IMS Collections
Borrowing Strength: Theory Powering Applications –
A Festschrift for Lawrence D. Brown
Vol. 6 (2010) 1–21

DOI: 10.1214/10-IMSCOLL601

False vs. missed discoveries, Gaussian decision theory, and the Donsker-Varadhan principle

Anirban DasGupta

Purdue University, USA

Abstract: The article touches on two themes in which Larry Brown has been interested, namely foundations, and mathematical analysis of Bayesian decision theory.

In the section on foundations, a new formulation of the problem of testing is given, and is theoretically explored. The formulation strikes a balance between false discoveries and missed discoveries. Basic higher order asymptotic theory for the α level that this formulation would imply is then worked out. Accuracy is investigated and examples are given.

In the section on Bayesian decision theory, first, the Brown identities are connected to a set of inequalities of elliptic boundary value problems. It is shown by four specific results that sometimes a result in that field can lead to new results in Bayesian decision theory, and sometimes a result in decision theory can give surprisingly useful information about a problem in that field. For instance, Brown identities can provide amazingly good estimates of best constants in the Nash inequalities over Sobolev spaces.

The article ends with a result on the modern theory of high dimensional Gaussian mean estimation. By means of a triangulation of the Brown identity, the Rayleigh-Ritz variational formula of boundary value problems, and the famous Donsker-Varadhan result connecting the Rayleigh-Ritz formula to the absorption time of a Brownian motion into the boundary of a smooth bounded open domain, we show that the minimax risk of estimating the Gaussian mean can be approximated by chasing a Brownian motion to the boundary of the parameter space. This link should be tested by simulation.

1. Foreword

Although my failing eyesight has limited my normal activities, when the editors of this volume kindly invited me to write an article, I knew that it was an honor that I could not pass. I first corresponded with Larry Brown around 1982, when I was a PhD student at the Indian Statistical Institute in Calcutta. In the 30 years that have followed, I have learned from and have been inspired by Larry Brown's signature humility, and his work in numerous areas, including decision theory and mathematical analysis of difficult Bayesian problems, foundations, asymptotics, and theoretical validation of methodologies. This article touches on two of these themes, which form an extremely small corner of Larry Brown's work. These are among the main topics on which I have myself corresponded with him, or have worked with him. It is my greatest pleasure to present these few calculations in honor of Professor Brown.

Here is a summary of what is in this article and where the ideas may or may not go. Section 2 gives a formulation of the problem of testing of hypotheses that attempts to strike a balance between false discoveries and missed discoveries. The traditional Neyman-Pearson formulation and the current work in the popular area of multiple testing place an asymmetric emphasis on false discoveries. In the symmetric formulation of Section 2 of this article, it is argued, via appropriate asymptotic expansions, that the traditional α levels are quite possibly too large in some problems and that the level α should go down to zero at a suitable rate, depending on the particular problem, and depending on which particular alternatives are practically significant in that problem. Mathematically, this requires the specification of a density g on the alternative. The asymptotic expansions show the rate at which α should go to zero, and this rate depends on the smoothness of g at the boundary. The exact asymptotic expansions should not be taken literally. Asymptotic expansions often produce ugly coefficients, and here too they do. The expansions should be separated from the formulation and the outcome that α should go to zero at some suitable rate, which is not universal, but depends on the problem. Coming from other angles, this has been argued in some of the Bayesian literature, for example [3]; also see [16]. Section 2.4 offers some thoughts on where this general approach might next go.

The final section, Section 3, goes back to where many of us, including Larry Brown himself, started. It is the mutual connection of various Bayesian techniques, tools, and theorems and risk based decision theoretic properties, such as minimaxity. Of course, admissibility is also mostly about Bayes, but it is not mentioned here. Diaconis and Holmes [21] is a very good place to look for new possibilities and good connections. The main thesis of Section 3 is that there are aspects of the Brown identity [8] that have unexplored connections to a very well developed area of analysis, namely inequalities of elliptic boundary value problems. Typically, such an inequality is of the form $\|f^{(k)}\|_{q,G} \leq K(\|f\|^{\alpha}_{p,G})(\|f^{(n)}\|^{\beta}_{r,G})$ for some suitably large class of Sobolev type spaces of functions f on some subset G of an Euclidean space, and for flexible p, q, r, k, n, with the restriction that $0 \leq k < n$. The inequality is supposed to hold with a universal constant $K = K(G, k, n, p, q, r)$. When the smallest possible universal constant has been found, one calls it the *best constant* for that inequality. Usually, best constants are very hard to find, and appear to be known analytically only in isolated cases.

The link of all these to Bayesian decision theory is through the Brown identity for Bayes risks. The Brown identity showed that the Bayes risk in a multivariate normal mean problem is related to the Fisher information of the marginal, which is essentially the square of the L_2 norm of the gradient of the square root of the marginal density. Once this identity is at hand, the inequalities of boundary value problems, such as the ones mentioned above, lead to a large collection of connections, and these connections go both ways. Section 3 shows that sometimes a known result in boundary value problems can lead to a result in decision theory, with the Brown identity being the link; and sometimes, quantities in Bayesian decision theory lead to information about a question in boundary value problems, for example, how small can the best constant in an inequality possibly be? Four such specific connections are laid out in Section 3, by using a generalized Heisenberg's uncertainty inequality, an inequality of the HELP type, the Nash inequality, and an inequality of Landau (although commonly ascribed to Kolmogorov and Landau). The number of such connections would be essentially unlimited, with the Brown identity always being the link. Diaconis and Saloff-Coste [22] have used Nash inequalities to study the speed of convergence of some finite Markov chains to stationarity. So, in some sense, there is a history of the Nash inequalities being linked to problems in mathematical statistics. Hopefully, someone will pursue these connections to bring further insights and other concrete results to the statistical community.

There is one aspect of Section 3 that would be of interest in the modern theory of high dimensional Gaussian mean estimation. One often assumes in such problems that the mean vector belongs to a suitably small convex compact domain Ω in some $\mathcal{R}^d$. The minimax risk over this domain is a quantity of interest. In some sense, the area seriously started with the work of David Donoho and Iain Johnstone [24]. Wasserman [38] gives an immensely readable review of the work up to that time. We show that the minimax risk should be theoretically and also numerically approximable by triangulating three things; namely, the Brown identity, and the Rayleigh-Ritz variational principle of boundary value problems, and the probabilistic connection of the Rayleigh-Ritz principle itself to the absorption time of a Brownian motion into the boundary of our domain. Theorem 7 gives an inequality. We conjecture that asymptotically, the inequality should be essentially an equality. There should be such a theorem.

2. A new formulation of testing

As much as the Neyman-Pearson theory of testing has had a tremendous impact on how we do and think about statistics, some apparent difficulties with its formulation have been widely noticed. The aspect that we focus on is the asymmetric nature of the formulation of the Neyman-Pearson theory vis-a-vis the null and the alternative hypotheses. In the Neyman-Pearson formulation, the null hypothesis has a special role. We first protect against false rejection of the null, and then try to do something about the false rejection of the alternative. In the terminology of modern multiple testing problems, in the Neyman-Pearson formulation we would consider preventing a false discovery our first goal, and preventing a missed discovery a secondary goal. One could potentially argue that there should be some balance between the two errors in the formulation of the problem.

At an even more classic level, in an elementary class, an instructor would test one hypothesis A against another, say B, usually with some simple minded data, and pick A as the null and end up rejecting A at a 5% level. Occasionally, an inquisitive student would later ask why was that particular hypothesis chosen as the null. It will often turn out, causing some embarrassment, that if the choices of the null and the alternative were reversed, then with that same data, and once again at the 5% level, the hypothesis B would now be rejected in favor of A. As a matter of fact, if the *true* parameter value is somewhere in between A and B, then this anomaly will occur with a very large probability, because under the true parameter value, the acceptance probability of either hypothesis, A or B, would be a large deviation type probability, and so typically it will be exponentially small.

Lehmann [34], TSH, 2nd Edition, comments tangentially about a symmetric formulation of testing problems, while pointing out possible formulations in which the well known discrepancies between Bayesian posterior probabilities and classical P-values will disappear ([3, 36]).

We present a symmetric formulation of the testing problem in this section, and then work out some asymptotic theory about it. The main point where this formulation differs from the traditional one is that the α level is not prespecified. It is determined from the symmetry of the formulation, and it will depend on n, the sample size. Furthermore, as n increases, the implied α level will decrease, and quite typically will converge to zero. The asymptotic calculations that we do here are directed towards understanding this main point of divergence from the traditional formulation.

