Communications of the Korean Statistical Society
2009, Vol. 16, No. 5, 813–820

A View on Extension of Utility-Based on Links with Information Measures

A.R. Hoseinzadeh[a], G.R. Mohtashami Borzadaran[1,b], G.H. Yari[c]

[a]Science and Research Branch, Islamic Azad University
[b]Department of Statistics, Ferdowsi University of Mashhad
[c]Iran University of Science and Technology

Abstract

In this paper, we review the utility-based generalization of the Shannon entropy and Kullback-Leibler information measure as the U-entropy and the U-relative entropy that was introduced by Friedman *et al.* (2007). Then, we derive some relations between the U-relative entropy and other information measures based on a parametric family of utility functions.

Keywords: Shannon entropy, Kullback-Leibler information measure, utility function, expected utility maximization, U-entropy, U-relative entropy.

1. Introduction

Distance measures between two probability distributions play an important role in probability theory and statistical inference. For the first time, the concept of measuring distance between two probability distributions was developed by Mahalanobis (1963). A class of measure which may not satisfy all the conditions of distance measures or metric space is called divergence measures. Divergence measures based on the idea of information-theoretic entropy was initially introduced in communication theory by Shannon (1948) and later in Cybernetics by Wiener (1949). Kullback and Leibler (1951) introduced relative entropy or the divergence measures between two probability distributions as a generalization of the Shannon entropy. During the half past century, various extensions of the Shannon entropy and Kullback-Leibler divergence measures are introduced by Renyi (1961), Ali and Silvey (1966), Havrda and Charvát (1967), Csiszar (1967), Sharma and Mittal (1977), Burbea and Rao (1982), Rao (1982), Kapur (1984), Vajda (1989), Lin (1991), Pardo (1993), Shioya and Da-te (1995).

Dragomir (2003) introduced the concept of p-Logarithmic and α-power divergence measures and derived a number of basic results.

Friedman and Sandow (2003) introduced a utility-based generalization of Shannon entropy and Kullback-Leibler information measures. Friedman *et al.* (2007) proved various properties for these generalized quantities similar to the Kullback-Leibler information measure.

In this paper, we review utility-based motivations and generalizations of the Shannon entropy and Kullback-Leibler information measures. Then, we introduce a version of the power utility function of the first kind and derive a number of its basic properties. Finally, we obtain a utility-based divergence measure by using of the power utility function of the first kind and find its links with other divergence measures.

[1] Corresponding authors: Associate Professor, Department of Statistics, Ferdowsi University of Mashhad, Also, member of Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad, Mashhad, Iran.
E-mail: gmb1334@yahoo.com

2. Preliminaries

Let X be a discrete random variable with support χ and probability mass function $p(x) = P(X = x)$, $x \in \chi$, Shannon entropy $H(p)$ is defined by:

$$H(p) = -\sum_{x\in\chi} p(x)\ln(p(x)).$$

Shannon entropy measures lack of uniformity (concentration of probabilities) under $p(x)$, $x \in \chi$.

Let $p(x)$ and $q(x)$ on the support χ be two probability mass functions of random variable X. The fundamental information measure for comparing the two distributions is given by Kullback-Leibler as:

$$K(p,q) = \sum_{x\in\chi} p(x)\ln\left(\frac{p(x)}{q(x)}\right).$$

It is well known that, $K(p,q)$ is nonnegative and zero if and only if, $p = q$.

There are various versions on generalization of Shannon entropy and Kullback-Leibler information measure. A general class of divergence measures is called the Csiszar f-divergence, that is introduced by Csiszar (1967) and is defined as:

$$D_f(p,q) = \sum_{x\in\chi} p(x) f\left(\frac{p(x)}{q(x)}\right),$$

where $f(x)$ is convex on $(0,\infty)$ such that, $f(1) = 0$ and strictly convex at $x = 1$. Csiszar f-divergence includes several divergences used in measuring the distance between two probability distributions. For $f(x) = -\ln(x)$, $f(x) = (x-1)^2$ and $f(x) = (1-x^{1-\beta})/(1-\beta)$, we have Kullback-Leibler, χ^2-divergence and β-class divergence measures, respectively.

