Probability in the Engineering and Informational Sciences, **24**, 2010, 549–559.
doi:10.1017/S0269964810000173

METHODS FOR SYMMETRIZING RANDOM VARIABLES

CHRISTOPHER S. WITHERS
Applied Mathematics Group
Industrial Research Limited
Lower Hutt, New Zealand
E-mail: c.withers@irl.cri.nz

SARALEES NADARAJAH
School of Mathematics
University of Manchester
Manchester M13 9PL, UK
E-mail: mbbsssn2@manchester.ac.uk

Let X be a random variable with nonsymmetric density $p(x)$. We give the symmetric density $q(x)$ closest to it in the sense of Kulback–Liebler and Hellinger distances. (All symmetries are around zero.) For the first distance, we show that $q(x)$ is proportional to the geometric mean of $p(x)$ and $p(-x)$. For example, a symmetrized shifted exponential is a centered uniform, and a symmetrized shifted gamma is a centered beta random variable. For the second distance, $q(x)$ is proportional to the square of the arithmetic mean of $p(x)^{1/2}$ and $p(-x)^{1/2}$. Sample versions are also given for each. We also give the optimal random function f such that $f(X)$ is symmetrically distributed and minimizes $|f(X) - X|$. Finally, we show how to optimize the Hellinger distance for vector X subject to supersymmetry and for scalar X subject to being monotone about zero in each half-line.

1. INTRODUCTION AND SUMMARY

The need for symmetric random variables arises in many areas of probability and statistics. For example, in the construction of symmetric random walks, in the construction of skew symmetric distributions (Azzalini [2]), in kernel density estimation (Silverman [8]), and in the construction of models for circular data (Fisher [4]). The need for symmetric random variables also arises in a variety of application areas. Particularly, in engineering and informational sciences, see, for example, Coffman, Flajolet, Flatto, and Hofri [3] and Al-Awadhi, Konsowa, and Najeh [1].

© Cambridge University Press 2010 0269-9648/10 $25.00

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

Given a real random variable X with a density $p(x)$ that is not symmetric about zero, there are many ways to obtain the density $q(x)$ of a real random variable Y that is a symmetrized form of X. A common method is to take

$$Y = \pm X \text{ with probability } \alpha = (1 - p_0)/2, \tag{1.1}$$

where $p_0 = P(X = 0)$. This gives

$$q(x) = [p(x) + p(-x)]/2, \qquad x \neq 0, \qquad P(Y = 0) = p_0;$$

that is, $q(x)$ is just the arithmetic mean of $p(x)$ and $p(-x)$. (If there is no atom at zero, this becomes $Y = \pm X$ with probability $1/2$.) In Section 5, we show that this choice of $Y = f(X)$ is in fact the optimal random function f independent of X that minimizes $E\,(f(X) - X)^2$ if $E\,X^2 < \infty$.

The aim of this article is to present some alternatives to minimize distances between the two densities. In Sections 2 and 3, we give $q(x)$ that minimizes the Kulback–Liebler distance (Kullback and Leibler [6]; Kullback [5]) defined by

$$I(q, p) = \int q(x) \ln[q(x)/p(x)]\, d\mu(x)$$

and the Hellinger distance (Nikulin [7]) defined by

$$2H(q, p)^2 = \int \left[q(x)^{1/2} - p(x)^{1/2}\right]^2 d\mu(x),$$

where μ is the common measure with respect to which the Radon–Nikodym derivatives p and q are defined.

In Section 2 we show that $q(x)$ minimizing $I(q, p)$ is proportional to the geometric mean of $p(x)$ and $p(-x)$. In Section 3 we show that $q(x)$ minimizing $H(q, p)$ is proportional to the square of the arithmetic mean of $p(x)^{1/2}$ and $p(-x)^{1/2}$. In Section 4 we give sample versions of Sections 2 and 3. In Section 6, we suppose that we have a random sample of size n, say $\mathbf{X} = (X_1, \ldots, X_n)$. We want to choose $\mathbf{Y} = (Y_1, \ldots, Y_n)$ symmetrically distributed around zero in the sense that $\{Y_1, \ldots, Y_n\} = \{-Y_1, \ldots, -Y_n\}$, to minimize the distance $|\mathbf{Y} - \mathbf{X}|$. We show that there are many solutions, each minimizer giving $|\mathbf{Y} - \mathbf{X}| = |\mathbf{X}|$. Section 7 gives similar results using a stronger form of symmetry. Section 8 gives similar results subject to the density being monotone on each half-line.

