University of Tartu
Institute of Mathematical Statistics

Risk Theory

Fall 2016

Kalev Pärna

Email: kalev.parna@ut.ee

These notes reflect the content of a course in Risk Theory given at the Institute of Mathematics and Statistics, UT. The course covers several basic topics related to mathematical treatment of risks in financial and actuarial world. The first major topic is ruin theory that analyzes certain random processes which model the wealth process of an insurance company. Next we consider basic elements of portfolio theory, including classical Markowitz model and CAPM model. The third main issue is the measurement of financial risk. We focus on Value-at-Risk (VaR) and related methodologies like expected shortfall.

Knowledge of basic concepts and facts of probability theory is a prerequisite for this course. Some knowledge of stochastic processes, especially Poisson and renewal processes, is also useful. Still, some more advanced results in these areas will be given and explained in due course. Basic rules of calculus and some matrix algebra are also used in this course.

This course is mainly based on following books:

- J. Grandell. *Aspects of Risk Theory.* Springer-Verlag, 1991.

- A.J. McNeil, R. Frey, P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press, 2005.

- E.J. Elton, M. J. Gruber. *Modern Portfolio Theory and Investment Analysis.* Wiley, 2003.

1 The Concept of Risk

1.1 The meaning of the word

Arabic word *risq* signifies "anything that has been given to you [by God] and from which you draw **profit**" and has connotations of a fortuitous (random) and **favorable** outcome.

The Latin *risicum* originally referred to the challenge that a barrier reef presents to a sailor and has connotations of an equally fortuitous but **unfavorable** event.

English word "risk" has definite negative associations:

- "run the risk of ..."
- " at risk" (= exposed to danger)

Webster's Dictionary (1981): Risk is 'the possibility of loss, injury, disadvantage, or destruction'

In more specialized literature 'risk' is also used as a *measure* of bad outcome. We can measure the chance (probability) of the bad (negative) outcome, its negativity (severity), or a combination of both.

Our definition: "**Risk is the possibility of an unfavorable event**"

In concrete fields 'risk' has more specific meaning. In business, the risk often means chance of loss of money. An investor loses money when the price of the stock or currency he has invested decreases. In insurance business typical risk is possibility of an big claim, or even possibility of the ruin (bankruptcy) of an insurance company as a result of many big claims that can not be covered by an insufficient premium flow. The study of ruin probabilities is our first major topic in this course.

Nowadays in almost all fields people face with risks: medicine, industry, ecology, security, defence, sports... It is not possible to avoid risks. The problem is not to decide whether to take the risk or not, but rather which risk to take(should we go from A to B by plane, train, car, or on foot...)

Risk are taken by individuals, organizations, and also by governments. (e.g. nationalization of Eesti Raudtee)

1.2 Risk analysis

'*Analysis* is the separation of a whole into its component parts: an examination of a complex, its elements and their relationships' (Concise Oxford, 1976).

Purpose of risk analysis: find out all possible outcomes related to the decision to be made.

The basic risk paradigm:

It is a decision problem in which there is a choice between just two options, one of which will have only one possible outcome (X = no change or status quo), whilst the other option has two possible outcomes (G = gain, L = loss).

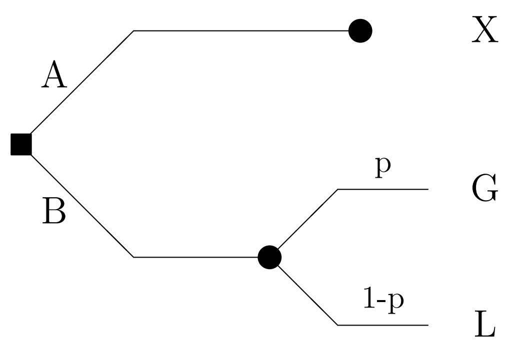

Figure 1: The basic risk paradigm

Examples:

- Investor : to leave money in the bank account (safe option), or to invest money in a new stock
- Doctor: to prescribe a known drug or to experiment with a new drug
- Advanced example: to marry or not to marry

Practical problems are much more complex, more outcomes (sometimes a continuum of possible outcomes), many decisions and processes together. For example, the design of a new chemical plant comprising numerous interconnected processes each one of which could cause the whole plant to fail.

1.3 Risk assessment

The evaluation and comparison of risks is often some form of cost-benefit analysis. It assumes estimation of both probabilities of outcomes and also their severities (magnitudes).

In the basic risk paradigm, when deciding in favor of A, the result is X. If one decides in favor of B, then the expected value is

$$E_B(V) = p\ G + (1 - p)\ L.$$

More generally,

$$E_B(V) = \sum_{i=1}^{n} p_i\ v_i.$$

The variance of outcome is often used as a *measure of risk* (or even synonym for 'risk'):

$$D_B(V) = \sum_{i=1}^{n} p_i\ (v_i - E_B(V))^2.$$

1.4 Risk management

Making practical decisions based on different risk measures.

Well-known financial risk management models:

- risk processes in insurance
- portfolio analysis
- value at risk methodologies
- credit credit scoring
- option pricing.

2 Risk processes

2.1 Stochastic processes

Definition 1. Stochastic process *(or random process) is a family of random variables $\{X(t) : t \in T\}$, where t is time parameter and T is the set of possible values of t.*

Usually $T = \{1, 2, ...\}$ (discrete time) or $T = [0, \infty)$ (continuous time). For each value of t, $X(T)$ is a random variable.

Counting process is a special case of stochastic processes. Let us consider an event A that happens from time to time at random time points $S_1, S_2, ...$ The number of occurrences of the event A within the time interval $[0, t]$ is called a **counting process**:

$$N(t) = \sharp\{i : \ S_i \in [0, t]\}.$$

Example: $N(t)$ is the number of claims on the insurance company during the time interval $[0, t]$.

Let us denote *waiting times* of the events by

$$T_i = S_i - S_{i-1}.$$

Definition 2. *A counting process $N(t)$ is called* **Poisson process** *if its waiting times $T_1, T_2, ...$ are independent random variables having exponential distribution, $T_i \sim Exp(\alpha), \forall i$. The parameter α is called the* intensity *of the Poisson process.*

Recall that the exponential distribution is defined by its density function

$$f(x) = \alpha \cdot e^{-\alpha x}, \ x \geq 0.$$

The expected value of an waiting time is

$$\mathbf{E}T_i = 1/\alpha.$$

It is a consequence of the definition above that for any fixed time t, the random variable $N(t)$ has Poisson distribution with parameter αt , i.e. $N(t) \sim Possson(\alpha t)$, hence its mean (expected) value is

$$\mathbf{E}N(t) = \alpha \cdot t.$$

It is seen that the higher the density α, the more times the event A happens (in average) during the time interval $[0, t]$.

2.2 Risk process

(general)

What is risk process? Safety loading. Some classical results in ruin theory

Risk process is a stochastic process for modeling the wealth of an insurance company.

Definition 3. Risk process *is a stochastic process defined by*

$$X(t) = ct - \sum_{k=1}^{N(t)} Z_k$$

where
$c > 0$ *- a constant called gross premium rate (the company receives* c *units of money per unit time),*
$N(t)$ *- the number of claims on the company during* $(0, t]$*,*
Z_k *- the size of claim* k*.*

At each time point $S_1, S_2, ...$ (the points where N grows) the company has to pay out a stochastic amount of money, and the company receives (deterministically) c units of money per unit time.

$X(t)$ is the profit of the company over the time interval $(0, t]$.

Normally an insurance company starts operating with some initial capital u. Minimum amount of the initial capital is given by regulators.

Ruin of the company means that starting with initial capital u the wealth $u+X(t)$ becomes negative at some time point t.

Ruin probability

$$\Psi(u) = P\{u + X(t) < 0 \text{ for some } t \in (0, \infty)\}.$$

Non-ruin probability: $\Phi(u) = 1 - \Psi(u)$.

Calculation of the ruin probability is the main task of ruin theory.

In practice, companies are often interested in knowing ruin probability during next 4-5 years. For a *finite time horizon* T the ruin probability is defined by

$$\Psi(u, T) = P\{u + X(t) < 0 \text{ for some } t \in (0, T])\}.$$

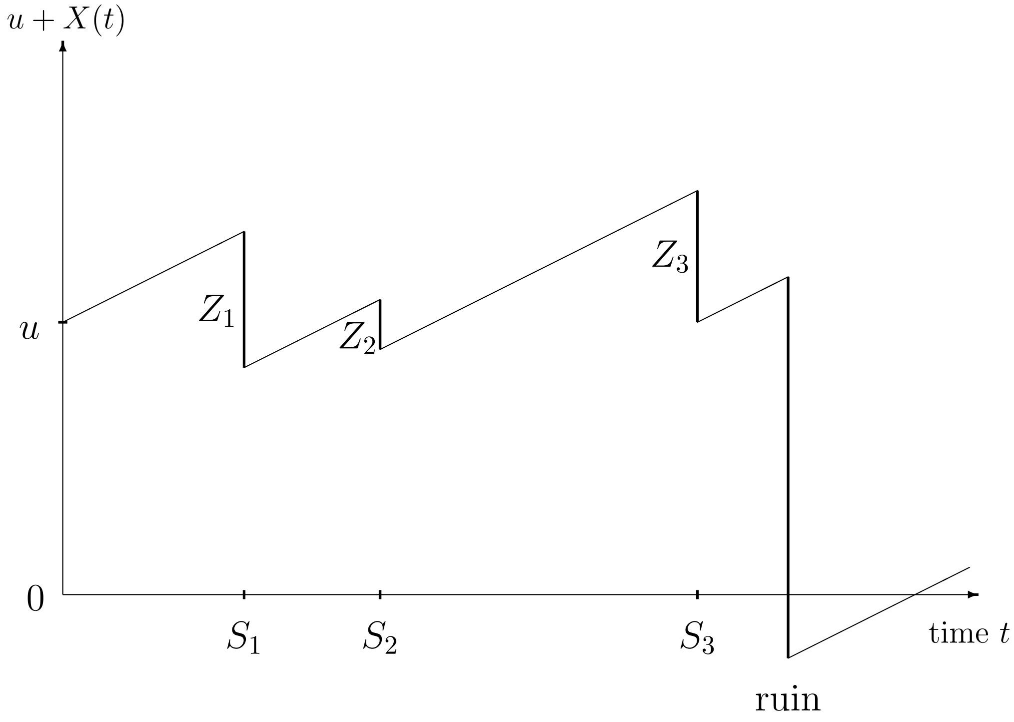

Figure 2: A trajectory of the risk process

Generally speaking, the finite time horizon case is more difficult to handle.

Study of collective risks (all risks of the portfolio of an insurance company are pooled) started over 100 years ago:

F. Lundberg 1903
H. Cramér 1930
H. Bühlmann 1970
H. Gerber 1979
J. Grandell 1991
S. Asmussen 1980 -
H.-J. Albrecher 2010
et al.

2.3 Classical risk process

Here we specify a particularly simple case of risk processes which will be our main subject in coming chapters 2-9.

Definition 4. *The risk process* $X(t) = ct - \sum_{k=1}^{N(t)} Z_k$ *is called* **classical risk process** *if*

- $\{Z_k\}_{k=1}^{\infty}$ *are i.i.d. random variables having common distribution function* $F(z)$ *with* $F(0) = 0$ *and mean value* $EZ_k = \mu$*,*
- $N(t)$ *is a homogenous Poisson process with intensity* α *and independent of* $\{Z_k\}$*.*

NB! We will mainly be interested in this type of risk processes.