We consider only the one dimensional case in this first paper. The multidimensional case should be considered, but the asymptotic expansions will be even harder. The case of one dimensional continuous exponential families should be similar to the developments here. The asymptotic approximations derived here tell the reader that the level α ought to go down to zero at a rate which, to the first order, is $(\frac{\log n}{n})^\alpha$, for a suitable explicit $\alpha > 0$. It has been sporadically commented on in the literature that α should depend on n in Neyman-Pearson testing. Apparently, Neyman himself thought so (private correspondence with the late Prem Puri).

2.1. An illustrative example

We start with an illustrative example.

Example 1 (Illustrative Example). Suppose on the basis of n iid observations $X_1, X_2, \ldots, X_n \sim N(\theta, \sigma^2)$, we wish to test $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$; σ^2 is assumed to be known. We may take θ_0 to be zero. The Neyman-Perason tests reject H_0 for large values of $\overline{X}$, and if we reject when $\overline{X} > \frac{\sigma z_\alpha}{\sqrt{n}}$, where z_α is the standard normal percentile $\Phi^{-1}(1-\alpha)$, then the test has type I error probability α, and type II error probability

$$\beta = \beta(\alpha, n, \theta) = \Phi\left(z_\alpha - \frac{\theta\sqrt{n}}{\sigma}\right), \quad \theta > 0.$$

For example, if $\alpha = .05$, $n = 20$, and $\sigma = 1$, then the type II error probability at $\theta = .5$ is $\Phi(1.645 - .5 \times \sqrt{20}) = \Phi(-.59) = .2776$. On the other hand, at any fixed alternative, eventually, i.e., as $n \to \infty$, the type II error probability $\beta(\alpha, n, \theta)$ will converge to zero, while by choice the type I error probability remains fixed at the specified α.

We want to look at symmetric formulations of the following kind. Consider *Pitman alternatives* $\theta = \frac{\Delta}{\sqrt{n}}$ first. Then the type II error probability is $\beta(\alpha, \Delta) = \Phi(z_\alpha - \frac{\Delta}{\sigma})$. Let $g : \mathcal{R}^+ \to \mathcal{R}^+$ be a density function, and let $\beta(\alpha, g) = \int_0^\infty \Phi(z_\alpha - \frac{\Delta}{\sigma})g(\Delta)d\Delta$. We seek a level $\alpha = \alpha(g)$ such that $\beta(\alpha, g) = \alpha$. Note that use of Pitman alternatives eliminates the role of n; the final α level depends only on the user's g. If we were to use $g(\Delta) = e^{-\Delta}$, then calculation reduces the level α to be the unique root of

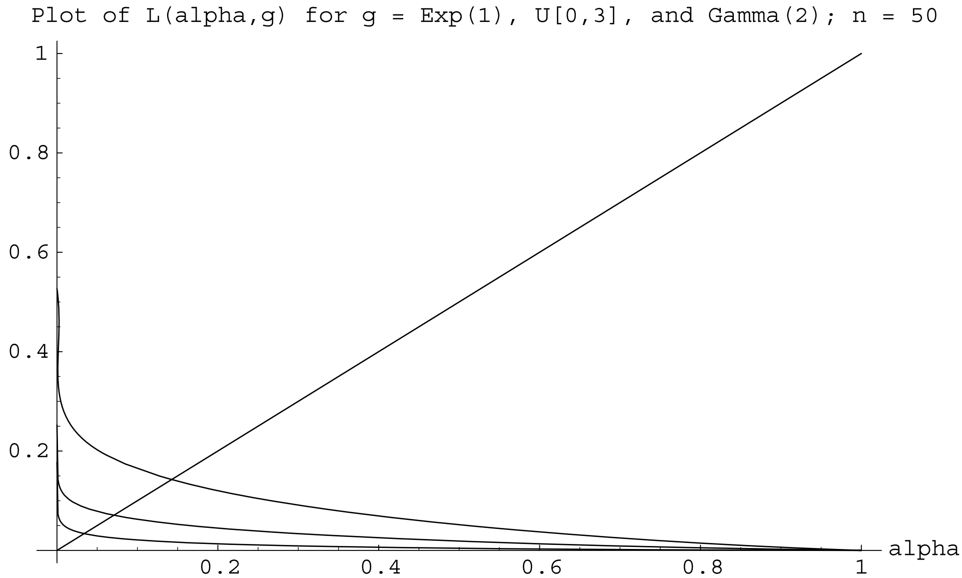
$$\begin{aligned} \alpha &= 1 - \alpha - \phi(z_\alpha)\frac{1 - \Phi(z_\alpha - \sigma)}{\phi(z_\alpha - \sigma)} \\ \Leftrightarrow \alpha &= \frac{1}{2} - \phi(z_\alpha)\frac{1 - \Phi(z_\alpha - \sigma)}{2\phi(z_\alpha - \sigma)}, \end{aligned} \tag{1}$$

where ϕ is the standard normal density.

For example, if $\sigma = 1$, then the unique root is $\alpha = .348835$. The α level works out to a large number with $g(\Delta) = e^{-\Delta}$ because g places a lot of emphasis near $\Delta = 0$. For choices of g that place less emphasis near zero, the α level will work out to a smaller number. It will always depend on the chosen g.

Let us see what happens if we consider fixed alternatives. If we take *fixed alternatives*, then the role of n will not be eliminated. Indeed, if we seek a level α such that $\alpha = L(\alpha, g) = \int_0^\infty \beta(\alpha, n, \theta)g(\theta)\, d\theta$, then on a small amount of algebra, the equation reduces to

$$\alpha = 1 - \frac{1}{\sqrt{n}}\int_0^\infty \left[1 - \Phi\left(z_\alpha - \frac{\theta}{\sigma}\right)\right] g\left(\frac{\theta}{\sqrt{n}}\right) d\theta.$$



This representation is going to be useful in seeing why the behavior of g near $\theta = 0$ is critical in determining the value (and more importantly, the order of magnitude) of the required level α.

Consider now, as an example, $g(\theta) = \theta e^{-\theta}$. This choice of g places less emphasis near the null value $\theta = 0$ than does $g(\theta) = e^{-\theta}$. The α levels, which are the roots of the equation $\alpha = \int_0^\infty \beta(\alpha, n, \theta) g(\theta)\, d\theta$, work out as follows for some selected values of n.

n	10	20	30	50	100	200	500
α	.088	.060	.047	.034	.021	.013	.0065

The values tell us that according to the symmetric formulation that we have laid out, with the fixed alternative approach, the choice of the α level should be smaller for larger n, and that the conventional 5% level is somewhat too large for large sample sizes. However, the example also suggests that the conventional 5% level is about right for sample sizes in the range $n \leq 50$, depending on g. The choice of g will be governed by the problem, namely which alternatives are practically important in that problem. A plot of $L(\alpha, g)$ for three choices of g illustrates the role of g in the α level that solves $\alpha = L(\alpha, g)$.

One may interpret the finding of this example and the plot to be suggestive of both Neyman's uncanny wisdom when he suggested the 5% convention, and at the same time, the criticism of the Bayesians that the 5% level is sometimes too low. This example sets the tone for our theorem below in the next section.

2.2. An asymptotic expansion

Theorem 1. *Let* $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, *where* $\sigma^2 > 0$ *is assumed to be known. For given* $\alpha, 0 < \alpha < 1$, *and* n, *consider the nonrandomized test that rejects* $H_0 : \theta = 0$ *in favor of* $H_1 : \theta > 0$ *if* $\overline{X} > \frac{\sigma z_\alpha}{\sqrt{n}}$. *Let* $\beta(\alpha, n, \theta), \theta > 0$ *be its type II* *error probability function, and g a probability density on $\mathcal{R}^+$. Assume the following conditions on g:*

1. $g(0) < \infty$.
2. *There is a nonnegative finite integer m such that $g^{(j)}(0) = 0$ for $j = 1, \ldots, m$, and $g^{(m+1)}(0+) \neq 0$.*
3. *g is $(m+2)$ times continuously differentiable on $(0, \infty)$ and $g^{(m+2)}$ is absolutely uniformly bounded on $(0, \infty)$.*