2. MINIMIZING THE KULBACK–LIEBLER DISTANCE

We begin with X being a discrete random variable.

THEOREM 2.1: *Consider a random variable X taking integer values. Set $p_i = P(X = i)$. Let Y be a symmetric random variable taking integer values. Set $q_i = P(Y = i)$.*

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

So, $q_{-i} = q_i$. Then the Kulback–Liebler distance $I(q,p)$ is minimized by

$$q_i \equiv \nu_i/d, \qquad i \geq 0, \tag{2.1}$$

where $\nu_i = (p_i p_{-i})^{1/2}$ and $d = \nu_0 + 2\sum_{j\geq 1}^{\infty} \nu_j$.

PROOF: We seek to minimize

$$I(q,p) = \sum_{i=-\infty}^{\infty} q_i \ln(q_i/p_i)$$

such that $q_{-i} = q_i \geq 0$ and $1 = \sum_{i=-\infty}^{\infty} q_i$; that is, we minimize

$$I(q,p) = q_0 \ln(q_0/p_0) + 2\sum_{i=1}^{\infty} q_i(\ln q_i - a_i), \tag{2.2}$$

where $2a_i = \ln(p_i p_{-i})$, so that $\exp(a_i) = \nu_i$ and $1 = q_0 + 2\sum_{i=1}^{\infty} q_i$. Taking λ as a Lagrange multiplier, we seek to solve $J_{.i} \equiv 0$, where $J_{.i}$ is the derivative with respect to q_i of

$$J(q) = I(q,p) + \lambda \sum_{i=-\infty}^{\infty} q_i.$$

So, $0 = J_{.0} = \ln(q_0/p_0) + 1 + \lambda$, [i.e., $q_0 = p_0 \exp(-1-\lambda)$]; and for $i \geq 1$, $0 = J_{.i} = 2(\ln q_i - a_i + 1 + \lambda)$ [i.e., $q_i = \exp(a_i - 1 - \lambda) = \nu_i \exp(-1-\lambda)$]. Additionally, λ is determined by

$$1 = q_0 + 2\sum_{i=1}^{\infty} q_i = p_0 \exp(-1-\lambda) + 2\sum_{i=1}^{\infty} \nu_i \exp(-1-\lambda);$$

that is,

$$\exp(1+\lambda) = p_0 + 2\sum_{i=1}^{\infty} \nu_i = d.$$

Substituting this completes our proof of (2.1). ■

Note 2.1: The condition $d \neq 0$ rules out the case, in which X is positive with probability 1. However, if X is nonnegative and $p_0 > 0$, it gives $q_i = \delta_{i,0}$, (i.e., $Y = 0$).

Example 2.1: Suppose that X is a two-point random variable on $\{-1, 1\}$. Then Y is a two-point random variable on $\{-1, 1\}$ with $q_1 = 1/2$. For $\nu_1 = (p_1 p_{-1})^{1/2}$, $d = 2\nu_1$.

Example 2.2: Let N be a positive integer and suppose that $X = X_0 - N$, where X_0 is a nonnegative random variable taking integer values. Set $r_j = P(X_0 = j)$,

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

$j = 0, 1, 2, \ldots$. So, $p_i = r_{N+i}$, $v_i = (r_{N+i} r_{N-i})^{1/2}$, $|i| \le N$, and $v_i = 0$ for $|i| > N$. For example, if X_0 is Poisson with mean θ, then $r_j = \theta^j \exp(-\theta)/j!$.

