Data Dispersion in Economics (I) --- Possibility of Restrictions

Alexander Harin

Modern University for the Humanities
109029, Moscow, Nizgorodskaya Street, Building 32
Tel: +7-915-400-9879 E-mail: aaharin@gmail.com

Abstract: The article is devoted to data dispersion in economics. The state of the art and the problems of the utility theory are briefly reviewed. Examples of typical paradoxes, which are analyzed by the utility theory, are given. Analogous paradoxes may concern problems such as the underweighting of high and the overweighting of low probabilities, risk aversion, the equity premium puzzle, etc. Introductory examples of the restrictions near the boundaries of finite numerical segments and of the probability scale are considered and analyzed.

JEL Classifications: C02, C44, D81, G22
Keywords: Economics, Utility, Probability, Uncertainty, Decisions, Restrictions

1. Introduction

1.1 Problems and Researches

This review is devoted to the well-known fundamental problems of utility theory. The St. Petersburg paradox, the Allais paradox and other problems were raised and discussed in the celebrated works Bernoulli (1738) and Allais (1953) and in a wealth of subsequent works. Kahneman and Thaler (2006) pointed out these problems have been still unexplained.

The review summarizes the particular results of more than 10-years researches on a new approach devoted to the reviewed item, those were presented in Russian journals and on scholarly conferences (see, e.g., Harin (2003, 2007, 2009a-2009b, 2010a-2010d, 2011a-2011b, 2012)), improves and refines the theorems.

1.2 The Idea of the New Approach

The new idea has a simple appearance but a considerable mathematical meaning:

In the presence of data dispersion, the restrictions, the forbidden zones can exist for the mean values near the boundaries of finite numerical segments and near the boundaries of the probability scale.

These restrictions can bias the results of experiments and can partially explain them.

1.3 Results and Literature

This bias can partially explain the abovementioned problems of utility theory and economics. Taking into account the bias may provide a multi-purpose means to partially explain some problems of utility theory. This bias may be taken into account in present utility theory models to improve their results.

The review represents the progress achieved due to the proofs of the theorems of existence of the restrictions near the boundaries of finite numerical segments and of the probability scale (Harin 2010a, 2010d, 2011a, 2011b) and their applications (Harin 2010b, 2010c, 2012) which were proved, developed and used in 2010-2012. These theorems prove and support the hypotheses and
assumptions that were used in the works of the topic from the first (Harin 2003) (in Russian) and (Harin 2005) (in English) to (Harin 2009b). The theorems may open a new direction to facilitate the explanation of the discussed problems.

The analysis of scolar literature based on (Harin 2005-2011, Harin 2007-2011a and Harin 2007-2011b) shows these theorems and the way of explanation of discussed problems are the original ones and there are no such new ideas both before and since (Kahneman and Thaler 2006).

2. The Old Problems of the Utility Theory

2.1 History

A man is the key subject of economic and economic theory. Decisions of a man are fundamental operations of them. Utility theory, as a branch of the economic theory, is specially devoted to the research of decisions of a man.

Bernoulli (1738) had given rise to researches of problems of the utility theory. Von Neumann and Morgenstern (1947) had provided promises of feasibility of correct and, naturally, rational fundamentals of the economic theory. But these promises were broken by Allais (1953). Other later works of various authors had shown that real man’s decisions are undoubtedly inconsistent with rational models. Moreover, these decisions are inconsistent with the probability theory. As pointed by Kahneman and Thaler (2006), these inconsistensies are still not overcomed by the economic theory.

There are numerous works devoted to this topic (see, e.g., Ellsberg (1961), Kahneman and Tversky (1979), Schoemaker (1982), Tversky and Wakker (1995), Prelec (1998), Gilboa (2010), Robson and Samuelson (2011)). A very brief review of some of these well-known problems will be given below.

2.2 Risky and Guaranteed Outcomes

Let us consider simple typical examples of these problems.

