The 12th Seminar on Probability and Stochastic Processes
Semnan University
31 August-1 September 2019
دوازدهمین سمینار احتمال و فرایندهای تصادفی
دانشگاه سمنان ۹-۱۰ شهریور ۱۳۹۸

Difference and similarity between differential entropy and discrete entropy

G. R. Mohtashami Borzadaran[1] [1]

[1]***Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran***

---

abstract: Shannon's discovery of the fundamental laws of data compression and transmission marks the birth of Information Theory. Classically, Shannon entropy was formalized over discrete probability distributions. This discrete idea extended to continuous case. So, is Shannon entropy and differential entropy are connected to each other or not is a question that will discuss here. Various entropies, Phi-divergence, AEP and maximum entropy will discuss in two cases, and answer to the question is continuous cases not the limit of discrete cases?

keyword: Shannon entropy, Differential entropy, Maximum entropy, asymptotic equipartition property (AEP), Phi-divergence, Kullback Leibler information.
Mathematics Subject Classification (2010):62B10, 62E10.

---

1 Introduction

Entropy was introduced in 1863 within the field of thermodynamics to give a mathematical expression to the second law of thermodynamics. The law was first formulated around 15 years earlier, in 1849, by Clausius. The information theory was developed in the context of the theory of communication to answer two fundamental

[1]Speaker: email@grmohtashami@um.ac.ir

questions : (i) What is the ultimate achievable data compression? (ii) What is the ultimate achievable rate of transmission of information? For answering these questions, Shannon [1948] laid the foundation of information theory through his paper "A mathematical theory of communication" where the continue and extension of the ideas of Nyquist in 1924 and Hartley [1928] which described a logarithmic measure for information content for equally likely messages to answer to the question "What happens when messages are not equally likely?" Many papers reviewed the art of Shannon like Chakraborty [2015] that showed generating discrete analogues of continuous probability distributions-A survey of methods and constructions. In this paper, answer to the question " Is continuous entropy the limit of discrete entropy?" and What about information measures ? are important parts of this note. We will study AEP and maximum entropy in continuous and discrete case.

2 Various entropies in continuous cases are not the limit of discrete cases

Shannon defined the entropy as $H(X) = -\sum_{i=1}^{\infty} p_i \log p_i$ and its continuous version is $h(X) = -\int_S f(x) \log f(x) dx$ which is called differential entropy. It is well-known that this integral exists iff the density function of the random variables is Riemann-integrable. Consider the continuous random variable X with a probability density function $f(x)$. Quantizing by dividing its range into bins of length Δ. Then, in accordance to the Mean Value Theorem, within each bin of size $[i\Delta; (i+1)\Delta]$, there exists a value $X^\Delta = x_i$ that satisfies $\int_{i\Delta}^{(i+1)\Delta} f(x)dx = f(x_i)\Delta$ where $p_i = f(x_i)\Delta$. Th Shannon entropy is

$$H(X^\Delta) = -\sum_{-\infty}^{\infty} p_i \log p_i = -\sum \Delta f(x_i) \log f(x_i) - \log \Delta$$

so,

$$H(X^\Delta) + \log \Delta \overset{\Delta \longrightarrow 0}{\longrightarrow} h(X),$$

and differential entropy is not the limit of discrete case.

$H_\alpha(X) = \frac{1}{1-\alpha}\Sigma_i p_i^\alpha$ is Renyi entropy in discrete case which the continuous version is $h_\alpha(X) = \frac{1}{1-\alpha}\int_S f(x)^\alpha dx$. If we apply the same as Shannon entropy, then $\lim_{\Delta\to 0}[H_\alpha(X^\Delta) + \log\Delta] = h_\alpha(X)$. For other extension of the Shannon entropy continuous cases are not the limit of discrete cases. If $(X,Y) \sim p(x,y)$ then $I(X,Y) = H(X) - H(X|Y)$ is mutual information, and $I(X^\Delta, Y^\Delta) = I(X,Y)$. For more examples and details see Sanei Tabbas et al. [2016]. Now a comperative tour on discrete and continuous version of entropy can be as below:

- As in the discrete case, the differential entropy depends only on the probability function of the random variable. Note that the differential entropy is defined only if the aforementioned integral and density function exist.

