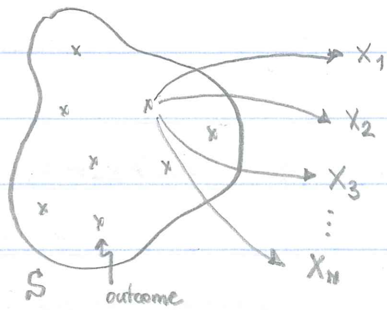
LECTURE 2



Let us consider a scenario where $N > 1$ random variables $X_1, X_2, \dots, X_N$ are defined on a given set S of outcomes

Ex #1: Each RV models a measurement of a given event (tetrode spike sorting)

Ex #2: Each RV conveys only one piece of information about the event (spike count pairs)

* Multivariate Distributions

We are interested in characterizing the vector: $X \triangleq [X_1 \; X_2 \dots X_N]^T$

First, we need to study how these RVs vary together $\Rightarrow$ Remember that:

- $\{a_i < X_i \leq b_i\}$ is an event
- $\{a_i < X_i \leq b_i\}$ is defined on S $\forall i \in \{1, 2, \dots, N\}$

Hence: $P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, \dots, a_N < X_N \leq b_N)$ is the probability of the intersection of N events on S $\Rightarrow$ In analogy with what we saw for a single RV, we need a mathematical function to describe the probability of just these intersections of events $\Rightarrow$ We define the JOINT PDF as:

$$f: (x_1, x_2, \dots, x_N) \in \mathcal{R}^N \longrightarrow \mathcal{R} \quad \text{such that:}$$

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2 \dots, a_N < X_N \leq b_N) = \int_{a_1}^{b_1}\int_{a_2}^{b_2} \cdots \int_{a_N}^{b_N} f(x_1, x_2, \dots, x_N)\, dx_1 \dots dx_N$$

Note: $S = \{-\infty \leq X_i \leq +\infty\}$ for any $i \in \{1, 2, \dots, N\}$

and $\underbrace{\int_{a_1}^{b_1} \cdots \int_{a_{i-1}}^{b_{i-1}} \int_{a_{i+1}}^{b_{i+1}} \cdots \int_{a_N}^{b_N}}_{N-1} f(x_1, \dots, x_N)\, dx_1 \dots dx_N$ is a function of x_i only

Hence, we have:

$$P(a_i < X_i \leq b_i) = P(a_i < X_i \leq b_i, -\infty \leq X_1 \leq +\infty, -\infty \leq X_2 \leq +\infty, \dots) =$$

$$= \int_{a_i}^{b_i} \Big[\underbrace{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{N-1} f(x_1, x_2, \dots x_N)\, dx_1 \dots dx_{i-1}\, dx_{i+1} \dots dx_N \Big] dx_i$$

$$= \int_{a_i}^{b_i} f_{x_i}(x_i)\, dx_i \Rightarrow f_{x_i}(x_i) \triangleq \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{N-1} f(x_1, \dots, x_N)\, dx_1 \dots dx_N$$

by definition of pdf

PDF of the RV X_i

The pdf of the RV X_i is now called "MARGINAL" pdf

Remember that two events A and B are independent iff $P(A \cap B) = P(A) \cdot P(B) \Rightarrow$

Therefore, we say that $X_1, X_2, \dots, X_N$ are INDEPENDENT RVs iff

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, \dots, a_N < X_N \leq b_N) = \prod_{i=1}^{N} P(a_i < X_i \leq b_i)$$

From this definition a condition on the joint pdf follows:

$$f(x_1, x_2, \dots, x_N) = \prod_{i=1}^{N} f_{X_i}(x_i)$$

We characterize the vector X in term of joint probability distribution because, in general, the RVs $X_1, X_2, \dots, X_N$ are NOT independent $\Rightarrow$ How do we measure the amount of dependency between RVs?