Above all, let us define an underestimation of probable outcomes as the negative deviation from the probability theory and an overestimation of probable outcomes as the positive one.

2.2.1 Gain at high probabilities

Suppose a scholar, who performs the experiment, offers you a choice of two outcomes:
A) A guaranteed gain, prize of $99 with the probability of 1 (or 100%);

or

B) A probable gain of $100 with the probability of 0.99 (or 99%), or nothing with the probability of 0.01 (or 1%).

(For the experiment accuracy, both $99 and $100 should be in $1 banknotes, i.e. 99 and 100 banknotes of $1)

The mean values for the guaranteed and probable outcomes are

$99\times100\% = \$99,$

$100\times99\% = \$99,$

Here, $99 = \$99.$

We evidently see that the mean values to win the probable gain and to get the guaranteed gain are precisely equal to each other.

But the well-determined experimental fact is: in similar experiments for gains at high probabilities the overwhelming majority of people choose the guaranteed gain instead of the
probable one (see, e.g., Tversky and Wakker (1995)). People underestimate probable outcomes and do not like risk. The deviation from the probability theory is negative.

The possible well-known explanation for gains at high probabilities by means of risk aversion cannot supply a uniform explanation, e.g., for this case and for losses at high probabilities and for gains at low probabilities (see below).

2.2.2 Losses at high probabilities

Suppose a scholar offers you a choice of two outcomes:
A) A guaranteed loss of -$99 with the probability 100%,

or

B) A probable loss of -$100 with the probability 99%, or no loss with the probability 1%.

The mean values for the guaranteed and probable outcomes are

\[-99 \times 100\% = -99,\]
\[-100 \times 99\% = -99,\]

Here, \(-99 = -99\).

We evidently see both the mean of the guaranteed loss and the mean of the probable loss are equal to -$99 and are precisely equal to each other.

But the well-determined experimental fact is: in similar experiments for losses at high probabilities the overwhelming majority of people choose the probable loss instead of the guaranteed one (see, e.g., Tversky and Wakker (1995)). People overestimate probable outcomes and like risk. The deviation from the probability theory is positive.

2.2.3 Gains at low probabilities

Suppose a scholar offers you a choice of two outcomes:
A) A guaranteed gain of $1 with the probability 100%,

or

B) A probable gain of $100 with the probability 1%, or nothing with the probability 99%.

The mean values for the guaranteed and probable outcomes are

\[1 \times 100\% = 1,\]
\[100 \times 1\% = 1,\]

Here, $1 = $1.

We evidently see both the mean of the guaranteed gain and the mean of the probable gain are $1 and are precisely equal to each other.

But the well-determined experimental fact is: in similar experiments for gains at low probabilities the overwhelming majority of people choose the probable gain instead of the guaranteed one (see, e.g., Tversky and Wakker (1995)). People overestimate probable outcomes and like risk. The deviation from the probability theory is positive.

2.2.4 Contrary choices

So the common choices are, in some sense, contrary to each other: In the case of gains at high probabilities - people do not like risk. In the cases of gains at low probabilities and of losses at high probabilities - people like risk. The choices are contrary. Similar scenarios give rise to a number of unexplained fundamental problems.

The essence of these problems is the comparison between risky and guaranteed outcomes.
So, the trouble in utility theory is the real personal attitude to risk cannot be captured by expected utility functional. The predominating opinions are this trouble is due to lack of describing power of expected utility theory and due to the nature of people is not fully rational. Nevertheless, a complementary opinion of some authoritative scholars exists also.

The opinion is the additional general factors should be taken into account as well.

2.3 Noise, Imprecision, Dispersion of Data

The final statement of Hey and Orme (1994) was "... we are tempted to conclude by saying that our study indicates that behavior can be reasonably well modeled (to what might be termed a "reasonable approximation") as "expected utility plus noise." Perhaps we should now spend some time thinking about the noise, rather than about even more alternatives to expected utility?". Butler and Loomes (2007) stated "... any successful descriptive theory of choice and valuation will need to allow in some way for the imprecision ...". See also Hey (2005) and Chay, McEwan and Urquiola (2005).

