



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Staudte, RG;Xia, A

Title:

Divergence from, and convergence to, uniformity of probability density quantiles

Date:

2018-05-03

Citation:

Staudte, R. G. & Xia, A. (2018). Divergence from, and convergence to, uniformity of probability density quantiles. *Entropy*, 20 (5), <https://doi.org/10.3390/e20050317>.

Persistent Link:


<https://hdl.handle.net/11343/270798>

License:

[CC BY](#)

Article

Divergence from, and Convergence to, Uniformity of Probability Density Quantiles

Robert G. Staudte ^{1,*}  and Aihua Xia ^{2,†}¹ Department of Mathematics and Statistics, La Trobe University, Bundoora, VIC 3086, Australia² School of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia; aihuaxia@unimelb.edu.au

* Correspondence: r.staudte@latrobe.edu.au; Tel.: +61-3-9479-2610

† These authors contributed equally to this work.

Received: 7 March 2018; Accepted: 19 April 2018; Published: 25 April 2018



Abstract: We demonstrate that questions of convergence and divergence regarding shapes of distributions can be carried out in a location- and scale-free environment. This environment is the class of probability density quantiles (pdQs), obtained by normalizing the composition of the density with the associated quantile function. It has earlier been shown that the pdQ is representative of a location-scale family and carries essential information regarding shape and tail behavior of the family. The class of pdQs are densities of continuous distributions with common domain, the unit interval, facilitating metric and semi-metric comparisons. The Kullback–Leibler divergences from uniformity of these pdQs are mapped to illustrate their relative positions with respect to uniformity. To gain more insight into the information that is conserved under the pdQ mapping, we repeatedly apply the pdQ mapping and find that further applications of it are quite generally entropy increasing so convergence to the uniform distribution is investigated. New fixed point theorems are established with elementary probabilistic arguments and illustrated by examples.

Keywords: convergence in L_r norm; fixed point theorem; Kullback–Leibler divergence; relative entropy; semi-metric; uniformity testing

MSC: Primary 62E10; Secondary 62F03

1. Introduction

For each continuous location-scale family of distributions with square-integrable density, there is a probability density quantile (pdQ), which is an absolutely continuous distribution on the unit interval. Members of the class of such pdQs differ only in *shape*, and the asymmetry of their shapes can be partially ordered by their Hellinger distances or Kullback–Leibler divergences from the class of symmetric distributions on this interval. In addition, the tail behaviour of the original family can be described in terms of the boundary derivatives of its pdQ. Empirical estimators of the pdQs enable one to carry out inference, such as robust fitting of shape parameter families to data; details are in [1].

The Kullback–Leibler directed divergences and symmetrized divergence (KLD) of a pdQ with respect to the uniform distribution on [0,1] is investigated in Section 2, with remarkably simple numerical results, and a map of these divergences for some standard location-scale families is constructed. The ‘shapeless’ uniform distribution is the center of the pdQ universe, as is explained in Section 3, where it is found to be a fixed point. A natural question of interest is to find the invariant information of the pdQ mapping, that is, the conserved information after the pdQ mapping is applied. To this end, it is necessary to repeatedly apply the pdQ mapping to extract the information. Numerical studies indicate that further applications of the pdQ transformation are generally entropy increasing,

so we investigate the convergence to uniformity of repeated applications of the pdQ transformation, by means of fixed point theorems for a semi-metric. As the pdQ mapping is not a contraction, the proofs of the fixed point theorems are through elementary probabilistic arguments rather than the classical contraction mapping principle. Our approach may shed light on future research in the fixed point theory. Further ideas are discussed in Section 4.

2. Divergences between Probability Density Quantiles

2.1. Definitions

Let \mathcal{F} denote the class of cumulative distribution functions (cdfs) on the real line \mathbb{R} and for each $F \in \mathcal{F}$ define the associated *quantile function* of F by $Q(u) = \inf\{x : F(x) \geq u\}$, for $0 < u < 1$. When the random variable X has cdf F , we write $X \sim F$. When the density function $f = F'$ exists, we also write $X \sim f$. We only discuss F as absolutely continuous with respect to Lebesgue measure, but the results can be extended to the discrete and mixture cases using suitable dominating measures.

Definition 1. Let $\mathcal{F}' = \{F \in \mathcal{F} : f = F' \text{ exists and is positive}\}$. For each $F \in \mathcal{F}'$, we follow [2] and define the *quantile density function* $q(u) = Q'(u) = 1/f(Q(u))$. Parzen called its reciprocal function $fQ(u) = f(Q(u))$ the *density quantile function*. For $F \in \mathcal{F}'$, and U uniformly distributed on $[0,1]$, assume $\kappa = \mathbb{E}[fQ(U)] = \int f^2(x) dx$ is finite; that is, f is square integrable. Then, we can define the continuous pdQ of F by $f^*(u) = fQ(u)/\kappa, 0 < u < 1$. Let $\mathcal{F}'^* \subset \mathcal{F}'$ denote the class of all such F .

