Probab. Theory Relat. Fields 129, 391–409 (2004)
Digital Object Identifier (DOI) 10.1007/s00440-004-0344-0

Oliver Johnson · Andrew Barron

# Fisher information inequalities and the central limit theorem

Received: 7 July 2003 / Revised version: 22 January 2004 /
Published online: 29 April 2004 – 

**Abstract.** We give conditions for an $O(1/n)$ rate of convergence of Fisher information and relative entropy in the Central Limit Theorem. We use the theory of projections in $L^2$ spaces and Poincaré inequalities, to provide a better understanding of the decrease in Fisher information implied by results of Barron and Brown. We show that if the standardized Fisher information ever becomes finite then it converges to zero.

## 1. Introduction

Bounds on Shannon entropy and Fisher information have long been used in proofs of central limit theorems, based on quantification of the change in information as a result of convolution, as in the papers of Linnik (1959), Shimizu (1975), Brown (1982), Barron (1986) and Johnson (2000). Each of these papers have a final step involving completeness or uniform integrability in which a limit is taken without explicitly bounding the information distance from the normal distribution.

The purpose of the present paper is to provide an explicit rate of convergence of information distances, under certain natural conditions on the random variables. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with mean 0, variance $\sigma^2$ and density function $p(x)$, satisfying Poincaré conditions (relating $L^2$ norms of mean zero functions to $L^2$ norms of the derivative), and let $\phi_{\sigma^2}(x)$ be the corresponding $N(0, \sigma^2)$ density. The relative entropy distance is

$$D(X) = \int p(x) \log \left( \frac{p(x)}{\phi_{\sigma^2}(x)} \right) dx.$$

In the case of random variables with differentiable densities, the Fisher information distance is

O. Johnson: Statslab, Wilberforce Road, Cambridge, CB3 0WB, UK.
e-mail: otj1000@cam.ac.uk

A. Barron: Department of Statistics, Yale University, PO Box 208290, New Haven, Connecticut 06520-8290, USA. e-mail: andrew.barron@yale.edu

OTJ is a Fellow of Christ's College, Cambridge, who helped support two trips to Yale University during which this paper was written.

*Mathematics Subject Classification (2000):* Primary: 62B10 Secondary: 60F05, 94A17

*Key words or phrases:* Normal convergence – Entropy – Fisher information – Poincaré inequalities – Rates of convergence

$$J(X) = \sigma^2 \mathbb{E}\left(\frac{d}{dx}\log p(X) - \frac{d}{dx}\log \phi_{\sigma^2}(X)\right)^2$$

which is related to the Fisher information $I(X) = \mathbb{E}[(d/dx \log p(X))^2]$ via $J(X) = \sigma^2 I(X) - 1$. This is an $L^2$ norm between derivatives of log-densities, and gives a natural measure of convergence, not implied by existing central limit theorems. Note that the quantities $D$ and $J$ are scale-invariant, that is, $D(aX) = D(X)$ and $J(aX) = J(X)$ for all non-zero $a$.

Let $U_n = (X_1 + \cdots + X_n)/\sqrt{\sigma^2 n}$ be the standardized sum of the random variables. Theorem 1.3 shows that $D(U_n) \leq 2RD(U_1)/n\sigma^2$ for all random variables with Poincaré constant $R$, and that $J(U_n) \leq 2RJ(U_1)/n\sigma^2$ for all random variables with absolutely continuous density function and finite Poincaré constant.

In examination of the Fisher information a central role is played by the score function $\rho(y) = (d/dy)\log p(y) = p'(y)/p(y)$. The score function of the sum of independent random variables can be expressed in terms of the score function of the individual random variables, via a conditional expectation, as has been used in demonstration of convolution inequalities for Fisher information and Shannon entropy (in the work of Stam (1959), Blachman (1965), and others).

In particular, if $Y_1$ and $Y_2$ are independent and identically distributed with score function $\rho$ then the score $\overline{\rho}(u)$ of the sum $Y_1 + Y_2$ is the projection of $(\rho(Y_1) + \rho(Y_2))/2$ onto the linear space of functions of $Y_1 + Y_2$, so by the Pythagorean identity and rescaling:

$$\frac{I(Y_1) + I(Y_2)}{2} - I\left(\frac{Y_1 + Y_2}{\sqrt{2}}\right) = 2\mathbb{E}\left(\overline{\rho}\,(Y_1 + Y_2) - \frac{\rho(Y_1) + \rho(Y_2)}{2}\right)^2 \quad (1)$$

(see Lemma 3.1 for details). Hence, since Equation (1) is positive, one deduces that the Fisher information is decreasing on the powers of two subsequence $U_{2^k}$ and hence convergent (as is the whole sequence $I(U_n)$ by subadditivity of $nI(U_n)$). Thus the difference sequence $I(U_{2^k}) - I(U_{2^{k+1}})$ tends to zero and the right side of Equation (1) is used to characterize this difference, which becomes the object of interest in identifying the normal limit.

Work that follows this approach includes Shimizu (1975), Brown (1982) and Barron (1986). However, in these papers the Fisher information is only examined for random variables for which there is added a possibly small independent normal perturbation. Previously, identification of the Fisher information limit for general random variables with finite Fisher information was unresolved. Continuing with the examination of Equation (1), we aim to establish the general Fisher information limit, and, in certain settings, to have explicit bounds on the distance from the limit.

As we have noted, the difference sequence tends to zero. Thus interest is in random variables $Y_1, Y_2$, with score functions for which the right side of Equation (1) is small. This expression measures the squared $L^2$ difference between a 'ridge function' (a function of the sum $Y_1 + Y_2$) and an additive function (a function of the form $g_1(Y_1) + g_2(Y_2)$). From calculus, in general, the only functions $f(y_1, y_2) = g_1(y_1) + g_2(y_2)$ that are both ridge and additive are the linear functions $g_1(y_1) = ay_1 + b_1$, $g_2(y_2) = ay_2 + b_2$, with $a, b_1, b_2$ constants, that is, the functions for which the derivatives $g_i'(y)$ are constant and equal.

Previous work, as in Lemma 3.1 of Brown (1982), (see also Barron (1986)) established the following.

**Lemma 1.1.** *For any $L^2$ functions $f$ and $g$ there exist some a,b such that:*

$$\mathbb{E}(g(Y_1) - aY_1 - b)^2 \leq \mathbb{E}\left(f(Y_1 + Y_2) - g(Y_1) - g(Y_2)\right)^2,$$

*when $Y_1, Y_2$ are independent identically distributed normals.*

Brown takes $g \in L^2(\phi)$ and considers the projection $f(s) = \mathbb{E}(g(Y_1) + g(Y_2)|Y_1 + Y_2 = s)$. For $Y_1, Y_2$ normal, the eigenfunctions of this projection are the Hermite polynomials, so he can use expansions in this orthogonal Hermite basis.

The main technique used in the present paper will generalize Lemma 1.1 to a wider class of random variables $Y_1, Y_2$. We consider random variables $Y_1$ and $Y_2$ independent identically distributed (IID) with absolutely continuous densities and finite Fisher information. The method used to prove Proposition 2.1 implies for certain ridge functions $f(y_1 + y_2)$, with closest additive function $g(y_1) + g(y_2)$ and a certain constant $\mu$, that:

$$\mathbb{E}(g'(Y_1) - \mu)^2 \leq I(Y_2)\mathbb{E}\left(f(Y_1 + Y_2) - g(Y_1) - g(Y_2)\right)^2. \tag{2}$$

Our (basis-free) proof starts with $f(Y_1 + Y_2)$, finds its additive part with $g(y_1) = \mathbb{E}_{Y_2} f(y_1 + Y_2)$ and recognises that $g'(y_1) = -\mathbb{E}_{Y_2} f(y_1 + Y_2)\rho(Y_2)$. A Cauchy-Schwarz inequality completes the proof as detailed in Section 2.

Hence if Equation (1) is small then $\rho$ is close to a function with derivative close to constant in $L^2(Y_1, Y_2)$. However, we would like to find an inequality where the left side depends on $g$ itself, rather than $g'$. Poincaré inequalities provide a relationship between $L^2$ norms on functions and the $L^2$ norms on derivatives.