The α and β-classes of divergence measures are:

$$H_\alpha(p,q) = \frac{1}{\alpha-1}\ln\left[E_p\left(\frac{p(X)}{q(X)}\right)^{\alpha-1}\right], \quad \alpha \neq 1,\ \alpha > 0$$

and

$$H_\beta(p,q) = \frac{1}{\beta-1}\left[E_p\left(\frac{p(X)}{q(X)}\right)^{\beta-1} - 1\right], \quad \beta \neq 1,\ \beta > 0,$$

respectively.

Friedman *et al.* (2007) introduced a decision theoretic, *i.e.*, utility-based, motivation and generalization for the Shannon entropy and Kullback-Leibler information measure and called them U-entropy and U-relative entropy.

2.1. Utility functions

Investor's subjective probabilities are numerical representations of his beliefs and information. His utilities are numerical representations of his tastes and preferences. Utility functions give us a way to measure investor's preferences for wealth and the amount of risk to undertake in the hope of attaining greater wealth. It also measures an investor's relative preference for different levels of total wealth.

Definition 1. *A utility function is a function of wealth $U(x)$ that is a strictly monotone increasing and strictly concave function of x.*

This implies that $U'(x) > 0$ and $U''(x) < 0$. It is easy to see that if $U(x)$ is a utility function, then for constant $a > 0$ and b, the function $W(x) = aU(x) + b$ is also a utility function. The utility functions of most people tend to be concave, at least for large gains or large losses. A person with a strictly concave utility function is called a risk averter. For any utility function $U(x)$, we consider the following function:

$$R(x) = -\frac{U''(x)}{U'(x)}.$$

This function is called absolute risk aversion function. It is obvious that $R(x) > 0$. It is also generally agreed in finance theory that for a utility function to be realistic with regard to economic behavior, its absolute risk aversion should be a decreasing function of wealth.

The most common utility functions are as follows:

- **Exponential** utility function with parameter $\alpha > 0$,

  $$U(x) = \frac{1 - e^{-\alpha x}}{\alpha}, \quad -\infty < x < \infty,$$

  note that the utility tends to the finite value $1/\alpha$ as $x \to \infty$.

- **Logarithmic** utility function with parameters $\alpha > 0, \beta, \gamma$,

  $$U(x) = \alpha \ln(x - \beta) + \gamma, \quad x > \beta.$$

- **Iso-Elastic** utility function is a class of utility functions as follows:

  $$U(x) = \begin{cases} \dfrac{x^{\alpha} - 1}{\alpha}, & \alpha < 1,\ \alpha \neq 0, \\ \ln(x), & \alpha = 0. \end{cases}$$

  This class has the property of $U(kx) = f(k)U(x) + g(k)$, for all $k > 0$ and for some function $f(k) > 0$ and $g(k)$ which are independent of x.

- **Linex** utility function with parameters $\alpha > 0,\ \beta > 0,\ \gamma > 0$, is introduced by Bell (1988, 1995), as:

  $$U(x) = \alpha x - \beta e^{-\gamma x}.$$

- **Power utility function of the first kind** is a two parametric family of utility functions as:

  $$U(x) = \frac{\alpha^{\beta+1} - (\alpha - x)^{\beta+1}}{(\beta + 1)\alpha^{\beta}}, \quad \frac{x}{\alpha} < 1,\ \frac{\beta}{\alpha} > 0.$$

  For $\beta = 1$ leads to quadratic utility function of the form, $U(x) = x - x^2/(2\alpha),\ x < \alpha$.

  By setting $a = 1, b = -U(1)$, $aU(x) + b$ is also a utility function as:

  $$U(x) = \frac{(\alpha - 1)^{\beta+1} - (\alpha - x)^{\beta+1}}{(\beta + 1)\alpha^{\beta}}, \quad \frac{x}{\alpha} < 1,\ \frac{\beta}{\alpha} > 0. \tag{2.1}$$

- **Power utility function of the second kind** is another class of utility functions as:

  $$U(x) = \frac{x^{1-\alpha} - 1}{1 - \alpha}, \quad x > 0,\ \alpha > 0,\ \alpha \neq 1.$$

  We note that $U(x) = \ln(x)$, is the limiting case of this utility as $\alpha \to 1$.

  See for example Gerber and Pafumi (1999), Conniffe (2007) and Johnson (2007).

