Measures of statistical dispersion based on Shannon and Fisher information concepts

Lubomír Kostal,* Petr Lansky, and Ondřej Pokora
Institute of Physiology of the Czech Academy of Sciences, Videnska 1083, 14220 Prague 4, Czech Republic

We propose and discuss two information-based measures of statistical dispersion of positive continuous random variables: the entropy-based dispersion and Fisher information-based dispersion. Although standard deviation is the most frequently employed dispersion measure, we show, that it is not well suited to quantify some aspects that are often expected intuitively, such as the degree of randomness. The proposed dispersion measures are not entirely independent, though each describes the quality of probability distribution from a different point of view. We discuss relationships between the measures, describe their extremal values and illustrate their properties on the Pareto, the lognormal and the lognormal mixture distributions. Application possibilities are also mentioned.

Keywords: Statistical dispersion, Entropy, Fisher information, Positive random variable

1. INTRODUCTION

In recent years, information-based measures of randomness (or “regularity”) of signals have gained popularity in various branches of science [1–4]. In this paper we construct measures of statistical dispersion based on Shannon and Fisher information concepts and we describe their properties and mutual relationships. The effort was initiated in [5], where the entropy-based dispersion was employed to quantify certain aspects of neuronal timing precision. Here we extend the previous effort by taking into account the concept of Fisher information (FI), which was employed in different contexts [2, 6–9]. In particular, FI about the location parameter has been employed in the analysis of EEG [8, 10], of the atomic shell structure [11], (together with Shannon entropy) or in the description of variations among the two-electron correlated wavefunctions [12].

The goal of this paper is to propose different dispersion measures and to justify their usefulness. Although the standard deviation is used ubiquitously for characterization of variability, it is not well suited to quantify certain “intuitively intelligible” properties of the underlying probability distribution. For example highly variable data might not be random at all if it only consists of “very small” and “very large” values. Although the probability density function (or histogram of data) provides a complete view, one needs quantitative methods in order to make a comparison between different experimental scenarios.

The methodology investigated here does not adhere to any specific field of applications. We believe, that the general results are of interest to a wide group of researchers who deal with positive continuous random variables, in theory or in experiments.

2. MEASURES OF DISPERSION

2.1. Generic case: standard deviation

We consider a continuous positive random variable (r.v.) T with a probability density function (p.d.f.) f(t) and finite first two moments. Generally, statistical dispersion is a measure of “variability” or “spread” of the distribution of r.v. T, and such a measure has the same physical units as T. There are different dispersion measures described in the literature and employed in different contexts, e.g., standard deviation, interquartile range, mean difference or the Lγ coefficient [13–16].

By far, the most common measure of dispersion is the standard deviation, σ, defined as

\[ \sigma = \sqrt{\mathbb{E}((T - \mathbb{E}(T))^2)}. \] (1)

The corresponding relative dispersion measure is obtained by dividing σ with \( \mathbb{E}(T) \). The resulting quantity is denoted as the coefficient of variation, \( C_V \),

\[ C_V = \frac{\sigma}{\mathbb{E}(T)}. \] (2)

The main advantage of \( C_V \) over \( \sigma \) is, that \( C_V \) is dimensionless and thus probability distributions with different means can be compared meaningfully.

From Eq. (1) follows, that \( \sigma \) (or \( C_V \)) essentially measures how off-centered (with respect to \( \mathbb{E}(T) \)) is the distribution of \( T \). Furthermore, since the difference \((T - \mathbb{E}(T))\) is squared in Eq. (1), it follows that \( \sigma \) is sensitive to outlying values [15]. On the other hand, \( \sigma \) does not quantify how random, or unpredictable, are the outcomes of r.v. \( T \). Namely, high value of \( \sigma \) (high variability) does not indicate that the possible values of \( T \) are distributed evenly [5].

