

Bounds on Variance for Symmetric Unimodal Distributions

Hye Won Chung

Brian M. Sadler

Alfred O. Hero

Abstract—We show a direct relationship between the variance and the differential entropy for the general class of symmetric unimodal distributions by providing an upper bound on variance in terms of entropy power. Combining this bound with the well-known entropy power lower bound on variance, we prove that for the general class of symmetric unimodal distributions the variance can be bounded below and above by the scaled entropy power. As differential entropy decreases, the variance is sandwiched between two exponentially decreasing functions in the differential entropy. This establishes that for the general class of symmetric unimodal distributions, the differential entropy can be used as a surrogate for concentration of the distribution.

Index Terms—Differential entropy, variance, symmetric unimodal distributions, information theoretic surrogates.

I. INTRODUCTION

In this paper, we establish a direct relationship between differential entropy and variance for the general class of symmetric unimodal distributions over \mathbb{R} . The variance of a random variable X having a distribution with density function $p(x)$ with mean $m \in \mathbb{R}$ is denoted as

$$\text{var}(X) = \int_{-\infty}^{\infty} (x - m)^2 p(x) dx, \quad (1)$$

and the differential entropy of X with density function $p(x)$ as

$$h(p) = - \int_{-\infty}^{\infty} p(x) \log p(x) dx. \quad (2)$$

Even though both variance and differential entropy quantify certain random properties in the variable X , there is no universal monotonic relation between (1) and (2).

The estimation counterpart to the Fano's inequality (Theorem 8.6.6 in [1]) establishes that

$$\frac{1}{2\pi e} e^{2h(p)} \leq \text{var}(X), \quad (3)$$

where equality is achieved for Gaussian distributions. This inequality shows that for general distributions, variance can approach 0 only if its differential entropy converges to $-\infty$.

The question is then whether there exists a generally applicable upper bound on variance in terms of differential entropy. The answer is negative and we can easily construct a counterexample. Consider the distribution

$$p(x) = \begin{cases} \epsilon & x \in [-t - \frac{1}{2\epsilon}, -t] \cup [t, t + \frac{1}{2\epsilon}] \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Hye Won Chung (hyechung@umich.edu) and Alfred O. Hero (hero@eecs.umich.edu) are with the EECS department at the University of Michigan; Brian M. Sadler (brian.m.sadler6.civ@mail.mil) is with the US Army Research Laboratory. This research was partially supported by US Army Research Laboratory grant W911NF-11-1-0391.

The differential entropy of this distribution is $h(p) = \log \frac{1}{\epsilon}$, which is independent of t , but the variance is $\text{var}(X) = t^2 + \frac{t}{2\epsilon} + \frac{1}{12\epsilon^2}$, which increases without bound in t . Thus in general there does not exist an upper bound on variance that is monotone in differential entropy, and, even if the differential entropy of a distribution goes to $-\infty$, the variance of this distribution can be strictly larger than a positive constant. Therefore, differential entropy cannot be a good surrogate for variance in all cases.

However, for certain distributions, including Gaussian and uniform, there does exist a monotonic relationship between variance and differential entropy. For a Gaussian distribution with mean m and variance σ^2 , denoted $\mathcal{N}(m, \sigma^2)$, the entropy power, defined as $e^{2h(p)}$, is proportional to the variance as

$$\sigma^2 = \frac{1}{2\pi e} e^{2h(p)}. \quad (5)$$

For a uniform distribution $p(x) = \text{unif}(m - \frac{1}{2\epsilon}, m + \frac{1}{2\epsilon})$, the variance is equal to $1/(12\epsilon^2)$ and the differential entropy is $\log(1/\epsilon)$. Thus, for uniform distributions, we have

$$\text{var}(X) = \frac{e^{2h(p)}}{12}. \quad (6)$$

Therefore, for these cases, the variance decreases exponentially as $h(p)$ decreases.

Moreover, in [2], an important observation was made that for a random variable X whose density function $p(x)$ is log concave, i.e.,

$$p(\alpha x + (1 - \alpha)y) \geq p(x)^\alpha p(y)^{1-\alpha} \quad (7)$$

for each $x, y \in \mathbb{R}$ and each $0 \leq \alpha \leq 1$, the variance can be not only lower bounded but also upper bounded in terms of the entropy power, $e^{2h(p)}$, as

$$\frac{e^{2h(p)}}{2\pi e} \leq \text{var}(X) \leq c_0 \cdot e^{2h(p)}, \quad (8)$$

for some constant $c_0 > 0$. Applications of log-concave densities for inference and modeling can be found in [3], [4]. This result is important in the sense that even though there is no monotonic relation between variance and entropy, for any log-concave distribution variance is guaranteed to approach 0 if and only if its differential entropy converges to $-\infty$.

In this paper, we establish the complementary result that such an upper bound on variance in terms of the entropy power extends to the general class of symmetric unimodal distributions. The symmetric and unimodal distributions have been widely studied in probability theory, statistics, signal

processing and machine learning, in particular, for estimation and testing [5], [6], [7], [8], [9]. There exist many symmetric unimodal distributions that are not log concave, e.g., the class of generalized Gaussian densities defined in (14), where $\theta < 1$. Thus our extended entropy upper bound on variance may have broad applicability. For example, in Bayesian sequential optimal design of experiments [10], [11], [12], successive entropy minimization is often proposed as a way to progressively concentrate the posterior distribution. When combined with the results of [2] our results provide additional justification for such approaches when the posterior is either log-concave or symmetric and unimodal.

