MaxEnt Upper Bounds for the Differential Entropy of Univariate Continuous Distributions

Frank Nielsen, *Senior Member, IEEE*, and Richard Nock

***Abstract*—We present a series of closed-form upper bounds of the differential entropy of univariate continuous distributions based on the maximum entropy principle. We apply those bounds to Gaussian mixture models, and study their tightness properties.**

***Index Terms*—Absolute geometric moment, absolute monomial exponential family (EF), differential entropy, maximum entropy principle, mixture model.**

I. Introduction

SHANNON's *differential entropy* [1] $h(X)$ of a *continuous random variable* X following a probability density function $p(x)$ (denoted by $X \sim p(x)$) on the support $\mathcal{X} = \{x \in \mathbb{R} : p(x) > 0\}$ quantifies the uncertainty [1] of X as follows:

$$h(X) = \int_{\mathcal{X}} p(x) \log \frac{1}{p(x)} \mathrm{d}x = -\int_{\mathcal{X}} p(x) \log p(x) \mathrm{d}x. \quad (1)$$

When the logarithm is expressed in basis 2, the entropy is measured in *bits*. When using the natural logarithm (basis e), the entropy is measured in *nats*. The differential entropy functional $h(\cdot)$ is *concave* [1], may be *negative,*[1] and may be potentially *infinite*[2] when the integral of (1) diverges.

Although closed-form formulas for the differential entropy are available for many statistical distributions (see [2]–[4]), the *differential entropy of mixtures* usually does not admit closed-form expressions [5], [6] because the log term in (1) transforms into an intractable *log-sum term* when dealing with mixture densities. Let us denote by $m(x) = \sum_{i=1}^k w_i p_i(x)$ the density of a mixture[3] $M \sim m(x)$ with k components $X_i \sim p_i(x)$, where $w \in \Delta_{k-1}$. Δ_{k-1} denotes the $(k-1)$-dimensional open probability simplex. That is, a mixture is a convex combination of component distributions $p_1(x), \ldots, p_k(x)$. We shall consider mixtures of Gaussians $X_i \sim N(\mu_i, \sigma_i)$ with component probability density functions $p_i(x) = p(x; \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{(x-\mu_i)^2}{2\sigma_i^2})$, where $\mu_i = E[X_i]$ is the expectation parameter and $\sigma_i = \sqrt{E[(X_i - \mu_i)^2]}$ the standard deviation of X_i. Mixtures provably allow flexible fine modeling of *arbitrary* smooth densities, and are thus often met in applications. Thus, to tackle the computation of the differential entropy of mixtures, various *approximation techniques* have been designed (see [7] and references therein). In practice, to estimate $h(X)$ with $X \sim p(x)$, one uses Monte-Carlo stochastic integration:

$$\hat{h}_s(X) = -\frac{1}{s} \sum_{i=1}^s \log p(x_i) \quad (2)$$

for $x_1, \ldots, x_s$ an independent and identically distributed (iid) set of variates sampled from $X \sim p(x)$. This estimator $\hat{h}_s(X)$ is consistent (i.e., $\lim_{s\to\infty} \hat{h}_s(X) = h(X)$, convergence in probability) provided that $E[\log^2 p(x)] < \infty$. In [8], the differential entropy of univariate Gaussian mixtures is both lower and upper bounded by using log-sum-exp inequalities. Moshksar and Khandani [9] considered isotropic spherical gaussian mixture models (GMMs with identical standard deviation), and used Taylor expansions to arbitrarily finely approximate the differential entropy of those particular isotropic GMMs. Interestingly, they mentioned the so-called *maximum entropy upper bound* [9] (MEUB) that relies on the fact that the continuous distribution with prescribed variance maximizing the differential entropy is the Gaussian distribution of same variance. Since the entropy of a univariate Gaussian $N(\mu, \sigma)$ is $\frac{1}{2}\log(2\pi e \sigma^2)$, we get the following upper bound.

Lemma 1 (Gaussian MaxEnt Upper Bound): The differential entropy of a continuous random variable X defined on the support $\mathcal{X} = \mathbb{R}$ is upper bounded by

$$\mathrm{MEUB}: h(X) \leq \frac{1}{2} \log \left(2\pi e V[X]\right) \quad (3)$$

where $V[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$ denotes the variance of X.