Sometimes reversed risk processes are of interest where $c < 0$ and $Z_k < 0$ (e.g. life annuity)

Let us calculate the expectation of the risk process.

Technical remark 1. (sum of random number of random variables):
Let $Z_1, Z_2, ..., Z_N$ be a random number of random variables with $\mathbf{E}Z_k = \mu$. If N is independent of $\{Z_k\}$, then

$$\mathbf{E}(\sum_{k=1}^{N} Z_k) = \mu \cdot \mathbf{E}N.$$

The proof is elementary (condition on N).

Assume now that the Poisson process $N(t)$ has intensity α, i.e. in average α claims arrive per unit time. Then $N(t)$ has Poisson distribution with parameter αt, hence $\mathbf{E}N(t) = \alpha t$, and we have

$$\mathbf{E}X(t) = ct - \mathbf{E}N(t) \cdot \mathbf{E}Z_k = (c - \alpha\mu)t.$$

The ratio

$$\rho = \frac{c - \alpha\mu}{\alpha\mu} = \frac{c}{\alpha\mu} - 1$$

is called **relative safety loading**.

Relative safety loading is an important parameter of the risk process. Large value of ρ means that the income flow (determined by c) significantly exceeds outgoing flow (determined by α and μ.)

- Normally $\rho > 0$ - the company is profitable in average.
- One can make $\rho < 0$ by reducing gross premium rate c, in order to win new customers.

Limit behavior of the risk process

What happens with the path (trajectory) of the risk process when $t \to \infty$? The cases $\rho > 0$ and $\rho < 0$ differ significantly.

Technical remark 2. (Strong Law of Large Numbers - SLLN)
Let $X_1, X_2, ...$ be IID random variables having expectation $\mathbf{E}X_k = a$ Then the convergence

$$\frac{\sum_{k=1}^{n} X_k}{n} \to a$$

takes place almost surely (i.e. with probability 1).

Corollary. A simple consequence from SLLN is that if $a > 0$ and $n \to \infty$, then it must be that $\sum_{k=1}^{n} X_k \to +\infty$. (Similarly, if $a < 0$, then $\sum_{k=1}^{n} X_k \to -\infty$.)

We now apply this to show that if $t \to \infty$ then the risk process $X(t) \to +\infty$ or $X(t) \to -\infty$, depending on whether $\rho > 0$ or $\rho < 0$. For that, consider first the values of t which coincide with one of the claim times $t = S_1, S_2, ...$ Each such t can be represented as an exact sum $t = \sum_{k=1}^{N(t)} T_k$. Now the risk process can be written as a sum of IID random variables

$$X(t) = ct - \sum_{k=1}^{N(t)} Z_k = \sum_{k=1}^{N(t)} (cT_k - Z_k).$$

Since the expected value of each summand is $\mathbf{E}(cT_k - Z_k) = c \cdot \mathbf{E}T_k - \mathbf{E}Z_k = c \cdot \frac{1}{\alpha} - \mu = \mu\rho$, the Corollary above applies showing that $X(t) \to +\infty$ or $-\infty$, depending on whether $\rho > 0$ or $\rho < 0$. (Note that average claim size $\mu > 0$ always.)

Some classical results

Filip Lundberg, Harald Cramér (1930's)

1. Ruin probability in the case of zero initial capital
$$\Psi(0) = \frac{1}{1+\rho} \tag{1}$$

2. Exact formula for the case of exponentially distributed claims, $Z_k \sim Exp(1/\mu)$,
$$\Psi(u) = \frac{1}{1+\rho} e^{\frac{-\rho u}{\mu(1+\rho)}} \tag{2}$$

3. Cramér - Lundberg asymptotic formula
$$\lim_{u\to\infty} e^{Ru}\Psi(u) = C, \tag{3}$$
with constants $R > 0$ (*Lundberg exponent*) and $C > 0$, both depending on μ, α, c, and F.

4. Lundberg inequality
$$\Psi(u) \le e^{-Ru}. \tag{4}$$

We try to prove all these results.

Homework Simulation of the risk process (MS Excel, R)

Technical remark 3. (Rules of conditioning)
Conditioning is a widely used toolkit in probabilistic reasoning. Its origin is the formula of total probability.

1. If $\{B_1, B_2, ...\}$ is a partition of Ω such that $P(B_i) > 0$ then
$$P(A) = \sum_i P(A|B_i)P(B_i).$$
Form this one we deduce, step by step, the following formulae.

2. Conditional expectation w.r.t. a partition $\{B_1, B_2, ...\}$
$$\mathbf{E}X = \sum_i \mathbf{E}(X|B_i)P(B_i).$$

3. In particular, if the partition is induced by a discrete r.v. Y with values y_i then $B_i = \{Y = y_i\}$ and we get the conditional expectation w.r.t. a random variable Y
$$\mathbf{E}X = \sum_i \mathbf{E}(X|Y = y_i)P(Y = y_i),$$
which can be written shortly as
$$\mathbf{E}X = \mathbf{E}[\mathbf{E}(X|Y)],$$
where $\mathbf{E}(X|Y)$ is conditional expectation of X given Y.

4. For continuous r.v. Y with density $f(y)$
$$P(A) = \int_{-\infty}^{+\infty} P(A|y)\ f(y)dy.$$

5. For arbitrary r.v. Y with distribution function $F(y)$
$$P(A) = \int_{-\infty}^{+\infty} P(A|y)dF(y).$$
(Lebesgue-Stiltjes integral)

6. More generally, if Y and Z are two independent r.v. with d.f. $F(y)$ and $G(z)$ then
$$P(A) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(A|y, z)dF(y)dG(z).$$

7. Similar formulae are valid for the expectation
$$\mathbf{E}(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathbf{E}(X|y, z)dF(y)dG(z).$$

3 Derivation of integral equation for ruin probability

We show here that the non-ruin probability $\Phi(u)$ satisfies an integral equation. To derive the equation, we will use the rule of conditioning, combined with a "renewal" argument.

Let S_1 be the time of the first claim. Then we have $X(S_1) = cS_1 - Z_1$. At time S_1 the risk process starts like again, with the only difference that now the initial capital is $cS_1 - Z_1$. We condition upon S_1 and Z_1, whose distribution functions are $F_{S_1}(s)$ (say) and $F(z)$, respectively:

$$\Phi(u) = \mathbf{E}P(nonruin|S_1, Z_1) = \int_0^\infty \int_0^\infty \Phi(u + cs - z)dF_{S_1}(s)dF(z),$$

where we also took into account that the ruin can not occur in $(0, S_1)$.

Since the distribution of S_1 is exponential, we can replace $dF_{S_1}(s) = \alpha e^{-\alpha s}ds$, and since large claims of size $z \geq u + cs$ imply ruin, we have

$$\Phi(u) = \int_0^\infty \alpha e^{-\alpha s} \int_0^{u+cs} \Phi(u + cs - z)dF(z)ds.$$

The change of variables $x = u + cs \Rightarrow ds = dx/c$ leads to

$$\Phi(u) = \frac{\alpha}{c} e^{\alpha u/c} \int_u^\infty \alpha e^{-\alpha x/c} \int_0^x \Phi(x - z)dF(z)dx.$$

Consiquently Φ is differentiable and differentation (using $(fg)' = f'g + fg'$ and $[\int_0^u f(x)dx]' = f(u)$) leads to

$$\begin{aligned}\Phi'(u) = & \frac{\alpha}{c} \cdot \frac{\alpha}{c} e^{\alpha u/c} \int_u^\infty \alpha e^{-\alpha x/c} \int_0^x \Phi(x - z)dF(z)dx \\ & - \frac{\alpha}{c} e^{\alpha u/c} \cdot e^{-\alpha u/c} \int_0^u \Phi(u - z)dF(z).\end{aligned}$$

The first term on the right hand side equals $\frac{\alpha}{c}\Phi(u)$ and hence we have

$$\Phi'(u) = \frac{\alpha}{c}\Phi(u) - \frac{\alpha}{c} \int_0^u \Phi(u - z)dF(z). \tag{5}$$

Replacing $dF(z) = -d(1 - F(z))$ and integrating by parts ($\int fdg = fg| - \int gdf$)

we have

$$\begin{aligned}
\Phi'(u) &= \frac{\alpha}{c}\Phi(u) + \frac{\alpha}{c}\int_0^u \Phi(u-z)d(1-F(z)) \\
&= \frac{\alpha}{c}\Phi(u) + \frac{\alpha}{c}\left[\Phi(0)(1-F(u)) - \Phi(u)\right] + \frac{\alpha}{c}\int_0^u \Phi'(u-z)(1-F(z))dz \\
&= \frac{\alpha}{c}\Phi(0)(1-F(u)) + \frac{\alpha}{c}\int_0^u \Phi'(u-z)(1-F(z))dz
\end{aligned}$$

Integrating over $(0,t)$ yields

$$\Phi(t) - \Phi(0) = \quad = \frac{\alpha}{c}\Phi(0)\int_0^t (1-F(u))du + \frac{\alpha}{c}\int_0^t \int_0^u \Phi'(u-z)(1-F(z))dzdu$$

(now change the order of integration in the double integral)

$$\begin{aligned}
&= \frac{\alpha}{c}\Phi(0)\int_0^t (1-F(u))du + \frac{\alpha}{c}\int_{z=0}^t (1-F(z))\int_{u=z}^t \Phi'(u-z)dudz \\
&= \frac{\alpha}{c}\Phi(0)\int_0^t (1-F(u))du + \frac{\alpha}{c}\int_{z=0}^t (1-F(z))[\Phi(t-z)-\Phi(0)]dz \\
&= \frac{\alpha}{c}\int_0^t (1-F(z))\Phi(t-z)dz
\end{aligned}$$

Thus we can write

$$\Phi(u) = \Phi(0) + \frac{\alpha}{c}\int_0^u \Phi(u-z)[1-F(z)]dz. \tag{6}$$

This is an **integral equation** since the unknown function $\Phi(u)$ stands under the integral sign. This equation can be used for several purposes.

4 Ruin probability with 0 capital

We assume that $\rho > 0$ i.e. the company is profitable. If the initial capital u is zero, the ruin probability takes very simple form (see the 'classical' result (1)). To show that, we first recall a result from integration theory indicating when it is possible to exchange the order of integration and limiting process.

Technical remark (Monotone Convergence Theorem)
Let $0 \le f_n(x) \uparrow$ and let $f_n(x)$ be integrable, $\forall n = 1, 2, \ldots$. Then there exists an integrable limit function $f(x) = \lim_n f_n(x)$ and the equality $\int f(x)dx = \lim_n \int f_n(x)dx$ holds.

