Theory Dec. (2012) 72:273–285
DOI 10.1007/s11238-010-9232-5

Aspects concerning entropy and utility

A. R. Hoseinzadeh · G. R. Mohtashami Borzadaran · G. H. Yari

Published online: 2 October 2010

Abstract Expected utility maximization problem is one of the most useful tools in mathematical finance, decision analysis and economics. Motivated by statistical model selection, via the principle of expected utility maximization, Friedman and Sandow (J Mach Learn Res 4:257–291, 2003a) considered the model performance question from the point of view of an investor who evaluates models based on the performance of the optimal strategies that the models suggest. They interpreted their performance measures in information theoretic terms and provided new generalizations of Shannon entropy and Kullback–Leibler relative entropy and called them U-entropy and U-relative entropy. In this article, a utility-based criterion for independence of two random variables is defined. Then, Markov's inequality for probabilities is extended from the U-entropy viewpoint. Moreover, a lower bound for the U-relative entropy is obtained. Finally, a link between conditional U-entropy and conditional Renyi entropy is derived.

G. R. Mohtashami Borzadaran is a member of Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad, Iran.

A. R. Hoseinzadeh
Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran
e-mail: hoseinzade1@yahoo.com

G. R. Mohtashami Borzadaran (✉)
Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran
e-mail: gmb1334@yahoo.com

G. H. Yari
Iran University of Science and Technology, Tehran, Iran
e-mail: Yari@iust.ac.ir

Keywords Expected utility maximization · U-entropy · U-relative entropy · Conditional relative U-entropy · Conditional U-entropy · Mutual U-information · Data processing inequality

1 Introduction and motivations

Statistical modeling is a critical tool in scientific research. A statistical model is a probability distribution that uses observed data to approximate the true distribution of probabilistic events. In practice, model selection and evaluation are central issues, and a crucial aspect is selecting the most appropriate model from a set of candidate models. Therefore, probabilistic model builders and model users often must choose the best model from some collection of models. There are a number of different model performance measures that arise in different contexts. For example, in the information-theoretic approach advocated by Akaike (1973), Akaike (1974), the Kullback and Leibler (1951) information discrepancy is considered as the basic criterion for evaluating the goodness of a model as an approximation to the true distribution that generates the data. The Akaike information criterion (AIC) was derived as an asymptotic approximate estimate of the Kullbackf́b-Leibler information discrepancy and provides a useful tool for evaluating models estimated by the maximum likelihood method. Burnham and Anderson (2002) provided a nice review and explanation of the use of AIC in the model selection and evaluation problems.

Probability models estimated by minimum relative entropy (MRE) methods are special in the sense that they are tailored to the risk preferences of logarithmic-family-utility investors. However, not all risk preferences can be expressed by utility functions in the logarithmic family. In the financial community, a substantial percentage, if not a majority, of practitioners implement utility functions outside the logarithmic family (see, for example, Morningstar 2002). This is not surprising that other utility functions may more accurately reflect the risk preferences of certain investors or possess important defining properties or optimality properties.

We note that these model performance measures are typically not motivated by considering the performance of an investor who relies on the models to make investment decisions. There is some question as to whether the aforementioned model performance measures are the best tools for a decision maker who uses the models to manage risk. According to the principle of expected utility maximization, a rational investor, when faced with a choice among a set of competing feasible investment alternatives, chooses the option with the greatest expected utility under the probability distribution that he believes (see, for example, Ingersoll 1987, Theorem 3, p. 31).

Motivated by the principle of expected utility maximization, Friedman and Sandow (2003a) considered the model performance question from the point of view of an investor who evaluates models based on the performance of the optimal strategies that the models suggest. In order to evaluate a particular model, they assumed that there is an investor who believes in the model. Also, they employed a utility function and a well-established concept in economics (see, for example, Von-Neumann and Morgenstern 1944). Under this new paradigm, the investor selects the model consistent with the highest estimated expected utility. They interpreted their performance measures in information theoretic terms and provided new generalizations of Shannon entropy and Kullback–Leibler relative entropy. Also, they showed that the relative performance measure is independent of the market prices if and only if the investor's utility function is a member of a logarithmic family that admits a wide range of possible risk aversions (see, for more details, Friedman and Sandow 2003a).

