# Probabilistic Graphical Models

## Lectures 19

Introduction to Sampling

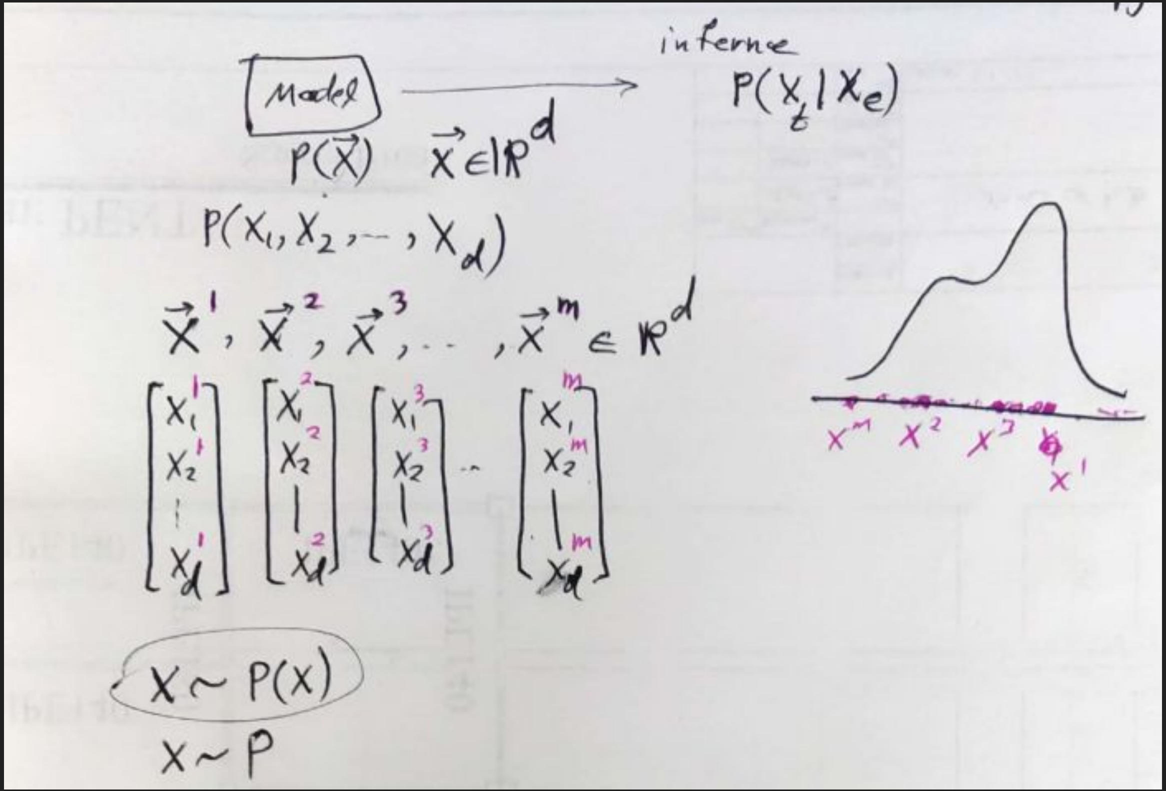
# Sampling

$P(X)$ $\quad\quad\quad\quad$ $X \in R^d$

Samples: $X^1, X^2, X^3, \ldots, X^m \in R^d$

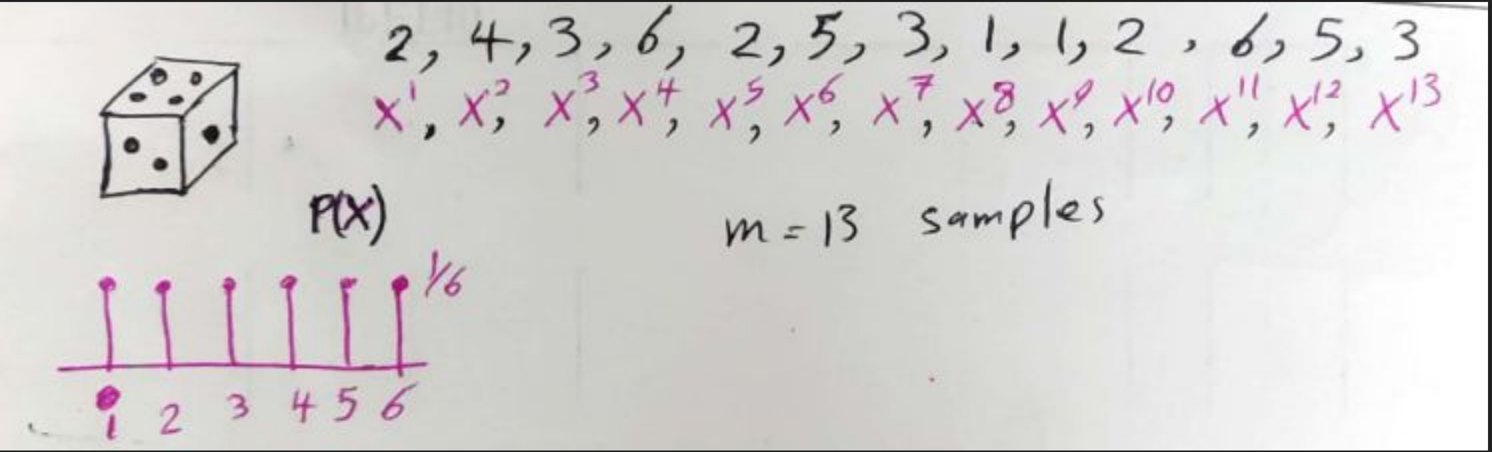
$X^1, X^2, X^3, \ldots, X^m \sim P(X)$

# Sampling

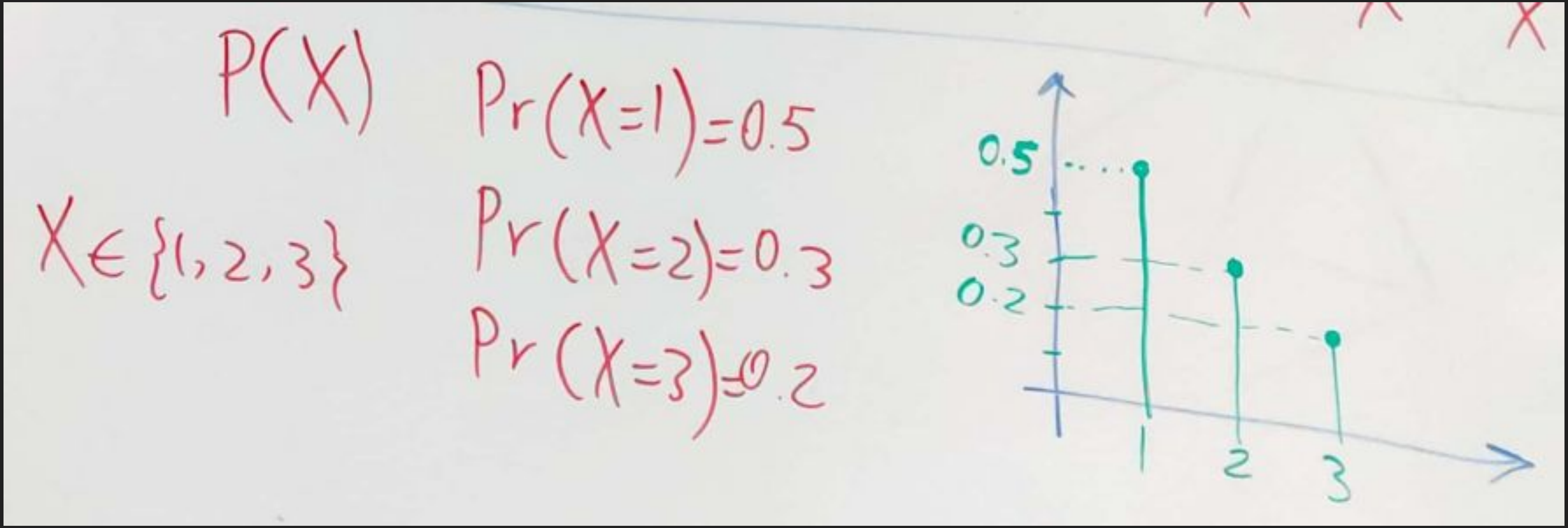


Model
inference
$P(X_t \mid X_e)$
$P(\vec{X})$ $\vec{X} \in \mathbb{R}^d$
$P(X_1, X_2, \dots, X_d)$
$\vec{X}^1, \vec{X}^2, \vec{X}^3, \dots, \vec{X}^m \in \mathbb{R}^d$
$\begin{bmatrix} X_1^1 \\ X_2^1 \\ \vdots \\ X_d^1 \end{bmatrix} \begin{bmatrix} X_1^2 \\ X_2^2 \\ \vdots \\ X_d^2 \end{bmatrix} \begin{bmatrix} X_1^3 \\ X_2^3 \\ \vdots \\ X_d^3 \end{bmatrix} \cdots \begin{bmatrix} X_1^m \\ X_2^m \\ \vdots \\ X_d^m \end{bmatrix}$
$X \sim P(X)$
$X \sim P$
$X^m$ $X^2$ $X^3$ $X^1$