(A) *Let $\Gamma(c, x)$ denote the incomplete Gamma function $\Gamma(c, x) = \int_x^\infty e^{-t} t^{c-1} dt$, c, $x > 0$. Then $\int_0^\infty \beta(\alpha, n, \theta) g(\theta) d\theta$ admits the asymptotic expansion*

$$
\begin{aligned}
&\int_0^\infty \beta(\alpha, n, \theta) g(\theta) d\theta \\
&= \frac{\sigma g(0)}{\sqrt{n}} [z_\alpha(1-\alpha) + \phi(z_\alpha)] + \frac{\sigma^{m+2} g^{(m+1)}(0+)}{n^{m/2+1}(m+1)!} \\
&\times \Bigg[z_\alpha^{m+2}(1-\alpha)\frac{1-m}{2} + z_\alpha^{m+1}\phi(z_\alpha) + z_\alpha^{m+2}(1-\alpha)\frac{m(m+1)}{2(m+2)} \\
&+ \frac{1}{\sqrt{2\pi}} \sum_{j=0}^{m+1} \binom{m+1}{j} 2^{\frac{j}{2}} z_\alpha^{m+1-j} \frac{\Gamma(1+\frac{j}{2})}{j+1} \\
&+ \frac{1}{\sqrt{2\pi}} \sum_{j=1}^{m+1} \binom{m+1}{j} (-1)^j 2^{\frac{j}{2}} z_\alpha^{m+1-j} \frac{\Gamma(1+\frac{j}{2}, \frac{z_\alpha^2}{2})}{1+j} \Bigg] \\
&+ O(n^{-(m+3)/2}).
\end{aligned} \tag{2}
$$

(B) *In particular,*

(a) *If $g(0) > 0$, then*

$$
\begin{aligned}
&\int_0^\infty \beta(\alpha, n, \theta) g(\theta)\, d\theta \\
&\quad = \frac{\sigma g(0)}{\sqrt{n}} [z_\alpha(1-\alpha) + \phi(z_\alpha)] + O(n^{-1}).
\end{aligned} \tag{3}
$$

(b) *If $g(0) = 0$, $g'(0+) > 0$, then*

$$
\begin{aligned}
&\int_0^\infty \beta(\alpha, n, \theta) g(\theta)\, d\theta \\
&= \frac{\sigma^2 g'(0+)}{n} \times \left[\frac{z_\alpha^2(1-\alpha)}{2} + \frac{z_\alpha \phi(z_\alpha)}{2} + \frac{z_\alpha}{\sqrt{2\pi}} + \frac{1}{4} - \frac{\alpha}{2} \right] + O(n^{-3/2}).
\end{aligned} \tag{4}
$$

Sketch of the Proof. The derivation of the expansion follows the usual technique of a pointwise Taylor expansion of the integrand, while ensuring that the error remains of the claimed order after integration.

Write

$$
\begin{aligned}
&\int_0^\infty \beta(\alpha,n,\theta)g(\theta)d\theta \\
&\quad = \int_0^\infty \left[1-\Phi\left(\frac{\theta\sqrt{n}}{\sigma}-z_\alpha\right)\right]g(\theta)\,d\theta \\
&\quad = \frac{\sigma}{\sqrt{n}}\int_{-z_\alpha}^\infty [1-\Phi(x)]g\left(\frac{\sigma(x+z_\alpha)}{\sqrt{n}}\right)dx \\
&\quad = \frac{\sigma}{\sqrt{n}}\int_{-z_\alpha}^\infty [1-\Phi(x)]\left[\sum_{j=0}^{m+1}\frac{(\frac{\sigma(x+z_\alpha)}{\sqrt{n}})^j}{j!}g^{(j)}(0)+\frac{(\frac{\sigma(x+z_\alpha)}{\sqrt{n}})^{m+2}}{(m+2)!}g^{(m+2)}(x^*)\right]dx
\end{aligned}
$$

(because the remainder term in the Taylor expansion has such a representation by the continuous differentiability of $g^{(m+2)}$)

$$
\begin{aligned}
&= \frac{\sigma}{\sqrt{n}}g(0)\int_{-z_\alpha}^\infty [1-\Phi(x)]\,dx \\
&\quad + \frac{\sigma}{\sqrt{n}}g^{(m+1)}(0+)\int_{-z_\alpha}^\infty [1-\Phi(x)]\frac{(\frac{\sigma(x+z_\alpha)}{\sqrt{n}})^{m+1}}{(m+1)!}\,dx \\
&\quad + O(n^{-(m+3)/2})
\end{aligned}
$$

(since $|g^{(m+2)}|$ is assumed to be uniformly bounded)

$$
\begin{aligned}
&= \frac{\sigma}{\sqrt{n}}g(0)\int_{-z_\alpha}^\infty [1-\Phi(x)]dx \\
&\quad + \frac{\sigma^{m+2}}{n^{\frac{m}{2}+1}(m+1)!}g^{(m+1)}(0+)\left[\sum_{j=0}^{m+1}\binom{m+1}{j}z_\alpha^{m+1-j}\int_{-z_\alpha}^\infty x^j[1-\Phi(x)]dx\right] \\
&\quad + O(n^{-(m+3)/2}).
\end{aligned} \tag{5}
$$

Now use the following integration formulas, which we do not derive here:

$$\int_{-z_\alpha}^\infty [1-\Phi(x)]\,dx = z_\alpha(1-\alpha)+\phi(z_\alpha);$$

For $j \geq 1$,

$$
\begin{aligned}
\int_{-z_\alpha}^\infty x^j[1-\Phi(x)]\,dx &= (-1)^j\frac{z_\alpha^{j+1}}{j+1}(1-\alpha)+\frac{1}{\sqrt{2\pi}}\frac{2^{\frac{j}{2}}\Gamma(1+\frac{j}{2})}{j+1} \\
&\quad + \frac{1}{\sqrt{2\pi}}\frac{(-1)^j 2^{\frac{j}{2}}\Gamma(1+\frac{j}{2},\frac{z_\alpha^2}{2})}{j+1}.
\end{aligned}
$$

Then, use the following special values for the terms below:

$$\sum_{j=0}^{m+1}\binom{m+1}{j}\frac{(-1)^j}{j+1}=\frac{1}{m+2}; \quad \text{and for } j=0, \quad \Gamma\left(1+\frac{j}{2},\frac{z_\alpha^2}{2}\right)=e^{-\frac{z_\alpha^2}{2}}.$$

On substituting all of these above expressions into the expression (5), the asymptotic expansion of the theorem follows on collecting the terms and simplification. This additional detail is omitted. □

2.3. Asymptotic approximation to the level α

The asymptotic expansion for $\int_0^\infty \beta(\alpha,n,\theta)g(\theta)\,d\theta$ is useful for finding an asymptotic approximation for the α level that solves the equation $\alpha = \int_0^\infty \beta(\alpha,n,\theta)g(\theta)d\theta$. We provide such an approximation in this section. The approximation is based on inverting an asymptotic expansion; the technique has been used before, e.g., in [28] for writing Cornish-Fisher expansions.

A word of caution and apology is in order. The asymptotic approximation to the α level does not come with an order for the error of the approximation. We have not proved any such rate. Therefore, although the approximation given below is quite accurate (as we will see), we have refrained from stating it as a theorem.

We first explain the method for deriving an asymptotic approximation to the α level in the case $g(0) > 0$. For reference below, we denote $\sigma g(0) = c$ and $\sigma g(0)\sqrt{2\pi} = a$. In this case, Theorem 1 says

$$\int_0^\infty \beta(\alpha,n,\theta)g(\theta)\,d\theta \approx \frac{c}{\sqrt{n}}[z_\alpha(1-\alpha)+\phi(z_\alpha)].$$

Setting $\alpha = \frac{c}{\sqrt{n}}[z_\alpha(1-\alpha)+\phi(z_\alpha)]$, and by transposition we get

$$\begin{aligned}
\alpha &\approx \frac{z_\alpha+\phi(z_\alpha)}{\frac{\sqrt{n}}{c}+z_\alpha}\\
\Rightarrow 1-\Phi(z_\alpha) &\approx \frac{z_\alpha+\phi(z_\alpha)}{\frac{\sqrt{n}}{c}+z_\alpha}\\
\Rightarrow \frac{\phi(z_\alpha)}{z_\alpha} &\approx \frac{z_\alpha+\phi(z_\alpha)}{\frac{\sqrt{n}}{c}+z_\alpha}\\
\Rightarrow z_\alpha^2 &\approx \frac{\sqrt{n}}{c}\phi(z_\alpha)
\end{aligned}$$

(on cancelling the $z_\alpha\phi(z_\alpha)$ term from two sides of the formal equation)

$$\Rightarrow z_\alpha^2 e^{\frac{z_\alpha^2}{2}} = \frac{\sqrt{n}}{c\sqrt{2\pi}} = \frac{\sqrt{n}}{a}.$$

A first approximation to the root of this equation is $z_\alpha^2 = \log n - 2\log\log n - 2\log a$, which, by plugging back into the basic equation $\alpha = \frac{\sigma g(0)}{\sqrt{n}}[z_\alpha(1-\alpha)+\phi(z_\alpha)]$, results after additional algebra, and collection of terms, in the first approximation for our required α level when $g(0) > 0$, namely

$$\boxed{(6)\qquad \alpha \approx \sigma g(0)\sqrt{\frac{\log n}{n}}\left[1+\sqrt{\frac{1}{2\pi\log n}}-\frac{\log\log n+\log a}{\log n}\right]}$$

where we recall that $a = \sigma g(0)\sqrt{2\pi}$.