Not all f are square-integrable, and this requirement for the mapping $f \rightarrow f^*$ means that \mathcal{F}'^* is a proper subset of \mathcal{F}' . The advantages of working with f^* s over f s are that they are free of location and scale parameters; they ignore flat spots in F and have a common bounded support. Moreover, f^* often has a simpler formula than f ; see Table 1 for examples.

Table 1. Quantiles of some distributions, their pdQs and divergences. In general, we denote $x_u = Q(u) = F^{-1}(u)$, but for the normal $F = \Phi$ with density ϕ , we use $z_u = \Phi^{-1}(u)$. The logistic quantile function is only defined for $u \leq 0.5$, but it is symmetric about $u = 0.5$. Lognormal(σ) represents the lognormal distribution with shape parameter σ . The quantile function for the Pareto is for the Type II distribution with shape parameter a , and the pdQ is the same for Type I and Type II Pareto models.

	$Q(u)$	$f^*(u)$	$I(\mathcal{U} : f^*)$	$I(f^* : \mathcal{U})$	$J(\mathcal{U}, f^*)$
Normal	z_u	$2\sqrt{\pi} \phi(z_u)$	0.153	0.097	0.250
Logistic	$\ln(u/(1-u))$	$6u(1-u)$	0.208	0.125	0.333
Laplace	$\ln(2u), u \leq 0.5$	$2 \min\{u, 1-u\}$	0.307	0.193	0.500
t_2	$\frac{2u-1}{\{2u(1-u)\}^{1/2}}$	$\frac{2^7\{u(1-u)\}^{3/2}}{3\pi}$	0.391	0.200	0.591
Cauchy	$\tan\{\pi(u-0.5)\}$	$2 \sin^2(\pi u)$	0.693	0.307	1.000
Exponential	$-\ln(1-u)$	$2(1-u)$	0.307	0.193	0.500
Gumbel	$-\ln(-\ln(u))$	$-4u \ln(u)$	0.191	0.116	0.307
Lognormal (σ)	$e^{\sigma z_u}$	$\frac{2\sqrt{\pi}}{e^{\sigma^2/4}} \phi(z_u) e^{-\sigma z_u}$	$\frac{\sigma^2}{4} + \frac{1}{2} - \ln(\sqrt{2})$	-	$\frac{1}{4} + \frac{3\sigma^2}{8}$
Pareto (a)	$(1-u)^{-1/a}$	$\frac{2a+1}{a} (1-u)^{1+1/a}$	$\frac{1+a}{a} - \ln(2 + \frac{1}{a})$	-	$\frac{(1+a)^2}{a(1+2a)}$
Power (b)	$u^{1/b}$	$\frac{2b-1}{b} u^{-1/b}$	$\frac{b-1}{b} - \ln(2 - \frac{1}{b})$	-	$\frac{(b-1)^2}{b(2b-1)}$

Remark 1. Given that a pdQ f^* exists for a distribution with density f , then so does the cdf F^* and quantile function $Q^* = (F^*)^{-1}$ associated with f^* . Thus, a monotone transformation from $X \sim F$ to $X^* \sim F^*$ exists; it is simply $X^* = m(X) = Q^*(F(X))$. For the Power(b) distribution of Table 1, $f_b^* = f_{b^*}$, where $b^* = 2 - 1/b$, so $m_b(x) = Q_b^*(F_b(x)) = x^{b/b^*} = x^{b^2/(2b-1)}$. For the normal distribution with parameters μ, σ , it is

$m_{\mu,\sigma}(x) = \Phi((x - \mu)/\sqrt{2}\sigma)$. In general, an explicit expression for Q^* that depends only on f or F (plus location-scale parameters) need not exist.

2.2. Divergence Map

Next, we evaluate and plot the [3] divergences from uniformity. The [3] divergence of density f_1 from density f_2 , when both have domain $[0,1]$, is defined as

$$I(f_1 : f_2) := \int_0^1 \ln(f_1(u)/f_2(u)) f_1(u) du = \mathbb{E}[\ln(f_1(U)/f_2(U)) f_1(U)],$$

where U denotes a random variable with the uniform distribution \mathcal{U} on $[0,1]$. The divergences from uniformity are easily computed through

$$I(\mathcal{U} : f^*) = - \int_0^1 \ln(f^*(u)) du = -\mathbb{E}[\ln(f^*(U))]$$

and

$$I(f^* : \mathcal{U}) = \int_0^1 \ln(f^*(u)) f^*(u) du = \mathbb{E}[\ln(f^*(U)) f^*(U)].$$

Kullback ([4], p. 6) interprets $I(f^* : \mathcal{U})$ as the mean evidence in one observation $V \sim f^*$ for f^* over \mathcal{U} ; it is also known as the *relative entropy* of f^* with respect to \mathcal{U} . The terminology directed divergence for $I(f_1 : f_2)$ is also sometimes used ([4], p. 7) with ‘directed’ explained in ([4], pp. 82, 85); see also [5] in this regard.

Table 1 shows the quantile functions of some standard distributions, along with their pdQs, associated divergences $I(\mathcal{U} : f^*), I(f^* : \mathcal{U})$ and symmetrized divergence (KLD) defined by $J(\mathcal{U}, f^*) := I(\mathcal{U} : f^*) + I(f^* : \mathcal{U})$. The last measure was earlier introduced in a different form by [6].

