Risk analysis and management

Telek Miklós
BME

October 24, 2023

Table of content

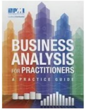
Theory – practice

- Theory: Miklós Telek
  I.L.117, phone: 2084, or `telek@hit.bme.hu`.

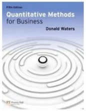
- Practice: András Mészáros
  I.B.115, phone: 3219, or `meszarosa@hit.bme.hu`.

Books

Finance and insurance industries build on involved mathematical models

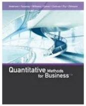


Business Analysis for Practitioners

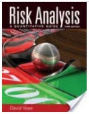


Quantitative Methods for Business

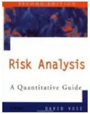
Quantitative Analysis for Management (13th



Quantitative Methods for Business



Risk Analysis - A Quantitative



Risk Analysis: A Quantitative ...

Table of content

Table of content

Summary of Linear algebra

- System of linear equations
  - 0, 1, or infinitely many solutions.
- Vectors, matrices
- Singular value decomposition (SVD),
  - solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ with the SVD of $\mathbf{A}$.
- Spectral decomposition,
  - iterative procedure for finding the dominant eigenvalue and eigenvector.

Summary of Linear algebra

- Commutativity of matrices
- Sylvester equation
  - *vec* operator, Kronecker product ($\otimes$),
  - $vec(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) vec(\mathbf{B})$,
- Matrix functions
  - definition,
  - spectral decomposition based interpretation.

Linear equation

Scalar linear equation: $ax = b$

- $a \neq 0 \longrightarrow$ single solution: $x = b/a$.
- $a = 0$
  - $b = 0 \longrightarrow$ infinite solutions: $x \in \mathbb{R}$,
  - $b \neq 0 \longrightarrow$ no solution.

System of linear equations

System of linear equations:

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$
$$a_{31}x_1 + a_{32}x_2 = b_3$$

That is

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

with

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Scalar description of the matrix equation:

$$\sum_{j=1}^{2} a_{ij}x_j = b_i, \quad \text{for } i = 1, 2, 3.$$

Table of content

Matrix properties

Matrix properties

- size,
- rank (number of independent rows/columns)
- singular values (numerically stable)

Square matrix properties

- determinant,
- eigenvalues, eigenvectors (numerically sensitive),
- inverse exists:
  - invertible, full rank, independent rows/columns, non-zero determinant, non-singular, . . .

Special matrices

Identity matrix: $\mathbf{I} = \{\delta_{ij}\}$,

where $\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$ is the Kronecker delta.

Diagonal matrix: $\mathbf{D} = \text{diag}\{d_1, \dots, d_n\}$,

Unitary matrix: $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}$ (if $\mathbf{U}$ is real)

For complex $\mathbf{U}$: $\mathbf{U}^H\mathbf{U} = \mathbf{U}\mathbf{U}^H = \mathbf{I}$,

where ^{H} is the conjugate transpose operator.

Commuting matrices

Commonly, $\mathbf{AB} \neq \mathbf{BA}$,

as a consequence several scalar identity fails for matrices, e.g.:

$$(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + \mathbf{AB} + \mathbf{BA} + \mathbf{B}^2 \neq \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2$$

$$\frac{d}{dx}(\mathbf{A}+x\mathbf{B})^2 = \mathbf{B}(\mathbf{A}+x\mathbf{B}) + (\mathbf{A}+x\mathbf{B})\mathbf{B} \neq 2(\mathbf{A}+x\mathbf{B})\mathbf{B}$$

Exceptions:

$\mathbf{A}, \mathbf{I}, \mathbf{A}^{-1}, \mathbf{A}^n$ for $n \in \mathbb{N}$ and all of their linear combinations, $\sum_{n=-\infty}^{\infty} c^n \mathbf{A}^n$, always commute.

The usual scalar identities hold for commuting matrices.

Singular value decomposition (SVD)

$$\begin{bmatrix} & \mathbf{A} & \end{bmatrix}_{n\times m} = \begin{bmatrix} & \mathbf{U} & \end{bmatrix}_{n\times n} \begin{bmatrix} & \Psi & \end{bmatrix}_{n\times m} \begin{bmatrix} & \mathbf{V} & \end{bmatrix}_{m\times m}$$

where **U** and **V** are unitary matrices,

$$\begin{bmatrix} & \Psi & \end{bmatrix}_{n\times m} = \begin{bmatrix} \sigma_1 & \\ & \sigma_n \end{bmatrix}_{n\times m} = \begin{bmatrix} \boxed{\mathbf{S}} & \\ & \end{bmatrix}_{n\times m}$$

is a matrix whose diagonal elements are the $\sigma_i \geq 0$ singular values.

Ψ is assumed to be ordered such that $\sigma_1 \geq \sigma_2 \geq \ldots$.

The r non-zero singular value form diagonal matrix **S** of size $r \times r$, where $r \leq \min(m, n)$.

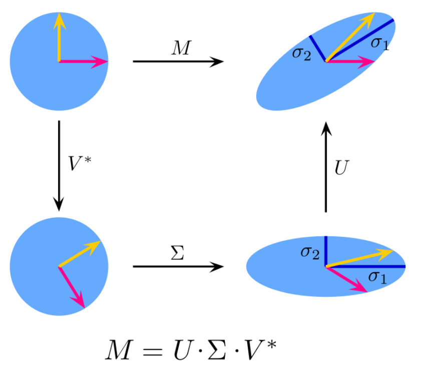
For the number of non-zero singular values we have

$$r = \text{rank}\mathbf{A} = \text{number of independent rows/columns of } \mathbf{A}$$

Singular value decomposition (SVD)

Graphical demonstration

(from `https://en.wikipedia.org/wiki/Singular_value_decomposition`)



Linear equations

$\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution

if $\mathbf{b}$ is a linear combination of the columns of $\mathbf{A}$.

That is

$$\operatorname{rank}\left[\begin{array}{c} \mathbf{A} \end{array}\right] = \operatorname{rank}\left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \end{array}\right]$$

Linear equations

If $\mathbf{A} = \mathbf{U}\Psi\mathbf{V}$ is the SVD of $\mathbf{A}$ then

$$\mathbf{A}_{n\times m} \;\; \mathbf{x}_{m\times 1} = \mathbf{b}_{n\times 1} \quad \Big| \cdot \mathbf{U}^T \text{ from left}$$

can be written as

$$\mathbf{U}^T\mathbf{A}\mathbf{x} = \mathbf{U}^T\mathbf{b}$$
$$\underbrace{\mathbf{U}^T\mathbf{U}}_{\mathbf{I}}\,\Psi\,\underbrace{\mathbf{V}\mathbf{x}}_{\mathbf{x}'} = \underbrace{\mathbf{U}^T\mathbf{b}}_{\mathbf{b}'}$$

that gives a transformed linear equation

$$\Psi_{n\times m} \;\; \mathbf{x}'_{m\times 1} = \mathbf{b}'_{n\times 1}$$
$$\begin{bmatrix} \boxed{\mathbf{S}} & \\ & \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}'_{\mathbf{1}} \\ \hline \mathbf{x}'_{\mathbf{2}} \end{bmatrix} = \begin{bmatrix} \mathbf{b}'_{\mathbf{1}} \\ \hline \mathbf{b}'_{\mathbf{2}} \end{bmatrix}.$$

with $\mathbf{x}' = \mathbf{V}\mathbf{x}$, $\mathbf{b}' = \mathbf{U}^T\mathbf{b}$ and block sizes $\mathbf{S}_{r\times r}$, $\mathbf{x}'_{\mathbf{1}\,r\times 1}$, $\mathbf{x}'_{\mathbf{2}\,m-r\times 1}$, $\mathbf{b}'_{\mathbf{1}\,r\times 1}$, $\mathbf{b}'_{\mathbf{2}\,n-r\times 1}$.

Linear matrix equations

The block decomposed version of the a transformed linear equation

$$\left[\begin{array}{c|c} \mathbf{S} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \cdot \left[\begin{array}{c} \mathbf{x}_1' \\ \hline \mathbf{x}_2' \end{array}\right] = \left[\begin{array}{c} \mathbf{b}_1' \\ \hline \mathbf{b}_2' \end{array}\right]$$

is

$$\begin{aligned} \mathbf{S}\mathbf{x}_1' + \mathbf{0}\mathbf{x}_2'{}_{m-r\times 1} &= \mathbf{b}_1' \\ \mathbf{0}\mathbf{x}_1' + \mathbf{0}\mathbf{x}_2' &= \mathbf{b}_2'{}_{n-r\times 1} \end{aligned}$$

- If $n - r > 0$ and $\mathbf{b}_2' \neq \mathbf{0}$ then no solution.
- If $\mathbf{b}_2' = \mathbf{0}$ and $m - r = 0$ then the single solution is $\mathbf{x} = \mathbf{V}^T \mathbf{S}^{-1} \mathbf{b}_1'$.
- If $\mathbf{b}_2' = \mathbf{0}$ and $m - r > 0$ then there are infinite solutions of dimension $m - r$.

Linear matrix equations

In some cases, a matrix of unknowns $\mathbf{X}$ and some matrices of coefficients form a linear matrix equation.

E.g., $\mathbf{AX} = \mathbf{B}$.

If $\exists \mathbf{A}^{-1}$ then $\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$ is the solution.

If $\nexists \mathbf{A}^{-1}$ then $\mathbf{AX} = \mathbf{B}$ needs to be transformed into standard linear equation form using

- *vec* operator,
- Kronecker product ($\otimes$),
- $vec(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) vec(\mathbf{B})$,

$$\begin{aligned} vec(\mathbf{AX}) &= vec(\mathbf{C}) \\ vec(\mathbf{AXI}) &= vec(\mathbf{C}) \\ \underbrace{\mathbf{I} \otimes \mathbf{A}}_{\mathbf{A}'} \underbrace{vec(\mathbf{X})}_{\mathbf{x}'} &= \underbrace{vec(\mathbf{C})}_{\mathbf{b}'} \end{aligned}$$

Linear matrix equations

In case of the Sylvester equation

$$\mathbf{AX} + \mathbf{XB} = \mathbf{C}$$

the same approach has to be applied also when $\exists \mathbf{A}^{-1}$ and $\exists \mathbf{B}^{-1}$:

$$\begin{aligned} vec(\mathbf{AX} + \mathbf{XB}) &= vec(\mathbf{C}) \\ vec(\mathbf{AXI} + \mathbf{IXB}) &= vec(\mathbf{C}) \\ \underbrace{(\mathbf{I} \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I})}_{\mathbf{A}'} \underbrace{vec(\mathbf{X})}_{\mathbf{x}'} &= \underbrace{vec(\mathbf{C})}_{\mathbf{b}'} \end{aligned}$$

Spectral decomposition

$\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}$ is the spectral decomposition of $\mathbf{A}$ when $\mathbf{U}^{-1} = \mathbf{V}$ and Λ is a block diagonal matrix composed of Jordan blocks $\mathbf{J_i}$

$$\Lambda = \begin{bmatrix} \mathbf{J_1} & & & \\ & \mathbf{J_2} & & \\ & & \ddots & \\ & & & \mathbf{J}_{\#\lambda} \end{bmatrix}_{n \times n}, \quad \mathbf{J_i} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}_{\#\lambda_i \times \#\lambda_i}.$$

If all Jordan blocks are of size one then $\#\lambda = n$,

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

and $\mathbf{A}$ is said to be *diagonalizable*.