By monotone convergence it follows from (6), as $u \to \infty$, that

$$\Phi(\infty) = \Phi(0) + \frac{\alpha\ \mu}{c}\Phi(\infty). \tag{7}$$

Show that $\Phi(\infty) = 1$. It suffices to show that for $\rho > 0$ the process $X(t)$ never attains the value $-\infty$, remaining always finite (then $u + X(t) > 0, \forall t$ and there will not be ruin). First recall that for $\rho > 0$ the paths of the risk process tend to infinity, $\lim_{t\to\infty} X(t) = +\infty$ a.s. It follows that there exists time $T = T(\omega)$ such that for all $t > T$ we have $X(t) > 0$. Hence there cannot be ruin *after* the time T. On the other side, *before* the time T (i.e. within the interval $[0, T]$) only a finite number of claims arrive a.s. (since $N(T)$ has Poisson distribution). Therefore, since each claim has finite size, the total sum to be payed out within $[0, T]$ is finite. Hence, with infinite initial capital, the process cannot ruin before the time T either. To conclude, the non-ruin probability $\Phi(\infty) = 1$. By inserting this into (7) we get

$$1 = \Phi(0) + \frac{\alpha\ \mu}{c},$$

which, together with the definition of ρ, gives the classical result (1):

$$\Psi(0) = \frac{1}{1+\rho} \quad \text{when } c > \alpha\mu.$$

5 Exponentially distributed claims

Consider the case when the claims are exponentially distributed, $Z_k \sim Exp(1/\mu)$. Our aim is to prove the classical result (2):

$$\Psi(u) = \frac{1}{1+\rho} e^{\frac{-\rho u}{\mu(1+\rho)}}$$

The starting point is the following equation, obtained in Section 3:

$$\Phi'(u) = \frac{\alpha}{c}\Phi(u) - \frac{\alpha}{c}\int_0^u \Phi(u-z)dF(z).$$

Since the claims' distribution F is exponential with mean value μ, we can replace $dF(z) = \frac{1}{\mu}e^{-z/\mu}dz$ to obtain

$$\Phi'(u) = \frac{\alpha}{c}\Phi(u) - \frac{\alpha}{c\mu}\int_0^u \Phi(u-z)e^{-z/\mu}dz.$$

Change of variables $u - z =: v,\ dz = -dv$ gives (using $-\int_u^0 \ldots = \int_0^u \ldots$)

$$\Phi'(u) = \frac{\alpha}{c}\Phi(u) - \frac{\alpha}{c\mu}\int_0^u \Phi(z)e^{-(u-z)/\mu}dz,$$

or

$$\Phi'(u) = \frac{\alpha}{c}\Phi(u) - \frac{\alpha}{c\mu}e^{-u/\mu}\int_0^u \Phi(z)e^{z/\mu}dz.$$

Differentiation by u (using rules $(fg)' = f'g + fg'$ and $(\int_a^u h(x)dx)'_u = h(u)$) leads to

$$\begin{aligned}
\Phi''(u) &= \frac{\alpha}{c}\Phi'(u) + \frac{1}{\mu}\Big(\frac{\alpha}{c}\Phi(u) - \Phi'(u)\Big) - \frac{\alpha}{c\mu}\Phi(u) \\
&= (\frac{\alpha}{c} - \frac{1}{\mu})\Phi'(u) = -\frac{\rho}{\mu(1+\rho)}\Phi'(u).
\end{aligned}$$

From here we have $(\ln \Phi'(u))' = \frac{\Phi''(u)}{\Phi'(u)} = -\frac{\rho}{\mu(1+\rho)}$, and , hence,

$$\ln \Phi'(u) = -\frac{\rho u}{\mu(1+\rho)} + C_1$$

from which

$$\Phi'(u) = C_2 e^{-\frac{\rho u}{\mu(1+\rho)}} \ \Rightarrow\ \Phi(u) = C_3 e^{-\frac{\rho u}{\mu(1+\rho)}} + C_4.$$

The constants C_3 and C_4 are defined by conditions $\Phi(\infty) = 1$ (giving $C_4 = 1$) and $\Phi(0) = 1 - 1/(1+\rho)$ (giving $C_3 = -1/(1+\rho)$. Therefore,

$$\Phi(u) = 1 - \frac{1}{1+\rho} e^{-\frac{\rho u}{\mu(1+\rho)}}$$

which is equivalent to the classical result (2).

Exercise. An insurance company is earning EUR 13200 per day (netto). It receives in average 20 claims per day with average claim size of EUR 600. Find the relative safety loading. Assuming that the claims sizes are exponentially distributed, calculate the ruin probability of the company at the time moment when its wealth is equal to EUR 25 000. Find the wealth of the company such that the probability of possible ruin in the future is less than 0,01%.

6 Cramér-Lundberg approximation to the ruin probability

Our aim here is to show the classical result (3), called Cramér-Lundberg approximation. We assume that $\rho > 0$, or $c > \alpha\mu$. The starting point is the equation (6):

$$\Phi(u) = \Phi(0) + \frac{\alpha}{c}\int_0^u \Phi(u-z)[1-F(z)]dz.$$

Form this and from classical result (1) we get the following:

$$\begin{aligned} 1-\Psi(u) &= 1 - \frac{\alpha\mu}{c} + \frac{\alpha}{c}\int_0^u (1-\Psi(u-z))[1-F(z)]dz \\ &= 1 - \frac{\alpha}{c}\left(\mu - \int_0^u [1-F(z)]dz + \int_0^u \Psi(u-z)[1-F(z)]dz\right) \end{aligned}$$

or , by using $\mu = \int_0^\infty [1-F(x)]dx$,

$$\Psi(u) = \frac{\alpha}{c}\int_u^\infty [1-F(z)]dz + \frac{\alpha}{c}\int_0^u \Psi(u-z)[1-F(z)]dz. \tag{8}$$

To solve for $\Psi(u)$ this 'renewal type' equation, we rely on the following.

Technical remark (Key renewal theorem)
Let $G(u)$ satisfy the following ('renewal type') equation:

$$G(u) = H(u) + \int_0^u G(u-x)dA(x), \tag{9}$$

where H is known, A is a given distribution function. Then the asymptotic solution is given by:

$$\lim_{u\to\infty} G(u) = \lim_{u\to\infty} H(u) + \frac{1}{\mu_A}\int_0^\infty H(u)du, \tag{10}$$

where μ_A is the expectation of the distribution A, $0 < \mu_A < \infty$.

However, since $\int_0^\infty \frac{\alpha}{c}[1-F(z)]dz = \frac{\alpha\mu}{c} < 1$, equation (8) is not directly of type (9) (for the function $\frac{\alpha}{c}[1-F(z)]$ to be regarded as a density of a distribution A, this integral must be equal to 1.) W. Feller overcame the difficulty by multiplying both sides of the equation (8) by e^{Ru}, where $R > 0$ is properly chosen constant(Lundberg exponent). We therefore *assume* that there exists a constant $R > 0$ such that

$$\frac{\alpha}{c}\int_0^\infty e^{Rz}[1-F(z)]dz = 1. \tag{11}$$

Then $\frac{\alpha}{c}e^{Rz}[1-F(z)]$ is the density of a proper density distribution. Multiplication of (8) by e^{Ru} yields

$$e^{Ru}\Psi(u) = \frac{\alpha}{c}e^{Ru}\int_u^\infty (1-F(z))dz + \frac{\alpha}{c}\int_0^u e^{R(u-z)}\Psi(u-z)e^{Rz}[1-F(z)]dz.$$

which is a proper renewal equation. From the key renewal theorem it then follows that

$$\lim_{u\to\infty} e^{Ru}\Psi(u) = \frac{C_1}{C_2}, \tag{12}$$

where

$$C_1 = \frac{\alpha}{c}\int_0^\infty e^{Ru}\int_u^\infty (1-F(z))dzdu \tag{13}$$

and

$$C_2 = \frac{\alpha}{c}\int_0^\infty ze^{Rz}(1-F(z))dz \tag{14}$$

provided that finite positive numbers R, C_1, C_2 exist.

(How the functions $G(u), H(u)$ and $A(x)$ should be specified when applying the Key renewal theorem?)

Let us now calculate C_1 and C_2.

Calculation of C_1. Change of the order of integration in (13) gives

$$C_1 = \frac{\alpha}{c}\int_0^\infty (1-F(z)\int_0^z e^{Ru}dudz.$$

Since $\int_0^z e^{Ru}du = \frac{1}{R}e^{Rz} - \frac{1}{R}$ one obtains

$$C_1 = \frac{\alpha}{Rc}\int_0^\infty e^{Rz}(1-F(z))dz - \frac{\alpha}{Rc}\int_0^\infty (1-F(z))dz = \frac{1}{R} - \frac{\alpha\mu}{Rc} = \frac{1}{R}\cdot\frac{\rho}{1+\rho}.$$

(Here we used relationships (11) and $\int_0^\infty (1-F(z))dz = \mu$.)

Calculation of C_2. First introduce the function $h(r) = \int_0^\infty e^{rz} dF(z) - 1$. Then we get from (11), using integration by parts, that

$$\frac{c}{\alpha} = \int_0^\infty e^{Rz}[1 - F(z)]dz = -\frac{1}{R} + \frac{1}{R}\int_0^\infty e^{Rz} dF(z) = \frac{h(R)}{R}$$

and we see that the Lundberg exponent R is the positive solution of the equation

$$h(r) = \frac{cr}{\alpha}. \tag{15}$$

Note also that $h'(R) = \int_0^\infty z e^{Rz} dF(z)$. Now, using $\int z e^{Rz} dz = (\frac{z}{R} - \frac{1}{R^2})e^{Rz}$ (antiderivative of ze^{Rz}), we integrate by parts

$$\begin{aligned} C_2 &= \frac{\alpha}{c}\int_0^\infty z e^{Rz}(1 - F(z))dz \\ &= \frac{\alpha}{c}[1 - F(z)]\left[(\frac{z}{R} - \frac{1}{R^2})e^{Rz}\right]_0^\infty + \frac{\alpha}{c}\int_0^\infty (\frac{z}{R} - \frac{1}{R^2})e^{Rz} dF(z) \\ &= \frac{\alpha}{c}\cdot\frac{1}{R^2} + \frac{\alpha}{c}\int_0^\infty (\frac{z}{R} - \frac{1}{R^2})e^{Rz} dF(z) \end{aligned}$$

where we have used $\lim_{z\to\infty}[1 - F(z)]e^{Rz} = 0$, $\lim_{z\to\infty}[1 - F(z)]z \cdot e^{Rz} = 0$ (both functions are integrable). Now, using expressions for $h(R)$ and $h'(R)$ above, we get

$$\begin{aligned} C_2 &= \frac{\alpha}{c}\left(\frac{1}{R^2} + \frac{h'(R)}{R} - \frac{h(R)+1}{R^2}\right) = \frac{\alpha}{c}\left(\frac{h'(R)}{R} - \frac{c}{\alpha R}\right) \\ &= \frac{\alpha\mu}{c}\frac{1}{R}\frac{1}{\mu}(h'(R) - c/\alpha) = \frac{1}{1+\rho}\frac{1}{R}\frac{1}{\mu}(h'(R) - c/\alpha). \end{aligned}$$

Now it only remains to substitute C_1 and C_2 into (12) to obtain the classical result (3),

$$\lim_{u\to\infty} e^{Ru}\Psi(u) = \frac{\rho\mu}{h'(R) - c/\alpha},$$

called the "Cramér - Lundberg approximation formula".

In practice, the formula can be used for estimation of ruin probabilities for large values of u. In that case,

$$\Psi(u) \approx \frac{\rho\mu}{h'(R) - c/\alpha}e^{-Ru}.$$

We see that with unlimited growth of initial capital the ruin probability tends to 0 at exponential rate. However, one has to remember that Cramér - Lundberg approximation works only if the claim distribution F satisfies the condition (11). Such distributions can not have 'heavy tails'. We will take a closer look at this condition in a later chapter.

Exercise

Find the Lundberg constant R for the case when all claim sizes are equal to 1 (such claims can be interpreted as winnings in a lottery with fixed prizes).

Exercise

Cramér-Lundberg approximation is precise in the case of exponentially distributed claims.

