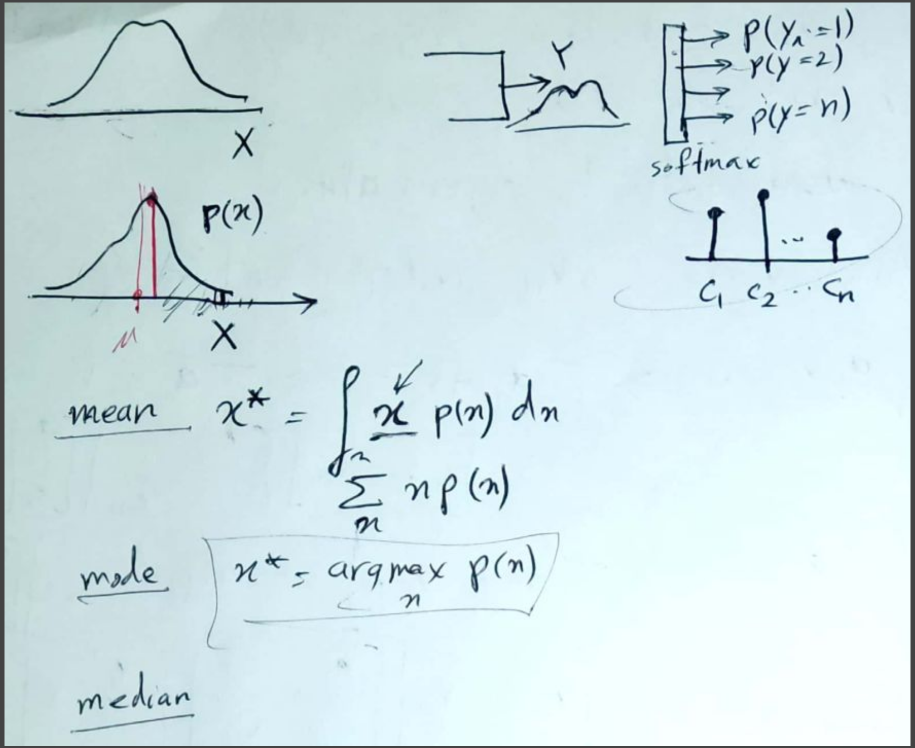
Mathematics for AI

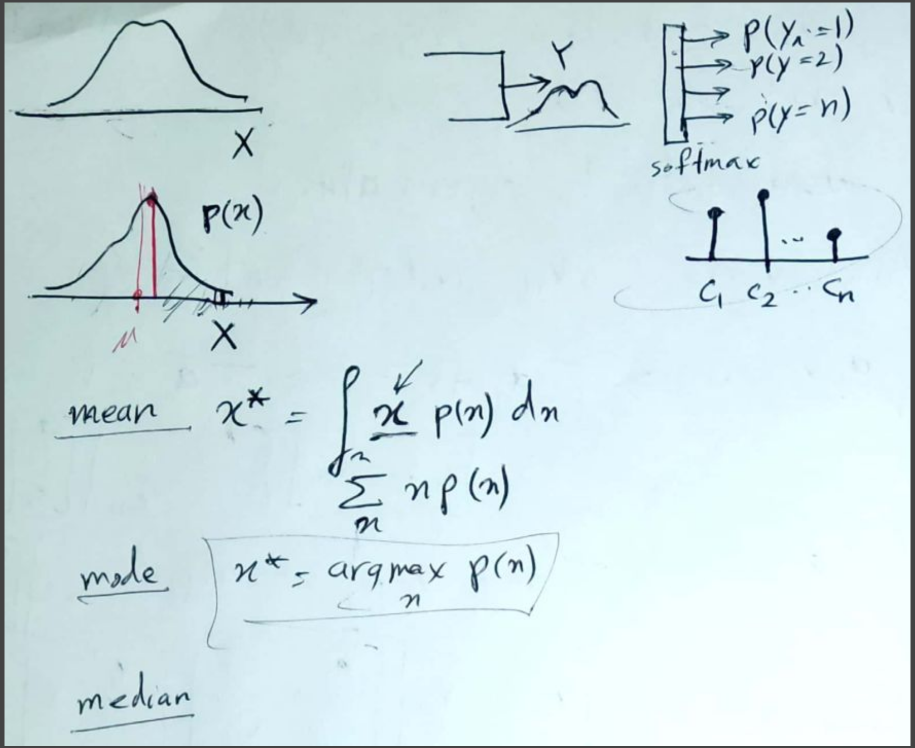
Lecture 23

Statistical Measures: mean, median, mode
variance, standard deviation, covariance matrix