We usually measure dependency per pairs of RVs. Numerous different types of measures can be considered. The most used measures are:

- COVARIANCE $\quad Cov(X,Y) \triangleq E\big((X-\mu_X)(Y-\mu_Y)\big)$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_x)(y-\mu_y)\, f(x,y)\, dx\, dy$$

$\mu_X \triangleq E(X)$; $\mu_Y \triangleq E(Y)$

Note this: X, Y - independent $\Rightarrow f(x,y) = f_X(x) f_Y(y) \Rightarrow Cov(X,Y) = 0$

$X = Y \Rightarrow Cov(X,Y) = \sigma_X^2$ - variance

Both facts are consistent with cov being a measure of dependency

However: $Cov(X,Y) = 0 \not\Rightarrow X, Y$ - independent

In fact, consider the following example:

$Y = X^2$

$f_x(x) = f_x(-x)$

$$\Rightarrow \mu_x = E(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx = \int_0^{+\infty} (-x) f_x(-x)\, dx + \int_0^{+\infty} x f_x(x)\, dx = 0$$

$$E(X^3) = \int_{-\infty}^{+\infty} x^3 f_X(x)\, dx = 0$$

$$cov(X,Y) = E\big(X(Y-\mu_Y)\big) = E(X^3) - \mu_Y E(X) = 0$$

- CORRELATION (Pearson's coefficient) $\quad \rho_{XY} \triangleq \dfrac{cov(X,Y)}{\sigma_X \sigma_Y}$

It is used instead of covariance because it is standardized and is not affected by X, and Y's own variability, i.e., it only depends on the joint variation of X, Y.

Moreover, it can be shown that : $-1 \le \rho_{XY} \le 1$

- The notion of "correlation" can be used to predict the value of a RV:

$Y \sim$ RV to be predicted

$X \sim$ RV to be used to predict Y

4

$g(X) = \alpha + \beta X$ - prediction function (*)

It can be shown that the prediction function of form (*) that minimizes the mean square (prediction) error $err = E\left((Y - g(X))^2\right)$ is given when:

$$\alpha = \mu_Y - \beta \mu_X$$

$$\beta = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$$

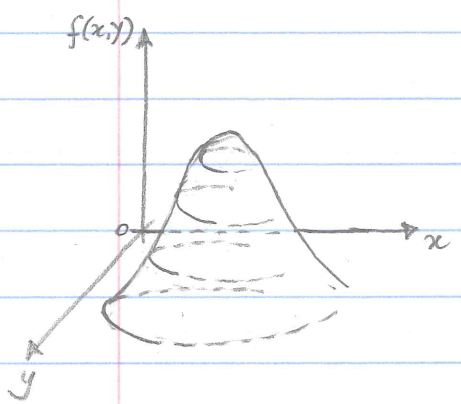
In this case, it can be shown that: $\rho_{XY} = 1 - \dfrac{E\left((Y - \alpha - \beta X)^2\right)}{\sigma_Y{}^2}$ (**)

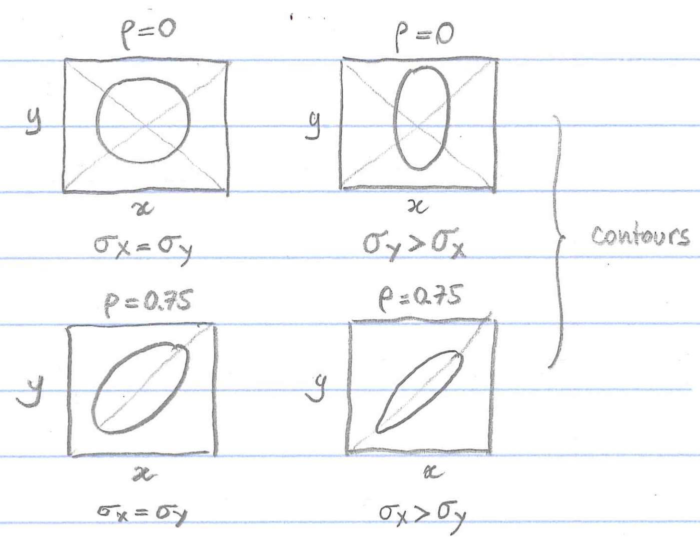
Formula (**) indicates that (i) $err = 0$ when $\rho_{XY} = 1$; (ii) err is maximized when $\rho_{XY} = 0$; (iii) ρ_{XY} is a measure of LINEAR association between X and Y