We consider the general class of symmetric unimodal densities that can be represented as a linear mixture

$$p(x) = \sum_{i=1}^n \alpha_i p_i(x) \quad \text{for } \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i = 1, \quad (9)$$

of either exponentially decreasing distributions with unbounded support, $p_i(x) \propto e^{-\beta_i |x-m|^{\theta_i}}$ for any $\beta_i > 0$, $\theta_i > 0$, $i = 1, \dots, n$, or uniform distributions with bounded support, $p_i(x) = \text{unif}\left(-\frac{1}{2\epsilon_i} + m, \frac{1}{2\epsilon_i} + m\right)$ for any $\epsilon_i > 0$, $i = 1, \dots, n$. We show that when the ratio between the maximum and minimum variances of the mixture components $p_i(x)$ is bounded, the variance of such a mixture distribution $p(x)$ can be bounded as

$$\frac{e^{2h(p)}}{2\pi e} \leq \text{var}(X) \leq \frac{c}{2\pi e} e^{2h(p)} \quad (10)$$

for some constant $c \geq 1$. Thus, for such symmetric unimodal distributions, as $h(p(X)) \rightarrow -\infty$, the variance is guaranteed to converge to 0. By constructing a counterexample, we also show that the ratio between the maximum and minimum variances of the mixture components must be bounded for the variance of symmetric unimodal distributions to be dominated by a monotonic function of the entropy power.

Note that, for sufficiently large order n , any symmetric unimodal distribution can be represented as a linear mixture of the form (9). This was established for the special case of uniform densities $p_i(x)$ by Feller in [13]. Specifically, when a symmetric unimodal distribution $p(x)$ with bounded support $\text{supp}(-s, s)$ is Lipschitz continuous, i.e., when there exists $\kappa > 0$ such that

$$\max_{|\delta| \leq h} |p(x) - p(x - \delta)| \leq \kappa h, \quad (11)$$

there exists a linear mixture of n -uniform distributions, denoted as $\bar{p}_n(x)$, such that the approximation error $\|p - \bar{p}_n\| = \left(\int_{-s}^s |p(x) - \bar{p}_n(x)|^2 dx\right)^{1/2}$ decreases on the order of $O(1/n)$, as does the difference between the variances of $p(x)$ and $\bar{p}_n(x)$:

$$\left| \int_{-s}^s (x-m)^2 p(x) dx - \int_{-s}^s (x-m)^2 \bar{p}_n(x) dx \right| = O(1/n) \quad (12)$$

for common mean m . Therefore, by combining (12) with (10), we can show that the variance of any Lipschitz continuous symmetric unimodal distribution with bounded support,

under the assumption of finite ratio between variances of the mixture components, can be bounded above and below by a constant scaling of entropy power in $\bar{p}_n(x)$ plus an error term decreasing on the order of $O(1/n)$, i.e.,

$$\frac{e^{2h(\bar{p}_n)}}{2\pi e} - \frac{c_1}{n} \leq \text{var}(X) \leq \frac{C \cdot e^{2h(\bar{p}_n)}}{2\pi e} + \frac{c_2}{n} \quad (13)$$

for some positive constants c_1 and c_2 . In Section II, we will state these results more rigorously and discuss the tightness of the bounds. We provide the proofs in the appendices.

II. UPPER BOUND ON THE VARIANCE OF SYMMETRIC UNIMODAL DISTRIBUTIONS WITH ENTROPY

In this section, we exhibit two classes of symmetric unimodal densities whose variance can be bounded above by its entropy power. By combining this upper bound with the entropy power lower bound on variance (3), we obtain a bound of the form (10).

A. Symmetric Unimodal Distributions with Unbounded Support

It is well known that for Gaussian distributions $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ with mean m and variance σ^2 , the entropy power and the variance have a direct monotonic relationship: $\sigma^2 = \frac{1}{2\pi e} e^{2h(p)}$. Therefore, the variance of Gaussian distributions decreases exponentially as the differential entropy decreases.

We are interested in finding such a monotonic relationship between variance and differential entropy for a larger class of distributions than the Gaussian. However, since the differential entropy cannot capture the concentration of density in general, it has no direct relevance to the variance of a general non-Gaussian distribution. What are the properties of the Gaussian distribution that enable a monotonic relationship between variance and entropy? The Gaussian distribution is a symmetric unimodal distribution for which the differential entropy is a function of only its variance. Moreover, since the Gaussian distribution has only one mode, as the distribution becomes more concentrated around its mode both variance and differential entropy monotonically decrease.

The unimodality and symmetry of the Gaussian distribution motivate the definition of a larger class of distributions for which variance has a monotonic relationship to its entropy power. Consider a symmetric unimodal class of generalized Gaussian densities

$$p(x) = \frac{1}{Z(\theta, \beta)} e^{-\beta |x-m|^\theta}, \quad \text{for } \beta, \theta > 0 \quad (14)$$

where Z is a normalizing constant: $Z(\theta, \beta) = \int_{-\infty}^{\infty} e^{-\beta |x-m|^\theta} dx$. The normalizing constant Z and the variance $\text{var}(X)$ of this distribution can be written in terms of β and θ as follows.