In particular, since the variance of a Gaussian mixture $M \sim \sum_{i=1}^k w_i p_i(x)$ is $V[M] = \sum_{i=1}^k w_i((\mu_i - \bar{\mu})^2 + \sigma_i^2))$ with $\bar{\mu} = \sum_{i=1}^k w_i \mu_i$), (3) yields the following corollary.

Manuscript received November 16, 2016; revised January 17, 2017; accepted February 5, 2017. Date of publication February 9, 2017; date of current version March 1, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ashish Pandharipande. (*Corresponding author: Frank Nielsen.*)

F. Nielsen is with École Polytechnique, Paris 91128, France, and also with Sony Computer Science Laboratories Inc., Tokyo 141-0022, Japan (e-mail: Frank.Nielsen@acm.org).

R. Nock is with Data61, Canberra, ACT 2601, Australia, with the Australian National University, Canberra, ACT 0200, Australia, and also with the University of Sydney, Camperdown, NSW 2006, Australia (e-mail: Richard.Nock@data61.csiro.au).

Digital Object Identifier 10.1109/LSP.2017.2666792

[1]For example, when $X \sim N(\mu, \sigma)$ is a Gaussian distribution of mean μ and standard deviation $\sigma > 0$, then $h(X) = \frac{1}{2}\log(2\pi e \sigma^2)$, and is therefore negative when $\sigma < \frac{1}{\sqrt{2\pi e}}$.

[2]For example, consider $X \sim p(x)$ with $p(x) = \frac{\log(2)}{x \log^2 x}$ for $x > 2$ (with support $\mathcal{X} = (2, \infty)$). Then, $h(X) = +\infty$. This result is to contrast with the fact that the entropy $H(X)$ of a discrete distribution X on a finite alphabet $\mathcal{X}$ is upper bounded by $\log|\mathcal{X}|$.

[3]Beware that the mixture random variable $M \neq \sum_i w_i X_i$. The probability density of a weighted sum of random variables is obtained by convolution of the densities.

Corollary 1 (GMM Gaussian MEUB): The differential entropy of a GMM $M \sim \sum_{i=1}^k w_i p_i(x)$ is upper bounded by

$$h(M) \leq \frac{1}{2}\log\left(2\pi e \sum_{i=1}^k w_i((\mu_i - \bar{\mu})^2 + \sigma_i^2))\right) \qquad (4)$$

where $\bar{\mu} = \sum_{i=1}^k w_i \mu_i$.

In this letter, we propose to further use the maximum entropy principle to derive a *collection* of MaxEnt upper bounds. Although our bounds will be mainly instantiated for GMMs, they apply to *any* univariate continuous (mixture) distribution. In particular, our MaxEnt upper bounds hold for mixtures of exponential families [10] that include GMMs and have always finite differential entropy.

II. MaxEnt Upper Bounds on Differential Entropy

Jaynes' MaxEnt principle [11], [12] defines a unique distribution satisfying a set of prescribed "moment constraints" that maximizes the differential entropy as follows:

$$\max_{p : X \sim p} h(X) \ : \ E[t_i(X)] = \eta_i, \quad i \in \{1, \ldots, D\}. \qquad (5)$$

When an iid dataset $x_1, \ldots, x_s$ is given, we may choose, for example, the raw geometric *sample moments* $\eta_i = \frac{1}{s}\sum_{j=1}^s {x_j}^i$ for setting the constraints $E[X^i] = \eta_i$ (taking $t_i(X) = X^i$). The distribution $p(x)$ maximizing the entropy under those moment constraints is *unique* and called the *MaxEnt distribution*. The constrained optimization of (5) is solved using the Lagrangian multipliers [1], [13]. It is well known [13], [14] that the MaxEnt distribution $p(x)$ belongs to a *parametric family* of distributions called an *EF* [15]. By abuse of notation, we also equivalently note $E_p[t_i(X)] = E[t_i(X)]$ for $X \sim p(x)$. An [15] EF admits the following canonical probability density function:

$$p(x; \theta) = \exp\left(\langle \theta, t(x)\rangle - F(\theta)\right) \qquad (6)$$

where $\langle a, b\rangle = a^\top b$ denotes the scalar product, θ the natural parameter vector, and $t(x) = (t_1(x), \ldots, t_D(x))$ the vector of sufficient statistics [15]. Thus, the MaxEnt distribution belongs to the EF with sufficient statistic vector $t(x)$, and the natural parameter θ of the unique MaxEnt distribution is given by the Lagrangian multipliers [1], [13]. It follows that we have $E[t(X)] = \eta$ for some unique random variable $X \sim p(x; \theta)$ with $\theta \in \Theta = \{\theta \ : \ \int \exp(\theta^\top t(x))\mathrm{d}x < \infty\}$ called the *natural parameter space* [15]. The quantity $F(\theta) = \log \int \exp(\theta^\top t(x))\mathrm{d}x$ is called the log-normalizer [15] since it allows to normalize the density to a probability: $\int_{\mathcal{X}} p(x; \theta)\mathrm{d}x = 1$. By construction, *any other distribution* with density $q(x)$ different from the MaxEnt distribution $p(x)$ and satisfying all the D moment constraints $E_q[t_i(X)] = \eta_i$ will have necessarily smaller entropy: $h(q(x)) \leq h(p(x))$ with $p(x) = p(x; \theta)$. In general, neither θ nor $F(\theta)$ may be available in closed forms, and thus need to be approximated numerically [13]. A distribution belonging to an EF can either be indexed by its natural parameter θ or equivalently by its moment parameter $\eta(\theta) = E[t(X)] = \nabla F(\theta)$ (see [15], [16]). That is, $p(x; \theta) = p(x; \eta)$ with $\eta = \nabla F(\theta)$. By construction, the MaxEnt distribution has moment parameter $\eta = (\eta_1, \ldots, \eta_D)$ of (5).

In the remainder, we upper bound the differential entropy of a continuous distribution by building a *collection* of upper bounds derived from MaxEnt distributions which admit closed-form expressions for their differential entropies. In practice, those bounds prove very useful for GMMs since the differential entropy of a GMM is not available in closed form [5], [8].

A. MaxEnt Upper Bounds From Geometric Absolute Moments

Consider the univariate uni-order ($D = 1$) EF of *Absolute Monomial EF* (AMEF) induced by the absolute value of a monomial of degree $l \in \mathbb{N}$ defined over the full support $\mathcal{X} = \mathbb{R}$:

$$p_l(x; \theta) = \exp\left(\theta |x|^l - F_l(\theta)\right) \qquad (7)$$

for $\theta < 0$ (with natural parameter space $\Theta = (-\infty, 0)$), sufficient statistic $t(x) = |x|^l$, and log-normalizer[4]

$$F_l(\theta) = \log \frac{2}{l}\Gamma\left(\frac{1}{l}\right) - \frac{1}{l}\log(-\theta) \qquad (8)$$

where log is the natural logarithm and $\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x)\mathrm{d}x$ the Gamma function that generalizes the factorial ($\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$). The Gamma function can be approximated finely, and in fact, even better, it is $\log \Gamma(u)$ that can be calculated quickly (see the numerical recipe in [20]), so that the log-normalizer of (8) can be calculated for any $l \in \mathbb{N}$ and $\theta < 0$. Since those AMEF densities are even functions ($p_l(-x; \theta) = p_l(x; \theta)$), their expectations are zero: $E[X] = 0$ for $X \sim p_l(x; \theta)$ for any $l \in \mathbb{N}$. The differential entropy $h_l(\theta) = h(p_l(x; \theta))$ admits the following closed-form formula using the moment parameter $\eta = E[t(X)]$:

$$h_l(\eta) = b_l + \frac{1}{l}\log \eta \qquad (9)$$

with $b_l = \log \frac{2\Gamma(\frac{1}{l})(el)^{\frac{1}{l}}}{l}$ a constant at fixed l, independent of the moment parameter η.

