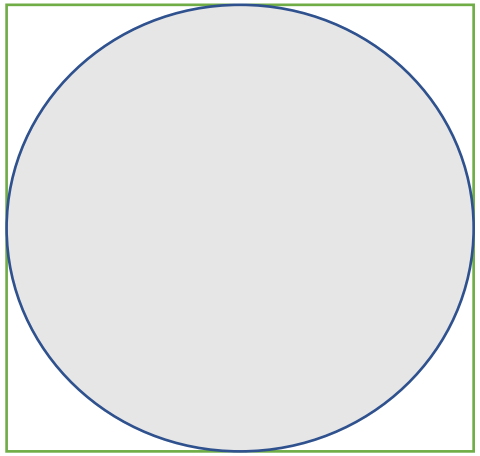
Probability and statistics, 2022, Computer Science Algorithmics, Undergraduate Course, Part II, lecturer: Mirosław Kutyłowski

2. Monte Carlo methods

Computing π on a desert:

Option 1: use Taylor series

Option 2: random experiment



General situation

- a subset *A* of all values of a random variable *X*,
- what is the probability *p* that *X* falls into *A*?

We run *n* independent experiments and get *n* values of *X*

Estimation:

$$\hat{p} = \widehat{\mathbf{P}}\{X \in A\} = \frac{\text{number of } X_1, \ldots, X_N \in A}{N},$$

$$\begin{aligned} \mathbf{E}(\hat{p}) &= \frac{1}{N}(Np) = p, \text{ and} \\ \mathrm{Std}(\hat{p}) &= \frac{1}{N}\sqrt{Np(1-p)} = \sqrt{\frac{p(1-p)}{N}}. \end{aligned}$$

How good it is this method?

We need guarantees like:

"the probability that $|p-\hat{p}| > \delta$
is at most ε "

To simplify the computation we can work on normal distribution:

$$\frac{N\hat{p} - Np}{\sqrt{Np(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} \approx \text{Normal}(0,1),$$

For normal distribution:

$$P\{|\hat{p}-p|>\varepsilon\}=P\left\{\frac{|\hat{p}-p|}{\sqrt{\frac{p(1-p)}{N}}}>\frac{\varepsilon}{\sqrt{\frac{p(1-p)}{N}}}\right\}\approx 2\Phi\left(-\frac{\varepsilon\sqrt{N}}{\sqrt{p(1-p)}}\right).$$

$$\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz, \text{ Standard Normal cdf}$$

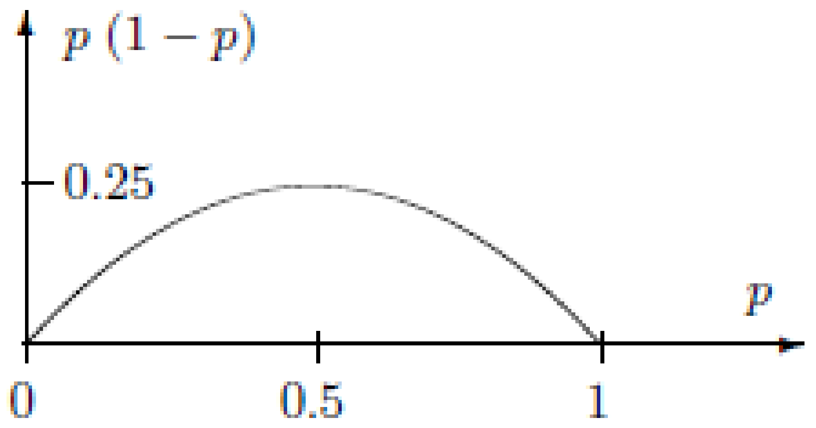
problem: we do not know *p* to perform this computation

A solution:

1) calculations for an intelligent guess for ***p***

2) taking the worst possible ***p*** :

    make ***p(1-p)*** as big as possible (it happens for ***p*=0.5**)



A solution:

approach 1:

$$2\Phi\left(-\frac{\varepsilon\sqrt{N}}{\sqrt{p^*(1-p^*)}}\right) \leq \alpha$$

approach 2:

$$2\Phi\left(-2\varepsilon\sqrt{N}\right) \leq \alpha$$

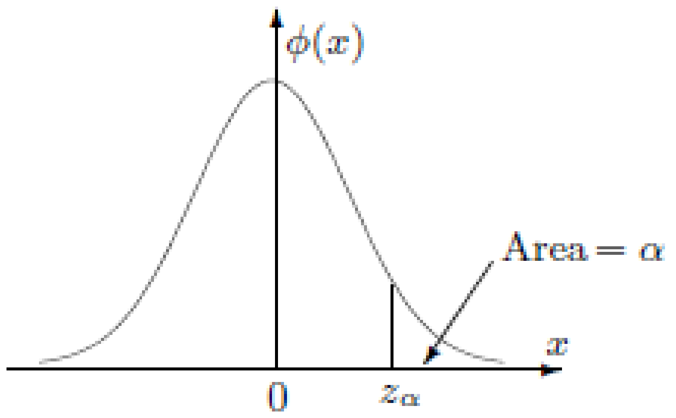
Finally:

approach 1:

$$N \geq p^*(1-p^*)\left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2$$

approach 2:

$$N \geq 0.25\left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2$$



Why we approximate by Normal distribution? Why it leads to reasonable results?

This is a very frequent approach in many situations!