Proposition 1: For distributions of the form $p(x) = \frac{1}{Z(\theta, \beta)} e^{-\beta |x-m|^\theta}$ for $\beta, \theta > 0$, the normalizing constant Z

and the variance $\text{var}(X)$ can be expressed as

$$Z(\theta, \beta) = 2\beta^{-\frac{1}{\theta}} \theta^{-1} \Gamma\left(\frac{1}{\theta}\right), \quad (15)$$

$$\text{var}(X) = \beta^{-\frac{2}{\theta}} \frac{\Gamma\left(\frac{3}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)}, \quad (16)$$

where Γ denotes the Gamma function $\Gamma(t) := \int_0^\infty x^{t-1} e^{-x} dx$ for $t > 0$.

Proof of Proposition 1: Appendix A ■

We show that for the class of generalized Gaussian distributions in (14) the variance has an exact monotonic relationship to the entropy power for a fixed θ . Note that this set of symmetric unimodal distributions is not log concave for $\theta < 1$ so the results of [2] do not apply.

Lemma 1: For symmetric unimodal distributions of the form $p(x) = \frac{1}{Z(\theta, \beta)} e^{-\beta|x-m|^\theta}$ for $\beta, \theta > 0$, the variance is proportional to the entropy power

$$\text{var}(X) = \frac{1}{A(\theta)} e^{2h(p)}, \quad (17)$$

where

$$A(\theta) = 4\theta^{-2} \frac{(\Gamma(1/\theta))^3}{\Gamma(3/\theta)} e^{2/\theta}. \quad (18)$$

Proof of Proposition 1: Appendix B ■

In Fig. 1, we plot $1/A(\theta)$ versus θ (black solid line). The minimum of $1/A(\theta)$ is achieved at $\theta = 2$ with the value $1/(2\pi e) \approx 0.0585$. Note that when $\theta = 2$, i.e., $p(x)$ is a Gaussian distribution, it can be shown that $1/A(\theta) = 1/(2\pi e)$ by using $\Gamma(3/2) = \frac{1}{2}\Gamma(1/2)$ and $(\Gamma(1/2))^2 = \pi$. As θ decreases below 2, $1/A(\theta)$ increases, and it diverges as $\theta \rightarrow 0$. On the other hand, as θ increases above 2, $1/A(\theta)$ increases and converges to $1/12 \approx 0.0833$.

From the Gamma function property that $\Gamma(1-z)\Gamma(z) = \frac{\pi}{\sin(\pi z)}$, it follows that $\lim_{z \rightarrow 0} z\Gamma(z) = 1$, i.e., $\Gamma(z) \sim \frac{1}{z}$ as $z \rightarrow 0$, since $\Gamma(1) = 1$ and $\lim_{z \rightarrow 0} \sin(\pi z) = \pi z + O(z^3)$. Thus, when $\theta \rightarrow \infty$, $A(\theta) \sim 4\theta^{-2} \frac{\theta^3}{\theta^3} \exp(2/\theta)$ and $\lim_{\theta \rightarrow \infty} 1/A(\theta) = 1/12$.

On the other hand, by using Stirling's formula, as $z \rightarrow \infty$, $\Gamma(z+1) \sim \sqrt{2\pi z} \left(\frac{z}{e}\right)^z$. Using this approximation and the property that $\Gamma(z+1) = z\Gamma(z)$, it can be shown that $\frac{1}{\theta}\Gamma\left(\frac{1}{\theta}\right) = \Gamma\left(\frac{1}{\theta}+1\right) \sim \sqrt{2\pi} \frac{1}{\theta} \left(\frac{1}{\theta e}\right)^{\frac{1}{\theta}}$ and $\frac{3}{\theta}\Gamma\left(\frac{3}{\theta}\right) = \Gamma\left(\frac{3}{\theta}+1\right) \sim \sqrt{2\pi} \frac{3}{\theta} \left(\frac{3}{\theta e}\right)^{\frac{3}{\theta}}$. Therefore, as $\theta \rightarrow 0$, i.e., $1/\theta \rightarrow \infty$,

$$A(\theta) \sim 8\sqrt{3}\pi\theta^{-1} e^{\frac{1}{\theta}(2-3\log 3)}, \quad (19)$$

which goes to 0 since $(2-3\log 3) < 0$. Therefore, as $\theta \rightarrow 0$, $1/A(\theta)$ diverges.

By using these properties of $1/A(\theta)$, we can bound the variance of the generalized Gaussian distribution $p(x) = \frac{1}{Z(\theta, \beta)} e^{-\beta|x-m|^\theta}$ with $\beta, \theta > 0$ in terms of a constant scaling of entropy power, $ce^{2h(p)}$, with c independent of θ when θ is known to be larger than some positive constant.

Corollary 1: For symmetric unimodal densities of the form $p(x) = \frac{1}{Z(\theta, \beta)} e^{-\beta|x-m|^\theta}$ with $\beta > 0$, when $\theta \geq 2$

