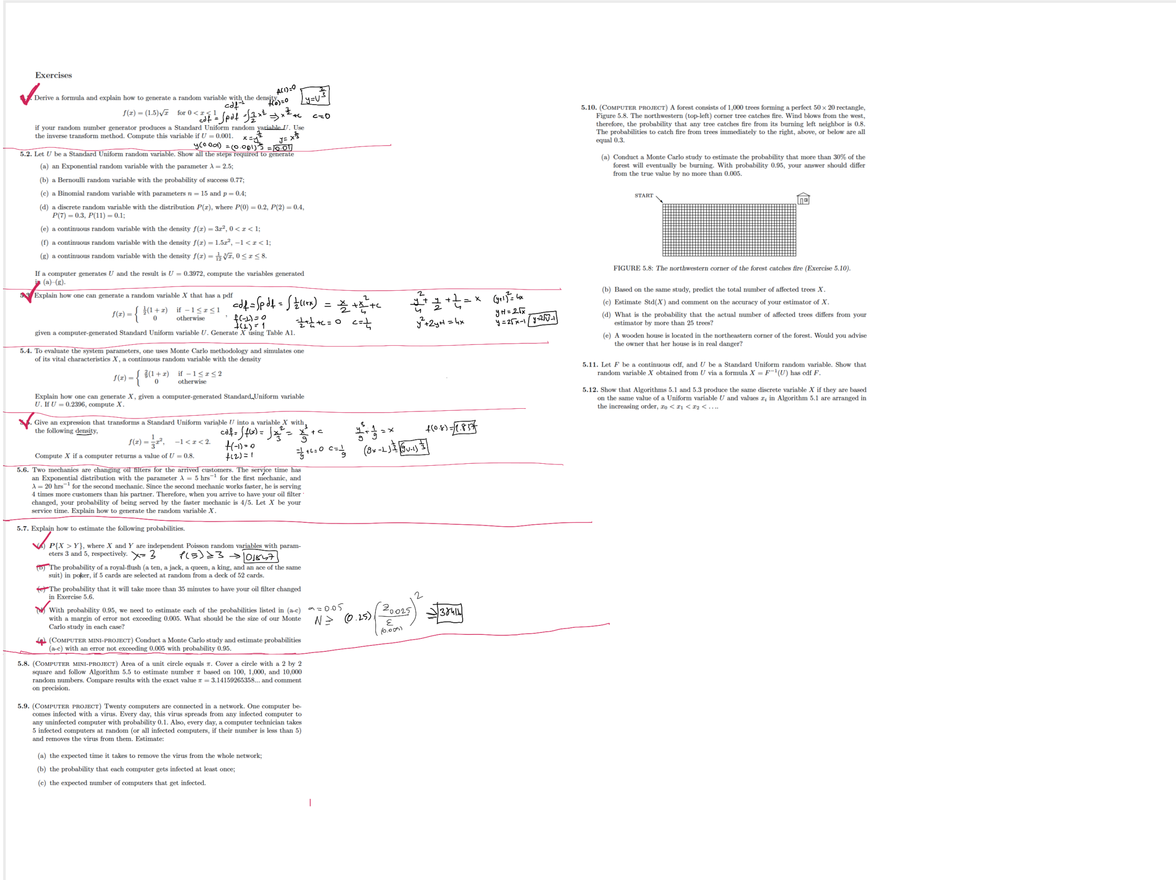
4 - monte carlo method



Exercises

5.1. Derive a formula and explain how to generate a random variable with the density

$$f(x) = (1.5)\sqrt{x} \quad \text{for } 0 < x < 1$$

if your random number generator produces a Standard Uniform random variable U. Use the inverse transform method. Compute this variable if $U = 0.001$.

cdf^{-1} $f(1)=0$ $f(0)=0$ $\boxed{y=U^{\frac{2}{3}}}$
$cdf = \int pdf = \int \frac{3}{2}x^{\frac{1}{2}} \Rightarrow x^{\frac{3}{2}} + c$ $c=0$
$x = y^{\frac{3}{2}}$ $y = x^{\frac{2}{3}}$
$y(0.001) = (0.001)^{\frac{2}{3}} = \boxed{0.01}$

5.2. Let U be a Standard Uniform random variable. Show all the steps required to generate

(a) an Exponential random variable with the parameter $\lambda = 2.5$;

(b) a Bernoulli random variable with the probability of success 0.77;

(c) a Binomial random variable with parameters $n = 15$ and $p = 0.4$;

(d) a discrete random variable with the distribution $P(x)$, where $P(0) = 0.2$, $P(2) = 0.4$, $P(7) = 0.3$, $P(11) = 0.1$;

(e) a continuous random variable with the density $f(x) = 3x^2$, $0 < x < 1$;

(f) a continuous random variable with the density $f(x) = 1.5x^2$, $-1 < x < 1$;

(g) a continuous random variable with the density $f(x) = \frac{1}{12}\sqrt[3]{x}$, $0 \le x \le 8$.

If a computer generates U and the result is $U = 0.3972$, compute the variables generated in (a)–(g).