To upper bound the entropy of *any arbitrary continuous distribution* X (let it be a mixture or not), we simply calculate the distribution lth raw absolute geometric moment $E[|X|^l]$, and deduce the following MaxEnt entropy upper bound U_l: $h(X) \leq h_l(E[|X|^l])$.

Theorem 1 (AMEF MaxEnt Upper Bounds): Let X be a continuous random variable with support $\mathcal{X} = (-\infty, \infty)$. We get the following collection of MaxEnt Upper Bounds (MEUBs) of the differential entropy of X:

$$U_l : h(X) \leq b_l + \frac{1}{l}\log(E[|X|^l]), \quad \forall l \in \mathbb{N} \qquad (10)$$

with $b_l = \log \frac{2}{l}\Gamma\left(\frac{1}{l}\right) + \frac{1}{l}(1 + \log l)$.

Using the identity $x\Gamma(x) = \Gamma(1+x)$, we may rewrite $b_l = \log 2\Gamma(1 + \frac{1}{l}) + \frac{1}{l}(1 + \log l)$. Note that for even integer l, the absolute geometric moments amount to geometric moments: $E[|X|^l] = E[X^l]$.

Let us give two AMEF MaxEnt distributions and their corresponding differential entropy formulas.

[4]This integral can be computed using any computer algebra systems (CASs) like `maxima` that can be downloaded at http://maxima.sourceforge.net/. CASs implement the semi-algorithm of Risch [17] for symbolic integration, and can therefore determine whether the integral admits a closed form or not in terms of elementary functions. See [18], [19] for recent developments.

TABLE I
FIRST TEN ABSOLUTE GEOMETRIC MOMENTS $E[|X|^l]$ OF $X \sim N(\mu,\sigma)$ WITH $\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t$ DENOTING THE ERROR FUNCTION

l	$E\left[\lvert X\rvert^l\right]$
1	$\sigma\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2})+\mu\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
2	$\mu^2+\sigma^2$
3	$(2\sigma^3+\mu^2\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2})+(\mu^3+3\mu\sigma^2)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
4	$\mu^4+6\mu^2\sigma^2+3\sigma^4$
5	$(8\sigma^5+9\mu^2\sigma^3+\mu^4\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2})+(\mu^5+10\mu^3\sigma^2+15\mu\sigma^4)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
6	$\mu^6+15\mu^4\sigma^2+45\mu^2\sigma^4+15\sigma^6$
7	$(48\sigma^7+87\mu^2\sigma^5+20\mu^4\sigma^3+\mu^6\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2})+(\mu^7+21\mu^5\sigma^2+105\mu^3\sigma^4+105\mu\sigma^6)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
8	$\mu^8+28\mu^6\sigma^2+210\mu^4\sigma^4+420\mu^2\sigma^6+105\sigma^8$
9	$(384\sigma^9+975\mu^2\sigma^7+345\mu^4\sigma^5+35\mu^6\sigma^3+\mu^8\sigma)\sqrt{\frac{2}{\pi}}\exp(-\frac{\mu^2}{2\sigma^2})+(\mu^9+36\mu^7\sigma^2+378\mu^5\sigma^4+1260\mu^3\sigma^6+945\sigma^8)\mathrm{erf}(\frac{\mu}{\sqrt{2}\sigma})$
10	$\mu^{10}+45\mu^8\sigma^2+630\mu^6\sigma^4+3150\mu^4\sigma^6+4725\mu^2\sigma^8+945\sigma^{10}$

1) Consider $l=2$. Since $\Gamma(\frac{3}{2})=\frac{\sqrt{\pi}}{2}$, we get $F_2(\theta)=\frac{1}{2}\log\frac{\pi}{-\theta}$. Thus, $p_2(x;\theta)=\sqrt{\frac{-\theta}{\pi}}\exp(\theta x^2)$. By setting $\theta=-\frac{1}{2\sigma^2}$, we get the canonical *standard Gaussian density*: $\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{x^2}{2\sigma^2})$. Since $\theta=-\frac{1}{2\sigma^2}$, we recover the usual entropy of a Gaussian: $h(p_2(x;\theta))=\sqrt{\frac{-\theta}{\pi}}\frac{\sqrt{\pi}}{2\sqrt{-\theta}}+\log\sqrt{\frac{\pi}{-\theta}}$. Plugging $\theta=-\frac{1}{2\sigma^2}$, we recover the formula $\frac{1}{2}\log(2\pi e\sigma^2)$.
2) Consider $l=1$. The MaxEnt distribution is the *standard Laplacian distribution* [14] with density written canonically as $p(x;\theta)=\exp(\theta|x|-\log(-\frac{2}{\theta}))$ with $F_1(\theta)=\log(-\frac{2}{\theta})$. The differential entropy can be expressed using *either* the natural or expectation parameter as $h(p(x;\theta))=1+\log(\frac{2}{-\theta})$ or $h(p(x;\eta))=1+\log(2\eta)$ with $\eta=F_1'(\theta)=-\frac{1}{\theta}$.