2.2. Dispersion measure based on Shannon entropy

For continuous r.v.’s the association between entropy and randomness is less straightforward than for discrete r.v.’s. The
(differential) entropy $h(T)$ of r.v. $T$ is defined as [17]

$$h(T) = -\int_0^\infty f(t) \ln f(t) \, dt.$$  \hspace{1cm} (3)

The value of $h(T)$ may be positive or negative, therefore $h(T)$ is not directly usable as a measure of statistical dispersion [5]. In order to obtain a properly behaving quantity, the entropy-based dispersion, $\sigma_h$, is defined as

$$\sigma_h = \exp[h(T)].$$  \hspace{1cm} (4)

The interpretation of $\sigma_h$ relies on the asymptotic equipartition property theorem [17]. Informally, the theorem states that almost any sequence of realizations of the random variable $T$ comes from a rather small subset (the typical set) in the $n$-dimensional space of all possible values. The volume of this subset is approximately $\sigma_h^n = \exp[nh(T)]$, and the volume is bigger for those random variables, which generate more diverse (or unpredictable) realizations. Further connection between $\sigma_h$ and $\sigma$ follows from the analogue to the entropy power concept [17]: $\sigma_h/e$ is equal to the standard deviation of an exponential distribution with entropy equal to $h(T)$.

Analogously to Eq. (2), we define the relative entropy-based dispersion coefficient, $C_h$, as

$$C_h = \frac{\sigma_h}{\Xi(T)}.$$  \hspace{1cm} (5)

The values of $\sigma_h$ and $C_h$ quantify how “evenly” is the probability distributed over the entire support. From this point of view, $\sigma_h$ is more appropriate than $\sigma$ when discussing randomness of data generated by r.v. $T$.

### 2.3. Dispersion measure based on Fisher information

The FI plays a key role in the statistical estimation of continuously varying parameters [18]. Let $X \sim p(x; \theta)$ be a family of r.v.'s defined for all values of parameter $\theta \in \Theta$, where $\Theta$ is an open subset of the real line. Let $\hat{\theta}(X)$ be an unbiased estimator of parameter $\theta$, i.e., $\mathbb{E}\{\hat{\theta}(X) - \theta\} = 0$. If for all $\theta \in \Theta$ and both $\varphi(x) \equiv 1$ and $\varphi(x) \equiv \varphi(x)$ the following equation is satisfied ([19, p.169] or [18, p.31]),

$$\frac{\partial}{\partial \theta} \int_X \varphi(x)p(x; \theta) \, dx = \int_X \varphi(x) \frac{\partial p(x; \theta)}{\partial \theta} \, dx,$$  \hspace{1cm} (6)

then the variance of the estimator $\hat{\theta}(X)$ satisfies the Cramer-Rao bound,

$$\text{Var}(\hat{\theta}(X)) \geq \frac{1}{J(\theta|X)},$$  \hspace{1cm} (7)

where

$$J(\theta|X) = \int_X \left[ \frac{\partial \ln p(x; \theta)}{\partial \theta} \right]^2 p(x; \theta) \, dx,$$  \hspace{1cm} (8)

is the FI about parameter $\theta$ contained in a single observation of r.v. $X$.

Exact conditions (the regularity conditions) under which Eq. (7) holds are stated slightly differently by different authors. In particular, it is sometimes required that the set $\{x : p(x; \theta) > 0\}$ does not depend on $\theta$, which is an unnecessarily strict assumption [18]. For any given p.d.f. $f(t)$ one may conveniently “generate” a simple parametric family by introducing a location parameter. The appropriate regularity conditions for this case are stated below.

The family of location parameter densities $p(x; \theta)$ satisfies

$$p(x; \theta) = p_0(x - \theta),$$  \hspace{1cm} (9)

where we consider $\Theta$ to be the whole real line and $p_0(x)$ is the p.d.f. of the “generating” r.v. $X_0$. Let the location family $p(x; \theta)$ be generated by the r.v. $T \sim f(t)$, thus $p(x; \theta) = f(x - \theta)$ and Eq. (8) can be written as

$$J(\theta|X) = \int_0^\infty \left[ \frac{\partial \ln f(x - \theta)}{\partial \theta} \right]^2 f(x - \theta) \, dx = \int_0^\infty \left[ \frac{\partial \ln f(t)}{\partial t} \right]^2 f(t) \, dt \equiv J(T),$$  \hspace{1cm} (10)

where the last equality follows from the fact that the derivatives of $f(x - \theta)$ with respect to $\theta$ or $x$ are equal up to a sign and due to the location-invariance of the integral (thus justifying the notation as $J(T)$). Since the value of $J(T)$ depends only on the “shape” of the p.d.f. $f(t)$, it is sometimes denoted as the FI about the random variable $T$ [17].

