Journal of Sciences, Islamic Republic of Iran 35(3): 233 - 241 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Bounds for the Varentropy of Basic Discrete Distributions and Characterization of Some Discrete Distributions

F. Goodarzi*

Department of Statistics, Faculty of Mathematical Sciences, University of Kashan, Kashan, Islamic Republic of Iran

Received: 9 July 2024 / Revised: 31 December 2024 / Accepted: 1 February 2025

Abstract

Given the importance of varentropy in information theory, and since a closed form cannot be derived for some discrete distributions, we aim to establish bounds for the varentropy of these distributions and introduce the past varentropy for discrete random variables. In this article, we first acquired lower and upper bounds for the varentropy of the Poisson, binomial, negative binomial, and hypergeometric distributions. Since the resulting upper bounds are expressed as squared logarithmic expectations, we provide an equivalent formulation using squared logarithmic difference coefficients. Similarly, we present lower bounds in terms of logarithmic difference coefficients. Furthermore, an upper bound is derived for the variance of a function of discrete reversed residual lifetime function. We also investigate inequalities involving moments of selected functions via the reversed hazard rate and characterize certain discrete distributions by the Cauchy-Schwarz inequality.

Keywords: Varentropy; Reversed hazard rate; Binomial transform; Cauchy-Schwarz inequality.

Introduction

If X is absolutely continuous with probability density function $f(x)$, then the entropy of X is given by

$$H(X) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx, \tag{1}$$

where $-\log f(X)$ is the information content of X. Notably, the existence of $H(X)$ is not guaranteed. When it exists, its values range belongs to $[-\infty, \infty]$, while the entropy of discrete random variables (RVs) does not take negative values.

It is noteworthy that the variance entropy (for short varentropy) of a RV X is given by

$$V(X) = \int_{-\infty}^{+\infty} f(x)[\log f(x)]^2 dx - [H(X)]^2. \tag{2}$$

The importance of this measure in the fields of mathematics and physics has been emphasized in various studies, such as those by (1), (2), and (3).

As an application of varentropy, we consider a system with complex network. A complex network, in reality, contains a large amount of information necessary to describe the system's behaviors. (1) stated that varentropy is utilized as a general measure of probabilistic uncertainty for a complex network in terms of the laws of thermodynamics. Next, we will mention the application of variance of entropy in computer science. One of the most significant threats internet users and cloud computing services face is denial-of-service (DDoS) attacks. The nonlinear time series model is employed to predict future network traffic states by (4) and used to predict the future values of entropy variance.

* Corresponding Author: Tel: 03145249136; Email: f-goodarzi@kashanu.ac.ir

Also, they determined prediction errors by comparing the actual variance of entropy and the predicted variance of entropy. (5) have derived an explicit formula of the varentropy measure for the invariant density of one-dimensional ergodic diffusion processes.

Furthermore, (3) found an optimal varentropy bound applicable to log-concave distributions. (2) obtained a sharp varentropy bound on Euclidean spaces for convex probability measures. Another method to compute a bound for varentropy is given in (6) and (7) via reliability theory. (8) proposed the concept of varentropy for doubly truncated RVs and extensively analyzed its theoretical properties. A method for computing varentropy measure for the order statistics is introduced by (9). (10) introduced the variance residual entropy measure. (11) and (12) obtained bounds for past varentropy for continuous RV. Also, (13) obtained a bound for residual varentropy of discrete RV. Moreover, (14) recently offered a few estimators for varentropy for a continuous RV. The lossless source coding research (15) stated that the source dispersion equals its varentropy for Markov sources.

Suppose X is a discrete RV supporting $S = \{0,1,\dots,b\}$, where b is an integer and $0 < b \leq \infty$. If we express the probability mass function (PMF) and distribution function of X by $p(x)$ and $F(x)$, respectively, then, in comparison with (1) and (2), the entropy and varentropy of a nonnegative discrete RV X are given as follows.

$$H(X) = -\sum_{x=0}^{\infty} p(x)\log p(x), \quad (3)$$

$$V(X) = \sum_{x=0}^{\infty} p(x)[\log p(x)]^2 - [H(X)]^2. \quad (4)$$

The entropy of a discrete RV is the average amount of information, measured in bits, gained from observing a single symbol.

Characterizations of distributions are essential to many researchers in applied fields. In particular, in reliability theory, given an RV that often denotes a unit's lifetime, aging functions are assigned to it and characterize this variable. Among the most used are reversed failure rate and reversed mean residual life.

One can define the reversed hazard rate of X as

$$\varphi(x) = P(X = x|X \leq x) = \frac{p(x)}{F(x)},$$

hence, $F(x)$ is specified as follows

$$F(x) = \prod_{t=x+1}^{b} (1 - \varphi(t)), \quad x = 0,\dots,b-1. \quad (5)$$

Also, the reversed mean residual lifetime is given by

$$r(x) = E(x - X|X < x) = \frac{1}{F(x-1)}\sum_{t=1}^{x} F(t-1), \quad (6)$$

with defining $r(0) = 0$. See (16) for more details.

Definition 1. (a) F is said to be decreasing reversed hazard rate (DRHR) if $\varphi(x)$ is decreasing in x.

(b) F is said to increase expected inactivity time (IEIT) if $r(x)$ increases in x.

To derive variance bounds for functions of RVs, we employ Chernoff's inequality. For a discrete RV X with PMF $p(x)$, $x = 0,1,2,\dots$, bounds for $Var[g(X)]$ can be obtained using the forward difference of $g(X)$. Notably, these bounds were derived utilizing the Cauchy-Schwarz (C-S) inequality. We utilize the following lemma to derive these bounds, as presented in (17).