In general, the differential entropy of a AMEF distribution of degree l is negative when $\eta<\frac{l^l}{le(2\Gamma(\frac{1}{l}))^l}$, and nonnegative otherwise. Note that when $l=2$, the AMEF is the Gaussian family, and $\Gamma(\frac{1}{2})=\sqrt{\pi}$, so that we recover the negative entropy condition $\eta=\sigma^2<\frac{1}{2\pi e}$ and therefore $\sigma<\frac{1}{\sqrt{2\pi e}}$, as already claimed in the introduction.

We also report the differential entropy of an AMEF distribution $p_l(x;\theta)$ indexed by its natural parameter θ:

$$h_l(\theta)=a_l-\frac{1}{l}\log(-\theta) \tag{11}$$

where $a_l=\log\frac{2}{l}\Gamma\left(\frac{1}{l}\right)+\frac{1}{l}$ is a constant independent of θ.

B. MaxEnt Upper Bounds for GMMs and Other Mixtures

In order to apply the upper bound U_l for a GMM with probability density function $m(x)=\sum_{i=1}^k w_i p(x;\mu_i,\sigma_i)$, we need to compute the absolute raw geometric moment for odd integer l, and the raw geometric moment for even integer l, and plug those values into formula (10). By linearity of the expectation operator, we have $E_{m(x)}[|X|^r]=\sum_{i=1}^k w_i E_{p(x;\mu_i,\sigma_i)}[|X|^r]$. We give in Appendix B closed-form formulas for the (absolute) raw geometric moments of an *uncentered* univariate normal distribution. Table I reports the first ten absolute geometric moments of $X\sim N(\mu,\sigma)$.

For $l=2$, we recover the well-known Gaussian MaxEnt upper bound [7] U_2:

$$\mathrm{MEUB}: h(X)\leq\frac{1}{2}\log\left(2\pi e\sum_{i=1}^k w_i((\mu_i-\bar{\mu})^2+\sigma_i^2)\right) \tag{12}$$

with $\bar{\mu}=\sum_{i=1}^k w_i\mu_i$.

For $l=1$, considering the formula of $E[|X|]$ in Table I, we get the *Laplacian MaxEnt upper bound* U_1.

Since the differential entropy of a distribution X belonging to a location family $X\sim p_0(x;\mu)$ *does not* change by translating the distribution (see Appendix A), we translate the mixture M so that $E[M]=0$ (thus aligning the mixture with the AMEF so that both their expectations are zero) before applying the MaxEnt upper bounds.

Note that this series of upper bounds U_l induced by MaxEnt AMEF distributions apply to any continuous distribution X provided that the absolute raw geometric moments are computable or can be efficiently upper bounded. This is in particular true for univariate mixtures of exponential families with a polynomial sufficient statistic $t(X)$. Indeed, for a distribution belonging to an EF, we calculate the moments from its moment generating function (MGF) [15]

