CS 131 Problem Set 2

Catherine Angangco, Joshua Frankie Rayo

April 6, 2014

1 Value-at-Risk

1.1 Preliminaries

The value at risk *VaR* of a continuous loss distribution modeled by a probability distribution function, say $p(x)$ at a given risk level $(1-\alpha) \in [0,1]$ is defined as the value at which the cumulative probability of $p(x)$ from $-\infty$ is equal to α. Mathematically speaking, we say:

$$VaR_\alpha = z \Leftrightarrow \int_{-\infty}^{z} p(x)dx = \alpha$$

A value z is the value-at-risk at a if and only if the area under the curve of $p(x)$ in the interval $(-\infty, z]$ is equal to α. We first observe the properties of the probability functions

$$p_1(x) = \frac{1}{\sqrt{6\pi}} e^{-\frac{x^2}{6}}$$

and

$$p_2(x) = \frac{1}{\gamma\sqrt{3\pi}} \left(1 + \frac{x^2}{3}\right)^{-2}, \gamma \approx 0.886226925453$$

where $p_1(x)$ is a Gaussian distribution function and $p_2(x)$ is a t distribution function.

- We know that these functions are symmetric to some line $x = \mu$, where μ is the mean of the distribution. When $\mu = 0$, its graph is symmetric to the line $x = 0$ or y-axis. It has the property

  $$\int_{m}^{n} p(x)dx = \int_{-n}^{-m} p(x)dx$$

  for $m \leq n$.

- The functions above are probability distribution functions. It follows that the sum of the probabilities in the sample space is equal to 1. Or equivalently,

  $$\sum_{\forall x} p(x) = \int_{-\infty}^{\infty} p(x)dx = 1 \tag{1}$$

- We redefine the limits of integration (i.e.) split equation (2) into two as follows:

  $$\int_{-\infty}^{\infty} p(x)dx = \int_{-\infty}^{0} p(x)dx + \int_{0}^{\infty} p(x)dx \tag{2}$$

- Since the graph of the function is symmetric to y-axis, the area under the curve in the interval $[0, \infty)$ is also equal to the area uder the curve in the interval $(-\infty, 0)$. Using equation (1) and (2),

  $$\int_{-\infty}^{0} p(x)dx = \int_{0}^{\infty} p(x)dx = 0.5 \tag{3}$$

Now, we verify if $\int_0^\infty p(x)dx = 0.5$ using Gauss-Laguerre quadrature. For $p_1(x)$, the Gauss-Laguerre quadrature form of the integral is

$$I_1 = \int_0^\infty e^{-x} \left(\frac{e^x}{\sqrt{6\pi}} e^{-\frac{x^2}{6}}\right) dx = \int_0^\infty e^{-x} \left(\frac{1}{\sqrt{6\pi}} e^{x-\frac{x^2}{6}}\right) dx$$

where the weighting function $w(x) = e^{-x}$ and $g_1(x) = \frac{1}{\sqrt{6\pi}} e^{x-\frac{x^2}{6}}$. By Gaussian quadratures approximation,

$$I_1 \approx \sum_{i=1}^{n} A_i g_1(x_i)$$

where i is the number of nodes for approximation. We use $n = 6$ nodes. The nodal abscissas and corresponding weights for $n = 6$ is given below.

i	x_i	A_i
1	0.222 847	0.458 964
2	1.188 932	0.417 000
3	2.992 736	0.113 373
4	5.775 144	0.010 399 2
5	9.837 467	0.000 261 017
6	15.982 874	0.000 000 898 548

To solve the integral,

$$\begin{aligned}
I_1 &= A_1 g_1(x_1) + A_2 g_1(x_2) + A_3 g_1(x_3) \\
&\quad + A_4 g_1(x_4) + A_5 g_1(x_5) + A_6 g_1(x_6) \\
I_1 &= 0.458964 \cdot 0.285454 + 0.417000 \cdot 0.597557 \\
&\quad + 0.113373 \cdot 1.03226 + 0.0103992 \cdot 0.285985 \\
&\quad + 0.000261017 \cdot 0.000426427 \\
&\quad + 8.98548 \times 10^{-7} \cdot 6.50687 \times 10^{-13} \\
I_1 &= 0.500198 \approx 0.5
\end{aligned}$$