Lemma 2. Let X be a nonnegative and integer-valued RV with probability function $p(x)$ *with support* $\{0,1,2,\dots\}$ and let its mean be μ. Additionally, let $g(X)$ be a real-valued function with $Var[g(X)] < \infty$. Then

$$\sigma^2 E^2[w(X)\Delta g(X)] \leq Var[g(X)] \leq \sigma^2 E[w(X)(\Delta g(X))^2], \quad (7)$$

where $\Delta g(x) = g(x+1) - g(x)$ and $w(x)$ satisfies

$$\sigma^2 p(x)w(x) = \sum_{k=0}^{x} (\mu - k)p(k). \quad (8)$$

The equality satisfies iff g is linear.

The layout of the article is as follows. In Section 1, we compare two sequences by the coefficient of variation for coding a discrete source of information with three symbols and also define past varentropy for discrete RVs and obtain it by past entropy of order ζ for the discrete case. In Section 2, we get an upper and lower bound for the varentropy of the binomial, Poisson, negative binomial, and hypergeometric distribution. An upper bound for the variance of a function of the discrete reversed residual life RV is obtained in Section 3. Furthermore, we characterize some distributions through functions that ensure reliability for discrete RVs.

Coefficient of Variation and Past Varentropy

The significance of entropy is widely recognized in information theory and various other fields. However, varentropy has received comparatively less attention. Notably, the discrete entropy (3) quantifies the average number of symbols needed to code an event generated by an information source governed by the PMF of X. Varentropy, on the other hand, quantifies the variability associated with this coding. If the entropy of two sources of information is identical, then, during coding, the number of digits needed for the codeword of a symbol is closer to the expected value for the source with the lower varentropy.

Example 1.1. *Suppose that X has the PMF* $p(0) = \frac{1}{2}$ *and* $p(1) = p(2) = \frac{1}{4}$. *Also, let Y have Poisson distribution with parameter λ; then, it is easily calculated that* $\lambda \approx 0.620675$, *we have* $H(Y) \approx 1.039721$ *and* $V(Y) = 0.515302$. *Moreover,* $H(X) \approx 1.039721$ *and*

$V(X) = 0.120112$; *hence, the coding process is better suited for sequences produced by X.*

In the process of coding a discrete source of information with three symbols with probabilities p, q, and $1-p-q$, the quantifiers entropy and varentropy, respectively, are stated as:

$$H(p,q) = H(X) = -p\log p - q\log q - (1-p-q)\log(1-p-q), \tag{9}$$

$$V(p,q) = V(X) = (p-p^2)(\log p)^2 + (q-q^2)(\log q)^2 + (1-p-q-(1-p-q)^2)(\log(1-p-q))^2 - 2pq\log p\log q - 2p(1-p-q)\log p\log(1-p-q) - 2q(1-p-q)\log q\log(1-p-q). \tag{10}$$

We sketch $H(X)$ and $SD(X) = \sqrt{V(X)}$ defined over p and q in Figure 1. Regarding p and q, seven limit cases have no varentropy. These points are $(0,0)$, $(0,0.5)$, $(1/3,1/3)$, $(0.5,0)$, $(0.5,0.5)$, $(0,1)$, and $(1,0)$. Notice that varentropy would be zero in case (1/3,1/3), with equiprobable sequences and maximum entropy.

Now, we want to check the maximum variability in the information content. For this purpose, we are looking into the behavior of $\frac{dSD(p,q)}{dp}$ and $\frac{dSD(p,q)}{dq}$. By setting these terms equal to zero, we have

$$(\log p)^2 + 2\log p - (\log(1-p-q))^2 - 2\log(1-p-q) + (2(p\log p + q\log q + (1-p-q)\log(1-p-q)))(\log(1-p-q) - \log p) = 0, \tag{11}$$

and

$$(\log q)^2 + 2\log q - (\log(1-p-q))^2 - 2\log(1-p-q)) + (2(p\log p + q\log q + (1-p-q)\log(1-p-q)))(\log(1-p-q) - \log q) = 0. \tag{12}$$

Note that the seven points mentioned earlier have infinite derivative values (singular points). Thus, we apply the Newton-Raphson algorithm to obtain approximate roots of derivatives given in (11) and (12) (see (18)). The values $p = 0.0616518191$ and $q = 0.0616518191$ were obtained with an initial value $(0.06,0.06)$ to start the algorithm. It is clear that the points $(0.0616518191, 0.8766963618)$ and $(0.8766963618, 0.0616518191)$ also maximize $SD(p,q)$ and their values is 0.8728128309.

Additionally, we consider the intersection curves of the two surfaces of Figure 1, where $H(p,q) = SD(p,q)$. The intersection areas can be shown in Figure 2 using the implicit plot function in Maple. For example, if $p = 0.2$, then entropy and the standard deviation of the entropy are equal for values of q equal to 0.06929839562 and 0.7307016044. The range between the curves in Figure 2, $SD(p,q)$ is less than $H(p,q)$, whereas, in the points outside of this region, for example, $(p,q) = (0.1,0.1)$, the entropy smaller than the standard deviation of the information content of RV. Now, considering the coefficient of variation of $-\log p(X)$, such that described as $CV(X) = \frac{SD(X)}{H(X)}$, if two sequences of symbols are generated by X and Y , a sequence with less coefficient of variation is more suitable for coding.

