ELSEVIER

Statistics & Probability Letters 34 (1997) 37-42

STATISTICS & PROBABILITY LETTERS

Converse Poincaré-type inequalities for convex functions

S.G. Bobkov[a,1], C. Houdré[b,*,2]

[a] Department of Mathematics, University of Syktyvkar, 167001 Syktyvkar, Russian Federation
[b] Center for Applied Probability, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received 1 January 1996; revised 1 April 1996

Abstract

Converse Poincaré-type inequalities are obtained within the class of smooth convex functions. This is, in particular, applied to the double exponential distribution.

AMS Classification: Primary 60E15; 60E99; secondary 26D10

Keywords: Poincaré inequality; Double exponential distribution

Let ν be the double exponential distribution on the real line of density $2^{-1}\exp(-|x|)$, $x \in \mathbb{R}$. One of the purposes of these notes is to prove that if ξ has distribution ν and if f is a convex function on the real line, then

$$\boldsymbol{E} f'(\xi)^2 \leqslant \mathbf{Var} f(\xi) \leqslant 4\boldsymbol{E} f'(\xi)^2, \tag{1}$$

with equality on the left–hand side for the function $f(x) = |x|$.

The inequality on the right-hand side in (1) belongs to the class of Poincaré inequalities. For the measure ν this second inequality is well-known (see, e.g., Klaassen, 1985) and valid without any convexity assumption. In the literature, one can also find a number of lower estimates for the variance of functions of various distributions ((Cacoullos, 1982, 1989; Cacoullos and Papathanasiou, 1992; Houdré and Kagan, 1995) ,...,). We will see here, that within the class of all convex functions, it is sometimes possible to estimate from below the variance of $f(\xi)$ by a quantity similar to the one appearing in Poincaré-type inequalities. This is in contrast to the fact that this is never possible within the class of all functions.

In fact, to consider a more general problem, let μ be an arbitrary non-atomic probability measure on the real line $\mathbb{R}$. Given $a \in \mathbb{R}$, let μ_a^- and μ_a^+ be, respectively, the left- and the right-conditional restriction of μ

* Corresponding author.

[1] Research supported in part by the ISF grant NZX000 and NZX300 and by the Russian Foundation for Basic Research Grant 96-01-00201.

[2] Research supported in part by an NSF Postdoctoral Fellowship.

PII S0167-7152(96)00163-0

to the half-lines $(-\infty,a]$ and $[a,+\infty)$, that is, for any Borel set A, let

$$\mu_a^-(A)=\frac{\mu(A\cap(-\infty,a])}{\mu((-\infty,a])},\qquad \mu_a^+(A)=\frac{\mu(A\cap[a,+\infty))}{\mu([a,+\infty))}.$$

The above definition makes sense when $a_0(\mu)<a<a_1(\mu)$, where $a_0(\mu)=\inf\operatorname{supp}(\mu)$, $a_1(\mu)=\sup\operatorname{supp}(\mu)$, and with supp denoting the support of the corresponding measure. Let $\mathbf{Var}(f,\mu)$ and $\mathbf{Var}(\mu)$ denote, respectively, the variance of a function f and of the identity function $i(x)=x$, with respect to μ. Throughout, it is also always assumed that μ has finite variance.

With these notations, the following gives a sufficient condition for a converse Poincaré inequality to hold within the class of convex functions.

Theorem 1. *Let the random variable ξ be distributed according to μ, and let*

$$\sigma^2(\mu)=\inf_{a_0(\mu)<a<a_1(\mu)}\min\,(\mathbf{Var}(\mu_a^-),\mathbf{Var}(\mu_a^+)).\tag{2}$$

