Chapter 1

Comparison of Random Variables.
Preferences of Individuals

This chapter concerns various rules of comparison and subsequent selection among risky alternatives.

1 COMPARISON OF RANDOM VARIABLES.
SOME PARTICULAR CRITERIA

1.1 Preference order

What do we usually do when we choose an investment strategy in the presence of uncertainty? Consciously or not, we compare random variables (r.v.’s) of the future income, corresponding to different possible strategies, and we try to figure out which of these r.v.’s is the “best”.

Suppose you are one of 2 million of people who buy a lottery ticket to win a single one million dollar prize. Your income is a random variable (r.v.)

$$\xi = \begin{cases} 
1,000,000 & \text{with probability } \frac{1}{2,000,000}, \\
0 & \text{with probability } 1 - \frac{1}{2,000,000}.
\end{cases}$$

If the ticket’s price is $1, then your random profit is $\xi - 1$. If you decide to buy the ticket, it means that, when comparing the r.v.’s $X = \xi - 1$ and $Y = 0$ (the profit if you do not buy), you have decided, maybe at an intuitive level, that $X$ is better for you than $Y$.

The fact that the mean value $E\{X\} = E\{\xi\} - 1 = \frac{1}{2} - 1 = -\frac{1}{2}$ is negative does not say that the decision is unreasonable. You pay for hope or for fun.

Suppose you buy auto insurance against a possible future loss $\xi$. Assume that with probability 0.9 the r.v. $\xi = 0$ (nothing happened), and with probability 0.1, the loss $\xi$ takes on values between zero and $2000$, and all these values are equally likely. In this case, $E\{\xi\} = 0.1 \cdot 1000 = 100$. If the premium $c$ you pay is equal, say, to $110$, it means that the loss of $\xi$ is worse for you than the loss of the certain amount $c = 110$. The fact that you pay $10$ more than the mean loss, again, does not necessarily mean that you made a mistake. The additional $10$ may be viewed as a payment for stability.

For the insurance company the decision in this case is, in a sense, the opposite. The company gets your premium $c$, and it will pay you a random payment $\xi$. The company
1. COMPARISON OF RANDOM VARIABLES

compares the r.v. $X = c - \xi$ with the r.v. $Y = 0$, and if the company signs the insurance contract, it means that it has decided that the random income $c - \xi$ is better than zero income.

In the reasoning above, we assumed that decision did not depend on the total wealth of the investor but just on the r.v.’s under comparison. Suppose that in the case of insurance you also take into account your total wealth or a part of it, which we denote by $w$. Then the r.v.’s under comparison are $w - \xi$ (your wealth if you do not insure the loss) and $w - c$ (your wealth if you do insure the loss for the premium $c$).

In the examples we considered, one of the variables under comparison was non-random. Certainly, this is not always the case. For example, if you decide to insure only half of the future loss $\xi$ for a lower premium $c'$, then the r.v.’s we should consider are $X = w - \xi$ (you do not buy an insurance) and $Y = w - \frac{\xi}{2} - c'$ (you insure half of the loss for $c'$).

Thus, we may consider rather arbitrary r.v.’s.

This chapter addresses various criteria for the comparison of risky alternatives. In general, we will talk about possible values of future income. While the criteria may vary, they usually have one feature in common. When choosing a possible investment strategy, we have competing interests: we want the income to be large, but we also want the risk to be low. As a rule, we can reach a certain level of stability only by sacrificing a part of the income – we should pay for stability. So, our decision becomes a trade-off between the possible growth and stability.

Now, let us consider the general framework where we deal with a fixed class $X = \{X\}$ of r.v.’s $X$. We assume that all r.v.’s $X$ from $X$ are defined on a same sample space $\Omega = \{\omega\}$ (see Section 0.1.3.1). That is, $X = X(\omega)$.

Defining a rule of comparison on the class $X$ means that for each pair $(X, Y)$ of r.v.’s from $X$, we should determine whether $X$ is better than $Y$, or $X$ is worse than $Y$, or these two random variables are equivalent for us.

Formally, it means that among all possible pairs $(X, Y)$ (the order of the r.v.’s in the pair $(X, Y)$ is essential), we specify a collection of those pairs $(X, Y)$ for which $X$ is preferable or equivalent to $Y$. In other words, “$X$ is not worse than $Y$”, and as a rule we will use the latter terminology. We will write it as $X \succsim Y$.

If $(X, Y)$ does not belong to the collection mentioned, we say “$X$ is worse than $Y$” or “$Y$ is better than $X$”, writing $X \prec Y$ or $Y \succ X$, respectively. If simultaneously $X \succsim Y$ and $Y \succsim X$, we say that “$X$ is equivalent to $Y$”, writing $X \simeq Y$.

Not stating it each time explicitly, we will always assume that the relation $\succsim$ satisfies the following two properties:

(i) for any $X$ and $Y$ from $X$, either $X \succsim Y$, or $Y \succsim X$ (as was mentioned, these relations may hold simultaneously);

(ii) for any $X$, $Y$, and $Z$ from $X$, if $X \succsim Y$ and $Y \succsim Z$, then $X \succsim Z$. This property is referred to as a transitivity property.

The rule of comparison so defined is called a preference order on the class $X$.

Before discussing examples, we state one general requirement on preference orders. This requirement is quite natural when we view $X$’s as the r.v.’s of future income.
The monotonicity property:

\[
\text{If } X, Y \in X \text{ and } P(X \geq Y) = 1, \text{ then } X \succeq Y. \tag{1.1.1}
\]

This requirement reflects the rule “the larger, the better”. If a r.v. (or a random income) \(X(\omega)\) is greater than or equal to a r.v. \(Y(\omega)\) for all \(\omega\)’s or, at least, with probability one, then for us \(X\) is not worse than \(Y\).

It makes sense to emphasize that in (1.1.1) we consider not all r.v.’s but only those from the class \(X\) under consideration. We will see later that this is an important circumstance.

It is natural to consider also

The strict monotonicity property:

\[
\text{If } X, Y \in X, \ P(X \geq Y) = 1, \text{ and } P(X > Y) > 0, \text{ then } X \succ Y. \tag{1.1.2}
\]

The significance of this requirement is also clear. If a random income \(X = X(\omega)\) is not smaller than \(Y = Y(\omega)\) with probability one, and with some positive probability \(X(\omega)\) is larger than \(Y(\omega)\), then we prefer \(X\) to \(Y\).

In this book, we will accept only preference orders which are monotone in the class of the r.v.’s under consideration. However, we will not always require strict monotonicity; see for example, the VaR criterion in Section 1.2.2. Nevertheless, if a rule of comparison is not strictly monotone, this says about some non-flexibility of this rule, and it makes sense, at least, to recheck to what extent it meets our goals.

EXAMPLE 1. Consider two r.v.’s, \(X = X(\omega)\) and \(Y = Y(\omega)\), defined on a sample space \(\Omega\) consisting of only two outcomes: \(\omega_1\) and \(\omega_2\). The probabilities of the outcomes, \(P(\omega_1)\) and \(P(\omega_2)\), are equal, say, to 1/2. We may view \(X, Y\) as the random income corresponding to two investment strategies, and \(\omega_1, \omega_2\) as two states of the future market. Let

\[
\begin{align*}
X(\omega_1) &= 1, X(\omega_2) = 3, \\
Y(\omega_1) &= 1, Y(\omega_2) = 2.
\end{align*}
\]

Clearly, \(X(\omega) \geq Y(\omega)\) for both \(\omega\)’s. So, if our preference order \(\succeq\) is monotone, for us \(X \succeq Y\), i.e., \(X\) is not worse than \(Y\), which is natural. Assume, however, that the order \(\succeq\) is monotone but is not strictly monotone. Then it may turn out that, though \(P(X > Y) = \frac{1}{2} > 0\), the r.v. \(X\) is equivalent to \(Y\). This means that we are indifferent whether to choose \(X\) or \(Y\).

The only way to interpret it, is to say that we need at most two units of money, and do not need more. If, as a matter of fact, it is not true, we should describe our preferences in a more flexible way. \(\Box\)

We say that an order \(\succeq\) is preserved, or completely characterized, by a function \(V(X)\) if for any \(X, Y \in X\),

\[
X \succeq Y \iff V(X) \geq V(Y). \tag{1.1.3}
\]
1. COMPARISON OF RANDOM VARIABLES

The symbol ⇔ means if and only if, also abbreviated iff.

The function $V(X)$ may be viewed as a measure of “quality” of $X$: the larger $V(X)$, the better $X$, and $X$ is better than $Y$ iff $V(X) \geq V(Y)$.

Suppose there exists a function $V(X)$ with the property (1.1.3). Then the monotonicity property may be restated as

$$\text{If } X, Y \in X \text{ and } P(X \geq Y) = 1, \text{ then } V(X) \geq V(Y).$$  (1.1.4)

The strict monotonicity is equivalent in this case to the property

$$\text{If } X, Y \in X, \text{ and } P(X \geq Y) = 1, \text{ and } P(X > Y) > 0, \text{ then } V(X) > V(Y).$$  (1.1.5)

Let us turn to examples.

1.2 Several simple criteria

We will talk about preferences of economic agents – separate individuals, companies, using also, for brevity, the term “investor”.

1.2.1 The mean-value criterion

The investor cares only about the mean values of r.v.’s, that is,

$$X \succeq Y \iff E\{X\} \geq E\{Y\}.$$  

In this case, the collection of all pairs $(X, Y)$ mentioned above is just the collection of all pairs $(X, Y)$ for which $E\{X\} \geq E\{Y\}$, and in (1.1.3), $V(X) = E\{X\}$.

Clearly, this criterion is strictly monotone; the reader is invited to show it on her/his own.

Note, however, that from the mean-value criterion’s point of view, for example, r.v.’s

$$X = \begin{cases} 100 & \text{with probability } 1/2, \\ 0 & \text{with probability } 1/2, \end{cases} \text{ and } Y = 50$$  \hspace{1cm} (1.2.1)

are equivalent. Certainly, it is a very primitive rule of comparison, but as we will see, sometimes quite reasonable comparison rules turn out to be close to the mean-value criterion.

1.2.2 Value-at-Risk (VaR)

Another term in use is the capital-at-risk criterion. For a r.v. $X$, denote by $q_\gamma = q_\gamma(X)$ its $\gamma$-quantile or the $100\gamma$th percentile. The reader is recommended to look up the rigorous definition in Section 0.1.3.4, and especially Fig.0-7 there. Non-formally, it is the largest number $x$ for which $P(X \leq x) \leq \gamma$. If the r.v. $X$ is continuous and its distribution function (d.f.) $F(x)$ is increasing, $q_\gamma$ is just the number for which $F(q_\gamma) = \gamma$. See also Fig.0.7a.

Let $\gamma$ be a fixed level of probability, viewed as sufficiently small. Assume that an investor does not take into consideration events whose probabilities are less than $\gamma$. Then for such an investor the worst, smallest conceivable level of the income is $q_\gamma$. 

Let, for instance, $\gamma = 0.05$. Then $q_{0.05}$ is the smallest value of the income among all values which may occur with 95% probability. One may say that $q_{0.05}$ is the value at 5% risk. Note that $q_{\gamma}$ may be negative, which corresponds to losses.

The VaR criterion is defined as

$$X \succeq Y \iff q_{\gamma}(X) \geq q_{\gamma}(Y),$$

i.e., we set $V(X) = q_{\gamma}(X)$ in (1.1.3).

In applications of VaR, for the $\gamma$-quantile of $X$, the notation $\text{VaR}_\gamma(X)$ is frequently used; we will keep the notation $q_{\gamma}(X)$.

The particular choice of $\gamma = 0.05$ is very common, but it has rather a psychological explanation: 0.01 is “too small”, while 0.1 is “too large”. As a matter of fact, whether a particular value of $\gamma$ should be viewed as small or not depends on the situation. We can view a probability of 0.05 as small if it is the probability that there will be a rain tomorrow. However, the same number should be considered very large if it is the probability of being involved in a traffic accident: it would mean that on the average you are likely to be involved in an accident one out of twenty times you are in traffic.

EXAMPLE 1. Let a r.v. $X$ (say, a random income) take values 0, 10, 20, 30 with probabilities 0.1, 0.3, 0.3, 0.3, respectively, and a r.v. $Y$ take on the same values with probabilities 0.07, 0.31, 0.31, 0.31. Then for $\gamma = 0.05$, we have $q_{\gamma}(X) = q_{\gamma}(Y) = 0$ (check on your own), and r.v.’s $X$ and $Y$ are equivalent under the VaR criterion. For $\gamma = 0.08$ we have $q_{\gamma}(X) = 0$, and $q_{\gamma}(Y) = 10$, that is, $X$ is worse than $Y$. So, the result of comparison depends on $\gamma$. □

The VaR criterion is monotone. Indeed, if $X \geq Y$ with probability one, then $P(X \leq x) \leq P(Y \leq x)$, so in this case $q_{\gamma}(X) \geq q_{\gamma}(Y)$. However, VaR is not strictly monotone.

EXAMPLE 2. Let $Y$ be uniform on $[0, 2]$, and

$$X = \begin{cases} Y & \text{if } Y \leq 1 \\ 2 & \text{if } Y > 1. \end{cases}$$

We see that $P(X > Y) = P(1 < Y < 2) = \frac{1}{2} > 0$. However, if $x \leq 1$, then $P(X \leq x) = P(Y \leq x)$, and hence $q_{\gamma}(X) = q_{\gamma}(Y)$ if $\gamma \leq \frac{1}{2}$. □

The fact that the VaR criterion is not strictly monotone does not provide sufficient grounds to reject the VaR. However, we should be aware that this is not a flexible criterion since it does not take into account all values of r.v.’s, as we saw in the example above.

EXAMPLE 3. Let $X$ be normal with mean $m$ and variance $\sigma^2$. Since the d.f. of $X$ is $\Phi(\frac{x-m}{\sigma})$ [see, e.g., Section 0.3.2.4], the $\gamma$-quantile of $X$ is the solution to the equation $\Phi(\frac{x-m}{\sigma}) = \gamma$. Denote by $q_{\gamma}$, the $\gamma$-quantile of the standard normal distribution, i.e., $\Phi(q_{\gamma}) = \gamma$. Then we can rewrite the equation mentioned as $\frac{x-m}{\sigma} = q_{\gamma}$, and

$$q_{\gamma}(X) = m + q_{\gamma}\sigma.$$

The coefficient $q_{\gamma}$ depends only on $\gamma$. Usually people choose $\gamma < 0.5$, and in this case $q_{\gamma} < 0$. For example, if $\gamma = 0.05$, then $q_{\gamma} \approx -1.64$ (see Table 2 in the Tables in the end of
1. COMPARISON OF RANDOM VARIABLES

the book), and the VaR criterion is preserved by the function $q_{\gamma}(X) \approx m - 1.64\sigma$. Criteria of the type

$$V(X) = m - k\sigma,$$

where $k$ is a positive number, are frequently used in practice, and not only for normal r.v.’s – as we will see in Section 1.2.5, maybe too frequently. The expression in (1.2.2) can be interpreted as follows. If we view $X$ as a future income and variance as a measure of riskiness, then we want the mean $m$ to be as large as possible and $\sigma$ as small as possible. This is reflected by the minus sign in (1.2.2). The number $k$ may be viewed as the weight we assign to variance.

EXAMPLE 4. There are $n = 10$ assets with random returns $X_1, \ldots, X_n$. The term “return” means that, if you invest $1 into, say, the first asset, you will get an income of $X_1$ dollars. For example, if the today price of a stock is $11, while the yesterday price was $10, the return for this period is 1.1. Note that the returns $X_i$’s may be less than one.

Assume that $X_1, \ldots, X_n$ are independent and their distributions are closely approximated by the normal distribution with mean $m$ and variance $\sigma^2$.

Let us compare two strategies of investing $\$n$ million dollars: either investing the whole sum into one asset, for example, into the first, or distributing the investment sum equally between $n$ assets. We proceed from the VaR criterion with $\gamma = 0.05$.

For the first strategy, the income will be the r.v. $Y_1 = nX_1 = 10X_1$. The mean $E\{Y_1\} = nm$, and $\text{Var}\{Y_1\} = n^2\sigma^2$, so to compute $q_{\gamma}(Y_1)$ we should replace in (1.2.2) $m$ by $nm$, and $\sigma$ by $n\sigma$. Replacing $q_{\gamma}$ by its approximate value $1.64$, we have

$$q_{\gamma}(Y_1) = mn - 1.64n\sigma = 10m - 16.4\sigma.$$

For the second strategy, the income is the r.v. $Y_2 = X_1 + \ldots + X_n$. Hence, $E\{Y_2\} = nm$, $\text{Var}\{Y_1\} = n\sigma^2$, and

$$q_{\gamma}(Y_2) = mn - 1.64\sqrt{n}\sigma \approx 10m - 5.2\sigma.$$  

Thus, the second strategy is preferable, which might be expected from the very beginning. Nevertheless, in the next example we will see that if the $X_i$’s have a distribution different from normal, we may jump to a different conclusion. □

EXAMPLE 5\textsuperscript{1}. There are ten independent assets such that investment into each with 99% probability gives 4% profit, and with 1% probability the investor loses the whole investment. Assume that we invest $\$10 million and compare the same two strategies as in Example 4. Let us again apply the VaR criterion with $\gamma = 0.05$.

If we invest all $\$10 million into the first asset, we will get $\$10.4 million with probability 0.99, and in the notation of the previous example $q_{\gamma}(10X_1) = 10.4$.

For the second strategy, the number of successful investments has the binomial distribution with parameters $p = 0.99$, $n = 10$. If the number of successes is $k$, the income is $k \times 1\text{million} \times 1.04$. The d.f. $F(x)$ of the income is given in the table below. The values of $F(x)$ are the values of the binomial d.f. with the parameters mentioned.

\textsuperscript{1}This example is very close to an example from [1] presented also in [6, p.14] with the corresponding reference.
1. Some Particular Criteria

The 0.05-quantile of this distribution is $9.36 < 10.4$. Therefore, following VaR, we should choose the first investment strategy.

Note, however, that if we choose as $\gamma$ a number slightly smaller than 0.01 – for example $\gamma = 0.0095$, then the result will be different. In this case, $q_{\gamma}(10X_1) = 0$, while the 0.0095-quantile of the distribution presented in the table, is again 9.36.

Certainly, the results of the comparison above should not be considered real recommendations. On the contrary, the last example indicates a limitation of the application of VaR, and shows that this criterion is quite sensitive for the choice of $\gamma$. □

The reader can find more about the VaR criterion, for example, in [57], [63], [61]. Some references may be found also in http://www.riskmetrics.com and http://www.gloriamundi.org.

1.2.3 An important remark: risk measures rather than criteria

This simple but important remark concerns the two criteria above and practically all other criteria we will consider in this chapter. The point is that we do not have to limit ourselves to using only one criteria each time. On the contrary, we can combine them.

For example, when considering a random income $X$, we may compute its expectation $E\{X\}$ and its quantile $q_{\gamma}(X)$. In this case, we will know what we can expect on the average, and what is the worst conceivable (or likely) outcome. When comparing two r.v.’s, we certainly may take into account both characteristics. How we will do this depends on our preferences. The simplest way is to consider the linear combination $\alpha E\{X\} + \beta q_{\gamma}(X)$, where $\alpha$ and $\beta$ play the role of weights we assign to the mean and to the quantile. The larger $\beta$, the more pessimistic we are.

Under such an approach to risk assessment, various functions $V(X)$ present not criteria but possible characteristics of the random income $X$. In this case, we call $V(X)$ a risk measure.

\[
\text{(Route 1 \Rightarrow page 73)}
\]

1.2.4 Tail conditional expectation (TCE) or Tail-Value-at-Risk (TailVaR)

Next we consider a modification of VaR. The reason that we do it is illustrated by the following example.
EXAMPLE 1. Consider two r.v.'s $X$ and $Y$ of the future income such that $X$ takes on values $-2, -1, 10, 20$ with probabilities $0.01, 0.02, 0.47, 0.5$. Similarly, $Y$ takes on values $-2\cdot10^6, -1, 10, 20$ with probabilities $0.01, 0.01, 0.48, 0.5$. The probabilities that the income will be negative in both cases are small: 3% and 2%. For $\gamma = 0.025$, we would have $q_\gamma(X) = -1$ and $q_\gamma(Y) = 10$. So, under the VaR criterion, $Y$ is preferable, which does not look natural. While we may neglect negative values of the income in the first case, this may be unreasonable in the second: a loss of $2$ million can be too serious to ignore, even if such an event occurs with a small probability of 1%. □

In situations as above, we speak about the possibility of large deviations, or a heavy tail of the distribution (for the term “tail”, see also Section 0.2.6). We compare the “tails” of different distributions in more detail in Section 2.1.1. For now, we introduce a criterion which involves the mean values of large deviations.

First, consider the function

$$V(X; t) = E\{X | X \leq t\}, \quad \text{(1.2.3)}$$

the mean value of $X$ given that the income $X$ did not exceed a level $t$.

[Formally, the right member of (1.2.3) is defined by

$$E\{X | X \leq t\} = \frac{1}{P(X \leq t)} \int_{-\infty}^{t} x dF(x) = \frac{1}{F(t)} \int_{-\infty}^{t} x dF(x), \quad \text{(1.2.4)}$$

where $F(x)$ is the d.f. of $X$. The formula covers the cases of discrete and continuous r.v.'s simultaneously, if we understand the integral above as in (0.2.1.5) from Section 0.2.1. We consider conditional expectations in detail in Chapter 4; for now, it is sufficient for us to use definition (1.2.4).]

If we are interested only in losses, then it suffices to consider $t \leq 0$. Then $V(X; t)$ is negative.

Note that in such situations, people often consider not the income but the losses directly, that is, instead of the r.v. $X$, the r.v. $\bar{X} = -X$. Negative values of $X$ correspond to positive values of $\bar{X}$ and vice versa. In this case, $E\{X | X \leq t\} = -E\{\bar{X} | \bar{X} \geq |t|\}$ if $t \leq 0$. The risk measure $E\{\bar{X} | \bar{X} \geq s\}$ is the expected value of the loss given that it has exceeded a level $s$. In insurance, it is called an expected policyholder deficit. See also Exercise 8.

Let us come back to $V(X; t)$ and take, as the level $t$, the $\gamma$-quantile $q_\gamma(X)$. Accordingly, we set

$$V_{\text{tail}}(X) = E\{X | X \leq q_\gamma(X)\},$$

and define the rule of comparison of r.v.'s by the relation

$$X \gtrsim Y \iff V_{\text{tail}}(X) \geq V_{\text{tail}}(Y).$$
EXAMPLE 2. Consider the r.v.’s from Example 1 for \( \gamma = 0.25 \). As we already saw, in this case, \( q_\gamma(X) = -1 \) and \( q_\gamma(Y) = 10 \). If the r.v. \( X_1 \leq -1 \), then it can take on only values \(-1\) and \(-2\). Hence,

\[
V_{\text{tail}}(X) = E\{X | X \leq -1\} = (-2)P(X = -2 | X \leq -1) + (-1)P(X = -1 | X \leq -1)
\]

\[
= (-2) \cdot \frac{0.01}{0.03} + (-1) \cdot \frac{0.02}{0.03} = -\frac{4}{3}.
\]

For \( Y \), computing in the same manner, we have

\[
V_{\text{tail}}(Y) = (\frac{-2 \cdot 10^6}{0.5} \cdot \frac{0.01}{0.5} + (-1) \cdot \frac{0.01}{0.5} + 10 \cdot \frac{0.48}{0.5}) = -39990.42,
\]

which is much less than \(-4/3\). So, \( X \gtrsim Y \). □

Now assume, for simplicity, that the income consists of a fixed non-random positive part and a random loss \( \xi \). Since the positive part is certain, we can exclude it from consideration, setting the income \( X = -\xi \). Let us denote by \( G(x) \) the d.f. of \( \xi \), and set \( G(x) = P(\xi > x) = 1 - G(x) \), the tail of the distribution of \( \xi \). Assume \( G(x) \) to be continuous.

For \( t \leq 0 \), we have

\[
P(X \leq t) = P(\xi \geq -t) = P(\xi \geq |t|) = \overline{G}(|t|),
\]

and

\[
E\{X | X \leq t\} = E\{-\xi | -\xi \leq t\} = -E\{\xi | \xi \geq |t|\} = \frac{-1}{P(\xi \geq |t|)} \int_{|t|}^{\infty} xdG(x)
\]

\[
= \frac{1}{G(|t|)} \int_{|t|}^{\infty} xd(1 - G(x)) = \frac{1}{G(|t|)} \int_{|t|}^{\infty} x d\overline{G}(x).
\]

Integration by parts implies that

\[
E\{X | X \leq t\} = \frac{1}{G(|t|)} \left( -|t|\overline{G}(|t|) - \int_{|t|}^{\infty} \overline{G}(x) dx \right) = - \left( |t| + \frac{1}{G(|t|)} \int_{|t|}^{\infty} \overline{G}(x) dx \right).
\]

EXAMPLE 3. (a) Let \( \overline{G}(x) = 1/(1 + x)^2 \) for \( x \geq 0 \). This is a particular case of the Pareto distribution that we consider in more detail in Section 2.1.1. The tail of this distribution is viewed as “heavy”.

To find \( q = q_\gamma \), we make use of (1.2.5) and write

\[
\gamma = P(X \leq q) = \overline{G}(|q|) = \frac{1}{(1 + |q|)^2},
\]

which implies

\[
|q_\gamma| = \frac{1}{\sqrt{\gamma}} - 1.
\]
1. COMPARISON OF RANDOM VARIABLES

From (1.2.6) it follows that

\[ |E\{X | X \leq q\}| = |q| + \frac{1}{G(|q|)} \int_{|q|}^{\infty} G(x)dx = |q| + (1 + |q|)^2 \int_{|q|}^{\infty} \frac{1}{(1+x)^2}dx \]

\[ = |q| + (1 + |q|)^2 \frac{1}{(1+|q|)} = |q| + (1 + |q|) = 2|q| + 1. \]

Substituting (1.2.8), we have

\[ |E\{X | X \leq q_{\gamma}\}| = 2 \sqrt{\gamma} - 1. \]

The absolute value above corresponds to the losses. For the (negative) income \( X \)

\[ V_{\text{tail}}(X) = E\{X | X \leq q_{\gamma}\} = 1 - \frac{2}{\sqrt{\gamma}}. \]

(b) Let now \( G(x) = e^{-x} \), that is, \( \xi \) is a standard exponential r.v. We can provide the same calculations, but we may avoid it, if we recall that \( \xi \) has the memoryless property (see, e.g., Section 0.3.2.2). By this property, if \( \xi \) has exceeded a level \( x \), the overshoot (over \( x \)) has the same distribution as the r.v. \( \xi \) itself. Since \( E\{\xi\} = 1 \), we can write

\[ E\{\xi | \xi \geq t\} = t + 1. \]

Hence,

\[ |E\{X | X \leq q\}| = E\{\xi | \xi \geq |q|\} = |q| + 1. \]

The reader can double check this, carrying out calculations similar to what we did above.