Note 2.2: The solution holds with no change if we replace the range of X and Y from the integers $N = \{i\}$ to a countable set of distinct values $\{x_i\}$ from some arbitrary space S in which $\{-x_i\}$ are defined. Then $p_i = P(X = x_i)$ and $q_i = P(Y = x_i)$. For example, S could be N^r for some $r \ge 1$.

Now, suppose that X is continuous.

THEOREM 2.2: *Suppose that X and Y are continuous random variables with densities $p(x)$ and $q(x)$, respectively, with respect to the Lebesgue measure. Suppose that $q(x)$ is symmetric about zero. Then the density $q(x)$ that minimizes the Kulback–Liebler distance $I(q, p)$ is*

$$q(x) \equiv v(x)/d,$$

where $v(x) = (p(x)p(-x))^{1/2}$ and $d = 2\int_0^\infty v(x)\,dx$.

PROOF: The result follows by a calculus of variations argument, or more simply from Theorem 2.1 by a limiting argument. ■

As for the discrete case, the condition $d \neq 0$ rules out the case where $P(X > 0) = 1$.

Example 2.3: Suppose that X_0 is a positive random variable with density $p_0(x)$ with respect to the Lebesgue measure and that $X = X_0 - N$, where $N > 0$. Then Y is a random variable on $[-N, N]$ with density $q(x) = v(x)/d$, where $v(x) = (p_0(N + x)p_0(N - x))^{1/2}$ and $d = 2\int_0^N v(x)\,dx$.

For example, if X_0 is exponential, then Y is uniform, and if X_0 is gamma, then Y is beta.

Example 2.4: Suppose that X_0 is an exponential random variable with mean 1 and that $X = X_0 - N$, where $N > 0$. Then Y is distributed uniformly on $[-N, N]$.

Example 2.5: Suppose that X_0 is an exponential random variable with mean 1 and that $X = X_0 - N$, where $N > 0$. Then Y is distributed uniformly on $[-N, N]$. (For, $v(x) = \exp(-N)I(|x| \le N)$, $d = 2N\exp(-N)$.)

Example 2.6: Suppose that X_0 is a gamma random variable with mean $\gamma > 0$ and that $X = X_0 - N$, where $N > 0$. Then Y is distributed as a beta random variable on $[-N, N]$ with parameters (α, α), where $\alpha = (\gamma + 1)/2$; that is, $Y = N(2B_{\alpha,\alpha} - 1)$, where $B_{\alpha,\beta}$ is a standard beta random variable with density $x^{\alpha-1}(1 - x)^{\beta-1}/B(\alpha, \beta)$. (For, $v(x)$ is

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

proportional to $(1 - x^2/N^2)^\alpha$ on $[-N, N]$ so that $q(x) = (1 - x^2/N^2)^{\alpha-1}/d_0$, where $d_0 = 2\int_0^N (1 - x^2/N^2)^{\alpha-1}\, dx = 2Nc_\gamma$ and

$$c_\gamma = \int_0^1 (1 - y^2)^{\alpha-1} dy = \int_0^{\pi/2} (\cos\theta)^\gamma d\theta = 2^\gamma B(\alpha, \alpha) = 2^\gamma \Gamma(\alpha)^2/\Gamma(2\alpha). \quad \textbf{(2.3)}$$

The result now follows.)

3. MINIMIZING THE HELLINGER DISTANCE

Again, we begin with X a discrete random variable.

THEOREM 3.1: *Consider a random variable X taking integer values. Set* $p_i = P(X = i)$. *Let Y be a symmetric random variable taking integer values. Set* $q_i = P(Y = i)$. *So* $q_{-i} = q_i$. *Then the Hellinger distance* $H(q, p)$ *is minimized by*

$$q_i \equiv \alpha_i^2/d, \quad i \geq 0,$$

where $\alpha_i = (p_i^{1/2} + p_{-i}^{1/2})/2$ *and* $d = p_0 + 2\sum_{j\geq 1}^\infty \alpha_j^2$.