Spectral decomposition

If $\mathbf{A}$ is diagonalizable then

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{V} = \sum_{i=1}^{n} \mathbf{u_i}\lambda_i\mathbf{v_i}$$

where $\mathbf{u_i}$ is the ith column of $\mathbf{U}$ and $\mathbf{v_i}$ is the ith row of $\mathbf{V}$.

Computing the spectral decomposition

- Solve the order n polynomial equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ $\lambda_1, \ldots, \lambda_n$ are its roots,
- for $i = 1, \ldots, n$ solve the linear equation $(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{u_i} = \mathbf{0}$,
- obtain $\mathbf{v_i}$ from $\mathbf{V} = \mathbf{U}^{-1}$.

Note that $\mathbf{v_i}\mathbf{u_j} = \delta_{ij}$ due to $\mathbf{VU} = \mathbf{I}$.

Iterative procedure for computing λ^* and u^*

The dominant eigenvalue, λ^*, and the related eigenvector u^* of $\mathbf{A}$ can be computed using the summation vector $\mathbf{s}$ and initial vector $\mathbf{u_{init}}$ as follows

Input: $\mathbf{u_{init}}, \mathbf{A}, \mathbf{s}$;
$\mathbf{u} = \mathbf{u_{init}}$;
repeat
  $\mathbf{u_{old}} = \mathbf{u}$;
  $c = \mathbf{s}^T \mathbf{u}$;
  $\mathbf{u} = \mathbf{A}\mathbf{u}/c$;
until $|\mathbf{u_{old}} - \mathbf{u}| < \epsilon$;
return : $c, \mathbf{u}$;

A potential initial setting is $\mathbf{s}^T = \{1, 1, \ldots, 1\}$ and $\mathbf{u_{init}}^T = \{1, 0, \ldots, 0\}$.

Evaluate the conditions when the procedure converges.

Matrix functions

If $\mathbf{A}$ is a square matrix and $f(x)$ is a scalar function with Taylor series $f(x) = \sum_{i=0}^{\infty} c_i x^i$ then

$$f(\mathbf{A}) \triangleq \sum_{i=0}^{\infty} c_i \mathbf{A}^i$$

If $\mathbf{A}$ is diagonalizable and $\mathbf{A} = \mathbf{U}\Lambda\mathbf{V}$ is its spectral decomposition then

$$f(\mathbf{A}) = \sum_{i=0}^{\infty} c_i \mathbf{A}^i = \sum_{i=0}^{\infty} c_i (\mathbf{U}\Lambda\mathbf{V})^i = \sum_{i=0}^{\infty} c_i \mathbf{U}\Lambda^i\mathbf{V}$$

$$= \mathbf{U} \sum_{i=0}^{\infty} c_i \begin{bmatrix} \lambda_1^i & & & \\ & \lambda_2^i & & \\ & & \ddots & \\ & & & \lambda_n^i \end{bmatrix} \mathbf{V} = \mathbf{U} \begin{bmatrix} f(\lambda_1) & & & \\ & f(\lambda_2) & & \\ & & \ddots & \\ & & & f(\lambda_n) \end{bmatrix} \mathbf{V}$$

Table of content

Table of content

Random variables

Independent random variables (RV)

- cumulated distribution function (CDF)
  $F_X(x) = Pr(X < x)$
  - discrete RV: probability mass function (PMF)
    $p_i = Pr(X = x_i)$
  - continuous RV: probability density function (PDF)
    $f_X(x) = \frac{d}{dx} F_X(x)$
- moments: $E(X^n)$
- and their descendants (e.g., variance (2nd cumulant):
  $\sigma_X^2 = E(X^2) - E(X)^2$, nth cumulant κ_X^n)
- The cumulants sums up: $\kappa_{X+Y}^n = \kappa_X^n + \kappa_Y^n$

Law of total probability

Law of total probability (LTP)

- $Pr(A) = \sum_i Pr(A|B_i)Pr(B_i)$,
  - discrete condition:

$$Pr(A) = \sum_i Pr(A|X = x_i)Pr(X = x_i)$$
$$= \sum_i Pr(A|X = x_i)p_i$$

  - continuous condition:

$$Pr(A) = \int_x Pr(A|X = x)f_X(x)dx$$

- $E(Y) = \sum_i E(Y|B_i)Pr(B_i)$.

Danger: $Pr(A|X = x) \to \lim_{\delta \to 0} Pr(A|x \leq X < x + \delta)$

Law of total probability

Application

- $E(g(Y)) = \sum_i E(g(Y)|B_i)Pr(B_i)$,
  - discrete condition:

$$E(g(Y)) = \sum_i E(g(Y)|X = x_i)p_i$$

  - continuous condition:

$$E(g(Y)) = \int_x E(g(Y)|X = x)f_X(x)dx$$

If $g(x) = x^n$ and $Y = X$ then $E(g(Y)) = E(X^n)$.

Table of content

Distributions

One-parameter

- Discrete
  - Bernoulli (on $\{0,1\}$)
  - Geometric
  - Poisson
- Continuous
  - Exponential

Two-parameter

- Discrete
  - Uniform
  - Binomial
- Continuous
  - Uniform
  - Normal

Transforms

Transforms:

- Characteristic function $\phi(t) = E(e^{itX}), t \in R$
- Moment generating function $M(t) = E(e^{tX}), t \in R$
- Cumulant generating function $K(t) = log(E(e^{tX})), t \in R$
- Probability generating function $G(z) = E(z^X), z \in C$
- Laplace transform $L(s) = E(e^{-sX}), s \in C$

Advantages:

- analytically tractable (due to convolution, linear operations)
- direct computation of moments
- inverse transformation (symbolic/numeric)

Dependent random variables

Dependent random variables (X, Y)

- cumulative distribution function (CDF) $F_{X,Y}(x,y) = Pr(X < x, Y < y)$
  - discrete RV: probability mass function (PMF) $p_{ij} = Pr(X = x_i, Y = y_j)$
  - continuous RV: probability density function (PDF) $f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y)$
- marginal distribution:

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y)$$
$$= \lim_{y\to\infty} Pr(X < x, Y < y) = Pr(X < x)$$

Dependent random variables

Dependent random variables (X, Y)

- conditional distribution $Pr(X < x | Y = y)$
  - discrete RV: $Pr(X = x_i | Y = y_j) = \frac{Pr(X=x_i, Y=y_j)}{Pr(Y=y_j)} = \frac{p_{ij}}{p_j}$
  - continuous RV: $f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{\int_z f_{X,Y}(z,y)dz}$
- joint moments: $E(X^n Y^m) = \int_x \int_y x^n y^m f_{X,Y}(x,y) dy dx$
- and their descendants (e.g., covariance: $E(XY) - E(X)E(Y)$, correlation)

Normal distribution

PDF of normal distribution with (μ, σ^2):

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

If X is normal distributed with (μ, σ^2), then $\hat{X} = \frac{X-\mu}{\sigma}$ is standard normal distributed.

PDF and CDF of standard normal distribution with $(\mu = 0, \sigma^2 = 1)$:

$$f_{\hat{X}}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi_{\hat{X}}(x) = \int_{y=-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

CDF of normal distribution with (μ, σ^2):

$$Pr(X < x) = \Phi_{\hat{X}}\left(\frac{x-\mu}{\sigma}\right).$$

Multivariate normal distribution

Probability density function

- $\mathbf{X} = \{X_1, \ldots, X_k\}^T$ is multivariate normal with *location* $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_k\}^T$ and *covariance matrix* $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ if its PDF is

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-k/2} \det(\boldsymbol{\Sigma})^{-1/2} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})}$$

where $E(X_i) = \mu_i$ and $E(X_i X_j) - E(X_i)E(X_j) = \sigma_{ij}$, that is $\sigma_{ii} = E(X_i X_i) - E(X_i)E(X_i) = Var(X_i)$.

In matrix form,
$E(\mathbf{X}) = \boldsymbol{\mu}$ and $E(\mathbf{X}\mathbf{X}^T) - E(\mathbf{X})E(\mathbf{X}^T) = \boldsymbol{\Sigma}$.

$\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ is symmetric, positive definite matrix (with positive eigenvalues).

Multivariate normal distribution

Construction of multivariate normal distribution

- Let $\mathbf{Z} = \{Z_1, \ldots, Z_k\}^T$ be composed of i.i.d. standard normal distributed RVs.
  That is $E(Z_i) = 0$, $Var(Z_i) = 1$,
  and $E(Z_i Z_j) = E(Z_i)E(Z_j)$ for $i \neq j$.

  In matrix form $E(\mathbf{Z}) = \mathbf{0}$ and $E(\mathbf{Z}\mathbf{Z}^T) - E(\mathbf{Z})E(\mathbf{Z}^T) = \mathbf{I}$,
  because for $i \neq j$, $E(Z_i Z_j) - E(Z_i)E(Z_j) = 0$ and
  $E(Z_i Z_i) - E(Z_i)E(Z_i) = Var(Z_i) = 1$.

Multivariate normal distribution

Construction of multivariate normal distribution

- Let $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$.

  $\mathbf{X}$ is multivariate normal distributed with *location* $\boldsymbol{\mu}$ and *covariance matrix* $\boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$, because

$$E(\mathbf{X}) = E(\boldsymbol{\mu} + \mathbf{A}\mathbf{Z}) = \boldsymbol{\mu} + \mathbf{A}\underbrace{E(\mathbf{Z})}_{\mathbf{0}} = \boldsymbol{\mu}$$

  and

$$\begin{aligned}
\boldsymbol{\Sigma} &= E(\mathbf{X}\mathbf{X}^T) - E(\mathbf{X})E(\mathbf{X}^T) \\
&= E((\boldsymbol{\mu} + \mathbf{A}\mathbf{Z})(\boldsymbol{\mu} + \mathbf{A}\mathbf{Z})^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= E(\boldsymbol{\mu}\boldsymbol{\mu}^T) + E(\boldsymbol{\mu}(\mathbf{A}\mathbf{Z})^T) + E(\mathbf{A}\mathbf{Z}\boldsymbol{\mu}^T) + E(\mathbf{A}\mathbf{Z}(\mathbf{A}\mathbf{Z})^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}(\mathbf{A}\underbrace{E(\mathbf{Z})}_{\mathbf{0}})^T + \mathbf{A}\underbrace{E(\mathbf{Z})}_{\mathbf{0}}\boldsymbol{\mu}^T + E(\mathbf{A}\mathbf{Z}\mathbf{Z}^T\mathbf{A}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= \mathbf{A}\underbrace{E(\mathbf{Z}\mathbf{Z}^T)}_{\mathbf{I}}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T.
\end{aligned}$$

Table of content

Table of content

Problem formulation

There is financial institution S (system) with C resources (currency) and N customers (investors).

