsymbol{X}])$$

Jensen's Inequality

If $g(x)$ is a convex function on R_X, and $E[g(X)]$ are finite, then

$$E[g(X)] \geq g(E[X]).$$

The **Law of Large Numbers** *(LLN)* *f*or **i.i.d.** random variables

- The **Weak Law of Large Numbers** (**WLLN**) states that if you repeat an experiment independently a large number of times and **average the result** (it is an estimation of the expected value), what you obtain should be **close to the expected value**.

The Weak Law of Large Numbers (WLLN)

Definition For i.i.d. random variables $X_1, X_2, \ldots, X_n$, the **sample mean**, denoted by $\overline{X}$, is defined as

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Another common notation for the sample mean is M_n. If the X_i's have CDF $F_X(x)$, we might show the sample mean by $M_n(X)$ to indicate the distribution of the X_i's.

The **Weak Law of Large Numbers** (WLLN)

For **sample mean** (**i.i.d**. random variables)

(by linearity of expectation)

$$E[\overline{X}] = \frac{EX_1 + EX_2 + \ldots + EX_n}{n}$$
$$= \frac{nEX}{n} \text{ (since } EX_i = EX)$$
$$= EX.$$

$$\mathrm{Var}(\overline{X}) = \frac{\mathrm{Var}(X_1 + X_2 + \ldots + X_n)}{n^2} \text{ (since } \mathrm{Var}(aX) = a^2\mathrm{Var}(X))$$

(since the X_i's are independent)

$$= \frac{\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_n)}{n^2}$$
$$= \frac{n\mathrm{Var}(X)}{n^2} \text{ (since } \mathrm{Var}(X_i) = \mathrm{Var}(X))$$
$$= \frac{\mathrm{Var}(X)}{n}.$$

The **Weak Law of Large Numbers** (WLLN)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with a finite expected value $EX_i = \mu < \infty$. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{X} - \mu| \geq \epsilon) = 0.$$

The **Central Limit Theorem** (CLT)

Observe (from Ex. 5.3): $$Z = \frac{X - \mu_X}{\sigma_X}$$

If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is a **random sample** **(i.i.d.**) of size n taken from a population (either finite or infinite) with mean μ_X and finite variance σ_X^2 and if $\bar{\boldsymbol{X}}$ is the **sample mean**, the limiting form of the distribution of

$$Z_n = \frac{\bar{\boldsymbol{X}} - \mu_{\bar{\boldsymbol{X}}}}{\sigma_{\bar{\boldsymbol{X}}}} = \frac{\bar{\boldsymbol{X}} - \mu_X}{\sigma_X/\sqrt{n}} = \frac{X_1 + X_2 + \cdots X_n - n\mu_X}{\sqrt{n}\sigma_X}$$

as $n \to \infty$, is the standard normal distribution with he zero-mean, unit variance random variable.

The **Central Limit Theorem** (CLT)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with expected value $EX_i = \mu < \infty$ and variance $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Then, the random variable

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sqrt{n}\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

