Academic Journal of Management Science Research, Vol.1 No.1, January 2016
pISSN 2414-6498| eISSN 2414-648X

Measuring Information Loss in Managerial Decision

Petr Volf

Institute of Information Theory and Automation, Prague, Czech Republic
Email: volf@utia.cas.cz

Abstract: Traditional decision theory dealing with uncertainty is usually considering criteria based on expected values, or, variantly, on selected quantiles of objective function. Both, however, take into account just rather small part of available information, in particular not counting with possible variability of involved random variables. That is why the criteria based simultaneously on a set of reasonable characteristics should be preferred. This leads to a multi-objective problem and solution based on an appropriate utility function. In the present paper we propose quantitative characteristics measuring information loss caused by reduction of information used in decision. Such measures can help us to find a trade-off between the decision problem complexity and its reasonable simplified re-formulation. This concept is illustrated on examples.

Keywords: stochastic optimization, utility function, newsvendor problem, Gini index, information loss.

---

1. Introduction

The decision problem under uncertainty, e.g. a stochastic optimization problem, is, as a rule, formulated in terms of expected cost function (c.f. Dupacova, Wets, 1984). Alternatives can be based on other, sometimes more convenient, characteristics of uncertain objective function. For instance, the Value-at-Risk, used in the area of risk management in finance or insurance, is defined as a quantile of the distribution of losses. Thus, criteria based on quantiles are often used in many applications, economic (for instance Kim and Powell, 2011) as well as engineering (Volf, 2013). The main advantage of such approach to a decision problem under uncertainty is the improved robustness to heavy tails and to influence of possible outliers in analyzed data, as well as in overcoming the problem of theoretical infiniteness of the mean.

However, the question of proper quantile (and other characteristics) selection arises immediately. Moreover, a criterion based on just one characteristic does not take in account possible variability of involved random variables, it uses just very narrow part of information available. In order to improve this limitation, we cannot avoid considering a kind of multiobjective decision problem taking into account both an optimal „mean" level as well as an uncertainty of decision result. When comparing possible consequences of different decisions, we are actually interested in ordering (ranking) of possible results along their „utility". It leads to formulation of a weighted joint criterion based on more or less subjective user's preferences, i.e. on his utility (or preference) function (c.f. Anand, 1993). This is actually the most standard approach to the multi-objective optimization problem solution. In the present contribution, except exploring the problems formulated above, we are mainly interested in quantification of how much information we neglect when building our decision on just one or a small set of characteristics of future scenario. It is actually one of objectives of the paper. Consequences of consideration just a reduced information instead the full scenario distribution will be studied and illustrated on several examples, either simple artificial or with a real origin.

In the case of complex future scenarios and dense space of them, the problem of decision is better formulated as a problem of stochastic optimization with appropriate objective function. As a rule, the objective function depends on a set of parameters we are able to specify and in such a way to determine resulting scenario. Hence, even in such a case it is desirable to be able to quantify the information loss caused by reduction of complete scenario characterization (e.g. in terms of probability distribution) to a small set of selected characteristics (the mean and selected quantiles, for instance). Again, reasonable quantitative measures of the information loss will be proposed.

The paper is organized as follows: In the next part a simple case of discrete-valued „game" is considered. Even here it could be shown how subjective preferences can influence the decision. Standard deviation and Gini index are used as the decision supporting characteristics and, simultaneously, as the measures of information loss. In Section 3 a more complex problem of stochastic optimization is studied, again from the point of view of optimal decision on the basis of selected criteria and preference function. Simultaneously, the measures of information loss are generalized, distances in L_1 and L_2 functional spaces are utilized. Illustration deals with the newsvendor problem, we again explore properties of solutions optimal with respect to different preference functions.

2. Decision Rule and Loss of Information

As it has already been said, in an ideal case we can try to find a uniformly best scenario and to decide to follow it. However, as a rule, such a scenario does not exist. Instead, we have to base our decision on one or more characteristics of the objective function. Then, in order to obtain an ordered scale of scenarios, possible values of selected criteria have to be combined in a convenient way corresponding to our (user's) preference. Naturally, in such an approach (which is reasonable) only a part of information is utilized. Hence, our concern is also to quantify, with the aid of a properly selected measure, how much information was neglected when applying certain decision rule. This question is actually the main interest of the present paper, in the next parts and examples we shall try to propose such „information loss" measures and illustrate their use.

