Chapter 4
Asymptotic Risk and Bayes Risk of Thresholding and Superefficient Estimates and Optimal Thresholding

Anirban DasGupta and Iain M. Johnstone

4.1 Introduction

The classic Hodges' estimator (Hodges, 1951, unpublished) of a one dimensional normal mean demolished the statistical folklore that maximum likelihood estimates are asymptotically uniformly optimal, provided the family of underlying densities satisfies enough regularity conditions. Hodges' original estimate is

$$T_n(X_1, \cdots, X_n) = \begin{cases} \bar{X}_n & \text{if} \quad |\bar{X}_n| > n^{-1/4} \\ 0 & \text{if} \quad |\bar{X}_n| \le n^{-1/4} \end{cases} \tag{4.1}$$

A more general version is

$$S_n(X_1, \ldots, X_n) = \begin{cases} \bar{X}_n & \text{if} \quad |\bar{X}_n| > c_n \\ a_n \bar{X}_n & \text{if} \quad |\bar{X}_n| \le c_n \end{cases} \tag{4.2}$$

Here, c_n, for the moment, is a general positive sequence and $0 \le a_n \le 1$. With squared error as the loss function, the risk of $\bar{X}_n$, the unique MLE, satisfies $nR(\theta, \bar{X}_n) \equiv 1$, and Hodges' original estimate T_n satisfies

$$\lim_{n\to\infty} n^\beta R(0, T_n) = 0 \; \forall \, \beta > 0,$$

while

$$\lim_{n\to\infty} \sup_\theta \; nR(\theta, T_n) = \infty.$$

Thus, at $\theta = 0$, Hodges' estimate is asymptotically infinitely superior to the MLE, while globally its peak risk is infinitely more relative to that of the MLE. *Superefficiency at $\theta = 0$ is purchased at a price of infinite asymptotic inflation in risk away from zero*. Hodges' example showed that the claim of the uniform asymptotic optimality of the MLE is false even in the normal case, and it seeded the development

A. DasGupta (✉) · I. M. Johnstone
Purdue University and Stanford University, USA
e-mail: dasgupta@purdue.edu

S. Lahiri et al. (eds.), *Contemporary Developments in Statistical Theory,*
Springer Proceedings in Mathematics & Statistics 68, DOI 10.1007/978-3-319-02651-0_4,

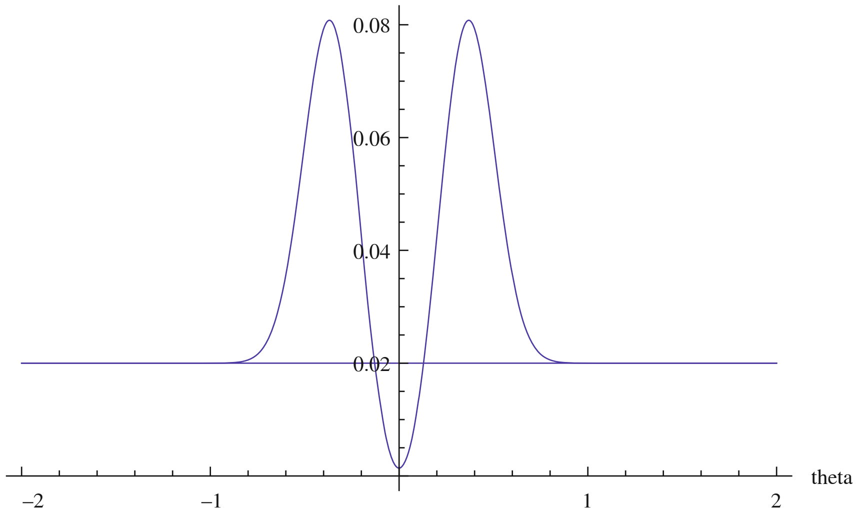

Fig. 4.1 Risk of Hodges' Estimate for n=50

of such fundamental concepts as regular estimates. It culminated in the celebrated *Hájek-Le Cam convolution theorem*. It probably, also had some indirect impact on the development and study of the now common *thresholding estimates* in *large p small n* problems, the most well known among them being the Donoho-Johnstone estimates (Donoho and Johnstone (1994)), although while the classic Hodges' estimate uses a small threshold ($n^{-1/4}$), the new thresholding estimates use a large threshold (Fig 4.1).

It is of course already well understood that the risk inflation of Hodges' estimate occurs *close to zero*, and that the worst inflation occurs in a neighborhood of small size. This was explicitly pointed out in Le Cam (1953):

$$\lim_{n\to\infty} \sup_{U_n} \sup_{\theta\in U_n} nR(\theta, T_n) = \infty,$$

where U_n denotes a general sequence of open neighborhoods of zero such that $\lambda(U_n)$, the Lebesgue measure of U_n, goes to zero; *we cannot have asymptotic superefficiency in nonvanishing neighborhoods*. Provided only that a competitor estimate sequence T_n has a limit distribution under every θ, i.e., $\sqrt{n}(T_n - \theta)$ has some limiting distribution L_θ, it must have an asymptotic pointwise risk at least as large as that of $\bar{X}$ at almost all θ:

$$\text{For almost all } \theta, \limsup_{n\to\infty} nR(\theta, T_n) \geq 1.$$

Indeed, a plot of the risk function of Hodges' estimate nicely illustrates these three distinct phenomena, *superefficiency at zero, inflation close to zero, worst inflation in a shrinking neighborhood*: Similar in spirit are the contemporary thresholding

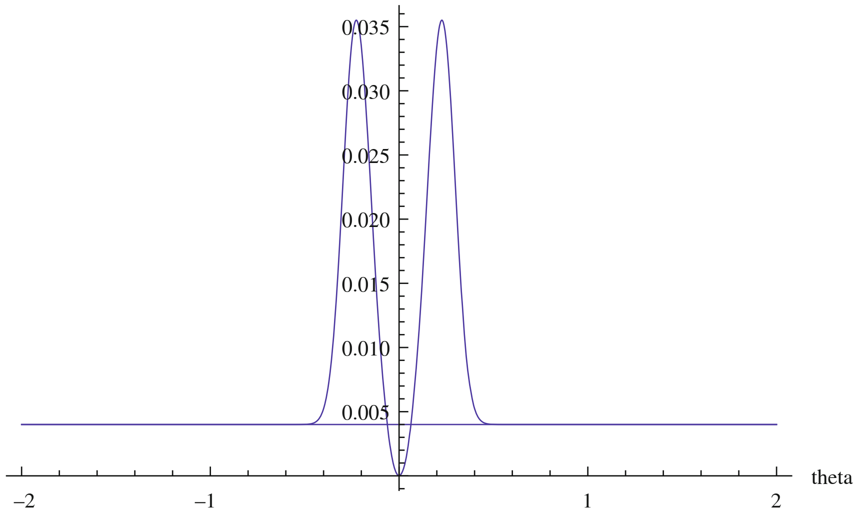

Fig. 4.2 Risk of Hodges' Estimate for n = 250

estimates of Gaussian means. Formally, given $X \sim N(\theta, 1)$, and $\lambda > 0$, the *hard thresholding estimate* is defined as

$$\hat{\theta}_\lambda = X \quad \text{if } |X| > \lambda$$
$$= 0 \quad \text{if } |X| \leq \lambda$$

Implicit in this construction is an underlying Gaussian sequence model

$$X_i \overset{indep.}{\sim} N(\theta_i, 1), i = 1, 2, \cdots, n,$$

and

$$\hat{\theta}_i = X_i I_{|X_i| > \lambda(n)}, \tag{4.3}$$

and $\lambda(n)$ often being asymptotic to $\sqrt{2 \log n}$, which is a first order asymptotic approximation (although not very accurate practically) to the expectation of the maximum of n iid $N(0, 1)$ observations. The idea behind this construction is that we expect nearly all the means to be zero (i.e., the observed responses are instigated by pure noise), and we estimate a specific θ_i to be equal to the observed signal only if the observation stands out among a crowd of roughly n pure Gaussian white noises. See Johnstone (2012) for extensive discussion and motivation (Fig 4.2).

The similarity between Hodges' estimate and the above hard thresholding estimate is clear. We would expect the hard thresholding estimate to manifest risk phenomena similar to that of Hodges' estimate: better risk than the naive estimate X_i itself if the true θ_i is zero, risk inflation if the true θ_i is adequately away from zero, and we expect that the finer details will depend on the choice of the threshold level λ. One may ask what is the *optimal* λ that suitably balances the risk gain at zero with the risk inflation away from zero.

Another commonality in the behavior of Hodges' estimate and the hard thresholding estimate is that if we take a prior distribution on the true mean that is very tightly concentrated near zero, then they ought to have smaller Bayes risks than the MLE, and the contrary is expected if we take an adequately diffuse prior.

It is meaningful and also interesting to ask if these various anticipated phenomena can be pinned down with some mathematical precision. The main contributions of this article are the following:

a) For the one dimensional Gaussian mean and superefficient estimates of the general form as in (4.2), we precisely quantify the behavior of the risk at zero (Eq. (4.10), Corollary 1.2.5).
b) We precisely quantify the risk at $\frac{k}{\sqrt{n}}$ for fixed positive k (Eq. (4.22)), and we show that the risk at $\frac{1}{\sqrt{n}}$ (which is exactly one standard deviation away from zero) is for all practical purposes equal to $\frac{1}{n}$, which is the risk of the MLE (Theorem 1.2.4, Corollary 1.2.5).
c) We show that in the very close vicinity of zero, the risk of superefficient estimates increases at an increasing rate, i.e., the risk is locally convex (Theorem 1.2.2).
d) We show that the global peak of the risk is *not* attained within $n^{-1/2}$ neighborhoods. In fact, we show that at $\theta = c_n$, the risk is much higher (Theorem 1.2.5, Eq. (4.26)), and that *immediately below* $\theta = c_n$, the risk is even higher. Precisely, we exhibit explicit and parsimonious shrinking neighborhoods U_n of $\theta = c_n$, such that

$$\liminf c_n^{-2} \sup_{\theta \in U_n} R(\theta, S_n) \geq 1. \tag{4.4}$$

(Theorem 1.2.6, Eq. (4.28)). Note that we can obtain the lower bound in (4.4) with an lim inf, rather than lim sup.