The customers can request $h_1, h_2, \ldots, h_N$ resource with probability $p_1, p_2, \ldots, p_N$, respectively.

The system risk is defined as

$$Pr(\text{aggregate request exceeds the resources}) = Pr\left(\sum_{i=1}^{N} Y_i h_i > C\right)$$

where Y_i is a Bernoulli RV with $Pr(Y_i = 1) = p_i$.

The main challenge and solution methods

The main challenge

- Timely response (real-time)
- Scaling: N is fairly large
- Computational complexity provided by the $\mathcal{O}(2^N)$ cases needs to be reduced.

Solution methods:

- Brute-force
- Large Deviation Theory (based on on-line tail approximation methods)
- Central limit theorem
- Statistical sampling
- Adaptive approximation

Problem variants

Problem variants:

- $h = h_1 = h_2 = \ldots = h_N$ and $p = p_1 = p_2 = \ldots = p_N$
- $h_1, h_2, \ldots, h_N$ are i.i.d. with PDF $f_h(x)$ and $p = p_1 = p_2 = \ldots = p_N$
- $h = h_1 = h_2 = \ldots = h_N$
- $h_1, h_2, \ldots, h_N$ are i.i.d. with PDF $f_h(x)$
- $p = p_1 = p_2 = \ldots = p_N$

Table of content

Brute force solution

Brute force solution:
LTP completely eliminating the randomness

$$Pr\left(\sum_{i=1}^{N} Y_i h_i > C\right)$$
$$= \sum_{y_1=0}^{1} \cdots \sum_{y_N=0}^{1} \prod_{j=1}^{N} Pr(Y_j = y_j)$$
$$\cdot Pr\left(\sum_{i=1}^{N} Y_i h_i > C \,\middle|\, Y_1 = y_1, \ldots, Y_N = y_N\right)$$
$$= \sum_{\forall \mathbf{y} \in \{0,1\}^N} Pr(\mathbf{y}) \cdot \underbrace{Pr\left(\mathbf{y}\mathbf{h}^T > C\right)}_{0 \text{ or } 1},$$

where $\mathbf{y} = \{y_1, \ldots, y_N\}$ and $\mathbf{h} = \{h_1, \ldots, h_N\}$.

Table of content

Markov inequality

Markov inequality:

$$Pr(X \geq a) \leq \frac{E(X)}{a}$$

where X is non-negative RV.

The distribution satisfying the equality is

$$\hat{X} = \begin{cases} 0 & \text{with probability } 1 - \frac{E(X)}{a}, \\ a & \text{with probability } \frac{E(X)}{a}. \end{cases}$$

Markov inequality

Proof for continuous non-negative X:

$$E(X) = \int_0^\infty x f_X(x) dx \geq \int_a^\infty x f_X(x) dx$$
$$\geq \int_a^\infty a f_X(x) dx = a \int_a^\infty f_X(x) dx$$
$$= a Pr(X \geq a)$$

Proof for general non-negative X:

$$E(X) = \int_0^\infty x dF_X(x) \geq \int_a^\infty x dF_X(x)$$
$$\geq \int_a^\infty a dF_X(x) = a \int_a^\infty dF_X(x)$$
$$= a(F_X(\infty) - F_X(a)) = a Pr(X \geq a)$$

Chebysev inequality

Chebysev inequality ($X \in \mathbb{R}$, $b \in \mathbb{R}^+$)

$$Pr(|X - E(X)| \geq b) \leq \frac{\sigma_X^2}{b^2}$$

Proof:

Let $Y = (X - E(X))^2$ and apply the Markov inequality for Y at b^2

$$Pr(Y \geq b^2) \leq \frac{E(Y)}{b^2}$$

$$Pr((X - E(X))^2 \geq b^2) \leq \frac{E((X - E(X))^2)}{b^2}$$

$$Pr(|X - E(X)| \geq b) \leq \frac{\sigma_X^2}{b^2}$$

Markov related inequalities

$g(x)$ is *non-negative*, *monotone increasing* for $x > a$, then

$$Pr(X \geq a) \leq \frac{E(g(X))}{g(a)}$$

Proof for continuous X:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx \overset{\text{non-neg.}}{\geq} \int_{a}^{\infty} g(x) f_X(x) dx$$

$$\overset{\text{mon. inc.}}{\geq} \int_{a}^{\infty} g(a) f_X(x) dx = g(a) \int_{a}^{\infty} f_X(x) dx$$

$$= g(a) Pr(X \geq a)$$

Moment inequalities

If $g(x) = x^n$ and $X \in \mathbb{R}^+$ then

$$Pr(X \geq a) \leq \frac{E(X^n)}{a^n}$$

If all $E(X^n)$ moments are known, then

$$Pr(X \geq a) \leq \min_{n \in \mathbb{N}^+} \frac{E(X^n)}{a^n}$$

If $g(x) = x^u$ with $u \in \mathbb{R}^+$ then

$$Pr(X \geq a) \leq \min_{u \in \mathbb{R}^+} \frac{E(X^u)}{a^u}$$

Central moment inequalities

If $g(x) = |x - \mu|^n$, $\mu = E(X)$ and $a > \mu$ then

$g(x)$ is monotone increasing for $x > a$ and

$$Pr(X \geq a) \leq \frac{E(|X - E(X)|^n)}{|a - \mu|^n} = \frac{E(|X - E(X)|^n)}{(a - \mu)^n}$$

where $E(|X - E(X)|^n)$ is the nth central moment of X.

If all central moments are known then

$$Pr(X \geq a) \leq \min_{n \in \mathbb{N}^+} \frac{E(|X - E(X)|^n)}{(a - \mu)^n}.$$

Similarly, if $g(x) = |x - \mu|^u$ with $u \in \mathbb{R}^+$ and $a > \mu$ then

$$Pr(X \geq a) \leq \min_{u \in \mathbb{R}^+} \frac{E(|X - E(X)|^u)}{(a - \mu)^u}.$$

Chernoff bound

If $g(x) = e^{sx}$ and $s > 0$ then

$$Pr(X \geq a) \leq \frac{E(e^{sX})}{e^{sa}},$$

where $M_X(s) = E(e^{sX})$ is the moment generating function.

If $M_X(s)$ is known then

$$Pr(X \geq a) \leq \min_{s \in \mathbb{R}+} \frac{M_X(s)}{e^{sa}} = \frac{M_X(s^*)}{e^{s^* a}},$$

where $s^* = \text{argmin}_{s \in \mathbb{R}+} \frac{M_X(s)}{e^{sa}}$.

Chernoff versus moment bounds

Let $B_C(a,s) = \frac{E(e^{sX})}{e^{sa}}$ and $B_M(a,u) = \frac{E(X^u)}{a^u}$ then

$$B_C(a,s) = \frac{E(e^{sX})}{e^{sa}} = e^{-sa} E\left(\sum_{n=0}^{\infty} \frac{s^n}{n!} X^n\right)$$

$$= e^{-sa} \sum_{n=0}^{\infty} \frac{s^n}{n!} E(X^n) = \sum_{n=0}^{\infty} \frac{(sa)^n}{n!} e^{-sa} \frac{E(X^n)}{a^n}$$

$$= \sum_{n=0}^{\infty} \underbrace{\frac{(sa)^n}{n!} e^{-sa}}_{\text{Poisson}(sa)\text{ weights}} B_M(a,n)$$

$\rightarrow$ the best *moment* bound is at $n^* = \lfloor s^* a + 0.5 \rfloor$

and $B_M(a, n^*) < B_C(a, s^*)$.

$\rightarrow$ the tightest moment-like bound is $B_M(a, s^* a)$.

Cantelli's inequality

For $a \in \mathbb{R}^+$ and $X \in \mathbb{R}$

$$Pr(X - E(X) \geq a) \leq \frac{\sigma_X^2}{\sigma_X^2 + a^2}$$

Proof

Let $Y = X - E(X)$, $u = \frac{\sigma_X^2}{a}$ and $\sigma_X^2 = E(X^2) - E(X)^2$

then $E(Y) = 0$, $E(Y^2) = \sigma_X^2$ and

$$Pr(Y \geq a) = Pr(Y + u \geq a + u) \leq Pr((Y+u)^2 \geq (a+u)^2)$$
$$\stackrel{Markov}{\leq} \frac{E((Y+u)^2)}{(a+u)^2} = \frac{E(Y^2 + 2uY + u^2)}{(a+u)^2}$$
$$= \frac{\sigma_X^2 + u^2}{(a+u)^2}\bigg|_{u=\frac{\sigma_X^2}{a}} = \frac{\sigma_X^2}{\sigma_X^2 + a}$$

Exercise: Which $g(x)$ provides the Cantelli's inequality?

Example

X is Binomial(n, p) with $p = 1/4$.

$$P(X \geq 3n/4) = ???$$

Method	order	opt	bound	$n = 100$
Markov	1	-	$\frac{1}{3}$	0.333
Moment	2	-	$\frac{3n+n^2}{9n^2}$	0.114
All moments	∞	+		$1.11 \cdot 10^{-24}$
Chebyshev	2	-	$\frac{3}{4n}$	0.0075
Cent. mom.	3	-	$\frac{3}{4n^2}$	0.000075
All cent. mom.	∞	+		$1.03 \cdot 10^{-24}$
Chernoff	∞	+	$3^{-\frac{n}{2}}$	$1.39 \cdot 10^{-24}$

For $n = 100$, $E(X) = np = 25$ and

$$P(X \geq 3n/4) = P(X \geq 75) = 1.4 \cdot 10^{-25}$$

Markov related inequalities

$\breve{g}(x)$ is *non-negative*, *monotone decreasing* for $x < a$ and $X \in \mathbb{R}$ then

$$Pr(X \leq a) \leq \frac{E(\breve{g}(X))}{\breve{g}(a)}$$

Proof for continuous X:

$$E(\breve{g}(X)) = \int_{-\infty}^{\infty} \breve{g}(x) f_X(x) dx \overset{\text{non-neg.}}{\geq} \int_{-\infty}^{a} \breve{g}(x) f_X(x) dx$$

$$\overset{\text{mon. dec.}}{\geq} \int_{-\infty}^{a} \breve{g}(a) f_X(x) dx = \breve{g}(a) \int_{-\infty}^{a} f_X(x) dx$$

$$= \breve{g}(a) Pr(X \leq a)$$

Chernoff lower bound

If $g(x) = e^{-sx}$ and $s > 0$ then

$$Pr(X \leq a) \leq \frac{E(e^{-sX})}{e^{-sa}},$$

where $L_X(s) = E(e^{-sX})$ is the Laplace transform of X.

