František Rublík

On the quadratic derivative of exponential probabilities

Aplikace matematiky, Vol. 25 (1980), No. 4, 267--272

Persistent URL: http://dml.cz/dmlcz/103860

ON THE QUADRATIC DERIVATIVE OF EXPONENTIAL PROBABILITIES

FRANTIŠEK RUBLÍK

(Received April 18, 1978)

INTRODUCTION

A statistician who needs to make a decision on measured data has to use different types of probability distributions. For example, the normal distribution is used in geodesy, the Poisson distribution is used for measuring the decay of nuclear particles and for measuring random electric signals, the lognormal distribution is used in geology. All these distributions belong to the class of exponential families, which is the reason for studying various properties of this class.

Some statistical procedures can be described in the same way for all parametric families of probabilities which satisfy certain regularity conditions. This has been done for classes which have quadratic differentiable root of likelihood ratio e.g. in [2], [3], [4] and [7]. The following assumptions for testing hypotheses on a parameter of a distribution are made in [7]. The distribution of a random variable is supposed to belong to a parametric class of probabilities $\mathscr{P} = \{P_\theta;\ \theta \in \Theta\}$, which are defined on $(X, \mathscr{B})$, are mutually absolutely continuous and Θ is an open subset of R^k. Moreover, there exists a σ-finite measure μ defined on $(X, \mathscr{B})$ such that for the densities

$$f_\theta(t) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\mu}(t) \tag{1}$$

and for each $\theta_0 \in \Theta$, the functions

$$\varphi_\theta(t) = \left(\frac{f_\theta(t)}{f_{\theta_0}(t)}\right)^{1/2} \tag{2}$$

have the quadratic derivative $\dot{\varphi}_{\theta_0}(t)$ with respect to the measure P_{θ_0}. This derivative is defined in [7] to be a measurable function $\dot{\varphi}_{\theta_0} : X \to R^m$, satisfying the condition

$$\lim_{h \in R^m, h \to 0} \int_X \left[\frac{\varphi_{\theta_0+h}(t) - 1 - h'\dot{\varphi}_{\theta_0}(t)}{\|h\|}\right]^2 \mathrm{d}P_{\theta_0}(t) = 0\,. \tag{3}$$

The symbol h' means a row vector and $h'z = h_1 z_1 + \ldots + h_m z_m$. Further, the covariance matrix

$$\Gamma(\theta_0) = 4 \operatorname{cov}(\dot{\varphi}_{\theta_0}) \tag{4}$$

of the vector $2\dot{\varphi}_{\theta_0}$ is supposed to be regular.

Examples of some distributions satisfying these conditions are in [7]. The purpose of this paper is to prove that the exponential probabilities satisfy these conditions, which is done in Theorem 1. As is shown in [5], though the class of exponential families is very large, many usual distributions assume the exponential form after some reparametrization only. The explicit formula for calculation of the quadratic derivative of a probability reparametrized into an exponential family is given in Theorem 2. The paper is concluded by an example of such a calculation.

MAIN RESULTS

We shall suppose in this part that $\mathscr{P} = \{P_\theta; \theta \in \Theta\}$ are probabilities on $(R^k, \mathscr{B}^k)$, $\Theta \subset R^k$ and the densities (1) have the form

$$f_\theta(t) = C(\theta) \exp(\theta' t), \tag{5}$$

where $\theta' t = \theta_1 t_1 + \ldots + \theta_k t_k$.

Theorem 1. *Let $\theta_0 \in \Theta$ and in accordance with (2) let*

$$\varphi_\theta(t) = \left[\frac{C(\theta)}{C(\theta_0)} \exp((\theta - \theta_0)' t)\right]^{1/2}. \tag{6}$$

Further, let θ_0 be an inner point of the set Θ.

(I) *Let $T_j(t) = t_j$ be the projection into the j-th coordinate. The function*

$$\dot{\varphi}_{\theta_0}(t) = \begin{bmatrix} \frac{1}{2}(t_1 - E_{\theta_0}(T_1)) \\ \vdots \\ \frac{1}{2}(t_k - E_{\theta_0}(T_k)) \end{bmatrix} \tag{7}$$

is the quadratic derivative of (6), i.e. it satisfies the relation (3).

