STOCHASTIC SIMULATION

Topics for LABS

Author:
Wojciech NIEMIRO

Warszawa, May 21, 2017

1 Introduction. Probability distributions

General remarks: One of the objectives of these excercises is to better understand Probability and Statistics via Simulations. Don't underestimate theoretical questions. If you are more advanced then help your colleagues. If you do not understand then ask. Remember that R is NOT C++ and try to code in a way recommended for R. Loops and if's should be replaced by clever functions and vector arithmetic (whenever possible). Please use similar notation, if suggested. I (almost) always denote random quantities in capital letters, and non-random in lower case. At the end of the labs, remember to save what is worth saving.

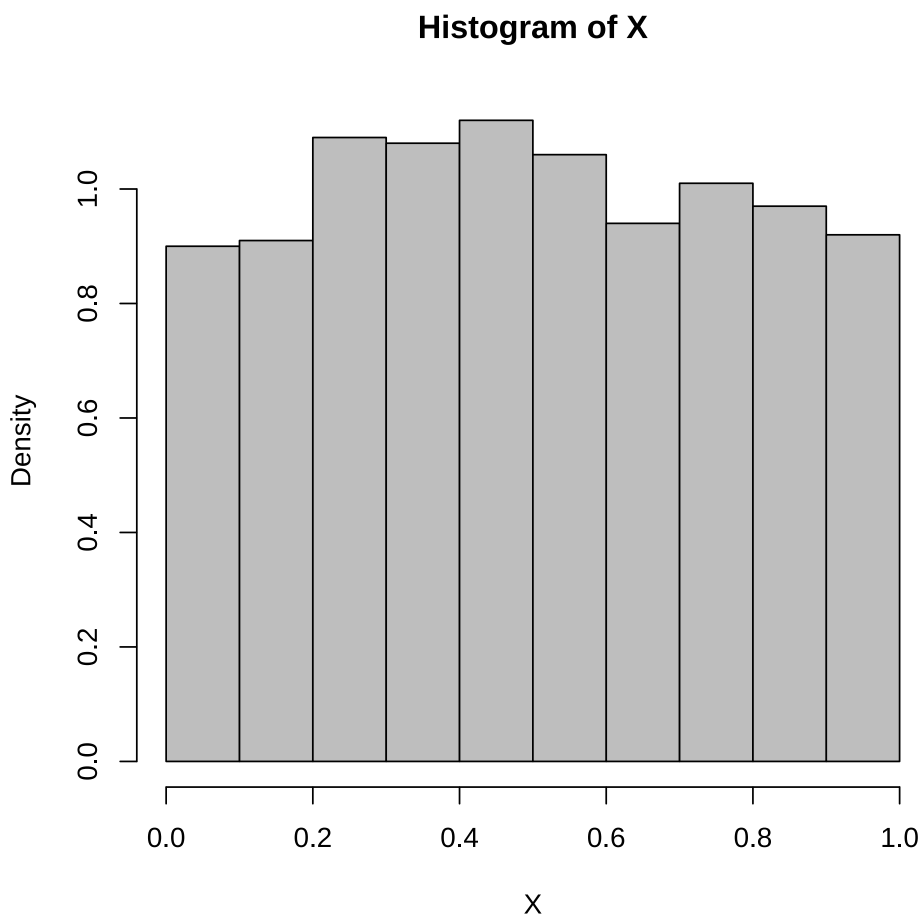
1. Make acquaintance with R (if you are not already familiar with R). Maybe R Studio? Find "Introduction to R". Learn how to use online help (?functionname).

2. Learn how to use built-in random generators, eg.

```
> n <- 10
> (X <- runif(n))
```

```
 [1] 0.50592413 0.02790791 0.16497038 0.16782016 0.44836931 0.0738059
 [7] 0.27427970 0.73863304 0.01180299 0.53957905
```

```
> n <- 1000
> X <- runif(n)
> hist(X,prob=TRUE,col="gray")
```



3. Choose one of the standard probability distributions (eg. Gamma or Normal). Draw a large sample (n iid random variables). Make a histogram. Superimpose the graph of the density (PDF) eg. via

```
> ?curve
```

Make a graph of the ECDF (what is this?) via

```
> ?ecdf
```

Superimpose the graph of the CDF. Compute empirical mean and variance and compare with known formulas. Experiment with various n.

