ASEN 5044, Fall 2024

Statistical Estimation for Dynamical Systems

Lecture 13:
Joint Distributions, Multivariate Expectations, Covariance, Correlation

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

Thursday 10/03/2024

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

Announcements

Midterm 1: out today

- **Week long take home exam posted**
- **Due Thurs 10/10/2024 on Gradescope by 11:59 pm [1 week!]**
- Open book/notes – **must complete by yourself (honor code applies)**
  - **TAs and I will provide answers to clarification questions only**
- **No quiz or homework this week**
- **No office hours next week for TAs or me**
  (email us or privately message us on Piazza if clarification questions)

- **HW #4 DUE TODAY!!!** (will post solutions Sat morning)
- Advanced Topic Lecture #2: to be posted tomorrow

Last Time…

- Sums of independent random variables
  - Central limit theorem

- Gaussian (Normal) random variables and PDFs/distributions

Today…

- Joint, marginal and conditional pdfs for continuous RVs
- Multivariate expected values
  - mean, covariance, correlation
- Random vectors
- Multivariate Normal (Gaussian) PDFs for random vectors

READ SIMON BOOK, CHAPTER 2.7

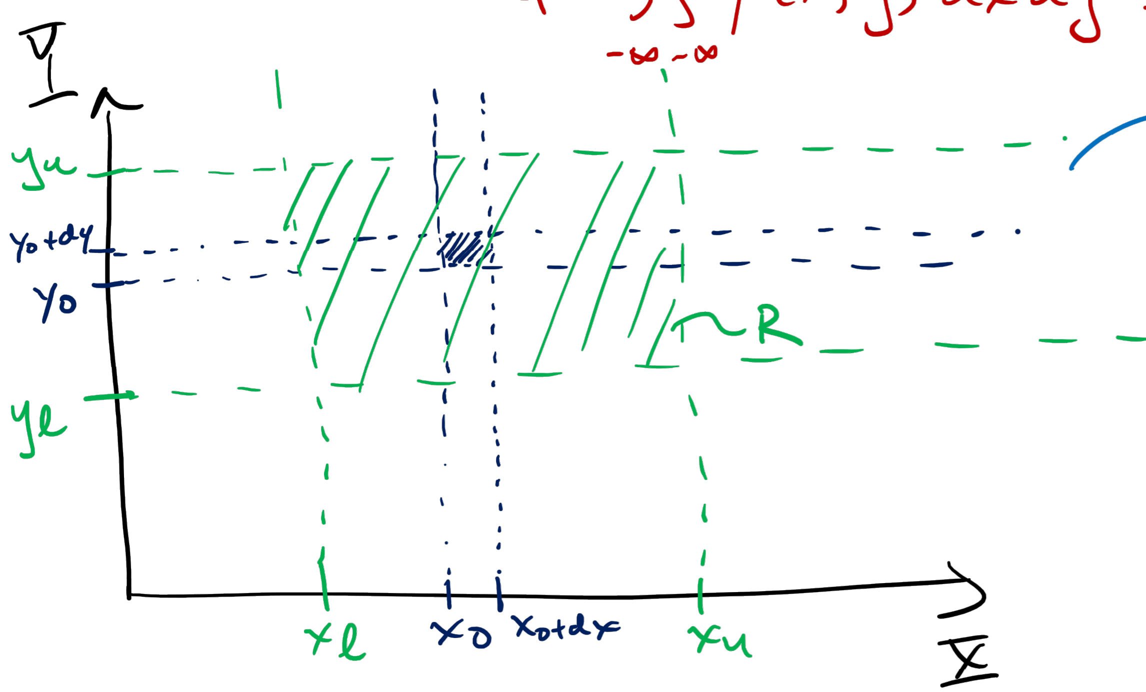
Joint Continuous Random Variables

- How to define joint probabilities for 2 or more continuous RVs?
- Need ***joint probability density functions*** (**joint pdfs**; also called **multivariate pdfs**)

Joint pdf for X & Y : $p_{X,Y}(x,y) = p(x,y) = p(x \,\&\, y)$: takes in 2 real #'s x & y, & returns a scalar density value

such that $P(\{x_0 \leq X \leq x_0 + dx\} \,\&\, \{y_0 \leq Y \leq y_0 + dy\}) = p(x_0, y_0)\, dx\, dy$