Suppose that the claims are exponentially distributed with mean μ , $Z_i \sim Exp(1/\mu)$. Prove that then the Cramér-Lundberg approximation is exact.

Hints:
1) show that $h(r) = \frac{\mu r}{1-\mu r}$
2) show that the Lundberg exponent satisfies $R = \frac{\rho}{\mu(1+\rho)}$
3) calculate $h'(R) = \mu(1+\rho)^2$
4) show that the right-hand term of Cramér-Lundberg formula verifies

$$\frac{\rho\mu}{h'(R) - c/\alpha} = \frac{1}{1+\rho}$$

5) Compare the result with the classical result (2). (Comment!)

7 Lundberg exponent

The Lundberg exponent was defined by (11) as a positive number $R > 0$ such that

$$\frac{\alpha}{c}\int_0^\infty e^{Rz}[1-F(z)]dz = 1. \tag{16}$$

From this it follows that $e^{Rz}[1-F(z)] \to 0$ as $z \to \infty$. Therefore, $1-F(z)$ must tend to 0 faster than e^{-Rz}, i.e. *the tail of F must be light.* A simple positive example here is the exponential distribution. (In fact, our Homework showed that for exponentially distributed claims the Lundberg exponent equals $R = \rho/\mu(1+\rho)$. However, we now show that many standard distributions do not satisfy this condition, e.g. Pareto, lognormal, or Weibull with shape parameter smaller than 1.

Three examples of heavy-tailed claim distributions

Example 1. The Pareto density is

$$f(z) = \alpha\frac{\beta^\alpha}{z^{\alpha+1}}, \quad z > \beta, \quad \alpha > 0.$$

The tail of the Pareto distribution is $1 - F(z) = \left(\frac{\beta}{z}\right)^\alpha$, which decreases with power speed, i.e. too slowly, since

$$e^{Rz}[1-F(z)] = e^{Rz}\left(\frac{\beta}{z}\right)^\alpha \to \infty \text{ for any } \ R > 0.$$

Hence, for the Pareto distribution the Lundberg exponent does not exist. Note that Pareto distribution is generally accepted as a good model for claim sizes in fire insurance.

Example 2. By definition, Z has log-normal distribution, $Z \sim LN(\mu,\sigma)$, if $\ln Z \sim N(\mu,\sigma)$. The log-normal density is

$$f(z) = \frac{1}{z\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\ln z-\mu)^2}{2\sigma^2}\right\}.$$

Let for simplicity $\mu = 0,\ \sigma = 1$. Then $\ln Z \sim N(0,1)$ and the distribution function of Z is obtained as

$$F(z) = P(Z < z) = P(\ln Z < \ln z) = \Phi(\ln z),$$

where $\Phi(\cdot)$ is the distribution function of $N(0,1)$. Since for large z, $1-\Phi(z) \sim \frac{1}{z}\varphi(z)$, the tail

$$1-F(z) = 1-\Phi(\ln z) \sim \frac{1}{\ln z}\varphi(\ln z) = \frac{1}{\ln z\sqrt{2\pi}}e^{-\frac{(\ln z)^2}{2}}.$$

Therefore

$$e^{Rz}[1-F(z)] \sim \frac{1}{\ln z\sqrt{2\pi}} e^{Rz-\frac{(\ln z)^2}{2}} > \frac{1}{\ln z\sqrt{2\pi}} e^{Rz/2}$$

for z large enough. However, the latter tends to ∞ for any $R > 0$, when $z \to \infty$. Hence, for the log-normal distribution the Lundberg exponent does not exist. Note that log-normal distribution is generally accepted as a good model for claim sizes in motor insurance.

Example 3. The Weibull distribution is defined by its density

$$f(z) = \frac{\gamma z^{\gamma-1}}{\beta^\gamma} \exp\left[-\left(\frac{z}{\beta}\right)^\gamma\right], \quad 0 \le z < \infty, \quad \beta > 0, \quad \gamma > 0.$$

Note that the case of $\gamma = 1$ reduces to the exponential distribution. The tail of the Weibull distribution is

$$1 - F(z) = \exp\left[-\left(\frac{z}{\beta}\right)^\gamma\right].$$

Therefore

$$e^{Rz}[1-F(z)] = e^{Rz} e^{-(z/\beta)^\gamma} = e^{Rz-(z/\beta)^\gamma},$$

which goes to $+\infty$ for $\gamma < 1$ and $-\infty$ for $\gamma \ge 1$. Hence, the Lundberg exponent exists only for $\gamma \ge 1$.

□

Now recall that R is the positive solution of the equation (15):

$$h(r) = \frac{cr}{\alpha},$$

where $h(r) = \int_0^\infty e^{rz} dF(z) - 1$. Multiplying both sides of (15) by α, we easily get the following equivalent condition:

$$\alpha + cr = \alpha \int_0^\infty e^{rz} dF(z). \tag{17}$$

Let us analyze the equation (17). At $r = 0$ both sides are equal to α. The left hand side (LHS) is a linear function with positive slope $c > 0$. The right hand side (RHS) is a continuous, monotonically increasing function (in r), with derivative at $r = 0$ equal to

$$\alpha \int_0^\infty z e^{rz} dF(z)|_{r=0} = \alpha \int_0^\infty z dF(z) = \alpha\mu < c,$$

due to $\rho > 0$. Since its second derivative

$$\alpha \int_0^\infty z^2 e^{rz} dF(z) > 0 \quad \text{for each} \quad r,$$

the RHS is also a convex function. For there to have a **positive** solution $R > 0$ to (17), the graph of the RHS must intersect the straight line on the LHS at some positive value of r. It can be ensured by the following assumption:

Assumption. We assume that there exists $r_0 > 0$ such that $h(r) \uparrow +\infty$ when $r \uparrow r_0$ (we allow for the possibility $r_0 = +\infty$).

Warning! As evidenced by P. Embrechts (and others), claim sizes should rather be modelled by heavy-tailed distributions. But then the moment generating function $\int_0^\infty e^{rz} dF(z)$ of Z will no longer exist and we cannot use neither Cramér-Lundberg approximation nor Lundberg inequality. There are some works on non-exponential upper bounds for the ruin probability, and there is much literature on the Cramér-Lundberg approximation when the claim sizes have heavy-tailed distribution (Asmussen, Embrechts-Klüppelberg-Mikosch).

8 Short overview of heavy-tailed distributions

As it has been mentioned above in several cases, good models for claim distributions encompasse heavy tails.

Heavy-tailed distributions are probability distributions whose tails are not exponentially bounded: that is, they have heavier tails than the exponential distribution. In many applications it is the right tail of the distribution that is of interest, but a distribution may have a heavy left tail, or both tails may be heavy.

There are three important subclasses of heavy-tailed distributions,

- the fat-tailed distributions,
- the long-tailed distributions,
- the subexponential distributions.

In practice, all commonly used heavy-tailed distributions belong to the subexponential class.

There is still some discrepancy over the use of the term heavy-tailed. There are two other definitions in use. However, the definition given below is the most general in use, and includes all distributions encompassed by the alternative definitions (e.g. log-normal that possess all their power moments, yet which are generally acknowledged to be heavy-tailed.)

Definition of heavy-tailed distribution

The distribution of a random variable X with distribution function F is said to have a heavy right tail if

$$\lim_{x\to\infty} e^{\lambda x}\,\Pr[X > x] = \infty \quad \text{for all } \lambda > 0.$$

This is also written in terms of the tail distribution function

$$\overline{F}(x) \equiv \Pr[X > x]$$

as

$$\lim_{x\to\infty} e^{\lambda x}\overline{F}(x) = \infty \quad \text{for all } \lambda > 0.$$

This is equivalent to the statement that the moment generating function of F, $MF(t) = \mathbf{E}(e^{tX})$, is infinite for all $t > 0$.

The definitions of heavy-tailed for left-tailed or two tailed distributions are similar.

Definition of fat-tailed distribution

The distribution of a random variable X is said to have a **fat tail** if

$$\Pr[X > x] \sim x^{-\alpha} \text{ as } x \to \infty, \qquad \alpha > 0.$$

That is, if X has a probability density function, $f_X(x)$,

$$f_X(x) \sim x^{-(1+\alpha)} \text{ as } x \to \infty, \qquad \alpha > 0.$$

Here the notation " $\sim$" means the asymptotic equivalence of functions. Some reserve the term "fat tail" for distributions only where $0 < \alpha < 2$ (i.e. only in cases with infinite variance).

Definition of long-tailed distribution

The distribution of a random variable X with distribution function F is said to have a **long right tail** if for all t > 0,

$$\lim_{x\to\infty} \Pr[X > x + t | X > x] = 1,$$

or equivalently

$$\overline{F}(x + t) \sim \overline{F}(x) \quad \text{as } x \to \infty.$$

This has the intuitive interpretation for a right-tailed long-tailed distributed quantity that if the long-tailed quantity exceeds some high level, the probability approaches 1 that it will exceed any other higher level: if you know the situation is bad, it is probably worse than you think.

All long-tailed distributions are heavy-tailed, but the converse is false, and it is possible to construct heavy-tailed distributions that are not long-tailed.

The class of long-tailed distributions is often denoted by $\mathcal{L}$.

Subexponential distributions

Subexponentiality is defined in terms of convolutions of probability distributions. For two independent, identically distributed random variables X_1, X_2 with common distribution function F the convolution of F with itself, F^{*2} is defined, using Lebesgue–Stieltjes integration, by:

$$\Pr[X_1 + X_2 \leq x] = F^{*2}(x) = \int_{-\infty}^{\infty} F(x - y)\, dF(y)$$

.

The n-fold convolution F^{*n} is defined in the same way. The tail distribution function $\overline{F}$ is defined as $\overline{F}(x) = 1 - F(x)$.

A distribution F on the positive half-line is **subexponential** if

$$\overline{F^{*2}}(x) \sim 2\overline{F}(x) \quad \text{as } x \to \infty.$$

This implies that, for any $n \geq 1$,

$$\overline{F^{*n}}(x) \sim n\overline{F}(x) \quad \text{as } x \to \infty. \tag{18}$$

This condition has a rather simple interpretation. Note that $n\overline{F}(x)$ is the tail of the distribution of the maxima of n random variables $X_1, \cdots, X_n$. Indeed, due to $(1-a)^n \sim 1 - na$ when $a \to 0$, we have

$$\Pr[\max(X_1, \ldots, X_n) > x] = 1 - F^n(x) = 1 - (1 - \overline{F}(x))^n \sim n\overline{F}(x) \tag{19}$$

Therefore, the probabilistic interpretation of (18) is that, for a sum of n independent random variables $X_1, \ldots, X_n$ with common distribution F,

$$\Pr[X_1 + \cdots + X_n > x] \sim \Pr[\max(X_1, \ldots, X_n) > x] \quad \text{as } x \to \infty.$$

This is often known as the principle of the **single big jump.**

The class of subexponential distributions is often denoted by $\mathcal{S}$.

All subexponential distributions are long-tailed, but examples can be constructed of long-tailed distributions that are not subexponential.

Common heavy-tailed distributions

All commonly used heavy-tailed distributions are subexponential.

Those that are one-tailed include:

- the Pareto distribution;
- the Log-normal distribution;
- the Lévy distribution;
- the Weibull distribution with shape parameter less than 1;
- the Burr distribution;
- the log-gamma distribution;
- the log-Cauchy distribution, sometimes described as having a ”super-heavy tail” because it exhibits logarithmic decay producing a heavier tail than the Pareto distribution.

