Kom 505E WEEK 6

Lecture Notes

01. 11. 2016

G.Ü.

Basics of Monte Carlo Sampling:

When a sum or an integral cannot be computed exactly, when no exact simplification is known, it is often possible to approximate it w/ Monte Carlo sampling.

The idea: View the sum or integral as if it was an expectation under some distribution; and approximate the expectation by a corresponding average.

Let $\underline{S} = \sum_x p(x) f(x)$

$\underline{S} = \int_{-\infty}^{\infty} f(x) p(x) dx$

$\boxed{= E_{p_x}[f(x)]}$

We can approx S by drawing N samples $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ from p, then form the empirical average: (sample mean)

$$\hat{S}_N = \frac{1}{N} \sum_{i=1}^{N} f(x^{(i)})$$

The estimator $\hat{S}$ is an unbiased estimator.
↳ (later)

This relies on us to easily sample from the base distrib. p(x), but doing so is not always possible ⟹

Importance Sampling:

Integrand: $p(x) f(x) \xrightarrow{\text{rewrite the integrand:}} q(x) \dfrac{p(x) f(x)}{q(x)}$,

where we now sample from $q(x)$, and average the function $g(x) = \dfrac{p(x) f(x)}{q(x)}$.

11.10 Real World Ex. Importance Sampling.

(Read).

X: time to failure

(Testing software, designed to fail once every 1000 years)

$P[X > \gamma] = ?$ for very large γ,

in a computer generate samples $x^1, \dots, x^n, \dots$

for γ very large, we hardly see an example where $x^{(i)} > \gamma$.

eg. $X \sim N(0,1)$, $P(X > 5) = Q(5) = \underline{2.8 \cdot 10^{-7}}$

$$P(X > 5) = \int_5^\infty p_X(x)\,dx = \int_{-\infty}^{\infty} \underbrace{\mathbb{I}_{(5,\infty)}(x)}_{\text{indicator fn}} p_X(x)\,dx$$

indicator fn $= \begin{cases} 1, & \in (5,\infty) \\ 0, & \text{o/w} \end{cases}$

$$= \int_{-\infty}^{\infty} \left[\mathbb{I}_{(5,\infty)}(x) \frac{p_X(x)}{p_Y(x)} \right] p_Y(x)\,dx$$

Idea: Generate Y's where $P(Y > 5)$ is more likely

$\Longrightarrow$

Using Monte Carlo simulations:

Generate M of Y's: $y^{(1)}, y^{(2)}, \ldots, y^{(M)}$

$$P(X>5) = \frac{1}{M}\sum_{i=1}^{M} g(y^i) = \frac{1}{M}\sum_{i=1}^{M} \mathbb{I}_{(5,\infty)}(y_i)\frac{P_X(y_i)}{P_Y(y_i)}$$

e.g. Choose $P_Y(y) = \lambda e^{-\lambda y} \underbrace{u(y)}$ step (Heaviside fn) $u(y) = \begin{cases} 1, & y \geq 0 \\ 0, & o/w \end{cases}$

$Y \sim \exp(1) \Rightarrow (\lambda \overset{?}{=} 1)$

$$P(Y>5) = \int_{-\infty}^{\infty} \underbrace{\left(\mathbb{I}_{(5,\infty)}(x) \cdot \frac{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}}{e^{-x}}\right)}_{g(x)} \underbrace{e^{-x}}_{P_Y(x)} dx$$

$$= E_Y[g(y)] \quad , \quad Y \sim \exp(\underset{\lambda=1.}{1})$$

$$P(Y>5) = 0.0067$$

Table 11.2 : e.g. $N = 10^4$ realizations, expect about 67 realizations to contribute to the sum! (compare to 0! samples w/ $N(0,1)$).

Sample mean:

$$P(Y>5) = \frac{1}{M}\sum_{i=1}^{M} g(y_i) = \frac{1}{M}\sum_i \mathbb{I}_{(5,\infty)}\frac{1}{\sqrt{2\pi}} e^{\left(-\frac{y_i^2}{2}+y_i\right)}$$

Read Sec 11.10

Estimating Mean & Variance from Data:

→ Read 6.8 from your textbook:

Sample Mean:

$$\hat{\mu}_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$E[\hat{\mu}_x] \to \text{unbiased estimator}$$

$$\xrightarrow[N\to\infty]{} \mu_x$$

Sample Variance:

$$\underbrace{\hat{\sigma}_x^2}$$

$$\hat{\sigma}_x = \frac{1}{N}\sum_{i=1}^{N}(x_i - \underset{\text{sample mean}}{\hat{\mu}_x})^2 \Rightarrow \text{(biased estimator).}$$

Unbiased sample variance: $\hat{\sigma}_x = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{\mu}_x)^2$

App 6-B: Matlab code to calculate these.

