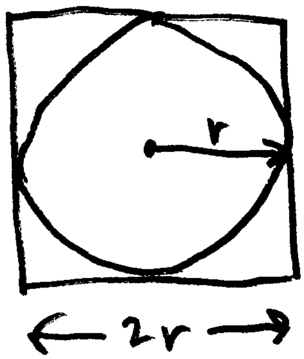
this time: joint, marginal,
next time: conditional PDFs

Case Study: physical estimation of π

STAT 131
5 May 20
(Lecture)

You have a circular wooden plate of radius r and a square wooden plate of dimension $(2r)$ on each side, rigged with electronic sensors that can record raindrop impacts (!) (Arduino). Put this apparatus out in a rainstorm and count: ①



$$P\left(\text{raindrop falls randomly inside circle} \,\middle|\, \text{raindrop falls randomly inside square} \right)$$

$$= \frac{\pi r^2}{(2r)^2} = \frac{\pi}{4} = \theta$$

$\theta \approx 0.79$

Let $X_i = \begin{cases} 1 & \text{if raindrop } i \text{ falls inside circle} \\ 0 & \text{else}^{(*)} \end{cases}$

Then physical context implies the probability model $\boxed{(X_i \mid \theta) \overset{\text{IID}}{\sim} \text{Bernoulli}(\theta), \; i = 1, \dots, n}$

(*) if raindrop i falls inside square but outside circle

Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ = proportion of raindrops inside circle. ②

This is an estimate of $\theta = \frac{\pi}{4}$; set $\bar{X}_n = \hat{\theta}$

Define $I_n = 4 \bar{X}_n = 4 \hat{\theta}$ | I_n is an estimate of π

$\boxed{Q:}$ How big does n need to be so that I_n gets π right to k significant digits with high probability? $\boxed{A:}$ We want n such that $P_F (| I_n - \pi | \le \epsilon_1) \ge 1 - \epsilon_2$ where ϵ_1 is a small positive error on the π scale and ϵ_2 is a small positive error on the probability scale.

The Arduino guy got $I_n = 3.1352$ with $n = 2{,}000$;

3.14159

his error $\epsilon_1 \doteq 0.0063$ / or was this to be expected? was he lucky?

later in the course we'll see that his approach was mathematically valid: he was using an important result called the Weak Law of Large Numbers (WLLN) (first proven by James Bernoulli in 1713, first stated without proof by Gerolamo Cardano in the late 1500s); the WLLN in this situation says that $\bar{X}_n \xrightarrow{P} \theta$, which means that $\bar{X}_n$ will get closer and closer to θ as $n \uparrow$ with probability approaching 1. However, he was extremely lucky to get within 0.0063 of π with only 2,000 raindrops; using a calculation that we'll see how to do later, to have at least 99% probability of getting π right to within 0.0063 he would have needed about 451,000 raindrops. See the R code on the course web page for a demo.

~~population~~ all possible raindrops inside circle?

~~sample~~ the observed raindrops ($1 = Y$, $0 = N$) inside circle?

~~repeated sampling~~ dataset possible values of Π_n

1_s & 0_s

mean $\theta = \frac{\pi}{4}$

IID

X_1 $x_1 = 1$

X_2 $x_2 = 0$

$\vdots$

X_n $x_n = 1$

$n = 2000$

mean $\bar{X}_n = ?$

3.18

3.09

$\vdots$

$M =$ big

$4\bar{X}_n = \Pi_n = ? = 4\hat{\theta}$

long run mean expected value of $\Pi_n = \pi$

long run SD standard error of Π_n.

