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

Robert G. Staudte*

La Trobe University

Aihua Xia†

University of Melbourne

21 February, 2017

Abstract

The probability density quantile (pdQ) carries essential information regarding shape and tail behavior of a location-scale family. Convergence of repeated applications of the pdQ mapping to the uniform distribution is investigated and new fixed point theorems are established. The Kullback-Leibler divergences from uniformity of these pdQs are mapped and found to be ingredients in power functions of optimal tests for uniformity against alternative shapes.

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

AMS subject classification: Primary 62E10; Secondary 62F03.

*Postal address: Department of Mathematics and Statistics, La Trobe University, VIC 3086, Australia. Email address: r.staudte@latrobe.edu.au.

†Corresponding author. Postal address: School of Mathematics and Statistics, University of Melbourne, VIC 3010, Australia. Email address: aihuaxia@unimelb.edu.au. Research supported by ARC Discovery Grant DP150101459.

1 Introduction

1.1 Background and summary

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 fitting shape parameter families to data; details are in Staudte (2017).

The Kullback-Leibler directed divergence and symmetrized divergence (KLD) of a pdQ with respect to the uniform distribution on [0,1] is investigated in Section 1.2, 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 2, where it is found to be a fixed point. We then investigate the convergence to uniformity of repeated applications of the pdQ transformation, by means of fixed point theorems for a semi-metric. In Section 3 power functions of hypothesis tests of uniformity against specified alternative shapes are shown to be dependent on the symmetrized Kullback-Leibler divergence. Further ideas are discussed in Section 4.

1.2 Definitions and divergence map

Let $\mathcal{F}$ denote the class of cumulative distribution functions (cdf s) on the real line 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 $X \sim f$ or $f \sim F$. We only discuss F absolutely continuous with respect to Lebesgue measure, but the results can be extended to the discrete and mixture cases using suitable dominating measures.

Definition 1.1 *Let $\mathcal{F}' = \{F \in \mathcal{F} : f = F'$ exists and is positive$\}$. For each $F \in \mathcal{F}'$ we follow Parzen (1979) and define the* quantile density *function $q(u) = Q'(u) = 1/f(Q(u))$. He called its reciprocal $fQ(u) = f(Q(u))$ 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 $f \to f^*$ means that $\mathcal{F}'^*$ is a proper subset of $\mathcal{F}'$. The advantages of working with f^*s over fs 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.

Next we evaluate and plot the Kullback & Leibler (1951) divergences from uniformity. The Kullback & Leibler (1951) 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, 1968, 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 $\mathcal{U}$. In Table 1 are shown 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) $J(\mathcal{U}, f^*) := I(\mathcal{U} : f^*) + I(f^* : \mathcal{U})$.

Definition 1.2 *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* $(s_1, s_2) = (\sqrt{I(f^* : \mathcal{U})}, \sqrt{I(\mathcal{U} : f^*)})$, *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 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 $d(\mathcal{U}, \mathcal{C}) = 1$. However, d can provide an informative measure of distance from uniformity.

In Figure 1 are shown 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 upper 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 $\nu \to \infty$. The Gamma