As in (1.2.7),

\[ \gamma = P(X \leq q) = G(|q|) = e^{-|q|}, \]

and \( |q| = \ln \frac{1}{\gamma} \). Thus,

\[ |E\{X | X \leq q_{\gamma}\}| = \ln \frac{1}{\gamma} + 1, \quad \text{and} \quad V_{\text{tail}}(X) = E\{X | X \leq q_{\gamma}\} = -\ln \frac{1}{\gamma} - 1. \]

(c) Let us compare the two cases above. In the first, \( P(\xi > x) = 1/(1+x)^2 \); in the second, \( P(\xi > x) = e^{-x} \). The former function converges to zero much more slowly than the latter. One may say that the tail in the former case is much “heavier”. It means that the probability to have essential losses is larger in the former case, and we should expect that under the TailVaR criterion the exponential distribution is better.

This is indeed the case for all \( \gamma \). To show it, we should prove that the difference

\[ \left( \frac{2}{\sqrt{\gamma}} - 1 \right) - \ln \frac{1}{\gamma} = 2 \frac{1}{\sqrt{\gamma}} - \frac{1}{\gamma} - 2 \]

is nonnegative for all \( \gamma \). Denote this difference by \( C(\gamma) \). Note that \( C(1) = 0 \), and \( C'(\gamma) = -1 + \frac{1}{\gamma^3/2} = -\frac{1}{\gamma^{3/2}} (\sqrt{\gamma} - 1) < 0 \) for all \( \gamma \in [0,1) \). Hence, \( C(\gamma) > 0 \) for all \( \gamma < 1 \). □
Next, we test the TailVaR on monotonicity. It turns out that in general, the TailVaR criterion is not monotone.

**EXAMPLE 4.** Though we are interested in losses, to make the example illustrative, we will consider non-negative r.v.’s $X = X(\omega)$ and $Y = Y(\omega)$. Subtracting from both r.v.’s a large number $c$, we may come to r.v.’s with negative values, but the result of comparison of $X - c$ and $Y - c$ will be the same as for $X$ and $Y$. Let the space of elementary outcomes $\Omega = \{\omega_1, \omega_2, \omega_3\}$, and the probabilities and the values of $X$ and $Y$ be as follows:

\[
\begin{array}{ccc}
\omega_1 & \omega_2 & \omega_3 \\
P(\omega) &=& 0.1 & 0.4 & 0.5 \\
X(\omega) &=& 0 & 10 & 20 \\
Y(\omega) &=& 0 & 10 & 10
\end{array}
\]

Clearly, $P(X \geq Y) = 1$ and $P(X > Y) = 0.5 > 0$, so it is quite reasonable to prefer $X$ to $Y$.

Set, however, $\gamma = 0.2$. Then, as we can see from the table (or by graphing the d.f.’s of $X$ and $Y$), the quantiles $q_\gamma(X) = q_\gamma(Y) = 10$. Now, $\text{VaR}_\gamma(X) = E\{X | X \leq 10\} = 0 \cdot \frac{1}{5} + 10 \cdot \frac{4}{5} = 8$, while $\text{VaR}_\gamma(Y) = E\{Y | Y \leq 10\} = E\{Y\} = 0 \cdot 0.1 + 10 \cdot 0.9 = 9$. Thus, with respect to the TailVaR, $Y$ is better than $X$, which contradicts common sense. □

Nevertheless, the TailVaR criterion arose as a result of reasonable argumentation. Therefore, it makes sense not to reject it but realize each time in what situations the monotonicity property is fulfilled.

First of all, note that the TailVaR is monotone in the class of continuous r.v.’s. An advice on how to show this is given in Exercise 6.

It may be also shown that if $\Omega$ is finite and all $\omega$’s are equiprobable, then under some mild conditions on r.v.’s or for a slightly modified criterion, monotonicity does take place. We consider it in more detail in Section 1.3.

### 1.2.5 The mean-variance criterion

This criterion is, in essence, the same as (1.2.2), but the motivation and derivation are different. Consider an investor expecting a random income $X$. Set $m_X = E\{X\}$, $\sigma_X^2 = \text{Var}\{X\}$. Suppose the investor identifies the riskiness of $X$ with its variance, and wishes the mean income $m_X$ to be as large as possible and the variance $\sigma_X^2$ – as small as possible. The quality of the r.v. $X$ for such an investor is determined by a function of $m_X$ and $\sigma_X^2$. In the simplest case, it is a linear function, and we can write it as

\[
V(X) = \tau m_X - \sigma_X^2,
\]

where the minus reflects the fact that the quality decreases as the variance increases.

The positive parameter $\tau$ plays the role of a weight assigned to $m_X$: larger $\tau$ indicates that the investor values mean more highly. This parameter is usually called a *tolerance to risk*. (We assigned a weight to the mean, not to the variance, merely following a tradition in Finance; it does not matter which parameter is endowed by a coefficient. For example, we can write $V(X) = \tau(m_X - \frac{1}{\tau} \sigma_X^2)$. Then a weight is assigned to the variance, while the factor $\tau$ in the very front does not change the comparison rule.)
The function \( V(X) \) in (1.2.9) preserves the corresponding preference order \( \succcurlyeq \) among r.v.'s:

\[
X \succcurlyeq Y \iff m_X - \sigma_X^2 \geq m_Y - \sigma_Y^2.
\]

Note that, when writing (1.2.9), we did not assume r.v.'s under consideration to be normal, which we did when deriving a similar criterion (1.2.2). We will see that this may cause problems. The mean-variance criterion is very popular, especially in Finance, and at first glance looks quite natural. However, in many situations the choice of such a criterion may contradict common sense.

EXAMPLE 1. Let \( X \) take on values from \([1, \infty)\), and \( P(X > x) = 1/x^\alpha \) for all \( x \geq 1 \) and some \( \alpha > 2 \). This is a version of the Pareto distribution we discuss in Section 2.1.1. The Pareto distribution, in different versions, is quite “popular” and is frequently used in many applications including actuarial modeling. It is not difficult to compute that

\[
m_X = \frac{\alpha}{\alpha - 1}, \quad \text{and} \quad \sigma_X^2 = \frac{\alpha}{(\alpha - 2)(\alpha - 1)^2}.
\]

The reader can do it right away or wait until Section 2.1.1.

Let \( Y \) be uniformly distributed on \([0, 1]\). Obviously, \( X \geq Y \) with probability one.

In accordance with (1.2.9),

\[
V(X) = \tau \cdot \frac{\alpha}{\alpha - 1} - \frac{\alpha}{(\alpha - 2)(\alpha - 1)^2},
\]

\[
V(Y) = \tau \cdot \frac{1}{2} - \frac{1}{12}
\]

(for the variance of the uniform distribution, see Section 0.3.2.1).

We see from (1.2.10) that, whatever \( \tau \) is, if \( \alpha \) approaches 2, the function \( V(X) \) converges to \(-\infty\). Consequently, for any \( \tau \), we can choose \( \alpha \) (and hence a r.v. \( X \)) such that \( V(X) < V(Y) \).

Thus, under the mean-variance criterion, \( X \) is worse than \( Y \), and consequently the criterion (1.2.9) is not monotone. \( \square \)

The linearity of the function in the r.-h.s. of (1.2.9) is not an essential circumstance, neither is the choice of particular r.v.'s \( X \) and \( Y \) in the last example. One may observe the same phenomenon for \( V(X) \) equal to almost any function \( g(m_X, \sigma_X^2) \).

To avoid cumbersome formulations, assume that \( g(x, y) \) is smooth. Since we want the mean to be large and the variance to be small, it is natural to assume that the partial derivatives

\[
g_1(x, y) = \frac{\partial}{\partial x} g(x, y) > 0, \quad g_2(x, y) = \frac{\partial}{\partial y} g(x, y) < 0.
\]

To make the proposition below simpler, we assume these functions are continuous.

**Proposition 1** For any r.v. \( Y \), there exists a r.v. \( X \) such that \( X \geq Y \) with probability one, while

\[
g(m_X, \sigma_X^2) < g(m_Y, \sigma_Y^2).
\]
\section*{Some Particular Criteria}

\textbf{Proof} uses the Taylor expansion for functions of two variables. The reader who is not familiar with it may skip this proof without an essential loss in understanding.

Let $E\{Y\} = m$, $Var\{Y\} = \sigma^2$, and a number $\varepsilon \in (0, 1)$. Set $X = Y + \varepsilon^2 Z_\varepsilon$, where the r.v. $Z_\varepsilon$ is independent of $Y$, and

$$Z_\varepsilon = \begin{cases} \varepsilon^{-3} \text{ with probability } \varepsilon^3, \\ 0 \text{ with probability } 1 - \varepsilon^3. \end{cases}$$

Obviously, $X \geq Y$ with probability one. We have $E\{Z_\varepsilon\} = 1$, $Var\{Z_\varepsilon\} = \varepsilon^{-3} - 1$. Hence, $E\{X\} = m + \varepsilon^2 \cdot 1 = m + \varepsilon^2$, and

$$Var\{X\} = \sigma^2 + \varepsilon^4(\varepsilon^{-3} - 1) = \sigma^2 + \varepsilon - \varepsilon^4.$$ 

Next, we use the small ‘o’ notation. The reader unfamiliar with it is highly recommended to look at the explanation of this simple and very convenient notation in Section 0.7.1. In this book, we use this notation repeatedly.

By the Taylor expansion for $g(x, y)$,

$$g(m_X, \sigma_X^2) = g(m + \varepsilon^2, \sigma^2 + \varepsilon - \varepsilon^4) = g(m, \sigma^2) + g_1(m, \sigma^2)\varepsilon^2 + g_2(m, \sigma^2)(\varepsilon - \varepsilon^4) + o(\varepsilon^2 + \varepsilon - \varepsilon^4) = g(m, \sigma^2) + g_2(m, \sigma^2)\varepsilon + o(\varepsilon).$$

Because $g_2(m, \sigma^2) < 0$ and the remainder $o(\varepsilon)$ is negligible for small $\varepsilon$, there exists $\varepsilon > 0$ such that $g_2(m, \sigma^2)\varepsilon + o(\varepsilon) < 0$. Hence, for such an $\varepsilon$

$$g(m_X, \sigma_X^2) < g(m, \sigma^2) = g(m_Y, \sigma_Y^2).$$

\textbf{\large ▲}

Proposition 1 is a strong argument against using variance as a measure of risk. However, if we restrict ourselves to a sufficiently narrow class of r.v.’s, the monotonicity property may hold.

In particular, this is true if we consider only normal r.v.’s because there are no two normal r.v.’s, $X$ and $Y$, with different variances and such that $X \geq Y$ with probability one. If we graph two normal densities with different variances, this fact will become understandable, at least, at a heuristic level. To show it rigorously, one may proceed as follows.

Assume that the normal r.v.’s $X$ and $Y$ mentioned exist. We have $P(X \leq x) = \Phi \left( \frac{x - m_X}{\sigma_X} \right)$, $P(Y \leq x) = \Phi \left( \frac{x - m_Y}{\sigma_Y} \right)$; see (0.3.2.17). Since $P(X \geq Y) = 1$, it is true that $P(X \leq x) \leq P(Y \leq x)$, and hence $\Phi \left( \frac{x - m_X}{\sigma_X} \right) \leq \Phi \left( \frac{x - m_Y}{\sigma_Y} \right)$. The function $\Phi(x)$ is strictly increasing, therefore it follows from the last inequality that $\frac{x - m_X}{\sigma_X} \leq \frac{x - m_Y}{\sigma_Y}$ for all $x$. This is certainly not true if $\sigma_X \neq \sigma_Y$, because two lines with different slopes intersect.
1. COMPARISON OF RANDOM VARIABLES

(Note that we have used above only one property of \( \Phi(x) \): that this function is strictly increasing. Consequently, we can construct similar examples of other classes of r.v.’s with the same property.)

Thus, there are no normal \( X \) and \( Y \) with different variances and such that \( P(X \geq Y) = 1 \). On the other hand, if \( \sigma_X = \sigma_Y \), the comparison is trivial, and for monotonicity to be true, it suffices to require \( g(x, y) \) above to increase in \( x \).

The case of normal r.v.’s is simple because the normal distribution is characterized only by two parameters: mean \( m \) and variance \( \sigma^2 \). Each normal distribution may be identified with a point \((m, \sigma)\) in a plane, and the rule of comparison will be equivalent to a rule of comparison of points in this plane.

If we consider a family of distributions with three or more parameters but still compare these distributions basing on their means and variances, we may come to paradoxes similar to what we saw above.

One well known example concerns the family of r.v.’s \( c + X_{a\nu} \), where \( c \) is a shift parameter and \( X_{a\nu} \) has the \( \Gamma \)-distribution with parameters \( a, \nu \) (see Section 0.3.2.3 and Fig.0.14 there). The distribution of \( c + X_{a\nu} \) is a \( \Gamma \)-distribution with a shift; it is widely used in many areas including insurance as we will see in this book repeatedly. The distribution is asymmetric, so it may be only very roughly characterized by its mean and variance. One may build an example of violation of the monotonicity property in the class of the r.v.’s defined above. The first such example was suggested by K. Borch [13].

1.3 On coherent measures of risk

In this section, we discuss some desirable properties of risk measures. It is important to emphasize, however, that we should not expect these properties to hold in all situations, especially simultaneously. These properties themselves have long been known, but recently they attracted a great deal of attention primarily due to the paper [6] which had given greater insight into the nature of some useful criteria. See also a further discussion in [7] and the monograph [30], and “an exposition for the lay actuary” with some examples in [87].

We describe properties below in terms of \( V(X) \) preserving \( \succeq \).

I. Subadditivity. For all \( X,Y \in \mathcal{X} \),

\[
V(X+Y) \geq V(X) + V(Y).
\] (1.3.1)

This requirement concerns the diversification of portfolios. Let us view \( X \) and \( Y \) as the random results of the investments into two assets, and \( V(X) \) and \( V(Y) \) as the values of the corresponding investments. Then the left member of (1.3.1) is the value of the portfolio consisting of the two investments mentioned, while the right member is the sum of the values of \( X \) and \( Y \), considered separately.

Note also that, if (1.3.1) is true for two r.v.’s, it is true for any number of r.v.’s.
Thus, under a preference with this property, it is reasonable to have many risks in one portfolio (when risks may, in a sense, compensate each other) rather than to deal with these risks separately.

II. Positive Homogeneity. For any \( \lambda \geq 0 \) and \( X \in \mathcal{X} \),
\[
V(\lambda X) = \lambda V(X).
\]

III. Translation Invariance. For any number \( c \) and \( X \in \mathcal{X} \),
\[
V(X + c) = V(X) + c.
\]

Properties II-III establish invariance with respect to the change of scale. For example, if we decide to measure income not in dollars but in cents, under the requirement II, the value of investment (if this value is measured in money units) should be multiplied by 100. If we add to a random income a certain amount \( c \), in accordance with III, the value of the income should increase by \( c \).

Note at once that the value of investment may be measured not only in money units. This is the case, for example, when we apply the utility theory which we discuss in detail in Section 3. Properties II-III are not so innocent as they might seem, and many criteria we consider later, do not satisfy them. However, if (II-III) hold, it certainly “makes life better”.

Since in this setup, in general, \( V(X) \) is not connected with some particular probability measure, we call \( V(X) \) monotone if \( V(X) \geq V(Y) \) when \( X(\omega) \geq Y(\omega) \) for all \( \omega \).

Criteria satisfying I-III together with the monotonicity property are called coherent.

Because the mean-variance criterion is monotone only in special situations, consider, as examples, the first three criteria from Section 1.2.

The mean-value function \( V(X) = E\{X\} \) satisfies all three criteria above, as is easy to see. (For example, \( E\{X + Y\} = E\{X\} + E\{Y\} \), and similarly one can check the other properties.) This is true not because this criterion is very good, but because it is very simple.

The VaR and TailVaR satisfy II-III. Assume, for simplicity, that \( X \) is a continuous r.v. Then \( P(X \leq q_\gamma(X)) = \gamma \). To show that \( q_\gamma(\lambda X) = \lambda q_\gamma(X) \), we should prove that \( P(\lambda X \leq \lambda q_\gamma(X)) = \gamma \). But it is obvious since \( \lambda \) cancels out.

To prove that, for example, \( V_{\text{tail}}(\lambda X) = \lambda V_{\text{tail}}(X) \), it suffices to write \( V_{\text{tail}}(\lambda X) = E\{\lambda X \mid \lambda X \leq q_\gamma(\lambda X)\} = \lambda E\{X \mid \lambda X \leq q_\gamma(X)\} = \lambda V_{\text{tail}}(\lambda X) \).

Property III is considered similarly.

It remains to check the main (and most sophisticated) property I. The VaR does not satisfy this property in general as is shown in

EXAMPLE 1. Let us revisit Example 1.2.2-5. Note that if Property I holds for two r.v.’s, then it holds for any number of r.v.’s. In Example 1.2.2-5, we computed that \( q_\gamma(X_1 + \ldots + X_{10}) = 9.36 \) if \( \gamma = 0.05 \). Since \( P(X_1 = 0) = 0.01 \), for the same \( \gamma \), the quantile \( q_\gamma(X_1) = 1.04 \). Then \( q_\gamma(X_1 + \ldots + q_\gamma(X_n) = 10 \cdot 1.04 = 10.4 > 9.36 \), and hence Property I does not hold. □

In general, the TailVaR criterion does not satisfy Property I either, and, as we know, it is not even monotone. Nevertheless, in Example 2c below, we consider some conditions under which both properties hold. As was mentioned in Section 1.2.4, in particular, it concerns the case where the space \( \Omega \) is finite and all \( \omega \)’s are equally likely.
Since in the scheme of equiprobable $\omega$’s, the r.v.’s themselves may assume various values, the requirement that all $\omega$’s are equally likely is not very strong, and the TailVaR criterion may prove to be efficient in many situations. Nevertheless, it is worthwhile to make the following two remarks.

The goal of the TailVaR criterion is to exclude, as far as it is possible, strategies which could lead to large losses, but it does not take into account possibilities of other values of income, large or moderate. One may say it is a pessimistic criterion.

Secondly, the TailVaR criterion and other criteria we considered are normative, that is, invented by people. These criteria are applied consciously by companies for explicitly stated goals and in explicitly designated situations. When we deal with separate people, the picture may be different. Real individuals are not always pessimistic, often make decisions at an intuitive level, and sometimes are quite sophisticated. To describe their behavior, we should proceed from qualitatively different principles. An introduction to the corresponding theory is given in Sections 3-4.

Next, we give an implicit representation of the whole class of criteria satisfying Properties I-III together with monotonicity. Consider r.v.’s $X = X(\omega)$ defined on a sample space $\Omega$. Denote by $E_P\{X\}$ the expected value of $X$ with respect to a probability measure $P$ defined on sets from $\Omega$.

It was shown in [6] that functions $V(X)$ satisfying all properties mentioned are functions which may be represented as

$$V(X) = \min_{P \in \mathcal{P}} E_P\{X\},$$

(1.3.2)

where $\mathcal{P} = \{P\}$ is a family of probability measures $P$ on $\Omega$. In other words, each function $V(\cdot)$ corresponds to a family $\mathcal{P}$, and vice versa.

For the reader familiar with the notion of infimum, note that in general the minimum above may be not attainable, and more rigorously, a necessary and sufficient condition is the existence of a family $\mathcal{P}$ such that $V(X) = \inf_{P \in \mathcal{P}} E_P\{X\}$.

EXAMPLE 2. (a) Let $P_0$ be the probability measure representing the “actual” probabilities of the occurrence of events $\omega$, and let $\mathcal{P}$ consist of only one measure $P_0$. Then (1.3.2) implies that $V(X) = E_{P_0}\{X\}$, and we deal with the mean-value criterion.

(b) Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be finite, and let $\mathcal{P}$ consist of all probability measures. Denote by $P(\omega)$ the probability of $\omega$ corresponding to measure $P$. The expected value with respect to $P$ is

$$E_P\{X\} = \sum_{i=1}^n X(\omega_i)P(\omega_i).$$

To minimize the last expression, we should choose $P$ which assigns the probability one to the minimum value of $X(\omega)$. For such a measure, $E_P\{X\} = \min_{\omega} X(\omega)$. So,

$$V(X) = \min_{\omega} X(\omega).$$

(c) Let again $\Omega = \{\omega_1, \ldots, \omega_n\}$. Assume that the “actual” probability measure $P_0$ assigns the equal probabilities $\frac{1}{n}$ to each $\omega$. We show that the TailVaR criterion admits the representation (1.3.2).
To make our reasoning simpler, consider only r.v.’s $X(\omega)$ taking different values for different $\omega$’s.

We fix a $\gamma$ and denote by $k = k(\gamma)$ the integer such that $\frac{k-1}{n} \leq \gamma < \frac{k}{n}$; that is, $k = |\gamma n| + 1$, where $[a]$ denotes the integer part of $a$. Consider all sets $A$ from $\Omega$ containing exactly $k$ points. Let $P_A(\omega)$ be the measure assigning the probability $\frac{1}{k}$ to each point from $A$, and zero probability to all other $n - k$ points. Let $\mathcal{P}$ consist of all such measures $P_A$.

Consider now a r.v. $X(\omega)$ and set $x_i = X(\omega_i)$. We assumed that $x_i$’s are different. Without loss of generality, we can suppose that $x_1 < x_2 < \ldots < x_n$, since otherwise we can renumerate the $\omega$’s. The reader is invited to verify that with respect to the original measure $P_0$, first, $q_\gamma(X) = x_k$, where $k = k(\gamma)$ chosen above, and second, that

$$\text{TailVar}(X) = E \{ X \mid X \leq q_\gamma \} = \frac{1}{k} (x_1 + \ldots + x_k). \quad (1.3.3)$$

On the other hand, for any $A$ consisting of $k$ points, say, points $\omega_{i_1}, \ldots, \omega_{i_k}$, $\frac{1}{k} (x_{i_1} + \ldots + x_{i_k}) = \frac{1}{k} (x_1 + \ldots + x_k) \geq E \{ X \mid X \leq q_\gamma \}$, because $x_1, \ldots, x_k$ are the $k$ least values of $X$. Thus, the minimum in (1.3.2) is attained at $P_{A_0}$, where $A_0 = \{ \omega_{i_1}, \ldots, \omega_{i_k} \}$, and this minimum is equal to $\text{TailVar}(X)$.

Note that if $x_i$’s are not different, formally we cannot reason as above since in this case (1.3.3) may be not true, and we may construct an example close to Example 1.2.4-4. However, in this case, we may modify the TailVar criterion itself defining it as in (1.3.3). In the case of different $x_i$’s, it will coincide with the “usual” TailVar.

In conclusion, note that we should not, certainly, restrict ourselves only to coherent measures. Often, it is reasonable to sacrifice some properties mentioned above in order to deal with more flexible characteristics of distributions. It concerns, in particular, criteria we consider in following sections of this chapter.

---

2 COMPARISON OF R.V.’S AND LIMIT THEOREMS OF PROBABILITY THEORY

In this section, we return to the mean-value and VaR criteria, and look at them from another point of view. We saw that they were not very sophisticated – especially the former, and do not satisfy all desirable properties – at least, the latter. Nevertheless, reasonable decision making rules may occur to be close to the criteria mentioned when the corresponding decision making acts are made repeatedly. To understand this, we should recollect rigorous statements of two limit theorems of Probability Theory.
2.1 A diversion to Probability Theory: two limit theorems

2.1.1 The Law of Large Numbers (LLN)

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) r.v.’s. Let $S_n = X_1 + \ldots + X_n$, and $\bar{X}_n = S_n / n$. Set $m = E\{X_i\}$, provided that it exists. It does not depend on $i$ since $X$’s are identically distributed. The LLN says that, though $X_n$ is random for each particular $n$, this randomness vanishes as $n$ gets larger, and $\bar{X}_n$ approaches $m$. The point here, however, is that since $\bar{X}_n$ is not a sequence of numbers but of random variables, the very notion of convergence should be defined properly.

Below, if for a sequence of r.v.’s $Y_n$ and a r.v. $Y$ we say that $P(Y_n \rightarrow Y)$, this means that $Y_n(\omega) \rightarrow Y(\omega)$ for all $\omega$’s from a set of probability one. In short, $Y_n \rightarrow Y$ with probability one. (See also Section 0.5.)

\textbf{Theorem 2 (The strong LLN)}

(a) Suppose that $E\{|X_i|\}$ is finite. Then

$$P(\bar{X}_n \rightarrow m) = 1. \quad (2.1.1)$$

(b) If for some $c$

$$P(\bar{X}_n \rightarrow c) = 1, \quad (2.1.2)$$

then $E\{|X_i|\}$ is finite, and $c = m$.

Thus, the difference $\bar{X}_n - m$ is vanishing as $n \rightarrow \infty$, which implies that the probability that $|\bar{X}_n - m|$ will be larger than any positive number $\varepsilon$, should be small for large $n$. Rigorously, this can be stated as the following

\textbf{Corollary 3 (The weak LLN)} Suppose that $E\{|X_i|\}$ is finite. Then for any $\varepsilon > 0$,

$$P\left(|\bar{X}_n - m| \geq \varepsilon\right) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \quad (2.1.3)$$

Usually the last assertion is written as

$$\bar{X}_n \xrightarrow{p} m,$$

$\bar{X}_n$ converges to $m$ in probability; see also Section 0.5.

Proofs of both facts may be found in many textbooks on Probability; see, e.g., [113] (with some additional conditions), [24], [35], [110], [120].

