

UPPER BOUND FOR VARIANCE OF FINITE MIXTURES OF POWER EXPONENTIAL DISTRIBUTIONS

HOK SHING KWONG — SARALEES NADARAJAH^c

(Communicated by Gejza Wimmer)

ABSTRACT. Both variance and entropy are commonly used measures for uncertainty. There exists many cases where variance is infinite and entropy is finite. In this note, we derive an upper bound illustrating the relationship between variance and entropy of random variables having a special class of distributions. We also derive an upper bound for k th absolute central moment proportional to entropy power for a special class of distributions.

©2022
Mathematical Institute
Slovak Academy of Sciences

1. Introduction

Both variance and differential entropy are commonly used metrics to measure uncertainty of random variables. While variance measures how a random variable is spread out from its expected value, differential entropy takes a different approach. Differential entropy measures uncertainty from an information approach, it measures the average amount of information of a random variable. Although variance and entropy are generally not the same, there is one case when variance and entropy are equivalent measures. When normal distribution is considered, the maximum entropy given variance is attained, this suggests that variance and entropy are related to some extent. For comprehensive reviews on entropy, we refer to [2] and [7]. See also [4].

When the underlying probability distribution is multimodal, variance becomes inefficient in representing the uncertainty of random variables. For example, consider a finite mixture of two equally weighted univariate normal distributions with unit variance and different means μ_1 and μ_2 . The variance of the overall distribution approaches ∞ as $|\mu_1 - \mu_2| \rightarrow \infty$. On the other hand, differential entropy remains stable in this situation, approaching $\frac{\log(2\pi e)}{2} + \log(2)$ as $|\mu_1 - \mu_2| \rightarrow \infty$. Therefore, differential entropy appears to be a more appropriate measure of uncertainty of multimodal random variables.

Let X denote a random variable with probability density function f . It is well known that given the value of differential entropy,

$$h(f) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx,$$

2020 Mathematics Subject Classification: Primary 62E99.

Keywords: Differential entropy, finite mixture distribution, generalized normal distribution, sharp bounds, variance.

^c Corresponding author.

the lower bound of variance of X is achieved by normal distribution. That is, there exists a sharp lower bound on variance proportional to entropy power $e^{2h(f)}$, as shown in (1):

$$\frac{1}{2\pi e} e^{2h(f)} \leq \text{Var}(X). \quad (1)$$

This result can be extended (see, for example, [6]) for lower bounds of the k th absolute central moment such that the lower bound of the k th absolute central moment is proportional to $e^{kh(f)}$ as shown in (2):

$$\frac{1}{k\Gamma(1/k)^k} e^{kh(f)} \leq E(|X - \mu|^k). \quad (2)$$

In this case, the lower bound is attained by power exponential distribution (or generalised normal distribution) with shape parameter $\beta = k$. The probability density function of power exponential distribution is

$$f(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta},$$

where $\beta > 0$ is the shape parameter, α is the scale parameter, and $-\infty < \mu < \infty$ is the location parameter.

In general, a similar form of upper bound is not known. There are many cases where the variance is not bounded given finite differential entropy. For instance, non-logconcave distributions, for example, Pareto distributions with shape parameter ≤ 2 , or finite mixture distributions, do not generally have an upper bound on variance given entropy. [1] considered a finite mixture of power exponential distributions with equal means; they showed that the variance has an upper bound proportional to entropy power, if the maximum ratio of variances between component distributions is finite.

Although multimodal distributions are generally not upper bounded by entropy power, we believe it would be useful to derive an upper bound under some specific conditions, giving further understanding between differential entropy and variance. In this note, we consider a special class of multimodal distributions, a multimodal finite mixture of power exponential distributions. Entropy of finite mixture distributions cannot be expressed in closed-form generally, and the evaluation relies on various approximation methods. For example, [3] proposed a method based on a Taylor-series expansion method to approximate entropy of Gaussian mixtures. We show that the variance has an upper bound proportional to $e^{2h(f)}$, if ratios of variances, and ratios of squared means between all component distributions are finite.

The main result on the upper bound of variance is given in Section 2. A numerical study is presented in Section 3. An extension for k th absolute central moments is given in Section 4. Conclusions are given in Section 5.

2. Upper bound on variance for multimodal distributions

Consider $f(x)$ specified as follows

$$f(x) = \sum_{i=1}^n \omega_i f_i(x) = \sum_{i=1}^n \omega_i \frac{\beta_i}{2\alpha_i\Gamma(1/\beta_i)} e^{-(|x-\mu_i|/\alpha_i)^{\beta_i}}, \quad (3)$$

where $\beta_i > 0$ are the shape parameters, α_i are the scale parameters, $-\infty < \mu_i < \infty$ are the location parameters, n is the number of components and $0 < \omega_i < 1$ are weights with $\omega_1 + \cdots + \omega_n = 1$.