$$\text{var}(X) \leq \frac{1}{12} e^{2h(p)}. \quad (20)$$

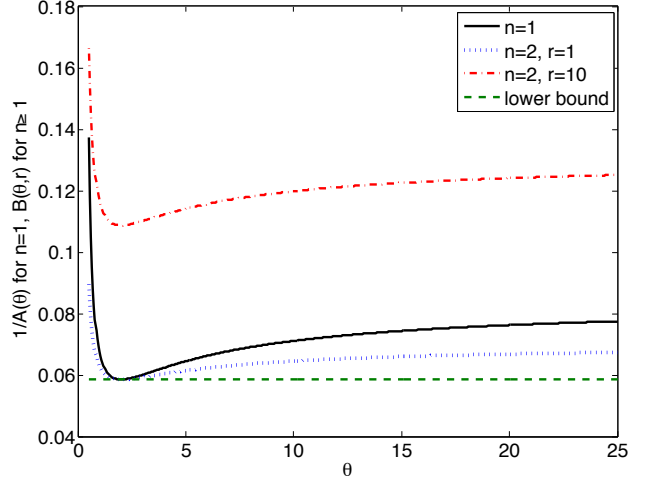


Fig. 1. The constant factor $1/A(\theta)$ in (18) vs. θ for $n = 1$ (black solid line); the constant factor $B(\theta, r)$ in (25) for $n = 2$ with mixture weights $\alpha_1 = \alpha_2 = 0.5$ and decaying orders $\theta_1 = 2$ and $\theta_2 = \theta$ when the ratio between variances $r = 1$ (blue dotted line) and when $r = 10$ (red dash-dot line). The lower bound $1/2\pi e$ is also shown (green dashed line).

When $\theta \geq 1/2$

$$\text{var}(X) \leq \frac{2e^4}{15} e^{2h(p)}. \quad (21)$$

We next consider a generalization of Lemma 1 for a much broader class of symmetric unimodal distributions. Let us consider a mixture distribution $p(x)$ composed of a finite number of exponentially decreasing distributions, $p_i(x) = \frac{1}{Z_i(\theta_i, \beta_i)} e^{-\beta_i|x-m|^{\theta_i}}$ with order $\theta_i > 0$ and $\beta_i > 0$ for $i = 1, \dots, n$, with mixture weights α_i , i.e.,

$$p(x) = \sum_{i=1}^n \alpha_i \left(\frac{1}{Z_i(\theta_i, \beta_i)} e^{-\beta_i|x-m|^{\theta_i}} \right) \quad (22)$$

where $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i = 1$. All the mixture components $p_i(x)$ have the same mean m to ensure unimodality of $p(x)$. Note that for each mixture component $p_i(x)$, the normalizing constant is $Z_i(\theta_i, \beta_i) = 2\beta_i^{-1/\theta_i} \theta_i^{-1} \Gamma(1/\theta_i)$ and the variance is $\sigma_i^2 := \beta_i^{-2/\theta_i} \frac{\Gamma(3/\theta_i)}{\Gamma(1/\theta_i)}$, as shown in Proposition 1. The variance of the symmetric unimodal mixture density (22) is

$$\text{var}(X) = \int_{-\infty}^{\infty} (x-m)^2 p(x) dx = \sum_{i=1}^n \alpha_i \sigma_i^2. \quad (23)$$

Under the mixture representation (22) we obtain an upper bound on the variance in terms of the entropy power.

Theorem 1: Let $p(x)$ be a symmetric unimodal density of the form $p(x) = \sum_{i=1}^n \alpha_i \left(\frac{1}{Z_i(\theta_i, \beta_i)} e^{-\beta_i|x-m|^{\theta_i}} \right)$ with a normalizing constant $Z_i(\theta_i, \beta_i)$ where $\beta_i, \theta_i > 0$, $\alpha_i \geq 0$, and $\sum_{i=1}^n \alpha_i = 1$. Assume that the ratio of component variances σ_i^2/σ_j^2 is bounded for all $i \neq j$. Then the variance of the density $p(x)$ satisfies

$$\frac{e^{2h(p)}}{2\pi e} \leq \text{var}(X) \leq B(\theta, r) e^{2h(p)}, \quad (24)$$

with

$$B(\theta, r) = M(r) \cdot \prod_{i=1}^n \left(\frac{1}{A(\theta_i)} \right)^{\alpha_i}, \quad (25)$$

for $\theta = (\theta_1, \dots, \theta_n)^T$, where

$$A(\theta) = 4\theta^{-2} \frac{(\Gamma(1/\theta))^3}{\Gamma(3/\theta)} e^{2/\theta}, \quad (26)$$

and, for $r := \max_{i,j \in \{1, \dots, n\}} \left\{ \frac{\sigma_i^2}{\sigma_j^2} \right\}$

$$M(r) := \frac{(r-1)r^{\frac{1}{r-1}}}{e \log r}. \quad (27)$$

Proof of Theorem 1: Appendix C ■

Corollary 2: For symmetric unimodal densities of the form

$p(x) = \sum_{i=1}^n \alpha_i \left(\frac{1}{\sqrt{2\pi} \beta_i} e^{-\beta_i |x-m|^{\theta_i}} \right)$, $\beta_i > 0$, $\alpha_i \geq 0$, and $\sum_{i=1}^n \alpha_i = 1$, where $\theta_i \geq 2$ for all i , the variance $\text{var}(X)$ is bounded as

$$\frac{e^{2h(p)}}{2\pi e} \leq \text{var}(X) \leq \frac{M(r)}{12} \cdot e^{2h(p)}. \quad (28)$$

When $\theta_i \geq 1/2$ for all i ,

$$\frac{e^{2h(p)}}{2\pi e} \leq \text{var}(X) \leq \frac{2e^4 M(r)}{15} \cdot e^{2h(p)}. \quad (29)$$

Equality in the variance lower bound in (24) is achieved if and only if $p(x)$ is a Gaussian distribution. The equality in the variance upper bound in (24) is met when all $p_i(x)$'s are the same distribution, i.e., $p_i(x) = p_j(x)$ for $\forall i \neq j$. Therefore, the upper and lower bounds become tight when all $p_i(x)$'s are the same Gaussian distribution.