Definition 2. Given pdQs f_1^*, f_2^* , let $d(f_1^*, f_2^*) := \sqrt{I(f_1^* : f_2^*) + I(f_2^* : f_1^*)}$. Then, d is a semi-metric on the space of pdQs; i.e., d satisfies all requirements of a metric except the triangle inequality. Introducing the coordinates $(s_1, s_2) = (\sqrt{I(\mathcal{U} : f^*)}, \sqrt{I(f^* : \mathcal{U})})$, we can define the distance from uniformity of any f^* by the Euclidean distance of (s_1, s_2) from the origin $(0,0)$, namely $d(\mathcal{U}, f^*)$.

Remark 2. This d does not satisfy the triangle inequality: for example, if \mathcal{U}, \mathcal{N} and \mathcal{C} denote the uniform, normal and Cauchy pdQs, then $d(\mathcal{U}, \mathcal{N}) = 0.5$, $d(\mathcal{N}, \mathcal{C}) = 0.4681$ but $d(\mathcal{U}, \mathcal{C}) = 1$; see Table 1 and Figure 1. However, d can provide an informative measure of distance from uniformity.

Figure 1 shows the loci of points (s_1, s_2) for some continuous shape families. The light dotted arcs with radii 1/2, 1 and 2 are a guide to these distances from uniformity. The large discs in purple, red and black correspond to \mathcal{U}, \mathcal{N} and \mathcal{C} . The blue cross at distance $1/\sqrt{2}$ from the origin corresponds to the exponential distribution. Nearby is the standard lognormal point marked by a red cross. The lower red curve is nearly straight and is the locus of points corresponding to the lognormal shape family.

The chi-squared(ν), $\nu > 1$, family also appears as a red curve; it passes through the blue cross when $\nu = 2$, as expected, and heads toward the normal disc as $\nu \rightarrow \infty$. The Gamma family has the same locus of points as the chi-squared family. The curve for the Weibull(β) family, for $0.5 < \beta < 3$, is shown in blue; it crosses the exponential blue cross when $\beta = 1$. The Pareto(a) curve is shown in black. As a increases from 0, this line crosses the arcs distant 2 and 1 from the origin for $a = (2\sqrt{2} + 1)/7 \approx 0.547$ and $a = (\sqrt{5} - 1)/2 \approx 1.618$, respectively, and approaches the exponential blue cross as $a \rightarrow \infty$.

The Power(b) or Beta($b, 1$) for $b > 1/2$ family is represented by the magenta curve of points moving toward the origin as b increases from 1/2 to 1, and then moving out towards the exponential blue cross as $b \rightarrow \infty$. For each choice of $\alpha > 0.5$, $\beta > 0.5$ the locus of the Beta(α, β) pdQ divergences lies

above the chi-squared red curve and mostly below the power(*b*) magenta curve; however, the U-shaped Beta distributions have loci above it.

The lower green line near the Pareto black curve gives the loci of root-divergences from uniformity of the Tukey(λ) with $\lambda < 1$, while the upper green curve corresponds to $\lambda \geq 1$. It is known that the Tukey(λ) distributions, with $\lambda < 1/7$, are good approximations to Student’s *t*-distributions for $\nu > 0$, provided λ is chosen properly. The same is true for their corresponding pdQs ([1], Section 3.2). For example, the pdQ of t_ν with $\nu = 0.24$ degrees of freedom is well approximated by the choice $\lambda = -4.063$. Its location is marked by the small black disk in Figure 1; it is of distance 2 from uniformity. The generalized Tukey distributions of [7] with two shape parameters also fill a large funnel shaped region (not marked on the map) emanating from the origin and just including the region bounded by the green curves of the Tukey symmetric distributions.

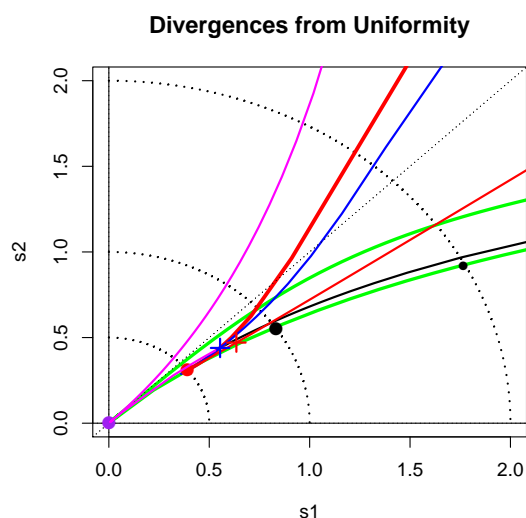


Figure 1. Divergence from uniformity. The loci of points $(s_1, s_2) = (\sqrt{I(U : f^*)}, \sqrt{I(f^* : U)})$ is shown for various standard families. The large disks correspond respectively to the symmetric families: uniform (purple), normal (red) and Cauchy (black). The crosses correspond to the asymmetric distributions: exponential (blue) and standard lognormal (red). More details are given in Section 2.2.