Next, consider the case when $g(0) = 0$ and $g'(0+) > 0$. We denote $\sigma^2 g'(0+) = d$, and $d\sqrt{\frac{\pi}{2}} = b$. From Theorem 1,

$$\int_0^\infty \beta(\alpha,n,\theta)g(\theta)\,d\theta \approx \frac{d}{n}\left[\frac{z_\alpha^2(1-\alpha)}{2}+\frac{z_\alpha\phi(z_\alpha)}{2}+\frac{z_\alpha}{\sqrt{2\pi}}+\frac{1}{4}-\frac{\alpha}{2}\right].$$

Once again, setting

$$\alpha = \frac{d}{n}\left[\frac{z_\alpha^2(1-\alpha)}{2} + \frac{z_\alpha\phi(z_\alpha)}{2} + \frac{z_\alpha}{\sqrt{2\pi}} + \frac{1}{4} - \frac{\alpha}{2}\right],$$

after some algebra, we get the approximation

$$z_\alpha^3 + \sqrt{\frac{2}{\pi}}z_\alpha^2 + \frac{z_\alpha}{2} \approx \left[\frac{2}{d}n + 1\right]\phi(z_\alpha).$$

A first approximation to the root of this equation is $z_\alpha^2 = 2\log n - 3\log\log n - \log(8b^2)$. Plugging this back into the defining equation $\alpha = \frac{d}{n}[\frac{z_\alpha^2(1-\alpha)}{2} + \frac{z_\alpha\phi(z_\alpha)}{2} + \frac{z_\alpha}{\sqrt{2\pi}} + \frac{1}{4} - \frac{\alpha}{2}]$, further calculations give the first approximation for α when $g(0) = 0, g'(0+) > 0$, as

$$\boxed{(7) \qquad \alpha \approx \frac{\sigma^2 g'(0+)\log n}{n}\left[1 + \frac{1}{\sqrt{\pi\log n}} - \frac{\frac{3}{2}\log\log n + \frac{1}{2}\log(8b^2) - \frac{1}{4}}{\log n}\right]}$$

where we recall that $b = \sigma^2 g'(0+)\sqrt{\frac{\pi}{2}}$.

Example 2 (Accuracy of the approximation of α). We investigate the accuracy of our two theoretical approximations derived above for the α level that makes $\int_0^\infty \beta(\alpha,n,\theta)g(\theta)\,d\theta = \alpha$. The exact value is found by programming in *Mathematica*; the approximations are found by using the two approximations derived immediately above in (6) and (7). The choices of g are $g(\theta) = e^{-\theta}$ and $g(\theta) = \theta e^{-\theta}$. The second choice places smaller emphasis for alternatives close to the null, and in that case the α level works out to numbers smaller than traditional levels, such as 5%, even for moderate n. The approximation to α is very accurate for $g(\theta) = \theta e^{-\theta}$, and is acceptable for $g(\theta) = e^{-\theta}$.

n	$g(\theta) = e^{-\theta}$ (Approximate in parentheses)	$g(\theta) = \theta e^{-\theta}$ (Approximate in parentheses)
50	.143(.173)	.0339(.0393)
100	.115(.141)	.0212(.0251)
200	.091(.111)	.0130(.0154)
500	.066(.080)	.0065(.0077)
1000	.051(.061)	.0038(.0045)
2500	.036(.043)	.0018(.0021)

We now mention the case of general m. Recall that m is the first integer such that $g^{(m+1)}(0+) \neq 0$. Denote

$$\frac{\sigma^{m+2}g^{(m+1)}(0+)}{(m+1)!} = c_m; \quad \frac{c_m\sqrt{2\pi}}{m+2} = a_m; \quad a_m^2(m+2)^{m+3} = \gamma_m.$$

Then, assuming also that $g(0) = 0$, by using Theorem 1, a formal first approximation to $\int_0^\infty \beta(\alpha,n,\theta)g(\theta)\,d\theta$ is

$$\int_0^\infty \beta(\alpha,n,\theta)g(\theta)\,d\theta$$
$$\approx \frac{c_m}{n^{m/2+1}}\left[\frac{z_\alpha^{m+2}(1-\alpha)}{m+2} + \frac{1-m}{2}z_\alpha^{m+1}\phi(z_\alpha) + \frac{z_\alpha^{m+1}}{\sqrt{2\pi}} + \frac{m+1}{4}z_\alpha^m - \frac{m+1}{2}z_\alpha^m\right].$$

This gives, on rearrangement of terms,

$$\alpha \approx \frac{\frac{z_\alpha^{m+2}}{m+2} + \frac{z_\alpha^{m+1}}{\sqrt{2\pi}} + \frac{m+1}{4} z_\alpha^m + \frac{1-m}{2} z_\alpha^{m+1}\phi(z_\alpha)}{\frac{n^{m/2+1}}{c_m} + \frac{z_\alpha^{m+2}}{m+2} + \frac{m+1}{2} z_\alpha^m}. \tag{8}$$

This gives a first approximation for z_α^2:

$$z_\alpha^2 \approx (m+2)\log n - (m+3)\log\log n - \log \gamma_m.$$

Now we plug this back into our basic equation. We expand the numerator and the denominator separately, combine them, and collect terms. Then we obtain an approximation to the required α level as

$$\boxed{\begin{aligned}\alpha \approx\ & \frac{\sigma^{m+2}(m+2)^{\frac{m}{2}} g^{(m+1)}(0+)}{(m+1)!}\left(\frac{\log n}{n}\right)^{\frac{m}{2}+1} \\ & \times \left[1 + \sqrt{\frac{m+2}{2\pi \log n}} - \frac{\frac{m+3}{2}\log\log n - \frac{m+1}{4} + \frac{1}{2}\log\gamma_m}{\log n}\right]\end{aligned}} \tag{9}$$

where γ_m is as defined in the above.

2.4. What next

It should be emphasized that the thesis of a symmetric formulation of the testing problem that we propose here should be separated from the asymptotic expansions. The expansions are not meant to be taken literally. They serve as a guide, and not as a prescription.

Second, the formulation and the results given here should have extensions to various situations. The most interesting of such extensions could be to the case of multiple testing. There has been an overwhelming amount of stress on limiting false discoveries. A symmetric approach to the multiple testing problem along the lines here could be both refreshing and important. A simple formulation could be to write a density $g(\theta_1, \theta_2, \ldots, \theta_p)$ for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ in the alternative, and make the expected number of false discoveries and missed discoveries the same. Other formulations, more sophisticated, should be quite possible.

The multidimensional case would be good to work out. The obvious question is whether the level α will depend on the dimension, if asymptotic expansions similar to the ones in this article are carried out in higher dimensions, and if so, exactly how.

It would also be good to work out some theory corresponding to our symmetric formulation when the test used is a nonparametric or a robust test, for example, the Wilcoxon test, or a similar rank based test. Highly interesting rank methodologies in multiple testing problems have most recently been invented in [30, 31], and it may be similarly useful to extend our symmetric formulation to nonparametric situations.

Textbook confidence intervals, such as the t confidence interval, or the Wald interval for a proportion often run into problems when used in inappropriate situations. For the t interval, the problems occur with skewed data (see [29], and [20]), and for the Wald interval for a proportion the problem is the inaccuracy of the central limit theorem [12]. It could be of some use to know how the coverage probabilities behave when the nominal coverage is $1-\alpha_n$, with α_n as in this article (i.e., α_n roughly like $\frac{\log n}{n}$, or some power of it). The calculations will need a new Edgeworth expansion.

3. The Brown identities, inequalities of boundary value problems, and the Donsker-Varadhan principle

Two identities on Bayesian estimation of a d-dimensional multivariate normal mean $\boldsymbol{\theta}$ were presented in an unconspicuous manner in [8]. The identities are now generally referred to as the *Brown Identities*. The first identity gives a representation for the mean of the posterior distribution of $\boldsymbol{\theta}$, and the second a representation for the Bayes risk in terms of the Fisher information operator. Here are the two identities in their simplest form.