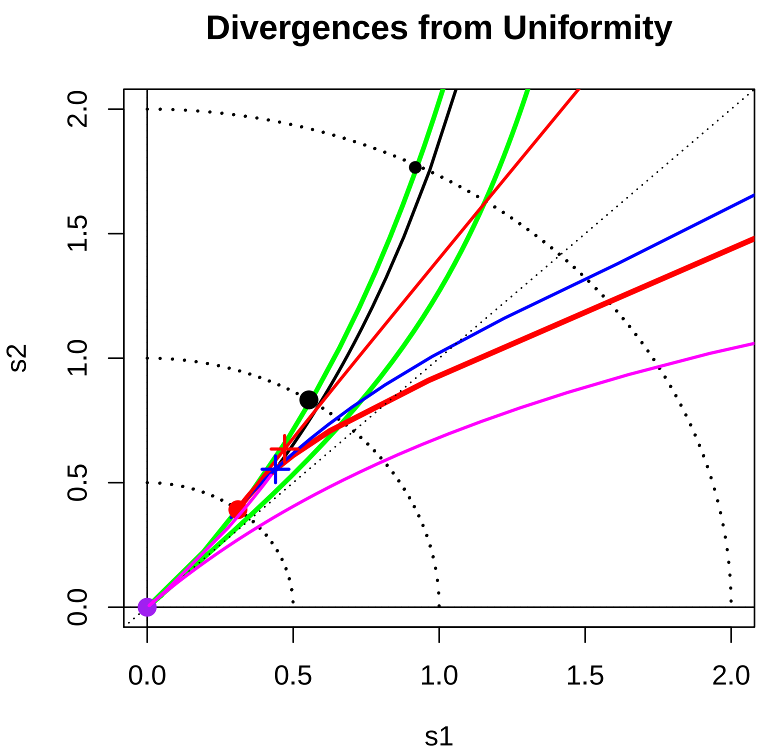

Figure 1: **Divergence from uniformity.** *The loci of points* $(s_1, s_2) = (\sqrt{I^*(\mathcal{U} : f)}, \sqrt{I^*(f : \mathcal{U})})$ *is shown for various standard families. The large disks correspond respectively to the symmetric families: uniform (purple), normal (red) and Cauchy (black). The crosses correspond to the asymmetric distributions: exponential (blue) and standard lognormal (red). More details are given in Section 1.2.*

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 $a \to \infty$.

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 $b \to \infty$. For each choice of $\alpha > 0.5$, $\beta > 0.5$ the locus of the Beta(α, β) pdQ divergences lies below the chi-squared upper red curve and mostly above the power(b) magenta curve; however, the U-shaped Beta distributions have loci below it.

The upper green line near the Pareto black curve gives the loci of root-divergences from uniformity of the Tukey(λ) with $\lambda < 1$, while the lower 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 (Staudte, 2017, Sec.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 distant 2 from uniformity. The generalized Tukey distributions of Freimer *et al.* (1988) 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.

2 Convergence of density shapes to uniformity via fixed point theorems

The transformation $f \to f^*$ of Definition 1.1 is quite powerful, removing location and scale and moving the distribution from the support of f to the unit interval. 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. Further, with n iterations $f^{(n+1)*} := (f^{n*})^*$ for $n \geq 2$, we would expect that f^{n*} converges to the uniform density as $n \to \infty$. An R script Team (2008) for finding repeated $*$-iterates of a given pdQ is available as Supplementary Online Material.

2.1 Conditions for convergence to uniformity

Definition 2.1 *Given $f \in \mathcal{F}'$, we say that f is of* $*$-order *n if $f^*, f^{2*}, \ldots, f^{n*}$ 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 2.2 *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 2.2: For each $i, n \geq 1$, provided all terms below are finite, we have the following recursive formula

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

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 1.1, $\kappa_1 < \infty$ means that $\kappa_0 < \infty$. Hence (i) follows from (2) with $n = m$. □

Next we investigate the involutionary nature of the $*$-transformation.

Proposition 2.3 *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 2.3: 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. □.

Proposition 2.3 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 f^{n*} has a limit and that the limit is the uniform distribution. It was hoped the classical machinery for convergence in Banach spaces (Luenberger, 1969, Ch.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., Bessenyei & Páles

(2017) and references therein), the fixed point Theorems 2.4, 2.5 and 2.6 shown below do not seem to be covered in these general studies. 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 $n \to \infty$.

Theorem 2.4 *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) $\frac{\mu_n \mu_{n+2}}{\mu_{n+1}^2} \to 1$ *as* $n \to \infty$.

Remark Notice that $\mu_n = \mathbb{E}\left\{f^*(U)^{n-1}\right\}$, $n \geq 1$, are the moments of the random variable $f^*(U)$ with $U \sim \mathcal{U}$, Theorem 2.4 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 (Feller, 1971, 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 (Feller, 1971, pp. 225–226), leading to stronger results in Theorem 2.5.

Proof of Theorem 2.4: It is obvious that (ii) implies (i).