- The notion of "correlation" can be used to define the shape of a joint pdf:

Two RVs (e.g., X and Y) have a BIVARIATE normal distribution $\overset{DEF}{\iff}$ $f(x,y) = \dfrac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho_{XY}^2}} e^{-\frac{1}{2} Q(x,y)}$

$$Q(x,y) \triangleq \frac{1}{1 - \rho_{XY}^2} \left[\frac{(x - \mu_X)^2}{\sigma_X{}^2} + \frac{(y - \mu_Y)^2}{\sigma_Y{}^2} - 2\rho_{XY} \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X \sigma_Y} \right]$$





In the bivariate normal distribution, the contours satisfy the equation:

$$Q(x,y) = c^*$$

with one value c^* for every contour $\Rightarrow$ Contours are ellipses with axes:

$$y = \frac{\sigma_y}{\sigma_x} x \qquad y = -\frac{\sigma_x}{\sigma_y} x$$

The concentration of the contours around the axes depends on ρ_{xy}, i.e., the concentration increases as $\rho_{xy} \to \pm 1$

Note: X, Y bivariate (a.k.a. "jointly") normal $\Rightarrow$ $X \sim N(\mu_X, \sigma_X^2)$ $Y \sim N(\mu_Y, \sigma_Y^2)$

$\not\Leftarrow$

The joint normality is a condition <u>stronger</u> than just having normal RVs

- MUTUAL INFORMATION $I(X,Y)$ - In order to give a definition for $I(X,Y)$, let us first introduce a few preliminary notions:

a) Let us assume that $f(x)$ and $g(x)$ are two continuous pdf such that:

f, g are both defined on an interval (a,b)
$f(x) > 0 \quad g(x) > 0 \quad \forall x \in (a,b)$
} In this case, a measure of the distance between f and g is:

$$D_{KL}(f,g) \triangleq \int_a^b f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$ - KULLBACK-LEIBER (KL) DIVERGENCE

b) $D_{KL}(\cdot,\cdot)$ satisfies a few useful conditions:

these are now functions of RV

- If X is a RV with $f(x)$ as pdf $\Rightarrow$ $D_{KL}(f,g) = E_X\left(\log f(X) - \log g(X)\right)$

$$= E_X\left(\log f(X)\right) - E_X\left(\log g(X)\right)$$

- $D_{KL}(g,f) \neq D_{KL}(f,g)$ in general, which means that $D_{KL}(\cdot,\cdot)$ captures (nonlinear) discrepancies between $f(\cdot)$ and $g(\cdot)$

- $f(\cdot) = \arg\min_{g(\cdot)} D_{KL}(f,g)$

- If $f(x)$ and $g(x)$ are normal pdfs with same variance σ^2 and mean μ_f and μ_g, respectively, then: $D_{KL}(f,g) = \dfrac{(\mu_f - \mu_g)^2}{\sigma^2}$

Let us assume that X and Y are two RVs with joint pdf $f(x,y)$ and marginal pdf $f_X(x)$ and $f_Y(y)$, respectively. Then we have:

MUTUAL INFORMATION $\quad I(X,Y) \triangleq D_{KL}(f, f_X \cdot f_Y) =$

$$\int_{x_a}^{x_b}\int_{y_a}^{y_b} f(x,y)\log\left(\frac{f(x,y)}{f(x)\,f_Y(y)}\right)dx\,dy$$

intervals wherein $f(\cdot,\cdot)$ is defined

Note: mutual information (MI) provides a measure of the distance of the joint distribution from independence.