Those that are two-tailed include:

- The Cauchy distribution, itself a special case of both the stable distribution and the t-distribution;
- The family of stable distributions, excepting the special case of the normal distribution within that family. Some stable distributions are one-sided (or supported by a half-line), see e.g. Lévy distribution. See also financial models with long-tailed distributions and volatility clustering.
- The t-distribution.
- The skew lognormal cascade distribution.

References

1. Asmussen, S. (2003). Applied Probability and Queues. Berlin: Springer.
2. Rolski, Schmidli, Scmidt, Teugels, Stochastic Processes for Insurance and Finance, 1999
3. Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance. Berlin: Springer.

9 Lundberg inequality

The Lundberg exponent plays an important role in the ruin theory. We next show the classical result (4).

Theorem 1. ***(Lundberg inequality)*** *For the classical risk process the probability of ruin $\Psi(u)$ satisfies*

$$\Psi(u) \le e^{-Ru}, -\infty < u < \infty, \tag{20}$$

where $R > 0$ is the Lundberg exponent.

Proof. Let A be the event that starting with the initial capital u the risk process will ruin. Then $\Psi(u) = P(A)$. Let us denote

$$A_n = \{\text{the ruin occurs as a result of first } n \text{ claims}\}, n = 1, 2, ...$$

and

$$A_0 = \{\text{the ruin occurs as a result of a negative initial capital}\}.$$

Let the corresponding probabilities be $\Psi_n(u) = P(A_n)$. Obviously, it is an increasing sequence of events: $A_n \subset A_{n+1},\ n = 0, 1, 2, ...$ Further on, as the ruin means that at least one of the events A_n occurs, then $A = \cup_n A_n$. By the continuity of probability we have $P(A) = \lim P(A_n)$, or $\Psi(u) = \lim \Psi_n(u)$. Therefore, it suffices to show that for each $n \ge 0$ the inequality

$$\Psi_n(u) \le e^{-Ru} \tag{21}$$

holds. For doing that, we use the method of mathematical induction. Show first that the inequality holds for $n = 0$ (induction basis). Since $\Psi_0(u) = 1$ for $u < 0$ (negative initial capital means the ruin) and $\Psi_0(u) = 0$ for $u \ge 0$, the inequality $\Psi_0(u) \le e^{-Ru}$ holds, as far as $R > 0$.

Show now that if the inequality (21) holds for an $n - 1$, then it also holds for n. To do that, we condition upon S_1 and Z_1, the time and the size of the first claim. If the first claim arrived at time $S_1 = s$ and its size was $Z_1 = z$, then after paying out the first claim, the company's capital is $u + cs - z$. Starting with this new capital at time $S_1 = s$, the event A_n defined above is equivalent to the ruin due to next $\le n - 1$ claims, and therefore the conditional probability $P(A_n|S_1 = s, Z_1 = z) = \Psi_{n-1}(u + cs - z)$. According to the rules of conditioning, we now average (integrate) such conditional probabilities over all possible values of S_1 and Z_1, while keeping in mind that S_1 has exponential distribution, $S_1 \sim Exp(\alpha)$, and that Z_1 has distribution F:

$$\Psi_n(u) = P(A_n) = \int_0^\infty \int_0^\infty P(A_n|S_1 = s, Z_1 = z) f_{S_1}(s) dF(z) ds.$$

As the exponential density is $f_{S_1}(s) = \alpha e^{-\alpha s}$, $s \geq 0$, we have, after substitution, that

$$\Psi_n(u) = \int_0^\infty \alpha e^{-\alpha s} \int_0^\infty \Psi_{n-1}(u + cs - z) dF(z) ds.$$

As we assume that $\Psi_{n-1}(u)$ satisfies (21), we have $\Psi_{n-1}(u+cs-z) \leq e^{-R(u+cs-z)}$ which gives

$$\Psi_n(u) \leq \int_0^\infty \alpha e^{-\alpha s} \int_0^\infty e^{-R(u+cs-z)} dF(z) ds = e^{-Ru} \cdot \alpha \int_0^\infty e^{-(\alpha+Rc)s} ds \cdot \int_0^\infty e^{Rz} dF(z).$$

Direct integration shows that

$$\int_0^\infty e^{-(\alpha+Rc)s} ds = \frac{1}{\alpha + Rc}$$

(here it is necessary to take into account that $\alpha + Rc > 0$), which gives the inequality

$$\Psi_n(u) \leq e^{-Ru} \cdot \frac{\alpha}{\alpha + Rc} \int_0^\infty e^{Rz} dF(z).$$

However, by the definition of the Lundberg exponent R, the second multiplier satisfies $\frac{\alpha}{\alpha+Rc} \int_0^\infty e^{Rz} dF(z) = 1$, therefore we have shown that

$$\Psi_n(u) \leq e^{-Ru}.$$

The proof is completed. □

In order to use the Lundberg inequality, it would be good to have bounds for the Lundberg exponent R.

Derivation of a lower bound for R.

Suppose that the claims are **bounded** above by some constant K, i.e. $Z \leq K$. Then for $z \leq K$ the inequality

$$e^{Rz} \leq \frac{z}{K} e^{RK} + (1 - \frac{z}{K}). \tag{22}$$

holds. Indeed, due to the well-known expansion formula

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots,$$

we have

$$\begin{aligned}
\frac{z}{K} e^{RK} + (1 - \frac{z}{K}) &= 1 - \frac{z}{K} + \frac{z}{K} + \frac{z}{K} RK + \frac{z}{K} \frac{R^2 K^2}{2!} + \frac{z}{K} \frac{R^3 K^3}{3!} + \ldots \\
&\geq 1 + zR + \frac{R^2 z^2}{2!} + \ldots \\
&= e^{Rz}.
\end{aligned}$$

Now recall the equation (17) which defines R, and apply (22):

$$\alpha + cR = \alpha \int_0^\infty e^{Rz} dF(z) = \alpha E(e^{RZ})$$
$$\leq \alpha \left(\frac{\mu}{K} e^{RK} + 1 - \frac{\mu}{K} \right).$$

From this we have:

$$cR \leq \frac{\alpha\mu}{K}(e^{RK} - 1),$$

or

$$\frac{e^{RK} - 1}{KR} \geq \frac{c}{\alpha\mu} \equiv 1 + \rho.$$

Now a small 'trick' i.e. the inequality $\frac{e^x - 1}{x} < e^x$ can be used to obtain

$$e^{RK} \geq 1 + \rho,$$

which gives

$$R > \frac{1}{K} \ln(1 + \rho).$$

Corollary. If the claims are bounded, $Z \leq K$, then the ruin probability satisfies

$$\Psi(u) \leq (1 + \rho)^{-u/K}. \tag{23}$$

Example

An insurance company is running business under following conditions:
- it earns in average EUR 50 000 per day from selling insurance policies ,
- it receives in average 10 claims per day,
- the average claim is EUR 4 000.
- the claim size is bounded from above by EUR 100 000.

1) Assuming that all conditions of the classical risk process are fulfilled, estimate the ruin probability for initial capitals EUR 10k, 100k, 1m, 5m, and 10m, using the Lundberg inequality.
2) Draw a graph of the upper bound of the ruin probability (with initial capital on the x-axis.)

10 Simple approximations of ruin probabilities

It is natural to try to find "simple" and "good" approximations of $\Psi(u)$.

A "simple approximation" of a ruin probability is an approximation using only some moments of the claim distribution and not the detailed tail behaviour of that distribution. Such approximations may be based on more or less ad hoc arguments and their merits can only be judged by numerical comparison. Others are based on limit theorems, and the limit procedure may give hints on their applicability. In that case numerical comparison may be needed in order to get information about the speed of convergence and — which is almost the same — their numerical accuracy.

The most successful simple approximation is certainly the *De Vylder approximation*, which is based on the idea to replace the risk process with a risk process with exponentially distributed claims such that the three first moments coincide. That approximation is known to work extremely well for "kind" claim distributions. The purpose of this chapter is to analyse the De Vylder approximation and other simple approximations from a more mathematical point of view.

Several such approximations have been proposed. The most famous approximation is, of course, the Cramér–Lundberg approximation (3):

$$\lim_{u\to\infty} e^{Ru}\Psi(u) = \frac{\rho\mu}{h'(R) - c/\alpha}.$$

This approximation, which goes back to Cramér (1930), works well in case of light tail claim distributions, and is very accurate for large values of u. The approximation requires that the tail of F decreases at least exponentially fast, and thus for instance the lognormal and the Pareto distributions are excluded. In order to include that last mentioned distributions it is usual to consider distributions F such that its itd F_I belongs to the class $\mathcal{S}$ of subexponential distribution. Then, as we have seen already,

$$\psi(u) \sim \frac{\overline{F_I}(u)}{\rho}, \quad u \to \infty,$$

does hold exactly. However, the latter approximation has a much slower speed of convergence than that of Cramér–Lundberg (see e.g. Grandell (1997, p. 222).

Both approximations above are practically somewhat difficult to apply, since they require full knowledge of the claim distribution. Notice that they apply for fixed values of ρ as $u \to \infty$. Thus those approximations may be looked upon as "large deviation" results and it is seen that the asymptotic behaviour of $\Psi(u)$ is very different.

We will concentrate on "simple" approximations, by which we mean that the approximations only depend on some moments of F. The simplest such approximation seems to be the **diffusion approximation:**

$$\Psi(u) \approx \Psi_D(u) := e^{-u\rho\frac{2\zeta_1}{\zeta_2}}.$$

where

$$\zeta_k = \mathbf{E}(Z_1^k)$$

are the moments of the claim size distribution F. (Note that $\zeta_1 = \mu$.) This approxiamtion goes back to Hadwiger (1940). It is nowadays derived by application of weak convergence of the compound Poisson process to a Wiener process, see for instance Grandell (1991). It may be used if ρ is small and u is large in such a way that u and ρ^{-1} are of the same order. In queuing theory is known as the "heavy traffic approximation". The numerical accuracy of this approximation is not very impressive. It is natural to regard the asymptotic behind the diffusion approximation as a "central limit" situation in the sense that many claims "co-operate" on almost equal terms. It seems that simple approximations can only be expected to work well in such a case. Similarly as one shall apply the central limit theorem with great care far out in the tails, one may suspect that simple approximations ought to be used mainly when the ruin probability is not too small.

De Vylder approximation

The De Vylder approximation, proposed by De Vylder (1978), is based on the simple, but ingenious, idea to replace the risk process X with a risk process $\tilde{X}$ with exponentially distributed claims such that

$$\mathbf{E}[X^k(t)] = \mathbf{E}[\tilde{X}^k(t)] \quad \text{for } k = 1, 2, 3.$$

The risk process $\tilde{X}$ is determined by the three parameters $(\tilde{\alpha}, \tilde{c}, \tilde{\mu})$ or $(\tilde{\alpha}, \tilde{\rho}, \tilde{\mu})$.