2.2. U-entropy and U-relative entropy

The principle of expected utility maximization states that a rational investor acts to select an investment which maximizes his expected utility of wealth based on the model he believes (see, for example, Theorem 3, p.31 of Ingersoll, 1987).

Friedman *et al.* (2007) used this to make utility-based generalization of the Shannon entropy and Kullback-Leibler information measure and gave the following definitions.

Definition 2. *Let X be a discrete random variable with finite support χ and $p(x)$, $q(x)$ be two probability measures on χ. The U-relative entropy from the probability measure $p(x)$ to probability measure $q(x)$ is given by:*

$$D_U(p,q) = \sup_{b \in \beta_\chi} \sum_{x \in \chi} p(x)\, U\left(\frac{b(x)}{q(x)}\right),$$

where, $U(x)$ is a utility function and

$$\beta_\chi = \left\{ b : \sum_{x \in \chi} b(x) = 1 \right\}.$$

We note that this optimality is yield for

$$b(x) = b_p^*(x) = q(x)(U')^{-1}\left(\frac{\lambda q(x)}{p(x)}\right), \tag{2.2}$$

where, λ is the solution of the following equation:

$$\sum_{x \in \chi} q(x)(U')^{-1}\left(\frac{\lambda q(x)}{p(x)}\right) = 1.$$

Definition 3. *The U-entropy of the probability measure $p(x)$ is given by:*

$$H_U(p) = U(|\chi|) - D_U\left(p, \frac{1}{|\chi|}\right).$$

In the special case, when $q(x)$ has discrete uniform distribution on the finite support χ, we consider $q(x) = 1/|\chi|$, where $|\chi|$ is the cardinality of the finite set χ. For $U(x) = \ln(x)$, the above definitions are reduced to Kullback - Leibler divergence measure and the Shannon entropy, respectively.

Theorem 1. *The generalized relative entropy, $D_U(p,q)$ and the generalized entropy, $H_U(p)$, have the following properties:*

(i) $D_U(p,q) \geq 0$ *with equality if and only if* $p = q$,

(ii) $D_U(p,q)$ *is a strictly convex function of* p,

(iii) $H_U(P) \geq 0$, $H_U(P)$ *is a strictly concave function of* p.

Proof: See Friedman *et al.* (2007). □

3. Main Results

We consider the power utility function of the first kind again. Setting $\alpha = -1/\theta$ and $\beta = \lambda/\theta$, in it leads to the following:

$$U_{\theta,\lambda}(x) = \frac{(1+\theta x)^{\frac{\theta-\lambda}{\theta}} - 1}{\theta - \lambda}, \quad 1+\theta x > 0,\ \lambda > 0,\ \theta \in R. \tag{3.1}$$

Note that for $a = 1$ and $b = -U_{\theta,\lambda}(1)$, $aU_{\theta,\lambda}(x) + b$ is also a utility function as:

$$U_{\theta,\lambda}(x) = \frac{(1+\theta x)^{\frac{\theta-\lambda}{\theta}} - (1+\theta)^{\frac{\theta-\lambda}{\theta}}}{\theta - \lambda}, \quad 1+\theta x > 0,\ \lambda > 0,\ \theta \in R. \tag{3.2}$$

The following properties are noticeable:

- The absolute risk aversion of (3.1) is $R(x) = \lambda/(1+\theta x)$. It is decreasing for reasonable value of $\theta > 0$, but the absolute risk aversion of the power utility function of the first kind is $R(x) = \beta/(\alpha - x)$, that is increasing for reasonable value of $\beta > 0$.
- $U(0) = 0$ and $U'(0) = 1$, therefore the graph of U passes through the origin with slope one.
- For any fixed θ, we have,
$$\lim_{\lambda\to\theta} U_{\theta,\lambda}(x) = \frac{1}{\theta}\ln(1+\theta x)$$
and for any fixed λ, we have,
$$\lim_{\theta\to\lambda} U_{\theta,\lambda}(x) = \frac{1}{\lambda}\ln(1+\lambda x),$$
where for the positive values of θ and λ, are logarithmic utility functions.
- For any fixed $\lambda > 0$, we obtain,
$$\lim_{\theta\to 0} U_{\theta,\lambda}(x) = \frac{1}{\lambda}\left(1 - e^{-\lambda x}\right),$$
where for the positive values of λ is an exponential utility function.
- It is easy to show that for fixed $\lambda > 0$, we have,
$$\lim_{\theta\to\pm\infty} U_{\theta,\lambda}(x) = x.$$
- The inverse function of $U_{\theta,\lambda}(x)$ as follows:
$$U^{-1}_{\theta,\lambda}(x) = \frac{\{1+(\theta-\lambda)x\}^{\frac{\theta}{\theta-\lambda}} - 1}{\theta}.$$
Therefore, by setting θ in place of $\lambda - \theta$ and x in place of $-x$ in $U_{\theta,\lambda}(x)$, we obtain, $U^{-1}_{\theta,\lambda}(x) = -U_{\lambda-\theta,\lambda}(-x)$ and $U_{\theta,\lambda}(-U_{\lambda-\theta,\lambda}(-x)) = x$.