Specifically, our calculations indicate that $\text{argmax}_\theta R(\theta, S_n) \approx c_n - \sqrt{\frac{\log(nc_n^2)}{n}}$, and $\sup_\theta R(\theta, S_n) \approx c_n^2 - 2c_n\sqrt{\frac{\log n}{n}}$ (Eq. (4.35)).
e) For normal priors $\pi_n = N(0, \sigma_n^2)$, we obtain exact closed form expressions for the Bayes risk $B_n(\pi_n, S_n)$ of S_n (Theorem 1.2.7, Eq. (4.45)), and characterize those priors for which $B_n(\pi_n, S_n) \leq \frac{1}{n}$ for all large n. Specifically, we show that $\sigma^2 = \frac{1}{n}$ acts in a very meaningful way as the boundary between $B_n(\pi_n, S_n) < \frac{1}{n}$ and $B_n(\pi_n, S_n) > \frac{1}{n}$ (Theorem 1.2.8).
More generally, we use the theory of regular variation to show the quite remarkable fact that for *general smooth* prior densities $\pi_n(\theta) = \sqrt{n}h(\theta\sqrt{n})$, *all Hodges type estimates are approximately equivalent in Bayes risk to the MLE* $\bar{X}$ and that the exact rate of convergence of the difference in Bayes risks is determined by whether or not $\text{Var}_h(\theta) = 1$ (Theorem 1.2.10, Eq. (4.64)). This theorem, in turn, follows from a general convolution representation for the difference in Bayes risks under general π_n (Theorem 1.2.9, Eq. (4.48)).
f) For the Gaussian sequence model, we obtain appropriate corresponding versions of a)-e) for hard thresholding estimates of the form (4.3).

g) We identify the specific estimate in the class (4.2) that minimizes an approximation to the global maximum of the risk subject to a guaranteed specified improvement at zero; this is usually called a *restricted minimax problem*. More precisely, we show that subject to the constraint that the percentage risk improvement at zero is at least $100(1-\epsilon_n)\%$, the global maximum risk is approximately minimized when $c_n = \sqrt{2\log\frac{1}{\epsilon_n}}$ (Eq. (4.38)).
h) We illustrate the various results with plots, examples, and summary tables.

Several excellent sources where variants of a few of our problems have been addressed include Hájek (1970), Johnstone (2012), Le Cam (1953, 1973), Lehmann and Romano (2005), van der Vaart (1997, 1998), and Wasserman (2005). Also, see DasGupta (2008) and lecture notes written by Jon Wellner and Moulinath Banerjee. Superefficiency has also been studied in some problems that do not have the LAN (locally asymptotically normal) structure; one reference is Jeganathan (1983).

If the variance σ^2 of the observations was unknown, estimates similar to Hodges' are easily constructed by hard thresholding the MLE whenever $\frac{|\bar{X}|}{s} \le c_n$, where s is the sample standard deviation. Some of its risk properties can be derived along the lines of this article. However, the optimal thresholding and global maximum risk problems are likely to be even more difficult.

4.2 Risk Function of Generalized Hodges Estimates

Consider generalized Hodges estimates of the form (4.2). We first derive an expression for the risk function of the estimate $S_n(X_1, \cdots, X_n)$. This formula will be repeatedly used for many of the subsequent results. This formula for the risk function then leads to formulas for its successive derivatives, which are useful to pin down finer properties of S_n.

4.2.1 Global Formulas

Theorem 1.2.1 *Let $n \ge 1$ and $X_1, \cdots, X_n$ iid $N(\theta, 1)$. Let $0 \le a_n \le 1$ and $c_n > 0$. For the estimate $S_n(X_1, \cdots, X_n)$ as in* (4.2)*, the risk function under squared error loss is given by*

$$R(\theta, S_n) = \frac{1}{n} + e_n(\theta),$$

where

$$e_n(\theta) = \left[\frac{a_n^2 - 1}{n} + (1-a_n)^2\theta^2\right]\left(\Phi(\sqrt{n}(c_n - \theta)) + \Phi(\sqrt{n}(c_n + \theta)) - 1\right)$$
$$+ \frac{2a_n(a_n - 1)\theta}{\sqrt{n}}\left(\phi(\sqrt{n}(c_n + \theta)) - \phi(\sqrt{n}(c_n - \theta))\right)$$

$$+\frac{1-a_n^2}{\sqrt{n}}\Big((c_n+\theta)\phi(\sqrt{n}(c_n+\theta))+(c_n-\theta)\phi(\sqrt{n}(c_n-\theta))\Big), \qquad (4.5)$$

where ϕ and Φ denote the density and the CDF of the standard normal distribution.

Proof Write $R(\theta, S_n)$ as

$$R(\theta, S_n) = E[(\bar{X}-\theta)^2 I_{|\bar{X}|>c_n}] + E[(a_n\bar{X}-\theta)^2 I_{|\bar{X}|\le c_n}]$$

$$= E[(\bar{X}-\theta)^2] + E[(a_n\bar{X}-\theta)^2 I_{|\bar{X}|\le c_n}] - E[(\bar{X}-\theta)^2 I_{|\bar{X}|\le c_n}]$$

$$= \frac{1}{n} + \int_{-\sqrt{n}(c_n+\theta)}^{\sqrt{n}(c_n-\theta)} [a_n(\theta+\frac{z}{\sqrt{n}})-\theta]^2\phi(z)dz - \frac{1}{n}\int_{-\sqrt{n}(c_n+\theta)}^{\sqrt{n}(c_n-\theta)} z^2\phi(z)dz$$

$$= \frac{1}{n} + T_1 + T_2 \qquad \text{(say)} \qquad (4.6)$$

On calculation, we get

$$T_1 = \Big[\frac{a_n^2}{n} + (1-a_n)^2\theta^2\Big]\Big(\Phi(\sqrt{n}(c_n-\theta)) + \Phi(\sqrt{n}(c_n+\theta)) - 1\Big)$$

$$-\frac{a_n^2}{\sqrt{n}}\Big((c_n+\theta)\phi(\sqrt{n}(c_n+\theta))+(c_n-\theta)\phi(\sqrt{n}(c_n-\theta))\Big)$$

$$+\frac{2a_n(a_n-1)\theta}{\sqrt{n}}\Big(\phi(\sqrt{n}(c_n+\theta))-\phi(\sqrt{n}(c_n-\theta))\Big), \qquad (4.7)$$

and

$$T_2 = \frac{1}{n}\Big(\Phi(\sqrt{n}(c_n-\theta)) + \Phi(\sqrt{n}(c_n+\theta)) - 1\Big)$$

$$-\frac{1}{\sqrt{n}}\Big((c_n+\theta)\phi(\sqrt{n}(c_n+\theta))+(c_n-\theta)\phi(\sqrt{n}(c_n-\theta))\Big) \qquad (4.8)$$

On combining (4.6), (4.7), and (4.8), and further algebraic simplification, the stated expression in (4.5) follows.

4.2.1.1 Behavior at Zero

Specializing the global formula (4.5) to $\theta = 0$, we can accurately pin down the improvement at zero.

Corollary 1.2.1 *The risk improvement of S_n over $\bar{X}$ at $\theta = 0$ satisfies*

$$e_n(0) = \frac{1}{n} - R(0, S_n) = \frac{2(1 - a_n^2)}{n}\phi(\sqrt{n}c_n)\Big[\frac{\Phi(\sqrt{n}c_n) - \frac{1}{2}}{\phi(\sqrt{n}c_n)} - \sqrt{n}c_n\Big] \tag{4.9}$$

Furthermore, provided that $\limsup_n |a_n| \leq 1$, and $\gamma_n = \sqrt{n}c_n \to \infty$,

$$R(0, S_n) = \frac{a_n^2}{n} + \sqrt{\frac{2}{\pi}}\,\frac{1 - a_n^2}{n}\gamma_n\, e^{-\gamma_n^2/2} + o(\frac{\gamma_n e^{-\gamma_n^2/2}}{n}) \tag{4.10}$$

Corollary 1.2.1 *can be proved by using* (4.5) *and standard facts about the* $N(0, 1)$ *CDF; we will omit these details.*

An important special case of Corollary 1.2.1 is the original Hodges' estimate, for which $c_n = n^{-1/4}$ and $a_n \equiv 0$. In this case, an application of Corollary 1.2.1 gives the following asymptotic expansion; it is possible to make this into a higher order asymptotic expansion, although it is not done here.

Corollary 1.2.2 *For Hodges' estimate T_n as in* (4.1),

$$R(0, T_n) = \sqrt{\frac{2}{\pi}}n^{-3/4}e^{-\frac{\sqrt{n}}{2}} + o(n^{-3/4}e^{-\frac{\sqrt{n}}{2}}) \tag{4.11}$$

In particular,

$$\lim_{n\to\infty} \frac{\log(nR(0, T_n))}{\sqrt{n}} = -\frac{1}{2} \tag{4.12}$$

We record the following corollary for completeness. Note that $\sqrt{n}c_n$ need not go to ∞ for superefficiency to occur, as shrinkage will automatically take care of it.

Corollary 1.2.3 *Suppose $\gamma_n = \sqrt{n}c_n \to \gamma, 0 < \gamma \leq \infty$. Then, S_n is superefficient at zero, i.e.,* $\limsup_n nR(0, S_n) < 1$ *iff* $\limsup_n |a_n| < 1$.

4.2.1.2 Local Convexity and Behavior in Ultrasmall Neighborhoods

For understanding the local shape properties of the risk function of S_n, it is necessary to understand the behavior of its derivatives. This is the content of the next result, which says in particular that the risk function of all generalized Hodges estimates is locally convex near zero. For these results, we need the following notation:

$$f_n(\theta) = (1 - a_n)^2\theta\left[2\Phi(\sqrt{n}(c_n + \theta)) - 1\right] \tag{4.13}$$

$$g_n(\theta) = (a_n - 1)\left[(1 + a_n)\sqrt{n}c_n^2 + \frac{2a_n}{\sqrt{n}} + 2\sqrt{n}c_n\theta\right]\phi(\sqrt{n}(c_n + \theta)) \tag{4.14}$$

Theorem 1.2.2 *For all n and θ,*

$$\frac{d}{d\theta}R(\theta, S_n) = f_n(\theta) - f_n(-\theta) + g_n(\theta) - g_n(-\theta) \tag{4.15}$$

In particular, $\frac{d}{d\theta}R(\theta, S_n)|_{\theta=0} = 0$, and provided that $|a_n| < 1$, $\frac{d^2}{d\theta^2}R(\theta, S_n) > 0$ in a neighborhood of $\theta = 0$. Hence, under the hypothesis that $|a_n| < 1$, $R(\theta, S_n)$ is locally convex near zero, and $\theta = 0$ is a local minima of $R(\theta, S_n)$.

Proof Proof of (4.15) is a direct calculation followed by rearranging the various terms. The calculation is not presented.

That the derivative of $R(\theta, S_n)$ at $\theta = 0$ is zero follows from symmetry of $R(\theta, S_n)$, or, also immediately from (4.15). We now sketch a proof of the local convexity property. Differentiating (4.15),

$$\frac{d^2}{d\theta^2}R(\theta, S_n) = f_n'(\theta) + f_n'(-\theta) + g_n'(\theta) + g_n'(-\theta). \tag{4.16}$$

Now, on algebra,

$$f_n'(\theta) = (1 - a_n)^2\left[2\Phi(\sqrt{n}(c_n + \theta)) - 1\right] + 2\theta(1 - a_n)^2\sqrt{n}\phi(\sqrt{n}(c_n + \theta))$$

and $g_n'(\theta) = 2(a_n - 1)\sqrt{n}c_n\phi(\sqrt{n}(c_n + \theta)) - n(c_n + \theta)\phi(\sqrt{n}(c_n + \theta))$

$$\times\left[2(a_n - 1)\sqrt{n}c_n\theta + \frac{2a_n(a_n - 1)}{\sqrt{n}} + (a_n^2 - 1)\sqrt{n}c_n^2\right] \tag{4.17}$$

On substituting (4.17) into (4.16), and then setting $\theta = 0$, we get after further algebraic simplification,

$$\frac{d^2}{d\theta^2}R(\theta, S_n)|_{\theta=0} = 4(1 - a_n)^2\left[\Phi(\sqrt{n}c_n) - \frac{1}{2} - \sqrt{n}c_n\phi(\sqrt{n}c_n)\right] + 2(1 - a_n^2)c_n^3 n^{3/2}\phi(\sqrt{n}c_n) \tag{4.18}$$

By simple calculus, $\Phi(x) - \frac{1}{2} - x\phi(x) > 0$ for all positive x. Therefore, on using our hypothesis that $|a_n| < 1$, from (4.18), $\frac{d^2}{d\theta^2}R(\theta, S_n)|_{\theta=0} > 0$. It follows from the continuity of $\frac{d^2}{d\theta^2}R(\theta, S_n)$ that it remains strictly positive in a neighborhood of $\theta = 0$, which gives the local convexity property.