Note: If we define the vector $W \triangleq [X\ Y]^T$, we have the following:

- pdf of W is $f(x,y)$. In fact: $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1$ and all the other features of pdfs are satisfied by f

- $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\cdot y\, f(x,y)\,dx\,dy \triangleq E_W(XY)$ by analogy with the definition of "mean" for a single RV

mean of the product of RVs X,Y

Therefore, we can write:

$$I(X,Y) = E_W\left(\log f(X,Y) - \log f_X(X) f_Y(Y)\right)$$

Ex.: X, Y are bivariate normal
$\rho_{XY} \triangleq$ correlation between X and Y $\Rightarrow I(X,Y) = -\frac{1}{2}\log\left(1-\rho_{XY}^2\right)$

$\Rightarrow I(X,Y) = 0$ when X, Y are independent and $I(X,Y) \to +\infty$ as the correlation between X and Y increases

More in general, far any two RVs X and Y whose joint pdf is $f(x,y) > 0$, we have:

X, Y independent $\Rightarrow I(X,Y) = 0$ and $I(\tilde{f}(X), \tilde{g}(X)) = I(X,Y)$ for any pair of one-to-one continuous transformation $\tilde{f}$ and $\tilde{g}$

$\Leftarrow$ This was not guaranteed by correlation

In case of correlation, this would be true only if $\tilde{f}$ and $\tilde{g}$ are linear

In summary, given a vector of RVs $X = [X_1\ X_2 \dots X_N]^T$, we can provide a multivariate probability distribution by defining the joint pdf. In addition, we can provide all the pairwise measures of dependency aggregated in one matrix:

$$\Sigma_I \triangleq \begin{bmatrix} * & m_{12} & \cdots & m_{1N} \\ m_{21} & * & & \vdots \\ \vdots & & \ddots & \vdots \\ m_{N1} & m_{N2} & \cdots & * \end{bmatrix}$$

$\Rightarrow$ This matrix can be envisioned as the adjanceny matrix of a graph whose N nodes are the RVs

$m_{ij} \triangleq$ measure of dependency (e.g., correlation, mutual information, etc.)
$* \triangleq$ conventional value assigned to m_{ii} $i = 1, 2, \dots, N$ if the chosen measure is not definite

8

In the special case when $m_{ij} = \text{cov}(X_i, Y_j)$ we have:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_N\rho_{1N} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ \sigma_N\sigma_1\rho_{N1} & \cdots & & \sigma_N^2 \end{bmatrix}$$ - Covariance Matrix

Note: Σ - symmetric

$\forall w \in \mathcal{R}^{N \times 1}$, $\text{Variance}(w^T X) = E\left(\left(w^T X - w^T \underline{\underline{\mu}}\right)^2\right) = w^T \Sigma w$

where, by definition, $\underline{\underline{\mu}}$ is the vector of means, i.e.:

$$\underline{\underline{\mu}} \triangleq \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_N) \end{bmatrix} \qquad w^T X \triangleq \sum_{i=1}^{N} w_i X_i \qquad w \triangleq \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}$$

Because variance of $w^T X$ is the variance of a random variable Y that is the linear combination of RVs, it must be: $\text{Variance}(w^T X) \geq 0$

$\Rightarrow w^T \Sigma w \geq 0 \quad \forall w \in \mathcal{R}^{N \times 1} \Rightarrow \Sigma$ is a positive semi-definite matrix

* Conditional Densities and Uncertainty

Another approach to the characterization of the vector $X = [X_1 \; X_2 \ldots X_N]^T$ exploits the notion of conditional probability. Consider this:

A, B - events $\quad P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Let us assume: $A = \{X \leq x\}$

$B = \{y < Y \leq y + \Delta y\} \Rightarrow P(X \leq x \mid y < Y \leq y + \Delta y) =$