How to represent a random variable with a single number? mean, median, and mode.



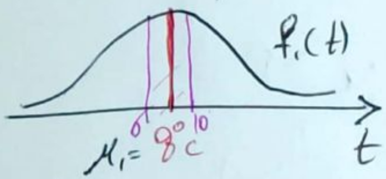
How to represent a random variable with a single number? mean, median, and mode.

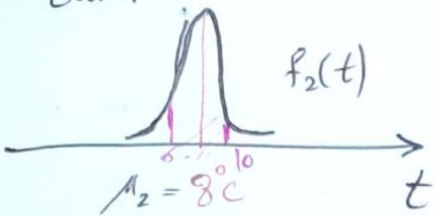


MAD, Variance, and standard deviation

K. N. Toosi
University of Technology

T: temprature tomorrow at 8am.





$\mu_1 = \mu_2 = 8^\circ c$

→ more certainty

با قطعیت بیشتر

$Pr(t \in 8^\circ \pm 2^\circ) = 0.1$

$Pr(t \in (8-9, 8+9)) = 0.95$

$t = 8 \pm 9$ with prob 0.95
with high prob

$Pr(t \in (8-2, 8+2)) = 0.8$

$Pr(t \in (8-2.5, 8+2.5)) = 0.95$

$t = 8 \pm 2.5$ with prob 0.95
〃 high prob

We need a measure of deviation from μ.

MAD, Variance, and standard deviation

We need a measure of deviation from μ.

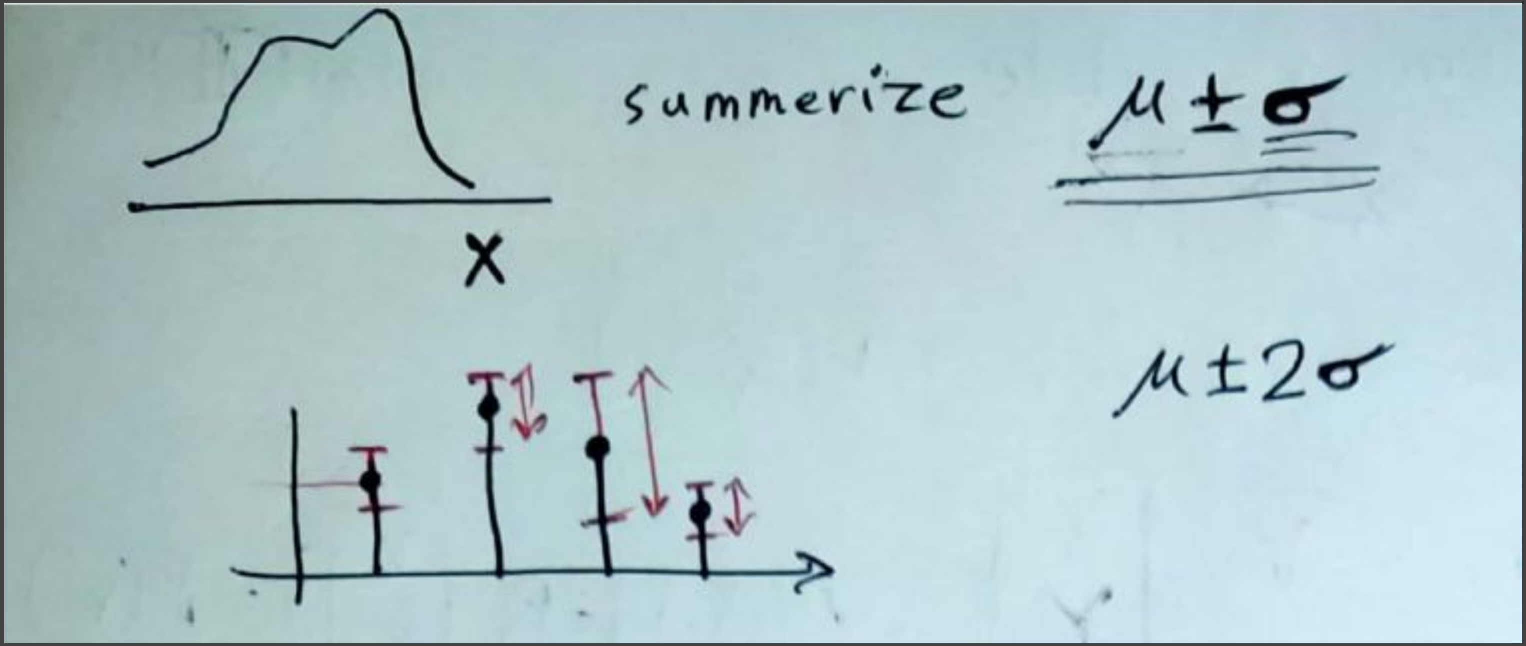
Expected distane from μ: $\int_x |x-\mu|\, p(x)\, dx$: MAD