**Definition 1.2.** *Given a random variable $Y$, define the Poincaré constant $R_Y$:*

$$R_Y = \sup_{g \in H_1(Y)} \frac{\mathbb{E}g^2(Y)}{\mathbb{E}g'(Y)^2},$$

*(where $H_1(Y)$ is the space of absolutely continuous functions $g$ such that* Var $g(Y) > 0$, $\mathbb{E}g(Y) = 0$ *and $\mathbb{E}g^2(Y) < \infty$), and the restricted Poincaré constant $R_Y^*$:*

$$R_Y^* = \sup_{g \in H_1^*(Y)} \frac{\mathbb{E}g^2(Y)}{\mathbb{E}g'(Y)^2},$$

*where $H_1^*(Y) = H_1(Y) \cap \{g : \mathbb{E}g'(Y) = 0\}$.*

For certain $Y$, $R_Y$ is infinite. However, $R_Y$ is finite for the normal and other log-concave distributions (see for example Klaasen (1985), Cacoullos (1982) and Borovkov and Utev (1984)). Since we maximise over a smaller set of functions, $R_Y^* \leq R_Y$. Further, for $Z \sim N(0, \sigma^2)$, $R_Z = \sigma^2$ and $R_Z^* = \sigma^2/2$, with $g(x) = x$ and $g(x) = x^2 - \sigma^2$ respectively achieving these values (one can show this by expanding $g$ in the Hermite basis).

Using Poincaré inequalities, extensions of Brown's inequality Lemma 1.1 hold (with a constant depending on $I(Y_1)$ and $R_{Y_1}$) for a wider class of random variables

than just normals. Since linear score functions correspond to the family of normal distributions, Equations (1) and (2) provide a means to prove the following Central Limit Theorems.

**Theorem 1.3.** *Given $X_1, X_2, \ldots$ IID and with finite variance $\sigma^2$, define the normalized sum $U_n = (\sum_{i=1}^n X_i)/\sqrt{n\sigma^2}$.*

*If $X_i$ have an absolutely continuous density with finite restricted Poincaré constant $R^*$ then*

$$J(U_n) \leq \frac{2R^*}{2R^* + (n-1)\sigma^2} J(X_1) \text{ for all } n. \tag{3}$$

*If $X_i$ have a density with finite Poincaré constant $R$, then*

$$D(U_n) \leq \frac{2R}{2R + (n-1)\sigma^2} D(X_1) \text{ for all } n. \tag{4}$$

*Proof.* Note that if $J(X_1)$ is infinite then (3) is trivially true, and similarly for $D(X_1)$ and (4). See Sections 2 and 3 for the proof of the Fisher information bound (3). Notice that for $X$ normal, $2R^* = \sigma^2$, so the 'closer to normal $X$ is', the closer the bound becomes to $J(X)/n$.

The relative entropy bound (4) is a corollary. Using an integral form of the de Bruijn identity (Lemma 1 of Barron (1986)), the relative entropy satisfies

$$D(X) = \int_0^1 \frac{J(\sqrt{t}X + \sqrt{1-t}Z)}{2t} dt, \tag{5}$$

where $Z$ is a normal independent of $X$, with the same mean and variance as $X$. Now, if $X$ has finite Poincaré constant $R$, then for each $t$, parts (v) and (vii) of Theorem 2 of Borovkov and Utev (1984) show that the $(\sqrt{t}X + \sqrt{1-t}Z)$ itself has Poincaré constant $\leq tR + (1-t)\sigma^2 \leq R$ (since part (vi) of the same Theorem gives that $\sigma^2 \leq R$). Moreover, for all $0 < t < 1$ the resulting density is absolutely continuous (even if the density of $X$ is not), so using $R^* \leq R$, expressions (3) and (5) imply the bound (4). □

*Note:* Instead of requiring $X_1$ to satisfy the stated conditions, it is enough for such bounds that the conditions are satisfied after some number of convolutions. Indeed, if $U_k$ has finite $J(U_k)$ and $R_{U_k}$ for some $k$ then for all $n \geq k$, we have $J(U_n) \leq (2R_{U_k} J(U_k) + 1)/\lfloor n/k \rfloor$.

The $O(1/n)$ rate of convergence of Theorem 1.3 is perhaps to be expected. For example if $X_i$ is exponentially distributed, and hence $U_n$ has a $\Gamma(n)$ distribution, then $J(U_n) = 2/(n-2)$, consistent with this. In fact, by extending the Cramér-Rao inequality we deduce the following lower bound.

**Lemma 1.4.** *Let $X_1, X_2, \ldots, X_n$ be IID with finite fourth moment and assume that their standardized sum $U_n$ has an absolutely continuous density and finite Fisher information. Writing $m_r(X)$ for the centered $r$th moment of $X$, then defining the skewness $s = m_3(X)/m_2(X)^{3/2}$ and excess kurtosis $k = m_4(X)/m_2(X)^2 - 3$:*

$$J(U_n) \geq \frac{s^2}{2n + k}.$$

*Proof.* The positivity of $\mathbb{E}(\rho_Y(Y) + f(Y))^2$ implies that

$$I(Y) = \mathbb{E}\rho_Y(Y)^2 \geq \mathbb{E}(2f'(Y) - f(Y)^2), \tag{6}$$

giving a whole family of bounds for random variables $Y$ and functions $f$ satisfying the identity $\mathbb{E}[\rho(Y)f(Y)] = -\mathbb{E}[f'(Y)]$ which we use here for linear and quadratic $f$. (This identity may be thought of as an integration by parts, valid under integrability conditions in the case that $p \times f$ is absolutely continuous – see also Lemma A.1 in the Appendix).

If in inequality (6) one takes a random variable $Y$ with $\mathbb{E}Y = 0$ and sets $f(y) = y/\sigma^2$ where $\sigma^2 = m_2(Y)$ is the variance of $Y$, then one deduces $I(Y) \geq 1/\sigma^2$, which we refer to as the Cramér-Rao bound.

A stronger lower bound is obtained in (6) by taking $f(y) = y/\sigma^2 + a(y^2 - \sigma^2)$. Then choosing the optimal $a$, which is $a = -m_3(Y)/(m_2(Y)(m_4(Y) - m_2(Y)^2))$, we deduce that for any $Y$

$$J(Y) = m_2(Y)I(Y) - 1 \geq \frac{m_3(Y)^2}{m_2(Y)(m_4(Y) - m_2(Y)^2)}.$$

Now taking $Y = U_n$, we relate its moments to the moments of $X$. Indeed, $m_2(U_n) = 1$, $m_3(U_n) = m_3(X)/(m_2(X)^{3/2}\sqrt{n}) = s/\sqrt{n}$ and $m_4(U_n) = m_4(X)/(nm_2(X)^2) + 3(n-1)/n = (k/n) + 3$. The result follows. □

Further, this $O(1/n)$ convergence is consistent with estimates of Berry–Esseen type which give a $O(1/\sqrt{n})$ rate of weak convergence. The following lemma shows the relationship between convergence in Fisher information and several weaker forms of convergence.

**Lemma 1.5.** *If $X$ is a random variable with density $f$, and $\phi$ is a standard normal, then:*

$$\sup_x |f(x) - \phi(x)| \leq \left(1 + \sqrt{\frac{6}{\pi}}\right)\sqrt{J(X)},$$

$$\int |f(x) - \phi(x)|dx \leq 2d_H(f, \phi) \leq \sqrt{2}\sqrt{J(X)},$$

*where $d_H(f, \phi)$ is the Hellinger distance $\left(\int |\sqrt{f(x)} - \sqrt{\phi(x)}|^2 dx\right)^{1/2}$.*

*Proof.* The first bound comes from Shimizu (1975). The second inequality tightens a bound of Shimizu. Since:

$$\sqrt{\phi(x)}\frac{\partial}{\partial x}\sqrt{\frac{f(x)}{\phi(x)}} = \frac{1}{2}\left(\frac{f'(x)}{\sqrt{f(x)}} + x\sqrt{f(x)}\right),$$

we deduce from the Poincaré inequality for $\phi$ that:

$$J(X) = 4\int \phi(x)\left(\frac{\partial}{\partial x}\sqrt{\frac{f(x)}{\phi(x)}}\right)^2 \geq 4\int \phi(x)\left(\sqrt{\frac{f(x)}{\phi(x)}} - \mu\right)^2 = 4(1 - \mu^2),$$

where $\mu = \mathbb{E}_\phi\sqrt{f/\phi}$, so $2d_H^2(f, \phi) = 2(2 - 2\mu) \leq 4(1 - \mu^2)$. □

Recent work by Ball et al. (2002) has also considered the rate of convergence of information. Their paper obtains similar results, but by a very different method, involving transportation costs and a variational characterisation of Fisher information.

