Probab. Th. Rel. Fields 76, 103–119 (1987)

Probability Theory and Related Fields

Optical Local Gaussian Approximation of an Exponential Family*

Enno Mammen

Universität Heidelberg, Sonderforschungsbereich 123, Im Neuenheimer Feld 294, D-6900 Heidelberg, Federal Republic of Germany

Summary. Under certain regularity conditions products $\mathscr{E}^n$ of an experiment $\mathscr{E}$ can be locally approximated by homoschedastic Gaussian experiments $\mathscr{G}_n$. $\mathscr{G}_n$ can be defined such that the square roots of the densities have nearly the same structure with respect to the L^2-geometry as in $\mathscr{E}^n$. The main result of this paper is that this choice of $\mathscr{G}_n$ is asymptotically optimal in the sense of minimizing the deficiency distance between $\mathscr{E}^n$ and $\mathscr{G}$ if $\mathscr{E}$ is a one-dimensional exponential family.

1. Introduction

Families of product measures fulfilling certain regularity conditions can be locally approximated by Gaussian experiments. This result is due to Wald (1943) and was further studied for instance by LeCam (1956, 1960, 1968), and Michel and Pfanzagl (1970), Pfanzagl (1972).

A natural statistical quantity for comparing two families of distributions is the *deficiency distance* due to LeCam (1964) which is based on the comparison of risk functions available in the two experiments. For sufficiently regular experiments $\mathscr{E}=(P_\theta\colon \theta\in\Theta)$ and $\mathscr{F}=(Q_\theta\colon \theta\in\Theta)$ the deficiency distance $\Delta(\mathscr{E},\mathscr{F})$ can be calculated as follows. Let

$$\delta(\mathscr{E},\mathscr{F})=\inf_K \sup_{\theta\in\Theta} \|KP_\theta-Q_\theta\|, \tag{1.1}$$

where the infimum is taken over all Markov kernels K between the measurable spaces on which (P_θ) and (Q_θ) respectively are defined. The deficiency distance is defined as

$$\Delta(\mathscr{E},\mathscr{F})=\delta(\mathscr{E},\mathscr{F})\vee\delta(\mathscr{F},\mathscr{E}).$$

In this paper we restrict ourselves to the special case of a one-dimensional

* This work has been supported by the Deutsche Forschungsgemeinschaft

exponential family (P_θ)

$$dP_\theta = \exp(\theta x - \psi(\theta))\, dP_0. \tag{1.2}$$

We choose a point θ_0 in the interior of the natural parameter space and we consider the local experiment

$$\mathscr{E}_n = (P_\theta : |\theta - \theta_0| \leqq c_n/\sqrt{n}), \tag{1.3}$$

where (c_n) is a sequence fulfilling for technical reasons:

$$c_n \to \infty, \qquad c_n = o(n^{1/6}). \tag{1.4}$$

For convenience of notation we assume

$$\theta_0 = 0, \qquad \psi'(0) = 0, \qquad \psi''(0) = 1. \tag{1.5}$$

Furthermore, let

$$\gamma_3 = \psi'''(0), \qquad \Theta_n = (-c_n/\sqrt{n}, c_n/\sqrt{n}). \tag{1.6}$$

Then functions $\mu_n(\cdot): \Theta_n \to \mathbb{R}$ can be chosen such that the Gaussian experiment $\mathscr{G}(n, \mu_n) = (N(\mu_n(\theta), 1): \theta \in \Theta_n)$ approximates $\mathscr{E}_n^n$

$$\Delta(\mathscr{E}_n^n, \mathscr{G}(n, \mu_n)) \to 0 \qquad (\text{for } n \to 0). \tag{1.7}$$

One possible choice of $\mu_n(\cdot)$ is based on the Hellinger distance H. We define

$$\mu_n^H(\theta) = 2 \operatorname{sgn}(\theta) \sqrt{n} H(P_0, P_\theta), \tag{1.8}$$

where

$$H^2(\mu, v) = \int (\sqrt{d\mu} - \sqrt{dv})^2 \tag{1.9}$$

denotes the Hellinger distance between two measures μ and v. This definition is motivated by the Hilbert space parametrized Gaussian aproximation

$$\tilde{\mathscr{G}}_n^H = (N(\xi_{\theta,n}, I): \theta \in \Theta_n), \tag{1.10}$$

where $\xi_{\theta,n} = 2\sqrt{n}\sqrt{dP_\theta/dP_0} \in L_2(P_0)$ and I is the identity operator (see Proposition 4 in Müller (1979) and see Millar (1979)). (This definition of $\tilde{\mathscr{G}}_n^H$ is formally correct only in the case where Θ_n is finite. Then the Gaussian measures may be defined on the finite dimensional linear space spanned by $\{\xi_{\theta,n}: \theta \in \Theta_n\}$.) $\tilde{\mathscr{G}}_n^H$ is parametrized such that in $\mathscr{E}_n^n$ and in $\tilde{\mathscr{G}}_n^H$ the square roots of the densities form nearly the same L^2-structure

$$\begin{aligned} H^2(N(\xi_{\theta,n}, I), N(\xi_{\tau,n}, I)) &= 2(1 - \exp(-\tfrac{1}{8} \|\xi_{\theta,n} - \xi_{\tau,n}\|^2_{L_2(P_0)})) \\ &= 2(1 - \exp(-\tfrac{1}{2} n H^2(P_\theta, P_\tau)) \\ &= H^2(P_\theta^n, P_\tau^n) + O(1/n) \end{aligned} \tag{1.11}$$

uniformly in $\theta, \tau \in \Theta_n$. For a more detailed discussion of $\tilde{\mathscr{G}}_n^H$ see LeCam (1986).

For the case of an exponential family $\mathscr{E}$ it can be checked that

$$|\mu_n^H(\theta) - \mu_n^H(\tau)| = 2\sqrt{n} H(P_\tau, P_\theta) + O(1/n) = \|\xi_{\theta,n} - \xi_{\tau,n}\|_{L_2(P_0)} + O(1/n) \tag{1.12}$$

uniformly for $\theta, \tau \in \Theta_n$. This implies

$$\Delta(\mathscr{G}_n^H, \tilde{\mathscr{G}}_n^H) = O(1/n) \tag{1.13}$$

with $\mathscr{G}_n^H = (N(\mu_n^H(\theta), 1) \colon \theta \in \Theta_n)$.

In this paper second order edgeworth expansions are used to compare different Gaussian approximations for an exponential family $\mathscr{E}_n$. The main result is that $\tilde{\mathscr{G}}_n^H$ (or $\mathscr{G}_n^H$ resp.) is asymptotically optimal in the sense of minimizing

$$\lim_{n \to \infty} \sqrt{n}\, \Delta(\mathscr{G}(n, \mu_n), \mathscr{E}_n^n)$$

(see Theorem 2). Initially the Hilbert-space parametrization has been introduced because of its clear mathematical structure which leads to a simplified statistical analysis of product experiments. This optimality result shows that the Hellinger distance has not only an asymptotic interpretation, but it has also a finite sample meaning.

