ASEN 5044, Fall 2024

Statistical Estimation for Dynamical Systems

Lecture 15: Mean Vectors and Covariance Matrices; Multivariate Gaussian Density Functions and Jointly Gaussian Random Vectors

Prof. Nisar Ahmed (Nisar.Ahmed@Colorado.edu)

Tuesday 10/08/2024

Ann and H.J. Smead
Aerospace Engineering Sciences
UNIVERSITY OF COLORADO BOULDER

Announcements

- **<u>Midterm 1: due Thurs 10/10 by 11:59 pm (Gradescope)</u>**
  - **No office hours today or this week (b/c of exam)**
- HW 5 to be posted Thurs 10/10 (to be due Fri 10/18)
- Quiz 5 out on Fri 10/11 (due Tues 10/15)
- Next advanced topic Lecture to be posted Fri 10/11

Last Time…

- Joint distributions for multiple random variables and operations
- Continuous multivariate expectations
  - means, cross-moment, covariance, correlation (normalized covariance)
  - examples for simple linear random function of Gaussian random variables

(last time) Useful (Scalar-Valued) Multivariate Moments and Expectations

- Expectation of XY (product of X and Y, i.e. "cross-moment"):

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot p(x,y)\,dx\,dy$$

If $X \perp\!\!\!\perp Y$: $E[xy] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, p(x) \cdot p(y)\,dx\,dy$

$$= \left(\int_{-\infty}^{\infty} x\,p(x)\,dx\right)\left(\int_{-\infty}^{\infty} y\,p(y)\,dy\right) = E[x]\cdot E[y] = E[xY]$$

$\cancel{\Leftrightarrow}$ if $\cancel{\text{and only if}}$ $X \perp\!\!\!\perp Y$

*Statement should read:
E[xy]=E[x]*E[y] IF X and Y are independent, i.e. it is sufficient (but NOT necessary) for X and Y to be independent for their cross-moment to split into the product of their means (this is actually different from the fact p(x,y)=p(x)*p(y) if and only if [iff] X and Y are independent)

If $E[xY] = 0$
Then we say that x & Y are orthogonal RVs

- Covariance of X and Y: how **linearly related** X and Y are to each other

$$\text{cov}(x,y) \triangleq E[(x - \underset{E[x]}{m_x})(y - \underset{E[Y]}{m_y})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)(y-m_y)\,p(x,y)\,dx\,dy$$

Foil out:

$$E[xY - x m_y - m_x y + m_x m_y] = E[xY] - \underset{m_x}{E[x]}m_y - m_x \underset{m_y}{E[y]} + m_x m_y = E[xY] - m_x m_y$$

- Correlation coefficient: covariance normalized by product of standard devs

$$\text{corr}(x,y) = \rho(x,y) \triangleq \frac{\text{cov}(x,y)}{\text{stdev}(x)\cdot\text{stdev}(y)} = \frac{E[(x-m_x)(y-m_y)]}{\sqrt{\text{var}(x)}\cdot\sqrt{\text{var}(y)}}$$

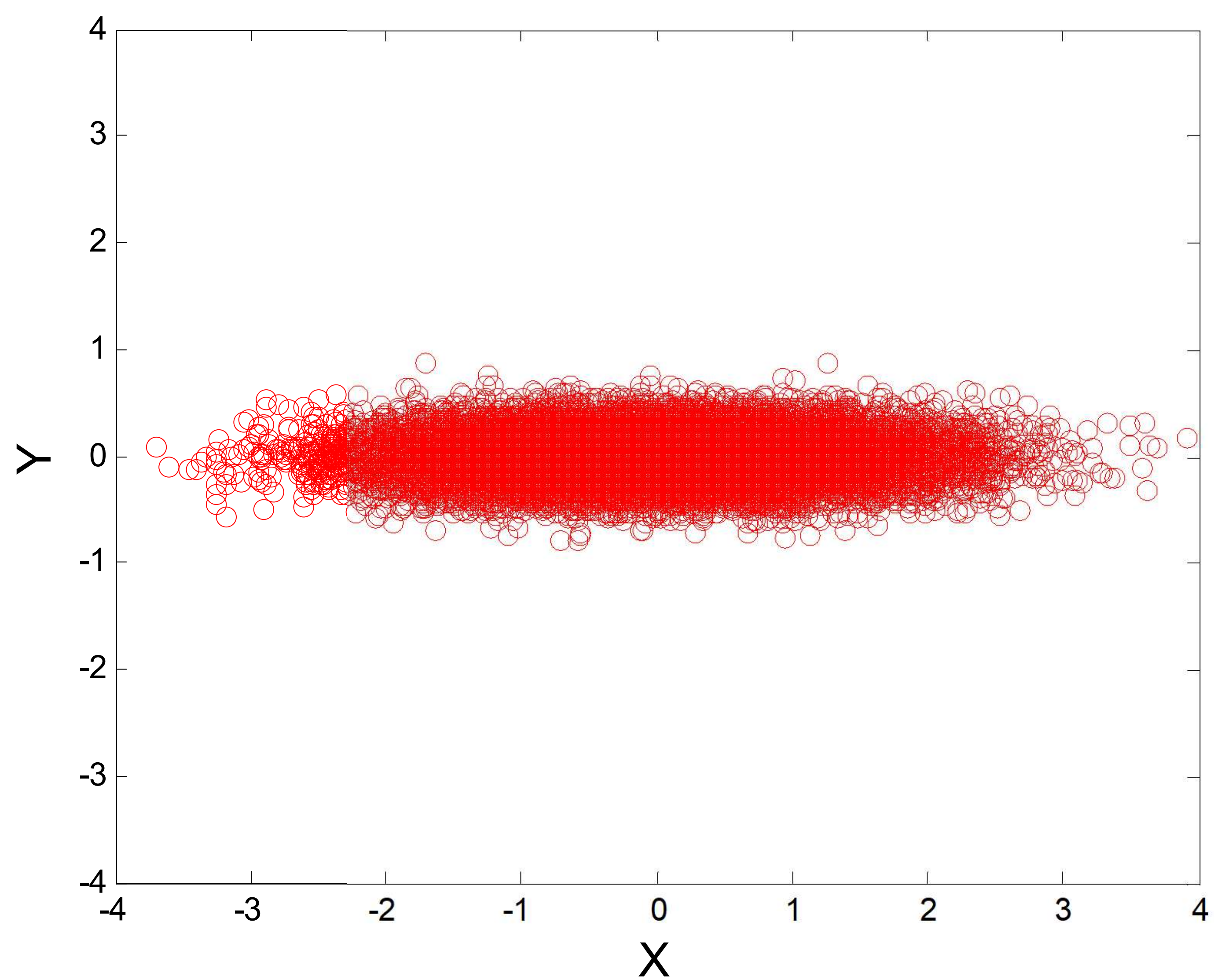
"normalized covariance of x & y"