Then, for any convex function f on the real line,

$$\mathbf{Var}f(\xi)\geqslant\sigma^2(\mu)\boldsymbol{E}f'(\xi)^2.\tag{3}$$

The property $\sigma^2(\mu)>0$ implies that $a_0(\mu)=-\infty$ and that $a_1(\mu)=+\infty$. Thus, the infimum in (2) is, in fact, taken over the whole real line. This can easily be seen by applying (3) to the functions $f(x)=(a-x)^+$, $f(x)=(x-a)^+$, and letting, respectively, $a\to a_0(\mu)$, $a\to a_1(\mu)$. In addition, we then have

$$\liminf_{a\to-\infty}\frac{1}{F(a)}\int_{-\infty}^{a}(a-x)^2\,\mathrm{d}F(x)>0,\tag{4}$$

$$\liminf_{a\to+\infty}\frac{1}{1-F(a)}\int_{a}^{+\infty}(x-a)^2\,\mathrm{d}F(x)>0,\tag{5}$$

where $F(x)=\mu((-\infty,x])$ is the distribution function of μ (and of the random variable ξ). We do not know if the properties (4)–(5) which are necessary for (3) to hold (up to a positive constant), imply that $\sigma^2(\mu)>0$. One can however see that (4) and (5) imply that the tails $F(-x)$, $1-F(x)$ are "big" and decrease at infinity rather slowly (at least as slowly as exponentials). In particular, the normal distribution function does not satisfy (4)–(5). Therefore, one cannot hope to extend (3) to the multidimensional case to get

$$\mathbf{Var}\,f(\xi_1,\ldots,\xi_n)\geqslant c\;\boldsymbol{E}|\nabla f(\xi_1,\ldots,\xi_n)|^2,$$

where $\{\xi_k\}$ is an i.i.d. sequence, f is an arbitrary smooth convex function on $\mathbb{R}^n$, ∇f is its gradient, and $c>0$ does not depend on the dimension n. Indeed, assuming $\boldsymbol{E}\xi_k=0$ and applying the above inequality to functions of the form $f(x)=g((x_1+\cdots+x_n)/\sqrt{n})$, we would obtain by the central limit theorem that

$$\mathbf{Var}\,g(\xi)\geqslant c\;\boldsymbol{E}g'(\xi)^2$$

for the class of all convex functions g and for ξ normal.

Note that a convex function f on $\mathbb{R}$ is differentiable except possibly on a countable set, and that in general one defines

$$|f'(x)|=\max\{|f'(x^-)|,\,|f'(x^+)|\}.$$

Of course, this is essential only for distributions F which have atoms. Denote by $\mathscr{F}_+$ the family of all non-decreasing, convex functions on the real line. The proof of Theorem 1 will rely on the following lemma of independent interest.

Lemma 1. *Given a random variable ξ with finite second moment and a constant $c > 0$, the following are equivalent:*

(a) $\mathbf{Cov}(f(\xi), g(\xi)) \geqslant c\boldsymbol{E} f'(\xi)g'(\xi)$ *for any $f, g \in \mathscr{F}_+$ such that $f(\xi)$ and $g(\xi)$ have finite second moment;*

(b) $\mathbf{Var} f(\xi) \geqslant c\boldsymbol{E} f'(\xi)^2$ *for any $f \in \mathscr{F}_+$;*

(c) $\mathbf{Var}(\xi - a)^+ \geqslant c\boldsymbol{P}\{\xi \geqslant a\}$ *for any a real.*

Proof. Clearly, (a) implies (b) which implies (c) (note also that (b) makes sense even if $f(\xi)$ has infinite second moment). To derive (a) from (c), one can assume that the distribution function F of ξ is continuous, and that the functions f and g in (a) are non-negative and vanish at $-\infty$. When this is the case, these functions can be represented as a mixture of functions of the form $f_a(x) = (x-a)^+$, and since $\mathbf{Cov}(f(\xi), g(\xi))$ is linear in f and in g, it suffices to establish (a) for such functions. Let $a \leqslant b$. By an integration by parts, we easily have

$$\begin{aligned}\mathbf{Cov}(f_a(\xi), f_b(\xi)) &= \int_b^{+\infty} (x-a)(x-b)\,\mathrm{d}F(x) - \int_a^{+\infty} (x-a)\,\mathrm{d}F(x) \int_b^{+\infty} (x-b)\,\mathrm{d}F(x) \\ &= \int_b^{+\infty} (2x-a-b)(1-F(x))\,\mathrm{d}x - \int_a^{+\infty} (1-F(x))\,\mathrm{d}x \int_b^{+\infty} (1-F(x))\mathrm{d}x.\end{aligned}$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}a}\mathbf{cov}(f_a(\xi), f_b(\xi)) = -\int_a^{+\infty} (1-F(x))\,\mathrm{d}x + (1-F(a)) \int_b^{+\infty} (1-F(x))\,\mathrm{d}x \leqslant 0,$$

