MATH 180A: INTRO TO PROBABILITY
(FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: §6.2-6.3

Next: §8.1-8.3

Midterm 2: Wednesday 8pm ← CENTR 101!

- covering Chapters 3-5. ~~Poisson Process~~
- seat assignment on TritonEd
- bring student ID
- bring 1 sheet (double-sided) hand-written notes.
- bring your A GAME!

Jointly Continuous Random Vectors

A random vector $\underline{X} = (X_1, \dots, X_n)$ has a pdf $f_{\underline{X}} : \mathbb{R}^n \to \mathbb{R}_+$ if, for "nice" subsets $B \subseteq \mathbb{R}^n$

$$\mathbb{P}(\underline{X} \in B) = \int_B f_{\underline{X}}(x_1, \dots, x_n)\, dx_1 \cdots dx_n .$$

(we say $X_1, X_2, \dots, X_n$ are jointly continuous.)

Properties: $f_X(\underline{x}) \geq 0$ for all $\underline{x} \in \mathbb{R}^n$

$$\int_{\mathbb{R}^n} f(\underline{x})\, d\underline{x} = 1 .$$

$\|\underline{x}\|^2 = x_1^2 + \cdots + x_n^2$

Eg. Standard Multivariate Normal $\quad f(\underline{x}) = (2\pi)^{-\frac{n}{2}} e^{-\|\underline{x}\|^2/2}$

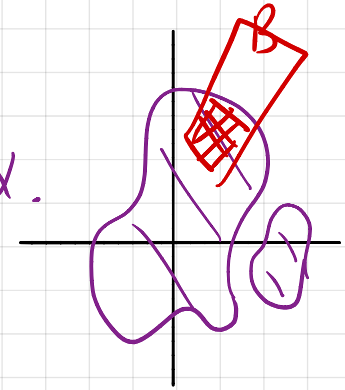
(n=2) $\quad f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} > 0$

$$\therefore \iint_{\mathbb{R}^2} f(x,y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2}\, dx\, dy = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-y^2/2}\Big(\underbrace{\int_{-\infty}^{\infty} e^{-x^2/2}\, dx}_{\sqrt{2\pi}}\Big) dy = \frac{(\sqrt{2\pi})^2}{2\pi} = 1 .$$

$\sqrt{2\pi}$

E.g. Uniform Probability

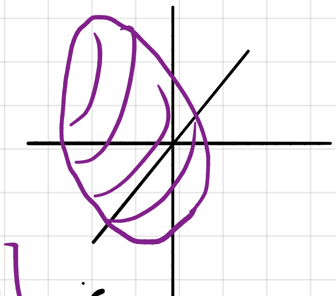
Let A be a bounded region in $\mathbb{R}^2$ with area α.



$$\mathbb{P}((X,Y) \in B) = \frac{\text{Area}(B \cap A)}{\text{Area}(A)}$$

Let V be a bounded region in $\mathbb{R}^3$ with volume ϑ

$\vdots$



A random vector $\underline{X} = (X_1, X_2)$ [resp. $\underline{X} = (X_1, X_2, X_3)$] is uniformly distributed in A [resp. V] if it is jointly continuous and has density

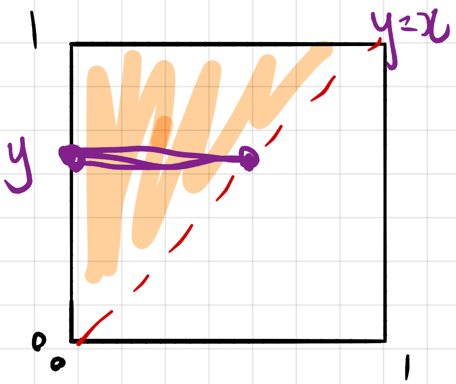
$$f_A(x_1,x_2) = \frac{1}{\alpha}\mathbb{1}((x_1,x_2)\in A) \quad \text{resp.} \quad f_V(x_1,x_2,x_3) = \frac{1}{\vartheta}\mathbb{1}((x_1,x_2,x_3)\in V)$$

$$\begin{cases} = 1 & \text{if } (x_1,x_2)\in A \\ = 0 & \text{if } (x_1,x_2)\notin A \end{cases}$$

$$\mathbb{P}((X,Y)\in B) := \iint_B f_A(x_1,x_2)\,dx_1dx_2 = \iint_B \frac{1}{\alpha}\mathbb{1}\{(x_1,x_2)\in A\}\,dx_1dx_2 = \iint_{A\cap B}\frac{1}{\alpha}\,dx_1dx_2$$

$$\iint_{A\cap B}\frac{1}{\alpha}\,dx_1dx_2 = \text{Area}(A\cap B)/\alpha.$$

E.g. Suppose (X, Y) has joint density $f(x,y) = \frac{3}{2}(xy^2+y)\,\mathbb{1}(0 \le x,y \le 1)$



$$\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = \int_0^1\int_0^1 \frac{3}{2}(xy^2+y)\,dx\,dy$$

$$= \frac{3}{2}\int_0^1 dy \left(y^2\cdot\frac{x^2}{2} + yx\right)\Big|_{x=0}^{1} = \frac{3}{2}\int_0^1 \left(\frac{1}{2}y^2 + y\right) dy$$

$$= \frac{3}{2}\left(\frac{1}{6}y^3 + \frac{1}{2}y^2\right)\Big|_{y=0}^{y=1}$$

$$= \frac{3}{2}\left(\frac{1}{6}+\frac{1}{2}\right) = 1.$$

Compute $\mathbb{P}(X < Y)$.