5.3. Explain how one can generate a random variable X that has a pdf

$$f(x) = \begin{cases} \frac{1}{2}(1+x) & \text{if } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases},$$

given a computer-generated Standard Uniform variable U. Generate X using Table A1.

$cdf = \int pdf = \int \frac{1}{2}(1+x) = \frac{x}{2} + \frac{x^2}{4} + c$
$f(-1) = 0$ $f(1) = 1$ $-\frac{1}{2} + \frac{1}{4} + c = 0$ $c = \frac{1}{4}$
$\frac{y^2}{4} + \frac{y}{2} + \frac{1}{4} = x$ $y^2 + 2y + 1 = 4x$
$(y+1)^2 = 4x$ $y+1 = 2\sqrt{x}$ $y = 2\sqrt{x} - 1$ $\boxed{y = 2\sqrt{U} - 1}$

5.4. To evaluate the system parameters, one uses Monte Carlo methodology and simulates one of its vital characteristics X, a continuous random variable with the density

$$f(x) = \begin{cases} \frac{2}{9}(1+x) & \text{if } -1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Explain how one can generate X, given a computer-generated Standard Uniform variable U. If $U = 0.2396$, compute X.

5.5. Give an expression that transforms a Standard Uniform variable U into a variable X with the following density,

$$f(x) = \frac{1}{3}x^2, \quad -1 < x < 2.$$

Compute X if a computer returns a value of $U = 0.8$.

$cdf = \int f(x) = \int \frac{x^2}{3} = \frac{x^3}{9} + c$
$f(-1) = 0$ $f(2) = 1$ $-\frac{1}{9} + c = 0$ $c = \frac{1}{9}$
$\frac{y^3}{9} + \frac{1}{9} = x$ $(9x-1)^{\frac{1}{3}}$ $\boxed{(9U-1)^{\frac{1}{3}}}$
$f(0.8) = \boxed{1.837}$

5.6. Two mechanics are changing oil filters for the arrived customers. The service time has an Exponential distribution with the parameter $\lambda = 5$ hrs^{-1} for the first mechanic, and $\lambda = 20$ hrs^{-1} for the second mechanic. Since the second mechanic works faster, he is serving 4 times more customers than his partner. Therefore, when you arrive to have your oil filter changed, your probability of being served by the faster mechanic is 4/5. Let X be your service time. Explain how to generate the random variable X.

5.7. Explain how to estimate the following probabilities.

(a) $P\{X > Y\}$, where X and Y are independent Poisson random variables with parameters 3 and 5, respectively. $X = 3$ $P(5) \ge 3 \rightarrow \boxed{0.1847}$

(b) The probability of a royal-flush (a ten, a jack, a queen, a king, and an ace of the same suit) in poker, if 5 cards are selected at random from a deck of 52 cards.

(c) The probability that it will take more than 35 minutes to have your oil filter changed in Exercise 5.6.

(d) With probability 0.95, we need to estimate each of the probabilities listed in (a-c) with a margin of error not exceeding 0.005. What should be the size of our Monte Carlo study in each case?

$\alpha = 0.05$ $N \ge (0.25)\left(\frac{z_{0.025}}{\varepsilon_{(0.005)}}\right)^2 \Rightarrow \boxed{38416}$

(e) (Computer mini-project) Conduct a Monte Carlo study and estimate probabilities (a-c) with an error not exceeding 0.005 with probability 0.95.

5.8. (Computer mini-project) Area of a unit circle equals π. Cover a circle with a 2 by 2 square and follow Algorithm 5.5 to estimate number π based on 100, 1,000, and 10,000 random numbers. Compare results with the exact value $\pi = 3.14159265358...$ and comment on precision.

5.9. (Computer project) Twenty computers are connected in a network. One computer becomes infected with a virus. Every day, this virus spreads from any infected computer to any uninfected computer with probability 0.1. Also, every day, a computer technician takes 5 infected computers at random (or all infected computers, if their number is less than 5) and removes the virus from them. Estimate:

(a) the expected time it takes to remove the virus from the whole network;

(b) the probability that each computer gets infected at least once;

(c) the expected number of computers that get infected.

5.10. (Computer project) A forest consists of 1,000 trees forming a perfect 50×20 rectangle, Figure 5.8. The northwestern (top-left) corner tree catches fire. Wind blows from the west, therefore, the probability that any tree catches fire from its burning left neighbor is 0.8. The probabilities to catch fire from trees immediately to the right, above, or below are all equal 0.3.

(a) Conduct a Monte Carlo study to estimate the probability that more than 30% of the forest will eventually be burning. With probability 0.95, your answer should differ from the true value by no more than 0.005.

START

FIGURE 5.8: *The northwestern corner of the forest catches fire (Exercise 5.10).*

(b) Based on the same study, predict the total number of affected trees X.

(c) Estimate Std(X) and comment on the accuracy of your estimator of X.

(d) What is the probability that the actual number of affected trees differs from your estimator by more than 25 trees?

(e) A wooden house is located in the northeastern corner of the forest. Would you advise the owner that her house is in real danger?

