Lab session 2: Prediction

MATH 350.011

Another important use of probabilistic simulation is for prediction. Generally speaking, in a prediction problem, a quantity of interest q is modeled as the output of a multivariate function, that is

$$q = g(X_1, X_2, \ldots, X_m)$$

where $X_1, X_2, \ldots, X_m$ are inputs and g is a known but complicated function.

To make prediction about q, people make measurements of the inputs. The measurements, however, is subject to noise and can not be accurate 100%. Thus, $X_1, X_2, \ldots, X_m$ are usually modeled by random variables.

The central question is: given the uncertainties in $X_1, X_2, \ldots, X_m$, what is the value of q and what is the uncertainty associated with it? The end-product of such prediction is something like Figure 1.

1 Example 1

A farmer wants to know the area of his rectangular field. He asks two probabilists to measure the dimension of the field, to which they did and give him the following summary: Let X, Y be the width and the length of the rectangle, then

$$X \sim \mathcal{N}(30, 9), \qquad Y \sim Exp(0.01)$$

where $\mathcal{N}(30, 9)$ is the normal distribution with mean 30 and variance 9, while $Exp(0.01)$ is the exponential distribution with $\lambda = 0.01$.

Use two R functions *rnorm* and *rexp* to sample 20000 samples of the area of the field $A = X \times Y$. Compute the mean, the standard deviation and produce a histogram of A.

Repeat the tasks with various functions of X, Y

- $g(X, Y) = X^2 + Y^2$
- $g(X, Y) = X^2 Y$

2 Part 2: Prey-predator model

In lots of examples, the function g is so complicated that you don't really know what it does. People refer to such cases as black-box predictions.

In the file *prediction.r*, I wrote a set of codes that implements the LotkaVolterra equations.

https://en.wikipedia.org/wiki/Lotka%E2%80%93Volterra_equations

The LV model describes the dynamics of biological systems in which two species interact, one as a predator and the other as prey.

Assume that the two parameters *Alpha* and *Beta* in the codes are not constant, but follow the following distributions

$$Alpha, Beta \sim \mathcal{N}(0.001, 10^{-6})$$

and that we are interested in P, the number of preys at the end of the simulation (that is, t=10 years).

Compute the mean, the standard deviation and produce a histogram of P.