For $p_2(x)$, the Gauss-Laguerre quadrature form of the integral is

$$I_2 = \int_0^\infty e^{-x} \left(\frac{e^x}{\gamma\sqrt{3\pi}} \left(1 + \frac{x^2}{3}\right)^{-2}\right), \gamma \approx 0.886226925453$$

where the weighting function $w(x) = e^{-x}$ and $g_2(x) = \frac{e^x}{\gamma\sqrt{3\pi}} \left(1 + \frac{x^2}{3}\right)^{-2}$. By Gaussian quadratures approximation,

$$I_2 \approx \sum_{i=1}^{n} A_i g_2(x_i)$$

We used again $n = 6$ nodes, and the nodal abscissas and weights above.

To solve the integral,

$$\begin{aligned}
I_2 &= A_1 g_2(x_1) + A_2 g_2(x_2) + A_3 g_2(x_3) \\
&\quad + A_4 g_2(x_4) + A_5 g_2(x_5) + A_6 g_2(x_6) \\
I_2 &= 0.458964 \cdot 0.444468 + 0.417000 \cdot 0.55761 \\
&\quad + 0.113373 \cdot 0.461408 + 0.0103992 \cdot 0.806524 \\
&\quad + 0.000261017 \cdot 6.22113 \\
&\quad + 8.98548 \times 10^{-7} \cdot 432.589 \\
I_2 &= 0.499229 \approx 0.5
\end{aligned}$$

And now we established equation (3), (2), and (1).

Going back to our original problem, take note that we are only estimating the value-at-risks for $\alpha \geq 0.5$, so we know that $z \geq 0$. Therefore we should use the relationship that

$$\begin{aligned}
\int_{-\infty}^{z} p(x)dx &= \int_{-\infty}^{0} p(x)dx + \int_{0}^{z} p(x)dx \\
&= 0.5 + \int_{0}^{z} p(x)dx \\
&= \alpha
\end{aligned}$$

So it only remains to compute for $\int_0^z p(x)dx$. We use composite Simpson's 1/3 rule with 6 points also..

The general integral $\int_a^b f(x)dx$ can be numerically solved using the composite Simpson's 1/3 rule:

$$\begin{aligned}
\int_a^b f(x)dx &= \frac{h}{3} f(a) + \frac{4h}{3} \sum_{i\ even} f(x_i) \\
&\quad + \frac{2h}{3} \sum_{i\ odd} f(x_i) + \frac{h}{3} f(b) \\
h &= \frac{b-a}{n}
\end{aligned} \tag{4}$$

where n is the number of panels. The next step is to model our integral $\int_0^z p(x)dx$ as a function, say $P(z)$ so that we can use root-finding method for the equation $0.5 + P(z) - \alpha = 0$.

1.2 VaR estimates using Gaussian distribution

We first divide the interval $[0, z]$ into five panels with 6 points with 5 panels: $\{0, z/5, 2z/5, 3z/5, 4z/5, 5z/5, z\}$. Using equation (4), the integral $\int_0^z \frac{1}{\sqrt{1\pi}} e^{-\frac{x^2}{6}} dx = \int_0^z p_1(x)dx$ is:

$$\begin{aligned}
\int_0^z p_1(x)dx &= \frac{h}{3} p_1(0) + \frac{4h}{3} p_1\left(\frac{z}{5}\right) + \frac{2h}{3} p_1\left(\frac{2z}{5}\right) \\
&= \frac{4h}{3} p_1\left(\frac{3z}{5}\right) + \frac{2h}{3} p_1\left(\frac{4z}{5}\right) + \frac{h}{6} p_1(z) \\
&= P_1(z) \\
h &= \frac{z}{5}
\end{aligned}$$

And now we try to estimate the VaR with $\alpha = 0.8$. Then we use Regula-Falsi method for the equation $0.5 + P_1(z) - 0.8 = P_1(z) - 0.3 = 0 = Q_1(z)$, with $z \in [0, 1]$ and $tol = 10^{-9}$. The file `PS2_1.sce` implements the root-finding method, the root converges to 1.567 692 7.