Unfortunately, Poincaré constants are not finite for all distributions $Y$. Indeed, as Borovkov and Utev (1984) point out, if $R_Y < \infty$, then by considering $g_n(x) = |x|^n$, we inductively deduce that all the moments of $Y$ are finite. From the Berry-Esseen Theorem (see for example Theorem 5.7 of Petrov (1995)) we know that only $(2+\delta)$th moment conditions are enough to ensure an explicit $O(1/n^{\delta/2})$ rate of weak convergence, for $0 < \delta \leq 1$. In Section 4 we describe a proof of Fisher information convergence under only second moment conditions, though without an explicit rate.

**Theorem 1.6.** *Given $X_1, X_2, \ldots$ IID with finite variance $\sigma^2$, define the normalized sum $U_n = (\sum_{i=1}^n X_i)/\sqrt{n\sigma^2}$. If $U_m$ has an absolutely continuous density and $J(U_m)$ is finite for some $m$ then*

$$\lim_{n\to\infty} J(U_n) = 0.$$

*Note:* This extends Lemma 2 of Barron (1986), which only holds when $X$ is of the form $Y + Z_\tau$.

Aside from the intrinsic interest that these theorems give by offering a strong form of the Central Limit Theorem, they can be used in problems ranging from quantum probability to statistics. Indeed, in estimation of the shift parameter of a quantum state, Theorem 1.6 is precisely what is used to prove Theorem 3 of Holevo (2003).

Another application is the demonstration of risk efficiency of certain parameter estimators, in particular, the best unbiased estimators of natural parameters in general exponential families. Indeed, suppose $X_i$ are IID real-valued with density function of the form $p(x|\eta) = e^{\eta x} h(x)/c_\eta$, with $\eta$ in the interior of the interval in which the normalizing constant $c_\eta$ is finite. Two score functions and associated Fisher informations arise in estimation. On one hand $(d/d\eta)\log p(x|\eta) = x - \mu$ where $\mu$ is the mean of $X$, so the Fisher information $I_\eta$ for $\eta$ is equal to the variance of $X$. On the other hand, the sum $S_n = X_1 + \cdots + X_n$ has a density function of a similar form $p_n(s|\eta) = e^{\eta s} h_n(s)/c_\eta^n$, for which $\hat{\eta}_n = -(d/ds)\log h_n(s)$ is the best unbiased estimator of $\eta$ (indeed it is the only unbiased estimator that is a function of the complete sufficient statistic $S_n$, see Casella and Berger (1990) pages 88,243–244). Consequently $\rho_n(s) = (d/ds)\log p_n(s|\eta)$ is the error $\eta - \hat{\eta}_n$, and its expected square $I(S_n)$ is equal to the mean squared error. In terms of the standardized information one has $I(S_n) = (1 + J(U_n))/(nI_\tau)$. Thus, in this setting, $J(U_n)$ characterizes the gap in the information inequality $\mathbb{E}(\hat{\eta}_n - \eta)^2 \geq 1/(nI_\tau)$ and, if this mean squared error is finite for some $n$, the following risk efficiency holds

$$\mathbb{E}(\hat{\eta}_n - \eta)^2 = \frac{1}{nI_\tau}(1 + o(1)).$$

Indeed, it is equivalent to our central limit theorem (Theorem 1.6).

## 2. Projection of ridge functions in $L^2$

Although the main application of the following Proposition will concern score functions, we present it as an abstract result concerning projection of ridge functions in $L^2(Y_1, Y_2)$ onto the space of additive functions. First note that for any $f$ with $\mathbb{E}f^2(Y_1 + Y_2)$ finite and $\mathbb{E}f(Y_1 + Y_2) = 0$ with $Y_1$ and $Y_2$ independent, if we form $g_1(u) = \mathbb{E}_{Y_2} f(u + Y_2)$ and $g_2(v) = \mathbb{E}_{Y_1} f(Y_1 + v)$, then $g_1(Y_1) + g_2(Y_2)$ is the projection onto the space of additive functions. Indeed, one has the Pythagorean relation

$$\begin{aligned}\mathbb{E}(f(Y_1 + Y_2) - h_1(Y_1) - h_2(Y_2))^2 \\ = \mathbb{E}(f(Y_1 + Y_2) - g_1(Y_1) - g_2(Y_2))^2 \\ + \mathbb{E}(g_1(Y_1) - h_1(Y_1))^2 + \mathbb{E}(g_2(Y_2) - h_2(Y_2))^2. \end{aligned} \tag{7}$$

**Proposition 2.1.** *Consider independent random variables $Y_1, Y_2$ with absolutely continuous densities and restricted Poincaré constants $R_1^*$ and $R_2^*$. Consider a function $f$ such that $\mathbb{E}f(Y_1 + Y_2)^2$ is finite and $\mathbb{E}f(Y_1 + Y_2) = 0$. Let $g_1(u) = \mathbb{E}_{Y_2} f(u + Y_2)$ and $g_2(v) = \mathbb{E}_{Y_1} f(Y_1 + v)$. There exist constants $\mu$, $\nu_1$ and $\nu_2$ such that for any $\beta \in [0, 1]$:*

$$\begin{aligned}\mathbb{E}\left(f(Y_1 + Y_2) - g_1(Y_1) - g_2(Y_2)\right)^2 \\ \geq \frac{1}{\overline{I}}\left(\frac{\beta}{R_1^*}\mathbb{E}\left(g_1(Y_1) - \mu Y_1 - \nu_1\right)^2 + \frac{(1-\beta)}{R_2^*}\mathbb{E}\left(g_2(Y_2) - \mu Y_2 - \nu_2\right)^2\right),\end{aligned}$$

*where $\overline{I} = (1 - \beta)I(Y_1) + \beta I(Y_2)$.*

*Proof.* We may assume that $\overline{I}$ is finite, otherwise the desired inequality is trivial. Having removed the additive part of $f$, we hope that what remains will be small in magnitude and we control its inner product with certain functions of the variables. Specifically we define

$$\begin{aligned} r_1(u) &= \mathbb{E}_{Y_2}\left[\left(f(u + Y_2) - g_1(u) - g_2(Y_2)\right)\rho_2(Y_2)\right], \\ r_2(v) &= \mathbb{E}_{Y_1}\left[\left(f(Y_1 + v) - g_1(Y_1) - g_2(v)\right)\rho_1(Y_1)\right], \end{aligned}$$

and show that we can control their norms. Indeed, by Cauchy-Schwarz, for any $u$:

$$r_1^2(u) \leq \mathbb{E}_{Y_2}\left(f(u + Y_2) - g_1(u) - g_2(Y_2)\right)^2 \mathbb{E}\rho_2^2(Y_2),$$

so taking expectations over $Y_1$, we deduce that

$$\mathbb{E}r_1^2(Y_1) \leq \mathbb{E}\left(f(Y_1 + Y_2) - g_1(Y_1) - g_2(Y_2)\right)^2 I(Y_2). \tag{8}$$

Also, we see that $\mathbb{E}r_1(Y_1) = 0$, since when $I(Y_2)$ is finite we have that $\mathbb{E}\rho_2(Y_2) = 0$ (one of the properties of score functions reviewed in Lemma A.1 of the appendix). Similarly,

$$\mathbb{E}r_2^2(Y_2) \leq \mathbb{E}\left(f(Y_1 + Y_2) - g_1(Y_1) - g_2(Y_2)\right)^2 I(Y_1). \tag{9}$$

Now in examining the function $r_1(u)$ further, we would like to exchange limits to see that $g_1(u) = \mathbb{E}[f(u+Y_2)]$ has derivative $g_1'(u) = -\mathbb{E}f(u+Y_2)\rho_2(Y_2)$, that is, $\int f(s)p(s-u)ds$ has derivative $\int f(s)p'(s-u)ds$, and we would like this $g_1$ to be absolutely continuous. One may achieve this exchange and the absolute continuity in certain cases, such as bounded $f$ (again see Lemma A.1 in the appendix). Hence we first assume that $f$ is bounded. We will relax that assumption at the end of the proof.

Setting $\mu_1 = -\mathbb{E}g_2(Y_2)\rho_2(Y_2) = -\mathbb{E}f(Y_1 + Y_2)\rho_2(Y_2)$, we recognize that $r_1(u)$ defined above simplifies to

$$r_1(u) = -\left(g_1'(u) - \mu_1\right).$$

Now $g_1(y_1) - \mu_1 y_1 - \nu_1$ has mean zero (with $\nu_1 = \mu_1 \mathbb{E}Y_1$), is absolutely continuous, and has derivative $-r_1(y_1)$, enabling use of the Poincaré inequality.