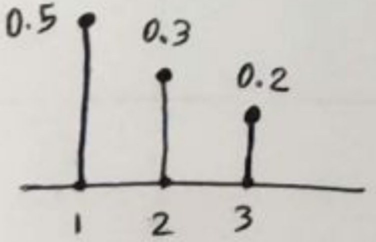
# Sampling - Example



# Sampling - Example



# Sampling - Example




$P(1) = Pr(X=1) = 0.5$

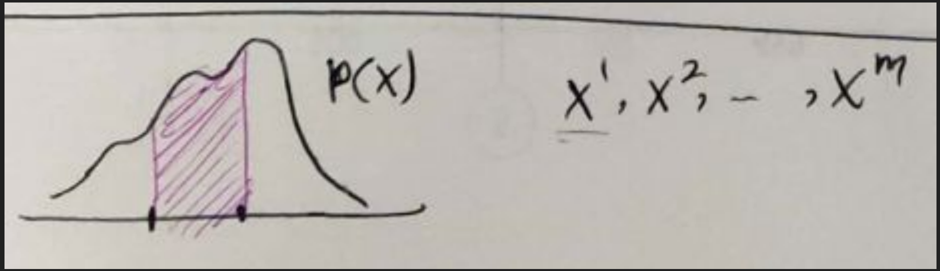
$P(2) = Pr(X=2) = 0.3$

$P(3) = Pr(X=3) = 0.2$

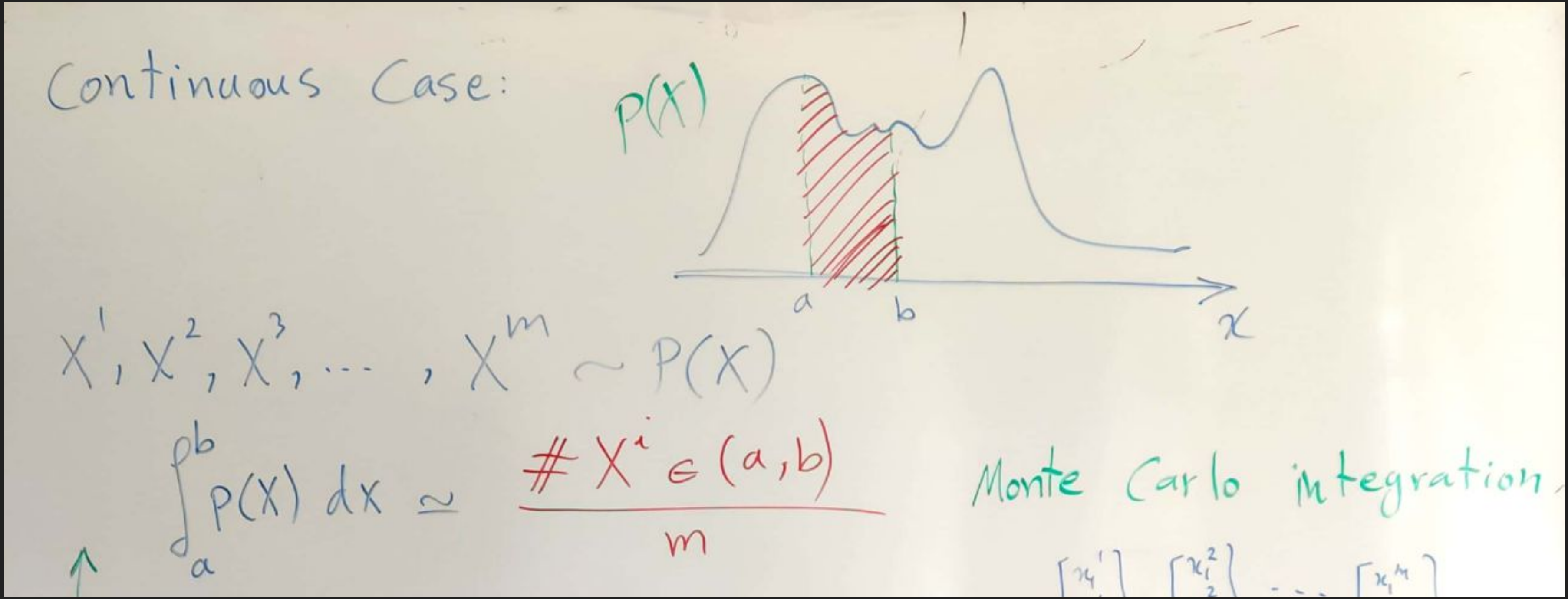
Sampler → 1 2 3

1 1 1 1 1 2 2 2 3 3 → not a valid sampler

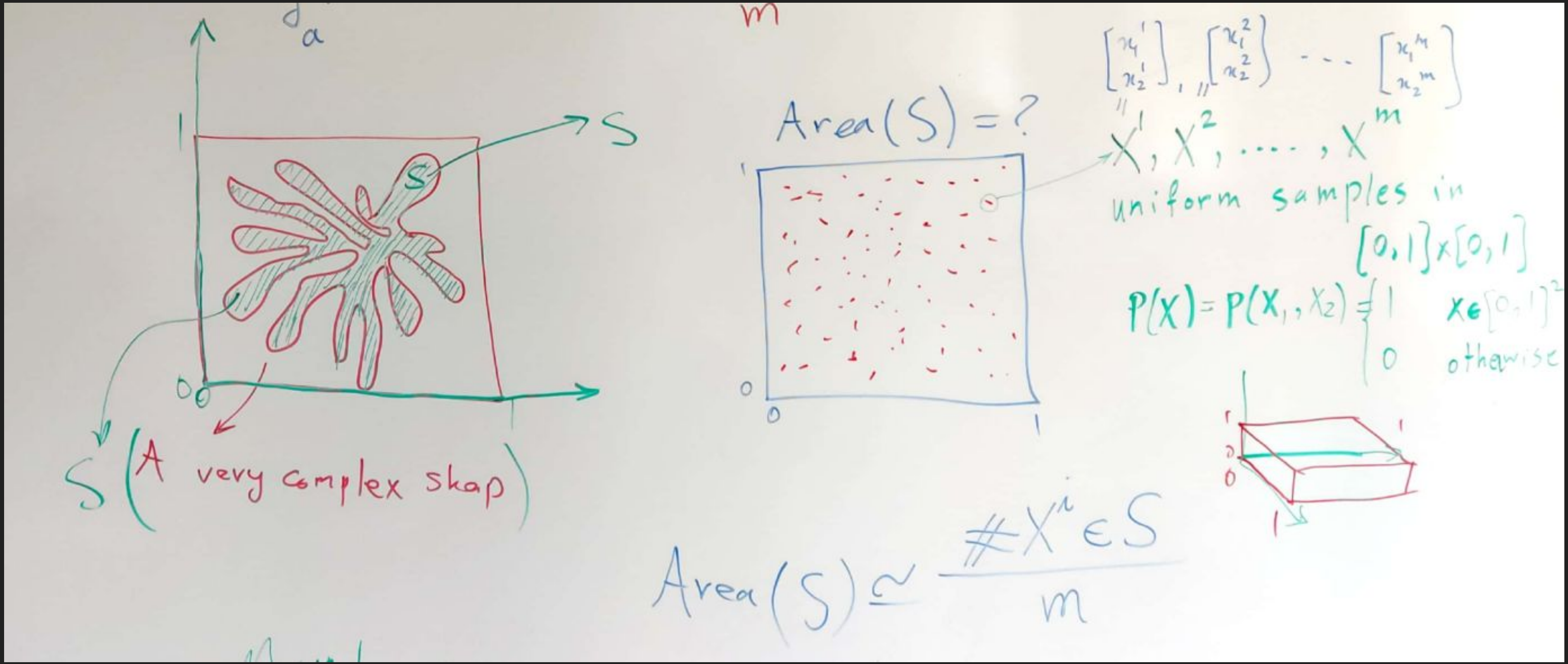
# Sampling - Example



# Sampling - Continuous Case



# Sampling - Continuous Case



# Sampling-based Approximate Inference

K. N. Toosi
University of Technology

$$P(X) = P\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = P(x, y)$$

$x \in \{1, 2, 3\}$

$y = \{0, 1, 2, 3, 4\}$

take samples $(x^1, y^1), (x^2, y^2), \ldots, (x^m, y^m)$

$$\left.\begin{aligned} P(x=1) &\simeq \frac{\# x^i = 1}{m} \\ P(x=2) &\simeq \frac{\# x^i = 2}{m} \\ P(x=3) &\simeq \frac{\# x^i = 3}{m} \end{aligned}\right\} \quad P(x) \simeq \checkmark$$

Inference $\quad P(x, y) \longrightarrow P(x)$

Exact Inference $\quad P(x) = \sum_y P(x, y)$

# Sampling-based Approximate Inference

Inference with query

$$p(x,y) \longrightarrow p(\;x \mid y=b)$$

samples $(x^1,y^1),(x^2,y^2),\dots,(x^m,y^m)$

$$p(x=a \mid y=b) = \frac{Pr(x=a, y=b)}{Pr(y=b)} \simeq \frac{\#(x^i=a,\ y^i=b)/m}{\#(y^i=b)/m}$$

$$\Rightarrow P(x=a \mid y=b) \sim \frac{\#(x^i=a,\ y^i=b)}{}$$

# Sampling-based Approximate Inference

K. N. Toosi
University of Technology

pgm 19 (II)

$$P\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \underline{p(x,y)}$$

$$\boxed{P_x(x)} = \sum_y p(x,y)$$

$$\begin{bmatrix}x^1\\y^1\end{bmatrix}, \begin{bmatrix}x^2\\y^2\end{bmatrix}, \dots, \begin{bmatrix}x^m\\y^m\end{bmatrix} \sim p(x,y)$$

$$P_x(1) \doteq \Pr(X=1) \simeq \frac{\#(x^i=1)}{m}$$