In this article, first, we review these generalized quantities and their properties. Then, a criterion for independence of two random variables is given from the conditional U-entropy viewpoint. Further, Markov's inequality for probabilities is extended from U-entropy perspective. Two lower bounds are obtained for the U-relative entropy and its special case, i.e., Kullback–Leibler information measure. Finally, a link between the conditional U-entropy and the conditional Renyi entropy is derived.

2 Preliminaries and background theory

Let X be a discrete random variable with support χ and probability distribution $p(x)$. The Shannon (1948) entropy of random variable X is defined by:

$$H(X) = -\sum_{x\in\chi} p(x) \ln p(x), \tag{1}$$

which is a measure of uncertainty of a discrete random variable. Clearly, in the discrete case $H(X) \geq 0$.

Suppose that $p(x)$ and $q(x)$ are two probability distributions on the common finite support χ. The relative entropy or Kullback–Leibler information measure between two probability distributions $p(x)$ and $q(x)$ is defined as:

$$K(p\|q) = \sum_{x\in\chi} p(x) \ln\left[\frac{p(x)}{q(x)}\right]. \tag{2}$$

The relative entropy is a measure of the distance between two probability distributions. It is well known that $K(p\|q) \geq 0$, the equality holds if and only if $p(x) = q(x)$ for all x in the common finite support χ.

Over the past 60 years, various generalizations of the Shannon entropy and Kullback–Leibler information measure were introduced, (see, for more details, Ullah 1996; Verdu 1998). A number of these generalization are closely related to the material in this article. For example, the Tsallis entropy, which was introduced by Cressie and Read (1984) and used for statistical decisions. A generalization of Kullback–Leibler information measure, relative Tsallis entropy, was introduced by Liese and Vajda (1987).

Renyi entropy is a one-parameter generalization of Shannon entropy. Despite its formal origin, Renyi entropy proved important in a variety of practical applications in coding theory (Campbell 1965; Aczel and Daroczy 1975; Lavenda 1998), statistical inference (Arimitsu and Arimitsu 2000, 2001), etc. Recently, Friedman and Sandow (2003a,b) and Friedman et al. (2004); Friedman et al. (2007) introduced utility-based generalizations of the Shannon entropy and Kullback–Leibler information measure.

Utility function is one of the most useful tools in decision analysis and economics. An investor's subjective probabilities numerically represent his beliefs and information, and his utilities numerically represent his tastes and preferences. Utility functions provide us with a method to measure an investor's preferences for wealth and the amount of risk he is willing to undertake in the hope of gaining greater wealth. Axiomatizations of expected utility theory have been provided several various ways by Von-Neumann and Morgenstern (1944), Herstein and Milnor (1953), Debreu (1960), and Fishburn (1989) among others. Axiomatizations of general rank-dependent utility have been provided by Nakamura (1995) and Abdellaoui (2002). Abbas (2003) introduced a relation between probability and utility based on the concept of a utility density function and showed the application of this relation via the maximum entropy principle. Expected utility maximization problems in mathematical finance and economics have been studied by Pikovsky and Karatzas (1996), Amendinger et al. (1998), Frittelli and Biagni (2005), and Gundel (2005).

