Metody probabilistyczne I statystyka, 2021
informatyka algorytmiczna, WIiT PWr

2-Monte Carlo methods

Part 2

Estimating means and standard deviations:

- **CLT: when computing the sum of iid random variables then the result converges to normal distribution**

- **However: the parameters of normal distribution depend on Exp and Var:**

$$\bar{X} = \frac{1}{N}(X_1 + \ldots + X_N)$$

$$\begin{aligned} \mathrm{E}(\bar{X}) &= \frac{1}{N}(\mathrm{E}X_1 + \ldots + \mathrm{E}X_N) = \frac{1}{N}(N\mu) = \mu, \quad \text{and} \\ \mathrm{Var}(\bar{X}) &= \frac{1}{N^2}(\mathrm{Var}X_1 + \ldots + \mathrm{Var}X_N) = \frac{1}{N^2}(N\sigma^2) = \frac{\sigma^2}{N}. \end{aligned}$$

Expected value:

$$\bar{X} = \underline{\frac{1}{N}(X_1 + \ldots + X_N)} \qquad \underbrace{\mathbf{E}(\bar{X}) = \mu}$$

So we have an *unbiased estimator*

Variance:

The situation is more complicated:

$$\operatorname{Var}(\bar{X}) \;=\; \frac{1}{N^2}\left(\operatorname{Var}X_1 + \ldots + \operatorname{Var}X_N\right) = \frac{1}{N^2}(N\sigma^2) = \frac{\sigma^2}{N}.$$

But we need to compute variance $\operatorname{Var}X_1$, **...**

Impossible, since we have only an estimator for the expected value

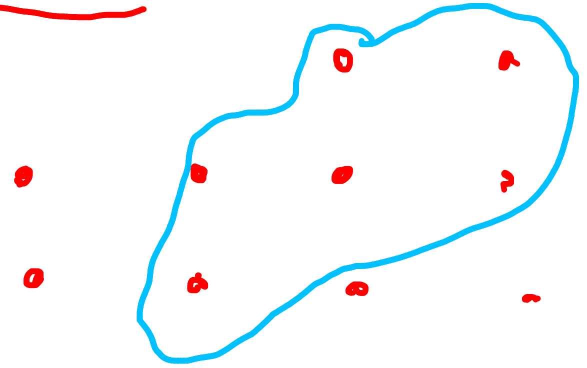
Solution (to be explained later) -- an unbiased estimator:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

$E(s^2) = \frac{\sigma^2}{N}$

Estimating volume:

Naïve approach: **take grid points and check how many of them fall into a set A**

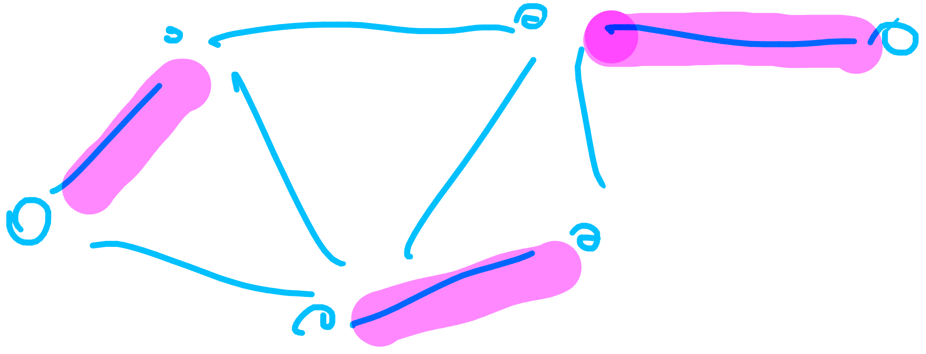


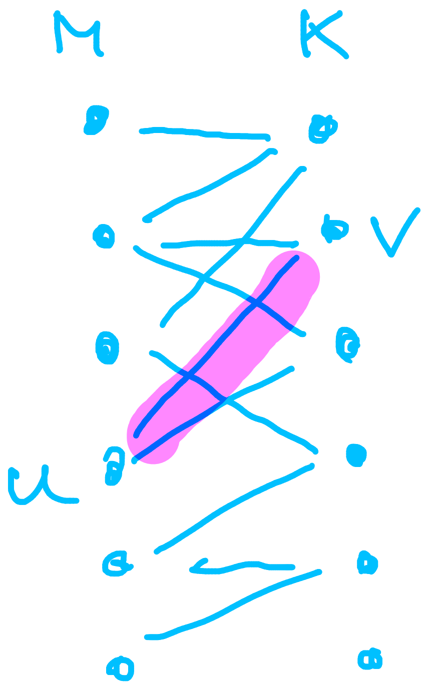
Problem cases:

(Very) Complicated cases:

Spaces where alone finding the elements as well as finding random elements is hard

Example: maximal matchings in a graph G that contain an edge (u,v)