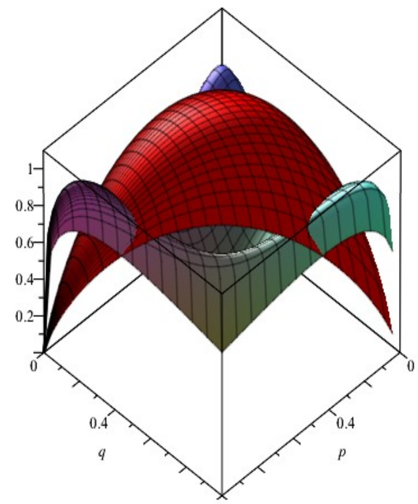

Figure 1. Plots of $H(X)$ and $SD(X)$ on p and q.

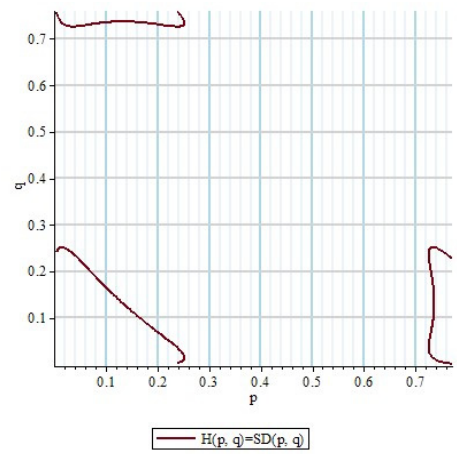

Figure 2. The curve of intersection of the two surfaces.

Rényi entropy of order ζ for a discrete RV is expressed as $H_\zeta(x) = \frac{1}{1-\zeta}\log\sum_x p^\zeta(x)$ for $\zeta \neq 1$.

$H_\zeta(X)$ is additionally named the spectrum of Rényi information. Rényi information and the loglikelihood are related via the gradient, $\dot{H}_\zeta(X)$, of the spectrum at $\zeta = 1$. A straightforward computation demonstrates, assuming that the differentiation operations are legitimate, that

$$\dot{H}_1(X) =$$

$$\lim_{\zeta\to 1}\frac{(1-\zeta)(\sum_x p^\zeta(x))^{-1}\sum_x (p^\zeta(x)\log p(x))+\log\sum p^\zeta(x)}{(1-\zeta)^2}$$
$$=-\frac{1}{2}\lim_{\zeta\to 1}\left\{\frac{\sum_x p^\zeta(x)\log^2 p(x)}{\sum_x p^\zeta(x)}-\left(\frac{\sum_x p^\zeta(x)\log p(x)}{\sum_x p^\zeta(x)}\right)^2\right\}=-\frac{1}{2}V(X). \quad (13)$$

Therefore, the varentropy is obtained as $V(X)=-2\dot{H}_1(X)$. In addition, the discrete past entropy is defined as

$$H(X;j)=-\sum_{x=0}^{j}\frac{p(x)}{F(j)}\log\left(\frac{p(x)}{F(j)}\right). \quad (14)$$

The past entropy of order ζ for a discrete case is expressed by

$$H_\zeta(X;j)=\frac{1}{1-\zeta}\log\left[\sum_{x=0}^{j}\left(\frac{p(x)}{F(j)}\right)^\zeta\right], \quad (15)$$

for $\zeta\neq 1$ and $\zeta>0$. It is well known that when ζ tends to 1, $H_\zeta(X;j)$ tends to $H(X;j)$. Similarly, also, we can show that $V(X;j)=-2\dot{H}_1(X;j)$, in which we call $V(X;j)$,

$$V(X;j)=\sum_{x=0}^{j}\frac{p(x)}{F(j)}\left(\log\frac{p(x)}{F(j)}\right)^2-(H(X;j))^2, \quad (16)$$

as the past varentropy.

Example 1.2. *If X is distributed geometrically with parameter p, then*

$$H_\zeta(X;j)=\frac{1}{1-\zeta}\log\left[\sum_{x=0}^{j}\frac{p^\zeta q^{x\zeta}}{(1-q^{j+1})^\zeta}\right]=\frac{1}{1-\zeta}\left\{\log\frac{p^\zeta}{1-q^\zeta}+log\frac{1-q^{(j+1)\zeta}}{(1-q^{j+1})^\zeta}\right\}, \quad (17)$$

where $q=1-p$ *and therefore,*

$$V(X;j)=-2\lim_{\zeta\to 1}\dot{H}_\zeta(X;j)=\frac{q(\ln q)^2}{(1-q)^2}-\frac{((j+1)\ln q)^2 q^{j+1}}{(1-q^{j+1})^2}. \quad (18)$$

It is observed that, for $j=0$, the past varentropy is zero and increases for j, as shown in Figure 3.

To estimate $V(X;j)$, we generate a sample of size $n=100$ from a geometric distribution with 1000 replicates. For this sample, we set $p_0=0.6$. Then, the Maximum Likelihood Estimator (MLE) for $\hat{p}$ is calculated to be 0.5978. For example by plugging $\hat{p}$ into (18) for j=1, the MLE of $V(X;j)$ is 0.1697.

Like the discrete case, (19) has previously obtained a relationship between varentropy and Rényi information for continuous RV. He expressed that varentropy can identify a distribution's shape, while the kurtosis measure is not applicable.