2.1.2 The Central Limit Theorem (CLT)

Let $E\{X_i^2\}$ be finite, and $\sigma^2 = \text{Var}\{X_i\}$. Since the $X$’s are i.i.d., we have $E\{S_n\} = mn$ and $\text{Var}\{S_n\} = \sigma^2 n$. Consider the normalized sum

$$S'_n = \frac{S_n - E\{S_n\}}{\sqrt{\text{Var}\{S_n\}}} = \frac{S_n - mn}{\sigma \sqrt{n}}.$$
It is worth emphasizing that the normalized r.v. $S_n^*$ is just the same sum $S_n$ considered in an appropriate scale: after normalization, $E\{S_n^*\} = 0$, and $Var\{S_n^*\} = 1$ (see also Example 0.2.7-1, and Exercise 10).

**Theorem 4** (The CLT) For any $x$,$$
\frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-u^2/2} du, \text{ the standard normal distribution function.}
$$

**Corollary 5** For any $a$ and $b$,

$$
P(a \leq S_n^* \leq b) \to \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-x^2/2} dx \text{ as } n \to \infty.
$$

In spite of its simple formulation, this is a very deep fact. It says that the influence of separate terms in a large sum is diminishing, vanishing, and for large $n$, the distribution of $S_n$ is getting close to a standard distribution independent of the distribution of the $X$’s.

Proofs, with use of different methods, and generalizations of this theorem may be found in many textbooks. Often, to make proofs more transparent, in textbooks some additional unnecessary conditions are imposed. Complete proofs with conditions close to necessary may be found, for example, in [24], [35], [110], [120].

Next, we apply these two theorems to a particular insurance problem.

### 2.2 A simple model of insurance with many clients

Consider an insurance company dealing with $n$ clients. Let $X_i$, $i = 1, \ldots, n$, be the random value of the payment to the $i$th client. We assume the $X$’s to be i.i.d., which may be interpreted as if the clients come from a homogeneous group. We keep all notations from Section 2.1.

Let $c = m + \varepsilon$, where $\varepsilon > 0$, be the premium for each client. Thus, we assume that the premium is, at least a bit, larger than $m$. The total profit of the company equals $nc - S_n$, where $S_n = X_1 + \ldots + X_n$. The probability that the company will not suffer a loss is equal to

$$
P(nc - S_n \geq 0) = P(S_n - mn \leq n\varepsilon) = P(|X_n - m| \leq \varepsilon) \leq P(|X_n - m| \leq \varepsilon)
$$

$$
= 1 - P(|X_n - m| > \varepsilon) \to 1 \text{ as } n \to \infty,
$$

for ANY arbitrarily small $\varepsilon > 0$, by Corollary 3.

Thus, if $n$ is “large”, for the company not to suffer a loss, the premium $c$ should be “just a little bit” larger than $m$. In this case, the company would prefer, with regard to each client, that the profit be $c - X_i$ rather than zero.

We see that for large $n$ the necessary premium $c$ is close to the expected value $m$, and accordingly the criterion of the choice of a premium is close to the mean-value criterion.
1. COMPARISON OF RANDOM VARIABLES

The “little bit” mentioned, that is, the value of $\varepsilon$, is one of the main objects of study in Actuarial Modeling, and we will return to it repeatedly. Here we will make just preliminary observations.

First, note that though $\varepsilon$ can be small, it cannot be zero. Indeed, assume $\sigma > 0$; otherwise $X$’s are not random (that is, assume just one value) and the situation is trivial. Then, if $\varepsilon = 0$,

$$P(nc - S_n \geq 0) = P(mn - S_n \geq 0) = P(S_n - mn \leq 0) = P\left(\frac{S_n - mn}{\sigma/\sqrt{n}} \leq 0\right) = P(S_n^* \leq 0) \to \Phi(0) = \frac{1}{2} \text{ as } n \to \infty,$$

by Theorem 4. So, in this case, the probability that the company will not suffer a loss is close only to $1/2$.

Note also that, since the limiting normal distribution is continuous, the probability for $S_n$ to be exactly equal to some value is close to zero. Therefore, the probability of making a profit and the probability of not suffering a loss asymptotically, for large $n$, are the same.

Let now $\varepsilon > 0$. The same CLT gives the first heuristic approximation for a reasonable value of $\varepsilon$. Assume that the company specifies the lowest acceptable level $\beta$ for the probability not to suffer a loss. For instance, the company wishes the mentioned probability to be not less than $\beta = 0.95$, in the worst case – to be equal to 0.95.

Set $\varepsilon = a\sigma/\sqrt{n}$, where the number $a$ is what we want to estimate. Let $c$ be the least acceptable premium for the company. Then

$$\beta = P(nc - S_n \geq 0) = P(S_n - mn \leq n\varepsilon) = P\left(\frac{S_n - mn}{\sigma/\sqrt{n}} \leq \frac{\varepsilon}{\sigma}\sqrt{n}\right) = P(S_n^* < a).$$

By the CLT, $P(S_n^* < a) \approx \Phi(a)$ for large $n$. Thus, $\beta \approx \Phi(a)$, and $a \approx q_{\beta s}$, the $\beta$-quantile of the standard normal distribution. So, the first, certainly rough, approximation for $c$ is

$$c \approx m + \frac{q_{\beta s}\sigma}{\sqrt{n}}. \quad (2.2.1)$$

For $\beta = 0.95$, we have $q_{0.95} = 1.64...$, and $c \approx m + \frac{1.64\sigma}{\sqrt{n}}$.

EXAMPLE 1. A special insurance pays $b = \$150$ to passengers of an airline in the case of a serious flight delay. Assume that for each of 10,000 clients who bought such an insurance, the probability of a delay is $p = 0.1$. In this case,

$$X_i = \begin{cases} b & \text{with probability } p, \\ 0 & \text{with probability } 1 - p, \end{cases}$$

$$m = bp = 15, \sigma = b\sqrt{p(1 - p)} = 45 \text{ (recall the formulas for the mean and the variance of a binomial r.v.)}.$$ Then for $\beta = 0.95$, by (2.2.1), $c \approx 15 + \frac{1.64\times 45}{100} \approx 15.74$. So, a premium of $16$ would be enough for the company. □
2. Comparison of R.V.’s and Limit Theorems

Note that the choice of \( c \) in (2.2.1) is closely related to the VaR criterion. For each premium \( c \), the company compares its random profit \( nc - S_n \) with the r.v. \( Y \equiv 0 \), the profit in the case when the company does not sell the insurance product. For the \( c \) chosen, up to normal approximation, \( \hat{\beta} = P(nc - S_n > 0) \) and hence \( P(nc - S_n \leq 0) = 1 - \hat{\beta} \). Thus, zero is the \( (1 - \hat{\beta}) \)-quantile for the r.v. \( nc - S_n \). On the other hand, \( Y \) takes on only one value – zero, and this singular value is the \( \gamma \)-quantile for any \( \gamma \), including \( \gamma = 1 - \hat{\beta} \). (See again the definition of quantile in Section 0.1.3 and Fig.7e there.)

Thus, for the least acceptable \( c \) in (2.2.1), \( (1 - \hat{\beta}) \)-quantiles of the r.v.’s \( nc - S_n \) and \( Y \) coincide, that is, \( nc - S_n \) is equivalent to \( Y \) in the sense of the VaR criterion. For \( c \) larger than the value in (2.2.1), \( nc - S_n \) will be better than \( Y = 0 \).

The approach based on limit theorems is, however, far from being universal. First of all, the acts of making decisions are not always repeated a large number of times. We can say so about an insurance company when it deals with a large number of clients, but a separate client may make decisions rarely enough, and the law of large numbers (LLN) in this case may not work well.

Second – and this is also very important – even when limit theorems formally might work, real people in real situations may proceed from preferences not connected with means or variances.

For example, when comparing r.v.’s as in (1.2.1) even repeatedly, people rarely consider such r.v.’s equivalent. Usually, the less risky alternative [as \( Y \) in (1.2.1)] is preferred to the more risky [as \( X \) in (1.2.1)]; see Section 3.4 for more detail. So, the LLN argument does not work here.

The same concerns the CLT. Assume, for instance, that an individual proceeds – perhaps unconsciously – from the same argument based on the CLT, as we used above. Then adding a small amount \( \varepsilon \) of money to \( X \) from (1.2.1) would have made a difference: the r.v. \( X + \varepsilon \) would have been better than \( Y = 50 \). However, usually people – not companies but separate individuals – do not exhibit such behavior.

We consider now an old and celebrated example when the application of the LLN leads to a conclusion that is inconsistent with usual human behavior.

2.3 St. Petersburg’s paradox

The problem below was first investigated by Daniel Bernoulli in his paper [11] published in 1738 when D. Bernoulli worked in Saint Petersburg. Consider a game of chance consisting of tossing a regular coin until a head appears. Suppose that, if the first head appears right away at the first toss, the payment equals 2, say, dollars if we update the problem to the present day. If the first head appears at the second toss, the payment equals 4, and so on; namely, if a head appears at the first time at the \( k \)th toss, the payment equals \( 2^k \). It is easy to see that the payment in this case is a r.v. \( X \) taking values 2, 4, 8, ..., \( 2^k \), ... with probabilities \( \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^k}, \ldots \), respectively, and \( E \{ X \} = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + \ldots = 1 + 1 + 1 + \ldots = \infty \).

By the LLN, this means that, if the game is played repeatedly, the payment equals \( \frac{X_1 + \ldots + X_n}{n} \rightarrow \infty \) as \( n \rightarrow \infty \).
Thus, in the long run, the average payment will be greater than ANY arbitrary large number. Let $C$ be the price for participating in the above game. We see that, if we had proceeded from the LLN and the mean-value criterion, we would have jumped to the conclusion that any, arbitrary large, price $C$ would not have been enough for a person who arrange such a game. Certainly, it does not reflect preferences of real people: most would agree to pay the random payment $X$ above, for example, for the price $C = 10,000$, or even $1000$, if $C$ is paid in advance. (Would the reader agree to arrange such a game for $10,000$?)

One solution to this paradox is based on the fact that in the particular problem we consider,

$$\frac{X_1 + \ldots + X_n}{n \log_2 n} \to 1, \text{ as } n \to \infty, \text{ with probability one.} \quad (2.3.1)$$

We skip a not very difficult but a bit lengthy proof of it; see, e.g., [35], [110].

The last relation may be interpreted as follows. Let $n$ be the total number of games to be played. Assume that the price for participating in a separate game depends on $n$ and is equal to $\log_2 n$. Then the total payment for participating in $n$ games is $n \log_2 n$. Relation (2.3.1) shows that in this case, for large $n$, the total payment for participating in the $n$ games is close to the total gain of the player. That is, the price is “fair”. This is a rather mathematical approach and is strongly connected with the problem under consideration.

Fortunately, D. Bernoulli did not know this fact and suggested another and general approach that proved to be very useful for comparison of risky alternatives. We consider it in the next section.

### 3 EXPECTED UTILITY

#### 3.1 Expected utility maximization (EUM)

##### 3.1.1 Utility function

D. Bernoulli proceeded from the simple observation that the “degree of satisfaction” of having capital, or in other words, the “utility of capital”, depends on the particular amount of capital in a nonlinear way. For example, if we give $1000$ to a person with a wealth of $1,000,000$, and the same $1000$ – to a person with zero capital, the former will feel much less satisfied than the latter.

To model this phenomenon, D. Bernoulli assumed that the satisfaction of possessing a capital $x$, or the “utility” of $x$, may be measured by a function $u(x)$ that, as a rule, is not linear. Such a function is called a *utility function*, or a utility of money function. The word “satisfaction” would possibly reflect the significance of the definition better, but the term “utility” has been already accepted.

The utility function, if it exists, can be viewed as a characteristic of the individual, as if the individual is endowed by this function; so to speak, it is “built into the mind”. To some extent, we can talk about the utility function of a company too. In this case, it reflects the preferences of the company.
D. Bernoulli himself suggested as a good candidate for the “natural” utility function $u(x) = \ln x$, assuming that the increment of the utility is proportional not to the absolute but to the relative growth of the capital. More specifically, if capital $x$ is increased by a small $dx$, then the increment of the utility, $du(x)$, is proportional to $dx/x$, that is,

$$
du = k\frac{dx}{x} \tag{3.1.1}
$$

for a constant $k$. The solution to this equation is $u(x) = k \ln x + C$, where $C$ is another constant. We will see soon that the values of $k$ and $C$ do not matter.

Consider now a random income $X$. In this case, the utility of the income is the r.v. $u(X)$. Bernoulli’s suggestion was to proceed from the expected utility $E\{u(X)\}$.

**EXAMPLE 1.** Assume that the utility function of the player in St. Petersburg’s paradox is $u(x) = \ln x$. Then the expected utility

$$
E\{u(X)\} = \sum_{k=1}^{\infty} u(2^k)2^{-k} = \sum_{k=1}^{\infty} \ln(2^k)2^{-k} = (\ln 2) \sum_{k=1}^{\infty} k2^{-k} = 2\ln 2,
$$

and, unlike $E\{X\}$, the expected utility is finite. (To realize that $\sum_{k=1}^{\infty} k2^{-k} = 2$, one may compute it directly, or observe that this is the expected value of the geometric r.v. with the parameter $p = 1/2$. See Section 0.3.1.3.) □

Next, we consider the general case. Clearly, we can restrict ourselves to non-decreasing utility functions, which reflects the rule “the larger, the better or at least not worse”.

### 3.1.2 Expected utility maximization (EUM) criterion

By definition, this criterion corresponds to the preference order $\succeq$ for which

$$
X \succeq Y \iff E\{u(X)\} \geq E\{u(Y)\} \tag{3.1.2}
$$

for a utility function $u$. Not stating it each time explicitly, we will always assume that $u(x)$ is defined on an interval (which may be the whole real line).

The relation (3.1.2) means that among two r.v.’s, we prefer the r.v. with the larger expected utility. In particular,

if $E\{u(X)\} = E\{u(Y)\}$, we say that $X \simeq Y$,

$X$ is equivalent to $Y$.

If $u(x)$ is non-decreasing (as we agreed), the rule (3.1.2) is monotone. (If $X \geq Y$ with probability one, then $u(X) \geq u(Y)$ with probability one too, which immediately implies that $E\{u(X)\} \geq E\{u(Y)\}$.) In Exercise 11, we discuss strict monotonicity.

The investor who follows (3.1.2) is called an expected utility maximizer (EU maximizer; we will use also the same abbreviation EUM when it does not cause misunderstanding).

It is worth emphasizing that when we are talking about an EU maximizer, we mean that the person’s preferences *may be described by* (3.1.2), or in other words that the person behaves as if she/he were an EU maximizer. However, this does not imply in any way that calculations in (3.1.2) are really going in the mind. A good image illustrating this was
suggested in [78]. A thrown ball exhibits a trajectory described as the solution to a certain equation, but no one thinks that the ball “has and itself solves” this equation. People do not get confused about the ball but they sure do about models of other people.

The first property of EUM criterion. The preference order (3.1.2) does not change if $u(x)$ is replaced by any function $u^\star(x) = bu(x) + a$, where $b$ is a positive and $a$ is an arbitrary number.

Indeed, if we replace in (3.1.2) $u$ by $u^\star$, then $b$ and $a$ will cancel.

Thus, $u$ may be defined up to a linear transformation, and the scale in which we measure utility may be chosen at our convenience. In particular, there is nothing wrong or strange if $u$ assumes negative values.

EXAMPLE 1. Consider (3.1.1) with $u(x) = k\ln x + C$ as above. We see now that constants $k$ and $C$ indeed do not matter, and we can restrict ourselves to $u(x) = \ln x$. □

EXAMPLE 2. Let $u(x) = -\frac{1}{1+x}, x \geq 0$; see Fig.1. Should the fact that $u(x)$ is negative for all $x$’s make us uncomfortable? Not at all. Consider $u^\star(x) = u(x) + 1 = \frac{x}{1+x}$. The new function is positive but reflects the same preference order. The sign of $u(x)$ does not matter; what matters when we compare $X$ and $Y$ is whether $E\{u(X)\}$ is larger than $E\{u(Y)\}$ or not. □

Consider an example of the comparison of r.v.’s.

EXAMPLE 3. (a) A reckless gambler. Let a gambler’s utility function $u(x) = e^x$. Negative $x$’s correspond to losses, and positive – to gains. The values of $u(x)$ for large negative $x$’s practically do not differ, while in the region of positive $x$’s the function $u(x)$ grows very fast; see Fig.2. We may interpret it as if the gambler is not concerned about possible losses and is highly enthusiastic about large gains. Note that the function $e^x$ is convex, and as we will see later, in Section 3.4, the convexity of the utility function corresponds to the inclination to risk.

Consider a game in which the gambler wins $a$ dollars with a probability of $p$, and loses the same amount $a$ with the probability $q = 1 - p$. So, we deal with $X = \pm a$ with the mentioned probabilities. Assume $p < q$.

In our case, $E\{u(X)\} = e^ap + e^{-a}q$. The gambler will participate in such a game if $X$ is better than a r.v. $Y \equiv 0$, which amounts to $E\{u(X)\} > u(0) = 1$. This is equivalent to $e^ap + e^{-a}q > 1$.

If we set $e^a = y$, the last inequality may be reduced to the quadratic inequality $py^2 - y + q > 0$. One root of the corresponding quadratic equality is one, the other is $q/p$. Since $y \geq 1$, the solution is $y > q/p$, and consequently, $a > \ln(q/p)$. Thus, the gambler is inclined to bet large stakes, and will participate in the game only if $a > \ln(q/p)$. For instance, if
3. Expected Utility

\[ u(x) \]

\[ g(a) \]

\[ g(a) \]

FIGURE 3.

(b) A cautious gambler. Consider now a gambler who views the loss of a unit of money as a disaster. What this unit of money is equal to, $1,000,000 or just $100, depends on the gambler. On the other hand, the gambler does not mind taking some risk and participating in a game with a moderate stake. The utility function of such a gambler may look as in Fig.3a: \( u(x) \to -\infty \) as \( x \to -1 \), and \( u(x) \) is growing as a convex function for positive \( x \)'s.

For instance, the function

\[
u(x) = \begin{cases} kx^2 & \text{for } x \geq 0 \\ \ln(1-x^2) & \text{for } -1 < x < 0 \end{cases}
\]

has a similar graph. We wrote \( x^2 \) in \( \ln(1-x^2) \) to make the function smooth at zero. The parameter \( k \) indicates the gambler’s inclination to risk. The larger \( k \), the steeper \( u(x) \) for positive \( x \)'s.

Consider the same r.v. \( X \) as in the example above. Then \( E\{u(X)\} = p \cdot ka^2 + q \cdot \ln(1-a^2) \).

Denote the r.-h.s. by \( g(a) \). The gambler will participate in the game if \( E\{u(X)\} > u(0) = 0 \), which amounts to \( g(a) > 0 \). The reader can readily verify that, if \( k \leq q/p \), the graph of \( g(a) \) looks as in Fig.3b, and \( g(a) \) does not assume positive values. In this case, the person we consider will refuse to play.

The graph for \( k > q/p \) is sketched in Fig.3c. In this case, \( g(a) \) is positive in a neighborhood of zero. The maximum of \( g(a) \) is attained at \( a_0 \) such that \( a_0^2 = 1 - \frac{q}{kp} \). For example, for \( p = \frac{1}{4} \), and \( k = 4 \), we get \( a_0 = \frac{1}{2} \), so the gambler’s optimal behavior is to bet half of the unit of money. □

The next notion we introduce is a certainty equivalent. First, note that any number \( c \) may be viewed as a r.v. taking only one value \( c \). Consider a preference order \( \succeq \) not necessarily connected with expected utility maximization. Assume that for a r.v. \( X \), we can find a number \( c = c(X) \) such that \( c \succeq X \) with respect to the order \( \succeq \). That is, \( c \) is equivalent to \( X \), or the decision maker is indifferent whether to choose \( c \) or \( X \). It may be said that the decision maker considers \( c \) an “adequate price of \( X \)”. 

\( p = \frac{1}{4} \), the lowest acceptable stake for the gambler is \( \ln 3 \approx 1.1 \).
1. COMPARISON OF RANDOM VARIABLES

The number \( c(X) \) so defined is called a certainty equivalent of \( X \).

Now let us consider an EU maximizer with a utility function \( u \). For such a person, in accordance with (3.1.2), the relation \( c \simeq X \) is equivalent to \( E \{ u(X) \} = E \{ u(c) \} \), and since \( c \) is not random, \( E \{ u(X) \} = u(c) \). If \( u \) is a one-to-one function, there exists the inverse function \( u^{-1}(y) \), and

\[
c(X) = u^{-1}(E \{ u(X) \}).
\]

EXAMPLE 4. (a) In the situation of Example 3a, \( u^{-1}(x) = \ln x \). So, the certainty equivalent \( c(X) = \ln(e^p + e^{-q}) \). For example, for \( p = \frac{1}{4} \) and \( a = 10 \), we would have \( c(X) = \ln(\frac{1}{4}e^{10} + \frac{3}{4}e^{-10}) \approx 8.614 \), which is close to 10. It is not surprising: the gambler does not care much about losses.

(b) Consider Example 3b for \( p = \frac{1}{4}, k = 4 \). Here the situation is quite different. The gambler bets \( a = a_0 = \frac{1}{4} \). In this case, \( E \{ u(X) \} = g \left( \frac{1}{4} \right) = \frac{1}{4} \cdot 4 \cdot \frac{1}{4} + \frac{3}{4} \cdot \ln \frac{3}{4} \approx 0.034 \). On the other hand, \( u^{-1}(y) = \sqrt{y/k} \) for positive \( y \)’s. So, in our case the certainty equivalent \( c(X) \approx \sqrt{0.034/4} \approx 0.0922 \). □

Note that the certainty equivalent of a certain number \( a \) is, of course, this number: \( c(a) = u^{-1}(E \{ u(a) \}) = u^{-1}(u(a)) = a \).

### 3.1.3 Some “classical” examples of utility functions

1. **Positive-power functions.** Let \( u(x) = x^\alpha \) for all \( x \geq 0 \) and some \( \alpha > 0 \); see Fig.4a.

   The expected utility in this case is considered only for positive r.v.’s, and \( E \{ u(X) \} = E \{ X^\alpha \} \), the moment of \( X \) of the order \( \alpha \). If \( \alpha = 1 \), then \( E \{ u(X) \} = E \{ X \} \), and the EUM criterion coincides with the mean-value criterion. For \( \alpha < 1 \) the function \( u(x) \) is concave (downward), for \( \alpha > 1 \) - convex (concave upward). We will see soon that this is strongly connected with the attitude of the investor to risk. For \( u(x) \) we are considering, the certainty equivalent of a r.v. \( X \) is \( c(X) = (E \{ X^\alpha \})^{1/\alpha} \). In the simplest case \( \alpha = 1 \), the certainty equivalent \( c(X) = E \{ X \} \).

   **EXAMPLE 1.** Let \( X = b > 0 \) or 0 with equal probabilities. Then \( c(X) = \left( \frac{1}{2} b^\alpha \right)^{1/\alpha} = 2^{-1/\alpha} b \). The smaller \( \alpha \) is, the smaller the certainty equivalent. We will interpret this fact later when we consider the notion of risk aversion. □

   **EXAMPLE 2.** Let \( X \) be uniform on \([0, b] \). Then \( c(X) = \left( \int_0^b x^{\alpha-1} dx \right)^{1/\alpha} = \left( \frac{1}{1+\alpha} b^\alpha \right)^{1/\alpha} = \left( \frac{1}{1+\alpha} \right)^{1/\alpha} b \). Because \((1+\alpha)^{1/\alpha}\) is decreasing in \( \alpha \), again the smaller \( \alpha \), the smaller the certainty equivalent. □

2. **Negative-power functions.** Next, consider \( u(x) = -1/x^\alpha \) for all \( x > 0 \) and some \( \alpha > 0 \); see Fig.4b. We again deal only with positive r.v.’s, and \( E \{ u(X) \} = -E \{ X^{-\alpha} \} \). The fact that \( u(x) \) is negative does not matter, but the fact that \( u(x) \to -\infty \), as \( x \to 0 \), is meaningful: now the investor is much more “afraid” of being ruined than in the
3. Expected Utility

$$u(x) = x^\alpha, \alpha < 1$$

$$u(x) = -x^{-\alpha}, \alpha > 0$$

FIGURE 4. Positive- and negative-power utility functions.

previous case when $u(x) \to 0$ as $x \to 0$. We see also that $u(x) \to 0$ as $x \to +\infty$, which may be interpreted as the saturation effect: the investor does not distinguish much large values of the capital. Compare it with the previous case of the positive power where $u(x) \to +\infty$ as $x \to +\infty$.

Both cases above may be described by the unified formula

$$u_\gamma(x) = \frac{1}{1-\gamma} x^{1-\gamma}, \gamma \neq 1. \quad (3.1.3)$$

In the case $\gamma < 1$, we have a positive power function (by the first property above, the absolute value of the multiplier $\frac{1}{1-\gamma}$ does not matter, only the sign does). For $\gamma > 1$, we deal with a negative power function.

3. The logarithmic utility function, $u(x) = \ln x, x > 0$, is in a sense intermediate between the two cases above and has been already discussed.

4. Quadratic utility functions. Consider $u(x) = 2ax - x^2$, where parameter $a > 0$; the multiplier 2 is written for convenience. Certainly, such a utility function is meaningful only for $x \leq a$ when the function is increasing. Hence, in this case, we consider only r.v.’s $X$ such that $P(X \leq a) = 1$. Negative values of $X$ are interpreted as the case when the investor loses or owes money. We have $E\{u(X)\} = 2aE\{X\} - E\{X^2\} = 2aE\{X\} + (E\{X\})^2 - Var\{X\}$. Thus, the expected utility is a quadratic function of the mean and the variance.

5. Exponential utility functions. Let $u(x) = -e^{-\beta x}$, where parameter $\beta > 0$, and the function is considered for all $x$’s. The graph is depicted in Fig.5. Since $u(x) \to 0$, as $x \to \infty$, faster than any power function, the saturation effect in this case is stronger than in Case 2. The expected utility $E\{u(X)\} = -E\{e^{-\beta X}\} = -M(-\beta)$, where $M(z) = E\{e^{zX}\}$.

The function $M(z)$ – we also use the notation $M_X(z)$ to emphasize that it depends on the choice of the r.v. $X$ – is the moment generating function of $X$. (See a definition and examples in Section 0.4.)
1. COMPARISON OF RANDOM VARIABLES

The exponential utility function.