PROOF: We seek to minimize

$$2H(q, p)^2 = \sum_{i=-\infty}^{\infty} \left(q_i^{1/2} - p_i^{1/2}\right)^2 = 2 - 2q_0^{1/2}\alpha_0 - 4\sum_{i=0}^{\infty} q_i^{1/2}\alpha_i$$

such that $q_{-i} = q_i \geq 0$ and $1 = \sum_{i=-\infty}^\infty q_i$. Taking λ as a Lagrange multiplier and differentiating $2H(q, p)^2 + \lambda(q_0 + 2\sum_{i=1}^\infty q_i)$ with respect to q_i gives for $i \geq 0$, $q_i \equiv \alpha_i^2/d$, where $d = \lambda^2$. However,

$$d = d\left(q_0 + 2\sum_{i=1}^{\infty} q_i\right) = \alpha_0^2 + 2\sum_{j\geq 1}^{\infty} \alpha_j^2.$$

The result follows. ■

Example 3.1: Suppose that X is a nonnegative random variable. Then $\alpha_0 = p_0^{1/2}$ and $\alpha_i = p_i^{1/2}/2$ for $i \geq 1$. So $q_0 = p_0/d$ and $q_i = p_i/(4d)$ for $i \geq 1$, where $d = P(X = 0) + 2^{-1}P(X > 0) = (1 + p_0)/2$. So, if X is a positive random variable, then $q_i \equiv p_i/2$. In contrast, we saw that for this case there does not exist a solution using the Kulback–Liebler distance.

Note 2.2 again applies. Now, suppose that X is continuous.

THEOREM 3.2: *Suppose that X and Y are continuous random variables with densities* $p(x)$ *and* $q(x)$, *respectively, with respect to the Lebesgue measure. Suppose that* $q(x)$

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

is symmetric about zero. Then the density $q(x)$ that minimizes the Hellinger distance $H(q,p)$ is

$$q(x) \equiv \alpha(x)^2/d,$$

where $\alpha(x) = (p(x)^{1/2} + p(-x)^{1/2})/2$ *and* $d = 2\int_0^\infty \alpha(x)^2\,dx$.

PROOF: The result follows by a calculus of variations argument, or more simply from Theorem 3.1 by a limiting argument. ■

Example 3.2: Suppose that X_0 is a nonnegative random variable with density $p_0(x)$ with respect to the Lebesgue measure and that $X = X_0 - N$, where $N > 0$. Then Y is a random variable on $[-N,N]$ with density $q(x) = \alpha(x)^2/d$, where $\alpha(x) = [p_0(N + x)^{1/2} + p_0(N - x)^{1/2}]/2$ and $d = 2\int_0^N \alpha(x)^2\,dx$.

For example, if X_0 is exponential, then Y has a cosh density.

Example 3.3: Suppose that X_0 is an exponential random variable with mean 1 and that $X = X_0 - N$, where $N > 0$. Then Y is distributed on $[-N,N]$ with density $q(x) = (\cosh x + 1)/d_0$, where $d_0 = N + \sinh N$. (For, $|x| < N$, $\alpha(x) = \exp(-N/2)\cosh(x/2)$, $q(x) = \exp(-N)\cosh(x/2)^2/d$, and $d = \exp(-N)d_0$, where $d_0 = 2\int_0^N \cosh(x/2)^2\,dx = \int_0^N (\cosh x + 1)\,dx = N + \sinh N$.)

Example 3.4: Suppose that X_0 is a gamma random variable with mean $\gamma > 0$ and that $X = X_0 - N$, where $N > 0$. Then Y is distributed on $[-N,N]$ with density

$$q(x) = \alpha_0(x)^2/d_0,$$

where

$$\alpha_0(x) = (N + x)^{\gamma-1}\exp(-x) + (N - x)^{\gamma-1}\exp(x) + 2(N^2 - x^2)^{(\gamma-1)/2},$$

$$d_0 = \exp(-N)\int_0^{2N} y^{\gamma-1}d^{-y}\,dy + N^\gamma c_\gamma$$

and c_γ is given by (2.3). ($2\alpha(x) = \exp(-N/2)\Gamma(\gamma)^{-1/2}\alpha_0(x)$ so that $2d = \exp(-N)\Gamma(\gamma)\,d_0$, where $d_0 = \int_0^N \alpha_0(x)^2\,dx = d_1 + d_2 + 2d_3$, $d_1 = \exp(N)\int_N^{2N} y^{\gamma-1}d^{-y}\,dy$, $d_2 = \exp(N)\int_0^N y^{\gamma-1}d^{-y}\,dy$ and $d_3 = n^\gamma c_\gamma$.)