Another characterization of the Hilbert space parametrization has been given in LeCam (1985). There it has been pointed out that under mild conditions if a product can be approximated by some Gaussian experiment it must also be approximiable by $\tilde{\mathscr{G}}_n^H$. This has an interpretation which is related to this paper: For finite n the approximation $\tilde{\mathscr{G}}_n^H$ could be expected to be accurate – compared with other Gaussian approximations – because only relatively weak conditions are required.

2. Results

Our first theorem states some asymptotic bounds for the accuracy of Gaussian approximations. $\mathscr{G}_n^H$ will be compared with the common approximations based on the asymptotic normality of the derivative of the log likelihood function or of the maximum likelihood estimate, respectively

$$\mathscr{G}_n^D = (N(\sqrt{n}\, \psi'(\theta), 1) \colon \theta \in \Theta_n)$$

$$\mathscr{G}_n^{ML} = (N(\sqrt{n}\, \theta, 1) \colon \theta \in \Theta_n).$$

Theorem 1. *Assume* (1.4) *and* (1.5). *Then*

$$\sqrt{\frac{2}{\pi e}}\, |\gamma_3| \frac{(\sqrt{c_n} - 1)^2}{\sqrt{n}} + o(1/\sqrt{n}) \leq \Delta(\mathscr{G}_n^i, \mathscr{E}_n^n) \leq \sqrt{\frac{2}{\pi e}}\, |\gamma_3| \frac{c_n}{\sqrt{n}} + o(1/\sqrt{n}) \quad \textit{for } i = D, ML. \tag{2.1}$$

$$\frac{1}{2}\sqrt{\frac{2}{\pi e}}\, |\gamma_3| \frac{c_n}{\sqrt{n}} + o(1/\sqrt{n}) \leq \Delta(\mathscr{G}_n^i, \mathscr{E}_n^n) \quad \textit{for } i = D, ML. \tag{2.2}$$

$$0.046 \frac{|\gamma_3|}{\sqrt{n}} + o(1/\sqrt{n}) \leq \Delta(\mathscr{G}_n^H, \mathscr{E}_n^n) \leq \frac{\ln(2)}{3\sqrt{2\pi}} \frac{|\gamma_3|}{\sqrt{n}} + o(1/\sqrt{n}), \quad \textit{where } \mathscr{G}_n^H = (N(\mu_n^H(\theta), 1) \colon \theta \in \Theta_n). \tag{2.3}$$

Table 1. The asymptotic bounds of Theorem 1 calculated for a binomial experiment

		$n=30$	$n=50$	$n=200$	$n=1000$
$\Delta(\mathscr{G}_n^i, \mathscr{E}_n^n)$	as. $\leqq$	0.38	0.38	0.38	0.38
$(i=D, ML)$	as. $\geqq$ (2.1)	0.08	0.10	0.12	0.19
	as. $\geqq$ (2.2)	0.19	0.19	0.19	0.19
$\Delta(\mathscr{G}_n^H, \mathscr{E}_n^n)$	as. $\leqq$	0.022	0.017	0.014	0.004
	as. $\geqq$	0.011	0.009	0.007	0.002

Because of $\frac{\ln(2)}{3\sqrt{2\pi}} \approx 0.0922$ the asymptotic upper bound in (2.3) is approximately twice the lower bound.

The asymptotic bounds of Theorem 1 are calculated in Table 1 for a binomial experiment with different numbers n of observations and parameter space

$$\Theta_n = \{\theta: 1/8 \leqq \psi'(\theta) \leqq 3/8\}.$$

The expansions leading to the bounds in Theorem 1 use the fact that for large n the skewness $\psi'''(\theta)$ is nearly constant in Θ_n. This is not the case in the example above where $\psi'''(\theta)$ varies from 0.5 to 2.3. Therefore the numbers of Table 1 should rather be interpreted as indicating the order of the deficiencies than as being exact bounds. Here the lower bound of (2.1) is very poor compared with the bound of (2.2). Furthermore, the approximation $\mathscr{G}_n^H$ is much more accurate than $\mathscr{G}_n^D$ or $\mathscr{G}_n^{ML}$ respectively. This is clear from (2.3) because only $\mathscr{G}_n^H$ leads to an approximation of order $1/\sqrt{n}$, whereas in the other two cases one gets approximations of the slower order $c_n/\sqrt{n}$. The main result of this paper is that the approximation $\mathscr{G}_n^H$ is optimal in a general sense.

Theorem 2. *Assume* (1.4) *and* (1.5). *For an arbitrary function* $\mu_n(\cdot): \Theta_n \to \mathbb{R}$ *define the Gaussian experiment* $\mathscr{G}_n = (N(\mu_n(\theta), 1): \theta \in \mathbb{R}_n)$. *Then*

$$\liminf_{n \to +\infty} \sqrt{n}\,\Delta(\mathscr{G}_n, \mathscr{E}_n^n) \geqq \lim_{n \to \infty} \sqrt{n}\,\Delta(\mathscr{G}_n^H, \mathscr{E}_n^n). \tag{2.4}$$

Our proof of Theorem 2 is based on

$$\delta(\mathscr{E}_n^n, \mathscr{G}_n^H) = \delta(\mathscr{G}_n^H, \mathscr{E}_n^n) + o(1/\sqrt{n}). \tag{2.5}$$

We will show that every replacement of $\mathscr{G}_n^H$ by a different Gaussian experiment will increase at least one of the two deficiencies in (2.5).

We expect that the Hellinger parametrization is asymptotically optimal (in the sense of Theorem 2) for a broader class of experiments than the relatively special case of taking i.i.d. observations of a one-dimensional exponential family. In a subsequent paper generalizations to experiments fulfilling Cramér's type conditions will be studied. There the proof will make use of the fact that up to second order ($o(1/\sqrt{n})$) such experiments are equivalent to mixtures of exponential families. Other generalizations can be proved straightforwardly. For instance the i.i.d. structure is not needed here. In the proof of Theorem 2 it has only been used that up to $o(1/\sqrt{n})$ $\mathscr{E}_n^n$ can be approximated by exponential families which are asymptotically Gaussian. Another modification of Theorem 2 would be to consider higher dimensional exponential families. In these two cases $\mathscr{G}_n^H$ has to be chosen – as in the one dimensional case – as a homoschedastic Gaussian experiment such that the respective L^2-geometries of the square roots of the densities (in $\mathscr{G}_n^H$ and in the considered sequence of experiments) agree up to order $o(1/\sqrt{n})$. Such a Gaussian experiment can easily be constructed, but in general a simple explicit form is not available as in the case of a one-dimensional parametric product experiment (see (1.8)).

The deficiency distance is based on the comparison of the risk functions for *all* decision problems with bounded loss functions. It should be pointed out that the approximation $\mathscr{G}_n^H$ is also asymptotically optimal if the comparison of $\mathscr{G}_n$ and $\mathscr{E}_n^n$ is based only on certain *subclasses* of decision problems, for instance if the number of possible decisions is bounded by a fixed constant. We expect that it suffices that the class of decision problems is closed under shifts of the parameterspace in the Hellinger parametrization (see the 6th step of the proof of Theorem 2). Here we want to formulate only another modification of Theorem 2 for binary experiments.