Bounds for Varentropy

Obtaining expressions for the entropy and varentropy of well-known distributions is significant in information

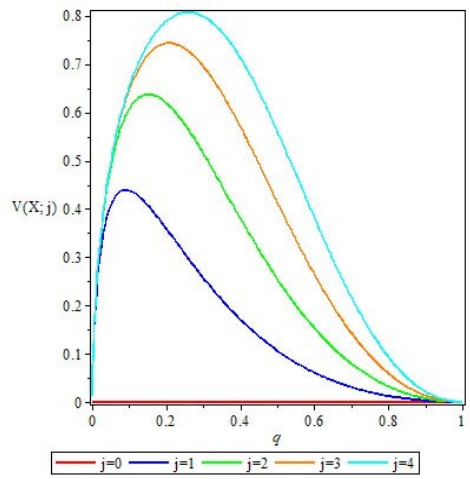

Figure 3. The curve of intersection of the two surfaces.

and communication theory, physics, probability and statistics, and economics. An exact expression and closed form for the varentropy were obtained for most distributions. Among these distributions, we can mention the uniform, Bernoulli, geometric, exponential, Beta, Cauchy, Cramér, F, gamma, Gumbel, Laplace, Lévy, logistic, log-logistic, lognormal, normal, parabolic, Pareto, power exponential, t-distribution, triangular, von Mises and Weibull distribution. However, for many distributions, there is no closed form and an explicit expression for the varentropy using elementary functions. In such cases, we can obtain an upper and lower bound for varentropy via the expectation of a function of a logarithmic function.

In this section, we find bounds for the varentropy of some nonnegative RVs. If X follows a discrete nonnegative RV, with variance σ^2, then by utilizing Lemma 2, we have

$$\sigma^2 E^2[w(X)\Delta\log p(X)]\leq Var[-\log p(X)]\leq \sigma^2 E[w(X)(\Delta\log p(X))^2]. \quad (19)$$

Example 2.1. *Suppose X has a binomial distribution distribution* $Bin(n,p)$ *then, since* $w(x)=\frac{n-x}{n(1-p)}$, *hence*

$$V(X)\leq np(1-p)\sum_{x=0}^{n}\frac{n-x}{n(1-p)}\left(-\log\frac{(n-x)p}{(x+1)(1-p)}\right)^2\binom{n}{x}p^x(1-p)^{n-x}=np(1-p)E_{n-1}\left[(\log\frac{(n-X)p}{(X+1)(1-p)})^2\right], \quad (20)$$

where E_{n-1} denotes expected value under the binomial distribution $Bin(n-1,p)$.

Likewise, we can derive a lower bound for

$Var[-\log p(X)]$ as follows

$$V(X) \geq np(1-p)E_{n-1}^2\left[\log\frac{(n-X)p}{(X+1)(1-p)}\right]. \quad (21)$$

If $n=1$, then, since $-\log p(x)$ has a linear relation with x, hence the upper and lower bounds are equal to varentropy $p(1-p)(-\log(\frac{p}{1-p}))^2$ in the Bernoulli distribution.

Example 2.2. *Assume X is distributed according to a Poisson distribution with parameter $\lambda = 1$, then since $w(x) = 1$, the upper bound for varentropy is given as*

$$V(X) \leq \lambda E\left[\log\frac{X+1}{\lambda}\right]^2. \quad (22)$$

When $n \to \infty$ and $p \to 0$ so that $np = \lambda$, the upper bound (21) and (22) are approximately equal. Also,

since $\log x \leq x-1$, we can obtain the upper bound $1+\frac{1}{\lambda}$ for varentropy of Poisson distribution

Conversely, the lower bound for $V(X)$ is computed as follows

$$V(X) \geq \lambda E^2\left[\log\frac{X+1}{\lambda}\right]. \quad (23)$$

In this section, we compute the equivalent expressions for the upper and lower bounds of varentropy according to series and integral expressions. To achieve a general expression for expectation of squared logarithmic function, that is, expressions like $E[\log^2(X+\omega)]$, we recall the ith forward difference of a function $g(\omega)$ is defined as

$$\Delta^i[g](\omega) := \sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}g(k+\omega), \quad (24)$$

where $\Delta^0[g](\omega) = g(\omega)$. Moreover, Newton series expansion of a function g around point ω is

$$g(k+\omega) = \sum_{i=0}^{\infty}\binom{x}{i}\Delta^i[g](\omega). \quad (25)$$

By considering $g(x) = (\log x)^2$ in equation (24), as (20) stated for $\log x$, we have

$$\Delta^i[\log]^2(\omega) = \sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}\log^2(k+\omega) = (-1)^{i+1}d_\omega(i+1), \quad (26)$$

where

$$d_\omega(i) = -\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\log^2(k+\omega),$$

so that, by using (26), we can obtain

$$(\log(x+\omega))^2 = \sum_{i=0}^{\infty}\binom{x}{i}(-1)^{i+1}d_\omega(i+1).$$

Now, to find the generating function for the coefficient d_ω, we use of Lerch transcendent such that it was recalled by (20), as follows:

$$\Phi(z,s,\omega) := \sum_{k=0}^{\infty}\frac{z^k}{(k+\omega)^s} = \frac{1}{\Gamma(s)}\int_0^{+\infty}\frac{t^{s-1}e^{-\omega t}}{1-ze^{-t}}dt. \quad (27)$$

By evaluating the second derivative of (27) for *s*, we have

$$\frac{d^2}{ds^2}\Phi(z,s,\omega) = \sum_{k=0}^{\infty}z^k(k+\omega)^{-s}(\log(k+\omega))^2 \quad (28)$$

and thus

$$\Phi''_\omega(z) := \frac{d^2}{ds^2}\Phi(z,s,\omega)|_{s=0} = \sum_{k=0}^{\infty}\log^2(k+\omega)z^k. \quad (29)$$

Next, putting $\omega = 1$ and using polylogarithm function $Li_s(z) := \sum_{k=1}^{\infty}z^k k^{-s}$, (29) can be written as

$$\Phi''(z) := \Phi''_1(z) = \frac{d^2}{ds^2}Li_s(z)/z|_{s=0}.$$

In fact, we have

$$\frac{d^2}{ds^2}Li_s(z)/z = \sum_{k=1}^{\infty}z^{k-1}(-\log k)^2k^{-s}. \quad (30)$$

At the same,

$$\Phi''(z) := \Phi''_1(z) = \sum_{k=0}^{\infty}z^k(\log(k+1))^2 = \sum_{k=1}^{\infty}z^{k-1}(\log(k))^2,$$

which is the equation of (30) for *s* = 0.