& $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\, dx\, dy = 1$



probability that (X,Y) fall into region $R : [x_l, x_u] \times [y_l, y_u]$

$$= P(x \,\&\, y \in R) = \iint_R p(x,y)\, dx\, dy$$

$$= \int_{x_l}^{x_u}\int_{y_l}^{y_u} p(x,y)\, dx\, dy$$

→ Can also define joint cdf!

$$\text{cdf}(x_0, y_0) = c(x_0, y_0) = P(\{x \leq x_0\} \,\&\, \{y \leq y_0\}) = \int_{-\infty}^{x_0}\int_{-\infty}^{y_0} p(x,y)\, dx\, dy$$

Marginal pdfs and Conditional pdfs

- As with joint probability tables discussed earlier, joint pdfs tell "the whole story"
- Analogous expressions exist for marginalization, conditioning, Bayes' rule, independence
- **Basically just need to replace summations with integrals**

*Marginal pdf of x: $\underbrace{p(x)}_{\text{fxn of } x \text{ only}} = \int_{-\infty}^{\infty} p(x,y)dy$ ("averaging" joint pdf p(x,y) w.r.t. y)

*Marginal pdf of y: $\underbrace{p(y)}_{\text{fxn of } y \text{ only}} = \int_{-\infty}^{\infty} p(x,y)dx$ ("averaging" joint pdf p(x,y) w.r.t. x)

*Conditional pdf of x given $y = y^*$: $p(x|\underbrace{y = y^*}_{\text{given}}) \equiv \frac{p(x, y = y*)}{p(y = y^*)} = \frac{p(x, y = y^*)}{\int_{-\infty}^{\infty} p(x, y = y^*)dx}$

by def: $p(y=y^*|x) = \frac{p(x,y=y^*)}{p(x)}$

*Bayes' Rule for pdfs:
(i.e. inversion of conditional pdfs)

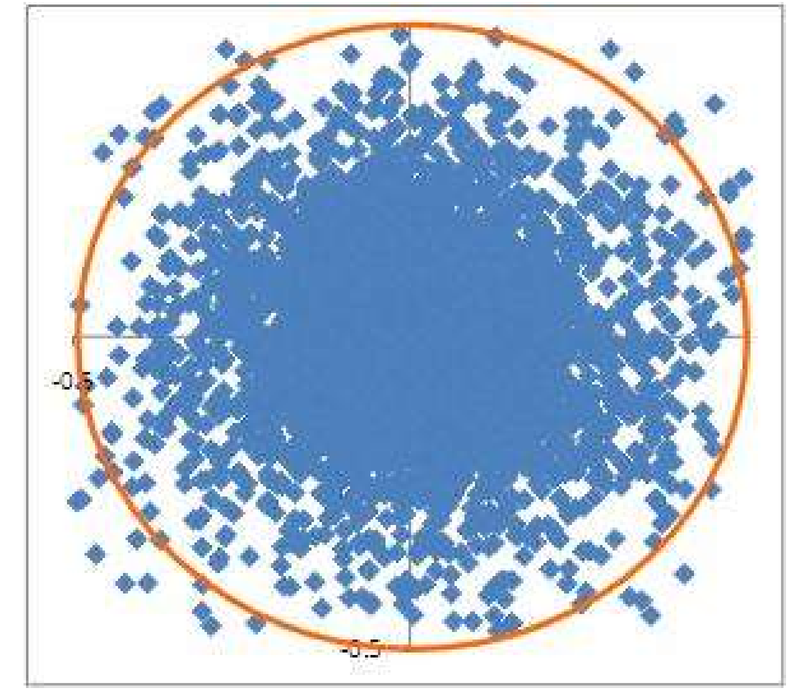
$$p(x|y = y^*) \equiv \frac{p(x)p(y = y*|x)}{p(y = y^*)} = \frac{p(x)p(y = y^*|x)}{\int_{-\infty}^{\infty} p(x)p(y = y^*|x)dx}$$

Independence: continuous RVs X and Y are independent if and only if $p(x,y) = p(x) \cdot p(y) \; \forall x, y$

All of the above readily extend to n-dimensional pdfs $p(x_1, x_2, \cdots, x_n)$

Example: Simple Bivariate Gaussian

- Suppose people throw darts at infinite board with x-y coord system at (0,0)
- Coordinate (x,y) of each dart hole is continuous 2-dim RV



