Probability and statistics, 2021, Computer Science Algorithmics,
Undergraduate Course, Part II, lecturer: Mirosław Kutyłowski

I. Generating Random Numbers for a given probability distribution

Chapter 5.2 in Byron

Goal:

- simulations (e.g., pharma industry, weather forecast, system testing ...)

Weather simulations:

Simulations for new chemical products, pharmaceuticals:

Physical sources: examples:

1) Electronics (bistable)

3) quantum generators

3) noise

Problems of physical sources:

1) bias

2) memory: dependence on history

3) external influence

deterministic random number generators:

DRNG: NIST, recommendations,

architecture: PRNG(seed) yields: bits

DRNG:

basic property: not distinguishable from coin flipping

what does it mean?:

NIST tests

left-or-right game

secure PRNG:

unpredictability forwards:

backwards:

realizations:

families of PRNG (based on residual arithmetic and algebraic expressions

$$a \cdot x^2 + b \cdot x + c \bmod p$$

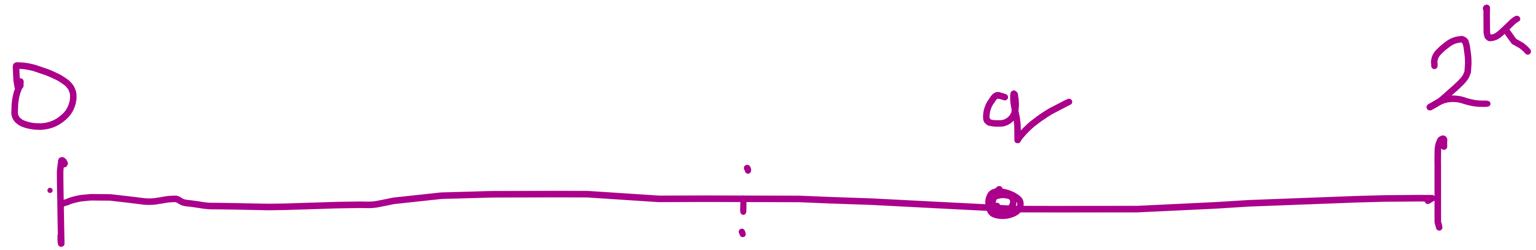
cryptographic generators: e.g. based on encryption

$$trunc(32, Enc_K(1)),\ trunc(32, Enc_K(2)),\ .\ .\ .$$

Problem: domain

We have a good generator for uniform distribution over n bit numbers

How to get a uniform distribution over integers in the range [0,q)?



1. choose $x \in [0, 2^k)$ at random
2. if $x \geq q$ then goto 1
3. output(x)

Problem: uniform versus non-uniform distribution

all good PRNG resources deliver the output that is uniform over some interval
e.g.: 32 bit nonnegative integers

Needed:
e.g. geometric distribution, Poisson, ...

single Bernoulli trial:

procedure:

1. **choose u uniformly at random in [0,1]**
2. **if u<p then output 0 else output 1**

n Bernoulli trials, number of successes:

n times:
    0: with probability p
    1: with probability 1-p
count the number of successes

stupid solution:
compute pbb according to formulas
...

n Bernoulli trials, number of successes:

n times:

0: with probability p

1: with probability 1-p

count the number of successes

procedure (in MATLAB):

```
n=20; p=0.34;
U=rand(n,1);
X=sum(U<p)
```

geometric distribution:

Bernoulli trials with pbb p of 0
output: the number of trials until 1 chosen

naive way: take mathematical formulas and then choose according to the probabilities

procedure (in MATLAB):

```
X=1;
while rand<p
X=X+1;
end;
X
```

arbitrary discrete distribution:

assume: n possible values, p(i) -probability of the ith value

approach: for each i=<n compute

$$a_i = \sum_{j<i} p_j$$

the results saved in a data structure D

procedure:

1. **u=random;**
2. **with D find i such that $a_i \leq u < a_{i+1}$**

Poisson distribution:

$Pr(X=i) = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$

Mat

```
lambda  =  5;                    % Parameter
U       =  rand;                 % Generated Uniform variable
i       =  0;                    % Initial value
F       =  exp(-lambda);         % Initial value, F(0)
while (U >= F);                  % The loop ends when U < F(i)
    F = F + exp(-lambda) * lambda^i/gamma(i+1);
    i = i + 1;
end;
X=i
```

the case of invertible CDF

Theorem 2 *Let X be a continuous random variable with cdf $F_X(x)$. Define a random variable $U = F_X(X)$. The distribution of U is Uniform(0,1).*

PROOF: First, we notice that $0 \leq F(x) \leq 1$ for all x, therefore, values of U lie in $[0, 1]$. Second, for any $u \in [0, 1]$, find the cdf of U,

$$
\begin{aligned}
F_U(u) &= \boldsymbol{P}\{U \leq u\} \\
&= \boldsymbol{P}\{F_X(X) \leq u\} \\
&= \boldsymbol{P}\{X \leq F_X^{-1}(u)\} && \text{(solve the inequality for } X\text{)} \\
&= F_X(F_X^{-1}(u)) && \text{(by definition of cdf)} \\
&= u && (F_X \text{ and } F_X^{-1} \text{ cancel})
\end{aligned}
$$

the case of invertible CDF - continuous distribution X

Procedure:

1. choose u uniformly at random in [0,1]

2. take

$$x := F_X^{-1}(u)$$

Example

Exponential distribution $F(x) = 1 - e^{\lambda x}$

Procedure:

1. **choose** u **uniformly at random in** $[0,1]$

2. **solve** $u = 1 - e^{\lambda x}$

that is $1 - u = e^{\lambda x}$

$$\ln(1-u) = \lambda x$$
$$x = \ln(1-u)/\lambda$$

Or simply $x = \ln(u)/\lambda$

Example –warning

Gamma distribution has complicated density function

$$F(t) = \int_0^t f(x)dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^t x^{\alpha-1} e^{-\lambda x} dx.$$

Inverting *F*?