Noise and imprecision are some of reasons those can lead to a dispersion of data. So, more generally, the above statements may be reformulated as an appeal to pay more attention to a dispersion of data.

2.4 Zones near the Boundaries

Let us consider the deviations from the probability theory in the experiments via the whole range of the probability scale.

According to a lot of experiments (see, e.g., Loewenstein and Prelec (1992), Tversky and Wakker (1995) and Prelec (1998):

For gains: the deviations are negative at high probabilities and positive at low probabilities.

For losses: the deviations are positive at high probabilities and negative at low probabilities.

So, at smooth plots, in the middle of the probability scale the deviations change the sign and, therefore, pass through zero. Hence, in the middle of the probability scale, the deviations are minimal (in the absolute value) and near the boundaries of the probability scale, the deviations are maximal (in the absolute value).

Hence, the zones near the boundaries are more preferable for research than the middle of the probability scale. So, a next step of the consideration is to pay more attention to zones near the boundaries of the probability scale.

So, the subsequent consideration will be confined at:
1) Comparisons of risky and guaranteed outcomes.
2) Risky outcomes near the boundaries of the probability scale.
3) We will consider the items 1) and 2) in the presence of non-zero dispersion of data.

3. Proofs of Possibility of Existence of Restrictions

It is enough to adduce only one example of existence of a phenomenon to prove the statement this phenomenon can exist. In other words, it is enough to adduce only one counterexample of existence of a phenomenon to disprove a statement this phenomenon cannot exist.

For purposes of this article, let us define a restriction for a mean value as a zone in the scale of possible locations of the value, such as the mean value (due to the dispersion of the value) cannot be located in this zone. Note, the value can be located in the restriction. So, the restriction takes place for the mean value only.
For purposes of this article, let us define a restriction for a probability estimation and probability as a zone in the probability scale, such as the probability estimation and probability (due to the dispersion of the density of the probability estimation) cannot be located in this zone. Note, the density of the probability estimation can be located in the restriction. So, the restriction takes place for the probability estimation and probability only.

Consider two simple but significant examples for two phenomena. The first one will correspond to the restrictions in numerical segments. The second one will correspond to the restrictions in the probability scale.

3.1 Restrictions in Numerical Segments

Several points may be considered on a numerical segment as an example. The example of two points (as the minimum number of points) is the simplest and the most clear one.

3.1.1 Two points

Suppose a numerical segment \([A; B]\) (see figure 1). Suppose two points are determined on this segment: a left point \(x_{\text{Left}}\) and a right point \(x_{\text{Right}}\) : \(x_{\text{Left}} < x_{\text{Right}}\). The coordinates of the middle, mean point may be calculated as \(M = (x_{\text{Left}} + x_{\text{Right}})/2\).

Suppose the points cannot escape outside the boundaries of this segment. This means \(A \leq x_{\text{Left}}\) and \(x_{\text{Right}} \leq B\).

Suppose the points cannot approach each other closer than a non-zero distance of two sigma \(2\sigma > 0\). This means \(x_{\text{Right}} \geq x_{\text{Left}} + 2\sigma\) or \(x_{\text{Left}} \leq x_{\text{Right}} - 2\sigma\). At that, \(M - x_{\text{Left}} = x_{\text{Right}} - M \geq \sigma > 0\).

For the sake of simplicity and obviousness, figures 1-3 represent a case: \(x_{\text{Right}} = x_{\text{Left}} + 2\sigma\) and \(x_{\text{Left}} = x_{\text{Right}} - 2\sigma\) and \(M - x_{\text{Left}} = x_{\text{Right}} - M = \sigma\).

![Figure 1](image)

**Figure 1.** A segment \([A, B]\). Left \(x_{\text{Left}}\), right \(x_{\text{Right}}\) and the middle point or mean point \(M\) on it

Notes about figure 1:

**Note 1:** Suppose two points are determined on a segment and they cannot approach each other closer than a non-zero distance.