Remark Consider now the case of original Hodges' estimate, for which $a_n = 0$ and $c_n = n^{-1/4}$. In this case, (4.18) gives us $\lim_{n\to\infty}\frac{d^2}{d\theta^2}R(\theta, T_n)|_{\theta=0} = 2$. Together with (4.11), we then have the approximation

$$R(\theta, T_n) \approx \sqrt{\frac{2}{\pi}}n^{-3/4}e^{-\frac{\sqrt{n}}{2}} + \theta^2 \tag{4.19}$$

for θ very close to zero. Of course, we know that this approximation cannot depict the subtleties of the shape of $R(\theta, T_n)$, because $R(\theta, T_n)$ is known to have turning points, which the approximation in (4.19) fails to recognize. *We will momentarily see that $R(\theta, T_n)$ rises and turns so steeply that (4.19) is starkly inaccurate in even $n^{-1/2}$ neighborhoods of zero.*

4.2.2 Behavior in $n^{-1/2}$ Neighborhoods

We know that the superefficient estimates T_n, or S_n have a much smaller risk than the MLE at zero, and that subsequently their risks reach a peak that is much higher than that of the MLE. Therefore, these risk functions must again equal the risk of the MLE, namely $\frac{1}{n}$ at some point in the vicinity of zero. We will now first see that reversal to the $\frac{1}{n}$ level happens within $n^{-1/2}$ neighborhoods of zero. A general risk lower bound for generalized Hodges estimates S_n would play a useful role for this purpose, and also for a number of the later results. This is presented first.

Theorem 1.2.3 *Consider the generalized Hodges estimate S_n.*

(i) Suppose $0 \le a_n \le 1$. Then, for every n and $0 \le \theta \le c_n$,

$$R(\theta, S_n) \ge \frac{a_n^2}{n} + (1-a_n)^2\theta^2\Big[\Phi(\sqrt{n}(c_n+\theta)) + \Phi(\sqrt{n}(c_n-\theta)) - 1\Big] \tag{4.20}$$

(ii) Suppose $\sqrt{n}c_n \to \infty$, and that $a, 0 \le a < 1$ is a limit point of the sequence a_n. Let $\theta_n = \frac{1}{(1-a)^2\sqrt{n}}$. Then, $\limsup_n nR(\theta_n, S_n) \ge a^2 + 1$.

Proof In expression (4.5) for $e_n(\theta)$, observe the following:

$$0 \le \Phi(\sqrt{n}(c_n+\theta)) + \Phi(\sqrt{n}(c_n-\theta)) - 1 \le 1;$$

$$\text{For } 0 \le \theta \le c_n,\ \phi(\sqrt{n}(c_n+\theta)) - \phi(\sqrt{n}(c_n-\theta)) \le 0;$$

$$\text{For } 0 \le \theta \le c_n,\ (c_n+\theta)\phi(\sqrt{n}(c_n+\theta)) + (c_n-\theta)\phi(\sqrt{n}(c_n-\theta)) \ge 0.$$

Therefore, by virtue of the hypothesis $0 \le a_n \le 1$, from (4.5),

$$R(\theta, S_n) \ge \frac{1}{n} + \frac{a_n^2-1}{n} + (1-a_n)^2\theta^2\Big[\Phi(\sqrt{n}(c_n+\theta)) + \Phi(\sqrt{n}(c_n-\theta)) - 1\Big]$$

$$= \frac{a_n^2}{n} + (1-a_n)^2\theta^2\Big[\Phi(\sqrt{n}(c_n+\theta)) + \Phi(\sqrt{n}(c_n-\theta)) - 1\Big],$$

as claimed in (4.20).

For the second part of the theorem, choose a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ converging to a. For notational brevity, we denote the subsequence as a_n itself. Then, (along this subsequence), and with $\theta_n = \frac{1}{(1-a)^2\sqrt{n}}$,

$$a_n^2 + (1-a_n)^2\theta_n^2\Big[\Phi(\sqrt{n}(c_n+\theta_n)) + \Phi(\sqrt{n}(c_n-\theta_n)) - 1\Big] \to a^2 + 1 \tag{4.21}$$

Since we assume for the second part of the theorem that $\sqrt{n}c_n \to \infty$, we have that $\theta_n \le c_n$ for all large n, and hence the lower bound in (4.20) applies. Putting together

(4.20) and (4.21), and the Bolzano-Weierstrass theorem, we have one subsequence for which the limit of $nR(\theta_n, S_n)$ is $\geq a^2+1$, and hence, $\limsup_n \; nR(\theta_n, S_n) \geq a^2+1$.

We will now see that if we strengthen our control on the sequence $\{a_n\}$ to require it to have a limit, and likewise require $\sqrt{n}c_n$ also to have a limit, then the (normalized) risk of S_n at $\frac{k}{\sqrt{n}}$ will also have a limit for any given k. *Furthermore, if the limit of a_n is zero and the limit of $\sqrt{n}c_n$ is ∞, which, for instance, is the case for Hodges' original estimate, then the risk of S_n at $\frac{1}{\sqrt{n}}$ is exactly asymptotic to the risk of the MLE, namely $\frac{1}{n}$. So, reversal to the risk of the MLE occurs, more or less, at $\theta = \frac{1}{\sqrt{n}}$.* The next result says that, but in a more general form.

Theorem 1.2.4 *Consider the generalized Hodges estimate S_n.*

(a) If $a_n \to a, -\infty < a < \infty$, and $\sqrt{n}c_n \to \gamma, 0 \leq \gamma \leq \infty$, then for any fixed $k \geq 0$,

$$\lim_{n\to\infty} nR(\frac{k}{\sqrt{n}}, S_n) = 1 + \Big[a^2 - 1 + k^2(1-a)^2\Big]\Big[\Phi(k+\gamma) - \Phi(k-\gamma)\Big]$$
$$+ 2a(a-1)k\Big[\phi(k+\gamma) - \phi(k-\gamma)\Big]$$
$$+ (1-a^2)\Big[(k+\gamma)\phi(k+\gamma) - (k-\gamma)\phi(k-\gamma)\Big], \tag{4.22}$$

with (4.22) being interpreted as a limit as $\gamma \to \infty$ if $\sqrt{n}c_n \to \infty$.

(b) In particular, if $a_n \to 0$ and $\sqrt{n}c_n \to \infty$, then, $\lim_{n\to\infty} nR(\frac{k}{\sqrt{n}}, S_n) = k^2$.

(c) If $a_n = 0$ for all n and $\sqrt{n}c_n \to \infty$, then for any positive k, we have the asymptotic expansion

$$nR(\frac{k}{\sqrt{n}}, S_n) = k^2 + \frac{1}{\sqrt{2\pi}} e^{-\gamma_n^2/2 - k^2/2}$$
$$\times \Big[(\gamma_n - k)e^{k\gamma_n} + (\gamma_n + k)e^{-k\gamma_n} - (k^2-1)\frac{e^{k\gamma_n}}{\gamma_n} - (k^2-1)\frac{e^{-k\gamma_n}}{\gamma_n}\Big]$$
$$+ O(\frac{e^{-\gamma_n^2/2 + k\gamma_n}}{\gamma_n^2}) \tag{4.23}$$

(d) If $a_n = 0$ for all n and $\sqrt{n}c_n \to \infty$, then for $k = 0$, we have the asymptotic expansion

$$nR(0, S_n) = \sqrt{\frac{2}{\pi}} e^{-\gamma_n^2/2}\Big[\gamma_n + \frac{2}{\gamma_n}\Big] + O(\frac{e^{-\gamma_n^2/2}}{\gamma_n^3}) \tag{4.24}$$

The plot below nicely exemplifies the limit result in part (b) of Theorem 1.2.4 Fig. 4.3.

The proofs of the various parts of Theorem 1.2.4 involve use of standard facts about the standard normal tail and rearrangement of terms. We omit these calculations. It follows from part (b) of this theorem, by letting $k \to \infty$ that for the original Hodges' estimate T_n, $\sup_\theta R(\theta, T_n) >> \frac{1}{n}$ for large n, in the following sense.

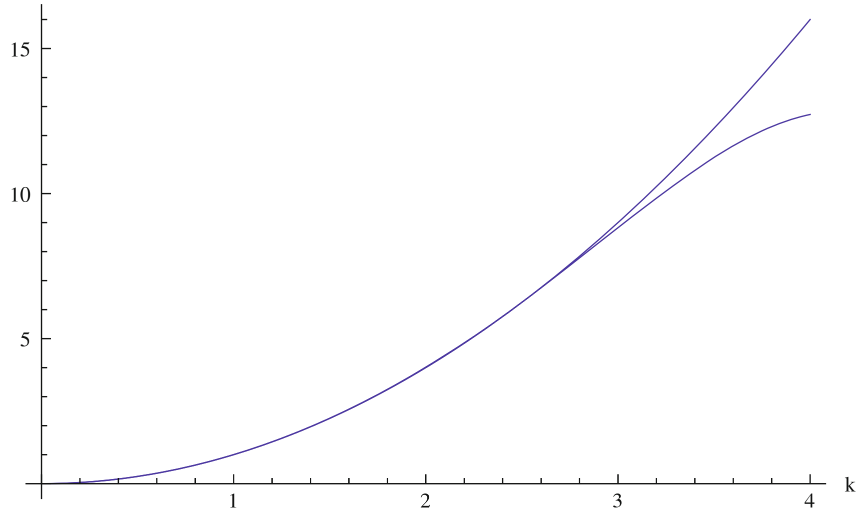

Fig. 4.3 Plot of n* Risk of Hodges' Estimate at k/sqrt(n) and k^2 for n = 500

Corollary 1.2.4 *If* $a_n \to 0$ *and* $\sqrt{nc_n} \to \infty$*, then* $\lim_n \left[\sup_\theta nR(\theta, S_n)\right] = \infty$. *On the other hand, part (c) and part (d) of the above theorem together lead to the following asymptotic expansions for the risk of Hodges' original estimate* T_n *at* $\theta = 0$ *and* $\theta = \frac{1}{\sqrt{n}}$. *We can see how close to* $\frac{1}{n}$ *the risk at* $\frac{1}{\sqrt{n}}$ *is, and the rapid relative growth of the risk near* $\theta = 0$ *by comparing the two expansions in the corollary below, which is also a strengthening of Corollary* 1.2.2.