To interpret $J(T)$ according to the Cramer-Rao bound in Eq. (7), the required regularity conditions on $f(t)$ are: $f(t)$ must be continuously differentiable for all $t > 0$ and $f(0) = f'(0) = 0$. The integral (10) may exist and be finite even if $f(t)$ does not satisfy these conditions, e.g., if $f(t)$ is differentiable almost everywhere or $f(0) \neq 0$. However, in such a case the value of $J(T)$ does not provide any information about the efficiency of the location parameter estimation [18].

The units of $J(T)$ correspond to the inverse of the squared units of $T$, therefore we propose the FI based dispersion measure, $\sigma_f$, as

$$\sigma_f = \frac{1}{\sqrt{J(T)}},$$  \hspace{1cm} (11)

Heuristically, $\sigma_f$ quantifies the change in the p.d.f. $f(t)$ subject to an infinitesimally small shift $\delta \theta$ in $t$, i.e, it quantifies the difference between $f(t)$ and $f(t - \delta \theta)$. Any peak, or generally “non-smoothness” in the shape of $f(t)$ decreases $\sigma_f$. Analogously to Eqs. (2) and (5) we define the relative dispersion coefficient $C_f$ as

$$C_f = \frac{\sigma_f}{\Xi(T)}.$$  \hspace{1cm} (12)

In this paper we do not introduce different symbols for $C_f$ in dependence on whether the Cramer-Rao bound holds or not. We evaluate $C_f$ whenever the integral in Eq. (10) exists and we comment on the regularity conditions in the text.
3. RESULTS

3.1. Extrema of variability

Generally, the value \( C_V \) can be any non-negative real number, \( 0 \leq C_V < \infty \). The lower bound, \( C_V \to 0 \), is approached by a p.d.f. highly peaked at the mean value, in the limit corresponding to the Dirac’s delta function, \( f(t) = \delta(t - \mathbb{E}(T)) \). There is, however, no unique upper bound distribution for which \( C_V \to \infty \). For example, the p.d.f. examples analyzed in the next section allow arbitrarily high values of \( C_V \) and yet their shapes are different.

3.2. Extrema of entropy and its relation to variability

The relation between \( C_V \) and entropy was investigated in a series of papers [5, 20, 21]. The results can be re-stated in terms of \( C_h \) as follows. From the definition of \( C_h \) by Eq. (5) and from the properties of entropy [17] follows, that \( 0 < C_h < e \). The lower bound, \( C_h \to 0 \), is not realized by any unique distribution. Highly-peaked (possibly multimodal) densities approach the bound and in the limit any discrete-valued distribution achieves it. From this fact follows, that the relationship between \( C_V \) and \( C_h \) is not unique, small \( C_V \) implies small \( C_h \) but not vice versa.

The maximum value of \( C_h \) is connected with the problem of maximum entropy (ME), which is well known in the literature, see e.g., [17, 22]. The goal is to find such a p.d.f., that maximizes the functional (3) subject to \( n \) constraints of the form \( \mathbb{E}(\alpha_i(T)) = \xi_i \), where \( \alpha_i(t) \) and \( \xi_i \) are known and \( i = 1, \ldots, n \). The ME p.d.f. satisfying these constraints can be written in the form [17]

\[
  f(t) = \frac{1}{Z(\lambda_1, \ldots, \lambda_n)} \exp \left[ \sum_{i=1}^{n} \lambda_i \alpha_i(t) \right],
\]

where the "partition function" \( Z(\lambda_1, \ldots, \lambda_n) \) is the normalization factor, \( Z(\lambda_1, \ldots, \lambda_n) = \int_0^\infty \exp \left[ \sum_{i=1}^{n} \lambda_i \alpha_i(t) \right] dt \). The introduced Lagrange multipliers, \( \lambda_i \), are related to the averages \( \xi_i \) as [22]

\[
  -\frac{\partial}{\partial \lambda_i} \ln Z(\lambda_1, \ldots, \lambda_n) = \xi_i.
\]