Let's assume X and Y are independent Gaussian RVs with:

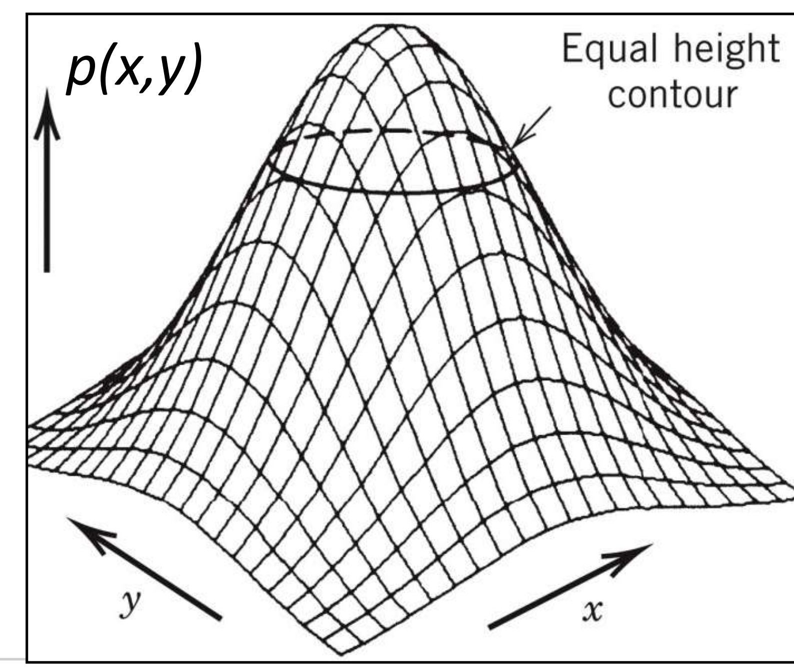
$$\left.\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \\ p(y) &= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left\{-\frac{y^2}{2\sigma^2}\right\} \end{aligned}\right\} \quad \sigma_x = \sigma_y = \sigma$$

$$\rightarrow p(x,y) = p(x) \cdot p(y) \text{ (since } X \text{ and } Y \textit{ assumed} \text{ independent)}$$

$$\rightarrow p(x,y) = \left(\frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left\{-\frac{x^2}{2\sigma^2}\right\}\right)\left(\frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left\{-\frac{y^2}{2\sigma^2}\right\}\right)$$

Returns a single scalar pdf value p(x,y) for any pair of x & y

$$= \frac{1}{2\pi\sigma^2} \cdot \exp\left\{-\frac{(x^2+y^2)}{2\sigma^2}\right\}$$



Expectation Operators and Expected Values for Multivariate pdfs

- Also extend in obvious way – use n-dimensional integrals, etc.
- Key trick is that functions being integrated could take in multiple arguments

Suppose you have RVs $X_1, X_2, \cdots, X_n$ with joint pdf $p(x_1, x_2, \cdots, x_n)$, then

$$E[g(x_1, x_2, \cdots, x_n)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n)p(x_1, x_2, \cdots, x_n)dx_n \cdots dx_2 dx_1$$

$$= \text{some (scalar) value (if } g(x_1, x_2, ..., x_n) \text{ maps to scalar outcome value)}$$

$\rightarrow$ Can also take expectation w.r.t. only *some* subset of the n RVs, e.g.

$$E[g(x_1, x_2, \cdots, x_n)]_{(x_2, x_3, \cdots, x_n)} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n)p(x_1, x_2, \cdots, x_n)dx_n \cdots dx_3 dx_2$$

$$= \text{some function of } x_1 \textit{ only } = \text{marginal expectation of } g(\cdot) \text{ w.r.t. } x_2, \cdots, x_n$$

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

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

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

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

Foil out:

$$E[xY - xm_y - m_x y + m_x m_y] = E[xY] - E[x]m_y - m_x E[y] + m_x m_y = E[xY] - m_x m_y$$

- Correlation coefficient: covariance normalized by product of standard devs

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

“normalized covariance of x & y”