Corollary 1.2.5 *For Hodges' estimate* T_n *as in* (4.1),

$$R(0, T_n) = \sqrt{\frac{2}{\pi}} e^{-\frac{\sqrt{n}}{2}} n^{-3/4}\left[1 + \frac{2}{\sqrt{n}}\right] + O(\frac{e^{-\frac{\sqrt{n}}{2}}}{n^{7/4}}); R(\frac{1}{\sqrt{n}}, T_n)$$
$$= \frac{1}{n} + \frac{1}{\sqrt{2\pi}} n^{-3/4} e^{-\frac{1}{2}(n^{1/4}-1)^2}\left[1 - n^{-1/4}\right] + O(\frac{e^{-\frac{1}{2}(n^{1/4}-1)^2}}{n^{3/2}}) \quad (4.25)$$

4.2.3 Behavior in c_n Neighborhoods

We saw in the previous section that reversal to the risk of the MLE occurs in $n^{-1/2}$ neighborhoods of zero. However, $n^{-1/2}$ neighborhoods are still too short for the risk to begin to approach its peak value. If $c_n >> \frac{1}{\sqrt{n}}$ and we expand the neighborhood of $\theta = 0$ to c_n neighborhoods, then the risk of S_n increases by factors of magnitude, and captures the peak value. We start with the risk of S_n at $\theta = c_n$ and analyze its asymptotic behavior.

Theorem 1.2.5 *Consider the generalized Hodges estimate S_n.*

(a) Suppose $0 \leq a_n \leq 1$ and that $\sqrt{n}c_n \to \infty$. Then, $\limsup_n c_n^{-2}R(c_n, S_n) \geq \frac{(1-\liminf_n a_n)^2}{2}$, and $\liminf_n c_n^{-2}R(c_n, S_n) \geq \frac{(1-\limsup_n a_n)^2}{2}$.

(b) If $a_n \to a, -\infty < a < \infty$, and $\sqrt{n}c_n \to \gamma, 0 \leq \gamma \leq \infty$, then

$$\lim_{n\to\infty} c_n^{-2}R(c_n, S_n) = \frac{1}{\gamma^2} + \Big[\frac{a^2-1}{\gamma^2} + (1-a)^2\Big]\Big[\Phi(2\gamma) - \frac{1}{2}\Big]$$

$$+\frac{2a(a-1}{\gamma}\Big[\phi(2\gamma) - \phi(0)\Big] + 2(1-a^2)\frac{\phi(2\gamma)}{\gamma}, \tag{4.26}$$

with (4.26) being interpreted as a limit as $\gamma \to \infty$ if $\sqrt{n}c_n \to \infty$.

Proof By (4.20),

$$R(c_n, S_n) \geq \frac{a_n^2}{n} + c_n^2(1-a_n)^2\left[\Phi(2\sqrt{n}c_n) - \frac{1}{2}\right]$$

$$\Rightarrow c_n^{-2}R(c_n, S_n) \geq (1-a_n)^2\left[\Phi(2\sqrt{n}c_n) - \frac{1}{2}\right] \tag{4.27}$$

Since $\sqrt{n}c_n \to \infty$, (4.24) implies that given $\epsilon > 0$, for all large enough n,

$$c_n^{-2}R(c_n, S_n) \geq (\frac{1}{2} - \epsilon)(1-a_n)^2$$

$$\Rightarrow \limsup_n c_n^{-2}R(c_n, S_n) \geq \limsup_n (\frac{1}{2} - \epsilon)(1-a_n)^2 = (\frac{1}{2} - \epsilon)(1 - \liminf_n a_n)^2.$$

Since $\epsilon > 0$ is arbitrary, this means $\limsup_n c_n^{-2}R(c_n, S_n) \geq \frac{(1-\liminf_n a_n)^2}{2}$; the lim inf inequality follows similarly.

4.2.3.1 Behavior Near c_n and Approach to the Peak

Theorem 1.2.6 *Consider the generalized Hodges estimate S_n. Suppose $a_n = 0$ for all n and $\gamma_n = \sqrt{n}c_n \to \infty$. Then, for any fixed $\alpha, 0 < \alpha \leq 1$, we have the asymptotic expansion*

$$c_n^{-2}R((1-\alpha)c_n, S_n) = (1-\alpha)^2 + \frac{\phi(\alpha\gamma_n)}{\alpha\gamma_n}(2\alpha - 1) + \frac{\phi((2-\alpha)\gamma_n)}{(2-\alpha)\gamma_n}(3-2\alpha)$$
$$+ O(\frac{\phi(\alpha\gamma_n)}{\gamma_n^3}) \tag{4.28}$$

Proof: Fix $0 < \alpha < 1$, and denote $\theta_n = (1-\alpha)c_n$. Using (4.5),

$$R(\theta_n, S_n) = \frac{1}{n} + \Big[(1-\alpha)^2 c_n^2 - \frac{1}{n}\Big]\Big[\Phi((2-\alpha)\gamma_n) - \Phi(-\alpha\gamma_n)\Big]$$
$$+ \frac{1}{\sqrt{n}}\Big[(2-\alpha)c_n\phi((2-\alpha)\gamma_n) + \alpha c_n \phi(\alpha\gamma_n)\Big]$$
$$\Rightarrow c_n^{-2}R(\theta_n, S_n) = \frac{1}{\gamma_n^2} + \Big[(1-\alpha)^2 - \frac{1}{\gamma_n^2}\Big]\Big[\Phi((2-\alpha)\gamma_n) - \Phi(-\alpha\gamma_n)\Big]$$
$$+ \frac{1}{\gamma_n}\Big[(2-\alpha)\phi((2-\alpha)\gamma_n) + \alpha\phi(\alpha\gamma_n)\Big] = \frac{1}{\gamma_n^2} + \Big[(1-\alpha)^2 - \frac{1}{\gamma_n^2}\Big]$$
$$\Big[1 - \frac{\phi((2-\alpha)\gamma_n)}{(2-\alpha)\gamma_n}(1 + O(\gamma_n^{-2})) - \frac{\phi(\alpha\gamma_n)}{\alpha\gamma_n}(1 + O(\gamma_n^{-2}))\Big]$$
$$+ \frac{(2-\alpha)\phi((2-\alpha)\gamma_n)}{\gamma_n} + \frac{\alpha\phi(\alpha\gamma_n)}{\gamma_n}$$
$$= (1-\alpha)^2 + \frac{\phi((2-\alpha)\gamma_n)}{\gamma_n}\Big[(2-\alpha) - \frac{(1-\alpha)^2}{2-\alpha}\Big]$$
$$+ \frac{\phi(\alpha\gamma_n)}{\gamma_n}\Big[\alpha - \frac{(1-\alpha)^2}{\alpha}\Big] + O(\frac{\phi(\alpha\gamma_n)}{\gamma_n^3}). \tag{4.29}$$

The theorem now follows from (4.29).

By scrutinizing the proof of Theorem 1.2.6, we notice that the constant α can be generalized to suitable sequences α_n, and this gives us a useful and more general corollary. Note that, indeed, the remainder term in the corollary below is $O(\frac{\phi(\alpha_n\gamma_n)}{\gamma_n})$, rather than $O(\frac{\phi(\alpha_n\gamma_n)}{\gamma_n^3})$.

Corollary 1.2.6 *Consider the generalized Hodges estimate S_n. Suppose $a_n = 0$ for all n and $\gamma_n = \sqrt{n}c_n \to \infty$. Let α_n be a positive sequence such that $\alpha_n \to 0, \alpha_n\gamma_n \to \infty$. Let $\theta_n = (1-\alpha_n)c_n$. Then we have the asymptotic expansion*

$$c_n^{-2}R(\theta_n, S_n) = (1-\alpha_n)^2 - \frac{\phi(\alpha_n\gamma_n)}{\alpha_n\gamma_n} + O(\frac{\phi(\alpha_n\gamma_n)}{\gamma_n}) \tag{4.30}$$

Remark Together, Theorem 1.2.5 and Corollary 1.2.6 enable us to make the following conclusion: *at $\theta = c_n, R(\theta, S_n) \approx \frac{c_n^2}{2} >> \frac{1}{n}$, which is the risk of the MLE, provided $\gamma_n = \sqrt{n}c_n \to \infty$. If we move slightly to the left of $\theta = c_n$, then the risk increases even more. Precisely, if we take $\theta = (1-\alpha_n)c_n$ with a very small α_n, then $R(\theta, S_n) \approx c_n^2$. We believe that this is the exact rate of convergence of the global maximum of the risk,* i.e.,

$$\lim_{n\to\infty} c_n^{-2} \sup_{-\infty<\theta<\infty} R(\theta, S_n) = 1. \tag{4.31}$$

4.2.3.2 Global Maximum of the Risk and Point of Maxima

Corollary 1.2.6 suggests a pathway to addressing the two related questions: what is an approximation to the point at which the global maximum of the risk is attained, and what is a higher order approximation to the value of the global maximum. In Eq. (4.31), if we use the two leading terms $(1-\alpha_n)^2 - \frac{\phi(\alpha_n\gamma_n)}{\alpha_n\gamma_n}$, we notice that $(1-\alpha)^2$ and $\frac{\phi(\alpha\gamma_n)}{\alpha\gamma_n}$ are both decreasing in α. Therefore, if we maximize $(1-\alpha)^2 - \frac{\phi(\alpha\gamma_n)}{\alpha\gamma_n}$ over α (in (0, 1)), it will give us an approximation to the global maximum of $R(\theta, S_n)$ and at the same time, an approximation to the point $\theta_n = (1-\alpha_n)c_n$ where the maximum is attained. *It must be understood that these two approximations are heuristic, because we do not have a proof that* $\sup_\theta R(\theta, S_n)$ *is attained at a point of the form* $(1-\alpha_n)c_n$ *with* α_n *as in Corollary 1.2.6.*

To maximize $(1-\alpha)^2 - \frac{\phi(\alpha\gamma_n)}{\alpha\gamma_n}$, we want to find the root of

$$0 = \frac{d}{d\alpha}\left[(1-\alpha)^2 - \frac{\phi(\alpha\gamma_n)}{\alpha\gamma_n}\right]$$

$$= 2(\alpha-1) + \phi(\alpha\gamma_n)\left[\gamma_n + \frac{1}{\alpha\gamma_n}\right] = 2(\alpha-1) + \gamma_n\phi(\alpha\gamma_n) + 0(\gamma_n\phi(\alpha\gamma_n))$$

$$\Rightarrow (1-\alpha) = \frac{\gamma_n}{2}\phi(\alpha\gamma_n)(1+0(1))$$

$$\Rightarrow -\alpha = \log\gamma_n - \frac{\alpha^2\gamma_n^2}{2} + O(1)$$

$$\Rightarrow \alpha^2\gamma_n^2 - 2\alpha - 2\log\gamma_n + O(1) = 0 \tag{4.32}$$