$$M(t(u))=E[\exp(\langle u,t(X)\rangle)]=\exp(F(\theta+u)-F(\theta)). \tag{13}$$

The geometric moments $E[t(X)^l]$ are given by the higher order derivatives $E[t(X)^l]=M^{(l)}(0)$. For uni-order exponential families, it follows that $E[t(X)]=F'(\theta)=\eta$. It follows that EFs have always all finite-order moments expressed using the higher order derivatives of the MGF. Thus, we can always explicitly calculate the geometric moments of the sufficient statistic $t(X)$ from the MGF provided that the log-normalizer $F(\theta)$ is available in closed form. For example, we may consider mixtures of Rayleigh distributions (with $t(x)=x^2$, see [15]) instead of GMMs, and get closed-form MaxEnt upper bounds. The density of a Rayleigh distribution is $p(x;\sigma)=\frac{x}{\sigma^2}\exp(-\frac{x^2}{2\sigma^2})$ for $x>0$ and scale parameter $\sigma>0$. The geometric raw moments of a Rayleigh mixture X is $E[X^l] = 2^{\frac{l}{2}}\Gamma(1+\frac{l}{2})\sum_{i=1}^k \sigma_i^l$. AMEFs can be considered on the support $(0,\infty)$ and Rayleigh MaxEnt upper bounds obtained.

C. Analysis of MaxEnt Upper Bounds

Since we derived an *infinite* collection of MaxEnt upper bounds U_l, we study whether all those bounds can be potentially useful or not. First, let us show that bound U_1 (Laplacian MaxEnt Upper Bound) maybe better than U_2 (Gaussian MaxEnt Upper Bound). We consider the *restricted case* of zero-centered GMMs [21]. In that case, we have $E[|X|] = \sqrt{\frac{2}{\pi}}\sum_{i=1}^k w_i\sigma_i = \sqrt{\frac{2}{\pi}}\bar{\sigma}$ (where $\bar{\sigma} = \sum_{i=1}^k w_i\sigma_i$), and therefore get the upper bound U_1 as

$$h(X) \leq U_1 = 1 + \log 2\sqrt{\frac{2}{\pi}}\left(\sum_{i=1}^k w_i\sigma_i\right). \tag{14}$$

This bound is strictly better than the traditional variance bound [9] $U_2 = \log\sqrt{2\pi e}\sqrt{\sum_{i=1}^k w_i\sigma_i^2}$ provided that $U_1 < U_2$. Note that when $l=2$ and $k=1$ that U_2 bounds matches precisely the entropy of the single-component Gaussian mixture. Let $\overline{\sigma}_1$ be the *arithmetic weighted mean* and $\overline{\sigma}_2 = \sqrt{\sum_{i=1}^k w_i\sigma_i^2}$ be the *quadratic mean* of the weighted standard deviation. Then, $U_1 < U_2$ if and only if

$$\log 2e\sqrt{\frac{2}{\pi}}\bar{\sigma_1} \leq \log\sqrt{2\pi e}\overline{\sigma_2}. \tag{15}$$

That is, we need to have $\frac{\overline{\sigma}_1}{\overline{\sigma}_2} \leq \frac{\pi}{2\sqrt{e}} \approx 0.9527$. Observe that the *weighted quadratic mean* dominates the *weighted arithmetic mean*, and therefore $\frac{\overline{\sigma}_1}{\overline{\sigma}_2} \leq 1$. Equality of arithmetic/quadratic means only happens when all the σ_i's coincide. For the degenerate case $k=1$ (single component $X = N(0,\sigma)$), the condition of $U_1 < U_2$ (U_1 tighter than U_2) writes as $\sigma > \frac{2\sqrt{e}}{\pi}$ (that is, $\sigma > 1.0496$).

We consider the following experiment repeated $t = 1000$ times: We draw of a GMM $X \sim m$ with $k=2$ components with $\mu_i, \sigma_i \sim_{\text{iid}} U(0,1)$ and $w_i \sim_{\text{iid}} U(0,1)$ chosen as uniform weights renormalized to 1, we recenter the GMM so that $\bar{\mu}' = 0$, and compute the stochastic approximation $\hat{h}$ of $h(X)$ (for $s = 10^6$ samples), and the first-order and second-order MEUBs $h_1(E_m[|X|])$ and $h_2(E_m[X^2])$. We report average approximations $|\frac{\widehat{h_i}-\hat{h}}{\hat{h}}|$, and the percentage of times MEUB $U_1(X) < U_2(X)$: Gaussian MEUB is on average 40% above $\hat{h}$ and Laplacian MEUB is on average 10% above $\hat{h}$. Laplacian MEUB bound beats the Gaussian MEUB 32.9% on average.