2.3. Uniformity Testing

There are numerous tests for uniformity, but as [8] points out, many are undermined by the common practice of estimating location-scale parameters of the null and/or alternative distributions when in fact it is assumed that these distributions are known exactly. In practice, this means that if a test for uniformity is preceded by a probability integral transformation including parameter estimates, then the actual levels of such tests will not be those nominated unless (often complicated and model-specific) adjustments are made. Examples of such adjustments are in [9,10].

Given a random sample of m independent, identically distributed (i.i.d.) variables, each from a distribution with density f , it is feasible to carry out a nonparametric test of uniformity by estimating the pdQ with a kernel density estimator \hat{f}_m^* and comparing it with the uniform density on $[0,1]$ using any one of a number of metrics or semi-metrics. Consistent estimators \hat{f}_m^* for f^* based on normalized reciprocals of the quantile density estimators derived in [11] are available and described in (Staudte [1], Section 2). Note that such a test compares an arbitrary uniform distribution with an arbitrary member of the location-scale family generated by f ; it is a test of shape only. Preliminary work suggests that such a test is feasible. However, an investigation into such omnibus nonparametric testing procedures, including comparison with bootstrap and other kernel density based techniques found in the literature, such as [12–17], is beyond the scope of this work.

3. Convergence of Density Shapes to Uniformity via Fixed Point Theorems

The transformation $f \rightarrow f^*$ of Definition 1 is quite powerful, removing location and scale and moving the distribution from the support of f to the unit interval. A natural question of interest is to find the information in a density that is invariant after the pdQ mapping is applied. To this end, it is necessary to repeatedly apply the pdQ mapping to extract the information. Examples suggest that another application of the transformation $f^{2*} := (f^*)^*$ leaves less information about f in f^{2*} and hence it is closer to the uniform density. Furthermore, with n iterations $f^{(n+1)*} := (f^{n*})^*$ for $n \geq 2$, it seems that no information can be conserved after repeated $*$ -transformation so we would expect that f^{n*} converges to the uniform density as $n \rightarrow \infty$. An R script [18] for finding repeated $*$ -iterates of a given pdQ is available as Supplementary Material.

3.1. Conditions for Convergence to Uniformity

Definition 3. Given $f \in \mathcal{F}'$, we say that f is of $*$ -order n if $f^*, f^{2*}, \dots, f^{n*}$ exist but $f^{(n+1)*}$ does not. When the infinite sequence $\{f^{n*}\}_{n \geq 1}$ exists, it is said to be of infinite $*$ -order.

For example, the Power(3/4) family is of $*$ -order 2, while the Power(2) family is of infinite $*$ -order. The χ^2_ν distribution is of finite $*$ -order for $1 < \nu < 2$ and infinite $*$ -order for $\nu \geq 2$. The normal distribution is of infinite $*$ -order.

We write $\mu_n := \int_{-\infty}^{\infty} \{f(y)\}^n dy$, $\kappa_n = \int_0^1 \{f^{n*}(x)\}^2 dx$, $n \geq 1$, and $\kappa_0 = \int_{-\infty}^{\infty} \{f(x)\}^2 dx$. The next proposition characterises the property of infinite $*$ -order.

Proposition 1. For $f \in \mathcal{F}'$ and $m \geq 1$, the following statements are equivalent:

- (i) $\mu_{m+2} < \infty$,
- (ii) $\mu_j < \infty$ for all $1 \leq j \leq m + 2$,
- (iii) $\kappa_j < \infty$ and $\kappa_j = \frac{\mu_j \mu_{j+2}}{\mu_{j+1}^2}$ for all $1 \leq j \leq m$.

In particular, f is of infinite $*$ -order if and only if $\mu_n < \infty$, $n \geq 1$.

Proof of Proposition 1. For each $i, n \geq 1$, provided all terms below are finite, we have the following recursive formula

$$v_{n,i} := \int \{f^{n*}(x)\}^i dx = \frac{1}{\kappa_{n-1}^i} v_{n-1,i+1}, \tag{1}$$

giving

$$\kappa_n = \frac{1}{\prod_{j=0}^{n-1} \kappa_j^{n+1-j}} \mu_{n+2}. \tag{2}$$

(i) \Rightarrow (ii) For $1 \leq j \leq m + 2$,

$$\begin{aligned} \mu_j &= \int_{-\infty}^{\infty} \{f(x)\}^j \mathbf{1}_{\{f(x) > 1\}} dx + \int_{-\infty}^{\infty} \{f(x)\}^j \mathbf{1}_{\{f(x) \leq 1\}} dx \\ &\leq \int_{-\infty}^{\infty} \{f(x)\}^{m+2} dx + \int_{-\infty}^{\infty} f(x) dx = \mu_{m+2} + 1 < \infty. \end{aligned}$$

(ii) \Rightarrow (iii) Use (2) and proceed with induction for $1 \leq n \leq m$.

(iii) \Rightarrow (i) By Definition 1, $\kappa_1 < \infty$ means that $\kappa_0 < \infty$. Hence, (i) follows from (2) with $n = m$. □

Next, we investigate the involutory nature of the $*$ -transformation.