Monte Carlo standard error as est. of π

IID

$n = 2000$

$\bar{X}_n = ?$

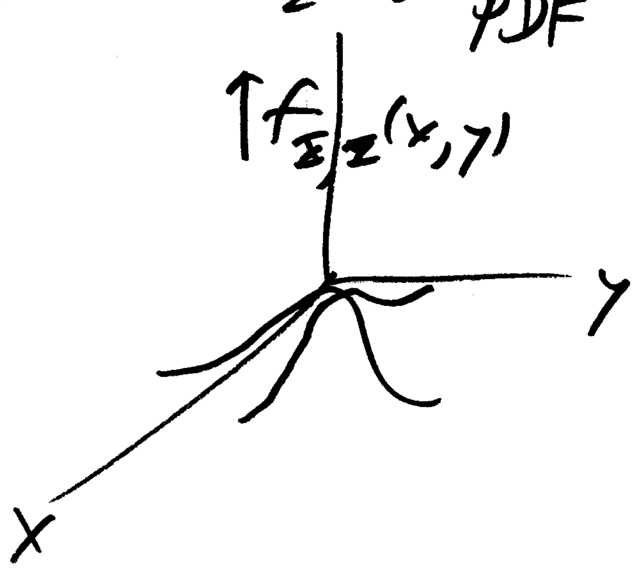
$\Pi_n = ?$

SD = standard deviation = expected amount by which Π_n will differ from π

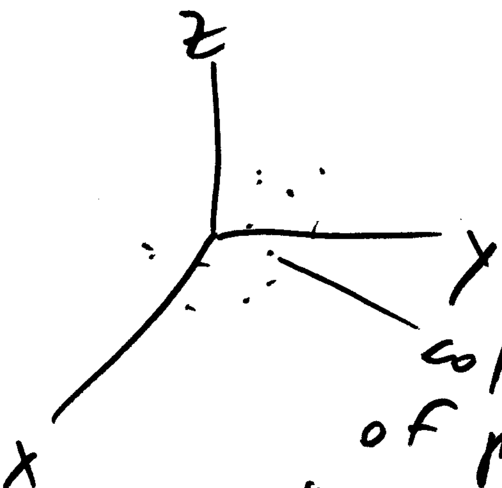
1-dimension

$f_X(x)$ PDF

2-D PDF



3-D PDF



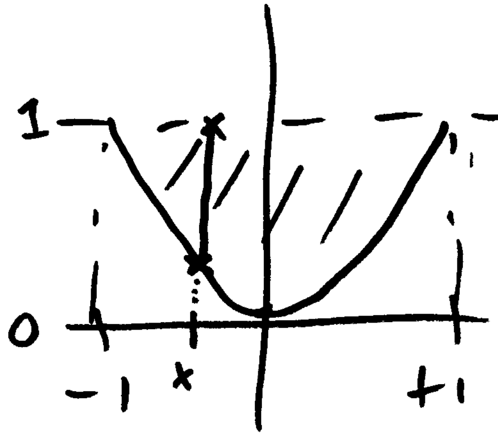
color of point proportional to $f_{X,Y,Z}(x,y,z)$

4+ r.v.

hard to visualize

$f_{X,Y,Z,W}(x,y,z,w)$

can learn something from $f_X(x)$, $f_Y(y)$, $f_Z(z)$, $f_W(w)$



$$f_{X,Y}(x,y) = \begin{cases} \frac{21}{4} x^2 y & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

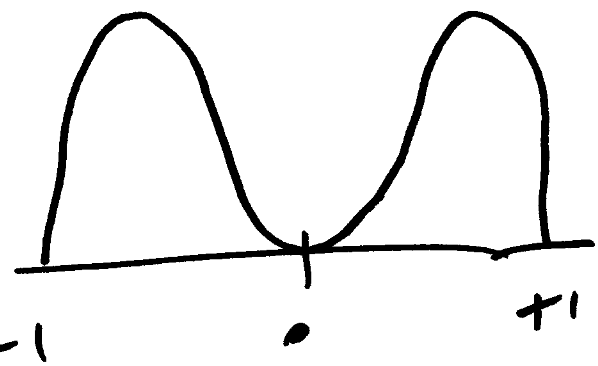
---

to extract the marginal for X, $f_X(x)$, first you pick an x (fix x) in $S_X = [-1, +1]$ and calculate