- Differential entropy retains many of the properties of its discrete counterpart, but with some important differences. Chief among these is the fact that differential entropy may take on any value and discrete entropy is always non-negative. Differential entropy represents not an absolute measure of uncertainty.

- The formula for the differential entropy certainly represents the natural extension to the continuous case, but this transition from discrete to continuous random variables must be handled carefully as we shall see.

- Differential entropy can be a function of variance. Sometimes the variance does not exist for example, Cauchy distribution and special case of Pareto distribution. So the differential entropy is a suitable measure instead of variance.

- The differential entropy is not always well defined. For example, the differential entropy of $f(x) = \frac{1}{2x[(\ln x)]^2}, 0 < x < e^{-1}, e < x < \infty$ is infinite. If $E[\ln(1+|X|)]$ is infinite, in particular if X has finite first and second moments, then h_X is well defined.

- The entropy power $N(X) = \frac{e^{\frac{2}{n}h(X)}}{2\pi e}$, when $h(X) = -\infty$ then $N(X) = 0$ and $N(aX) = a^2 N(X)$. It can be one of the advantage of the power entropy. Also,

$N(X) \leq \sigma^2$ with equality iff $X \sim N(\mu, \sigma^2)$, $N(\Sigma_{i=1}^n a_i X_i) \geq \Sigma_{i=1}^n a_i^2 N(X_i)$ and $h(\Sigma_{i=1}^n a_i X_i) \geq \Sigma_{i=1}^n a_i^2 h(X_i), \Sigma_{i=1}^n a_i^2 = 1$.

- If $Y = g(X)$ is differential function of X, then $h(Y) \leq h(X) + E[\log|\frac{dg(x)}{dx}|]$ with equality iff g has inverse. For $g(X) = aX + b$ then $h(Y) = h(X) + \ln|a|$ and $h(AX) = h(X) + \log det|A|$. In discrete case $H(Y) = H(X)$.

3 Phi-divergence in continuous case is the limit of discrete case?

Since 1948 up to now, many important variant measures of Shannon entropy and its extension to divergence measure have been introduced. For example, Renyi introduced an entropy of order α in Renyi [1961] and in Harvda and Charvat [1967], proposed the entropy of order s and studied some of its mathematical properties. However concavity is not preserved though monotonicity and unification of entropy measures by means of (h, ϕ) entropy, for different values of (h, ϕ) functions:

$$H_\phi^h(f(x)) = h(\sum_{i=1}^{n} \phi(f(x_i))), \tag{3.1}$$

where either $\phi : [0, \infty) \to R$ is convex(concave) and $h : R \to R$ is decreasing (increasing). In here, the same as other entropies continuous case is the limit of discrete case where it is general case. Some examples of the (h, ϕ) entropy measures are presented in Table 1.

The Shannon entropy is concave and extensive, the Renyi entropy $H(\alpha)$ is extensive but nonconcave (for all $\alpha > 0$; concave only for$\alpha \in [0, 1]$).Tsallis form is nonextesive but it is concave (for $\alpha > 0$). Sharma and Mittal entropy is nonextesive and nonconcave. The later measure is determined by two parameters and contains the other three measures as particular cases.

The principle of maximum entropy states that, subject to precisely stated prior data (such as moments,...) the probability distribution which best represents the current

Table 1: Some of entropy measures

entropy measure	h	ϕ
Shannon	x	$-xlogx$
Renyi	$\frac{1}{1-\alpha}logx$	x^{α}
Harvda and Charvat (fitted for binary variables)	x	$\frac{1}{1-s}(x^s - x)$
Tsallis	x	$\frac{1}{1-s}(x^s - x)$
Sharma and Mittal	$\frac{e^{(s-1)x-1}}{s-1}$	$-xlogx$
Arimoto	$\frac{x^t-1}{2^{t-1}-1}$	$x^{\frac{1}{t}}$
Taneja	$-2^{r-1}x$	x^{rlogx}
Ferreri	$(1+\frac{1}{\lambda})log(1+\lambda) - \frac{x}{\lambda}$	$(1+\lambda x)log(1+\lambda x)$
Varmma	$\frac{1}{m-r}logx$	x^{r-m+1}

state of knowledge is the one with the largest entropy. The assertion of the MEP is that the most unbiased probability distribution is the maximum entropy distribution satisfying the constraints. The form of distribution with maximum entropy is usually expressed as exponential function of constraints. Let p, q be two probability density functions, our aim is calculating the value of the relations between them.