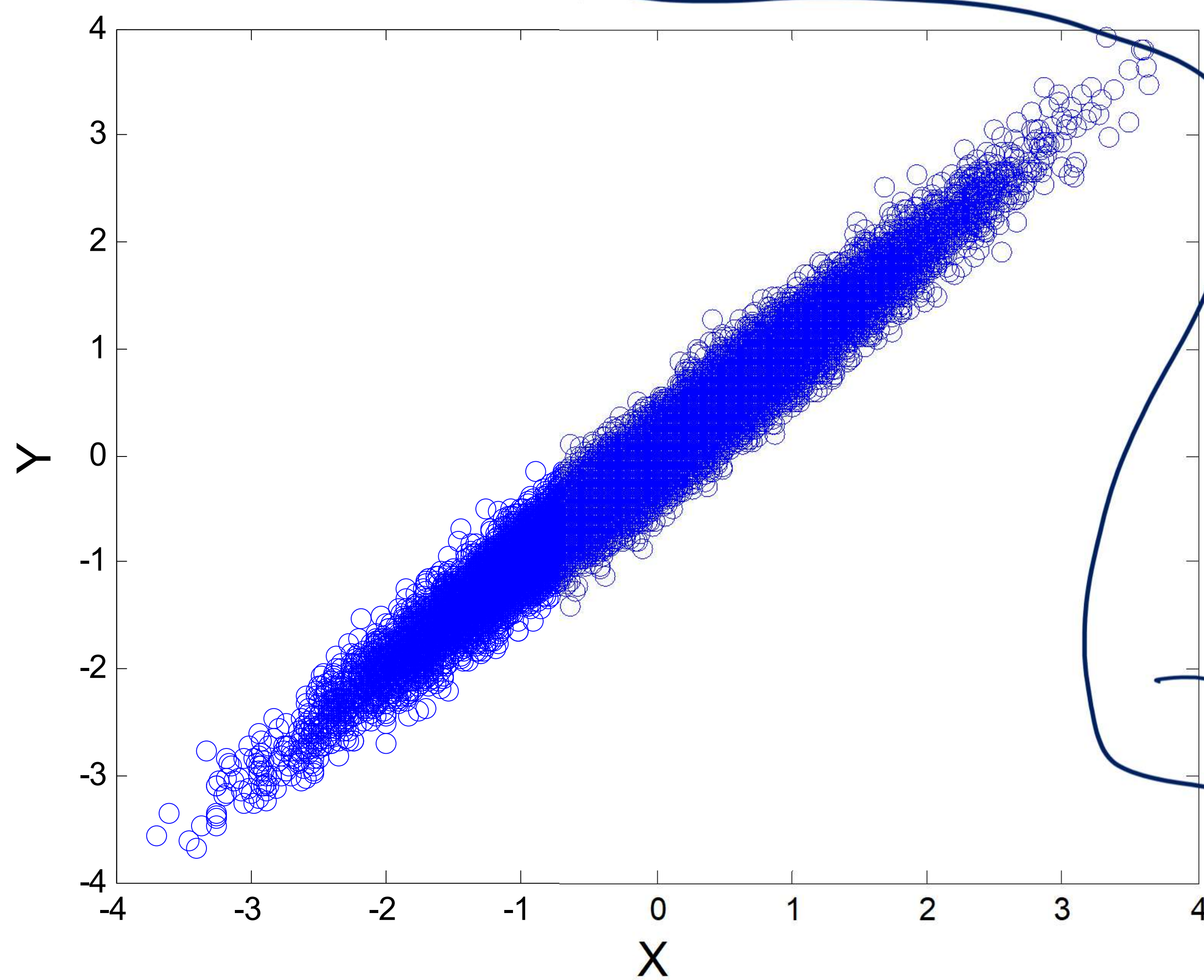
where $\rho \in [-1, 1]$

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

$\mathrm{var}(y) = E[(y-m_y)^2]$

Example: Sample Covariances and Correlations

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



Sample expected values (Matlab)