It is well known [17], that the distribution maximizing the entropy on \([0, \infty)\) for given \( \mathbb{E}(T) \) is the exponential distribution,

\[
  f(t) = \frac{1}{\mathbb{E}(T)} \exp \left[ -\frac{t}{\mathbb{E}(T)} \right],
\]

and entropy \( h(T) = 1 + \ln \mathbb{E}(T) \). Thus the upper bound, \( C_h = e \), is unique: it is achieved only if \( f(t) \) is exponential. For the exponential distribution holds \( C_V = 1 \), however, non-exponential distributions may have \( C_V = 1 \) too (see the next section). In other words, the maximum of \( C_h \) does not correspond to any exclusive value of \( C_V \). This fact highlights the main difference between these two measures: the variability (described by \( C_V \)) and randomness (described by \( C_h \)) are not interchangeable notions. High variability (overdispersion), \( C_V > 1 \), results in decreased randomness for many common distributions, see Fig. 2, although there are exceptions, e.g., the Pareto distribution discussed later.

In order to find the ME distribution on \([0, \infty)\) given both \( \mathbb{E}(T) \) and \( C_V \), we first realize that the problem is equivalent to finding the ME distribution given \( \mathbb{E}(T) \) and \( \mathbb{V}(T^2) \). Applying the Lagrange formalism results in a p.d.f. based on the Gaussian, with the probability of all negative values aliased onto the positive half-line,

\[
  f(t) = \frac{1}{Z} \exp \left[ -\frac{(t - \alpha)^2}{2\beta^2} \right],
\]

where

\[
  Z = \beta \sqrt{\frac{\pi}{2}} \left[ 1 + \text{erf} \left( \frac{\alpha}{\sqrt{2}\beta} \right) \right].
\]

The density in Eq. (16) is also known as the density of the folded normal r.v. [23]. The parameters \( \alpha, \beta > 0 \), and \( \mathbb{E}(T) \), \( C_V \) are related as

\[
  \mathbb{E}(T) = \beta + \frac{\beta^2}{Z} \exp \left( -\frac{\alpha^2}{2\beta^2} \right),
\]

\[
  C_V = \beta \sqrt{\frac{\alpha^2}{2\beta^2} - \frac{\alpha}{Z} \exp \left( \frac{\alpha^2}{2\beta^2} - \frac{\beta^2}{Z^2} \times \left[ 1 - \alpha \exp \left( \frac{\alpha^2}{2\beta^2} \right) + \frac{\beta^2}{Z^2} \right] \right)}.
\]

The entropy and FI can be calculated for Eq. (16) to be

\[
  h(T) = \frac{1}{2} - \frac{\alpha}{2Z} \exp \left( -\frac{\alpha^2}{2\beta^2} \right) + \ln Z,
\]

\[
  J(T) = \frac{1}{\beta^2} \left[ 1 - \frac{\alpha}{Z} \exp \left( -\frac{\alpha^2}{2\beta^2} \right) \right].
\]

Note, that \( C_V \) in Eq. (19) is limited to \( C_V \in (0,1) \), and therefore the p.d.f. in Eq. (16) provides a solution to the ME problem only in this range. The density of the ME distribution given by Eq. (16) is shown for different values of \( C_V \) in Fig. 1. Although it is not possible to express \( \alpha, \beta \) in terms of \( \mathbb{E}(T), C_V \) from Eqns. (18) and (19), we obtain all distinct shapes (neglecting the scale) of the folded normal density by fixing, e.g., \( \beta = 1 \) and varying \( \alpha \in (-\infty, \infty) \), since \( \lim_{\alpha \to -\infty} C_V(\alpha) = 1 \), \( \lim_{\alpha \to \infty} C_V(\alpha) = 0 \) and noting that \( C_V(\alpha) \) is monotonically decreasing. In the limit \( C_V = 1 \) the density in Eq. (16) becomes exponential, and for \( C_V > 1 \) there is no unique ME distribution. However, we can always construct a p.d.f. with \( C_V > 1 \), which is arbitrarily close to the exponential p.d.f., e.g., almost-exponential with a small peak located at some large value of \( t \). Therefore, the maximum value of entropy is \( 1 + \ln \mathbb{E}(T) - \varepsilon \) for \( C_V > 1 \), where \( \varepsilon > 0 \) can be arbitrarily small. The corresponding \( C_h \) is shown in Fig. 2.
3.3. Extrema of Fisher information and its relation to entropy