2.1 Utility function, U-entropy and U-relative entropy

Let $\Re$ be a set of outcomes and Ω be the class of all probability distributions P on the set $\Re$. The essential requirement is that $\Re$ must be a well-defined set of elements. We shall not distinguish in our notation between a particular outcome $r_0 \in \Re$ and the degenerate probability distribution $P_0 \in \Omega$ which yields the reward r_0 with probability one. Therefore, we suppose that the set Ω contains all the elements of $\Re$ through degenerate probability distributions. Consider two probability distributions $P_1 \in \Omega$ and $P_2 \in \Omega$. We write $P_1 \prec P_2$ to indicate that P_2 is preferred to P_1, $P_1 \preceq P_2$ to indicate that P_1 is not preferred to P_2, and $P_1 \sim P_2$ to indicate that P_1 and P_2 are equivalent. The preference relation $\prec$ satisfies the Von Neumann and Morgenstern axioms. The definition of a utility function is as follows.

Definition 1 A real-value function U defined on the set $\Re$ is said to be a utility function if it has the following property: Let $P_1 \in \Omega$ and $P_2 \in \Omega$ be any two distributions such that both $E(U|P_1)$ and $E(U|P_2)$ exist. Then $P_1 \preceq P_2$ if and only if $E(U|P_1) \leq E(U|P_2)$, in which $E(U|P)$ is the expected utility under the probability distribution of P.

For all $r \in \Re$, the number $U(r)$ is called the utility of r. Also, for any distribution $P \in \Omega$, the number $E(U|P)$, when it exists, is often called simply the utility of P. Hence, the utility of a probability distribution is equal to the expected utility of the outcome that will be received under that distribution. We consider, throughout this article, an expected utility maximizing investor with a utility function, $U(.)$, that is strictly monotone increasing, strictly concave and twice differentiable (see, for more details, Friedman et al. 2007).

Definition 2 Let $p(x)$ and $q(x)$ be two probability distributions defined over a common finite support χ. The U-relative entropy from the probability distribution $p(x)$ to probability distribution $q(x)$ is given by:

$$D_U(p\|q) = \sup_{b(x)\in\beta_\chi} \sum_{x\in\chi} p\left(x\right) U\left(\frac{b\left(x\right)}{q\left(x\right)}\right), \tag{3}$$

where $U(.)$ is a utility function and

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

Note that $D_U(p\|q)$ is optimized when

$$b(x) = b_p^*\left(x\right) = q\left(x\right) \left(U'\right)^{-1} \left(\frac{\lambda q\left(x\right)}{p\left(x\right)}\right), \tag{4}$$

in which, λ is the solution of the following equation:

$$\sum_{x\in\chi} q\left(x\right) \left(U'\right)^{-1} \left(\frac{\lambda q\left(x\right)}{p\left(x\right)}\right) = 1. \tag{5}$$

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

$$H_U\left(X\right) = U\left(|\chi|\right) - D_U\left(p \parallel \frac{1}{|\chi|}\right), \tag{6}$$

where $q(x) = \frac{1}{|\chi|}$ and $|\chi|$ is the cardinality of the finite set χ.

For logarithmic utility function, via simple calculations, we have $b_p^*\left(x\right) = p(x)$. Therefore, Definitions 2 and 3 are reduced to the Kullback–Leibler information measure and Shannon entropy, respectively. Hence, these present generalizations for the Kullback–Leibler information measure and Shannon entropy.

Friedman et al. (2007) defined the conditional relative U-entropy, the conditional U-entropy, and the mutual U-information similar to the classical concepts of information theory. The following definitions generalize the concepts of the conditional relative entropy, the conditional entropy, and the mutual information, respectively.

