Lecture 7: Basic Sampling

Professor Ilias Bilionis

Sampling the categorical

PREDICTIVE
SCIENCE LABORATORY

Example: Sampling from the Bernoulli distribution

$$X \sim \text{Bernoulli}(\theta); \quad X = \begin{cases} 1, & \text{w/pr. } \theta \\ 0, & \text{otherwise} \end{cases}$$

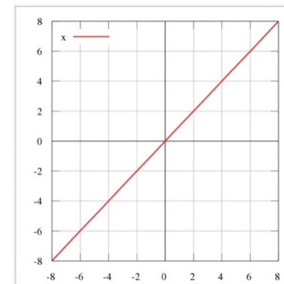
To sample from it, we do the following steps:

Algorithm:

- Sample a uniform number u. $\sim U([0,1])$
- If $u \leq \theta$, then set $x = 1$.
- Otherwise, set $x = 0$.

$U \sim U([0,1])$

$$X = \begin{cases} 1, & \text{if } U \leq \theta \\ 0, & \text{otherwise} \end{cases}$$



Graph of the identity function on the real numbers

$$p(X=1) = P(U \leq \theta) = \underbrace{F_U(\theta)}_{} = \theta$$

$$p(X=0) = 1 - \theta$$

cdf of standard uniform is the identity function

Sampling discrete distributions

K : possibilities

$0, 1, 2, \dots, K-1$ - enumerating *

$p_0, p_1, p_2, \dots, p_{K-1}$

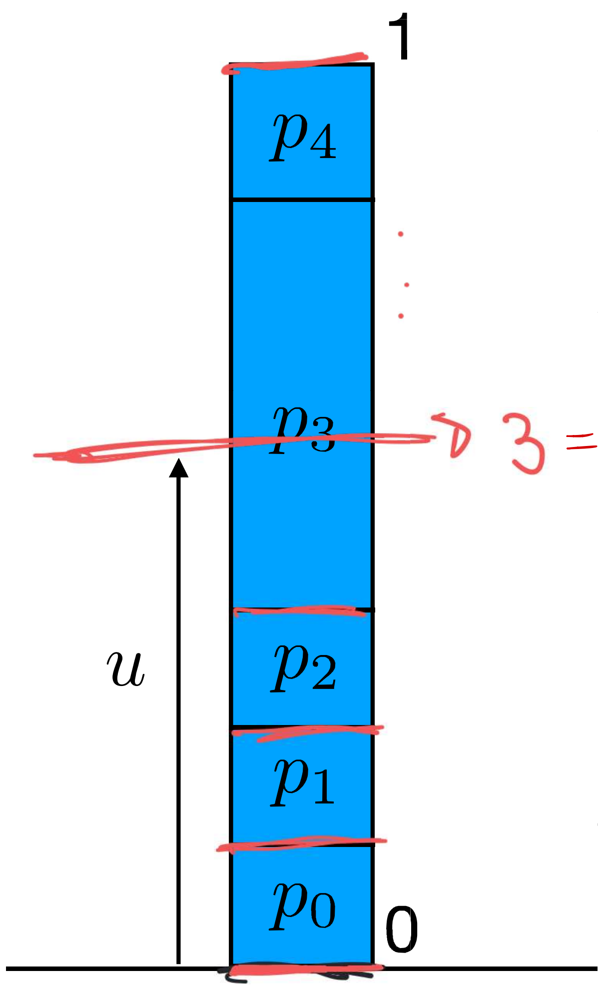
- Consider a generic discrete random variable taking different values, with probability:

$$\boxed{p(X=k)=p_k.}$$

$$X \sim \text{Categorical}(p_0, p_1, \dots, p_{K-1}) \; ; \; X = \begin{cases} 0, & \text{w/ pr. } p_0 \\ 1, & \text{w/ pr. } p_1 \\ \vdots & \\ K-1, & \text{w/ pr. } p_{K-1} \end{cases}$$

PREDICTIVE SCIENCE LABORATORY

Sampling Discrete Distributions



- Draw a uniform number u. $\sim U([0,1])$

- Find j such that:

3 = j here

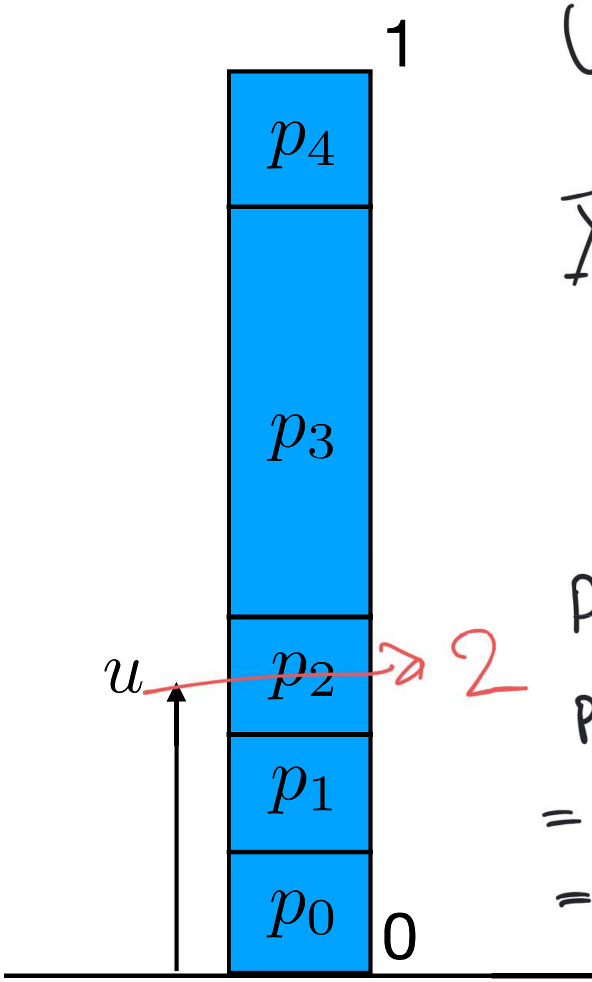
$$\sum_{k=0}^{j-1} p_k \leq u < \sum_{k=0}^{j} p_k$$

- j is your sample

Sampling Discrete Distributions

$$U \sim U([0,1])$$

$$X = \begin{cases} 0, & \text{if } U \le p_0 \\ 1, & \text{if } p_0 \le U \le p_0+p_1 \\ 2, & \text{if } p_0+p_1 \le U \le p_0+p_1+p_2 \\ \vdots & \\ K-1, & \text{otherwise} \end{cases}$$



Sample 1

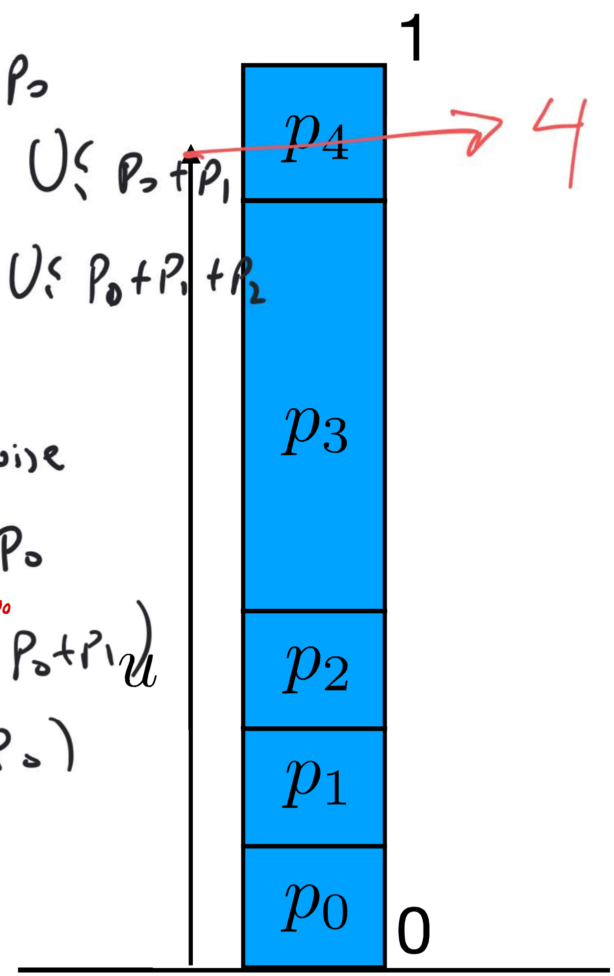
$$P(X=0) = \underbrace{P(U \le p_0)}_{\text{cdf evaluated at } p_0} = p_0$$

$$P(X=1) = P(p_0 \le U \le p_0+p_1)$$

$$= F_U(p_0+p_1) - F_U(p_0)$$

$$= \cancel{p_0} + p_1 - \cancel{p_0} = p_1$$

$$P(X=2) = \ldots p_2$$



Sample 2

$\therefore$ X has same pdf as a categorical