4. SAMPLE VERSIONS

Suppose that we have observations $X_1,\ldots,X_n$ that we wish to symmetrize. What happens when we apply the methods of Sections 2 and 3 with $p(x)$ replaced by the sample density? Denote their empirical distribution by $F_n(x)$. Let $I(A)$ be the indicator function, equal to 1 or 0 for A true or false, respectively.

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

First, suppose that $X_1, \ldots, X_n$ take on only a countable set of distinct values $\{x_i\}$ in some space S in which $\{-x_i\}$ are defined. Set

$$N_i = \sum_{j=1}^{n} I(X_j = x_i), \quad p_i = N_i/n.$$

Then conditional on the observations, Theorems 2.1 and 3.1 give symmetric random variables Y with $q_i = P(Y = x_i)$. For Theorem 3.1, Y is in terms of $\alpha_i = (N_i)^{1/2} + N_{-i}^{1/2})/(2n)$ provided that $d > 0$, which is always true.

For Theorem 2.1, Y is in terms of $\nu_i = (N_i N_{-i})^{1/2}/n$ provided that $d > 0$. This is not an empirical distribution since $n\nu_i$ are generally not integers, although one could approximate it by the empirical distribution with M_i observations at x_i, where M_i is the nearest integer to nq_i.

THEOREM 4.1: *We have*

$$P(d = 0) = \sum\{(p_{S_1 1} + p_{S_2 2} + \cdots)^n \ : \ S_1, S_2, \ldots = \pm 1\} = NP^n(1 + o(1))$$

as $n \to \infty$, *where*

$$P = \max\{p_{S_1 1} + p_{S_2 2} + \cdots \ : \ S_1, S_2, \ldots = \pm 1\}$$

and

$$N = \sum\{I(p_{S_1 1} + p_{S_2 2} + \cdots = P) \ : \ S_1, S_2, \ldots = \pm 1\}.$$

PROOF: Consider the case where the possible values are $x_i = -2, -1, 0, 1, 2$. Then $d = 0$ if and only if either $N_{-2} + N_{-1} = n$, or $N_{-2} + N_1 = n$, or $N_{-1} + N_2 = n$, or $N_1 + N_2 = n$ and this occurs with probability $(p_{-2} + p_{-1})^n + (p_{-2} + p_1)^n + (p_{-1} + p_2)^n + (p_1 + p_2)^n$. Additionally, $\sum_i P_i^n = NP^n(1 + o(1))$, where $P = \max_i P_i$ and $N = \sum_i I(P_i = P)$. The proof of the general case is identical. ■

Now, consider the case when $N_i \equiv 1$. (This holds with probability 1 if the observations are a random sample from a continuous distribution.) Then using Theorem 2.1, with probability 1 $\nu_i = d = 0$ so that q_i does not exist. However, using Theorem 3.1, with probability 1, $\alpha_i = 1/(2n)$ and Y corresponds to the empirical distribution of $\{\pm X_1, \ldots, \pm X_n\}$.