that is, $\mathbf{Cov}(f_a(\xi), f_b(\xi))$ is non-increasing in a, while the right-hand side of (a), $(c\boldsymbol{E} f_a'(\xi) f_b'(\xi) = c(1 - F(b)))$, does not depend on a. Therefore, the inequality (a) is true for all $a \leqslant b$ if and only if it is true for all $a = b$, in which case it becomes (c). The lemma is proved. □

Proof of Theorem 1. Let $\mathbf{Var}(f, \mu)$ be finite (otherwise, there is nothing to prove). Also, and without loss of generality, f is assumed to have a finite global minimum, say, at a point a (otherwise, one can approximate f by the sequence of convex functions $f_n(x) = \max(f(x), -n)$, and then letting $n \to \infty$ in (3) with f_n gives (3) for f). As noted before, we can also assume that $a_0(\mu) = -\infty$, $a_1(\mu) = +\infty$. Now, if ν and λ are two probability measures, and if $\boldsymbol{E}_\nu f$ and $\boldsymbol{E}_\lambda f$, are the respective expectations of f, we have the identity,

$$\mathbf{Var}(f, p\nu + (1-p)\lambda) = p\mathbf{Var}(f, \nu) + (1-p)\mathbf{Var}(f, \lambda) + p(1-p)|\boldsymbol{E}_\nu f - \boldsymbol{E}_\lambda f|^2.$$

Putting $\nu = \mu_a^-$, $\lambda = \mu_a^+$, and $p = F(a)$, we obtain

$$\mathbf{Var}(f, \mu) \geqslant F(a)\mathbf{Var}(f, \mu_a^-) + (1-F(a))\mathbf{Var}(f, \mu_a^+). \tag{6}$$

By assumption, f is non-decreasing on $[a, +\infty)$ and non-increasing on $(-\infty, a]$. Hence, Lemma 1 applied to (f, μ_a^+) and (f, μ_a^-) gives

$$\mathbf{Var}(f, \mu_a^+) \geqslant c^+(a) \int f'(x)^2\,\mathrm{d}\mu_a^+(x), \tag{7}$$

$$\mathbf{Var}(f, \mu_a^-) \geqslant c^-(a) \int f'(x)^2\,\mathrm{d}\mu_a^-(x), \tag{8}$$

where the optimal values of $c^+(a)$ and $c^-(a)$ are given by

$$c^+(a) = \inf_{b \geqslant a} \frac{\mathbf{Var}((x-b)^+, \mu_a^+)}{\mu_a^+([b, +\infty))}, \qquad c^-(a) = \inf_{b \leqslant a} \frac{\mathbf{Var}((b-x)^+, \mu_a^-)}{\mu_a^-((-\infty, b])}.$$

Now, for any $b \geqslant a$,

$$\begin{aligned}
&\frac{\mathbf{Var}((x-b)^+, \mu_a^+)}{\mu_a^+([b,+\infty))} \\
&= \frac{1-F(a)}{1-F(b)}\left[\frac{1}{1-F(a)}\int_b^{+\infty}(x-b)^2\,\mathrm{d}F(x) - \left(\frac{1}{1-F(a)}\int_b^{+\infty}(x-b)\,\mathrm{d}F(x)\right)^2\right] \\
&= \frac{1}{1-F(b)}\int_b^{+\infty}(x-b)^2\,\mathrm{d}F(x) - \frac{1}{(1-F(a))(1-F(b))}\left(\int_b^{+\infty}(x-b)\,\mathrm{d}F(x)\right)^2 \\
&\geqslant \frac{1}{1-F(b)}\int_b^{+\infty}(x-b)^2\,\mathrm{d}F(x) - \left(\frac{1}{1-F(b)}\int_b^{+\infty}(x-b)\,\mathrm{d}F(x)\right)^2 \\
&= \mathbf{Var}((x-b)^+, \mu_b^+) = \mathbf{Var}(\mu_b^+),
\end{aligned}$$