If $L_X(s)$ is known then

$$Pr(X \leq a) \leq \min_{s \in \mathbb{R}^+} \frac{L_X(s)}{e^{-sa}},$$

Table of content

Central limit theorem (CLT)

"sum of i.i.d. rv-s converges to normal distribution"

Sample average:

$$S_n = \frac{X_1 + \ldots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n}$$

It converges to $\lim_{n \to \infty} S_n = E(X)$.

But how fast does it converge?

How many samples needed to approximate $E(X)$.

Variance of S_n:

$$Var(S_n) = \sum_{i=1}^{n} Var\left(\frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{Var(X_i)}{n^2} = \frac{Var(X)}{n}$$

Central limit theorem (CLT)

$$\lim_{n\to\infty} S_n - E(X) = 0$$

$$\lim_{n\to\infty} n(S_n - E(X)) =??$$

$$\lim_{n\to\infty} \sqrt{n}(S_n - E(X)) \stackrel{d}{=} N(0, \sigma_X^2)$$

Equivalently

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} X_i - nE(X) \right) \stackrel{d}{=} N(0, \sigma_X^2)$$

Lyapunov's Central limit theorem (LCLT)

Generalization of the CLT:

The sum of independent but *differently* distributed rv-s, $\sum_{i=1}^{n} X_i$, also converges to normal distribution with mean $\sum_{i=1}^{n} E(X_i)$ and variance $\sum_{i=1}^{n} Var(X_i)$, if

$$\lim_{n\to\infty} \left(\sum_{i=1}^{n} \sigma_i^2\right)^{-\delta} \sum_{i=1}^{n} E(|X_i - \mu_i|^{2+\delta}) = 0$$

for $\forall \delta > 0$, where $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$.

Hard to check condition, hard to predict convergence speed.

Application of LCLT

Application of LCLT for the cash-flow problem:

- Let $X = \sum_{i=1}^{N} Y_i h_i$, than $E(X) = \sum_{i=1}^{N} p_i h_i$ and

$$Var(X) = \sum_{i=1}^{N} Var(Y_i) h_i^2 = \sum_{i=1}^{N} (p_i - p_i^2) h_i^2.$$

- Let Z is normal distributed with mean $\mu = E(X)$ and variance $\sigma^2 = Var(X)$.
- Let $\hat{Z} = \frac{Z-\mu}{\sigma}$, i.e. $\hat{Z}$ is standard normal distributed.

Than

$$Pr\left(\sum_{i=1}^{N} Y_i h_i > C\right) = Pr(X > C) \approx Pr(Z > C)$$
$$= Pr\left(\hat{Z} > \frac{C-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{C-\mu}{\sigma}\right).$$

Example

B_i is Bernoulli with $p = 1/4$.

$X = \sum_{i=1}^{n} B_i$ is Binomial(n, p) with $p = 1/4$.

$$P(X \geq 3n/4) = ???$$

For $n = 100$, $E(X) = np = 25$ and

$$Var\,(X) = 100\ Var\,(B) = 100\left(\frac{1}{4} - \frac{1}{16}\right)$$

$$P(X \geq 75) \approx 1 - \Phi\left(\frac{75 - E(X)}{\sqrt{Var\,(X)}}\right) = 1.34 \cdot 10^{-25}$$

while the exact results is

$$P(X \geq 3n/4) = P(X \geq 75) = 1.4 \cdot 10^{-25}.$$

In this case the CLT underestimates the risk!!!

Table of content

Sampling

The complexity of the risk analysis problem is $\mathcal{O}(2^N)$.

Shall we approximate the result based on partial information (sampling)?

$$\text{risk} = Pr\left(\sum_{i=1}^{N} Y_i h_i > C\right) = Pr\left(\mathbf{Y}\mathbf{h}^T > C\right)$$

$$= \sum_{\forall \mathbf{y}\in\{0,1\}^N} Pr(\mathbf{y}) \cdot Pr\left(\mathbf{y}\mathbf{h}^T > C\right)$$

$$= \sum_{\forall \mathbf{y}\in\mathcal{C}} Pr(\mathbf{y}) \cdot Pr\left(\mathbf{y}\mathbf{h}^T > C\right) + \sum_{\forall \mathbf{y}\in\bar{\mathcal{C}}} Pr(\mathbf{y}) \cdot \underbrace{Pr\left(\mathbf{y}\mathbf{h}^T > C\right)}_{0\le\cdot\le 1},$$

where $\mathbf{y} = \{y_1, \ldots, y_N\} \in \{0,1\}^N$ and $\mathcal{C} \subset \{0,1\}^N$.

$$\sum_{\forall \mathbf{y}\in\mathcal{C}} Pr(\mathbf{y}) \cdot Pr\left(\mathbf{y}\mathbf{h}^T > C\right) \le \text{risk}$$

$$\le \sum_{\forall \mathbf{y}\in\mathcal{C}} Pr(\mathbf{y}) \cdot Pr\left(\mathbf{y}\mathbf{h}^T > C\right) + 1 - \sum_{\forall \mathbf{y}\in\mathcal{C}} Pr(\mathbf{y}).$$

Li-Silvester method

Li, V. and J. Silvester. "Performance Analysis of Networks with Unreliable Components." IEEE Trans. Commun. 32 (1984): 1105-1110.

For a given complexity, $c = |\mathcal{C}|$, the tightest bounds are obtained when $\sum_{\forall \mathbf{y} \in \mathcal{C}} Pr(\mathbf{y})$ is maximal.

Order the $\mathbf{y}$ vectors with decreasing probabilities:

$Pr(\mathbf{y}^{(1)}) \geq Pr(\mathbf{y}^{(2)}) \geq \ldots \geq Pr(\mathbf{y}^{(c)}) \geq \ldots \geq Pr(\mathbf{y}^{(2^N)})$ and bound the risk based on the c most probable samples

$$\sum_{i=1}^{c} Pr(\mathbf{y}^{(i)}) \cdot Pr\left(\mathbf{y}^{(i)}\mathbf{h}^T > C\right) \leq \text{risk}$$

$$\leq \sum_{i=1}^{c} Pr(\mathbf{y}^{(i)}) \cdot Pr\left(\mathbf{y}^{(i)}\mathbf{h}^T > C\right) + 1 - \sum_{i=1}^{c} Pr(\mathbf{y}^{(i)}).$$

Problem: Efficient generation of the ordered $\mathbf{y}$ *vectors.*

Example

Same as before: $N = 100$, $p = 1/4$

The difference between the lower and upper bounds by the Li-Silvester method is $\Delta = 1 - \sum_{\forall \mathbf{y} \in \mathcal{C}} Pr(\mathbf{y})$

(independent of $\mathbf{h}$ and C)

	$p = 1/4$		$p = 1/100$
samples	Δ	$1 - \Delta$	Δ
1	~ 1	$3.2 \cdot 10^{-13}$	0.63
101	~ 1	$1.1 \cdot 10^{-11}$	0.26
5051	~ 1	$1.9 \cdot 10^{-10}$	0.079
166751	~ 1	$2.1 \cdot 10^{-9}$	0.018
$\vdots$			

Still very slow convergence.

Table of content

Random sampling

Monte Carlo simulation:

Generate random $\mathbf{y}$ samples according to the distribution of $\mathbf{y}$ and check if $\mathbf{y}\mathbf{h}^T > C$

If the generated samples are $\mathbf{y}_{rnd}^{(1)}, \mathbf{y}_{rnd}^{(2)}, \ldots, \mathbf{y}_{rnd}^{(S)}$ then

$$\text{risk} \approx \eta = \frac{1}{S} \sum_{s=1}^{S} \underbrace{\mathcal{I}\left(\mathbf{y}_{rnd}^{(s)}\mathbf{h}^T > C\right)}_{B_s}.$$

η is the sample average of S i.i.d. rv: $B_s = \begin{cases} 1 & \text{risk} \\ 0 & 1 - \text{risk} \end{cases}$

As discussed with CLT:

$$E(\eta) = E(B_s) = \text{risk},$$

$$Var\left(\eta\right) = \sum_{s=1}^{S} Var\left(\frac{B_s}{S}\right) = \sum_{s=1}^{S} \frac{Var\left(B_s\right)}{S^2} = \frac{Var\left(B\right)}{S},$$

with $Var\left(B\right) = \text{risk} - \text{risk}^2$.

Monte Carlo simulation

Algorithm:

- Sample generation
  - Generate S samples such that the elements of $\mathbf{y}_{rnd}^{(s)}$ are independent and $Pr(\mathbf{y}_{rnd}^{(s)}\mathbf{e_i}^T = 1) = p_i$ for all $i \leq N$ and $s \leq S$
- Risk estimation

$$\text{risk} \approx \eta = \frac{1}{S}\sum_{s=1}^{S}\mathcal{I}\left(\mathbf{y}_{rnd}^{(s)}\mathbf{h}^T > C\right).$$

Table of content

Stratified sampling

We need to explore $\{0,1\}^N$.

Decompose $\{0,1\}^N$ to I disjoint subsets $\mathcal{C}_1,\ldots,\mathcal{C}_I$ (that is $\bigcup_{i=1}^{I}\mathcal{C}_i=\{0,1\}^N$ and $\mathcal{C}_i\bigcap\mathcal{C}_j=\emptyset$ for $i\neq j$). Than by the LTP

$$\begin{aligned}\text{risk} &= Pr\left(\mathbf{y}\mathbf{h}^T > C\right)\\ &= \sum_{i=1}^{I} Pr\left(\mathbf{y}\mathbf{h}^T > C|\mathbf{y}\in\mathcal{C}_i\right) Pr\left(\mathbf{y}\in\mathcal{C}_i\right) = \sum_{i=1}^{I} \text{risk}_i\ p_i,\end{aligned}$$

where $p_i = Pr\left(\mathbf{y}\in\mathcal{C}_i\right)=\sum_{\forall\mathbf{y}\in\mathcal{C}_i} Pr(\mathbf{y})$ and $\text{risk}_i = Pr\left(\mathbf{y}\mathbf{h}^T > C|\mathbf{y}\in\mathcal{C}_i\right)$ is the risk in set $\mathcal{C}_i$.