(II) *If for each vector $h \in R^k$ there is a number $\beta \in (0, 1)$ such that the probabilities $P_{\theta_0 + \beta h}$, P_{θ_0} are different, then the matrix $\Gamma(\theta_0)$ defined by (4) and (7) is regular.*

Proof. Since θ_0 is an inner point of Θ, there is $\varepsilon > 0$ such that $U = \{\theta \in R^k; \|\theta - \theta_0\| < \varepsilon\}$ is a subset of Θ. According to Theorem 9, Chapter 2 in [5], the function

$$q(z_1, \ldots, z_k) = \int_{R_k} \exp\left(\sum_1^k z_j t_j\right) \mathrm{d}\mu(t) \tag{8}$$

of complex variables $z_1, \ldots, z_k$, has derivatives of all orders on the set

$$D = \{(\xi_1 + i\eta_1, \ldots, \xi_k + i\eta_k)\,; \quad (\xi_1, \ldots, \xi_k) \in U\}$$

and these derivatives may be computed by differentiating under the integration sign. Since the set D is homeomorphic to $U \times R^k$ and the topological product of connected spaces is connected according to Theorem 6.1.4 in [1], the set D is connected and open, and Hartog's theorem (cf. [6] p. 277) implies that the derivatives are continuous. Thus, for $\theta \in \Theta$,

$$\text{(9)} \qquad \frac{\partial q(\theta_1, \ldots, \theta_k)}{\partial \theta_1^{r_1} \ldots \partial \theta_k^{r_k}} = \int t_1^{r_1} \ldots t_k^{kr} \exp\left(\sum_{j=1}^{k} \theta_j t_j\right) \mathrm{d}\mu(t)$$

which means that the function $\varphi_{\theta_0+h} - 1 - h'\varphi_{\theta_0}$ belongs to $L_2(P_{\theta_0})$ provided $\|h\|$ is sufficiently small. Further, the relation (9) implies that

$$\frac{\partial C(\theta)}{\partial \theta_j} = -C(\theta)\, E_\theta(T_j)\,, \quad \frac{\partial \varphi_\theta(t)}{\partial \theta_j} = \tfrac{1}{2}\varphi_\theta(t)\,[t_j - E_\theta(T_j)]\,,$$

$$\frac{\partial^2 \varphi_\theta(t)}{\partial \theta_s\, \partial \theta_j} = \tfrac{1}{2}\varphi_\theta(t)\,[\tfrac{1}{2}(t_s - E_\theta(T_s))\,(t_j - E_\theta(T_j)) - \operatorname{cov}_\theta(T_s, T_j)]$$

where $\operatorname{cov}_\theta(T_s, T_j)$ is the covariance of T_s, T_j with respect to the measure P_θ. Hence, by means of Taylor's theorem for functions of k variables, we obtain

$$\varphi_{\theta_0+h}(t) = 1 + \sum_{j=1}^{k} \tfrac{1}{2}(t_j - E_{\theta_0}(T_j))\, h_j + \sum_{s,j=1}^{k} \frac{1}{2} \frac{\partial^2 \varphi_{\tilde\theta}(t)}{\partial \theta_s\, \partial \theta_j} h_s h_j\,,$$

$$\tilde\theta = \theta(h, t) \in \overline{\theta_0, \theta_0 + h}\,.$$

This equality means that

$$\text{(10)} \qquad \|h\|^{-1}\, |\varphi_{\theta_0+h}(t) - 1 - \sum_{j=1}^{k} \tfrac{1}{2}(t_j - E_{\theta_0}(T_j))\, h_j| \leqq \|h\| \sum_{s,j=1}^{k} Q_{s,j}(h, t)$$

$$Q_{sj}(h,, t) = |\varphi_{\tilde\theta}(t)|\, [|t_s - E_{\tilde\theta}(T_s)|\, |t_j - E_{\tilde\theta}(T_j)| + |\operatorname{cov}_{\tilde\theta}(T_s, T_j)|]\,.$$

