PARTIAL DIFFERENTIAL EQUATIONS AND EXPECTATION IDENTITIES: WITH APPLICATIONS*

by

L. Brown
University of Pennsylvania

A. DasGupta
Purdue University
University of Pennsylvania

L. R. Haff
University of California
San Diego

W. E. Strawderman
Rutgers University

Technical Report #00-04

Department of Statistics
Purdue University
West Lafayette, IN USA

April 2000

PARTIAL DIFFERENTIAL EQUATIONS AND EXPECTATION IDENTITIES: WITH APPLICATIONS*

by

L. Brown
University of Pennsylvania

A. DasGupta
Purdue University
University of Pennsylvania

L. R. Haff
University of California
San Diego

W. E. Strawderman
Rutgers University

Abstract

By using the telegraph partial differential equation, the Fokker-Planck equation, and Itô's theorem, we derive a collection of expectation identities covering various univariate and multivariate distributions, and obtain various applications in statistics, probabilty, and mathematics. A particular expectation identity is shown to be almost equivalent to Stein's identity and is also shown to be characteristic of the normal distribution, in a suitable sense. Applications include exact and approximate expressions and bounds for moments, improvement and reversal of Jensen's inequality, characterizing unbiased estimability to solutions of PDEs, applications to decision theory and Bayesian statistics, deriving some new and some known properties of harmonic, polyharmonic and subharmonic functions, and obtaining a method of counting matchings in graphs by use of Stein's identity and the heat equation. Illustrative examples are given.

*Research supported by the National Science Foundation & National Security Agency

1. INTRODUCTION

In 1981, Charles Stein published a simple but greatly useful identity that has now come to be known as Stein's identity. The simplest version of the identity says that if $X \sim N(\mu, 1)$, then for sufficiently smooth functions $g(x), E((X-\mu)g(X)) = E(g'(X))$; see Stein(1981). The identity may be called an expectation identity. In this article, we present a series of such expectation identities covering various probability distributions, both univariate and multivariate, and show various mutual connections between the identities, and outline a rather large collection of applications of these identities in a range of areas in statistics, probability, and mathematics. It seems there are potentials for additional applications and refinements also.

The article starts with deriving the various identities; most of the identities are presented in Section 2. The identities come from one of three sources : the telegraph equation of electromagnetic transmission, the Fokker-Planck equation of homogeneous diffusions, and Itô's theorem in stochastic calculus. Some of the identities can be derived from more than one source, indicating a connection between the approaches. The number of identities presented in all is somewhat large. But in some sense, identities (3), (4), (15) and (27) are special. The various applications in the later sections generally stem from one of these four identities. As we just remarked, there are some technical connections between some of these identities. Actually, something more is true. The identity (3), which we have called the heat equation identity, is almost equivalent to the Stein identity. And the heat equation identity is characteristic of the normal distribution, in a suitable sense. These results are available in Section 4.

The applications are grouped according to area in the remaining sections. The applications that are presented by us in the present draft can be broadly classified into the following areas :

a) deriving exact moment formulas and analytical lower and upper bounds;

b) approximate but apparently practically useful computation of expectations of complex statistics, by simultaneous use of our identities and Bahadur representations and Hajek projections of these statistics;

c) connecting unbiased estimation to elliptic partial differential equations;

d) applications to decision theory, specifically, establishing inadmissibilty results and a Stein inequality (as opposed to a Stein identity) for spherically symmetric t distributions;

e) applications to Bayesian statistics, specifically, establishing lower bounds on Bayes risks in the spirit of Brown-Gajek-Borovkov-Sakhanienko, and establishing a connection between oscillations of a Bayes estimate and its Bayes risk;

f) applications to deriving mathematical and statistical properties of harmonic, subharmonic and polyharmonic functions, and specifically, some general zero correlation results;

g) applications in graph theory, and specifically, establishing a connection between counting perfect matchings in graphs and the heat equation identity.

These and other applications and illustrative examples are presented according to the area of application in Section 3 and Sections 5 through 10.

The results we have outlined in this manuscript developed over a period of several years. Perhaps the most interesting aspect is that there appears to be a broad range of applications, and it seems likely that there should be other applications. We are thankful to Persi Diaconis and Joe Eaton for their suggestions, questions, and advice.

2. NOTATION AND IDENTITIES

As stated in the introduction, the article presents a series of expectation identities and one of the goals is to show the mutual connections and the commonalities between the identities and the various methods to derive them. Of course, application has to be a goal and with practical applications in mind, some other identities covering other practically important problems will be presented in later sections. However, most of our family of identities are presented together in this section to clarify the connections. First the notation is explained. We recommend that the reader refers back to the notation as a particular result is stated.

2.1. Notation

a. $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$: will generally mean a probability density function on $\mathbb{R}^p, p \geq 1 : \underset{\sim}{\mu}, t$ are to be understood as parameters, with $\underset{\sim}{\mu}$ in $\mathbb{R}^p$ and $t > 0$.

b. $g(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$: will generally denote a nonstochastic k dimensional function, $k \geq 1$;

For any scalar function $g(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$, g_t will denote $\frac{\partial}{\partial t} g$, g_{tt} will denote $\frac{\partial^2}{\partial t^2} g$, $\Delta_x g$ will denote the gradient vector with respect to $\underset{\sim}{x}$, $\nabla_x \cdot g$ will denote divergence, and Δ_x will denote the Laplacian. Similar meanings will apply to $\nabla_\mu g$, $\nabla'_\mu \cdot g$, and $\Delta_\mu g$. Also, H_g will denote the Hessian matrix of g, with respect to $\underset{\sim}{x}$. If $\underset{\sim}{x}$ is scalar, $g^{(n)}(x)$ will as usual mean the n^{th} derivative of g.

c. $B(\underset{\sim}{a}, r)$: will denote a sphere in p dimensions with radius r and center at $\underset{\sim}{a}$; $\partial B(\underset{\sim}{a}, r)$ will denote the boundary of $B(\underset{\sim}{a}, r)$ and $\int\limits_{\partial B} u d\sigma$ will denote the surface integral of a given function u on $\partial B(\underset{\sim}{a}, r)$.

d. $||\underset{\sim}{y}||$ will denote Euclidean norm and $\underset{\sim}{y}'\underset{\sim}{z}$ will denote inner product. $I = I_p$ will denote the $p \times p$ identity matrix and $\underset{\sim}{e}_i$ the i^{th} unit vector.

e. W_t: will denote k-dimensional Wiener process.

f. $E_{\underset{\sim}{\mu}, t}$ will denote expectation and often will be written as just E; similarly, var$(\cdot)$ will stand for variance and cov for covariance.

g. $h(\underset{\sim}{\mu}, t)$: will generally denote a parametric function and $\delta(\cdot)$ an estimate of a parametric function.

h. $R(\underset{\sim}{\mu}, t, \delta)$: will generally denote the risk function of an estimate under squared error loss; also, $\pi(\underset{\sim}{\mu})$ will denote a prior density for $\underset{\sim}{\mu}$ and $r(t, \pi) = r(t, \pi, \delta)$ the Bayes risk of an estimate δ.

i. $\phi(\cdot)$ will denote, as usual, the standard normal density and $\Phi(\cdot)$ the standard normal CDF; also, H_n will denote the n^{th} Hermite polynomial.

j. $\Gamma(\alpha, y)$ will denote $\int\limits_y^\infty e^{-u} u^{\alpha-1} du$ and $\gamma(\alpha, y)$ will denote $\int\limits_0^y e^{-u} u^{\alpha-1} du$.

k. $N_p(\underset{\sim}{\mu}, t\Sigma_0)$ will denote a p-dimensional normal distribution with mean vector $\underset{\sim}{\mu}$ and covariance matrix $t\sum_0$; $C_p(\underset{\sim}{\mu}, t\sum_0)$ will denote a p-dimensional elliptically symmetric Cauchy distribution with location parameter $\underset{\sim}{\mu}$ and scale matrix $t\sum_0$; $\chi^2(m)$ will denote a central chisquare distribution with m degrees of freedom; $t_p(m)$ will denote the p-dimensional t distribution with m degrees of freedom defined as the distribution of $\frac{\sqrt{m}\underset{\sim}{Z}}{\sqrt{\underline{Y}}}$ if $\underset{\sim}{Z} \sim N_p(\underset{\sim}{0}, I), \underline{Y} \sim \chi^2(m)$ and $Y, \underset{\sim}{Z}$ are independent; the distribution of $\underset{\sim}{\mu} + \frac{\sqrt{m}\underset{\sim}{Z}}{\sqrt{\underline{Y}}}$ will be denoted as $t_p(m, \underset{\sim}{\mu})$.

2.2. Identities From the Telegraph Equation

2.2.1. The Telegraph Equation. The telegraph equation in $\mathbb{R}^p$ is a partial differential equation of the form

$$\Delta_x p = \alpha p_{tt} + \beta p_t + \gamma p, \tag{1}$$

where α, β are constants and γ is a function of t. Our motivation for considering the telegraph equation is that many important probability density functions satisfy the telegraph equation for suitable choices of α, β, and γ. We will see many such examples. What does that give us? It turns out that if a particular probability density function $p(\underset{\sim}{x}, t)$ satisfies the telegraph equation (1), then for a statistic $g(\underset{\sim}{x})$, by simply multiplying both sides of (1) by $g(\underset{\sim}{x})$ and integrating, one formally gets an expectation identity:

$$E(\Delta_x g) = \alpha \frac{\partial^2}{\partial t^2} E(g) + \beta \frac{\partial}{\partial t} E(g) + \gamma E(g). \tag{2}$$

To make the formalism of (2) rigorous, it will be necessary to impose conditions on the function g. Once one finds an expectation identity (2), one then attempts to find applications. In brief, this is the agenda. We may note here that the origin of the telegraph equation (1) is in the area of electromagnetic transmission of signals, α, β, γ having to do with resistance, capacitance, inductance, and leakage of the cable. See Folland (1992).

2.2.2. Densities Satisfying the Telegraph Equation

As stated above, many common density functions satisfy the telegraph equation. Some particular cases are given below.

Proposition 1.

a. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$. Then the density of $\underset{\sim}{x}$ satisfies the heat equation, i.e., the telegraph equation with $\alpha = \gamma = 0$ and $\beta = 2$;

b. Let $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$ denote the density of any Gaussian convolution, i.e., let $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t) = \int \frac{1}{(2\pi t)^{\frac{p}{2}}} e^{-\frac{1}{2t}(\underset{\sim}{x}-\underset{\sim}{\mu}-\underset{\sim}{z})'(\underset{\sim}{x}-\underset{\sim}{\mu}-\underset{\sim}{z})} f(\underset{\sim}{z}) d\underset{\sim}{z}$. Then $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$ also satisfies the heat equation;

c. Let $\underset{\sim}{x} \sim C_p(\underset{\sim}{\mu}, tI)$. Then the density of $\underset{\sim}{x}$ satisfies the wave equation, i.e., the telegraph equation with $\beta = \gamma = 0$ and $\alpha = -1$;

d. Let $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$ denote the density of any Cauchy convolution, i.e., $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t) = \int \frac{}{(||\underset{\sim}{x}-\underset{\sim}{\mu}-\underset{\sim}{z}||^2+pt^2)^{\frac{p+1}{2}}} f(\underset{\sim}{z}) d\underset{\sim}{z}$. Then $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$ also satisfies the wave equation;

e. Let $f(x)$ be a twice differentiable one dimensional density. Then the mixture density $p(x,t) = wf(x-t) + (1-w)f(x+t)$ satisfies the telegraph equation with $\beta = \gamma = 0$ and $\alpha = 1$;

f. Let $p(x, \mu, t)$ denote the one dimensional Exponential density, i.e., $p(x, \mu, t) = te^{-t(x-\mu)}$, $x > \mu$. Then $p(x, \mu, t)$ satisfies the telegraph equation with $\alpha = \beta = 0$ and $\gamma(t) = t^2$.

Proof: The proof of Proposition 1 involves straightforward calculations and is omitted.

2.2.3. Resulting Identities

Proposition 1 leads to expectation identities for the corresponding densities. The formal identities are presented below.

Theorem 1.

a. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$. Let $g(\underset{\sim}{x}, \mu)$ be a twice continuously differentiable function and suppose $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ and $||\nabla g(\underset{\sim}{x}, \underset{\sim}{\mu})||$ are $O(e^{c||\underset{\sim}{x}||})$ for some $0 \le c < \infty$. Then

$$\frac{\partial}{\partial t} E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = \frac{1}{2} E(\Delta_x g(\underset{\sim}{x}, \underset{\sim}{\mu})) \tag{3}$$

Identity (3) will be referred to as the **Heat Equation Identity**.

b Let $\underset{\sim}{x}$ be a Gaussian convolution in the sense of Proposition 1. Let $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ be as in part a. Then identity (3) holds.

c. Let $\underset{\sim}{x} \sim C_p(\underset{\sim}{\mu}, tI)$. Let $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ be twice continuously differentiable and suppose $g(\underset{\sim}{x}, \underset{\sim}{\mu}) = O(||\underset{\sim}{x}||^{1-\epsilon})$ for some $\epsilon > 0$. Then

$$\frac{\partial^2}{\partial t^2} E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = -E\Delta_x(g(\underset{\sim}{x}, \underset{\sim}{\mu})). \tag{4}$$

Identity (4) will be referred to as the **Wave Equation Identity.**

d. Let $\underset{\sim}{x}$ be a Cauchy convolution in the sense of Proposition 1. Let $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ be as in part c. Then identity (4) holds.

e. Let x have the mixture density $wf(x-t) + (1-w)f(x+t)$ and suppose f is twice differentiable and suppose $\lim_{|x|\to\infty} g(x)f'(x) = \lim_{|x|\to\infty} g'(x)f(x) = 0$. Then

$$\frac{\partial^2}{\partial t^2} E(g(x)) = E(g''(x)). \tag{5}$$

Identity (5) will be referred to as the **Mixture Identity.**

f. Let x have the Exponential density in the sense of Proposition 1. Let $g(x, \mu)$ be twice continuously differentiable and suppose $g(x, \mu) = O(x^k)$ for some $0 \leq k < \infty$. Then

$$E(g(x,\mu)) = g(\mu,\mu) + \frac{1}{t} g_x(\mu,\mu) + \frac{1}{t^2} E(g_{xx}(x,\mu)). \tag{6}$$

Identity (6) will be referred to as the **Exponential Identity**.

2.2.4. The Proofs

The proofs of the various parts of Theorem 1 are similar. We will only present the proof of part a.