X, Y are RVs

$$= \frac{P(X \le x,\ y < Y \le y + \Delta y)}{P(y < Y \le y + \Delta y)}$$

By definition of cdf and pdf of a RV, we have:

$$P(X \le x \mid y < Y \le y + \Delta y) = F_X(x \mid y < Y \le y + \Delta y) = \int_{-\infty}^{x} f_x(w \mid y < Y \le y + \Delta y)\, dw$$

$$P(y < Y \le y + \Delta y) = F_Y(y + \Delta y) - F_Y(y) = \int_{y}^{y+\Delta y} f_Y(v)\, dv$$

$$P(X \le x,\ y < Y \le y + \Delta y) = \int_{-\infty}^{x} \int_{y}^{y+\Delta y} f(w, v)\, dw\, dv$$

where: $f(x,y)$ - joint pdf of X and Y

$f_X(x)$, $f_Y(y)$ - marginal pdf of X and Y, respectively

Hence, by dropping the external integral, we have:

$$f_x(x \mid y < Y \le y + \Delta y) = \frac{\int_y^{y+\Delta y} f(x, v)\, dv}{\int_y^{y+\Delta y} f_Y(v)\, dv} \underset{\Delta y \text{ small}}{\cong} \frac{f(x,y)\,\Delta y}{f_Y(y)\,\Delta y}$$

$$f_x(x \mid y) \triangleq \lim_{\Delta y \to 0} f_x(x \mid y < Y \le y + \Delta y) = \frac{f(x,y)}{f_Y(y)}$$ - Conditional Probability Density of X given $Y = y$

It is also written as: $f_{x|Y}(x \mid y)$

Note: $f(x,y) = f_{x|Y}(x \mid y)\, f_Y(y)$ $\Rightarrow$ The joint pdf is equivalent to the pdf of a compound process of first drawing Y with marginal pdf f_Y, and then drawing X with conditional pdf $f_{X|Y}$

Also, if X and Y are independent, then: $f_{X|Y}(x|y) = f_X(x)$.

The notion of conditional pdf can be used to study one RV as a function of the other. Specifically, let us consider:

$$E_X(X|Y=y) \underset{\substack{\uparrow \\ \text{by definition} \\ \text{of mean}}}{=} \int_{-\infty}^{+\infty} x\, f_{x|Y}(x|y)\,dx \Rightarrow E_X(X|Y=y) \underset{\substack{\uparrow \\ \text{It is a function} \\ \text{of } y}}{=} \eta(y) \Rightarrow \eta(Y) \text{ is a RV}$$

Therefore, we can prove that:

$$\left.\begin{aligned} E_Y(\eta(Y)) &= E_Y(E_X(X|Y)) = E_X(X) \\ E_Y(P(X\leq x|Y)) &= P(X\leq x) = F_X(x) \\ \sigma_Y^2(\eta(Y)) &= \sigma_X^2 - E_Y(\sigma_X^2(X|Y)) \end{aligned}\right\}$$

These results indicate that, by conditioning X to Y, we obtain an estimate that, on average, converges to the actual X

This condition often applies to experiments involving biosignals. For instance, consider the case: an experiment is run $N>1$ times, each time with slightly different conditions. For each experimental trial, the recorded signal is itsef random (i.e, two trials run under the same conditions would still return two different signals) $\Rightarrow$ Hence, we have:

$Y_i \triangleq$ a measure computed out of the signal recorded in trial i $\Rightarrow$ Y_i is a RV affected by both signal and trial variability

$\Rightarrow X_i \triangleq E(Y_i)$ is a RV that only depends on trial variability

Based on the stated results, we can write:

$$\underbrace{\sigma_{X_i}^2\left(E_{Y_i}(Y_i|X_i)\right)}_{\text{variance across trials}} = \underbrace{\sigma_{Y_i}^2}_{\text{intra-trial variance}} - E_{X_i}\left(\sigma_{Y_i}^2(Y_i|X_i)\right)$$

Note: the function $\eta(y) = E_x(X|Y=y)$ is called the REGRESSION of X on Y and can be, in general, nonlinear. However, if X and Y are two bivariate normal RVs with $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ and $\rho_{XY} \triangleq$ correlation coefficient between X and Y, then:

$$\eta(y) = \mu_X + \underbrace{\rho_{XY}\frac{\sigma_X}{\sigma_Y}}_{\triangleq \beta_{X|Y}}(y - \mu_Y)$$