In Exercise 14, we show that the certainty equivalent

\[ c(X) = -\frac{1}{\beta} \ln(M_X(-\beta)) \]

Consider a negative \( \beta \), setting \( \beta = -a \) for some \( a > 0 \). Then

\[ c(X) = \frac{1}{a} \ln(M_X(a)) = \frac{1}{a} \ln(E\{e^{aX}\}) \]

This is the Masset criterion popular in Economics. When \( X \) is a loss, the same expression appears as the premium for the coverage of \( X \) in accordance with the so-called exponential principle. We consider it later in Section 11.1 and in Exercise 11.6. In particular, we compare there the cases \( \beta > 0 \) and \( \beta < 0 \).

EXAMPLE 3. Let \( X \) be distributed exponentially with parameter \( a \). Then \( M_X(z) = \frac{a}{a - z} \). (See Section 0.4.) Now calculations lead to \( c(X) = \frac{1}{\beta} \ln([a + \beta] - \ln a) \). □

In the case of exponential utility, EU maximization has an important property stated in

**Proposition 6** Let \( u(x) = -e^{-\beta x} \) and, under the EUM criterion with this utility function, \( X \succsim Y \). Then \( w + X \succsim w + Y \) for any number \( w \).

The number \( w \) above may be interpreted as the initial wealth, and \( X \) and \( Y \) – as random incomes corresponding to two investment strategies. Proposition 6 claims that in the exponential utility case, the preference relation between \( X \) and \( Y \) does not depend on the initial wealth.

**Proof** is straightforward. By definition, \( w + X \succsim w + Y \) iff \( E\{u(w+X)\} \geq E\{u(w+Y)\} \). For the particular \( u \) above, \( E\{u(w+X)\} = -E\{e^{-\beta[w+X]}\} = -e^{-\beta w}E\{e^{-\beta X}\} \), and the same is true for \( Y \). So, in the last inequality, the common multiplier \(-e^{-\beta w}\) cancels out, and the validity of the relation \( E\{u(w+X)\} \geq E\{u(w+Y)\} \) does not depend on \( w \). Hence, if this relation is true for \( w = 0 \), it is true for all \( w \). □
3. Expected Utility

3.2 Utility and insurance

Consider an individual with a wealth of \( w \), facing a possible random loss \( \xi \). Assume that the individual is an EU maximizer with a utility function \( u(x) \). What premium \( G \) would the individual be willing to pay to insure the risk?

The individual’s wealth after paying the premium will become \( X = w - G \), while if she/he does not buy the insurance, the wealth will equal the r.v. \( Y = w - \xi \).

Then in accordance with the principle (3.1.2), a premium \( G \) will be acceptable for the person under consideration only if

\[
u(w - G) \geq E\{u(w - \xi)\}.
\]

(3.2.1)

For the maximal accepted premium \( G_{\text{max}} \),

\[
u(w - G_{\text{max}}) = E\{u(w - \xi)\}.
\]

(3.2.2)

\( G_{\text{max}} \) exists if, say, \( u \) is continuous and increasing; we skip formalities.)

EXAMPLE 1. Let \( u(x) = 2x - x^2 \), \( w = 1 \), and let \( \xi \) be uniformly distributed on \([0, 1]\).

Because \( w - \xi = 1 - \xi \leq 1 \), we deal only with \( x \)'s for which \( u(x) \) increases. Let \( y = w - G \).

By (3.2.1),

\[
2y - y^2 \geq 2E\{(1 - \xi)\} - E\{(1 - \xi)^2\}.
\]

Observing that \( 1 - \xi \) is also uniformly distributed on \([0, 1]\) (show it!), we have

\[
2y - y^2 \geq 2\frac{1}{2} - \frac{1}{3} = \frac{2}{3}.
\]

We are interested in \( y \leq 1 \). As is easy to verify, for the last inequality to be true, we should have \( y \geq 1 - \frac{1}{\sqrt{3}} \). Hence, any acceptable premium \( G \leq \frac{1}{\sqrt{3}} \), and \( G_{\text{max}} = \frac{1}{\sqrt{3}} \approx 0.57 \).

In the example we consider, the loss is positive with probability one. In Exercise 16, we provide similar calculations for the case when \( \xi = 0 \) with probability 0.9, and \( \xi \) is uniformly distributed on \([0, 1]\) with probability 0.1. This corresponds to the typical situations when the loss equals zero with a large probability. Nevertheless, it is worth noting that situations when the loss is a positive (or practically positive) r.v. are not rare, especially when we deal with an aggregate loss concerning a large group of clients. For example, if a university provides medical insurance for its employees as one insurance contract, the total loss may be considered a positive r.v. The same remark concerns other examples in this section. □

Next, we consider not an insured but an insurer. The latter offers the complete coverage of a loss \( \xi \) for a premium \( H \) which, in general, may be different from \( G \) above. Assume that the insurer is an EU maximizer with a utility function \( u_1(x) \) and a wealth of \( w_1 \). (Actually, it is more natural to interpret \( w_1 \) as an additional reserve kept by the insurer to fulfill its obligations.) Following a similar logic, we obtain that an acceptable premium \( H \) for the insurer must satisfy the inequality

\[
u_1(w_1) \leq E\{u_1(w_1 + H - \xi)\},
\]

(3.2.3)

and hence for the minimal accepted premium \( H_{\text{min}} \)

\[
u_1(w_1) = E\{u_1(w_1 + H_{\text{min}} - \xi)\}.
\]

(3.2.4)
EXAMPLE 2. Let $u_1(x) = x^\alpha$, $w_1 = 1$, and $\xi$ be the same as in Example 1. Taking again into account that $\eta = 1 - \xi$ is uniformly distributed on $[0,1]$, we derive from (3.2.3) that
\[ 1 \leq E\{ (H + 1 - \xi)^\alpha \} = E\{ (H + \eta)^\alpha \} = \int_0^1 (H + x)^\alpha dx = \frac{1}{\alpha + 1} [(H + 1)^{\alpha+1} - H^{\alpha+1}]. \]
Hence, $H_{\min}$ is a solution to the equation
\[ (H + 1)^{\alpha+1} - H^{\alpha+1} = \alpha + 1. \]
For example, when $\alpha = 1/2$, it is easy to calculate – using even a simple calculator – that $H_{\min} \approx 0.52$. □

Clearly, for a premium $P$ to be acceptable for both sides, the insurer and the insured, we should have
\[ H_{\min} \leq P \leq G_{\max}. \]
Hence, if $H_{\min} > G_{\max}$, insurance is impossible. If $H_{\min} \leq G_{\max}$, the premium will be chosen from the interval $[H_{\min}, G_{\max}]$. For instance, in the situation of Examples 1-2, we have $0.52 \leq P \leq 0.57$.

If for example, the insurer has a sort of monopoly in the market, the premium will be close to $G_{\max}$. In the case of competition or if a law imposes restrictions on the size of premiums, we can expect the premium to be closer to $H_{\min}$.

It is worth emphasizing that in general and in the examples above, premiums depend on the initial wealth $w$. Now we consider the special case of exponential utility, when premiums do not depend on wealth.

Let $u(x) = -e^{-\beta x}$. Due to Proposition 6, we can set $w = 0$ in (3.2.1)-(3.2.4). Hence, (3.2.2) is equivalent to $-e^{\beta G_{\max}} = -E\{ e^{\beta \xi} \}$, and $G_{\max} = \frac{1}{\beta} \ln(E\{ e^{\beta \xi} \})$. We see in the r.-h.s. the moment generating function (m.g.f.) $M_{\xi}(z) = E\{ e^{z \xi} \}$. Thus,
\[ G_{\max} = \frac{1}{\beta} \ln M_{\xi}(\beta). \] (3.2.5)
The same formula is true for the insurer: in a similar way, we derive from (3.2.4) that in the case $u_1(x) = -e^{-\beta_1 x}$,
\[ H_{\min} = \frac{1}{\beta_1} \ln M_{\xi}(\beta_1). \] (3.2.6)

It may be proved (see Exercise 32 and an advice there for detail) that the r.-h.s. of (3.2.5) is non-decreasing in $\beta$. Consequently, for $H_{\min} \leq G_{\max}$, we should require $\beta_1 \leq \beta$. This fact will be interpreted when we consider the notion of risk aversion.

EXAMPLE 3. Assume that the random loss $\xi$ may be well approximated by a normal r.v. with mean $m$ and variance $\sigma^2$. In this case, the m.g.f. $M(z) = \exp\{ m z + \sigma^2 z^2 / 2 \}$; see Section 0.4.3. The reader is invited to provide simple calculations leading in this case to
\[ G_{\max} = m + \beta \frac{\sigma^2}{2}, \quad H_{\min} = m + \beta_1 \frac{\sigma^2}{2}. \] (3.2.7)
The answer looks nice and natural: the larger $\beta$ and/or the variance, the more the premium exceeds the expected value of the loss. For $H_{\text{min}} \leq G_{\text{max}}$, we indeed should have $\beta_1 \leq \beta$. □

The same logic can be applied to more complicated forms of insurance. Let, for example, a client be willing to insure only half of a possible loss $\xi$. Then the corresponding equation for the maximal premium will be

$$E\{u(w - \frac{1}{2} - \xi - G_{\text{max}})\} = E\{u(w - \xi)\}. \quad (3.2.8)$$

In conclusion, it is worth noting that the expected utility analysis can work well when we deal rather with the preferences of individual clients. This does not mean that we cannot apply the EUM criterion to the description of the behavior of companies, but one should do it with caution. As we will see in later chapters, the behavior of companies may be determined by principles qualitatively different from those based on expected utility.

### 3.3 How to determine the utility function in particular cases

In Section 3.5.5, we will see that in the EUM case, when one considers r.v.’s taking only $n$ fixed values, to completely determine the preference order, it suffices to specify $n - 1$ equivalent distributions. At least theoretically it may be done by questioning the individual.

Another way is to determine certainty equivalents, which may be illustrated by the following

EXAMPLE 1. We believe that Chris is an EU maximizer, and we try to determine his utility function $u(x)$. In view of the first property from Section 3.1.2, the scale in measuring utility does not matter, so we can set, say, $u(0) = 0$ and $u(100) = 1$, where money is measured in convenient units (for example, not in $1$ but in $100$).

You create a game with prizes $X = 100$ and $0$ each with probability $1/2$. You invite Chris to pay $50$ to play, but he refuses. You reduce the price to $49$, $48$, and so on, up to the moment when Chris starts to hesitate. Assume that it happens at the price $c = 40$. Then we can view $c$ as the certainty equivalent of $X$. This means that $u(c) = E\{u(X)\} = \frac{1}{2}u(100) + \frac{1}{2}u(0) = \frac{1}{2}$. Hence $u(40) = 0.5$, and we know the value of $u(x)$ at one more point. You can continue such a process, for example, figuring out how much Chris values a r.v. $X_1 = 100$ and $40$ with equal probabilities. Assume that Chris’s answer is $60$. Then $u(60) = \frac{1}{2}u(100) + \frac{1}{2}u(40) = \frac{3}{4}$, etc.

Similar questioning may involve insurance premiums. Suppose, for example, that Chris’s initial wealth is 100 units of money. (To make an example meaningful we should certainly assume that the units are substantially larger than $1$.) Assume that, when facing a possible loss of the whole wealth with a probability of 0.1, Chris is willing to pay a premium of at most 25 to insure the loss. In view of (3.2.2), it means that $u(75) = \frac{1}{10}u(0) + \frac{9}{10}u(100) = 9/10$. □

Unfortunately, in real life it works not so well as in nice theoretical examples. The problem is not in mathematical modeling but in making results of such an inquiry reliable, reflecting the real preferences of the individual. This is a psychological rather than mathematical question. The difficulty is that answers depend on the situation, on the form in
1. COMPARISON OF RANDOM VARIABLES

which the questions are asked, whether the questioning involves real money or the experiment is virtual, and on many other psychological and social issues. These problems are beyond the scope of this book on mathematical modeling. For a corresponding discussion see, e.g., [53], [69], [81], [135], [137], and references therein.

3.4 Risk aversion

3.4.1 A definition

Below, by the symbol \( Z_\varepsilon \) we will denote a r.v.

\[
Z_\varepsilon = \begin{cases} 
\varepsilon & \text{with probability } 1/2, \\
-\varepsilon & \text{with probability } 1/2,
\end{cases}
\]

where \( \varepsilon > 0 \). We will talk about the risk aversion of an individual with a preference order \( \succeq \) if the following condition holds.

**Condition Z:** For any r.v. \( X \), any \( \varepsilon > 0 \), and any r.v. \( Z_\varepsilon \) independent of \( X \), it is true that \( X \succeq X + Z_\varepsilon \).

Condition Z reflects the rule “the less stable, the worse”. An investor with preferences satisfying this property would not accept an offer resulting in either an additional income with probability 1/2 or a loss of the same amount and with the same probability.

It is important to emphasize that Condition Z concerns an arbitrary preference order, not only the EUM criterion.

An individual whose preference order satisfies Condition Z is called a risk averter. If \( X \preceq X + Z_\varepsilon \) for any \( X \), any \( \varepsilon > 0 \), and any \( Z_\varepsilon \) independent of \( X \), then we call such an individual a risk lover.

The fact that we consider in Condition Z a non-strict relation \( \succeq \) is not essential. We do it to avoid below some superfluous constructions. Formally, the above definition does not exclude the case when an individual is simultaneously a risk averter and a risk lover, that is, \( X \asymp X + Z_\varepsilon \) for all \( X \) and \( \varepsilon \). In this case, we say that the individual is risk neutral.

Certainly, a person may be neither a risk averter nor a risk lover. For example, it may happen that for some particular \( X \) and \( \varepsilon \), it is true that \( X \succeq X + Z_\varepsilon \), and for another r.v., say, \( X^* \), it may turn out that \( X^* \preceq X^* + Z_\varepsilon \).

Next, we consider the EUM criterion and figure out when this particular criterion satisfies Condition Z.

**Proposition 7** Let \( \succeq \) be a EUM order defined in (3.1.2). Then Condition Z holds iff \( u(x) \) is concave.

We will prove this proposition in the end of this section; now we turn to examples and comments.

Usually we deal with smooth utility functions, so to check whether an EU maximizer with a utility function \( u \) is a risk averter, it suffices to check the second derivative \( u'' \).

For example, for \( u = x^\alpha \), we have \( u''(x) = \alpha(\alpha - 1)x^{\alpha - 2} \). Thus, \( u''(x) < 0 \) for \( \alpha < 1 \), which corresponds to the risk aversion case, while for \( \alpha > 1 \) we deal with a risk lover.
The case $\alpha = 1$ when $E\{u(X)\} = E\{X\}$ may be assigned to both types: the person is risk neutral. Other utility functions are considered in Exercise 25.

Whether a person is a risk averter or a risk lover (or neither) depends, of course, not only on her/his personality but on the particular situation. You may be a risk averter in routine life but if you have decided to spend some time in a casino, you are definitely a risk lover.

There is also strong evidence based on experiments that many people incline to behave as risk averters when concerned with future gains (positive values of $X$), and as risk lovers when facing losses.

For example, a person may choose $500$ for sure rather than $1,000$ with probability $1/2$. However, the same person may prefer to take a risk of losing $1,000$ with probability $1/2$ rather than to lose (only) $500$ for sure. A utility function in this case may look as in Fig.6.

Certainly, the utility function may be more complicated or – better to say – more sophisticated. For example, in the region of moderate $x$’s the function may be concave and in the region of large income values – convex.

The following inequality clarifies why the concavity of utility functions is relevant to risk aversion.

3.4.2 Jensen’s inequality

We assume all expectations below to be finite.

**Proposition 8** Let $X$ be a r.v. (with a finite expectation). Then, if $u(x)$ is concave,

$$E\{u(X)\} \leq u(E\{X\}).$$

(3.4.1)

If $u$ is convex (concave upward),

$$E\{u(X)\} \geq u(E\{X\}).$$

(3.4.2)

The proof is relegated to Section 3.4.4.

Being purely mathematical assertions, inequalities (3.4.1)-(3.4.2) are relevant to the basic question of insurance: why it is possible.

Assume that a client of an insurance organization is a EU maximizer and consider relation (3.2.2). If the client is a risk averter (which is natural to assume since the client is willing to pay to insure the risk), then $u(x)$ is concave, and by Jensen’s inequality,

$$u(w - G_{\text{max}}) = E\{u(w - \xi)\} \leq u(E\{w - \xi\}) = u(w - E\{\xi\}).$$

Since $u$ is non-decreasing, it implies that $w - G_{\text{max}} \leq w - E\{\xi\}$, or

$$G_{\text{max}} \geq E\{\xi\}.$$

Thus, the maximum premium the client agrees to pay is larger than (or, for the boundary case, equals) the average coverage of the risk, $E\{\xi\}$. 
So, the company will get on the average more than it will pay, which means that the company can function.

To the contrary, if the client had been a risk lover, from Jensen’s inequality it would have followed that $G_{\text{max}} \leq E\{\xi\}$, and insurance would have been impossible.

**EXAMPLE 1.** Consider Example 3.2-1. We computed $G_{\text{max}}=\frac{1}{\sqrt{3}}\approx 0.57$, while $E\{\xi\}=0.5$.

The same argument may be applied to the certainty equivalent of a r.v. $X$ (see Section 3.1.1). The inverse of an increasing function is increasing. From this and (3.4.1) it follows that if $u$ is concave, then $c(X) = u^{-1}(E\{u(X)\}) \leq u^{-1}(u(E\{X\})) = E\{X\}$. Thus,

$$c(X) \leq E\{X\}.$$  

In the case of risk aversion, $c(X) \leq E\{X\}$.

For the risk lover a similar argument leads to $c(X) \geq E\{X\}$.

**EXAMPLE 2.** Let $X$ be exponential with parameter $a$, and $u(x) = -e^{-\beta x}$, $\beta > 0$. The function is concave, and the person with such a utility function is a risk averter. Continuing the calculations from Example 3.1.3-3, we have

$$c(X) = \frac{1}{\beta} [\ln(a+\beta) - \ln(a)] = -\frac{1}{\beta} (\ln a)[1 - \frac{\ln(a+\beta)}{\ln a}] = \frac{1}{\beta} (\ln a)[1 - \frac{\ln(a+\beta)}{\ln a}]. \quad (3.4.3)$$

Let $X$ be “large”: formally let $a \to 0$. Then $E\{X\} = \frac{1}{a} \to \infty$. Since the third factor in (3.4.3) converges to one, $c(X) \sim \frac{1}{\beta} \ln\left(\frac{1}{a}\right)$. (The relation $u \sim v$ means $\frac{u}{v} \to 1$.) Thus, in our case

$$c(X) \sim \frac{1}{\beta} \ln(E\{X\}).$$

Since $\ln x$ is much smaller than $x$ for large $x$’s, the certainty equivalent is much smaller than the mean value $E\{X\}$. We interpret this as saying that the individual is a strong risk averter.

**Route 1 ⇒ page 125**

### 3.4.3 How to measure risk aversion in the EUM case

We use below the Calculus notation $o(\varepsilon)$ for a function $o(\varepsilon)$ such that $\frac{o(\varepsilon)}{\varepsilon} \to 0$ as $\varepsilon \to 0$. The reader who is not familiar with this very convenient and simple notation is highly recommended to look at the detailed explanation in Section 0.7.1.

Let $x > 0$ be a fixed capital. Its certainty equivalent is the same number $x$. Let us consider the r.v. $X + Z_\varepsilon$ and compute its certainty equivalent for small $\varepsilon$. 
Lemma 9 Suppose the second derivative \( u'' \) exists and is continuous, and \( u'(x) > 0 \) for \( x \) chosen above. Then the certainty equivalent

\[
c(x + Z_{\epsilon}) = x - \frac{1}{2} R(x) \epsilon^2 + o(\epsilon^2),
\]

where

\[
R(x) = -\frac{u''(x)}{u'(x)}.
\]

The proof is given in Section 3.4.4; now we will discuss the significance of (3.4.4).

By definition of \( o(\epsilon^2) \), we view the third term in (3.4.4) as negligible with respect to the second.

The function \( R(x) \) may be considered a characteristic of the concavity of \( u \). In particular, if \( u \) is concave (risk aversion!), then \( R(x) \geq 0 \).

If \( u \) is concave, by Proposition 8, the r.v. \( x + Z_{\epsilon} \preceq x \), and for the corresponding certainty equivalents we have

\[
c(x + Z_{\epsilon}) \leq x.
\]

The difference \( x - c(x + Z_{\epsilon}) \) may be viewed as a “price for risk”, a “measure of riskiness”. By Lemma 9, this measure is proportional to \( R(x) \) up to the negligible remainder \( o(\epsilon^2) \). The characteristic \( R(x) \) is called an absolute risk aversion function, or the Arrow-Pratt index of risk aversion.

For a r.v. \( X \), the expectation \( E\{R(X)\} \) may be called an expected absolute risk aversion. If \( x \) is measured in dollars, the dimension of \( R(x) \) is dollar\(^{-1}\). We define the relative risk aversion function as

\[
R_r(x) = \frac{|x|R(x)}{x}.
\]

This function does not have dimension. We call \( E\{R_r(X)\} \) an expected relative risk aversion.

EXAMPLE 1. Let \( u(x) = -e^{-\beta x} \). Then, as is easy to calculate by substituting into (3.4.5), the absolute risk aversion function \( R(x) = \beta \) and does not depend on \( x \). In light of this, formulas (3.2.7) from Example 3.2-3 look very nice and understandable: the larger the risk aversion characteristics \( \beta \) and \( \beta_1 \), the more the premiums \( G_{\text{max}} \) and \( H_{\text{min}} \) exceed the expected value \( m \). The differences \( G_{\text{max}} - m \) and \( H_{\text{min}} - m \) are proportional to \( \beta \) and \( \beta_1 \), respectively. The insurance is possible if the insurer is not more risk averse than the insured \((\beta_1 \leq \beta)\). This reflects reality. The insurance company can afford to be less risk averse. It deals with many clients, the number of those is usually essentially larger than the number of claims, and payments are compensated by premiums. The corresponding rigorous model will be considered in Section 2.3.

EXAMPLE 2. Let \( u(x) = x^\alpha, x \geq 0, \alpha > 0 \). Then \( R(x) = (1 - \alpha)/x \) (compute on your own), and the relative aversion \( R_r(x) = 1 - \alpha \). It is non-negative iff \( u \) is concave \((\alpha < 1)\).

For \( u(x) = -x^{-\alpha}, x > 0, \alpha > 0 \), we have \( R(x) = (1 + \alpha)/x \), and the relative aversion \( R_r(x) = 1 + \alpha \) and is positive for all \( \alpha > 0 \) (which is consistent with the fact that \( u \) is concave for all \( \alpha \)).
1. COMPARISON OF RANDOM VARIABLES

In Examples 3.1.3-1 and 2, we considered two types of a particular r.v. $X$. In the first example, $X$ took on two values, $b > 0$ and 0, with equal probabilities; in the second, $X$ was uniform on $[0, b]$. The certainty equivalents proved to be $c(X) = 2^{-1/\alpha}b$ and $c(X) = (1 + \alpha)^{-1/\alpha}b$, respectively. In both cases, the less the risk aversion $\alpha$, the less the certainty equivalent. □

EXAMPLE 3. For $u(x) = \ln(x)$, $x > 0$, we have $R(x) = 1/x$, and $R_r(x) = 1$. □

In Exercise 36, the reader is invited to prove that the cases considered above exhaust all cases with constant risk aversion functions.

EXAMPLE 4. Let $u(x) = x / (1 + |x|)$. The graph is given in Fig.7a. We see that the person with such a utility function is a risk averter in the region of gains and a risk lover in the case of losses (negative x’s). We observe also the saturation effect in both sides for $x \to \pm\infty$. So, we expect that at $\pm\infty$ the absolute aversion should vanish. It is true since, as is easy to compute, $R(x) = 2/(1 + |x|)$ for $x > 0$, and $R(x) = -2/(1 + |x|)$ for $x < 0$. See Fig.7b.

Since $u''(0)$ does not exist, $R(0)$ is not defined. However, we can write that $R_r(x) = |x|R(x) = 2x/(1 + |x|)$ for $x \neq 0$. Then, by continuity, we may set $R_r(0) = 0$. □

3.4.4 Proofs

To make constructions below simpler, we assume additionally that functions $u(x)$ are continuous. Since we deal with concave functions which are continuous on any open interval, it is a very minor assumption.

Proof of Proposition 7. Sufficiency. Note that if $u(x)$ is concave, then setting $\lambda = \frac{1}{2}$ in the definition (0.7.3.1), we get that for any $x_1, x_2$ from the domain of $u(x)$,

$$
\frac{u(x_1) + u(x_2)}{2} \leq u \left( \frac{x_1 + x_2}{2} \right) .
$$

(3.4.6)

Now, let $F_X(x)$ and $F_Z(x)$ be the d.f. of $X$ and $Z_\epsilon$, respectively. Since $X$ and $Z_\epsilon$ are
implies continuity on each open interval. This fact is traced to Jensen’s paper \[62\]; see also, e.g., which we have assumed. This proves necessity.

\[ \text{convex.} \]

As a matter of fact this assumption is not necessary, and we made it for simplicity.\[u(x)\]

The function \(u\) is assumed to be continuous. Note that as a matter of fact this assumption is not necessary, and we made it for simplicity.\[u(x)+\epsilon\]

By virtue of (3.4.6), the r.v. in \{\cdot\} above is not greater than \(u(X)\). Hence,

\[ E\{u(X+Z_\epsilon)\} \leq E\{u(X)\}, \]

and, consequently, \(X+Z_\epsilon \preceq X\).

The reader familiar with the notion of conditional expectation could just condition on \(X\), writing \(E\{u(X+Z_\epsilon)\} = E\{E\{u(X+Z_\epsilon)\} \mid X\} = E\{\frac{1}{2}u(X+\epsilon) + \frac{1}{2}u(X-\epsilon)\}\). Since we are going to consider the notion mentioned later in Chapter 3, we preferred the former proof.

**Necessity.** Let \(x_1, x_2\) be two numbers such that \(x_1 < x_2\). Set \(x_0 = \frac{x_1 + x_2}{2}, \epsilon = \frac{x_2 - x_1}{2}\), and \(X \equiv x_0\). By definition of risk aversion, \(E\{u(X+Z_\epsilon)\} \leq E\{u(X)\}\). For the particular \(X\) above, it means that \(\frac{1}{2}u(x_0+\epsilon) + \frac{1}{2}u(x_0-\epsilon) \leq u(x_0)\). By the choice of \(x_0\) and \(\epsilon\), this implies (3.4.6). The last property is called midconcavity, and formally it does not imply concavity, that is, \(0.7.3.1\) for all \(\lambda \in [0,1]\). However, it does if \(u(x)\) is continuous (see, e.g., \([52]\)), which we have assumed. This proves necessity.