Using (27) and generating the function of the binomial transform, we get

$$D_\omega(z) = -\sum_{i=1}^{\infty}z^i\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\log^2(k+\omega) = \sum_{k=0}^{\infty}\sum_{i=k+1}^{\infty}(-1)^{k+1}z^i\binom{i-1}{k}\log^2(k+\omega) = \sum_{k=0}^{\infty}\log^2(k+\omega)(-1)^{k+1}\frac{z^{k+1}}{(1-z)^{k+1}} = \frac{-z}{1-z}\Phi''_\omega(\frac{-z}{1-z}).$$

Consider now the coefficient sequence $(-d_\omega(i+1))_{i=0}^{\infty}$, that is, the binomial transform of the sequence $(\log^2(k+\alpha))_{k=0}^{\infty}$. Let

$D_\omega(z) := \sum_{i=0}^{\infty}d_\omega(i)z^i$,

be the generating function for $(d_\omega(j))_{j=0}^{\infty}$, where $d_\omega(0)$ is defined as 0.

$$E\log^2(X+\omega) = \sum_{i=0}^{\infty}E[\frac{(X)_i}{i!}](-1)^{i+1}d_\omega(i+1) = \sum_{i=1}^{\infty}(-1)^i q(i-1)d_\omega(i).$$

The moment generating function $M(t)$ of the Poisson distribution is $\exp(\lambda(e^t-1))$, so as given in Theorem 1 in (20), we have

$$Q(z) = M(\log(z+1)) = e^{\lambda z} = \sum_{i=0}^{\infty}\frac{\lambda^i}{i!}z^i,$$

and hence $q(i) = \frac{\lambda^i}{i!}$. Furthermore, by using the equation

$Elog(X+\omega) = \sum_{i=1}^{\infty} (-1)^i q(i-1)c_\omega(i) = \int_0^{+\infty} \frac{e^{-t}-e^{-\omega t}M(-t)}{t}dt,$

in (20), where $c_\omega(i) = -\sum_{k=0}^{i-1} (-1)^k \binom{i-1}{k}\log(k+\omega)$, the lower bound of varentropy of Poisson distribution is given as follows

$$\lambda E^2\left[\log\frac{X+1}{\lambda}\right] = \lambda\left\{\sum_{i=1}^{\infty}\frac{(-\lambda)^{i-1}}{(i-1)!}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\log\left(\frac{k+1}{\lambda}\right)\right]\right\}^2 = \lambda\left\{\int_0^{\infty}\frac{e^{-t}}{t}(1-e^{\lambda(e^{-t}-1)})dt - \log\lambda\right\}.$$

Moreover, the upper bound for the varentropy of Poisson distribution is

$$\lambda E\left[\log\frac{X+1}{\lambda}\right]^2 = \lambda\left\{\sum_{i=2}^{\infty}\frac{(-\lambda)^{i-1}}{(i-1)!}\left[\sum_{k=0}^{i-1}(\log^2(k+1) - 2\log\lambda\log(k+1))\right] + \log^2(\lambda)\right\}$$
$$= \lambda\left\{\sum_{i=1}^{\infty}\frac{(-\lambda)^{i-1}}{(i-1)!}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}(\log(\frac{k+1}{\lambda}))^2\right]\right\}.$$

Example 2.3. *Let X follow a negative binomial distribution with a PMF* $p(x) = \binom{x+r-1}{r-1}p^r q^{x-r}$ *for* $x =$ 0, 1, … . *Then, since* $w(x) = p(1+\frac{x}{r})$, *the upper bound for varentropy is computed by*

$$V(X) \le \frac{r(1-p)}{p^2}\sum_{x=0}^{\infty}p(1+\frac{x}{r})(\log\frac{(r+x)(1-p)}{x+1})^2\binom{x+r-1}{r-1}p^r(1-p)^x = \frac{r(1-p)}{p^2}E_{r+1}\left[\log\frac{(1-p)(r+X)}{X+1}\right]^2, \quad (31)$$

where E_{r+1} *is the expected value of negative binomial distribution with parameters* parameters $r+1$ and p.

The lower bound for the distribution is determined as

$$V(X) \ge \frac{r(1-p)}{p^2}\left(E_{r+1}\left[\log\frac{(1-p)(r+X)}{X+1}\right]\right)^2. \quad (32)$$

It is trivial that if X has a geometric distribution with parameter p, *then varentropy is equal to the upper and lower bounds given in* (31) *and* (32) *for r* = 1 *and hence* Var[−log$p$(X)] $=\frac{1-p}{p^2}(\log(1-p))^2$.