Definition 3.1. *If $\phi(.)$ is convex function, for two pdf p, q the Csiszar divergence measure Csiszár [1963] is define:*

$$D_{\phi}(P||Q) = \sum_{\chi} q(x)\phi(\frac{p(x)}{q(x)}), \tag{3.2}$$

for different forms of ϕ , Csiszar's measure results the different divergence measures. Some of them are summarized in Table 2.

$D_{\phi}(P||Q)$ in continuous case is the limit of discrete case, which is applicable for others special case via the arguments that has done for differential entropy. Tsallis distributions can be derived from the Shannon maximum entropy setting, by incorporating a constraint on the divergence between the distribution and another. He also expressed that the problem of minimization of the Kullback-Leibler under parameterizing the Shannon entropy, and maximization of the Shannon entropy under parameterizing the Kullback-Leibler lead to the same solutions. Here we are using the generalized measure entropy ($H(\phi)$) and the generalized ϕ divergence

Table 2: Some ϕ Divergences

Measure divergence	$\phi(x)$
Kullback-Leibler divergence	$-xlogx$
Triangular divergence	$\frac{2}{(x+1)^2}$
Batacharya divergence	$\sqrt{x}$
Jeffreys divergence	$(x-1)ln(x)$
Harmoonic divergence	$\frac{2x}{x+1}$
Tsallis divergence	$\frac{x^\alpha - x}{\alpha-1}$
Hellinger divergence	$\frac{(1-\sqrt{x})^2}{2}$
χ^2-divergence	$(x-1)^2$
α-divergence	$\begin{cases} \frac{4}{1-\alpha^2(1-x^{\frac{1+\alpha}{2}})} & \alpha \neq +1,-1 \\ xlogx & \alpha = 1 \\ -logx & \alpha = -1. \end{cases}$

$D_\phi(p||q)$ instead of the Shannon entropy and Kullback-Leibler measures respectively. The problem of maximization of the $H(\phi)$ under parameterizing the $D_\phi(p||q)$ and the problem of minimization of the $D_\phi(p||q)$ under parameterizing the $H(\phi)$ have been checked. We also show that these two problems lead to the same solutions in general and through substituting any of the measures from the above tables, as particular cases, we show that the above problem is satisfied.

4 AEP and centeral limit theorem related to entropy in continuous and discrete cases

In information theory, the asymptotic equipartition property (AEP) is a general property of the output samples of a stochastic source. It is fundamental to the concept of typical set used in theories of compression (see Cover and Thomas 2006). Analogue to weak law of large numbers $\frac{1}{n}\log\frac{1}{p(X_1,X_2,....,X_n)} \xrightarrow{P} H(X)$ where $X_1, X_2, ..., X_n$ are iid random variables, so $p(X_1, X_2, ..., X_n) \approx \exp\{-nH(X)\}$ with high probability. The typical set $A_\epsilon^{(n)}$ is the set of sequences $(x_1, x_2, ..., x_n) \in \chi^n$ where $2^{-n(H(X)+\epsilon)} \leq p(X_1, X_2, ..., X_n) \leq 2^{-n(H(X)-\epsilon)}$.

- If $(x_1, x_2, ..., x_n) \in A_\epsilon^{(n)}$, then $H(X) - \epsilon \leq -\frac{1}{n}\log p(X_1, X_2, ..., X_n) \leq H(X) + \epsilon$.
- $P(A_\epsilon^{(n)}) \geq 1 - \epsilon$ for n sufficiently large.

- $|A_\epsilon^n| \leq 2^{n(H(X)+\epsilon)}$ and $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X)-\epsilon)}$ for n sufficiently large where $|A|$ denotes the number of elements in the set A.