Likewise, use $r_2(v) = -(g_2'(v) - \mu_2)$, with $\mu_2 = -\mathbb{E}g_1(Y_1)\rho_1(Y_1) = -\mathbb{E}f(Y_1 + Y_2)\rho_1(Y_1)$. In fact, $\mu_1$ and $\mu_2$ are equal (using Lemma 3.1 they both equal $-\mathbb{E}f(S)\overline{\rho}(S)$ where $S = Y_1 + Y_2$), so we refer to them both as $\mu$.

Adding $\beta$ times Equation (8) to $(1-\beta)$ times Equation (9), and using the Poincaré inequalities, we deduce the result for bounded functions $f$.

We can deal with the general case of $f \in L^2$ with $\mathbb{E}f(S) = 0$, by considering the truncation $f(s)\mathbb{I}(|f(s)| \leq c)$ and subtracting its mean $a_c$ to give $f_c(s) = f(s)\mathbb{I}(|f(s)| \leq c) - a_c$. This gives absolutely continuous functions of the form $g_c(u) = \mathbb{E}[f_c(u+Y)]$ with $Y$ either $Y_1$ or $Y_2$, for which the lower bound holds in terms of the $L^2$ norms of such $g_c$, using the constant $\mu_c = -\mathbb{E}f_c(S)\bar{\rho}(S)$. Then, by Cauchy-Schwarz, the $L^2$ norms of $f - f_c$ and $g - g_c$ are all less than twice the $L^2$ norm of $f(s)\mathbb{I}(|f(s)| > c)$, which tends to zero as $c \to \infty$, and likewise $\mu - \mu_c$ also tends to zero, as $c \to \infty$. So the desired inequality holds for general $f \in L^2$ with $\mathbb{E}f(S) = 0$. □

*Note:* this inequality holds in general, for any $Y_1, Y_2$ with finite Fisher information, whereas previous such expressions have only held in the case of $Y_i \sim U_i + Z_\tau$, for some $U_i$ and for $Z_\tau$ a $N(0,\tau)$ independent of $U_i$.

*Note:* this inequality allows for independent random variables that are not identically distributed. Armed with it, one may provide Central Limit Theorems giving information convergence to the normal for random variables satisfying a uniform Lindeberg-type condition (see also Johnson (2000)). In certain cases we can provide a rate of convergence.

*Note:* we can produce a similar expression using a similar method for finite-dimensional random vectors $\mathbf{Y}_1, \mathbf{Y}_2$, where $\rho_i = (\partial/\partial x_i)(\log p(x))$ will be the $i$th component of the score vector function $\boldsymbol{\rho}$. Similar analysis in this case can give an alternative proof of the Theorems in Johnson and Suhov (2001).

## 3. Rate of convergence

In this section, we prove Theorem 1.3. If $Y_1, Y_2$ have finite restricted Poincaré constants $R_1^*, R_2^*$ then we can extend Lemma 1.1 from the case of normal $Y_1, Y_2$

to more general distributions, providing an explicit rate of convergence of Fisher information. We can apply Proposition 2.1 because the score functions of sums can be expressed as $L^2$ projections.

**Lemma 3.1.** *Let $S = Y_1 + Y_2$ where $Y_1$ and $Y_2$ are independent and suppose $Y_2$ has an absolutely continuous density with score function $\rho_2$ with finite Fisher information $I(Y_2) = \mathbb{E}\rho_2^2(Y_2)$. Then $S$ has an absolutely continuous density with score function*

$$\overline{\rho}(s) = \mathbb{E}[\rho_2(Y_2)|S = s], \text{ and } I(S) \leq I(Y_2).$$

*Moreover for independent random variables $Y_1$ and $Y_2$ with absolutely continuous densities and score functions $\rho_1$ and $\rho_2$, writing $\overline{\rho}$ for the score function of $S$:*

$$\frac{I(Y_1) + I(Y_2)}{2} - I\left(\frac{Y_1 + Y_2}{\sqrt{2}}\right) = 2\mathbb{E}\left(\overline{\rho}(S) - \frac{\rho_1(Y_1) + \rho_2(Y_2)}{2}\right)^2.$$

*Proof.* First, by the device of Lemma A.1 in the appendix, to show that the proposed $\overline{\rho}(s)$ is the score function (and that $S$ has an absolutely continuous density) we show for every bounded test function $T(u + S)$ that $\mathbb{E}T(u + S)$ has derivative $-\mathbb{E}[T(u + S)\bar{\rho}(S)]$. For any such bounded $T$ define $T_2(v) = \mathbb{E}T(v + Y_1)$ so that $T_2(u + Y_2) = \mathbb{E}[T(u + S)|Y_2]$. Then, since the indicated property holds for $Y_2$ we have that $\mathbb{E}[T(u + S)\overline{\rho}(S)] = \mathbb{E}[T(u + S)\rho_2(Y_2)] = \mathbb{E}[T_2(u + Y_2)\rho_2(Y_2)] = -(d/du)\mathbb{E}T_2(u + Y_2) = -(d/du)\mathbb{E}T(u + S)$.

Secondly, if both random variables have the indicated properties, then $\overline{\rho} = \mathbb{E}[(\rho_1(Y_1) + \rho_2(Y_2))/2|S = s]$. Thus by the Pythagorean identity, the result follows, on rescaling: $\rho_{aX}(x) = \rho_X(x/a)/a$ and $J(aX) = J(X)/a^2$. □

**Proposition 3.2.** *Consider $Y_1, Y_2$ IID with absolutely continuous densities, variance $\sigma^2$ and restricted Poincaré constant $R^*$. Then*

$$J\left(\frac{Y_1 + Y_2}{\sqrt{2}}\right) \leq J(Y_1)\left(\frac{2R^*}{\sigma^2 + 2R^*}\right).$$

*Proof.* The claim is trivial if $J$ is infinite, and is merely $J((Y_1 + Y_2)/\sqrt{2}) \leq J(Y_1)$ if $R^*$ is infinite (which is covered by Lemma 3.1). So suppose now that $J$ and $R^*$ are finite.

Without loss of generality, suppose $Y_i$ have mean 0 and variance 1, since we can just rescale, using $R^*_{aX} = a^2 R^*_X$. Write $J$ and $I$ for the standardardized and non-standardized Fisher information of $Y$, and $J'$ and $I'$ for the corresponding quantities for $(Y_1 + Y_2)/\sqrt{2}$.

Let $f(s) = 2\overline{\rho}(s) = \sqrt{2}\widetilde{\rho}((Y_1 + Y_2)/\sqrt{2})$, where $\overline{\rho}$ is the score of the sum $S = Y_1 + Y_2$ and $\tilde{\rho}$ is the score of the standardized sum $(Y_1 + Y_2)/\sqrt{2}$, and let $g(u) = \mathbb{E}f(u + Y_2)$. By Lemma 3.1 and (7):

$$\begin{aligned} &J(Y_1) - J\left(\frac{Y_1 + Y_2}{\sqrt{2}}\right) \\ &\quad = \mathbb{E}\left(\widetilde{\rho}\left(\frac{Y_1 + Y_2}{\sqrt{2}}\right) - \frac{g(Y_1) + g(Y_2)}{\sqrt{2}}\right)^2 + \mathbb{E}(\rho_1(Y_1) - g(Y_1))^2. \end{aligned}$$

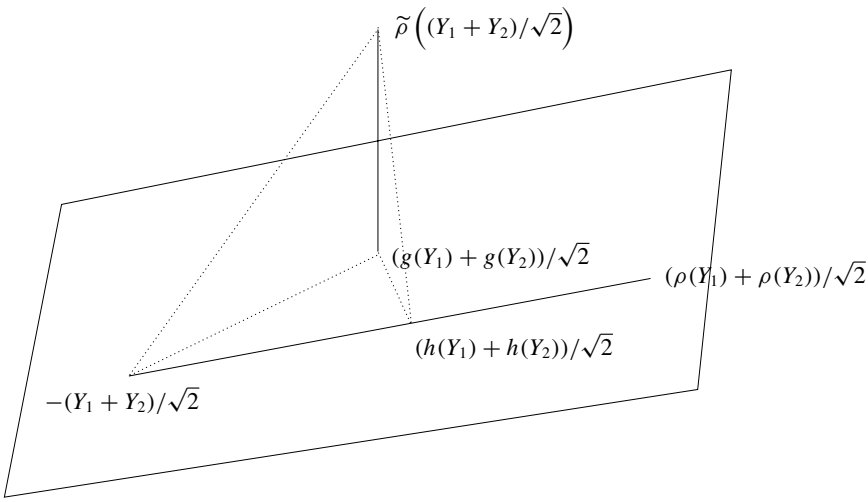


**Fig. 1.** Role of projections

Now, consider the projection of $\widetilde{\rho}$ into the space of additive functions, shown as a plane in Figure 1, where $(h(Y_1)+h(Y_2))/\sqrt{2}$ is the closest point to $\widetilde{\rho}$ on the line between $-(Y_1+Y_2)/\sqrt{2}$ and $(\rho(Y_1)+\rho(Y_2))/\sqrt{2}$, so that $\mathbb{E}(g(Y_1)+Y_1)^2 \geq \mathbb{E}(h(Y_1)+Y_1)^2$.