Proposition 2. Let f^* be a pdQ and assume f^{2*} exists. Then, $f^* \sim \mathcal{U}$ if and only if $f^{2*} \sim \mathcal{U}$.

Proof of Proposition 2. For $r > 0$, we have

$$\int_0^1 |f^{2*}(u) - 1|^r du = \frac{1}{\kappa_1^r} \int_0^1 |f^*(x) - \kappa_1|^r f^*(x) dx. \tag{3}$$

If $f^*(u) \sim \mathcal{U}$, then $\kappa_1 = 1$ and (3) ensures $\int_0^1 |f^{2*}(u) - 1|^r du = 0$, so $f^{2*}(u) \sim \mathcal{U}$.

Conversely, if $f^{2*}(u) \sim \mathcal{U}$, then using (3) again gives $\int_0^1 |f^*(x) - \kappa_1|^r f^*(x) dx = 0$. Since $f^*(x) > 0$ a.s., we have $f^*(x) = \kappa_1$ a.s. and this can only happen when $\kappa_1 = 1$. Thus, $f^* \sim \mathcal{U}$, as required. \square

Proposition 2 shows that the uniform distribution is a fixed point in the Banach space of integrable functions on $[0,1]$ with the L_r norm for any $r > 0$. It remains to show that f^{n*} has a limit and that the limit is the uniform distribution. It was hoped that the classical machinery for convergence in Banach spaces ([19], Chapter 10) would prove useful in this regard, but the $*$ -mapping is not a contraction. For this reason, although there are many studies of fixed point theory in metric and semi-metric spaces (see, e.g., [20] and references therein), the fixed point Theorems 1, 2 and 3 shown below do not seem to be covered in these general studies. Moreover, our proofs are purely probabilistic and non-standard in this area. For simplicity, we use $\xrightarrow{L_r}$ to stand for the convergence in L_r norm and $\xrightarrow{\mathbb{P}}$ for convergence in probability as $n \rightarrow \infty$.

Theorem 1. For $f \in \mathcal{F}'$ with infinite $*$ -order, the following statements are equivalent:

- (i) $f^{n*} \xrightarrow{L_2} 1$;
- (ii) For all $r > 0$, $f^{n*} \xrightarrow{L_r} 1$;
- (iii) $\frac{\mu_n \mu_{n+2}}{\mu_{n+1}^2} \rightarrow 1$ as $n \rightarrow \infty$.

Remark 3. Notice that $\mu_n = \mathbb{E}\{f^*(U)^{n-1}\}$, $n \geq 1$, are the moments of the random variable $f^*(U)$ with $U \sim \mathcal{U}$. Theorem 1 says that the convergence of $\{f^{n*} : n \geq 1\}$ is purely determined by the moments of $f^*(U)$. This is rather puzzling because it is well known that the moments do not uniquely determine the distribution ([21], p. 227), meaning that different distributions with the same moments have the same converging behaviour. However, if f is bounded, then $f^*(U)$ is a bounded random variable so its moments uniquely specify its distribution ([21], pp. 225–226), leading to stronger results in Theorem 2.

Proof of Theorem 1. It is clear that (ii) implies (i).

(i) \Rightarrow (iii): By Proposition 1, $\kappa_n = \frac{\mu_n \mu_{n+2}}{\mu_{n+1}^2}$. Now,

$$\int_0^1 \{f^{n*}(x) - 1\}^2 dx = \kappa_n - 1, \tag{4}$$

so (iii) follows immediately.

(iii) \Rightarrow (ii): It suffices to show that $f^{n*} \xrightarrow{L_r} 1$ for any integer $r \geq 4$. To this end, since for $a, b \geq 0$, $|a - b|^{r-2} \leq a^{r-2} + b^{r-2}$, we have from (4) that

$$\int_0^1 |f^{n*}(x) - 1|^r dx \leq \int_0^1 (f^{n*}(x) - 1)^2 (f^{n*}(x)^{r-2} + 1) dx = v_{n,r} - 2v_{n,r-1} + v_{n,r-2} + \kappa_n - 1, \tag{5}$$

where, as before, $v_{n,r} = \int_0^1 \{f^{n*}(x)\}^r dx$. However, applying (1) gives

$$v_{n,r} = \frac{\mu_{n+r}}{\kappa_{n-1}^r \kappa_{n-2}^{r+1} \dots \kappa_0^{n+r-1}}$$

and (2) ensures

$$\mu_{n+r} = \kappa_{n+r-2} \kappa_{n+r-3}^2 \dots \kappa_0^{n+r-1},$$

which imply

$$v_{n,r} = \kappa_{n+r-2} \kappa_{n+r-3}^2 \dots \kappa_n^{r-1} \rightarrow 1$$

as $n \rightarrow \infty$. Hence, it follows from (5) that $\int_0^1 |f^{n*}(x) - 1|^r dx \rightarrow 0$ as $n \rightarrow \infty$, completing the proof. \square

We write $\|g\| = \sup_x |g(x)|$ for each bounded function g .

Theorem 2. *If f is bounded, then*

- (i) for all $n \geq 0$, $\|f^{(n+1)*}\| \leq \|f^{n*}\|$ and the inequality becomes equality if and only if $f^{n*} \sim \mathcal{U}$;
- (ii) $f^{n*} \xrightarrow{L_r} 1$ for all $r > 0$.