We can calculate the three first moments (using characteristic function of $X(t)$, for example):

$$\begin{aligned}
\mathbf{E}[Xt)] &= (c - \alpha\mu)t = \rho\alpha\zeta_1 t, \quad \text{(as we already know)} \\
\mathbf{E}[X^2(t)] &= \alpha\zeta_2 t + (\rho\alpha\zeta_1 t)^2, \\
\mathbf{E}[X^3(t)] &= -\alpha\zeta_3 + 3(\rho\alpha\zeta_1 t)(\alpha\zeta_2 t) + (\rho\alpha\zeta_1 t)^3.
\end{aligned}$$

Respective moments of the process $\tilde{X}(t)$ can also be calculated:

By equating the moments of the two processes we see that the parameters $(\tilde{\alpha}, \tilde{\rho}, \tilde{\mu})$ must satisfy

$$\rho\alpha\zeta_1 t = \tilde{\rho}\tilde{\alpha}\tilde{\mu}, \quad \alpha\zeta_2 = 2\tilde{\alpha}\tilde{\mu}^2, \quad \alpha\zeta_3 = 6\tilde{\alpha}\tilde{\mu}^3,$$

and we get

$$\tilde{\mu} = \frac{\zeta_3}{3\zeta_2}, \quad \tilde{\rho} = \frac{2\zeta_1\zeta_3}{3\zeta_2^2}\rho.$$

Thus we are led to the De Vylder's approximation

$$\Psi(u) \approx \Psi_{DV}(u) := \frac{1}{1+\tilde{\rho}} e^{-\frac{\tilde{\rho}u}{\tilde{\mu}(1+\tilde{\rho})}}.$$

Many other approximations have been introduced. However, the De Vylder's approximation is considered as very simple and very often surprisingly precise.

11 Further generalizations of the classical risk processes

The classical risk process studied in the previous sections is a simplified model of the wealth of an insurance company. Next some possible extensions are shown that introduce more realistic features into the model.

A. The premiums may depend on the result of the insurance business. For example, it is natural to make the safety loading smaller if the risk business attains a large value.

B. Inflation and interest may be included in the model.

C. The claim arrival process may be described by a more general process than the Poisson process:

1. Non-homogeneous Poisson processes

2. Cox processes

3. Renewal processes

The classical result (1) remains valid for all three cases 1) - 3). Classical results (1)-(4) remain (basically) valid for renewal processes.

Also, the *finite time horizon* $T < \infty$ is of interest in many cases. Then the ruin probability is defined as

$$\Psi(u, T) = P\{u + X(t) < 0 \text{ for some } t \in (0, T\}.$$

However, then the formulae are more complex.

12 Introduction to other financial risks

Financial institutions have become very sophisticated and scientific in their analysis, assessment and management of their financial risks. Banks and funds are looking for quantitative risk analysts who are able to estimate risks numerically and who also manage techniques for hedging the risks.

Types of financial risks:

- Credit risk estimates potential losses due to the inability of a counterparty to meet its obligations. One has to account for credit risk when deciding whether to give credit or whether to extend an existing credit.
- An operational risk is, as the name suggests, a risk arising from execution of a company's business functions. It is a very broad concept which focuses on the risks arising from the people, systems and processes through which a company operates. It also includes other categories such as fraud risks, legal risks, physical or environmental risks. Most important tool to fight with operational risk is to maintain tight control (keeping time schedule, planned duration, budgeted costs).
- Liquidity risk is associated with the inability of a firm to fund illiquid assets. (E.g. you can have a real estate property but nobody is ready to buy it - you can not turn it into the money).
- Market risk involves the uncertainty of earnings resulting from changes in market conditions such as the **asset prices**, **exchange rates**, **interest rates**, **volatility**, and market **liquidity**. Market risk can be absolute or relative. **Absolute market risk** estimates total loss expressed in currency terms, e.g. Dollars at Risk. Trading managers focus on how much they can lose over a relatively short time horizon such as one day. This is called **DEaR**, Daily earnings at Risk. In some cases the investment horizon is longer, such as a month. Then the term **VaR** (Value at Risk) is used for a measure of potential losses. **Relative market risk** measures the potential for under performance, i.e. estimated tracking error, against a benchmark. The investment management industry (funds, investment banks etc.) uses this version of market risk.

Of course, investors also have to face

- legal risks
- political risks.

12.1 The principle of diversification

Should we invest five millions into one single stock or five different stocks, one million per stock?

Everybody has heard that

"Never put all eggs into one basket".

But why?

Suppose 5 eggs are to be transported from one place to another.

Let us compare two strategies:

Strategy A: Put all eggs into one basket.

Strategy B: Put each egg into a new basket.

In financial terms, you have 5 millions to invest and there is a choice between two strategies: to buy shares of one single stock, or to buy shares of five different stocks.

To answer, we assume that any basket achieves the destination with probability p and breaks down with probability $q = 1 - p$. Also, assume that the baskets behave independently from each other.

Let X be the number of 'successful' eggs that reach the destination. We compare strategies A and B via the expectation of X and its variance.

We see that the two strategies produce equal expected values of X:

$$E_A X = E_B X = 5p.$$

However, the strategies differ in the variance of X:

$$D_A X = 25p\,(1-p)$$
$$D_B X = 5p\,(1-p).$$

In case of strategy B, the variance is 5 times less. With strategy B, we also have intermediate values $1, 2, 3, 4$ that can be useful: in order to bake a cake, you do not need all 5 eggs! We have less uncertainty (less risk) with strategy B. *The risk can be measured by the variance.*

13 Markowitz portfolio theory

Modern portfolio theory started after a paper by Harry M. Markowitz appeared in 1956. His ideas were developed further on by Sharpe, Miller, Mossin, Lintner a.o. (Nobel Prize in Economics 1990 - Markowitz, Miller, Sharpe). The aim is to find an investment strategy that enables high return with a low risk. More precisely, a suitable compromise between the expected return of the portfolio and its risk should be found by an investor.

The following short overview of the modern portfolio theory is mainly based on:
S. Roman. Introduction to the Mathematics of Finance. Springer, 2004;
E. J. Gruber. Modern Portfolio Theory and Investment Analysis. 5th ed., Wiley, 1995.

13.1 Return

Consider a stock with its price P_t at time t. Most often, the time unit is a day, or a year. The percent return (or simply, *return*) at time t is the relative change of the price of the stock:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

The logarithmic return is defined by

$$R_t = ln(1 + r_t).$$

Since for small x, $ln(1+x) \approx x$, we have that $R_t \approx r_t$. In practice, the percentage return is calculated, but in theoretical developments both are used depending on the problem and simplicity.

We start with the case of two stocks.

13.2 Portfolios with two stocks

Consider two stocks A and B.

Let their returns be r_A and r_B.

Expected returns are denoted by $e_A = E(r_A),\ e_B = E(r_B)$.

The variances of the returns by $\sigma_A^2 = D(r_A)$ and $\sigma_B^2 = D(r_B)$.

The covariance of r_A and r_B is $cov(r_A, r_B) = E(r_A - e_A)(r_B - e_B)$.

The correlation coefficient between r_A and r_B is

$$\rho_{AB} = \frac{cov(r_A, r_B)}{\sigma_A \cdot \sigma_B}.$$

Toolbox: The basic tool for further developments is the following simple formula of the variance of the sum of two random variables:

$$D(aX + bY) = a^2 DX + b^2 DY + 2ab \cdot cov(X, Y),$$

or

$$D(aX + bY) = a^2 DX + b^2 DY + 2ab \cdot \rho\, \sigma_X \sigma_Y,$$

where σ_X and σ_Y are the standard deviations of X and Y, and ρ is the correlation coefficient between X and Y.

Now consider a portfolio (denoted by p) consisting of the stocks A and B. Suppose the investor has invested w_A percent of the money into the stock A and w_B percent into the stock B. The numbers $w_A,\ w_B$ are called the weights of the stocks in the portfolio. So the portfolio (p) itself is characterized by the weights $w_A,\ w_B$.

Class exercise

Show that the (percent) return of the portfolio satisfies

$$r_p = w_A r_A + w_B r_B,$$

i.e. the portfolio return is equal to the weighted average of the returns of its composite assets.

Let us now calculate the expected return e_p of the portfolio

$$e_p = E(r_p) = E(w_A r_A + w_B r_B) = w_A e_A + w_B e_B$$

and its variance

$$\sigma_p^2 = D(r_p) = D(w_A r_A + w_B r_B) = w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2\rho\ w_A w_B\ \sigma_A \sigma_A.$$

Class exercise

Expected returns for stocks A and B are 10 percent and 20 percent (per year), respectively. The standard deviations of the returns are 2 and 5 percent (resp.). The correlation coefficient between the returns is -0,5. Find the expected returns and standard deviations of returns for six portfolios composed from A and B with weights of A as follows: $w_A = 0; 0,2; 0,4; 0,6; 0,8; 1$. Draw the corresponding graph with standard deviation on the x-axis and expected returns on the y-axis.

The analysis of the figure:

Portfolio frontier

Frontier portfolio is any portfolio on the portfolio line.

Short selling makes the graph 'longer'

13.3 Portfolios consisting of n stocks

We generalize the portfolio analysis from the case of two stocks to the case of arbitrary number (n) of stocks. Assume that short selling is allowed.

Let us introduce the following vector notation:

$$\begin{aligned} r &= (r_1, \dots, r_n)^T, \\ e &= (e_1, \dots, e_n)^T, \\ w &= (w_1, \dots, w_n)^T, \\ \mathbb{1} &= (1, \dots, 1)^T, \\ \mathbb{0} &= (0, \dots, 0)^T, \\ V &= (\sigma_{ij}), \end{aligned}$$

where $\sigma_{ij} = cov(r_i, r_j)$.

Similarly to Homework 3, the return of the portfolio is

$$r_p = w_1 r_1 + ... + w_n r_n = w^T r = r^T w.$$

Therefore, the expected return and the variance of the portfolio:

$$e_p = E(r_p) = w_1 e_1 + ... + w_n e_n = w^T e = e^T w, \quad (24)$$

$$\sigma_p^2 = D(r_p) = D(w_1 r_1 + ... + w_n r_n) = \sum_{i,j} w_i w_j \sigma_{ij} = w^T V w. \quad (25)$$

Markowitz (1956) formulated the following portfolio optimization problem: given a value of the expected return e_p of the portfolio, find the weights $(w_1, \dots, w_n)$ that minimize the variance σ_p^2 of the return of portfolio (risk of the portfolio). Mathematically:

$$\begin{aligned} \frac{1}{2} w^T V w &\to min_w \\ w^T e &= e_p \\ w^T \mathbb{1} &= 1. \end{aligned}$$

This optimization problem with two constraints can be solved by the method of Lagrange multipliers. Firstly the Lagrange functional is composed:

$$L = \frac{1}{2} w^T V w + \lambda(e_p - w^T e) + \gamma(1 - w^T \mathbb{1}),$$

where λ and γ are new variables (Lagrange multipliers.) Secondly, the partial derivatives of L are equated to zero:

$$\begin{cases} \frac{\partial L}{\partial w} = \mathbb{0}, \\ \frac{\partial L}{\partial \lambda} = 0, \\ \frac{\partial L}{\partial \gamma} = 0. \end{cases}$$

Toolbox: Matrix derivatives

Let $y = f(x)$ be a scalar function with the vector argument $x = (x_1, \ldots, x_n)$. Define the vector of partial derivatives

$$\frac{dy}{dx} = (\frac{\partial y}{\partial x_1}, \ldots, \frac{\partial y}{\partial x_n})^T.$$

Special cases:

1. Linear form
$$y = a_1 x_1 + \ldots + a_n x_n = a^T x.$$
Then
$$\frac{dy}{dx} = a$$
2. Quadratic form
$$y = x^T A x,$$
where A is an $n \times n$ matrix of constants, $A = (a_{ij})$. Then
$$\frac{dy}{dx} = 2Ax.$$