As a matter of fact we do not have to assume continuity. The utility function \(u\) is non-decreasing, and hence it is bounded on any interval where it is defined. It is known that in this case midconcavity implies continuity on each open interval. This fact is traced to Jensen’s paper \([62]\); see also, e.g., \([52]\).\n
**Proof of Proposition 8.** Let \(u\) be concave on an interval \(I\), and let \(X\) take values from \(I\). Set \(m = E\{X\}\). If \(X\) takes only one value corresponding to one of the endpoints of \(I\), Jensen’s inequality becomes an equality, and proof is trivial. Assume that it is not so. Then \(m\) is an interior point of \(I\), and by Proposition 0.9, there exists a number \(c\) such that

\[ u(x) - u(m) \leq c(x-m) \quad \text{for any } x \in I. \quad (3.4.7) \]

(We can include the endpoints since \(u\) is assumed to be continuous. Note that as a matter of fact this assumption is not necessary, and we made it for simplicity.)

Setting \(x = X\), we write \(u(X) - u(m) \leq c(X-m)\). Computing the expectations of both sides, we have \(E\{u(X)\} - u(m) \leq c(m-m) = 0\), which amounts to (3.4.1).

To prove (3.4.2), it suffices to consider the function \(-u(x)\) which is concave if \(u(x)\) is convex.\n
**Proof of Lemma 9.** By Taylor’s expansion, \(u(x+Z_\epsilon) = u(x) + u'(x)Z_\epsilon + \frac{1}{2}u''(x)Z_\epsilon^2 + o(Z_\epsilon^2)\). Note also that \(E\{Z_\epsilon\} = 0, E\{Z_\epsilon^2\} = \epsilon^2\). Hence, \(E\{u(x+Z_\epsilon)\} = u(x) + \frac{1}{2}u''(x)\epsilon^2 + o(\epsilon^2)\).

The function \(u'\) is continuous, and \(u'(x) > 0\). Hence, \(u'(y) > 0\) for \(y\)'s from a neighborhood of \(x\). Then we can consider \(u^{-1}(y)\) and apply the Taylor expansion. Since,
1. COMPARISON OF RANDOM VARIABLES

\((u^{-1}(y))' = 1/u'(u^{-1}(y))\), we have \(u^{-1}(y + s) = u^{-1}(y) + (u^{-1}(y))'s + o(s) = u^{-1}(y) + \frac{1}{u'(u^{-1}(y))}s + o(s)\) for small \(s\). Eventually, for small \(\varepsilon\), the certainty equivalent

\[
c(x + Z_\varepsilon) = u^{-1}(E\{u(x + Z_\varepsilon)\}) = u^{-1}\left(u(x) + \frac{1}{2}u''(x)\varepsilon^2 + o(\varepsilon^2)\right)
\]

\[
= u^{-1}(u(x)) + \frac{1}{u'(u^{-1}(u(x)))}\left(\frac{1}{2}u''(x)\varepsilon^2 + o(\varepsilon^2)\right) + o\left(\frac{1}{2}u''(x)\varepsilon^2 + o(\varepsilon^2)\right)
\]

\[
= x + \frac{1}{2}u''(x)\varepsilon^2 + \frac{1}{u'(x)}o(\varepsilon^2) + o\left(\frac{1}{2}u''(x)\varepsilon^2 + o(\varepsilon^2)\right).
\]

The last two terms are \(o(\varepsilon^2)\). ■

3.5 A new perspective: EUM as a linear criterion

3.5.1 Preferences on distributions

Let us recall the original Probability Theory framework with a sample space \(\Omega = \{\omega\}\) and a probability measure \(P(A)\) on it. Random variables \(X = X(\omega)\) are functions on \(\Omega\).

Let \(F_X(B) = P(X \in B)\), where \(B\) is a set from the real line. As in Section 0.1.3.1, we call the function of sets \(F_X(B)\) the distribution of \(X\). In particular, the d.f. \(F_X(x) = F((−\infty, x])\).

Certainly, if we know \(F_X(B)\) for all \(B\), we know \(F_X(x)\) for all \(x\). The converse assertion is also true: knowing the d.f. \(F(x)\), we can determine \(F_X(B)\) for any set \(B\). (See (0.2.1.6) and explanations on this point in Sections 0.1.3.1, 0.1.3.3.)

We fix a class \(X = \{X\}\) of r.v.’s \(X\) and consider a preference order \(\succeq\) on \(X\).

Suppose that, when comparing r.v.’s, we take into account not the structure of r.v.’s (that is, how they depend on \(\omega\)) but only the information about the possible values of r.v.’s and the corresponding probabilities. In other words, we proceed only from distributions \(F_X\) and compare not r.v.’s themselves but rather their distributions. For example, this is the case in the EUM framework. Indeed, \(E\{u(X)\}\) is completely determined by the distribution \(F_X\), and therefore, when comparing r.v.’s, as a matter of fact, we compare the corresponding distributions.

In the situation we described, instead of the preference order on the class \(X = \{X\}\), it suffices to consider a preference order \(\succeq\) on the set \(\mathcal{F} = \{F_X\}\) of the distributions of the r.v.’s \(X\) from \(X\). This order is determined by the rule

\[
F_X \succeq F_Y \iff X \succeq Y.
\] (3.5.1)

Because any distribution and its d.f. completely determine each other, it does not matter where we define the preference rule \(\succeq\): among distributions or distribution functions. Below, when it does not cause misunderstanding, we use the symbol \(F\) for both. The reader can even understand the word ‘distribution’ as ‘d.f.’

Conversely, if we agreed to compare only distributions, and if we defined somehow a preference order on a class \(\mathcal{F} = \{F\}\) of distributions \(F\), then we have defined, by the same relation (3.5.1), the preference order among all r.v.’s having the distributions from \(\mathcal{F}\).
3. Expected Utility

Next, we specify in terms of distributions the rule “the larger, the better”. In Section 1.1, we introduced the natural monotonicity property: if \( P(X \geq Y) = 1 \), then \( X \succeq Y \). However, if we proceed only from distributions, such a rule does not cover all situations when \( X \) is “obviously better” than \( Y \).

**EXAMPLE 1.** Let \( \Omega \) consist of two outcomes, \( \omega_1 \) and \( \omega_2 \), and \( P(\omega_1) = 1/3 \), \( P(\omega_2) = 2/3 \). Let

\[
X(\omega_1) = 10, \quad X(\omega_2) = 20, \\
Y(\omega_1) = 20, \quad Y(\omega_2) = 10.
\]

For example, we may view \( X \) and \( Y \) as the prices for two stocks and \( \omega_1, \omega_2 \) as two possible states of the future financial market. With a positive probability of 1/3 the value of \( X \) will be less than the value of \( Y \). Nevertheless, because \( P(\omega_1) < P(\omega_2) \), if for us it is not important which \( \omega \) will occur, but merely what income we will have, we will certainly prefer \( X \) to \( Y \). □

The purpose of the next definition is to take such cases into account.

We say that the distribution \( F \) dominates the distribution \( G \) in the sense of the first stochastic dominance (FSD) if

\[
F(x) \leq G(x) \quad \text{for any } x. \tag{3.5.2}
\]

(That is, whatever \( x \) is, the probability to have an income not larger than \( x \), is not larger for the distribution \( F \).)

Certainly, if \( P(X \geq Y) = 1 \), then \( F_X(x) \leq F_Y(x) \) for any \( x \). Indeed, \( F_X(x) = P(X \leq x) \leq P(Y \leq x) = F_Y(x) \). The converse assertion is not true.

**EXAMPLE 2.** Let us revisit Example 1. Let \( F(x) \) and \( G(x) \) be the d.f.’s of \( X \) and \( Y \), respectively. Their graphs are given in Fig.8. We see that (3.5.2) is true, though \( P(X < Y) = 1/3 > 0 \). □

In the case of the comparison of distributions, the rule “the larger, the better” is reflected
A preference order $\succeq$ on a set $\mathcal{F} = \{F\}$ is said to be monotone with respect to the first stochastic dominance (FSD), if $F \succeq G$ for any pair of distributions $F, G \in \mathcal{F}$ with property (3.5.2).

For brevity, in this case we will also say that $\succeq$ satisfies the FSD rule.

A preference order $\succeq$ is said to be strictly monotone with respect to the FSD, if $F \succ G$ (i.e., $F$ is better than $G$), once (3.5.2) is true, and at least for one $x$ the inequality in (3.5.2) is strict.

**EXAMPLE 3.** Consider the situation of Example 2. If a preference order $\succeq$ satisfies the FSD rule, then for the distributions $F_X$ and $F_Y$ from this example, we have $F_X \succeq G_Y$. □

**EXAMPLE 4.** Ann’s future income $X$ has the exponential distribution with a mean of $m > 0$, and Paul’s income $Y$ is uniformly distributed on $[0, 1]$. For what $m$ is Ann’s position better in the sense of the FSD rule?

Let $F(x)$ and $G(x)$ be the respective d.f.’s. We should figure out when (3.5.2) is true. We have $F(x) = 1 - e^{-x/m}$ for $x \geq 0$, and $G(x) = x$ for $x \in [0, 1]$. Furthermore, $F(0) = G(0) = 0$, $F(1) < 1$, and $G(1) = 1$; see Fig.9.

Since $F(x)$ is concave, it may coincide with $G(x)$ at no more than one point besides the origin. The derivative $F'(x) = e^{-x/m}/m$, and $G'(x) = 1$.

Hence, if $m \geq 1$, then $F'(0) \leq G'(0)$, and $F'(x) < F'(0) \leq 1$. From this it follows that $F(x) < G(x)$ for all $x > 0$; see Fig.9ab.

If $m < 1$, then $F'(0) > 1$, and $F(x) > G(x)$ for $x$’s from some interval; see Fig.9c.

Thus, if we follow the natural FSD rule, for $m \geq 1$ we prefer $X$ to $Y$. For $m < 1$, the comparison is not so obvious, and to make a decision, we should determine the preference order in more detail. □
3.5.3 The second stochastic dominance

This section concerns a property reflecting the rule “the riskier, the worse”. We say that \( F \) dominates \( G \) in the sense of the second stochastic dominance (SSD) if

\[
\int_{-\infty}^{t} [F(x) - G(x)]dx \leq 0 \quad \text{for any } t, \tag{3.5.3}
\]

provided that the integrals above are finite. Clearly, if \( F \) dominates \( G \) in the sense of the FSD, that is, (3.5.2) is true, then (3.5.3) is also true. (Since in this case, the integrand in (3.5.3) is non-positive.)

**EXAMPLE 1.** To clarify (3.5.3), let us recall the definition of risk aversion from Section 3.4.1. Let a r.v. \( X \) have a distribution \( F \), and let \( G_{\varepsilon} \) be the distribution of the r.v. \( X_{\varepsilon} = X + Z_{\varepsilon} \), as it was defined in Section 3.4.1. Then

\[
G_{\varepsilon}(x) = P(X_{\varepsilon} \leq x) = P(X_{\varepsilon} \leq x \mid Z_{\varepsilon} = \varepsilon) \frac{1}{2} + P(X_{\varepsilon} \leq x \mid Z_{\varepsilon} = -\varepsilon) \frac{1}{2} = \frac{1}{2} [P(X + \varepsilon \leq x) + P(X - \varepsilon \leq x)] = \frac{1}{2} [F(x - \varepsilon) + F(x + \varepsilon)].
\]

The distribution \( F \) dominates \( G_{\varepsilon} \) in the sense of the SSD. Indeed, set \( Q(t) = \int_{-\infty}^{t} F(x)dx \).

Note that \( \int_{-\infty}^{t} F(x+\varepsilon)dx = \int_{-\infty}^{t+\varepsilon} F(x)dx = Q(t+\varepsilon) \), and similarly \( \int_{-\infty}^{t} F(x-\varepsilon)dx = Q(t-\varepsilon) \).

Then

\[
\int_{-\infty}^{t} [F(x) - G_{\varepsilon}(x)]dx = Q(t) - \frac{1}{2} \left[ Q(t - \varepsilon) + Q(t + \varepsilon) \right].
\]

To show that the last expression is non-positive for any \( t \), it remains to prove that \( Q(t) \) is a convex function. Assume, for simplicity, that \( F(x) \) has the density \( f(x) = F'(x) \). Then \( Q'(t) = F(t), Q''(t) = f(t) \geq 0 \), since \( f \) is a density. As a matter of fact, the smoothness of \( F \) is not necessary. \( \Box \)

A preference order \( \succsim \) on a set \( \mathcal{F} = \{F\} \) is said to be monotone with respect to the SSD, if \( F \succsim G \) for any pair of distributions \( F, G \in \mathcal{F} \) with property (3.5.3).

This is a stronger requirement on the order \( \succsim \) than the monotonicity with respect to the FSD: if \( \succsim \) is monotone with respect to the SSD, then it is monotone with respect to the FSD. (Not vice versa! Let us double check the logic of the implication. Assume that \( \succsim \) is monotone with respect to the SSD. Let \( F \) dominate \( G \) in the sense of the FSD. Then (3.5.3) is true. Then \( F \succsim G \). Hence, \( \succsim \) is monotone with respect to the FSD also.)

We say that an individual is a risk averter in the sense of the SSD, if her/his preference order \( \succsim \) is monotone with respect to the SSD.

From Example 1 it follows that if somebody is a risk averter in the sense of the SSD, then she/he is a risk averter in the sense of Section 3.4.1.
3.5.4 The EUM criterion

We return to the EUM criterion. For a r.v. \( X \), the expected value \( E\{u(X)\} \) may be written as

\[
E\{u(X)\} = \int_{-\infty}^{\infty} u(x)dF(x),
\]  

(3.5.4)

where \( F(x) \) is the distribution function of \( X \).

In more detail, this formula is discussed in Section 0.1.3. In particular, if \( X \) has a probability density function \( f(x) \), then \( dF(x) = f(x)dx \), and the integral above may be understood in the “usual” way as \( \int_{-\infty}^{\infty} u(x)f(x)dx \).

If \( X \) is discrete and assumes values \( x_1, x_2, \ldots \) with probabilities \( f_1, f_2, \ldots \), respectively, then we set \( dF(x_j) = f_j \) for all \( j \), and \( dF(x) = 0 \) for all other \( x \)'s. This will lead to

\[
\int_{-\infty}^{\infty} u(x)dF(x) = \sum_j u(x_j)f_j,
\]

that is, to the definition of expected value in the discrete case.

As was already mentioned, since \( E\{u(X)\} \) is completely determined by the d.f. \( F \) of \( X \), we may restrict ourselves to the corresponding order \( \succeq \) on a set \( \mathcal{F} = \{ F \} \) of distributions.

Let us fix a utility function \( u \) and set

\[
U(F) = \int_{-\infty}^{\infty} u(x)dF(x),
\]  

(3.5.5)

assuming that the last integral is finite for all \( F \in \mathcal{F} \).

We will call \( U(F) \) a utility functional. The word functional is used in Mathematics when the argument in a function (as in \( U(F) \) ) is not a number but an object of a more general nature (as a distribution \( F \)). It is convenient to use this common mathematical term here in order to distinguish the utility functional \( U(F) \) from the utility function \( u(x) \).

The EUM preference order \( \succeq \) in \( \mathcal{F} \) may be defined as

\[
F \succeq G \iff U(F) \geq U(G),
\]

that is, \( \succeq \) is preserved by \( U \).

Certainly, the criterion we have defined is the same criterion as above, just presented in terms of distributions.

Consider the difference \( U(F) - U(G) \). Integrating by parts, we have

\[
U(F) - U(G) = \int_{-\infty}^{\infty} u(x)d[F(x) - G(x)] = \int_{-\infty}^{\infty} [G(x) - F(x)]du(x).
\]  

(3.5.6)

(The differences \( F(\infty) - G(\infty) = 1 - 1 = 0 \), \( F(-\infty) - G(-\infty) = 0 \); the limits \( u(x)[F(x) - G(x)] \) at \( \pm \infty \) equal zero; see, e.g., (0.2.6.5) and the argumentation there.)

Let \( F \) dominate \( G \) in the sense of the FSD, and the utility function \( u(x) \) be non-decreasing. Then \( G(x) - F(x) \geq 0 \), \( du(x) \geq 0 \), and (3.5.6) implies that \( U(F) \geq U(G) \). Thus, we can state the following:

If \( u(x) \) is non-decreasing, then \( U(F) \) is monotone with respect to the FSD.

The above condition is also necessary. Indeed, let \( x_1 > x_2 \), a r.v. \( X_1 \) assume just one value \( x_1 \), and a r.v. \( X_2 \) assume only one value \( x_2 \). Let \( F \) and \( G \) be the d.f.'s of \( X_1 \) and
X_2$, respectively. Then $F$ dominates $G$ in the sense of the FSD (show why). If $U(F)$ is monotone with respect to the FSD, then $U(F) \geq U(G)$. On the other hand, in the EUM case, $U(F) = E\{u(X_1)\} = u(x_1)$, and $U(G) = u(x_2)$. So, $u(x_1) \geq u(x_2)$.

Consider now risk aversion. For simplicity, assume that $u$ is sufficiently smooth. Let us integrate (3.5.6) by parts one time more, writing it in the following way:

$$U(F) - U(G) = \int_{-\infty}^{\infty} [G(x) - F(x)]u'(x)dx = \int_{-\infty}^{\infty} u'(x) \left( \int_{-\infty}^{x} [G(s) - F(s)]ds \right) dx$$

$$= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{x} [F(s) - G(s)]ds \right) u''(x)dx.$$

Suppose that $F$ dominates $G$ in the sense of the SSD and the utility function $u(x)$ is concave. Then $\int_{-\infty}^{x} [F(s) - G(s)]ds \leq 0$, $u''(x) \leq 0$, and hence $U(F) \geq U(G)$. Thus, we have come to the following rule.

If $u(x)$ is concave, then $U(F)$ is monotone with respect to the SSD.

It may be proved that in the case of EUM, the definitions of risk aversion in the sense of Condition Z from Section 3.4.1 and in the sense of the SSD are equivalent. In view of Proposition 7, from this it follows, in particular, that the above concavity condition is also necessary.

3.5.5 Linearity of the utility functional

Let $F_1$ and $F_2$ be two distributions, and let a number $\alpha \in [0,1]$. We call the distribution $F^{(\alpha)}$ a mixture of distributions $F_1, F_2$ if

$$F^{(\alpha)} = \alpha F_1 + (1 - \alpha) F_2,$$

that is, the d.f. $F^{(\alpha)}(x) = \alpha F_1(x) + (1 - \alpha) F_2(x)$, and in general, $F^{(\alpha)}(B) = \alpha F_1(B) + (1 - \alpha) F_2(B)$.

The reader is recommended to look up this notion and comments on it in Section 0.1.3.5. In particular, we mentioned there that the r.v. $X$ having the distribution $F^{(\alpha)}$ may be represented as

$$X = \begin{cases} X_1 & \text{with probability } \alpha, \\ X_2 & \text{with probability } 1 - \alpha, \end{cases}$$

where $X_1, X_2$ have the distributions $F_1, F_2$, respectively. For a particular example, see also Exercise 39.

Now, let us consider the concept of mixture from a geometric point of view.

EXAMPLE 1. Consider r.v.’s taking only three fixed values, $x_1, x_2, x_3$, such that $x_1 < x_2 < x_3$. Any distribution $F$ of such a r.v. may be identified with the probability vector $(p_1, p_2, p_3)$, where $p_i$ is the probability of the value $x_i$. Since $p_1 + p_2 + p_3 = 1$, ...
1. COMPARISON OF RANDOM VARIABLES

The main property of the EUM criterion:

\[ U(\alpha F_1 + (1 - \alpha) F_2) = \alpha U(F_1) + (1 - \alpha) U(F_2). \]  (3.5.8)

That is, the utility of the mixture is equal to the mixture of the utilities, or, in other terms, \( U \) is a linear functional. (More precisely, functionals for which (3.5.8) holds only for \( \alpha \in [0, 1] \) are called affine; we keep the more explicit term ‘linear’.)

To prove (3.5.8), it suffices to replace \( F \) in (3.5.5) by the mixture (3.5.7) and to write...
EXAMPLE 2. Consider again the $\Delta$-scheme for $n = 3$. Any rule of comparison of distributions in this particular case amounts to a rule of comparison of points in $\Delta$. Identifying distributions $F$ with vectors $(p_1, p_2, p_3)$, we write the expectation $U(F)$ as

$$U(F) = u(x_1)p_1 + u(x_2)p_2 + u(x_3)p_3 = u(x_1)p_1 + u(x_2)(1 - p_1 - p_3) + u(x_3)p_3$$

where $a = u(x_1) - u(x_2)$, $b = u(x_3) - u(x_2)$, $h = u(x_2)$.

Thus, $U(F)$ is a linear function of $p_1$ and $p_3$. □

Let us fix an order $\succeq$ on $\mathcal{F}$. We call a set $F \subseteq \mathcal{F}$ of distributions an equivalence set, or equivalence class, if $F \succeq G$ for any $F, G \in \mathcal{F}$. In other words, any equivalence set contains only distributions equivalent to each other.

We see from (3.5.8) that under an EUM criterion, such a set is linear in the sense that if $F, G \in \mathcal{F}$, then any mixture $\alpha F + (1 - \alpha)G$ also belongs to $\mathcal{F}$. Indeed, let $U(F) = U(G)$. Then $U(\alpha F + (1 - \alpha)G) = \alpha U(F) + (1 - \alpha)U(G) = \alpha U(F) + (1 - \alpha)U(F) = U(F)$.

EXAMPLE 3. In the $\Delta$-scheme, $n = 3$, for points $(p_1, p_3)$ from $\Delta$ to be equivalent, the corresponding values of $U(F) = ap_1 + bp_3 + h$ should be equal to the same constant. In other words, an equivalence set is a set of points

$$ap_1 + bp_3 + h = d,$$

where $d$ is a constant. This is a line, or more precisely, the part of the line (3.5.9) lying in $\Delta$. This part is, certainly, a segment; see Fig.11a. All equivalence lines are parallel with the slope

$$k = \frac{a}{b} = \frac{u(x_2) - u(x_1)}{u(x_3) - u(x_2)}. \quad (3.5.10)$$

EXAMPLE 4. Let $u(x)$ be strictly increasing. Consider in $\Delta$ two parallel lines with the slope $k$ from (3.5.10). Points in each line are equivalent, points lying in different lines are not. Which line is better?

Answer: The higher. Intuitively it is clear that the closer to the best point $(0, 1)$, the better. For a rigorous proof, consider a point $p^0 = (0, p_3^0)$ in the vertical axis and denote by $F^0$ the corresponding distribution (see Fig.11a). Clearly, $U(F^0) = bp_3^0 + h = [u(x_3) - u(x_2)]p_3^0 + u(x_2)$. The function $u$ is increasing, and hence $u(x_3) > u(x_2)$ because $x_3 > x_1$. Then $U(F^0)$ is increasing in $p_3^0$, and the higher the point $p^0$ is, the better it is. On the other hand, the higher $p_3^0$, the higher the line starting from this point with the slope $k$. □

EXAMPLE 5. Ann is an EU maximizer. You ask her to compare two random variables both taking the same three values $x_1, x_2, x_3$; the first r.v. — with probabilities $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, respectively,
1. COMPARISON OF RANDOM VARIABLES

An equivalence line

FIGURE 11.

the second – with probabilities \( \left( \frac{1}{3}, \frac{1}{5}, \frac{1}{2} \right) \). It has turned out that Ann found these distributions equivalent for her. Is this information enough to completely determine Ann’s preferences among ALL random variables taking the same values?

Answer: Yes. Since the points mentioned are equivalent, the line going through these points and having the slope \( \frac{1}{2} - \frac{1}{4} = 3 \) is an equivalence line. Hence, \( k = 3 \).

Consider two other points, \( p_1 \) and \( p_2 \), in \( \Delta \), and draw the line \( l \) with the same slope \( k \) going through \( p_1 \); see Fig.11b. If the point \( p_2 \) turns out to be below \( l \), then \( p_2 \) is worse than \( p_1 \); if \( p_2 \) is above \( l \), then \( p_2 \) is better.

Certainly, the particular probabilities in Example 5 do not matter: \( \text{in the } \Delta\text{-scheme for } n = 3, \text{ in order to completely determine the preference order, it suffices to find two equivalent distributions.} \)

EXAMPLE 6. Assume that in the previous example, the values \( x_1, x_2, x_3 \) are equally spaced: \( x_2 - x_1 = x_3 - x_2 \). For instance, \( x_1 = 0, x_2 = 20, x_3 = 40 \). Then, for a concave \( u(x) \), we have \( k \geq 1 \). Indeed, in our case, \( x_2 = (x_1 + x_3)/2 \), and by property (3.4.6),

\[
 u(x_2) \geq \frac{u(x_1) + u(x_3)}{2}
\]

This implies that \( u(x_2) - u(x_1) \geq u(x_3) - u(x_1) \). Hence, in the risk aversion case, the slope of equivalence lines corresponds to an angle at least 45°.

Certainly, the reader certainly sees that the above reasoning may be easily generalized to the \( \Delta \)-scheme for \( n > 3 \). In this case, equivalence sets are planes for \( n = 4 \), and hyper-planes in \( \mathbb{R}^{n-1} \) for \( n > 4 \). We come to a general conclusion:

To completely determine the preference order among all distributions concentrated at \( n \) fixed points, it is enough to know \( n - 1 \) equivalent distributions.

In Exercise 42, we justify it rigorously. Risk aversion is again connected with the position of equivalence hyper-planes; we skip details.

3.5.6 An axiomatic approach

D. Bernoulli did not derive the EUM criterion from some original assumptions: he simply suggested it and gave an argument for why this criterion seems natural. The modern
3. Expected Utility

approach to such problems is more sophisticated. We do not point out a good solution from the very beginning, but first establish desirable properties of the solution, called axioms. After that, we try to figure out which solutions satisfy the properties established.

In Utility Theory such an approach was first applied in 1944, more than 200 years after D. Bernoulli’s paper [11], by J. von Neumann and O. Morgenstern in [90].

**The basic axiom.** Consider a preference order \( \succcurlyeq \) on a set \( \mathcal{F} = \{F\} \) of distributions \( F \).