Let

$$r_{\sigma^2} = \max_{i,j} \frac{\sigma_i^2}{\sigma_j^2}, \quad r_{\mu^2} = \max_{i,j} \frac{\mu_i^2}{\mu_j^2}, \quad M(r) = \frac{(r-1)r^{\frac{1}{r-1}}}{e \log r}.$$

THEOREM 2.1 (Mitrinovic & Vasic, 1970, p. 79). *If a_i are strictly positive quantities, there exists an inequality between weighted geometric mean and weighted arithmetic mean as follows:*

$$\prod_{i=1}^n a_i^{\omega_i} \geq \frac{1}{M(r)} \sum_{i=1}^n \omega_i a_i,$$

where

$$r = \max_{i,j} \left(\frac{a_i}{a_j} \right).$$

The value of $M(r)$ is monotonically increasing with r , i.e., the difference between weighted geometric mean and weighted arithmetic mean is increasing.

PROPOSITION 2.1.1. *Let X have the probability density function (3) with $n = 1$, $\alpha_1 = \alpha$, $\beta_1 = \beta$ and $\mu_1 = \mu$. The k th absolute central moment of X is*

$$\begin{aligned} E(|X - \mu|^k) &= \int_{-\infty}^{\infty} |x - \mu|^k \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta} dx \\ &= \frac{\alpha^{k+1}}{\beta} \frac{\beta}{\alpha\Gamma(1/\beta)} \int_0^{\infty} y^{(k+1)/\beta-1} e^{-y} dy = \frac{\alpha^k \Gamma(\frac{k+1}{\beta})}{\Gamma(1/\beta)}. \end{aligned}$$

PROPOSITION 2.1.2. *Let X have the probability density function (3) with $n = 1$, $\alpha_1 = \alpha$, $\beta_1 = \beta$ and $\mu_1 = \mu$. The differential entropy of X is*

$$\begin{aligned} h(f) &= - \int_{-\infty}^{\infty} \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta} \log \left[\frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x-\mu|/\alpha)^\beta} \right] dx \\ &= - \left[\log \left(\frac{\beta}{2\alpha\Gamma(1/\beta)} \right) - \frac{1}{\Gamma(1/\beta)} \int_0^{\infty} y^{1/\beta} e^{-y} dy \right] = \log \left(\frac{2\alpha\Gamma(1/\beta)}{\beta} \right) + \frac{1}{\beta}. \end{aligned}$$

PROPOSITION 2.1.3. *From Propositions 2.1.1 and 2.1.2, we have the following relationship between differential entropy and the k th absolute central moment of X :*

$$e^{kh(f)} = \frac{2^k \alpha^k \Gamma(1/\beta)^k}{\beta^k} e^{k/\beta} = E(|X - \mu|^k) A(k, \beta),$$

where

$$A(k, \beta) = \frac{2^k e^{k/\beta} \Gamma(1/\beta)^{k+1}}{\beta^k \Gamma(\frac{k+1}{\beta})}.$$

LEMMA 2.1.1. *Let X have the probability density function (3). Due to concavity of $h(f)$, the entropy of finite mixture distributions has the following lower bound:*

$$h(f) \geq \sum_{i=1}^n \omega_i h(f_i).$$

Proof. The result follows because

$$\begin{aligned} h(f) - \sum_{i=1}^n \omega_i h(f_i) &= - \int_{-\infty}^{\infty} f(x) \log f(x) dx + \sum_{i=1}^n \omega_i \int_{-\infty}^{\infty} f_i(x) \log f_i(x) dx \\ &= - \left[\sum_{i=1}^n \omega_i \int_{-\infty}^{\infty} f_i(x) \log \frac{f(x)}{f_i(x)} dx \right] = \sum_{i=1}^n \omega_i D_{KL}(f_i || f) \geq 0, \end{aligned}$$

where

$$D_{KL}(f_i || f) = \int_{-\infty}^{\infty} f_i(x) \log \frac{f(x)}{f_i(x)} dx. \quad \square$$

THEOREM 2.2. *Let X have the probability density function (3). An upper bound of variance of X can be written proportional to entropy power as follows:*

$$\text{Var}(X) \leq \left[M(r_{\sigma^2}) + M(r_{\mu^2}) \prod_{i=1}^n (r_i^2)^{\omega_i} \right] \left[\prod_{i=1}^n \frac{1}{A(2, \beta_i)^{\omega_i}} \right] e^{2h(f)},$$

where $\mu_i^2 = \sigma_i^2 r_i^2$.