Central Limit Theorem : behavior of the sum of independent random variables:

$$S_n = X_1 + \ldots + X_n,$$

Let $\mu = \mathrm{E}(X_i)$ and $\sigma = \mathrm{Std}(X_i)$ for all $i = 1, \ldots, n$.

$$\mathrm{Var}(S_n) = n\sigma^2 \to \infty,$$

$$\mathrm{Var}(S_n/n) = \mathrm{Var}(S_n)/n^2 = n\sigma^2/n^2 = \sigma^2/n \to 0,$$

$$
\begin{aligned}
\mu &= \text{expectation, location parameter} \\
\sigma &= \text{standard deviation, scale parameter} \\
f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty \\
\mathbf{E}(X) &= \mu \\
\mathrm{Var}(X) &= \sigma^2
\end{aligned}
$$

Theorem 1 (Central Limit Theorem) *Let $X_1, X_2, \ldots$ be independent random variables with the same expectation $\mu = \mathbf{E}(X_i)$ and the same standard deviation $\sigma = \mathrm{Std}(X_i)$, and let*

$$S_n = \sum_{i=1}^{n} X_i = X_1 + \ldots + X_n.$$

As $n \to \infty$, the standardized sum

$$Z_n = \frac{S_n - \mathbf{E}(S_n)}{\mathrm{Std}(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a Standard Normal random variable, that is,

$$F_{Z_n}(z) = \boldsymbol{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right\} \to \Phi(z) \qquad (4.18)$$

for all z.

Important properties:

- it does not matter which probability distribution has X
the result is always the normal distribution

- convergence is strong: "in probability"

Important properties:

Proof: there are elementary ones but ... an elegant and really convincing argument is the one with generating functions

idea: transformation to a strange form of a power series where:

-- the first coefficient is zero (as the expected value of normalized X is 0)

-- the 2nd coefficient does not disappear and is normalized

-- the higher coefficients converge to 0 with N

-- for normal distribution, everything disappears right away

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 Expectation and Variance:

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

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

Expected value:

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

So we have an ***unbiased estimator***

Variance:

The situation is more complicated:

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

But we need to compute variance $\mathrm{Var}X_1$ **, …**

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

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

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

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

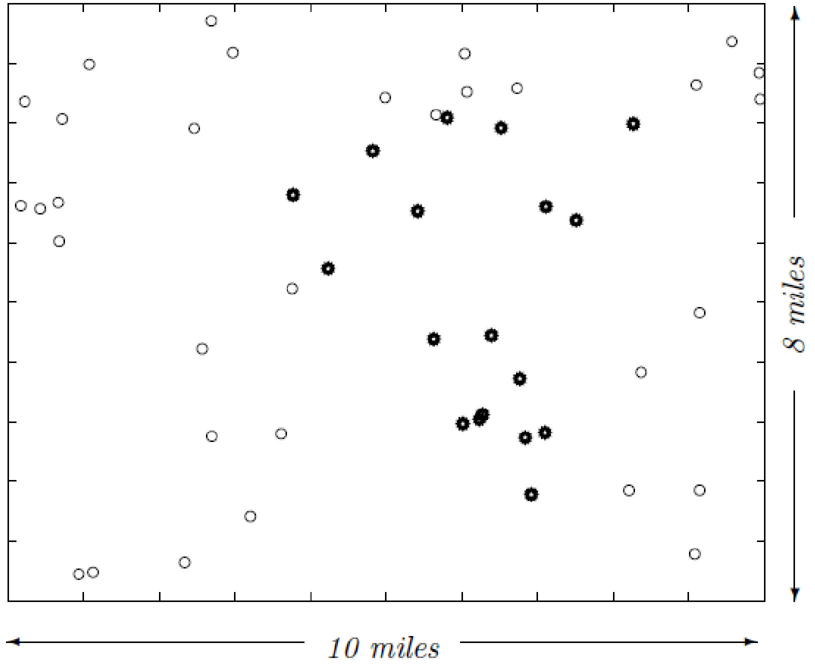
Example: maximal matchings in a graph G that contain an edge (u,v)

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

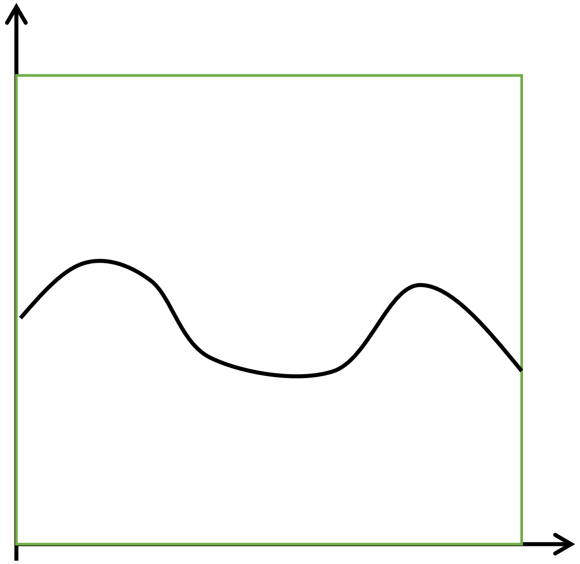
Volume of A = E[$\frac{1}{N}(X_1 + \ldots + X_N)$]

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:

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

```
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

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

↑

$$g^2 \le g \text{ for } 0 \le g \le 1.$$