Definition 4 The conditional relative U-entropy from the conditional probability distribution $p(y|x)$ to the conditional probability distribution $q(y|x)$ is given by:

$$\begin{aligned} D_U(p(y|x)\|q(y|x)) &= \sum_{x\in\chi} p_X\left(x\right) D_U(p(y|X=x)\|q(y|X=x)) \\ &= \sum_{x\in\chi} p_X\left(x\right) \sup_{b(y|x)\in\beta_y} \sum_{y\in y} p\left(y|x\right) U\left(\frac{b\left(y|x\right)}{q\left(y|x\right)}\right), \end{aligned}$$

where $p_X(x)$ is the marginal probability distribution of random variable X and

$$\beta_y = \left\{ b(y|x) : \sum_{y \in y} b\,(y|x) = 1 \right\}.$$

Definition 5 The conditional U-entropy, $H_U(Y|X)$, is defined by:

$$\begin{aligned} H_U(Y|X) &= \sum_{x \in \chi} p_X\,(x)\, H_U(Y|X = x) \\ &= U(|y|) - \sum_{x \in \chi} p_X\,(x) \sup_{b(y|x) \in \beta_y} \sum_{y \in y} p\,(y|x)\, U(|y|b(y|x)). \end{aligned}$$

Definition 6 Consider two random variables X and Y with a joint probability distribution $p(x, y)$ and marginal probability distributions $p_X(x)$ and $p_Y(y)$. The mutual U-information, $I_U(X; Y)$, is defined as:

$$I_U(X; Y) = D_U(p(x, y) \| p_X(x) p_Y(y)).$$

2.2 Properties of U-entropy and U-relative entropy

Now, we review some properties of $D_U(p\|q)$ and $H_U(X)$ established by Friedman et al. (2007), in the same manner as classical information theory.

Theorem 1 *The generalized relative entropy, $D_U(p\|q)$, and the generalized entropy, $H_U(X)$, 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) $H_U(X) \geq 0$*, and* $H_U(X)$ *is a strictly concave function of* p.
(iv) $D_U(p(x)\|q(x)) \leq D_U(p(x, y)\|q(x, y))$.
(v) $D_U(p(y|x)\|q(y|x)) \leq D_U(p(x, y)\|q(x, y))$.

According to Friedman et al. (2007), the conditional U-entropy and the mutual U-information have the following properties:

- The inequality $H_U(Z|X, Y) \leq H_U(Z|X)$ holds, which induces $H_U(Y|X) \leq H_U(Y)$. Therefore, conditioning reduces U-entropy.
- If the random variables X, Y, Z form a Markov chain $X \to Y \to Z$, then $H_U(Z|X, Y) = H_U(Z|Y)$, $H_U(Z|X) \geq H_U(Z|Y)$ and $H_U(X|Z) \geq H_U(X|Y)$.
- $I_U(X; Y) = I_U(Y; X)$ and $I_U(X; Y) \leq I_U(X; Y, Z)$.
- Data Processing Inequality: If the random variables X, Y, Z form a Markov chain $X \to Y \to Z$, then $I_U(X; Y) \geq I_U(X; Z)$ and if $Z = g(Y)$, we have $I_U(X; Y) \geq I_U(X; g(Y))$.

3 Main results

We note that a number of properties that hold for classical quantities of information theory do not retain anymore. For example, the chain rule for U-relative entropy does not retain, i.e.,

$$D_U(p(x, y)\|q(x, y)) \neq D_U(p(x)\|q(x)) + D_U(p(y \mid x)\|q(y \mid x)).$$

Also, the relationships between U-entropy and mutual U-information do not retain, i.e.

$$I_U(X; Y) \neq H_U(X) - H_U(X \mid Y) \neq H_U(Y) - H_U(Y \mid X).$$

In this section, some new results are obtained as the novelty of this article for which its special cases are discussed in classical information theory.

In the classical information theory, a criterion for independence of two random variables X and Y is the mutual information, i.e., $I(X; Y)$ that is zero if and only if X and Y are independent random variables. Moreover,

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X),$$

however, as we mentioned, similar relations in utility approach do not hold. Now, we consider $H_U(X) - H_U(X|Y)$ and introduce a criterion for independence of two random variables.