Corollary. *Let* (θ_n), (τ_n) *be two sequences in* $\mathbb{R}$ *with* $\theta_n = O(1/\sqrt{n})$, $\tau_n = O(1/\sqrt{n})$, *and* $\sqrt{n}(\theta_n - \tau_n) \to \text{const.}$ *for* $n \to \infty$. *Let* $\mathscr{G}_n$ *denote the binary experiment* $\mathscr{G}_n = (N(0,1), N(m_n, 1))$, *where* m_n *is an arbitrary sequence. Then*

$$\liminf_{n\to\infty} \sqrt{n}\,\Delta(\mathscr{F}_n^n, \mathscr{G}_n) \geqq \lim_{n\to+\infty} \sqrt{n}\,\Delta(\mathscr{F}_n^n, \mathscr{G}_{n,2}^H), \tag{2.6}$$

where

$$\mathscr{F}_n = (P_{\theta_n}, P_{\tau_n}),$$
$$\mathscr{G}_{n,2}^H = (N(0,1), N(2\sqrt{n}\,H(P_{\theta_n}, P_{\tau_n}), 1)).$$

Before going into the proofs of Theorems 1 and 2 in the next section we now give a heuristic explanation of the optimality of $\mathscr{G}_n^H$.

Firstly let

$$P_{\theta,n} = \mathscr{L}\left(1/\sqrt{n}\sum_{i=1}^{n} X_i \,|\, P_\theta^n\right) \quad \text{for } \theta \in \Theta_n. \tag{2.7}$$

In the case where P_θ is a nonlattice distribution a first order Edgeworth expansion yields the following approximation for $P_{\theta,n}$

$$dQ_{\theta,n} = \phi(x - \sqrt{n}\,\theta)\left(1 + \frac{\gamma_3}{6\sqrt{n}}(x^3 - 3x - (\sqrt{n}\,\theta)^3)\right) d\lambda. \tag{2.8}$$

Here λ denotes the Lebesgue measure. In the sense of deficiency distance this approximation is also valid in the lattice case:

Proposition. *Assume $c_n = o(n^{1/6})$ and (1.5). Then*

$$\sup_{\theta \in \Theta_n} \| L P_{\theta,n} - Q_{\theta,n} \| = o(1/\sqrt{n}) \tag{2.9}$$

$$\sup_{\theta \in \Theta_n} \| L^* Q_{\theta,n} - P_{\theta,n} \| = o(1/\sqrt{n}), \tag{2.10}$$

where for some β with $1/3 < \beta < 1/2$ the kernel L is defined by

$$L(x, dy) = n^\beta \phi(n^\beta (y - x))\, dy$$

and L^* is the dual kernel

$$L \times P_{0,n} = L P_{0,n} \times L^*.$$

(For a measure P and a kernel K the measure $K \times P$ is defined by

$$K \times P(A \times B) = \int_B K(x, A)\, P(dx).)$$

The proposition entails that up to order $o(1/\sqrt{n})$ the experiments $\mathscr{E}_n^n$ and $(Q_{\theta,n}: \theta \in \Theta_n)$ are asymptotically equivalent (measured by the deficiency distance).

We consider the following (deterministic) transformations T_b of the smoothed data (or of the data in the experiment $(Q_{\theta,n}: \theta \in \Theta_n)$ resp.).

$$T_b(x) = x + \gamma_3 \frac{b}{\sqrt{n}} (x^2 - 1). \tag{2.11}$$

The following lemma gives an approximation for

$$Q_{\theta,n,b} = \mathscr{L}(T_b(X) \mid Q_{\theta,n}).$$

Lemma. *Assume $c_n = o(n^{1/6})$. Then*

$$\sup_{\theta \in \Theta_n} \| Q_{\theta,n,b} - R_{\theta,n,b} \| = o(1/\sqrt{n}), \tag{2.12}$$

where

$$\frac{dR_{\theta,n,b}}{d\lambda} = \phi(x - t) + \frac{\gamma_3}{6\sqrt{n}} (1 + 6b)\, s(x - t) + \frac{\gamma_3}{\sqrt{n}} (2b + \tfrac{1}{2})\, t\, v(x - t)$$
$$+ \frac{\gamma_3}{\sqrt{n}} (b + \tfrac{1}{2})\, t^2 m(x - t).$$

Here

$$t = \sqrt{n}\,\theta \quad \text{and} \quad s(y) = (y^3 - 3y)\,\phi(y), \quad v(y) = (y^2 - 1)\,\phi(y), \quad m(y) = y\,\phi(y).$$

The term in the definition of $R_{\theta,n,b}$ which contains $s(x - t)$ (or $v(x - t)$ or $m(x - t)$ resp.) can be interpreted as a small difference of $R_{\theta,n,b}$ in the skewness (or in the variance or in the mean resp.) compared with $N(\sqrt{n}\,\theta, 1)$. If t is of order c_n (which according to (1.4) is supposed to tend to ∞) then the skewness term is of order $1/\sqrt{n}$ (if $b \neq -1/6$) and the variance term is of order $c_n/\sqrt{n}$ (if $b \neq -1/4$) and the term related to the mean is of order $c_n^2/\sqrt{n}$ (if $b \neq -1/2$).

The proposition and the lemma imply that

$$\Delta(\mathscr{E}_n^n, (R_{\theta,n,b}: \theta \in \Theta_n)) = o(1/\sqrt{n}) \qquad \text{for } b \in \mathbb{R}. \tag{2.13}$$

After these preparatory calculations we turn now to the problem of determing the parametrization $\mu_n(\theta)$ of the optimal Gaussian approximation $\mathscr{G}(n, \mu_n)$. Suppose that (the optimal) $\mu_n(\cdot)$ is a smooth function

$$\mu_n(\theta) = t + \alpha \frac{\gamma_3}{\sqrt{n}} t^2 + o(1/\sqrt{n}) \tag{2.14}$$

uniformly in $t = \sqrt{n}\,\theta \in [-c_n, c_n]$ for a fixed constant α. Then

$$\sup_{\theta \in \Theta_n} \|N(\mu_n(\theta), 1) - S_{\theta,n,\alpha}\| = o(1/\sqrt{n})$$

where

$$\frac{dS_{\theta,n,\alpha}}{d\lambda} = \phi(x-t) + \frac{\alpha\gamma_3}{\sqrt{n}} t^2 m(x-t).$$

With (2.13) this gives for $b \in \mathbb{R}$

$$\begin{aligned} \Delta(\mathscr{E}_n^n, \mathscr{G}(n, \mu_n)) &= \Delta((R_{\theta,n,b}: \theta \in \Theta_n), (S_{\theta,n,\alpha}: \theta \in \Theta_n)) + o(1/\sqrt{n}) \\ &\leq \sup_{\theta \in \Theta_n} \|R_{\theta,n,b} - S_{\theta,n,\alpha}\| + o(1/\sqrt{n}). \end{aligned} \tag{2.15}$$

For the special choice $b = b(\alpha) = \alpha - 1/2$ the term related to the change of the mean coincide in $dS_{\theta,n,\alpha}/d\lambda$ and $dR_{\theta,n,b}/d\lambda$. Especially the upper bound of (2.15) is then of order $c_n/\sqrt{n}$ (instead of $c_n^2/\sqrt{n}$ as for other choices of b). Furthermore – as stated in the following theorem – for $b = b(\alpha)$ the upper and lower bound in (2.15) differ only by a term of order $o(c_n/\sqrt{n})$.