From Eqns. (10) and (12) follows $C_J > 0$. Similarly to $C_h$, the lower bound is not achieved by a unique distribution, since any continuous, highly peaked density (possibly multimodal) approaches it. Determination of the maximum value of $C_J$ is, however, more difficult. In the following we solve the problem of $C_J$ maximization (FI minimization) subject to $\xi = \mathbb{E}(T)$, both when the regularity conditions hold and when they do not, see Fig. 1.

It is convenient [1, 2] to rewrite the FI functional by employing the real probability amplitude $u(t) = \sqrt{f(t)}$, so that Eq. (10) becomes

$$J(T) = 4 \int_T u'(t)^2 \, dt,$$

where $u'(t) = du(t)/dt$. The extrema of FI satisfies the Euler-Lagrange equation

$$\frac{\partial L}{\partial u} - \frac{d}{dt} \frac{\partial L}{\partial u'} = 0,$$

where the Lagrangian $L$ is

$$L = \int_T u'(t)^2 \, dt + \lambda_1 \left[ \int_T u(t)^2 \, dt - 1 \right] + \lambda_2 \left[ \int_T tu(t)^2 \, dt - \xi \right],$$

and the multiplicative constants resulting from the substitution $f \to u$ are contained in Lagrange multipliers $\lambda_1, \lambda_2$. Substituting from Eq. (24) into Eq. (23) results in the differential equation

$$u''(t) - u(t)[\lambda_1 - \lambda_2 t] = 0.$$

The solution to this equation can be written as [24]

$$u(t) = C_1 Ai \left( \frac{\lambda_1 + \lambda_2 t}{\lambda_2^{2/3}} \right) + C_2 Bi \left( \frac{\lambda_1 + \lambda_2 t}{\lambda_2^{2/3}} \right),$$

where $C_1, C_2$ are constants and $Ai(\cdot), Bi(\cdot)$ are the Airy functions. Since the integrability of the solution is required, it must be $C_2 = 0$. The remaining parameters $\lambda_1, \lambda_2, C_1$ are determined by requiring that $\int_0^\infty f(t) \, dt = 1$, that the mean equals $\xi$ and from the regularity conditions ($f(0) = f'(0) = 0$). Due to the presence of the Airy function, these parameters must be determined by numerical means. The resulting p.d.f. can be written as

$$f(t) = \frac{1}{Z_1} \exp \left( \frac{b_1 t}{\mathbb{E}(T)} \right),$$

where $Z_1$ is the normalizing constant, and $a_1 \equiv -2.3381, b_1 \equiv 1.5587$. The expression for FI of this p.d.f. can be obtained by integrating Eq. (22) and by combining Eq. (25) with the constraint values,

$$J(T) = -4 \left( \frac{b_1^2 + a_1 b_1^2}{\mathbb{E}(T)^2} \right) = \frac{7.5744}{\mathbb{E}(T)^2}.$$
4. APPLICATIONS

4.1. Lognormal and Pareto distributions

Both lognormal and Pareto distributions appear in a broad range of scientific applications [25]. The lognormal distribution is found in the description of, e.g., concentration of elements in the Earth’s crust, distribution of organisms in environment or in human medicine, see [26] for a review. The Pareto distribution is often described as an alternative model in situations similar as in the lognormal case, e.g., the sizes of human settlements, sizes of particle or allocation of wealth among individuals [27, 28]. Another common aspect of lognormal and Pareto distributions is, that both can be derived from exponential transforms of common distributions: normal and exponential.