where $\rho \in [-1, 1]$

where $\text{var}(x) = E[(x-m_x)^2]$
$\text{var}(y) = E[(y-m_y)^2]$

(last time) Example: Sample Covariances and Correlations

- Consider $x \sim N(0,1)$ with 20,000 samples
- Evaluate $y = 0*x + 0.2*e$, where error $e \sim N(0,1)$



Sample values (Matlab)
$$cov(x,y) = 0.0016$$
$$corr(x,y) = \rho = 0.0082$$

No linear information between X and Y
(i.e. there is no <u>linear</u> dependence)

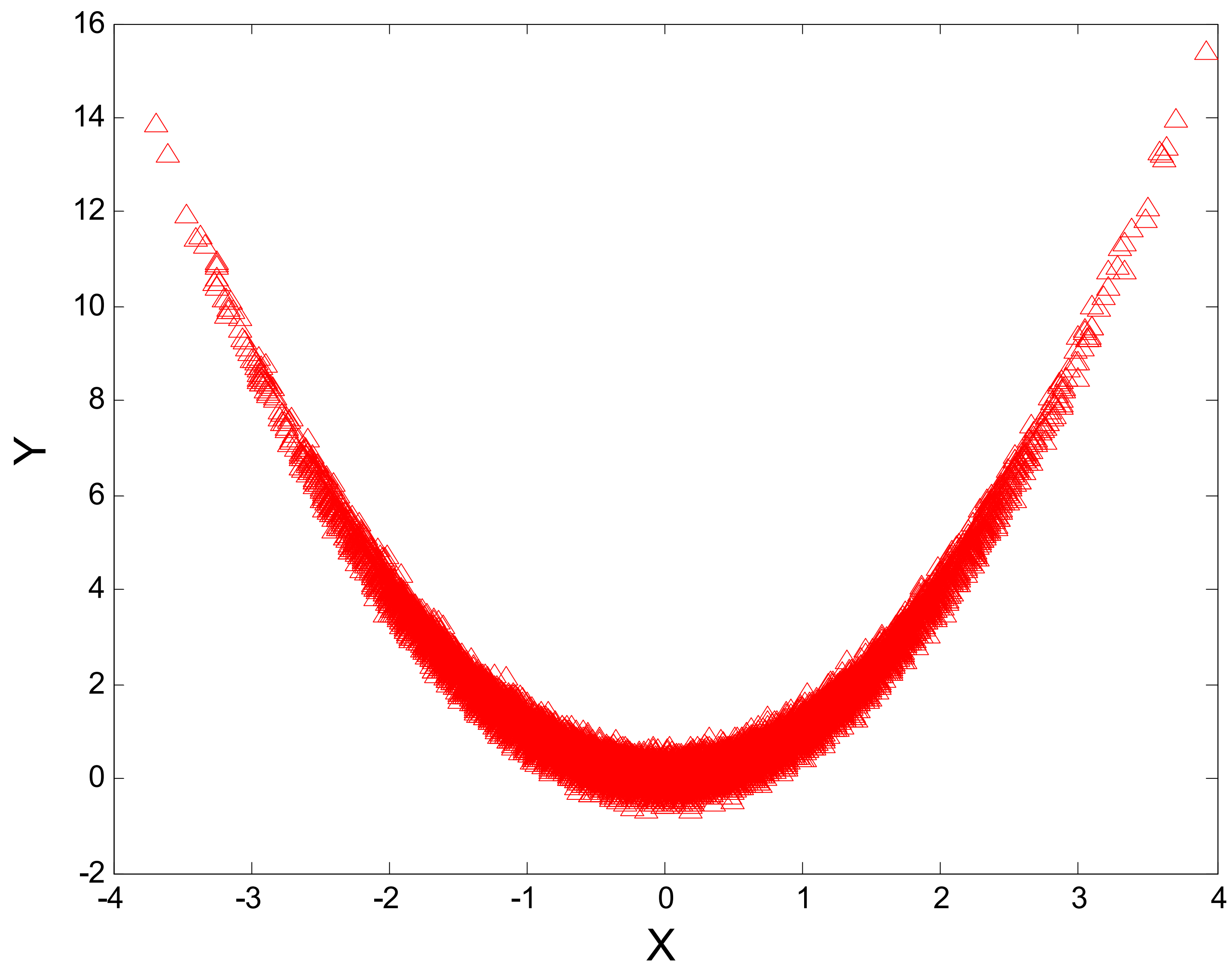
Cov & corr $\rightarrow 0$ as $N \rightarrow \infty$