Proof. From Theorem 2.1, Proposition 2.1.3 and Lemma 2.1.1, we can write

$$\begin{aligned} e^{2h(f)} &\geq e^{2 \sum_{i=1}^n \omega_i h(f_i)} = \prod_{i=1}^n e^{2h(f_i) \omega_i} = \prod_{i=1}^n A(2, \beta_i)^{\omega_i} (\sigma_i^2)^{\omega_i} \\ &\geq \frac{1}{M(r_{\sigma^2})} \left[\prod_{i=1}^n A(2, \beta_i)^{\omega_i} \right] \left[\sum_{i=1}^n \omega_i \sigma_i^2 \right], \end{aligned}$$

implying that

$$\sum_{i=1}^n \omega_i \sigma_i^2 \leq \left[M(r_{\sigma^2}) \prod_{i=1}^n \frac{1}{A(2, \beta_i)^{\omega_i}} \right] e^{2h(f)}. \quad (4)$$

Also

$$e^{2h(f)} \geq \prod_{i=1}^n A(2, \beta_i)^{\omega_i} (\sigma_i^2)^{\omega_i}$$

can be rewritten as follows

$$e^{2h(f)} \geq \prod_{i=1}^n A(2, \beta_i)^{\omega_i} \left(\frac{\mu_i^2}{r_i^2} \right)^{\omega_i} \geq \left[\prod_{i=1}^n A(2, \beta_i)^{\omega_i} \left(\frac{1}{r_i^2} \right)^{\omega_i} \right] \frac{1}{M(r_{\mu^2})} \left[\sum_{i=1}^n \omega_i \mu_i^2 \right],$$

implying that

$$\sum_{i=1}^n \omega_i \mu_i^2 \leq M(r_{\mu^2}) \left[\prod_{i=1}^n \frac{1}{A(2, \beta_i)^{\omega_i}} (r_i^2)^{\omega_i} \right] e^{2h(f)}.$$

Therefore,

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 = \sum_{i=1}^n \omega_i \sigma_i^2 + \sum_{i=1}^n \omega_i \mu_i^2 - \left(\sum_{i=1}^n \omega_i \mu_i \right)^2 \\ &\leq \sum_{i=1}^n \omega_i \sigma_i^2 + \sum_{i=1}^n \omega_i \mu_i^2 \leq \left[M(r_{\sigma^2}) + M(r_{\mu^2}) \prod_{i=1}^n (r_i^2)^{\omega_i} \right] \left[\prod_{i=1}^n \frac{1}{A(2, \beta_i)^{\omega_i}} \right] e^{2h(f)}. \quad \square \end{aligned} \quad (5)$$

3. Special cases

It is easy to see that $r_i^2 \rightarrow 0$ when $\mu_i \rightarrow 0$, and (5) reduces to the original upper bounds shown in [1], i.e., (4). The variance of X becomes equal to (5) if all component distributions are identical with zero mean, i.e., $\sigma_i^2 = \sigma_j^2$, $\beta_i = \beta_j$ for all $i, j = 1, \dots, n$ and $\mu_i^2 = 0$ for $i = 1, \dots, n$. In this case, the variance equals to $\frac{1}{A(2, \beta)} e^{2h(f)}$.

If only $\sigma_i^2 = \sigma_j^2$ for all $i, j = 1, \dots, n$ then (5) reduces to

$$\left[1 + M(r_{\mu^2}) \prod_{i=1}^n (r_i^2)^{\omega_i} \right] \left[\prod_{i=1}^n \frac{1}{A(2, \beta_i)^{\omega_i}} \right] e^{2h(f)}.$$

If only $\beta_i = \beta_j$ for all $i, j = 1, \dots, n$ then (5) reduces to

$$\left[M(r_{\sigma^2}) + M(r_{\mu^2}) \prod_{i=1}^n (r_i^2)^{\omega_i} \right] \left[\frac{1}{A(2, \beta)} \right] e^{2h(f)}.$$

If only $\mu_i = \mu_j$ for all $i, j = 1, \dots, n$ then (5) reduces to

$$\left[M(r_{\sigma^2}) + \prod_{i=1}^n (r_i^2)^{\omega_i} \right] \left[\prod_{i=1}^n \frac{1}{A(2, \beta_i)^{\omega_i}} \right] e^{2h(f)}.$$

If only $\mu_i = 0$ for all $i = 1, \dots, n$ then (5) reduces to

$$M(r_{\sigma^2}) \left[\prod_{i=1}^n \frac{1}{A(2, \beta_i)^{\omega_i}} \right] e^{2h(f)}.$$

4. Upper bound on the k th absolute central moments for unimodal distributions

The variance upper bound for finite mixture of generalized normal distributions can be easily extended for k th absolute central moment. We consider a finite mixture of generalized normal distributions in (3).