For a mixture distribution of the form (22), possibly with $p_i(x) \neq p_j(x)$ for some pairs of (i, j) , the constant scale $B(\theta, r)$ in (25) is greater than or equal to $1/2\pi e$ since the geometric mean of $1/A(\theta_i) \geq 1/(2\pi e)$, $\forall \theta_i$, with orders $\{\alpha_i\}$, is $\prod_{i=1}^n (1/A(\theta_i))^{\alpha_i} \geq 1/(2\pi e)$, and $M(r) \geq 1$ with $\lim_{r \rightarrow 1} M(r) = 1$.

When all θ_i 's of $p_i(x)$ are lower bounded by some positive constant, we can simplify the upper bound on variance in Theorem 1 only in terms of $M(r)$, as stated in Corollary 2. When $\theta_i \geq 2$ for all i 's, the maximum ratio between the upper bound and the lower bound is equal to $\frac{2\pi e}{12} M(r)$ since $\prod_{i=1}^n (1/A(\theta_i))^{\alpha_i} \leq 1/12$. When $\theta_i \geq 1/2$ for all i 's, since $\prod_{i=1}^n (1/A(\theta_i))^{\alpha_i} \leq (2e^4)/15$, the maximum ratio between the upper bound and the lower bound becomes $\frac{4\pi e^5}{15} M(r)$.

In Fig. 1, we compare the constant factor $B(\theta, r)$ in the upper bound for three different mixture distributions of the form (22). The constant in the lower bound, $1/2\pi e$, is also plotted as the green dashed line. When $n = 1$, or equivalently, when all $p_i(x)$'s are the same distribution with $\theta = \theta_i$, $\forall i$, the factor $B(\theta, r) = 1/A(\theta)$. This case is denoted by the black solid line in Fig. 1. In the figure, we compare this case with that of two different mixture distributions when $n = 2$ and the mixture weights $\alpha_1 = \alpha_2 = 0.5$. Specifically, consider the case when a Gaussian distribution with $\theta_1 = 2$ is mixed with another distribution with $\theta_2 = \theta$. If the variance of the two distributions are the same, i.e., $r = 1$, the resulting $B(\theta, r)$ is the geometric mean of $1/A(2) = 1/(2\pi e)$ and $1/A(\theta) \geq 1/(2\pi e)$, and thus $B(\theta, r)$ is smaller than $1/A(\theta)$ for all $\theta \neq 2$ and equal to $1/A(\theta)$ at $\theta = 2$, as shown by the blue dotted line. On the other hand, when the ratio between

the variance of two mixture components is $r = 10$, $B(\theta, r)$ is increased by a factor of $M(10) \approx 1.86$ compared to the case when $r = 1$, as shown by the red dash-dot curve.

B. Symmetric Unimodal Distributions with Bounded Support

Among distributions with bounded support, the uniform distribution specifies a direct relationship between its variance and its entropy power: for a uniform distribution $p(x) = \text{unif}(m - \frac{1}{2\epsilon}, m + \frac{1}{2\epsilon})$, the variance is equal to $1/(12\epsilon^2)$ and the differential entropy is $\log(1/\epsilon)$. Therefore, for uniform distributions, we have

$$\text{var}(X) = \frac{e^{2h(p)}}{12}. \quad (30)$$

Moreover, a linear mixture of uniform distributions can represent more general symmetric unimodal distributions with bounded support. Let us consider a linear mixture of uniform distributions $p_i(x) = \text{unif}(m - \frac{1}{2\epsilon_i}, m + \frac{1}{2\epsilon_i})$, $\epsilon_i > 0$, $i = 1, \dots, n$, which are all centered at the mean $m \in \mathbb{R}$, i.e., $p(x) = \sum_{i=1}^n \alpha_i \cdot p_i(x)$, with mixture weights $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i = 1$. The variance of this distribution $\text{var}(X)$ is

$$\text{var}(X) = \frac{1}{12} \sum_{i=1}^n \alpha_i \frac{1}{\epsilon_i^2}, \quad (31)$$

since the variance of each uniform distribution is $\frac{1}{12\epsilon_i^2}$. The differential entropy of each mixture component $p_i(x)$ is

$$h(p_i) = \log(1/\epsilon_i). \quad (32)$$

We prove an upper bound on $\text{var}(X)$ in terms of the entropy power for such uniform mixtures. By combining with (3), the variance can be bounded below and above in terms of the entropy power.

Theorem 2: Assume that $p(x)$ is a bounded support symmetric unimodal density of the form $p(x) = \sum_{i=1}^n \alpha_i p_i(x)$ where $p_i(x) = \text{unif}(m - \frac{1}{2\epsilon_i}, m + \frac{1}{2\epsilon_i})$ for $\epsilon_i > 0$. Also, assume that $r := \max_{i,j \in \{1, \dots, n\}} \left\{ \frac{\epsilon_i^2}{\epsilon_j^2} \right\}$ is bounded. Then the variance of X is bounded as

$$\frac{e^{2h(p)}}{2\pi e} \leq \text{var}(X) \leq \frac{M(r)}{12} e^{2h(p)}, \quad (33)$$

where

$$M(r) := \frac{(r-1)r^{\frac{1}{r-1}}}{e \log r}. \quad (34)$$

Proof of Theorem 2: Appendix D ■

Remark 1: A more general result can be established for arbitrary Lipschitz continuous symmetric unimodal density of the form (11) where $\bar{p}_n(x) = \sum_{i=1}^n \alpha_i p_i(x)$. The bounds in this theorem hold up to $O(1/n)$ as indicated in (13).