since $1-F(a) \geqslant 1-F(b)$, and $(x-b)^+ = x-b \pmod{\mu_b^+}$. Thus,

$$c^+(a) \geqslant \min_{b \geqslant a} \mathbf{Var}(\mu_b^+) \geqslant \sigma^2(\mu),$$

where $\sigma^2(\mu)$ is defined by (2). In the same way, $c^-(a) \geqslant \sigma^2(\mu)$. Using these estimates in (7)–(8) and then in (6) gives (3). Theorem 1 is proved. □

It is clear from Lemma 1 that the optimal constant c in (b) can be found from (c). However, we would like to mention another way of finding this constant when the random variable ξ is exponentially distributed with density $\exp(-x), x > 0$.

Theorem 2. *Let the random variables ξ, η and ζ be independent, exponentially distributed random variables. Then, for any absolutely continuous functions f, g such that $f(\xi)$ and $g(\xi)$ have finite second moments,*

$$\mathbf{Cov}(f(\xi), g(\xi)) = \mathbf{E} f'(\xi+\eta) g'(\xi+\zeta). \tag{9}$$

In particular, under the additional assumption $f \in \mathscr{F}_+$, we have

$$\mathbf{Var} f(\xi) \geqslant \mathbf{E} f'(\xi)^2. \tag{10}$$

Proof. Both sides of (9) are bilinear in f and g, hence, it suffices to verify the equality for the functions $f(x) = \exp(\mathrm{i}tx)$, $g(x) = \exp(\mathrm{i}sx)$. But for such functions, and if $\varphi_\xi(t)$ is the characteristic function of ξ, (9) becomes

$$\varphi_\xi(t+s) - \varphi_\xi(t)\varphi_\xi(s) = -ts\varphi_\xi(t+s)\varphi_\xi(t)\varphi_\xi(s).$$

This identity can easily be verified directly since $\varphi_\xi(t) = 1/(1-\mathrm{i}t)$. To prove (10), we have from (9) and for $f = g$:

$$\mathbf{Var} f(\xi) = \int_0^{+\infty} \left(\mathbf{E} f'(\xi+t)\right)^2 \mathrm{e}^{-t}\,\mathrm{d}t \geqslant \int_0^{+\infty} f'(t)^2 \mathrm{e}^{-t}\,\mathrm{d}t,$$

since f' is non-negative and non-decreasing. Theorem 2 follows.

We are now ready to prove the result corresponding to the inequality (1).

Theorem 3. *Let the random variable ξ have a double exponential distribution. Then, for any convex function f on the real line,*

$$\boldsymbol{E} f'(\xi)^2 \leqslant \mathbf{Var} f(\xi) \leqslant 4\boldsymbol{E} f'(\xi)^2, \tag{11}$$

with equality on the left-hand side for the function $f(x) = |x|$.

Proof. Recall that $\mathrm{d}v(x)/\mathrm{d}x = 2^{-1}\exp(-|x|)$, $x \in \mathbb{R}$. By symmetry, $\mathbf{Var}(v_a^-) = \mathbf{Var}(v_{-a}^+)$, so we need only to show that $\mathbf{Var}(v_a^+) \geqslant 1$, for all a real. When $a \geqslant 0$, v_a^+ is the one-sided exponential distribution, hence, $\mathbf{Var}(v_a^+) = 1$, and it only remains to consider the case $a \leqslant 0$. To perform some computations, we find it convenient to work with the distribution function $F_a(x) = v_a^+((a,x])$. F_a is simply a shift of v_a^+, and thus $\mathbf{Var}(v_a^+) = \mathbf{Var}(F_a)$. Clearly, F_a has density $\mathrm{e}^{-|a+x|}/(2-\mathrm{e}^a)$, $x \geqslant 0$. Next, we use the elementary formulas

$$\int x\mathrm{e}^x\,\mathrm{d}x = (x-1)\mathrm{e}^x, \qquad \int x\mathrm{e}^{-x}\,\mathrm{d}x = -(x+1)\mathrm{e}^{-x},$$