Expected squared distance from μ: variance

$$\sigma^2 = \int_x (x-\mu)^2 p(x)\, dx = E\{(x-\mu)^2\}$$

Standard deviation $\sqrt{\sigma^2} = \sigma$

Summerize a distribution with mean and std

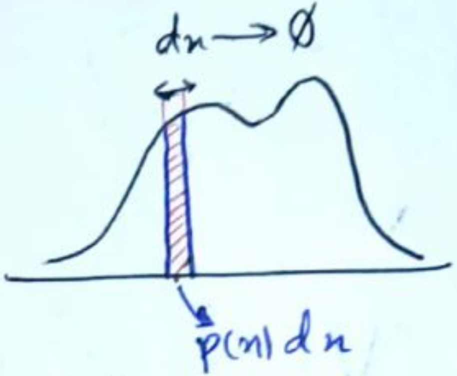


Expectation

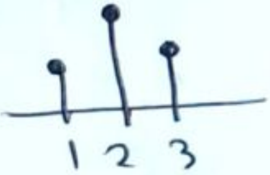
Cont: $E(X) = \int_x x\, p(x)\, dx$

Disc: $E(X) = \sum_x x\, p(x)$

$$E_X\{f(x)\} = \int f(x)\, \underbrace{\overbrace{p(x)}^{\text{prob. density}}\, dx}_{\text{probability}}$$



$$\sum f(x)\, \underbrace{p(x)}_{\text{probabitiy}}$$



$f(1)p(1) + f(2)p(2) + f(3)p(3)$

$\sigma^2 = E\{(x-\mu)^2\}$

Mean, Variance, and MAD as expectations

K. N. Toosi
University of Technology

$$E(X) = \int_x x\, p(x)\, dx$$

$$E(X) = \sum x\, p(x)$$

$$\sigma^2 = E\{(X-\mu)^2\}$$

$$MAD = E\{|X-\mu|\}$$

Expectation is a linear operator

Expectation is a linear operator.

$$E\{ f(X) + g(X) \} = \int \left(\alpha f(x) + \beta g(x) \right) p(x)\, dx$$

$$= \int \alpha f(x)\, p(x)\, dx + \int \beta g(x)\, p(x)\, dx$$

$$= \alpha E\{ f(X) \} + \beta E\{ g(X) \}$$

Example: alternative formulation for variance

K. N. Toosi
University of Technology

$$\sigma^2 = E\{(X-\mu)^2\} = E\{X^2 - 2X\underline{\mu} + \mu^2\}$$

$$= E\{X^2\} - 2\mu\, E(X) + \mu^2 E\{1\}$$

$$= E\{X^2\} - 2\underbrace{\mu \cdot \mu}_{\mu^2} + \mu^2 = E(X^2) - \mu^2$$

$$= E(X^2) - E\{X\}^2$$

What about 2D and ND random vectors?

