Stochastic Processes, Session 2. Group Work —

Monte Carlo Simulation of Random Vectors, Expectation and Conditional Expectation

Troels Pedersen, Aalborg University, September, 2016

1 Session 2: Group Exercises

Go through the exercises below. Allow yourself time to reflect over your results and discuss them with other students! Use the book and notes for inspiration and for further information.

1.1 Means and Covariances

The first exercise is for you to practice computation with means and covariances. Consider two random vectors:

$$\mathbf{X} \sim \mathcal{N}\left(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}3 & 1\\1 & 2\end{bmatrix}\right) \qquad \text{and} \qquad \mathbf{Y} \sim \mathcal{N}\left(\begin{bmatrix}-3\\5\end{bmatrix}, \begin{bmatrix}2 & -1\\-1 & 2\end{bmatrix}\right) \tag{1}$$

Assume that $\mathbf{X}$ and $\mathbf{Y}$ are independent and compute the following entities:

$$E[10\mathbf{X}], \qquad E[\mathbf{X}+\mathbf{Y}], \qquad E[\mathbf{X}^T\mathbf{Y}], \qquad E[\mathbf{X}\mathbf{Y}^T],$$

$$E\left[\begin{bmatrix}2 & 2\\0 & 1\end{bmatrix}\mathbf{X}\right], \qquad \text{Cov}\left[\begin{bmatrix}2 & 2\\0 & 1\end{bmatrix}\mathbf{X}\right] \qquad \text{Cov}(\mathbf{X}+\mathbf{Y})$$

Hint: Keep your derivation in vector/matrix form for as long as you possible can. This saves a lot of writing, time, and potential errors. In fact the above exercise can be done without splitting vectors into its entries at all!

1.2 Ten useful Matlab Commands for Monte Carlo Simulation

Familiarize yourself with the following Matlab commands:

```
help   rand       randn   mean      var
plot   histogram  stem    scatter   cdfplot
```

Read the help and try the commands out. Explain what the commands do and how they work to your colleagues. These are just a few of the relevant commands, but you will get far with these!

1.3 Random Numbers?

Restart Matlab (really, close it and open it again). Run `randn(1)` and write down the result. Then restart Matlab and run `randn(1)` once more. Compare and discuss the two results. When could the observed behavior be wanted or unwanted? Read about "seeding" the random number generators in the documentation.

1.4 An Average is not the Mean!

Let $X_1, \ldots, X_N$ be independent Gaussian random variables with zero mean an variance one and define the average

$$S_N = \frac{1}{N}\sum_{n=1}^{N} X_n$$

The average S_N can be used as an estimate of the mean of a random variable.

- Derive an expression for the mean and variance of S_N (recall that $\mathrm{Var}(V + W) = \mathrm{Var}(V) + \mathrm{Var}(W)$ for independent V and W). How do these depend on the number of samples N?
- Simulate S_N for $N = 1, \ldots, 1000$ and plot the result. Notice that this can be done in a simple manner using the Matlab function `mean` in a `for` loop.
- Discuss the result. Do you think that the command name `mean` is appropriate?

1.5 Law of Large Numbers — Monte Carlo Simulation

The law of large numbers tell us that if an expectation exists, we can approximate this by averaging a large number of independent samples[1]:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} g(X_n) = E[g(X)]$$

This can be used to approximate expected values by computer simulation considering a finite N. This approach is called Monte Carlo simulation. Monte Carlo simulation is particularly relevant if expectation is difficult or impossible to carry out. To see how this works out, we test the idea using a toy example, where we actually can compute the expectation by hand and compare this to the simulation.

- Let $X \sim \mathcal{U}(-1, 1)$ and compute the expectation $E[X^3]$ by direct integration.

[1] Mathematicians and careful engineers will need to discuss in which sense the limit is to be interpreted. Actually there are two different laws valid under slightly different assumptions. The weak one where the convergence is in probability and the strong one where the convergence is "almost surely". Essentially, they both say that for very large samples the mean is close to the expectation.

- Now make a Monte Carlo simulation by drawing independent realizations of X^3 and taking the appropriate average. How large should N be to achieve a sufficient accuracy?

- Try to make a similar simulation for another function g for which you cannot compute the mean directly. Try also to redo the experiment with a different distribution for X.

- The term "Monte Carlo simulation" is used for a broad range of methods, including the type of method described above. Wikipedia has a long and informative article about the topic — check this article out!

1.6 Simulation of a Dice

Write a program to simulate a fair dice. This can be done in one line of code using the `rand` function!

1.7 Joint, marginal and conditional pdf

In this exercise, we highlight the relation between the joint, marginal and conditional pdfs. We consider three scalar random variables:

$$X \sim \mathcal{U}(0,3), \qquad Y \sim \mathcal{U}(0,1), \qquad \text{and} \qquad Z = X + Y$$

where X and Y are independent.

- Draw a large number of joint realizations of X, Y and Z. Investigate scatter-plots of the pairs $[X, Y]$ and $[Z, X]$.

- Write up and sketch the following pdfs: $p_X, p_Y, p_{X,Y}$, and $p_{X,Z}$.

- Write up the conditional pdfs $p_{Y|X}$ and $p_{X|Y}$ discuss their form.

- Derive the conditional pdf $p_{Z|X}$. The easiest way to do this is to consider the sketch the joint pdf $p_{X,Z}$ and then consider the graphical interpretation of the conditional random variable $Z|X = x$, namely that we are "given a particular value x of X".

- Compute the conditional expectation $E[Z|X]$ (it is best to do this graphically first and then to formalize). This is an intuitive guess (or estimate) of the value of Z based upon an observation of X.

- Now repeat the previous two questions with $p_{X|Z}$. Note that the form the conditional pdf is in this case more complicated, however, the conditional expectation can be obtained graphically by inspection of your drawing of the joint pdf.