$$f_X(x) = \int_{S_{X,Y}} f_{X,Y}(x,y)\, dy$$

$$= \int_{x^2}^{1} \frac{21}{4} x^2 y \, dy$$

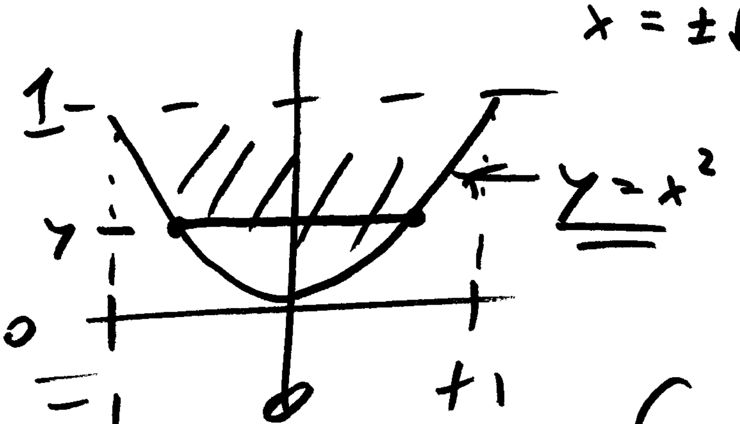
$$= \begin{cases} \frac{21}{8} x^2 (1 - x^4) & \text{for } -1 \le x \le 1 \\ 0 & \text{else} \end{cases}$$



$f_X(x)$

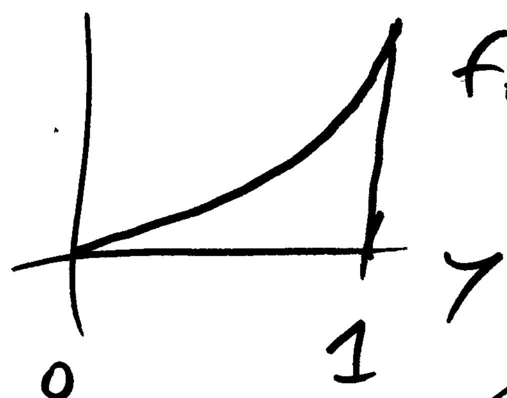
$f_Y(y) = \begin{cases} \frac{7}{2} y^{5/2} & \text{for } 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$