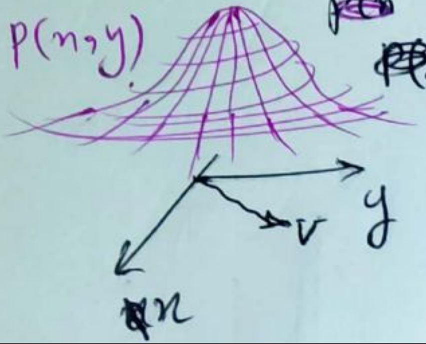
K. N. Toosi
University of Technology

Two or more random variable MA23 III

$\cancel{X_1, X_2, \ldots, X_n}$ $\cancel{X, Y}$

X, Y random variables $P(\underline{\underline{X, Y}})$

Random Vector $\vec{V} = \begin{bmatrix} X \\ Y \end{bmatrix}$ $P(X, Y) = P\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = P(\vec{V})$



Mean $\underset{p(x,y)}{E}(\vec{V}) \in \mathbb{R}^2$

Mean of an 2D distribution

$$\mu = E(\mathbf{V}) = \int \vec{v}\, p(\vec{v})\, d\vec{v} = \iint \begin{bmatrix} x \\ y \end{bmatrix} p\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) dx\, dy$$

$$= \int_x \int_y \underset{\substack{\downarrow \\ \in \mathbb{R}^2}}{\begin{bmatrix} x \\ y \end{bmatrix}} \underset{\substack{\downarrow \\ \mathbb{R}}}{p(x, y)}\, dx\, dy = \mu \in \mathbb{R}^2$$

Mean of an 2D distribution is equal to the vector of means of each variable

$\in \mathbb{R}$ $\quad$ $\mathbb{R}$

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \qquad \mu_x = \int_x \int_y x \, p(x,y) \, dx \, dy$$

$$= \int_x x \left(\int_y p(x,y) \, dy \right) dx$$

$$= \int_x x \, p(x) \, dx = E(X)$$

$$\mu = E\{V\} = \underset{p(x,y)}{E} \left\{ \begin{bmatrix} X \\ Y \end{bmatrix} \right\} = \begin{bmatrix} E_{p(x)}\{X\} \\ E_{p(y)}\{Y\} \end{bmatrix} \in \mathbb{R}^2$$

Variance for 2D distributions

K. N. Toosi
University of Technology

$X, Y \qquad \mu = E\left\{\begin{bmatrix} X \\ Y \end{bmatrix}\right\} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$