An approximation to the root of the quadratic Eq. (4.32) is

$$\alpha = \frac{\sqrt{2\log\gamma_n}}{\gamma_n}, \tag{4.33}$$

which results in the following two heuristic approximations:

Conjecture *In the class of estimates*

$$S_n(X_1,\ldots,X_n) = \begin{matrix} \bar{X}_n & \text{if} & |\bar{X}_n| > c_n \\ 0 & \text{if} & |\bar{X}_n| \le c_n \end{matrix}, \tag{4.34}$$

one has,

$$\text{argmax}_{-\infty<\theta<\infty} R(\theta, S_n) \approx c_n - \sqrt{\frac{\log(nc_n^2)}{n}}; \quad \sup_{-\infty<\theta<\infty} R(\theta, S_n) \approx c_n^2 - \frac{2c_n\sqrt{\log n}}{\sqrt{n}}. \tag{4.35}$$

Example 1.2.1 We look at the credibility of (4.35) for Hodges' original estimate T_n, for which $c_n = n^{-1/4}$. In that case, (4.35) says that the global maximum of $R(\theta, T_n)$ should be approximately $\frac{1}{\sqrt{n}} - \frac{2\sqrt{\log n}}{n^{3/4}}$ and it should be attained at $\theta_n \approx n^{-1/4} - \sqrt{\frac{\log n}{2n}}$. We show in the following table the exact global maximum (computed numerically), the risk at c_n and at θ_n and the approximation to the maximum risk as claimed in (4.35). For very large n, our conjecture appears to work out almost exactly. Otherwise, it does not.

n	Exact Maximum	$R(c_n, T_n)$	$R(\theta_n, T_n)$	Approx. (1.35)
100	0.0558	0.0550	0.0112	0.0357
2500	0.0126	0.0102	0.0073	0.0042
100000	0.0025	0.0016	0.0021	0.0020
250000	0.0016	0.0010	0.0014	0.0014
10^6	0.0008	0.0005	0.0008	0.0008

4.2.3.3 Optimal Thresholding

The approximation laid out in (4.35) enables us to pose and give a solution to another relevant question: what is an optimal choice of the thresholding parameter (sequence) c_n? Obviously, this calls for a definition of optimal thresholding. We adopt the definition of controlled minimaxity. Here is an explanation, and then a formal mathematical definition.

It is clear that the choice of the thresholding parameter affects two key quantities in the problem, the risk at zero, and the maximum risk. For instance, as an extreme, if we choose $c_n = 0$, then the risk at zero is zero, but the maximum risk is infinity. Thus, there is a trade-off between $R(0, S_n)$ and $\sup_\theta R(\theta, S_n)$, and the thresholding parameter c_n influences both of them, but in opposite directions. It seems reasonable to ask for the sequence c_n that minimizes $\sup_\theta R(\theta, S_n)$ subject to a guaranteed percentage improvement in risk over the MLE at $\theta = 0$. More precisely, the question is: which sequence c_n minimizes $\sup_\theta R(\theta, S_n)$ subject to the constraint $n|e_n(0)| \geq 1 - \epsilon_n$, where, $e_n(\theta) = R(\theta, S_n) - \frac{1}{n}$. Thus, in this formulation we seek the thresholding estimate that is minimax subject to a risk gain of at least $100(1 - \epsilon_n)\%$ at zero; ϵ_n is supposed to be user provided. Such restricted minimax formulations have been proposed and studied in other problems before; one reference is Bickel (1983).

From (4.9) and (4.35), we wish to

$$\text{minimize } \gamma_n^2 - 2\gamma_n\sqrt{\log n} \text{ subject to } H(\gamma_n) = \Phi(\gamma_n) - \frac{1}{2} - \gamma_n\phi(\gamma_n) \geq \frac{1 - \epsilon_n}{2}$$

The unconstrained minimum of $\gamma_n^2 - 2\gamma_n\sqrt{\log n}$ is $\gamma_n = \sqrt{\log n}$. If $H(\sqrt{\log n}) \geq \frac{1-\epsilon_n}{2}$ (which approximately corresponds to $\epsilon_n \geq \frac{1}{\sqrt{n}}$, then the solution to our problem is $\gamma_n = \sqrt{\log n}$. Otherwise, since $H(x)$ is increasing in x for positive x, i.e., increasing in x for $x > 0$, it follows that the sequence γ_n that solves the constrained minimum problem is the root of the equation

$$\Phi(\gamma_n) - \frac{1}{2} - \gamma_n\phi(\gamma_n) = \frac{1-\epsilon_n}{2} \tag{4.36}$$

$$\Leftrightarrow 1 - \Phi(\gamma_n) + \gamma_n\phi(\gamma_n) = \frac{\epsilon_n}{2}$$

$$\Leftrightarrow \phi(\gamma_n)\Big[\gamma_n + O(\frac{1}{\gamma_n})\Big] = \frac{\epsilon_n}{2}$$

$$\Leftrightarrow \sqrt{\frac{\pi}{2}} e^{\frac{\gamma_n^2}{2}} \frac{\gamma_n}{\gamma_n^2 + O(1)} = \frac{1}{\epsilon_n} \tag{4.37}$$

A first approximation to the root of (4.36) is $\gamma_n = \sqrt{2\log\frac{1}{\epsilon_n}}$. Plugging the first approximation back into (4.36), a higher order approximation is

$$\gamma_n^2 = 2\log\frac{1}{\epsilon_n} + 2\log\left(\sqrt{2\log\frac{1}{\epsilon_n}}\right) = 2\log\frac{1}{\epsilon_n} + \log\log\frac{1}{\epsilon_n} + O(1),$$

which gives

$$\gamma_n = \sqrt{2\log\frac{1}{\epsilon_n} + \log\log\frac{1}{\epsilon_n} + O(1)} = \sqrt{2\log\frac{1}{\epsilon_n}}\Big[1 + \frac{\log\log\frac{1}{\epsilon_n}}{4\log\frac{1}{\epsilon_n}} + o\left(\frac{\log\log\frac{1}{\epsilon_n}}{\log\frac{1}{\epsilon_n}}\right)\Big]$$

$$= \sqrt{2\log\frac{1}{\epsilon_n}} + \frac{\log\log\frac{1}{\epsilon_n}}{2\sqrt{2\log\frac{1}{\epsilon_n}}} + o\left(\frac{\log\log\frac{1}{\epsilon_n}}{\sqrt{\log\frac{1}{\epsilon_n}}}\right)$$

We propose finally the following thresholding sequence:

$$\gamma_n = \sqrt{n}c_n = \sqrt{\log n}, \quad \text{if } \epsilon_n \geq \frac{1}{\sqrt{n}},$$

$$\gamma_n = \sqrt{n}c_n = \sqrt{2\log\frac{1}{\epsilon_n} + \frac{\log\log\frac{1}{\epsilon_n}}{2\sqrt{2\log\frac{1}{\epsilon_n}}}}, \quad \text{if } \epsilon_n < \frac{1}{\sqrt{n}} \tag{4.38}$$

Example 1.2.2 The recommended thresholding sequence in (4.38) depends on the specification of ϵ_n. We work out the form of c_n for four choices of ϵ_n. Suppose, independent of n, we want a fixed percentage risk improvement $100(1-\epsilon)\%$ at zero. Then, $\epsilon_n \equiv \epsilon$, which, by (4.38), leads to

$$c_n = \sqrt{\frac{\log n}{n}}$$

Thus, a fixed percentage risk improvement at zero leads to $c_n \sim \sqrt{\frac{\log n}{n}}$.

Suppose we want the percentage risk improvement at zero to increase with n at a polynomial rate, $\epsilon_n = n^{-\beta}, \beta > \frac{1}{2}$. Then, (4.38) leads to

$$c_n = \frac{\sqrt{2\beta \log n}}{\sqrt{n}} + \frac{\log \log n}{2\sqrt{2\beta n \log n}} + O(\frac{1}{\sqrt{n \log n}}).$$

Thus, for polynomial growth in the percentage risk improvement at zero, still, the recommended thresholding sequence $c_n \sim \sqrt{\frac{\log n}{n}}$, but with a constant in front that is > 1.

Next, suppose we want the percentage risk improvement at zero to increase at a subexponential rate, namely, $\epsilon_n = e^{-\beta\sqrt{n}}, \beta > 0$. Then, (4.38) leads to

$$c_n = \sqrt{2\beta} n^{-1/4} + \frac{\log n}{4\sqrt{2\beta} n^{3/4}}.$$

Thus, for subexponential growth in the percentage risk improvement at zero,we get $c_n \sim n^{-1/4}$. *Compare this with Eq. (4.11) which describes the percentage risk improvement at zero of Hodges' original estimate T_n. Interestingly, his choice of $c_n = n^{-1/4}$ matches to the first order the recommended sequence we just derived above.*

Finally, suppose we want the percentage risk improvement at zero to increase at the fully exponential rate, namely, $\epsilon_n = e^{-\beta n}, \beta > 0$. Then, (4.38) leads to

$$c_n = \sqrt{2\beta} + \frac{\log n}{2\sqrt{2\beta} n}.$$

Thus, for exponential growth in the percentage risk improvement, we get $c_n \sim c$, a constant.

4.2.4 Comparison of Bayes Risks and Regular Variation

Since the risk functions of the MLE and thresholding estimates S_n cross, it is meaningful to seek a comparison between them by using Bayes risks. Because of the intrinsic specialty of the point $\theta = 0$ in this entire problem, it is sensible to consider priors that are symmetric about zero. Purely for technical convenience, we only consider normal priors here, $N(0, \sigma_n^2)$, and we ask the following question: how should σ_n behave for the thresholding estimate to have (asymptotically) a smaller Bayes risk than the MLE? It turns out that certain interesting stories emerge in answering the question, and we have a fairly complete answer to the question we have posed.

We start with some notation. Let $\pi = \pi_n$ denote a prior density and $B_n(S_n, \pi)$ the Bayes risk of S_n under π. Let also $B_n(\pi)$ denote the Bayes risk of the Bayes rule under π. Then,

$$B_n(S_n, \pi) = \int R(\theta, S_n)\pi(\theta)d\theta = \frac{1}{n} + \int e_n(\theta)\pi(\theta)d\theta \tag{4.39}$$

and

$$B_n(\pi) = \frac{1}{n} - \frac{1}{n^2}\int \frac{(m'(x))^2}{m(x)}dx, \tag{4.40}$$

where $m(x) = m_n(x)$ denotes the marginal density of $\bar{X}$ under π. In the case where $\pi = \pi_n$ is the $N(0, \sigma_n^2)$ density, $B_n(\pi) = \frac{\sigma_n^2}{n\sigma_n^2+1}$.

4.2.4.1 Normal Priors

We use (4.5) to write a closed form formula for $B_n(S_n, \pi)$; it is assumed until we specifically mention otherwise that henceforth $\pi = N(0, \sigma_n^2)$, and for brevity, we drop the subscript and write σ^2 for σ_n^2.