Sample allocation scheme $S_1,\ldots,S_I$ ($\sum_{i=1}^{I} S_i = S$) then risk_i is approximated based on the series of random samples $\mathbf{y}_{rnd_i}^{(s)}\in\mathcal{C}_i$, $s=1,\ldots,S_i$ as

$$\text{risk}_i \approx \eta_i = \frac{1}{S_i}\sum_{s=1}^{S_i}\mathcal{I}\left(\mathbf{y}_{rnd_i}^{(s)}\mathbf{h}^T > C\right).$$

Stratified sampling

Simple Algorithm:

- Sample generation
  - For $i = 1, \ldots, I$, generate S_i samples such that $\mathbf{y} \in \mathcal{C}_i$
- Risk estimation

$$\text{risk} \approx \eta = \sum_{i=1}^{I} p_i \eta_i = \sum_{i=1}^{I} p_i \underbrace{\frac{1}{S_i} \sum_{s=1}^{S_i} \mathcal{I}\left(\mathbf{y}_{rnd_i}^{(s)} \mathbf{h}^T > C\right)}_{\text{sample average in } \mathcal{C}_i}.$$

Stratified sampling

Sampling in layers:

Let $\#\mathbf{y}$ be the number of ones in $\mathbf{y}$ and $\mathcal{C}_i = \{\mathbf{y} : \#\mathbf{y} = i\}$.

Layer 0:

$Pr(\mathbf{y} = \mathbf{0}) = \prod_{j=1}^{N}(1 - p_j)$

$p_{\mathbf{0}}^{(0)} = Pr(\mathbf{y} = \mathbf{0}|\mathbf{y} \in \mathcal{C}_0) = 1$

Layer 1:

$Pr(\mathbf{y} = e_i) = \frac{p_i}{1-p_i}\prod_{j=1}^{N}(1 - p_j)$

$p_i^{(1)} = Pr(\mathbf{y} = e_i|\mathbf{y} \in \mathcal{C}_1) = \frac{\frac{p_i}{1-p_i}\prod_{j=1}^{N}(1-p_j)}{\sum_{k=1}^{N}\frac{p_k}{1-p_k}\prod_{j=1}^{N}(1-p_j)} = \frac{\frac{p_i}{1-p_i}}{\sum_{k=1}^{N}\frac{p_k}{1-p_k}}$

- Sample generation in $\mathcal{C}_1$ according to $p_i^{(1)}$

Task: Compute the sample distribution in $\mathcal{C}_2$

Stratified sampling

Approximating the error of stratified sampling

η_i is the sample average of S_i i.i.d. rv: $B_s^{(i)} = \begin{cases} 1 & \text{risk}_i \\ 0 & 1 - \text{risk}_i \end{cases}$

That is $\eta_i = \frac{\sum_{s=1}^{S_i} B_s^{(i)}}{S_i}$, where $E(B_s^{(i)}) = \text{risk}_i$ and $Var\left(B_s^{(i)}\right) = E(B_s^{(i)}) - E(B_s^{(i)})^2 = \text{risk}_i - \text{risk}_i^2$.

$$E(\eta_i) = \frac{1}{S_i}\sum_{s=1}^{S_i} E(B_s^{(i)}) = \frac{1}{S_i}\sum_{s=1}^{S_i} \text{risk}_i = \text{risk}_i,$$

$$Var(\eta_i) = Var\left(\frac{\sum_{s=1}^{S_i} B_s^{(i)}}{S_i}\right) = \frac{1}{S_i^2}\sum_{s=1}^{S_i} Var(B_s^{(i)})$$

$$= \frac{1}{S_i^2}\sum_{s=1}^{S_i} \text{risk}_i - \text{risk}_i^2 = \frac{\text{risk}_i - \text{risk}_i^2}{S_i}.$$

Stratified sampling

Using the data of the strata and $\eta = \sum_{i=1}^{I} p_i \eta_i$ we can compute the mean and variance of η.

$$E(\eta) = E\left(\sum_{i=1}^{I} p_i \eta_i\right) = \sum_{i=1}^{I} p_i E(\eta_i) = \sum_{i=1}^{I} p_i \text{risk}_i = \text{risk},$$

$$Var\left(\eta\right) = Var\left(\sum_{i=1}^{I} p_i \eta_i\right) = \sum_{i=1}^{I} p_i^2 Var\left(\eta_i\right) = \sum_{i=1}^{I} p_i^2 \frac{\text{risk}_i - \text{risk}_i^2}{S_i}.$$

Stratified sampling

Optimal sample allocation

$$Var_S = \min_{\substack{S_1,\ldots,S_I \\ \sum_{i=1}^{I} S_i = S}} Var(\eta) = \min_{\substack{S_1,\ldots,S_I \\ \sum_{i=1}^{I} S_i = S}} \sum_{i=1}^{I} p_i^2 \frac{\text{risk}_i - \text{risk}_i^2}{S_i}.$$

For $I = 2$ and $s_i = \frac{S_i}{S}$, $c_i = \frac{p_i^2(\text{risk}_i - \text{risk}_i^2)}{S}$ for $i = 1, 2$,

$$Var_S = \min_{\substack{s_1,s_2 \\ s_1+s_2=1}} \frac{c_1}{s_1} + \frac{c_2}{s_2}.$$

Its minimum is obtained at $\frac{\sqrt{c_1}}{s_1} = \frac{\sqrt{c_2}}{s_2}$ that is $s_i = \frac{\sqrt{c_i}}{\sqrt{c_1}+\sqrt{c_2}}$.

Interpretation of c_i: $\underbrace{p_i^2}_{\text{importance}} \underbrace{(\text{risk}_i - \text{risk}_i^2)}_{\text{uncertainty}}$

Stratified sampling

Optimal sample allocation

$$Var_S = \min_{\substack{S_1,\dots,S_I \\ \sum_{i=1}^I S_i = S}} Var(\eta) = \min_{\substack{S_1,\dots,S_I \\ \sum_{i=1}^I S_i = S}} \sum_{i=1}^{I} p_i^2 \frac{\mathrm{risk}_i - \mathrm{risk}_i^2}{S_i}.$$

Let $s_i = \frac{S_i}{S}$, $c_i = \frac{p_i^2(\mathrm{risk}_i - \mathrm{risk}_i^2)}{S}$ for $i = 1, \dots, I$, then

$$Var_S = \min_{\substack{s_1,\dots,s_I \\ \sum_{i=1}^I s_i = 1}} \sum_{i=1}^{I} \frac{c_i}{s_i}.$$

Its minimum is obtained at $\frac{\sqrt{c_1}}{s_1} = \dots = \frac{\sqrt{c_I}}{s_I}$ that is

$$s_i = \frac{\sqrt{c_i}}{\sum_{j=1}^I \sqrt{c_j}}.$$

Stratified sampling

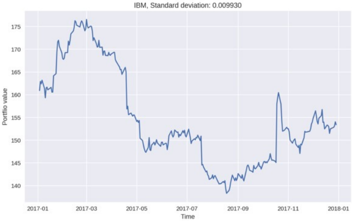
Approaches when the variance is not known.

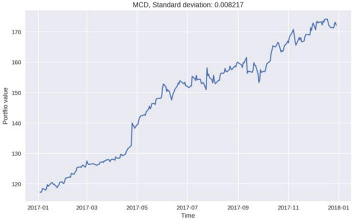
- variance free: $s_i = p_i$
- estimation/processing:
  approximate the variance based on the first S^* samples
- adaptive method:
  start with $s_i = p_i$
  in each step maintain $E(B^{(i)})$, $Var\left(B^{(i)}\right)$, and update s_i.

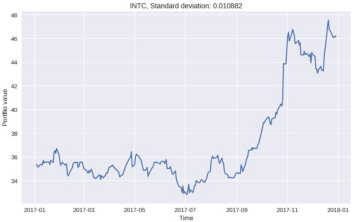
Table of content

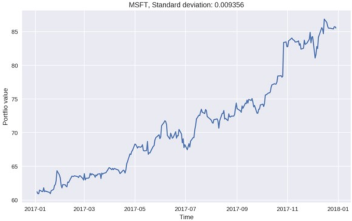
Portfolio

Which one to buy?









Portfolio

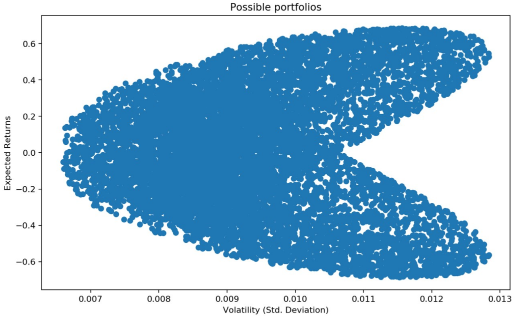
Which one to buy?

Portfolio: $[-0.043, 0.24, 0.29, -0.42]$

Portfolio

Possible answers:

- the one with the highest expected increase,
- the one with the least risk,



- combine income and risk:
  - e.g., minimize the risk for a given expected income,
  - e.g., maximize the income with a given risk level.

Model

- $s_i(t)$ - the price of asset i at time t,
- $r_i(t) = s_i(t) - s_i(t-1)$ - the profit of asset i at time t,

Assumption:

$\mathbf{r}(t) = \{r_1(t), \ldots, r_N(t)\}^T$ is time stationary, multi-dimensional normal distributed with location $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

To recap:

$E(r_i(t)) = \mu_i$ and $E((r_i(t) - \mu_i)(r_j(t) - \mu_j)) = \sigma_{ij}$ for $\forall t$.

Model

Portfolio:

w_i - amount of asset i, $\mathbf{w} = \{w_1, \ldots, w_N\}^T$.

- market value at time t:
  $p(t) = \sum_{i=1}^{N} w_i s_i(t) = \mathbf{w}^T \mathbf{s}(t)$
- income at time t:
  $x(t) = \sum_{i=1}^{N} w_i r_i(t)$
- expected income at time t (independent of t):

$$E(x(t)) = E\left(\sum_{i=1}^{N} w_i r_i(t)\right) = \sum_{i=1}^{N} w_i \mu_i = \mathbf{w}^T \boldsymbol{\mu}$$

- risk at time t (independent of t):
  variance of $x(t)$.