**Axiom 10** (the Independence Axiom). Let \( F, G \) and \( H \) belong to \( \mathcal{F} \), and \( F \succsim G \). Then for any \( \alpha \in [0, 1] \),

\[
\alpha F + (1 - \alpha)H \succeq \alpha G + (1 - \alpha)H.
\]

The axiom sounds quite plausible: if you mix \( F \) and \( G \) with the same distribution \( H \), then the relation between the mixtures will be the same as for the original distributions \( F \) and \( G \). (We will see, however, in Section 4 that this is far from being always true.)

A geometric illustration is given in Fig. 12. Let a “point” \( F \) be “better” than \( G \). Consider the “segments” connecting \( F \) and \( H \), and \( G \) and \( H \). Let us choose two points, one in each segment, in a way that their positions between \( F \) and \( H \), and \( G \) and \( H \), respectively, would be in the same “proportion”. Then these points are in the same relation as the original points \( F \) and \( G \).

**EXAMPLE 1.** John, when comparing two random variables,

\[
X_1 = \begin{cases} 
$100 \text{ with probability } 0.2, \\
$0 \text{ with probability } 0.8, 
\end{cases} \quad \text{and} \quad X_2 = \begin{cases} 
$50 \text{ with probability } 0.4, \\
$0 \text{ with probability } 0.6, 
\end{cases}
\]

has decided that for him \( X_2 \succ X_1 \). (John seems to be a strong risk averter). After that John is offered to play one of two games. In both games, a coin will be tossed, and in the case of a head, John will get $100. In the case of a tail, in the first game, John will get a random prize amounting to the r.v. \( X_1 \), while in the second game – to the r.v. \( X_2 \). Thus, eventually, the prizes for the games are

\[
Y_1 = \begin{cases} 
$100 \text{ with probability } 0.5 + 0.5 \cdot 0.2 = 0.6, \\
$0 \text{ with probability } 0.4 
\end{cases} \quad \text{and} \quad Y_2 = \begin{cases} 
$100 \text{ with probability } 0.5 \\
$50 \text{ with probability } 0.5 \cdot 0.4 = 0.2 \\
$0 \text{ with probability } 0.3 
\end{cases}
\]

respectively. If John, consciously or not, follows the Independence Axiom, he would prefer \( Y_2 \) to \( Y_1 \). □

**Proposition 11** The EUM criterion satisfies the Independence Axiom.

**Proof** is practically immediate. In the EUM case, if \( F \succsim G \), then \( U(F) \geq U(G) \). Hence, by (3.5.8),

\[
U(\alpha F + (1 - \alpha)H) = \alpha U(F) + (1 - \alpha)U(H) \geq \alpha U(G) + (1 - \alpha)U(H) = U(\alpha G + (1 - \alpha)H).
\]

Consequently, \( \alpha F + (1 - \alpha)H \succeq \alpha G + (1 - \alpha)H \). □
The main result of classical utility theory is that under some additional, more technical, conditions, the converse claim is also true:

The Independence Axiom implies the EUM principle.

We skip here details and a rigorous formulation; for a detailed exposition see, e.g., [78].

Linearity and continuity lead to EUM. Consider a somewhat weaker proposition illustrating, to a certain extent, the situation.

We saw that the utility functional (3.5.5) was linear, that is,

$$U(\alpha F_1 + (1 - \alpha) F_2) = \alpha U(F_1) + (1 - \alpha) U(F_2).$$

The question we discuss here is whether any linear utility functional admits the representation (3.5.5). (That is, instead of the Independence Axiom, we consider the linearity property itself.) The answer is "yes, if $U(F)$ is in a certain sense continuous".

To specify what this means, first note that convergence of distributions may be defined in different ways. We consider two.

1. **Weak convergence.** We say that a sequence $F_n$ converges to a distribution $F$ weakly, and write it as $F_n \xrightarrow{w} F$, if $F_n(x) \rightarrow F(x)$ for any $x$ at which $F(x)$ is continuous.

2. **Convergence for all sets.** We say that a sequence $F_n$ converges to a distribution $F$ for all sets, and write it as $F_n \xrightarrow{c} F$, if $F_n(B) \rightarrow F(B)$ for any $B$.

For more detail on convergence of r.v.'s and distributions, see Section 0.5.

Clearly, convergence in the later case implies convergence in the former.

Accordingly, we consider two definitions of continuity of a functional $U(F)$.

**Condition C1.** (Weak continuity): $U(F_n) \rightarrow U(F)$ provided $F_n \xrightarrow{w} F$.

**Condition C2.** (Continuity with respect to convergence for all sets): $U(F_n) \rightarrow U(F)$ provided $F_n \xrightarrow{c} F$.

Condition C1 is stronger than Condition C2 since convergence $U(F_n) \rightarrow U(F)$ in the former case takes place under a weaker (!) requirement on the convergence of $F_n$.

**Theorem 12** Suppose $U(F)$ is defined on the set of all distributions and is linear, i.e., (3.5.8) is true. Then, if Condition C2 holds, there exists a bounded function $u(x)$ such that

$$U(F) = \int_{-\infty}^{\infty} u(x) dF(x).$$

(3.5.11)

If Condition C1 holds, the function $u(x)$ in (3.5.11) is continuous.

(A proof may be found in [37], [110].)
4 NON-LINEAR CRITERIA

The EUM approach may be considered as a first approximation to the description of people’s preferences. Over the years, there has been a great deal of discussion about the adequacy of this approach. Many experiments have been provided and a number of examples have been suggested, showing that the EUM approach is far from being efficient in all situations. The existence of such examples is not surprising; on the contrary, it would have been surprising if the behavior of such sophisticated (and sometimes strange) creatures as human beings had been always well described by simple linear functions. This section concerns some elements of modern utility theory.

4.1 Allais’ paradox

The following example considered by M. Allais [2] is probably the most famous. Though being contrived, it is very illustrative. Consider distributions $F_1, F_2, F_3, F_4$ of a random income with values $0, 10$ millions, or $30$ millions. The corresponding probabilities are given in the following table.

<table>
<thead>
<tr>
<th></th>
<th>$0$</th>
<th>$10$ million</th>
<th>$30$ million</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.01</td>
<td>0.89</td>
<td>0.1</td>
</tr>
<tr>
<td>$F_3$</td>
<td>0.9</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>$F_4$</td>
<td>0.89</td>
<td>0.11</td>
<td>0</td>
</tr>
</tbody>
</table>

Apparently, a majority of people would prefer $F_1$ to $F_2$, reasoning as follows. Ten million dollars is a lot of money, for ordinary people as inconceivable as thirty million. So, it is better to get ten for sure than to go in pursuit of thirty million at the risk of receiving nothing (even if the probability of this is very small). Thus, $F_1 \succ F_2$.

The situation with $F_3$ and $F_4$ is different. Now the probabilities of receiving nothing are large – and hence one should be ready to lose, and these probabilities are practically the same. Then it is reasonable to choose the variant with the larger prize. So, $F_3 \succ F_4$.

Let us consider the mixtures $\frac{1}{2}F_1 + \frac{1}{2}F_3$ and $\frac{1}{2}F_2 + \frac{1}{2}F_4$. If the preference $\succ$ had been preserved by a utility functional (3.5.4), in the light of the linearity property (3.5.8), we would have had $\frac{1}{2}F_1 + \frac{1}{2}F_3 \succ \frac{1}{2}F_2 + \frac{1}{2}F_4$.

However, as a matter of fact, as is easy to calculate, $\frac{1}{2}F_1 + \frac{1}{2}F_3 = \frac{1}{2}F_2 + \frac{1}{2}F_4$.

Next, we address several directions in which the EUM criterion can be generalized. (In particular, the situation in Allais’s paradox may be described by using the schemes below; we will skip concrete calculations.)

Note also that examples in this section do not aim to justify criteria we are introducing; justification comes from empirical evidence and qualitative reasoning based on axioms we will discuss. The goal of these examples is more modest – to demonstrate how criteria
introduced will work if we accept them, and what new features they can raise in comparison with the classical EUM criterion.

4.2 Weighted utility

The criterion below is based on two functions: $u(x)$ which we view as a utility function, and a non-negative function $w(x)$ which we call a weighting function. Consider the functional

$$W(F) = \frac{\int_{-\infty}^{\infty} u(x)w(x)dF(x)}{\int_{-\infty}^{\infty} w(x)dF(x)},$$

(4.2.1)

assuming that the denominator in (4.2.1) is not zero.

Note that if $F$ is the distribution of a r.v. $X$, then (4.2.1) may be rewritten as

$$W(F) = \frac{E\{u(X)w(X)\}}{E\{w(X)\}}.$$

The difference between the classical expected utility scheme and the last case is that now we assign to different values $x$ weights $w(x)$. If all weights $w(x) \equiv 1$, the denominator in (4.2.1) equals one, and we deal with the EUM case.

Since we compare here rather distributions than r.v.’s themselves, it is convenient to define a preference order $\succeq$ on the set $\mathcal{F} = \{F\}$ of distributions $F$ for which $W(F)$ is well defined. Let an order $\succeq$ be preserved by the functional $W$:

$$F \succeq G \iff W(F) \geq W(G).$$

When comparing r.v.’s, we will say that $X \succeq Y$ if $F_X \succeq G_Y$, where $F_X$ and $G_Y$ are the distributions of $X$ and $Y$, respectively.

Following tradition, we denote by $\delta_c$ the distribution of a non-random $X \equiv c$. Then the certainty equivalent $c(F)$ is defined as a number $c$ such that $\delta_c \simeq F$.

In the particular case of this section, $W(\delta_c) = u(c)w(c)/w(c) = u(c)$. Thus, for the certainty equivalent $c = c(F)$ of a r.v. $X$ with a distribution $F$, we have $u(c) = W(F)$, and

$$c(F) = u^{-1}(W(F)).$$

(4.2.2)

EXAMPLE 1. Let $u(x) = x^\alpha$, and $w(x) = x^\beta$. We assume $\alpha > 0$. As to the parameter $\beta$, depending on the situation, it may be either positive (the larger a value, the larger its weight), or negative (the larger a value, the less its weight), or zero. In the last case, we deal with the EUM criterion.

For a positive r.v. $X$ with a distribution $F$, by definition (4.2.1),

$$W(F) = \frac{E\{X^{\alpha+\beta}\}}{E\{X^\beta\}}.$$

(4.2.3)

In particular, for $\alpha = \beta = 1$, we have

$$W(F) = \frac{E\{X^2\}}{E\{X\}} = \frac{E\{X^2\}}{(E\{X\})^2}E\{X\}.$$
4. Non-Linear Criteria

It is well known that \( (E\{X\})^2 \leq E\{X^2\} \) (see, for example, Exercise 32b, or recall that \( \text{Var}(X) = E\{X^2\} - (E\{X\})^2 \geq 0 \)). Hence, \( W(F) \geq E\{X\} \). Because in the last case \( u(x) = x \) and, consequently, \( u^{-1}(x) = x \), the certainty equivalent \( c(F) = W(F) \geq E\{X\} \), so we deal with a risky person. It is not surprising since large values have large weights. □

EXAMPLE 2. Let \( X \) in the previous example be uniformly distributed on \([0,d]\). The reader is invited to verify that in this case, for \( \beta > -1 \),

\[
W(F) = \frac{1 + \beta}{1 + \alpha + \beta} d^\alpha.
\]

Following (4.2.2), for the certainty equivalent we have

\[
c(F) = (W(F))^{1/\alpha} = \left( \frac{1 + \beta}{1 + \alpha + \beta} \right)^{1/\alpha} d.
\] (4.2.4)

If \( \beta = 0 \), we come to the result of Example 3.1.3-2, which is not surprising since in this case \( w(x) \equiv 1 \), and we deal with expected utility. The larger \( \beta \), the greater \( c(F) \), and this is also understandable: the larger \( \beta \), the larger weights for large \( x \)'s, so the certainty equivalent must grow as \( \beta \) increases. □

Next, we generalize the model from Section 3.2. Consider the maximal premium a client with a wealth of \( w \) is willing to pay to insure a risk \( \xi \). As in Section 3.2, we write that

\[
w - G_{\text{max}} \simeq w - \xi.
\] (4.2.5)

Since the r.-h.s. of (4.2.5) is certain, (4.2.5) is equivalent to \( w - G_{\text{max}} = c(F_{w-\xi}) \), where as usual, the symbol \( F_X \) stands for the distribution of a r.v. \( X \). So,

\[
G_{\text{max}} = w - c(w - \xi) = w - u^{-1}(W(F_{w-\xi})).
\]

EXAMPLE 3. As in Examples 1-2, let \( u(x) = x^\alpha \) and \( w(x) = x^\beta \). Let \( \xi \) be uniformly distributed on \([0,d]\), and \( w = d \). Then the r.v. \( w - \xi \) is also uniformly distributed on \([0,d]\), and it follows from (4.2.4) that

\[
G_{\text{max}} = w - c(w - \xi) = d \left( 1 - \left( \frac{1 + \beta}{1 + \alpha + \beta} \right)^{1/\alpha} \right).
\]

The classical EUM case corresponds to \( \beta = 0 \). We see that for \( \beta > 0 \), the maximum premium becomes smaller. □

Next, we discuss the linearity issue (see Section 3.5.5). If we replace \( F \) in (4.2.1) by a mixture \( F^{(\alpha)} = \alpha F_1 + (1 - \alpha) F \) (compare with Section 3.5.5), we get

\[
W(F^{(\alpha)}) = \frac{\alpha \int_{-\infty}^{\infty} u(x)w(x)dF_1(x) + (1 - \alpha) \int_{-\infty}^{\infty} u(x)w(x)dF_2(x)}{\alpha \int_{-\infty}^{\infty} w(x)dF_1(x) + (1 - \alpha) \int_{-\infty}^{\infty} w(x)dF_2(x)}.
\] (4.2.6)

In general, this quantity is certainly not equal to \( \alpha W(F_1) + (1 - \alpha) W(F_2) \) (see also Exercise 55). Hence, in this case, the linearity property and the Independence Axiom do not hold.
EXAMPLE 4. Consider the Δ-scheme described in Section 3.5.5 with r.v.’s taking only three fixed values \(x_1, x_2, x_3\). Let \(x_1 < x_2 < x_3\). Any distribution \(F\) of such a r.v. is identified with the probability vector \((p_1, p_2, p_3)\), where \(p_1\) is the probability of the value \(x_1\). Consider points in \(\Delta = \{(p_1, p_3) : p_1, p_3 \geq 0, p_1 + p_3 \leq 1\}\). Setting \(p_2 = 1 - p_1 - p_3\) we have

\[
W(F) = \frac{u(x_1)w(x_1)p_1 + u(x_2)w(x_2)p_2 + u(x_3)w(x_3)p_3}{w(x_1)p_1 + w(x_2)p_2 + w(x_3)p_3} = \frac{ap_1 + bp_3 + h}{\tilde{a}p_1 + \tilde{b}p_3 + \tilde{h}},
\]

where

\[
a = u(x_1)w(x_1) - u(x_2)w(x_2), b = u(x_3)w(x_3) - u(x_2)w(x_2), h = u(x_2)w(x_2),
\]

\[
\tilde{a} = w(x_1) - w(x_2), \tilde{b} = w(x_3) - w(x_2), \tilde{h} = w(x_2).
\]

Since we consider only distributions for which \(\int_{-\infty}^{\infty} w(x)dF(x) > 0\), we consider only those points in \(\Delta\) for which the denominator in (4.2.7) is positive. All points for which \(W(F)\) equals a constant \(d\) are points for which \(ap_1 + bp_3 + h = d(\tilde{a}p_1 + \tilde{b}p_3 + \tilde{h})\), or

\[
(a - d\tilde{a})p_1 + (b - d\tilde{b})p_3 + h - d\tilde{h} = 0.
\]

This is a line, or more precisely the part of the line (4.2.8) lying in \(\Delta\). However, the slope of this line depends on \(d\), so lines corresponding to different \(d\)’s are not parallel (!).

Thus, although the Independence Axiom and the corresponding linearity property are not true in this case, we have a sort of the linearity property when we consider equivalent distributions.

\[
\text{An equivalence segment}
\]

It is interesting that all lines (4.2.8) intersect at the same point. This point is the intersection of two lines: \(l_0\) defined by the equation \(ap_1 + bp_3 + h = 0\) (points where \(W(F) = 0\)), and \(l_w\) defined by \(\tilde{a}p_1 + \tilde{b}p_3 + \tilde{h} = 0\) (points where the denominator in (4.2.7) is zero, and hence \(W(F)\) is not defined). Advice on how to show it rigorously is given in Exercise 54, a typical picture is depicted in Fig.13.

**FIGURE 13.**

EXAMPLE 5. In the Δ-scheme, let \(x_1 = 0, x_2 = 1, x_3 = 2\); \(u(x) = \sqrt{x}\), and \(w(x) = 1/(1 + x)\). Then, as is easy to verify, \(a = -\frac{1}{2}, b = \sqrt{\frac{7}{1}} - \frac{1}{2} \approx -0.029, h = \frac{1}{2}, \tilde{a} = \frac{1}{2}, \tilde{b} = -\frac{1}{6} \) and \(\tilde{h} = \frac{1}{2}\). Thus, up to the third digit, the point of the intersection of all lines (4.2.8) is the intersection of the lines \(-0.5p_1 - 0.029p_3 + 0.5 = 0\) and \(0.5p_1 - 0.166p_3 + 0.5 = 0\). The approximate solution is the point \((0.704, 5.111)\). \(\square\)

The next scheme generalizes the approach of this section.

4.3 Implicit or comparative utility

4.3.1 Definitions and examples

In this section, we consider not a utility function \(u(x)\) but a function \(v(x, y)\) which we will call an *implicit utility or comparative utility* function, and interpret it as a function
indicating to what extent income \( x \) is preferable to income \( y \). One may say that \( v(x, y) \) is the comparative utility of \( x \) with respect to \( y \). In light of this, we assume \( v(x, x) = 0 \), \( v(x, y) \geq 0 \) if \( x \geq y \), and \( v(x, y) \leq 0 \) if \( x \leq y \). Sometimes one can assume \( v(x, y) = -v(y, x) \) but in general it may be false: \( x \) may be “better” than \( y \) to a smaller extent than \( y \) is “worse” than \( x \); see also Example 3 below.

It is natural to assume that \( v(x, y) \) is non-decreasing in \( x \) and is non-increasing in \( y \), which again reflects the property “the larger, the better”.

EXAMPLE 1. Let

\[
v(x, y) = \frac{x - y}{1 + |x| + |y|}. \tag{4.3.1}
\]

In this case, for small \( x \) and \( y \) the comparative utility almost equals the difference \( x - y \), while for large \( x \)'s and \( y \)'s the measure \( v(x, y) \) may be viewed as a relative difference: \( x - y \) is divided by \( 1 + |x| + |y| \). □

We define the certainty equivalent of a r.v. \( X \) as a solution to the equation

\[
E\{v(X, c)\} = 0, \tag{4.3.2}
\]

provided that this solution exists and is unique. The interpretation is clear: \( c(X) \) is the certain amount whose comparative utility with respect to \( X \) equals zero on the average.

EXAMPLE 2. Let \( v(x, y) \) be given by (4.3.1), and let \( X = d > 0 \) or 0 with equal probabilities. Then (4.3.2) is equivalent to \( \frac{1}{2} v(d, c) + \frac{1}{2} v(0, c) = 0 \). Obviously, \( c \) should be between \( d \) and 0. So, \( c \geq 0 \) and the equation may be written as

\[
\frac{1}{2} \frac{d - c}{1 + d + c} + \frac{1}{2} \frac{0 - c}{1 + 0 + c} = 0.
\]

This is a quadratic equation. Its positive solution is

\[
c = \frac{\sqrt{1 + 2d} - 1}{2}.
\]

As is easy to verify, \( c \) above is less than \( E\{X\} = d/2 \), so we have a sort of risk aversion. For large \( d \), we have \( c \sim \sqrt{d}/2 \), which is much smaller than \( d/2 \). In Exercise 56, we compute the maximal accepted premium. □

Once we have defined what is certainty equivalent in this case, we can define the corresponding preference order by the rule

\[
X \succ Y \iff c(X) \geq c(Y).
\]

First of all, note that this scheme includes the classical EU maximization as a particular case. Indeed, let \( v(x, y) = u(x) - u(y) \), where \( u \) is a utility function. Assume that \( u \) is increasing, so its inverse \( u^{-1} \) exists. In this case, (4.3.2) implies \( E\{u(X)\} - E\{u(c)\} = 0 \), and since \( c \) is certain, we have \( E\{u(X)\} = u(c) \). Hence, \( c(X) = u^{-1}(E\{u(X)\}) \), as in the classical case. Because \( u^{-1}(x) \) is increasing, the relation \( c(X) \geq c(Y) \) is equivalent to the relation \( E\{u(X)\} \geq E\{u(Y)\} \).
1. COMPARISON OF RANDOM VARIABLES

Furthermore, the weighted utility scheme is also a particular case of the comparative utility. To show it, set
\[ v(x, y) = w(x)[u(x) - u(y)]. \]
In this case, for the certainty equivalent \( c \) of a r.v. \( X \) with a distribution \( F \) we write
\[
0 = E\{v(X, c)\} = \int_{-\infty}^{\infty} v(x, c) dF(x) = \int_{-\infty}^{\infty} w(x)[u(x) - u(c)] dF(x)
= \int_{-\infty}^{\infty} u(x) w(x)dF(x) - u(c) \int_{-\infty}^{\infty} w(x)dF(x).
\]
From this it follows that \( u(c) = W(F) \), where \( W(F) \) is the same as in (4.2.1). If \( u(x) \) is strictly increasing, then \( c = u^{-1}(W(F)) \). Since \( u^{-1} \) is also strictly increasing, \( c(F) \geq c(G) \)
iff \( W(F) \geq W(G) \).

Let now \( v(x, y) \) be concave with respect to \( x \). Consider a r.v. \( X \) and set \( m = E\{X\} \) and \( c = c(X) \). By definition, \( v(m, m) = 0 \). Then, by Jensen’s inequality,
\[ v(m, m) = 0 = E\{v(X, c)\} \leq v(E\{X\}, c) = v(m, c). \]
Thus, \( v(m, m) \leq v(m, c) \). Since \( v(x, y) \) is non-increasing in \( y \), this implies that
\[ c(X) \leq E\{X\}. \]
A good example for such a function \( v \) is
\[ v(x, y) = g(x - y), \]
where \( g(s) \) is a concave increasing function such that \( g(0) = 0 \). Note that in this case we should not expect \( v(x, y) = -v(x, y) \).

**EXAMPLE 3 (a jealous person).** Let Mr. J.’s implicit utility function \( v(x, y) = g(x - y) \), where
\[
g(s) = \begin{cases} 
\frac{s}{1+s} & \text{if } s \geq 0, \\
0 & \text{if } s < 0.
\end{cases}
\]

The function \( g(s) \) is concave, its graph is given in Fig.14.

**FIGURE 14.**

Mr. J. may be characterized as pretty jealous. Assume that \( x \) is Mr. J.’s wealth, and he compares it with a Mr. A.’s wealth \( y \). If \( x \) is much larger than \( y \), the comparative utility \( v(x, y) \) is, nevertheless, not large, and \( v(x, y) \to 1 \) as \( x - y \to \infty \). (Mr. J. does not think that his wealth is much more valuable than that of Mr. A.)

On the other hand, if \( x \) is much smaller than \( y \), the comparative utility is negative with a large absolute value, and \( v(x, y) \to -\infty \), as \( x - y \to -\infty \). (Now Mr. J. considers himself much less happy than Mr. A.)

Let a r.v. \( X = d > 0 \) or \( 0 \) with equal probabilities. In this case, equation (4.3.2) is reduced to \( \frac{1}{2}g(d - c) + \frac{1}{2}g(-c) = 0 \), which leads to
\[
\frac{d - c}{1+d-c} = c.
\]
This is a quadratic equation. The solution that lies between \(d\) and 0 is
\[
c = c(X) = \frac{d}{1 + (d/2) + \sqrt{1 + (d^2/4)}}.
\]
The denominator is greater than two. So, \(c < (d/2) = E\{X\}\). □

4.3.2 In what sense the implicit utility criterion is linear

As we saw, the equation (4.3.2) may be written as
\[
\int_{-\infty}^{\infty} v(x, c) dF(x) = 0, \tag{4.3.3}
\]
where \(F\) is the d.f. of \(X\). Consequently, the solution to this equation depends not on the r.v. \(X\) itself but on its d.f. \(F\). That is, \(c(X)\) is a function (or functional) of \(F\), and it should not cause confusion if we use also the notation \(c(F)\) defining it as a solution to (4.3.3).

Consider a set \(\mathcal{F} = \{F\}\) of distributions. Assume that the function \(v(x, y)\) is such that \(c(F)\) exists and is unique for each \(F \in \mathcal{F}\). Let us define a preference order \(\succeq\) in \(\mathcal{F}\) by
\[
F \succeq G \iff c(F) \geq c(G).
\]

**Proposition 13** Let \(F \simeq G\) (\(F\) is equivalent to \(G\).) Then for any \(\alpha \in [0, 1]\) the mixture \(F^{(\alpha)} = \alpha F + (1 - \alpha)G \simeq F\).

**Proof** is short. If \(F \simeq G\), then \(F\) and \(G\) have the same certainty equivalent. Denote it by \(c\). By definition (4.3.3),
\[
\int_{-\infty}^{\infty} v(x, c) dF(x) = 0, \quad \text{and} \quad \int_{-\infty}^{\infty} v(x, c) dG(x) = 0.
\]
Then
\[
\int_{-\infty}^{\infty} v(x, c) dF^{(\alpha)}(x) = \alpha \int_{-\infty}^{\infty} v(x, c) dF(x) + (1 - \alpha) \int_{-\infty}^{\infty} v(x, c) dG(x) = 0.
\]
Hence, \(c\) is the certainty equivalent for \(F^{(\alpha)}\) too. ■

Geometrically, it means that equivalent points still lie in the same line, but equivalency lines may not be parallel.

**Example 1.** Consider again the \(\Delta\)-scheme for \(n = 3\). In this case, equation (4.3.3) may be written as
\[
v(x_1, c)p_1 + v(x_2, c)p_2 + v(x_3, c)p_3 = 0.
\]
Since \(p_2 = 1 - p_1 - p_3\), we can rewrite it as
\[
a(c)p_1 + b(c)p_3 + h(c) = 0, \tag{4.3.4}
\]
where \(a(c) = v(x_1, c) - v(x_2, c)\), \(b(c) = v(x_3, c) - v(x_2, c)\), \(h(c) = v(x_2, c)\).
1. COMPARISON OF RANDOM VARIABLES

Let us compare this with what we did in Section 3.5.5, where \( a, b, \) and \( h \) did not depend on \( c \). For a fixed \( c \), all points that satisfy (4.3.4) lie in a line, and the slope of this line depends on \( c \). Unlike the case of weighted utility, the dependence of the slope on \( c \) is rather arbitrary, so we should not expect that all these lines intersect at the same point. A typical picture is given in Fig.15.