Now we see that the relation (3) (cf. (7)) will be proved, if we find functions $f_{sj}(t) \in L_2(P_{\theta_0})$ such that $Q_{s,j}(h, t) \leqq f_{sj}(t)$ for all t and all h with $\|h\|$ sufficiently small. But the mentioned continuity of the quantities (9) implies that the functions $C(\theta)$, $E_\theta(T_s)$, $\operatorname{cov}_\theta(T_s, T_j)$ are continuous at $\theta = \theta_0$, hence for $\delta > 0$ sufficiently small and $\|\tilde\theta - \theta_0\| \leqq \delta$ we have

$$|C(\theta_0) - C(\tilde\theta)| < 1\,, \quad |E_{\theta_0}(T_j) - E_{\tilde\theta}(T_j)| < 1\,,$$

$$|\operatorname{cov}_{\theta_0}(T_s, T_j) - \operatorname{cov}_{\tilde\theta}(T_s, T_j)| < 1$$

for all s, j. Thus for $h \in R^k$, $\|h\| \leqq \delta$ we may write

$$(11) \qquad 0 \leqq Q_{s,j}(h, t) \leqq |\varphi_{\bar\delta}(t)|\, q_{sj}(t)\,,$$

$$q_{sj}(t) = (|t_s - E_{\theta_0}(T_s)| + 1)(|t_j - E_{\theta_0}(T_j)| + 1) + 1 + |\mathrm{cov}_{\theta_0}(T_s, T_j)|$$

where (9) and the Hölder-Minkowski inequality imply that

$$(12) \qquad q_{sj}(t) \in L_4(P_{\theta_0})\,.$$

Further, if $\|h\| \leqq \delta/2k$, then

$$(13) \qquad |\varphi_{\bar\delta}(t)| \leqq \left(\frac{C(\theta_0) + 1}{C(\theta_0)}\right)^{1/2} \prod_{j=1}^{k} \exp\left(\frac{\delta}{4k}\,|t_j|\right).$$

Since

$$(14) \qquad \exp\left(\frac{\delta}{4k}\,|t_j|\right) = \max\left\{\exp\left(\frac{\delta}{4k}\,t_j\right),\ \exp\left(-\frac{\delta}{4k}\,t_j\right)\right\}$$

and the points $(\theta_0(1), \ldots, \theta_0(i-1), \theta_0(i) + \bar\delta, \theta_0(i+1), \ldots, \theta_0(k))$ belong to U for $\bar\delta \in \{\delta, -\delta\}$ and $i = 1, \ldots, k$, the function (14) belongs to $L_{4k}(P_{\theta_0})$. The generalized Hölder-Minkowski inequality implies that the right hand side of (13) belongs to $L_4(P_{\theta_0})$, which together with (11) and (12) completes the proof.

(II) If the matrix $\Gamma(\theta_0)$ is singular, then there exists a non-zero vector $h \in R^k$ such that

$$\sum_{j=1}^{k} h_j(T_j(t) - E_{\theta_0}(T_j)) = 0 \quad \mathrm{mod}\ P_{\theta_0}\,.$$

We may assume without loss of generality that the points $\theta_0 + \beta h, \beta \in \langle 0, 1\rangle$ belong to Θ and for $\beta \in (0, 1)$ we obtain

$$\frac{\mathrm{d}P_{\theta_0 + \beta h}}{\mathrm{d}P_{\theta_0}}(t) = \frac{C(\theta_0 + \beta h)}{C(\theta_0)} \exp\left(\beta h' E_{\theta_0}(T)\right] \quad \mathrm{mod}\ P_{\theta_0}\,,$$

which contradicts the assumptions.

Theorem 2. *Let a class of probabilities $\{Q_\gamma; \gamma \in \Gamma\}$, where $\Gamma \subset R^m$ is an open subset, be defined on $(X, \mathscr{B})$ by the densities*

$$f_\gamma(x) = \frac{\mathrm{d}Q_\gamma(x)}{\mathrm{d}\nu}\,,$$

where ν is a σ-finite measure. Let $\mathscr{P} = \{P_\theta; \theta \in \Theta\}$ be the class of probabilities defined by the densities (5) and let $\Theta = \{\theta \in R^k; \int \exp(\theta' t)\, \mathrm{d}\mu(t) < \infty\}$. If $T : X \to R^k$ is such a measurable transformation that $\mu(A) = \nu(T^{-1}A)$ and if $L : \Gamma \to \Theta$ is such a mapping that

(I) $f_\gamma(x) = C[L(\gamma)] \exp[L(\gamma)'\, T(x)] \bmod \mu$ *for each γ belonging to some neighbourhood V of γ_0;*

(II) $L'(\gamma) = (L_1(\gamma), \ldots, L_k(\gamma))$ *and the functions* $L_1, \ldots, L_k$ *have all partial derivatives of the first order on* V, *which are continuous at* γ_0;