Model

Risk at time t:

$$\sigma^2(t) = E\left(\left(x(t) - E(x(t))\right)^2\right) = E\left(\left(\sum_{i=1}^{N} w_i\left(r_i(t) - \mu_i\right)\right)^2\right)$$

$$= E\left(\left(\sum_{i=1}^{N} w_i\left(r_i(t) - \mu_i\right)\right)\left(\sum_{j=1}^{N} w_j\left(r_j(t) - \mu_j\right)\right)\right)$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} w_i E\left(\left(r_i(t) - \mu_i\right)\left(r_j(t) - \mu_j\right)\right) w_j$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} w_i \sigma_{ij} w_j = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$$

It is independent of t.

Portfolio optimization

Minimize the risk for a given expected income (b):

$$\mathbf{w}_{opt} = \underset{\mathbf{w}\,:\,\mathbf{w}^T\boldsymbol{\mu}=b}{\operatorname{argmin}}\ \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$$

Let $\boldsymbol{\Sigma} = \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \mathbf{x}_i^T$ be the spectral decomposition of $\boldsymbol{\Sigma}$ (symmetric), such that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$ and $\mathbf{x}_i^T \mathbf{x}_j = \delta_{ij}$.

$$\mathbf{w}^T\boldsymbol{\mu} = \sum_{i=1}^{N} \underbrace{\mathbf{w}^T\mathbf{x}_i}_{v_i} \underbrace{\mathbf{x}_i^T\boldsymbol{\mu}}_{\hat{\mu}_i} = \sum_{i=1}^{N} v_i \hat{\mu}_i = \mathbf{v}^T\hat{\boldsymbol{\mu}}$$

$$\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} = \sum_{i=1}^{N} \mathbf{w}^T\mathbf{x}_i\lambda_i\mathbf{x}_i^T\mathbf{w} = \sum_{i=1}^{N} v_i\lambda_i v_i = \mathbf{v}^T\boldsymbol{\Lambda}\mathbf{v}$$

Transformed problem (quadratic optimization with linear constraint):

$$\mathbf{v}_{opt} = \underset{\mathbf{v}\,:\,\mathbf{v}^T\hat{\boldsymbol{\mu}}=b}{\operatorname{argmin}}\ \mathbf{v}^T\boldsymbol{\Lambda}\mathbf{v} = \underset{\mathbf{v}\,:\,\sum_{i=1}^{N} v_i\hat{\mu}_i=b}{\operatorname{argmin}} \sum_{i=1}^{N} \lambda_i v_i^2$$

Portfolio optimization

Reverse problem definition:

Maximize the expected income for a given risk (r):

$$\mathbf{w}_{opt} = \underset{\mathbf{w}\,:\,\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}=r}{\operatorname{argmax}} \mathbf{w}^T\boldsymbol{\mu}$$

Transformed problem (linear optimization with quadratic constraint):

$$\mathbf{v}_{opt} = \underset{\mathbf{v}\,:\,\mathbf{v}^T\boldsymbol{\Lambda}\mathbf{v}=r}{\operatorname{argmax}} \mathbf{v}^T\hat{\boldsymbol{\mu}} = \underset{\mathbf{v}\,:\,\sum_{i=1}^N \lambda_i v_i^2=r}{\operatorname{argmax}} \sum_{i=1}^N v_i\hat{\mu}_i$$

Portfolio optimization

Modified optimization problem:

$$\mathbf{w}_{opt} = \underset{\mathbf{w}\,:\,||\mathbf{w}||_2=1}{\text{argmin}}\ \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$$

where $||\mathbf{w}||_2 = \sqrt{\sum_{i=1}^N w_i^2} = \sqrt{\mathbf{w}^T\mathbf{w}}$.

$$\mathbf{w}^T\mathbf{w} = \sum_{i=1}^N \underbrace{\mathbf{w}^T\mathbf{x}_i}_{v_i} \underbrace{\mathbf{x}_i^T\mathbf{w}}_{v_i} = \sum_{i=1}^N v_i v_i = \mathbf{v}^T\mathbf{v}$$

Transformed problem (linear optimization in v_i^2):

$$\mathbf{v}_{opt} = \underset{\mathbf{v}\,:\,\mathbf{v}^T\mathbf{v}=1}{\text{argmin}}\ \mathbf{v}^T \boldsymbol{\Lambda} \mathbf{v} = \underset{\mathbf{v}\,:\,\sum_{i=1}^N v_i^2=1}{\text{argmin}} \sum_{i=1}^N \lambda_i v_i^2$$

Optimal solution is $\mathbf{v}_{opt}^T = \{1, 0, \ldots, 0\}$, $\mathbf{w}_{opt} = \mathbf{x}_1$.

Obtaining $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

Form the samples

- $s_i(t)$ - the price of asset i at time t,
- $r_i(t) = s_i(t) - s_i(t-1)$ - the profit of asset i at time t,

the sample mean vector and sample covariance matrix are

$$\tilde{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}(t),$$

$$\tilde{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}(t)\mathbf{r}(t)^T - \tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T.$$

Computing the required eigenvector

How to obtain $\mathbf{x}_1$, the eigenvector of the minimal eigenvalue of $\boldsymbol{\Sigma}$ by the iterative procedure providing the maximal eigenvalue/eigenvector?

- Apply the iterative procedure for $\boldsymbol{\Sigma}^{-1}$.
- In 2 steps:
  - compute λ_N by the iterative procedure for $\boldsymbol{\Sigma}$,
  - apply the iterative procedure for $\lambda_N \mathbf{I} - \boldsymbol{\Sigma}$.

On the fly approximation

Form the $\mathbf{r}(t) = \{r_1(t), \ldots, r_N(t)\}^T$ time stationary, multi-dimensional normal distributed samples for $t = 1, 2, \ldots, T$ with location $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, compute

Input: $\mathbf{w_{init}}, \mathbf{r}(t), T, \eta$; ($\eta$ – convergence speed)
$\mathbf{w} = \mathbf{w_{init}}$; $\mathbf{s} = \mathbf{0}$; $y_s = 0$
for $t = 1$ to T **do**
    $\mathbf{s} = \mathbf{s} + \mathbf{r}(t)$;
    $\mathbf{v} = \mathbf{r}(t) - \mathbf{s}/t$;
    $y = \mathbf{w}^T \mathbf{v}$;
    $y_s = y_s + y^2$;
    $\mathbf{w} = \mathbf{w} + \eta y(\mathbf{v} - y\mathbf{w})$;
end for
return : $\mathbf{s}/T$, y_s/T, $\mathbf{w}$;

where

- $\mathbf{s}/T$ approximates the mean $\boldsymbol{\mu}$,
- y_s/T approximates the dominant eigenvalue of $\boldsymbol{\Sigma}$,
- $\mathbf{w}$ approximates the dominant eigenvector of $\boldsymbol{\Sigma}$.

On the fly approximation

Assuming, $\mathbf{v} = \mathbf{r}(t) - \boldsymbol{\mu}$ and using $y = \mathbf{w}^T\mathbf{v} = \mathbf{v}^T\mathbf{w}$,

the expected change of $\mathbf{w}$ is

$$E\left(\eta y(\mathbf{v} - y\mathbf{w})\right) = \eta E\left(\mathbf{v}\,\underbrace{\mathbf{v}^T\mathbf{w}}_{y} - \underbrace{\mathbf{w}^T\mathbf{v}}_{y}\,\underbrace{\mathbf{v}^T\mathbf{w}}_{y}\,\mathbf{w}\right)$$

$$= \eta\,\underbrace{E\left((\mathbf{r}(t) - \boldsymbol{\mu})(\mathbf{r}(t) - \boldsymbol{\mu})^T\right)}_{\boldsymbol{\Sigma}}\,\mathbf{w}$$

$$- \eta\mathbf{w}^T\,\underbrace{E\left((\mathbf{r}(t) - \boldsymbol{\mu})(\mathbf{r}(t) - \boldsymbol{\mu})^T\right)}_{\boldsymbol{\Sigma}}\,\mathbf{w}\mathbf{w}$$

$$= \eta\boldsymbol{\Sigma}\mathbf{w} - \eta\,\underbrace{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}_{c\,:\,\text{scalar}}\,\mathbf{w}$$

$$= \eta\left(\boldsymbol{\Sigma}\mathbf{w} - c\mathbf{w}\right).$$

On the fly approximation

A necessary condition for the convergence of the iteration is that the expected change of $\mathbf{w}$ converges to $\mathbf{0}$.

It holds when

$$\mathbf{\Sigma w} - c\mathbf{w} = 0,$$

that is c and $\mathbf{w}$ are eigenvalue and eigenvector pair of $\mathbf{\Sigma}$ and $\mathbf{w}^T\mathbf{w} = 1$.

$\mathbf{w}^T\mathbf{w} = 1$, because

$$c = \mathbf{w}^T \underbrace{\mathbf{\Sigma w}}_{c\mathbf{w}} = c\mathbf{w}^T\mathbf{w}.$$

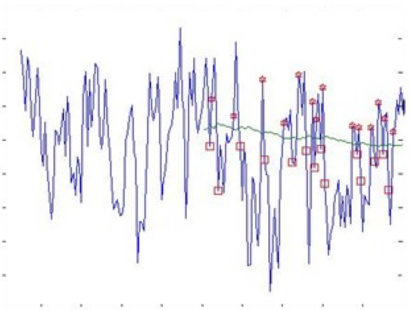
Table of content

Mean reversion

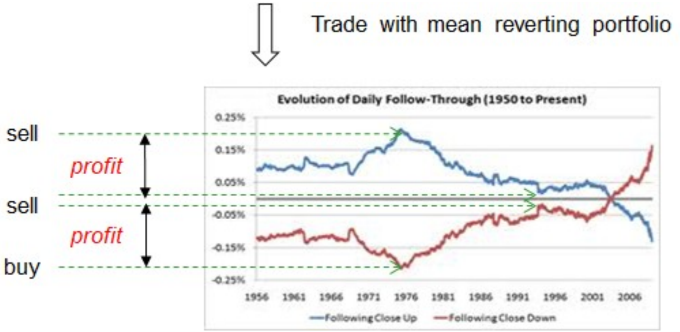
The tendency of a market variable (such as an interest rate) to revert back to some long-run average level.

A potential economic explanation for interest rate:

- increased interest rate,
- economic slowdown,
- low demand for funds,
- interest rates decreases.



Trade with the mean reverting portfolio



- far above/below mean: sell/buy
- back to mean from above/below: buy/sell

Table of content

Ornstein-Uhlenbeck model

Assume

- $s_i(t)$: the price of asset i at time t,
- w_i: amount of asset i,
- $p(t)$ market value at time t: $p(t) = \sum_{i=1}^{N} w_i s_i(t) = \mathbf{w}^T \mathbf{s}(t)$.

Mathematical model for continuous time behaviour (described by a stochastic differential equation)

$$dp(t) = \lambda(\mu - p(t))dt + \sigma dW(t),$$

where

- μ: is the mean (long time average),
- λ: mean reversion coefficient (the force to return to the mean),
- $W(t)$: Wiener process (normalized noise),
- σ: volatility (volume of noise).