4. Verify that your sample is indeed from the desired distribution. Use Kolmogorov-Smirnov test

```
> ?ks.test
```

Generate $m = 10000$ samples of size $n = 100$ and apply Kolmogorov-Smirnov test to every sample. Collect p-values and analyse them. What is the probability distribution of p-value, if the null hypothesis is true?

5. Similar exercise with a discrete distribution. Roll a die $n = 1000$ times. You can use

```
> ?sample
```

or just discretize Uniform. Analyse the results. Count the results and make a graph. The suggested functions are

```
> ?table
> ?barplot
```

Use the goodness-of-fit χ^2 test. Compute the test statistic and p-value using known formulas (Statistics 1). Compare with those automatically computed by

```
> ?chisq.test
```

6. Generate a large sample $X_1, \dots, X_n, \dots$ from a Uniform distribution. Make a graph of S_n/n versus n, where $S_n = X_1 + \cdots + X_n$. Hint: functions

```
> ?cumsum
> ?plot
```

What result do you expect? Recall SLLN (Probability 1).

7. Generate a large number m of samples of (moderate) size n from a Uniform distribution. Make a histogram of m "independent copies" of S_n. What result do you expect? Recall CLT (Probability 1). Superimpose a graph of appropriate Normal density.

8. Write a function which computes $X = U_1 + \cdots + U_{12} - 6$, where U_is are iid Uniform$(0, 1)$ (by now you should be able to write this function as a single line). Why $n = 12$? Generate n copies of X and analyse.

9. Generate a large number m of samples of size n (odd) from a Uniform distribution. Make a histogram of m medians. What result do you expect? Superimpose a graph of appropriate density (Statistics 1). What is the expectation and the variance of the median? (Statistics 1 or Wiki).

10. What is the limit distribution of the median if $n \to \infty$? Illustrate graphically. GUESS the theorem: if $U_1, \dots, U_n$ are iid $\text{Uniform}(0,1)$ then

$$b_n(\text{med}(U_1, \dots, U_n) - a) \to \text{Normal}(0,1).$$

Find appropriate a and b_n. "Confirm" experimentally.

11. Generate a large number m of samples of size n from a Uniform distribution. Make a histogram of m minima. What result do you expect? Superimpose a graph of appropriate density (Statistics 1). What is the expectation and the variance of the minimum? (Statistics 1 or Wiki). PROVE the theorem: if $U_1, \dots, U_n$ are iid $\text{Uniform}(0,1)$ then

$$b_n \min(U_1, \dots, U_n) \to \text{ what?}$$

Find appropriate b_n. "Confirm" experimentally.

12. • GUESS the theorem: if $U_1, \dots, U_n$ are iid $\text{Uniform}(0,1)$ then

$$b_n(\min(U_1, \dots, U_n) + 1 - \min(U_1, \dots, U_n)) \to \text{ what?}$$

Find appropriate b_n. "Confirm" experimentally.

• The same for

$$b_n(\min(U_1, \dots, U_n) - 1 + \min(U_1, \dots, U_n)) \to \text{ what?}$$

13. Let $X_1, \dots, X_n$ be iid with $\mathbb{E}X_i = 0$ (choose such a distribution to your taste) and $S_n = X_1 + \cdots + X_n$. Define

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(S_i > 0).$$

What is the "meaning" of T_n in plain English (or in plain Polish: po prostu...)? Make a graph of a "sample path" $(0, S_1, S_2, \dots, S_n)$ and look at the corresponding value of T_n. Explore experimentally the probability distribution of T_n for large n. GUESS the theorem: if $n \to \infty$ then

$$T_n \to \text{ what?}$$

Remarks: If you do not know the answer, you are not likely to guess without the help of simulation (well, if you can guess then you are exceptionally bright). Simulation helps. By now, you should be accustomed to the methodology of exploring a probability distribution via simulation: recall the previous questions.