Theorem 3. *Assume* (1.4), (1.5) *and* (2.14). *Then*

$$\begin{aligned} \Delta(\mathscr{E}_n^n, \mathscr{G}(n, \mu_n)) &= \sup_{\theta \in \Theta_n} \|R_{\theta,n,b(\alpha)} - S_{\theta,n,\alpha}\| + o(c_n/\sqrt{n}) \\ &= |4\alpha - 1| \frac{\gamma_3 c_n}{\sqrt{n}} \sqrt{\frac{2}{\pi e}} + o(c_n/\sqrt{n}). \end{aligned}$$

Especially for the Hellinger parametrization one can easily check that

$$\mu_n^H(\theta) = 2\sqrt{n} H(P_0, P_\theta) = t + \frac{\gamma_3}{4\sqrt{n}} t^2 + o(1/\sqrt{n}).$$

That means $\alpha=1/4$. According to Theorem 3 only for $\alpha=1/4$ the accuracy of the Gaussian approximation is of order $o(c_n/\sqrt{n})$ – given that the Gaussian approximation is smoothly parametrized. The reason is that only for $b=b(1/4)=-1/4$ the transformation T_b is asymptotically variance stabilizing (the variance term in $dR_{\theta,n,b}/d\lambda$ vanishes). Heuristically this connection between variance stabilization and Hellinger parametrization may be explained by the fact that the parametrization $\xi_{\theta,n}$ of the Gaussian approximation $\mathscr{G}_n^H$ (see (1.10)) does not depend on θ_0 and the local neighborhood Θ_n. Or – more explicitly – consider a binomial experiment $(B(n,p)\colon p\in(0,1))$. Then the arc sin transformation is known to be asymptotically variance stabilizing

$$\mathscr{L}\left(2\sin^{-1}\left(\sqrt{\frac{1}{n}x}\right)\middle| B(n,p)\right)\approx N\left(2\sin^{-1}(\sqrt{p}),\frac{1}{n}\right).$$

The asymptotic mean is closely related to the Hellinger parametrization.

$$\begin{aligned}
2\sin^{-1}(\sqrt{p})-2\sin^{-1}(\sqrt{q})&=2\cos^{-1}(\sqrt{q})-2\cos^{-1}(\sqrt{p})\\
&=2\sphericalangle\left(\begin{pmatrix}\sqrt{q}\\ \sqrt{1-q}\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix}\right)-2\sphericalangle\left(\begin{pmatrix}\sqrt{p}\\ \sqrt{1-p}\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix}\right)\\
&=2\sphericalangle\left(\begin{pmatrix}\sqrt{q}\\ \sqrt{1-q}\end{pmatrix},\begin{pmatrix}\sqrt{p}\\ \sqrt{1-p}\end{pmatrix}\right)\\
&=2\cos^{-1}(\sqrt{p}\sqrt{q}+\sqrt{1-p}\sqrt{1-q})\\
&=2\cos^{-1}(1-\tfrac{1}{2}H^2(B(1,p),B(1,q)))\\
&=2H(B(1,p),B(1,q))+o(H(B(1,p),B(1,q)))\\
&\quad\text{for } 0<q<p<1.
\end{aligned}$$

3. Proofs

Proof of the proposition. Because of $\|LQ_{\theta,n}-Q_{\theta,n}\|=o(1/\sqrt{n})$ it is enough to prove for (2.9)

$$\sup_{\theta\in\Theta_n}\|LQ_{\theta,n}-LP_{\theta,n}\|=o(1/\sqrt{n}). \tag{3.1}$$

Let $\chi_{\theta,n}(t)$, $\rho_{\theta,n}(t)$ be the characteristic functions of $P_{\theta,n}$ and $Q_{\theta,n}$ respectively. If $\omega_n(t)=\exp(-1/2\, n^{-2\beta}t^2)$, then $\omega_n\chi_{\theta,n}$, $\omega_n\rho_{\theta,n}$ are the characteristic functions of $LP_{\theta,n}$ and $LQ_{\theta,n}$ respectively.

$\omega_n\chi_{\theta,n}$ is an element of $L^1(\mathbb{R})$. Therefore $LP_{\theta,n}$ is absolutely continuous with respect to the Lebesgue measure. The density may be called $f_{\theta,n}$.

Put

$$h_{\theta,n}(x)=x^2\left(f_{\theta,n}(x)-\frac{dLQ_{\theta,n}}{d\lambda}(x)\right).$$

Then

$$\begin{aligned}\hat{h}_{\theta,n}(t) &= \int h_{\theta,n}(x)\exp(itx)\,dx\\ &= -\partial_t^2[\omega_n(t)(\chi_{\theta,n}(t)-\rho_{\theta,n}(t))]\\ &= -\omega_n''(t)(\chi_{\theta,n}(t)-\rho_{\theta,n}(t))-2\omega_n'(t)(\chi_{\theta,n}'(t)-\rho_{\theta,n}'(t))\\ &\quad -\omega_n(t)(\chi_{\theta,n}''(t)-\rho_{\theta,n}''(t)).\end{aligned}$$

Theorem 9.9 in Bhattacharya and Ranga Rao (1976) entails that there exist constants c_1, c_2 such that for $\theta\in\Theta_n$ and $\ell=0, 1, 2$:

$$\partial_t^\ell(\chi_{\theta,n}(t)-\rho_{\theta,n}(t))\leqq\frac{c_1}{n}(|t|^{4-\ell}+|t|^{6+\ell})\exp(-t^2/4) \qquad \text{for } |t|\leqq c_2\sqrt{n}.$$

This gives with some constants c_3, c_4 for $\theta\in\Theta_n$

$$\begin{aligned}|h_{\theta,n}(x)| &= |(2\pi)^{-1}\int\exp(-itx)\,\hat{h}_{\theta,n}(t)\,dt|\\ &= |(2\pi)^{-1}(\int_{|t|\leqq c_2\sqrt{n}}\ldots+\int_{|t|>c_2\sqrt{n}}\ldots|\\ &\leqq c_3/n+c_4\int_{|t|>c_2\sqrt{n}}\exp(-1/2n^{-2\beta}t^2)=o(1/\sqrt{n}).\end{aligned}$$

Analogous arguments show

$$\left|f_{\theta,n}(x)-\frac{dLQ_{\theta,n}}{d\lambda}(x)\right|=o(1/\sqrt{n}).$$

This entails (3.1).