Analogously, we can define: $\zeta(x) \triangleq E_Y(Y|X=x)$ and have:

$$\zeta(x) = \mu_Y + \underbrace{\rho_{XY}\frac{\sigma_Y}{\sigma_X}}_{\triangleq \beta_{Y|X}}(x - \mu_X)$$

$$\Rightarrow \rho_{XY} = \beta_{X|Y}\frac{\sigma_Y}{\sigma_X} = \beta_{Y|X}\frac{\sigma_X}{\sigma_Y} \Rightarrow \rho_{XY} = \text{sign}(\beta_{Y|X})\sqrt{\beta_{Y|X}\cdot\beta_{X|Y}}$$

Another useful application of conditional distribution is when the uncertainty of $Y|X$ is less than the uncertainty of Y. Let us use the variance as a measure of uncertainty and let us assume:

$$\left.\begin{array}{l}\zeta(x) = E_Y(Y|X=x) \\ \zeta(x) = \mu_Y + \beta_{Y|X}(x - \mu_X)\end{array}\right. \Rightarrow \sigma_{Y|X}^2 = \underbrace{(1-\rho_{XY}^2)}_{<1}\sigma_Y^2 \Rightarrow \text{The reduction of uncertainty is } \rho_{XY}^2\sigma_Y^2$$

As a result, $\rho_{XY}^2 = \dfrac{\sigma_Y^2 - \sigma_{Y|X}^2}{\sigma_Y^2}$ is a measure of the information about Y supplied by X

Note: $\log \sigma_{Y|X} = \log \sigma_Y - \left(-\frac{1}{2}\log(1-\rho_{XY}^2)\right)$

$$H(X) \triangleq \int_{-\infty}^{+\infty} -f_X(x)\log f_X(x)\,dx$$ - Entropy of the RV X (it measures the disorder in the distribution of X)

$$H(X|Y=y) = -\int_{-\infty}^{+\infty} f_{X|Y}(x|y) \log f_{X|Y}(x|y)\, dx = \eta(y)$$

$$\Downarrow$$

$$H(X|Y) \triangleq E_Y(\eta(Y)) = -\int_{-\infty}^{+\infty} f_Y(y) \left(\int_{-\infty}^{+\infty} f_{X|Y}(x|y) \log f_{X|Y}(x|y)\, dx \right) dy$$

Because $f(x,y) = f_{X|Y}(x|y)\, f_Y(y)$, it is easy to show that:

$H(X|Y) = H(X) - I(X,Y) \Rightarrow$ The MI is the average amount (over Y) by which the entropy of X decreases given the additional information $Y=y$

* Sequences of Random Variables

So far, we have assumed that all the RVs in the vector $X = [X_1\, X_2 \dots X_N]^T$ are defined on S and used simultaneously (i.e., we have focused on the intersection of events, each event being defined on one RV in X). Consider now the following case:

X - RV on S with $E_X(X) = \mu_X$ $\quad E_X((X-\mu_X)^2) = \sigma_X^2$

An experiment is repeated N independent times and each time a sample of X is obtained $\Rightarrow [X(\zeta_1)\ X(\zeta_2) \dots X(\zeta_N)]^T$

these are numbers

$\Rightarrow$ If we consider now the set S^N and define:

$$X_1 : (\zeta_1\, \zeta_2 \dots \zeta_N) \longrightarrow X(\zeta_1) \in \mathcal{R}$$
$$X_2 : (\zeta_1\, \zeta_2 \dots \zeta_N) \longrightarrow X(\zeta_2) \in \mathcal{R}$$
$$\vdots$$
$$X_N : (\zeta_1\, \zeta_2 \dots \zeta_N) \longrightarrow X(\zeta_N) \in \mathcal{R}$$

$\Rightarrow$ Then we have that $[X_1\, X_2 \dots X_N]^T$ is a vector of RVs defined on S^N

Moreover, these RVs satisfy the following conditions:

- cdf of X_i is: $F_{X_i}(x) = F_X(x) \quad \forall x \in \mathcal{R}$
- $X_1, X_2, \dots, X_N$ are independent (as so are the N consecutive experiments