Proof of Theorem 1:

a. The $N_p(\underset{\sim}{\mu}, tI)$ density will be denoted as $p(\underset{\sim}{x}, \underset{\sim}{\mu}, t)$ in this proof.

Step 1. By an interchange of the order of differentiation and integration and by use of the heat equation (part a. of Proposition 1),

$$\frac{\partial}{\partial t} E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = \frac{1}{2} \int (g(\underset{\sim}{x}, \underset{\sim}{\mu})) \Delta_x p(\underset{\sim}{x}, \underset{\sim}{\mu}, t) d\underset{\sim}{x}. \tag{7}$$

Step 2. By Green's second identity, for any sphere $B(\underset{\sim}{0}, r)$,

$$\int_{B(\underset{\sim}{0},r)} (g\Delta_x p - p\Delta_x g) d\underset{\sim}{x} = \int_{\partial B(\underset{\sim}{0},r)} (g\nabla p - p\nabla g)' n d\sigma, \tag{8}$$

where n denotes the unit outer normal.

Step 3. By Schwartz's inequality and the fact that $||n|| = 1$,

$$\int_{\partial B(\underset{\sim}{O},r)} (g\nabla p - p\nabla g)' n d\sigma$$

$$\leq \int_{\partial B(\underset{\sim}{O},r)} |g| \, ||\nabla p|| d\sigma + \int_{\partial B(\underset{\sim}{O},r)} p||\nabla g|| d\sigma$$

$$\leq A e^{cr} \int_{\partial B(\underset{\sim}{O},r)} ||\nabla p|| d\sigma + B e^{cr} \int_{\partial B(\underset{\sim}{O},r)} p d\sigma, \tag{9}$$

for some constants $0 \leq A, B, c < \infty$, by the assumptions made on g.

Step 4. From (7), write $\frac{\partial}{\partial t} E(g(\underset{\sim}{x}, \underset{\sim}{\mu}))$ as

$$\frac{\partial}{\partial t} E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = \frac{1}{2} \int (p \Delta_x g) d\underset{\sim}{x} + \frac{1}{2} \int (g \Delta_x p - p \Delta_x g) d\underset{\sim}{x}. \tag{10}$$

Step 5. Finally,

$$\int (g \Delta_x p - p \Delta_x g) d\underset{\sim}{x}$$

$$= \lim_{r \to \infty} \int_{B(\underset{\sim}{O},r)} (g \Delta_x p - p \Delta_x g) d\underset{\sim}{x}$$

$$= 0 \quad \text{by} \quad (8) \text{ and } (9).$$

2.3. Identities From the Fokker-Planck Equation

2.3.1. The Fokker-Planck Equation.

Consider a homogeneous Markov process $\{X_t, t \geq 0\}$ with state space equal to the real line and the transition probability $P(\mu, t, B) = P(X_{s+t} \in B | X_s = \mu)$, where B is a Borel set in the real line. Suppose that the process $\{X_t\}$ satisfies the three conditions

$$\left.\begin{aligned} &\lim_{t\to 0} \tfrac{1}{t} P_x(|X_t - x| > \delta) = 0 \text{ for any } \delta > 0, \\ &\lim_{t\to 0} \tfrac{1}{t} E_x(X_t - x) = b(x), \\ &\lim_{t\to 0} \tfrac{1}{t} E_x(X_t - x)^2 = \sigma^2(x), \end{aligned}\right\} \quad (11)$$

for suitable finite functions $b(\cdot)$ and $\sigma^2(\cdot)$. That is, the Markov process $\{X_t\}$ is actually a homogeneous diffusion with drift $b(\cdot)$ and diffusion coefficient $\sigma^2(\cdot)$. If the transition probability measure $P(\mu, t, \cdot)$ has a density, say $p(\mu, t, x)$, then it satisfies the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(b(x)p) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(\sigma^2(x)p). \quad (12)$$

In fact, all derivatives in (12) are assumed to exist and be continuous in (t, x). (12) is also known as the Kolmogorov forward differential equation. And so the question is what would (12) give us. It turns out that for appropriate choices of the drift $b(\cdot)$ and the diffusion coefficient $\sigma^2(\cdot)$, the corresponding transition density p is often a statistically important density; we will see a number of examples. Therefore, for such a density p, and a statistic, $g(x)$, one formally gets the expectation identity

$$\frac{\partial}{\partial t} E_\mu(g(x)) = E_\mu(g'(x)b(x)) + \frac{1}{2} E_\mu(g''(x)\sigma^2(x)). \quad (13)$$

In a statistical context, μ and t in the identity (13) will play the role of parameters. The multivariate analog of (12) for vector diffusion processes is

$$\frac{\partial p}{\partial t} = -p\nabla.\underset{\sim}{b} - \underset{\sim}{b}'\nabla p + \frac{1}{2}\sum_i\sum_j \frac{\partial^2}{\partial x_i \partial x_j}(\sigma_{ij}^2(\underset{\sim}{x})p) \quad (14)$$

for appropriate $\underset{\sim}{b}_{p\times 1}$ and $\Sigma_{p\times p} = ((\sigma_{ij}))$. The interest of (14) in the context of this article is on two grounds; first, (14) will result in interesting expectation identities for interesting statistical models, and second, many of these identities will be seen to also follow from the telegraph equation identity (2), thus showing the mutual connections.

For existence and methods of solutions of the Kolmogorov differential equation for given $\underset{\sim}{b}$ and Σ, one may see Feller (1936) and Risken (1984), among many sources.

2.3.2. Densities Satisfying the Fokker-Planck Equation

We now present some common density functions satisfying the Fokker-planck equation (14).

Proposition 2.

a. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$. Then the density of $\underset{\sim}{x}$ satisfies the Fokker-Planck equation with $\underset{\sim}{b} = \underset{\sim}{0}$ and $\Sigma = I$;

b. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\Sigma_0)$. Then the density of $\underset{\sim}{x}$ satisfies the Fokker-Planck equation with $\underset{\sim}{b} = \underset{\sim}{0}$ and $\Sigma = \Sigma_0$;

c. Let x have a density in the Pearson family of distributions with parameters. Then the density of X satisfies the Fokker-Planck equation with $b = b(x) = \alpha + \beta x$ and $\sigma^2 = \sigma^2(x) = a + bx + cx^2$;

d. Let $x \geq 0$ have the density $\frac{x}{\mu\sqrt{2\pi t}}(e^{-\frac{1}{2t}(x-\mu)^2} - e^{-\frac{1}{2t}(x+\mu)^2})$,

where $\mu > 0$. The density of x satisfies the Fokker-Planck equation with $b = 1$ and $\sigma^2 = \sigma^2(x) = x$.

e. Let $x \sim N(\mu + \lambda t, \sigma_0^2 t)$. Then the density of x satisfies the Fokker-Planck equation with $b = \lambda$ and $\sigma^2 = \sigma_0^2$.

Proof: Again, the proof of Proposition 2 involves straightforward calculations and is omitted.

2.3.3. Resulting Identities

The formal expectation identities resulting from Proposition 2 are presented below.

Theorem 2.

a. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$ and let $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ be as in part a. of Theorem 1. Then identity (3) holds.

b. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\Sigma_0)$ and let $g(\underset{\sim}{x}, \underset{\sim}{\mu}, \Sigma_0)$ be as in part a. above. Then

$$\frac{\partial}{\partial t} Eg(\underset{\sim}{x}, \underset{\sim}{\mu}, \Sigma_0) = \frac{1}{2} Etr(H_g \Sigma_0). \tag{15}$$

Identity (15) will be referred to as the **Generalized Heat Equation Identity**.

c. Let x have a density in the Pearson family of distributions with parameters.

Identity (16) will be referred to as the **Pearson Family Identity**.

d. Let $x \geq 0$ have the density $\frac{x}{\mu\sqrt{2\pi t}}(e^{-\frac{1}{2t}(x-\mu)^2} - e^{-\frac{1}{2t}(x+\mu)^2})$, where $\mu > 0$.

Let $g(x, \mu)$ and $g_x(x, \mu)$ be $O(e^{cx})$ for some $0 \leq c < \infty$. Then

$$\frac{\partial}{\partial t}E(g(x, \mu)) = E(g_x(x, \mu) + \frac{1}{2}xg_{xx}(x, \mu)). \tag{17}$$

e. Let $x \sim N(\mu + \lambda t, \sigma_0^2 t)$ and let $g(x, \mu, \lambda, \sigma_0) = O(e^{c|x|})$ for some $0 \leq c < \infty$.

Then

$$\frac{\partial}{\partial t}E(g(x, \mu, \lambda, \sigma_0)) = \lambda E(g_x(x, \mu, \lambda, \sigma_0)) + \frac{\sigma_0^2}{2}E(g_{xx}(x, \mu, \lambda, \sigma_0)). \tag{18}$$

More generally, if $x \sim N(\mu(t), \sigma_0^2(t))$, then for g as above,

$$\frac{\partial}{\partial t}E(g(x)) = \mu'(t)E(g'(x)) + \sigma_0(t)\sigma_0'(t)E(g''(x)) \tag{19}$$

Identity (19) will be referred to as the **Heteroscedastic Normal Identity.**

Proof: The proofs of all parts of Theorem 2 follow from the corresponding parts of Proposition 2 and are omitted.

2.4. Identities From Itô's Theorem

2.4.1. Itô's Formula.

Given the drift function and the diffusion coefficient, the Kolmogorov differential equation determines the transition density of a homogeneous diffusion under certain conditions. However, solving the equation is simply very hard. A more direct approach, initiated by Paul Lèvy, was given in Itô (1951). Itô showed that a diffusion is driven by a stochastic differential equation and a smooth function of a diffusion is driven by another stochastic differential equation.

Proposition 3 (Itô's Theorem). Let m, k, p be fixed positive integers. Let W_t denote an m-dimensional Wiener process and let $b(t, w) : (0, \infty) \times \Omega \to \mathbb{R}^p$ be a nonanticipating stochastic process. Let $G(t,w) = ((G_{ij}))p \times m$ be another matrix valued nonanticipating process such that for every t, $\int_0^t G_{ij}(s,w)ds < \infty$ *a.s.* for all pairs i, j.

Suppose x_t is a p-dimensional stochastic process satisfying the stochastic differential equation $dx_t = b(t)dt + G(t)dW_t$. If $g(t, \underset{\sim}{x}) : (0,\infty) \times \mathbb{R}^p \to \mathbb{R}^k$ has two continuous partial derivatives, then $Y_t = g(t, x_t)$ satisfies the following differential equation:

$$dY_t = g_x(t,x_t)G(t)dW_t + \{g_t(t,x_t) + g_x(t,x_t)b(t) + \frac{1}{2}\sum_{i,j=1}^{p}\sum g_{x_i x_j}(t,x_t)((GG'))_{ij}\}dt. \quad (20)$$

In the above,

$g_t = \frac{\partial}{\partial t} g$ is a k-dimensional vector;

$g_x = ((\frac{\partial}{\partial x_j} g_i))$ is a $k \times p$ dimensional matrix;

$g_{x_i x_j} = \frac{\partial^2}{\partial x_i \partial x_j} g$ is also a k-dimensional vector.

2.4.2. Resulting Identities

We can now give the following general expectation identity.

Theorem 3. With the notation of Proposition 3,

$$(\frac{d}{dt} Eg(t,x_t)) = E\left[g_t(t,x_t) + g_x(t,x_t)b(t) + \frac{1}{2}\sum_{i,j=1}^{p}\sum g_{x_i x_j}(t,x_t)((GG'))_{ij}\right]. \quad (21)$$

In particular, if $k = 1$ (i.e. if g is a scalar function), then

$$\frac{d}{dt} E(g(t,x_t)) = E[g_t(t,x_t) + (\nabla_x g)'b \; + \frac{1}{2} tr(H_g GG')]. \quad (22)$$

Proof: From (20),

$$g(t,x_t) - g(0,x_0)$$
$$= \int_0^t g_x(s,x_s)G(s)dW_s + \int_0^t [g_s(s,x_s) + g_x(s,x_{sx})b(s) + \frac{1}{2}\sum_{i,j=1}^{p}\sum$$
$$gx_i x_j(s,x_s)((GG'))_{ij}]ds \quad (23)$$

Note that, in (23), $\int_0^t g_x(s,x_s)G(s)dW_s$ has expectation zero by virtue of independent increments of the paths of a Wiener process and so (21) follows from (23).

Theorem 3 is a general result which is now applied to interesting special cases to cover some statistically interesting models.

Theorem 4.

a. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$ and let $g(t, \underset{\sim}{x}, \underset{\sim}{\mu})$ be twice continuously differentiable. Then subject to the existence of each integral,

$$\frac{d}{dt}E(g(t, \underset{\sim}{x}, \underset{\sim}{\mu})) = E(g_t(t, \underset{\sim}{x}, \underset{\sim}{\mu}) + \frac{1}{2}\Delta_x g(t, \underset{\sim}{x}, \underset{\sim}{\mu})). \tag{24}$$

Identity (24) will be referred to as the **Wiener Process Identity.**

b. Let $x \sim t\chi^2(p)$ and let $g(x)$ be twice continuously differentiable. Then, subject to the existence of each integral,

$$\frac{d}{dt}E(g(x)) = pE(g'(x)) + 2E(xg''(x)). \tag{25}$$

Identity (25) will be referred to as the **Chi-square Identity.**

c. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\Sigma_0)$ and let $g(t, \underset{\sim}{x}, \underset{\sim}{\mu}, \Sigma_0)$ be twice continuously differentiable. Then subject to the existence of each integral,

$$\frac{d}{dt}E(g(t, \underset{\sim}{x}, \underset{\sim}{\mu}, \Sigma_0)) = E(g_t(t, \underset{\sim}{x}, \underset{\sim}{\mu}, \Sigma_0) + \frac{1}{2}tr(H_g\Sigma_0)). \tag{26}$$

Identity (26) will be referred to as the **Scaled Wiener Process Identity.**

Proof:

a. In Theorem 3, take $k = 1, b = 0$, and $G = I_p$.

b. Part b. follows from part a. on taking $\underset{\sim}{\mu} = \underset{\sim}{0}$ and $g(t, \underset{\sim}{x}, \underset{\sim}{\mu})$ to be $g(\sum_{i=1}^{p} x_i^2)$.

c. In Theorem 3, take, $k = 1, b = 0$, and $G = \Sigma_0^{\frac{1}{2}}$.

2.5. Two Other Important Problems

Two statistically important cases not covered by the preceding sections are handled here. These are: a. the case where $x_1, \ldots, x_n$ are iid univariate normal with the mean and the variance being both treated as unknown parameters, and b. the case of time series data where $x_1, \ldots, x_n$ have a common mean μ but are not independent. In fact, a single general result covers both cases and we present that below.