Check & study Unbiased estimator,
proof of unbiased sample variance estimator

$$E[\underbrace{\hat{\sigma}_x}] \xrightarrow[N\to\infty]{} \sigma_x^2$$

6.9 Data Compression: Sending

Send symbol sequence w/ symbols A, B, C, D.

AAAA BB ... AC AAA ...

A typical seq. 50 letters DDA ...

$$\left.\begin{array}{l} N_A = 43 \\ N_B = 4 \\ N_C = 1 \\ N_D = 2 \end{array}\right\}$$

eg. use fixed length codes

$$\left.\begin{array}{l} A = 00 \\ B = 01 \\ C = 10 \\ D = 11 \end{array}\right\} \text{2 bits/letter.}$$

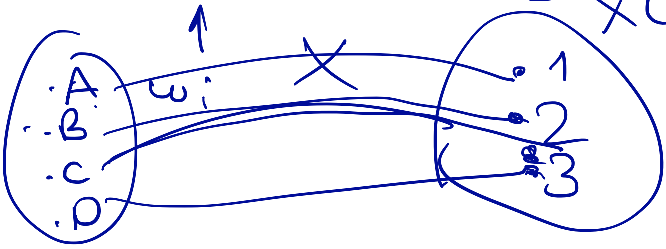
⇒ For a better compression, alternative code:

$A = 0$
$B = 10$
$C = 110$
$D = 111$

compute average length (bits)/letter ?

Define a discrete r.v. that measures length of the codeword:

$S = \{A, B, C, D\}$



$$X(\omega_i) = \begin{cases} 1, & \omega_1 = A \\ 2, & \omega_2 = B \\ 3, & \omega_3 = C, D \end{cases}$$

$P_X[\omega_i] = ?$

$$P_X[k] = \begin{cases} 7/8, & k = 1 \\ 1/16, & k = 2 \\ 1/32, & k = 3, 4 \end{cases}$$

$E_X[X] = \sum k p_X[k]$
$= 1.1875$ bits/letter

Compression ratio: $2/(1.18) = 1.68$.

Shannon entropy: → $H = \sum_k p[k] \ln \frac{1}{p[k]}$
(What is the best $E[X]$ that can be achieved?)

$= 0.73$ bits/letter.

Shannon showed this is the minimum average code length

Moments: $E[X^n] = \int_{-\infty}^{\infty} x^n p_X(x)\,dx$

Centralized Moments: $E[(X-\mu_X)^n]$: nth central moment

$$= \int_{-\infty}^{\infty} (x-\mu_X)^n p_X(x)\,dx$$

nth moment exists if $E[|X|^n] < \infty$.

Property

* If it is known that $E[X^s]$ exists, then $E[X^r]$ exists for $r < s$. (see Prob. 6.23)

eg. Cauchy pdf $\Rightarrow E[X]$ does not exist.

$\therefore$ all higher moments do not exist!

Ex: Moments of an exponential r.v.

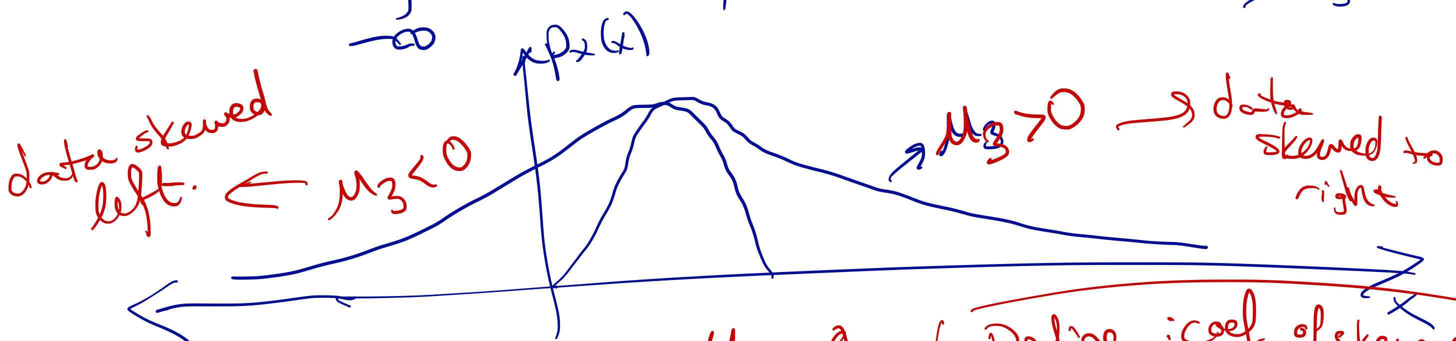
$$E[X^n] = \frac{n}{\lambda} E[X^{n-1}]$$