□

Using the rules of matrix derivatives, we have

$$\frac{\partial L}{\partial w} = Vw - \lambda e - \gamma \mathbb{1} = \mathbb{0}, \quad (26)$$
$$\frac{\partial L}{\partial \lambda} = e_p - w^T e = 0, \quad (27)$$
$$\frac{\partial L}{\partial \gamma} = 1 - w^T \mathbb{1} = 0. \quad (28)$$

This is a system of $n+2$ linear equations with the same number of unknowns (n components of w, α, γ). To solve it, we assume that V is positive definite, i.e., $x^T V x > 0$ for each $x \neq 0$. Then the inverse V^{-1} exists and multiplication (from left) of both sides of (26) by V^{-1} gives:

$$w = \lambda V^{-1} e + \gamma V^{-1} \mathbb{1}. \quad (29)$$

By substituting w into (37) and (28) one obtains:

$$\begin{cases} \lambda e^T V^{-1} e + \gamma e^T V^{-1} \mathbb{1} = e_p, \\ \lambda \mathbb{1}^T V^{-1} e + \gamma \mathbb{1}^T V^{-1} \mathbb{1} = 1. \end{cases}$$

or

$$\begin{cases} \lambda B + \gamma A = e_p, \\ \lambda A + \gamma C = 1, \end{cases}$$

with

$$A = e^T V^{-1}\mathbb{1}, \quad B = e^T V^{-1} e, \quad C = \mathbb{1}^T V^{-1}\mathbb{1}.$$

This is a system of two linear equations and the substitution rule gives

$$\begin{aligned} \lambda &= \frac{Ce_p - A}{D} \\ \gamma &= \frac{B - Ae_p}{D}, \end{aligned}$$

where $D = BC - A^2$. Finally, to get optimal weights of the portfolio, substitute λ and γ into (29). The solution (denoted by w_p) can be expressed in the form:

$$w_p = g + he_p, \tag{30}$$

where the vectors $g,\ h$ are given by

$$g = \frac{1}{D}(BV^{-1}\mathbb{1} - AV^{-1}e) \tag{31}$$

$$h = \frac{1}{D}(CV^{-1}e - AV^{-1}\mathbb{1}). \tag{32}$$

Interpretation: g is the weight vector of the portfolio whose expected return is set to 0; h describes the change of the weight vector when the expected return e_p is increased by 1 unit.

Let us now calculate the variance of the return of the optimal portfolio, σ_p^2. Due to (25) we have

$$\sigma_p^2 = w_p^T V w_p.$$

Substitution of w_p from (30) gives (after simple calculations)

$$\sigma_p^2 = \frac{C}{D}\left(e_p - \frac{A}{C}\right)^2 + \frac{1}{C} \tag{33}$$

(check it!)

Here one recognizes the equation of an hyperbola on the (σ_p, e_p)-plane (see next Figure). This hyperbola is called *portfolio frontier* and each point on it corresponds to an portfolio called *frontier portfolio*. Frontier portfolios on the upper branch of the hyperbola are called *efficient* portfolios, on lower branch - *inefficient portfolios*.

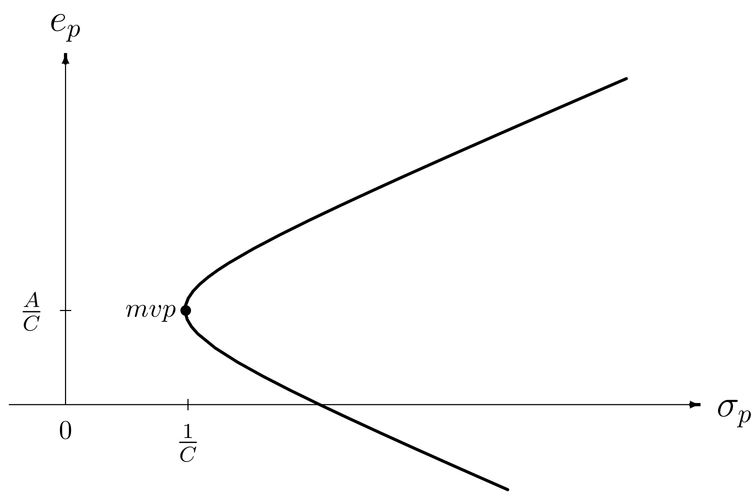

Figure 3: The portfolio frontier is an hyperbola - a geometrical presentation of the portfolios with minimum risk for a given expected return. The upper branch contains *efficient* portfolios, the lower branch - *inefficient* portfolios. The interior of the hyperbola is called "Markowitz bullet".

Summary: After an investor has fixed his expected return e_p, the formula (30) shows how to compose the optimal portfolio which minimizes the variance of return (risk). However, the investor may find the variance of the return, σ_p, too large (or too small) for him. It leads to a different choice of e_p. In fact, the investor has to decide which combination of expected return and variance suits him best. The choice depends on his level of *risk-aversion*.

Minimum variance portfolio (mvp)
From (33) it is seen that minimum portfolio variance is achieved when e_p is taken equal to $e_p = A/C$. Then $\sigma_p^2 = 1/C =: \sigma_{mvp}^2$.

The covariance between two frontier portfolios
Let p and q be two frontier portfolios. To calculate the covariance of returns of the portfolios p and q we start from

$$cov(r_p, r_q) = w_p^T V w_q.$$

Applying (30), we have

$$cov(r_p, r_q) = (g + he_p)^T V(g + he_q).$$

Exercise: Substitute g, h from (31) and (32), and show the equality

$$cov(r_p, r_q) = \frac{C}{D}\left(e_p - \frac{A}{C}\right)\left(e_q - \frac{A}{C}\right) + \frac{1}{C} \tag{34}$$

Note that in case of $p = q$ the formula (34) reduces to (33), as required.

14 Introducing risk-free assets into the portfolio

The discussion in the previous section, the Markowitz portfolio theory, was based upon the assumption that the portfolio can only be composed from risky assets (stock, e.g.) In fact, there are always riskless assets available, e.g. short-term government bills, bank account etc. The inclusion of a risk-free asset into the model is the basic factor that turns the Markowitz theory into so called *Capital Asset Pricing Model (CAPM)*. This innovation is generally regarded as the contribution of William Sharpe, although Lintner and Mossin developed similar theories at about the same time, in 1960's.

14.1 Combination of risky portfolio and riskless asset

Interestingly, an investor can improve his or her risk/expected return balance by investing partially in a portfolio of risky assets and partially in a risk-free asset. Let us see why this is true.

Consider any portfolio A consisting of n risky assets available on the market of interest. The portfolio A can be imagined as a point inside the Markowitz bullet created by these assets. Let r_A be the return on the portfolio A, $e_A = E(r_A)$ its expected return, and σ_A^2 the variance of the return (risk on A). Denote by r_f the return on an riskless asset. The term 'riskless' means that r_f is a constant. Compose now a new portfolio c (**c**ombined) by investing the fraction X of original funds in the portfolio A and the fraction $1 - X$ in the riskless asset. We allow for X to take any non-negative value. A value $X > 1$ corresponds to the borrowing of additional money (at risk-free rate r_f) and investing it in the portfolio A. As the return of the complete portfolio is $r_c = Xr_A + (1 - X)r_f$, its expected return is

$$e_c = Xe_A + (1 - X)r_f \tag{35}$$

and the risk of the combination, since r_f is a constant, is

$$\sigma_c^2 = D(r_c) = D(Xr_A + (1 - X)r_f) = D(Xr_A) = X^2\sigma_A^2.$$

Hence

$$\sigma_c = X\sigma_A.$$

Solving this for X gives

$$X = \frac{\sigma_c}{\sigma_A}.$$

Substituting this expression for X into (35) yields

$$e_c = \frac{\sigma_c}{\sigma_A}e_A + (1 - \frac{\sigma_c}{\sigma_A})\ r_f.$$

Rearranging terms,

$$e_c = r_f + \frac{e_A - r_f}{\sigma_A} \cdot \sigma_c. \tag{36}$$

Note that this is the equation of a straight line on (σ, e)–plane. Therefore, all combinations of riskless lending (or borrowing) and the portfolio A lie on a straight line in risk-expected return space. The intercept of the straight line is r_f and the slope is $(e_A - r_f)/\sigma_A$. This line is shown in the Figure below. It starts from the point $(0, r_f)$ and passes through the point (σ_A, e_A) inside the Markowitz bullet. Note that to the left of the point A we have combinations of lending money and portfolio A, whereas to the right of the point A we have combinations of borrowing money and portfolio A.

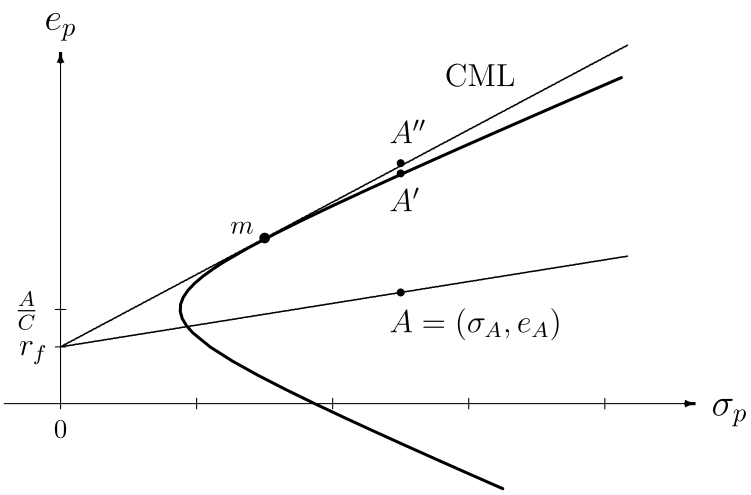

Figure 4: Capital Market Line. The symbol 'm' denotes the market portfolio.

Figure: The risk and expected return of combinations of an risky portfolio A and a riskless asset. Highest expected return (given the risk) is obtained when A is chosen as the tangency point m between the line and the efficient frontier.

It is understood that given a risk level σ_c, any rational investor prefers the portfolio which gives maximum expected return at that risk level. Therefore, the portfolio A is outperformed by another portfolio A' which lies on the upper branch of the portfolio frontier, has the same risk level σ_A but possesses higher expected return (see the figure). The portfolio A', in turn, is outperformed by a portfolio A'' which lies on the tangent line starting from the point $(0, r_f)$ and passes through the tangent point m. In our figure, A'' is located on the right of the tangent portfolio m, meaning that the portfolio A'' is a combination of 1) the portfolio m and 2) additional money borrowed at rate r_f and invested in m. We conclude that *rational investors only invest in a portfolio from the tangent line* and the investors only differ in fractions X and $1 - X$ they place in m and in riskless asset.

The portfolio m is called **market portfolio**, and the tangent line is called **Capital Market Line** (CML). To repeat, the whole CML consists of portfolios that can be combined from the portfolio m and risk-free asset by varying their fractions X and $1 - X$ (note that $X > 1$ is allowed, corresponding to the borrowing money at rate r_f and investing in m).

The term 'market portfolio' is well justified. Indeed, assuming that all investors are rational and knowing that a rational investor only invests in m (and partly in risk-free asset), the weights $(w_1, \ldots, w_n)$ used for investing in n risky assets are the same for all investors. Therefore the total money invested in the market follows the same structure $(w_1, \ldots, w_n)$ i.e. m becomes proportional to market capitalizations of n assets.