Workaround:

a random variable with Gamma distribution α is is a sum of α independent random variables with exponential distribution

---

the case of invertible CDF - discrete distribution X

Procedure:

1. **choose u uniformly at random in [0,1]**
2. **take**

$$x := \min\{x : F(x) > u\}$$

so: $x =: F^{-1}(u)$

Example: geometric distribution

$$F(x) = 1 - (1-p)^x,$$

Procedure:
Find the smallest *x* such that

$$1 - (1-p)^x > U$$

Solution:

$$X = \left\lceil \frac{\ln(1-U)}{\ln(1-p)} \right\rceil$$

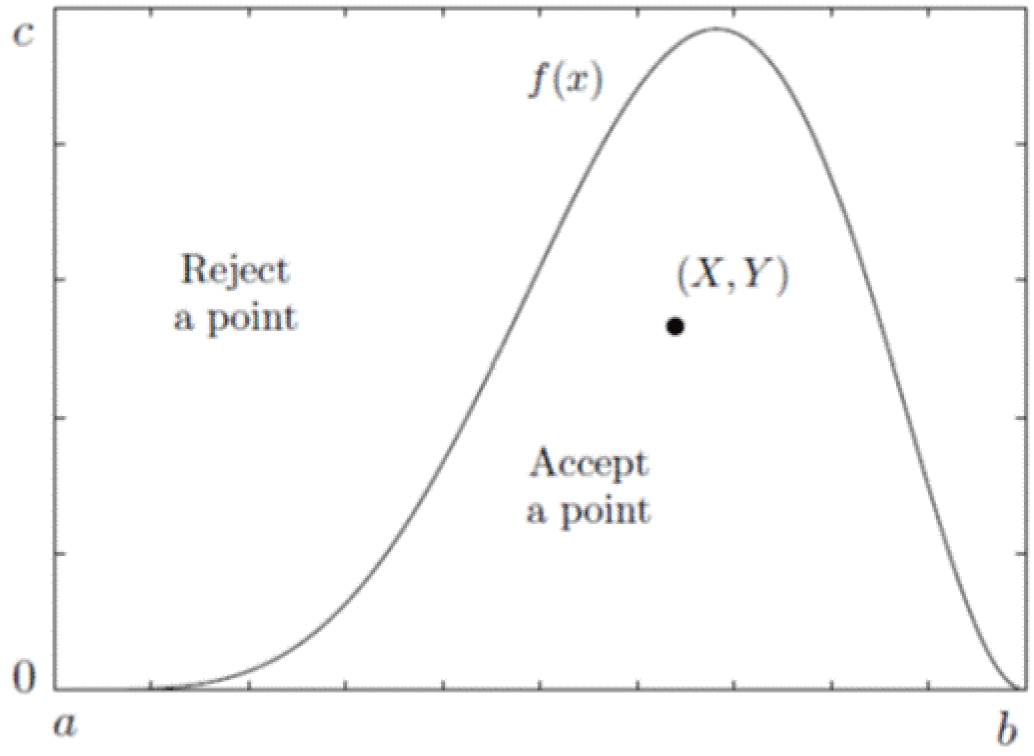
Rejection method:

1. sample ***X*** and ***Y*** uniformly at random

2. if **Y>f(X)**, then goto 1

3. Output **X**

then density of ***X = f(X)*** !



Pict. from Byron

Application:

distributions where density is computable (e.g. Beta distribution)

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } 0 \le x \le 1.$$

but computing cdf is hard (numeric computations of the integral)

```
alpha=5.5; beta=3.1; a=0; b=1; c=2.5;
X=0; Y=c;                    % Initial values
while Y > gamma(alpha+beta)/gamma(alpha)/gamma(beta)...
          * X.^(alpha-1) .* (1-X).^(beta-1);
   U=rand; V=rand; X=a+(b-a)*U; Y=c*V;
end; X
```

Poisson distribution

An example of a clever approach tailored to the particular case

$$P(x) \quad = \quad e^{-\lambda}\frac{\lambda^x}{x!}, \; x = 0, 1, 2, \ldots$$

but.. also can be understood as the number of rare events in an interval of time, where the time between events is exponential

Pragmatic computation:

1. Obtain Uniform variables $U_1, U_2, \ldots$ from a random number generator.
2. Compute Exponential variables $T_i = -\frac{1}{\lambda}\ln(U_i)$.
3. Let $X = \max\{k : T_1 + \ldots + T_k \leq 1\}$.