$$T = \{(x,y) : x < y\}$$

$$= \mathbb{P}((X,Y) \in T) = \iint_T \frac{3}{2}(xy^2+y)\,dx\,dy$$

$$= \frac{3}{2}\int_0^1 dy \int_0^y dx\,(xy^2+y) = \frac{3}{2}\int_0^1 dy \left(y^2\cdot\frac{x^2}{2} + yx\right)\Big|_{x=0}^{y}$$

$$= \frac{3}{2}\int_0^1 dy \left(y^2\cdot\frac{y^2}{2} + y^2\right)$$

$$= \frac{3}{2}\left(\frac{1}{10}y^5 + \frac{1}{3}y^3\right)\Big|_0^1 = \frac{3}{2}\left(\frac{1}{10}+\frac{1}{3}\right)$$

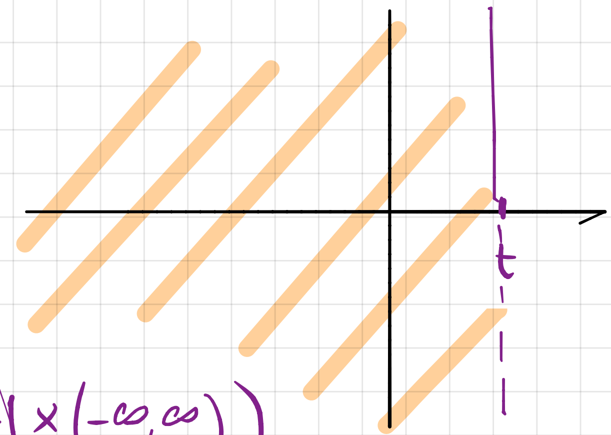
$$= 13/20.$$

Marginals

Let $\underline{X} = (X_1, X_2, \dots, X_n)$ be a jointly continuous random vector, with joint density $f_{\underline{X}}$. The density of X_j is

$$f_{X_j}(t) = \int_{\mathbb{R}^{n-1}} f_{\underline{X}}(x_1, x_2, \dots, x_{j-1}, t, x_{j+1}, \dots, x_n)\, dx_1 \cdots dx_{j-1}\, dx_{j+1} \cdots dx_n .$$

(Integrate out all but the j^{th} variable)



Proof $(n=2, j=1)$ $\quad \underline{X} = (X, Y)$

$$\mathbb{P}(X \le t) = \mathbb{P}(X \le t, Y \in \mathbb{R}) = \mathbb{P}\big((X,Y) \in (-\infty, t] \times (-\infty, \infty)\big)$$

$$F_X(t) = \int_{-\infty}^{t} dx \int_{-\infty}^{\infty} dy\, f_{(X,Y)}(x,y) = \int_{-\infty}^{t} g(x)\,dx \qquad g(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)\,dy$$

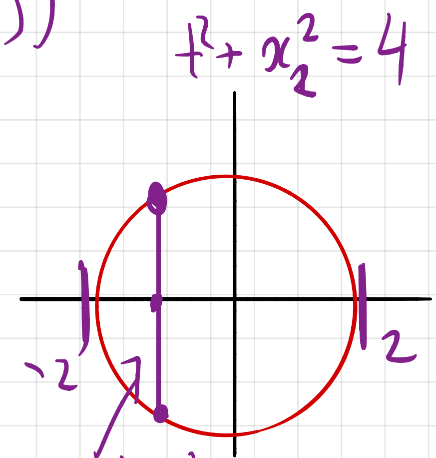
$$\therefore \; f_X(t) = \frac{d}{dt} F_X(t) = \frac{d}{dt} \int^{t} g(x)\,dx = g(t)$$

FTC

Eg. Suppose $\underline{X}$ is uniformly distributed on the disk of radius 2.
Find the marginal density of X_1 $\quad (\underline{X}=(X_1,X_2))$

$$f_{\underline{X}}(x_1,x_2)=\frac{1}{4\pi}\mathbb{1}_{\mathbb{D}_2}(x_1,x_2)$$

$$f_{X_1}(t)=\int_{-\infty}^{\infty} f_{\underline{X}}(t,x_2)\,dx_2=\frac{1}{4\pi}\int_{-\infty}^{\infty}\underbrace{\mathbb{1}_{\mathbb{D}_2}(t,x_2)}\,dx_2$$