(i) $\Rightarrow$ (iii): By Proposition 2.2, $\kappa_n = \frac{\mu_n \mu_{n+2}}{\mu_{n+1}^2}$. 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

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

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

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

and (2) ensures

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

which imply

$$\nu_{n,r} = \kappa_{n+r-2} \kappa_{n+r-3}^2 \cdots \kappa_n^{r-1} \to 1$$

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

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

Theorem 2.5 *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 2.5: 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 2.3 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 $\kappa_n \to 1$ as $n \to \infty$. In fact, if $\kappa_n \not\to 1$, 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

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

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

Theorem 2.6 *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) $\mu_n \mu_{n+2} \mu_{n+1}^{-2} \to 1$ as $n \to \infty$.

Proof of Theorem 2.6: 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 $\kappa_{n_{k_i}} \to b$ as $i \to \infty$. It remains to show that $b = 1$. To this end, for $\delta > 1$, we have

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

(i*) ensures that

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

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

2.2 Examples of convergence to uniformity

The main results in section 2.1 cover all the standard distributions with infinite $*$-order in Johnson *et al.* (1994), Johnson *et al.* (1995). In fact, as observed in the Remark after Theorem 2.4 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 $\{f^{n*}:\ n \ge 1\}$ does not converge to the uniform distribution. Here we give a few examples to show that the main results in section 2.1 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 $b \ge 1$. As f_b is bounded for $b \ge 1$, Theorem 2.5 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 2.5 says that f^{n*} converges to the uniform distribution as $n \to \infty$. 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 2.5 yields that the sequence $\{f_a^{n*}\}_{n\ge1}$ converges to the uniform distribution as $n \to \infty$.

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 as shown in Figure 1. 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 $n \ge 2$. However, as f is bounded, using Theorem 2.5, we can conclude that f^{n*} converges to the uniform distribution as $n \to \infty$.

Example 5: Normal distribution.

Although it is possible to obtain $\{f^{n*}\}$ by induction and then derive directly that f^{n*} converges to the uniform distribution as $n \to \infty$, one can easily see that the pdf is bounded and so Theorem 2.5 can be employed to get the same conclusion.

Example 6:

Let $f(x) = -\ln x$, $x \in (0,1)$, then $\mu_n = n!$ and $\kappa_n = \frac{n+2}{n+1} \to 1$ as $n \to \infty$, so we have from Theorem 2.4 that for any $r > 0$, f^{n*} converges in L_r norm to constant 1 as $n \to \infty$.

3 Testing for uniformity

The larger the value of $d(\mathcal{U}, f^*)$, the easier it should be to discriminate between $\mathcal{U}$ and f^*. This is indeed the case, as we now demonstrate.

3.1 Power of tests for detecting non-uniform shapes

The connection between Kullback-Leibler divergences and the Neyman-Pearson Lemma is well-known, see text and references in Eguchia & Copas (2006). The following material on power functions for tests for uniformity, while contextual to comparing shapes, may prove useful in other situations.

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 $\widehat{f}^*_m$ and comparing it with the uniform density on [0,1] using any one of a number of metrics. Consistent estimators $\widehat{f}^*_m$ for f^* based on normalized reciprocals of the quantile density estimators derived in Prendergast & Staudte (2016) are available and described in Staudte (2017, Section 2). However an investigation into such omnibus nonparametric testing procedures, and comparison with other kernel density based techniques in Bowman (1992); Fan (1994) and Pavia (2015), is beyond the scope of this work.

Given the large number of tests for uniformity that are available, see text and references in Stephens (2006), one may well ask why introduce new ones? The *doyen* of goodness-of-fit testing Stephens (2006) provides the answer:

> *Since transformations are often used to produce a set of uniforms, it might be appropriate to conclude with some cautionary remarks on when uniformity is*

not to be expected. This will be so, for example, when the U set is derived from the PIT and when some parameters, unknown in the distribution, are replaced by estimates. In this situation, even when the estimates are efficient, the U set will be superuniform, giving much lower values of, say, EDF statistics, than if the set were uniform; this remains so even as the sample size grows bigger.