Theorem 2 (The Brown identities). *Let* $\mathbf{X} \sim N_d(\boldsymbol{\theta}, \sigma^2 I_d)$, *and let* P *be a prior probability distribution on* $\boldsymbol{\theta}$. *Let*

$$m(x) = \int_{\mathcal{R}^d} \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{1}{2\sigma^2}(x-\theta)'(x-\theta)}\, dP(\theta)$$

denote the marginal density function of $\mathbf{X}$. *Then,*

$$(a) \quad E(\boldsymbol{\theta}|\mathbf{X}) = \mathbf{X} + \sigma^2 \frac{\nabla m(\mathbf{X})}{m(\mathbf{X})}, \tag{10}$$

where ∇m *is the gradient vector of* m.

$$(b) \quad r(P) = E_m(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2) = d\sigma^2 - \sigma^4 I(m), \tag{11}$$

where $I(g) = \int_{\mathcal{R}^d} \frac{\|\nabla g\|^2}{g}$ *is the Fisher information operator on the class of functions*

$$\mathcal{G} = \{g : g \geq 0,\ \text{and is absolutely continuous}\}.$$

The first applications of the Brown identities were in establishing admissibility; see, for example, [8] and [9]. A second spate of applications came in deriving bounds on Bayes and minimax risks; see, for example, [4, 5, 6, 10, 11, 37].

Brown identities were applied in classical asymptotic theory in [20] for deriving stochastic expansions, and hence cube root central limit theorems for the difference of the MLE and the posterior mean of a one dimensional normal mean for priors with smooth densities. The methods used there should apply to higher dimensions, although they have not been done. It would be good to derive these stochastic expansions in the multidimensional case, using the Brown identity.

The applicability of Brown identities to classical asymptotics indicated that the full potential of the Brown identities has not yet been exploited. A few results in this section will hopefully give further credence to this thesis. The applications indicated here are of an entirely different nature.

Theorems 4 and 5 show that there is a mutual give and take between the Brown identity for Bayes risks and the seemingly completely unrelated problem of inequalities in boundary value problems, and in particular the problem of best constants in such inequalities. At the least, it seems interesting that the extremely hard problem of finding best constants in variational inequalities can be usefully approached by using mathematical statistics as a tool.

Theorem 7, on the other hand, shows that a triangulation of the Brown identity, a variational representation of the principal eigenvalue of the Laplacian over smooth domains, and a Brownian motion representation by M.D. Donsker and S.R.S. Varadhan of that same principal eigenvalue leads to a result in Gaussian mean estimation. The result is a bound on the minimax risk in terms of absorption times of a Brownian motion into the boundary of the domain in which the Gaussian mean lies.

3.1. Variational inequalities and Bayes risk bounds

For the specific links that we show in the examples and in Theorems 4 and 5, the following analytic inequalities would be used. They are stated together for easy reference.

Theorem 3. (a) **(Heisenberg uncertainty principle)** *Let* p, q, r *be constants, with* $p > 1$, $q = \frac{p}{p-1}$, $r > -1$. *Let* $f : \mathcal{R}_+ \to \mathcal{R}$ *be such that* $f \in \mathcal{C}^1(\mathcal{R}_+)$, *and each of* $x^{\frac{r}{p}} f$, $x^{\frac{r+1}{p-1}} f$, $f' \in L_p(\mathcal{R}_+)$. *Then,*

$$\int_0^\infty x^r |f|^p \le \frac{p}{r+1}\Big(\int_0^\infty x^{q(r+1)}|f|^p\Big)^{1/q}\Big(\int_0^\infty |f'|^p\Big)^{1/p}.$$

(b) **(Kolmogorov-Landau inequality)** *Let* $1 \le p, q, r \le \infty$ *and* $0 \le k < n$ *be such that*

(i) k, n *are integers;*
(ii) $\frac{n-k}{p} + \frac{k}{r} \ge \frac{n}{q}$.

Let

$$\alpha = \frac{n-k+q^{-1}-r^{-1}}{n+p^{-1}-r^{-1}}, \quad \beta = 1-\alpha.$$

Let $f : \mathcal{R} \to \mathcal{R}$ *be such that* $f \in L_{p,\mathcal{R}}, f^{(k)} \in L_{q,\mathcal{R}}, f^{(n)} \in L_{r,\mathcal{R}}$. *Then, there exists a universal constant* $K = K(n,k,p,q,r)$ *such that*

$$\|f^{(k)}\|_q \le K\big(\|f\|_p\big)^\alpha\big(\|f^{(n)}\|_r\big)^\beta.$$

(c) **(HELP inequality)** *Let* $f : \mathcal{R} \to \mathcal{R}$ *be such that* f, f' *are absolutely continuous, and* $f, f', f'' \in L_2(\mathcal{R})$. *Then for any real number* τ, *there exists a universal constant* $K(\tau)$ *such that*

$$\Big(\int_{-\infty}^\infty [(f'(x))^2 - \tau f^2(x)]\,dx\Big)^2 \le K(\tau)\int_{-\infty}^\infty f^2(x)\,dx \int_{-\infty}^\infty [f''(x)+\tau f(x)]^2\,dx.$$

In particular, $K(\tau)$ *may be taken to be 1 for all* $\tau \in \mathcal{R}$.

(d) **(The Nash inequality)** *Let* $d \ge 1$ *and let* $f : \mathcal{R}^d \to \mathcal{R}$ *be such that* $f \in L_1(\mathcal{R}^d)$ *and* f *belongs to the Sobolev space*

$$W^{1,2}(\mathcal{R}^d) = \{u : u, \|\nabla u\| \in L_2(\mathcal{R}^d)\}.$$

Then there exists a universal constant $C = C_{N,d}$ *such that*

$$\Big(\int_{\mathcal{R}^d} f^2\Big)^{1+2/n} \le C_{N,d}\Big(\int_{\mathcal{R}^d}\|\nabla f\|^2\Big)\Big(\int_{\mathcal{R}^d}|f|\Big)^{4/n}.$$

For the inequalities in part (c) and part (d) in this theorem, we refer respectively to Benammar et al. [2], and Carlen and Loss [15]. For the inequality in part (a), we refer to Zwillinger [39], and for the inequality in part (b), we refer to Babenko [1] and Hardy, Littlewood and Pólya [32].

We start with a simple, and yet, effective illustrative example of the interplay of the Heisenberg uncertainty principle and the second Brown identity.

Example 3 (Heisenberg's inequality and Bayes risk upper bounds). Let $X \sim N(\theta, 1)$ and suppose $\theta \sim P$, a probability distribution on $\mathcal{R}$. Suppose that under P, θ and $-\theta$ have the same distribution (that is, the distribution of θ is symmetric about zero). As above, let $m(x)$ denote the marginal density of X. Note that $m(x)$ is an even function of x. Using $f(x) = \sqrt{m(x)}$ in Heisenberg's uncertainty principle, with $p = 2$, and writing $r + 1 = \alpha > 0$, we have for each $\alpha > 0$ such that $E_P(|\theta|^{2\alpha}) < \infty$,

$$\left(\int_0^\infty \frac{(m')^2}{4m}\right)^{1/2}\left(\int_0^\infty x^{2\alpha}m\right)^{1/2} \geq \frac{\alpha}{2}\int_0^\infty x^{\alpha-1}m$$

$$\Leftrightarrow \frac{1}{2}\left(\frac{1}{2}\int_{-\infty}^\infty \frac{(m')^2}{m}\right)^{1/2}\left(\frac{1}{2}\int_{-\infty}^\infty |x|^{2\alpha}m\right)^{1/2} \geq \frac{\alpha}{2}\frac{1}{2}\int_{-\infty}^\infty |x|^{\alpha-1}m$$

$$\Leftrightarrow I(m)\int_{-\infty}^\infty |x|^{2\alpha}m \geq \alpha^2\left(\int_{-\infty}^\infty |x|^{\alpha-1}m\right)^2 \tag{12}$$

$$\Leftrightarrow r(P) \leq 1 - \frac{\alpha^2\left[E_m|X|^{\alpha-1}\right]^2}{E_m|X|^{2\alpha}}, \tag{13}$$

where $r(P)$ denotes the Bayes risk under our prior distribution P and E_m denotes expectation under the marginal density of X. By taking an infimum of the rhs over α, we get a Bayes risk upper bound that involves only moment calculations under the marginal:

$$\boxed{r(P) \leq 1 - \sup_{\{\alpha>0: E_P(|\theta|^{2\alpha})<\infty\}} \frac{\alpha^2\left[E_m|X|^{\alpha-1}\right]^2}{E_m|X|^{2\alpha}}.} \tag{14}$$

Note that in the special case when the prior distribution P is normal, the bound is exact; i.e., the ultimate lower bound given in the line above coincides with the exact Bayes risk. But the bound itself is for any symmetric prior distribution P.