2.1 Measuring information loss

Theory of information offers a number of information measures (or, rather, measures of uncertainty), for instance the entropy or a set of more general diversity indices, see for instance Jost (2006). To quantify a relative loss of information caused by a misspecified distribution of probability, the Kullback-Leibler or other divergences are available (c.f. Vajda, 1989). All mentioned measures, however, have one disadvantage (from our point of view). They are based solely on probabilities assuming that the domain of values is fixed. Therefore, we shall prefer here the standard deviation and the Gini coefficient (index), two well known measures of variability. Non-standardized Gini index of a random variable X equals the mean absolute deviation, $GI^*=E|X-EX|$, while the standard deviation $std(X)$ equals the square root from the variance $E[(X-EX)^2]$. For non-negative variables both indices can be standardized (relativized), to $GI=GI^*/EX$ or to the coefficient of variation $CV=std(X)/EX$. As we shall work with both positive and negative values (gain, loss), non-standardized versions will be employed. In fact, we shall use them as a measure of information loss caused by a reduction of the whole distribution of a scenario to just one representative characteristic, EX. Their application will be specified in examples.

Let us remark here that while $min_a\ E[(X-a)^2]$ is reached at $a=EX$, while $min_a\ E|X-a|$ is attained at the median of X. From this point of view, it is interesting that Gini index was defined with respect to EX instead to median. Such an alternative definition can have also certain practical advantages in cases of strongly non-symmetric distribution of X or of distribution with heavy tails.

2.2 Example I

Let us imagine a simple game consisting of two possible scenarios. In one, we shall loose -1 (unit) for sure, in the second we can either gain $+5$ or loose -5, both with probability 0.5. Hence, when deciding on the basis of the expected value, we should prefer the second one, as its $EX=0$. However, both measures of variability, $std(X)$ and Gini index, are 0 in the first case and equal 5 in the second, indicating that the decision based only on EX may not be satisfactory. Which other decision rules could be reasonable in such a case? Let us first consider two extremal criteria, namely max(max(gain)) and max(min(loss)). In other words, the former prefers high gain, neglecting the risk of loss (no matter how large). In our „game" described above it will prefer the second scenario. On the contrary, the latter criterion wishes to minimize possible losses, not taking into account chances to gain, hence the first scenario will be preferred. From this example it is seen that some approaches correspond to behavior which could be called „the inclination to risk", while others correspond rather to „the aversion to risk". We may see here that even in such a simple case the decision is not easy. Decision rule could be based on some a-priory selected (formulated) strategy. Let us recall here the notion of utility function, $U(x)$, say. It is nothing else than a model function giving different weights to possible results in scenario, expressing actually our preferences. Essentially, we may restrict ourselves to either concave or convex functions. Naturally, both are assumed to be non-decreasing and having $U(0)=0$ (i.e. there is no utility from nonthing). Final criterion can then be based on the expected utility of a scenario, i.e. on its weighted mean $E(U(X))$. This is again in accordance with the fact that each multicriterial decision rule corresponds to a decision based on certain combination (i.e. weighted mean) of single criteria. In this context, the rule based on simple EX corresponds to expected utility under linear utility function.

In fact, people commonly do not solve problems on the basis of a-priory formulated utility function. This is rather an ex-post construction, a model of prevailing behavior, attemting to explain it (and to predict it for the future). A natural question arises how an appropriate utility function should be constructed. As a rule, it can be derived from results of observation or survey mapping people's behaviour, its quantification could be rather interesting and challenging problem of (statistical) modeling. Our aim here is just to show possible consequences of certain ordering of preferences, though in the real people's decision making this ordering may be just implicite and intuitive.

Let us return to the example. Figure 1 shows two examples of utility functions, one concave and the second convex, and their values at *-5, 5* and *-1*. Expected utility of the first scenario equals *U(-1),* while the second yields *(U(-5)+U(5))/2.* Expected utilities for both functions are denoted by stars, it is seen that conveniently chosen concave function leads to preference of the first scenario, i.e. also corresponds to the aversion to risk behavior.