$\rightarrow$ No linear info b/w x & Y

$\rightarrow x \perp\!\!\!\perp Y$

(last time) Example: Sample Covariances and Correlations

- Consider $x \sim N(0,1)$ with 20,000 samples
- Evaluate $y = x^2 + 0.2*e$, where error $e \sim N(0,1)$



Clarification from live lecture discussion: Since X is NOT independent of Y, E[XY] does NOT equal a product of two separate integrals over X and Y (indeed: easy to show E[XY] = E[X^3], which implies integral doesn't split). However, since E[XY]=E[X^3]=0 *anyway*, this gives the same result as E[X]*E[Y]=0*1 = 0 and hence cov(X,Y)=0, so X and Y are still uncorrelated, i.e. X and Y are still not linearly dependent, even though they are probabilistically dependent. This shows that X and Y being (probabilistically) independent is NOT a necessary condition for X and Y to be uncorrelated (even though it is a sufficient condition) - hence the directionality of the "if" statement in the definition of E[XY] in slide 9 (i.e. "only if" does NOT apply, and so we can't say "iff")

Sample values (Matlab)

$$cov(x,y) = 0.0095$$
$$std(x) = \sqrt{0.9911}$$
$$std(y) = \sqrt{2.0119}$$
$$corr(x,y) = \rho = 0.0067$$

Covariance and correlation → 0
as we get more and more samples…

→ X and Y are not linearly related,
but they are not independent!!

(last time) Difference Between Dependence and Correlation

VERY IMPORTANT:

- If X and Y are independent $\Rightarrow$ X and Y are uncorrelated *(Independence is sufficient for uncorrelatedness)*

- If X and Y are uncorrelated $\nRightarrow$ X and Y are independent *(Independence is NOT NECESSARY for uncorrelatedness)*

Today…

- Mean vectors and covariance matrices
- Multivariate Gaussian PDFs
- Marginals of multivariate Gaussian PDFs
- Linear function with Gaussian noise sampling problem revisited
    - Empirical vs. Theoretical Expected Values

READ SIMON BOOK, CHAPTER 3.1

N-dimensional Random Vectors

- To deal with n-dimensional dynamic state vectors and state space models with noise, need to work with **random vectors**
- Stack continuous scalar random vars on top of each other with a well-defined joint pdf

Given n scalar random vars $X_1, X_2, \cdots, X_n$, define the *random vector*

$$\vec{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \text{ with realizations } \vec{X} = \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ and (joint) pdf } p(\vec{X}) = p(X_1, \cdots, X_n).$$

$\rightarrow$ Define the *mean vector* as

$$\vec{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} = E[\vec{X}] = \int_{-\infty}^{\infty} \vec{x} \cdot p(\vec{x}) d\vec{x}$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot p(x_1, \cdots, x_n) dx_n \cdots dx_1,$$

where *marginal mean* of x_i is $m_i = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i \cdot p(x_1, \cdots, x_n) dx_n \cdots dx_1.$

Covariance Matrices

- Consider n-dimensional set of RVs $X_1, X_2, \ldots, X_n$ with joint pdf $p(X_1, X_2, \ldots, X_n)$
- Can assemble all variances and covariances into the nxn ***covariance matrix***

$$C \overset{\text{def.ned as}}{\triangleq} E[(\vec{x}-\vec{m})(\vec{x}-\vec{m})^T] = \int_{-\infty}^{\infty} (\vec{x}-\vec{m})(\cdots)^T p(\vec{x}) d\vec{x} = E[\vec{x}\vec{x}^T] - \vec{m}\vec{m}^T$$

$$\rightarrow C = \begin{bmatrix} E[(x_1-m_1)^2] & E[(x_1-m_1)(x_2-m_2)] & \cdots & E[(x_1-m_1)(x_n-m_n)] \\ E[(x_2-m_2)(x_1-m_1)] & E[(x_2-m_2)^2] & \cdots & E[(x_2-m_2)(x_n-m_n)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(x_n-m_n)(x_1-m_1)] & E[(x_n-m_n)(x_2-m_2)] & \cdots & E[(x_n-m_n)^2] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{x_1}^2 & c_{12} & \cdots & c_{1n} \\ c_{21} & \sigma_{x_2}^2 & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & \sigma_{x_n}^2 \end{bmatrix}$$

$$\text{where } \sigma_{x_i}^2 = \sigma_i^2 = \text{var}(x_i) = E[(x_i-m_i)^2] \text{ (=scalar)}$$

$$\text{and } c_{ij} = c_{ji} = E[(x_i-m_i)(x_j-m_j)] \text{ (=scalar)}$$

$$= \rho_{ij} \cdot \sigma_{x_i} \cdot \sigma_{x_j} \text{ (=scalar)}$$

(symmetric square positive (semi)definite matrix)

The Multivariate Normal (Gaussian) PDF

- Random vars $X_1, \ldots, X_n$ are said to be ***jointly normal (Gaussian)*** if their joint pdf is

Square of Mahalanobis distance $= \|(\vec{x}-\vec{m})\|^2_{C^{-1}}$
[special case of a quadratic form that has many applications in statistics]

$$p(\vec{X}) = \frac{1}{(2\pi)^{\frac{n}{2}}|C|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\underbrace{\underbrace{(\vec{x}-\vec{m})^T}_{1\times n}\underbrace{C^{-1}}_{n\times n}\underbrace{(\vec{x}-\vec{m})}_{n\times 1}}_{1\times 1}\right\}$$

Scalar output density

$$= \mathcal{N}(\vec{m}, C)$$

from vector $\vec{x}$ & mean vector $\vec{m}$ & covar matrix C!