As we saw, and as we see now again, the Independence Axiom (IA) does not hold in this case. A general theory which we do not present here shows that IA may be replaced by the following weaker Axiom 14 (The Betweenness Axiom.) Let \( F, G \in \mathcal{F} \), and \( F \simeq G \). Then for any \( \alpha \in [0,1] \),

\[
\alpha F + (1-\alpha) G \simeq G.
\]

We see that, unlike the IA, the Betweenness Axiom (BA) deals not with the case when \( F \succeq G \) but only with equivalent \( F \) and \( G \). The fact that BA is weaker than IA is presented in Proposition 15 If IA holds, then BA holds too.

Proof. Set \( H = G \) in the formulation of the Independence Axiom 10. Let the IA hold, and let \( F \succeq G \). Then \( \alpha F + (1-\alpha) G \succeq \alpha G + (1-\alpha) G = G \), that is, \( \alpha F + (1-\alpha) G \succeq G \). Now let \( F \simeq G \). Then \( F \succeq G \), and \( G \succeq F \) simultaneously. Consequently, \( \alpha F + (1-\alpha) G \succeq G \) and \( \alpha F + (1-\alpha) G \preceq G \), simultaneously. That is, \( \alpha F + (1-\alpha) G \simeq G \).

It proves that together with some more technical assumptions,

The Betweenness Axiom implies the implicit (or comparative) utility principle.

4.4 Rank Dependent Expected Utility

4.4.1 Definitions and examples

The next approach essentially differs from those of previous sections. For simplicity, we restrict ourselves to non-negative r.v.’s. Consider probability distributions \( F \) on \([0, \infty)\) and two functions: \( u(x) \) viewed again as a utility function, and a function \( \Psi(p) \) defined on \([0,1]\), which we call a transformation function. We assume \( \Psi(p) \) to be non-decreasing, \( \Psi(0) = 0 \), \( \Psi(1) = 1 \). We consider an individual (or investor) whose preferences are preserved by the function (or functional)

\[
R(F) = \int_0^\infty u(x)d\Psi(F(x)). \tag{4.4.1}
\]

The transformation (or weighting) function \( \Psi \) reflects the attitude of the individual to different probabilities. The individual, when perceiving information about the distribution
4. Non-Linear Criteria

Let us show that $F_\Psi(x)$ is indeed a distribution function. First, since we consider non-negative r.v.’s, we have $F(0) = 0$. Second, due to properties of $\Psi$, the function $F_\Psi(x) = \Psi(F(x))$ is non-decreasing. Moreover, $F_\Psi(0) = \Psi(F(0)) = \Psi(0) = 0$, and $F_\Psi(\infty) = \Psi(F(\infty)) = \Psi(1) = 1$. Hence, $F_\Psi(x)$ is a d.f.

Note also that in (4.4.1), we transform a distribution function, not a density. The latter transformation would lead to non-desired consequences. For example, a transformation $\Psi(f(x))$ of a density $f$ may be not a density, since $\int_0^x \Psi(f(x))dx$ might not equal one; so we would no longer deal with a probability distribution.

The quantity (4.4.1) is referred to as a Rank Dependent Expected Utility (RDEU). Another term in use is distorted expectation; see, e.g., [30].

The corresponding preference order $\succsim$ is preserved by the function $R(F)$, that is,

$$F \succsim G \iff R(F) \geq R(G).$$

A simple example is $\Psi(p) = p^\beta$. If $\beta = 1$, the subject perceives $F$ as it is, and deals with the “usual” expected utility. If $\beta < 1$, the investor overestimates the probability for the income to be less that a fixed value ($p^\beta > p$): the investor is “security-minded”. In the case $\beta > 1$, the investor underestimates the probability mentioned, being “potential-minded” (or opportunity-minded).

**EXAMPLE 1.** Let $X$ be uniformly distributed on $[0, 1]$. Its distribution function

$$F(x) = \begin{cases} 
0 & \text{if } x < 0, \\
1 & \text{if } x > 1, \\
x & \text{if } x \in [0, 1],
\end{cases}$$

and

$$F_\Psi(x) = \begin{cases} 
0 & \text{if } x < 0, \\
x^\beta & \text{if } x \in [0, 1],
1 & \text{if } x > 1.
\end{cases}$$

Then the density of the transformed distribution,

$$f_\Psi(x) = F'_\Psi(x) = \begin{cases} 
0 & \text{if } x < 0, \\
\beta x^{\beta - 1} & \text{if } x \in [0, 1],
0 & \text{if } x > 1.
\end{cases}$$

For example, if $\beta > 1$, the density $f_\Psi(x)$ is increasing, and while for the original distribution all values of $X$ are equally likely, in the “investor’s mind” it is not so: smaller values are less likely.

To the contrary, if $\beta < 1$, the density $f_\Psi(x) \to \infty$ (!) as $x \to 0$, that is, the investor strongly overestimates the probability to get nothing.

The case $\beta = 0$ corresponds to an “absolutely pessimistic” investor: $F_\Psi(x) = 0$ for $x < 0$, and $= 1$ for $x > 0$, that is, $F_\Psi$ is the distribution of a r.v. $X \equiv 0$. In this case, the investor expects that she/he will get nothing for sure.

**EXAMPLE 2.** Assume that an investor does not distinguish small values of the income. For instance, hoping for an income equal to $100,000$ on the average, the investor considers income values of $100$ or $1$ as too small and, consciously or not, identifies them with zero income.
Denote by $F$ the distribution of the investor’s income and assume that the investor identifies with zero all values which are less then the $\gamma$-quantile $q_\gamma(F)$ for some small fixed $\gamma$.

Suppose the same is true for “inconceivable” large values. For instance, the same investor may consider $\$1$ million or $\$10$ million an improbable luck, and (consciously or not) identify these numbers. More precisely, choosing for simplicity the same level $\gamma$, assume that the investor identifies with $q_{1-\gamma}(F)$ all values which are larger than $q_{1-\gamma}(F)$.

In both cases we may talk about the existence of a perception threshold. The situation we consider may be described by the truncation transforming function

$$
\Psi(p) = \begin{cases} 
\gamma & \text{if } p < \gamma, \\
 p & \text{if } \gamma \leq p < 1 - \gamma, \\
 1 & \text{if } p \geq 1 - \gamma.
\end{cases}
$$

In this case,

$$
F_\psi(x) = \begin{cases} 
\gamma & \text{if } 0 \leq x < q_\gamma(F), \\
 F(x) & \text{if } q_\gamma(F) \leq x < q_{1-\gamma}(F), \\
 1 & \text{if } x \geq q_{1-\gamma}(F).
\end{cases}
$$

Hence,

$$
R(F) = u(0) \cdot \gamma + \int_{q_\gamma(F)}^{q_{1-\gamma}(F)} u(x) dF(x) + u(q_{1-\gamma}(F)) \cdot \gamma. \tag{4.4.2}
$$

The functional (4.4.2) is not linear and should be distinguished from the naive criterion when truncation is carried out at a fixed, perhaps large, value not depending on $F$. □

EXAMPLE 3. Let $F$ be the distribution of a r.v. taking only two values, say, $a$ and $b > a$ with respective probabilities $p$ and $1 - p$. Then $F(x) = 0$ if $x \in [0, a)$, $F(x) = p$ if $x \in [a, b)$, and $F(x) = 1$ if $x \in [b, \infty)$. Consequently, $\Psi(F(x)) = 0$ if $x \in [0, a)$, $\Psi(F(x)) = \Psi(p)$ if $x \in [a, b)$, and $\Psi(F(x)) = 1$ if $x \in [b, \infty)$. Then, by (4.4.1),

$$
R(F) = u(a)\Psi(p) + u(b)[1 - \Psi(p)]. \tag{4.4.3}
$$

In this case, $\Psi(\cdot)$ “transforms” just one probability $p$.

Note also that if a r.v. $X \equiv c$, then its d.f. $\delta_c(x) = 1$ or $0$ for $x \geq 0$, and $x < 0$, respectively. Hence, $\Psi(\delta_c(x))$ also equals $1$ or $0$ for $x \geq 0$, and $x < 0$, respectively, and by (4.4.1) or (4.4.4),

$$
R(\delta_c) = u(c). \tag{4.4.4}
$$

□

EXAMPLE 4. Let a person having, for instance, the utility function $u(x) = \sqrt{x}$, choose between one of two retirement plans: either the annual pension is equal to $X = \$100,000$, or it is equal to r.v. $Y = \$50,000$ or $\$200,000$ with probabilities $1/2$.

For the numbers above, the expected utility criterion leads to a slight preference for the latter plan ($E\{u(X)\} \approx 316$ and $E\{u(Y)\} \approx 335$), which does not looks very realistic. One would expect most people to choose the plan $X$.

On the other hand, by (4.4.4), we have $R(F_X) = u(10^5)$ and, by (4.4.3), $R(F_Y) = u(5 \cdot 10^4)\Psi(1/2) + u(2 \cdot 10^2)[1 - \Psi(1/2)]$. 

1. COMPARISON OF RANDOM VARIABLES
4. Non-Linear Criteria

It is easy to calculate that \( R(F_X) > R(F_Y) \), that is, the individual would prefer \( X \), if \( \Psi(1/2) > 0.59 \). This means that such a person slightly overestimates the probability of the unlucky event to get $50,000 (since this probability equals 1/2). So, one can expect \( \Psi(p) \) to be concave for large \( p \)'s. Certainly, this naive example is given merely for illustration. □

For the certainty equivalent \( c = c(F) \) of a distribution \( F \), we have \( u(c) = R(F) \), and
\[
c(F) = u^{-1}(R(F)). \tag{4.4.5}
\]

**EXAMPLE 5.** Let \( F \) be the uniform distribution on \([0, b]\), \( u(x) = x^\alpha \), and \( \Psi(p) = p^\beta \).
Then, by (4.4.5),
\[
c(F) = \left( \int_0^b x^\alpha d(x/b)^\beta \right)^{1/\alpha} = \left( \frac{\beta^\alpha}{\beta + \alpha} \right)^{1/\alpha} = \left( \frac{\beta}{\beta + \alpha} \right)^{1/\alpha} b.
\]
For \( \beta = 1 \), we have \( c(F) = [1/(1 + \alpha)]^{1/\alpha} b \), which corresponds to the EUM case considered in Example 3.1.3-2. For \( \beta > 1 \), the certainty equivalent gets larger; for \( \beta < 1 \) – smaller. It makes sense: for \( \beta > 1 \), the individual underestimates the probabilities of “bad events”, so the certainty equivalent is larger in comparison with the case when the probabilities mentioned are perceived correctly. The case \( \beta < 1 \) is the opposite. □

4.4.2 Application to insurance

Consider the insurance model from Section 3.2, keeping the same notation for the wealth \( w \), the random loss \( \xi \), and the premium \( G \). Following the same logic, we see that for \( G \) to be acceptable, the certain quantity \( w - G \) should be preferable to the r.v. \( w - \xi \).

For the maximal accepted premium \( G_{\text{max}} \), the r.v. \( w - G_{\text{max}} \) should be equivalent to \( w - \xi \).

In the RDEU framework, this means that
\[
u(w - G_{\text{max}}) = R(F_{w-\xi}), \tag{4.4.6}
\]
where, as usual, \( F_X \) denotes the distribution of a r.v. \( X \).

**EXAMPLE 1.** As in Example 3.2-1, let \( u(x) = 2x - x^2 \), \( w = 1 \), and the r.v. \( \xi \) be uniformly distributed on \([0, 1]\). Let \( \Psi(p) = p^\beta \). Since \( 1 - \xi \) is also uniformly distributed on \([0, 1]\),
\[
R(F_{1-\xi}) = \int_0^1 (2x - x^2) dx^\beta = C_\beta,
\]
where \( C_\beta = \frac{\beta(3 + \beta)}{(1 + \beta)(2 + \beta)} \). Let \( y = w - G_{\text{max}} \). Then \( 2y - y^2 = C_\beta \). As in Example 3.2-1, we obtain from this that
\[
G_{\text{max}} = \sqrt{1 - C_\beta} = \sqrt{\frac{2}{(1 + \beta)(2 + \beta)}}.
\]
If \( \beta = 0 \) (the absolutely pessimistic investor from Example 4.4.1-1), \( G_{\text{max}} \) is equal to one, that is, to the maximal possible loss. (The investor feels that the maximal loss will happen.) The larger \( \beta \), the less \( G_{\text{max}} \), which is natural. For \( \beta = 1 \), we have \( G_{\text{max}} = 1/\sqrt{3} \) as in the expected utility case in Example 3.2-1. □
1. COMPARISON OF RANDOM VARIABLES

4.4.3 Further discussion and the main axiom

Next, we consider some possible forms of the transformation function $\Psi(p)$. To clarify the classification below, note that when saying that a subject underestimates the probability of an event, we mean that the subject perceives the likelihood of this event to be less than it really is. In the extreme case, the subject neglects the possibility of such an event. The four cases we discuss below are illustrated in Fig.16.

- $\Psi(p) \geq p$ and is concave. For any certain level of income, the subject overestimates the probabilities that the income will not reach this level. The subject is “security minded”.
- $\Psi(p) \leq p$ and is convex: the opposite case. The subject is “potential-minded”.
- $\Psi(p)$ is S-shaped. The subject underestimates the probabilities of very large and very small values and, consequently, proceeds from moderate values of income.
- $\Psi(p)$ is inverse-S-shaped: “cautiously hopeful”. Roughly speaking, the subject overestimates the probabilities of “very large” and “very small” values.

Many experiments presented in the literature testify to the inverse S-shaped pattern (see, e.g., [79]). However, one should realize that these experiments, usually dealing with students, concern one-time gains or investments, and the amounts of money involved are not
4. Non-Linear Criteria

large. In such situations, it is not surprising that people count to some extent on large values of the income, overestimating real probabilities of their occurrence.

In long run investment, when dealing with significant amounts of money and in situations when these amounts really matter for the investor (say, in the case of a retirement plan), the investor may exhibit a different behavior, proceeding from moderate values of the income rather than from the possibilities of large deviations. In such situations, an S-shaped transformation may be more adequate.

An interesting theory on possible forms of the transformation function $\Psi$ may be found, e.g., in [79] and [97].

In conclusion, we discuss the main axiom connected with RDEU.

Consider an investor with a preference order $\succsim$ and two d.f.’s, $F(x)$ and $G(x)$. Assume that $F(x) = G(x)$ for all $x$’s from a set $A$, and suppose that for the investor, $F \succsim G$.

Now, assume that we change $F(x)$ and $G(x)$ in a way such that

- all changes concern only values of $F(x)$ and $G(x)$ at $x$’s from $A$,
- when changing $F(x)$ and $G(x)$, we keep them equal to each other for $x \in A$.

If for any $F$ and $G$, after such a change, the investor continues to prefer $F$ to $G$, we say that the investor’s preferences satisfy the ordinal independence axiom. The idea of such axioms (if you change the common part, the relation does not change) is referred to as the sure-thing-principle. Formally, the above axiom may be stated as follows.

For any two d.f.’s, $F(x)$ and $G(x)$, such that $F(x) = G(x)$ for all $x$’s from some set $A$, consider two other d.f.’s, $\tilde{F}(x)$ and $\tilde{G}(x)$, such that $\tilde{F}(x) = \tilde{G}(x)$ and $\tilde{G}(x) = G(x)$ for all $x \notin A$, and $\tilde{F}(x) = \tilde{G}(x)$ for $x \in A$; see also Fig.17.

Consider the set $\mathcal{F} = \{F\}$ of all distributions $F$ and a preference order $\succsim$ on $\mathcal{F}$.

**Axiom 16 (The ordinal independence).** For any pair $F, G$ from $\mathcal{F}$, if $F \succsim G$, then for any distributions $\tilde{F}$ and $\tilde{G}$ with the mentioned above properties, $\tilde{F} \succsim \tilde{G}$. 

![FIGURE 17.](image-url)
One can prove that along with some more technical assumptions,

The ordinal independence axiom implies the RDEU principle.

To conclude the whole Section 4, it is worth pointing again to one special feature of all models we considered: we identified random variables with their distributions. In other words, two r.v.’s with the same distributions are viewed in these models as equivalent. Much evidence has been accumulated showing that this is not always the case. Sometimes people distinguish outcomes with the same probabilities if the ways leading to these outcomes are different. The question touched on is connected with the so called coalescing property in the modern theory of gambles; see, for example, [79]. This interesting issue, however, is beyond the scope of this book.

4.5 Remarks

The historical remarks below are far from being comprehensive. There is a very rich literature on various risk measures and criteria of comparison of risky alternatives. Many achievements of the modern theory are reflected in the monographs by M. Denuit, J. Dhaene, M. Goovaerts and P. Kaas [30], P. Fishburn [36], R.D. Luce [79], J. Quiggin [99], P. Wakker [131]. The reader can find there also historical notes and a rich bibliography. A bibliography with comments may be found also in P.Wakker’s web-site http://www1.fee.uva.nl/creed/wakker/refts/rfrncs.htm

To the author’s knowledge, the weighted utility concept was first considered by S.H. Chew and K.R. MacCrimmon (1979, [22], [23]) and H. Bühlmann (1980, [19]). The theory with corresponding axioms was later developed by S.H. Chew (1989, [21]). Other axioms leading to close criteria were considered by P. Fishburn (1988, [36]).

The implicit or comparative utility and the betweenness axiom or axioms quite similar to it were suggested and explored independently by A.Ya. Kiruta (1980, [70]), S.A. Smolyak (1983, [124]; see also [125]) and E. Dekel (1986, [29]). Various results on criteria close to comparative utility were considered in the mentioned monograph by P. Fishburn [36], and on the betweenness axiom – in the mentioned S.H. Chew’s paper [21].

The full RDEU model, including a set of axioms, was first suggested by J. Quiggin [100], though some earlier works of J. Quiggin had already contained some relevant ideas; see references in J. Quiggin [99]. Some models including weighting functions, say, in the case of binary gambles were considered earlier in the prospect theory of D. Kahneman and A. Tversky [65]. A special case of RDEU was independently considered in the “dual model” of M. Yaari (1987, [136]) and developed further by A. Roell (1987, [107]). To the author’s knowledge, an axiomatic system for the most general case including continuous distributions, was considered by P. Wakker (1994, [131]).
5. **OPTIMAL PAYMENT FROM THE STANDPOINT OF THE INSURED**

5.1 **Arrow’s theorem**

We consider here the following problem.

An individual with a wealth of $w$ is facing a random loss $\xi$ with a mean $m > 0$. To protect her/himself against at least a part of the risk, the individual appeals to an insurer.

The insurer, having many clients, when specifying the corresponding premium, proceeds merely from the mean value of the future payment. For example, if the mean payment is $\lambda$, the insurer agrees to sell the coverage for the premium $G = (1 + \theta)\lambda$ for a fixed $\theta > 0$. The coefficient $\theta$ is called a *relative security loading coefficient*. For instance, if $\theta = 0.1$, the insurer adds 10% to the mean payment. In following chapters, we will consider the characteristic $\theta$ repeatedly and in detail. Here, for us it is only important that there is a strict correspondence between $G$ and $\lambda$, and once $\lambda$ is given, the premium $G$ is fixed.

If the coverage is complete, the mean payment is equal to the mean loss, that is, $\lambda = m$. In particular, if the company proceeds from a security loading $\theta$, then the premium equals $G_{\text{complete}} = (1 + \theta)m$.

However, such a premium may be too large for the individual (or she/he may be just not willing to pay it). In this case, the individual buys non-complete coverage with a mean payment $\lambda < m$. In this case, the policy is specified by a payment function $r(x)$, the amount that the insurer will pay if the loss $\xi$ assumes a value $x$. Since the coverage is not complete, $r(x) \leq x$.

As we assumed, the insurer requires only one condition on $r(x)$ to hold:

$$E\{r(\xi)\} = \lambda. \quad (5.1.1)$$

In this case, the premium will be $G = (1 + \theta)\lambda$, and the individual can choose any $r(x)$ provided that (5.1.1) is true. The question is which $r(x)$ is the best.

Before considering examples of possible payment functions, note that (5.1.1) may be rewritten as follows.

Assume that $r(x)$ is non-decreasing, and $r(0) = 0$; both assumptions are quite natural. Let $F(x)$ be the d.f. of $\xi$. Then, by virtue of (0.2.2.1),

$$E\{r(\xi)\} = \int_0^\infty (1 - F(x))dr(x).$$

The significance of the “differential” $dr(x)$ is the same as $dF(x)$. Thus, condition (5.1.1) may be rewritten as

$$\int_0^\infty (1 - F(x))dr(x) = \lambda. \quad (5.1.2)$$

Consider particular examples of the payment function $r(x)$.

- *Proportional insurance* or *quota share insurance*: $r(x) = kx, k \leq 1$. Then $E\{r(\xi)\} = E\{k\xi\} = km$, and (5.1.1) implies that $k = \lambda/m$. 

• **Excess-of-loss or stop-loss insurance.** We will call it also *insurance with a deductible*. In this case,

\[
r(x) = r_d(x) = \begin{cases} 
0 & \text{if } x \leq d, \\
x - d & \text{if } x > d, 
\end{cases}
\]  

(5.1.3)

where the number \(d\) is called a deductible. In this case, payment is carried out only if the loss exceeds the level \(d\), and if it happens, the insurer pays the overshoot. The term excess-of-loss is used when such a rule concerns each contract separately, stop-loss – when it concerns the whole risk portfolio.

Inserting (5.1.3) into (5.1.2), we have

\[
\int_d^\infty (1 - F(x)) \, dx = \lambda. 
\]  

(5.1.4)

The last relation is an equation for \(d\) given \(\lambda\). Simple particular examples are relegated to Exercise 59.

• **Insurance with a limit coverage.** In this case,

\[
r(x) = \begin{cases} 
x & \text{if } x \leq s, \\
s & \text{if } x > s, 
\end{cases}
\]

where \(s\) is the maximum the insurer will pay. Again using (5.1.2), we see that restriction (5.1.1) may be written as

\[
\int_0^s (1 - F(x)) \, dx = \lambda, 
\]

which is an equation for \(s\).

We return to the optimization problem. Assume that the preferences of the individual are preserved by a function \(U(F)\). For the reader who skipped Section 4 on non-linear criteria, we start with the EUM case where

\[
U(F) = \int_0^\infty u(x) dF(x), 
\]  

(5.1.5)

and \(u\) is a non-decreasing utility function. In the end of the section, we consider the non-linear case.

Denote by \(F_r(x)\) the distribution function of the r.v.

\[
X_r = w - G - \xi + r(\xi), 
\]

the wealth of the individual under the choice of a payment function \(r(x)\). Our goal is to find a function \(r\) for which \(F_r(x)\) is the best. More rigorously, we are looking for a function \(r^*\) which maximizes the function \(Q(r) = U(F_r)\).

It is remarkable that under certain conditions, the optimal payment function does not depend on the particular form of the utility function \(u(x)\) and on the premium \(G\). More precisely, \(r^*\) has the type (5.1.3) with the deductible \(d\) specified in (5.1.4).
5. Optimal Payment from The Standpoint of The Insured

For a fixed \( \lambda \in (0, m] \), consider the set of all function \( r(x) \) satisfying (5.1.2), that is, the set

\[
\mathcal{R}_\lambda = \{ r(x) : E\{r(\xi)\} = \lambda \}.
\]

The theorem below belongs to K. Arrow; see, e.g., [4], [5].

**Theorem 17** Let \( u(x) \) in (5.1.5) be concave, and \( r^*(x) = r_d(x) \), where \( d \) satisfies (5.1.4). Then for any function \( r(x) \) from \( \mathcal{R}_\lambda \).

\[
Q(r) \leq Q(r^*).
\]

Thus,

The optimal payment is the same for ANY concave utility function.

**Proof of Theorem 17.** As we know from Calculus, the l.-h.s. of (5.1.4), as a function of \( d \), is continuous. By the general formula (0.2.2.2), this function is equal to \( m \) at \( d = 0 \), and it converges to zero as \( d \to \infty \). Hence, for any \( \lambda \in (0, m] \), there exists a number \( d \) for which the l.-h.s. of (5.1.4), that is, \( E\{r_d(\xi)\} \), is equal to \( \lambda \).

In particular, this means that \( r_d \in \mathcal{R}_\lambda \).

(As a matter of fact, since \( m > 0 \), the number \( d \) for which (5.1.4) is true, is unique. Indeed, let \( x_0 \) be the smallest point at which \( F(x) = 1 \). If \( F(x) < 1 \) for all \( x \)'s, we set \( x_0 = \infty \). Because \( m > 0 \), the point \( x_0 > 0 \). Then \( F(x) < 1 \) for all \( x < x_0 \), and \( 1 - F(x) > 0 \) on \([0, x_0)\). Consequently, the l.-h.s. of (5.1.4) is strictly decreasing for \( d \in [0, x_0) \).

Now, let us fix \( r \in \mathcal{R}_\lambda \) and set \( F(x) = F(r)(x) \). Denote by \( F^*(x) \) the d.f. of the r.v. \( X^* = X_{(r^*)} \), i.e., the final wealth in the case (5.1.3) with \( d \) satisfying (5.1.4). Set \( a = w - G - d \), and \( b = w - G \).

\[
F(x) \leq F^*(x) \text{ for } x \geq a.
\]

A typical picture is given in Fig.18.

We need also the following relations. Because \( E\{r(\xi)\} = E\{r^*(\xi)\} \), we have \( E\{X_{(r)}\} = E\{X^*\} \). Therefore, integrating by parts, we get that

\[
\int_{-\infty}^{b} [F^*(z) - F(z)]dz = \int_{-\infty}^{b} zd[F(z) - F^*(z)] = E\{r(\xi)\} - E\{r^*(\xi)\} = 0.
\]
1. COMPARISON OF RANDOM VARIABLES

(How to prove that \( \lim_{z \to -\infty} z[F(z) - F^*(z)] = 0 \) is shown in Section 0.2.6.) From (5.1.8) it follows that \( \int_{-\infty}^{x} [F^*(z) - F(z)] \, dz = -\int_{x}^{b} [F^*(z) - F(z)] \, dz \) for \( x \leq b \). Then, in view of (5.1.7), for \( a \leq x \leq b \),

\[
\int_{-\infty}^{x} [F^*(z) - F(z)] \, dz \leq 0.
\] (5.1.9)

On the other hand, since \( F^*(x) = 0 \) for \( x < a \), inequality (5.1.9) is true for \( x < a \) also, and hence it is true for all \( x \leq b \). Note also that, since \( F^*(z) - F(z) = 1 - 1 = 0 \) for \( x > b \), eventually (5.1.9) is true for all \( x \)'s.