Proof of Theorem 2. It follows from (4) that $\kappa_n \geq 1$ and the inequality becomes equality if and only if $f^{n*} \sim \mathcal{U}$.

(i) Let Q^{n*} be the inverse of the cumulative distribution function of f^{n*} , then $f^{(n+1)*}(u) = \frac{f^{n*}(Q^{n*}(u))}{\kappa_n} \leq \frac{\|f^{n*}\|}{\kappa_n}$, giving $\|f^{(n+1)*}\| \leq \frac{\|f^{n*}\|}{\kappa_n} \leq \|f^{n*}\|$. If $f^{n*} \sim \mathcal{U}$, then Proposition 2 ensures that $f^{(n+1)*} \sim \mathcal{U}$, so $\|f^{(n+1)*}\| = \|f^{n*}\|$. Conversely, if $\|f^{(n+1)*}\| = \|f^{n*}\|$, then $\kappa_n = 1$, so $f^{n*} \sim \mathcal{U}$.

(ii) It remains to show that $\kappa_n \rightarrow 1$ as $n \rightarrow \infty$. In fact, if $\kappa_n \not\rightarrow 1$, since $\kappa_n \geq 1$, there exist a $\delta > 0$ and a subsequence $\{n_k\}$ such that $\kappa_{n_k} \geq 1 + \delta$, which implies

$$\frac{\mu_{n_k+2}}{\mu_{n_k+1}} = \prod_{i=0}^{n_k} \kappa_i \geq (1 + \delta)^k \rightarrow \infty \text{ as } k \rightarrow \infty. \tag{6}$$

However, $\frac{\mu_{n_k+2}}{\mu_{n_k+1}} \leq \|f\| < \infty$, which contradicts (6). \square

Theorem 3. *For $f \in \mathcal{F}'$ with infinite $*$ -order such that $\{\mu_n \mu_{n+2} \mu_{n+1}^{-2} : n \geq 1\}$ is a bounded sequence, then the following statements are equivalent:*

- (i*) $f^{n*} \xrightarrow{\mathbb{P}} 1$;
- (ii) For all $r > 0$, $f^{n*} \xrightarrow{L_r} 1$;
- (iii) $\mu_n \mu_{n+2} \mu_{n+1}^{-2} \rightarrow 1$ as $n \rightarrow \infty$.

Proof of Theorem 3. It suffices to show that (i*) implies (iii). Recall that $\kappa_n = \mu_n \mu_{n+2} \mu_{n+1}^{-2}$. For each subsequence $\{\kappa_{n_k}\}$, there exists a converging sub-subsequence $\{\kappa_{n_{k_i}}\}$ such that $\kappa_{n_{k_i}} \rightarrow b$ as $i \rightarrow \infty$. It remains to show that $b = 1$. To this end, for $\delta > 1$, we have

$$\begin{aligned} & \int_0^1 |f^{(n_{k_i}+1)*}(x) - 1| \mathbf{1}_{\{|f^{(n_{k_i}+1)*}(x) - 1| \leq \delta\}} dx \\ &= \frac{1}{\kappa_{n_{k_i}}} \int_0^1 |f^{(n_{k_i})*}(x) - \kappa_{n_{k_i}}| f^{(n_{k_i})*}(x) \mathbf{1}_{\{|f^{(n_{k_i})*}(x) - \kappa_{n_{k_i}}| \leq \delta \kappa_{n_{k_i}}\}} dx. \end{aligned} \tag{7}$$

(i*) ensures that

$$|f^{(n_{k_i}+1)*} - 1| \xrightarrow{\mathbb{P}} 0, \quad |f^{(n_{k_i})*} - \kappa_{n_{k_i}}| \xrightarrow{\mathbb{P}} |1 - b|, \quad \mathbf{1}_{\{|f^{(n_{k_i})*}(x) - \kappa_{n_{k_i}}| \leq \delta \kappa_{n_{k_i}}\}} \xrightarrow{\mathbb{P}} 1$$

as $i \rightarrow \infty$, so applying the bounded convergence theorem to both sides of (7) to get $0 = |1/b - 1|$, i.e., $b = 1$. \square

Remark 4. *We note that not all distributions are of infinite $*$ -order so the fixed point theorems are only applicable to a proper subclass of all distributions.*

3.2. Examples of Convergence to Uniformity

The main results in Section 3.1 cover all the standard distributions with infinite $*$ -order in [22,23]. In fact, as observed in the Remark after Theorem 1 that the convergence to uniformity is purely determined by the moments of $f^*(U)$ with $U \sim \mathcal{U}$, we have failed to construct a density such that $\{f^{n*} : n \geq 1\}$ does not converge to the uniform distribution. Here, we give a few examples to show that the main results in Section 3.1 are indeed very convenient to use.

Example 1. *Power function family.*

From Table 1, the Power(b) family has density $f_b(x) = bx^{b-1}$, $0 < x < 1$, so it is of infinite $*$ -order if and only if $b \geq 1$. As f_b is bounded for $b \geq 1$, Theorem 2 ensures that f_b^{n*} converges to the uniform in L_r for any $r > 0$.

