Risk analysis and management

Telek Miklós
BME

October 28, 2019
Table of content

Administrative details

Linear algebra

Probability theory

Probability distributions

Statistical cash-flow management problem

Tail approximation methods

Central limit theorem

Sampling methods

Next lesson
Theory – practice

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

► Practice: Attila Ceffer
  I.b.116, phone: 3875, or ceffer@hit.bme.hu.
Summary of Linear algebra

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

- Commutativity of matrices
- Sylvester equation
  - \( vec \) operator, Kronecker product (\( \otimes \)),
  - \( vec(ABC) = (C^T \otimes A) vec(B) \),
- Matrix functions
  - definition,
  - spectral decomposition based interpretation.
Linear equation

Scalar linear equation: $ax = b$

- $a \neq 0 \rightarrow$ single solution: $x = b/a$.
- $a = 0$
  - $b = 0 \rightarrow$ infinite solutions: $x \in \mathbb{R}$,
  - $b \neq 0 \rightarrow$ 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

\[ Ax = b \]

with

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

Scalar description of the matrix equation:

\[
\sum_{j=1}^{2} a_{ij}x_j = b_i, \quad \text{for } i = 1, 2, 3.
\]
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: \( I = \{\delta_{ij}\} \),

where \( \delta_{ij} = \begin{cases} 
1, & i = j, \\
0, & i \neq j,
\end{cases} \)

is the Kronecker delta.

Diagonal matrix: \( D = \text{diag}\{d_1, \ldots, d_n\} \),

Unitary matrix: \( U^T U = U U^T = I \)
Commuting matrices

Commonly, $AB \neq BA$,
as a consequence several scalar identity fails for matrices, e.g.:

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

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

Exceptions:

$A, I, A^{-1}, A^n$ for $n \in \mathbb{N}$ and all of their linear combinations,

$\sum_{n=-\infty}^{\infty} c^n A^n$, always commute.

The usual scalar identities hold for commuting matrices.
Singular value decomposition (SVD)

\[
\begin{bmatrix}
A
\end{bmatrix}_{n \times m} = \begin{bmatrix}
U
\end{bmatrix}_{n \times n} \begin{bmatrix}
\Psi
\end{bmatrix}_{n \times m} \begin{bmatrix}
V
\end{bmatrix}_{m \times m}
\]

where \( U \) and \( V \) are unitary matrices,

\[
\begin{bmatrix}
\Psi
\end{bmatrix}_{n \times m} = \begin{bmatrix}
\sigma_1 \\
\sigma_2 \\
\end{bmatrix}_{n \times m}
\]

is a diagonal matrix composed of the \( \sigma_i \geq 0 \) singular values. \( \Psi \) is assumed to be ordered such that \( \sigma_1 \geq \sigma_2 \geq \ldots \).

If the number of non-zero singular values are \( r \) then

\[
r = \text{rank}A = \text{number of independent rows/columns of } A
\]
Ax = b has a solution if b is a linear combination of the columns of A.

That is

\[ \text{rank } \begin{bmatrix} A \\ \hline \end{bmatrix} = \text{rank } \begin{bmatrix} A & b \end{bmatrix} \]
Linear equations

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

$$A_{n \times m} \ x_{m \times 1} = b_{n \times 1}$$

can be written as

$$\Psi_{n \times m} \ x'_{m \times 1} = b'_{n \times 1}$$

$$\begin{bmatrix} S \\ \end{bmatrix} \cdot \begin{bmatrix} x'_1 \\ x'_2 \\ \end{bmatrix} = \begin{bmatrix} b'_1 \\ b'_2 \\ \end{bmatrix}.$$ 

with $x' = Vx$, $b' = U^Tb$ and block sizes $S_{r \times r}$, $x'_1_{r \times 1}$, $x'_{2m-r \times 1}$, $b'_{1r \times 1}$, $b'_{2n-r \times 1}$.