Now, by using equation (44) in (20), we can obtain an equivalent expression for the lower bound (32). We first have

$$E_{r+1}\left[\log\frac{(1-p)(r+X)}{X+1}\right] = \sum_{i=1}^{\infty}(-\frac{1-p}{p})^{j-1}\binom{i+r-1}{i-1}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\left\{\log(k+r)+\log\left(\frac{k+1}{1-p}\right)\right\}\right],$$

and therefore

$$Var[-\log p(X)] \ge \frac{r(1-p)}{p^2}\left(\sum_{i=1}^{\infty}(-\frac{1-p}{p})^{i-1}\binom{i+r-1}{i-1}\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\left\{\log\frac{(1-p)(k+r)}{k+1}\right\}\right]\right)^2.$$

Also

$$Var[-\log p(X)] \le \frac{r(1-p)}{p^2}\left(\sum_{i=1}^{\infty}(-\frac{1-p}{p})^{i-1}\binom{i+r-1}{i-1}\right)\left[\sum_{k=0}^{i-1}(-1)^k\binom{i-1}{k}\left\{\log\frac{(1-p)(k+r)}{k+1}\right\}^2\right]$$

Example 2.4. *Let X have a hypergeometric distribution with PMF* $p(x) = \frac{\binom{m}{x}\binom{n-m}{r-x}}{\binom{n}{r}}$, $max(0, r-n+m) \le x \le min(r,m)$. *Then, since* $w(x) = \frac{n(n-1)(m-x)(r-x)}{(n-m)(n-r)mr}$, *the upper bound for varentropy is computed by*

$$Var[-\log p(X)] \le \sigma^2\sum_{x=0}^{r}\frac{n(n-1)(m-x)(r-x)}{(n-m)(n-r)mr}\log(\frac{(m-x)(r-x)}{(x+1)(n-m-r+x+1)})\frac{\binom{m}{x}\binom{n-m}{r-x}}{\binom{n}{r}} = \sigma^2 E_{m-1,n-2;r-1}\left[\log(\frac{(m-X)(r-X)}{(X+1)(n-m-r+X+1)})\right]^2,$$

where $E_{m-1,n-2;r-1}$ denotes expected value under the hypergeometric distribution with parameters $m-1$, $n-2$, and $\sigma^2 = \frac{rm}{n}(1-\frac{m}{n})\frac{n-m}{n-1}$.

Characterization by Cauchy-Schwarz Inequality

(13) attained an upper bound for the variance of a function of the residual lifetime RV and characterized the type III and type I discrete Weibull distributions and the geometric distribution with the help of C-S inequality. Here, we derive a bound for the variance of a function of RV $X_x = (x-X|X<x)$ and characterize some distributions using inequalities involving the expectation of functions of reversed hazard rate.

The subsequent theorem gives an upper bound for $Var[g(X_x)]$ and characterizes the right-truncated geometric distribution.

Theorem 3.1. *Let X be a discrete and nonnegative RV with PMF* $p(x)$ *and distribution function* $F(x)$. *Suppose g is a function such that its forward difference is* $\Delta g(x)$ *then*

$$Var[g(X_x)] \le E\left[(\Delta g(X_{(x)}))^2(\frac{1}{\varphi(x-X_{(x)})})(r(x-\right.$$

$X_{(x)}+1)-r(x)+X_{(x)}-1)\Big].$ (33)

Proof. We know

$$P(X_{(x)}=t)=\frac{P(X=x-t)}{P(X<x)},\qquad t=1,\dots,x.$$

By applying Lemma 2 and noting that $E[X_{(x)}]=r(x)$, it follows that

$$\sum_{k=t}^{x}(k-r(x))\frac{P(X=x-k)}{P(X<x)}=\frac{1}{P(X<x)}[\sum_{k=t}^{x}kP(X=x-k)-r(x)\sum_{k=t}^{x}P\{X=x-k\}]$$
$$=\frac{P\{X<x-t+1\}}{P\{X<x\}}[r(x-t+1)+t-1-r(x)]$$
$$=\frac{1}{\varphi(x-t)}[r(x-t+1)+t-1-r(x)]\frac{P\{X=x-t\}}{P\{X<x\}},\qquad (34)$$

and again using Lemma 2 and replacing the right-hand side of (34) in inequality (7), we obtain

$$Var[g(X_x)]\le\sum_{t=1}^{x}[\Delta g(t)]^2\frac{1}{\varphi(x-t)}[r(x-t+1)+t-1-r(x)]P\{X_{(x)}=t\}$$
$$=E\left\{[\Delta g(X_{(x)})]^2\frac{1}{\varphi(x-X_{(x)})}[r(x-X_{(x)}+1)+X_{(x)}-1-r(x)]\right\},$$

Let $g(t)=-\log\frac{p(x-t)}{F(x-1)}$, then $\Delta g(t)=-\log\frac{p(x-t-1)}{p(x-t)}=\log(1-\eta_{x-t-1})$, hence

$$Var[-\log p(X_{(x)})]\le E\left\{[\log(1-\eta_{x-X_{(x)}-1})]^2\frac{1}{\varphi(x-X_{(x)})}[r(x-X_{(x)}+1)+X_{(x)}-1-r(x)]\right\}.\qquad (35)$$

Under Lemma 2, the above equality holds iff $g(t)=-\log p(x-t)+\log F(x-1)$ is linear in t, which is equivalent to $\log p(x-t)$ being linear in t. Consequently, $\log p(x-t)=a_1t+b_1$ for some constants a_1 and b_1, and therefore $p(y)=e^{-a_1y}e^{a_1x+b_1}=de^{-a_1y}$ for $y=0,\dots,x-1$, where $d=e^{a_1x+b_1}$ is a constant.

We thus conclude that the equality holds in (35), iff $p(x)=\frac{c-1}{c^y-1}c^x, x=0,\dots,y-1,\ c>0$, which is the right truncated geometric distribution.