$x = \pm\sqrt{y}$

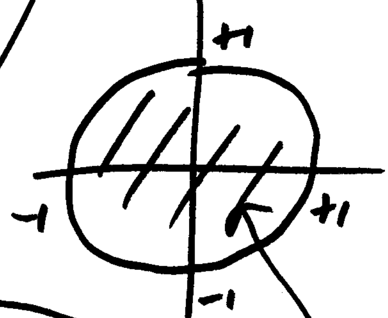


fix a $y \in [0, 1]$

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y \, dx$$



$f_Y(y)$



$S_X = [-1, +1]$

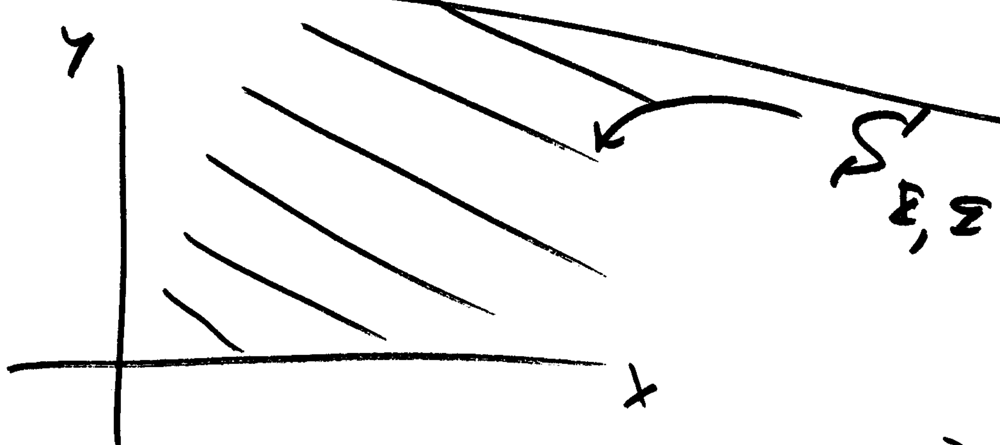
$S_Y = [-1, +1]$

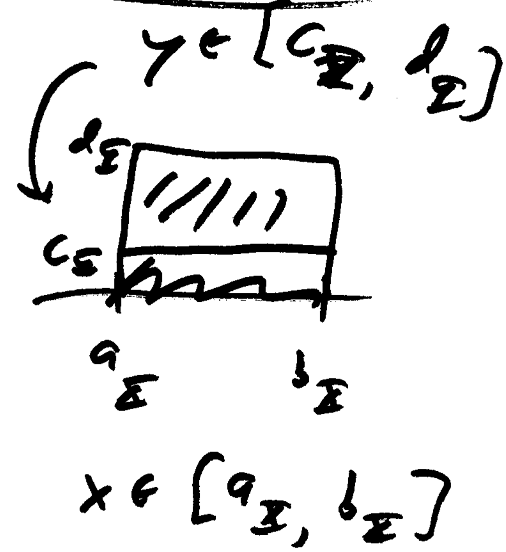
$S_{X,Y} = \{(x, y) : 0 \le x^2 + y^2 \le 1\}$

to see if (x, y) is in $S_{X,Y}$, it's not enough to check that

$x \in S_X$ and $y \in S_Y$; you need to directly see if $(x,y) \in S'_{X,Y}$; so X and Y are **dependent** in this joint PDF

$$f_{X,Y}(x,y) = \begin{cases} k\, e^{-(x+2y)} & \text{for } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{else} \end{cases}$$





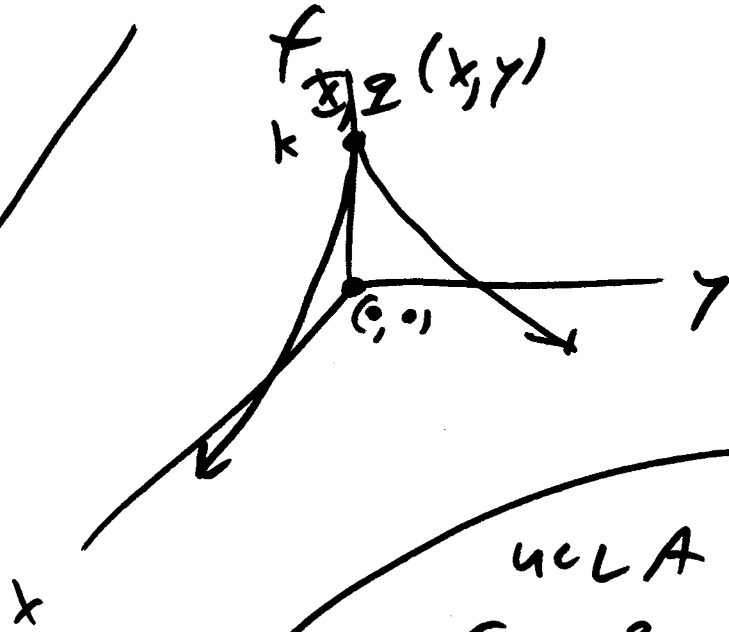
generalized rectangle:

$(a, \infty) \times (b, \infty)$ or

$(a, \infty) \times (-\infty, b)$ or

$(-\infty, a) \times (b, \infty)$ or

$(-\infty, a) \times (-\infty, b)$



$$f_{X,Y}(x,y) = \begin{cases} k e^{-(x+2y)} & \text{for } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{else} \end{cases}$$

UCLA survey: relationship between gender & marijuana legalization preference (MLP)

$\binom{Y}{N}$ MLP | gender $\binom{F}{M}$

$n = 104$