$$P(X=2 \mid Y=3) = \frac{P(X=2, Y=3)}{P(Y=3)} \simeq \frac{\frac{\#(X^i=2,\, y^i=3)}{m}}{\frac{\#(y^i=3)}{m}}$$

$$\Rightarrow \underline{P(X=2 \mid Y=3)} \simeq \frac{\#(X^i=2,\, y^i=3)}{\#(y^i=3)}$$

# Empirical Expectation

Expectation: $P(X) \xrightarrow{\sim \text{ samples}} x^1, x^2, \dots, x^m$

$$P(X) \begin{cases} \text{discrete} & E(X) = \sum_x x\, p(x) \\ \text{continuous} & E(X) = \int x\, p(x)\, dx \end{cases} \Bigg\} \quad E(X) \simeq \frac{\sum_{i=1}^{m} x^i}{m}$$

for any function $h(x)$

$$P(X) \begin{cases} \text{discrete} & E(h(x)) = \sum_x h(x)\, P(X) \\ \text{continuous} & E(h(x)) = \int h(x)\, P(X)\, dx \end{cases} \Bigg\} \quad E(X) \simeq \frac{\sum_{i=1}^{m} h(x^i)}{m}$$

# Monte-Carlo Integration

K. N. Toosi
University of Technology

$$h(x) = 1(x \in S) \Rightarrow E(h(x)) = \int_S p(x)\,dx \approx \frac{\sum_{i=1}^{m} 1(x^i \in S)}{m} = \frac{\#(x^i \in S)}{m}$$

monte carlo integration

$P(X)$ : uniform $\rightarrow$ computing area

$P(x,y)$

$$\left.h(x,y) = 1(x=a)\right\} E(h(x,y)) = E(1(x=a)) = \sum_{x,y} 1(x=a)\,p(x,y)$$

inference

$$= p(x=a)$$

# Monte Carlo Integration

K. N. Toosi
University of Technology

$$\int f(x)\,dx \simeq ?$$

1. write $f(x)$ in the form of $f(x) = h(x)\,\overbrace{p(x)}^{\text{P.D.F}}$
2. Take samples $x^1, x^2, \ldots, x^m$ from $p(x)$
3. $\int f(x)\,dx = \int h(x)\,p(x)\,dx \simeq \frac{1}{m}\sum_{i=1}^{m} h(x^i)$

Importance Sampling: How to decompose $f(x)$ into $h(x)\,p(x)$

K. N. Toosi
University of Technology

# Inference as Expectation

$$p(x, y)$$

$$p_x(x) = \sum_y p(x, y)$$

$$\Pr(X=a) = p_x(a) \qquad \overset{h(x,y)}{\cancel{h(x)}} = 1(X=a)$$

$$E\{h(x)\} = \sum_x \sum_y \underset{h(x,y)}{\cancel{h(x)}}\, p(x, y) = \sum_x \sum_y 1(X=a)\, p(x, y)$$