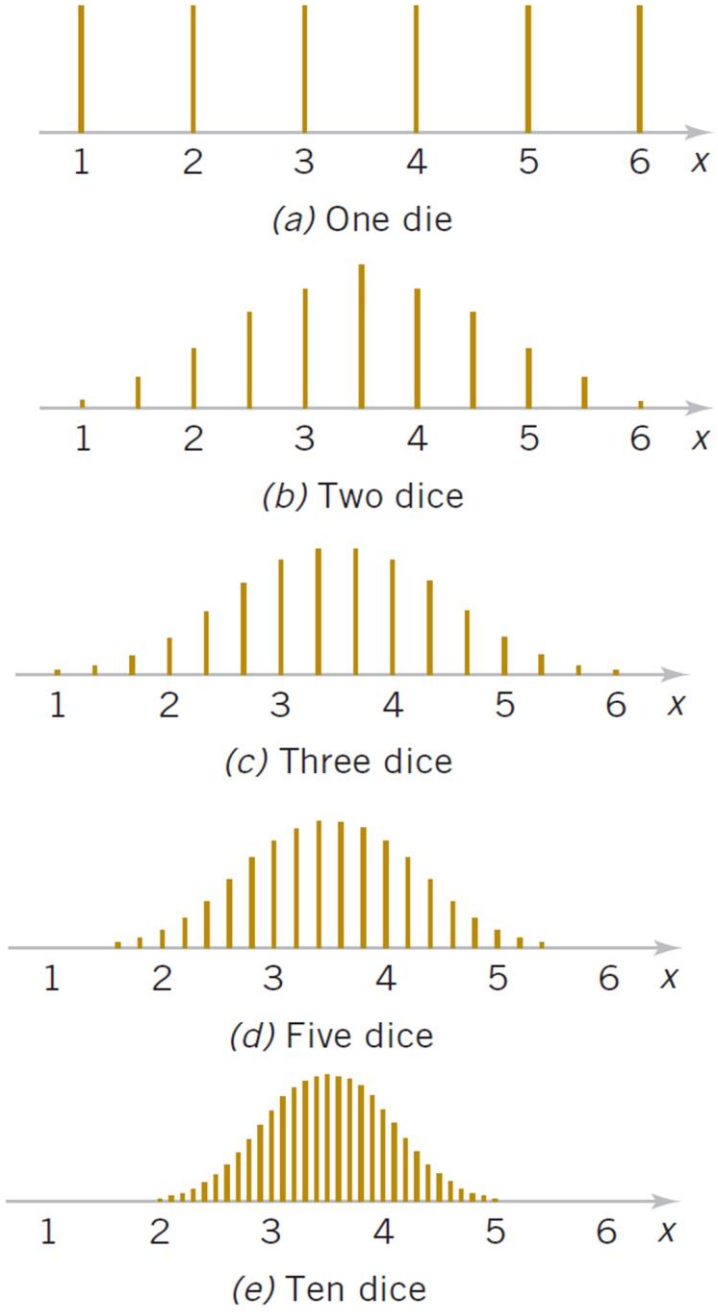
$$\lim_{n\to\infty} P(Z_n \leq x) = \Phi(x), \qquad \text{for all } x \in \mathbb{R},$$

where $\Phi(x)$ is the standard normal CDF.

The **Central Limit Theorem** (CLT)

For **sample mean**

Distributions of average scores from throwing dice.
Source: [Adapted with permission from Box, Hunter, and Hunter (1978).]



Z_n is the normalized num of n independent *Bernoulli*(p) random variables.

Assumptions:

- $X_1, X_2 \ldots$ are iid Bernoulli(p).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

$$EX_i = p$$

$$\mathrm{Var}(X_i) = p(1-p)$$

$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

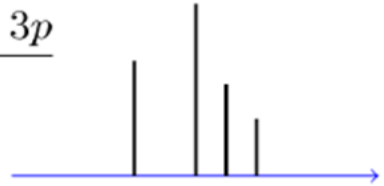
PMF of Z_1

$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

PMF of Z_2

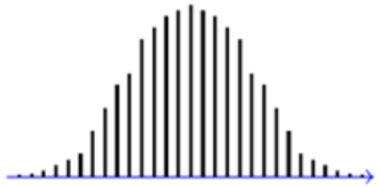
$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

PMF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$

PMF of Z_{30}



Z_n is the normalized sum of n independent *Uniform*(0, 1) Random Variables.

Assumptions:

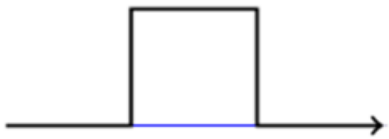
- $X_1, X_2 \ldots$ are iid Uniform(0,1).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$.

$$EX_i = \frac{1}{2}$$

$$\mathrm{Var}(X_i) = \frac{1}{12}$$

$$Z_1 = \frac{X_1 - \frac{1}{2}}{\sqrt{\frac{1}{12}}}$$

PDF of Z_1

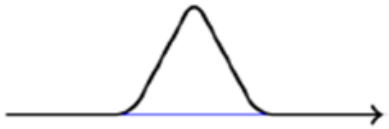


$$Z_2 = \frac{X_1 + X_2 - 1}{\sqrt{\frac{2}{12}}}$$

PDF of Z_2

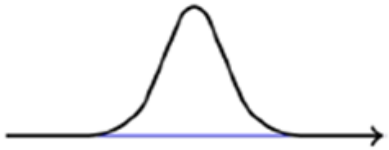
$$Z_3 = \frac{X_1 + X_2 + X_3 - \frac{2}{3}}{\sqrt{\frac{3}{12}}}$$

PDF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - \frac{30}{2}}{\sqrt{\frac{30}{12}}}$$

PDF of Z_{30}



Exercise 5.9

- In a random experiment a dice is rolled twice. The output of the experiment is the arithmetic mean of the two values produced by the dice.
- Determine the PMF of this experiment.