We show now (2.10). According to the definition of L^* one has:

$$L^*(x,dy)=\frac{n^\beta\,\phi(n^\beta(y-x))\,P_{0,n}(dy)}{\int n^\beta\,\phi(n^\beta(z-x))\,P_{0,n}(dz)}=\frac{n^\beta\,\phi(n^\beta(y-x+\theta n^{1/2-2\beta}))\,P_{\theta,n}(dy)}{\int n^\beta\,\phi(n^\beta(z-x+\theta n^{1/2-2\beta}))\,P_{\theta,n}(dz)}.$$

Therefore

$$L_\theta\times P_{\theta,n}=L_\theta P_{\theta,n}\times L^*$$

where

$$L_\theta(x,dy)=n^\beta\,\phi(n^\beta(y-x+\theta n^{1/2-2\beta}))\,dy.$$

Because of

$$\theta n^{1/2-2\beta}=o(n^{1/6-1/2+1/2-2/3})=o(n^{-1/2})$$

it follows from the first part of the proof that

$$\sup_{\theta\in\Theta_n}\|L_\theta P_{\theta,n}-Q_{\theta,n}\|=o(1/\sqrt{n}).$$

But this shows (2.10) because of

$$\|L^*Q_{\theta,n}-P_{\theta,n}\|=\|L^*Q_{\theta,n}-L^*L_\theta P_{\theta,n}\|\leqq\|L_\theta P_{\theta,n}-Q_{\theta,n}\|.$$

Proof of Theorem 1. The upper bound in (2.1) can be proved as indicated in the previous section. To prove the lower bound in (2.1) consider the following

Bayes decision problem. Given the uniform distribution on $\frac{1}{\sqrt{n}}(c_n - 2\sqrt{c_n}, c_n)$ as a priori measure construct a confidence interval of length $2/\sqrt{n}$. The loss function is ± 1 according to a true or false decision. The difference δ_n^D of the minimal Bayes risks in $\mathscr{G}_n^D$ and $\mathscr{E}_n^n$ is

$$\delta_n^D = -2 \cdot \frac{\gamma_3}{6\sqrt{n}} \cdot \frac{1}{2\sqrt{c_n}} \int_{c_n - 2\sqrt{c_n}+1}^{c_n - 1} \int_{x-1}^{x+1} \phi(x-t)\, 3t((x-t)^2 - 1)\, dt\, dx + o(1/\sqrt{n})$$

$$= \frac{2\gamma_3}{\sqrt{n}} \cdot \frac{1}{2\sqrt{c_n}} \int_{c_n - 2\sqrt{c_n}+1}^{c_n - 1} x \cdot \frac{1}{\sqrt{2\pi e}}\, dx + o(1/\sqrt{n})$$

$$= \sqrt{\frac{2}{\pi e}} \frac{\gamma_3}{\sqrt{n}} (\sqrt{c_n} - 1)^2 + o(1/\sqrt{n}).$$

Because of $|\delta_n^D| \leqq \Delta(\mathscr{G}_n^D, \mathscr{E}_n^n)$ this shows the lower bound in (2.1) for $i = D$. The proof for $i = ML$ is similar. The lower bound in (2.2) is based on the comparison of binary Bayes risks. The lower bound in (2.3) can be obtained in the context of the following Bayes decision problem: The a priori measure is the uniform distribution on Θ_n. For $a, b > 0$ the loss function $\ell_n(\theta, d)$ is defined by

$$\ell_n(\theta, d) = \begin{cases} -1 & \text{if } \theta - a/\sqrt{n} \leqq d \leqq \theta \\ 1 & \text{if } \theta < d \leqq \theta + b/\sqrt{n} \\ 0 & \text{elsewhere.} \end{cases}$$

It can be proved, that a suitable choice of a, b leads to the lower bound in (2.3).

To prove the upper bound in (2.3) we show that for $b = -1/4$ and $\varepsilon_n = (2 \ln(2) - 3) \frac{\gamma_3}{12\sqrt{n}}$

$$\|\delta_{\varepsilon_n} * R_{\theta,n,b} - N(s, 1)\| = \frac{\ln(2)}{3\sqrt{2\pi}} \frac{|\gamma_3|}{\sqrt{n}} + o(1/\sqrt{n}). \tag{3.2}$$

This shows (see (2.13)):

$$\delta(\mathscr{E}_n^n, \mathscr{G}_n^H) \leqq \frac{\ln(2)}{3\sqrt{2\pi}} \frac{|\gamma_3|}{\sqrt{n}} + o(1/\sqrt{n}). \tag{3.3}$$

Proof of (3.2).

$$\|\delta_{\varepsilon_n} * R_{\theta,n,b} - N(s, 1)\|$$

$$= \int \left| \phi(x-s) \left\{ -\frac{\gamma_3}{12\sqrt{n}}((x-s)^3 - 3(x-s)) + \frac{\gamma_3}{12\sqrt{n}}(2\ln 2 - 3)(x-s) \right\} \right| dx$$

$$+ o(1/\sqrt{n})$$

$$= \int \left| \phi(x) \frac{\gamma_3}{12\sqrt{n}} (x^3 - 2\ln(2)x) \right| dx + o(1/\sqrt{n}) = \frac{\ln(2)}{3\sqrt{2\pi}} \frac{|\gamma_3|}{\sqrt{n}} + o(1/\sqrt{n}).$$

Similarly it can be seen that (3.3) holds for $\delta(\mathcal{G}_n^H, \mathcal{E}_n^n)$. This proves the upper bound in (2.3).

Proof of Theorem 2. For the present we assume $c_n \geqq \sqrt{\ln(n)}$. We devide the proof in several steps:

In the first step we will show that $\mathcal{E}_n^n$ is second order equivalent to an experiment which is a translation experiment after the parameter transformation $s = \mu_n^H(\theta)$. This reduces the asymptotic comparison of $\mathcal{E}_n^n$ and $\mathcal{G}_n^H$ to the comparison of two translation experiments. If the parameter space of these two translation experiments is enlarged to $\mathbb{R}$ the deficiency distance changes only by an amount of order $o(1/\sqrt{n})$. This will be proved in the second step. Then results of Torgersen (1972) for the comparison of translation experiments can be applied to prove

$$\delta(\mathcal{E}_n^n, \mathcal{G}_n^H) = \delta(\mathcal{G}_n^H, \mathcal{E}_n^n) + o(1/\sqrt{n})$$

(see (2.5)) (third step). The deficiency distance between two experiments is equivalent to the maximal difference between (minimal) Bayes risks in these two experiments (see LeCam (1964)). In the 4th and 5th step it will be shown that for an asymptotic calculation of $\delta(\mathcal{E}_n^n, \mathcal{G}_n^H)$ and $\delta(\mathcal{G}_n^H, \mathcal{E}_n^n)$ it suffices to consider *one fixed* Bayes decision problem respectively which is formulated in the Hellinger parametrization $s = \mu_n^H(\theta)$. This simplifies essentially the treatment of error terms. In the 6th step all Bayes decision problems are considered which are generated by these two Bayes decision problems by a shift of the parameter space. It will be shown that every replacement of $\mathcal{G}_n^H$ by another Gaussian experiment would increase the difference of (minimal) Bayes risks for at least one of these decision problems. This proves then the statement of the theorem.