$$= \sum_x 1(x=a) \boxed{\sum_y p(x, y)}$$

$$= \sum_x 1(x=a)\, p_x(x)$$

$$= p_x(a) \simeq$$

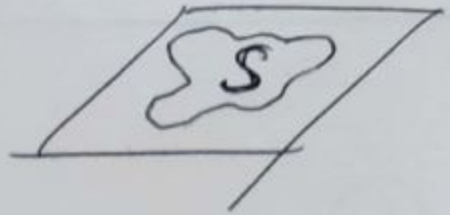
$$p_x(a) = \Pr(X=a) \simeq \frac{1}{m} \sum_{i=1}^{m} h(x^i, y^i)$$

$$= \frac{1}{m} \sum_{i=1}^{m} 1(x^i = a) = \frac{\#(x^i = a)}{m}$$

# Area as Expectation

$$h(x) = 1(x \in S)$$

$$E\{h(x)\} = \int_S p(x)\, dx$$

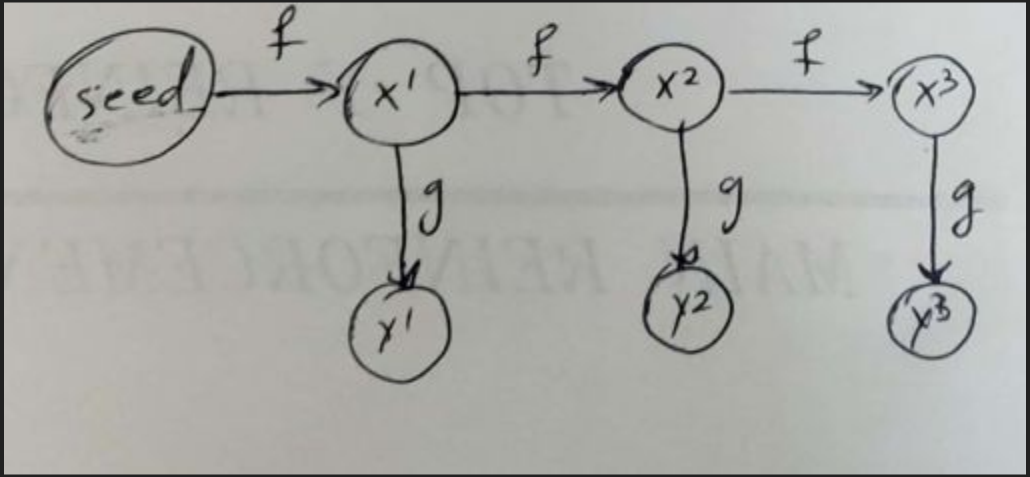


# How to generate samples from p(x)?

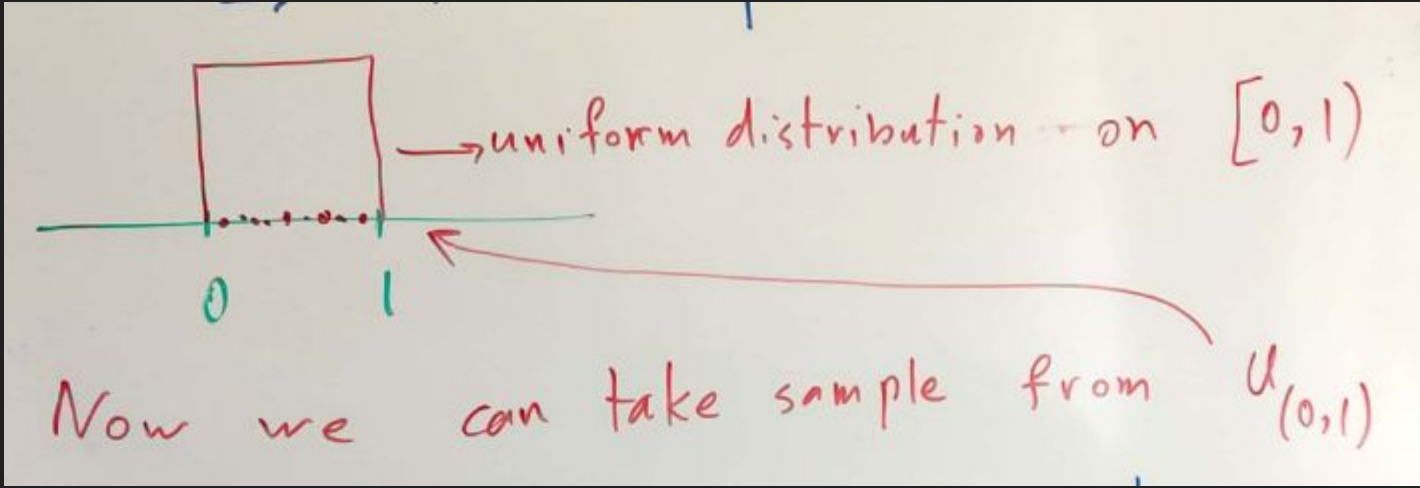
K. N. Toosi
University of Technology

# Pseudo-random sample generation

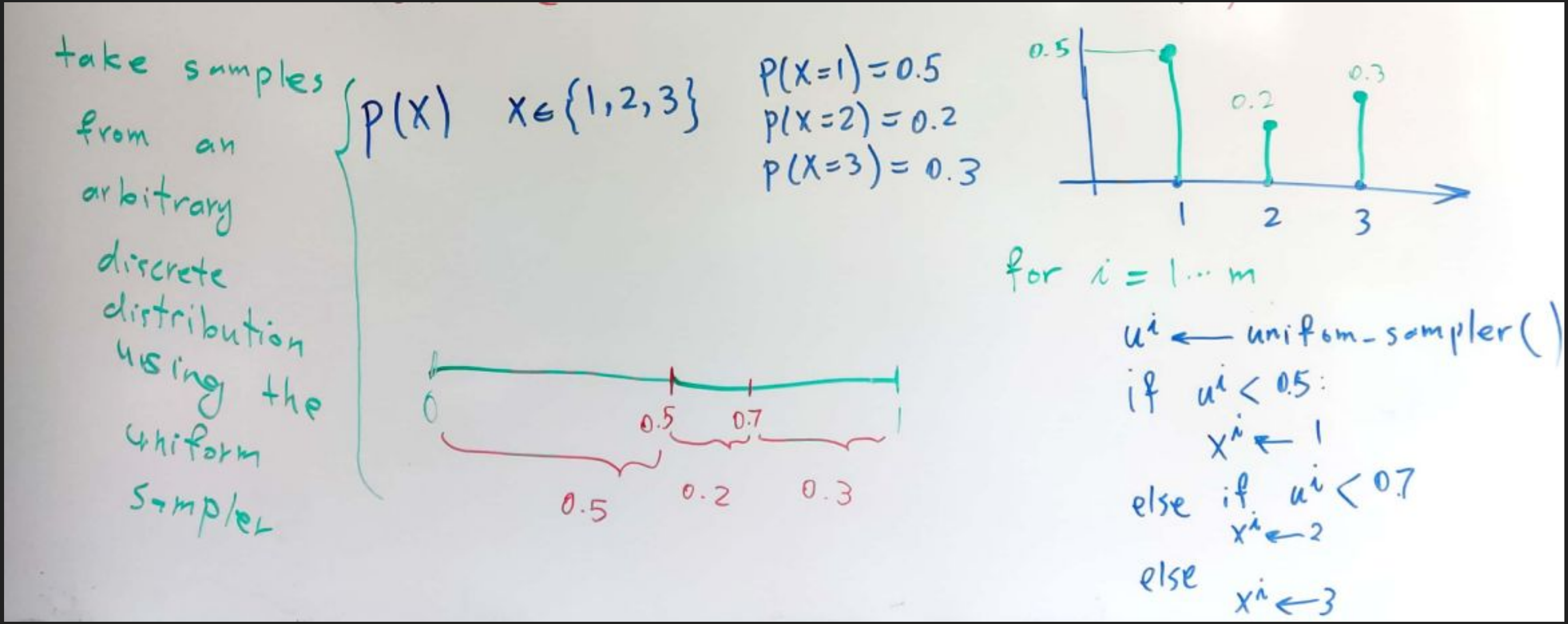
- Pseudo-random number generators
- Samples appear random
- Generated using a deterministic algorithm
- have statistical properties that make them behave like random