Exercise 5.9: solution

Index	Output	possible values of each dice	P(x)
1	1.0	(1,1)	1/36
2	1.5	(1,2) (2,1)	2/36
3	2.0	(1,3) (3,1) (2,2)	3/36
4	2.5	(1,4) (4,1) (2,3) (3,2)	4/36
5	3.0	(1,5) (5,1) (2,4) (4,2) (3,3)	5/36
6	3.5	(1,6) (6,1) (2,5) (5,2) (3,4) (4,3)	6/36
7	4.0	(2,6) (6,2) (3,5) (5,3) (4,4)	5/36
8	4.5	(6,3) (3,6) (4,5) (5,4)	4/36
9	5.0	(4,6) (6,4) (5,5)	3/36
10	5.5	(5,6) (6,5)	2/36
11	6.0	(6,6)	1/36

Exercise 5.10

- An electronics company manufactures resistors that have a **mean** resistance of 100 ohms and a **standard deviation** of 10 ohms. The distribution of resistance is normal.

- Find the probability that a **random sample** of n = 25 resistors will have an **average** resistance of fewer than 95 ohms.

Exercise 5.10:solution

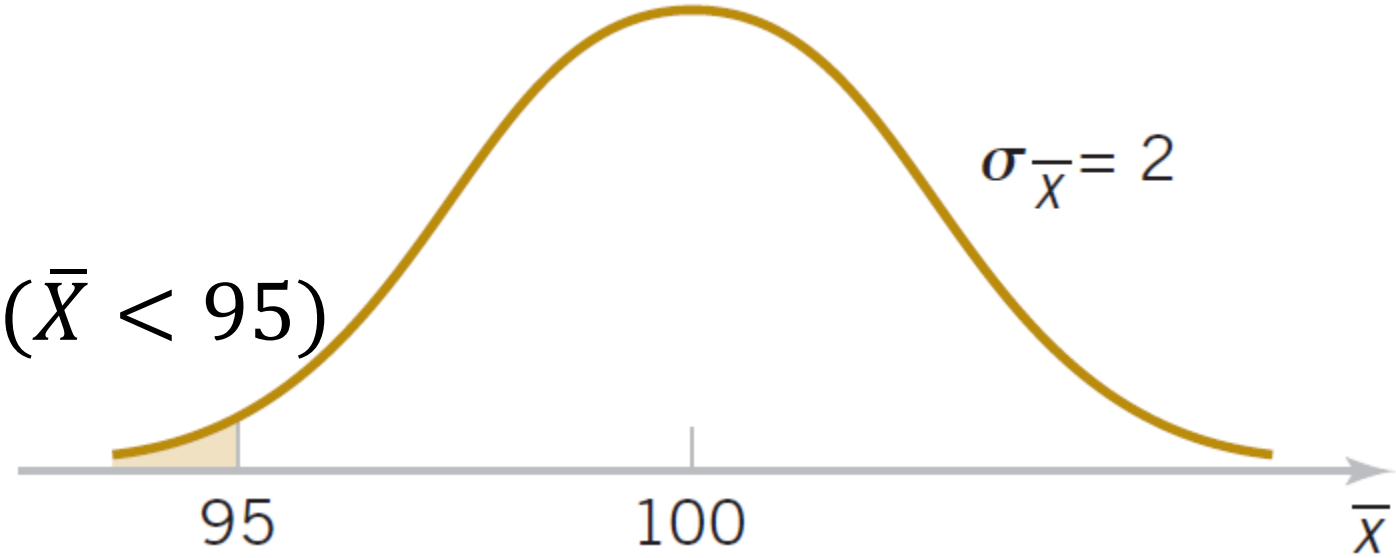
It is a mean sample!