The lognormal p.d.f., parametrized by the mean value and coefficient of variation, is

$$f_{\ln}(t) = \frac{1}{t\sqrt{2\pi\ln(1+C_V^2)}} \times \exp\left(-\frac{1}{8} \left[\ln(1+C_V^2) + 2\ln(t/\mathbb{E}(T))\right]^2\right) \frac{\ln(1+C_V^2)}{\ln(1+C_V^2)}.$$  \hspace{1cm} (35)

The coefficients $C_h$ and $C_f$ of the lognormal distribution can be calculated to be,

$$C_h = \sqrt{2\pi e} \sqrt{\frac{\ln(1+C_V^2)}{1+C_V^2}},$$ \hspace{1cm} (36)

$$C_f = \left[\ln(1+C_V^2)\right]/\left[1+C_V^2\right]^{1/2}.$$ \hspace{1cm} (37)

The dependencies of $C_h$ and $C_f$ on $C_V$ are shown in Fig. 2a, b. We see, that both $C_h$ and $C_f$ as functions of $C_V$ show a “∩” shape with maximum for $C_V = \sqrt{e - 1} \approx 1.31$ (for $C_h$) and around $C_V \approx 0.55$ (for $C_f$), confirming that each of the proposed dispersion coefficients provides a different point of view. The max $C_h$ p.d.f., Eq. (16), exists only for $C_V \leq 1$, for $C_V > 1$ the upper bound $C_h = 1$ is shown in Fig. 2a. Note, that the max $C_h$ distribution does generally not satisfy the regularity conditions, since $f(0) \neq 0$.

The dependence of $C_f$ on $C_h$ is shown in Fig. 2c. We observe, that $C_h$ and $C_f$ indeed do not describe the same qualities of the distribution, since for the lognormal distribution a single $C_h$ value does not correspond to a single $C_f$ value (and vice versa). In the lognormal case, the dependence between $C_h$ and $C_f$ forms a closed loop, where $C_h = C_f = 0$ for both $C_V \to 0$ and $C_V \to \infty$. In other words, both $C_h$ and $C_f$ fail to
distinguish between very different p.d.f. shapes \((C_V \to 0 \text{ or } C_V \to \infty)\).

The p.d.f. \(f_p(t)\) of the Pareto distribution is

\[
f_p(t) = \begin{cases} 
0, & t \in (0, b) \\
abla a^t^{-a-1}, & t \in [b, \infty) 
\end{cases}
\]

with parameters \(a > 0\) and \(b > 0\) (the expression in terms of \(E(T), C_V\) is cumbersome). Note, that \(E(T)\) exists only if \(a > 1\) and \(\text{Var}(T)\) only if \(a > 2\), thus we restrict ourselves to the case \(a > 1\) if only \(C_h\) and \(C_J\) are to be evaluated, and additionally to \(a > 2\) if \(C_V\) is required. From Eq. (38) follows that \(f_p(t)\) is not differentiable at \(t = b\), thus \(J(T)\) cannot be interpreted in terms of the Cramer-Rao bound, although \(J(T)\) is finite for all \(a > 0\).

The parameters \(a\) and \(b\) are related to \(E(T)\) and \(C_V\) by

\[
a = 1 + \frac{\sqrt{1 + C_V^2}}{C_V},
\]

\[
b = E(T) \left[ 1 + C_V^2 - C_V \sqrt{1 + C_V^2} \right].
\]

The coefficients \(C_h\) and \(C_J\) of the Pareto distribution can be expressed in terms of parameter \(a\) as

\[
C_h = \frac{a - 1}{a} \exp \left( \frac{1}{a} \right),
\]

\[
C_J = \frac{a - 1}{\sqrt{a^3(1 + a)}}.
\]

Both \(C_h\) and \(C_J\) have non-zero limit as \(C_V \to \infty\), namely

\(C_h = \sqrt{3}/4 \approx 1.12\) and \(C_J = 1/(3\sqrt{2}) \approx 0.2357\). However, while \(C_h\) as a function of \(C_V\) is monotonously increasing, \(C_J\) attains its maximum value, \(\max C_J \approx 0.2361\) for \(C_V \approx 2.3591\) (Fig. 2). The monotoneous shape of \(C_h\) versus the non-monotonous shape of \(C_J\) in dependence on \(C_V\) is a significant qualitative difference in the behavior of \(C_h\) and \(C_J\), although the effect is numerically very small. The shape of the dependence between \(C_h\) and \(C_J\) forms a closed loop if both \(a > 2\) and \(2 \geq a > 1\) regions are added together, since \(C_h = C_J = 0\) occurs for both \(C_V \to 0\) and \(a \to 1\) (\(C_V\) does not exist).