Further, we know that $h$ corresponds to the value of $\lambda$ which minimises:

$$\mathbb{E}\left(\widetilde{\rho}\left(\frac{Y_1+Y_2}{\sqrt{2}}\right)-\left(\lambda\left(\frac{\rho(Y_1)+\rho(Y_2)}{\sqrt{2}}\right)-(1-\lambda)\left(\frac{Y_1+Y_2}{\sqrt{2}}\right)\right)\right)^2.$$

Since in general $\mathbb{E}(U-\lambda V)^2$ is minimised at $\lambda = \mathbb{E}UV/\mathbb{E}V^2$, in this case the minimising $\lambda = J'/J$, so $h$ is $J'/J$ of the way along the line. This tells us that $\mathbb{E}(h(Y)+Y)^2 = (J'/J)^2\mathbb{E}(\rho(Y)+Y)^2 = J'^2/J$.

Overall then, we deduce that $\mathbb{E}(g(Y_1)+Y_1)^2 \geq J'^2/J$, and by Pythagoras,

$$\mathbb{E}\left(\widetilde{\rho}\left((Y_1+Y_2)/\sqrt{2}\right)-(g(Y_1)+g(Y_2))/\sqrt{2}\right)^2 \leq J'-J'^2/J.$$

Now applying Proposition 2.1 to the left side of the above equation, we can see that the factor of $I$ in the denominator of the inequality that follows will actually cancel, simplifying the expression.

$$\begin{aligned} J'-J'^2/J &\geq \mathbb{E}\left(\widetilde{\rho}\left(\frac{Y_1+Y_2}{\sqrt{2}}\right)-\frac{g(Y_1)+g(Y_2)}{\sqrt{2}}\right)^2 \geq \frac{\mathbb{E}(g_1(Y_1)-\mu Y_1)^2}{2R^*I} \\ &= \frac{\mathbb{E}(g_1(Y_1)+Y_1)^2+(-\mu-1)^2}{2R^*I} \geq \left(\frac{J'^2}{J}+J'^2\right)\frac{1}{2R^*I} = \frac{J'^2}{2R^*J} \end{aligned}$$

since $\mu = -I'$ and since $\mathbb{E}Y_1g_1(Y_1) = -1$, so rearranging, we obtain the result.

□

For the rest of this section suppose $X_1, \ldots, X_n$ are IID with absolutely continuous density with finite Fisher information $I(X)$ and let $U_n$ be the standardized sum of $X_1 + \cdots + X_n$. A more careful analysis generalises Proposition 3.2, to obtain Theorem 1.3 by performing successive projections onto smaller additive spaces. For a given mean zero function $f \in L^2$, define a series of functions by $f_n = f$, and for $m < n$:

$$f_m\left(\frac{X_1 + \cdots + X_m}{\sqrt{n}}\right) = \mathbb{E}_{X_{m+1}} f_{m+1}\left(\frac{X_1 + \cdots + X_m + X_{m+1}}{\sqrt{n}}\right).$$

Further, define $g(u) = \sqrt{n}\mathbb{E} f\left(\frac{X_1+\cdots+X_{n-1}+u}{\sqrt{n}}\right)$. Note that since the random variables are IID each $g(X_i)$ is the result of integrating out all the other variables. Likewise $f_i$ arises from integrating out $n - i$ of the random variables.

At step $i$, we approximate the function $f$ by $f_i((X_1 + \cdots + X_i)/\sqrt{n})$ plus a sum of $g(X_j)$ for $j > i$, which is the best approximation onto the linear space of such partially additive functions.

**Lemma 3.3.** *Defining the squared distance between successive projections to be*

$$t_i = \mathbb{E}\left(f_i\left(\frac{X_1 + \cdots + X_i}{\sqrt{n}}\right) - f_{i-1}\left(\frac{X_1 + \cdots + X_{i-1}}{\sqrt{n}}\right) - \frac{1}{\sqrt{n}} g(X_i)\right)^2,$$

*then there exists a constant $\mu$ such that for $X_i$ IID and with finite restricted Poincaré constant $R^*$:*

$$t_i \geq \frac{(i-1)}{nI(X)R^*}\mathbb{E}(g(X) - \mu X)^2.$$

*Proof.* Work with the function

$$r(z) = \mathbb{E}\left(f_i\left(\frac{X_1 + \cdots + X_{i-1} + z}{\sqrt{n}}\right) - f_{i-1}\left(\frac{X_1 + \cdots + X_{i-1}}{\sqrt{n}}\right) - \frac{1}{\sqrt{n}} g(z)\right) \times (\rho(X_1) + \cdots + \rho(X_{i-1}))$$

in two different ways. Firstly, we apply Cauchy-Schwarz to $r(z)^2$, and take expected values, to deduce that $\mathbb{E}r(X)^2 \leq t_i(i-1)I(X)$.

Secondly, writing $\overline{\rho}$ for the score of $X_1 + \cdots X_{n-1}$, the function $g(z)$ has derivative $g'(z) = -\mathbb{E}f((X_1 + \cdots + X_{n-1} + z)/\sqrt{n})\overline{\rho}(X_1 + \cdots + X_{n-1}) = -\mathbb{E}f((X_1 + \cdots + X_{n-1} + z)/\sqrt{n})\rho(X_1)$ (by Lemma 3.1). We consider this as an iterated integral, first integrating out the variables $X_i$ through $X_{n-1}$, so that the result is $-\mathbb{E}f_i(X_1 + \cdots + X_{i-1} + z)/\sqrt{n})\rho(X_1)$, and hence

$$r(z) = -\left(\frac{i-1}{n}\right)(g'(z) - \mu).$$

Putting these together, the result follows, using the Poincaré inequalities. □

**Lemma 3.4.** *For $X_i$ IID, the sum of these squared distances $t_i$ is $s_m = \sum_{i=1}^m t_i$, where*

$$s_m = \mathbb{E}\left(f_m\left(\frac{X_1 + \cdots + X_m}{\sqrt{n}}\right) - \sum_{i=1}^m \frac{g(X_i)}{\sqrt{n}}\right)^2.$$

*Proof.* Since $s_m = \mathbb{E}f_m^2 - (m/n)\mathbb{E}g^2$, and since $t_m = \mathbb{E}f_m^2 - \mathbb{E}f_{m-1}^2 - (1/n)\mathbb{E}g^2$, we can rearrange to obtain:

$$s_m = (t_m + \mathbb{E}f_{m-1}^2 + (1/n)\mathbb{E}g^2) - (m/n)\mathbb{E}g^2 = t_m + s_{m-1},$$

so summing the telescoping sum, the result follows. □

Combining Lemma 3.3 and Lemma 3.4, we deduce that:

$$s_n \geq \sum_{i=1}^{n} \frac{(i-1)}{nI(X)R^*}\mathbb{E}(g(X) - \mu X)^2 = \frac{(n-1)}{2I(X)R^*}\mathbb{E}(g(X) - \mu X)^2. \tag{10}$$

*Proof of Theorem 1.3.* Assume that $X$ has variance 1, and write $J'$ for $J(U_n)$, $J$ for $J(X)$ and take $f = \sqrt{n}\rho_n$ (where $\rho_n$ is the score of $U_n$) with corresponding $g$ and $\mu$ as defined above. As before we know that $\mathbb{E}(g(X) + X)^2 \geq J'^2/J$ and $s_n = \mathbb{E}\left(\rho_n - \sum g(X_i)/\sqrt{n}\right)^2 \leq J'(1 - J'/J)$. Hence by Equation (10), we deduce that:

$$\begin{aligned} J'(1 - J'/J) \geq s_n &\geq \frac{(n-1)}{2R^*I(X)}\mathbb{E}(g(X) - \mu X)^2 \\ &\geq \frac{(n-1)}{2R^*I(X)}\left(\frac{J'^2}{J} + J'^2\right) = \frac{(n-1)}{2R^*}\frac{J'^2}{J}. \end{aligned}$$

Thus, in general, rescaling gives:

$$J(U_n) \leq \frac{2R^*}{2R^* + (n-1)\sigma^2}J(X),$$

and the result follows. □

## 4. Convergence of Fisher information

The remainder of this paper will show how we can prove Theorem 1.6 which implies convergence of Fisher information (though without such an attractive rate of convergence), even if the Poincaré constants are not finite. We will need uniform control over the tails of the Fisher information, and then will bound it on the rest of the region using the projection arguments of Section 2. Recall that for $I(X)$ finite, the density of $X$ is bounded (see Lemma A.1).