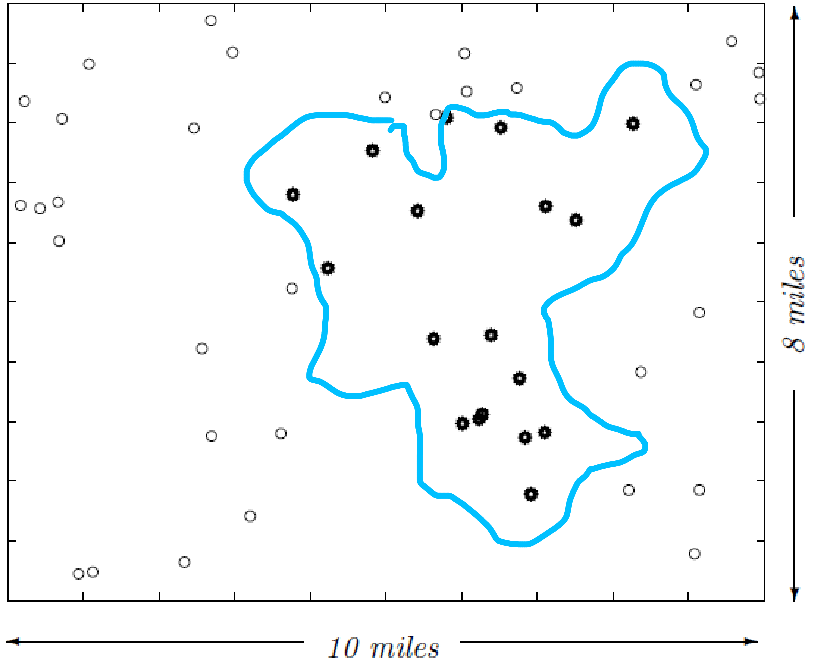


General approach:

1. N random variables, Y(i) is an element of the space chosen with uniform probability
2. X(i)=1 iff Y(i) belongs to A, otherwise X(i)=0

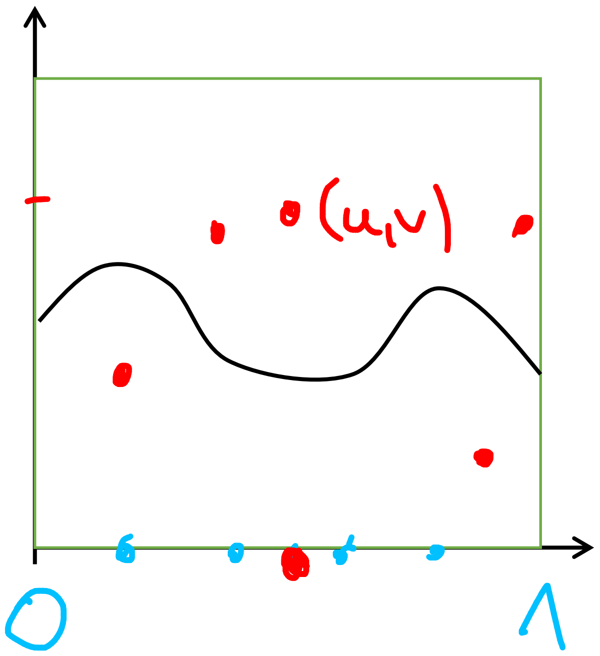
Volume of A = E[$\frac{1}{N}(X_1 + \ldots + X_N)$]

Easier than interpretation of a picture and drawing boundaries:



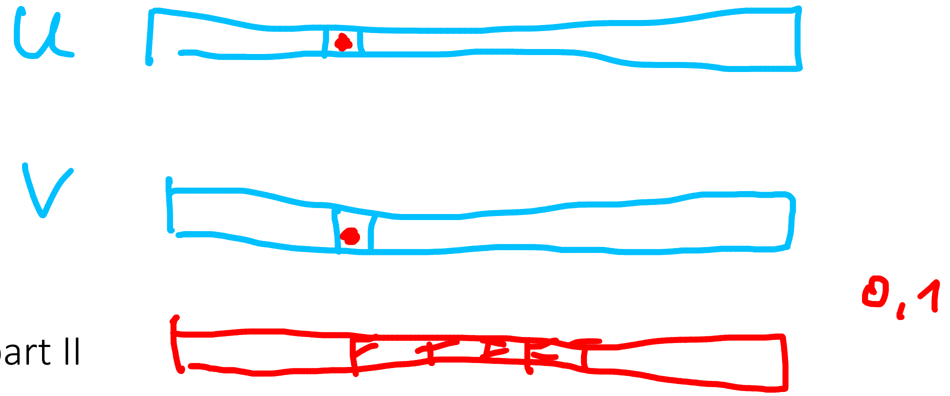
Monte Carlo integration:

```
N  =  1000;              % Number of simulations
U  =  rand(N,1);         % (U,V) is a random point
V  =  rand(N,1);         % in the bounding box
I  =  mean( V < g(U) )   % Estimator of integral I
```



Accuracy:

$$\mathrm{Std}\left(\hat{\mathcal{I}}\right) = \sqrt{\frac{\mathcal{I}(1-\mathcal{I})}{N}}$$



Monte Carlo integration - improved:

$$\mathcal{I} = \underbrace{\int_a^b g(x)dx}_{} = \underbrace{\int_a^b \frac{g(x)}{f(x)} f(x)\, dx}_{} = \mathbf{E}\left(\frac{g(X)}{f(X)}\right)$$

```
N = 1000;                          % Number of simulations
Z = randn(N,1);                    % Standard Normal variables
f = 1/sqrt(2*Pi) * exp(-Z.^2/2);   % Standard Normal density
Iest = mean( g(Z)./f(Z) )          % Estimator of ∫_{-∞}^{∞} g(x) dx
```

Accuracy:

choose f such that g(X)/f(X) is nearly constant
then variance of a random variable R=g(X)/f(X) is small

→ so the average has smaller variation as well

For f=1

$\int g(x)dx$

$$\sigma^2 = \operatorname{Var} R = \operatorname{Var} g(X) = \underbrace{\mathbf{E} g^2(X) - \mathbf{E}^2 g(X)} = \int_0^1 g^2(x)dx - \mathcal{I}^2 \leq \mathcal{I} - \mathcal{I}^2,$$

$g^2 \leq g$ for $0 \leq g \leq 1.$