1.3 VaR estimates using t distribution

Using equation (4), the integral

$$\int_0^z \frac{1}{\gamma\sqrt{3\pi}} \left(1 + \frac{x^2}{3}\right)^{-2} = \int_0^z p_2(x)dx$$

is:

$$\begin{aligned}
\int_0^z p_2(x)dx &= \frac{h}{3} p_2(0) + \frac{4h}{3} p_2\left(\frac{z}{5}\right) + \frac{2h}{3} p_2\left(\frac{2z}{5}\right) \\
&= \frac{4h}{3} p_2\left(\frac{3z}{5}\right) + \frac{2h}{3} p_2\left(\frac{4z}{5}\right) + \frac{h}{6} p_2(z) \\
&= P_2(z) \\
h &= \frac{z}{5}
\end{aligned}$$

And now we try to estimate the VaR with $\alpha = 0.8$. Then we use Regula-Falsi method for the equation $0.5 + P_2(z) - 0.8 = P_2(z) - 0.3 = 0 = Q_2(z)$, with $z \in [0, 1]$ and $tol = 10^{-9}$. The root converges to 1. 053 385 1. We estimate the VaR with $\alpha \in [0.8, 0.99]$ at intervals of 0.01 for the two models. Results are displayed in a tabular and graphical form when the `PS2_1.sce` is run.

2 Naive Fourier Series Approximation

2.1 Romberg Integration

The periodic function contained three variables with integrals. Given parameters f (function) and p (interval), Romberg's was used for the integral approximation. For this case parameters used were: $f(x) = e^{-|x|}$ and $p = 1$

Romberg's integration involves using two previous integrations to form the next one. The generalization is as follows:

$$R_{ij} = \frac{R_{i-1,j-1} - 4^{i-1} R_{i-1,j}}{1 - 4^{i-1}}$$

Stopping condition used:

$$|R_{i+1,i+1} - R_{i,i}| < 10^{-6}$$

This is done for a_o,a_i,and b_i. Given the parameters above, the Romberg table for each of the variables are as follows:

	i	j	$R_{i,j}$	Value
a_0	22	22	0.99999	0.99999
a_i	22	22	0.99999	0.99999
b_i	2	2	0.00000	0.00000

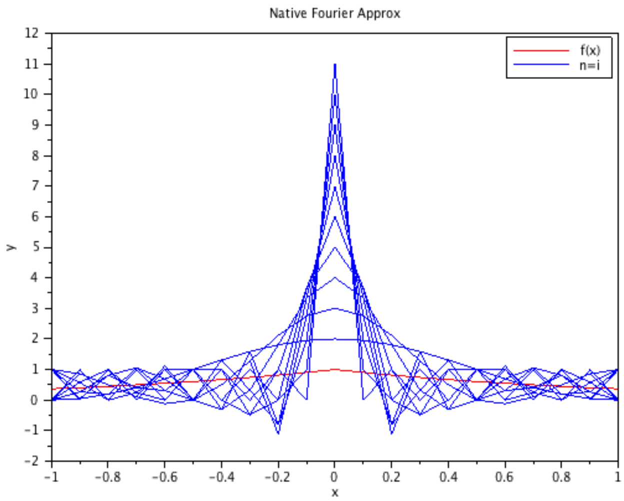
2.2 Fourier Series Approximation

The three approximated integrals were used in the Fourier Series Approximation equation:

$$f(x) = \frac{a_o}{2} + \sum a_i \cos\left(\frac{i\pi x}{p}\right) + \sum b_i \sin\left(\frac{i\pi x}{p}\right) Thexrangeusedwa$$

2.3 Approximation Convergence

To see the convergence Fourier Approx. was done with several n. This case was from 1 to 10. The plot was overlayed with the original $f(x)$ in red.