$\hookrightarrow E[X] = \frac{1}{\lambda}$

3rd Central Moment: Used as a measure of symmetry

$$\mu_3 = \int_{-\infty}^{\infty} (x-\mu_X)^3 p_X(x)\,dx$$

: for Normal distrib $\mu_3 = 0$



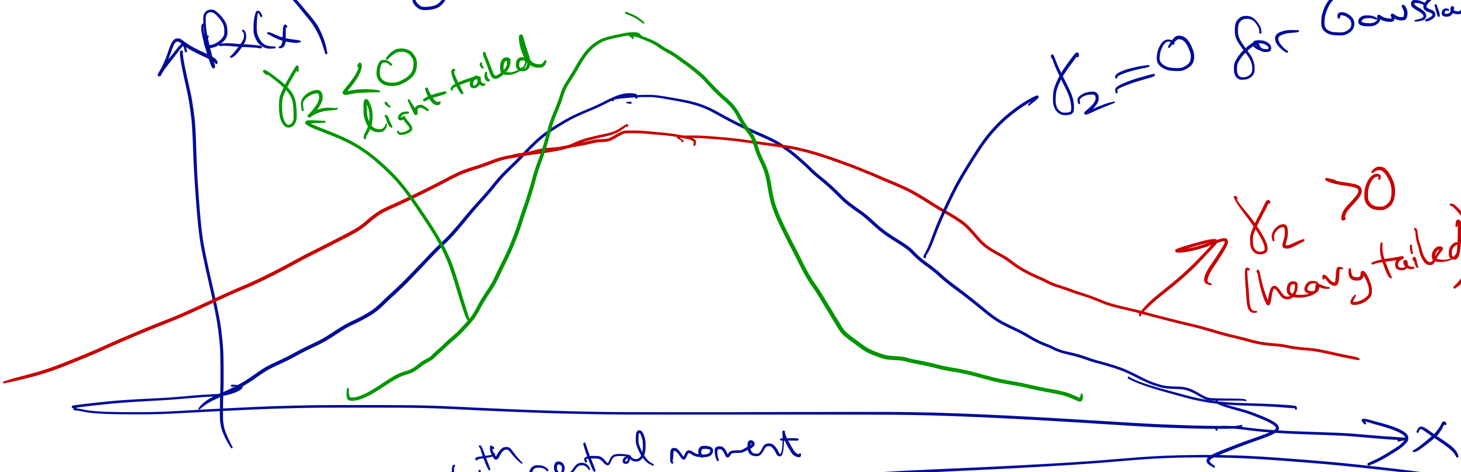
Any sym. distrib: $\mu_3 = 0$.

Define: coef. of skewness

$$\gamma_1 \triangleq \frac{\mu_3}{\sigma^3}$$

4th Central moment: compares a given distrib. w/ Gaussian.

whether heavy / light tailed relative to normal distrib



4th central moment

Define $\boxed{\gamma_2 \triangleq \frac{\mu_4}{\sigma^4} - 3}$ $\Rightarrow$ $\mu_4 = \int_{-\infty}^{\infty} (x-\mu)^4 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

coef. of excess (kurtosis)

eg. for Gaussian $= 3\sigma^4$.

* Moments of interest are generally confined to the first few, eg. 4.

Q: Is the pdf of an r.v. uniquely described by its moments?

In General: No. b/c not all moments need exist.

(eg. Cauchy: valid pdf w/o moments.)

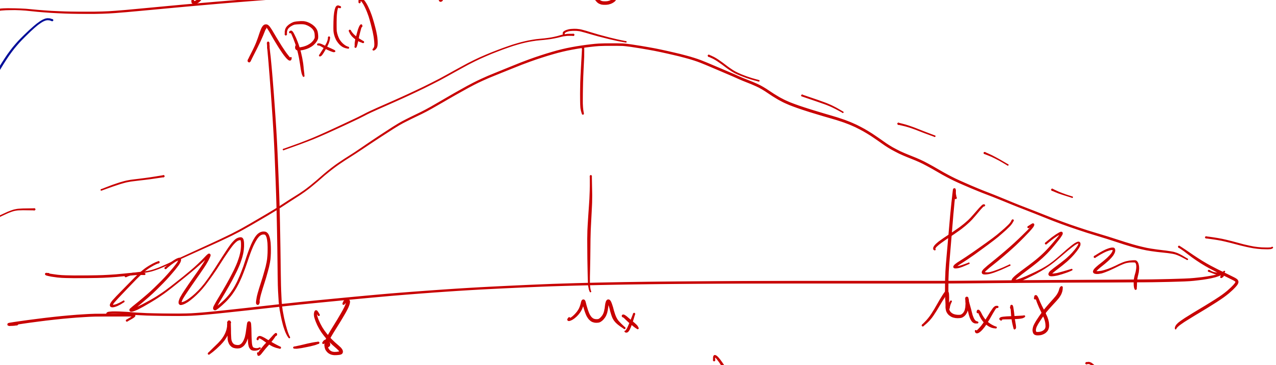
But Gaussian r.v. uniquely described by its 1st 2 moments