$cov(x,y) = 0.9916$

$std(x) = \sqrt{0.9911} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

$x_i \sim$ sampled from $N(0,1)$

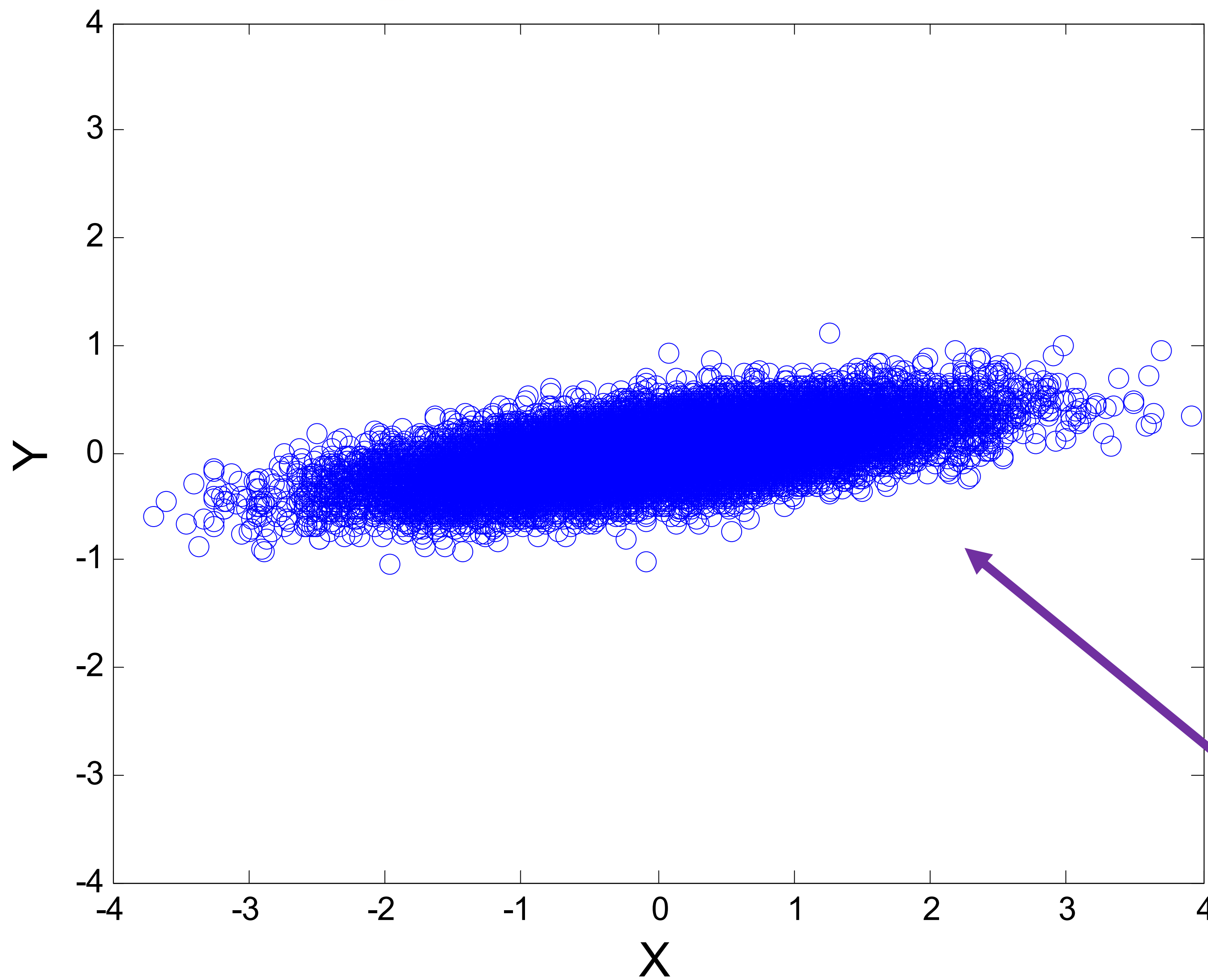
$std(y) = \sqrt{1.329}$

$corr(x,y) = \rho = \frac{0.9916}{\sqrt{0.9911}\sqrt{1.329}} = 0.9801$ $(N = 20{,}000)$

(as we will show later (next lecture): this plot is same as plotting samples from joint pdf of $p(x,y)$ since via chain rule: $p(x,y) = \underbrace{p(x)}_{\text{sample } x_i \text{ first}} \cdot \underbrace{p(y|x)}_{\text{sample } y_i \text{ given } x_i \text{ via}}$