Let us proceed to a direct proof. Assuming for simplicity that \( u \) is sufficiently smooth, and integrating by parts, we have

\[
U(F^*) - U(F) = \int_{-\infty}^{b} u(x)d[F^*(x) - F(x)] = \int_{-\infty}^{b} [F(x) - F^*(x)]u'(x)dx
\]

\[
= \int_{-\infty}^{b} u'(x) \left( \int_{-\infty}^{x} [F(z) - F^*(z)] \, dz \right). \tag{5.1.10}
\]

Making use of (5.1.8), we integrate by parts in (5.1.10) one more time, which leads to

\[
U(F^*) - U(F) = \int_{-\infty}^{b} \left( \int_{-\infty}^{x} [F(z) - G(z)] \, dz \right) u''(x)dx. \tag{5.1.11}
\]

Since \( u''(x) \leq 0 \), (5.1.11) and (5.1.9) implies that \( U(F^*) \geq U(F) \). ■

5.2 A generalization

Now, let us realize that (5.1.9) means that \( F^* \) dominates \( F \) in the sense of the second stochastic dominance (SSD); see Section 3.5.3. Consequently, we have proved above the following much more general theorem.

**Theorem 18** The payment function \( r^*(x) \) is optimal for any preference order \( \succeq \) which is monotone with respect to the SSD.

**Example 1.** Consider the case of implicit utility described in Section 4.3. Let \( v(x, y) \) be a given function as it is defined in this section, and let the preference order be preserved by the certainty equivalent \( c(F) \) defined in (4.3.3). Assume that \( v(x, y) \) is concave with respect to \( x \) and non-increasing in \( y \). We show that the corresponding preference order is monotone with respect to the SSD.

To simplify calculations, assume that \( v(x, y) \) is sufficiently smooth. For two distributions, \( F \) and \( G \), similarly to (5.1.10)-(5.1.11), we have

\[
\int_{-\infty}^{\infty} v(x, c)d[F(x) - G(x)] = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{x} [F(z) - G(z)] \, dz \right) v''_{xx}(x, c)dx,
\]
Find the Solve Exercises 2a,c for the case of the TailVaR criterion. Let r.v.'s It is known that, if \( v'' \leq 0 \) and \( F \) dominates \( G \) in the sense of the SSD, then

\[
\int_0^\infty v(x,c) d[F(x) - G(x)] \geq 0
\]

(5.2.1) for all \( c \). Let \( c = c(F) \). Then \( \int_0^\infty v(x,c(F)) dF(x) = 0 \), and hence \( \int_0^\infty v(x,c(F)) dG(x) \leq 0 \). On the other hand, by definition, \( \int_0^\infty v(x,c(G)) dG(x) = 0 \). Since \( v(x,y) \) is non-increasing in \( y \), from this it easily follows that \( c(G) \leq c(F) \), and hence \( F \succcurlyeq G \). □

The idea of the above proof of Theorem 17 and Theorem 18 belong to C.Gollier and H.Schlesinger; see [45] and also [116]. Other generalizations of Arrow’s theorem (mostly using optimization technique) may be found, e.g., in papers by E.Karni [68], A.Raviv [102], and I.Zilcha and S.H.Chew [139]. See also references therein.

6 EXERCISES

Sections 1 and 2

1. Find the 0.2-quantile of a r.v. \( X \) taking values 0, 3, 7 with probabilities 0.1, 0.3, 0.6, respectively.

2. This exercise concerns the VaR criterion with a parameter \( \gamma \). R.v.’s \( X \) with or without indices correspond to an income.

   (a) Let \( X_1 \) take on values 0, 1, 2, 3 with probabilities 0.1, 0.3, 0.4, 0.2, respectively, and \( X_2 \) take on the same values with probabilities 0.1, 0.4, 0.2, 0.3. Find all \( \gamma \)'s for which \( X_1 \succcurlyeq X_2 \).

   (b) Let r.v.’s \( X_1 = 1 - \xi_1 \), \( X_2 = 1 - \xi_2 \), where the loss \( \xi_1 \) is uniform on \([0, 1]\), and \( \xi_2 \) is exponential with \( E\{\xi_2\} = m \). When is the relation \( X_1 \succcurlyeq X_2 \) true for all \( \gamma \)? Let \( m = 1/2 \). Find all \( \gamma \)'s for which \( X_1 \succcurlyeq X_2 \).

   (c) In Exercise 2b, let \( \xi_1 \) be uniform on \([0, 3]\), and let \( \xi_2 \) be uniform on \([1, 2]\). Find all \( \gamma \)'s for which \( X_1 \succcurlyeq X_2 \).

3. Solve Exercises 2a,c for the case of the TailVaR criterion.

4. It is known that, if \( X_1, \ldots, X_n \) are independent exponential r.v. with unit means, then \( S_n = X_1 + \ldots + X_n \) has the \( \Gamma \)-distribution with parameters \((1,n)\). We will prove this fact in Section 2.2.2.

Let \( n = 10 \), and let r.v.’s \( X_1, \ldots, X_{10} \) be defined as above. Suppose that \( Y_i = 1.1 \cdot X_i \) for \( i = 1, \ldots, 10 \), represent the returns for the investments into 10 independent assets. (For the notion of “return” see Example 1.2.2-4.) Thus, since \( E\{X_i\} = 1 \), an investment into each asset gives 10% profit on the average. Proceeding from the VaR criterion, figure out what is more profitable: to invest $10 into one asset, or split it between 10 assets. You are recommended to use Excel (or another software); the corresponding
1. COMPARISON OF RANDOM VARIABLES

command in Excel for quantiles is =GAMMAINV\((p, v, 1/a)\) where \(a\) is the scale parameter.

5. Consider two assets. The investment of a unit of money into the \(i\)th asset leads to an income of \(1 + \xi_i\) units; \(i = 1, 2\). Assume that \(\xi\)'s have a joint normal distribution, \(E\{\xi_i\} = 0, \text{Corr}\{\xi_1, \xi_2\} = \rho\). Suppose that you invest some money into each asset, \(K_i\) is the profit of the investment into the \(i\)th asset, and \(K\) is the total profit. Prove that

\[ q_\gamma(K) = \sqrt{q_\gamma^2(K_1) + q_\gamma^2(K_2)} + 2\rho q_\gamma(K_1)q_\gamma(K_2). \]

The last formula is relevant to the JP Morgan RiskMetics\(^{TM}\) methodology; see, e.g. [61]. Some references may be found also in http://www.riskmetrics.com and http://www.gloriamundi.org.

6. (a) Proceeding from (1.2.4) and using integration by parts, prove that

\[ E\{X | X \leq t\} = \frac{1}{F(t)} \left( tF(t) - \int_{-\infty}^{t} F(x)dx \right) = t - \frac{1}{F(t)} \int_{-\infty}^{t} F(x)dx. \]

where \(F(x)\) is the d.f. of \(X\).

(b) Show that, if \(F(t)\) is continuous and \(q_\gamma\) is the \(\gamma\)-quantile of \(X\), then

\[ V_{\text{tail}}(X) = E\{X | X \leq q_\gamma\} = q_\gamma - \frac{1}{\gamma} \int_{-\infty}^{q_\gamma} F(x)dx. \quad (6.1) \]

Show that it may be false if \(X\) is not a continuous r.v.

(c) Show that \(V_{\text{tail}}(X)\) is monotone in the class of continuous r.v.'s. (Advice: You can either use (6.1) or observe that in the continuity case, the conditional d.f. \(P(X \leq x | X \leq q_\gamma) = \frac{1}{\gamma} F(x)\) for \(x \leq q_\gamma\) and = 1 otherwise.)

(d) Sketch a typical graph of a continuous \(F(x)\). Consider \(\gamma\) for which \(q_\gamma < 0\) and point out the region in the graph whose area equals \(\gamma |V_{\text{tail}}(X)|\).

7. Take real data on the daily stock prices for the stocks of two companies for one year from, say, http://finance.yahoo.com or another similar site. For different values of \(\gamma\), compare the performance of the companies using the VaR and the TailVaR criteria. (The absolute values of the prices should have no effect on results. The analysis should be based on returns, that is, on the ratios of the prices on the current and the previous days. For the notion of “return” see Example 1.2.2-4.) Estimate the mean return for each company. Try to characterize and compare the performance of the companies, taking into account the all three characteristics mentioned.

8. If we interpret \(X\) as an income, then the r.v. \(\tilde{X} = -X\) may be interpreted as a loss. Considering only r.v.'s whose d.f.'s are strictly increasing, do the following.

(a) Prove that \(q_\gamma(X) = -q_{1-\gamma}(\tilde{X})\). Show it graphically.
Consider, instead of (1.2.3), the function \( \tilde{V}(\hat{X}; s) = E\{\hat{X} | \hat{X} \geq s\} \), the mean value of the loss given that it has exceeded a level \( s \). Show that \( \tilde{V}(\hat{X}; s) = -V(X; -s) \) for any \( s \), and \( \tilde{V}(\hat{X}; s) = |V(X; -s)| \) for all \( s \geq 0 \). Give a heuristic explanation.

(c) Consider the criterion preserved by the risk measure \( \tilde{V}_{\text{tail}}(\hat{X}) = E\{\hat{X} | \hat{X} \geq q_{1-\gamma}(\hat{X})\} \).

Show that \( X \gtrless Y \Leftrightarrow \hat{X} \gtrless \hat{Y} \Leftrightarrow \tilde{V}_{\text{tail}}(\hat{X}) \leq \tilde{V}_{\text{tail}}(\hat{Y}) \).

9. For some \( V(X) \) and a family \( \mathcal{P} \), suppose that (1.3.2) is true. Show that the monotonicity property and Properties I-III from Section 1.3 are fulfilled. (The converse assertion is more difficult to prove, but the sufficiency of the representation (1.3.2) is understandable. Recall that \( \min_{x}\{f(x) + g(x)\} \geq \min_{x}f(x) + \min_{x}g(x) \).)

10. Make sure that you indeed understand why, if \( E\{X\} = m \) and \( \text{Var}\{X\} = \sigma^2 \neq 0 \), then for the normalized r.v. \( X^* = (X - m)/\sigma \), we have \( E\{X^*\} = 0 \), \( \text{Var}\{X^*\} = 1 \).

Section 3

11. Show that, if \( u(x) \) is strictly increasing, then the rule (3.1.2) is strictly monotone. (Advice: Consider \( E\{u(X)\} - E\{u(Y)\} = E\{u(X) - u(Y)\} \).)

12. Graph all utility functions from Section 3.1.3.

13. Consider a r.v. \( X \) such that \( P(X > x) = x^{-1} \) for \( x \geq 1 \). This is a particular case of the Pareto distribution which we will consider in Section 2.1.1.3 in detail. Does \( X \) have a finite expected value? Find the certainty equivalent of \( X \) for \( u(x) = \sqrt{x} \).

14. Write formulas for the certainty equivalents for the cases 2, 3, 5 from Section 3.1.3.

15. Let \( X \) be exponential, and \( u(x) = -e^{-bx} \) (see Section 3.1.3). Show that for the certainty equivalent \( c(X) \), we have \( c(X) = \frac{E(X)}{E[X]} \to 0 \) as \( E\{X\} \to 0 \), that is, \( c(X) \approx E\{X\} \) if \( E\{X\} \) is small. Interpret it.

16. Repeat calculations of Example 3.2-1 for the case when \( \xi = 0 \) with probability 0.9, and is uniformly distributed on \([0, 1]\) with probability 0.1.

17. An EUM customer of an insurance company has a total wealth of 100 (in some units) and is facing a random loss \( \xi \) distributed as follows: \( P(\xi = 0) = 0.9, P(\xi = 50) = 0.05, P(\xi = 100) = 0.05 \). (a) Let the utility function of the customer be \( u(x) = x - 0.005x^2 \) for \( 0 \leq x \leq 100 \). Graph it. Is the customer a risk aveter?

(b) What would you say in the case \( u(x) = x + 0.005x^2 \)

(c) For the case 17a, find the maximal premium the customer would be willing to pay to insure his wealth against the loss mentioned. First, set the equation clearly and explain it, then solve. Is the premium you found greater or less than \( E\{X\} \)? Might you predict it from the beginning?
1. COMPARISON OF RANDOM VARIABLES

(d) Find the minimal premium which an insurance company would accept to cover the risk mentioned, if the company’s preferences are characterized by the utility function \( u_1(x) = \sqrt{x} \), and the company takes 300 as an initial wealth. Is the premium you found greater or less than \( E(\xi) \). Might you predict it?

(e) Solve Exercise 17d for the case when the r.v. \( \xi \) is uniformly distributed on \([0, 100]\).

(f) Solve Exercise 17c for \( u(w) = 200x - x^2 + 349 \). (Advice: Look at this function attentively before starting calculations.)

18. Find the maximal premium the customer would be willing to pay to insure half of the loss in the situations of Example 3.2-1.

19. Give an explicit example when the maximal acceptable premium for a customer does depend on the initial wealth.

20. Take real data on the daily stock prices for the stocks of two companies for one year from, say, http://finance.yahoo.com or another similar site. Considering a particular utility function, for instance, \( u(x) = -e^{-\beta x} \) for some \( \beta \), determine which company is better for an EU maximizer with this utility function. (Advice: Look at the comment in Exercise 7. To estimate the expected value \( E\{u(X)\} \), where \( X \) is a random return, we can use the usual estimate \( \frac{1}{n}[E\{u(X_1)\} + \ldots + E\{u(X_n)\}] \), where \( X_1, \ldots, X_n \) is the time series from your data. Excel is convenient for such calculations.) Add to your analysis the characteristics considered in Example 7. Try to describe the performance of the companies, taking into account all characteristics you computed.

21. Is Condition Z from Section 3.4.1 a requirement on (a) distribution functions, or (b) random variables, or (c) the preference order of the person under consideration?

22. Is Condition Z from Section 3.4.1 based on the concept of expected utility?

23. Is it true that for an expected utility maximizer to be a risk averter, his/her utility function should have a negative second derivative? Give an example. (Advice: Look up the definition of concavity.)

24. Let \( u(x) = x \) for \( x \in [0, 1] \), and \( u(x) = \frac{1}{2} + \frac{1}{2}x \) for \( x \geq 1 \). Is an EU maximizer with this utility function a risk averter?

25. Check for risk aversion the criteria with utility functions from Section 3.1.3.

26. Let the utility function of a person be \( u(x) = e^{ax} \), \( a > 0 \). Graph it. Is the person a risk averter or a risk lover? Show that in this case, the comparison of risky alternatives does not depend on the initial wealth.

27. Let \( X \) be exponential, and \( u(x) = -e^{-\beta x} \) (see Section 3.1.3). Show that the certainty equivalent \( c(X) \to 0 \), as \( \beta \to \infty \). Interpret it in terms of risk aversion.

28. In Examples 3.1.3-1 and 2, compare the expected values and the certainty equivalents for different values of \( \alpha \) including the case \( \alpha \to 0 \). Interpret results.
29. Let \( u_1(x) \) and \( u_2(x) \) be John’s and Mary’s utility functions, respectively.

(a) How do John’s and Mary’s preferences differ if \( u_1(x) = 2u_2(x) + 3 \)?
(b) Let \( u_1(x) = \sqrt{x} \) and \( u_2(x) = x^{1/3} \). Who is more averse to risk? In what sense?
(c) Let \( u_1(x) = -1/\sqrt{x} \) and \( u_2(x) = -1/x^{1/3} \). Who is more afraid of being ruined (having a zero income)? Who is more averse to risk?

30. Let Michael be an EUM with the utility function \( u(x) = -\exp\{-\beta x\} \) and \( \beta = 0.001 \). (For this value of \( \beta \), the values of expected utility in this problem will be in a natural scale.) Michael compares stocks of two mutual funds. The today price for each is $100 per share. Michael believes that in a year the price for the first stock will be on the average either 10% higher or 10% lower with equal probabilities, while for the second stock 10% up or down should be replaced by a slightly higher figure, approximately, 11%. (a) Which mutual fund is “better” for Michael? Do we need to calculate something? (b) Now, assume that the second fund invites all people who buy 100 shares to a dinner valued at $k. Which \( k \) would make difference?

31. Provide calculations to obtain (3.2.6).

32. (a) It is known that for any positive r.v. \( X \), the function \( n(s) = (E\{X^s\})^{1/s} \) is non-decreasing in \( s \). Using this fact, prove that the r.h.s. of (3.2.5) is non-decreasing in \( \beta \). (Advice: set \( \xi = \ln(X) \) and write \( X = e^\xi \).)
(b) Using Jensen’s inequality, prove that indeed \( n(s) \) above is non-decreasing. (Advice: We should prove that, if \( s < t \), then \( n(s) \leq n(t) \), which is equivalent to \( E\{X^s\} \leq (E\{X^t\})^{s/t} \). Write \( X^t = (X^s)^{s/t} = u(X^s) \), where \( u(x) = x^{s/t} \). Figure out whether \( u(x) \) is concave for \( s < t \).)

33. Write the risk aversion function for the utility function (3.1.3) setting \( u_1(x) = \ln x \).

34. Let \( R(x) \) and \( R_a(x) \) be absolute and relative risk aversion functions for a utility function \( u(x) \). Find the corresponding risk aversion functions for \( u^a(x) = bu(x) + a \), where \( a, b \) are constants. Interpret the result in the light of the first property from Section 3.1.2.

35. Find the absolute risk aversion for \( u(x) = e^{\beta x} \), \( \beta > 0 \). Interpret the fact that the risk aversion characteristic is negative.

36. Prove that, up to linear transformation, only exponential utility functions have a constant absolute risk aversion, and only power utility functions and the logarithm have a constant relative risk aversion. (Advice: You should consider equations \( u''(x) = cu' \), and \( xu''(x) = cu'(x) \).)

37. Consider two r.v.’s both taking values 1, 2, 3, 4. For the first r.v. the respective probabilities are 0.1, 0.2, 0.5, 0.2, for the second, 0.1, 0.3, 0.3, 0.3. Which r.v. is better in the EUM-risk-aversion case? Justify the answer.
1. COMPARISON OF RANDOM VARIABLES

38. Consider a r.v. taking values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$, respectively. Let the $x_i$’s be equally spaced, that is, $x_{i+1} - x_i$ equals the same number for all $i$. Assume that for a particular $i$, we replaced probabilities $p_{i-1}, p_i, p_{i+1}$ by probabilities $p_{i-1} + \Delta, p_i - 2\Delta, p_{i+1} + \Delta$ where a positive $\Delta \leq p_i/2$. Has the new distribution become worse or better in the EUM-risk-aversion case? Justify the answer.

39. John agreed with the following strange payment for a job done. A regular coin will be tossed. In the case of a head, John will be paid $200. In the case of a tail, a die will be rolled. If the die shows one or two, John will be paid $300, otherwise – nothing. Determine the random payment considering the mixture of the distributions of r.v.’s $X_1 = 200$, and $X_2 = 300$ or $0$ with probabilities $1/3$ and $2/3$, respectively.

40. (a) Is the FSD rule defined in Section 3.5.2 a requirement on distributions or on the preference order?

(b) Does the FSD rule concern only the EUM criterion or all preference orders?

(c) Answer the same questions regarding the SSD.

41. Which criteria from Section 1.2 satisfy the FSD rule?

42. Show rigorously that in the general $\Delta$-scheme from Section 3.5.5, (a) equivalence sets are planes or hyper-planes; (b) to determine completely the preference order, it suffices to determine $n - 1$ equivalent points.

43. Fig.’s 19abcd depict equivalence curves in the $\Delta$-scheme for the distributions of r.v.’s taking three values. Which figures correspond to EUM, and which do not?

44. Consider the distributions of all random variables taking only values 0, 10, 20. We identify any such a distribution $F$ with the vector of probabilities $(p_1, p_2, p_3)$. Let $F_1 = (0.1, 0.5, 0.4), F_2 = (0.2, 0.2, 0.6)$, and $F_3 = (0.1, 0.8, 0.1)$.

(a) Mark all corresponding points in the $(p_1, p_3)$-plane picture.

(b) Find a distribution $F_4$ such that the assertion $F_1 \simeq F_2, F_3 \simeq F_4$ would not contradict the independence axiom.
6. Exercises

45. Consider the probability distributions of ALL random variables taking values 0, 10, 20, and 40.

   (a) Identify such distributions with points in $\mathbb{R}^3$. What region in $\mathbb{R}^3$ do we consider in this case? To what distributions do boundary points of this region correspond?

   (b) Assume that you are an EU maximizer. Fix a distribution $F_0$, and consider ALL distributions $F \simeq F_0$. Where do all points corresponding to these distributions lie?

   (c) Assume that the following three distributions are equivalent:

   $$(0.2, 0.3, 0.1, 0.4), \ (0.3, 0, 0, 0.7), \ (0, 0.6, 0.4, 0).$$

   Find one more distribution equivalent to the mentioned.

46. Consider the distributions from Exercise 44. You are an EUM, and your utility function is increasing. What is better, $F_1$ or $F_3$? Find $G_1 = \frac{1}{2}(F_1 + F_2)$, $G_2 = \frac{1}{2}(F_3 + F_2)$. Mark both points in the $\Delta$-scheme picture. Which point is better?

47. Consider an EUM in the situation of Exercise 44. Assume that distributions $(0.1, 0.5, 0.4)$ and $(0.2, 0.3, 0.5)$ are equivalent. Find one more distribution which is equivalent to these distribution. Find all such distributions.

48. You know that Fred is an EUM, and for him “the larger, the better”. You ask Fred to compare the following two random variables: both take on values 0, 20, 40; the first – with probabilities $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, respectively, and the second – with probabilities $(\frac{1}{3}, \frac{1}{6}, \frac{1}{2})$. It turns out that Fred considers these distributions equivalent for him.

   (a) Is this information enough in order to predict the result of comparison by Fred of ANY two random variables with the values mentioned?

   (b) Which distribution would Fred prefer: $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ or $(\frac{3}{5}, \frac{1}{10}, \frac{3}{10})$?

   (c) Is Fred risk lover? If yes, justify the answer. If no, does this mean that he is a risk averter?

Section 4

49. Consider Fig.’s 19abcd from Exercise 43 depicting equivalence curves in the $\Delta$-scheme for the distributions of r.v.’s taking three values. Which figures correspond to axioms you know. Identify these axioms.

50. Consider distributions $F_1$, $F_2$, and $F_3$ from Exercise 44. Find a non-trivial and meaningful example of a distribution $F_4$ such that the relations $F_1 \simeq F_2$, $F_3 \simeq F_4$ would not contradict the betweenness axiom.

51. Consider probability distribution of the income taking values 0, 20, and 40 with probabilities $p_1, p_2, p_3$, respectively. Specify all points in the plane $(p_1, p_3)$, which correspond to the distributions under consideration.
Assume that for Jane equivalence curves on the diagram mentioned are curves given by the formula
\[ p_3 = c + e^{p_1}, \quad \text{where} \quad -e \leq c \leq 0. \]

(a) Graph these curves, and realize why we need the condition \(-e \leq c \leq 0\).
(b) Is Jane a EUM? Do her preferences meet the Betweenness Axiom?
(c) Since we consider only the narrow class of distributions concentrated on the set \(0, 20, 40\), we cannot figure out whether Jane is a risk lover or not. Nevertheless, reasoning heuristically, give an argument that we should not expect that Jane is a risk lover.
(d) Jane follows the rule “the larger the better”. Which distribution is better for her: \((0.1, 0.5, 0.4)\) or \((0.2, 0.2, 0.6)\) ?

52. In Example 4.3.1-3, find the certainty equivalent of a r.v. \(X = d > 0\) or \(0\) with probabilities \(p\) and \(q\), respectively. Analyze and interpret the case \(p\) close to one.

53. Let \(g(s)\) defined in Section 4.3 equal \(1 - e^{-s}\) for \(s > 0\), and \(s\) for \(s \leq 0\). Graph \(g(s)\). Let \(c\) be the certainty equivalent of a r.v. \(X\). Is it true that \(c \leq E\{X\}\)? Estimate the certainty equivalent for \(X\) equal to \(1\) or \(0\) with equal probabilities.

54. Show that all lines (4.2.8) intersect at one point. (Advice: Consider two lines, \(l_1\) and \(l_2\), corresponding to two different values \(d_1\) and \(d_2\). They may intersect only at a point in \(L_\infty\) where \(W(F)\) is not defined since otherwise at this point \(W(F)\) would have taken two different values.)

55. Show that \(W(F^{(\alpha)})(F)\) in (4.2.6) is not equal, in general, to \(\alpha W(F_1) + (1 - \alpha)W(F_2)\). (Hint: The answer does not require long calculations. Consider, for example \(\alpha = 0.5\) and the case \(f_{-\infty}^\infty w(x) dF_2(x) = 2 \int_{-\infty}^\infty w(x) dF_1(x)\).)

56. In the situation of Example 4.3.1-2, find the maximal premium \(G_{\max}\) assuming \(w = d\) and the loss to have the same distribution as \(X\) in this example.

57. Let an investor follow the RDEU criterion with \(\Psi(p) = 1 - (1 - p)^\beta\). In other words, \(\Psi\) behaves as a power function for \(p\) close to one, that is, for \(p\)'s corresponding to large values of r.v.’s. Let \(X\) be an exponential r.v. with a distribution \(F\). Show that in this case, the transformation of \(F\) corresponds to dividing \(X\) by \(\beta\).

58. Find \(G_{\max}\) in the situation of Example 4.4.2-1 in the case when \(P(\xi = 0) = 0.9\), \(P(\xi = 1) = 0.1\).

Section 5

59. Find the deductible \(d\) in (5.1.4) for \(\lambda = m/2\) for two cases: (a) \(\xi\) is exponential, \(E\{\xi\} = m\); (b) \(\xi\) is uniform on \([0, 2m]\). (So, the expected values are the same.) Compare results. Explain the difference from a heuristic point of view.

60. Which regions in Fig.18 have the same area?