Remark 3.2. In Theorem 3.1, if X is a nonnegative RV and F is *DRHR,* then since the *DRHR* property implies the *IEIT* property (21), then

$$Var[g(X_{(x)})]\le E\left[(\Delta g(X))^2(\frac{1}{\varphi(X)}-1)(X-1)\right].\qquad (36)$$

Next, we aim to characterize certain distributions. Throughout the theorems presented below, we assume that Z is a discrete RV with a finite support $S=\{0,1,\dots,b\}$.

Given that $E(\frac{1}{\varphi(Z)})=b+1-E(Z)$, we can derive a useful lower bound for $E[\varphi(Z)]$, as presented in the next theorem.

Theorem 3.3. For any nonnegative discrete RV Z,

$$E[\frac{1}{\varphi(Z)}]\ge\frac{1}{E(\varphi(Z))}.\qquad (37)$$

The equality satisfies iff for constant θ

$$F(z)=\begin{cases}(1-\theta)^{b-z}, & z=0,1,\dots,b,\quad 0<\theta<1,\quad b<\infty,\\ 1, & x\ge b.\end{cases}\qquad (38)$$

Proof. To achieve (37), we make use of C-S inequality. The equality in (37) satisfies iff there's a positive constant A so that, for all $z\in\{0,1,\dots,b\}$,

$$\frac{\sqrt{P(Z=z)}}{\sqrt{\varphi(z)}}=A\sqrt{\varphi(z)P(Z=z)},$$

which is equivalent to $\varphi(z)=\theta=constant$. Now, using (5), we have a (38).

Theorem 3.4. Let Z be a nonnegative discrete RV. Then

$$E[\frac{\varphi(Z)}{Z}]\ge\frac{2}{b(b+1)-E(Z(Z-1))},$$

with equality iff Z distributed as,

$$F(z)=\begin{cases}\prod_{t=z+1}^{b}(1-\theta t), & z=0,1,\dots,b-1,\quad 0<\theta<\frac{1}{b},\\ 1, & z\ge b.\end{cases}\qquad (39)$$

where θ is a constant.

Proof. By the C-S inequality, we have

$$1=(\sum_{z=0}^{b}P\{Z=z\}\sqrt{\frac{zF(z)}{zF(z)}})^2\le\sum_{z=0}^{b}\frac{P^2\{Z=z\}}{zF(z)}\sum_{z=0}^{b}zF(z)=\sum_{z=0}^{b}\frac{\varphi(z)}{z}P\{Z=z\}(\sum_{z=0}^{b}z-\sum_{z=0}^{b}zP\{Z>z\}).\qquad (40)$$

Now, since

$$\sum_{z=0}^{b}zP\{Z>z\}=E\left(\frac{Z(Z-1)}{2}\right),$$

(40) reduces to

$$1\le E(\frac{\varphi(Z)}{Z})[\frac{b(b+1)}{2}-E\left(\frac{Z(Z-1)}{2}\right)],$$

and the desired result is obtained. The equality is gotten iff there's some positive constant θ so that

$$\frac{P(Z=z)}{\sqrt{zF(z)}}=\theta\sqrt{zF(z)}.$$

It follows that $\varphi(z)=\theta z$, which, using equation (5), again satisfies iff Z has distribution given in equation (39).

The following two theorems derive lower bounds for $E(Z\varphi(Z))$.

Theorem 3.5. *Let Z be a discrete RV with $E(Z\varphi(Z))<\infty$ and $E(\frac{1}{Z\varphi(Z)})<\infty$. Then*

$$E[\frac{1}{Z\varphi(Z)}]\geq\frac{1}{E(Z\varphi(Z))}, \quad (41)$$

and equality holds iff Z is distributed as

$$F(z)=\begin{cases}\frac{(b-\theta)!z!}{b!(z-\theta)!}, & z=\theta,\dots,b-1,\quad \theta=1,2,\dots,b-1\\ 1, & z\geq b.\end{cases}$$

Proof. As in the proof of Theorem 3.4, the result is stablished.

Theorem 3.6. For any nonnegative discrete RV Z,

$$E(Z\varphi(Z))\geq\frac{2E^2(Z)}{b(b+1)-E(Z(Z-1))}. \quad (42)$$

The equality satisfies iff Z has the distribution function (38).

Proof. By the C-S inequality, we find that

$$(E(Z))^2\leq E(\frac{Z}{\varphi(Z)})E(Z\varphi(Z))$$
$$=[\sum_{z=0}^{b}zF(z)]E(Z\varphi(Z))$$
$$=[\frac{b(b+1)-E(Z(Z-1))}{2}]E[Z\varphi(Z)],$$

and thus (42) is obtained.

The equality satisfies iff there exists some nonnegative constant A so that, for all $z\in\{0,\dots,b\}$,

$$\sqrt{\frac{z}{\varphi(z)}}=A\sqrt{z\varphi(z)}.$$

This implies that $\varphi(z)=\theta=$ constant, and therefore the result is obtained.

Now, we proceed to compare the bounds utilized for $E(Z\varphi(Z))$ in inequalities (41) and (42).

Assume Z follows a discrete uniform distribution with support on $\{1,\dots,b\}$. In this case, $\varphi(z)=\frac{1}{z}$,and consequently, the lower bound in (41) becomes

$$\frac{1}{E(\frac{1}{Z\varphi(Z)})}=1.$$

Besides that, the lower bound (42) is

$$\frac{2(E(Z))^2}{b(b+1)-E(Z(Z-1))}=\frac{3(b+1)}{2(2b+1)}.$$

Accordingly, for the distribution, we deduce that the bound (41) is superior to the bound (42) for $b>1$.

Theorem 3.7. Let Z be a nonnegative discrete RV. Then

$$E[c^{-Z}\varphi(Z)]\geq\frac{(E[c^{-Z}])^2(c-1)}{cE(c^{-Z})-c^{-b}},$$

for constant $c\neq 1$, where equality satisfies iff Z has the distribution function given in equation (38).