2.5.1. The Identity

Theorem 5. Let $c, m \geq 0$. Suppose $x \sim N(\mu, ct), y \sim t\chi^2(m)$, and suppose $x, \underline{y}$ are independent. Let $g(x, y, \mu)$ satisfy the following conditions;

i. g is twice continuously differentiable in x and once continuously differentiable in y;

ii. g, g_x are each $0(e^{a|x|}y^k)$ for some $0 \leq a, k < \infty$. Then

$$\frac{\partial}{\partial t}E(g(X,Y,\mu)) = \frac{c}{2}E(g_{xx}(X,Y,\mu)) + \frac{1}{t}E(Yg_y(X,Y,\mu)). \tag{27}$$

Identity (27) will be referred to as the **Canonical Normal Identity**.

Proof:

Step 1. Let $p_1(x, \mu, t)$ denote the density of x and let $p_2(y, t)$ denote the density of y. Then,

$$\frac{\partial}{\partial t}p_1 = \frac{c}{2}\frac{\partial^2}{\partial x^2}p_1, \tag{28}$$

$$\text{and} \quad \frac{\partial}{\partial t}p_2 = \frac{y}{2t^2}p_2 - \frac{m}{2t}p_2 = -\frac{y}{t}\frac{\partial}{\partial y}p_2 - \frac{p_2}{t}, \tag{29}$$

on some calculations.

Step 2. Since x, y are independent, the joint density is given by

$$p(x,y,\mu,t) = p_1(x,\mu,t)p_2(y,t).$$

Multiplying both sides by $g(x, y, \mu)$, (27) follows on integration after some algebra.

Remark. If $x_1, \ldots, x_n$ are iid $N(\mu, t)$, then (x, y) are to be understood as the jointly sufficient statistic $(\overline{x}, \sum_{i=1}^{n}(x_i - \overline{x})^2)$. On the other hand, if $x_1, \ldots, x_n$ are jointly normal each with mean μ and the covariance matrix $t\Sigma_0$, then (x, y) are to be understood as $\left(\frac{\underset{\sim}{1}'\Sigma_0^{-1}\underset{\sim}{x}}{\underset{\sim}{1}'\Sigma_0^{-1}\underset{\sim}{1}}, \underset{\sim}{x}'\Sigma_0^{-1}\underset{\sim}{x} - \frac{(\underset{\sim}{1}'\Sigma_0^{-1}\underset{\sim}{x})^2}{\underset{\sim}{1}'\Sigma_0^{-1}\underset{\sim}{1}}\right)$, again the jointly sufficient statistic. Note that c is to be taken as $\frac{1}{\underset{\sim}{1}'\Sigma_0^{-1}\underset{\sim}{1}}$ in this case.

3. FROM PDES TO UNBIASED ESTIMATION

Of the large number of expectation identities presented in Section 2, some particular ones have some interesting implications in the theory of unbiased estimation. The typical result we will present will either characterize an unbiasedly estimable parametric function or characterize parametric functions unbiasedly estimable by statistics of relevant natural form. We would note here that there is a body of literature on existence of unbiased estimates in special and general problems. One may see in particular Brown and Liu (1993) and Doss and Sethuraman (1989).

3.1. Multivariate Normal and Multivariate Cauchy

First we state a convention subsequently assumed in the results of this section.

Convention. For any result specific to a given distribution in this section, by a statistic $g(\underset{\sim}{X})$ we shall mean a function $g(\underset{\sim}{X})$ which satisfies the smoothness and growth conditions previously imposed on g in Section 2 for the relevant expectation identity to hold.

3.1.1. Multivariate Normal

Theorem 6. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$. Let $h = h(\underset{\sim}{\mu}, t)$ be a twice continuously differentiable parametric function.

a. If $h(\underset{\sim}{\mu}, t)$ has an unbiased estimate $g(\underset{\sim}{x})$, then h must satisfy the heat equation $\frac{\partial}{\partial t}h = \frac{1}{2}\Delta_\mu h$.

b. Conversely, if h satisfies the heat equation and if $\lim\limits_{t\to 0+} h(\underset{\sim}{\mu}, t) = g(\underset{\sim}{\mu})$ exists, then $g(\underset{\sim}{x})$ is an unbiased estimate of h provided $E(g(\underset{\sim}{x}))$ exists.

Proof:

a. If $g(\underset{\sim}{x})$ is an unbiased estimate of $h(\underset{\sim}{\mu}, t)$, then, by the Heat Equation Identity (3),

$$\frac{d}{dt}h = \frac{1}{2}E(\Delta_x g). \tag{30}$$

However, due to the location parameter structure, by a change of variable one has

$$\begin{aligned} E(\Delta_x g) &= \int (\Delta_x g)(\underset{\sim}{x}) p(\underset{\sim}{x}, \underset{\sim}{\mu}, t) d\underset{\sim}{x} \\ &= \int (\Delta_x g)(\underset{\sim}{x} + \underset{\sim}{\mu}) p(\underset{\sim}{x}, \underset{\sim}{0}, t) d\underset{\sim}{x} \end{aligned}$$

$$= \int (\Delta_\mu g)(\underset{\sim}{x} + \underset{\sim}{\mu}) p(\underset{\sim}{x}, \underset{\sim}{0}, t) d\underset{\sim}{x}$$

$$= \Delta_\mu \left(\int g(\underset{\sim}{x} + \underset{\sim}{\mu}) p(\underset{\sim}{x}, \underset{\sim}{0}, t) d\underset{\sim}{x} \right)$$

$$= \Delta_\mu h; \tag{31}$$

substitution into (30) gives part a.

b. Part b. is a restatement of the result that a solution of the problem $\frac{\partial}{\partial t} h = \Delta_\mu h$ under the initial value condition $\lim_{t \to 0+} h(\mu, t) = g(\mu)$ is $h(\mu, t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}(\mu - x)^2} g(x) dx$; see Folland (1992).

We will now describe some interesting consequences of the above result. Parts b. and c. can be anticipated, but we give a formal proof.

Corollary 1.

a. Let $h = h(\underset{\sim}{\mu})$ be a twice continuously differentiable function of $\underset{\sim}{\mu}$. h is unbiasedly estimable only if h is harmonic, in which case it is self-estimable, i.e., $E(h(\underset{\sim}{x})) = h(\underset{\sim}{\mu})$, provided $E(h(\underset{\sim}{x}))$ exists.

b. Let $h = h(\underset{\sim}{\mu})$ be a twice continuously differentiable function of $\underset{\sim}{\mu}$. Suppose h is unbiasedly estimable. Then,

   i. If h is radial, i.e., $h = h(||\underset{\sim}{\mu}||)$, then h must be a constant;

   ii. If h is bounded, then h must be a constant;

   iii. If h is integrable, then $h \equiv 0$;

   iv. If $|h(\underset{\sim}{\mu})| \leq a + b||\underset{\sim}{\mu}||$ for some $a, b \geq 0$, then h must be linear; similarly, if $|h(\underset{\sim}{\mu})| \leq a + b||\underset{\sim}{\mu}||^2$ for some $a, b \geq 0$, then h must be a quadratic.

c. Let $h = h(t)$ be once continuously differentiable. h is unbiasedly estimable only if it is a constant function.

d. Let $p = 1$ and let $h = h(\mu, t)$ be a function of the form $f(\mu + c\sqrt{t})$ for some constant $c \neq 0$. If f is twice continuously differentiable and $f'(0) \neq 0$, then h is not unbiasedly estimable. In particular, the quantiles of x are not unbiasedly estimable.

e. Let $p = 1$ and let $h = h(\mu, t)$ be a bivariate polynomial in the mean μ and the standard deviation $\sqrt{t}$ (i.e., $h(\mu, t) = \sum_{j=0}^{k} c_j \mu^j (\sqrt{t})^{k-j}$ for some k and constants c_j). Then $h(\mu, t)$ is unbiasedly estimable if and only if $h(\mu, t)$ is a multiple of $E(x^k)$.

Proof:

a. This follows form part a. of Theorem 6, for $\frac{\partial}{\partial t} h = 0$, and hence $\Delta_\mu h = 0$. On the other hand, $\lim_{t \to 0+} h(\underset{\sim}{\mu}) = h(\underset{\sim}{\mu})$ and so h is self-estimable.

b. For i. by part a., it follows that the function $h(\cdot)$ must satisfy the differential equation $h''(z) + (p-1)\frac{h'(z)}{z} = 0$ at all $z > 0$ if $h(||\underset{\sim}{\mu}||)$ is unbiasedly estimable.

For $p = 2$, this makes $h(z) = a \log z + b$ for $z > 0$ and for $p \geq 3$, this makes $h(z) = az^{2-p} + b$ for $z > 0$, and so one cannot have $h(||\underset{\sim}{\mu}||)$ to be in $\mathcal{C}^2(\mathbb{R}^p)$ unless h is a constant.

ii. follows from part a. and the fact that the only bounded harmonic functions are constants;

iii. note that an integrable harmonic function must be identically zero; see Rudin (1974) and also Proposition 8.1 in Axler, Bourdon and Ramey (1992).

Finally, for iv., by taking $\text{A} = \max(a, b)$, we have, respectively, $|h(\underset{\sim}{\mu})| \leq \text{A}(1 + ||\underset{\sim}{\mu}||^i), i = 1, 2$, and the assertion follows from the fact that a harmonic function with this property is necessarily a polynomial of degree i (e.g., see Axler, Bourdon and Ramey (1992)).

c. Again, as $\Delta_\mu h = 0$ now, $\frac{\partial}{\partial t} h$ is also 0.

d. Suppose $f(\mu + c\sqrt{t})$ was unbiasedly estimable. By Theorem 6,

$$\frac{c}{2\sqrt{t}} f'(\mu + c\sqrt{t}) = \frac{\partial}{\partial t} h = h_{\mu\mu} = f''(\mu + c\sqrt{t})$$

$$\Rightarrow \quad cf'(\mu + c\sqrt{t}) = 2\sqrt{t} f''(\mu + c\sqrt{t}). \tag{32}$$

Consider now (μ, t) lying on the one-dimensional curve $\mu = -c\sqrt{t}$. Then, from (32), $cf'(0) = 2\sqrt{t} f''(0)$. This forces $f''(0)$ to be not 0, implying t to be $\frac{(cf'(0))^2}{4(f''(0))^2}$, a constant, thus a contradiction.

Since the quantiles of x are of the form $\mu + \Phi^{-1}(p)\sqrt{t}$ for some p, it follows that they are not unbiasedly estimable.

e. Let $h(\mu,t) = \sum_{j=0}^{k} c_j \mu^j (\sqrt{t})^{k-j}$ be unbiasedly estimable and suppose without loss of generality that $c_k = 1$. By Theorem 6,

$$\frac{\partial}{\partial t} h = \frac{k}{2} c_0 t^{\frac{k}{2}-1} + c_1 \frac{k-1}{2} t^{\frac{k-3}{2}} \mu + \sum_{i=2}^{k} c_i \mu^i \frac{k-i}{2} (\sqrt{t})^{k-i-2}$$

$$= \frac{1}{2} \frac{\partial^2}{\partial \mu^2} h$$

$$= \frac{1}{2} \sum_{i=0}^{k-2} (i+1)(i+2) c_{i+2} \mu^i (\sqrt{t})^{k-i-2}. \tag{33}$$

By comparing coefficients of the powers of μ on the two sides of (33), one gets $c_{k-1} = c_{k-3} = \ldots = 0$ and $c_{k-2j} = \frac{(k-2j+1)(k-2j+2)}{2j} c_{k-2j+2}, j \geq 1$. As $c_k = 1$, it follows that $c_{k-2j} = \frac{k!}{(k-2j)!2^j j!} = \binom{k}{2j} \frac{(2j)!}{2^j j!} = \binom{k}{2j} E(\frac{X-\mu}{\sqrt{t}})^{2j} (\sqrt{t})^{2j}$. Hence, $h(\mu,t) = E(X^k)$ if $c_k = 1$.

Remark. Theorem 6 is easily generalized to the case $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\Sigma_0)$. For instance, $h(\underset{\sim}{\mu})$ is unbiasedly estimable if and only if $h(\Sigma_0^{\frac{1}{2}} \underset{\sim}{\mu})$ is harmonic.

Example 1. By part a. of Corollary 1, smooth harmonic functions of $\underset{\sim}{\mu}$ are unbiasedly self estimable. Simple nonlinear examples are $h(\mu_1, \mu_2) = e^{\mu_1}(a \text{ Sin } \mu_2 + b \text{ Cos } \mu_2), 3\mu_1\mu_2^2 - \mu_1^3$, or any "quadratic contrast" $\sum_{i=1}^{p} a_i \mu_i^2$ in p dimensions with $\sum_{i=1}^{p} a_i = 0$.

3.1.2. Multivariate Cauchy

Proposition 4. Let $\underset{\sim}{x} \sim C_p(\underset{\sim}{\mu}, tI)$. Let $h(\underset{\sim}{\mu}, t)$ be a twice continuously differentiable parametric function.

a. If $h(\underset{\sim}{\mu}, t)$ has an unbiased estimate $g(\underset{\sim}{x})$, then h must satisfy the wave equation $\frac{\partial^2}{\partial t^2} h + \Delta_\mu h = 0$.

b. Conversely, if h satisfies the wave equation and if $\lim_{t \to 0+} h(\underset{\sim}{\mu}, t) = g(\underset{\sim}{\mu})$ exists, then $g(\underset{\sim}{x})$ is an unbiased estimate of h provided $E(g(\underset{\sim}{x}))$ exists.

Proof: The argument is similar to that of Theorem 6 and is omitted.

Analogous to Corollary 1, one gets the following consequences of Proposition 4.

Corollary 2.

a. Parts a., b., and c. of Corollary 1 hold.

b. Let $p = 1$ and let $h(\mu, t)$ be a bivariate polynomial in μ and t (not $\sqrt{t}$) as defined in Corollary 1. Then $h(\mu, t)$ is not unbiasedly estimable.

Remark: It is therefore the case that harmonic functions $h(\underset{\sim}{\mu})$ of $\underset{\sim}{\mu}$ are unbiasedly estimable under both models as long as $E(h(\underset{\sim}{x}))$ exists. Theorem 6 and Proposition 4 show that there are also plenty of functions unbiasedly estimable under one model, but not under the other model.

3.1.3. The Canonical Normal Case

Recall from Section 2.5 that the canonical normal case is when we have $x \sim N(\mu, ct), y \sim t\chi^2(m)$ and x, y are independent. Some interesting assertions can be made regarding unbiased estimability in this case also. They are driven by identity (27).