Toward this agenda, the following formulas are used; for reasons of space, we will not provide their derivations.

$$\int \Phi(\sqrt{n}(c_n \pm \theta))\frac{1}{\sigma}\phi(\frac{\theta}{\sigma})d\theta = \Phi(\frac{\sqrt{n}c_n}{\sqrt{1+n\sigma^2}}) \tag{4.41}$$

$$\int \phi(\sqrt{n}(c_n \pm \theta))\frac{1}{\sigma}\phi(\frac{\theta}{\sigma})d\theta = \frac{\sigma e^{-nc_n^2/(2(1+n\sigma^2))}}{\sqrt{2\pi}\sqrt{1+n\sigma^2}} \tag{4.42}$$

$$\int \theta\phi(\sqrt{n}(c_n \pm \theta))\frac{1}{\sigma}\phi(\frac{\theta}{\sigma})d\theta = \mp\frac{\sigma^2 nc_n e^{-nc_n^2/(2(1+n\sigma^2))}}{\sqrt{2\pi}(1+n\sigma^2)^{3/2}} \tag{4.43}$$

$$\int \theta^2\Phi(\sqrt{n}(c_n \pm \theta))\frac{1}{\sigma}\phi(\frac{\theta}{\sigma})d\theta = \sigma^2\Big[\Phi(\frac{\sqrt{n}c_n}{\sqrt{1+n\sigma^2}}) - \frac{\sigma^2 n^{3/2}c_n e^{-nc_n^2/(2(1+n\sigma^2))}}{\sqrt{2\pi}(1+n\sigma^2)^{3/2}}\Big] \tag{4.44}$$

By plugging (4.41), (4.42), (4.43), (4.44) into $\int e_n(\theta)\frac{1}{\sigma}\phi(\frac{\theta}{\sigma})d\theta$, where the expression for $e_n(\theta)$ is taken from (4.5), additional algebraic simplification gives us the following closed form expression.

Theorem 1.2.7

$$B_n(S_n, \pi) = \frac{1}{n} + \int e_n(\theta)\pi(\theta)d\theta,$$

with

$$\begin{aligned}\int e_n(\theta)\pi(\theta)d\theta &= \frac{1-a_n^2}{n} - (1-a_n)^2\sigma^2 \\ &+ \Big[2(1-a_n)^2\sigma^2 - \frac{2(1-a_n^2)}{n}\Big]\Phi\left(\frac{\sqrt{n}c_n}{\sqrt{1+n\sigma^2}}\right) \\ &- \frac{\sqrt{n}c_n}{\sqrt{1+n\sigma^2}}\phi\left(\frac{\sqrt{n}c_n}{\sqrt{1+n\sigma^2}}\right)\Big[\frac{2n(1-a_n)^2\sigma^4}{1+n\sigma^2} \\ &+ \frac{2(1-a_n)^2\sigma^2}{1+n\sigma^2} - \frac{2(1-a_n^2)}{n}\Big] \qquad (4.45)\end{aligned}$$

Theorem 1.2.7 leads to the following more transparent corollary.

Corollary 1.2.7 Consider the generalized Hodges estimate S_n with $a_n \equiv 0$. Then

$$\int e_n(\theta)\pi(\theta)d\theta = \frac{n\sigma^2 - 1}{n}\Big[2\Phi(\frac{\gamma_n}{\sqrt{1+n\sigma^2}}) - \frac{1+n\sigma^2}{1+\sigma^2}\frac{\gamma_n}{\sqrt{1+n\sigma^2}}\phi(\frac{\gamma_n}{\sqrt{1+n\sigma^2}})\Big] \qquad (4.46)$$

In particular, if $\sigma^2 = \frac{1}{n}$, then whatever be the thresholding sequence c_n, $B_n(S_n, \pi) = \frac{1}{n}$, i.e., S_n and the MLE $\bar{X}$ have the same Bayes risk if $\theta \sim N(0, \frac{1}{n})$. By inspecting (4.46), we can make more general comparisons between $B_n(S_n, \pi)$ and $\frac{1}{n} = B_n(\bar{X}, \pi)$ when $\sigma^2 \neq \frac{1}{n}$. It turns out that $\sigma^2 = \frac{1}{n}$ acts in a very meaningful sense as a boundary between $B_n(S_n, \pi) < B_n(\bar{X}, \pi)$ and $B_n(S_n, \pi) > B_n(\bar{X}, \pi)$. We will now make it precise. In this analysis, it will be useful to note that once we know whether $\sigma^2 >$ or $< \frac{1}{n}$, by virtue of formula (4.46), the algebraic sign of $\triangle_n(\pi) = B_n(S_n, \pi) - B_n(\bar{X}, \pi)$ is determined by the algebraic sign of $\eta_n = 2\Phi(\frac{\gamma_n}{\sqrt{1+n\sigma^2}}) - \frac{1+n\sigma^2}{1+\sigma^2}\frac{\gamma_n}{\sqrt{1+n\sigma^2}}\phi(\frac{\gamma_n}{\sqrt{1+n\sigma^2}})$.

Theorem 1.2.8 *Provided the thresholding sequence c_n satisfies $c_n \to 0, \gamma_n = \sqrt{n}c_n \to \infty$,*

(a) $\triangle_n(\pi) < 0$ for all large n if $\sigma^2 = \frac{c}{n} + o(\frac{1}{n})$ for some $c, 0 \le c < 1$.
(b) $\triangle_n(\pi) > 0$ for all large n if $\sigma^2 = \frac{c}{n} + o(\frac{1}{n})$ for some $c, c > 1$.
(c) $\triangle_n(\pi) = 0$ for all n if $\sigma^2 = \frac{1}{n}$.
(d) If $n\sigma^2 \to 1$, then in general $\triangle_n(\pi)$ oscillates around zero.
(e) If $n\sigma^2 \to \infty$, then $\triangle_n(\pi) < 0$ for all large n.

Proof We indicate the proof of part (a). In this case, $n\sigma^2 - 1 < 0$ for all large n. On the other hand,

$$\Phi(\frac{\gamma_n}{\sqrt{1+n\sigma^2}}) \to 1;\ \frac{1+n\sigma^2}{1+\sigma^2} \to 1+c;\ \frac{\gamma_n}{\sqrt{1+n\sigma^2}}\phi(\frac{\gamma_n}{\sqrt{1+n\sigma^2}}) \to 0.$$

Therefore, $\eta_n \to 1 > 0$, and hence, for all all large n, $\triangle_n(\pi) < 0$. The other parts use the same line of argument and so we do not mention them.

4.2.4.2 General Smooth Priors

We now give an asymptotic expansion for $\triangle_n = B_n(S_n, \pi) - B_n(\bar{X}, \pi)$ for general smooth prior densities of the form $\pi(\theta) = \pi_n(\theta) = \sqrt{n}h(\theta\sqrt{n})$, where h is a fixed sufficiently smooth density function on $(-\infty, \infty)$. It will be seen below that scaling by $\sqrt{n}$ is the right scaling to do in π_n, similar to our finding that in the normal case, $\sqrt{n}\,\theta \sim N(0,1)$ acts as a boundary between $\triangle_n < 0$ and $\triangle_n > 0$. We introduce the following notation

$$q(z) = \int_0^z (t^2-1)h(t)dt - h'(z), -\infty < z < \infty;\ \ w(z) = -\frac{d}{dz}\log q(z). \quad (4.47)$$

The functions $q(z)$ and $\log q(z)$ will play a pivotal role in the three main results below, Theorem 1.2.9, Proposition 1.2.1, and Theorem 1.2.10. Note that $q(z) \equiv 0$ if $h = \phi$, the standard normal density. For general h, q can take both positive and negative values, and this will complicate matters in the analysis that follows.

We will need the following assumptions on h and q. *Not all of the assumptions are needed for every result below.* But we find it convenient to list all the assumptions together, at the expense of some generality.

Assumptions on h

(1) $h(z) < \infty \forall z$.
(2) $h(-z) = h(z) \forall z$.
(3) $\int_{-\infty}^{\infty} z^2\, h(z)dz < \infty$.
(4) h is twice continuously differentiable, and $h'(z) \to 0$ as $z \to \infty$.
(5) q is ultimately decreasing and positive.
(6) $\log q$ is absolutely continuous, ultimately negative, and ultimately concave or convex.
(7) $\liminf_{z\to\infty} \frac{d}{dz}\log q(z) > -\infty$.

The first result below, Theorem 1.2.9, is on a unified convolution representation and some simple asymptotic order results for the Bayes risk difference $\triangle_n = B_n(S_n, \pi) - B_n(\bar{X}, \pi)$. A finer result on the asymptotic order of $\triangle_n$ is the content of Theorem 1.2.10. *In the result below, (4.49) and (4.50) together say that the first order behavior of $\triangle_n$ is determined by whether or not $Var_h(\theta) = 1$. If $Var_h(\theta) \neq 1$, then $\triangle_n$ converges at the rate $\frac{1}{n}$; but if $Var_h(\theta) = 1$, then $\triangle_n$ converges at a rate faster than $\frac{1}{n}$. This provides greater insight into the result of part (c) of Theorem 1.2.8.*

Theorem 1.2.9 *Consider generalized Hodges estimates S_n of the form (1.2) with $a_n \equiv 0$. Let h be a fixed density function satisfying the assumptions (1)-(4) above and let $\pi(\theta) = \pi_n(\theta) = \sqrt{n}h(\theta\sqrt{n}), -\infty < \theta < \infty$. Then we have the identity*

$$\triangle_n = \frac{2}{n}(q * \phi)(\gamma_n) = \frac{2}{n}\int_{-\infty}^{\infty} q(z)\phi(\gamma_n - z)dz$$
$$= \frac{2}{n}\int_{0}^{\infty} q(z)\Big[\phi(\gamma_n - z) - \phi(\gamma_n + z)\Big]dz \tag{4.48}$$

In particular, if $q \in \mathcal{L}_1$, then

$$n\triangle_n \to 0, \text{ i.e., } \triangle_n = o(\frac{1}{n}), \tag{4.49}$$

and if $q(z) \to c \neq 0$ as, $z \to \infty$, then

$$n\triangle_n \to 2c, \text{ i.e., } \triangle_n = \frac{2c}{n} + o(\frac{1}{n}). \tag{4.50}$$