The next theorem gives a lower bound on Bayes risks by using the HELP inequality in Theorem 3. The lower bound involves only the first two derivatives of the density of the prior distribution, and is, therefore, easy to compute.

Theorem 4 (HELP inequality and Bayes risk lower bounds). *Let $X \sim N(\theta, 1)$ and let θ have an absolutely continuous prior distribution P with a density g such that*

(i) g is logconcave;

(ii) $\dfrac{|g'|}{g} \leq \alpha < \infty$; $\quad \dfrac{|g''|}{g} \leq \beta < \infty$.

Then,

$$r(P) \geq 1 - \frac{(5\alpha^2 + 8\beta) + \sqrt{25\alpha^4 + 80\beta^2 + 80\alpha^2\beta}}{2}. \tag{15}$$

Remark. Note that it is possible that the lower bound turns out to be a negative number, in which case the bound is not useful. So, a better way to think of this bound is to think of it for general n, when the bound will read like

$$\boxed{(16) \qquad r_n(P) \geq \frac{1}{n} - \frac{(5\alpha^2 + 8\beta) + \sqrt{25\alpha^4 + 80\beta^2 + 80\alpha^2\beta}}{2n^2}}$$

and so for large enough n, the bound will not be vacuous.

Proof of Theorem 4. We use the HELP inequality with the special choice $\tau = 0$ and use $f = \sqrt{m}$. The following facts will be used during the proof, and we collect them together for easy reference:

(a) By familiar arguments, $\sup_x \frac{|m'(x)|}{m(x)} \leq \sup_\theta \frac{|g'(\theta)|}{g(\theta)} \leq \alpha$, and $\sup_x \frac{|m''(x)|}{m(x)} \leq \sup_\theta \frac{|g''(\theta)|}{g(\theta)} \leq \beta$.

(b) Since the logconcavity of g implies that P is strongly unimodal, and since strong unimodality is closed under convolution, $m(x)$ is also a logconcave function.

(c) Therefore,

$$(\log m)'' = \frac{m''}{m} - \frac{(m')^2}{m^2} \leq 0. \tag{17}$$

(d) The first two derivatives of $f = \sqrt{m}$ are

$$f' = \frac{m'}{2\sqrt{m}}; \quad f'' = \frac{m''}{2\sqrt{m}} - \frac{(m')^2}{4m^{3/2}}. \tag{18}$$

(e) Moreover, we can rewrite f'' as

$$\begin{aligned} f'' &= \sqrt{m}\left[\frac{m''}{2m} - \frac{(m')^2}{4m^2}\right] \\ &= \sqrt{m}\left[\frac{m''}{2m} - \frac{(m')^2}{2m^2} + \frac{(m')^2}{4m^2}\right]. \end{aligned} \tag{19}$$

We now use the HELP inequality in part (c) of Theorem 3, with $\tau = 0$. Then, the lhs of the inequality is $(\int_{-\infty}^{\infty} (f'(x))^2\,dx)^2 = \frac{I^2(m)}{16}$. On the rhs, $\int_{-\infty}^{\infty} f^2(x\,dx = 1$, and

$$\int_{-\infty}^{\infty} [f''(x)]^2\,dx$$

$$= \int_{-\infty}^{\infty} \left[\left\{\frac{m''}{2m} - \frac{(m')^2}{2m^2}\right\}^2 m + \frac{(m')^4}{16m^4} m + 2\frac{(m')^2}{4m^2}\left\{\frac{m''}{2m} - \frac{(m')^2}{2m^2}\right\} m\right] dx \tag{20}$$

$$\leq \int_{-\infty}^{\infty} \left[\left\{\frac{1}{4}\left(\frac{m''}{m}\right)^2 + \frac{1}{4}\left(\frac{m'}{m}\right)^4 - \frac{m''(m')^2}{2m^3}\right\} m + \frac{1}{16}\left(\frac{m'}{m}\right)^4 m\right] \tag{21}$$

(because, by fact (c) stated above, the cross product term in the line before is ≤ 0)

$$\begin{aligned} &\leq \left[\frac{1}{4}\beta^2 + \frac{1}{4}\alpha^2 I(m) + \frac{\beta}{2} I(m) + \frac{1}{16}\alpha^2 I(m)\right] \\ &= \left[\frac{1}{4}\beta^2 + \left\{\frac{5\alpha^2}{16} + \frac{\beta}{2}\right\} I(m)\right]. \end{aligned} \tag{22}$$

Now, plugging this into the HELP inequality, we obtain

$$\begin{aligned}\frac{I^2(m)}{16} &\le \frac{1}{4}\beta^2 + \left[\frac{5\alpha^2}{16} + \frac{\beta}{2}\right] I(m) \\ \Rightarrow I^2(m) &\le 4\beta^2 + [5\alpha^2 + 8\beta] I(m).\end{aligned} \tag{23}$$

This last inequality is a quadratic inequality in $I(m)$, and by solving for the larger root of the corresponding quadratic equation, we get

$$I(m) \le \frac{(5\alpha^2+8\beta)+\sqrt{25\alpha^4+80\beta^2+80\alpha^2\beta}}{2}, \tag{24}$$

and therefore by the Brown identity,

$$r(P) = 1 - I(m) \ge 1 - \frac{(5\alpha^2+8\beta)+\sqrt{25\alpha^4+80\beta^2+80\alpha^2\beta}}{2}, \tag{25}$$

which is the stated lower bound on the Bayes risk $r(P)$. □

3.2. The Brown identity and the best constant in the Nash inequality

The result that we next present shows the reverse side of a synergistic relationship, namely that it will show that the best constant problem in the Nash inequality can benefit from relating it to the Brown identity for Bayes risks. Additionally, it will be seen that the estimate of the best constant that the Brown identity produces is surprisingly accurate, while also being explicit and simple. In contrast, the apparently common estimate of the best constant cited in analysis [15] involves the first positive zero of the Bessel function of the first kind, J_ν, with $\nu = \frac{d}{2}$, and there are no formulas for the first positive root.

Theorem 5. *Let $X \sim N_d(\boldsymbol{\theta}, I_{d\times d})$ and suppose $\boldsymbol{\theta}$ has a strongly unimodal prior probability distribution P. Let $h(x) = \log m(x)$ and $L(x) = L_{d\times d} = (\frac{-\partial^2 h}{\partial x_i \partial x_j})$. Let $\lambda_1(x)$ be the smallest eigenvalue of $L(x)$. Then the best constant in the d-dimensional Nash inequality satisfies*

$$\boxed{C_{N,d} \ge \sup_{\{P:P \text{ strongly unimodal}\}} \left[\frac{[\inf_x \lambda_1(x)]^2}{4\pi^2 [\sup_x m(x)]^{2/d} [d - r(P)]}\right].} \tag{26}$$

In particular, by taking P to be the $N(\mathbf{0}, I_{d\times d})$ distribution,

$$C_{N,d} \ge \frac{1}{2\pi d}. \tag{27}$$

Discussion. The estimate of the best constant given in [15] is

$$C_{N,d} \le \frac{2(\frac{d}{2}+1)^{1+2/d}}{d w_d^{2/d} k_d^2}, \tag{28}$$

where $w_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ is the volume of the d-dimensional unit ball, and k_d is the first positive zero of the Bessel function $J_{\frac{d}{2}}(x)$. The following short numerical table illustrates the surprising effectiveness of the lower bound $C_{N,d} \ge \frac{1}{2\pi d}$. The lower bound

$\frac{1}{2\pi d}$ can be improved by using strongly unimodal priors P that are nonnormal. The main purpose of Theorem 3 is to show the connection of the best constant to the supremum over strongly unimodal priors of the quantity on the rhs in the theorem. However, we do not know which prior P, if any, will result in the Carlen-Loss upper bound. In any case, the simple bound $\frac{1}{2\pi d}$ is already impressively accurate; here is a table.

d	First positive zero of $J_{d/2}$	Carlen-Loss upper bound	Theorem 5 lower bound
1	π	.170979	.159155
2	3.83171	.086721	.079578
3	4.49341	.058515	.053052
4	5.13562	.044344	.039789
5	5.76346	.035799	.031831
6	6.38016	.030074	.026526
10	8.77148	.018508	.015916