  Now we consider power utility function of the first kind and introduce a parametric divergence measure that is linked with other important divergence measures.

Theorem 2. *If the utility function be as in (2.1), then the corresponding utility divergence measure is given by:*

$$D_U(p,q) = \frac{(\alpha-1)^{\beta+1}}{(\beta+1)\alpha^\beta}\left[1-\left\{\sum_{x\in\mathcal{X}} q(x)\left(\frac{q(x)}{p(x)}\right)^{\frac{1}{\beta}}\right\}^{-\beta}\right].$$

Proof: Consider the utility function (2.1), then we have,

$$(U')^{-1}(x) = \alpha\left(1-\sqrt[\beta]{x}\right).$$

By setting $(U')^{-1}(x)$ in (2.2), we obtain:

$$b_p^*(x) = \alpha q(x)\left\{1-\left(\frac{\lambda q(x)}{p(x)}\right)^{\frac{1}{\beta}}\right\},$$

in which,

$$\lambda = \left\{\frac{\alpha}{\alpha-1}\sum_{x\in\mathcal{X}} q(x)\left(\frac{q(x)}{p(x)}\right)^{\frac{1}{\beta}}\right\}^{-\beta}.$$

After some calculation, we obtain:

$$D_U(p,q) = \frac{(\alpha-1)^{\beta+1}}{(\beta+1)\alpha^\beta}\left[1-\left\{\sum_{x\in\mathcal{X}} q(x)\left(\frac{q(x)}{p(x)}\right)^{\frac{1}{\beta}}\right\}^{-\beta}\right].$$

□

Corollary 1. *Let $\alpha=\beta$ and $\beta\to\infty$, we have,*

$$\lim_{\beta\to\infty} D_U(p,q) = \frac{1}{e}\left(1-e^{K(p,q)}\right) = \frac{1}{e}ID(p,q),$$

where $ID(p,q)$, is the ID Index as mentioned in Soofi et al. (1995).

Corollary 2. *If $\beta=-2$ and $\alpha=-1/2(1+\sqrt{5})$ then,*

$$D_U(p,q) = 1-\left(\sum_{x\in\mathcal{X}}\sqrt{p(x)q(x)}\right)^2 = 1-\{D_B(p,q)\}^2,$$

where, $D_B(p,q)$, is the Bhattacharyya distance measure.

Corollary 3. *When $\alpha=-1$ and $\beta\to-1$ we have,*

$$\lim_{\beta\to-1} D_U(p,q) = K(p,q),$$

in which $K(p,q)$, is the Kullback-Leibler information measure.

Corollary 4. *$\beta=1$ and $\alpha=2+\sqrt{2}$, implies that,*

$$D_U(p,q) = \frac{D_{\chi^2}(p,q)}{D_{\chi^2}(p,q)+1},$$

in which,

$$D_{\chi^2}(p,q) = \sum_{x\in\chi} \frac{\{p(x)-q(x)\}^2}{p(x)},$$

is the χ^2-divergence measure.