In any case, if $\text{Var}_h(\theta) < \infty$, and $h' \in \mathcal{L}_\infty$, then, for every fixed n,

$$|n\triangle_n| \le 1 + \text{Var}_h(\theta) + ||h'||_\infty. \tag{4.51}$$

Proof Using (4.5) and the definition of $\pi(\theta)$,

$$\triangle_n = \int_{-\infty}^{\infty} e_n(\theta)\pi_n(\theta)d\theta$$
$$= \int_{-\infty}^{\infty}(\theta^2 - \frac{1}{n})\Big[\Phi(\gamma_n + \theta\sqrt{n}) + \Phi(\gamma_n - \theta\sqrt{n}) - 1\Big]\sqrt{n}h(\theta\sqrt{n})d\theta$$
$$+ \frac{1}{\sqrt{n}}\int_{-\infty}^{\infty}\Big[(c_n + \theta)\phi(\gamma_n + \theta\sqrt{n}) + (c_n - \theta)\phi(\gamma_n - \theta\sqrt{n})\Big]\sqrt{n}h(\theta\sqrt{n})d\theta$$
$$= \frac{1}{n}\Big(\int_{-\infty}^{\infty}(z^2 - 1)\Big[\Phi(\gamma_n + z) + \Phi(\gamma_n - z) - 1\Big]h(z)dz$$
$$+ \int_{-\infty}^{\infty}\Big[(\gamma_n + z)\phi(\gamma_n + z) + (\gamma_n - z)\phi(\gamma_n - z)\Big]h(z)dz\Big)$$
$$= \frac{1}{n}\Big(\int_{-\infty}^{\infty}(z^2 - 1)\Big[2\Phi(\gamma_n + z) - 1\Big]h(z)dz + 2\int_{-\infty}^{\infty}(\gamma_n + z)\phi(\gamma_n + z)h(z)dz\Big)$$
$$= \frac{2}{n}\Big(\int_{-\infty}^{\infty}(z^2 - 1)h(z)\Phi(\gamma_n + z)dz + \int_{-\infty}^{\infty}(\gamma_n + z)\phi(\gamma_n + z)h(z)dz\Big)$$
$$- \frac{1}{n}\int_{-\infty}^{\infty}(z^2 - 1)h(z)dz$$
$$= \frac{2}{n}\Big(\int_{-\infty}^{\infty}(z^2 - 1)h(z)\Phi(\gamma_n + z)dz - \int_{-\infty}^{\infty}\Phi(\gamma_n + z)h''(z)dz\Big)$$
$$- \frac{1}{n}\int_{-\infty}^{\infty}(z^2 - 1)h(z)dz$$

(by twice integrating by parts the integral $\int_{-\infty}^{\infty}(\gamma_n+z)\phi(\gamma_n+z)h(z)dz$)

$$= \frac{2}{n}\int_{-\infty}^{\infty}\Big[(z^2-1)h(z)-h''(z)\Big]\Phi(\gamma_n+z)dz - \frac{1}{n}\int_{-\infty}^{\infty}(z^2-1)h(z)dz$$

$$= \frac{2}{n}\int_{-\infty}^{\infty}q'(z)\Phi(\gamma_n+z)dz - \frac{1}{n}\int_{-\infty}^{\infty}(z^2-1)h(z)dz$$

$$= \frac{2}{n}\Big(q(z)\Phi(\gamma_n+z)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}q(z)\phi(\gamma_n+z)dz\Big) - \frac{1}{n}\int_{-\infty}^{\infty}(z^2-1)h(z)dz$$

$$= \frac{2}{n}\int_{0}^{\infty}(z^2-1)h(z)dz - \frac{2}{n}\int_{-\infty}^{\infty}q(z)\phi(\gamma_n+z)dz - \frac{1}{n}\int_{-\infty}^{\infty}(z^2-1)h(z)dz$$

(refer to (4.47))

$$= -\frac{2}{n}\int_{-\infty}^{\infty}q(z)\phi(\gamma_n+z)dz$$

(since $2\int_0^{\infty}(z^2-1)h(z)dz = \int_{-\infty}^{\infty}(z^2-1)h(z)dz$)

$$= \frac{2}{n}\int_{-\infty}^{\infty}q(z)\phi(\gamma_n-z)dz$$

$$= \frac{2}{n}\int_{0}^{\infty}q(z)\Big[\phi(\gamma_n-z)-\phi(\gamma_n+z)\Big]dz \qquad (4.52)$$

(since $q(-z)=-q(z)$ for all z), and this gives (4.48). (4.49), (4.50), and (4.51) follow on application of the dominated convergence theorem and the triangular inequality, and this establishes the theorem.

Remark Eq. (4.48) is a pleasant general expression for the Bayes risk difference $\triangle_n$ and what is more, has the formal look of a convolution density. One might hope that techniques from the theory of convolutions can be used to assert useful things about the asymptotic behavior of $\triangle_n$, via (4.48). We will see that indeed this is the case.

Before embarking on further analysis of $\triangle_n$, we need to keep two things in mind. First, the function $q(z)$ is usually a signed function and, therefore, we are not dealing with convolutions of probability measures in (4.48). This adds a bit of additional complexity into the analysis. Second, it does not take too much to fundamentally change the asymptotic behavior of $\triangle_n$. In the two pictures below, we have plotted $\int_0^{\infty}q[z]\Big[\phi(\gamma-z)-\phi(\gamma+z)\Big]dz$, for two different choices of the (probability density) function h. In the first picture, h is a standard Laplace (double exponential) density, while in the second picture, h is a Laplace density scaled to have variance exactly equal to 1. We can see that just a scale change changes both the asymptotic (in γ) sign and shape of $\triangle_n$ (refer to (4.49) and (4.50) as well). Thus, in our further analysis of $\triangle_n$ by exploiting the formula in (4.48), we must remain mindful of small changes in h that can make big changes in (4.48).

For future reference, we record the following formula.

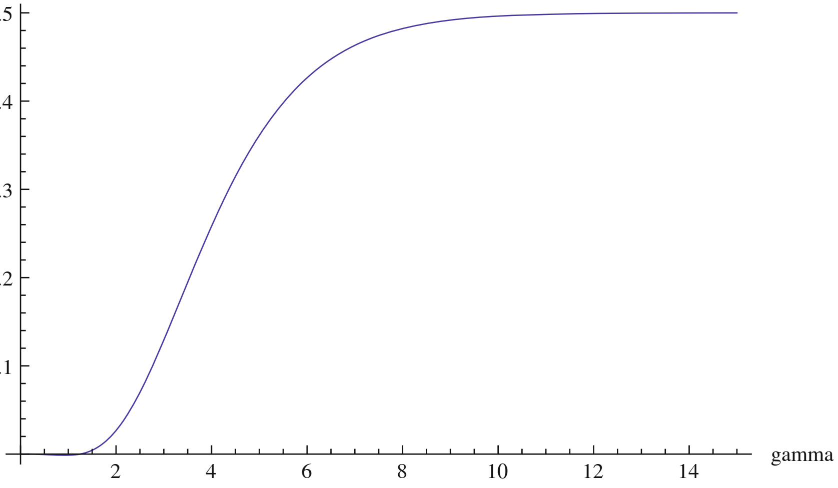

Fig. 4.4 Plot of (48) for a Standard Double Exponential h

If $h(t) = \frac{1}{2\sigma}e^{-|t|/\sigma}$, then (for $z > 0$),

$$q(z) = \sigma^2 - \frac{1}{2} + (\alpha_0 + \alpha_1 z + \alpha_2 z^2)e^{-z/\sigma}, \tag{4.53}$$

where

$$\alpha_0 = \frac{1}{2} + \frac{1}{2\sigma^2} - \sigma^2, \quad \alpha_1 = -\sigma, \quad \alpha_2 = -\frac{1}{2}$$

Thus, if $2\sigma^2 \neq 1$, then q acts asymptotically like a nonzero constant; but if $2\sigma^2 = 1$, then asymptotically q dies. This affects the asymptotic sign and shape of the convolution expression (4.48), and explains why the two pictures below look so different. Fig. 4.4 and 4.5

The next technical proposition will be useful for our subsequent analysis of (4.48) and $\triangle_n$. For this proposition, we need two special functions.

For $-\infty < p < \infty$, by $D_p(z)$ we denote the *parabolic cylinder function* which solves the differential equation $u'' + (p + \frac{1}{2} - \frac{z^2}{4})u = 0$. For $-\infty < a < \infty$ and $c \neq 0, -1, -2, \cdots$, $M(a, c, z)$ (also often written as ${}_1F_1(a, c, z)$) denotes the *confluent hypergeometric function* $\sum_{k=0}^{\infty} \frac{(a)_k}{(c)_k} \frac{z^k}{k!}$. We have the following proposition.

Proposition 1.2.1 *Let $k \geq 0$ be an integer and a a nonnegative real number. Then, for any real number μ,*

$$\int_0^\infty z^k e^{-az}\phi(\mu - z)dz = \frac{k! e^{-\mu^2/2}}{2^{k/2+1}}\Big[\frac{M(\frac{k+1}{2}, \frac{1}{2}, \frac{(\mu-a)^2}{2})}{\Gamma(\frac{k+2}{2})} + \sqrt{2}(\mu - a)\frac{M(\frac{k+2}{2}, \frac{3}{2}, \frac{(\mu-a)^2}{2})}{\Gamma(\frac{k+1}{2})}\Big] \tag{4.54}$$

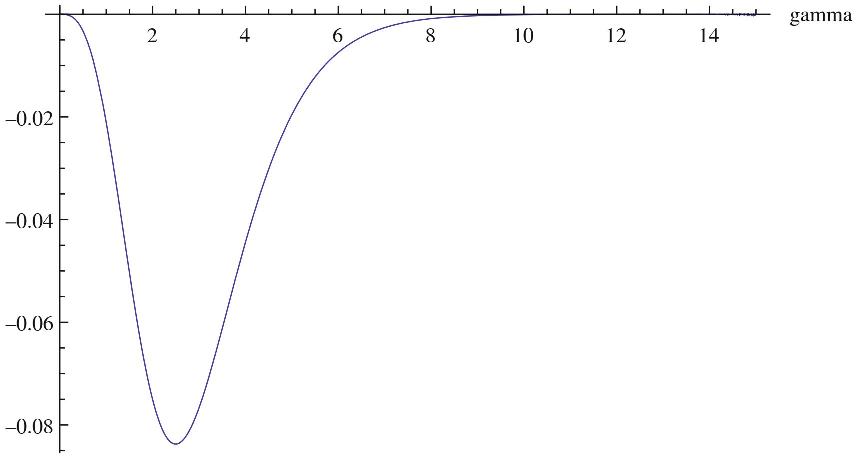

Fig. 4.5 Plot of (48) for a Scaled Double Exponential h

and, as $\gamma \to \infty$,

$$\int_0^\infty z^k e^{-az}\Big[\phi(\gamma - z) - \phi(\gamma + z)\Big]dz \sim e^{a^2/2} e^{-a\gamma} \gamma^k, \tag{4.55}$$

(in the sense that the ratio of the two sides converges to 1 as $\gamma \to \infty$)

Proof To obtain (4.54), write for any real number μ,

$$\int_0^\infty z^k e^{-az} \phi(\mu - z) dz = \frac{e^{-\mu^2/2}}{\sqrt{2\pi}} \int_0^\infty z^k e^{(\mu - a)z - z^2/2} dz, \tag{4.56}$$

and first, use the integration formula

$$\int_0^\infty z^k e^{-bz - z^2/2} dz = k! e^{b^2/4} D_{-k-1}(b) \tag{4.57}$$

(pp 360, Gradshteyn and Ryzhik (1980)) Next, use the functional identity

$$D_p(z) = 2^{p/2} e^{-z^2/4} \Big[\frac{\sqrt{\pi}}{\Gamma(\frac{1-p}{2})} M(-\frac{p}{2}, \frac{1}{2}, \frac{z^2}{2}) - \frac{\sqrt{2\pi} z}{\Gamma(-\frac{p}{2})} M(\frac{1-p}{2}, \frac{3}{2}, \frac{z^2}{2}) \Big] \tag{4.58}$$

(pp 1018, Gradshteyn and Ryzhik (1980))

Substituting (4.57) and (4.58) into (4.56), we get (4.54), on careful algebra.