In practice this means that if a test for uniformity is preceded by a probability integral transformation (PIT) 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 are in Lockhart *et al.* (1986) and Schader & Schmid (1997).

In this section we study the simpler problem of testing the null hypothesis of uniformity $H_0 : f^*(u) = 1$ for all $0 < u < 1$ (denoted $\mathcal{U}$) against a specified alternative $H_1 : f^* = f_1^*$. This test will give us a standard by which to judge the power of any future nonparametric test when the specific alternative holds. Given a vector of $\mathbf{X} = (X_1, \ldots, X_m)$ of i.i.d. variables from f^*, and realized values $\mathbf{x} = (x_1, \ldots, x_m)$ the Neyman-Pearson (NP) test rejects H_0 in favor of H_1 when the product $\prod_1^m f_1^*(x_i)$ is large, or equivalently, when

$$l_{\mathbf{x}} = \sum_1^m \ln(f_1^*(x_i)) \geq c_{m,\alpha} \ , \tag{8}$$

where $c_{m,\alpha}$ is chosen to achieve level α. In general this critical point will be difficult to determine, but for the normal pdQ alternative $f_1^*(u) = 2\sqrt{\pi}\,\varphi(z_u)$ we have: $l_{\mathbf{x}} = m\ln(\sqrt{2}) - \frac{1}{2}\sum_1^m\{\Phi^{-1}(x_i)\}^2$, where φ is the density of the standard normal. Under the null hypothesis $\sum_1^m\{\Phi^{-1}(U_i)\}^2 \sim \chi_m^2$, so an equivalent test to (8) would reject the null hypothesis of uniformity in favor of normality when $\sum_1^m\{\Phi^{-1}(x_i)\}^2 \leq \chi_m^2(\alpha)$. This is the most powerful level-α test of these simple hypotheses based on m observations.

Returning to a general alternative pdQ we can find asymptotically most powerful level-α tests based on the fact that $l_{\mathbf{X}}$ is a sum of i.i.d. random variables with common mean $\mu_0 = \mathbb{E}_0[\ln(f_1^*(U))]$ and variance $\sigma_0^2 = \mathrm{Var}_0[\ln(f_1^*(U))]$, which we assume are finite. By virtue of a CLT, the large-sample NP test rejects H_0 at asymptotic level α when

$$(l_{\mathbf{x}}/\sqrt{m} - \sqrt{m}\mu_0)/\sigma_0 \geq z_{1-\alpha} \ . \tag{9}$$

To obtain an expression for the large sample power of this test, let $\mu_1 = \mathbb{E}_1[\ln(f_1^*(X))]$ and $\sigma_1^2 = \mathrm{Var}_1[\ln(f_1^*(X))]$, again assumed to be finite; then the asymptotic power of the test (9) against f^* based on m observations is readily found to be:

$$\Pi_m(f_1^*) = \Phi\left(\sqrt{m}\,\frac{(\mu_1 - \mu_0)}{\sigma_1} + z_\alpha\,\frac{\sigma_0}{\sigma_1}\right) \ . \tag{10}$$

Table 1: **Quantiles of some distributions, their pdQs and quantities relevant to the asymptotic power function (10).** *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* $u = 0.5$*. LN represents the standard lognormal distribution. The quantile function for the Pareto is for the Type II distribution with shape* $a = 1$*, and the pdQ is the same for Type I and Type II Pareto models.*