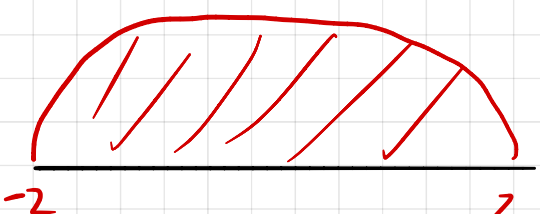
Constant,

$-\sqrt{4-t^2}\le x_2\le\sqrt{4-t^2}$

$-2\le t\le 2$

$$=\frac{1}{4\pi}\int_{-\sqrt{4-t^2}}^{\sqrt{4-t^2}}dx_2=\frac{1}{4\pi}\left(2\sqrt{4-t^2}\right)$$

$$=\begin{cases}\dfrac{\sqrt{4-t^2}}{2\pi} & -2\le t\le 2\\[2mm] 0 & |t|>2\end{cases}$$



"Semicircle Law"

$\mathbb{E}$xpectation

If $\underline{X}$ is a random vector in $\mathbb{R}^n$ with joint density $f_{\underline{X}}: \mathbb{R}^n \to \mathbb{R}_+$, and $g: \mathbb{R}^n \to \mathbb{R}$, then

$$\mathbb{E}(g(\underline{X})) = \int_{\mathbb{R}^n} g(\underline{x}) f_{\underline{X}}(\underline{x})\, d\underline{x}$$

Eg. $\underline{X} = (X, Y)$, $f_{\underline{X}}(x,y) = \frac{3}{2}(xy^2 + y)\, \mathbb{1}(0 \le x, y \le 1)$. Find $\mathbb{E}(X^2 Y)$.

$$\mathbb{E}(X^2 Y) = \iint_{\mathbb{R}^2} x^2 y\, f_{\underline{X}}(x,y)\, dx\, dy = \int_0^1 \int_0^1 x^2 y \cdot \frac{3}{2}(xy^2 + y)\, dx\, dy$$

Exercise

$$= \frac{25}{96}.$$

CAUTION !

If $X_1, X_2, \ldots, X_n$ are discrete random variables, then

$\underline{X} = (X_1, \ldots, X_n)$ is a (jointly) discrete random vector.

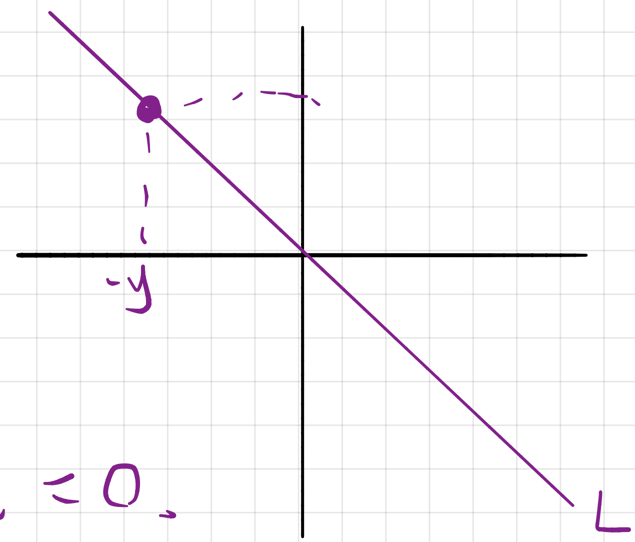
BUT

Just because $X_1, \ldots, X_n$ are (separately) continuous random variables does <u>not</u> necessarily imply that $\underline{X} = (X_1, \ldots, X_n)$ has a joint density!

E.g. $X \sim N(0,1)$, $Y = -X \sim N(0,1)$

$$\mathbb{P}(Y \le t) = \mathbb{P}(-X \le t) = \mathbb{P}(X \ge -t) = 1 - \mathbb{P}(X < -t) = 1 - \Phi(-t) = \Phi(t)$$

Must have $\mathbb{P}(X = -Y) = 1$.



If $f_{(X,Y)}$ existed,

$$\mathbb{P}(X = -Y) = \iint_L f_{(X,Y)}(x,y)\,dx\,dy$$

$$\int_{-\infty}^{\infty} dy \underbrace{\int_{-y}^{-y} f_{(X,Y)}(x,y)\,dx}_{=0} = 0.$$

Joint Distributions & Independence

6.3

Suppose $\underline{X} = (X_1, \dots, X_n)$ is jointly continuous. Then

$X_1, \dots, X_n$ are independent iff $f_{\underline{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$.

($n=2$: $\quad f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$.)

Pf. ($\Longleftarrow$) $\quad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}\big((X,Y) \in A \times B\big)$

$$= \iint_{A \times B} f_X(x) f_Y(y)\,dx\,dy$$

$$= \int_A f_X(x)\,dx \int_B f_Y(y)\,dy = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B)$$

$\therefore$ X, Y independent.

($\Longrightarrow$) NEXT TIME.