Example 2. *Exponential distribution.*

Suppose $f(x) = e^{-x}$, $x > 0$. f is bounded, so Theorem 2 says that f^{n*} converges to the uniform distribution as $n \rightarrow \infty$. By symmetry, the same result holds for $f(x) = e^{-x}$, $x < 0$.

Example 3. *Pareto distribution.*

The Pareto(a) family, with $a > 0$, has $f_a(x) = ax^{-a-1}$ for $x > 1$, which is bounded, so an application of Theorem 2 yields that the sequence $\{f_a^{n*}\}_{n \geq 1}$ converges to the uniform distribution as $n \rightarrow \infty$.

Example 4. *Cauchy distribution.*

The pdQ of the Cauchy density is given by $f^*(u) = 2 \sin^2(\pi u)$, $0 < u < 1$, see Table 1; it retains the bell shape of f . It follows that $F^*(t) = t - \sin(2\pi t)/(2\pi)$, for $0 < t < 1$. It seems impossible to obtain an analytical form of f^{n*} for $n \geq 2$. However, as f is bounded, using Theorem 2, we can conclude that f^{n*} converges to the uniform distribution as $n \rightarrow \infty$.

Example 5. *Skew-normal.*

A skew-normal distribution [17,24] has the density of the form

$$f(x) = 2\phi(x)\Phi(\alpha x), \quad x \in \mathbb{R},$$

where $\alpha \in \mathbb{R}$ is a parameter, ϕ and Φ , as before, are the density and cdf of the standard normal distribution. When $\alpha = 0$, f is reduced to the standard normal so it is possible to obtain its $\{f^{n*}\}$ by induction and then derive directly that f^{n*} converges to the uniform distribution as $n \rightarrow \infty$. However, the general form of skew-normal densities is a lot harder to handle and one can easily see that the density is bounded and so Theorem 2 can be employed to conclude that f^{n*} converges to the uniform distribution as $n \rightarrow \infty$.

Example 6. Let $f(x) = -\ln x$, $x \in (0, 1)$. Then, $\mu_n = n!$ and $\kappa_n = \frac{n+2}{n+1} \rightarrow 1$ as $n \rightarrow \infty$, so we have from Theorem 1 that, for any $r > 0$, f^{n*} converges in L_r norm to constant 1 as $n \rightarrow \infty$.

4. Discussion

The pdQ, transformation from a density function f to f^* extracts the important information of f such as its asymmetry and tail behaviour and ignores the less critical information such as gaps, location and scale, and thus provides a powerful tool in studying the shapes of density functions. We found the directed divergences from uniformity of the pdQs of many standard location-scale families and used them to make a map locating each shape family relative to others and giving its distance from uniformity. It would be of interest to find the pdQs of other shape families, such as the skew-normal of Example 5; however, a simple expression for this pdQ appears unlikely given the complicated nature of its quantile function. Nevertheless, the [25] skew-normal family should be amenable in this regard because there are explicit formulae for both its density and quantile functions. To obtain the information conserved in the pdQ transformation, we repeatedly applied

the transformation and found the limiting behaviour of repeated applications of the pdQ mapping. When the density function f is bounded, we showed that each application lowers its modal height and hence the resulting density function f^* is closer to the uniform density than f . Furthermore, we established a necessary and sufficient condition for f^{n^*} converging in L_2 norm to the uniform density, giving a positive answer to a conjecture raised in [1]. In particular, if f is bounded, we proved that f^{n^*} converges in L_r norm to the uniform density for any $r > 0$. The fixed point theorems can be interpreted as follows. As we repeatedly apply the pdQ transformation, we keep losing information about the shape of the original f and will eventually exhaust the information, leaving nothing in the limit, as represented by the uniform density, which means no points carry more information than other points. Thus, the pdQ transformation plays a similar role to the difference operator in time series analysis where repeated applications of the difference operator to a time series with a polynomial component lead to a white noise with a constant power spectral density ([26], p. 19). We conjecture that every almost surely positive density g on $[0, 1]$ is a pdQ of a density function, hence uniquely representing a location-scale family. This is equivalent to saying that there exists a density function f such that $g = f^*$. When g satisfies $\int_0^1 \frac{1}{g(t)} dt < \infty$, one can show that the cdf F of f can be uniquely (up to location-scale parameters) represented as $F(x) = H^{-1}(H(1)x)$, where $H(x) = \int_0^x \frac{1}{g(t)} dt$ (Professor A.D. Barbour, personal communication). The condition $\int_0^1 \frac{1}{g(t)} dt < \infty$ is equivalent to saying that f has bounded support and it is certainly not necessary, e.g., $g(x) = 2x$ for $x \in [0, 1]$ and $f(x) = e^x$ for $x < 0$ (see Example 2 in Section 3.2).

5. Conclusions

In summary, the study of shapes of probability densities is facilitated by composing them with their own quantile functions, which puts them on the same finite support where they are absolutely continuous with respect to Lebesgue measure, and thus amenable to metric and semi-metric comparisons. In addition, we showed that further applications of this transformation, which intuitively reduces information and increases the relative entropy, is generally valid but requires a non-standard approach for proof. Similar results are likely to be obtainable in the multivariate case. Further research could investigate the relationship between relative entropy and tail-weight or distance from the class of symmetric pdQs.