2 Rejection sampling

1. (A simple Bayesian model) Consider the following 2-stage sampling: first draw $\theta \sim \text{Uniform}(0,1)$ and then $X \sim \text{Bin}(n, \theta)$. θ is a parameter of the probability distribution of observed random variable X, and is itself a random variable with the uniform prior distribution. Choose eg. $n = 9$.

   • Discover experimentally the marginal distribution of X. Compute this distribution analytically.

   • Discover experimentally the posterior distribution of θ given $X = 3$ (say), using ABC. Compute the posterior analytically. Compare the empirical and theoretical distribution.

- Try the 2-stage sampling in the opposite order: first draw X from the marginal and then θ from the posterior. Verify that you get θ distributed according to the prior.

Remark: ABC means "Approximate Bayesian Computation". This is essentially a straightforward application of rejection sampling in Bayesian models. One samples θ from the *prior*, then samples X and uses rejection method to obtain samples from the *posterior* (by selecting those θ which correspond to $X = x$).

2. The following algorithm is a simple example of rejection sampling:

```
repeat
  Gen X ~ U(0,1)
  Gen U ~ U(0,1)
until U < X
return X
```

What is the distribution of X at the output?

Someone (a rather prominent someone) proposed a "more efficient version" of this algorithm:

```
Gen U ~ U(0,1)
repeat
  Gen X ~ U(0,1)
until U < X
return X
```

What is the distribution of X at the output? Compute theoretically and check the answer experimentally.

3. Let $X_1, \ldots, X_n, \ldots$ be iid with probability distribution F. Define times and values of *records* by $T_1 = 1$, $R_1 = X_1$, $T_k = \min\{n > T_{k-1} : X_n > R_{k-1}\}$ and $R_k = X_{T_k}$. Generate a sequence (having the joint distribution equal to that) of record values $R_1, \ldots, R_k$ for a big k.

- Using brute force, for $F = \mathrm{Ex}(1)$ (exponential distribution).
- In a civilized way, for the exponential F.
- For $F = \mathrm{Uniform}(0, 1)$, using transformation of the previous sequence.
- Compute 1-dimensional cdf $\mathbb{P}(R_k \leq x)$ theoretically (using induction) and compare with the results of your experiment.

3 Composition, alias and other special methods

1. Write a function which generates n iid variates with the Laplace distribution,

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}.$$

You may use composition or put $X = W_1 - W_2$, W_i iid exponential.

2. Write a function which generates n iid variates with the distribution with density given by

$$f(x) \propto \int_1^\infty y^{-n} e^{-xy} dy.$$

You may use composition: generate (X, Y) with density $f(x, y) \propto y^{-n} e^{-xy}$. As a by-product compute the norming constant.

3. (*Alias*) Consider the following algorithm.

```
Gen X ~ U(0,1)
Gen U ~ U(0,1)
if U < X then return X else return 1 - X
```

What is the probability distribution at the output? Write a code in R and verify the answer experimentally.

4. (*Theoretical question*) Prove that if $W_1, \dots, W_k$ are independent and have exponential distributions $X_i \sim \mathrm{Ex}(\lambda_i)$ then $\min(W_1, \dots, W_k) \sim \mathrm{Ex}(\sum \lambda_i$ and if J is the subscript for which the minimum is attained then $\mathbb{P}(J = j) = \lambda_j / \sum \lambda_i$. *Hint:* First consider the case $k = 2$.

4 Generating multidimensional random variables

1. The famous Marsaglia's algorithm for generating from the normal distribution is the following:

```
repeat
  Gen U,V ~ U(-1,1)
  R^2 := U^2 + V^2
until R^2 < 1
R := ···?···
X := RU; Y := RV
return (X,Y)
```

- Complete the missing line "$R := \cdots ? \cdots$". At the output we should obtain two independent normally distributed variables.
- Write an R code. Check correctness (1-dim marginals, Kolmogorov-Smirnov test, moments).
- Plot $(X_i, Y_i) \sim N(0, 1)$, $i = 1, \dots, n$. Superimpose contours of the 2-dim normal density. Use functions `contour` and `outer`.