Theorem 7. Let $c, m > 0$ and suppose $x \sim N(\mu, ct), y \sim t\chi^2(m)$, and x, y are independent. Let $h(\mu, t)$ be twice continuously differentiable in μ and once in t.

a. If $h(\mu, t) = h(\mu)$, then it is unbiasedly estimable by a function $g(x)$ of x alone if and only if h is linear in μ.

b. If $h(\mu, t) = h(\mu)$, then it can be unbiasedly estimated by the extended class of functions $g_1(x) + g_2(y)$ if and only if h is a quadratic in μ. Furthermore, $g_1(x)$ has to be a quadratic in x and $g_2(y)$ has to be linear in y.

c. If $h(\mu, t) = h_1(\mu) + h_2(t)$, then it can be unbiasedly estimated by a function $g_1(x) + g_2(y)$ only if h_1 is a quadratic in μ. Furthermore, $g_1(x)$ has to be a quadratic in x, but there is no further constraint on $g_2(y)$.

Proof: For each part, the key step is to rewrite the canonical normal identity (27) as

$$\frac{\partial}{\partial t}E(g(x, y)) = \frac{c}{2}\,\frac{\partial^2}{\partial \mu^2}E(g(x, y)) + \frac{1}{t}E(y\,\frac{\partial}{\partial y}\,g(x, y)). \tag{34}$$

a. This is an immediate consequence of the above identity (34).

b. Again, by (34),

$$0 = \frac{c}{2}h''(\mu) + \frac{1}{t}E(yg_2'(y))$$

$$\Rightarrow \frac{1}{t}E(yg_2'(y)) = -\frac{c}{2}h''(\mu). \tag{35}$$

The RHS of (35) is a function of μ and the LHS a function of t. Consequently, each side is a constant function, which shows h to be a quadratic.

Hence, from (35) again, $E(y\ g_2'(y)) = at$ for some constant 'a', which by standard completeness arguments forces $g_2'(y)$ to be a constant.

Now, therefore, for constants $\alpha, \beta, \gamma, \eta, \delta$,

$$h(\mu) = \alpha\mu^2 + \beta\mu + \gamma = E(g_1(x)) + E(\delta y + \eta), \tag{36}$$

and hence $g_1(x)$ is a quadratic in x by another completeness argument.

c. The proof of this is quite similar to that of b. and we shall omit it.

4. Analysis of the Heat Equation Identity

The heat equation identity (3) of Section 2 showed that under certain conditions on a function $g(\underset{\sim}{x})$, $\frac{\partial}{\partial t}E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = \frac{1}{2}E(\Delta_x g(\underset{\sim}{x}, \underset{\sim}{\mu}))$ if $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$. Stein (1973, 1981) showed that for a vector-valued function $\underset{\sim}{h}(\underset{\sim}{x}, \underset{\sim}{\mu}))$ satisfying certain conditions, $E((\underset{\sim}{x} - \underset{\sim}{\mu})'\underset{\sim}{h}(\underset{\sim}{x}, \underset{\sim}{\mu})) = tE(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{x}, \underset{\sim}{\mu}))$ if $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$; this is known as Stein's identity. It is known that Stein's identity characterizes the normal distribution in an appropriately precise sense; one may see and Diaconis and Zabell (1991). The following two questions, therefore, emerge naturally:

Question 1. Does the heat equation identity characterize the normal distribution in any precise sense?

Question 2. Is there a connection between Stein's identity and the heat equation identity?

We shall now address these two questions.

4.1. Characterization of the Normal Distribution

The characterization results below are for the one dimensional case. For part a., extension to the multivariate spherically symmetric case is apparent. For part b., however, we do not presently have a multivariate analog.

Theorem 8. Let $\mathcal{C}^m(\mathbb{R})$ be the class of m times continuously differentiable functions and $\mathcal{C}_0^\infty(\mathbb{R})$ be the class of infinitely differentiable functions with compact support.

a. Suppose x has a location scale parameter density $\frac{1}{\sqrt{t}}p(\frac{x-\mu}{\sqrt{t}})$ and $p(\cdot)$ is in $\mathcal{C}^2(\mathbb{R})$. Suppose the heat equation identity $\frac{d}{dt}E(g(x)) = \frac{1}{2}E(g''(x))$ holds for all g in $\mathcal{C}_0^\infty(\mathbb{R})$. Then $p(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

b. Let $x \sim p(x|t) = e^{-\frac{T(x)}{t}}\beta(t)h(x)$, where $h(x) > 0, T(x) \geq 0, t > 0, \beta$ belongs to $\mathcal{C}^1(\mathbb{R}_t)$, and the functions T, h belong to $\mathcal{C}^2(\mathbb{R})$. Suppose $\frac{d}{dt}E(g(x)) = \frac{1}{2}E(g''(x))$ for all g in $\mathcal{C}_0^\infty(\mathbb{R})$. Then $p(x|t)$ is the density of $N(\mu, t)$ for some constant μ.

Proof:

Step 1. We shall take μ to be 0.

a. Let g be an element of $\mathcal{C}_0^\infty(\mathbb{R})$. By hypothesis,

$$
\begin{aligned}
2\frac{d}{dt}Eg(x) &= -2\int g(x)\left\{\frac{1}{2t^{3/2}}p\left(\frac{x}{\sqrt{t}}\right) + \frac{x}{2t^{3/2}}\frac{1}{\sqrt{t}}p'\left(\frac{x}{\sqrt{t}}\right)\right\}dx \\
&= Eg''(x) \\
&= \int g''(x)\frac{1}{\sqrt{t}}p\left(\frac{x}{\sqrt{t}}\right)dx \qquad \forall t > 0
\end{aligned}
$$

Step 2. Therefore,

$$
\begin{aligned}
&-\int g(x)\left\{\frac{1}{t}p\left(\frac{x}{\sqrt{t}}\right) + \frac{x}{t^{3/2}}p'\left(\frac{x}{\sqrt{t}}\right)\right\}dx \\
&= \int g''(x)p\left(\frac{x}{\sqrt{t}}\right)dx \\
&= \int g(x)\frac{1}{t}p''\left(\frac{x}{\sqrt{t}}\right)dx \text{ (on integration by parts) } \forall t > 0. \qquad (37)
\end{aligned}
$$

Hence, by using $t = 1$,

$$
\begin{aligned}
&\int g(x)\{p''(x) + xp'(x) + p(x)\} = 0 \text{ for all } g \text{ in } \mathcal{C}_0^\infty(\mathbb{R}) \\
&\Rightarrow\ p''(x) + xp'(x) + p(x) = a.e. \text{ (and hence everywhere)} \qquad (38)
\end{aligned}
$$

Step 3. One solution of (38) is $p_1(x) = \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. By a direct application of Abel's identity (see, e.g., pp. 1132 in Gradsheyn and Ryzhik (1980)) one sees that a second linearly independent solution is $p_2(x) = \phi(x) \int_0^x e^{\frac{u^2}{2}} du$. Hence the general solution of (38) is of the form $\phi(x)(a + b \int_0^x e^{\frac{u^2}{2}} du)$, of which only the case $a = 1,\ b = 0$ corresponds to a probability density. This completes part a.

b. Step 1. Following the first few lines of part a., one gets after some algebra,

$$h(x)(T'(x))^2 - tT''(x)h(x) - 2tT'(x)h'(x) + t^2 h''(x)$$
$$= 2h(x)T(x) + 2t^2 \frac{\beta'(t)}{\beta(t)} h(x) \qquad \forall t, \forall n. \tag{39}$$

Step 2. On letting $t \to 0$, one therefore gets:

$$(T'(x))^2 - 2T(x) = 0, \tag{40}$$

and hence $T(x) = \frac{(x-\mu)^2}{2}$ for some constant μ

Step 3. Substituting $T(x) = \frac{(x-\mu)^2}{2}$ in (39) and setting $x = \mu$, one now gets:

$$-th(\mu) + t^2 h''(\mu) = 2t^2 \frac{\beta'(t)}{\beta(t)} h(\mu) \qquad \forall t > 0$$
$$\Rightarrow 2t \frac{\beta'(t)}{\beta(t)} = t \frac{h''(\mu)}{h(\mu)} - 1 \qquad \forall t > 0. \tag{41}$$

Step 4. From (41), it follows on separation of variables that it *must* be the case that $h''(\mu) = 0$ and consequently, $\frac{\beta'(t)}{\beta(t)} = -\frac{1}{2t}$, i.e., $\beta(t) = \frac{k}{\sqrt{t}}$ for some constant k.

Step 5. Since we already have $T(x) = \frac{(x-\mu)^2}{2}$, this now forces $h(x)$ to be a constant and $p(x|t)$ to be the $N(\mu, t)$ density. This completes b.

4.2. Relation to the Stein Identity

We show that the heat equation identity (3) is equivalent to Stein's identity in one dimension and in more than one dimension, they are equivalent if Stein's $\underset{\sim}{h}(\underset{\sim}{x}, \underset{\sim}{\mu})$ function is the gradient ∇g of some function g.

Theorem 9.

a. For every $p \geq 1$, Stein's identity $\Rightarrow$ Identity (3).

b. For $p = 1$, Identity (3) $\Rightarrow$ Stein's identity.

c. For $p > 1$, Identity (3) $\Rightarrow$ Stein's identity if Stein's $\underset{\sim}{h}(\underset{\sim}{x}, \underset{\sim}{\mu}) = \nabla_x g(\underset{\sim}{x}, \underset{\sim}{\mu})$ for some g satisfying the growth conditions in part a. of Theorem 1.

Proof:

a. The proof given here is for $p = 1$, but with just a notational change, the same proof works for $p > 1$.

Step 1. Given g as in identity (3), define $h(x, \mu) = g_x(x, \mu)$. Then by Stein's identity,

$$tE(g_{xx}(x, \mu)) = tE(h_x(x, \mu)) = E((x - \mu)h(x, \mu)). \tag{42}$$

Step 2. However,

$$\begin{aligned} &E((x - \mu)h(x, \mu)) \\ &= \int (x - \mu)\frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2t}(x-\mu)^2} g_x(x, \mu)dx \\ &= \sqrt{t}\int \frac{1}{\sqrt{2\pi}}ze^{-\frac{1}{2}z^2} g_x(\mu + z\sqrt{t}, \mu)dz, \end{aligned} \tag{43}$$

by a change of variable.

Step 3. Write (43) as

$$2t\int \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\frac{\partial}{\partial t}g(\mu + z\sqrt{t}, \mu)dz = 2t + \frac{\partial}{\partial t}\int \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}g(\mu + z\sqrt{t}, \mu)dz \tag{44}$$

Step 4. Now, make the change the change of variable back to $x = \mu + z\sqrt{t}$, yielding (44)$= 2t\frac{\partial}{\partial t}E(g(x, \mu))$, hence establishing identity (3).

b. Step 1. Given h as in Stein's identity, define $g(x, \mu) = \int\limits_0^x h(u, \mu)du$.

Step 2. Thus, $E(g(x))$

$$= \int\limits_0^\infty\int\limits_0^x \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dudx - \int\limits_{-\infty}^0\int\limits_x^0 \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dudx$$

$$= \int\limits_0^\infty\int\limits_u^\infty \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dxdu - \int\limits_{-\infty}^0\int\limits_{-\infty}^u \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dxdu$$

$$= \int_0^\infty \left\{1 - \Phi\left(\frac{u-\mu}{\sqrt{t}}\right)\right\} h(u)du - \int_{-\infty}^0 \Phi\left(\frac{u-\mu}{\sqrt{t}}\right) h(u)du. \tag{45}$$

Step 3. By the heat equation identity,

$$\begin{aligned} \frac{d}{dt}E(g(x)) &= \frac{1}{2}E(g_{xx}(x)) \\ &= \frac{1}{2}E(h_x(x)). \end{aligned} \tag{46}$$

Step 4. Therefore, by (45),

$$\begin{aligned} \frac{1}{2}E(h_x(x)) &= \frac{d}{dt}\left[\int_0^\infty \left\{1 - \Phi\left(\frac{u-\mu}{\sqrt{t}}\right)\right\} h(u)du - \int_{-\infty}^0 \Phi\left(\frac{u-\mu}{\sqrt{t}}\right) h(u)du\right] \\ &= \int_{-\infty}^\infty \frac{u-\mu}{2t^{3/2}}\phi\left(\frac{u-\mu}{\sqrt{t}}\right) h(u)du, \end{aligned} \tag{47}$$

on differentiation.

But (47) $= \frac{1}{2t}E((x-\mu)h(x))$, yielding Stein's identity.

c. The same argument as in part b. applies on using the multivariate analog of the fundamental theorem of calculus, i.e., if $\underset{\sim}{x} \in \mathbb{R}^p$, if $\underset{\sim}{h} = \nabla g$ for some g in $\mathcal{C}^1(\mathbb{R}^p)$, and if $\sigma : [0,1] \to \mathbb{R}^p$ is a $\mathcal{C}^1$ path joining $\underset{\sim}{0}$ and $\underset{\sim}{x}$, then the line integral of h along σ satisfies $\int_\sigma h \cdot dS = g(\sigma(1) - \sigma(0))$ (see, e.g., Marsden and Tromba (1996)).

5. FIRST APPLICATIONS OF THE HEAT EQUATION IDENTITY

We now start to provide various applications of identity (3) and its generalization (15), the generalized heat equation identity. First we shall present a moment formula. This formula is simple to derive but an important one, in the sense that a very large number of the applications in this section and the subsequent sections will all flow from this moment formula.

5.1. A General Moment Formula

The moment formula given immediately below is for a function $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ when $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI), p \geq 1$. Since the moment formula is derived from the heat equation identity, g has to meet the assumptions of that identity. g has to satisfy one additional technical assumption that will almost always hold in applications.