**Note 2:** More generally: Suppose a quantity is determined on a finite segment and the dispersion of the quantity is not less than a non-zero constant.

One can easily see two types of zones can exist on the segment for the mean point \(M\):

The mean point \(M\) can be located only in the zone which may be named "allowed" (Figure 2).

The mean point \(M\) cannot be located in the zones which may be named "forbidden" (Figure 3).

3.1.2 Allowed zone

Due to the conditions of the example, the left point \(x_{\text{Left}}\) may not be located more left than the left boundary of the segment \(A \leq x_{\text{Left}}\) and the right point \(x_{\text{Right}}\) may not be located more right than the right boundary of the segment \(x_{\text{Right}} \leq B\). For \(M\) we have \(M - x_{\text{Left}} = x_{\text{Right}} - M = \sigma\).

The allowed zone for \(M\) equals to \((B - A) - 2\sigma\). It is less than the segment on \(2\sigma\). If the distance \(2\sigma\) between the left \(x_{\text{Left}}\) and right \(x_{\text{Right}}\) points is non-zero then the difference between the allowed zone and the segment is non-zero also.

~ 63 ~
So, the mean point $M$ cannot be located in any position of the segment.

3.1.3 Restrictions. Forbidden zones

If $A \leq x_{Left} \leq B$ and $x_{Right} - x_{Left} \geq 2\sigma$, then there are the restrictions of $\sigma$ (One sigma) between the mean point and the boundaries of the segment. So, the mean point $M$ cannot be located in two zones located near the boundaries. These zones may be named restrictions or forbidden zones.

So, the non-zero restrictions exist between the allowed zone of the mean $M$ and the boundaries of the segment. The width of every restriction is equal to $\sigma$. If the distance $2\sigma$ between the left $x_{Left}$ and right $x_{Right}$ points is non-zero, then the forbidden zones, restrictions for $M$ are non-zero also.

Readers of this article can doubtlessly suggest such or better examples for more complex cases.

3.2 Restrictions in the Probability Scale

Consider a classical example: an aiming firing at a target.

Suppose a round target (Figure 4) of the diameter $2L$.

For the obviousness suppose (Figure 5) the dispersion of hits is uniformly distributed in a zone of the diameter $2\sigma$ (See an example of the normal distribution, e.g., in Harin (2010a)).

Consider two cases:

1) The diameter $2\sigma_{Small}$ of the zone of dispersion of hits is considerably less than the diameter $2L$ of a target (Small dispersion).

2) The diameter $2\sigma_{Large}$ of the zone of dispersion of hits is considerably more than the diameter $2L$ of a target (Large dispersion).
Figure 5. Dispersion of hits is uniformly distributed in a zone of the diameter $2\sigma$.

Notes about this figure:

Note 1: This is only a simplified example (See an example of the normal distribution, e.g., in Harin (2010a)).

Note 2: The case 1) represents the case of small diameter $2\sigma_{Small}$ of the zone of dispersion of hits. The case 2) represents the case of large diameter $2\sigma_{Large}$ of the zone of dispersion of hits.

Suppose the point of aiming may be varied between the center of the target and a point which is outside the target.

3.2.1 Small dispersion

The case, when the diameter $2\sigma_{Small}$ of the zone of dispersion of hits is considerably less than the diameter $2L$ of the target, is drawn on the figure 6.

Figure 6. Firing for the small dispersion of hits

Notes: The diameter $2\sigma_{Small}$ of the zone of dispersion of hits is considerably less than the diameter $2L$ of a target.

At the condition of the small dispersion of hits, the maximum possible probability of hit in the target can be equal to 1 (can reach the boundary of the probability scale).

When the point of aiming is varied between the center of the target and a point which is outside the target, the probability of hit in the target is varied from 1 to 0. There are no restrictions in the probability scale.