Proof of Theorem 5. Once again, the function f to use is $f = \sqrt{m}$. We apply the Nash inequality to this choice of f. However, notice that we must now deal with the L_1 norm of $\sqrt{m}$ in order to apply the Nash inequality. It is for this purpose that the strong unimodality of P will be helpful. Here is how we deal with the L_1 norm of $\sqrt{m}$. In the following, we use the notation $u = \operatorname{argmax}(h) = \operatorname{argmax}(\log m)$. Then,

$$\begin{aligned}
\int_{\mathcal{R}^d} \sqrt{m}\,dx &= \int_{\mathcal{R}^d} e^{\log\sqrt{m}}\,dx = \int_{\mathcal{R}^d} e^{\frac{1}{2}h(x)}\,dx \\
&= \int_{\mathcal{R}^d} e^{\frac{1}{2}[h(u)+(x-u)'\nabla h(u)-\frac{1}{2}(x-u)'L(x^*)(x-u)]}\,dx \\
&= \sqrt{m(u)}\int_{\mathcal{R}^d} e^{-\frac{1}{4}(x-u)'L(x^*)(x-u)}\,dx \\
&\le \sqrt{m(u)}\int_{\mathcal{R}^d} e^{-\frac{1}{4}(\inf_x \lambda_1(x))(x-u)'(x-u)}\,dx \\
&= \sqrt{\sup_x m(x)}\left(\frac{4\pi}{\inf_x \lambda_1(x)}\right)^{d/2}.
\end{aligned} \tag{29}$$

We plug this bound on $\int_{\mathcal{R}^d}|f|$ in the Nash inequality (see part (d) of Theorem 3). Since $\int_{\mathcal{R}^d} f^2 = 1$, and $\int_{\mathcal{R}^d}\|\nabla f\|^2 = \frac{1}{4}I(m)$, the inequality given in Theorem 5 now follows on simple manipulation, which is omitted. $\square$

We end with a result connecting the marginal density of X to the Fisher information of the density of the prior distribution P. The result is obtained by putting together three ingredients, namely the Brown identity, the Borovkov-Sakhanienko lower bound on Bayes risks, and the Kolmogorov-Landau inequality in part (b) of Theorem 3. The result essentially says that if the density of the prior distribution is flat, so that its Fisher information is small, then the marginal density must also be flat, because the supremum of the marginal density will be small. This is not surprising; but the result below succeeds in quantifying the flatness of the marginal density. We should add that the best constant K cited in the result below is known in the Russian literature (see pp. 80 in [1]); but we could not acquire the article. Here is what the result says.

Theorem 6. *Let $X \sim N(\theta, 1)$ and let $\theta \sim P$. Assume that P is absolutely continuous with an absolutely continuous density g. Let K denote the best constant $K(1, 0, 2, \infty, 2)$ in the Kolmogorov-Landau inequality. Then,*

$$m(x) \leq \frac{K^2}{2}\sqrt{\frac{I(g)}{1+I(g)}} \quad \textit{for all } x. \tag{30}$$

Proof. Choosing $k = 0$, $n = 1$, $q = \infty$, $p = r = 2$ in the Kolmogorov-Landau inequality, and using $f = \sqrt{m}$, we have

$$\sqrt{\sup m} = \sup(\sqrt{m}) \leq K\left(\frac{1}{2}\sqrt{I(m)}\right)^{1/2}. \tag{31}$$

By the Brown identity and the Borovkov-Sakhanienko lower bound (also available in pp. 1581, [10]),

$$1 - I(m) \geq \frac{1}{1+I(g)}. \tag{32}$$

Plugging this into (31), on a little manipulation, one gets

$$\sup m \leq \frac{K^2}{2}\big(I(m)\big)^{1/2} \leq \frac{K^2}{2}\sqrt{\frac{I(g)}{1+I(g)}}.$$

□

3.3. Brown identity, minimax risk over bounded domains, and chasing a Brownian motion

We finish by drawing a connection between the minimax risk in estimating a d-dimensional Gaussian mean constrained to lie in a bounded domain Ω, and the time to absorption of a d-dimensional Brownian motion into the boundary of Ω. We derive an inequality on the minimax risk, and then state a conjecture. Minimax estimation of Gaussian means constrained to lie in some bounded domain now has a huge literature. The first two key papers are [4], and [17]. For later developments, see [24, 25, 23, 38, 13, 14], and [19], among numerous others.

To derive our inequality on the minimax risk, we will borrow two results from probability and analysis. It would be helpful to gather these two results first.

Suppose $\Omega \subset \mathcal{R}^d$ is a bounded open domain, with a *smooth* boundary. The result we borrow requires the boundary to be sufficiently smooth for the Dirichlet problem on $\bar{\Omega}$ to be solvable (the solid ellipsoids, solid cubes, etc., all will work).

Let c be a bounded and Hölder continuous real valued function on $\mathcal{R}^d$. If a function u and a real valued constant λ solve the equation

$$\frac{1}{2}\triangle u + cu = \lambda u \text{ in } \Omega;$$
$$u \text{ continuous in } \bar{\Omega},\ u = 0 \text{ on } \partial\Omega,$$

then λ is called an eigenvalue of $\frac{1}{2}\triangle + c$ corresponding to the eigenfunction u.

It is known that there is a sequence of eigenvalues

$$\infty > \lambda_0 > \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots;$$

the principal eigenvalue λ_0 admits the Rayleigh-Ritz representation:

$$\lambda_0 = -\inf_{\{h\geq 0:\, h\in L_2(\Omega)\}} \int_\Omega \left[ch^2 + \|\nabla h\|^2\right]. \tag{33}$$

This result will be used by us. We refer to [18] for this result.

But λ_0 also has a Brownian motion connection. Take an $x \in \Omega$, and consider a d-dimensional Brownian motion $W_d(t), t \geq 0$, starting at x. Let τ be the stopping time

$$\tau = \inf\{t \geq 0 : W_d(t) \notin \Omega\}.$$

Then,

$$\lambda_0 = \lim_{t\to\infty} \frac{1}{t} \log \sup_x E_x\left[e^{\int_0^t c(W_d(s))ds} I_{\{\tau > t\}}\right]. \tag{34}$$

We refer to Donsker and Varadhan [26, 27] for this result.

With this background, we now proceed to draw the connection between the minimax risk and τ. Suppose then that $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N_d(\theta, \sigma^2 I_d)$. Suppose $\boldsymbol{\theta} \in \bar{\Omega}$, a domain as specified in the above. Consider the minimax risk $\rho_{n,\Omega} = \inf_\delta \sup_\theta E_\theta \|\delta(X_1, X_2, \ldots, X_n) - \theta\|^2$. We recall the notation $r_n(P)$ for Bayes risk under a prior probability distribution P. Then,

$$\begin{aligned}
\rho_{n,\Omega} &= \sup_{\{P:\text{ P prob. measure on } \bar{\Omega}\}} r_n(P) \\
&= \sup_P \left[d\frac{\sigma^2}{n} - \frac{\sigma^4}{n^2} I(m)\right] \quad \text{(Brown identity)} \\
&= d\frac{\sigma^2}{n} - \frac{\sigma^4}{n^2} \inf_P I(m) \\
&\geq d\frac{\sigma^2}{n} - \frac{\sigma^4}{n^2} \inf_{\{P:\ P \text{ has a density g}\}} I(m) \\
&\geq d\frac{\sigma^2}{n} - \frac{\sigma^4}{n^2} \inf_g I(g)
\end{aligned}$$

(because $I(m) \leq I(g)$, Fisher information being a convex operator; Huber [33] proves this convexity property)

$$\begin{aligned}
&= d\frac{\sigma^2}{n} + \frac{\sigma^4}{n^2}\left(-\inf_g \int \frac{\|\nabla g\|^2}{g}\right) \\
&= d\frac{\sigma^2}{n} + \frac{4\sigma^4}{n^2}\left[-1 - \inf_{\{h\geq 0: \|h\|_2 = 1\}} \int \{-h^2 + \|\nabla h\|^2\}\right]
\end{aligned}$$

(by writing $h = \sqrt{g}$)

$$= d\frac{\sigma^2}{n} - \frac{4\sigma^4}{n^2} + \frac{4\sigma^4}{n^2} \lim_{t\to\infty} \frac{1}{t} \log \sup_x E_x\left[e^{\int_0^t (-1)ds} I_{\{\tau > t\}}\right]$$

(by (34)).

We have thus derived an inequality on the minimax risk, and it is now stated formally.