Proposition 5. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI), p \geq 1$. Let $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ satisfy the assumptions of identity (3) and in addition assume that $E(|\Delta_x g(\underset{\sim}{x}, \underset{\sim}{\mu})|) < \infty$. Then

$$E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + e(\underset{\sim}{\mu}, t),$$

$$\text{where } \; e(\underset{\sim}{\mu}, t) = e_g(\underset{\sim}{\mu}, t) = \frac{1}{4\pi^{\frac{p}{2}}} \int (\Delta_x g) ||\underset{\sim}{x} - \underset{\sim}{\mu}||^{2-p} \Gamma(\frac{p}{2} - 1, \frac{||\underset{\sim}{x} - \underset{\sim}{\mu}||^2}{2t}) d\underset{\sim}{x}. \tag{48}$$

Proof:

Step 1. By identity (3), for $s > 0$, $\frac{\partial}{\partial s} E_{\underset{\sim}{\mu},s}(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = \frac{1}{2} E_{\underset{\sim}{\mu},s}(\Delta_x g)$ and it follows that for every fixed $\underset{\sim}{\mu}$, $\frac{\partial}{\partial s} E_{\underset{\sim}{\mu},s}(g(\underset{\sim}{x}, \underset{\sim}{\mu}))$ is continuous in s. Therefore it can be integrated to yield, by the Fundamental Theorem of calculus:

$$\begin{aligned} & E_{\underset{\sim}{\mu},t}(g(\underset{\sim}{x}, \underset{\sim}{\mu})) - g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) \\ &= \int_0^t \frac{\partial}{\partial s} E_{\underset{\sim}{\mu},s}(g(\underset{\sim}{x}, \underset{\sim}{\mu})) ds \\ &= \frac{1}{2} \int_0^t E_{\underset{\sim}{\mu},s}(\Delta_x g) ds. \end{aligned} \tag{49}$$

Step 2. Now,

$$\begin{aligned} & \int_0^t E_{\underset{\sim}{\mu},s}(\Delta_x g) ds \\ &= \int_0^t \int \frac{1}{(2\pi s)^{\frac{p}{2}}} e^{-\frac{1}{2s}||\underset{\sim}{x} - \underset{\sim}{\mu}||^2} (\Delta_x g) d\underset{\sim}{x} ds \\ &= \frac{1}{(2\pi)^{\frac{p}{2}}} \int (\Delta_x g) \int_0^t \frac{e^{-\frac{1}{2s}||\underset{\sim}{x} - \underset{\sim}{\mu}||^2}}{s^{p/2}} ds d\underset{\sim}{x}. \end{aligned} \tag{50}$$

It is for this application of Fubini's theorem that the additional assumption $E|\Delta_x g| < \infty$ is needed.

Step 3. Now if one transforms s to say $u = \frac{1}{s}$ in the inner integral, then formula (48) follows after a few lines of algebra.

The generalization to $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\Sigma_0)$ is the following.

Proposition 6. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\Sigma_0), p \geq 1$. Let $g(\underset{\sim}{x}, \underset{\sim}{\mu}, \Sigma_0)$ satisfy the assumptions of identity (15) and in addition assume that $E(|tr(H_g\Sigma_0)|) < \infty$. Then

$$E(g(\underset{\sim}{x}, \underset{\sim}{\mu}, \Sigma_0)) = g(\underset{\sim}{\mu}, \underset{\sim}{\mu}, \Sigma_0) + e(\Sigma_0, \underset{\sim}{\mu}, t),$$

where $$e(\Sigma_0, \underset{\sim}{\mu}, t) = e_g(\Sigma_0, \underset{\sim}{\mu}, t) = \frac{1}{4\pi^{\frac{p}{2}}|\Sigma_0|^{\frac{1}{2}}} \int tr(H_g\Sigma_0)((\underset{\sim}{x} - \underset{\sim}{\mu})'\Sigma_0^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu}))1 - \frac{p}{2}$$
$$\Gamma\left(\frac{p}{2} - 1, \frac{(\underset{\sim}{x} - \underset{\sim}{\mu})'\Sigma_0^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu})}{2t}\right) d\underset{\sim}{x}. \tag{51}$$

5.2. Reduction to Useful Forms

We will now show that the general moment formula (48) reduces to more useful forms. The reduction presented below is as follows:

a. for p =1,2, and 3, formula (48) will be used to establish a lower as well as an upper bound on $E(g(\underset{\sim}{x}, \underset{\sim}{\mu}))$ for general g as in Proposition 5. Among the immediate applications of these bounds are an improvement as well as reversal of Jensen's inequality for convex functions;

b. for $p \geq 4$, formula (48) will in fact be reduced to a considerably more useful exact form if p is even. This latter exact formula (formula (61) below) for $E(g(\underset{\sim}{x}, \underset{\sim}{\mu}))$ is a surprising reduction and leads to lower and upper bounds again. We will see in Section 6.1 the application of these to construction of simple to use but fully analytical bounds on the risks of formal Bayes estimates of $\underset{\sim}{\mu}$. The case of odd $p \geq 5$ will not be reported here, but the bounds (slightly complex) can be obtained similarly.

5.2.1. A Technical Lemma

Lemma 1. Let $z > 0$.

a. For $p \leq 3, \Gamma\left(\frac{p}{2} - 1, z\right) \geq \frac{e^{-z} z^{\frac{p}{2}-1}}{z+2-\frac{p}{2}}$; (52)

b. For $p = 1, \Gamma\left(\frac{p}{2} - 1, z\right) \leq \frac{2e^{-z}}{\sqrt{z}}$; (53)

For $p = 2, \Gamma\left(\frac{p}{2} - 1, z\right) \leq e^{-z}(1 + |\log z|)$; (54)

For $p = 3, \Gamma\left(\frac{p}{2} - 1, z\right) \leq \frac{e^{-z}}{\sqrt{z}}$; (55)

c. If $p \geq 4$ and is even, $\Gamma\left(\frac{p}{2} - 1, z\right) = \left(\frac{p}{2} - 2\right)! e^{-z} \sum_{i=0}^{\frac{p}{2}-2} \frac{z^i}{i!}$. (56)

Proof:

a. Use the representation that for $\alpha < 1$,

$$\Gamma(\alpha, z) = \frac{e^{-z} z^{\alpha}}{\Gamma(1-\alpha)} \int_0^{\infty} \frac{e^{-t} t^{-\alpha}}{z+t} dt \tag{57}$$

(e.g., see pp. 941 in Gradshteyn and Ryzhik (1980))

$$= e^{-z} z^{\alpha} \cdot E\left(\frac{1}{T+z} \Big/ T \sim \text{Gamma}(1-\alpha)\right)$$

$$\geq \frac{e^{-z} z^{\alpha}}{E(T) + z} = \frac{e^{-z} z^{\alpha}}{z + 1 - \alpha}.$$

b. The case $p = 1$ is immediate.

If $p = 2, \Gamma(\frac{p}{2} - 1, z) = \int_z^{\infty} \frac{e^{-t}}{t} dt$. If $z \geq 1$, this is evidently $\leq e^{-z}$. If $z < 1$,

$$\int_z^\infty \frac{e^{-t}}{t} dt = \int_z^1 \frac{e^{-t}}{t} dt + \int_1^\infty \frac{e^{-t}}{t} dt$$

$$\leq e^{-z} \cdot |\log z| + \int_1^\infty \frac{e^{-t}}{t} dt$$

$$= e^{-z}|\log z| + e^{-z} \cdot e^{z} \int_1^\infty \frac{e^{-t}}{t} dt$$

$$\leq e^{-z}|\log z| + e^{-z} \cdot e \cdot \int_1^\infty \frac{e^{-t}}{t} dt$$

$$\leq e^{-z}(|\log z| + 1).$$

For $p = 3$, $\Gamma(\frac{p}{2} - 1, z) = \int_z^\infty \frac{e^{-t}}{\sqrt{t}} dt \leq \frac{e^{-z}}{\sqrt{z}}$ (see inequality (1.05) on pp. 67 in Olver (1997)).

c. This is just a well known representation of the Poisson CDF.

5.2.2. Upper and Lower Bounds

We provide below upper bounds on $E(g(\underset{\sim}{x}, \underset{\sim}{\mu}))$ for fairly general smooth functions and lower bounds if the function g is subharmonic, i.e, $\Delta_x g \geq 0$ (convex for $p = 1$). The interesting thing is that the bounds only involve the second order derivatives and yet they are not Taylor expansion results: normality is playing a role. Also, the bounds for $p = 1$ provide a reversal as well as an improvement of Jensen's inequality for expectations of convex functions; the reversal is part a. and the improvement is in part b.

Theorem 10. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, tI), p \geq 1$, and let $g(\underset{\sim}{x}, \underset{\sim}{\mu})$ be any function as in Proposition 5.

a. If $p = 1$,

$$E(g(x, \mu)) \leq g(\mu, \mu) + t \cdot E(|g_{xx}(x, \mu)|). \tag{58}$$

If $p = 2$,

$$E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) \leq g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + \frac{t}{2} \cdot E\left(\left|\Delta_x g\left(\underset{\sim}{x}, \underset{\sim}{\mu}\right)\right|\left(1 + \left|\log \frac{\|\underset{\sim}{x} - \underset{\sim}{\mu}\|^2}{2t}\right|\right)\right). \tag{59}$$

If $p = 3$

$$E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) \le g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + t \cdot E\left(\frac{|\Delta_x g(\underset{\sim}{x}, \underset{\sim}{\mu})|}{\frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{t}}\right). \tag{60}$$

If $p \ge 4$ and is even one has the equality

$$E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) = g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + t \cdot \left(\frac{p}{2} - 2\right)!$$

$$\left\{\sum_{j=1}^{\frac{p}{2}-1} \frac{2^{j-1}}{(\frac{p}{2} - j - 1)!} E\left(\frac{\Delta_x g(\underset{\sim}{x}, \underset{\sim}{\mu})}{\left(\frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{t}\right)^j}\right)\right\} \tag{61}$$

b. If $p = 1$ and g is known to be convex,

$$E(g(x, \mu)) \ge g(\mu, \mu) + t \cdot E\left(\frac{g_{xx}(x, \mu)}{\frac{(x-\mu)^2}{t} + 3}\right). \tag{62}$$

If $p = 2$ and g is subharmonic,

$$E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) \ge g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + t \cdot E\left(\frac{\Delta_x g(\underset{\sim}{x}, \underset{\sim}{\mu})}{\frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{t} + 2}\right). \tag{63}$$

If $p = 3$ and g is subharmonic,

$$E(g(\underset{\sim}{x}, \underset{\sim}{\mu})) \ge g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + t \cdot E\left(\frac{\Delta_x g(\underset{\sim}{x}, \underset{\sim}{\mu})}{\frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{t} + 1}\right). \tag{64}$$

Proof: The bounds for $p = 1, 2, 3$ and the equality of $p \ge 4$ all follow on combining the basic moment formula (48) with Lemma 1; we omit the calculational details.

5.2.3. Two Short Examples

Although the more substantive applications are postponed till later sections, we will present two examples briefly to create a context for the bounds of Theorem 10.

Example 2. Marginal Density in Bayesian Statistics.
Suppose $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$ and we want to estimate $\underset{\sim}{\mu}$. A recently popular prior is the t-prior with density

$$\pi(\underset{\sim}{\mu}) = \frac{\Gamma(\frac{\alpha+p}{2})}{(\alpha\pi)^{p/2}\Gamma(\frac{\alpha}{2})} \; \frac{1}{(1 + \frac{\underset{\sim}{\mu}'\underset{\sim}{\mu}}{\alpha})^{\frac{\alpha+p}{2}}}; \tag{65}$$

see Berger (1986).

The marginal density of $\underset{\sim}{x}$ is just

$$m(\underset{\sim}{x}) = \int \frac{1}{(2\pi t)^{\frac{p}{2}}} e^{-\frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{2t}} \pi(\underset{\sim}{\mu}) d\underset{\sim}{\mu}, \tag{66}$$

which is therefore $E(\pi(\underset{\sim}{Y}))$ when $\underset{\sim}{Y} \sim N_p(\underset{\sim}{x}, tI)$. $\pi(\cdot)$ has all the properties needed and so Theorem 10 is applicable. For instance, for specificity if we choose $\alpha = 1$ (i.e., the prior is a Cauchy prior) and $p = 3$, then on calculations, the Laplacian of π is $\frac{12(\underset{\sim}{\mu}'\underset{\sim}{\mu}-1)}{\pi^2(1+\underset{\sim}{\mu}'\underset{\sim}{\mu})^4}$. And so, if we apply (60), then we have, uniformly in $\underset{\sim}{x}, m(\underset{\sim}{x}) \leq \pi(\underset{\sim}{x}) + \frac{12t}{\pi^2}$, a simple bound (of the correct order).

Example 3. Improving on Jensen's Inequality. Of course, in general, for a convex function g, one only can assert that $E(g(X)) \geq g(E(X))$. (62) says that due to normality, we can say more. To be specific take $Z \sim N(0,1)$ and a symmetric convex function $g(z) = f(z^2)$. Thus, $g''(z) = 2(f'(z^2) + 2z^2 f''(z^2))$. So if $f'(t) \geq a \geq 0$ and $f''(t) \geq b \geq 0$, then $g''(z) \geq 2(a + 2bz^2)$ and so, if we apply (62), then we get

$$\begin{aligned} E(g(Z)) \geq & g(0) + 2E\frac{a + 2bZ^2}{Z^2 + 3} \\ & = g(0) + 4(b + (\phi(\sqrt{3}) - 1)e^{\frac{3}{2}}\sqrt{\frac{\pi}{6}}(66 - a)) \\ & = g(0) + .54a + .766, \end{aligned} \tag{67}$$

as can be seen by exact computation of $E(\frac{1}{W+k})$ for a chisquare(1) random variable W. This is quite an improvement over what we can get from Jensen's inequality.

6. APPLICATIONS TO DECISION THEORY

The heat equation identity (3) and the canonical normal identity (15) are now used to study mean squared errors of estimates of location parameters in p dimensions, $p \geq 1$. Two principal directions are pursued:

i) provide bounds and an approximation to the mean squared error of an estimate of a multivariate normal mean. The approximation is compared to one that would arise naturally from Stein's (1981) unbiased estimate of risk;

ii) prove, by using the canonical normal identity, the inadmissibility of $\underset{\sim}{x}$ for estimating the location parameter of a multivariate t distribution in 3 or more dimensions. Along the way, we show that Stein's identity for the normal distribution holds as an inequality for t distributions and his unbiased estimate of risk is now an upwardly biased estimate of risk, under a condition.

6.1. Estimation of a Multivariate Normal Mean

For simplicity of presentation and notation, only the case $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$ is presented. First we give lower and upper bounds on the risk of an estimate $\underset{\sim}{\delta}(\underset{\sim}{x})$.

6.1.1. Bounds and Approximations for the Risk Function

Theorem 11.

Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI), p \geq 1$ and let $\underset{\sim}{\delta}(\underset{\sim}{x})$ be any three times differentiable estimate of $\underset{\sim}{\mu}$ such that $||\underset{\sim}{\delta}(\underset{\sim}{x})||$ and $\Delta\delta_i(\underset{\sim}{x})$ for each i are $0(e^{c||\underset{\sim}{x}||})$ for some $0 \leq c < \infty$. Then,

$$\begin{aligned} &E_{\underset{\sim}{\mu},t}||\underset{\sim}{\delta}(\underset{\sim}{x}) - \underset{\sim}{\mu}||^2 \\ &= ||\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}||^2 + \int_0^t E_{\underset{\sim}{\mu},s}\Bigg\{ \sum_{i=1}^{p} ||\nabla\delta_i(\underset{\sim}{x})||^2 \\ &\qquad + \sum_{i=1}^{p}(\delta_i(\underset{\sim}{x}) - x_i)\Delta\delta_i(\underset{\sim}{x}) + s \cdot \Delta(\nabla \cdot \underset{\sim}{\delta}(\underset{\sim}{x}))\Bigg\} ds \end{aligned} \tag{68}$$

As noted in Section 2, in the above $\nabla \cdot \underset{\sim}{\delta}(\underset{\sim}{x})$ denotes the divergence

$$\sum_{i=1}^{p} \frac{\partial}{\partial x_i}\delta_i(\underset{\sim}{x}) \text{ of } \underset{\sim}{\delta}.$$

We defer the proof of Theorem 11 in order to first construct bounds and an approximation for the risk $E||\underset{\sim}{\delta}(\underset{\sim}{x}) - \underset{\sim}{\mu}||^2$ by using the result (68) of Theorem 11. The approximation will be tested on an example. The following notation is used.

Notation. Given $\underset{\sim}{\delta}(\underset{\sim}{x})$ as in Theorem 11, denote

$$\left.\begin{aligned} \mathcal{D}_1(\underset{\sim}{x}) &= \sum_{i=1}^{p} ||\nabla\delta_i(\underset{\sim}{x})||^2 + \sum_{i=1}^{p}(\delta_i(\underset{\sim}{x}) - x_i)\Delta\delta_i(\underset{\sim}{x}) \\ \mathcal{D}_2(\underset{\sim}{x}) &= \Delta(\nabla \cdot \underset{\sim}{\delta}(\underset{\sim}{x})) \end{aligned}\right\} \tag{69}$$

Corollary 3. For $\underset{\sim}{\delta}(\underset{\sim}{x})$ as in Theorem 11,

$$\begin{aligned} &||\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}||^2 + t \cdot \inf_{\underset{\sim}{x}} \mathcal{D}_1(\underset{\sim}{x}) + \frac{t^2}{2} \cdot \inf_{\underset{\sim}{x}} \mathcal{D}_2(\underset{\sim}{x}) \\ &\leq E_{\underset{\sim}{\mu},t}||\underset{\sim}{\delta}(\underset{\sim}{x}) - \underset{\sim}{\mu}||^2 \\ &\leq ||\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}||^2 + t \cdot \sup_{\underset{\sim}{x}} \mathcal{D}_1(\underset{\sim}{x}) + \frac{t^2}{2} \cdot \sup_{\underset{\sim}{x}} \mathcal{D}_2(\underset{\sim}{x}). \end{aligned} \tag{70}$$

Note that (70) is immediate from Theorem 11. For most common estimates, the bounds in (70) are nontrivial (i.e., not $\pm\infty$). For any linear estimate $c\underset{\sim}{x} + \underset{\sim}{\mu}_0$, the lower and upper bounds coincide with the exact risk. The average, i.e., $\frac{\text{lower bound in (70)+upper bound in (70)}}{2}$, has served well in a few cases we tried. One such case will be reported as an example. The approximation to the risk function is presented next.

By Theorem 11, the risk function equals $||\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}||^2 + \int_0^t E_{\underset{\sim}{\mu},s}\{\mathcal{D}_1(\underset{\sim}{x}) + s\mathcal{D}_2(\underset{\sim}{x})\}ds$. In applications, the observation denoted here as $\underset{\sim}{x}$ will correspond to the sample mean of n observations and so the variance parameter $t = O(\frac{1}{n})$. The "first approximation" to the risk function replaces $E_{\underset{\sim}{\mu},s}\{\mathcal{D}_1(\underset{\sim}{x}) + s\mathcal{D}_2(\underset{\sim}{x})\}$ by $\mathcal{D}_1(\underset{\sim}{\mu}) + s\mathcal{D}_2(\underset{\sim}{\mu})$. In other words, the first approximation is

$$\begin{aligned} R(\underset{\sim}{\mu}, t, \underset{\sim}{\delta}) &\approx ||\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}||^2 + t\mathcal{D}_1(\underset{\sim}{\mu}) + \frac{t^2}{2} \cdot \mathcal{D}_2(\underset{\sim}{\mu}), \\ &= R_a(\underset{\sim}{\mu}, t, \underset{\sim}{\delta})(\text{ say}). \end{aligned} \tag{71}$$

Note that the approximation $R_a(\underset{\sim}{\mu}, t, \underset{\sim}{\delta})$ is not an estimate based on $\underset{\sim}{x}$ for the true risk as Stein's unbiased estimate of risk is. But still there is a basis for comparison. Recall that the Stein unbiased estimate is given as

$$R(\underset{\sim}{\mu}, t, \delta) = E_{\underset{\sim}{\mu},t}\{pt + ||\underset{\sim}{\delta}(\underset{\sim}{x}) - \underset{\sim}{x}||^2 + 2t\nabla \cdot (\underset{\sim}{\delta}(\underset{\sim}{x}) - \underset{\sim}{x})\}. \tag{72}$$

So an approximation analogous to (71) will be

$$\begin{aligned} R(\underset{\sim}{\mu}, t, \underset{\sim}{\delta}) &\approx ||\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}||^2 + 2t\nabla \cdot (\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}) + pt, \\ &= R_b(\underset{\sim}{\mu}, t, \underset{\sim}{\delta})(\text{ say}). \end{aligned} \tag{73}$$

Inspection of (71) and (73) shows that R_b involves terms up to the linear order in t and the first partials of $\underset{\sim}{\delta}$ in $\underset{\sim}{\mu}$ and R_a involves terms up to the order t^2 and the third order partials of $\underset{\sim}{\delta}$ in $\underset{\sim}{\mu}$. They are analogously constructed but different in form. We now compare R_a and R_b in an illustrative example.

6.1.2. Numerical Illustration

Example 4. Let $x_1, \ldots x_n \overset{iid}{\sim} N(\mu, 1)$ so that the sufficient statistic $\overline{x} \sim N(\mu, t)$ with $t = \frac{1}{n}$. As a fair test case, we take $n = 16$, a sample size that is neither large nor too small. To have our notation consistent with the results above, we will denote $\overline{x}$ by simply x. As the estimate $\delta(x)$, we will use a Bayes estimate. The prior is the double exponential prior with density $\pi(\mu) = \frac{1}{2}e^{-|\mu|}$. This prior has several quite special properties: it satisfies the Brown-Hwang (1982) gradient property; it arises as an asymptotically default proper prior in a natural way, as shown in Delampady et al (2000); and it is one of the rare nonconjugate priors for which the Bayes estimate can be written in a closed form.

Indeed, the Bayes estimate $\delta(X) = X + t\frac{m'(X)}{m(X)}$, where $m(x) = e^{\frac{t}{2}-x}\Phi(\frac{x-t}{\sqrt{t}}) + e^{\frac{t}{2}+x}$ $(1 - \Phi(\frac{x+t}{\sqrt{t}}))$ is the marginal density of X under this prior; see Brown (1986). Thus both $R_a(\mu, t, \delta)$ and $R_b(\mu, t, \delta)$ are available in closed form. The exact risk function $R(\mu, t, \delta)$ is not available in closed form, but is computable by numerical integration of Stein's unbiased estimate of risk (or by simulation also). First we present a small table. Figure 1 shows in more detail that $R_a(\mu, t, \delta)$ is a considerably better approximation to the true risk function than $R_b(\mu, t, \delta)$. The difference is especially visible for μ near 0. In this example, our proposal R_a seems to work well.

μ	Exact Risk	$\mathbf{R_a}$	$\mathbf{R_b}$
0	.0469	.0464	.0402
.25	.0509	.0497	.0489
.5	.0586	.0586	.0615
.75	.0639	.0649	.0659
1	.0660	.0663	.0664
2	.0664	.0664	.0664
5	.0664	.0664	.0664

We now indicate the proof of Theorem 11.

6.1.3. The Proof.

Proof of Theorem 11: The proof directly uses the heat equation identity (3).

Step 1. By identity (3), if one lets $g(\underset{\sim}{x}, \underset{\sim}{\mu}) = ||\underset{\sim}{\delta}(\underset{\sim}{x}) - \underset{\sim}{\mu}||^2$, then,

$$R(\underset{\sim}{\mu}, t, \delta) = ||\underset{\sim}{\delta}(\underset{\sim}{\mu}) - \underset{\sim}{\mu}||^2 + \frac{1}{2}\int_0^t E_{\underset{\sim}{\mu},s}(\Delta_x g)ds. \tag{74}$$

Step 2. Now,

$$\frac{\partial}{\partial x_j} g = 2\sum_{i=1}^{p}(\delta_i - \mu_i)\frac{\partial}{\partial x_j}\delta_i,$$

and
$$\frac{\partial^2}{\partial x_j^2} g = 2\{\sum_{i=1}^{p}(\frac{\partial}{\partial x_j}\delta_i)^2 + (\delta_i - \mu_i)\frac{\partial^2}{\partial x_j^2}\delta_i\},$$

and hence
$$\Delta_x g = 2\sum_{i=1}^{p}||\nabla\delta_i||^2 + 2\sum_{i=1}^{p}(\delta_i - \mu_i)\Delta\delta_i$$

$$= 2\sum_{i=1}^{p}||\nabla\delta_i||^2 + 2\sum_{i=1}^{p}(\delta_i - x_i)\Delta\delta_i + 2\sum_{i=1}^{p}(x_i - \mu_i)\Delta\delta_i \tag{75}$$

Step 3. By Stein's identity,

$$E_{\underset{\sim}{\mu},s}((x_i - \mu_i)\Delta\delta_i) = s\sum_{j=1}^{p}\frac{\partial}{\partial x_i}\frac{\partial^2}{\partial x_j^2}\delta_i,$$

and hence
$$E_{\underset{\sim}{\mu},s}(\sum_{i=1}^{p}(x_i - \mu_i)\Delta\delta_i)$$

$$= s\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{\partial}{\partial x_i}\frac{\partial^2}{\partial x_j 2}\delta_i$$

$$= s\sum_{j=1}^{p}\frac{\partial^2}{\partial x_j^2}\sum_{i=1}^{p}\frac{\partial}{\partial x_i}\delta_i$$

$$= s \cdot \Delta(\nabla \cdot \underset{\sim}{\delta}). \tag{76}$$

Step 4. Substituting (75) and (76) into (74) yields the result (68) of Theorem 11.

6.2. A Stein-Inequality for t Distributions

Diaconis and Zabell (1991) showed that the identity $E((X-\mu)h(X)) = E(h'(X))$ cannot hold for all $C_c^1(\mathbb{R})$ functions except when $X \sim N(\mu, 1)$. Interestingly, the inequality $E((T-\mu)h(T)) \geq \frac{m}{m-1}E(h'(T))$ does hold if $T \sim t(\mu, m)$ and if $h(\cdot)$ is monotone nondecreasing, as we show below. In fact, we give a multidimensional version. This can be called a Stein inequality for t distributions. The previously derived identity (27) is the key. We shall also see how an upwardly biased estimate of risk and thence certain inadmissibility results follow from this inequality.

6.2.1. The Inequality

Lemma 2. Let $\underset{\sim}{Z} \sim N_p(\underset{\sim}{0}, \frac{t}{n}I), n, p \geq 1, Y \sim t\chi^2(m), m+p > 2$, and suppose Y and $\underset{\sim}{Z}$ are independent. Let $T_0 = \frac{\sqrt{mn}\underset{\sim}{Z}}{\sqrt{Y}}$. Suppose $\underset{\sim}{h}: \mathbb{R}^p \to \mathbb{R}^p$ has the following properties:

$\underline{\text{i}}$ $\underset{\sim}{h} = \nabla f$ for some scalar function f

$\underline{\text{ii}}$ $\nabla \cdot h = \sum_{i=1}^{p} \frac{\partial}{\partial x_i} h_i(\underset{\sim}{x}) \leq 0.$

$\underline{\text{a}}$ Then, $E(\underset{\sim}{T}_0'\underset{\sim}{h}(\underset{\sim}{T}_o)) \leq \frac{m}{m+p-2}E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)).$ (77)

$\underline{\text{b}}$ If for a given $\underset{\sim}{\mu}, \underset{\sim}{T} = \underset{\sim}{T}_0 + \underset{\sim}{\mu}$, where $\underset{\sim}{T}_0$ is as defined above, then,

$$E((\underset{\sim}{T} - \underset{\sim}{\mu})'\underset{\sim}{h}(\underset{\sim}{T})) \leq \frac{m}{m+p-2}E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T})). \tag{78}$$

Proof: The proof uses the multidimensional version of the canonical normal identity (27). In the notation of the present lemma, the multidimensional version is:

Step 1. If a scalar function $g(\underset{\sim}{z}, y)$ is twice continuously differentiable in $\underset{\sim}{z}$ and once in y, and if $g, \nabla_z g$ are each $O(e^{a\|\underset{\sim}{z}\|}y^k)$ for some $0 \leq a, k < \infty$, then

$$\frac{\partial}{\partial t}E(g(\underset{\sim}{Z}, Y)) = \frac{1}{2n}E(\Delta_z g(\underset{\sim}{Z}, Y)) + \frac{1}{t}E((Y)g_y(\underset{\sim}{Z}, Y)). \tag{79}$$

Step 2. Let $\underset{\sim}{h}$ and f be as in the statement of the lemma, i.e., $\underset{\sim}{h} = \nabla f$. For this f, define a function $g(\underset{\sim}{z}, y)$ as $g(\underset{\sim}{z}, y) = f(\frac{\sqrt{mn}}{\sqrt{y}}\underset{\sim}{z})$. The multivariate identity (79) will be applied to this g. Note that the distribution of g does not depend on t, and so $\frac{\partial}{\partial t}E(g(\underset{\sim}{Z}, Y)) = 0$.

Step 3. In a straightforward manner,

$$\Delta_z g(\underset{\sim}{z}, y) = \frac{mn}{y} \nabla \cdot \underset{\sim}{h}(\underset{\sim}{t}_0),$$

$$\text{and} \quad g_y(\underset{\sim}{z}, y) = -\frac{1}{2y} \sum_{i=1}^{p} t_{0i} h_i(\underset{\sim}{t}_0), \tag{80}$$

where we have used $\underset{\sim}{t}_0$ to denote $\frac{\sqrt{mn}\underset{\sim}{z}}{\sqrt{y}}$.