3.2.2 Large dispersion, Restrictions

The case, when the diameter $2\sigma_{Large}$ of the zone of dispersion of hits is considerably more than the diameter $2L$ of the target, is drawn on the figure 7.
Figure 7. Firing for the large dispersion of hits

Note: The diameter $2\sigma_{\text{Large}}$ of the zone of dispersion of hits is considerably more than the diameter $2L$ of the target.

At the condition of the large dispersion of hits (exactly speaking at the condition the diameter $2\sigma_{\text{Large}}$ of the zone of dispersion of hits is more than the diameter $2L$ of a target), the maximum possible probability of hit in the target cannot be equal to 1.

So, the situation for the probability for this case is drawn on the figure 8.

Figure 8. Restriction for the probability: Allowed zone and forbidden zone

Notes: See the example of two restrictions for two boundaries in Harin (2011c).

The value $P_{\text{AllowedMax}}$ of the maximal allowed probability of the allowed zone $[0, P_{\text{AllowedMax}}]$ may be estimated as the ratio of the mean number of the hits in the target to the total number of the hits. In this particular case, when the distribution of hits is supposed to be uniform, this ratio equals to the ratio of the area of hits scattering to the area of the target

$$P_{\text{AllowedMax}} = \frac{S_{\text{Target}}}{S_{\text{Hits Large}}} = \frac{\pi L^2}{\pi \sigma_{\text{Large}}^2} = \frac{L^2}{\sigma_{\text{Large}}^2}.$$

If $L < \sigma_{\text{Large}}$,

Then $P_{\text{AllowedMax}} < 1$.

In this particular case, the probabilities of hit in the target, that are larger than $P_{\text{AllowedMax}}$ are impossible. The allowed probabilities of hit in the target belong to the allowed zone $[0, P_{\text{AllowedMax}}]$. 

~ 66 ~
The value of the restriction $R_{\text{Restriction}}$ may be estimated as the difference between unit and the maximal allowed probability $P_{\text{AllowedMax}}$ of hit in the target

$$R_{\text{Restriction}} = 1 - P_{\text{AllowedMax}} > 0.$$ 

and, if $L<\sigma_{\text{Large}}$, then $R_{\text{Restriction}}$ is a positive nonzero quantity. At the conditions of the figure 7, it is evident the probability $P_{\text{AllowedMax}}$ cannot be more, then $0.5-0.7$ (50%-70%) and the restriction $R_{\text{Restriction}}$ is as more as $0.3-0.5$ (30%-50%).

4. Reformulated Examples of Problems of the Utility Theory and Their Partial Solution

4.1 General Consideration

Let us reformulate the above examples of problems from the chapter 2.2 of the utility theory to the terms of the chapter 3.2 figure 7.

We will replace the conditions of the examples from the chapter 2.2 by the following conditions from the above illustrated example of the figure 7 of the chapter 3.2:

The "guaranteed outcome" is replaced by "to fire point-blank in the target." The "probable outcome with the probability 0.99 (99%)" is replaced by "to hit the target with the probability 0.99 (99%)."

Further. The scholar keeps in mind the small dispersion (Actually, he considers an ideal, not real situation). But people may conversely keep in mind the real, the large dispersion. At the conditions of the chapter 3.2, especially at the conditions of the figure 7, it is evident the restriction $R_{\text{Restriction}}$ is as more as $0.3-0.5$ (30%-50%). Let us suppose the restriction $R_{\text{Restriction}}$ is more than, say, $0.03$ (3%). At this condition the probability $P_{\text{AllowedMax}}$ cannot be as more as $0.97$ (97%), that is $P_{\text{AllowedMax}} < 97\%$.

If the whole real experience of the people proves them that in real life $P_{\text{AllowedMax}} < 97\%$, then the people keep in mind real $P_{\text{AllowedMax}} < 97\%$ though the scholar keeps in mind $P=99\%$. Moreover, if the scholar says $P=99\%$, the people nevertheless keep in mind real $P_{\text{AllowedMax}} < 97\%$, e.g. supposing that $P=99\%$ is the ideal probability without the real large dispersion.