Proof. By utilizing the C-S inequality, we have

$$(E[c^{-Z}])^2\leq E[c^{-Z}\varphi(Z)]E[\frac{c^{-Z}}{\varphi(Z)}].$$

Besides that, since

$$E\left[\frac{c^{-Z}}{\varphi(Z)}\right]=\sum_{z=0}^{b}c^{-z}F(z)=\sum_{y=0}^{b}\sum_{z=y}^{b}c^{-z}P\{Z=y\}$$
$$=\sum_{y=0}^{b}\frac{c^{-y}-c^{-b-1}}{1-c^{-1}}P\{Z=y\}$$
$$=\frac{cE[c^{-Z}]-c^{-b}}{c-1}.$$

Thus, the result is obtained. The equality holds iff there exists some nonnegative constant A so that, for all $z\in\{0,\dots,b\}$,

$$\sqrt{\frac{c^{-z}}{\varphi(z)}}=A\sqrt{c^{-z}\varphi(z)}.$$

This concludes that $\varphi(z)$ is a constant, and the result is obtained.

Results

In this work, we first introduced the past varentropy for discrete RVs. Then, we obtained bounds for the varentropy of some discrete distributions. In the following, by considering the resulting upper bounds, the squared logarithmic expectation, we obtained an expression for the bounds in terms of the squared logarithmic difference coefficients $d_\omega(j)$. In future work, we propose obtaining similar results for continuous distributions using logarithmic and log-gamma expectations. Moreover, we evaluated an upper bound for $Var[g(X_x)]$ and derived bounds for the expected values of specific functions in reliability theory.

Acknowledgments

I would like to express my sincere gratitude to Professor Mohsen Mohammadzadeh for his invaluable suggestion, which improved this article.

References

1. Jiang J, Wang R., Pezeril M. and Wang QA. Application of varentropy as a measure of probabilistic uncertainty for complex networks, Science

Bulletin. 2011; 56: 3677–3682.
2. Li J, Fradelizi M. and Madiman M. Information concentration for convex measures, IEEE International Symposium on Information Theory, Barcelona. 2016; 1128-1132.
3. Fradelizi M, Madiman M. and Wang L. Optimal concentration of information content for logconcave densities. In C. Houdré, D. Mason, P. Reynaud-Bouret & J. Rosiński (eds.), High Dimensional Probability VII. Progress in Probability, vol. 71, Cham, Springer. 2016; 45-60.
4. Gupta BB and Badve OP. GARCH and ANN-based DDoS detection and filtering in cloud computing environment, International Journal of Embedded Systems. 2017; 9: 391-400.
5. De Gregorio A and Iacus SM. On Rényi information for ergodic diffusion processes, Information Sciences. 2009; 179: 279-291.
6. Goodarzi F, Amini M and Mohtashami Borzadaran GR. Characterizations of continuous distributions through inequalities involving the expected values of selected functions, Applications of Mathematics. 2017(a); 62: 493–507.
7. Goodarzi F, Amini M and Mohtashami Borzadaran GR. On lower bounds for the variance of functions of random variables, Applications of Mathematics. 2021; 66: 767–788.
8. Sharma A and Kundu C. Varentropy of doubly truncated random variable, Probability in the Engineering and Informational Sciences. 2022; 37(3): 852–871.
9. Maadani S, Mohtashami Borzadaran GR and Rezaei Roknabadi AH. Varentropy of order statistics and some stochastic, Communication in Statistics-Theory and Methods. 2022; 51: 6447-6460.
10. Goodarzi F, Amini M and Mohtashami Borzadaran GR. Some results on upper bounds for the variance of functions of the residual life random variables, Journal of Computational and Applied Mathematics. 2017(b); 320, 30-42.
11. Goodarzi F, Amini M and Mohtashami Borzadaran GR. On upper bounds for the variance of functions of the inactivity time, Statistics and Probability Letters. 2016; 117: 62–71.
12. Buono F, Longobardi M. Varentropy of past lifetimes, Mathematical Methods of Statistics. 2022; 31: 57-73.
13. Goodarzi F, Characterizations of some discrete distributions and upper bounds on discrete residual varentropy, Journal of the Iranian Statistical Society. 2022; 21(2): 233–250.
14. Alizadeh Noughabi H and Shafaei Noughabi M. Varentropy estimators with applications in testing uniformity, Journal of Statistical Computation and Simulation. 2023; 93: 2582-2599.
15. Kontoyiannis I and Verdú. S. Optimal lossless data compression: non-asymptotics and asymptotics, IEEE Transactions on Information Theory. 2014; 60: 777-795.
16. Nair NU and Sankaran PG. Characterizations of discrete distributions using reliability concepts in reversed time, Statistics and Probability Letters. 2013; 83: 1939–1945.
17. Cacoullos T and Papathanasiou V. Characterizations of distributions by variance bounds, Statistics and Probabability Letters. 1989; 7: 351-356.
18. Kelley CT. Solving nonlinear equations with Newtonś method, SIAM, Philadelphia, 2003.
19. Song K.-S. Rényi information, loglikelihood and an intrinsic distribution measure, Journal of Statistical Planning and Inference. 2001; 93: 51-69.
20. Cheraghchi M. Expressions for the entropy of basic discrete distribution, IEEE Transactions on Information Theory. 2019; 65: 3999-4009.
21. Gupta L. Properties of reliability functions of discrete distributions, Communication in Statistics-Theory and Methods. 2015; 44: 4114-4131.