→ completely defined by $\vec{m} \in \mathbb{R}^{n\times 1}$ [mean vector] & $C \in \mathbb{R}^{n\times n}$ (sym, pos def) [covar. matrix]

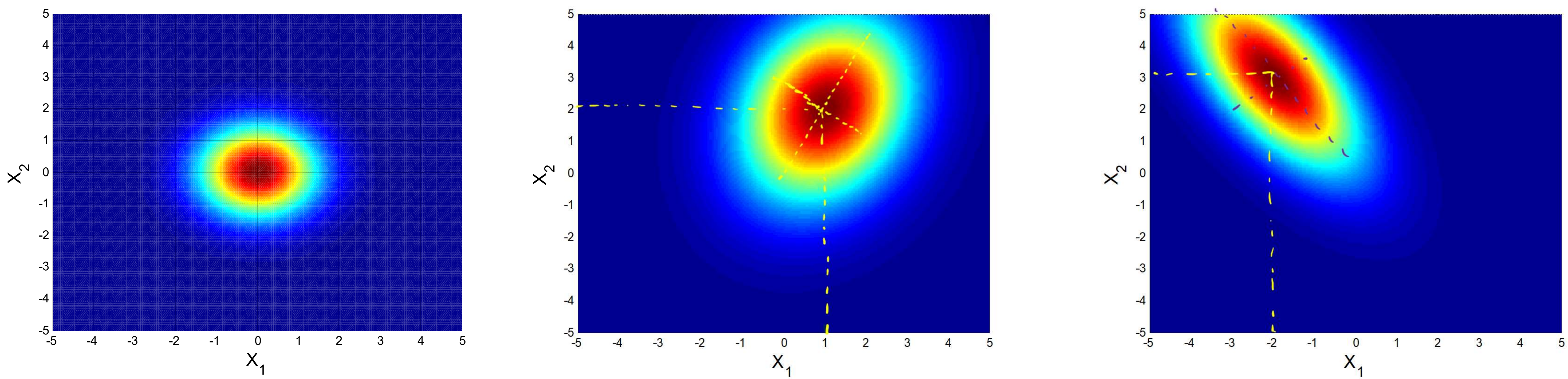
→ If $\vec{x} \sim N(\vec{m}, C)$, then $\vec{x}$ is said to be a Gaussian random vector

→ Matlab: can compute $p(\vec{x})$ for any given $\vec{x}$, $\vec{m}$ & C using "mvnpdf.m"

Example: 2D Multivariate Gaussian PDFs

- Special multivariate expectations: means, covariance, correlation
- Multivariate Normal (Gaussian) PDFs for random vectors

$$p(\vec{X}) = \frac{1}{(2\pi)^{\frac{n}{2}}|C|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{m})^T C^{-1}(\vec{x}-\vec{m})\right\} = \mathcal{N}(\vec{m}, C)$$



$$\vec{m}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ C_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \vec{m}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ C_2 = \begin{bmatrix} 2 & 0.6 \\ 0.6 & 4 \end{bmatrix} \qquad \vec{m}_3 = \begin{bmatrix} -2 \\ 3 \end{bmatrix},\ C_3 = \begin{bmatrix} 2 & -1.8 \\ -1.8 & 4 \end{bmatrix}$$

Looking Inside the Multivariate Gaussian Terms

- Consider the expansion for a general bivariate normal pdf:

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},\ \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},\ \vec{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix},$$

$$C = E[(x-\vec{m})(\ldots)^T] = \begin{bmatrix} E[(x_1-m_1)^2] & E[(x_1-m_1)(x_2-m_2)] \\ E[(x_2-m_2)(x_1-m_1)] & E[(x_2-m_2)^2] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$\rightarrow$ easy to show: $|C| = (1-\rho^2)\sigma_1^2\sigma_2^2$,

$$\text{and } C^{-1} = \begin{bmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} & \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} \\ \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} & \frac{1}{(1-\rho^2)\sigma_2^2} \end{bmatrix}$$

$\rightarrow$ plug $|C|$ and C^{-1} into the definition of $p(\vec{X}) = \mathcal{N}_{\vec{x}}(\vec{m}, C)$ from prev slides:

$\rightarrow x_1 \not\perp x_2$ b/c of coupling when $\rho \neq 0$

$$p(\vec{X}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{ \underbrace{-\frac{1}{2(1-\rho^2)} \cdot \left[\frac{(x_1-m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2} \right]}_{-\frac{1}{2} * \text{ square of Mahalanobis distance for } n=2 \text{ dims}} \right\}$$