* In general, the mean & variance do not provide enough info to characterize the pdf. eg. $\mathcal{N}(0,1)$ & Laplacian $\left(\frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}\right)$

but give us bounds on certain probabilities

$\Longrightarrow$

Chebyshev Inequality:



$$\underbrace{P(|x-\mu_x| > \gamma)}_{\text{Valid for any continuous r.v.}} \le \frac{Var(X)}{\gamma^2} \quad \text{for any } \gamma.$$

For any continuous r.v.:

For any distrib. w/ mean μ_x and var σ^2:

$$P(|x-\mu_x| > 3\sigma) \le \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9} = \boxed{\underline{0.111}}$$

for any distrib.

eg. for Gaussian: we can calculate an exact value

for $P(|X-\mu| \ge 3\sigma) = 1 - \int_{-3\sigma}^{3\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x)^2/2\sigma^2} dx$

$\mu = 0$

Let $\sigma = 1$

$$= 1 - \frac{2}{\sqrt{2\pi}} \int_0^3 e^{-x^2/2} dx$$

$$= 1 - 2\,\mathrm{erf}\,3 = \underline{0.002}$$

Chebyshev bound is very conservative: 0.111 vs 0.002

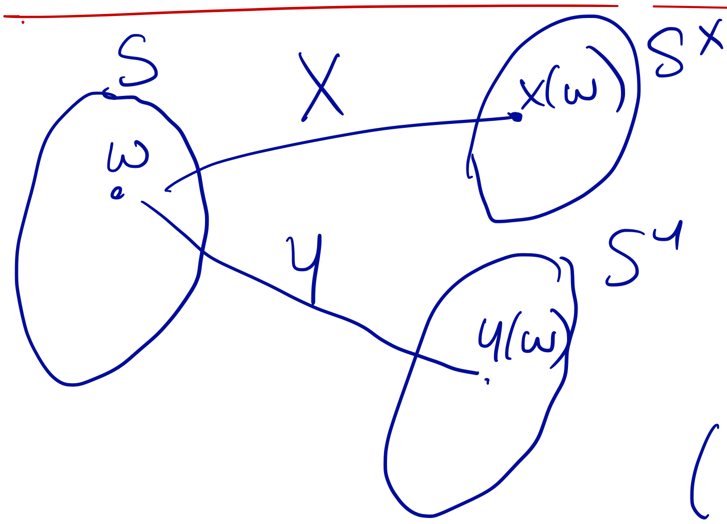
It must take the worst-case distrib. into account. (General) (Gaussian)

Cheb. ineq.:

Used to set limit on prob. of very rare events.

Useful in situations where you have no knowledge of the distrib. itself other than its mean & variance.

Ch. 7 & Ch 12 Multiple Discrete / Continuous Random Variables



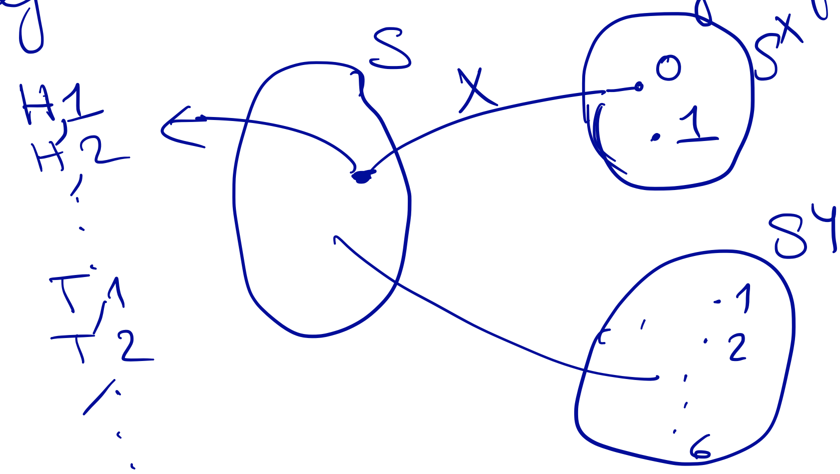
$S^X \subseteq \mathbb{R}$

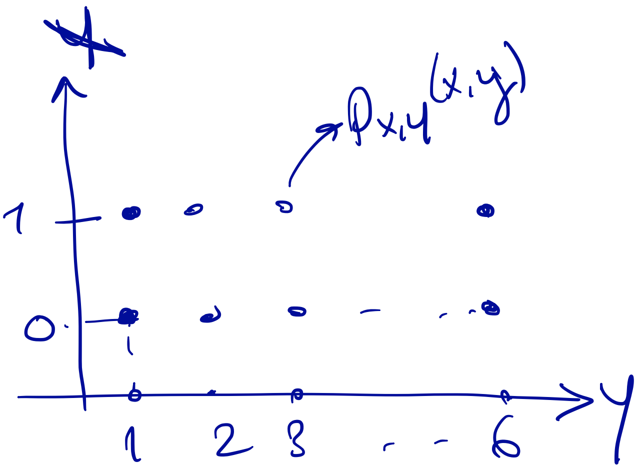
$S_Y \subseteq \mathbb{R}$ — a random vector