2. Let X, Z be independent $N(0, 1)$ and $Y = X + 2Z$. Generate a sample from the probability distribution of (X, Y). Use e.g. the method of conditional distributions. Plot (X_i, Y_i), $i = 1, \dots, n$. Superimpose contours of the 2-dim normal density. Plot the graph of $r(x) = \mathbb{E}(Y|X = x)$. Compare with the results of applying function `lsfit` to the sample.

3. Write a function `rmult.norm(n,V)` which produces n iid random vectors, each with d-dim normal distribution $\mathrm{N}(0, V)$, where V is a $d \times d$ variance-covariance matrix. Redo the previous excercise using your new function.

Hint: there is a function `chol`. The most convenient format of the output is a `matrix` of dimension $n \times d$. You can transform the whole matrix using one matrix multiplication `%*%`.

Remark: There *is* a built-in function `mvrnorm`. You can compare with your function.

4. Let $\Phi \sim \text{Uniform}(0, 2\pi)$, $\Psi \sim \text{Uniform}(-\pi, \pi)$ and let

$$X = \cos\Psi\cos\Phi; Y = \cos\Psi\sin\Phi; Z = \sin\Psi.$$

Check experimentally if (X, Y, Z) is uniformly distributed on the 3-dim sphere or not.

5. Let $\Phi \sim \text{Uniform}(0, 2\pi)$, $Z \sim \text{Uniform}(-1, 1)$ and let

$$X = \sqrt{1 - Z^2}\cos\Phi; Y = \sqrt{1 - Z^2}\sin\Phi; Z = \sin\Psi.$$

Check experimentally if (X, Y, Z) is uniformly distributed on the 3-dim sphere or not.

6. Write a function which generates n iid wectors with the uniform distribution on the $(d-1)$-dimensional sphere.

7.
   - Generate a large sample of 3-dim vectors (X_1, X_2, X_3) with uniform distributon over S^2. Look at 1-dim marginals. What is the distrubution of, say X_1?
   - Generate a large sample of 4-dim vectors (X_1, X_2, X_3, X_4) with uniform distributon over S^3. Look at 2-dim marginals. What is the distrubution of, say (X_1, X_2)?
   - Guess the conclusion of the following theorem: If $(X_1, \ldots, X_d) \sim \text{Uniform}(S^{d-1})$ then $(X_1, \ldots, X_{d-2}) \sim \cdots ? \cdots$.

8. Generate a sample (X_i, Y_i), $i = 1, \ldots, n$ from 2-dim Cauhy distribution (Student's t-distribution with 1 df). Compare with a sample (X_i, Y_i), where X_i and Y_i are independent with 1-dim Cauchy. You may also draw contours of the 2-dim densities in both cases.

9. Write a function which generates n iid vectors with Dirichlet d-dim distribution.
   - Use Beta generation.
   - Use Gamma generation.

10. Simulate the vector of order statistics $X_{1:n} \le \cdots \le X_{n:n}$ (ordered sample) from a probability distribution F.
    - Using brute force, for $F = \text{Uniform}(0, 1)$ (uniform distribution).
    - For $F = \text{Ex}(1)$, using transformation of the previous sequence.
    - Compute 1-dimensional pdf $f_{X_{k:n}}(x)$ theoretically and compare with the results of your experiment.
    - Use a Beta generator to simulate some chosen sub-vector of order statistics, e.g. $(X_{2:5}, X_{4:5})$, without generating all order statistics. Compute the covariance $\text{Cov}(X_{2:5}, X_{4:5})$ theoretically. Check experimentally.

Bigger project Write a generator for *Multinomial* distribution using simulation of order statistics, similar to a *Binomial* generator explained in the lecture notes.

Warning: It might be that a reasonable code should be written in C++, and not in R.