	$Q(u)$	$f^*(u)$	$-\mu_0$	μ_1	σ_0	σ_1	$\mu_1 - \mu_0$	$\frac{\mu_1-\mu_0}{\sigma_1}$	$\frac{\sigma_0}{\sigma_1}$
Normal	z_u	$2\sqrt{\pi}\,\varphi(z_u)$	0.153	0.097	0.707	0.354	0.250	0.707	2.000
Logistic	$\ln(u/(1-u))$	$6u(1-u)$	0.208	0.125	0.843	0.393	0.333	0.848	2.143
Laplace	$\ln(2u),\ u \leq 0.5$	$2\min\{u, 1-u\}$	0.307	0.193	1.000	0.500	0.500	1.000	2.000
t_2	$\frac{2u-1}{\{2u(1-u)\}^{1/2}}$	$\frac{2^7\{u(1-u)\}^{3/2}}{3\pi}$	0.391	0.200	1.264	0.463	0.591	1.276	2.728
Cauchy	$\tan\{\pi(u-0.5)\}$	$2\sin^2(\pi u)$	0.693	0.307	1.814	0.538	1.000	1.857	3.369
Exp.	$-\ln(1-u)$	$2(1-u)$	0.307	0.193	1.000	0.500	0.500	1.000	2.000
Gumbel	$-\ln(-\ln(u))$	$-4u\ln(u)$	0.191	0.116	0.803	0.381	0.307	0.806	2.109
LN	e^{z_u}	$\frac{2\sqrt{\pi}}{e^{1/4}}\,\varphi(z_u)\,e^{-z_u}$	0.403	0.222	1.225	0.500	0.625	1.250	2.449
Pareto	$(1-u)^{-1}$	$3\,(1-u)^2$	0.901	0.432	2.000	0.667	1.333	2.000	3.000

We need α, m, $l_{\mathbf{x}}$, μ_0, σ_0^2 to carry out the test (9); and we also need μ_1, σ_1^2 to compute the asymptotic power (10). Notice that the distances from the origin in Figure 1 of Section 1.2 are based on the directed divergences $I(\mathcal{U} : f_1^*) = -\mathbb{E}_0[\ln(f_1^*(X))] = -\mu_0$ and $I(f_1^* : \mathcal{U}) = \mathbb{E}_1[\ln(f_1^*(X))] = \mu_1$, so the symmetrized divergence KLD is $J(\mathcal{U}, f_1^*) = \mu_1 - \mu_0$. Thus the power function (10) is non-decreasing in the KLD or its square root, the distance $d(\mathcal{U}, f_1^*)$ between null and alternative.

Some examples of μ_0, μ_1, $\mu_1 - \mu_0$, $(\mu_1 - \mu_0)/\sigma_1$ and σ_0/σ_1 are given in Table 1. Note the particularly simple values for the symmetrized divergence $J(\mathcal{U}, f_1^*) = \mu_1 - \mu_0$. Distributions with shapes 'visually far' from uniformity have large values in the two right-most columns, so that the test will more easily detect them. For the normal alternative, the asymptotic power (10) at level α is $\Pi_m(f_1^*) = \Phi(\sqrt{m/2} + 2\,z_\alpha)$, which exceeds α when $m > 2z_{1-\alpha}^2$.

There are other situations where the asymptotic power functions are monotone increasing functions of the KLD with many one-sample examples in Kulinskaya *et al.* (2008, 2010, 2014), one-parameter exponential families Morgenthaler & Staudte (2012), two-sample bi-