The typical sequences have short description of length$\approx nH(.)$.
The set $A_\epsilon^{(n)}$ of joint typical sequences of $\{(x^n, y^n)\}$ w.r.t. the distribution $p(x,y)$ is given by

$$A_\epsilon^{(n)} = \{(x^n, y^n) \in \chi^n \times \kappa^n / |-\frac{1}{n}\text{log}p(x^n) - H(X)| < \epsilon, |-\frac{1}{n}\text{log}p(y^n) - H(Y)| < \epsilon, |-\frac{1}{n}\text{log}p(x^n, y^n) - H(X,Y)| < \epsilon\},$$

where $p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i)$. Hence
(1) $P\{(x^n, y^n) \in A_\epsilon^{(n)}\} \longrightarrow 1$ as $n \longrightarrow \infty$.
(2) $|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(X,Y)-\epsilon)}$ for n sufficiently large.
(3) If $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n))$ then

$$P\{(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(X,Y)-3\epsilon)},$$

and for n sufficiently large

$$P\{(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\} \geq (1-\epsilon)2^{-n(I(X,Y)+3\epsilon)},$$

The above suggests there are about $2^{nI(X,Y)}$ distinguishable signals X^n where $I(X,Y)$ is the mutual information. It can be obtained via a continuous view by similar arguments but a little different to discrete cases as below:
Let $X_1, X_2, ..., X_n$ be a sequences of iid random variables with density $f(x)$, then $\frac{1}{n}\text{log}\frac{1}{f(X_1,X_2,...,X_n)} \xrightarrow{P} h(X)$. For $\epsilon > 0$ and any n the typical set in continuous case is defined as $A_\epsilon^{(n)} = \{(x_1, x_2, ..., x_n) \in S^n, \quad |\frac{1}{n}\text{log}\frac{1}{f(X_1,X_2,...,X_n)} - h(X)| \leq \epsilon\}$ where $f(x_1, x_2, ..., x_n) = \Pi_{i=1}^n f(x_i)$ and the volume is defined as $vol(A) = \int_A dx_1 dx_2....dx_n$.
The following properties are valid for continuous cases:
(1) $P(A_\epsilon^{(n)}) \geq 1-\epsilon$ for n sufficiently large.
(2) $vol(A_\epsilon^{(n)}) \leq 2^{n(h(X)+\epsilon)}$ and $vol(A_\epsilon^{(n)}) \geq (1-\epsilon)2^{n(h(X)-\epsilon)}$ for n sufficiently large.

One of the important application of AEP is in channel capacity for example "source channel coding theorem (Cover and Thomas 2006 page 220).

5 Maximum Entropy

One of the most important problems in statistical inference is the estimation of unknown distribution. The maximum entropy and minimum divergence approaches for estimating the probability density functions has been widely used in many research areas. The problem of maximizing entropy (MEPD) subject to some constraints such as moments has been studied by many authors (in continuous or discrete cases). For example, see Jaynes Jaynes [1957] and Kagan et al. [1973].

- The inverse Gaussian distribution is MEPD, when the arithmetic, geometric and harmonic means are prescribed.
- The Pearson type-V distribution is MEPD, when the geometric and harmonic means are prescribed.
- If $E(X)$ and $E(\ln X)$ are the constraints, one can obtain two parameter Weibull distribution as a MEPD.

The maximizers of another entropy, the Tsallis [1988] entropy have a high interest in many applied fields, namely economy, biology, and physics. The maximum Tsallis distributions have encountered a large success because of their remarkable agreement with experimental data. Some works have been done in the subject of Tsallis entropy maximization with inequality measures constraints. Yaghoobi et al. [2014], Khosravi et al. [2015], Khosravi et al. [2017], Eliazar and Sokolov [2010a], Eliazar and Sokolov [2010b] and Nakhaei et al. [2016] obtained families of Lorenz curves by Shannon and Tsallis entropy maximization under mean and Gini index and generalized Gini index constraints. One of the most well-known integral functionals that has been studied in variational calculus is the Lagrange functional.

$$L(y) = \int_a^b G(y(x), y'(x), x)dx, \tag{5.1}$$

where the given function G is continuous and has continuous first partial derivatives in each of its arguments. The basic variational problem can be stated as follows: Let $L(y)$ be a functional of the form (5.1) defined on the set of functions $y(x)$ which have continuous first derivatives in $[a,b]$ and satisfy the boundary conditions $y(a)=A$, $y(b)=B$. Then, a necessary condition for $L(y)$ to have an extremum for a given function $y(x)$, is that $y(x)$ satisfy Euler's equation:

$$\frac{\partial G}{\partial y}-\frac{d}{dx}\frac{\partial G}{\partial y'}=0. \tag{5.2}$$