3 Halley's Comet

4 Yeast Growth Modelling

4.1 Preliminaries

Recall the Runge-Kutta order 4 method for numerical solution of first-order differential equation. Given a differential equation

$$\frac{dy}{dt} = f(t, y)$$

the numerical solution of the differential equation with RK4 is

$$\begin{aligned} x_1 &= f\left(t_n, y_n\right) \\ x_2 &= f\left(t_n + \frac{\Delta t}{2}, y_n + x_1\frac{\Delta t}{2}\right) \\ x_3 &= f\left(t_n + \frac{\Delta t}{2}, y_n + x_2\frac{\Delta t}{2}\right) \\ x_4 &= f\left(t_n + \Delta t, y_n + x_3\Delta t\right) \\ x &= \frac{x_1 + 2x_2 + 2x_3 + x_4}{6} \\ y_{n+1} &= y_n + x\Delta t \end{aligned} \tag{5}$$

Where f is the differential equation, y is the dependent variable, t is the dependent variable, x is the approximated slope and Δt is the increment for the independent variable.
Also recall the Gauss-Newton optimization for nonlinear curve fitting. The optimum parameter vector $\mathbf{P} = [p_1\, p_2\, ...p_n]$ for a parameterized function $f_p(x)$ to fit with n points (x_n, y_n) can be obtained using the iteration

$$\begin{aligned} \mathbf{P_{new}} &= \mathbf{P_{old}} - \delta \\ \delta &= \mathrm{GE}\left(\mathbf{J_r^T J_r}, \mathbf{J_r^T r}\right) \\ \mathbf{r} &= \begin{bmatrix} f_p(x_1) - y_i \\ \vdots \\ f_p(x_n) - y_n \end{bmatrix} \end{aligned} \tag{6}$$

where $\mathbf{r}$ is the residual matrix and $\mathbf{J_r}$ is the Jacobian of the residual matrix. The iteration will stop if $\|x_new - x_old\|_2 \leq tol$.

4.2 Logistic Model

The logistic growth model of yeast is given by

$$\frac{dP}{dt} = kP(C - P) = f(t, P)$$

where P is the current population. Variables k and C are the parameters of the differential equation. Now, we obtain the solution of this model using RK4 with $\Delta t = 1$ using equation (5). See next page:

$$
\begin{aligned}
x_1 &= kP_n(C - P_n) \\
x_2 &= k\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\left(C - \left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right) \\
x_3 &= k\left(C - P_n - \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\left(P_n + \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right) \\
x_4 &= k\left(C - P_n - k\left(C - P_n - \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\left(P_n + \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\right)\right) \\
&\quad \left(P_n + k\left(C - P_n - \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\left(P_n + \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\right)\right) \\
x &= \frac{x_1 + 2x_2 + 2x_3 + x_4}{6}
\end{aligned}
$$

$$
\begin{aligned}
P_{n+1} &= P_n + \frac{k}{6}P_n(C - P_n) + \frac{2k}{6}\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\left(C - \left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right) \\
&\quad + \frac{2k}{6}\left(C - P_n - \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\left(P_n + \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right) \\
&\quad + \frac{k}{6}\left(C - P_n - k\left(C - P_n - \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\left(P_n + \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\right)\right) \\
&\quad \left(P_n + k\left(C - P_n - \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\right)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\left(P_n + \frac{1}{2}k\left(C - P_n - \frac{1}{2}kP_n(C - P_n)\left(P_n + \frac{1}{2}kP_n(C - P_n)\right)\right)\right)\right)
\end{aligned}
$$

And this is the function we want to optimize, using different values of P_n (independent variable) and P_{n+1} (dependent variable). Take note that the current population is dependent on the previous population. In this case, P_{n+1} is a function of P_n parameterized by k and C. We let $P_{n+1} = f_{<k,C>}(P_n)$. Given the appendix data, we form the residual matrix.