For (4.55), we use the asymptotic order result

$$M(\alpha, \beta, z) \sim e^z z^{\alpha - \beta} \frac{\Gamma(\beta)}{\Gamma(\alpha)}, \quad z \to \infty \tag{4.59}$$

(see, for example, pp 255-259 in Olver (1997))

Use of (4.59) in (4.54) with $\mu = \mp\gamma$, and then subtraction, leads to the asymptotic order result that as $\gamma \to \infty$,

$$\int_0^\infty z^k e^{-az}\Big[\phi(\gamma - z) - \phi(\gamma + z)\Big]dz = \frac{k!e^{a^2/2}}{2^{k/2+1}}$$
$$\times\Big\{e^{-a\gamma}(\frac{(\gamma - a)^2}{2})^{k/2}\frac{\sqrt{\pi}}{\Gamma(\frac{k+1}{2})\Gamma(\frac{k+2}{2})}$$
$$+\sqrt{2}(\gamma - a)e^{-a\gamma}(\frac{(\gamma - a)^2}{2})^{k/2-1/2}\frac{\frac{1}{2}\sqrt{\pi}}{\Gamma(\frac{k+1}{2})\Gamma(\frac{k+2}{2})}\Big\}\times(1 + o(1))$$
$$+\Big\{\sqrt{2}(\gamma + a)e^{a\gamma}(\frac{(\gamma + a)^2}{2})^{k/2-1/2}\frac{\frac{1}{2}\sqrt{\pi}}{\Gamma(\frac{k+1}{2})\Gamma(\frac{k+2}{2})}$$
$$-e^{a\gamma}(\frac{(\gamma + a)^2}{2})^{k/2}\frac{\sqrt{\pi}}{\Gamma(\frac{k+1}{2})\Gamma(\frac{k+2}{2})}\Big\}\times(1 + o(1)) = \frac{k!e^{a^2/2}\sqrt{\pi}}{2^{k+1/2}\Gamma(\frac{k+1}{2})\Gamma(\frac{k+2}{2})}$$
$$\Big[e^{-a\gamma}\frac{(\gamma - a)^k}{\sqrt{2}} + e^{-a\gamma}\frac{(\gamma - a)^k}{\sqrt{2}} - e^{a\gamma}\frac{(\gamma + a)^k}{\sqrt{2}} + e^{a\gamma}\frac{(\gamma + a)^k}{\sqrt{2}}\Big]\times(1 + o(1))$$
$$= \frac{k!e^{a^2/2}\sqrt{\pi}}{2^k\Gamma(\frac{k+1}{2})\Gamma(\frac{k+2}{2})}e^{-a\gamma}(\gamma - a)^k \times(1 + o(1)) \tag{4.60}$$

In (4.60), by using the *Gamma duplication formula*

$$\Gamma(z + 1/2) = \sqrt{\pi}2^{1-2z}\frac{\Gamma(2z)}{\Gamma(z)},$$

we get

$$\int_0^\infty z^k e^{-az}\Big[\phi(\gamma - z) - \phi(\gamma + z)\Big]dz$$

$$= e^{a^2/2}e^{-a\gamma}(\gamma - a)^k \times(1 + o(1)) = e^{a^2/2}e^{-a\gamma}\gamma^k \times(1 + o(1)), \tag{4.61}$$

as claimed in (4.55).

Remark The real use of Proposition 1.2.1 is that by using (4.54), we get an *exact analytical formula* for $\triangle_n$ in terms of the confluent hypergeometric function. If all we care for is the asymptotic order result (4.55), then we may obtain it in a less complex way. Indeed, by using techniques in Feller (1971, pp 442-446) and Theorem 3.1 in Berman (1992), we can conclude that $\int_0^\infty z^k e^{-az}\phi(\gamma - z)dz = \gamma^k e^{-a\gamma}\int_{-\infty}^\infty e^{(a-\frac{k}{\gamma})t}\phi(t)dt \times(1 + o(1))$, and (4.55) follows from this.

Corollary 1.2.8 *Consider generalized Hodges estimates of the form (4.2) with* $a_n \equiv 0$*. Let* $h(\theta) = \frac{1}{2\sigma} e^{-|\theta|/\sigma}$ *and* $\pi(\theta) = \pi_n(\theta) = \sqrt{n}h(\theta\sqrt{n})$*. Then,*

$$\triangle_n = \frac{2\sigma^2 - 1}{n}(1 + o(1)), \quad \text{if } \mathrm{Var}_h(\theta) \neq 1 \Leftrightarrow 2\sigma^2 - 1 \neq 0, \tag{4.62}$$

and,

$$\triangle_n = -e\frac{\gamma_n^2 e^{-\gamma_n\sqrt{2}}}{n}(1 + o(1)), \quad \text{if } \mathrm{Var}_h(\theta) = 1 \Leftrightarrow 2\sigma^2 - 1 = 0 \tag{4.63}$$

This corollary follows by using the formula in (4.53) and the result in (4.55). *Notice that the critical issue in determining the rate of convergence of* $\triangle_n$ *to zero is whether or not* $Var_h(\theta) = 1$.

As indicated previously, we can generalize the result on the asymptotic order of the Bayes risk difference $\triangle_n$ to more general priors. The important thing to understand is that Theorem 1.2.9 (more precisely, (4.48)) gives a representation of $\triangle_n$ in a convolution form. Hence, we need to appeal to results on orders of the tails of convolutions. The right structure needed for such results is that of *regular variation*. We state two known results to be used in the proof of Theorem 1.2.10 as lemmas.

Lemma 1.2.1 *(Landau's Theorem) Let* U *be a nonnegative absolutely continuous function with derivative* u*. Suppose* U *is of regular variation of exponent* $\rho \neq 0$ *at* ∞*, and that* u *is ultimately monotone and has a finite number of sign-changes. Then* u *is of regular variation of exponent* $\rho - 1$ *at* ∞*.*

Lemma 1.2.2 (Berman (1992)) *Suppose* $p(z)$ *is a probability density function on the real line, and* $q(z)$ *is ultimately nonnegative, and that* $w(z) = -\frac{d}{dz}\log q(z)$, $v(z) = -\frac{d}{dz}\log p(z)$ *exist and are functions of regular oscillation, i.e., if* $z, z' \to \infty$, $\frac{z}{z'} \to 1$*, then* $\frac{f(z)}{f(z')} \to 1$ *if* $f = w$ *or* v*. If, moreover,* $\liminf_{z\to\infty} \frac{d}{dz}\log q(z) > \liminf_{z\to\infty} \frac{d}{dz}\log p(z)$*, then,* $\int_{-\infty}^{\infty} q(z)p(\gamma - z)dz = q(\gamma)\int_{-\infty}^{\infty} e^{-zw(\gamma)}p(z)dz\,(1 + o(1))$*, as* $\gamma \to \infty$*.*

We now present the following general result.

Theorem 1.2.10 *Suppose assumptions (1)-(7) hold true and if* $-\log q(z)$ *is a function of regular variation of some exponent* $\rho \neq 0$ *at* $z = \infty$*. Then,*

$$\triangle_n = \frac{2q(\gamma_n)e^{\frac{1}{2}\left[w(\gamma_n)\right]^2}}{n}(1 + o(1)), \quad \text{as } n \to \infty. \tag{4.64}$$

Proof By assumption (6), $w(z)$ is ultimately monotone, and by assumption (5), $w(z)$ is ultimately positive. By hypothesis, $-\log q(z)$ is a function of regular variation. Therefore, all the conditions of *Landau's theorem* (Lemma 1.2.1) are satisfied, and hence it follows that $w(z)$ is also a function of regular variation at ∞. This will imply, by well known local uniformity of convergence for functions of regular variation that if $z, z' \to \infty$, and $\frac{z}{z'} \to 1$, then $\frac{w(z)}{w(z')} \to 1$. By assumption (7), we have

$\limsup_{z\to\infty} w(z) < \infty = \limsup_{z\to\infty} \frac{d}{dz} - \log\phi(z)$. Hence, we can now appeal to Lemma 1.2.2 to coclude that

$$\int_{-\infty}^{\infty} q(z)\phi(\gamma_n - z)dz = q(\gamma_n)\int_{-\infty}^{\infty} e^{-zw(\gamma_n)}\phi(z)dz\,(1+o(1))$$
$$= q(\gamma_n)e^{\frac{1}{2}\left[w(\gamma_n)\right]^2}(1+o(1))$$

(by completing the squares), and hence, by (4.48),

$$\triangle_n = \frac{2}{n}\int_{-\infty}^{\infty} q(z)\Big[\phi(\gamma_n - z) - \phi(\gamma_n + z)\Big]dz$$
$$= \frac{2}{n}\int_{-\infty}^{\infty} q(z)\phi(\gamma_n - z)dz\,(1+o(1))$$
$$= \frac{2q(\gamma_n)e^{\frac{1}{2}\left[w(\gamma_n)\right]^2}}{n}(1+o(1)), \tag{4.65}$$

as claimed.

Acknowledgements It is a pleasure to thank Jyotishka Datta for his very gracious help with formatting the paper.

References

Banerjee M (2006) Superefficiency, contiguity, LAN University Michigan Lecture Notes
Bickel PJ (1983) Minimax estimation of the mean of a normal distribution subject to doing well at a point. In: Recent Advances in Statistics, A Festschrift for H. Chernoff, M. Rizvi, J. Rustagi and D. Siegmund eds., 511-528, Academic Press, NY
DasGupta A (2008) Asymptotic theory of statistics and probability. Springer, New York
DasGupta A (2011) Probability for statistics and machine learning: fundamentals and advanced topics. Springer, New York
Donoho D, Johnstone IM (1994) Ideal spatial adaptation by wavelet shrinkage. Biom 81: 425-455
Hájek J (1970) A characterization of limiting distributions of regular estimates, Z. Wahr verw Geb 14: 323-330
Jeganathan P (1983) Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. Sankhyā Ser A 45:66-87
Johnstone IM (2012) Function estimation and Gaussian sequence models, Cambridge University Press, Forthcomiing
Le Cam L (1953) On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. University California Publications 1: 277-330
Le Cam L (1973) Sur les contraintes imposees par les passages a la limite usuels en statistique, Proc. 39th session International Statistical Institute, XLV, 169-177
Lehmann EL, Romano J (2005) Testing statistical hypotheses, 3rd edn. Springer, New York
van der Vaart A (1997) Superefficiency, in Festschrift for Lucien Le Cam, D. Pollard, E. Torgersen, G. Yang, Eds. Springer, New York
van der Vaart A (1998) Asymptotic statistics. Cambridge University Press, Cambridge
Wasserman L (2005) All of Nonparametric Statistics. Springer, New York
Wellner J (2012) University Washington Lecture Notes