$$\int x^2\mathrm{e}^x\,\mathrm{d}x = (x^2-2x+2)\mathrm{e}^x, \qquad \int x^2\mathrm{e}^{-x}\,\mathrm{d}x = -(x^2+2x+2)e^{-x},$$

to find

$$\begin{aligned}(2-\mathrm{e}^a)\int_0^{-a} x\,\mathrm{d}F_a(x) &= \int_0^{-a} x\mathrm{e}^{(a+x)}\,\mathrm{d}x = \mathrm{e}^a(x-1)\mathrm{e}^x|_0^{-a}\\ &= \mathrm{e}^a\left[-(a+1)\mathrm{e}^{-a}+1\right] = -(a+1)+\mathrm{e}^a\end{aligned}$$

$$\begin{aligned}(2-e^a)\int_{-a}^{\infty} x\,\mathrm{d}F_a(x) &= \int_{-a}^{\infty} x\mathrm{e}^{-(a+x)}\,\mathrm{d}x = \mathrm{e}^{-a}(-(x+1))\mathrm{e}^{-x}|_{-a}^{\infty}\\ &= 1-a.\end{aligned}$$

Thus,

$$\int_0^\infty x\,\mathrm{d}F_a(x) = \frac{-2a+\mathrm{e}^a}{2-\mathrm{e}^a}.$$

Moreover,

$$(2-\mathrm{e}^a)\int_0^{-a} x^2\,\mathrm{d}F_a(x) = \mathrm{e}^a(x^2-2x+2)\mathrm{e}^x|_0^{-a} = (a^2+2a+2)-2\mathrm{e}^a,$$

$$(2-\mathrm{e}^a)\int_{-a}^{\infty} x^2\,\mathrm{d}F_a(x) = -\mathrm{e}^{-a}(x^2+2x+2)\mathrm{e}^{-x}|_{-a}^{\infty} = a^2-2a+2,$$

and thus,

$$\int_0^\infty x^2 dF_a(x) = \frac{2a^2+4+\mathrm{e}^a}{2-\mathrm{e}^a}.$$

Hence,

$$\mathbf{Var}(F_a) = \frac{2a^2+4+\mathrm{e}^a}{2-\mathrm{e}^a} - \left(\frac{-2a+\mathrm{e}^a}{2-\mathrm{e}^a}\right)^2 = \frac{\mathrm{e}^{2a}-2a^2\mathrm{e}^a+4a\mathrm{e}^a-8\mathrm{e}^a+8}{(2-\mathrm{e}^a)^2}.$$

At this point, one can verify that $\mathbf{Var}(F_a) \to 2 = \mathbf{Var}(\nu)$, as $a \to -\infty$, and that $\mathbf{Var}(F_0) = 1$. Finally,

$$\begin{aligned}\mathbf{Var}(F_a) \geqslant 1 &\iff e^{2a} - 2a^2e^a + 4ae^a - 8e^a + 8 \geqslant (2 - e^a)^2 \\ &\iff -2a^2e^a + 4ae^a - 4e^a + 4 \geqslant 0 \\ &\iff (a^2 - 2a + 2)e^a \leqslant 2 \\ &\iff t^2 + 2t + 2 \leqslant 2e^t = 2(1 + t + t^2/2 + \cdots),\end{aligned}$$

where $t = -a$. This last inequality is certainly true since $t \geqslant 0$. The left inequality of Theorem 3 is proved.

References

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

Cacoullos, T. (1989), Dual Poincaré-type inequalities via the Cramer–Rao and the Cauchy–Schwarz inequalities and related characterizations, in: Y.Dodge, ed., *Statistical Data Analysis and Inference* (Elsevier, Amsterdam) 239–249.

Cacoullos, T. and V. Papathanasiou (1992), Lower variance bounds and a new proof of the central limit theorem, *J. Multivariate Anal.* **43**, 173–184.

Houdré, C. and A. Kagan (1995), Variance inequalities for functions of Gaussian variables, *J. Theoret. Probab.* **8**, 23–30.

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