**Definition 4.1.** *Given a function $\psi$, we define the following class:*

$$\mathcal{C}_\psi = \{X : \mathbb{E}X = 0, \sigma^2 = \mathbb{E}X^2 < \infty, \sigma^2\mathbb{E}\rho(X)^2\mathbb{I}(|X| \geq \sigma T) \leq \psi(T) \text{ for all } T.\}$$

In the remainder of the section, we will assume that the common variance of the random variables is equal to 1.

**Lemma 4.2.** *For $X_1, X_2, \ldots$ IID with finite variance and finite $I(X)$, then $U_m \in \mathcal{C}_\psi$ for all $m$ where $\psi(T) = \mathbb{E}\rho(X)^2\mathbb{I}(|X| \geq T) + C/T^{1/2}$.*

*Proof.* We use the notation that $p$ and $\rho$ stand for the density and score function of a single $X$, and $p_r$ for the density of $X_1 + \cdots X_r$. We know that $U_m$ has score function $\rho_m(u) = \mathbb{E}\left(\sum_i \rho(X_i) \middle| U_m = u\right)/\sqrt{m}$, so by the conditional version of Jensen's inequality

$$\rho_m(u)^2 \leq \mathbb{E}(\rho(X_1)^2 | U_m = u) + (m-1)\mathbb{E}(\rho(X_1)\rho(X_2)|U_m = u). \tag{11}$$

Consider the two terms of Equation (11) separately, firstly writing $W$ for $X_2 + \cdots + X_m$:

$$\begin{aligned}
&\mathbb{E}_{U_m}\mathbb{E}[\rho(X_1)^2|U_m]\mathbb{I}(|U_m| \geq T) \\
&\quad \leq \mathbb{E}\rho(X_1)^2 \left(\mathbb{I}(|X_1| \geq T, |U_m| \geq T) + \mathbb{I}(|X_1| < T, |U_m| \geq T)\right) \\
&\quad \leq \mathbb{E}\rho(X_1)^2 \left(\mathbb{I}(|X_1| \geq T) + \mathbb{I}(|W| \geq T(\sqrt{m}-1))\right) \\
&\quad \leq \mathbb{E}\rho(X)^2 \mathbb{I}(|X| \geq T) + \frac{I(X)(m-1)}{T^2(\sqrt{m}-1)^2}
\end{aligned}$$

Then for any $u$, writing $q_m$ for the density of $U_m$

$$\begin{aligned}
&\mathbb{E}(\rho(X_1)\rho(X_2)|U_m = u) \\
&\quad = \iint \frac{\sqrt{m}p(v)p(w)p_{m-2}(u\sqrt{m}-v-w)}{q_m(u)}\rho(v)\rho(w)dvdw \\
&\quad = \frac{\sqrt{m}}{q_m(u)} \int p_{m-2}(u\sqrt{m}-x) \int \frac{\partial p}{\partial v}(v)\frac{\partial p}{\partial x}(x-v)dvdx.
\end{aligned}$$

So on integrating the second term of Equation (11) we obtain $q'_m(-T) - q'_m(T)$ and we need a function $\psi'$ such that for all $T$:

$$|q'_m(T)| \leq \psi'(|T|) \tag{12}$$

For all $m$, $q_m(x) \leq \sqrt{I(U_m)} \leq \sqrt{I}$, so that

$$\begin{aligned}
q'_{2m}(u) &= 2\int q'_m(v)q_m(u\sqrt{2}-v)dv \\
&\leq 2^{3/4}\left(\int \frac{q'_m(v)^2}{q_m(v)}dv\right)^{1/2}\left(\int \sqrt{2}q_m(v)q_m^2(u\sqrt{2}-v)dv\right)^{1/2} \\
&\leq 2^{3/4}\sqrt{I}\left(\sqrt{I}\int \sqrt{2}q_m(v)q_m(u\sqrt{2}-v)dv\right)^{1/2} \\
&\leq (2I)^{3/4}\sqrt{q_{2m}(u)}
\end{aligned}$$

(a similar bound will hold for $q_{2m+1}$) and

$$q_m(u) \leq \int_u^\infty |q'_m(v)|dv \leq \left(\int_u^\infty \frac{q'_m(v)^2}{q_m(v)}dv\right)^{1/2}\left(\int_u^\infty q_m(v)dv\right)^{1/2} \leq \frac{\sqrt{I}}{u},$$

we deduce that Equation (12) holds, with $\psi'(T) = 2^{3/4}I/T^{1/2}$. Note that under a $(2+\delta)$th moment condition, we obtain $\psi'(T) = C/T^{(2+\delta)/4}$. □

By results of Brown (1982), we know that under a finite variance condition, there exists $\theta(T)$ such that $\mathbb{E}X^2\mathbb{I}(|X| \geq \sigma T) \leq \theta(T)$. If in addition, $\mathbb{E}|X|^{2+\delta}$ is finite for some $\delta$, the Rosenthal inequality implies that $\mathbb{E}|U_n|^{2+\delta}$ is uniformly bounded, so we can take $\theta(T) = 1/T^{\delta}$.

The other ingredient we require is a bound on the Poincaré constant $R^T_{U_n}$ (which is the Poincaré constant of a random variable $Y$ with the distribution of $U_n$ conditioned on the event $|U_n| \leq T$). This will be used in the proof of Theorem 1.6, since $\mathbb{E}[g(U_n)^2\mathbb{I}(|U_n| \leq T)]/\mathbb{E}[g'(U_n)^2\mathbb{I}(|U_n| \leq T)] = \mathbb{E}[g(U_n)^2|(|U_n| \leq T)]/\mathbb{E}[g'(U_n)^2|(|U_n| \leq T)] = \mathbb{E}g(Y)^2/\mathbb{E}g'(Y)^2 \leq R^T_{U_n}$.

**Lemma 4.3.** *If $I(X)$ is finite then there exist $R(T)$ and $N(T)$ such that for all $T$, $R^T_{U_n} \leq R(T)$ for $n \geq N(T)$ (that is, for all $T$ the sequence $(R^T_{U_n})_{n\geq 1}$ is bounded).*

*Proof.* Write the total variation distance between $f_n$ (the density of $U_n$) and the standard normal density as $d_n = \sup_A |f_n(A) - \phi(A)|$ (which tends to zero). Since $f_n$ is bounded then:

$$\begin{aligned}
|f_{2n}(x) - \phi(x)| &\leq \sqrt{2}\left|\int f_n(\sqrt{2}x - y)(f_n(y) - \phi(y))dy\right| \\
&\quad + \sqrt{2}\left|\int \phi(\sqrt{2}x - y)(f_n(y) - \phi(y))dy\right| \\
&\leq 2\sqrt{2}\,(\|f_n\|_\infty + \|\phi\|_\infty)\int (f_n(y) - \phi(y))I(f_n(y) \geq \phi(y))dy \\
&\leq 2\left(\sqrt{2I} + \sqrt{1/\pi}\right)d_n
\end{aligned}$$

Now, for given $T$, take

$$N(T) = 2\min\left\{m : \left(\sqrt{2I} + \sqrt{1/\pi}\right)d_n \leq \phi(T)/2 \text{ for all } n \geq m\right\}.$$

This implies that $f_n(x) \geq \phi(T)/2$, for $x \in [-T, T]$ and $n \geq N(T)$, so $R(T) = 2/\phi(T)$ means

$$-\int_{-T}^{x} y f_n(y)dy \leq R(T) f_n(x), \quad \text{for } 0 \geq x \geq -T \tag{13}$$

$$\int_{x}^{T} y f_n(y)dy \leq R(T) f_n(x), \quad \text{for } 0 \leq x \leq T, \tag{14}$$

since the LHS of (13) and (14) is always less than 1. Now Equations (13) and (14) are precisely the conditions under which Theorem 1 of Borovkov and Utev (1984) proves that the random variable has Poincaré constant $R(T)$, so we are done. □

Combining these two results gives the following.