Wiener process:

- independent increments,
- $W(t+\Delta) - W(t)$ is $N(0,\Delta)$ normal distributed.

Ornstein-Uhlenbeck model

Integrating the stochastic differential equation:

$$p(t) = p(0)e^{-\lambda t} + \mu(1 - e^{-\lambda t}) + \int_{s=0}^{t} \sigma e^{-\lambda(t-s)} dW(s).$$

From which

$$E(p(t)) = p(0)e^{-\lambda t} + \mu(1 - e^{-\lambda t}).$$

I.e. $E(p(t))$ exponentially converges to the mean with rate λ.

Limiting behaviour:

$$\lim_{t\to\infty} p(t) \sim N(\mu, \frac{\sigma^2}{2\lambda}).$$

Ornstein-Uhlenbeck model

Relation of the integral and differential forms:

$$p(t) = p(0)e^{-\lambda t} + \mu(1 - e^{-\lambda t}) + e^{-\lambda t}\int_{s=0}^{t} \sigma e^{\lambda s}\frac{dW(s)}{ds}ds$$

$$\begin{aligned}\frac{d}{dt}p(t) = & -\lambda p(0)e^{-\lambda t} + \lambda\mu e^{-\lambda t} - \lambda e^{-\lambda t}\int_{s=0}^{t} \sigma e^{\lambda s}\frac{dW(s)}{ds}ds \\ & + e^{-\lambda t}\sigma e^{\lambda t}\frac{dW(t)}{dt} \\ = & \underbrace{-\lambda p(0)e^{-\lambda t} - \lambda\mu(1 - e^{-\lambda t}) - \lambda e^{-\lambda t}\int_{s=0}^{t} \sigma e^{\lambda s}\frac{dW(s)}{ds}ds}_{-\lambda p(t)} \\ & + \sigma\frac{dW(t)}{dt} + \lambda\mu\end{aligned}$$

Table of content

Autoregressive model

Assume $s_i(t)$ is the price of asset i at time step t.

Mathematical model for asset prices in discrete time instants:

$$\begin{aligned} \mathbf{s}(t) - \boldsymbol{\mu} &= \mathbf{A}(\mathbf{s}(t-1) - \boldsymbol{\mu}) + \boldsymbol{\omega}(t) && \text{evolution} \\ \mathbf{s}(t) - \mathbf{s}(t-1) &= (\mathbf{I} - \mathbf{A})(\boldsymbol{\mu} - \mathbf{s}(t-1)) + \boldsymbol{\omega}(t) && \text{OU diff. form} \\ \mathbf{s}(t) &= \mathbf{A}\mathbf{s}(t-1) + (\mathbf{I} - \mathbf{A})\boldsymbol{\mu} + \boldsymbol{\omega}(t) && \text{AR(1) form} \end{aligned}$$

where

- $\mathbf{A}$: modification of prices in one time step.
- $\boldsymbol{\omega}(t)$: noise in time step t.

AR(1) model, because only $\mathbf{s}(t-1)$ affects $\mathbf{s}(t)$ (directly).

Autoregressive model

Assumptions

- $\mathbf{s}(t)$ is stationary, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $E((\mathbf{s}(t)-\boldsymbol{\mu})(\mathbf{s}(t)-\boldsymbol{\mu})^T) = \boldsymbol{\Sigma}$ for $\forall t$.
- $\boldsymbol{\omega}(t)$ is multivariate normal with mean $\mathbf{0}$ and covariance $\boldsymbol{\Theta}$

Condition of stability: $sp(\mathbf{A}) < 1$

Covariance relation based on the evolution form:

$$\begin{aligned} Var\left(\mathbf{s}(t)-\boldsymbol{\mu}\right) =& Var\left(\mathbf{A}(\mathbf{s}(t-1)-\boldsymbol{\mu})\right) + Var\left(\boldsymbol{\omega}(t)\right) \\ E((\mathbf{s}(t)-\boldsymbol{\mu})(\mathbf{s}(t)-\boldsymbol{\mu})^T) =& \mathbf{A}E((\mathbf{s}(t-1)-\boldsymbol{\mu})(\mathbf{s}(t-1)-\boldsymbol{\mu})^T)\mathbf{A}^T \\ &+ E(\boldsymbol{\omega}(t)\boldsymbol{\omega}(t)^T) \\ \boldsymbol{\Sigma} =& \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T + \boldsymbol{\Theta} \end{aligned}$$

Autoregressive model

Market value of portfolio $\mathbf{w}$ at time step t:

$$p(t) = \mathbf{w}^T \mathbf{s}(t).$$

Mean of portfolio $\mathbf{w}$ (independent of t):

$$\mu = E(p(t)) = \mathbf{w}^T E(\mathbf{s}(t)) = \mathbf{w}^T \boldsymbol{\mu}.$$

Market value of the autoregressive model

$$p(t) = \mathbf{w}^T \mathbf{s}(t) = \underbrace{\mathbf{w}^T \mathbf{A}\mathbf{s}(t-1)}_{\text{past effect}} + \underbrace{\mathbf{w}^T (\mathbf{I} - \mathbf{A})\boldsymbol{\mu}}_{\text{constant}} + \underbrace{\mathbf{w}^T \boldsymbol{\omega}(t)}_{\text{noise}},$$

and its variance

$$\begin{aligned} Var(p(t)) &= \underbrace{Var(\mathbf{w}^T \mathbf{s}(t))}_{\sigma} = E\left((\mathbf{w}^T (\mathbf{s}(t) - \boldsymbol{\mu}))^2\right) \\ &= \underbrace{Var(\mathbf{w}^T \mathbf{A}\mathbf{s}(t-1))}_{\sigma_{past}} + \underbrace{Var(\mathbf{w}^T \boldsymbol{\omega}(t))}_{\sigma_{noise}} \\ &= E\left((\mathbf{w}^T \mathbf{A}(\mathbf{s}(t-1) - \boldsymbol{\mu}))^2\right) + \sigma_{noise}. \end{aligned}$$

Optimal portfolio

Predictability factor of portfolio $\mathbf{w}$

$$v(\mathbf{w}) = \frac{\sigma_{past}}{\sigma} = \frac{\sigma_{past}}{\sigma_{past} + \sigma_{noise}} = \frac{Var(\mathbf{w}^T \mathbf{A}\mathbf{s}(t-1))}{Var(\mathbf{w}^T \mathbf{s}(t))}$$

$$= \frac{E(\mathbf{w}^T \mathbf{A}(\mathbf{s}(t-1)-\boldsymbol{\mu})(\mathbf{s}(t-1)-\boldsymbol{\mu})^T \mathbf{A}^T \mathbf{w})}{E(\mathbf{w}^T (\mathbf{s}(t)-\boldsymbol{\mu})(\mathbf{s}(t)-\boldsymbol{\mu})^T \mathbf{w})}$$

$$= \frac{\mathbf{w}^T \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T \mathbf{w}}{\mathbf{w}^T \boldsymbol{\Sigma}\mathbf{w}}$$

- $v(\mathbf{w})$ large – noise is small $\rightarrow$ $p(t)$ is predictable
- $v(\mathbf{w})$ small – noise is large $\rightarrow$ $p(t)$ is unpredictable

Optimal portfolio:

$$\mathbf{w}_{opt} = \underset{\mathbf{w}}{\operatorname{argmax}}\, v(\mathbf{w}) = \underset{\mathbf{w}}{\operatorname{argmax}} \frac{\mathbf{w}^T \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T \mathbf{w}}{\mathbf{w}^T \boldsymbol{\Sigma}\mathbf{w}}$$

Optimal portfolio

$$\mathbf{w}_{opt} = \underset{\mathbf{w}}{\text{argmax}} \frac{\mathbf{w}^T \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T \mathbf{w}}{\mathbf{w}^T \boldsymbol{\Sigma}\mathbf{w}}$$

Let $\boldsymbol{\Sigma} = \mathbf{B}^T \boldsymbol{\Lambda}\mathbf{B}$ be the spectral decomposition of $\boldsymbol{\Sigma}$ (symmetric), such that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$.

$$\mathbf{w}^T \boldsymbol{\Sigma}\mathbf{w} = \underbrace{\mathbf{w}^T \mathbf{B}^T \boldsymbol{\Lambda}^{1/2^T}}_{\mathbf{v}^T} \underbrace{\boldsymbol{\Lambda}^{1/2}\mathbf{B}\mathbf{w}}_{\mathbf{v}} = \mathbf{v}^T \mathbf{v}$$

$$\mathbf{w}^T \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T \mathbf{w} = \underbrace{\mathbf{w}^T \mathbf{B}^T \boldsymbol{\Lambda}^{1/2^T}}_{\mathbf{v}^T} \underbrace{\boldsymbol{\Lambda}^{-1/2}\mathbf{B}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T \mathbf{B}^T \boldsymbol{\Lambda}^{-1/2^T}}_{\hat{\mathbf{A}}} \underbrace{\boldsymbol{\Lambda}^{1/2}\mathbf{B}\mathbf{w}}_{\mathbf{v}}$$

$$= \mathbf{v}^T \hat{\mathbf{A}}\mathbf{v}$$

Transformed problem with $\mathbf{v} = \boldsymbol{\Lambda}^{1/2}\mathbf{B}\mathbf{w}$:

$$\mathbf{v}_{opt} = \underset{\mathbf{v}}{\text{argmax}} \frac{\mathbf{v}^T \hat{\mathbf{A}}\mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \underset{\mathbf{v}\,:\,||\mathbf{v}||_2 = 1}{\text{argmax}} \mathbf{v}^T \hat{\mathbf{A}}\mathbf{v}$$

Table of content

Model identification

Model: $\mathbf{s}(t) - \boldsymbol{\mu} = \mathbf{A}(\mathbf{s}(t-1) - \boldsymbol{\mu}) + \boldsymbol{\omega}(t)$,
with $\mathbf{s}(t) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\omega}(t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta})$

Auto covariance matrix:

$$\mathbf{R}(k) = E((\mathbf{s}(t) - \boldsymbol{\mu})(\mathbf{s}(t-k) - \boldsymbol{\mu})^T)$$

is asymmetric in general ($\mathbf{R}(k) = \mathbf{R}(-k)^T$),

but $\mathbf{R}(0) = \boldsymbol{\Sigma}$ is symmetric.