The equality in the upper bound on variance is achievable when $\epsilon_i = \epsilon_j$ for $\forall i \neq j$, i.e., when $r = 1$, since $\lim_{r \rightarrow 1} M(r) = 1$. The ratio between the upper and the lower bound, which is proportional to $M(r)$, is increasing in r and as $r \rightarrow \infty$, $M(r) \sim \frac{r}{e \log r}$, which diverges.

C. Necessity of Bounded Variance Ratio

In both Theorem 1 and Theorem 2, we assumed boundedness of the ratio r between the maximum and minimum variances of the mixture components, i.e., $\max_{i,j \in \{1, \dots, n\}} \{\sigma_i^2/\sigma_j^2\}$ in Theorem 1 and $\max_{i,j \in \{1, \dots, n\}} \{\epsilon_i^2/\epsilon_j^2\}$ in Theorem 2 are bounded. Here we show by a counterexample that this assumption is *necessary* for existence of an upper bound on variance that is a scaled entropy power of the form $ce^{2h(p)}$ for some constant $c > 0$. When the ratio between the variances of the mixture components becomes unbounded, then for the mixture distribution $p(x)$ no such scale factor c exists. The following example illustrates this point. Consider a bounded support symmetric unimodal distribution composed of two uniform distributions of the form

$$p(x) = \sum_{i=1}^2 \alpha_i \cdot \text{unif}(-1/(2\epsilon_i), 1/(2\epsilon_i)) \quad (35)$$

where $\alpha_i > 0$ and $\sum_{i=1}^2 \alpha_i = 1$ for $\epsilon_1 > \epsilon_2 > 0$. The variance of this distribution is equal to

$$\text{var}(X) = \frac{1}{12} (\alpha_1/\epsilon_1^2 + \alpha_2/\epsilon_2^2), \quad (36)$$

and the differential entropy of this distribution is

$$h(p) = -(\alpha_1 + \alpha_2\epsilon_2/\epsilon_1) \log(\alpha_1\epsilon_1 + \alpha_2\epsilon_2) - (1/\epsilon_2 - 1/\epsilon_1) \alpha_2\epsilon_2 \log(\alpha_2\epsilon_2). \quad (37)$$

When $\epsilon_1/\epsilon_2 \rightarrow \infty$, the limit of the differential entropy becomes

$$\lim_{\epsilon_1/\epsilon_2 \rightarrow \infty} h(p) = -\alpha_1 \log \epsilon_1 - \alpha_2 \log \epsilon_2 + H_B(\alpha_1) \quad (38)$$

where $H_B(\alpha_1) = -\alpha_1 \log \alpha_1 - (1 - \alpha_1) \log(1 - \alpha_1)$. Then, the entropy power becomes

$$\lim_{\epsilon_1/\epsilon_2 \rightarrow \infty} e^{2h(p)} = e^{2H_B(\alpha_1)} (1/\epsilon_1^2)^{\alpha_1} (1/\epsilon_2^2)^{\alpha_2}. \quad (39)$$

In order to find an upper bound on the variance in (36) with the entropy power in (39), we need an upper bound on the arithmetic mean $(\alpha_1/\epsilon_1^2 + \alpha_2/\epsilon_2^2)$ in terms of the geometric mean $(1/\epsilon_1^2)^{\alpha_1} (1/\epsilon_2^2)^{\alpha_2}$. However, if $\epsilon_2 \rightarrow 0$ for a fixed ϵ_1 , since the arithmetic mean increases on the order of $\Theta(1/\epsilon_2^2)$ while the geometric mean on the order of $\Theta(1/\epsilon_2^{2\alpha_2})$, for $\alpha_2 < 1$ the variance increases much faster than $e^{2h(p)}$ so that it cannot be bounded above by any constant scaling of entropy power. On the other hand, if $\epsilon_1 \rightarrow \infty$ for a fixed ϵ_2 , the arithmetic mean, which is proportional to the variance, is approximately α_2/ϵ_2^2 , which is a constant, but the geometric mean, which is proportional to $e^{2h(p)}$, goes to 0 on the order of $\Theta(1/\epsilon_1^{2\alpha_1})$ for $\alpha_1 < 1$. Therefore, for both cases satisfying $\epsilon_1/\epsilon_2 \rightarrow \infty$, the variance of $p(x)$ cannot be bounded above by a constant scaling of $e^{2h(p)}$. This example shows that a bounded ratio between variances of mixture components is necessary in order for an upper bound on the variance of a symmetric unimodal distribution to exist that is a constant scaling of entropy power.

III. CONCLUSIONS

We established upper bounds on the variance of symmetric unimodal distributions in terms of entropy power. We considered symmetric unimodal mixture densities both for the case of mixtures with unbounded support (mixtures of generalized Gaussian densities with common mean) and mixtures with bounded support (mixtures of uniform densities with common mean). The tightness of the upper bound on variance depends on the ratio between the maximum and minimum variances of the mixture components. By constructing a counterexample, we showed that a bounded ratio between variances of mixture components is necessary in order for an upper bound on variance of symmetric unimodal mixture densities to exist that is a constant scaling of entropy power.