- If $n - r > 0$ and $b'_2 \neq 0$ then no solution.
- If $b'_2 = 0$ and $m - r = 0$ then the single solution is $x = V^TS^{-1}b'_1$.
- If $b'_2 = 0$ and $m - r > 0$ then there are infinite solutions of dimension $m - r$. 
Linear matrix equations

In some cases, a matrix of unknowns \( X \) and some matrices of coefficients form a linear matrix equation.

E.g., \( AX = B \).

In this case, if \( \exists A^{-1} \) then \( X = A^{-1}B \) is the solution.

This approach is not applicable for the Sylvester equation

\[
AX + XB = C
\]

We transform it into standard linear equation form using

- \( vec \) operator,
- Kronecker product (\( \otimes \)),
- \( vec(ABC) = (C^T \otimes A) vec(B) \),

\[
vec(AX + XB) = vec(AXI + IXB) = (I \otimes A + B^T \otimes I)vec(X) = vec(C)
\]
Spectral decomposition

\[ \mathbf{A} = \mathbf{U} \Lambda \mathbf{V} \] is the spectral decomposition of \( \mathbf{A} \) when \( \mathbf{U}^{-1} = \mathbf{V} \) and \( \Lambda \) 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 \( A \) is diagonalizable then

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

where \( u_i \) is the \( i \)th column of \( U \) and \( v_i \) is the \( i \)th row of \( V \).

Computing the spectral decomposition

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

Note that \( v_i u_j = \delta_{ij} \) due to \( VU = I \).
Iterative procedure for computing $\lambda^*$ and $u^*$

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

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

Evaluate the conditions when the procedure converges.
Matrix functions

If $A$ 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

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

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

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

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

Administrative details

Linear algebra

Probability theory
  Probability distributions
  Statistical cash-flow management problem
  Tail approximation methods
  Central limit theorem
  Sampling methods
  Next lesson

Next lesson
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):
  \[ \sigma_X^2 = E(X^2) - E(X)^2, \text{ cumulants} \]
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: $Pr(A|X = x) \rightarrow \lim_{\delta \to 0} Pr(A|x \leq X < x + \delta)$
Law of total probability

Application

$E(g(Y)) = \sum_i E(g(Y) \mid B_i) Pr(B_i)$,

- discrete condition:

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

- continuous condition:

$$E(g(Y)) = \int_{x} E(g(Y) \mid X = x) f_X(x) dx$$

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

Administrative details
Linear algebra

**Probability theory**

Probability distributions

Statistical cash-flow management problem
Tail approximation methods
Central limit theorem
Sampling methods
Next lesson
Distributions

One-parameter

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

- Continuous
  - Exponential

Two-parameter

- Discrete
  - Uniform
  - Binomial

- Continuous
  - Uniform
  - Normal
### Transforms

**Transforms:**

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

**Advantages:**

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

Dependent random variables \((X, Y)\)

- **cumulated 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: \(E(X^nY^m)\)

- and their descendants (e.g., covariance: \(E(XY) - E(X)E(Y)\), correlation)
Normal distribution

PDF of normal distribution with \((\mu, \sigma^2)\):

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

If \(X\) is normal distributed with \((\mu, \sigma^2)\), 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 \((\mu, \sigma^2)\):

\[
Pr(X < x) = \Phi_{\hat{X}} \left( \frac{x - \mu}{\sigma} \right).
\]
Multivariate normal distribution

Probability density function

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

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

where $E(X_i) = \mu_i$, $E(X_iX_j) - E(X_i)E(X_j) = \sigma_{ij}$ and $\Sigma$ is non-singular positive definite matrix.