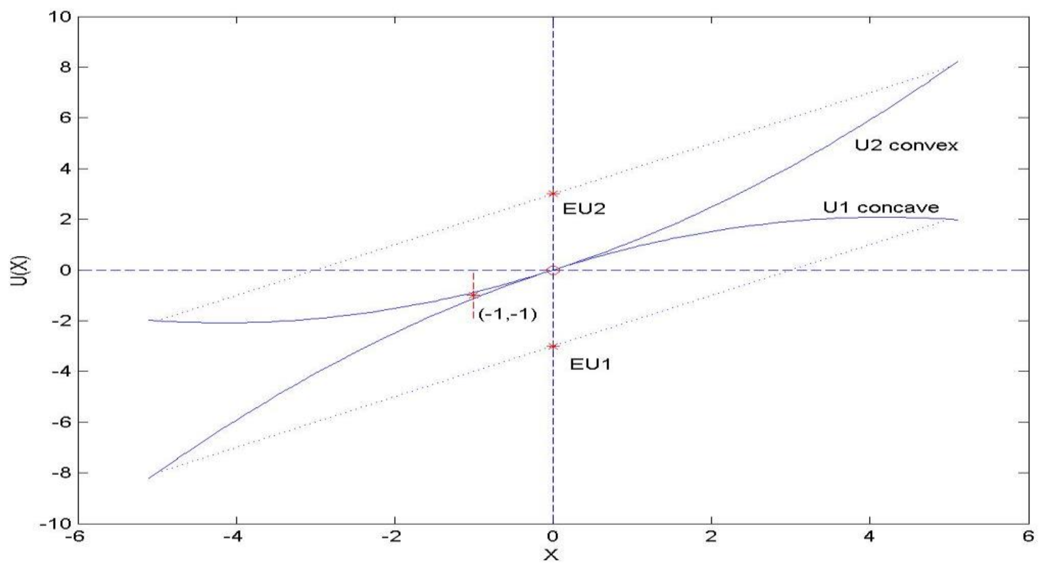

Figure 1. Example of two utility functions (concave and convex) and corresponding expected utilities

3. Relation to Stochastic Optimization Problem

In the sequel we shall deal with a more complex (and more interesting, too) situation than in the preceding part and example. Let us imagine a case with a large set of possible scenarios and correspondingly large, in many instances even "dense", continuous, scale of possible decisions. Optimization problems are then as a rule formulated in terms of the stochastic optimization. In such a situations it is even more important to take into account all aspects af solutions. Typically, a slightly sub-optimal solution (with respect to one attribute, e.g. expectation) could be much better as regards another characteristic (e.g. range, uncertainty, variability) of chosen scenario.

Hence, let us consider an optimization problem with objective function $\Phi(y;\, v)$; where v are input variables from certain feasibility set V and values y are random, results of a random variable (or vector) Y with distribution function F. Standardly, corresponding stochastic optimization problem is formulated as $\sup_v E_F\, \Phi(Y;\, v)$, where E_F stands for the expectation w.r. to F. If F is known, we actually deal with a „deterministic" optimization case. It has been said already that the criterion based on averaging does not take in account possible variability of r.v. Y. That is why alternatives may consider optimization of other characteristics, for instance the quantiles of random criterion $Z(v) = \Phi(Y;\, v)$. As there arises a problem of which quantile should be selected, it is reasonable to utilize a kind of multi-objective optimization procedure considering simultaneously several characteristics, and also the variability of solution (measured by variance or by certain inter-quantile range). It is well known that a multi-objective optimization task can be re-formulated as a task to optimize certain wighted sum of particular criteria. We are then facing again to the question how the weights should be selected. It is again a matter of individual preferences. Simultaneously, we are also interested in quantification of amount of information neglected by reducing the whole scenarion distribution to its several characteristics. Illustration of these ideas is the sense of the following example.

3.1 Example II

We shall consider a very simple case of the newsvendor model. Let D be a random variable with known probability distribution representing the demand (of units of certain commodity), further let each unit be sold for price q and purchased for price c, and S be the number of units stocked, i.e. purchased to be sold. In the simplest case c and q are fixed, we search the solution to the optimal stocking quantity S which maximizes profit $Z=q*min(S,D)-c*S$. As both D and therefore Z are random, we have a choice which criterion to maximize. It is well known and can be shown easily that to get $max_s\, EZ$, optimal solution is $S=S_E=((q-c)/q)$-quantile of distribution of D, while for maximization of the α-quantile of Z optimal stock equals $S=S_\alpha$ = α-quantile of D. Computation of resulting distribution of profit Z, hence also of achieved values of EZ and of quantiles of *Z*, is then rather easy task. See for instance Petruzzi, Dada (1999) and also Kim, Powell (2011). Hence, it is possible and interesting to study dependence of distribution of *Z* and values of its characteristics, including its variance, on chosen stock amount *S*. This is actually purpose of the example, and, in particular, of Figure 2.