(III) *there exist numbers* $\delta_i < 0 < \delta_i^*$ *such that the points* $(L_1(\gamma_0), \ldots, L_{i-1}(\gamma_0), L_i(\gamma_0) + \bar{\delta}_i, L_{i+1}(\gamma_0), \ldots, L_k(\gamma_0))$ *belong to* Θ *for* $\bar{\delta}_i \in \{\delta_i, \delta_i^*\}$ *and for* $i = 1, \ldots, k$; *then the function*

$$p_\gamma(x) = \left(\frac{f_\gamma(x)}{f_{\gamma_0}(x)}\right)^{1/2}$$

has the quadratic derivative $\dot{p}_{\gamma_0}$ *at* $\gamma = \gamma_0$, *and*

$$\dot{p}_{\gamma_0}(x) = J(\gamma_0)\begin{bmatrix} \frac{1}{2}(T_1(x) - E_{\gamma_0}(T_1)) \\ \vdots \\ \frac{1}{2}(T_k(x) - E_{\gamma_0}(T_k)) \end{bmatrix}, \tag{15}$$

where $J(\gamma_0)$ *is the jacobian of the mapping* L *at the point* γ_0. *Moreover, if the point* $\theta_0 = L(\gamma_0)$ *satisfies the condition* (II) *of the preceding theorem and the rank of the matrix* $J(\gamma_0)$ *is* k, *then the matrix*

$$\Gamma(\gamma_0) = 4E_{\gamma_0}(\dot{p}_{\gamma_0}(x)\, \dot{p}_{\gamma_0}(x)')$$

is regular.

Proof. Since Θ is a convex set by Lemma 7, Chapter II in [5], it follows from the condition III of the theorem that $\theta_0 = L(\gamma_0)$ is an inner point of Θ, therefore the function $\dot{\varphi}_{\theta_0}$ defined by the formula (7) is the quadratic derivative of the function φ_θ (cf. (6)) with respect to P_{θ_0}. The rest of the proof follows from the fact that the functions $L_1, \ldots, L_k$ are differentiable, the relation (3) holds and that the matrix J is of full rank.

As an example of the particular situation described in Theorem 2, let $\Gamma = Rx(0, \infty)$ and for $\gamma = (M, d) \in \Gamma$ let

$$f_\gamma(x) = (2\pi\, \mathrm{d})^{-1/2} \exp\left(-\frac{(x - M)^2}{2d}\right), \quad x \in R$$

i.e. $\{Q_\gamma; \gamma \in \Gamma\}$ are the normal distributions with the mean M and the dispersion d. If we denote

$$T'(x) = (x, -x^2), \quad L(\gamma)' = \left(\frac{M}{\mathrm{d}}, \frac{1}{2\mathrm{d}}\right)$$

and define a measure μ on $(R^2, \mathscr{B}^2)$ by the formula

$$\mu(A) = \nu(T^{-1}A),$$

where ν is the Lebesque measure on the line, the formula (15) has the form

$$\dot{p}_\gamma(x) = \begin{bmatrix} \dfrac{x - M}{2\mathrm{d}} \\ \dfrac{(x - M)^2 - \mathrm{d}}{4\mathrm{d}^2} \end{bmatrix}.$$

References

[1] *R. Engelking:* An Outline of General Topology (in Polish). Warszawa 1968.

[2] *R. A. Johnson, G. G. Roussas:* Asymptotically most powerful tests in Markov processes. Ann. Math. Stat. 40 (1969), 1207–1215.

[3] *R. A. Johnson, G. G. Roussas:* Asymptotically optimal tests in Markov processes. Ann. Math. Stat. 41 (1970), 918–938.

[4] *R. A. Johnson, G. G. Roussas:* Applications of Contiguity to Multiparameter Hypotheses Testing. Proceedings of the Sixth Berkeley Symposium on Math. Stat. and Prob.

[5] *E. L. Lehman:* Testing Statistical Hypotheses. New York 1959.

[6] *F. Leja:* Complex Functions (in Polish). Warszawa 1971.

[7] *G. G. Roussas:* Контигуальность вероятностных мер. Москва 1975.

Súhrn

O KVADRATICKÝCH DERIVÁCIÁCH POMERU VIEROHODNOSTÍ

František Rublík

V článku sa dokazuje, že exponenciálne triedy pravdepodobností majú kvadraticky diferencovateľný pomer vierohodností a odvádzajú sa explicitné formuly pre túto kvadratickú deriváciu.

Author's address: RNDr. *František Rublík*, Ústav merania a meracej techniky SAV, Patrónka-Dúbravská cesta, 885 27 Bratislava.