In signal processing, adaptive sensing, and machine learning, information theoretic surrogates such as Kullback-Leibler divergence, entropy, and Fisher information have been widely adopted in place of task-specific cost functions such as mean squared error or probability of classification error. Since such task-specific cost functions are often intractable, information theoretic surrogates are used as natural objectives for developing waveforms or sensor selection strategies to collect and/or filter information. The results reported in this paper can be used to justify the use of differential entropy as a surrogate for the mean squared error in such applications when the posterior is symmetric and unimodal such as the generalized Gaussian distributions.

APPENDIX A

PROOF OF PROPOSITION 1

The normalizing constant Z is $Z(\theta, \beta) = \int_{-\infty}^{\infty} e^{-\beta|x-m|^\theta} dx = 2 \int_0^{\infty} e^{-\beta x^\theta} dx$. Let $\beta x^\theta = y$. Then, $\beta \theta x^{\theta-1} dx = dy$ and $x^{\theta-1} = (y/\beta)^{\frac{\theta-1}{\theta}}$. Thus, Z can be written in terms of y as

$$Z(\theta, \beta) = 2 \int_0^{\infty} \beta^{-\frac{1}{\theta}} \theta^{-1} y^{-1+\frac{1}{\theta}} e^{-y} dy = 2\beta^{-\frac{1}{\theta}} \theta^{-1} \Gamma\left(\frac{1}{\theta}\right). \quad (40)$$

In a similar way,

$$\begin{aligned} \text{var}(X) &= \frac{1}{Z(\theta, \beta)} \int_{-\infty}^{\infty} (x-m)^2 e^{-\beta|x-m|^\theta} dx \\ &= \frac{2}{Z(\theta, \beta)} \int_0^{\infty} x^2 e^{-\beta x^\theta} dx \\ &= \frac{2}{Z(\theta, \beta)} \beta^{-\frac{3}{\theta}} \theta^{-1} \int_0^{\infty} y^{-1+\frac{3}{\theta}} e^{-y} dy \\ &= \frac{2}{Z(\theta, \beta)} \beta^{-\frac{3}{\theta}} \theta^{-1} \Gamma\left(\frac{3}{\theta}\right) = \beta^{-\frac{2}{\theta}} \frac{\Gamma\left(\frac{3}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)}. \end{aligned} \quad (41)$$

APPENDIX B

PROOF OF LEMMA 1

For $p(x) = \frac{1}{Z(\theta, \beta)} e^{-\beta|x-m|^\theta}$ where $\beta, \theta > 0$, the differential entropy can be directly calculated as follows.

$$h(p) = \log Z(\theta, \beta) + \frac{2\beta}{Z(\theta, \beta)} \int_0^{\infty} x^\theta e^{-\beta x^\theta} dx. \quad (42)$$

Let $\beta x^\theta = y$. Then, $\beta\theta x^{\theta-1}dx = dy$ and $x = (y/\beta)^{\frac{1}{\theta}}$. The differential entropy becomes

$$\begin{aligned} h(p) &= \log Z(\theta, \beta) + \frac{2\beta}{Z(\theta, \beta)} \int_0^\infty \beta^{-1}\theta^{-1}(y/\beta)^{\frac{1}{\theta}} e^{-y} dy \\ &= \log Z(\theta, \beta) + \frac{2\beta^{-\frac{1}{\theta}}\theta^{-1}}{Z(\theta, \beta)} \int_0^\infty y^{\frac{1}{\theta}} e^{-y} dy \\ &= \log Z(\theta, \beta) + \frac{2\beta^{-\frac{1}{\theta}}\theta^{-1}}{Z(\theta, \beta)} \Gamma\left(1 + \frac{1}{\theta}\right). \end{aligned} \quad (43)$$

By using the normalizing constant Z in (40) and the variance in (41) as well as the property that $\Gamma(1+z) = z\Gamma(z)$ for $z > 0$, the entropy power can be written in terms of variance as follows.

$$\begin{aligned} e^{2h(p)} &= 4\beta^{-\frac{2}{\theta}}\theta^{-2} (\Gamma(1/\theta))^2 e^{2/\theta} \\ &= A(\theta) \cdot \text{var}(X) \end{aligned} \quad (44)$$

APPENDIX C

PROOF OF THEOREM 1

Consider the generalized Gaussian mixture distribution $p(x) = \sum_{i=1}^n \alpha_i p_i(x)$ where $p_i(x) = \frac{1}{Z_i(\theta_i, \beta_i)} e^{-\beta_i|x-m|^{\theta_i}}$ for $\beta_i, \theta_i > 0$. Using the concavity of the differential entropy $h(p)$ in distribution $p(x)$,

$$h(p) \geq \sum_{i=1}^n \alpha_i h(p_i), \quad (45)$$

and thus

$$e^{2h(p)} \geq e^{\sum_{i=1}^n 2\alpha_i h(p_i)} = \prod_{i=1}^n \left(e^{2h(p_i)}\right)^{\alpha_i}. \quad (46)$$

As shown in Lemma 1, for $p_i(x) = \frac{1}{Z_i(\theta_i, \beta_i)} e^{-\beta_i|x-m|^{\theta_i}}$,

$$\sigma_i^2 = \frac{1}{A(\theta_i)} \cdot e^{2h(p_i)}, \quad (47)$$

and thus

$$e^{2h(p)} \geq \left(\prod_{j=1}^n A(\theta_j)^{\alpha_j}\right) \cdot \left(\prod_{i=1}^n (\sigma_i^2)^{\alpha_i}\right). \quad (48)$$