Theorem 2 *Let (X, Y) be a pair of discrete random variable with a joint probability distribution $p(x, y)$ and marginal probability distributions $p_X(x)$ and $p_Y(y)$. Define*

$$\delta_U(X; Y) = H_U(X) - H_U(X|Y),$$

(i) *For logarithmic utility function, $\delta_U(X; Y)$ is reduced to the mutual information $I(X; Y)$ which agrees with classical information theory.*
(ii) *$\delta_U(X; Y) \geq 0$ with equality if and only if X and Y are independent.*

Proof

(i) For $U(x) = \ln x$, $H_U(X)$ and $H_U(X|Y = y)$ are reduced to $H(X)$ and $H(X|Y = y)$, respectively, hence $\delta_U(X; Y)$ is reduced to $I(X; Y)$.

(ii) Note that conditioning reduces U-entropy, i.e., $H_U(X|Y) \leq H_U(X)$, therefore $\delta_U(X;Y) \geq 0$. Using definitions $H_U(X)$ and $H_U(X|Y)$ we have,

$$\begin{aligned} H_U(X) - H_U(X|Y) &= U(|\chi|) - D_U\left(p_X(x) \| \frac{1}{|\chi|}\right) - \sum_{y \in y} p_Y(y)\, H_U(X|Y=y) \\ &= U(|\chi|) - D_U\left(p_X(x) \| \frac{1}{|\chi|}\right) \\ &\quad - \sum_{y \in y} p_Y(y) \left(U(|\chi|) - D_U\left(p(x|y) \| \frac{1}{|\chi|}\right)\right) \\ &= \sum_{y \in y} p_Y(y)\, D_U\left(p(x|y) \| \frac{1}{|\chi|}\right) - D_U\left(p_X(x) \| \frac{1}{|\chi|}\right). \end{aligned}$$

If X and Y are independent, then $p(x|y) = p_X(x)$ leads to $\delta_U(X;Y) = 0$. If $\delta_U(X;Y) = 0$, then for logarithmic utility function we have $\delta_U(X;Y) = I(X;Y) = 0$, therefore X and Y are independent random variables.

□

Corollary 1 *If the random variables X, Y, Z form a Markov chain $X \rightarrow Y \rightarrow Z$, then $\delta_U(Z; X, Y) = \delta_U(Z; Y)$.*

In the classical information theory, Markov's inequality for probabilities states that if X is a discrete random variable with probability distribution $p(x)$, then for all $0 < d < 1$,

$$Pr\{p(X) \leq d\} \ln \frac{1}{d} \leq H(X). \tag{7}$$

In the next theorem this property is extended to U-entropy.

Theorem 3 *Let X be a discrete random variable with finite support χ and probability distribution $p(x)$ and let $U(.)$ be a positive utility function. Then, for all $0 < d < 1$,*

$$Pr\{p(X) \leq d\} \leq \frac{U(|\chi|) - H_U(X)}{U\left(\frac{|\chi|}{d \sum_{x \in \chi} \frac{1}{p(x)}}\right)}. \tag{8}$$

Proof Using the definition of U-entropy we have:

$$H_U(X) = U(|\chi|) - \sup_{b \in \beta_\chi} \sum_{x \in \chi} p(x)\, U(b(x)|\chi|).$$

Let $b(x) = \frac{1}{p(x)\sum_{x\in\chi}\frac{1}{p(x)}}$, it is obvious that $b(x) \in \beta_\chi$. Thus, we can write:

$$H_U(X) \le U(|\chi|) - \sum_{x\in\chi} p(x)\, U\left(\frac{|\chi|}{p(x)\sum_{x\in\chi}\frac{1}{p(x)}}\right)$$
$$= U(|\chi|) - E\left[U\left(\frac{|\chi|}{p(X)\sum_{x\in\chi}\frac{1}{p(x)}}\right)\right].$$

Hence,

$$E\left[U\left(\frac{|\chi|}{p(X)\sum_{x\in\chi}\frac{1}{p(x)}}\right)\right] \le U(|\chi|) - H_U(X).$$