Example: Sample Covariances and Correlations

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



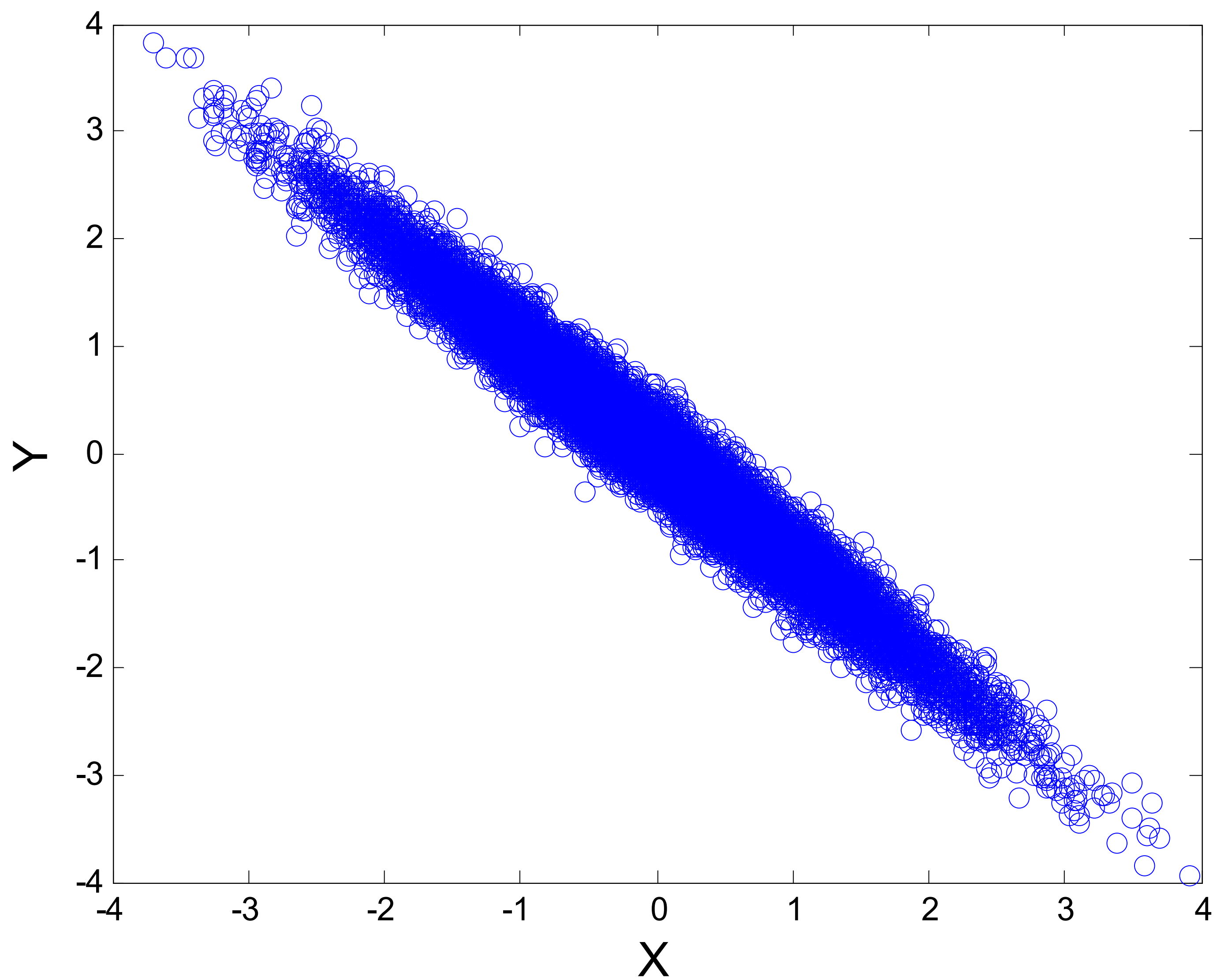
$$\text{Sample values (Matlab)}$$
$$corr(x,y) = \rho = 0.6051$$

Not a perfect fit between X and Y, but definitely some information about X from Y (and vice-versa)

[like samples from some joint pdf p(x,y) although we don't actually know what that joint p(x,y) is yet – all we have/know right now is
p(x)~N(0,1),
p(e)~ N(0,1),
and y = ax + de]