14.2 Derivation of the Capital Market Line

Our aim here is to derive the weights corresponding to the market portfolio m. By definition, market portfolio maximizes the slope of the straight line (36). Therefore we seek to find weights $w = (w_1, \ldots, w_n)$ of n risky assets in A, subject to the constraint that $\sum_{i=1}^{n} w_i = 1$, which solves the following maximization problem:

$$\begin{aligned} s := \frac{e_A - r_f}{\sigma_A} &\quad \longrightarrow \quad max_w \\ w^T \mathbb{1} &= 1. \end{aligned}$$

Since

$$e_A = w^T e$$

and

$$\sigma_A = (w^T V w)^{1/2},$$

the slope s can be expressed as

$$s = \frac{w^T e - r_f}{(w^T V w)^{1/2}}.$$

Using Lagrange multipliers, we first define

$$L = \frac{w^T e - r_f}{(w^T V w)^{1/2}} + \lambda(1 - w^T \mathbb{1}).$$

Now take partial derivatives and set the results to zero. Rules of matrix derivation give

$$\begin{aligned}\frac{\partial L}{\partial w} &= \frac{e(w^TVw)^{1/2} - (w^TVw)^{-1/2}Vw(w^Te - r_f)}{w^TVw} - \lambda\mathbb{1} = \mathbb{0},\\ \frac{\partial L}{\partial \lambda} &= w^T\mathbb{1} - 1 = 0.\end{aligned}$$

By substituting back $w^Te = e_A$ and $(w^TVw)^{1/2} = \sigma_A$ into the first equation, it is easy to obtain

$$\sigma_A^2 e - Vw(e_A - r_f) = \sigma_A^3\lambda\mathbb{1}. \tag{37}$$

Multiplication both sides by w^T from the left and using $w^T\mathbb{1} = 1$ yields

$$\sigma_A^2 e_A - \sigma_A^2(e_A - r_f) = \sigma_A^3\lambda.$$

Hence

$$\sigma_A^2 r_f = \sigma_A^3\lambda,$$

from which we obtain

$$\lambda = \frac{r_f}{\sigma_A}.$$

Substitute this expression for λ into (37) to get

$$\sigma_A^2 e - Vw(e_A - r_f) = \sigma_A^2 r_f\mathbb{1}.$$

Rearranging and multiplication of both sides by V^{-1} gives

$$\frac{e_A - r_f}{\sigma_A^2} w = V^{-1}(e - r_f\mathbb{1}). \tag{38}$$

We can calculate the factor $\frac{e_A - r_f}{\sigma_A^2}$ by multiplying (38) from the left by $\mathbb{1}^T$. As $\mathbb{1}^T w = 1$, we have

$$\frac{e_A - r_f}{\sigma_A^2} = \mathbb{1}^T V^{-1}(e - r_f\mathbb{1}) =: \delta.$$

Now, solving (38) for w gives

$$w = \frac{1}{\delta}V^{-1}(e - r_f\mathbb{1}).$$

As this defines weights of the market portfolio m, we denote the solution by w_m:

$$w_m = \frac{1}{\delta}V^{-1}(e - r_f\mathbb{1}). \tag{39}$$

We will use this formula of the market portfolio to derive the CAPM model.

We obtain some further insight into (39) by considering a simple case of independent assets. In such a case $V^{-1} = diag(\frac{1}{\sigma_i^2})$ and hence

$$w_i = \frac{1}{\delta} \cdot \frac{e_i - r_f}{\sigma_i^2},$$

showing that the higher the expected return of the asset i (normed by its risk σ_i^2), the larger is its weight in the portfolio.

As we saw, rational investors only invest in the market portfolio m (and partly in r_f, the weights of m should track the market capitalization. In practise, however, it is not the case. There are several factors why investors are not always 'rational': 1) there are regulators, 2) investors can not afford trading all assets available. Even big investors make choice of 20-40 assets among several hundreds. 3) Different investors use different estimates for the model parameters (means, covariances,. ..).

15 Capital asset pricing model

The models in the two preceding sections assume that the covariance matrix V is known. However, financial institutions normally follow n=150 ... 250 securities, and the number of parameters (covariances) to be estimated, n(n+1)/2 is about 20 thousand! Hence the models that simplify all analysis are welcome. Capital asset pricing model was proposed independently by Sharpe, Lintner and Mossin in 1960-s.

More on Market Portfolio

According to our theory all rational investors will invest in the market portfolio, along with some amount of risk-free asset. This has some profound consequences for this portfolio. First, the market portfolio must contain all assets on the market for if an asset is not in the market portfolio, no one will want to purchase it and so the asset will die out.

Since the market portfolio contains all assets, the portfolio has no specific risk- this risk has been diversified out. Thus all risk associated with the market portfolio is systematic risk.

In practice, the market portfolio can be approximated by a much smaller number of assets. Studies have indicated that a portfolio can achieve a degree of diversification approaching to that of a true market portfolio if it contains a well-chose set of perhaps 20-40 securities. We will use the term market portfolio to refer to an unspecified portfolio that is highly diversified and thus can be considered as essentially free of unsystematic risk.

The risk-return of an asset compared with the market portfolio.

Let us consider any particular asset k in the market portfolio. Our idea is to use the best linear predictor to approximate the return r_k on asset k by a linear function on the return r_m of the entire market portfolio, i.e.

$$r_k = \alpha_k + \beta_k r_m + \varepsilon,$$

where, as we know from the theory of simple linear regression analysis,

$$\begin{aligned} \beta_k &= \frac{cov(r_k, r_m)}{\sigma_m^2}, \\ \alpha_k &= e_k - \beta_k e_m, \\ e_k &= E(r_k) \end{aligned}$$

and ε is the error of the model (residual random variable).

Consider two different cases - large betas and small betas.

Figure 1. A large beta and different magnitudes of error.
The market risk is "magnified" in the asset risk.

Figure 2. A small beta and different magnitudes of error.
The asset risk is relatively insensitive with respect to the market risk, the market risk is "demagnified" in the asset risk.

It is also seen from the graphs in the figures above that there is another factor that contributes to the asset's risk, a factor that has nothing to do with the market risk - it is the error. The larger the error ε (measured, e.g., by its variance σ_ε^2), the larger the uncertainty in the asset's expected return.

Now let us express this mathematically.

The risk

Recall, first, that the error term of linear regression is not correlated with the regressor. In our case it means that

$$cov(r_m, \varepsilon) = 0.$$

Therefore the risk associated with the asset k is (since α_k is constant)

$$\begin{aligned}\sigma_k^2 = D(r_k) &= D(\alpha_k + \beta_k r_m + \varepsilon) \\ &= D(\beta_k r_m + \varepsilon) \\ &= \beta_k^2 \sigma_m^2 + \sigma_\varepsilon^2\end{aligned}$$

Thus, the risk of the asset k consists of two components - quantity $\beta_k^2\sigma_m^2$, called **systematic risk**, and quantity σ_ε^2, called **unsystematic** or **unique** or **idiosyncratic risk**. Systematic risk is proportional to the market risk, with a proportionality factor of β_k^2.

It turns out that, when adding an asset to a *diversified* portfolio, the unique risk of that asset is canceled out by other assets in the portfolio. Hence, the unique risk should not be considered when evaluating the risk-return performance of the asset and so the asset's beta becomes the crucial point for the risk-return analysis of an asset. Next we will justify this viewpoint.

The expected return

Consider the expected return of the market portfolio

$$e_m = e^T w_m,$$

where e is the vector of expected returns on all n assets, $e = (e_1, \ldots, e_n)$.

The expected return on an individual asset k can be written as

$$e_k = e^T \mathbb{1}_k,$$

where the vector 1_k is defined as

$$\mathbb{1}_k = (0, \ldots, 0, 1, 0, \ldots, 0)^T.$$

To relate these two quantities, we need an expression for e.

Recall that, according to (39), the weights of the market portfolio m verify

$$w_m = \delta V^{-1}(e - r_f \mathbb{1}),$$

where δ is a constant. Solving for e gives

$$e = \delta V w_m + r_f \mathbb{1}.$$

We can now write

$$
\begin{aligned}
e_m = e^T w_m &= (\delta V w_m + r_f \mathbb{1})^T w_m \\
&= \delta w_m^T V w_m + r_f \mathbb{1}^T w_m \\
&= \delta \sigma_m^2 + r_f.
\end{aligned}
$$

In the same way

$$
\begin{aligned}
e_k = e^T 1_k &= (\delta V w_m + r_f \mathbb{1})^T 1_k \\
&= \delta w_m^T V 1_k + r_f \mathbb{1}^T 1_k \\
&= \delta \cdot cov(r_k, r_m) + r_f.
\end{aligned}
$$

Therefore we can express the beta of asset k in terms of expected returns:

$$\beta_k = \frac{cov(r_k, r_m)}{\sigma_m^2} = \frac{\delta(e_k - r_f)}{\delta(e_m - r_f)} = \frac{e_k - r_f}{e_m - r_f}.$$

Finally, solving this for e_k gives

$$e_k = \beta_k(e_m - r_f) + r_f.$$

These important formulae are collected in the following theorem.

Theorem. *The expected return and risk of an asset k in the market portfolio is related to the asset's beta as follows:*

$$e_k = \beta_k(e_m - r_f) + r_f \tag{40}$$

and

$$\sigma_k^2 = \beta_k^2 \sigma_m^2 + \sigma_\varepsilon^2. \tag{41}$$

The most important observation here is that asset's expected return depends only on the asset's systematic risk $\beta_k^2 \sigma_m^2$ (through its beta) and not on its unique risk σ_ε^2. This justifies considering only the term $\beta_k^2 \sigma_m^2$ in assessing the asset's risk relative to the market portfolio.

The graph of the line in equation (40) is called **security market line (SML)**. The equation shows that the expected return of an asset is equal to the return of the risk-free asset plus the **risk premium** $\beta_k(e_m - r_f)$ of the asset. The security market line is shown in the figure below.

Under normal conditions the slope of SML, i.e. $e_m - r_f$ is positive. Hence large betas imply large expected returns and vice versa. This makes sense - the more (systematic) risk in an asset, the higher should be its expected return under market equilibrium.

The security market line is shown in the following figure.

Figure. The security market line.

Under market equilibrium, all assets should ideally lie on the SML line. Recall that this assumes rational investors, no investment restrictions, and other conditions - the conditions that are never fulfilled entirely. Therefore, in practice the assets are scattered around the SML, in accordance to their estimated risk and returns. However, the SML has an 'attractive power' which can be explained as follows. If an asset is returning less than the market feels is reasonable with respect to the asset's perceived risk (the points below the SML), then no one will buy that asset and its price will decline, thus increasing the asset's future return. Similarly, if the asset is returning more than the market feel is required by the asset's level of risk (the points above the SML), then more investors will buy the asset, thus raising its price and lowering its expected return. Thus there is a tendency for all assets to move closer to SML.

Example. Suppose that the riskfree rate is 3% and that the market portfolio's risk is 8%. Then the slope of SML is $e_m - r_f = 0.08 - 0.03 = 0.05$, and the security market equation is

$$e_k = 0.05\beta + 0.03$$

We can now compute expected returns under market equilibrium. For example, the value of beta $\beta_1 = 1$ gives the expected return $e_1 = 0.08$, equal to that of the market portfolio. Such an asset has the same systematic risk as the whole market (or market portfolio). However, if the beta is less than 1, say $\beta_2 = 0.8$, then the asset has a smaller risk than the market portfolio. Therefore, the market will sustain a lower expected return than that of the market portfolio, namely $e_2 = 0.05 \times 0.8 + 0.03 = 0.07 < 0.08 = e_m$.