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X(\omega) \\ Y(\omega) \end{bmatrix}$$

(X, Y) : jointly distributed random variab.

eg.* Simultaneous tossing of a coin & die :

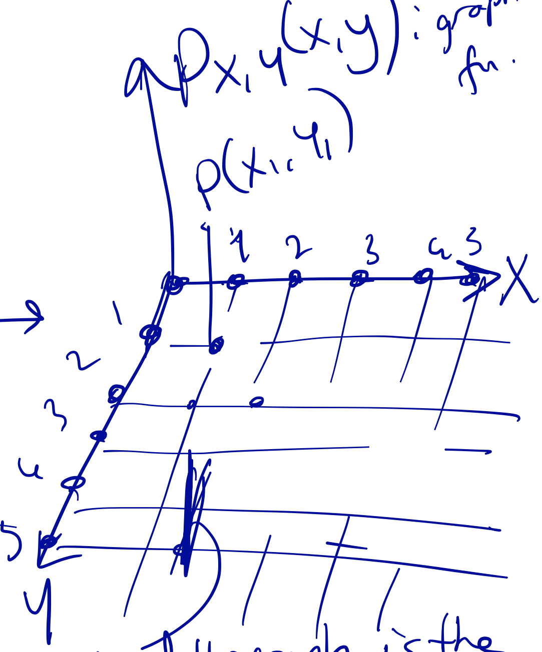


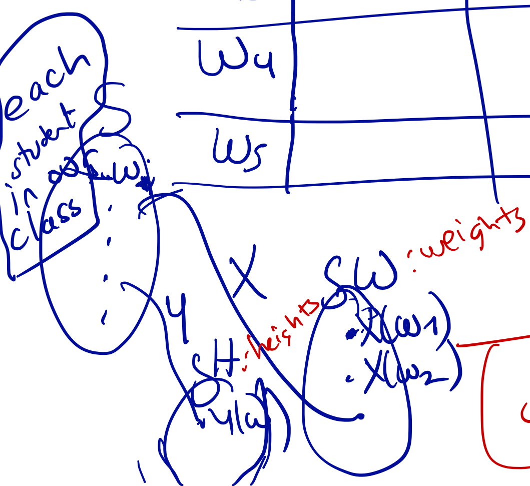


Ex: Student Height & Weight attributes

W \ H	H_1	H_2	H_3	H_4	H_5
W_1	0.08				
W_2					
W_3					
W_4					
W_5					

$p_{X,Y}(x,y)$: 2D graph fn.





joint. prob. space :

$$\boxed{S^{X,Y} = S^X \times S^Y}$$

height of the graph is the joint prob. = prob that a student has a weight & height in the given bin.

Define joint pmf: $p_{X,Y}(x_i, y_j) \triangleq P(X = x_i, Y = y_j)$

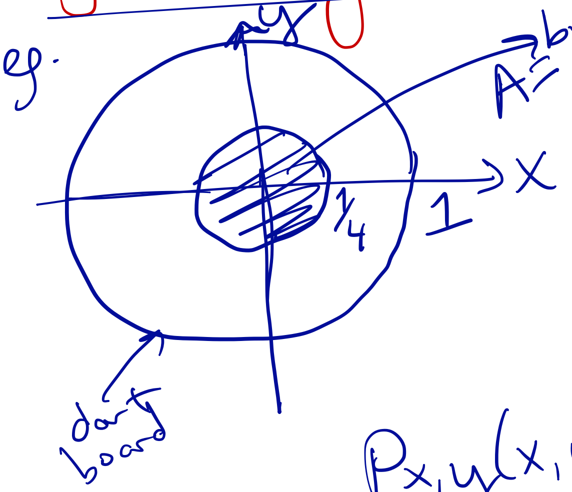
pmf properties:

(1) $0 \le p_{X,Y} \le 1$

(2) $\sum_{i,j} p_{X,Y}(x_i, y_j) = 1$

Cont. Joint pdf:

eg.



A = bull's eye region.