Step 4. Therefore, by identity (79),

$$E\left(\sum_{i=1}^{p} T_{0i} h_i(\underset{\sim}{T}_0)\right) = mt \cdot E\left(\frac{\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)}{Y}\right). \tag{81}$$

Step 5. Treating t as a parameter as merely a technical device, we see that $Y + n\underset{\sim}{Z}'\underset{\sim}{Z}$ is a complete sufficient statistic for t, and so by Basu's (1956) theorem, T_0 and $Y + n\underset{\sim}{Z}'\underset{\sim}{Z}$ are independent.

Hence,

$$\begin{aligned} E\left(\sum_{i=1}^{p} T_{0i} h_i(\underset{\sim}{T}_0)\right) &= mt \cdot E\left(\frac{\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)}{Y}\right) \\ &\le mt \cdot E\left(\frac{\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)}{Y + n\underset{\sim}{Z}'\underset{\sim}{Z}}\right) \\ &\quad \text{(since by assumption } \nabla \cdot \underset{\sim}{h} \le 0) \\ &= mt \cdot E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)) \cdot E\left(\frac{1}{Y + n\underset{\sim}{Z}'\underset{\sim}{Z}}\right) \\ &= \frac{m}{m+p-2} E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)), \end{aligned} \tag{82}$$

where the last line is a consequence of the $\chi^2(m+p)$ distribution for $\frac{Y+n\underset{\sim}{Z}'\underset{\sim}{Z}}{t}$.

(82) proves part $\underset{\sim}{a}$ of Lemma 2. Part $\underset{\sim}{b}$ follows from part $\underset{\sim}{a}$.

Corollary 4. Let $f : \mathbb{R}^p \to \mathbb{R}^1$ be superharmonic and define $h = \nabla f$. Then

$$E((\underset{\sim}{T} - \underset{\sim}{\mu})'\underset{\sim}{h}(\underset{\sim}{T})) \le \frac{m}{m+p-2} E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T})).$$

Corollary 4 increases the applicability of Lemma 2 by demonstrating how to construct the function $\underset{\sim}{h}$ as in Lemma 2. Its proof is just a restatement of the Laplacian property $\Delta f \leq 0$.

6.3. Proving Inadmissibility from Biased Estimates of Risk

Lemma 2 and Corollary 4 permit construction of an upwardly biased estimate of the risk of an estimate $\underset{\sim}{T} + \underset{\sim}{h}(\underset{\sim}{T})$ of the location parameter $\underset{\sim}{\mu}$, analogous to Stein's unbiased estimate of risk in the normal case. The biased estimate converges pointwise to Stein's unbiased estimate as the degrees of freedom, m, of T tend to infinity. By suitable selection of the function $\underset{\sim}{h}(\underset{\sim}{T})$, uniform domination over $\underset{\sim}{T}$ still follows, although the risk estimate is biased.

Explicit estimates dominating $\underset{\sim}{T}$ for $p \geq 3$ are known; however, unlike the common methods that use the mixture structure (normal mixture) and covariance inequalities, Lemma 2 permits one to follow a more direct route squarely in the spirit of Stein (1981) for the normal case; see Cellier and Fourdrinier (1995, Proposition 2.3.1) for another context. In addition, the famous Stein superharmonicity result for the normal case also follows for the t case for all $p \geq 3$.

Proposition 7. A Biased Estimate of Risk. Let $\underset{\sim}{T} \sim t_p(m, \underset{\sim}{\mu})$ and let $\underset{\sim}{h}(\underset{\sim}{T})$ be any function satisfying inequality (78). For $m > 2$,

$$\begin{aligned} & E(||\underset{\sim}{T} + \underset{\sim}{h}(\underset{\sim}{T}) - \underset{\sim}{\mu}||^2) - E(||\underset{\sim}{T} - \underset{\sim}{\mu}||^2) \\ & \leq E(||\underset{\sim}{h}||^2 + \frac{2m}{m+p-2} \nabla \cdot h). \end{aligned} \tag{83}$$

Proposition 7 is evident because $\underset{\sim}{h}$ is assumed to satisfy inequality (78). The domination result to follow from this is given next.

Proposition 8. Let $\underset{\sim}{T} \sim t_p(m, \underset{\sim}{\mu}), p \geq 3, m > 2$. Let $\underset{\sim}{h} = \nabla f$ for some scalar function f and suppose $||\underset{\sim}{h}||^2 + \frac{2m}{m+p-2} \nabla \cdot \underset{\sim}{h} \leq 0$. Then $\underset{\sim}{T} + \underset{\sim}{h}(\underset{\sim}{T})$ dominates $\underset{\sim}{T}$ in risk for all $\underset{\sim}{\mu}$. In particular, the following special results hold:

a. An estimate of the form $(1 - \frac{r(||T||^2)}{||T||^2})\underset{\sim}{T}$ dominates $\underset{\sim}{T}$ if $r(\cdot)$ is differentiable, monotone nondecreasing, and $0 \leq r(z) \leq \frac{2m}{m+p-2}(p-2)$, or more generally, if

$$r^2(z) - \frac{2m}{m+p-2}(p-2)r(z) - \frac{4m}{m+p-2} z r'(z) \leq 0,$$

for all $z > 0$;

b. An estimate of the form $\underset{\sim}{T} + \nabla \log m(\underset{\sim}{T})$ dominates $\underset{\sim}{T}$ if $m : \mathbb{R}^p \to \mathbb{R}^1$ is a positive superharmonic function.

Proof: The general statement that $\underset{\sim}{T} + \underset{\sim}{h}(\underset{\sim}{T})$ dominates $\underset{\sim}{T}$ if $\underset{\sim}{h} = \nabla f$ and $||\underset{\sim}{h}||^2 + \frac{2m}{m+p-2}\nabla \cdot h \leq 0$ is an immediate consequence of Proposition 7 and Lemma 2. The special cases a and b both follow on calculation of $||\underset{\sim}{h}||^2 + \frac{2m}{m+p-2}\nabla \cdot \underset{\sim}{h}$ for $\underset{\sim}{h}(\underset{\sim}{T})$ of the respective forms in a and b; we omit the calculation.

6.4 Discussion

Evidently, the condition $||\underset{\sim}{h}||^2 + \frac{2m}{m+p-2}\nabla \cdot \underset{\sim}{h} \leq 0$ is a stronger condition than $||\underset{\sim}{h}||^2 + 2\nabla \cdot \underset{\sim}{h} \leq 0$. On the other hand, Proposition 8 does not make additional other assumptions, as in Brandwein and Strawderman (1990, pp. 363 and 1991, Theorem 2.1 and Example 2.1). Another positive feature is that the Stein superharmonicity result for the normal case is given for the t case also (part b. in Proposition 8) for $p \geq 3$.

Most of all, the route adopted is directly in the spirit of Stein, as described in section 6.3. So, on balance, there are both pros and cons of the methods presented here.

7. APPLICATION TO BAYESIAN STATISTICS

The expectation identities presented by us can also be usefully exploited to study Bayes risks in point estimation problems. Specifically, the expectation identities lead to identities and bounds for Bayes risks. We will show only the normal case, i.e. only such results that follow from the heat equation and the canonical normal identity for ease of presentation. The Bayes risk identities relate the Bayes risk to oscillations of the Bayes estimate; the Bayes risk bounds show methods to bound the Bayes risk from below by expressions similar to those in the now classic Borovkov–Brown–Gajek–Sakhanienko lower bounds for Bayes risk. In fact, as we shall see, in one case our lower bound is exactly the one previously obtained by these authors, by entirely different methods. We will give a more thorough discussion of this issue after the results are presented.

7.1. A Bayes Risk Identity

Proposition 9. Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\sum), p \geq 1$, and suppose $\underset{\sim}{\mu}$ has a prior G with posterior mean δ_G. Then the Bayes risk $r(t,G) = r(t,G,\delta_G)$ of the Bayes estimate δ_G satisfies the identity

$$\frac{d}{dt} r(t,G) = E_m[\sum_{k=1}^{p} (\nabla \delta_{G,k})' \sum (\nabla \delta_{G,k})], \tag{84}$$

where $\delta_{G,k}$ is the kth coordinate of δ_G and E_m denotes marginal expectation.

Corollary 5. Let $x \sim N(\mu, t)$ and let $\mu \sim G$. Then the Bayes risk $r(t,G)$ satisfies

$$\frac{d}{dt} r(t,G) = E_m \delta_G'(x)^2. \tag{85}$$

Since the proof of (84) is essentially the same as that of (85), we will just prove Corollary 5.

Proof of Corollary 5: Consider the function $g(x,\mu,t) = (\delta_G(x,t) - \mu)^2$. By the expectation identity (3),

$$\begin{aligned} \frac{d}{dt} r(t,G) &= \frac{d}{dt} \int E_\mu(g(x,\mu,t)) dG(\mu) \\ &= \frac{1}{2} E_G E_\mu (g_x''(x,\mu,t)) + 2 \int \int (\delta_G(x,t) - \mu) \frac{\partial}{\partial t} \delta_G(x,t) p(x|\mu,t) dx dG(\mu); \quad (86) \end{aligned}$$

now note that the second term in (86) is zero because $\delta_G(x,t)$ is $E(\mu|x)$. In addition, from the definition of g,

$$\begin{aligned} g_x'(x,\mu,t) &= 2(\delta_G(x,t) - \mu)\delta_G'(x,t) \\ \Rightarrow g_x''(x,\mu,t) &= 2(\delta_G'(x,t))^2 + 2(\delta_G(x,t) - \mu)\delta_G''(x,t). \end{aligned}$$

Of these, $E_G E_\mu((\delta_G(x,t) - \mu)\delta_G''(x,t)) = 0$ again, and so from (86), $\frac{d}{dt} r(t,G) = E_m(\delta_G'(x,t))^2$.

Remark: A minor but immediate consequence of Proposition 9 is that for any prior G, the Bayes risk is always an increasing function of t. Of course, this increasingness in t will follow from simply comparison of experiments results. But Proposition 9 goes further by laying out explicitly what $\frac{d}{dt} r(t,G)$ equals, not just that it is > 0.

7.2. Bayes Risk Bounds

We will now show how one can obtain lower bounds on $r(t, G)$ by using the identity of Proposition 9. The derivation of the bound, as we shall now see, may seem to be strange! Not only shall we use the apparently new identity (85), but a well known old Bayes risk identity for the normal case, namely the Brown identity for Bayes risk (Brown (1971, 1986); see also Lehmann and Casella (1998)). The proof manipulates the tautology that if two formulas exactly represent the same quantity (in this case the Bayes risk $r(t, G)$), then the expressions implied by the two formulas must be the same. As regards the lower bound itself, perhaps the comment most worth making is that the bound is the classic Borovkov–Brown–Gajek–Sakhanienko bound (Corollary 2.3 in Brown and Gajek (1986)), but the method is different. For example, we never use any Cramer-Rao type inequalities in our proof. There must be some connections, it seems. The bound in Proposition 10 below is attained when G is a normal prior.

Proposition 10 Let $x \sim N(\mu, t)$ and let $\mu \sim G$. Then

$$r(t, G) \geq \frac{t}{1 + tI(G)} \tag{87}$$

where $I(g)$ denotes the Fisher information of G. (Note that (87) is formally valid even if $I(G) = \infty$)

Proof: Step 1. The Bayes estimate $\delta_G(x)$ itself has the representation

$$\delta_G(x, t) = x + t\frac{m'(x)}{m(x)}$$

$$\Rightarrow \delta'_G(x, t) = 1 + t\frac{m''(x)}{m(x)} - t\frac{(m'(x))^2}{m^2(x)}.$$

Note: we should write m_t for m, but the 't' is being suppressed

Step 2. By Step 1 and Corollary 5,

$$
\begin{aligned}
&\frac{d}{dt}r(t,G)\\
&= E_m(\delta_G'(x,t))^2\\
&\geq [E_m(\delta_G'(x,t))]^2\\
&= (1-tI(m))^2, (as \int_{-\infty}^{\infty} m''(x)dx = 0)
\end{aligned}
$$

where $I(m)$ is the Fisher information of the marginal.

Step 3. The Brown identity says

$$r(t,G) = t - t^2 I(m).$$

Step 4. By Step 2 and Step 3,

$$
\begin{aligned}
&\frac{d}{dt}r(t,G)\\
&= \frac{d}{dt}(t - t^2 I(m)) && \text{(Step 3)}\\
&= 1 - 2tI(m) - t^2\frac{d}{dt}I(m)\\
&\geq (1-tI(m))^2 && \text{(Step 2)}\\
&= 1 - 2tI(m) + t^2 I^2(m)\\
\Rightarrow &-\frac{d}{dt}I(m) \geq I^2(m)\\
\Rightarrow &\frac{d}{dt}\frac{1}{I(m)} \geq 1 \quad \forall t > 0
\end{aligned}
$$

Step 5. Therefore, $\forall t > 0$,

$$
\begin{aligned}
\frac{1}{I(m_t)} &= \int_0^t \frac{d}{ds}\frac{1}{I(m_s)}ds + \frac{1}{I(G)}\\
&\geq t + \frac{1}{I(G)}\\
\Rightarrow I(m_t) &\leq \frac{I(G)}{tI(G)+1}.
\end{aligned}
$$

Step 6. Using the Brown identity (Step 3) again,

$$r(t, G) \geq t - \frac{t^2 I(G)}{tI(G) + 1} = \frac{t}{tI(G) + 1},$$

completing the proof.

7.3. Discussion

Bounds similar to the one in Proposition 10 are obtainable by our methods for the case $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, t\Sigma)$ by using Proposition 9 and the Brown identity for $r(t, G)$ and $\delta_G(x, t)$ for the $N_p(\underset{\sim}{\mu}, t\Sigma)$ case. These would be even more useful compared to the univariate case described in Proposition 10.

8. APPLICATIONS TO HARMONIC, SUBHARMONIC AND POLYHARMONIC FUNCTIONS

We previously saw in Section 3 how our expectation identities connect unbiased estimation to harmonicity. Now we will describe further implications and applications of these identities for harmonic, subharmonic and polyharmonic functions. The results are of two general types:

i) We show that certain very well known properties of harmonic functions follow from an easy application of the heat equation identity;

ii) We present statistical properties of harmonic, subharmonic and polyharmonic functions. For example, if $\underset{\sim}{X} \sim N_p(\underset{\sim}{0}, tI)$, how does the variance of a harmonic function behave; what can we say about the covariance between two harmonic functions, etc. For polyharmonic functions, specifically, we are able to give a general explicit formula for their covariance with $||\underset{\sim}{X}||^{2m}$. These exact formulas, we believe, may have concrete applications to the topic of nonlinear principal components. Now we present the results.