Marginal Distributions of Multivariate Gaussian PDFs

Suppose random vecs $\vec{x}_1 \in \mathbb{R}^{n_1}$ and $\vec{x}_2 \in \mathbb{R}^{n_2}$ are *jointly Gaussian*, such that

$$\vec{X} = \begin{bmatrix} \vec{X}_1 \\ \vec{X}_2 \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times 1} \quad \text{and} \quad p(\vec{X}) = p\left(\begin{bmatrix} \vec{X}_1 \\ \vec{X}_2 \end{bmatrix}\right) = \mathcal{N}(\vec{m}, C),$$

where $\vec{m} = \begin{bmatrix} \vec{m}_1 \\ \vec{m}_2 \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times 1}$ and $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$

$\vec{m}_1 \in \mathbb{R}^{n_1},\ \vec{m}_2 \in \mathbb{R}^{n_2}$

$$C_{11} = E[(\vec{x}_1 - \vec{m}_1)(...)^T] \in \mathbb{R}^{n_1\times n_1} \text{ (posdef)}$$
$$C_{22} = E[(\vec{x}_2 - \vec{m}_2)(...)^T] \in \mathbb{R}^{n_2\times n_2} \text{ (posdef)}$$
$$C_{12} = E[(\vec{x}_1 - \vec{m}_1)(\vec{x}_2 - \vec{m}_2)^T] \in \mathbb{R}^{n_1\times n_2}$$
$$C_{21} = C_{12}^T = E[(\vec{x}_2 - \vec{m}_2)(\vec{x}_1 - \vec{m}_1)^T] \in \mathbb{R}^{n_2\times n_1}$$

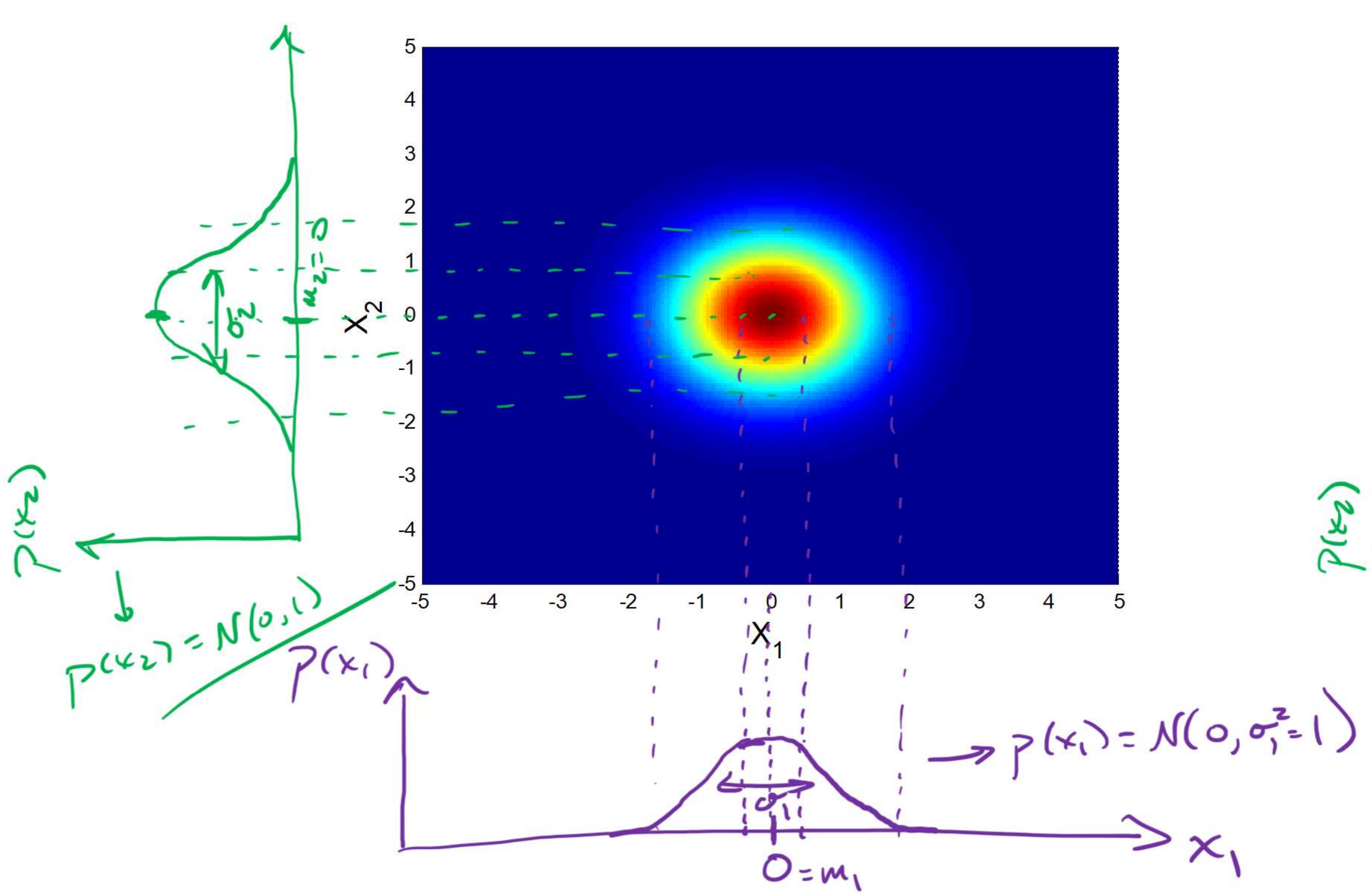
***FACT**: the vector marginal pdfs $p(\vec{X}_1)$ & $p(\vec{X}_2)$ are respectively given by:

$$\underbrace{p(\vec{X}_1) = \int_{-\infty}^{\infty} p(\vec{X})d\vec{X}_2 = \mathcal{N}_{\vec{X}_1}(\vec{m}_1, C_{11})}_{\text{only keep rows of } \vec{m} \text{ and rows/cols of } C \text{ corresp. to } \vec{x}_1!} \qquad \underbrace{p(\vec{X}_2) = \int_{-\infty}^{\infty} p(\vec{X})d\vec{X}_1 = \mathcal{N}_{\vec{X}_2}(\vec{m}_2, C_{22})}_{\text{only keep rows of } \vec{m} \text{ and rows/cols of } C \text{ corresp. to } \vec{x}_2!}$$