$\Rightarrow$ $[X_1 \dots X_N]^T$ is called a "random sample" and RVs $X_1 \dots X_N$ are said "i.i.d." (independent and identically distributed)

Now, let us suppose that the distribution of X (e.g., μ_X and σ_X^2) must be estimated by using the (sample) distribution of $X_1, X_2, \dots, X_N$, which have been computed on batches on N experiments (N-tuple). We can solve this problem by using the following results:

$$\hat{X} \triangleq \frac{1}{N} \sum_{i=1}^{N} X_i \qquad \text{(it is a RV on } S^N\text{)}$$

Theorem If $(X_1, X_2, \dots, X_N)$ are i.i.d. $\Rightarrow$

$$E_{\hat{X}}(\hat{X}) = \mu_X$$
$$E_{\hat{X}}((\hat{X} - \mu_X)^2) = \frac{\sigma_X^2}{N}$$

Note: This theorem indicates that combining experiments in N-tuple allows to estimate the mean μ_X of X with a variance that is a fraction of σ_X^2. Moreover, if the independence is NOT satisfied (e.g. $Cov(X_i, X_j) = \rho\sigma_X^2$ with $\rho > 0 \quad \forall i,j$), then we have: $E_{\hat{X}}((\hat{X} - \mu_X)^2) = \frac{\sigma_X^2}{N} + \frac{N-1}{N}\rho\sigma_X^2$

Definitions Because we consider the (sample) distributions of $X_1, X_2, \dots, X_N$, we will have N (sample) cdfs: $F_{X_1}(x), F_{X_2}(x), \dots, F_{X_N}(x)$. We say:

$(X_1, X_2, \dots, X_N)$ converges in DISTRIBUTION to X $\overset{DEF}{\iff} \lim\limits_{N \to \infty} F_{X_N}(x) = F_X(x) \quad \forall x \in \mathcal{R}$

$(X_1, X_2, \dots, X_N)$ converges in PROBABILITY to a constant $c^* \in \mathcal{R}$ $\overset{\text{DEF}}{\Longleftrightarrow}$ $X_N \xrightarrow{D} X$ where X is a RV such that $P(X = c^*) = 1$

Note: notation "$X_N \xrightarrow{D} X$" means "convergence in distribution to X".

LLN

<u>Theorem</u> If $(X_1, X_2, \dots, X_N)$ are i.i.d. $\Rightarrow$ The sequence $(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_N)$ with $\hat{X}_i \triangleq \sum_{j=1}^{i} \frac{1}{i} X_j$ converges in probability to μ_X

CLT

<u>Theorem</u> If $(X_1, X_2, \dots, X_N)$ are i.i.d. $\Rightarrow$ Denoted with: $Z_i \triangleq \frac{\sqrt{i}}{\sigma_X}(\hat{X}_i - \mu_X)$ where $\hat{X}_i \triangleq \sum_{j=1}^{i} \frac{1}{i} X_j$, we have that the sequence $(Z_1, Z_2, \dots, Z_N)$ converges in distribution to $N(0,1)$

MCLT

<u>Theorem</u>

- $\bar{\bar{X}} \triangleq \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_m \end{bmatrix}$ where $\bar{X}_j = \frac{1}{N} \sum_{i=1}^{N} X_{ij}$ and X_{ij} are RVs for any i and j
- $\Sigma_1 \triangleq \text{Cov}(\bar{\bar{X}})$ and $\Sigma_1 > 0$ (i.e., positive definite)
- Let us define: $Z_N(w) \triangleq \sqrt{N}\, w^T \Sigma_1^{-1/2} (\bar{\bar{X}} - \bar{\bar{\mu}})$

vector of means of $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m$

$\Rightarrow$ The sequence $(Z_1(w), \dots Z_N(w))$ converges in distribution to $N(0,1)$ $\forall w \in \mathcal{R}^{m \times 1}$ and $w \neq \underline{0}$

$\square$

<u>References:</u>

Textbook: ch 4 (Bayesian Estimators NOT included)

ch 6