*Proof of Theorem 1.6.* Using the method of Proposition 2.1, with $f = \sqrt{2}\rho_{2n}$, a function $g$ and constants $\mu, \nu$ are identified such that:

$$\begin{aligned}
&J(U_n) - J(U_{2n}) \\
&\quad = \mathbb{E}(\rho_n(U_n) - g(U_n))^2 + \mathbb{E}\left(\rho_{2n}(U_{2n}) - \frac{1}{\sqrt{2}}g(U_n) - \frac{1}{\sqrt{2}}g(U_n')\right)^2 \\
&\quad \geq \mathbb{E}(\rho_n(U_n) - g(U_n))^2 \mathbb{I}(|U_n| \leq T) \\
&\qquad + \frac{1}{2R_{U_n}^T I(U_n)} \mathbb{E}(g(U_n) - \mu U_n - \nu)^2 \mathbb{I}(|U_n| \leq T) \\
&\quad \geq \frac{1}{1 + 2R_{U_n}^T I(U_n)} \mathbb{E}(\rho_n(U_n) - \mu U_n - \nu)^2 \mathbb{I}(|U_n| \leq T),
\end{aligned}$$

since $ax^2 + by^2 \geq ab(x-y)^2/(a+b)$. Now $\mu = -I(U_{2n})$, and $\nu = -\mathbb{E}(g(U_n) - \mu U_n)\mathbb{I}(|U_n| < T)$. The standardized Fisher information involves the best linear approximation to the score. That is $J(U_n) = \inf_{a,b} \mathbb{E}(\rho_n(U_n) - aU_n - b)^2 \leq \mathbb{E}(\rho_n(U_n) - \mu U_n - \nu)^2$, so that

$$\begin{aligned}
J(U_n) &\leq \mathbb{E}(\rho_n(U_n) - \mu U_n - \nu)^2 \\
&= \mathbb{E}(\rho_n(U_n) - \mu U_n - \nu)^2 \left(\mathbb{I}(|U_n| \leq T) + \mathbb{I}(|U_n| > T)\right) \\
&\leq (1 + 2R_{U_n}^T I(U_n))(J(U_n) - J(U_{2n})) \\
&\qquad + \mathbb{E}(\rho_n(U_n) - \mu U_n - \nu)^2 \mathbb{I}(|U_n| > T),
\end{aligned}$$

and hence by Lemmas 4.2 and 4.3, for some function $\zeta(T)$ such that $\zeta(T) \to 0$ as $T \to \infty$:

$$J(U_n) \leq (1 + 2R_{U_n}^T I(U_n))(J(U_n) - J(U_{2n})) + \zeta(T).$$

For any $\epsilon$ we can find $T_0$ such that $\zeta(T_0) \leq \epsilon$, for all $n \geq N(T_0)$, then $(1 + 2R_{U_n}^T I(U_n))(J(U_n) - J(U_{2n}) \leq (1 + 2R(T_0)I)(J(U_n) - J(U_{2n}) \leq \epsilon$ for $n$ sufficiently large. □

The result that if $J(U_n)$ is finite for some $n$ then it tends to zero mirrors the main theorem of Barron (1986), that if $D(U_n)$ is finite for some $n$ then it tends to zero.

## A. Appendix: some score function properties

**Lemma A.1.** *Suppose $Y$ is a real-valued random variable with probability density function $p$.*

*A. If the density function is absolutely continuous and has finite Fisher information $I(Y) = \mathbb{E}\rho^2(Y)$, where the score function $\rho(y)$ is defined by $p'(y)/p(y)$ wherever $p$ is positive and differentiable (and elsewhere, in a $P_Y$-null set, is defined arbitrarily, say equal to 0), then*

*1. the density $p$ is bounded,*

*2. it has bounded variation $\int |p'(y)|dy = \mathbb{E}|\rho(Y)| \leq \sqrt{I(Y)}$,*

3. *for every bounded function $f$, the function $g(u) = \mathbb{E}f(u+Y)$ is absolutely continuous on the line and has derivative $g'(u) = -\mathbb{E}f(u+Y)\rho(Y)$, Lebesgue almost everywhere,*
4. *more generally, for functions with $\mathbb{E}f^2(u+Y)$ bounded as a function of $u$ in an interval, $g(u) = \mathbb{E}f(u+Y)$ is absolutely continuous in the interval with derivative $g'(u) = -\mathbb{E}f(u+Y)\rho(Y)$, Lebesgue almost everywhere,*
5. *$\mathbb{E}\rho(Y) = 0$, if $\mathbb{E}Y^2$ is finite then $\mathbb{E}Y\rho(Y) = -1$, and if $\mathbb{E}Y^4$ is finite then $\mathbb{E}Y^2\rho(Y) = -2\mathbb{E}Y$.*

B. *In the converse direction, if for some function $\rho(y)$ with finite expected square, the random variable satisfies (A3) for all bounded functions $f$, then a version of the density function $p$ is absolutely continuous and satisfies $p'(y) = \rho(y)p(y)$ Lebesgue almost everywhere.*

*Note:* Conclusions (A3) and (A4) may be regarded as exchanges of integral and derivative in $g(u) = \int f(s)p(s-u)du$, using smoothness of $p$ rather than smoothness of $f$. It shows that the operation of differentiation of the expectation of $f(u+Y)$, for classes of general $f$, consists of taking an inner product with the score function. Conclusion (A5) may be regarded as applications of integrations by parts or may be deduced from (A4) as shown here. These identities and properties of score functions are key tools in our projection inequalities.

*Proof of Lemma A.1.* Part A. Let $p$ be absolutely continuous and let $p'$ be a function defined to equal the derivative of $p$ where it exists and set arbitrarily on the Lebesgue null set where $p$ is not differentiable. Now at points of differentiability if $p(y)$ is 0 then $p'(y)$ must also be 0 since the density is non-negative. Consequently, letting $C = \{y : p(y) > 0\}$, we have $\int |p'(y)|dy = \int_C |p'(y)|dy$ equalling $\int |\rho(y)|p(y)dy = \mathbb{E}|\rho(Y)| \le \sqrt{\mathbb{E}|\rho(Y)|^2} = \sqrt{I(Y)}$. Thus when $I = I(Y)$ is finite, the variation $\int |p'(y)|dy$ is finite. Absolute continuity yields $p(u) = \int_{-\infty}^{u} p'(y)dy$, so that we have the bound $p(u) \le \int |p'(y)|dy \le \sqrt{I}$. This verifies conclusions (A1) and (A2).

Given a function $f$, let $g(u) = \mathbb{E}f(u+Y)$ and $h(u) = -\mathbb{E}f(u+Y)\rho(Y)$. We are to show, under conditions on $f$, that $g$ is absolutely continuous with derivative determined by $h$. Toward that end, consider intervals $[v, w]$, and the integral $\int_v^w h(u)du$, which entails the integration of $f(u+y)p'(y)$ for $u$ in the interval and $y$ on the line. Cauchy-Schwarz demonstrates the integrability $\int_v^w [\int |f(u+y)||p'(y)|dy]du \le \int_v^w (\mathbb{E}f^2(u+Y))^{1/2} I du$ where the indicated conditions on $f$ provide the required local integrability. Now $\int_v^w h(u)du = -\int_v^w [\int f(s)p'(s-u)ds]du$ and by Fubini we may exchange the order of this integration to obtain that this is $-\int f(s)[\int_v^w p'(s-u)du]ds$ which by the absolute continuity of $p$ is $-\int f(s)(p(s-v) - p(s-w))ds = g(w) - g(v)$, so $g(u)$ is absolutely continuous with derivative a.e. provided by $h(u) = -\mathbb{E}f(u+Y)\rho(Y)$. This demonstrates conclusions (A3) and (A4).