5. MEAN SQUARE OPTIMAL RANDOM FUNCTIONS

Here, we follow up the idea in (1.1) to obtain a random function f that minimizes $E\ (Y - X)^2$, where $Y = f(X)$ and f is independent of X. We show that the solution is in fact (1.1) assuming that $m_2 < \infty$, where $m_r = E\ X^r$. Consider first the discrete

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

case of Theorem 2.1. Set $f_i = f(i)$. Set $r_j = P(f_i = j)$. We require that $r_{-j} = r_j$. Since $E\,(f_i - i)^2 = \sum_{j=-\infty}^{\infty}(j-i)^2 r_j$, we seek to choose $\{r_j\}$ to minimize

$$E\,(f(X) - X)^2 = \sum_{i=-\infty}^{\infty} p_i E\,(f_i - i)^2 = \sum_{j=-\infty}^{\infty} r_j R_j = R_0 + 2\sum_{j=1}^{\infty} r_j R'_j,$$

where

$$R_j = \sum_{i=-\infty}^{\infty} p_i (j-i)^2 = E\,(j-X)^2 = j^2 - 2jm_1 + m_2,$$

$$2R'_j = R_j + R_{-j} = 2j^2 + 2m_2,$$

and $1 = \sum_{j=-\infty}^{\infty} r_j = r_0 + 2\sum_{j=0}^{\infty} r_j$. Substituting for r_0 using the last equation, we seek to minimize

$$E\,(f(X) - X)^2 = m_2 + 2\sum_{j=1}^{\infty} j^2 r_j$$

subject to $r_j \geq 0$ and $(1 - r_0)/2 = \sum_{j=1}^{\infty} r_j \leq 1/2$. The solution is clearly $r_1 = 1/2$ and $r_j = 0$ for $j \neq 0$; that is, among the random function f such that $f(x)$ and $f(-x)$ have the same distribution, the one that minimizes $E\,(f(X) - X)^2$ is $f(x) = \pm x$ with probability $1/2$. The general case follows from the discrete case by a limiting argument.

6. MINIMIZING |Y − X|

Here, we suppose that we have a random sample of size n from X_0, say $\mathbf{X} = (X_1, \ldots, X_n)$. We want to choose $\mathbf{Y} = (Y_1, \ldots, Y_n)$ symmetrically distributed around zero in the sense that $\{Y_1, \ldots, Y_n\} = \{-Y_1, \ldots, -Y_n\}$, to minimize the distance $|\mathbf{Y} - \mathbf{X}|$.

For example, if $n = 2$, then Y_1 minimizing $(X_1 - Y_1)^2 + (X_2 + Y_1)^2$ is $Y_1 = X_1 - X_2$ giving $|\mathbf{X} - \mathbf{Y}|^2 = |\mathbf{X}|^2$.

If $n = 3$, then we need one Y value to be zero, say $Y_3 = 0$. Now, choose Y_1 to minimize $|\mathbf{X} - \mathbf{Y}|^2 = (X_1 - Y_1)^2 + (X_2 - Y_2)^2 + (X_3 - Y_3)^2$ with $Y_2 = -Y_1$ giving $Y_1 = X_1 - X_2$ as earlier and $|\mathbf{X} - \mathbf{Y}|^2 = |\mathbf{X}|^2$. Clearly, this minimum is not altered by a different choice of the zero Y_j.

Now, suppose that $n = 4$. We matching pairs in $\binom{4}{2} = 6$ ways. For example, taking $Y_2 = -Y_1$ and $Y_4 = -Y_3$, we minimize $|\mathbf{X} - \mathbf{Y}|^2 = (X_1 - Y_1)^2 + (X_2 + Y_1)^2 + (X_3 - Y_3)^2 + (X_4 + Y_4)^2$ with $Y_1 = X_1 - X_2$ and $Y_2 = X_3 - X_4$ again giving $|\mathbf{X} - \mathbf{Y}|^2 = |\mathbf{X}|^2$. Again, this minimum is not altered by a different choice of Y_j.

Clearly, these conditions also apply for general n; that is, for $n = 2N$ even, the choice $Y_j = X_{2j-1} - X_{2j}$, $j = 1, 2, \ldots, N$, and $Y_{j+N} = -Y_j$ will give the minimum $|\mathbf{X} - \mathbf{Y}|^2 = |\mathbf{X}|^2$. For $n = 2N$ even, the choice as earlier but with $Y_{2N+1} = 0$ again gives $|\mathbf{X} - \mathbf{Y}|^2 = |\mathbf{X}|^2$.