**4.2. Example: lognormal mixture**

Finally, we analyze a more complex example, a mixture of two distributions of the same type. The mixture models are met in diverse situations, e.g., in modeling of populations composed of subpopulations, in neuronal coding of odorant mixtures \[29\] or in the description spiking activity of bursting neurons \[30, 31\].

Recently, Bhumbra and Dyball \[32\] have successfully employed a mixture of two lognormal distributions to describe the neuronal firing in supraoptic nucleus.

The p.d.f. of the lognormal mixture model is

\[
f_m(t) = p f_{ln}(t; \mu_1, C_{V_1}) + (1 - p) f_{ln}(t; \mu_2, C_{V_2}),
\]

where \(0 < p < 1\) gives the weight of mixture components, and \(f_{ln}(t; \mu, C_V)\) is the lognormal density parametrized by the mean \(\mu\) and \(C_V\) given by Eq. (35). The lognormal mixture does not allow to express \(C_h\) or \(C_J\) in a closed form. Numerical evaluation of the involved integrals is more convenient in terms of a logarithmically transformed r.v. \(X = \ln T\), since \(X\) is described by a mixture of two normals. Let the density of the r.v. \(X\) be denoted as \(g_m(x)\), then

\[
g_m(x) = p\phi(x, m_1, s_1) + (1 - p)\phi(x, m_2, s_2),
\]

where \(\phi(x, m, s) = \exp[-(x - m)^2/(2s^2)]/\sqrt{2\pi}s^2\) is the density of the normal distribution with mean \(m\) and variance \(s^2\).

The mean value, \(\mu = E(T)\), and \(C_V\) of the random variable \(T\) can be expressed as

\[
\mu = p\exp \left( m_1 + \frac{s_1^2}{2} \right) + (1 - p)\exp \left( m_2 + \frac{s_2^2}{2} \right),
\]

\[
C_V = \frac{1}{\mu} \left[ \frac{p \exp \left( 2m_1 + 2s_1^2 \right)}{\exp \left( 2m_2 + 2s_2^2 - \mu^2 \right)} \right]^{1/2}.
\]

Since it holds \(f_m(t) = g_m(\ln t)/t\) and \(dx = dt/t\), the entropy \(h(T)\), given by Eq. (3), can be expressed by employing \(g_m(x)\) as

\[
h(T) = h(X) + E(X)
\]

thus Eq. (10) can be written in terms of \(g_m(x)\) as

\[
J(T) = \int_{-\infty}^{\infty} \left[ \frac{1}{g_m(x)} \frac{d g_m(x)}{d x} \right]^2 e^{-2x} g_m(x) \, dx.
\]
Figure 2. Relationships between $C_V$, $C_h$ and $C_J$ for the lognormal, Pareto and $C_h$-maximizing distribution. The max $C_h$ density is unique for $C_V \leq 1$, for $C_V > 1$ only the upper bound can be given. The dependence of $C_h$ on $C_V$ for the lognormal has a global maximum, while for the Pareto distribution $C_h$ grows monotonously. For all distributions holds $C_h \to 0$ as $C_V \to 0$. For the lognormal distribution the dependence of $C_J$ on $C_V$ resembles a scaled version of the $C_h$-$C_V$ dependence. For the Pareto distribution the $C_J$-$C_V$ dependence shows a global maximum at $C_V \approx 2.36$, contrary to the monotonicity of the $C_h$-$C_V$ dependence. This confirms that “smoothness” and “evenness” of the distribution are different notions, although, e.g., $C_J \equiv 0$ for $C_V \to 0$ for all distributions. The Pareto distribution with parameter $1 < a \leq 2$ is added to the $C_h$-$C_J$ dependence plot, since for this case both $C_h$ and $C_J$ can be calculated but $C_V$ is undefined.