Example: Sample Covariances and Correlations

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



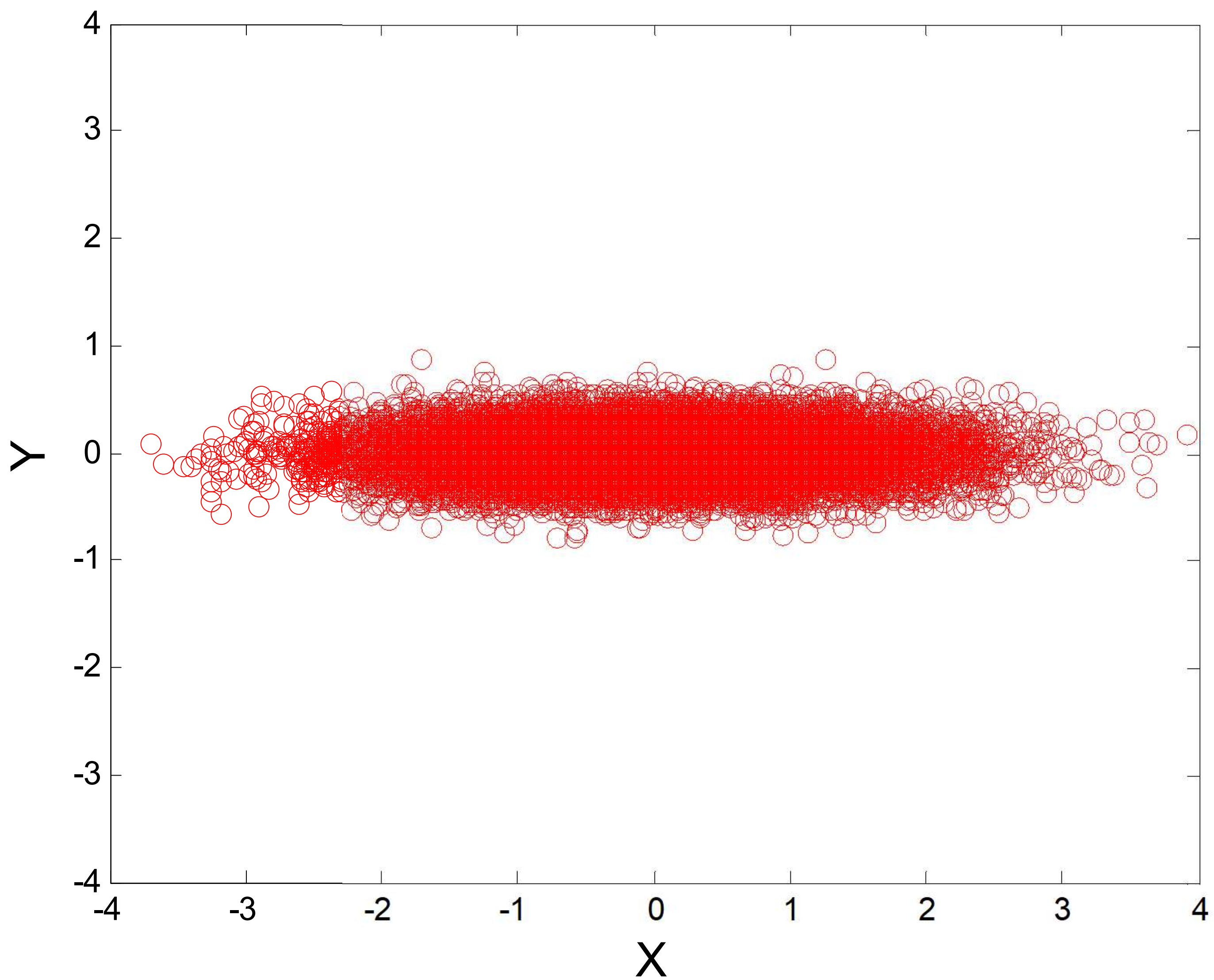
Sample values (Matlab)