The generalized entropy $\int_{-\infty}^{\infty}\phi[f(x)]dx$ is considered and we intend to find the maximum generalized entropy under the general constraints. Since $-H_\phi(f)$ is a convex functional, the problem is

$$\begin{aligned}&\min_f \quad \int_{-\infty}^{\infty}-\phi[f(x)]dx\\ &s.t.\\ &\qquad \int_{-\infty}^{\infty}D_i[F(x),f(x),x]dx=\theta_i\ ,\ \ i=1,2,...,m.\end{aligned} \tag{5.3}$$

Using the fact that target function and constraints are convex, necessary and sufficient condition to distribution function F with density function f has maximum ϕ-entropy under the constraints (5.3) is to satisfy the equation

$$\begin{aligned}&\frac{\partial}{\partial F}\left(\sum_{i=1}^{m}\lambda_iD_i-\phi(f)\right)-\frac{d}{dx}\frac{\partial}{\partial f}\left(\sum_{i=1}^{m}\lambda_iD_i-\phi(f)\right)=0\\ \Leftrightarrow\quad &\sum_{i=1}^{m}\lambda_i\left(\frac{\partial D_i}{\partial F}-\frac{d}{dx}\frac{\partial D_i}{\partial f}\right)+f'\phi''(f)=0,\end{aligned} \tag{5.4}$$

where $\lambda_1,\lambda_2,...,\lambda_m$ are Lagrange multipliers and depend on $\theta_1,\theta_2,...,\theta_m$. From solving equation (5.4) we obtain the maximum ϕ-entropy under the constraints (5.3). Equation (5.4) give us a generel rule to obtain maximum entropy distributions under different constraints. In the following we refer to some special cases of this problem.

- In equation (5.4), suppose $\phi(x)=-x\log(x)$ and $D_1(F,f,x)=f(x)$. In this case, the maximum Shannon entropy distributions can be obtained under different constraints.

- **Moment constraint:** If $D_i(F, f, x) = g_i(x)f(x)$, $i = 2, ..., m$, the results of Kagan et al. [1973] are obtained as

$$f(x) = A\exp[-\lambda_1 g_1(x) - \lambda_2 g_2(x) - ... - \lambda_m g_m(x)],$$

where A, λ_1, λ_2, ..., λ_m are to be obtained by using the constraints.

- **Mean and Gini index constraints:** If $D_2(F, f, x) = xf(x)$ and $D_3(F, f, x) = [F(x) - 1]^2$, then the results of Eliazar and Sokolov [2010a] are obtained as

$$\bar{F}(x) = \frac{1}{\sigma\exp(\rho x) + (1-\sigma)}, \; x \geq 0,$$

where $\bar{F}(x)$ is survival function and σ and ρ depend on the constraints.

- **Mean and Pietra index constraints:** If $D_2(F, f, x) = xf(x)$ and $D_3(F, f, x) = \max(0, x - \theta_2)f(x)$, then the results of Eliazar and Sokolov [2010b] are obtained:

$$f(x) = \begin{cases} c_1\exp(\alpha x) & \text{if } 0 < x < \mu, \\ c_2\exp(-\beta x) & \text{if } \mu < x < \infty, \end{cases}$$

where α and β are real exponents depending on the constraints.

- **Mean and generalized Gini index constraints:** If $D_2(F, f, x) = xf(x)$ and $D_3(F, f, x) = [F(x) - 1]^\nu$, then the maximum entropy distribution is

$$\bar{F}(x) = \left(\frac{1}{c_1\exp(c_2 x) + (1-c_1)}\right)^{\frac{1}{\nu-1}}, \; x \geq 0,$$

where c_1 and c_2 depend on the constraints. Khosravi et al. [2015] showed this distribution model the income data considered better than alternative models.

- In equation (5.4), suppose $\phi(x) = \frac{1-x^\alpha}{\alpha-1}, \alpha > 0, \alpha \neq 1$ and $D_1(F, f, x) = f(x)$. If $D_2(F, f, x) = xf(x)$ and $D_3(F, f, x) = [F(x) - 1]^\nu$, Khosravi et al. [2017] showed the generalized Pareto distribution have maximum Tsallis entropy.