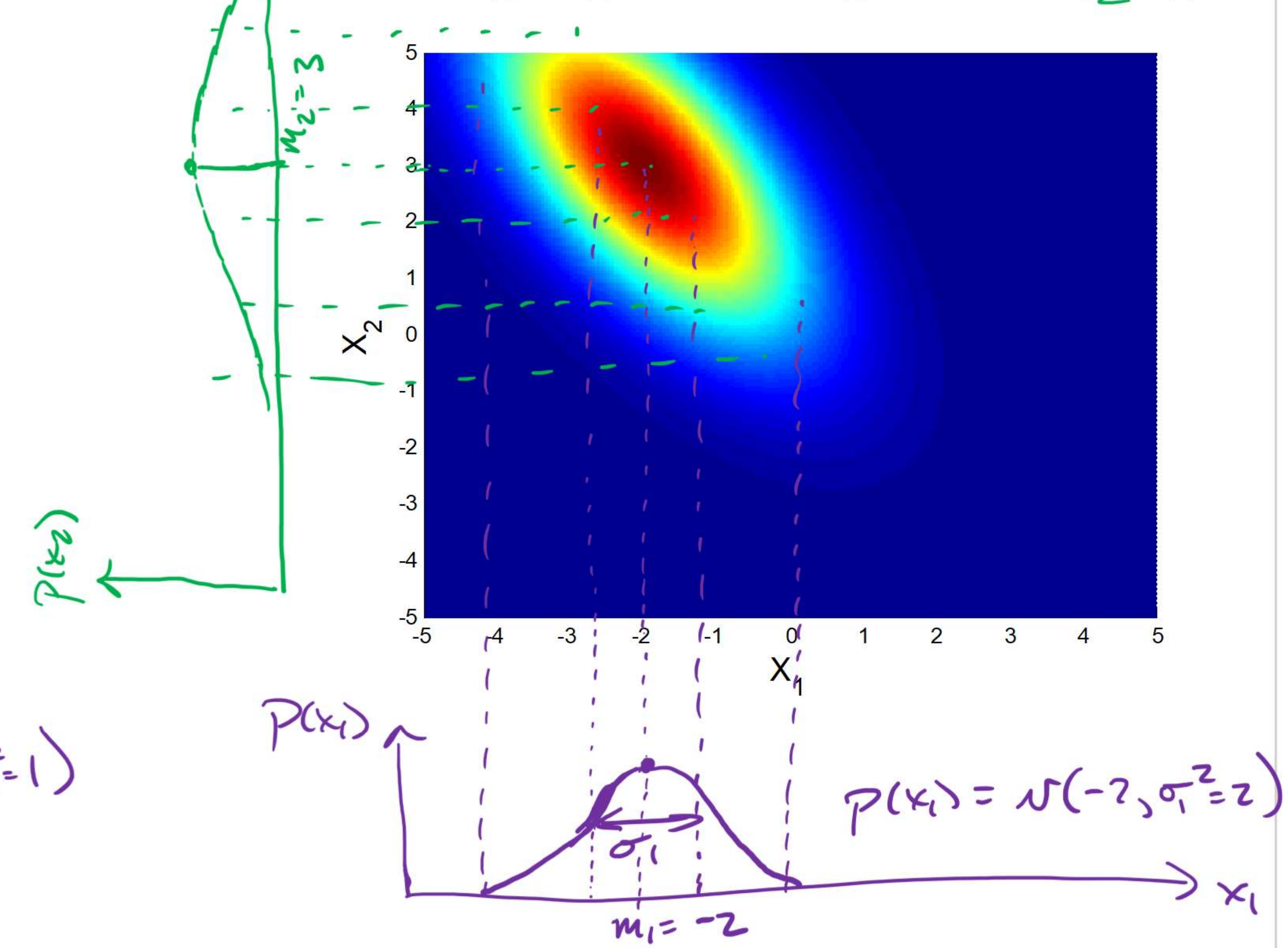
Example: Marginal PDFs of Multivariate Gaussian

- Consider the marginal distributions of two bivariate normal pdfs from before

$$\vec{m}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \; C_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