In the numerical example of the newsvendor model, we have selected the normal distribution for *D, D~N (μ=150, σ=20)*, prices were set to *c=10, q=15*. Then *0.1, 0.25, 0.5, 0.75,* and *0.9* quantiles of this normal distribution equal *124.3690, 136.5102, 150.0000, 163.4898*, and *175.6310*, respectively. It means that these values are also levels of stock S_α maximizing corresponding α-quantiles of profit *Z*. Further, the ratio *(q-c)/q=1/3*, hence the optimal stock level maximizig *EZ* equals S_E = *141.3855*, corresponding to *1/3* quantile of distribution of *D.*

Figure 2 shows dependence of considered five quantiles of profit *Z* on *S*. It contains also the graph of mean values *EZ* and curves $EZ-2\sigma_Z$, $EZ+2\sigma_Z$, where σ_Z denotes standard deviation of *Z*. Though the value S_E= *141.3855* guarantees maximal possible expected profit, simultaneously it is seen that the variance (and therefore uncertainty) of profit really achieved is rather large. However, let us consider just slightly sub-optimal *S*, for instance *S=130*. Expected profit is then just slightly smaller, approximately from *600* decreasing to *580,* simultaneously standard deviation of distribution of *Z* is smaller significantly (approximately twice). It is also seen how for even smaller levels of stock (less than *120*, say) the distribution of possible profit becomes to be rather narrow, but, naturally, all considered distribution characteristics are already lower.

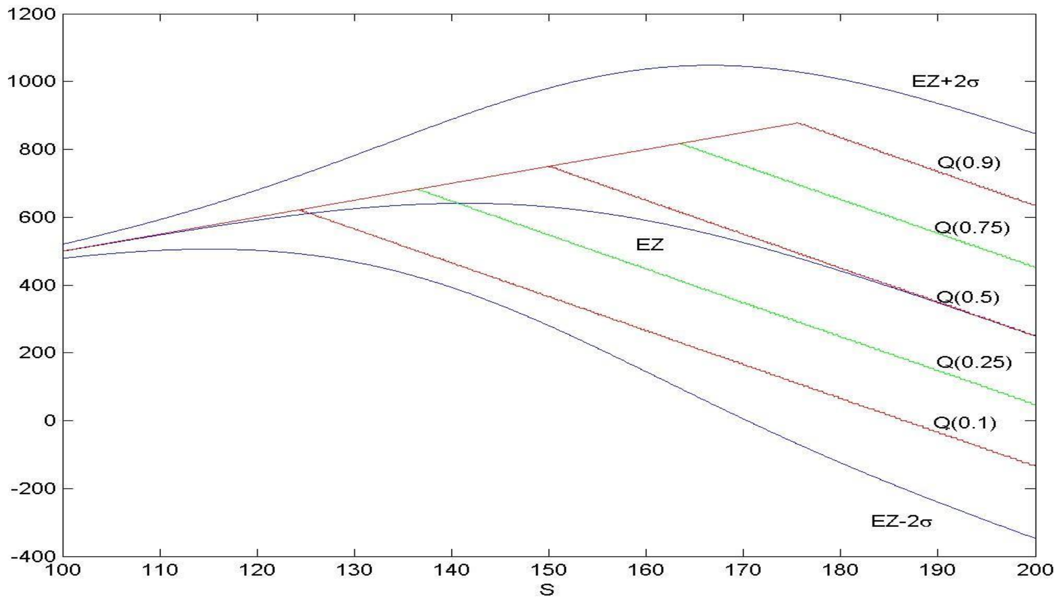

Figure 2. Dependence of characteristics of profit *Z* on purchased stock amount *S*

Figure 2 shows main characteristics of probability distribution of profit under different possible scenarios corresponding to different selections of *S*. The example itself again indicates how nonsatisfactory can be decision based only on one, though representative, characteristics. However, in the end it is necessary to reach some decision, i.e. to formulate a reasonable decision rule. Again, a quite natural way consists in a trade-of among several criteria, in other words in formulation of a multi-criterial problem expressed with the aid of weighted combination of single criteria. Consequently, we are facing once more the question of convenient choice of weights. As the objective of the present study is not to propose one “best” choice (which actually does not exists), the aim is the exploration of different cases and possibilities, and we shall study and compare consequences of several different solutions. That is why four different sets of weights, combining single criteria based on *0.1, 0.25, 0.5, 0.75,* and *0.9* quantiles of Z to one complex criterion, were taken into account, namely:

W1: *0.1, 0.15, 0.2, 0.25, 0.3.*
W2: *0.3, 0.25, 0.2, 0.15, 0.1.*
W3 uniform, i.e. all weights equal to *0.2.*
W4: *0.1, 0.2, 0.4, 0.2, 0.1.*

Figure 3 displays shapes of these four combined characteristics in dependence on *S*, hence we can trace also values of *S* maximizing each of them. As regards optimal stocks *S* for these combinations, from the form of quantile curves (broken linear, see Figure 2) it follows that extremal point has to be at one from S_α. As indicated also by Figure 3, they are at *S1=150, S2=136.5, S3=136.5, S4= 150*, respectively. The first combination with weights *W1* prefers higher gain with small respect to uncertainty of the result. Priority of the second combination, on the contrary, is smaller variability and suficiently sure lower

bounds. Weights *W3* does not prefer any from considered single criteria and means a „golden mean“, while the last combination is closest to the median criterion, taking into account also other characteristics od profit distribution.
After such a simplification of the decision problem reducing the set of characteristics to just several composite critera, e.g. the four defined above, we are still facing the question which of them to apply. However, each criterion reflects certain preferences and should be used in accordance with them. In fact the set of weights substitutes here the instrument of utility function. The choice of weights is without doubt related to user’s inclination or aversion to risk, instead of utility function such a construction could be called „the preference score“.

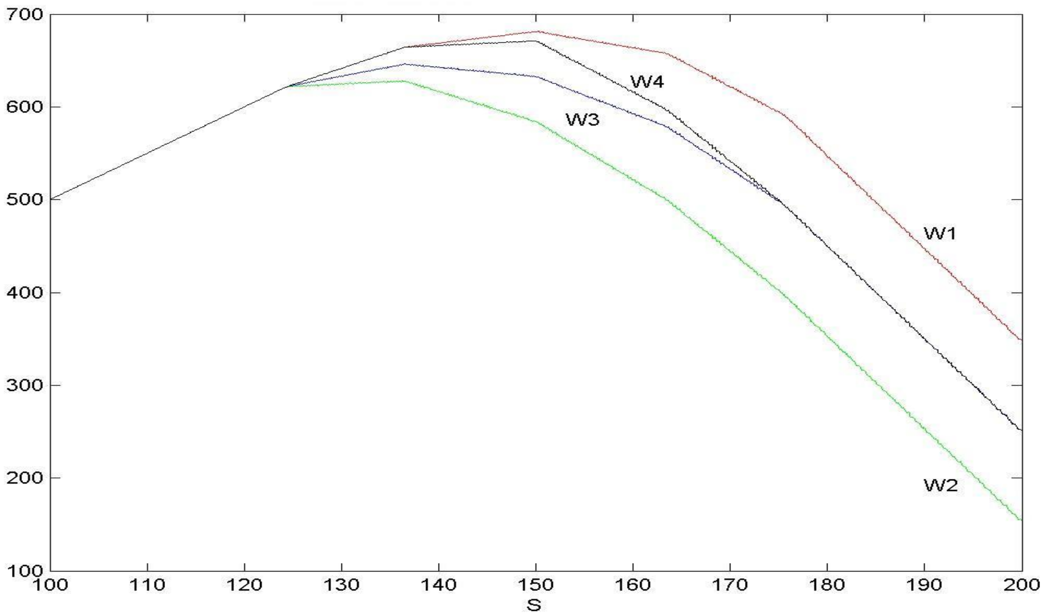

Figure 3. Four combinations of quantiles of profit *Z* and their dependence on *S*