1. Step. We show

$$\Delta(\mathcal{E}_n^n, \bar{\mathcal{E}}_n) = o(1/\sqrt{n}),$$

where

$$\bar{\mathcal{E}}_n = (M_{s,n} : s = \mu_n^H(\theta),\ \theta \in \Theta_n),$$

$$dM_{s,n} = \phi(x-s)\left(1 - \frac{\gamma_3}{12\sqrt{n}}((x-s)^3 - 3(x-s))\right) d\lambda.$$

This can be seen by the lemma using

$$\sup_{\theta \in \Theta_n} |\mu_n^H(\theta) - \sqrt{n}\theta| = O(c_n^2/\sqrt{n}).$$

2. Step. We prove

$$\delta(\bar{\mathcal{E}}_n, \mathcal{G}_n^H) = \delta(\tilde{\mathcal{E}}_n, \mathcal{G}) + o(1/\sqrt{n}),$$

$$\delta(\mathcal{G}_n^H, \bar{\mathcal{E}}_n) = \delta(\mathcal{G}, \tilde{\mathcal{E}}_n) + o(1/\sqrt{n}),$$

where

$$\tilde{\mathscr{E}}_n = (M_{s,n}: s \in \mathbb{R})$$
$$\mathscr{G} = (N(s, 1): s \in \mathbb{R}).$$

To prove this statement we use the following slightly changed version of Theorem 3 in Mammen (1983).

Proposition. *Let $\mathscr{F}_i = (Q^i_\theta: \theta \in \Theta)$ be two experiments with $\Theta \subset \mathbb{R}$. Assume that for two positive constants a and ε there exist in $\mathscr{F}_1$ resp. $\mathscr{F}_2$ two estimates $\hat{\theta}_1, \hat{\theta}_2$ with*

$$Q^i_\theta(\|\hat{\theta}_i - \theta\| > a) \leqq \varepsilon \quad \textit{for } i = 1, 2 \textit{ and } \theta \in \Theta.$$

Then for $b > (9/2)\, a$ and $(i, j) = (1, 2)$ or $(2, 1)$ the following holds:

$$\delta_b(\mathscr{F}_i, \mathscr{F}_j) \leqq \delta(\mathscr{F}_i, \mathscr{F}_j) \leqq \delta_b(\mathscr{F}_i, \mathscr{F}_j) + \frac{8a}{2b - a} \Delta(\mathscr{F}_1, \mathscr{F}_2) + 12\varepsilon$$

where

$$\delta_b(\mathscr{F}_i, \mathscr{F}_j) = \sup_{\tau \in \Theta} \delta((Q^i_\theta: \|\theta - \tau\| \leqq b/2), (Q^j_\theta: \|\theta - \tau\| \leqq b/2)).$$

In $\tilde{\mathscr{E}}_n$ resp. $\mathscr{G}$ there exist the following estimates

$$N(s, 1)(\|x - s\| > \sqrt{\ln(n)}) = o(1/\sqrt{n}),$$
$$M_{s,n}(\|x - s\| > \sqrt{\ln(n)}) = o(1/\sqrt{n}).$$

Furthermore $\tilde{\mathscr{E}}_n$ and $\mathscr{G}$ are translation experiments. Therefore:

$$\delta_b(\tilde{\mathscr{E}}_n, \mathscr{G}) = \delta((M_{s,n}: \|s\| \leqq b/2), (N(s, 1): \|s\| \leqq b/2))$$
$$\delta_b(\mathscr{G}, \tilde{\mathscr{E}}_n) = \delta((N(s, 1): \|s\| \leqq b/2); (M_{s,n}: \|s\| \leqq b/2)).$$

3. Step. We show now that

$$\delta(\mathscr{E}^n_n, \mathscr{G}^H_n) = \delta(\mathscr{G}^H_n, \mathscr{E}^n_n) + o(1/\sqrt{n}).$$

According to the second step it suffices to prove

$$\delta(\tilde{\mathscr{E}}_n, \mathscr{G}) = \delta(\mathscr{G}, \tilde{\mathscr{E}}_n) + o(1/\sqrt{n}). \tag{3.4}$$

$\tilde{\mathscr{E}}_n$ and $\mathscr{G}$ are translation experiments. By Theorem 1 of Torgersen (1972) this implies the existence of a measure ν_n with:

$$\delta(\tilde{\mathscr{E}}_n, \mathscr{G}) = \sup_{s \in \mathbb{R}} \|M_{s,n} * \nu_n - N(s, 1)\| = \|M_{0,n} * \nu_n - N(0, 1)\|.$$

We will show, that for $\nu^*_n(A) = \nu_n(-A)$ the following holds:

$$\delta(\tilde{\mathscr{E}}_n, \mathscr{G}) = \|N(0, 1) * \nu^*_n - M_{0,n}\| + o(1/\sqrt{n}). \tag{3.5}$$

This implies then

$$\delta(\tilde{\mathscr{E}}_n, \mathscr{G}) \geqq \delta(\mathscr{G}, \tilde{\mathscr{E}}_n) + o(1/\sqrt{n}). \tag{3.6}$$

Proof of (3.5). Firstly

$$\|N(0,1)*\nu_n - N(0,1)\| = \|M_{0,n}*\nu_n - N(0,1)\| + O(1/\sqrt{n}) = O(1/\sqrt{n})$$

because of $\Delta(\mathscr{E}_n^n, \mathscr{G}_n^H) = O(1/\sqrt{n})$ (see Theorem 1). This shows

$$\nu_n(\|X\| > d) = o(1) \qquad \text{for } d > 0$$

and therefore

$$\|G*\nu_n - G\| = o(1),$$

where G is the measure which has the density $x^3 - 3x$ with respect to $N(0,1)$. With this we get

$$\begin{aligned}\delta(\tilde{\mathscr{E}}_n, \mathscr{G}) &= \|M_{0,n}*\nu_n - N(0,1)\| \\ &= \left\|N(0,1)*\nu_n - N(0,1) - \frac{\gamma_3}{12\sqrt{n}}\, G*\nu_n\right\| + o(1/\sqrt{n}) \\ &= \left\|N(0,1)*\nu_n - N(0,1) - \frac{\gamma_3}{12\sqrt{n}}\, G\right\| + o(1/\sqrt{n}) \\ &= \left\|N(0,1)*\nu_n^* - N(0,1) + \frac{\gamma_3}{12\sqrt{n}}\, G\right\| + o(1/\sqrt{n}) \\ &= \|N(0,1)*\nu_n^* - M_{0,n}\| + o(1/\sqrt{n}).\end{aligned}$$

The inequality reversed to (3.6) can be shown similarly.