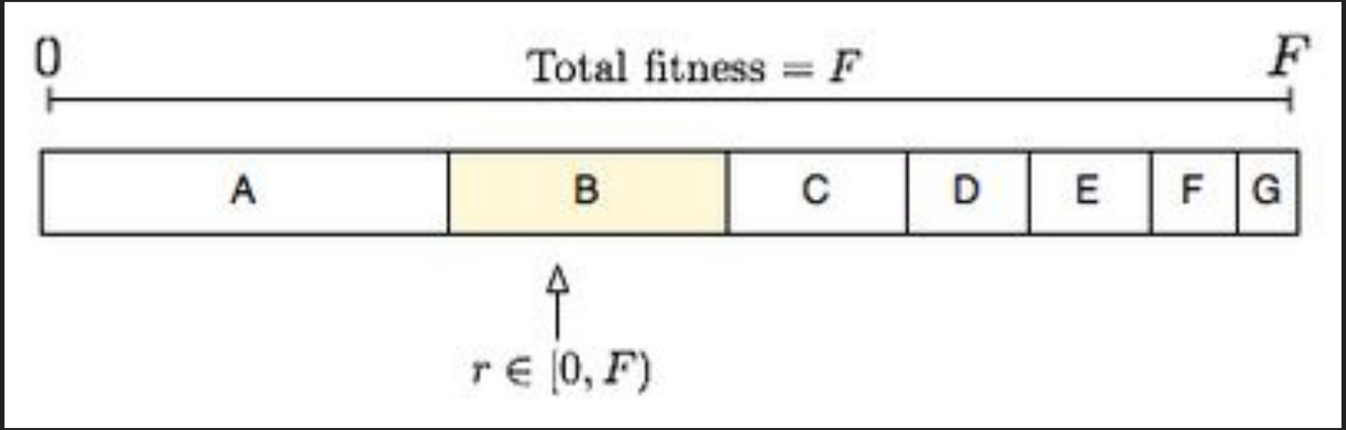
# Uniform Sampler



# Sampling discrete distributions



# Fitness proportionate selection

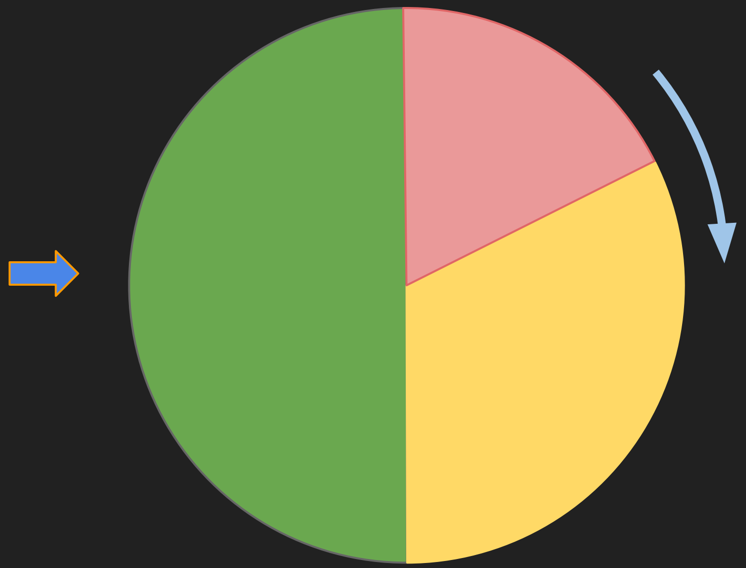



https://en.wikipedia.org/wiki/Fitness_proportionate_selection

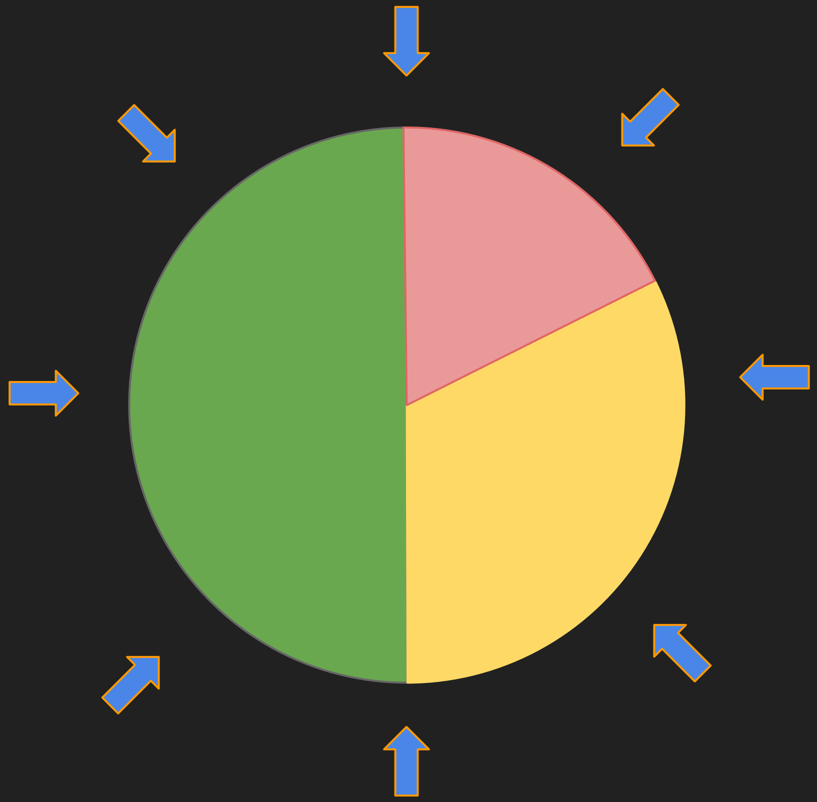