To show that all bounds U_l are potentially useful (meaning $U_l < U_{l-1}$), we consider the following two-component mixture model: $X \sim m(x) = \frac{1}{2}p(x; -\frac{1}{2}, 10^{-5}) + \frac{1}{2}p(x; \frac{1}{2}, 10^{-5})$. Then, experimentally $U_{i+1}(X) \leq U_i(X)$ for $i \leq 37$ (we get NaN numerical errors when computing U_{38} using the closed-form formula, see [22]).

III. Conclusion

We considered the novel parametric family of AMEFs, and reported a closed-form differential entropy formula for these AMEFs. We then considered a collection of MEUBs derived from the AMEFs for an arbitrary continuous random variable that requires to calculate the raw geometric absolute moments (Theorem 1). We show how to apply those MEUBs for the case of GMMs. In particular, we show that the Laplacian MaxEnt upper bound may potentially be tighter than the traditionally used Gaussian MaxEnt upper bound for GMMs. This collection of MaxEnt bounds prove useful in practice since the differential entropy of mixtures does not admit a closed-form formula [5].

A source code for reproducible research with test experiments is available [23].

Appendix A
Differential Entropy of a Location-Scale Distribution

Let $p(x;\mu,\sigma) = \frac{1}{\sigma}p_0(\frac{x-\mu}{\sigma})$ denote the density of a *location-scale distribution* on the full support $\mathbb{R}$, where $\mu \in \mathbb{R}$ denotes the *location parameter* and $\sigma > 0$ the *dispersion parameter*. For example, a normal distribution has location parameter as its mean and dispersion parameter as its standard deviation. Let us prove that the differential entropy $h(X)$ is $h(X_0) + \log\sigma$ with $X \sim p(x;\mu,\sigma)$ and $X_0 \sim p_0(x)$, a quantity always independent of the location parameter μ. We shall make use of a change of variable $y = \frac{x-\mu}{\sigma}$ (with $\mathrm{d}y = \frac{\mathrm{d}y}{\sigma}$) in the integral to obtain

$$\begin{aligned} h(X) &= \int_{x=-\infty}^{+\infty} -\frac{1}{\sigma}p_0\left(\frac{x-\mu}{\sigma}\right)\left(\log\frac{1}{\sigma}p_0\left(\frac{x-\mu}{\sigma}\right)\right)\mathrm{d}x, \\ &= \int_{y=-\infty}^{+\infty} -p_0(y)(\log p_0(y) - \log\sigma), &(16) \\ &= h(X_0) + \log\sigma. &(17) \end{aligned}$$

Appendix B
Raw Absolute Geometric Moments of a Noncentered Normal Distribution

The calculations of $E[|X|^l]$ for $l \in \mathbb{N}$ and $X \sim N(\mu,\sigma)$ are reported in details in [22]. Let $n!!$ denote the double factorial: $n!! = \prod_{k=0}^{\lceil\frac{n}{2}\rceil-1}(n-2k) = \sqrt{\frac{2^{n+1}}{\pi}}\Gamma(\frac{n}{2}+1)$ (with $0!! = 1$ and $(-1)!! = 1$). Then, we have for even integer l:

$$E[|X|^l] = E[X^l] = \sum_{i=0}^{\lfloor\frac{l}{2}\rfloor}\binom{l}{2i}\mu^{l-2i}\sigma^{2i}(2i-1)!! \tag{18}$$

For odd integer l, we have

$$E[|X|^l] = \sum_{i=0}^{l}\binom{l}{i}\mu^{l-i}\sigma^i\left(I_i\left(-\frac{\mu}{\sigma}\right) - (-1)^i I_i\left(\frac{\mu}{\sigma}\right)\right) \tag{19}$$

where $I_i(a)$ is defined using the following recursive formula:

$$I_i(a) = \frac{1}{\sqrt{2\pi}}\left(a^{i-1}\exp\left(-\frac{1}{2}a^2\right)\right) + (i-1)I_{i-2}(a) \tag{20}$$

with the terminal recursion cases $I_0(a) = 1 - \Phi(a)$ (where $\Phi(x) = \int_{-\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x$ denotes the cumulative distribution function) and $I_1(a) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}a^2)$). Equivalent expressions may be obtained using the error function or the complementary error function by using the following identities: $\Phi(x) = \frac{1}{2}(1+\mathrm{erf}(\frac{x}{\sqrt{2}})) = \frac{1}{2}\mathrm{erfc}(-\frac{x}{\sqrt{2}})$.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[2] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, *Handbook of Differential Entropy*. London, U.K.: Chapman & Hall/CRC, 2013.