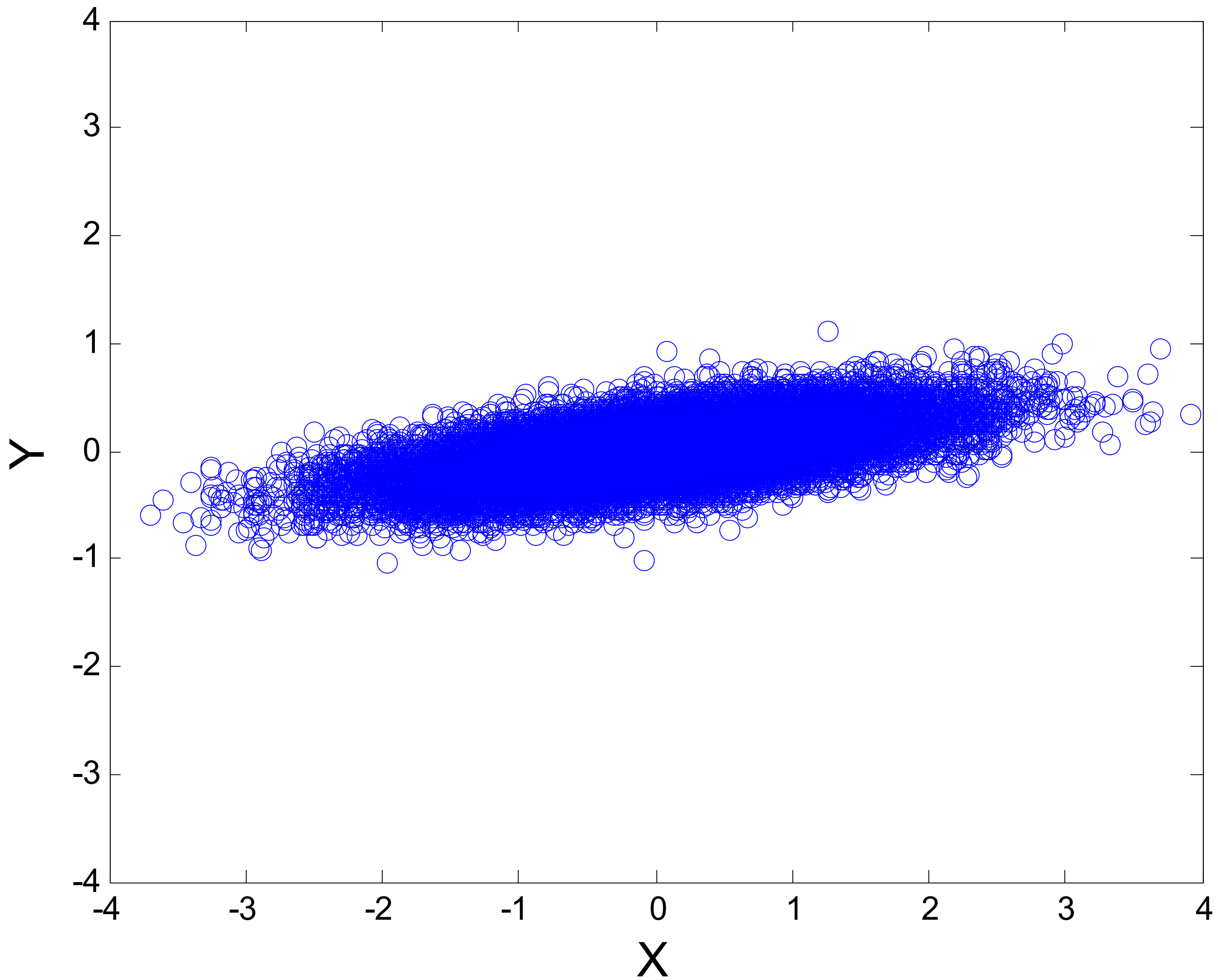


$$\vec{m}_3 = \begin{bmatrix} -2 \\ 3 \end{bmatrix}, \; C_3 = \begin{bmatrix} 2 & -1.8 \\ -1.8 & 4 \end{bmatrix}$$



Revisiting the Gaussian Sampling Experiment

- Consider x ~ N(0,1) with 20,000 samples
- Evaluate y = 0.15*x + 0.2*e, where error e ~ N(0,1)



<u>-Experimental sample expectation values:</u>

cov(x,y) = 0.1509

var(y) = 0.2503

var(x) = 0.9955

ρ(x,y)= 0.6051

-If we are given nothing but p(X) and equation for y in terms of x, how to compute *<u>theoretical expectations</u>*??:

- cov(x,y) = ???
- var(y) = ???
- ρ(x,y)= ???
- mean(y) = ???

Revisiting the Gaussian Sampling Experiment

- Consider scalars $x \sim \mathcal{N}(\bar{x}, \sigma_x^2)$, $e \sim \mathcal{N}(0, \sigma_e^2)$, where x and e are indep., and $y = cx + de$ (c and d = constants)

(i) $m_y = \bar{y} = E[y] = E[cx + de] \underset{\text{linearity of } E}{=} c \cdot E[x] + d \cdot E[e] = c\bar{x} + d \cdot 0 = \boxed{c\bar{x} = \bar{y} = E[y]}$ ⊛

(ii) $cov(x,y) \underset{\text{by def.}}{=} E[(x-\bar{x})(y-\bar{y})] \underset{\text{plug in def. of } y}{=} E[(x-\bar{x})(cx+de-\bar{y})]$

$= E[\, x\{cx+de\} - x\bar{y} - \bar{x}\{cx+de\} + \bar{x}\bar{y}\,]$

$=$ linearity of E & algebra (do in HW 5) $\longrightarrow cov(x,y) = E[xy] - \bar{x}\bar{y}$

where $E[xy] = E[x\{cx+de\}] = E[cx^2 + d \cdot e \cdot x] = c \cdot E[x^2] + \cancel{d \cdot E[e \cdot x]}$