$cov(x,y) = -0.9899$

$corr(x,y) = \rho = -0.9807$

Example: Sample Covariances and Correlations

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

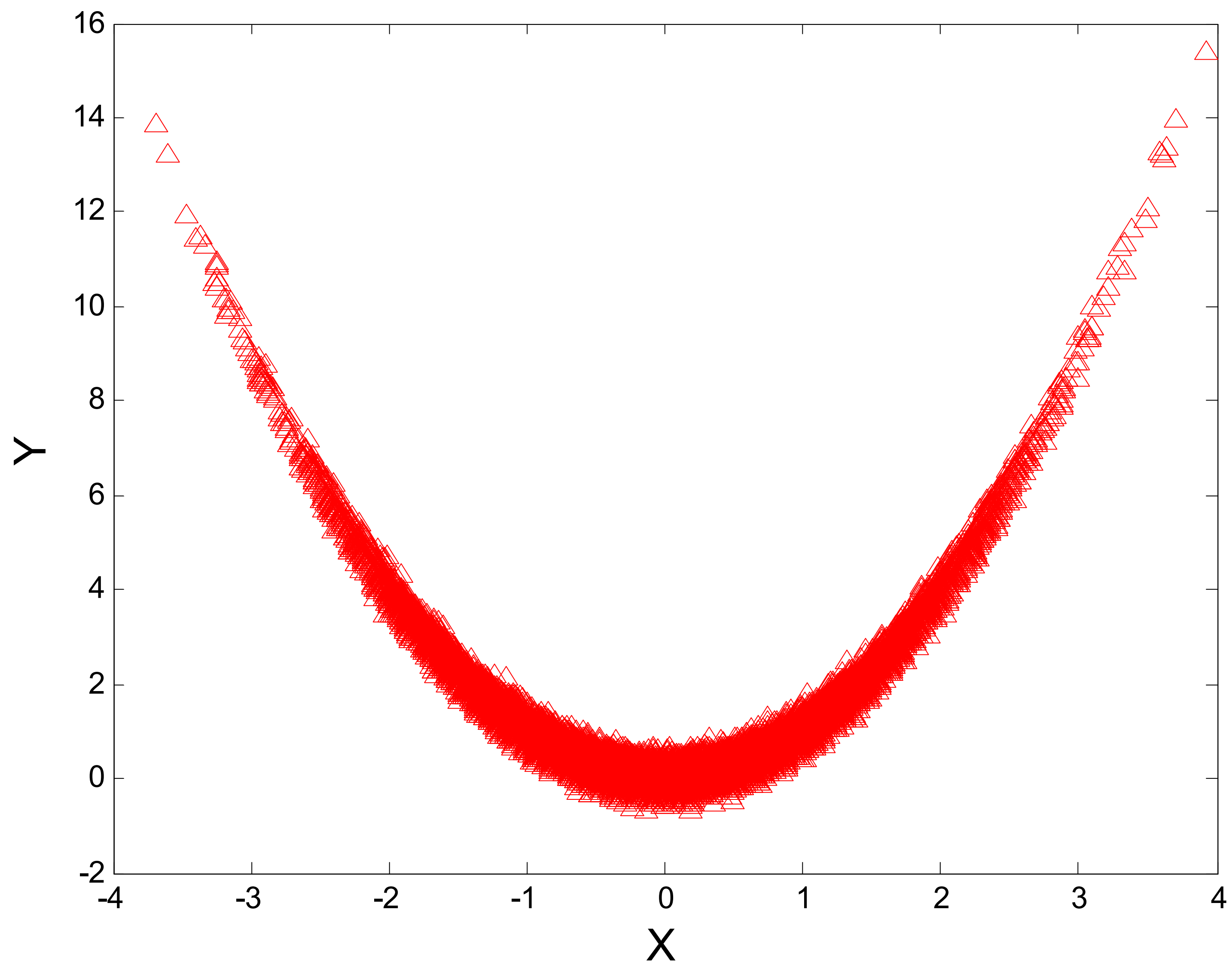
Cov & corr → 0 as N → ∞

→ No linear info b/w X & Y

→ $X \perp\!\!\!\perp Y$

Example: Sample Covariances and Correlations

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

$$\begin{aligned}
&\text{Sample values (Matlab)}\\
&cov(x,y) = 0.0095\\
&std(x) = \sqrt{0.9911}\\
&std(y) = \sqrt{2.0119}\\
&corr(x,y) = \rho = 0.0067
\end{aligned}$$

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

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

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 $\not\Rightarrow$ X and Y are independent

(Independence is NOT NECESSARY for uncorrelatedness)