Corollary 5. *By setting $\beta = -1/k$ and $\alpha = -1$, we have,*

$$D_U(p,q) = \frac{k}{(1-k)\sqrt[k]{2}}\left(1 - e^{\frac{k-1}{k}H_k(p,q)}\right), \quad k \neq 1,\ k > 0,$$

in which, $H_k(p,q)$ is the k-Renyi divergence measure with the following form:

$$H_k(p,q) = \frac{1}{k-1}\ln\left\{E_p\left(\frac{p(X)}{q(X)}\right)^{k-1}\right\}, \quad k \neq 1,\ k > 0.$$

Corollary 6. *If the utility function be as in (3.2), then the corresponding utility divergence measure is given by:*

$$D_U(p,q) = \frac{(1+\theta)^{\frac{\theta-\lambda}{\theta}}}{\theta-\lambda}\left[\left\{\sum_{x\in\chi} q(x)\left(\frac{q(x)}{p(x)}\right)^{\frac{-\theta}{\lambda}}\right\}^{\frac{\lambda}{\theta}} - 1\right].$$

Proof: Via the similar arguments in proof of Theorem 2, on choosing $\theta = -1/\alpha$ and $\lambda = \beta/\alpha$ the divergence measures simply is proved. □

Conclusions

In this paper, we derive a version of the power utility function of the first kind and its basic properties. Also, we obtain a utility-based divergence measure by using of the power utility function of the first kind and find relationships with some other divergence measures.

Acknowledgments

The authors thank the referees and editorial board for their helpful comments which improved the presentation of this paper.

References

Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another, *Journal of the Royal Statistical Society, Series B*, **28**, 131–142.

Bell, D. E. (1988). One switch utility functions and a measure of risk, *Management Science*, **34**, 1416–1424.

Bell, D. E. (1995). A contextual uncertainty condition for behavior under risk, *Management Science*, **41**, 1145–1150.

Burbea, J. and Rao, C. R. (1982). On the convexity of some divergence measures based on entropy functions, *IEEE Transactions on Information Theory*, **28**, 489–495.

Conniffe, D. (2007). Generalized means of simple utility functions with risk aversion, *Paper Presented at Irish Economics Association Conference*.

Csiszar, J. (1967). Information type measures of differences of probability distribution and indirect observations, *Studia Scientifica Materia Hungary*, **2**, 299–318.

Dragomir, S. S. (2003). On the p-Logarithmic and α-Power divergence measures in information theory, *PanAmerican Mathematical Journal*, **13**, 1–10.

Friedman, C., Huang, J. and Sandow, S. (2007). A utility-based approach to some information measures, *Entropy*, **9**, 1–26.

Friedman, C. and Sandow, S. (2003). Model performance measures for expected utility maximizing investors, *International Journal of Theoretical and Applied Finance*, **6**, 355–401.

Gerber, H. U. and Pafumi, G. (1999). Utility functions: From risk theory to finance, *North American Actuarial Journal*, **2**, 74–100.

Havrda, J. H. and Charvát, F. (1967). Quantification method classification process: Concept of structural α - entropy, *Kybernetika*, **3**, 30–35.

Ingersoll, J. (1987). *Theory of Financial Decision Making*, Rowman & Littlefield, New York.

Johnson, T. C. (2007). Utility functions, C2922 E.

Kapur, J. N. (1984). A comparative assessment of various measures of directed divergence, *Advances in Management Studies*, **3**, 1–16.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *The Annals of Mathematical Statistics*, **22**, 79–86.

Lin, J. (1991). Divergence measures based on the Shannon entropy, *IEEE Transactions on Information Theory*, **37**, 145–151.

Mahalanobis, P. C. (1936). On the generalized distance in statistics, In *Proceedings of the National Institute of Sciences of India*, **2**, 49–55.

Pardo, L. (1993). R_h^{φ} - Divergence statistics in applied categorical data analysis with stratified sampling, *Utilitas Mathematica*, **44**, 145–164.

Rao, C. R. (1982). Diversity and dissimilarity coefficients: A unified approach, *Theoretic Population Biology*, **21**, 24–43.

Renyi, A. (1961). On measures of entropy and information, In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 547–561.

Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379–423.

Sharma, B. D. and Mittal, D. P. (1977). New non-additive measures of relative information, *Journal Combined Information Systems Science*, **2**, 122–132.

Shioya, H. and Da-te, T. (1995). A generalization of Lin divergence and the derivative of a new information divergence, *Electronics and Communications in Japan*, **78**, 37–40.

Soofi, E. S., Ebrahimi, N. and Habibullah, M. (1995). Information distinguish ability with application to analysis of failure data, *Journal of the American Statistical Association*, **90**, 657–668.

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

Wiener, N. (1949). *Cybernetics*, John Wiley & Sons, New York.

Received January 2009; Accepted April 2009