Construction of multivariate normal distribution

- If $Z = \{Z_1, \ldots, Z_k\}^T$ is composed of i.i.d. standard normal distributed RVs, then $X = \mu + AZ$ multivariate normal $(\mu, \Sigma)$ distributed with $\Sigma = AA^T$. 
Table of content

Administrative details

Linear algebra

Probability theory

Probability distributions

Statistical cash-flow management problem

Tail approximation methods

Central limit theorem

Sampling methods

Next lesson
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
Special cases:

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

Brute force solution:
LTP completely eliminating the randomness

\[ Pr \left( \sum_{i=1}^{N} Y_i h_i > C \right) \]

\[ = \sum_{y_1=0}^{1} \ldots \sum_{y_N=0}^{1} \prod_{j=1}^{N} Pr(Y_j = y_j) \]

\[ \cdot Pr \left( \sum_{i=1}^{N} Y_i h_i > C \bigg| Y_1 = y_1, \ldots, Y_N = y_N \right) \]

\[ = \sum_{\forall y \in \{0,1\}^N} Pr(y) \cdot Pr \left( y h^T > C \right)_{0 \text{ or } 1}, \]

where \( y = \{y_1, \ldots, y_N\} \) and \( h = \{h_1, \ldots, h_N\} \).
Table of content

Administrative details

Linear algebra

Probability theory

Probability distributions

Statistical cash-flow management problem

Tail approximation methods

Central limit theorem

Sampling methods

Next lesson
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 $X$:

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

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

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

$$Pr(X \geq a) \leq \min_{s \in \mathbb{R}^+} \frac{M_X(s)}{e^{sa}},$$
Chernoff versus moment bounds

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

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

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

$$= \sum_{n=0}^{\infty} \frac{(sa)^n}{n!} e^{-sa} B_M(n)$$

Poisson$(sa)$ weights

→ the best moment bound is at $n^* = \lfloor sa + 0.5 \rfloor$

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

→ the tightest moment-like bound is $B_M(sa)$.
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

$$Pr(Y \geq a) = Pr(Y + u \geq a + u) \leq Pr((Y + u)^2 \geq (a + u)^2)$$

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 $p = 1/4$.

\[ P(X \geq 3n/4) = \text{???} \]

<table>
<thead>
<tr>
<th>Method</th>
<th>order</th>
<th>opt</th>
<th>bound</th>
<th>$n = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markov</td>
<td>1</td>
<td>-</td>
<td>$\frac{1}{3}$</td>
<td>0.333</td>
</tr>
<tr>
<td>Moment</td>
<td>2</td>
<td>-</td>
<td>$\frac{3n+n^2}{9n^2}$</td>
<td>0.114</td>
</tr>
<tr>
<td>All moments</td>
<td>$\infty$</td>
<td>+</td>
<td>$\frac{3n+n^2}{9n^2}$</td>
<td>1.11 $\cdot 10^{-24}$</td>
</tr>
<tr>
<td>Chebyshev</td>
<td>2</td>
<td>-</td>
<td>$\frac{3}{4n}$</td>
<td>0.0075</td>
</tr>
<tr>
<td>Cent. mom.</td>
<td>3</td>
<td>-</td>
<td>$\frac{3}{4n^2}$</td>
<td>0.000075</td>
</tr>
<tr>
<td>All cent. mom.</td>
<td>$\infty$</td>
<td>+</td>
<td>$\frac{3}{4n^2}$</td>
<td>1.03 $\cdot 10^{-24}$</td>
</tr>
<tr>
<td>Chernoff</td>
<td>$\infty$</td>
<td>+</td>
<td>$3^{-\frac{n}{2}}$</td>
<td>1.39 $\cdot 10^{-24}$</td>
</tr>
</tbody>
</table>

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

$\tilde{g}(x)$ is non-negative, monotone decreasing for $x < a$ then

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

Proof for continuous $X$:

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

$$\geq \int_{-\infty}^{a} \tilde{g}(x)f_X(x)dx$$

mon. dec.

$$\geq \int_{-\infty}^{a} \tilde{g}(a)f_X(x)dx = \tilde{g}(a)\int_{-\infty}^{a} f_X(x)dx$$

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

Administrative details

Linear algebra

Probability theory

Probability distributions

Statistical cash-flow management problem

Tail approximation methods

Central limit theorem

Sampling methods

Next lesson
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 of \( E(X) \).