[3] F. Nielsen and R. Nock, "Entropies and cross-entropies of exponential families," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 3621–3624.

[4] F. Nielsen and R. Nock, "A closed-form expression for the Sharma-Mittal entropy of exponential families," *J. Phys. A, Math. Theor.*, vol. 45, no. 3, 2012, Art. no. 032003.

[5] S. Watanabe, K. Yamazaki, and M. Aoyagi, "Kullback information of normal mixture is not an analytic function," (in Japanese), IEICE, Tokyo, Japan, *Tech. Rep. NC2004–50*, 2004, pp. 41–46.

[6] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, "Calculation of differential entropy for a mixed Gaussian distribution," *Entropy*, vol. 10, no. 3, pp. 200–206, 2008.

[7] F. Nielsen and K. Sun, "Guaranteed bounds on the Kullback-Leibler divergence of univariate mixtures," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1543–1546, Nov. 2016.

[8] F. Nielsen and K. Sun, "Guaranteed bounds on information-theoretic measures of univariate mixtures using piecewise log-sum-exp inequalities," *Entropy*, vol. 18, no. 12, 2016, Art. no. 442. [Online]. Available: http://www.mdpi.com/1099–4300/18/12/442

[9] K. Moshksar and A. K. Khandani, "Arbitrarily tight bounds on differential entropy of Gaussian mixtures," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3340–3354, Jun. 2016.

[10] M. Shaked, "On mixtures from exponential families," *J. Roy. Stat. Soc. B*, vol. 42, pp. 192–198, 1980.

[11] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, 1957, Art. no. 620.

[12] E. T. Jaynes, "Information theory and statistical mechanics II," *Phys. Rev.*, vol. 108, no. 2, 1957, Art. no. 171.

[13] A. Mohammad-Djafari, "A Matlab program to calculate the maximum entropy distributions," in *Maximum Entropy and Bayesian Methods*. New York, NY, USA: Springer, 1992, pp. 221–233.

[14] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis (Adaptive and Learning Systems for Signal Processing, Communications, and Control)*. New York NY, USA: Wiley, 2001.

[15] L. Brown, *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory (IMS Lecture Notes)*. Hayward, CA, USA: Inst. Math. Statist., 1986. [Online]. Available: https://books.google.com/books?id=hmMLShr43PUC

[16] F. Nielsen and V. Garcia, "Statistical exponential families: A digest with flash cards," 2009. [Online]. Available: https://arxiv.org/abs/0911.4863

[17] R. H. Risch, "The solution of the problem of integration in finite terms," *Bull. Amer. Math. Soc.*, vol. 76, no. 3, pp. 605–608, 1970.

[18] M. Bronstein, *Symbolic Integration I: Transcendental Functions*, vol. 1. New York, NY, USA: Springer, 2013.

[19] J. H. Davenport, "Complexity of integration, special values, and recent developments," in *Proc. 5th Int. Conf.*, 2016, pp. 485–491.

[20] W. H. Press, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[21] L. Trailovic and L. Y. Pao, "Variance estimation and ranking of Gaussian mixture distributions in target tracking applications," in *Proc. 41st IEEE Conf. Decision Control*, 2002, vol. 2, pp. 2195–2201.

[22] F. Nielsen and R. Nock, "A series of maximum entropy upper bounds of the differential entropy," 2016. [Online]. Available: https://arxiv.org/abs/1612.02954

[23] F. Nielsen and R. Nock, "Java$^{\mathrm{TM}}$ source code for maxent upper bounds on the differential entropy of Gaussian mixture models." 2017. [Online]. Available: https://www.lix.polytechnique.fr/nielsen/MEUB/