One step auto covariance matrix:

$$\begin{aligned}
\mathbf{R} = \mathbf{R}(1) &= E((\mathbf{s}(t) - \boldsymbol{\mu})(\mathbf{s}(t-1) - \boldsymbol{\mu})^T) \\
&= E((\mathbf{A}(\mathbf{s}(t-1) - \boldsymbol{\mu}) + \boldsymbol{\omega}(t))(\mathbf{s}(t-1) - \boldsymbol{\mu})^T) \\
&= E((\mathbf{A}(\mathbf{s}(t-1) - \boldsymbol{\mu}))(\mathbf{s}(t-1) - \boldsymbol{\mu})^T) + E(\boldsymbol{\omega}(t)(\mathbf{s}(t-1) - \boldsymbol{\mu})^T) \\
&= \mathbf{A}E((\mathbf{s}(t-1) - \boldsymbol{\mu})(\mathbf{s}(t-1) - \boldsymbol{\mu})^T) + \mathbf{0} \\
&= \mathbf{A}\boldsymbol{\Sigma}
\end{aligned}$$

Model identification

Model: $\mathbf{s}(t)-\boldsymbol{\mu}=\mathbf{A}(\mathbf{s}(t-1)-\boldsymbol{\mu})+\boldsymbol{\omega}(t)$.

Observations: $\mathbf{s}(t)$ for $t=1,\ldots,T$.

Model identification

- $\tilde{\boldsymbol{\mu}}=\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}(t)$
- $\tilde{\boldsymbol{\Sigma}}=\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{s}(t)\mathbf{s}^T(t)\right)-\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T$ (symmetric)
- $\tilde{\mathbf{R}}=\left(\frac{1}{T-1}\sum_{t=2}^{T}\mathbf{s}(t)\mathbf{s}^T(t-1)\right)-\tilde{\boldsymbol{\mu}}\tilde{\boldsymbol{\mu}}^T$
- $\tilde{\mathbf{A}}=\tilde{\mathbf{R}}\tilde{\boldsymbol{\Sigma}}^{-1}$
- $\tilde{\boldsymbol{\Theta}}=\tilde{\boldsymbol{\Sigma}}-\tilde{\mathbf{A}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{A}}^T=\tilde{\boldsymbol{\Sigma}}-\tilde{\mathbf{R}}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{R}}^T$

Mean reverting portfolio trading

Assuming $\mathbf{s}(t) = \mathbf{A}\mathbf{s}(t-1) + \boldsymbol{\omega}(t)$,
with $\mathbf{s}(t) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\omega}(t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta})$

Input: T, $\mathbf{s}(1), \mathbf{s}(2), \ldots$;
Compute $\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{A}}, \tilde{\boldsymbol{\Theta}}$ from $\mathbf{s}(1), \ldots, \mathbf{s}(T)$;
Compute $\mathbf{v}_{opt}$ from $\tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{A}}, \tilde{\boldsymbol{\Theta}}$;
$\mathbf{w}_{opt} = \mathbf{B}^T {\boldsymbol{\Lambda}^{1/2}}^T \mathbf{v}_{opt}$; $\mu = \mathbf{w}_{opt}^T \tilde{\boldsymbol{\mu}}$;
$Short = TRUE$;
for $t = T+1$ to ∞ **do**
  if $Short$ && $\mathbf{w}_{opt}^T \mathbf{s}(t) < \mu - \Delta$ **then**
    BUY; $Short = FALSE$;
  end if
  if $\overline{Short}$ && $\mathbf{w}_{opt}^T \mathbf{s}(t) > \mu + \Delta$ **then**
    SELL; $Short = TRUE$;
  end if
end for

Table of content

Table of content

Binomial Options Pricing Model

Assumptions

- discrete time (lattice based),
- price can take 2 new values in each step (up-down).

General binary tree:

- $S_{i+1}^u = S_i u_i$, $S_{i+1}^d = S_i d_i$, step dependent up/down ratio.
- $p_i = Pr(\text{up at step } i)$, step dependent up/down probability.
  $1 - p_i = Pr(\text{down at step } i)$

There are 2^N leaves of the tree.

Leaves are characterized by the binary vector $\mathbf{y} = \{y_1, \ldots, y_N\}$ with $y_i = 1$ indicating the upper price in step i.

The price at leaf $\mathbf{y}$ is $S_{\mathbf{y}} = S_0 \prod_{i:y_i=1} u_i \prod_{i:y_i=0} d_i$

Binomial tree

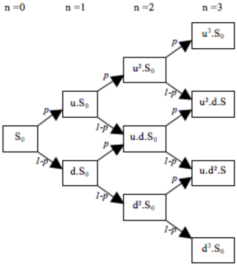
Binomial tree:

- $u_i = u$, $d_i = 1/u$, with step independent up/down ratio.
- $p_i = p$ step independent up/down probability.

There are $N + 1$ leaves of the tree.

Leaves are characterized by the number of steps with upper prices, $\hat{n}$.

The price at leaf $\hat{n}$ is $S_{\mathbf{y}} = S_0 u^{\hat{n}} d^{N-\hat{n}} = S_0 u^{2\hat{n}-N}$.



Binomial tree

To approximate the continuous distributed asset prices at time t in n steps with risk free rate r and volatility σ let

- $1+\bar{r}=e^{rt/n}$, per step risk free rate,
- $u=e^{\sigma\sqrt{t/n}}$, $d=1/u$, per step up/down price change,
- $p=\frac{e^{rt/n}-d}{u-d}$, per step up probability.

Binomial Options Pricing Model

- The strike price is X,
- the asset price at time t is S_t,
- the call option price at time t is C_t.

Time line:

- present – root,
- future before maturity – internal nodes,
- maturity time – leaves.

Option valuation is a three-step process:

- price tree generation (from root to leaves),
- calculation of option value at each leaf node: $C_{\text{leaf}} = \max(S_{\text{leaf}} - X, 0)$,
- sequential calculation of the option value at each preceding node (from leaves to root).

Calculation of the option value at a node

Risk neutrality assumption:

- today's asset price represents the expected asset value discounted at the risk free rate,
  that is $S = \frac{pS_u+(1-p)S_d}{1+\bar{r}} = \frac{puS+(1-p)dS}{1+\bar{r}}$
  from which $p = \frac{(1+\bar{r})-d}{u-d}$,
- today's call value represents the expected call value discounted at the risk free rate,
  that is $C = \frac{pC_u+(1-p)C_d}{1+\bar{r}}$.

Arbitrage-free pricing (delta-hedging):

- compute the portfolio for which both outcome (up/down) results the same pay off:

$$\Delta S_u - B(1+\bar{r}) = C_u$$
$$\Delta S_d - B(1+\bar{r}) = C_d,$$

- Solve these equations for Δ and B, and $C = \Delta S - B$.

The assumptions provide identical option value.

Arbitrage-free pricing (detailed)

Input:

- Current and future asset prices: S, S_u/S_d,
- Strike price: X, Risk free rate: $1+\bar{r}$,
- Future option prices: C_u/C_d,

Output: Current option prices C.

Hedging

- Make a future value independent portfolio:
  Δ asset and call option
- Future value of this portfolio:
  - in case of S_u : $\Delta S_u - C_u$
  - in case of S_d : $\Delta S_d - C_d$
- From the identity of the two cases

$$\Delta = \frac{C_u - C_d}{S_u - S_d}$$

- The current value of this port folio is: $\Delta S - C$
- The discounted future value of the portfolio is

$$B = \frac{\Delta S_u - C_u}{1+\bar{r}} = \frac{\Delta S_d - C_d}{1+\bar{r}}$$

- From the identity of the last two: $C = \Delta S - B$.

Calculation of the option value at a node

Vanilla options:

- European option: option can be exercised on the maturity date only,
- American option: option can be exercised any time up to the maturity date,

Option value at a node

- European option: $C_{Eur} = C$,
- American option: $C_{Am} = \max(C, S - X)$,

where C is computed as above, S is the asset value at the node and X is the strike price.

Table of content

Black–Scholes Options Pricing Model

- The strike price is X,
- the asset price at time t is S_t.

Payoff of the call option at maturity is $\max(S_t - X, 0)$.

S_t is a random variable with CDF $F(x) = Pr(S_t < x)$ and PDF $f(x)$. The expected payoff of the call option at maturity is

$$\Omega = \int_{x=0}^{\infty} \max(x - X, 0) f(x) dx$$

$$= \int_{x=0}^{X} \underbrace{\max(x - X, 0)}_{0} f(x) dx + \int_{x=X}^{\infty} \underbrace{\max(x - X, 0)}_{x-X} f(x) dx$$

$$= \int_{X}^{\infty} x f(x) dx - X \int_{X}^{\infty} f(x) dx$$

$$= E(S_t | S_t > X)(1 - F(X)) - X(1 - F(X))$$

$$= (E(S_t | S_t > X) - X)(1 - F(X)).$$

Black–Scholes Options Pricing Model

- The risk free interest rate is r,
- the (current) price of the call option is C.

The expected payoff of the call option at maturity is Ω.

The discounted expected payoff with risk free interest rate r is Ωe^{-rt}.

That is the current price of the call option $C = \Omega e^{-rt}$.

Black–Scholes Options Pricing Model

Under the assumptions of the BS model

- ...,
- the underlying process follows a geometric Brownian motion with constant drift and volatility.

S_t is a lognormal distributed with parameters μt and $\sigma\sqrt{t}$, where σ is referred to as volatility of S_t.

That is, $\log S_t$ is normal distributed with mean μt and standard deviation $\sigma\sqrt{t}$.

Consequently, its PDF is $f(x) = \frac{1}{x\sigma\sqrt{2\pi t}} e^{-\frac{(\log x - \mu t)^2}{2t\sigma^2}}$ and $E(S_t) = e^{t(\mu+\sigma^2/2)}$.

The mean of the discounted future price is S_0, that is $S_0 = e^{-rt}E(S_t) = e^{t(\mu+\sigma^2/2-r)}$.

Substituting $f(x)$ into the expected payoff expression for Ω gives the BS formula (after some algebra).

Table of content

Next lesson

Still to add:

- algorithm for computing Y vectors of decreasing probability,
- algorithm for the optimal mean reverting portfolio on the fly.

Table of content

Test problems

- You are given **h**, **p** ($N \sim 3$) and C. Compute the *risk* by
  - brute force,
  - CLT,
  - Li-Silvester by n samples,
  - Tail approximation (Markov, Chebysev, Chernoff, moment, ...),
  - Monte Carlo (samples are given),
  - stratified sampling (samples are given).
- You are given $\mathbf{\Sigma}$ (2×2) and $\boldsymbol{\mu}$ of the portfolio problem. Compute
  - the spectral decomposition of $\mathbf{\Sigma}$,
  - min risk portfolio with unit norm,
  - min risk portfolio with income b,
  - the risk of portfolio $\mathbf{w}$.
- You are given the parameters of the mean reverting portfolio problem. Define and compute elements of the optimal portfolio.