Variance of \( S_n \):

$$V ar \left( S_n \right) = \sum_{i=1}^{n} V ar \left( \frac{X_i}{n} \right) = \sum_{i=1}^{n} \frac{V ar \left( X_i \right)}{n^2} = \frac{V ar \left( X \right)}{n}$$
Central limit theorem (CLT)

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

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

\[
\lim_{n \to \infty} \sqrt{n}(S_n - E(X)) \overset{d}{=} N(0, \sigma^2_X)
\]

Equivalently

\[
\lim_{n \to \infty} \frac{1}{\sqrt{n}} \left( \sum_{i=1}^{n} X_i - nE(X) \right) \overset{d}{=} N(0, \sigma^2_X)
\]
Application of CLT

Application of CLT for tail approximation:

- $Y = \sum_{i=1}^{n} X_i$ is assumed to be normal distributed,
- $\mu = E(Y) = \sum_{i=1}^{n} E(X_i)$ and
- $\sigma^2 = Var(Y) = \sum_{i=1}^{n} Var(X_i)$

then

$$Pr(Y > C) \approx 1 - \Phi \left( \frac{C - \mu}{\sigma} \right)$$
Example

$X$ 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

Administrative details
Linear algebra
Probability theory

Probability distributions
Statistical cash-flow management problem
Tail approximation methods
Central limit theorem
Sampling methods
Next lesson

Sampling
Next lesson
**Sampling**

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

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

\[
\text{risk} = Pr \left( \sum_{i=1}^{N} Y_i h_i > C \right)
\]

\[
= \sum_{\forall y \in \{0,1\}^N} Pr(y) \cdot Pr(yh^T > C)
\]

\[
= \sum_{\forall y \in C} Pr(y) \cdot Pr(yh^T > C) + \sum_{\forall y \in \bar{C}} Pr(y) \cdot \underbrace{Pr(yh^T > C)}_{0 \leq \cdot \leq 1},
\]

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

\[
\sum_{\forall y \in C} Pr(y) \cdot Pr(yh^T > C) \leq \text{risk}
\]

\[
\leq \sum_{\forall y \in C} Pr(y) \cdot Pr(yh^T > C) + 1 - \sum_{\forall y \in \bar{C}} Pr(y).
\]
Li-Sylvester method

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

Order the $y$ vectors with decreasing probabilities:

$$Pr(y^{(1)}) \geq Pr(y^{(2)}) \geq \ldots \geq Pr(y^{(c)}) \geq \ldots \geq Pr(y^{(2^N)})$$

and bound the risk based on the $c$ most probable samples

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

*Problem:* Efficient generation of the ordered $y$ vectors.
Random sampling

Monte Carlo simulation:

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

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

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

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

As discussed with CLT:

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

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

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

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

Decompose it to \( I \) disjoint subsets \( C_1, \ldots, C_I \) such that
\[
\bigcup_{i=1}^{I} C_i = \{0, 1\}^N \quad \text{and} \quad C_i \cap C_j = \emptyset \quad \text{for} \ i \neq j .
\]

Let \( p_i = \sum_{\forall y \in C_i} Pr(y) \) then \( \sum_{i=1}^{I} p_i = 1 \).

Sample allocation scheme \( S_1, \ldots, S_I \) \( \sum_{i=1}^{I} S_i = S \) then
\[
\text{risk} = Pr(\mathbf{y} \mathbf{h}^T > C) = \sum_{i=1}^{I} p_i \text{risk}_i
\]

where \( \text{risk}_i \) is approximated based on the series of random samples \( \mathbf{y}_{\text{rnd}_i}^{(s)} \in C_i, \ s = 1, \ldots, S_i \) as
\[
\text{risk}_i \approx \eta_i = \frac{1}{S_i} \sum_{s=1}^{S_i} \mathcal{I} \left( \mathbf{y}_{\text{rnd}_i}^{(s)} \mathbf{h}^T > C \right).
\]
Stratified sampling

Algorithm:

- Sample generation
  - generate samples according to the distribution of $y$,
  - classify the samples according to $y \in C_i$,
  - until the number of samples in $C_i \geq S_i$, $\forall i$.