- We can have similar results for discrete cases where the coefficients need more care via difference equation and calculation with software like Matlab but basically discrete case leads to complicated forms.

Divergence measures are used to evaluate distance. Kullback and Leibler [1951] introduced the first divergence measure as a measure of information. The Kullback-Leibler divergence, also known as the relative entropy, between two probability density functions f and g is defined as

$$D_{KL}(f, g) = \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx. \tag{5.5}$$

By extension of Kullback-Leibler divergence, different divergence measures have been introduced (Csiszár [1963]) as :

$$D_\phi(f, g) = \int_{-\infty}^{\infty} g(x)\phi\left(\frac{f(x)}{g(x)}\right) dx,$$

where $\phi : [0, \infty) \to \mathbb{R}$ is a convex function such that $\phi(1) = 0$. The minimum divergence principle is another method of estimating distributions when we have constraints. Using the similar arguments in MEPD we can find the solution for this case also:

$$\begin{aligned} &\frac{\partial}{\partial F}\left(\sum_{i=1}^{m} \lambda_i D_i + g\phi\left(\frac{f}{g}\right)\right) - \frac{d}{dx}\frac{\partial}{\partial f}\left(\sum_{i=1}^{m} \lambda_i D_i + g\phi\left(\frac{f}{g}\right)\right) = 0 \\ \Leftrightarrow \quad &\sum_{i=1}^{m} \lambda_i \left(\frac{\partial D_i}{\partial F} - \frac{d}{dx}\frac{\partial D_i}{\partial f}\right) - \frac{d}{dx}\phi'\left(\frac{f}{g}\right) = 0. \end{aligned} \tag{5.6}$$

Generally, the equation (5.6) for different functions $g(x)$ and ϕ leads to a complex equation that is difficult to solve. But in a special case, If we intend to minimize Kullback-Leibler distance under the constraints on mean and generalized Gini index, that is assuming $\phi(x) = x\log(x)$, $D_1(F, f, x) = f(x)$, $D_2(F, f, x) = xf(x)$ and $D_3(F, f, x) = [F(x) - 1]^{\nu}$, from equation (5.6) we have

$$2\lambda_3 \bar{F}(x) + \frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)} + \lambda_2 = 0. \tag{5.7}$$

assuming that $\frac{g'(x)}{g(x)} = k$, where k is a constant (for example when prior distribution is exponential), the solution of equation (5.7) is

$$\bar{F}(x) = \left(\frac{1}{c_1 \exp(c_2 x) + (1 - c_1)}\right)^{\frac{1}{\nu-1}}, \ x \geq 0,$$

where c_1 and c_2 depend on θ_1 and θ_2.

MEPD in discrete case, is similar to continuous case but sometimes leads to finding the results via numerical study.

Let X be a random variable that its values concentrated on Z with pmf

$$f(k) = \frac{1}{1 + 2s(p)} p^{|k|^m}, \quad k \in Z, s(p) = \sum_{k=1}^{\infty} p^{|k|^m}, \quad m = 1, 2, ... \tag{5.8}$$

is a symmetric discrete distribution where $m = 1$ ($m = 2$) implies discrete Laplace (discrete normal) distribution.

$$H(X) = \ln(1 + 2s(p)) - \ln p \frac{2ps'(p)}{1 + 2s(p)}$$

is the entropy of (5.8). The maximum entropy probability distribution (MEPD) under the constraints

$$p_k = p_{-k}, k \in Z, \sum_{k=1}^{\infty} p_k = 1, \quad \sum_{k=1}^{\infty} |k|^m p_k = 1$$

is (5.8) where $p_k = f(k)$.

Among various distributions described in (5.8) $m = 1$ ($m = 2$) discrete Laplace (discrete normal) distributions which are MEPD can be also classified as generated to preserve the maximum entropy property of their continuous counterpart.

Under the constraints

$$EX = \mu, EX_{(k)} = \alpha_k, x_{(k)} = x(x-1)...(x-k+1),$$

where k is even and X has integer support. The bilateral polynomial power series has MEPD where $k = 2, 4$ implies discrete normal (Kemp 1997) and discrete quartic (Mohtashami Borzadaran [2000]) distributions respectively.