4. Step. There exists a constant d with

$$\sqrt{n}\,\Delta(\mathscr{E}_n^n, \mathscr{G}_n^H) \to d \qquad \text{for } n \to \infty.$$

Put $\mathscr{B}$ the set of all Bayes decision problems b consisting of a a priori measure π with finite support $\{s_i : i = 1, \ldots, k\}$, of a finite decision space $\{1, \ldots, \ell\}$, and of a loss function L, absolutely bounded by 1: $|L(i,j)| \leqq 1$ for $i = 1, \ldots, k$, and $j = 1, \ldots, \ell$. Furthermore set

$$\begin{aligned}\Delta_n(b) = &\int \inf_{1 \leqq j \leqq \ell} \sum_{1 \leqq i \leqq k} \pi_i L(i,j)\, \phi(x - s_i)\, dx \\ &- \int \inf_{1 \leqq j \leqq \ell} \sum_{1 \leqq i \leqq k} \pi_i L(i,j)\, \phi(x - s_i)\left(1 + \frac{1}{\sqrt{n}}\, g(x - s_i)\right) dx\end{aligned}$$

where

$$\pi_i = \pi(\{s_i\})$$

$$g(y) = \frac{\gamma_3}{12}(y^3 - 3y).$$

Then $\Delta_n(b)$ is the difference of the minimal Bayes risk to the Bayes decision problem b in $\mathscr{G}$ and $\tilde{\mathscr{E}}_n$.

According to LeCam (1964), Torgersen (1970) therefore

$$\Delta(\mathscr{E}_n^n, \mathscr{G}_n^H) = \sup_{b\in\mathscr{B}} |\Delta_n(b)| + o(n^{-1/2}).$$

For $\Delta_n(b)$ a simple inequality holds:

$$\Delta_n(b) \geqq d(b)/\sqrt{n} \tag{3.7}$$

where

$$d(b) = \sum_{1\leqq i\leqq k,\, 1\leqq j\leqq \ell} \pi_i L(i,j) \int_{A_j} g(x-s_i)\,\phi(x-s_i)\,dx$$

$$A_{j_0} = \{x\colon h_{j_0}(x) \leqq h_j(x) \quad \text{for } j=1, \ldots, \ell \text{ and}$$
$$\tilde{h}_{j_0}(x) < \tilde{h}_j(x) \quad \text{for every } j=1, \ldots, \ell \text{ with } h_{j_0}(x) = h_j(x)\}$$
$$h_j(x) = \sum_{1\leqq i\leqq k} \pi_i L(i,j)\,\phi(x-s_i)$$
$$\tilde{h}_j(x) = \sum_{1\leqq i\leqq k} \pi_i L(i,j)\, g(x-s_i)\,\phi(x-s_i).$$

The 3. Step and (3.7) imply:

$$\begin{aligned}\Delta(\mathscr{E}_n^n, \mathscr{G}_n^H) &= \delta(\mathscr{E}_n^n, \mathscr{G}_n^H) + o(1/\sqrt{n})\\ &= \sup_{b\in\mathscr{B}} (-\Delta_n(b)) + o(1/\sqrt{n})\\ &\leqq 1/\sqrt{n} \sup_{b\in\mathscr{B}} (-d(b)) + o(1/\sqrt{n}).\end{aligned} \tag{3.8}$$

Furthermore, using the dominated convergence theorem, one gets

$$\sqrt{n}\Delta_n(b) \to d(b) \quad \text{for } n\to\infty.$$

This and (3.8) shows

$$\sqrt{n}\Delta(\mathscr{E}_n^n, \mathscr{G}_n^H) \to \sup_{b\in\mathscr{B}} (-d(b)) \qquad \text{for } n\to\infty.$$

5. *Step.* For every $\varepsilon>0$ there exists Bayes decision problems b and b^* which differ only by $s_i = -s_i^*$ with

$$\begin{aligned}\Delta_n(b) + \varepsilon/\sqrt{n} &\geqq \Delta(\mathscr{E}_n^n, \mathscr{G}_n^H)\\ -\Delta_n(b^*) + \varepsilon/\sqrt{n} &\geqq \Delta(\mathscr{E}_n^n, \mathscr{G}_n^H)\end{aligned} \tag{3.9}$$

for n large enough.

This follows immediately from $d(b) = -d(b^*)$ for all b and b^* which differ only by $s_i = -s_i^*$ and from the considerations of the last step.

Furthermore the statement of this step is also valid if we drip the assumption taken at the beginning of the proof that $c_n \geqq \sqrt{\ln(n)}$, because it suffices to look at Bayes decision problems which are fixed and which do not depend on n.

6. *Step*. We show now, that the assumption

$$\liminf_{n\to\infty} \sqrt{n}\,\Delta(\mathscr{G}_n, \mathscr{E}_n^n) \leqq \lim_{n\to\infty} \sqrt{n}\,\Delta(\mathscr{G}_n^H, \mathscr{E}_n^n) - 2\varepsilon \tag{3.10}$$

for a $\varepsilon > 0$ leads to a contradiction.

Firstly (3.10) implies $\Delta(\mathscr{G}_n, \mathscr{G}_n^H) = O(1/\sqrt{n})$, and therefore:

$$\sup\{\sqrt{n}\,|\mu_n(\theta+\tau) - \mu_n(\theta) - \mu_n^H(\theta+\tau) + \mu_n^H(\theta)| : \theta, \theta+\tau \in \Theta_n, n \in \mathbb{N}\} < +\infty \quad \text{for every } \tau > 0. \tag{3.11}$$

Now take $s = \mu_n^H(\theta)$ in $\mathscr{G}_n$ and $\mathscr{G}_n^H$ as new parameters. Call the new experiments $\tilde{\mathscr{G}}_n$ and – as above – $\mathscr{G}$. Choose b and b^* according to the 5. Step. For $h \in \mathbb{R}$ b_h and b_h^* are the Bayes decision problems which differ from b and b^* only by $s_{i,h} = s_i + h$ and $s_{i,h}^* = s_i^* + h$ respectively. Put

$$\delta_n(s) = \sqrt{n}(\mu_n(\theta) - \mu_n^H(\theta)),$$

where θ is chosen such that $s = \mu_n^H(\theta)$. Denote the minimal Bayes risk in an experiment $\mathscr{F}$ according to a Bayes decision problem b by $\rho(b, \mathscr{F})$.

Then the following holds:

$$\begin{aligned} \rho(b_h, \tilde{\mathscr{G}}_n) &= \rho(b_h, \mathscr{G}) + 1/\sqrt{n} \sum_{1\leqq i\leqq k} e_i\,\delta_n(s_i+h) + o(1/\sqrt{n}). \\ \rho(b_h^*, \tilde{\mathscr{G}}_n) &= \rho(b_h^*, \mathscr{G}) - 1/\sqrt{n} \sum_{1\leqq i\leqq k} e_i\,\delta_n(-s_i+h) + o(1/\sqrt{n}), \end{aligned} \tag{3.12}$$

where

$$e_i = \sum_{1\leqq j\leqq \ell} \pi_i L(i,j) \int_{A_j} (x-s_i)\,\phi(x-s_i)\,dx.$$

(3.12) can be followed by

$$\begin{aligned} \sum_{1\leqq i\leqq k} e_i &= 0 \\ \delta_n(s_i+h) - c_{h,n} &= O(1) \\ \delta_n(-s_i+h) - c_{h,n}^* &= O(1) \end{aligned} \tag{3.13}$$

for $i \leqq k$ and for a $c_{h,n}, c_{h,n}^* \in \mathbb{R}$ (because of (3.11)).