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

7. A STRONGER FORM OF SYMMETRY

The symmetry required so far is that $-Y$ has the same distribution as Y. This holds, for example, for the r-variate normal with mean 0. Here, we assume that $\mathbf{Y}$ lies in R^r, but we require the stronger symmetry that for $1 \le j \le r$, $\mathbf{Z}_j$ has the same distribution as $\mathbf{Y}$, where $\mathbf{Z}_j$ is $\mathbf{Y}$ with its jth element Y_j replaced by $-Y_j$. For convenience, let us call this *supersymmetry*. For example, for the r-variate normal with mean 0 this stronger condition only holds if its covariance is diagonal.

Supersymmetry implies that the density $q(\mathbf{x})$ is defined uniquely by its values in just one of the 2^r quadrants formed by R^r. For example, if $\mathbf{X}$ and $\mathbf{Y}$ take on only r-variate integer values, then

$$1 = q_0 + \sum_{\mathbf{i} \in N^r} q_{\mathbf{i}} = q_0 + 2^r \sum_{\mathbf{i} \in S_1} q_{\mathbf{i}},$$

where $S_1 = \{\mathbf{i} \in N^r : i_1 \ge 1, i_2 \ge 0, \ldots, i_r \ge 0\}$ is the first quadrant minus $\mathbf{0} \in N^r$. Similarly, (2.2) becomes

$$I(q,p) = q_0 \ln(q_0/p_0) + 2^r \sum_{\mathbf{i} \in S_1}^{\infty} q_{\mathbf{i}}(\ln q_{\mathbf{i}} - a_{\mathbf{i}}),$$

where

$$a_{\mathbf{i}} = 2^{-r} \sum_{s_1, \ldots, s_r = \pm 1} \ln p_{s_1 i_1, \ldots, s_r i_r}.$$

The analogue of Theorem 2.1 is as follows.

THEOREM 7.1: *Consider a random variable* $\mathbf{X}$ *taking vector integer values* $\mathbf{i} = (i_1, \ldots, i_r) \in N^r$. *Set* $p_{\mathbf{i}} = P(\mathbf{X} = \mathbf{i})$. *Let* $\mathbf{Y}$ *be a supersymmetric random variable taking vector integer values in* N^r. *Set* $q_{\mathbf{i}} = P(\mathbf{Y} = \mathbf{i})$. *Then the Kulback–Liebler distance* $I(q,p)$ *is minimized by*

$$q_{\mathbf{i}} \equiv \nu_{\mathbf{i}}/d, \qquad \mathbf{i} \ge 0,$$

where

$$\nu_{\mathbf{i}} = \left(\prod_{s_1, \ldots, s_r = \pm 1} p_{s_1 i_1, \ldots, s_r i_r} \right)^{1/(2r)}, \qquad d = p_0 + 2^r \sum_{\mathbf{j} \in S_1} \nu_{\mathbf{j}}.$$

PROOF: The proof is similar to that of Theorem 2.1 and is therefore omitted. ■

Similarly, we have the following analogue of Theorem 3.1.

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

THEOREM 7.2: *Consider random variables X and Y and probabilities* $p_{\mathbf{i}}$ *and* $q_{\mathbf{i}}$ *as in Theorem* 7.1. *Then the Hellinger distance* $H(q,p)$ *is minimized by*

$$q_{\mathbf{i}} \equiv \alpha_{\mathbf{i}}^2/d, \qquad \mathbf{i} \in S_1,$$

where

$$\alpha_{\mathbf{i}} = 2^{-r} \sum_{s_1,\ldots,s_r=\pm 1} p_{s_1 i_1,\ldots,s_r i_r}, \qquad d = p_0 + 2^r \sum_{\mathbf{j}\in S_1}^{\infty} \alpha_{\mathbf{j}}^2.$$

Continuous versions of these theorems can now be obtained as in Sections 2 and 3.

8. OPTIMIZING SUBJECT TO MONOTONICITY OR CONVEXITY

Another type of constraint is not symmetry but being monotone about zero, referred to as *double monotonicity*.