Now, using Markov's inequality leads to:

$$Pr\left[p(X) \le d\right] = Pr\left[U\left(\frac{|\chi|}{p(X)\sum_{x\in\chi}\frac{1}{p(x)}}\right) \ge U\left(\frac{|\chi|}{d\sum_{x\in\chi}\frac{1}{p(x)}}\right)\right]$$
$$\le \frac{E\left[U\left(\frac{|\chi|}{p(X)\sum_{x\in\chi}\frac{1}{p(x)}}\right)\right]}{U\left(\frac{|\chi|}{d\sum_{x\in\chi}\frac{1}{p(x)}}\right)} \le \frac{U(|\chi|) - H_U(X)}{U\left(\frac{|\chi|}{d\sum_{x\in\chi}\frac{1}{p(x)}}\right)},$$

and the proof is completed. □

Corollary 2 *For* $U(x) = \ln\frac{x}{|\chi|}$ *and* $0 < d < \left[\sum_{x\in\chi}\frac{1}{p(x)}\right]^{-1/2}$, *the inequality* (8) *is reduced to* (7).

Now, we derive a link between $D_U(p\|q)$ *and* $K(q\|p)$ *based on a parametric family of exponential utility function. For a parametric family of utility functions Hoseinzadeh et al. (2009) derived some links between* $D_U(p\|q)$ *and other divergence measures.*

Theorem 4 *If* $U(x) = 1 - e^{-\alpha x}$, $\alpha > 0$, *then*

$$D_U(p\|q) = 1 - e^{-K(q\|p)-\alpha}, \tag{9}$$

where

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

Proof Considering the exponential utility function $U(x) = 1 - e^{-\alpha x}$, we obtain $(U')^{-1}(x) = \frac{-1}{\alpha}\ln\frac{x}{\alpha}$ and using (4) we achieve,

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

in which, $\ln \lambda + \ln \alpha = -\alpha - K\,(q\|p)$. After some simple calculations the proof is completed. □

For the power utility function, $U(x) = \frac{x^{1-\beta}-1}{1-\beta}, \beta \geq 0, \beta \neq 1$, Friedman et al. (2007) proved

$$D_U(p\|q) = \frac{\left[\sum_{x\in\chi} p(x)^{\frac{1}{\beta}} q(x)^{1-\frac{1}{\beta}}\right]^{\beta} - 1}{1-\beta}. \tag{10}$$

This measure is the Tsallis relative entropy.

Now, suppose that $p(x)$ and $q(x)$ are two probability distributions on the common finite support χ. In the following theorem we give a lower bound for $D_U(p\|q)$.

Theorem 5 *Let $p(x)$ and $q(x)$ be two probability distributions defined over a common finite support χ. Then*

$$D_U(p\|q) \geq \max\left\{\sum_{x\in\chi} p(x)U\left(\frac{p(x)}{\max p(x)}\right), \sum_{x\in\chi} p(x)U\left(\frac{p(x)}{\max q(x)}\right)\right\}. \tag{11}$$

Proof Using the definition of U-relative entropy we can write,

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

By setting $b(x) = \frac{p(x)q(x)}{\sum_{x\in\chi} p(x)q(x)} \in \beta_\chi$, we obtain

$$D_U\,(p\|q) \geq \sum_{x\in\chi} p(x)U\left(\frac{p(x)}{\sum_{x\in\chi} p(x)q(x)}\right).$$

Since $p(x)$ and $q(x)$ are two probability distributions, the following inequalities obviously hold

$$\frac{p(x)}{\sum_{x\in\chi} p(x)q(x)} \geq \frac{p(x)}{\max p(x)},$$
$$\frac{p(x)}{\sum_{x\in\chi} p(x)q(x)} \geq \frac{p(x)}{\max q(x)}.$$

On the other hand $U(.)$ is a strictly increasing function, so the proof is completed. □