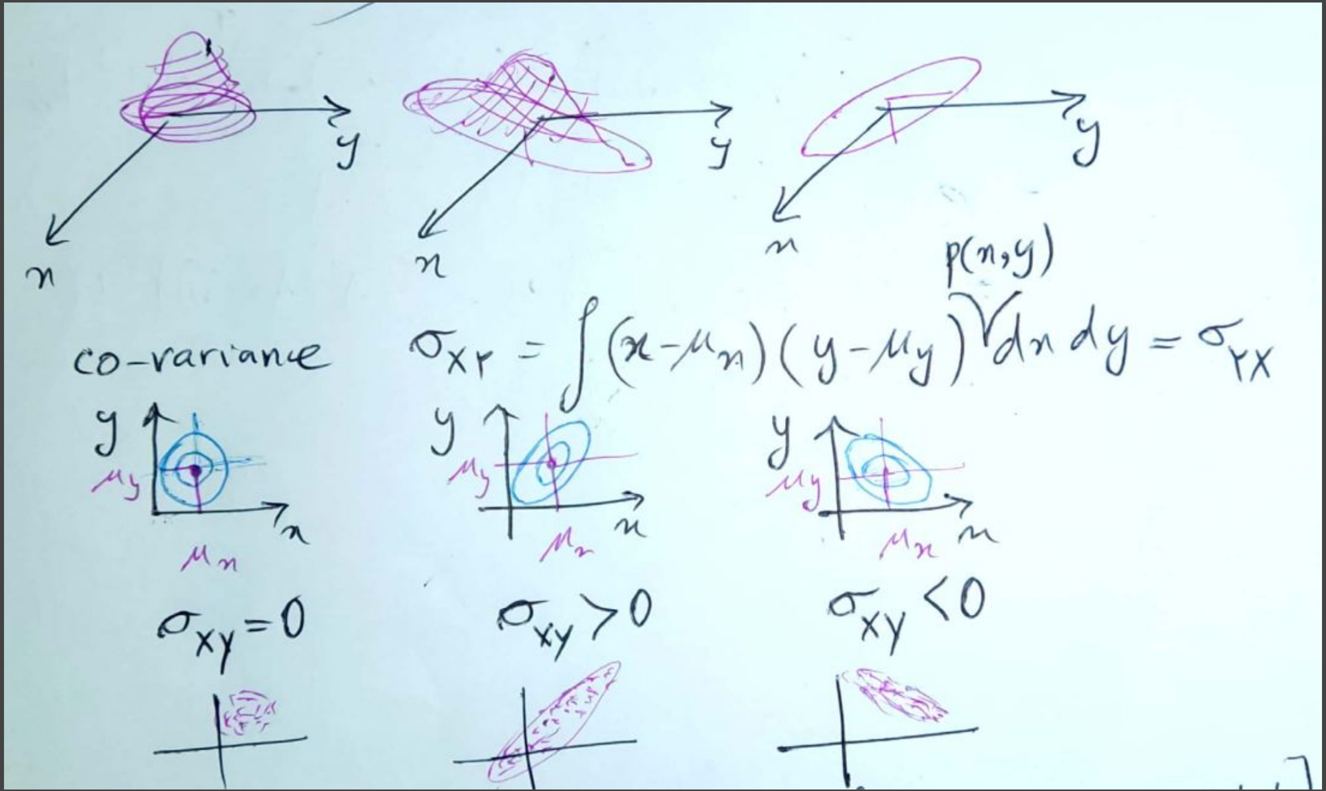
MA23 I

$(x-\mu_x)^2$

$$\sigma_X^2 = \iint (x-\mu_x)^2 \, \cancel{dxdy} \, p_{X,Y}(x,y)\, dx\,dy = \int \cancel{a}\, \underline{p_X}(x)\, dx$$

$$\sigma_Y^2 = \iint (y-\mu_y)^2 \, p_{X,Y}(x,y)\, dx\,dy = \int (y-\mu_y)^2 p_Y(y)\, dy$$

Covariance



y
x
p(x,y)
co-variance
$\sigma_{XY} = \int (x-\mu_x)(y-\mu_y)\, dx\, dy = \sigma_{YX}$
μ_y
μ_x
$\sigma_{xy} = 0$
$\sigma_{xy} > 0$
$\sigma_{xy} < 0$

The Covariance Matrix

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \iint (x-\mu_x)^2 \, p(x,y)\,dx\,dy & \iint (x-\mu_x)(y-\mu_y)\, p(x,y)\,dx\,dy \\ \iint (x-\mu_x)(y-\mu_y)\, p(x,y)\,dx\,dy & \iint (y-\mu_y)^2\, p(x,y)\,dx\,dy \end{bmatrix}$$

covariance matrix

$$= \iint \begin{bmatrix} (x-\mu_x)^2 & (x-\mu_x)(y-\mu_y) \\ (x-\mu_x)(y-\mu_y) & (y-\mu_y)^2 \end{bmatrix} p(x,y)\,dx\,dy$$

$$= \iint \underbrace{\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}}_{2\times 1} \underbrace{\begin{bmatrix} x-\mu_x , & y-\mu_y \end{bmatrix}}_{1\times 2} p(x,y)\,dx\,dy = \iint \underbrace{\underbrace{(V-\mu)}_{2\times 1}\underbrace{(V-\mu)^T}_{1\times 2}}_{2\times 2} \underbrace{p(V)}_{1\times 1}\,dV \in \mathbb{R}^{2\times 2}$$