5 Markov chains

1. Markov chain on state space $\mathcal{X} = \{1, 2\}$ has the transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

   - Generate a Markov chain trajectory $X_0, X_1, \dots, X_n$ and draw a graph (using e.g. `plot(...,type="s")`. Suggested values $\alpha = 0.35$, $\beta = 0.70$, $n = 100$.
   - Experimentally estimate the stationary distribution π. *Hint:* Use the SLLN for Markov chains and compute π using one long trajectory (suggested value $n = 100000$).
   - Compute π theoretically and compare.
   - Experimentally estimate $\mathrm{Var} M_n = \mathrm{Var}\left(\frac{1}{n}\sum_{i=0}^{n-1} X_i\right)$. Let $\alpha = \beta$, so $\pi(1) = \pi(2)$ in this experiment. Let $X_0 \sim \pi$ so that the chain is strictly stationary. *Hint:* To examine the distribution of M_n, you have to generate this random variable many times, say $m = 10000$ times. So you have to generate m Markov chains of length n each. Try $n = 10, 100, 1000, 2000, \dots$. Take care of the efficiency of your code to ensure that the simulation time is not outrageously long.
   - For $\alpha = \beta = 0$ the theoretical result is obvious. Compare with the the results of simulations.
   - For $\alpha = \beta = 0.2$, check that $\mathrm{Var} M_n \sim \sigma^2_{\mathrm{as}}/n$ for $n \to \infty$ for some constant σ^2_{as} (called "asymptotic variance"). Find an approximate value of σ^2_{as} experimentally. Compare with $\sigma^2_{\mathrm{st}} = \mathrm{Var}_\pi X_1$ (called "stationary variance").
   - The same for $\alpha = \beta = 0.8$.
   - Experimentally verify that

$$\sigma^2_{\mathrm{as}} = \sigma^2_{\mathrm{st}} \frac{1+\lambda}{1-\lambda}, \qquad \lambda = 1 - \alpha - \beta.$$

2. (*The Ehrenfests' model*) A box is divided into 2 parts, with a tiny hole in a divider. There are r particles in the box. Time is discrete: $n = 0, 1, \dots$. At time n one of the particles, chosen at random (with equal probability $1/r$ each) moves through the hole to the other part of the box (if it is in the left part at time n, then it will be in the right part at time $n+1$ and vice versa).

   - Describe the evolution as a Markov chain (the state of the process can be the number of particles in the left part). Write the transition matrix. Simulate the process.
   - What is the stationary distribution of the chain? Guess, compute and verify experimentally.

3. (*Gambler's ruin*) You play a series of games until you are broke or your opponent is broke. In every game you win 1 EUR with probability p or you lose 1 EUR with probability $1-p$. Initially you have a EUR and your opponent has b EUR.

   - Simulate the process. Compute the probability of winning. Compute the probability distribution of the length of game.
   - Compute analytically the probability of winning. Check if the result is consistent with simulation.

4. (*To drown or not to drown?*) Swimming in the ocean, you drift away from a beach at constant speed v. At random moments $T_1, T_2, \ldots$ people on the beach throw a lifeline. The distances the lifeline reaches from the beach are $X_1, X_2, \ldots$ in subsequent trials. Will you eventually catch the lifeline (does one of the X_is exceed your current distance from the beach)? Assume that times T_i are a Poisson process and that $X_1, X_2, \ldots$ are iid variables with probability distribution F.

   - Experiment with various Poisson intensities and probability distributions Fs.
   - What assumption about F ensures that the probability of eventual drowning is zero?

5. Markov chain on state space $\mathcal{X} = \mathbb{R}$ is defined by $X_0 \sim \nu$,

$$X_{n+1} = \alpha X_n + W_{n+1},$$

   where $W_1, W_2, \ldots$ are iid $N(0, 1)$ variables. Repeat the same series of experiments as in the previous example:

   - Generate a Markov chain trajectory $X_0, X_1, \ldots, X_n$ and draw a graph (using e.g. `plot(...,type="l")`. Suggested values $\alpha = 0.3$, $\alpha = 0.95$, $n = 500$.
   - Experimentally estimate the stationary distribution π. *Hint:* Use the SLLN for Markov chains and compute π using one long trajectory (suggested value $n = 100000$).
   - Compute π theoretically and compare.
   - Experimentally estimate $\mathrm{Var} M_n = \mathrm{Var}\left(\frac{1}{n}\sum_{i=0}^{n-1} X_i\right)$. Let $X_0 \sim \pi$ so that the chain is strictly stationary.
   - For $\alpha = 0$ the theoretical result is obvious. Compare with the the results of simulations.
   - For $\alpha = 0.9$, check that $\mathrm{Var} M_n \sim \sigma^2_{\mathrm{as}}/n$ for $n \to \infty$ for and find an approximate value of the "asymptotic variance" σ^2_{as}. Compare with the "stationary variance" $\sigma^2_{\mathrm{st}} = \mathrm{Var}_\pi X_1$.
   - The same for $\alpha = -0.9$.
   - Guess how the asymptotic variance depends on α: Experimentally verify that

$$\sigma^2_{\mathrm{as}} = \sigma^2_{\mathrm{st}} \cdot \frac{1 + \alpha}{1 - \alpha}.$$

6. Markov chain on a state space $\mathcal{X} = [0, 1]$ has the following transition rule. If $X_n = x$ then X_{n+1} is chosen from the uniform distribution on the longer of two intervals: either $[0, x]$ or $[x, 1]$.

   - Simulate this process.
   - Find the stationary distribution (experimentally and theoretically).

6 Markov jump processes

7 Importance sampling

8 MCMC for hierarchical Bayesian models

1. Consider a Bayesian model of variance components in a radically simplified version:

   - Likelihood: $Y_1, \dots, Y_n \sim_{\text{iid}} N(\theta, \sigma^2)$.
   - Prior: $\theta \sim N(\mu, v^2)$ and $\sigma^{-2} \sim Gamma(p, r)$ independent.

   Construct a Gibbs Sampler targeting the posterior $\pi(\theta, \sigma^{-2} | Y_1 = y_1, \dots, Y_n = y_n)$. Suggested data: $n = 5$, sufficient statistics $\bar{y} = 0$, $\sum(y_i - \bar{y})^2 = 5$, hyperparameters $\mu = 5$, $v^2 = 5$, $p = r = 0$ (improper prior for σ^{-2}).

   - Sketch a trajectory of μ.
   - Sketch a trajectory of σ^{-2}.
   - Sketch a 2-dim trajectory of (μ, σ^{-2}).
   - Plot a large 2-dim sample generated by the GS.
   - Plot the level lines of the theoretical target density and compare with the sample. *Hint:* To plot a level lines of a density you don't need to know the norming constant.

2. Find the "`faithful`" data in `R`.

   - Fit a Gaussian mixture model with 2 components to `faithful$waiting` using a Gibbs Sampler.
   - Fit a Gaussian mixture model with 2 components to `faithful$eruptions`.
   - Fit a 2-dim Gaussian mixture model with 2 components to both the variables of `faithful` together.

9 MCMC for Markov random fields

1. Ising model: $\pi(x) \propto \exp[-\beta H(x)]$, where $x = (x_t) \in \{-1, 1\}^{\{1,\dots,d\}^2}$,

   $$H(x) = -J \sum_{s \sim t} x_s x_t + h \sum_s x_s.$$

   Simulate Metropolis chain targeting π. Recall that the acceptance probability for a move $x_s \to -x_s$ is

   $$a_s(x_s, -x_s) = \exp\left[-2\beta \left(x_s \left(J \sum_{t: t \sim s} x_t + h \right) \right)_+ \right].$$

   Code a random scan Metropolis or a systematic scan Metropolis.

   - Code a random scan Metropolis or a systematic scan Metropolis.

- Check correctness. For suggested values $d = 50$, $\beta = 1$, $J = 0.2$, $h = 0$ and "+1 boundary conditions", I obtained the following avarage values of sufficient statistics $S(x) = \sum_s x_s$ and $SN(x) = \sum_{s\sim t} x_s x_t$:

$$\sum_x S(x)\pi(x) \approx 80; \qquad \sum_x SN(x)\pi(x) \approx 1093.$$

- Let's compare the running times for different codes, assuming that every site is visited around 10000 times.
- Derive the formula for a full conditional distribution, $\pi(x_s = 1|x_{-s})$.
- Code a Gibbs Sampler for the Ising model. Compare with the Metropolis.