Corollary 3 *For exponential utility function with parameter $\alpha = \max q(x)$, Theorem 4 is reduced to:*

$$K(q\|p) \geq -\max q(x) - \ln\left[\sum_{x\in\chi} p(x)e^{-p(x)}\right].$$

In the classical information theory, the conditional Shannon entropy for random variable Y, given X, with conditional probability distribution $p(y|x)$ is defined as:

$$H(Y|X) = \sum_{x \in \chi} p_X(x) H(Y|X = x) = -\sum_{x \in \chi} \sum_{y \in y} p_X(x) p(y|x) \ln p(y|x),$$

in which $p_X(x)$ is the marginal probability distribution of X. Using the conditional Shannon entropy, Cachin (1997) gave the following definition for conditional Renyi entropy:

$$H_\alpha(Y|X) = \frac{1}{1-\alpha} \sum_{x \in \chi} p_X(x) \ln \sum_{y \in y} p^\alpha(y|x).$$

In the next theorem, we derive a link between the conditional U-entropy and the conditional Renyi entropy based on the power utility function.

Theorem 6 *For the power utility function $U(x) = \frac{x^{1-\beta}-1}{1-\beta}, \beta \geq 0, \beta \neq 1$, with parameter $\beta = \frac{1}{\alpha}$,*

$$H_U(Y|X) = |y|^{1-\frac{1}{\alpha}} H_\alpha(Y|X) + O(\alpha).$$

Proof Using the definition of $H_U(Y|X)$, we have,

$$H_U(Y|X) = \sum_{x \in \chi} p_X(x) \, H_U(Y|X = x),$$

where,

$$H_U(Y|X = x) = U(|y|) - D_U\left(p(y|x) \| \frac{1}{|y|}\right).$$

Considering power utility function and relation (10), after some calculations we obtain:

$$H_U(Y|X) = \frac{|y|^{1-\beta}}{1-\beta} \left\{ 1 - \sum_{x \in \chi} p_X(x) \left[\sum_{y \in y} p^{\frac{1}{\beta}}(y|x) \right]^\beta \right\}.$$

Using the expansion:

$$\left[\sum_{y \in y} p^{\frac{1}{\beta}}(y|x) \right]^\beta = 1 + \beta \ln \left[\sum_{y \in y} p^{\frac{1}{\beta}}(y|x) \right] + O\left(\frac{1}{\beta}\right),$$

and setting $\beta = \frac{1}{\alpha}$, lead to the proof is completed. □

4 Conclusions and future work

Motivated by expected utility maximization in decision analysis and mathematical finance, we reviewed generalizations of the Shannon entropy and Kullback–Leibler information measure. We used the conditional U-entropy to present a criterion for independence of two random variables. Also, we derived a link between the conditional U-entropy and the conditional Renyi entropy. Moreover, we obtained two lower bounds for $D_U(p\|q)$ and its special case, i.e., Kullback–Leibler information measure. Markov's inequality for probabilities was extended to U-entropy. In the next step, we intend to find inequalities concerning utility-based information measures and other properties taking Dragomir (2003) and Taneja and Kumar (2004) works into account.

Acknowledgments The authors thank the Editor and anonymous referees for their helpful comments, which have greatly improved the presentation of this paper.

References

Abbas, A. (2003). Entropy methods in decision analysis. Ph.D. Thesis, Department of Management Science and Engineering, Stanford University.

Abdellaoui, M. (2002). A genuine rank-dependent generalization of the von Neumann–Morgenstern expected utility theorem. *Econometrica, 70*, 717–736.

Aczel, J., & Daroczy, Z. (1975). *On measures of information and their characterization*. New York: Academic Press.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716–723.

Amendinger, J., Imkeller, P., & Schweizer, M. (1998). Additional logarithmic utility of an insider. *Stochastic Processes and Their Applications, 75*, 263–286.