The Covariance Matrix

K. N. Toosi
University of Technology

$$V = \begin{bmatrix} X \\ Y \end{bmatrix} \qquad \mu = E\{V\}$$

$$\Sigma = E\left\{ (V-\mu)(V-\mu)^T \right\}$$

$\Sigma \downarrow$ covariance matrix

N-dimensional random vectors

K. N. Toosi
University of Technology

n random variables:

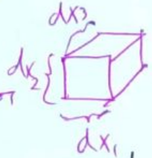
$X_1 , X_2 , \dots , X_n$ joint dist $p(\underbrace{X_1, X_2, \dots, X_n}_{X})$

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Random Vector

$$p(\vec{X}) = p\left(\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}\right) = p(X_1, X_2, \dots, X_n)$$

$$\vec{\mu}_X = E\{X\} = \int \cancel{X}\, \vec{X}\; p(\vec{X})\, d\vec{X}$$



$dx = dx_1 dx_2 \cdots dx_n$

$$= \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \vdots \\ \mu_{X_n} \end{bmatrix} \in \mathbb{R}^n$$

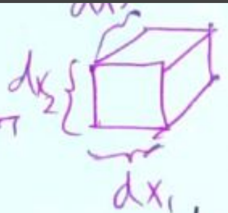
$$Var(X_i) = E\{(X_i - \mu_{X_i})^2\}$$

mean, variance, and covariance of N-D random vectors

K. N. Toosi
University of Technology

Random Vector

$$\vec{\mu}_X = E\{X\} = \int \cancel{x}\ \vec{x}\ p(\vec{x})\, d\vec{x}$$



$$dx = dx_1\, dx_2 \cdots dx_n$$

$$= \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \vdots \\ \mu_{X_n} \end{bmatrix} \in \mathbb{R}^n$$

$$Var(X_i) = E\{(X_i - \mu_{X_i})^2\}$$

$$Cor(X_i, X_j) = E\{(X_i - \mu_{X_i})(X_j - \mu_{X_j})\}$$

The Covariance matrix of N-D random vectors

$$\Sigma = E\{(X-\vec{\mu}_x)(X-\vec{\mu}_x)^T\} \in \mathbb{R}^{n\times n}$$

Covariance matrix

$$= \int \underbrace{\underbrace{(\vec{X}-\vec{\mu}_x)}_{n\times 1}\underbrace{(\vec{X}-\vec{\mu}_x)^T}_{1\times n}}_{n\times n} \underbrace{p(\vec{X})}_{1\times 1}\, d\vec{X} \in R^{n\times n}$$

$$\Sigma_{ii} = Var(X_i) \qquad \Sigma_{ij} = Cov(X_i, X_j)$$

Properties of the Covariance matrix

K. N. Toosi
University of Technology

MA

1- $\Sigma^T = \Sigma$

2- Σ is positive semi-definite

$$y^T \Sigma y = y^T \left(\int (x-\mu)(x-\mu)^T p(x)\, dx \right) y$$

$$= \int y^T (x-\mu)(x-\mu)^T y \; \underline{p(x)}\, dx$$

$$= \int \underbrace{\|(x-\mu)^T y\|^2}_{\geq 0} \; \underbrace{p(x)}_{\geq 0} \, dx \geq 0$$

Positive Definiteness

K. N. Toosi
University of Technology

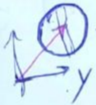
2- Σ is positive semi-definite

$$y^T \Sigma y = y^T \left(\int (x-\mu)(x-\mu)^T p(x)\, dx \right) y$$

$$= \int y^T (x-\mu)(x-\mu)^T y \; p(x)\, dx$$

$$= \int \underbrace{\|(x-\mu)^T y\|^2}_{\geq 0} \; \underbrace{p(x)}_{\geq 0} \, dx \geq 0$$

in many cases Σ is positive definite

$p(x) > 0$  $\rightarrow p(x) \not> 0$