Supplementary Materials: An R script entitled `StaudteXiaSupp.R`, which is available online at www.mdpi.com/1099-4300/20/5/317/s1, enables the reader to plot successive iterates of the pdQ transformation on any standard probability distribution available in R.

Author Contributions: Section 2 is mostly due to Robert Staudte and Section 3 is mostly due to Aihua Xia. Both authors contributed equally to the other three sections.

Acknowledgments: The authors thank the three reviewers for their critiques and many positive suggestions. The authors also thank Peter J. Brockwell for helpful commentary on an earlier version of this manuscript. The research by Aihua Xia is supported by an Australian Research Council Discovery Grant DP150101459. The authors have not received funds for covering the costs to publish in open access.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Staudte, R. The shapes of things to come: Probability density quantiles. *Statistics* **2017**, *51*, 782–800. [CrossRef]
2. Parzen, E. Nonparametric statistical data modeling. *J. Am. Stat. Assoc.* **1979**, *7*, 105–131. [CrossRef]
3. Kullback, S.; Leibler, R. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
4. Kullback, S. *Information Theory and Statistics*; Dover: Mineola, NY, USA, 1968.
5. Abbas, A.; Cadenbach, A.; Salimi, E. A Kullback–Leibler View of Maximum Entropy and Maximum Log-Probability Methods. *Entropy* **2017**, *19*, 232. [CrossRef]

6. Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A* **1946**, *186*, 453–461. [[CrossRef](#)]
7. Freimer, M.; Kollia, G.; Mudholkar, G.; Lin, C. A study of the generalized Tukey lambda family. *Commun. Stat. Theory Methods* **1988**, *17*, 3547–3567. [[CrossRef](#)]
8. Stephens, M. Uniformity, Tests of. In *Encyclopedia of Statistical Sciences*; John Wiley & Sons: Hoboken, NJ, USA, 2006; Volume 53, pp. 1–8, doi:10.1002/0471667196.ess2810.pub2.
9. Lockhart, R.; O'Reilly, F.; Stephens, M. Tests of Fit Based on Normalized Spacings. *J. R. Stat. Soc. B* **1986**, *48*, 344–352.
10. Schader, M.; Schmid, F. Power of tests for uniformity when limits are unknown. *J. Appl. Stat.* **1997**, *24*, 193–205. [[CrossRef](#)]
11. Prendergast, L.; Staudte, R. Exploiting the quantile optimality ratio in finding confidence Intervals for a quantile. *Stat* **2016**, *5*, 70–81. [[CrossRef](#)]
12. Dudewicz, E.; Van Der Meulen, E. Entropy-Based Tests of Uniformity. *J. Am. Stat. Assoc.* **1981**, *76*, 967–974. [[CrossRef](#)]
13. Bowman, A. Density based tests for goodness-of-fit. *J. Stat. Comput. Simul.* **1992**, *40*, 1–13. [[CrossRef](#)]
14. Fan, Y. Testing the Goodness of Fit of a Parametric Density Function by Kernel Method. *Econ. Theory* **1994**, *10*, 316–356. [[CrossRef](#)]
15. Pavia, J. Testing Goodness-of-Fit with the Kernel Density Estimator: GoFKernel. *J. Stat. Softw.* **2015**, *66*, 1–27. [[CrossRef](#)]
16. Noughabi, H. Entropy-based tests of uniformity: A Monte Carlo power comparison. *Commun. Stat. Simul. Comput.* **2017**, *46*, 1266–1279. [[CrossRef](#)]
17. Arellano-Valle, R.; Contreras-Reyes, J.; Stehlik, M. Generalized Skew-Normal Negentropy and Its Application to Fish Condition Factor Time Series. *Entropy* **2017**, *19*, 528. [[CrossRef](#)]
18. R Core Team. *R Foundation for Statistical Computing*; R Core Team: Vienna, Austria, 2008; ISBN 3-900051-07-0.
19. Luenberger, D. *Optimization by Vector Space Methods*; Wiley: New York, NY, USA, 1969.
20. Bessenyei, M.; Páles, Z. A contraction principle in semimetric spaces. *J. Nonlinear Convex Anal.* **2017**, *18*, 515–524.
21. Feller, W. *An Introduction to Probability Theory and Its Applications*; John Wiley & Sons: New York, NY, USA, 1971; Volume 2.
22. Johnson, N.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; John Wiley & Sons: New York, NY, USA, 1994; Volume 1.
23. Johnson, N.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; John Wiley & Sons: New York, NY, USA, 1995; Volume 2, ISBN 0-471-58494-0.
24. Azzalini, A. A Class of Distributions which Includes the Normal Ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
25. Jones, M.; Pewsey, A. Sinh-arcsinh distributions. *Biometrika* **2009**, *96*, 761–780. [[CrossRef](#)]
26. Brockwell, P.; Davis, R. *Time Series: Theory and Methods*; Springer: New York, NY, USA, 2009.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).