Arimitsu, T., & Arimitsu, N. (2000). Tsallis statistics and fully developed turbulence. *Journal of Physics A: Mathematical and General, 33*(27), L235–L241.

Arimitsu, T., & Arimitsu, N. (2001). Analysis of turbulence by statistics based on generalized entropies. *Physica A, 295*, 177–194.

Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference* (2nd ed.). New York: Springer.

Cachin, C. (1997). Entropy measures and unconditional security in cryptography. Ph.D. Thesis, Swiss Federal Institute of Technology, Zurich.

Campbell, L. L. (1965). A coding theorem and Renyis entropy. *Information and Control, 8*, 423–429.

Cressie, N., & Read, T. (1984). Multinomial goodness of fit tests. *Journal of the Royal Statistical Society Series B, 46*(3), 440–464.

Debreu, G. (1960). Topological methods in cardinal utility theory. In K. J. Arrow, S. Karlin & P. Suppes (Eds), *Mathematical methods in the social sciences* (pp. 16–26). Stanford, CA: Stanford University Press

Dragomir, S. S. (2003). On the p-Logarithmic and α-Power divergence measures in information theory. *Panamerican Mathematics Journal, 13*(3), 1–10.

Fishburn, P. C. (1989). Non-transitive measurable utility for decision under uncertainty. *Journal of Mathematical Economics, 18*, 187–207.

Friedman, C., & Sandow, S. (2003a). Learning probabilistic models: An expected utility maximization approach. *Journal of Machine Learning Research, 4*, 257–291.

Friedman, C., & Sandow, S. (2003b). Model performance measures for expected utility maximizing investors. *International Journal of Theoretical and Applied Finance, 6*(4), 355.

Friedman, C., Huang, J., & Sandow, S. (2004). Model performance measures for leveraged investors. *International Journal of Theoretical and Applied Finance, 7*(5), 541.
Frittelli, M., & Biagni, S. (2005). Utility maximization in incomplete markets for unbounded processes. *Finance and Stochastics, 9*, 493–517.
Friedman, C., Huang, J., & Sandow, S. (2007). A utility-based approach to some information measures. *Entropy, 9*, 1–26.
Gundel, A. (2005). Robust utility maximization for complete and incomplete market models. *Finance and Stochastics, 9*, 151–176.
Herstein, I. N., & Milnor, J. (1953). An axiomatic approach to measurable utility. *Econometrica, 21*, 291–297.
Hoseinzadeh, A. R., Mohtashami Borzadaran, G. R., & Yari, G. H. (2009). A view on extension of utility-based on links with information measures. *Communications of the Korean Statistical Society, 16*(5), 1–8.
Ingersoll, J. (1987). *Theory of financial decision making*. New York: Rowman and Littlefield.
Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79–86.
Lavenda, B. H. (1998). The analogy between coding theory and multifractals. *Journal of Physics A: Mathematical and General, 31*, 5651–5660.
Liese, F., & Vajda, I. (1987). *Convex statistical distances*. Leipzig: Teubner.
Morningstar. (2002). The new morningstar ratingTM methodology. http://www.morningstar.dk/downloads/MRARdefined.pdf.
Nakamura, Y. (1995). Rank dependent utility for arbitrary consequence spaces. *Mathematical Social Science, 29*, 103–129.
Pikovsky, I., & Karatzas, I. (1996). Anticipative portfolio optimization. *Advances in Applied Probability, 28*, 1095–1122.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423.
Taneja, I. J., & Kumar, P. (2004). Relative information of type s, Csiszar's f-divergence, and information inequalities. *Information Sciences, 166*, 105–125.
Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *Journal of the Statistical Planning and Inference, 49*, 137–162.
Verdu, S. (1998). Fifty Years of Shannon Theory. *IEEE Transactions on Information Theory, Special Commemorative Issue, 44*(6), 2057–2078.
Von-Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (2nd ed., 1947; 3rd ed., 1953). Princeton, NJ: Princeton University Press