$$\mathbf{r} = \begin{bmatrix} f_{<k,C>}(1.03519) - 1.06937 \\ f_{<k,C>}(1.06937) - 1.12796 \\ f_{<k,C>}(1.12796) - 1.06937 \\ \vdots \\ f_{<k,C>}(1.32817) - 1.32817 \\ f_{<k,C>}(1.32817) - 1.33305 \\ f_{<k,C>}(1.33305) - 1.33305 \end{bmatrix} \mathbf{J_r} = \begin{bmatrix} \frac{\partial}{\partial k} f_{<k,C>}(1.03519) & \frac{\partial}{\partial C} f_{<k,C>}(1.03519) \\ \frac{\partial}{\partial k} f_{<k,C>}(1.06937) & \frac{\partial}{\partial C} f_{<k,C>}(1.06937) \\ \frac{\partial}{\partial k} f_{<k,C>}(1.12796) & \frac{\partial}{\partial C} f_{<k,C>}(1.12796) \\ \vdots & \vdots \\ \frac{\partial}{\partial k} f_{<k,C>}(1.32817) & \frac{\partial}{\partial C} f_{<k,C>}(1.32817) \\ \frac{\partial}{\partial k} f_{<k,C>}(1.32817) & \frac{\partial}{\partial C} f_{<k,C>}(1.32817) \\ \frac{\partial}{\partial k} f_{<k,C>}(1.33305) & \frac{\partial}{\partial C} f_{<k,C>}(1.33305) \end{bmatrix}$$

To obtain the Jacobian matrix of R, we used the Scilab function `derivative(fxn, var, order)`. The Scilab script `PS2_4.sce` performs the Gauss-Newton optimization with $P_{old} = [1\ 1]^T$ and $tol = 1 \times 10^{-9}$. The optimum parameters $[k, C]$ converge to $[0.1615294, 1.3374996]$. Therefore the optimum logistic model for the given appendix is

$$\boxed{\frac{dP}{dt} = 0.1615294P(1.3374996 - P)}$$

And now, we obtain the general solution of the logistic growth model given above with $\Delta t = 0.05$. The result is shown in Scilab console. Take note that the numerical solution obtained is somewhat different from the appendix table. It is because we used optimal parameters for the logistic model, similar to curve fitting.

4.3 Gompertz Model

The Gompertz growth model of yeast is given by

$$\frac{dP}{dt} = kP\ln(C - P)$$

Similar to the logistic growth model, the function we want to optimize is (see next page):

$$
\begin{aligned}
P_{n+1} = {} & P_n + \frac{k}{6} P_n \ln(C - P_n) + \frac{2k}{6} \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \ln \left(C - \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \right) \\
& + \frac{2k}{6} \ln \left(C - P_n - \frac{1}{2} k \ln \left(C - P_n - \frac{1}{2} k P_n \ln(C - P_n) \right) \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \right) \left(P_n + \frac{1}{2} k \ln \left(C - P_n - \frac{1}{2} k P_n \ln(C - P_n) \right) \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \right) \\
& + \frac{k}{6} \ln \left(C - P_n - k \ln \left(C - P_n - \frac{1}{2} k \ln \left(C - P_n - \frac{1}{2} k P_n \ln(C - P_n) \right) \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \right) \left(P_n + \frac{1}{2} k \ln \left(C - P_n - \frac{1}{2} k P_n \ln(C - P_n) \right) \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \right) \right) \\
& \left(P_n + k \ln \left(C - P_n - \frac{1}{2} k \ln \left(C - P_n - \frac{1}{2} k P_n \ln(C - P_n) \right) \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \right) \left(P_n + \frac{1}{2} k \left(C - P_n - \frac{1}{2} k P_n \ln(C - P_n) \left(P_n + \frac{1}{2} k P_n \ln(C - P_n) \right) \right) \right) \right)
\end{aligned}
$$

Solving the optimum parameters is also the same as done in the logistic model. The optimum parameters $[k, C]$ converge to $[0.1615294, 1.3374996]$. Therefore the optimum logistic model for the given appendix is

$$\frac{dP}{dt} = (0.2824938 + 0.0037822i)P\ln((-2.6785793 - 6.3601969i) - P)$$

You may wonder why this happened. The Jacobian matrix has complex numbers in it. One reason is that Scilab encounters negative values and computes the natural logarithm of it, resulting to complex numbers. Then the optimum parameters and the numerical solutions are now complex numbers.

And now, we obtain the general solution of the Gompertz growth model given above with $\Delta t = 0.05$. The result is shown in Scilab console. Again, take note that the numerical solution obtained is different from the appendix table.

One may conclude that the Gompertz model is not appropriate to describe the growth of yeast.