3.2 Generalized measures of information loss

In section 2.1 we have recalled two indices convenient for measuring amount of information lost by reducing our knowledge on objective variable (scenario) *X* to only one, though representative, characteristics (as a rule, to *EX*). Now we deal with a more complex case. It can be described as a reduction of a full probability distribution (which can be of continuous type) of the objective variable to a subset of (characteristic) values. In terms of distribution function, initial „true“ distribution function F_1 is reduced to (substituted by) a distribution function F_2. To quantify the size of such reduction, we propose to utilize two following measures,

$$G(F_1;F_2) = \int_{-\infty}^{\infty} |F_1(x)\text{-}F_2(x)|\,dx \quad \text{and} \quad C(F_1;F_2) = \left[\int_{-\infty}^{\infty} (F_1(x) - F_2(x))^2\,dx\right]^{1/2}, \tag{1}$$

actually representing distances in L_1 and L_2 functional spaces. Simultaneously, the first can be regarded as a generalization of the Gini index, because in the case that F_2 is degenerated to one point *EX*, *G* coincides with *GI**. The second, however, equals *std(X)* only in particular cases, for instance when we deal just with Bernoulli random variable *X*. Then *X=1* with probability *p* and *X=0* with *1-p*, its F_1 jumps from *0* to *1-p* at *x=0* and then jumps from *1-p* to *1* at *x=1*. Further, let us reduce *X* to its expectation *EX=p*, hence corresponding $F_2(x)=0$ for $x<p$ and $=1$ for $x\geq p$. It is easy to compute that C^2 from (1) really equals $var(X) = p(1-p)$. *G* then equals corresponding Gini index E|*X-EX*| = *2p(1-p)*.

3.3 Example II continued

Our intention is now to apply both distances defined above to situation of Example II, in order to quantify an information loss caused by reduction of the whole distribution F_1 of variable *Z*, at different *S*, to just five quantiles (*0.1, 0.25, 0.5, 0.75, 0.9*). Hence, we shall consider four variants of distribution F_2 given by four sets of weights. Let us recall that $Z=q*min(S,D)-c*S$, where $D\sim N(\mu=150,\ \sigma=20)$, parameters (prices) *c=10*, *q=15*. Plots on Figures 4 and 5 display both distances *G* and *C* as functions of *S*, for all four sets of weights (here taken as probabilities of five values corresponding to five quantiles of distribution of *Z*, thus forming distribution function F_2). Further, in both figures, the highest curve shows indices *G* and *C* computed along (1) and quantifying information loss when F_2 is concentrated at only one value, *EZ*.