Figure 3. Top row: lognormal mixture with variable weight of the components, $p \in [0,1]$, in the direction of arrows ($m_1 = -1$, $m_2 = -0.5$, $s_1 = 0.2$ and $s_2 = 1$). Although $E(T)$ decreases with $p$, $C_V$ exhibits maximum at $p = 0.6$. While the relationship between $C_V$ and $C_J$ is non-unique, $C_h$ describes $C_V$ uniquely (although the reverse statement is not true). Bottom row: lognormal mixture with increasing separation between the mean values of the logarithmically transformed components, $m_2 \in [-1,2]$, in the direction of arrows ($p = 0.2$, $m_1 = -1$, $s_1 = 0.2$ and $s_2 = 0.5$). Although the mean value $E(T)$ and $C_V$ increase and $C_J$ decreases monotonically, the shape of $C_h$ is unimodal with maximum at $C_V \approx 0.69$. The example shows the specific sensitivity of $C_J$ to modes ($C_J$ decreases as the modes become more apparent), while $C_h$ is sensitive to the overall spread (at $C_V \approx 0.69$ the bimodal distribution is more evenly distributed than for any other value of $C_V$).
In the second example, shown in Fig. 3 (bottom row), we vary the parameter $m_2$. Both $\mathbb{E}(T)$ and $C_V$ increase monotonically with increasing $m_2$. While $C_J$ decreases monotonically, $C_h$ shows a unimodal behavior with maxima around $C_V \approx 0.69$. Thus, although the shape of $f_m(t)$ becomes increasingly bimodal with growing $m_2$ (i.e., $C_J$ decreases), at the same time the distribution becomes more spread (or equiprobable, thus $C_h$ increases) up to a point $C_V \approx 0.69$. From that point on, the bimodality becomes too strong and decreases the evenness (or equiprobability) of the distribution and both $C_h$ and $C_J$ decrease. The increasing tendency of the density to become multimodal (decreasing $C_J$) may result in more unpredictable outcomes of the random variable $T$ (increasing $C_h$). Both these examples show, that $C_h$ and $C_J$ describe different aspects of the p.d.f. shape.

5. DISCUSSION AND CONCLUSIONS

We propose and discuss two measures of statistical dispersion for continuous positive random variables: the entropy-based dispersion ($C_h$) and the Fisher information-based dispersion ($C_J$). Both $C_h$ and $C_J$ describe the overall spread of the distribution differently than the coefficient of variation. While $C_h$ is most sensitive to the concentration of the probability mass (the predictability of random variable outcomes), $C_J$ is sensitive to the modes of the p.d.f. or any non-smoothness in the p.d.f. shape in general. The difference between $C_h$ and $C_J$ is further demonstrated by the fact, that the distributions maximizing their values are not the same. On the other hand, we do not claim that $C_h$ (or $C_J$) is “more informative” than $C_V$ due to taking into account, e.g., higher moments of the distribution. For example, one can find different distributions with equal $C_V$’s but differing $C_h$’s, and vice-versa, distributions with equal $C_h$’s but differing $C_V$’s, see Fig 2a.

It is also important to emphasize what is the benefit of employing the proposed measures once the full distribution function (and therefore a complete description of the situation) is known. The answer is, that it is often required to compare (or “categorize”) individual distributions according some specific property, i.e., to assign a number to each function. The advantage of employing the newly proposed measures lies in the possibility to describe p.d.f. qualities from different points of view, that might be of interest in various applications, see e.g., [1, 2, 9]. The parametrical estimates of the proposed coefficients (for both simulated and experimental data from olfactory neurons) were treated in detail in [33]. However, it is natural to ask for the non-parametric versions, which are arguably more valuable in practice. Lansky and Dittelevsen [34] discussed the disadvantages of the “classical” $C_V$ estimator based on sample mean and deviation, proposing solutions especially for the problem of biasedness. Non-parametric reliable estimates of the entropy (and thus of $C_h$) are well known [35, 36]. Recently, Kostal and Pokora [37] employed the maximum penalized likelihood method of Good and Gaskins [38] to jointly estimate $C_h$ and $C_J$ from simulated data.

Acknowledgements. This work was supported by the Institute of Physiology RVO:67985823, the Centre for Neuroscience P304/12/G069 and by the Grant Agency of the Czech Republic projects P103/11/0282 and P103/12/P558.