- Note that the sampling distribution of $\bar{X}$ is Gaussian with mean $\mu_X = 100$ ohms and a standard deviation of

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

- Therefore, the desired probability corresponds to the shaded area in figure on the right.

$P(\bar{X} < 95)$



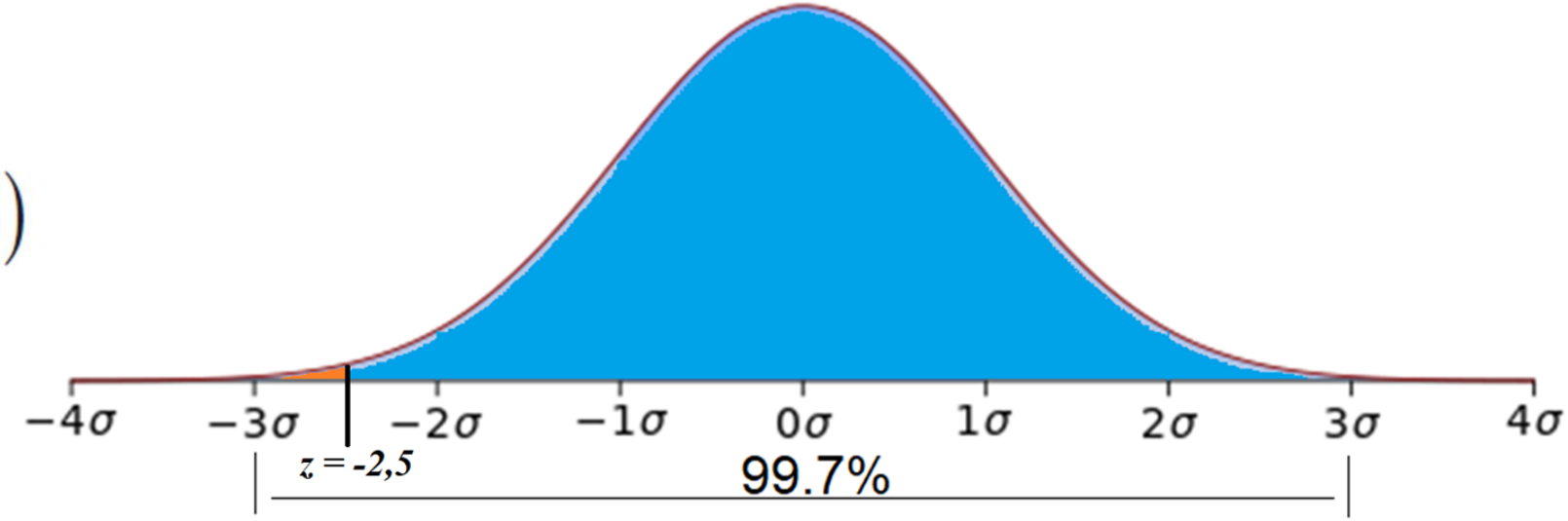
Exercise 5.10:solution

- Using the **Central Limit Theorem** (CLT) knowing $\mu_X = 100\Omega$ and $\sigma_X = 10\Omega$ and standardizing the point $\bar{X} = 95$ in

$$Z = Z_{n=25} = \frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} = \frac{95 - 100}{10/\sqrt{25}} = -2.5$$

and therefore,

$$P(\bar{X} < 95) = P(Z < -2.5)$$
$$= 0.0062$$



Exercise 5.10:solution

- Original problem: $\boldsymbol{X} \sim \mathcal{N}(x|\boldsymbol{\mu}_X, \sigma_X^2) \Longrightarrow \boldsymbol{X} \sim \mathcal{N}(x|100, 10^2)$
- Transformation after using the Central Limit Theorem (CLT):

$$\mathrm{Z} \sim \mathcal{N}(\mathrm{z}|\mu_{\bar{X}} = 0, (\sigma_{\bar{X}})^2 = 1) \Longrightarrow P(\bar{X} < 95) = P(Z < -2.5)$$
$$= 0.0062$$

Practical Conclusion: This example shows that if the distribution of resistance is normal with mean 100 ohms and standard deviation of 10 ohms, finding a **random sample** of 25 resistors with a **sample mean** less than 95 ohms is a **rare event**. If this actually happens, it casts doubt as to whether the true mean is really 100 ohms or if the true standard deviation is really 10 ohms.