$= E[e] \cdot E[x]$ b/c $x \perp\!\!\!\perp e$

$= 0 \cdot \bar{x} = 0$, $0 = \bar{e}$

$= c \cdot E[x^2]$

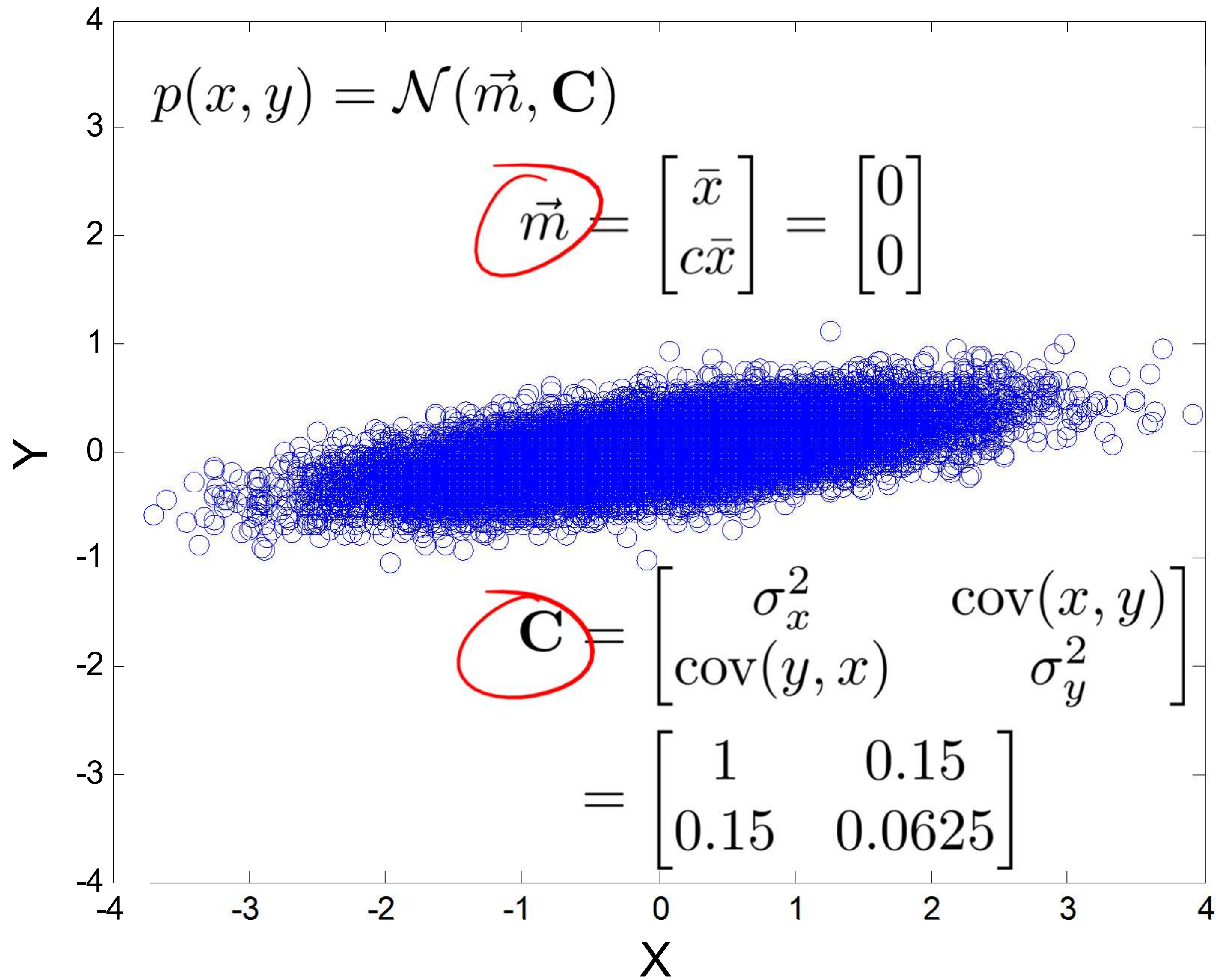
$var(x) = E[x^2] - \bar{x}^2 \rightarrow E[x^2] = var(x) + \bar{x}^2 = \sigma_x^2 + \bar{x}^2$

$= \boxed{c \cdot [\sigma_x^2 + \bar{x}^2] = E[xy]}$ $\Rightarrow \boxed{cov(x,y) = c\sigma_x^2}$ ⊛

Similarly: can show that $\boxed{\sigma_y^2 = var(y) = E[(y-\bar{y})^2] = c^2\sigma_x^2 + d^2\sigma_e^2}$ ⊛

Revisiting the Gaussian Sampling Experiment

- Consider $x \sim N(0,1)$ with 20,000 samples → xbar = 0, var(x) = 1
- Evaluate $y = 0.15*x + 0.2*e$, where error $e \sim N(0,1)$ → c = 0.15, d = 0.2



$$p(x,y) = \mathcal{N}(\vec{m}, \mathbf{C})$$

$$\vec{m} = \begin{bmatrix} \bar{x} \\ c\bar{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \sigma_x^2 & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.15 \\ 0.15 & 0.0625 \end{bmatrix}$$

Experimental sample expectation values:

cov(x,y) = 0.1509

var(y) = 0.0627

var(x) = 0.9955

ρ(x,y)= 0.6051

Exact theoretical expectation values:

cov(x,y) = $c\sigma^2_x$ = 0.15*1 = **0.15**

var(y) = $\sigma^2_y = c^2\sigma^2_x + d^2\sigma^2_e = (0.15)^2 + (0.2)^2 =$ **0.0625**

var(x) = σ^2_x = 1

ρ(x,y)= cov(x,y) /[stdv(x)*stdv(y)]= **0.6**