$P(A) = P\left(\sqrt{x^2+y^2} \le \frac{1}{4}\right)$

$= \dfrac{\text{Area}(\text{circle } r = 1/4)}{\text{Area}(\text{dart})_{r=1}}$

$= \dfrac{\pi \cdot 1/16}{\pi r^2} = 1/16.$

$$\underline{p_{X,Y}(x,y)} = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \le 1 \\ 0, & \text{o/w} \end{cases}$$

$$P(\underbrace{\text{bull's eye}}_{A}) = \int_A p_{X,Y}(x,y)\,dx\,dy = \int_{\{(x,y):\, x^2+y^2 \le \frac{1}{16}\}} \frac{1}{\pi}\,dx\,dy = \frac{1}{16}.$$

joint pdf properties:

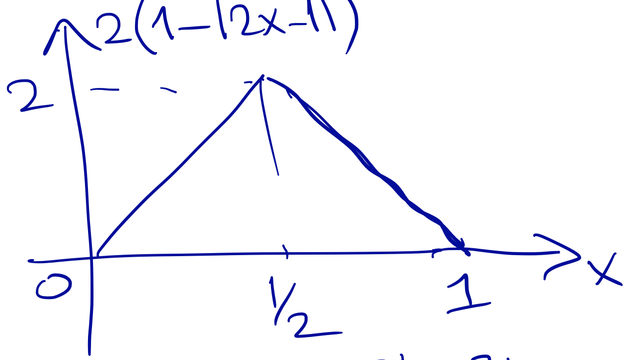
(1) $p_{X,Y}(x,y) \ge 0$

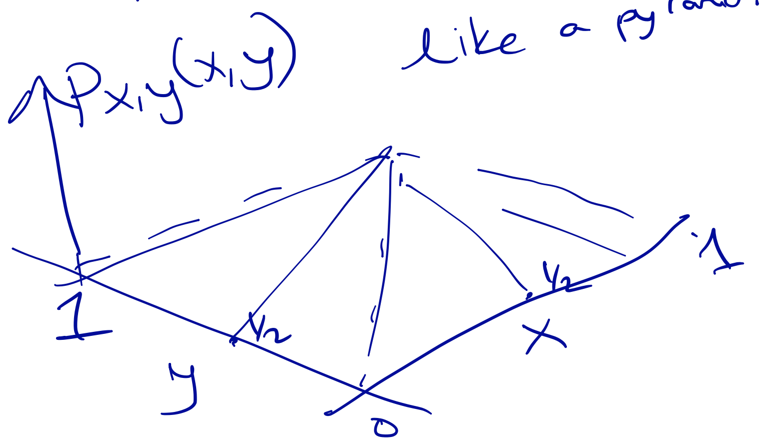
(2) $\underbrace{\iint}_{\underbrace{S^X \times S^Y}_{} \subseteq \mathbb{R}^2} p_{X,Y}(x,y)\,dx\,dy = 1.$

Ex: $p_{X,Y}(x,y) = \begin{cases} 4\,(1-|2x-1|)(1-|2y-1|), & 0\le x\le 1,\ 0\le y\le 1 \\ 0, & o/w. \end{cases}$

exercise: Check $p_{X,Y}$ integrates to 1. ✓

Let $A = \{\frac{1}{4}\le x\le\frac{3}{4},\ \frac{1}{4}\le y\le\frac{3}{4}\}$

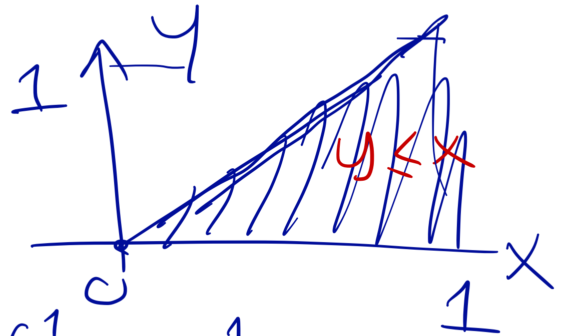




like a pyramid.

$$P(A) = \int_{1/4}^{3/4}\int_{1/4}^{3/4} p_{X,Y}(x,y)\,dx\,dy = 9/16.$$

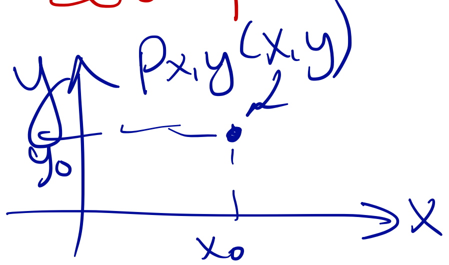
$P(Y\le X) = ?$



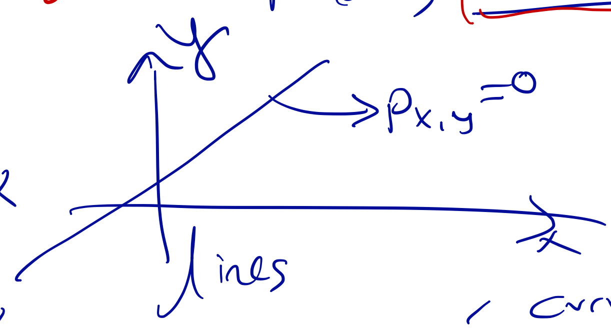
$$= \int_0^1\int_y^1 p_{X,Y}(x,y)\,dx\,dy = \int_0^1 dy\int_y^1 p_{X,Y}(x,y)\,dx$$

Be careful with integration limits!