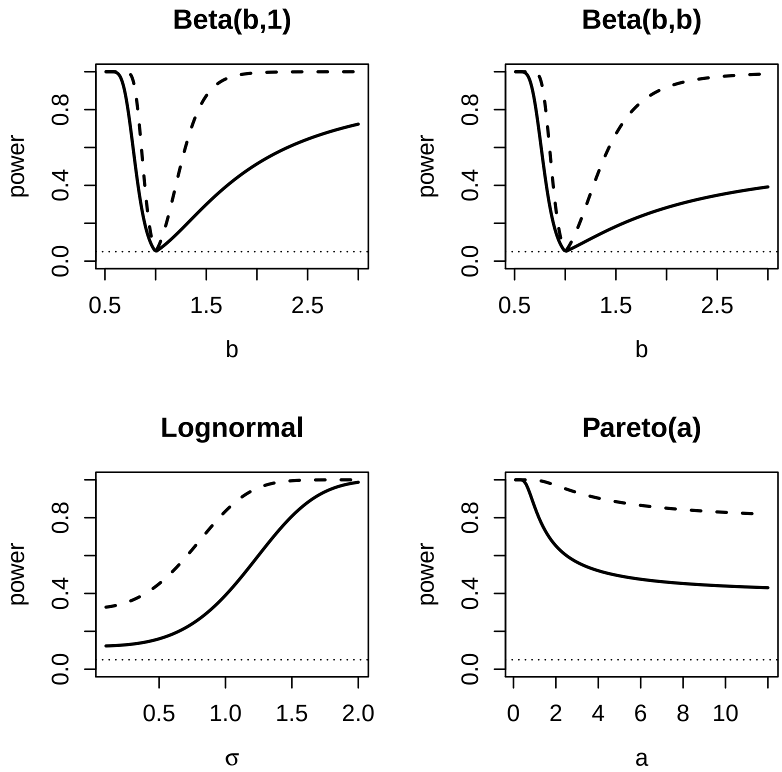

Figure 2: The top left plot shows the asymptotic power function for level 0.05 tests of uniformity against alternative the Beta(b,1) pdQ when $m = 25$ (solid line) and $m = 100$ (dashed line). The plots for the symmetric Beta distribution are on its right, with the same sample sizes; these power functions show that it is harder to detect the symmetric ones for all $b \neq 1$. In the bottom two plots, the sample sizes are much smaller, $m = 9$ solid lines and $m = 16$ dashed lines.

nomial tests Prendergast & Staudte (2014) and non-central chi-squared and non-central F families arising in tests for equivalence Morgenthaler & Staudte (2016). This reveals a general phenomenon but no meta-theorem containing these results is yet available.

3.2 Examples of power functions for shape families

The power functions of testing uniformity against the pdQs of four shape families, are shown in Figure 2. The first two models, power function model Beta$(b, 1)$ and the symmetric Beta(b, b) model for $b > 0.5$ contain the null hypothesis. Their respective power functions (10) for a level 0.05 test of uniformity based on $m = 25$ and 100 observations are shown in the top two plots. Similar plots for alternatives Lognormal(σ) and Pareto(a) families are also shown for much smaller sample sizes 9 and 16 indicating that small samples will likely detect these alternative shapes.

The plots in Figure 2 require the null and alternative means and variances of the test statistic, and were obtained by numerical integration. In the case of the Beta(b,1) model exact results are derived as follows. the quantile function is $Q(b) = u^{1/b}$ and for $b > 1/2$ its density is square integrable, leading to the pdQ $f_1^*(u) = (2 - \frac{1}{b})\, u^{1-\frac{1}{b}}$. The log-likelihood for one observation $X = x$ required in (9) is $l_x(b) = \ln(2 - 1/b) + (1 - 1/b)\ln(x)$. Thus $\mu_0(b) = \mathbb{E}_0[l_{\mathbf{X}}(b)] = \ln(2-1/b) + 1/b - 1$ and $\mu_1(b) - \mu_0(b) = (1-1/b)^2/(2-1/b)$. Further $\sigma_0(b) = \{\text{Var}_0[l_{\mathbf{X}}(b)]\}^{1/2} = |1 - 1/b|$ and $\sigma_1(b) = \{\text{Var}_1[l_{\mathbf{X}}(b)]\}^{1/2} = |1 - 1/b|/(2 - 1/b)$. Hence $(\mu_1(b) - \mu_0(b))/\sigma_1(b) = |1-1/b|$ and $\sigma_0(b)/\sigma_1(b) = 2 - 1/b$. The power function (10) is therefore $\Pi_m(f_1^*(b)) = \Phi(\sqrt{m}|1 - 1/b| + (2 - 1/b)\, z_\alpha) \qquad \text{for} \qquad b > 1/2.$

4 Summary and 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 distributional 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. We also found the most powerful tests of uniformity against alternative shapes and showed that their power functions are monotone increasing in the distances on the map.

In terms of 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 Staudte (2017). 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 polynomial component lead to a white noise with a constant power spectral density

(Brockwell & Davis, 2009, p. 19).

We conjecture that every almost surely positive density g on $[0,1]$ is a pdQ of a density function, hence uniquely represents 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 2.2).

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.

Acknowledgments: The authors thank Professor P.J. Brockwell for helpful commentary on an earlier version of this manuscript. This research is supported by ARC Discovery Grant DP150101459.