Continuous version of these results can be obtained similarly.

6 Conclusion

The most properties of differential entropy and Shannon entropy achieved via similar arguments but we can see some diffences in their properties.

References

Chakraborty, S. (2015). Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. *Journal of Statistical Distributions and Applications*, **2 (6)**, 1-30.

Csiszár I. (1963). Eine informations theoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, **8**, 85–108.

Eliazar. I. and Sokolov, M. (2010a). Maximization of statistical heterogeneity: from Shannon's entropy to Gini's index. *Physica A*, **389**, 3023–3038.

Eliazar, I. and Sokolov, M. (2010b). Measuring statistical heterogeneity: The Pietra index. *Physica A*, **389**, 117–125.

Harvda, J. and Charvat, F. (1967). Quantification method of classification processes: Concept of structural αentropy. *Kybernetika*, **3,** 30-35.

Hartley, R. V. L. (1928). Transmission of information. *Bell System Technical Journal*, **7**, 535-563.

Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review*, **106**, 620–630.

Kagan, A. M., Linnik, Y. M. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, John Wiley, New York.

Kemp, A.W., 1997. Characterizations of a discrete normal distribution.*J. Statist. Plann. Inference,* **63,** 223–229.

Khosravi Tanak, A., Mohtashami Borzadaran, G. R. and Ahmadi, J. (2015). Entropy maximization under the constraints on the generalized Gini index and its application in modeling income distributions. *Physica A: Statistical Mechanics and its Applications*, **438**, 657–666.

Khosravi Tanak, A., Mohtashami Borzadaran, G. R. and Ahmadi, J. (2017). Maximum Tsallis entropy with generalized Gini and Gini mean difference indices constraints. *Physica A: Statistical Mechanics and its Applications*, **471**, 554–560.

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

Mohtashami Borzadaran, G. R. (2000). Characterization of Pearsonian and bilateral power series distributions via maximum entropies. *Baysian Infer. Maximum Entropy Meth. Sci. Eng.*, 145–150.

Nakhaei Rad, N., Mohtashami Borzadaran, G. R. and Yari, G. H. (2016). Maximum entropy estimation of income share function from generalized Gini index. *Journal of Applied Statistics*, **43**, 2910–2921.

Renyi, A. (1961). On measures of entropy and information. *Proc. Berekeley Symposium, Statist. Probability*, **1,** 547-561.

Sanei Tabass, M., Mohtashami Borzadaran, G. R. and Amini, M. (2016). Renyi entropy in coutinuos case is not limit of discrete case. *Mathematical Sciences and Applications e-Notes*, **4(1)**, 113-117.

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

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, **52**, 479–487.

Yaghoobi Avval Riabi, M., Mohtashami Borzadaran, G. R. and Yari, G. (2014). Tsallis Maximum Entropy Lorenz Curves. *Journal of Statistical Research of Iran*, **11**, 41–56.

The 12th Seminar on Probability and Stochastic Processes
Semnan University
31 August-1 September 2019
دوازدهمین سمینار احتمال و فرایندهای تصادفی
دانشگاه سمنان ۹-۱۰ شهریور ۱۳۹۸

Bayesian Nonparametric Estimation of Distribution Function for Length-Biased Data

Najmeh Nakhaei Rad [1] [1], ***Vahid Fakoor***[2]

[1] ***Department of Mathematics and Statistics, Mashhad Branch, Islamic Azad University, Mashhad, Iran.***

[2] ***Departmen oft Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad,Iran.***

---

abstract: The length-biased sampling occurs when an appropriate sampling scheme is absent. Then units are chosen at a rate proportional to their length. As a result, the greater values have more chances to be selected. When observations are not coming directly from the distribution of interest, but from a length biased version, Cox (1969) proposed an estimator for distribution function which plays the same role as the empirical distribution for direct data. In this paper, by using Bayesian nonparametric approach the estimation of distribution function is derived under length bias and its consistency is also discussed. A simulation study is presented as well as a real data example to illustrate obtained results.

keyword: Bayesian nonparametric inference; Dirichlet process; Length-biased data; Monte Carlo method.

Mathematics Subject Classification (2010): 99X99, 99X99.

---

[1]Speaker: najme.rad@gmail.com