- Risk estimation

  \[
  \text{risk} \approx \sum_{i=1}^{I} p_i \frac{1}{S_i} \sum_{s=1}^{S_i} \mathcal{I} \left( y_{r_{nd_i}}^{(s)} h^T > C \right). 
  \]
Stratified sampling

Approximating the error of stratified sampling

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

Using \( E(\eta^2) = \text{Var}(\eta) + E(\eta)^2 \):

\[
E(\eta) = \sum_{i=1}^{I} p_i E(\eta_i) = \sum_{i=1}^{I} p_i E(B^{(i)}) = \text{risk},
\]

\[
E(\eta^2) = \sum_{i=1}^{I} p_i E(\eta_i^2)
\]

\[
\text{Var}(\eta) = \left( \sum_{i=1}^{I} p_i \left( \text{Var}(\eta_i) + E(\eta_i)^2 \right) \right) - E(\eta)^2
\]
Stratified sampling

Approximating the error of stratified sampling

\[
Var(\eta) = \left( \sum_{i=1}^{I} p_i \left( Var(\eta_i) + E(\eta_i)^2 \right) \right) - E(\eta)^2
\]

\[
= \left( \sum_{i=1}^{I} p_i \left( \sum_{i=1}^{S_i} \frac{Var(B_s^{(i)})}{S_i^2} + E(B^{(i)})^2 \right) \right) - E(\eta)^2
\]

\[
= \left( \sum_{i=1}^{I} p_i \left( \frac{Var(B^{(i)})}{S_i} + E(B^{(i)})^2 \right) \right) - E(\eta)^2
\]

\[
= \sum_{i=1}^{I} p_i Var\left(\frac{B^{(i)}}{S_i}\right) + \sum_{i=1}^{I} p_i E(B^{(i)})^2 - \left( \sum_{i=1}^{I} p_i E(B^{(i)}) \right)^2
\]

with \( Var(B^{(i)}) = E(B^{(i)}) - E(B^{(i)})^2 = \text{risk}_i - \text{risk}_i^2 \).
Stratified sampling

Optimal sample allocation

\[ Var_S = \min_{s_1, \ldots, s_I} \sum_{i=1}^{I} \frac{p_i Var(B^{(i)})}{s_i} \] 

For \( I = 2 \) and \( s_i = S_i / S \), \( c_i = p_i Var(B^{(i)}) / S \) for \( i = 1, 2 \),

\[ Var_S = \min_{s_1, s_2} \frac{c_1}{s_1} + \frac{c_2}{s_2} \]

Its minimum is obtained at \( \frac{c_1}{s_1} = \frac{c_2}{s_2} \) that is \( s_i = \frac{c_i}{c_1 + c_2} \).
Stratified sampling

Optimal sample allocation

\[ Var_S = \min_{s_1, \ldots, s_I} \sum_{i=1}^{I} p_i \text{Var} \left( B^{(i)} \right) \frac{s_i}{S_i}. \]

General case \( s_i = S_i / S \), \( c_i = p_i \text{Var} \left( B^{(i)} \right) / S \) for \( i = 1, \ldots, I \),

\[ Var_S = \min_{s_1, \ldots, s_I} \sum_{i=1}^{I} \frac{c_i}{S_i}. \]

Its minimum is obtained at \( \frac{c_1}{s_1} = \ldots = \frac{c_I}{s_I} \) that is \( s_i = \frac{c_i}{\sum_{j=1}^{I} c_j} \).
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 $E(B^{(i)})$, $Var(B^{(i)})$, and update $s_i$. 
Table of content

Administrative details
Linear algebra
Probability theory
  Probability distributions
Statistical cash-flow management problem
  Tail approximation methods
  Central limit theorem
  Sampling methods
Next lesson
Next lesson