# Sampling discrete distributions

K. N. Toosi
University of Technology

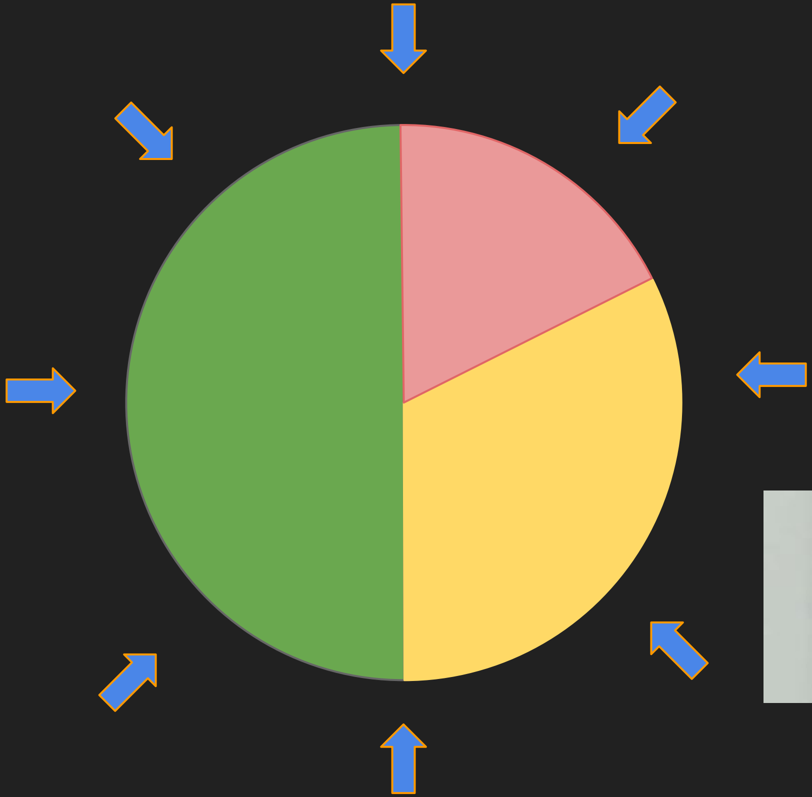
# roulette wheel selection

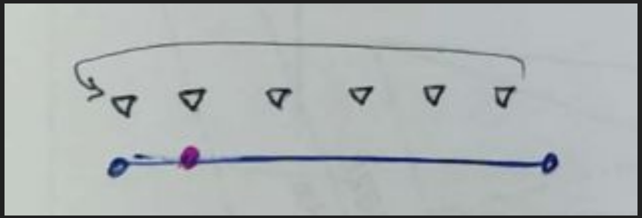


# roulette wheel selection

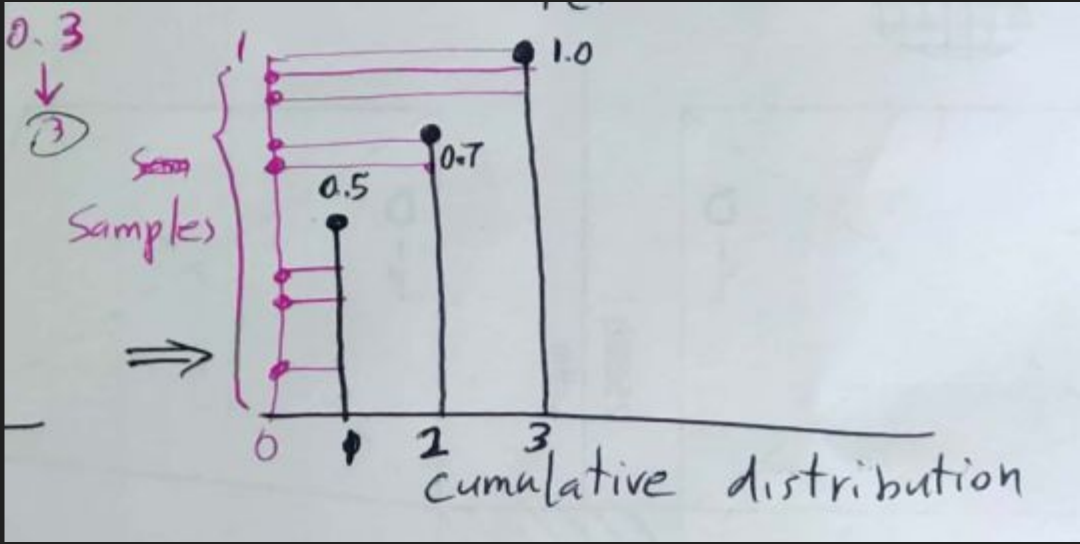


# roulette wheel selection





# Cumulative Distribution View



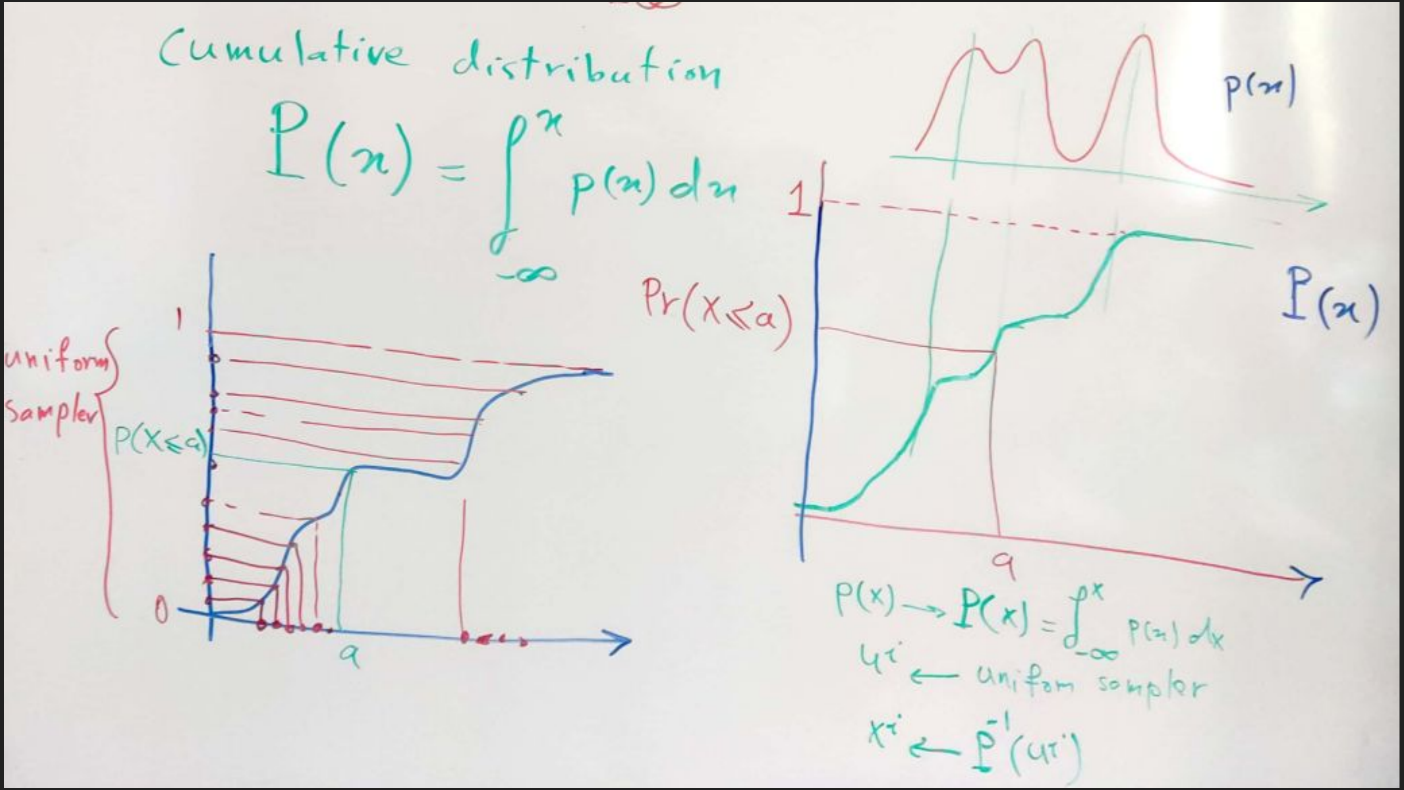
# Sampling from continuous distributions

sampling from arbitrary 1D continuous distro


$x^i$ sample

$$Pr(x^i \leq a) = S = \int_{-\infty}^{a} p(x)\, dx$$

# Sampling from continuous distributions



# i.i.d samples

- Independent
- Identically distributed