By using the reverse power mean inequality shown in [14], a lower bound on the geometric mean of $\{\sigma_i^2\}$ with orders $\{\alpha_i\}$ in terms of the arithmetic mean of $\{\sigma_i^2\}$ with orders $\{\alpha_i\}$ is given by

$$\sum_{i=1}^n \alpha_i \sigma_i^2 \leq M(r) \prod_{i=1}^n (\sigma_i^2)^{\alpha_i} \quad (49)$$

where $M(r)$ is defined as in (27) and $r := \max_{i,j \in \{1, \dots, n\}} \left\{ \frac{\sigma_i^2}{\sigma_j^2} \right\}$.

Since the variance of the mixture distribution $p(x) = \sum_{i=1}^n \alpha_i p_i(x)$ is $\text{var}(X) = \sum_{i=1}^n \alpha_i \sigma_i^2$, by combining (48) and (49),

$$\begin{aligned} \text{var}(X) &= \sum_{i=1}^n \alpha_i \sigma_i^2 \leq M(r) \prod_{i=1}^n (\sigma_i^2)^{\alpha_i} \\ &\leq M(r) \left(\prod_{i=1}^n \left(\frac{1}{A(\theta_i)}\right)^{\alpha_i}\right) e^{2h(p)}. \end{aligned} \quad (50)$$

APPENDIX D

PROOF OF THEOREM 2

Consider the uniform mixture distribution $p(x) = \sum_{i=1}^n \alpha_i p_i(x)$ where $p_i(x) = \text{unif}\left(m - \frac{1}{2\epsilon_i}, m + \frac{1}{2\epsilon_i}\right)$ for $\epsilon_i > 0$. By using the concavity of differential entropy, for $p(x) = \sum_{i=1}^n \alpha_i p_i(x)$ where $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i = 1$,

$$h(p) \geq \sum_{i=1}^n \alpha_i h(p_i) = \sum_{i=1}^n \alpha_i \log \frac{1}{\epsilon_i}, \quad (51)$$

which is equivalent to,

$$e^{2h(p)} \geq \prod_{i=1}^n \left(\frac{1}{\epsilon_i}\right)^{\alpha_i}. \quad (52)$$

By using the reverse power mean inequality shown in [14], the following lower bound on the weighted geometric mean of $\{1/\epsilon_i^2\}$ holds, in terms of the weighted arithmetic mean of $\{1/\epsilon_i^2\}$

$$\sum_{i=1}^n \alpha_i \frac{1}{\epsilon_i^2} \leq M(r) \prod_{i=1}^n \left(\frac{1}{\epsilon_i^2}\right)^{\alpha_i} \quad (53)$$

where $M(r)$ is defined as in (34) and $r := \max_{i,j \in \{1, \dots, n\}} \left\{ \frac{\epsilon_i^2}{\epsilon_j^2} \right\}$.

Since the variance of the mixture distribution $p(x)$ is $\text{var}(X) = \frac{1}{12} \sum_{i=1}^n \alpha_i \frac{1}{\epsilon_i^2}$, by combining (52) and (53),

$$\begin{aligned} \text{var}(X) &= \frac{1}{12} \sum_{i=1}^n \alpha_i \frac{1}{\epsilon_i^2} \leq \frac{M(r)}{12} \prod_{i=1}^n \left(\frac{1}{\epsilon_i^2}\right)^{\alpha_i} \\ &\leq \frac{M(r)}{12} e^{2h(p)}. \end{aligned} \quad (54)$$

REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [2] S. Bobkov and M. Madiman, "The entropy per coordinate of a random vector is highly constrained under convexity conditions," *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 4940–4954, 2011.
- [3] G. Walther, "Inference and modeling with log-concave distributions," *Statistical Science*, pp. 319–327, 2009.
- [4] M. Bagnoli and T. Bergstrom, "Log-concave probability and its applications," *Economic theory*, vol. 26, no. 2, pp. 445–469, 2005.
- [5] B. P. Rao, "Estimation of a unimodal density," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 23–36, 1969.
- [6] E. J. Wegman, "Maximum likelihood estimation of a unimodal density, ii," *The Annals of Mathematical Statistics*, pp. 2169–2174, 1970.
- [7] P. Groeneboom, "Estimating a monotone density," *Department of Mathematical Statistics*, no. R 8403, pp. 1–14, 1984.
- [8] L. Birge, "Estimation of unimodal densities without smoothness assumptions," *The Annals of Statistics*, pp. 970–981, 1997.
- [9] J. A. Hartigan and P. Hartigan, "The dip test of unimodality," *The Annals of Statistics*, pp. 70–84, 1985.
- [10] M. H. DeGroot, "Uncertainty, information, and sequential experiments," *The Annals of Mathematical Statistics*, pp. 404–419, 1962.
- [11] D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 1, pp. 1–14, 1996.
- [12] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for target localization," *Information Theory, IEEE Transactions on*, vol. 60, no. 4, pp. 2233–2252, 2014.
- [13] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, 2008, vol. 2.
- [14] W. Specht, "Zur theorie der elementaren mittel," *Mathematische Zeitschrift*, vol. 74, no. 1, pp. 91–98, 1960.