So, let us suppose people keep in mind $P_{\text{AllowedMax}} < 97\%$ irrespective of what keep in mind or says the scholar.

4.2 Gain at High Probabilities

Suppose a scholar offers you a choice of two outcomes:

A) A guaranteed gain of $99$ if somebody fires point-blank in the target, with the probability 1 (or equivalently 100%);

or

B) A probable gain of $100$ if somebody hits the target with the probability 0.99 (or equivalently 99%), or nothing if somebody hits beside the target with the probability 0.01 (or equivalently 1%).

From the point of view of the scholar, the probable gain has the probability 99% and the mean values for the guaranteed and probable outcomes are

$$99 \times 100\% = 99,$$

$$100 \times 99\% = 99,$$

Here, $99 = 99$

and both outcomes are equally preferable.
But from the point of view of the people, the maximal real probable gain has the probability less than 97% and the mean values for the guaranteed and probable outcomes are

\[ 99 \times 100\% = 99, \]
\[ 100 \times 97\% = 97, \]
Here, $99 > 97$

and the guaranteed outcome is more preferable. This takes place in the paradox also.

### 4.3 Loss at High Probabilities

Suppose a scholar offers you a choice of two outcomes:

A) A guaranteed loss of $99 if somebody fires point-blank in the target, with the probability 1 (or equivalently 100%);

\[ \text{or} \]

B) A probable loss of $100 if somebody hits the target with the probability 99%, or no loss if somebody hits beside the target with the probability 1%.

From the point of view of the scholar, the probable loss has the probability 99% and the mean values for the guaranteed and probable outcomes are

\[ -99 \times 100\% = -99, \]
\[ -100 \times 99\% = -99, \]
Here, $-99 = -99$

and both outcomes are equally preferable.

But from the point of view of the people, the maximal real probable loss has the probability less than 97% and the mean values for the guaranteed and probable outcomes are

\[ -99 \times 100\% = -99, \]
\[ -100 \times 97\% = -97, \]
Here, $-99 < -97$

and the probable outcome is more preferable. This takes place in the paradox also.

### 4.4 Partial Solution of the Problems

We see the conditions can exist, at which the restrictions, caused by a large dispersion, may provide a partial solution of the discussed problems.

Namely, a scholar, who performs the experiment, may keep in mind the small dispersion and propose probabilities that are very close to the boundary of the probability scale. Conversely, if the whole real experience of the people proves the dispersion is usually large, then people may keep in mind the large dispersion. This large dispersion may cause restrictions for the maximal attainable probability and partially explain some of the discussed paradoxes.

So, the essence of the part of the above paradoxes may be caused by the difference of the points of view of the scholar, who performs the experiment, and the people those answer the questions of the scholar.

The scholar keeps in mind the small dispersion, which is ideal, unreal. So, he asks about a probability, which does not exist in reality. Conversely, the people keep in mind the real, the large dispersion. So, they correct the unreal probability to the real one and answer about the real probability. This real probability is shifted, biased with respect to the ideal, unreal probability of the scholar.
5. Conclusion

In this article, as in the first part of the whole review, we have shown two statements:

(1) In the condition of a non-zero dispersion of data, the non-zero restrictions can exist near the boundaries of numerical segments;

(2) In the condition of a non-zero dispersion of data, the non-zero restriction can exist near the boundary 1 (100%) of the probability scale.

The considered examples show the conditions can exist, in which the restriction of the second statement may provide a partial solution of some of the discussed problems.

The second article, as the final part of the whole review, will be titled admittedly “Data dispersion in economics (II) -- Inevitability and consequences of restrictions”. It will be devoted to theorems of existence, which should specify conditions, namely the conditions of the dispersion of data, in which the restrictions must exist and to consequences of such restrictions.

References