THEOREM 8.1: *Consider a random variable X taking integer values with* $p_i = P(X = i)$. *Let Y be a random variable taking integer values with* $q_i = P(Y = i)$ *convex on* $(-\infty, 0]$ *and convex on* $(0, \infty]$. *Then the Hellinger distance*

$$H^2 = 1 - \sum_i (p_i q_i)^{1/2},$$

such that $q_{i+1} - 2q_i + q_{i-1} \geq 0$ *for* $i \leq -1$ *and for* $i \geq 1$ *(or alternatively* q_i *non-increasing as* $|i|$ *increases) is minimized by*

$$q_i = \begin{cases} a_1 & \text{if } 0 \leq i < K_1 \\ a_2 & \text{if } -K_2 \leq i < 0, \end{cases} \tag{8.1}$$

where $A_j = a_j^{1/2}$ *are given by*

$$A_j = W_j K_j^{-1}/S^{1/2}, \tag{8.2}$$

where

$$S = \sum_{j=1}^{2} W_j^2 K_j^{-1}, \qquad W_1 = \sum_{i=0}^{K_1-1} p_i^{1/2}, \qquad W_2 = \sum_{i=-K_2}^{-1} p_i^{1/2}.$$

Furthermore, K_1 *and* K_2 *are chosen to maximize*

$$W_1/K_1 = \sum_{i=0}^{K_1-1} p_i^{1/2}/K_1 \tag{8.3}$$

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press

and

$$W_2/K_2 = \sum_{i=-K_2}^{-1} p_i^{1/2}/K_2, \tag{8.4}$$

respectively. We will refer to Y as a double monotonicity random variable corresponding to X.

PROOF: We first maximize $\sum_{i=0}^{\infty}(p_iq_i)^{1/2}$ such that $\sum_{i=0}^{\infty} q_i = Q_1$ say and q_i is nonincreasing for $i \geq 0$. This gives $q_i = a_1 I(0 \leq i < K_1)$, where $a_1K_1 = Q_1$. Similarly, we maximize $\sum_{i=-\infty}^{-1}(p_iq_i)^{1/2}$ such that $\sum_{i=-\infty}^{-1} q_i = Q_2 = 1 - Q_1$ and q_i is nondecreasing for $i < 0$. This gives $q_i = a_2 I(-K_2 \leq i < 0)$, where $a_2K_2 = Q_2$. So, (8.1) follows. Now, maximize $\sum_{j=1}^{2} A_jW_j$ such that $\sum_{j=1}^{2} a_jK_j = \sum_{j=1}^{2} A_j^2K_j = 1$ with respect to A_1 and A_2. By Lagrange multipliers, this gives $W_j \equiv \lambda A_jK_j$, and so (8.2) follows. Finally, K_1 and K_2 must be chosen to maximize $\sum_{j=1}^{2} A_jW_j = (\sum_{j=1}^{2} W_j^2/K_j)^{1/2}$; so (8.3) and (8.4) follow. ■

Finding the K_1 and K_2 in (8.3) and (8.4), respectively, is not trivial. A continuous version of Theorem 8.1 can be obtained as in Sections 2 and 3.

References

1. Al-Awadhi, F., Konsowa, M., & Najeh, Z. (2009). Commute times and the effective resistances of random trees. *Probability in the Engineering and Informational Sciences* 23: 649–660.
2. Azzalini, A. (1985). A class of distributions include the normal ones. *Scandinavian Journal of Statistics* 12: 171–178.
3. Coffman, E.G., Flajolet, P., Flatto, L., & Hofri, M. (1998). The maximum of a random walk and its application to rectangle packing. *Probability in the Engineering and Informational Sciences* 12: 373–386.
4. Fisher, N.I. (1993). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.
5. Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.
6. Kullback, S. & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* 22: 79–86.
7. Nikulin, M.S. (2001). Hellinger distance. In M. Hazewinkel (ed.), *Encyclopaedia of mathematics*. Amsterdam: Kluwer Academic.
8. Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman & Hall.

https://doi.org/10.1017/S0269964810000173 Published online by Cambridge University Press