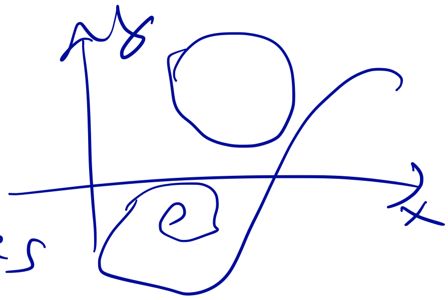
- Zero-probability events in 2D? Cont. joint pdfs: Now; 0-area events have 0 prob.



points,



lines



, curves

Marginal pdfs (pmfs) from joint pdfs.

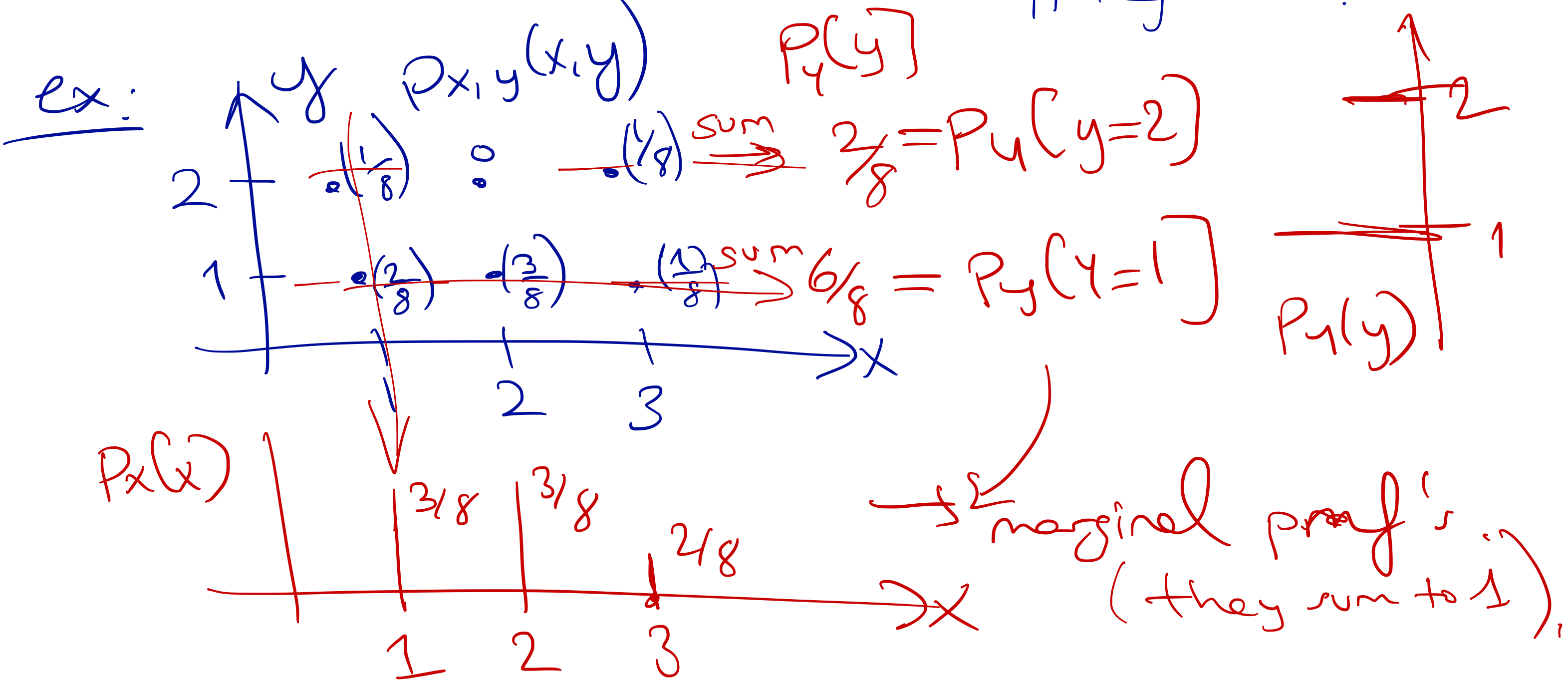
$$P_X(x_i) = P_{X,Y}(X = x_i, Y \in S^Y)$$

$$= \sum_{j=1}^{\infty} P_{X,Y}(x_i, y_j)$$

Marginalization: $P_X(x) = \int_{-\infty}^{\infty} P_{X,Y}(x,y)\,dy$

Given joint pdf (pmf) $\Longrightarrow$ obtain marginals by integration.

ex:



Joint CDF: $F_{X,Y}(x,y) = P(X \le x, Y \le y)$

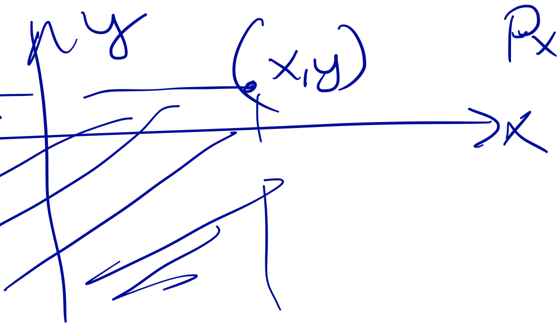
$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} P_{X,Y}(t,u)\,dt\,du$$

Given joint CDF: joint pdf is:

$$P_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y)$$

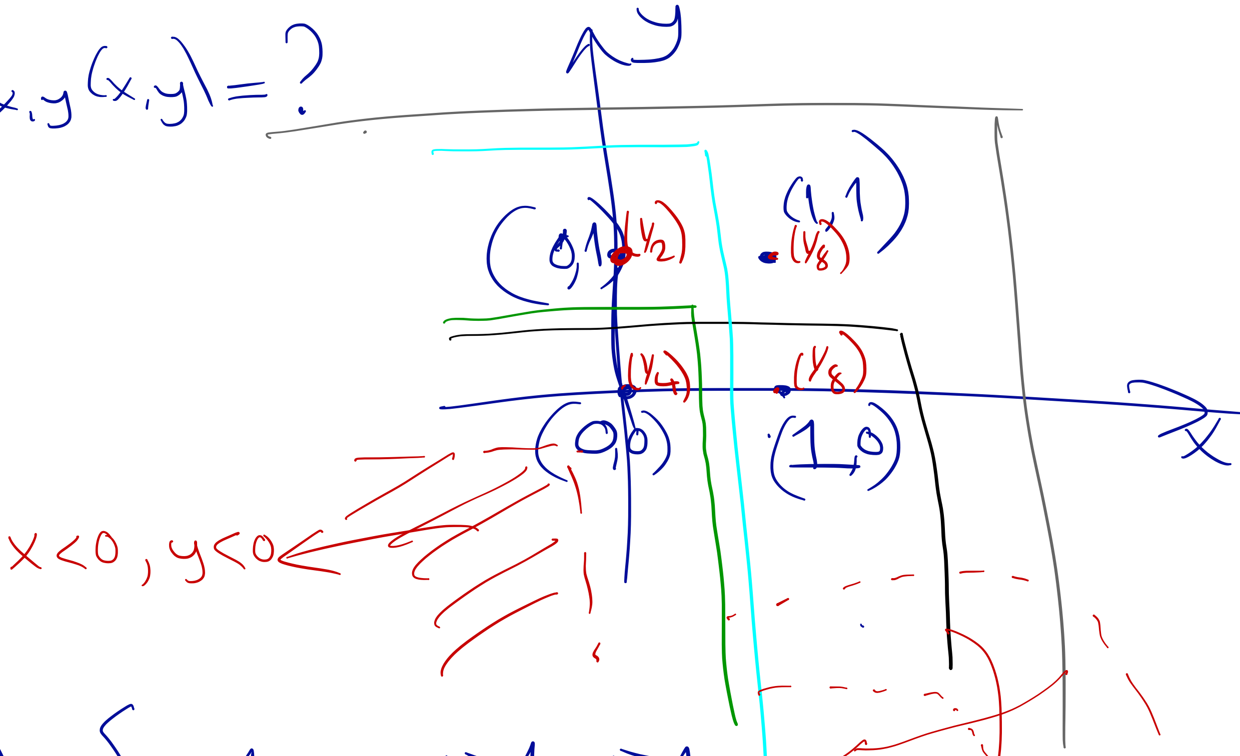
Define Regions in calculating $F_{X,Y}(x,y)$

$\{(X,Y): X \le x, Y \le y\}$



ex: $p_{X,Y}(x,y) = \begin{cases} 1/8, & x=1, y=0 \\ 1/8, & x=1, y=1 \\ 1/4, & x=0, y=0 \\ 1/2, & x=0, y=1 \end{cases}$

CDF: $F_{X,Y}(x,y) = ?$



$$F_{X,Y}(x,y) = \begin{cases} 1, & x \geq 1, y \geq 1 \\ 3/8, & x \geq 1, 0 < y < 1 \\ 3/4, & 0 < x < 1, y \geq 1 \\ 1/4, & 0 < x < 1, 0 < y < 1 \\ 0, & o/w \end{cases}$$

CDF Properties:

1. $0 \leq F_{X,Y} \leq 1$
2. $F(-\infty, -\infty) = 0$

   $F(\infty, \infty) = 1$
3. monotonically non-decreasing
4. Right continuous.