Theorem 7. *Let* $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N_d(\boldsymbol{\theta}, \sigma^2 I_d)$. *Suppose* $\boldsymbol{\theta} \in \bar{\Omega} \subset \mathcal{R}^d$, *where* Ω *is a bounded open domain such that the Dirichlet problem on* $\bar{\Omega}$ *is solvable. Let* $W_d(t), t \geq 0$, *be a* d*-dimensional Brownian motion starting at* $x, x \in \Omega$, *and*

$$\tau = \inf\{t \geq 0 : W_d \notin \Omega\}.$$

Then the minimax risk of estimation of θ *satisfies*

$$\begin{aligned}\rho_{n,\Omega} &= \inf_{\delta} \sup_{\boldsymbol{\theta} \in \bar{\Omega}} E_{\boldsymbol{\theta}} \left[\|\delta(X_1, X_2, \ldots, X_n) - \boldsymbol{\theta}\|^2 \right] \\ &\geq d\frac{\sigma^2}{n} - \frac{4\sigma^4}{n^2} + \frac{4\sigma^4}{n^2} \lim_{t \to \infty} \frac{1}{t} \log \sup_x E_x \left[e^{-t} I_{\{\tau > t\}} \right].\end{aligned} \tag{35}$$

The conjecture is that, asymptotically, i.e., for large n, *the inequality should be an approximate equality. In that case, we would be able to approximate the minimax risk in general dimensions over general smooth bounded domains by chasing a Brownian motion to the boundary of the domain.* Such a link should be quite interesting. It will also enable one to numerically approximate the minimax risk over arbitrary smooth bounded domains by using the following algorithm:

Step 1 Pick a (large) t and K random (or finely spaced) points $\mathbf{x}_i$, $i = 1, 2, \ldots, K$ from Ω.
Step 2 Pick a (large) number N, and simulate a d-dimensional Brownian motion starting at the origin N times.
Step 3 Shift it to start at $\mathbf{x}_i$. Evaluate τ_i, the time to absorption of the simulated path into the boundary of Ω.
Step 4 Count $M(\mathbf{x}_i)$, the number of simulated paths for which $\tau_i > t$.
Step 5 Find $c(t) = \max_{1 \leq i \leq K} \frac{M(\mathbf{x}_i)}{N}$.
Step 6 Approximate the minimax risk as

$$\rho_{n,\Omega} \approx \frac{d\sigma^2}{n} - \frac{8\sigma^4}{n^2} - \frac{4\sigma^4}{n^2 t} \log(c(t)).$$

Step 7 Pick a larger t and repeat.

Acknowledgements

I am grateful to an anonymous referee for catching a misstatement in an earlier draft of this paper. I am also thankful to Tony Cai, Peter Hall and Steve Stigler for their comments and for providing me with some references. Elyse Gustafson and Geri Mattson gave me invaluable help and I appreciate their gestures very much.

References

[1] Babenko, V. F. (1998). Inequalities for norms of intermediate derivatives and some applications. In *Recent Progress in Inequalities* (G. V. Milovanović, ed.) 77–96. Kluwer, Boston. MR1610003

[2] Benammar, M. et al. (1998). A catalogue of HELP and HELP-type integral and series inequalities. In *Recent Progress in Inequalities* (G. V. Milovanović, ed.) 127–160. Kluwer, Boston. MR1609931

[3] Berger, J. and Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of P-values and evidence. *J. Amer. Statist. Assoc.* **82**(397) 112–139. With comments and a rejoinder. MR883340

[4] Bickel, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9**(6) 1301–1309. MR630112

[5] Bickel, P. J. (1983). Minimax estimation of the mean of a normal distribution subject to doing well at a point. In *Recent Advances in Statistics* 511–528. Academic Press, New York. MR736544

[6] Bickel, P. J. and Collins, J. (1983). Minimizing Fisher information over mixtures of distributions. *Sankhyā, Ser. A* **45**(1) 1–19. MR749349

[7] Borovkov, A. A. and Sakhanienko, A. U. (1980). On estimates of the expected quadratic risk. *Probab. Math. Statist.* **1** 185–195. MR626310

[8] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903. MR0286209

[9] Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics, III, Vol. I* (J. Berger and S. S. Gupta, eds.) 205–230. Academic Press, New York. MR705290

[10] Brown, L. D. and Gajek, L. (1990). Information inequalities for Bayes risk. *Ann. Statist.* **18**(4) 1578–1594. MR1074424

[11] Brown, L. D. and Low, M. (1991). Information inequality bounds on the minimax risk. *Ann. Statist.* **19**(1) 329–337. MR1091854

[12] Brown, L., Cai, T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30**(1) 160–201. MR1892660

[13] Cai, Tony and Low, M. G. (2004). Minimax estimation of convex functionals over nonconvex parameter spaces. *Ann. Statist.* **32** 552–576. MR2060169

[14] Cai, Tony and Low, M. G. (2005). Nonparametric estimation over shrinking neighborhoods: Superefficiency and adaptation. *Ann. Statist.* **33** 184–213. MR2157801

[15] Carlen, E. and Loss, M. (1993). Sharp constant in Nash's inequality. *Internat. Math. Res. Notices* **7** 301–305. MR1230297

[16] Casella, G. and Berger, R. L. (1987). Reconciling Bayesian and frequentist evidence in the one sided testing problem. *J. Amer. Statist. Assoc.* **82**(397) 106–111. With comments and a rejoinder. MR883339

[17] Casella, G. and Strawderman, W. E. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9**(4) 870–878. MR619290

[18] Courant, R. and Hilbert, D. (1953). *Methods of Mathematical Physics, Vol. I.* Interscience, New York. MR0065391

[19] DasGupta, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. *Sankhyā, Ser. A* **47**(3) 326–332. MR863726

[20] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer, New York.

[21] Diaconis, P. and Holmes, S. (1996). Are there still things to do in Bayesian Statistics? *Erkenntnis* **45**(2/3) 145–158. MR1453370

[22] Diaconis, P. and Saloff-Coste, L. (1996). Nash inequalities for finite Markov chains. *J. Theoret. Probab.* **2** 459–510. MR1385408

[23] Donoho, D. and Liu, R. M. (1991). Geometrizing rates of convergence. *Ann. Statist.* **19** 668–701.

[24] Donoho, D. and Johnstone, I. M. (1994). Minimax risk over ℓ_p balls with

ℓ_q error. *Probab. Theory Related Fields* **99**(2) 277–303. MR1278886

[25] Donoho, D. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26** 879–921. MR1635414

[26] Donsker, M. D. and Varadhan, S. R. S. (1975). On a variational formula for the principal eigenvalue for operators with a maximum principle. *Proc. Natl. Acad. Sci. USA* **72** 780–783. MR0361998

[27] Donsker, M. D. and Varadhan, S. R. S. (1976). On the principal eigenvalue of second-order elliptic differential operators. *Comm. Pure Appl. Math.* **29** 595–621. MR0425380

[28] Hall, P. (1983). Inverting an Edgeworth expansion. *Ann. Statist.* **11**(2) 569–576. MR696068

[29] Hall, P. (1987). Edgeworth expansion for Student's t-statistic under minimal moment conditions. *Ann. Probab.* **15**(3) 920–931. MR893906

[30] Hall, P. and Miller, H. (2009). Using the bootstrap to quantify the authority of an empirical ranking. *Ann. Statist.* **37**(68) 3929–3959. MR2572448

[31] Hall, P. and Miller, H. (2010). Modelling the variability of rankings. *Ann. Statist.* To appear.

[32] Hardy, G., Littlewood, J. and Pólya, G. (1952). *Inequalities.* Cambridge University Press, Cambridge. MR0046395

[33] Huber, P. J. (1981). *Robust Statistics.* Wiley, New York. MR606374

[34] Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York. MR852406

[35] Mitrinović, D. S. and Vasić, P. M. (1970). *Analytic Inequalities.* Springer-Verlag, Berlin. MR0274686

[36] Oh, H. S. and DasGupta, A. (1999). Comparison of the P-value and posterior probability. *J. Statist. Plan. Inf.* **76**(1/2) 93–107. MR1673342

[37] Vidakovic, B. and DasGupta, A. (1995). Lower bounds on Bayes risks for estimating a normal variance. *Canad. J. Statist.* **23**(3) 269–282. MR1363592

[38] Wasserman, L. (2006). *All of Nonparametric Statistics.* Springer, New York. MR2172729

[39] Zwillinger, D. (1992). *Handbook of Integration.* Jones and Bartlett, Boston. MR1174295