5.11. Let F be a continuous cdf, and U be a Standard Uniform random variable. Show that random variable X obtained from U via a formula $X = F^{-1}(U)$ has cdf F.

5.12. Show that Algorithms 5.1 and 5.3 produce the same discrete variable X if they are based on the same value of a Uniform variable U and values x_i in Algorithm 5.1 are arranged in the increasing order, $x_0 < x_1 < x_2 < \ldots$.

- generating random variables

<u>1- discrete distributions</u>

```
#include <iostream>
#include <stdlib.h>
#include <time.h>
```

- uniform(0,1):

rand()%100000/100000.0

- bernoulli:

\+ 1s and 0s
p=0.8
(rand()%100000/100000.0)<p
matlab: U = rand; X = (U < p)
O(1)

- binomial:

+ sum of n independent bernoulli
+ # of succeses in n -> basically random numbers [0,n]
for (i=0;i<n;i++) {if ((rand()%100000/100000.0)<p) sum++; }
matlab: n = 20; p = 0.68; U = rand(n,1); X = sum(U < p)
O(n)

- geometric:

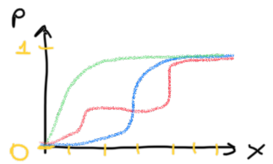
+ # of trials until first success
+ if p is too small result is high-> it can be infinite as well
+ results around 1/p
first_succ=1;   while ((rand()%100000/100000.0)>p)   first_succ++;
matlab: X = 1; while rand > p; X = X+1; end; X
O(1/p) is average, it can be infinite

- negative binomial:

+ sum of n independent geometric
+ results around 1/p * k
int i,first_succ, k=20 , sum = 0; srand(time(NULL)); double p =0.1; for (i=0 ;i<k;i++) { first_succ=1; while ((rand()%100000/100000.0)>p) first_succ++; sum += first_succ; }
matlab:
O(k/p) on average

2- continuous distributions

inverse transform method



→ cumulative distribution function
↳ non-decreasing between [0,1]
$cdf(x) = u \ (0 \le u \le 1)$
↓
$\boxed{\text{theorem}}$ distribution of u is uniform
does not depend on x

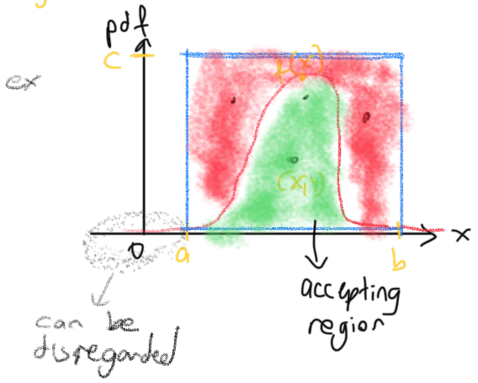
so
$cdf^{-1}(u) = x$ → will give you a number that is inside your distribution

matlab:
pkg load statistics
norminv(rand,180,3)
rand -> x, random numbers

180 -> average
10 -> standard deviation

rejection method → when inverse of cdf is not available, and we have pdf

ex



- find a, b, c such that $0 \le f(x) \le c$, $a \le x \le b$
- generate (X, Y) random uniform such that $a \le X \le b$, $0 \le Y \le c$

↳ if $Y > f(X)$ reject, else accept

monte carlo methods

generate many random variables from a distribution and estimate μ, σ, p..

Monte carlo estimators → $E(\hat{p}) = p$
↳ $Std(\hat{p}) = \sqrt{\frac{p(1-p)}{N}}$

- estimate p^* (intelligent guess)

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

Φ → inverse z

$z_\alpha = \Phi^{-1}(1-\alpha)$

- bound $p(1-p)$ by largest possible value → 0.25 if it is not given

$0 \le p \le 1$ → $p(1-p) \le 0.25$

$$2\Phi(-2\varepsilon\sqrt{N}) \le \alpha$$

size of N

→ to simulate $P\{|\hat{p} - p| > \varepsilon\} \le \alpha$

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

→ 0.25 → if p^* is preliminary estimator of p

Example 5.14 (SHARED COMPUTER). The following problem does not have a simple analytic solution (by hand), therefore we use the Monte Carlo method.

different p than what is wanted

A supercomputer is shared by 250 independent subscribers. Each day, each subscriber uses the facility with probability 0.3. The number of tasks sent by each active user has Geometric distribution with parameter 0.15, and each task takes a Gamma(10, 3) distributed computer time (in minutes). Tasks are processed consecutively. What is the probability that all the tasks will be processed, that is, the total requested computer time is less than 24 hours? Estimate this probability, attaining the margin of error ± 0.01 with probability 0.99.

ε

$\alpha = 1 - 0.99 = 0.01$

$N \geq 0.25 \left(\frac{z_{\frac{0.01}{2}}}{0.01} \right)^2$

$z = 0.05 \rightarrow \left(\frac{2.5758}{0.01} \right)^2 \cdot 0.25$

$N \geq 16587.2$

$E(\bar{x}) = \mu$

$var(\bar{x}) = \frac{\sigma^2}{N}$