How to apply the Central Limit Theorem (CLT)

1. Write the random variable of interest, Y, as the sum of n i.i.d. random variable X_i's:

$$Y = X_1 + X_2 + \ldots + X_n.$$

2. Find EY and $\mathrm{Var}(Y)$ by noting that

$$EY = n\mu, \qquad \mathrm{Var}(Y) = n\sigma^2,$$

where $\mu = EX_i$ and $\sigma^2 = \mathrm{Var}(X_i)$.

3. According to the CLT, conclude that $\frac{Y-EY}{\sqrt{\mathrm{Var}(Y)}} = \frac{Y-n\mu}{\sqrt{n}\sigma}$ is approximately standard normal; thus, to find $P(y_1 \leq Y \leq y_2)$, we can write

$$P(y_1 \leq Y \leq y_2) = P\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma} \leq \frac{Y - n\mu}{\sqrt{n}\sigma} \leq \frac{y_2 - n\mu}{\sqrt{n}\sigma}\right)$$
$$\approx \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma}\right).$$

CLT - Continuity correction for discrete random variables

Let $X_1, X_2, \cdots, X_n$ be independent discrete random variables and let

$$Y = X_1 + X_2 + \cdots + X_n.$$

Suppose that we are interested in finding $P(A) = P(l \leq Y \leq u)$ using the CLT, where l and u are integers. Since Y is an integer-valued random variable, we can write

$$P(A) = P(l - \frac{1}{2} \leq Y \leq u + \frac{1}{2}).$$

It turns out that the above expression sometimes provides a better approximation for $P(A)$ when applying the CLT. This is called the continuity correction and it is particularly useful when X_i's are Bernoulli (i.e., Y is binomial).

Convergence for Sequences of Random Variables

- Convergence in probability distribution function.
- Convergence in probability density function.
- Convergence in expected value (mean).
- Almost sure convergence.

Convergence in probability distribution function.

A sequence of random variables $X_1, X_2, X_3, \cdots$ converges **in distribution** to a random variable X, shown by $X_n \xrightarrow{d} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

for all x at which $F_X(x)$ is continuous.

Convergence in probability density function.

A sequence of random variables $X_1, X_2, X_3, \cdots$ converges **in probability** to a random variable X, shown by $X_n \xrightarrow{p} X$, if

$$\lim_{n\to\infty} P\big(|X_n - X| \geq \epsilon\big) = 0, \qquad \text{for all } \epsilon > 0.$$

Convergence in expected value.

Let $r \geq 1$ be a fixed number. A sequence of random variables X_1, X_2, X_3, $\cdots$ converges **in the rth mean** or **in the L^r norm** to a random variable X, shown by $X_n \xrightarrow{L^r} X$, if

$$\lim_{n\to\infty} E\left(|X_n - X|^r\right) = 0.$$

If $r = 2$, it is called the **mean-square convergence**, and it is shown by $X_n \xrightarrow{m.s.} X$.

Almost sure convergence.

A sequence of random variables $X_1, X_2, X_3, \cdots$ converges **almost surely** to a random variable X, shown by $X_n \xrightarrow{a.s.} X$, if

$$P\left(\left\{s \in S : \lim_{n\to\infty} X_n(s) = X(s)\right\}\right) = 1.$$

The **Strong Law of Large Numbers** (SLLN)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with a finite expected value $EX_i = \mu < \infty$. Let also

$$M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Then $M_n \xrightarrow{a.s.} \mu$.

Continuous mapping theorem.

Theorem Let $X_1, X_2, X_3, \cdots$ be a sequence of random variables. Let also $h : \mathbb{R} \mapsto \mathbb{R}$ be a <u>continuous</u> function. Then, the following statements are true:

1. If $X_n \xrightarrow{d} X$, then $h(X_n) \xrightarrow{d} h(X)$.
2. If $X_n \xrightarrow{p} X$, then $h(X_n) \xrightarrow{p} h(X)$.
3. If $X_n \xrightarrow{a.s.} X$, then $h(X_n) \xrightarrow{a.s.} h(X)$.

The End of Module 5.