(3.9), (3.10), and (3.12) imply that for every $h \in \mathbb{R}$:

$$\begin{aligned} \liminf_{n\to\infty} \sum_{1\leqq i\leqq k} e_i\,\delta_n(s_i+h) &\leqq -\varepsilon \\ \liminf_{n\to\infty} \sum_{1\leqq i\leqq k} e_i\,\delta_n(-s_i+h) &\leqq -\varepsilon. \end{aligned} \tag{3.14}$$

Without loss of generality we assume $s_{i+1}-s_i=\text{const.}=\delta$ for $i\leqq k-1$. Put $f_i=e_i+e_{k-i+1}$ for $i\leqq k$ and $x_{\ell,n}=\delta_n(s_1+(\ell-1)\delta)$ for $\ell\in\mathbb{Z}$:

$$\liminf_{n\to\infty}\sum_{1\leqq i\leqq k} f_i\, x_{\ell+i,n}\leqq -2\varepsilon. \tag{3.15}$$

Because of $\sum_{1\leqq i\leqq k} f_i=0$, and $f_i=f_{k-i}$ for $i\leqq k$, there exist $(\alpha_i)_{i\in\mathbb{Z}}$ with

$$\begin{aligned} f_i&=\alpha_{i-1}-2\alpha_i+\alpha_{i+1} && \text{for } 1\leqq i\leqq k\\ \alpha_i&=0 && \text{for } i\leqq 1 \text{ or } i\geqq k. \end{aligned}$$

With this notation one gets:

$$\sum_i f_i\, x_{\ell+i,n}=\sum_i \alpha_{i-1}\, x_{\ell+i,n}-2\sum_i \alpha_i\, x_{\ell+i,n}+\sum_i \alpha_{i+1}\, x_{\ell+i,n}=y_{\ell+1,n}-2y_{\ell,n}+y_{\ell-1,n}$$

for $y_{\ell,n}=\sum_i \alpha_i\, x_{i+\ell,n}$.

Putting this into (3.15) one gets

$$\liminf_{n\to\infty} y_{\ell+1,n}-2y_{\ell,n}+y_{\ell-1,n}\leqq -2\varepsilon \tag{3.16}$$

for $\ell\in\mathbb{Z}$.

But according to (3.11) the slope of $y_{\ell,n}$ (as a function of ℓ) is bounded (uniformly in n). This contradicts (3.16).

Proof of Theorem 3. Firstly

$$\begin{aligned}\sup_{\theta\in\Theta_n}\|R_{\theta,n,b(\alpha)}-s_{\theta,n,\alpha}\|&=\frac{\gamma_3}{\sqrt{n}}\int\left|\frac{6\alpha-2}{6}\,s(x)+c_n(2\alpha-\tfrac12)\,v(x)\right|dx+o(1/\sqrt{n})\\ &=\frac{\gamma_3}{\sqrt{n}}\,c_n(2\alpha-\tfrac12)\int|v(x)|\,dx+o(c_n/\sqrt{n})\\ &=|4\alpha-1|\,\frac{\gamma_3\,c_n}{\sqrt{n}}\sqrt{\frac{2}{\pi e}}+o(c_n/\sqrt{n}).\end{aligned}$$

Because of (2.15) it remains to prove

$$\Delta(\mathscr{E}_n^n,\mathscr{G}(n,\mu_n))\geqq|4\alpha-1|\,\frac{\gamma_3\,c_n}{\sqrt{n}}\sqrt{\frac{2}{\pi e}}+o(c_n/\sqrt{n}).$$

But this can be done as in the proof of the first inequality of (2.1) of Theorem 1.

Proof of the corollary. Using Proposition 1 one yields:

$$\delta(\mathscr{F}_n^n,\mathscr{G}_n^H)=\delta(\mathscr{G}_n^H,\mathscr{F}_n^n)+o(1/\sqrt{n}). \tag{3.17}$$

Furthermore, because the deficiency is determined by the difference of Bayes risks, the following holds:

$$\delta(\mathscr{F}_n^n, \mathscr{G}_n) \underset{(\leqq)}{\geqq} \delta(\mathscr{F}_n^n, \mathscr{G}_n^H) \quad \text{and} \quad \delta(\mathscr{G}_n, \mathscr{F}_n^n) \underset{(\geqq)}{\leqq} \delta(\mathscr{G}_n^H, \mathscr{F}_n^n) \quad \text{if} \quad m_n \underset{(\leqq)}{\geqq} 2\sqrt{n}\, H(P_{\theta_n}, P_{\tau_n}). \tag{3.18}$$

(3.18) and (3.17) show (2.6).

References

Bhattacharya, R.N., Ranga Rao, R.: Normal approximation and asymptotic expansions. New York: Wiley 1976

LeCam, L.: On the asymptotic theory of estimates and testing hypotheses. In: Proceedings of the Third Berkeley Sympos., Math. probab. statist., **I**, 129–156. University of California (1956)

LeCam, L.: Locally asymptotically normal families of distributions. University of California Publications in Statistic **3**, 37–98 (1960)

LeCam, L.: Sufficiency and approximate sufficiency. Ann. Math. Statist. **35**, 1419–1455 (1964)

LeCam, L.: Théorie asymptotique de la décision statistique. Les Presses de l'Université de Montréal (1968)

LeCam, L.: Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré **21**, 225–287 (1985)

LeCam, L.: Asymptotic methods in statistical decision theory. Berlin Heidelberg New York: Springer 1986

Mammen, E.: Die statistische Information zusätzlicher Beobachtungen. Inaugural-Dissertation, Heidelberg (1983)

Michel, R., Pflanzagl, J.: Asymptotic normality. Metrika **16**, 188–205 (1970)

Millar, P.W.: Asymptotic minimax theorems for the sample distribution function. Z. Wahrscheinlichkeitstheor. Verw. Geb. **48**, 233–252 (1979)

Müller, D.W.: Asymptotically multinomial experiments and the extension of a theorem of Wald. Z. Wahrscheinlichkeitstheor. Verw. Geb. **50**, 179–204 (1979)

Pflanzagl, J.: Further results on asymptotic normality I. Metrika **18**, 174–198 (1972)

Torgersen, E.N.: Comparison of translation experiments. Ann. Math. Statist. **43**, 1383–1399 (1972)

Torgersen, E.N.: Comparison of experiments when the parameter space is finite. Z. Wahrscheinlichkeitstheor. Verw. Geb. **16**, 219–249 (1970)

Wald, A.: Test of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Am. Math. Soc. **54**, 426–482 (1943)

Received January 1, 1985; in revised form April 10, 1987