Take $f(u+Y)$ equal to 1, $u+Y$, or $(u+Y)^2$ so the corresponding $g(u)$ is 1, $u + \mathbb{E}Y$, and $u^2 + 2u\mathbb{E}Y + \mathbb{E}Y^2$, respectively, with derivatives, 0, 1, and $2u + 2\mathbb{E}Y$. Under the respective stated conditions $\mathbb{E}f^2(u+Y)$ is locally bounded (bounded in finite intervals), and hence by conclusion (A4) these derivatives match $-\mathbb{E}\rho(Y)$,

$-\mathbb{E}[(u+Y)\rho(Y)]$, and $-\mathbb{E}[(u+Y)^2\rho(Y)]$, respectively, for a.e. $u$. Expanding the quadratic and using each conclusion in turn, we find $\mathbb{E}\rho(Y) = 0$, $\mathbb{E}Y\rho(Y) = -1$, and $\mathbb{E}Y^2\rho(Y) = -2\mathbb{E}Y$, proving (A5).

For Part B, suppose statement (A3) holds for some function $\rho(Y)$ with finite expected square. Take $f(s) = \mathbb{I}(s \le 0)$. Then $g(u) = \mathbb{E}f(u+Y) = \mathbb{P}(Y \le -u)$ is the cumulative distribution function, so being absolutely continuous its derivative provides the density function. Moreover statement (A3) expresses its derivative as $g'(u) = -\mathbb{E}f(u+Y)\rho(Y) = -\int_{-\infty}^{-u} \rho(y)p(y)dy$ for Lebesgue almost every $u$. Thus setting $p'(y) = \rho(y)p(y)$ we have that $\int_{-\infty}^{y} p'(u)du$ provides an absolutely continuous version of the density and that $\rho$ is indeed its score function. □

*Examples:* The one-sided exponential density, due to its discontinuity at 0, is not absolutely continuous. Its convolution with itself, $\Gamma(2)$, is absolutely continuous, with a jump in the derivative at 0, and its linear behavior at 0+ leads to an unbounded score function with infinite Fisher information. Convolutions of three or more exponentials yields the $\Gamma(n)$ density ($n \ge 3$) which is absolutely continuous and has an unbounded score function, yet finite Fisher information, so the conclusions of Lemma A.1 hold for these. The two-sided exponential (Laplace) density (though it has a jump in the derivative, that jump occurs where the density is positive) is absolutely continuous with finite Fisher information.

*Note:* The differentiability of $g(u) = \mathbb{E}f(u+Y)$ for absolutely continuous densities $p$ and certain conditions on the functions $f$ is related to the notion of weak differentiability of $p$ studied in Fabian and Hannan (1977), Brown and Gajek (1990), and Lehmann and Casella (1998). They require differentiability of $g$ for the slightly larger class of all $f$ with $\mathbb{E}f^2(u+Y)$ finite for $u$ in a given set. Potentially that condition could have been used, however it is slightly more stringent than the absolute continuity of $p$, and as the examples indicate, difficulties can occur with infinite $\mathbb{E}f^2(u+Y)$ at the edge of the support of the density, for relevant score functions $f$. Such behaviour can interfere with absolute continuity of the resulting $g$ which we want to appeal to at a step in our analysis. Indeed, what happens for densities that approach zero in certain intervals is that even when $\mathbb{E}f^2(Y_1+Y_2)$ is finite, the local Lebesgue integrability of $\sqrt{\mathbb{E}f^2(u+Y)}$ (sought for showing that $g$ is absolutely continuous) can fail because that integrability does not have the advantage of the factor $p(u)$ to ameliorate the affect of unbounded $\sqrt{\mathbb{E}f^2(u+Y)}$. So with either condition on $p$ (absolute continuity or weak differentiability) one would need to use the denseness of certain subclasses of functions in $L^2$ to obtain the inequalities. Once recognized, it led us in the present manuscript to use the more general condition (of absolute continuity of $p$) together with an argument which truncates the magnitude of $f$ for general $f$ in $L^2$. Nonetheless, for many unbounded score functions playing the role of $f$, no such truncation is required.

It is pedagogically of note that the analysis underlying the inequalities of this paper may also be carried out using classical sophomore level calculus, under the simpler (though not as full in generality) assumption that the density is continuously differentiable with finite Fisher information (as an extended Riemann integral), and

can be presented at that level in few lectures if one is willing to not belabour the exchanges of integrals and derivatives.

## References


Ball, K., Barthe, F., Naor, A.: Entropy jumps in the presence of a spectral gap. Duke Math. J. **119**, 41–63 (2002)

Barron, A.R.: Entropy and the central limit theorem. Ann. Probab. **14**, 336–342 (1986)

Blachman, N.M.: The convolution inequality for entropy powers. IEEE Trans. Inform. Theory **11**, 267–271 (1965)

Borovkov, A.A., Utev, S.A.: On an inequality and a related characterisation of the normal distribution. Theory Probab. Appl. **28**, 219–228 (1984)

Brown, L.D.: A proof of the Central Limit Theorem motivated by the Cramér-Rao inequality. In: G. Kallianpur, P.R. Krishnaiah, J.K. Ghosh, (eds), Statistics and Probability: Essays in Honour of C.R. Rao, North-Holland, New York, 1982, pp. 141–148

Brown, L.D., Gajek, L.: Information inequalities for the Bayes risk. Ann. Statist. **18**, 1578–1594 (1990)

Cacoullos, Th.: On upper and lower bounds for the variance of a function of a random variable. Ann. Probab. **10**, 799–809 (1982)

Casella, G., Berger, R.L.: Statistical inference. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990

Fabian, V., Hannan, J.: On the Cramér-Rao inequality. Ann. Statist. **5**, 197–205 (1977)

Gnedenko, B.V., Kolmogorov, A.N.: Limit distributions for sums of independent random variables. Addsion- Wesley, Cambridge, Mass, 1954

Gross, L.: Logarithmic Sobolev inequalities. Amer. J. Math., **97**, 1061–1083 (1975)

Holevo, A.S.: Asymptotic estimation of shift parameter of a quantum state. Preprint, 2003. quant-ph/0307225

Johnson, O.T.: Entropy inequalities and the Central Limit Theorem. Stochastic Process Appl. **88**, 291–304 (2000)

Johnson, O.T., Suhov, Y.M.: Entropy and random vectors. J. Statist Phys. **104**, 147–167 (2001)

Klaasen, C.A.J.: On an inequality of Chernoff. Ann. Probab. **13**, 966–974 (1985)

Lehmann, E., Casella, G.: *Theory of point estimation*. Springer Texts in Statistics. Second edition, Springer-Verlag, New York, 1998

Linnik, Y.V.: An information-theoretic proof of the Central Limit Theorem with the Lindeberg Condition. Theory Probab. Appl. **4**, 288–299 (1959)

Petrov, V.V.: *Limit Theorems of Probability: Sequences of Independent Random Variables*. Oxford Science Publications, Oxford, 1995

Prohorov, Y.V.: On a local limit theorem for densities. Doklady Akad. Nauk SSSR (N.S.), **83**, 797—800 (1952) In Russian.

Shimizu, R.: On Fisher's amount of information for location family. In: G.P.Patil et al, (eds), Statistical Distributions in Scientific Work, Vol. **3**, Reidel, 1975, pp. 305–312

Stam, A.J.: Some inequalities satisfied by the quantities of information of Fisher and Shannon. Inform. and Control **2**, 101–112 (1959)


**Added in Proof:** Here we comment further on central limit theorems for random variables without finite Poincaré constants. In the proof of Lemma 4.3 we appeal to convergence of the densities in $L^1$ together with finite $J$, to deduce $L^\infty$ convergence of the densities. This $L^\infty$ convergence is used in a demonstration of a lower bound on the densities in bounded intervals, which is an ingredient in our demonstration of Theorem 1.6. $L^1$ and $L^\infty$ convergence results are available in Prohorov (1952) and Gnedenko and Kolmogorov (1954), respectively. Alternatively, one can prove what is needed by noting that finite $J$ implies finite relative entropy $D$, indeed $D \leq J/2$, so that $D(U_n)$ tends to zero by Barron (1986), which

gives $L^1$ convergence as a corollary. Our point is that this method provides a proof of Theorem 1.6, that $J(U_n)$ tends to zero if and only if it is ever finite, based solely on consideration of entropy and information.

As for the inequality $D(X) \leq J(X)/2$, it is a consequence of an inequality by Stam (1959) who shows (based on convolution inequalities for entropy and the de Bruijn identity) that $(2\pi e)e^{-2H(X)} \leq I(X)$, which is equivalent to $D(X) \leq (1/2)\log(1 + J(X))$. In later developments $D(X) \leq J(X)/2$ is called a log-Sobolev inequality (Gross (1975)).