References

Bessenyei, M., & Páles, Z. 2017. A contraction principle in semimetric spaces. *Journal of nonlinear and convex analysis*, **18**(3), 515–524.

Bowman, A.W. 1992. Density based tests for goodness-of-fit. *J. Statist. Comp. & and Sim.*, **40**, 1–13.

Brockwell, P.J., & Davis, R.A. 2009. *Time Series: Theory and Methods.* Springer-Verlag.

Eguchia, S., & Copas, J. 2006. Interpreting kullbackleibler divergence with the neymanpearson lemma. *J. Multiv. Analysis*, **97**, 2034–2040.

Fan, Y. 1994. Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory*, **10**, 316–356.

Feller, W. 1971. *An introduction to probability theory and its applications, vol. 2.* New York: John Wiley & Sons, Inc.

Freimer, M., Mudholkar, G.S., Kollia, G., & Lin, C.T. 1988. A study of the generalized Tukey lambda family. *Communications in Statistics - Theory and Methods*, **17**, 3547–3567.

Johnson, N.L., Kotz, S., & Balakrishnan, N. 1994. *Continuous univariate distributions*. Vol. 1. New York: John Wiley & Sons.

Johnson, N.L., Kotz, S., & Balakrishnan, N. 1995. *Continuous univariate distributions*. Vol. 2. New York: John Wiley & Sons. ISBN 0-471-58494-0.

Kulinskaya, E., Morgenthaler, S., & Staudte, R.G. 2008. *Meta Analysis: a Guide to Calibrating and Combining Statistical Evidence*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons.

Kulinskaya, E., Morgenthaler, S., & Staudte, R.G. 2010. Variance stabilizing the difference of two binomial proportions. *The American Statistician*, **64**(4), 350–356.

Kulinskaya, E., Morgenthaler, S., & Staudte, R.G. 2014. Combining Statistical Evidence. *Int. Statist. Inst. Rev.*, **82**(2), 214–242.

Kullback, S. 1968. *Information Theory and Statistics*. Mineola, NY: Dover.

Kullback, S., & Leibler, R.A. 1951. On information and sufficiency. *Ann. Math. Statist.*, **22**, 79–86.

Lockhart, R.A., O'Reilly, F.J., & Stephens, M.A. 1986. Tests of fit based on normalized spacings. *J. R. Statist. Soc. B*, **48**, 344–352.

Luenberger, D.G. 1969. *Optimization by Vector Space Methods*. New York, NY: Wiley.

Morgenthaler, S., & Staudte, R.G. 2012. Advantages of variance stabilization. *Scandinavian Journal of Statistics*, **39**, 714–728.

Morgenthaler, S., & Staudte, R.G. 2016. Indicators of evidence for equivalence. *Entropy*, **18**, 291.

Parzen, E. 1979. Nonparametric statistical data modeling. *Journal of the American Statistical Association*, **7**, 105–131.

Pavia, J.M. 2015. Testing goodness-of-fit with the kernel density estimator: Gofkernel. *J. Statist. Soft.*, **66**, 1–27.

Prendergast, L.A., & Staudte, R.G. 2014. Better than you think: interval estimators of the difference of binomial proportions. *Journal of Statistical Planning and Inference*, **148**, 38–48. DOI: 10.1016/j.jspi.2013.11.012.

Prendergast, L.A., & Staudte, R.G. 2016. Exploiting the quantile optimality ratio in finding confidence intervals for a quantile. *Stat*, **5**(1), 70–81.

Schader, M., & Schmid, F. 1997. Power of tests for uniformity when limits are unknown. *Journal of Applied Statistics*, **24**, 193–205.

Staudte, R.G. 2017. The shapes of things to come: probability density quantiles. *Statistics: a Journal of Theoretical and Applied Statistics*, **51**, 782–800.

Stephens, M.S. 2006. Uniformity, tests of. *Pages 1–8 of: Encyclopedia of Statistical Sciences*, vol. 53. John Wiley & Sons. DOI: 10.1002/0471667196.ess2810.pub2.

Team, R Development Core. 2008. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.