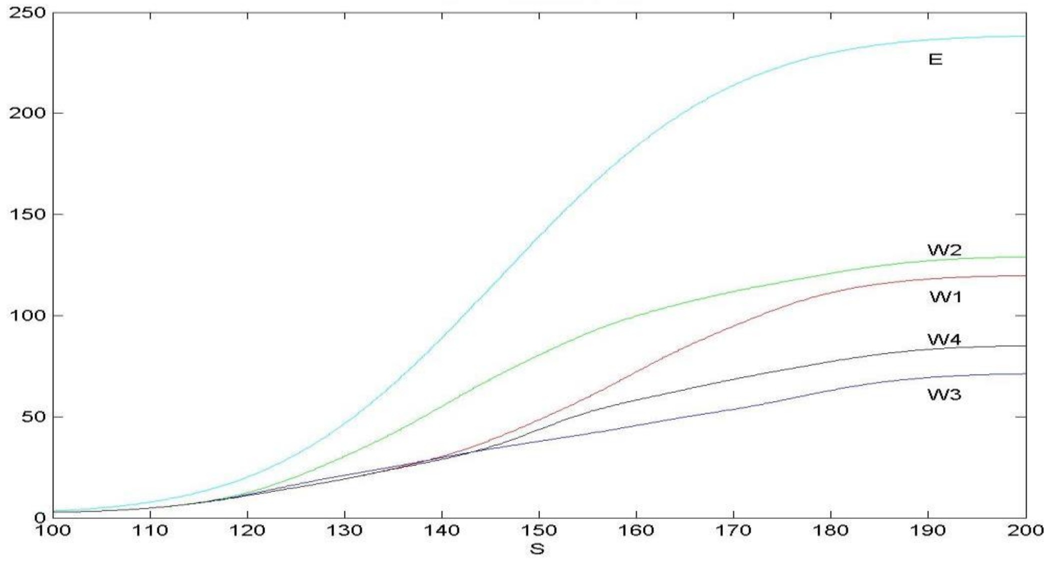

Figure 4. *G* – measures of information loss

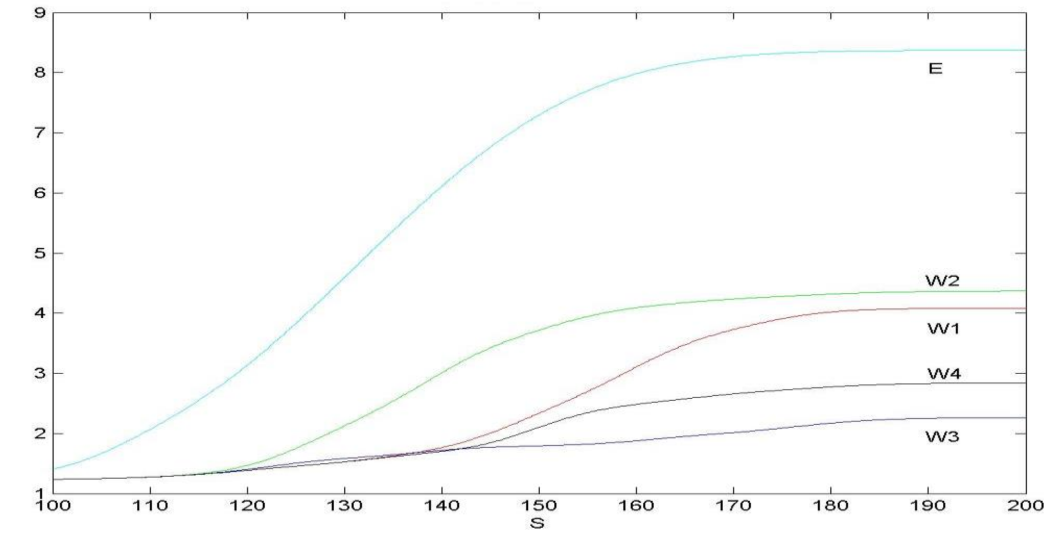

Figure 5. *C* – measures of information loss

Both graphs demonstrate that the reduction has larger consequences at larger *S*, i.e. in regions where the distribution of *Z* is richer, implying larger uncertainty of result of decision. It is also seen how the simultaneous use of several characteristics (here five quantiles) leads to much smaller information loss than when the decision is based on only one value (*EZ*). Both measures of lost information are smallest for the case with equal weights, *W3*. Quite different scales of both indices then follow from their different definitions.

4. Concluding Remarks

The main objective of the paper was to propose measures quantifying how much information is in fact neglected when decision (in a stochastic decision problem) is based just on one or several (though representative) characteristics of future scenario. The concept of measures was based on Gini index and standard deviation expresing the uncertainty, variability. of values of objective function after a decision is chosen. As the generalization of these measures, we proposed to use distances in L_1 and L_2 functional spaces and to employ them for quantifying distance between distribution function of full scenario and its reduction to just several characteristics. The performance of proposed criteria was studied. With the aid of examples, including a simple version of the newsvendor model, applicability and usefulness of presented information loss measures was demonstrated.

Acknowledgments

P. Volf would like to thank to Czech Scientific Foundation Project No 13-14445S.

References

Anand P. (1993). Foundations of Rational Choice Under Risk. Oxford University Press, Oxford.

Dupacova J. and Wets R. (1984). Asymptotic behaviour of statistical estimates and optimal solutions of stochastic optimization problems, Ann. Statist., 16: 1517–1549.

Jost L (2006). Entropy and diversity. Oikos, **113**: 363–375.

Kim J.H. and Powell W. (2011). Quantile optimization for heavy-tailed distributions using asymmetric signum functions. Working Paper, Princeton University. Retrieved January 12, 2016 from http://castlelab.princeton.edu/Papers/Kim%20Powell%20-%20Quantile%20Optimization%20June252011.pdf

Konak A., Coit D.W. and Smith A.E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. Reliability Engineering and System Safety, 91: 992–1007.

Petruzzi N.C. and Dada M. (1999). Pricing and the newsvendor problem: A review with extensions. Operations Research 47(2): 183–194.

Vajda I. (1989). Theory of Statistical Inference and Information. Kluwer Academic Publishers, Dordrecht.

Volf P. (2013). On quantile optimization problem with censored data. In Proceedings of the 31st International Conference on Mathematical Methods in Economics 2013. College of Polytechnics, Jihlava, CZ, 1004–1009.

Authors' Profile

Petr Volf received the M.S. degree in Statistics and Econometrics from the Faculty of Mathematics and Physics of the Charles University in Prague, PhD degree in Theoretical Cybernetics in the IITA of Czech Academy of Sciences, where he is at present working as a senior researcher. Simultaneously, the author is lecturing as an associate professor at the Technical University in Liberec, Czech Republic.