THEOREM 4.1. *Let X have the probability density function (3) with $\mu_i = \mu$ for all $i = 1, \dots, n$. An upper bound for the k th absolute moment of X can be written proportional to entropy power as follows:*

$$E(|X - \mu|^k) \leq M(r_\xi) \left[\prod_{i=1}^n \frac{1}{A(\beta_i, k)^{\omega_i}} \right] e^{kh(f)}.$$

Proof. As shown in Proposition 2.1.1, the k th absolute central moment given α, β is given by $\frac{\alpha^k \Gamma(\frac{k+1}{\beta})}{\Gamma(\frac{1}{\beta})}$. Let $\xi_i(k)$ denote the k th absolute central moment of i th component distribution. Then

$$\begin{aligned} e^{kh(f)} &\geq e^{k \sum_{i=1}^n \omega_i h(f_i)} = \prod_{i=1}^n e^{kh(f_i) \omega_i} = \prod_{i=1}^n A(\beta_i, k)^{\omega_i} \xi_i(k)^{\omega_i} \\ &\geq \frac{1}{M(r_\xi)} \left[\prod_{i=1}^n A(\beta_i, k)^{\omega_i} \right] \left[\sum_{i=1}^n \omega_i \xi_i(k) \right], \end{aligned}$$

implying that

$$\sum_{i=1}^n \omega_i \xi_i(k) \leq M(r_\xi) \left[\prod_{i=1}^n \frac{1}{A(\beta_i, k)^{\omega_i}} \right] e^{kh(f)}, \quad \text{where } r_\xi = \max_{i,j} \frac{\xi_i(k)}{\xi_j(k)}. \quad (6)$$

Therefore, using (6), we have the following bound:

$$E(|X - \mu|^k) \leq M(r_\xi) \left[\prod_{i=1}^n \frac{1}{A(\beta_i, k)^{\omega_i}} \right] e^{kh(f)}. \quad \square$$

A lower bound can also be obtained in the case of Theorem 4.1. Using (2), we obtain

$$\frac{1}{k\Gamma(1/k)^k e} \sum_{i=1}^n w_i e^{kh(f_i)} \leq E(|X - \mu|^k). \quad (7)$$

If $k = 2$, (7) reduces to

$$\frac{1}{2\pi e} \sum_{i=1}^n w_i e^{2h(f_i)} \leq \text{Var}(X).$$

5. Conclusions

In this note, we have derived upper bounds on variance for a special class of multimodal distributions under some constraints. Closeness of the upper bound under different conditions have been presented. We have also derived upper bounds on the k th absolute central moment for a special class of unimodal distributions.

Acknowledgement. The authors would like to thank the Editor and the referee for careful reading and comments which greatly improved the paper.

REFERENCES

- [1] CHUNG, H. W.—SADLER, B. M.—HERO, A. O.: *Bounds on variance for unimodal distributions*, IEEE Transactions on Information Theory **63** (2017), 6936–6949.
- [2] COVER, T. M.—THOMAS, J. A.: *Elements of Information Theory*, 2nd ed., Wiley, New York, 2006.
- [3] HUBER, M. F.—BAILEY, T.—DURRANT-WHYTE, H.—HANEBECK, U. D.: *On entropy approximation for Gaussian mixture random vectors*, In: Proceedings of 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008, pp. 181–188.
- [4] KUMAR, P.—HOODA, D. S.: *On generalized measures of entropy and dependence*, Math. Slovaca **58** (2008), 377–386.
- [5] MITRINOVIC, D. S.—VASIC, P. M.: *Analytic Inequalities*, Vol. 1, Springer, New York, 1970.
- [6] ZAMIR, R.—FEDER, M.: *On universal quantization by randomized uniform/lattice quantizers*, IEEE Transactions on Information Theory **38** (1992), 428–436.
- [7] ZIDEK, J. V.—VAN EEDEN, C.: *Uncertainty, entropy, variance and the effect of partial information*, Lecture Notes–Monograph Series **42** (2003), 155–167.

Received 28. 11. 2020

Accepted 15. 6. 2021

*Department of Mathematics
University of Manchester
Manchester M13 9PL
UK*

E-mail: hokshing.kwong@postgrad.manchester.ac.uk
saralees.nadarajah@manchester.ac.uk