8.1. Unboundedness of Nonconstant Harmonic Functions

The result stated below is very well known, and a myriad of other proofs can be given (e.g. by using properties of the Brownian motion). But we wish to indicate that it also follows from an application of the moment formula (48).

Proposition 11.

a) Let g be a nonconstant $\mathcal{C}_2$ harmonic function in the plane. Then g is unbounded.

b) Let g be a nonconstant $\mathcal{C}_2$ harmonic function in $\mathbb{R}^p, p \geq 3$. If,

$$\text{either } \liminf_{||x|| \to 0} ||x|| \quad ||\nabla g(x)|| > 0$$

$$\text{or } \liminf_{||x|| \to \infty} ||x|| \quad ||\nabla g(x)|| > 0,$$

then g is unbounded.

Proof:

a) For the purpose of this proof, consider $\underset{\sim}{X} \sim N_2(\underset{\sim}{0}, tI)$. By formula (48),

$$\begin{aligned} E_t g^2(\underset{\sim}{X}) &= g^2(\underset{\sim}{0}) + \frac{1}{4\pi} \int (\Delta g^2) \Gamma(0, \frac{||\underset{\sim}{x}||^2}{2t}) d\underset{\sim}{x} \\ &= g^2(\underset{\sim}{0}) + \frac{1}{2\pi} \int (||\nabla g||^2 + g \Delta g) \Gamma(0, \frac{||\underset{\sim}{x}||^2}{2t}) d\underset{\sim}{x} \\ &= g^2(\underset{\sim}{0}) + \frac{1}{2\pi} \int ||\nabla g||^2 \Gamma(0, \frac{||\underset{\sim}{x}||^2}{2t}) d\underset{\sim}{x}, \end{aligned} \tag{88}$$

as g is by assumption harmonic.

From (88), by Fatou's lemma,

$$\begin{aligned} \liminf_{t \to \infty} E_t \ g^2(\underset{\sim}{X}) &\geq g^2(\underset{\sim}{0}) + \frac{1}{2\pi} \int ||\nabla g||^2 \liminf_{t \to \infty} \Gamma(0, \frac{||\underset{\sim}{x}||^2}{2t}) d\underset{\sim}{x} \\ &= \infty \end{aligned} \tag{89}$$

because $\lim\limits_{t \to \infty} \Gamma(0, \frac{||\underset{\sim}{x}||^2}{2t}) = \infty$ a.e. and $||\nabla g||^2 > 0$ on a set of positive Lebesgue measure, again by the assumptions on g. (89) forces g to be unbounded.

b) Once again, by formula (48), if $\underset{\sim}{X} \sim N_p(\underset{\sim}{0}, tI)$,

$$E_t g^2(\underset{\sim}{X}) = g^2(\underset{\sim}{0}) + \frac{1}{2\pi^{p/2}} \int \frac{||\nabla g||^2 ||\underset{\sim}{x}||^2}{||\underset{\sim}{x}||^p} \Gamma(\frac{p}{2} - 1, \frac{||\underset{\sim}{x}||^2}{2t}) d\underset{\sim}{x}$$

$$\Rightarrow \liminf_{t \to \infty} E_t g^2(\underset{\sim}{X}) \geq g^2(\underset{\sim}{0}) + \frac{\Gamma(\frac{p}{2}) - 1}{2\pi^{p/2}} \int \frac{||\nabla g||^2 ||\underset{\sim}{x}||^2}{||\underset{\sim}{x}||^p} d\underset{\sim}{x}. \tag{90}$$

Now suppose $\liminf_{||\underset{\sim}{x}||\to\infty} ||\underset{\sim}{x}||\ ||\nabla g(\underset{\sim}{x})|| = c > 0$. Then from (90), one gets $\liminf_{t\to\infty} E_t g^2(\underset{\sim}{X}) = \infty$ because $\int_{||\underset{\sim}{x}||>M} \frac{1}{||\underset{\sim}{x}||^p} d\underset{\sim}{x} = \infty$ for any M. If $\liminf_{||\underset{\sim}{x}||\to 0} ||(\underset{\sim}{x})||\ ||\nabla g(\underset{\sim}{x})|| > 0$, the proof is the same, and this establishes part b).

8.2. Mean and Variance Properties of Harmonic and Subharmonic Functions

The results in this section say that the mean and the variance of harmonic and subharmonic functions under the $N(\underset{\sim}{0}, tI)$ distribution satisfy certain monotonicity and convexity properties and these properties follow from our expectation identity (3).

Proposition 12. Let g be a $\mathcal{C}_2$ function on $\mathbb{R}^p$. Let also $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, tI)$.

a. g is harmonic if and only if $E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x}) \equiv g(\underset{\sim}{\mu})$.

b. Suppose g also satisfies the weak growth condition $g(\underset{\sim}{x}) \leq ce^{\alpha||\underset{\sim}{x}||}$ for some $c, \alpha > 0$. Then g is subharmonic if and only if $E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x})$ is nondecreasing in t.

c. If g is harmonic, then $\text{Var}_{\underset{\sim}{\mu},t} g(\underset{\sim}{x})$ is nondecreasing in t.

d. If g is subharmonic, $E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x})$ is convex in $\log t$ when $p = 2$, and convex in t^{2-p} for $p > 2$.

Proof:

a. By formula (48), $E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x}) \equiv g(\underset{\sim}{\mu})$ if g is harmonic. Conversely, if $E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x}) \equiv g(\underset{\sim}{\mu})$, then by the heat equation identity, $E_{\underset{\sim}{\mu},t}(\Delta g(\underset{\sim}{x})) \equiv 0$; in particular, $E_{\underset{\sim}{\mu},t=1}(\Delta g(\underset{\sim}{X})) = 0\ \ \forall \underset{\sim}{\mu}$. By the completeness property of the $N(\underset{\sim}{\mu}, I)$ family, $\Delta g = 0$ a.e. and hence $\Delta g \equiv 0$ as g is in $\mathcal{C}_2$ by assumption. So g is harmonic.

b. If g is subharmonic, then by the heat equation identity, $\frac{\partial}{\partial t} E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x}) \geq 0$ and so $E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x})$ is nondecreasing in t. Conversely, if $E_{\underset{\sim}{\mu},t} g(\underset{\sim}{x})$ is nondecreasing in t, then by the heat equation identity, $E_{\underset{\sim}{\mu},t}(\Delta g(\underset{\sim}{x})) \geq 0\ \forall \underset{\sim}{\mu}, \forall t > 0$. By letting $t \to 0, \Delta g(\underset{\sim}{\mu}) \geq 0\ \forall \underset{\sim}{\mu}$ provided we can show that $\lim_{t\to 0} E_{\underset{\sim}{\mu},t}(\Delta g(\underset{\sim}{x})) = \Delta g(\underset{\sim}{\mu})$. This is true by the dominated convergence theorem under the growth condition imposed in part b. So g is subharmonic.

c. Since g is harmonic, by part a.

$$\begin{aligned}
\text{Var}_{\underset{\sim}{\mu},t}\, g(\underset{\sim}{x}) &= E_{\underset{\sim}{\mu},t}\, g^2(\underset{\sim}{x}) - g^2(\underset{\sim}{\mu}) \\
\Rightarrow \frac{\partial}{\partial t} \text{Var}_{\underset{\sim}{\mu},t}\, g(\underset{\sim}{x}) &= \frac{\partial}{\partial t} E_{\underset{\sim}{\mu},t}\, g^2(\underset{\sim}{x}) \\
&= \frac{1}{2} E_{\underset{\sim}{\mu},t}(\Delta g^2(\underset{\sim}{x})) \text{ (by the heat equation identity)} \\
&= E_{\underset{\sim}{\mu},t}(||\nabla g(\underset{\sim}{x})||^2)(\because \Delta g^2 = 2g\Delta g + ||\nabla g||^2 = ||\nabla g||^2) \\
&\geq 0,
\end{aligned}$$

implying that $\text{Var}_{\underset{\sim}{\mu},t}\, g(\underset{\sim}{X})$ is nondecreasing in t.

d. We give the proof for the case $p = 2$; for $p > 2$, the proof is more or less exactly the same.

Denote $E_{\underset{\sim}{\mu},t}\, g(\underset{\sim}{x})$ (for given $\underset{\sim}{\mu}$) by $f(t)$. To show that f is convex in $\log t$, we need to show that $f(e^s)$ is convex in s.

From formula (48),

$$f(t) = g(\underset{\sim}{\mu}) + \frac{1}{4\pi^{p/2}} \int (\Delta g)\Gamma(0, \frac{||\underset{\sim}{x} - \underset{\sim}{\mu}||^2}{2t}) d\underset{\sim}{x}$$

and so it is enough to show that $\Gamma(0, \frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{2e^s})$ is convex in s. By straightforward differentiation, $\Gamma(0, \frac{c}{e^s})$ is convex in s for $c \geq 0$, and this completes the proof.

8.3. Covariance Properties

Harmonic and polyharmonic functions satisfy a number of interesting covariance properties under the $N(\underset{\sim}{0}, tI)$ distribution. The case of harmonic functions is presented separately first as harmonic functions are more popular in some sense. For polyharmonic functions, we will present a general formula for their covariance with $||\underset{\sim}{X}||^{2m}$. What is special about this formula that it is worth presenting? Roughly speaking, what is special is that the covariance can be found by **only** knowing the iterated Laplacians of the polyharmonic function at $\underset{\sim}{0}$, i.e., across the entire class a global quantity can be found exactly by only knowing a few local numbers. We will explain it precisely during the derivation.

Proposition 13. Let g, h be C_2 harmonic functions and let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, tI)$.

a. $\text{Cov}_{\underset{\sim}{\mu}=0,t}(||\underset{\sim}{x}||^2, g(\underset{\sim}{x})) \equiv 0 \quad \forall t$

b. $\text{Cov}_{\underset{\sim}{\mu},t}(g(\underset{\sim}{x}), h(\underset{\sim}{x})) \geq 0 \quad \forall \underset{\sim}{\mu}, t$ if and only if $(\nabla g)'(\underset{\sim}{x})\nabla h(\underset{\sim}{x}) \geq 0 \quad \forall \underset{\sim}{x}$.

Proof:

a. The proof uses both the Stein identity and the heat equation identity. First,

$$\begin{aligned} \text{Cov}_{\underset{\sim}{0},t}(||\underset{\sim}{x}||^2, g(\underset{\sim}{x})) &= E_{\underset{\sim}{0},t}(||\underset{\sim}{x}||^2 g(\underset{\sim}{x})) - ptE_{\underset{\sim}{0},t}g(\underset{\sim}{x}) \\ &= E_{\underset{\sim}{0},t}(||\underset{\sim}{X}||^2 g(\underset{\sim}{x})) - ptg(\underset{\sim}{0}), \end{aligned}$$

as g is harmonic. Hence,

$$\begin{aligned} \frac{d}{dt}\text{Cov}_{\underset{\sim}{0},t}(||\underset{\sim}{x}||^2, g(\underset{\sim}{x})) &= \frac{d}{dt}E_{\underset{\sim}{0},t}(||\underset{\sim}{x}||^2 g(\underset{\sim}{x})) - pg(0) \\ &= \frac{1}{2}E_{\underset{\sim}{0},t}(\Delta(||\underset{\sim}{x}||^2 g(\underset{\sim}{x}))) - pg(0). \end{aligned} \tag{91}$$

$$\begin{aligned} \text{But} \quad \Delta(||\underset{\sim}{x}||^2 g(\underset{\sim}{x})) &= ||\underset{\sim}{x}||\Delta g + 2pg(\underset{\sim}{x}) + 4\Sigma x_i \frac{\partial}{\partial x_i} g \\ &= 2pg(\underset{\sim}{x}) + 4\Sigma x_i \frac{\partial}{\partial x_i} g, \text{since} \quad \Delta g = 0. \end{aligned}$$

$$\begin{aligned} \text{Therefore,} \quad E_{\underset{\sim}{0},t}(\Delta(||\underset{\sim}{x}||^2 g(\underset{\sim}{x}))) &= 2pE_{\underset{\sim}{0},t}g(\underset{\sim}{x}) + 4E_{\underset{\sim}{0},t}(\Sigma x_i \frac{\partial}{\partial x_i} g) \\ &= 2pE_{\underset{\sim}{0},t}g(\underset{\sim}{x}) + 4tE_{\underset{\sim}{0},t}(\Sigma \frac{\partial^2}{\partial x_i^2} g) \\ &\quad \text{(by Stein's identity)} \\ &= 2pg(\underset{\sim}{0}), \end{aligned} \tag{92}$$

as g is harmonic.

Substitution of (92) into (91) yields

$$\frac{d}{dt} \text{Cov}_{\underset{\sim}{0},t}(||\underset{\sim}{x}||^2, g(\underset{\sim}{x})) \equiv 0.$$

Hence, $\text{Cov}_{\underset{\sim}{0},t}(||\underset{\sim}{x}||^2 g(\underset{\sim}{x}))$ is a constant and this constant must be 0 as can be seen by the dominated convergence theorem on letting $t \to 0$.

b. The structural distinction now is that we have $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, tI)$ and the result says that the only way two harmonic functions g and h can have a nonnegative correlation under all $\underset{\sim}{\mu}$ and t is that the inner product of their gradient vectors is always nonnegative (if we take $g(\underset{\sim}{X}) = \underset{\sim}{c}'\, \underset{\sim}{X}$ and $h(\underset{\sim}{X}) = \underset{\sim}{d}'\underset{\sim}{X}$, two linear functions, we see immediately that we must have $\underset{\sim}{c}'\underset{\sim}{d} \geq 0$ for g, h to have nonnegative correlation).

The proof of part b) is essentially the same as of part a).

For example, for the if part,

$$\begin{aligned}
&\frac{d}{dt}\text{Cov}_{\underset{\sim}{\mu},t}(g(\underset{\sim}{x}), h(\underset{\sim}{x})) \\
&= \frac{d}{dt}[E_{\underset{\sim}{\mu},t}(g(\underset{\sim}{x})h(\underset{\sim}{x})) - g(\underset{\sim}{\mu})h(\underset{\sim}{\mu})] \\
&= \frac{1}{2}E_{\underset{\sim}{\mu},t}[\Delta(gh)] \\
&= \frac{1}{2}E_{\underset{\sim}{\mu},t}[h\Delta g + g\Delta h + 2(\nabla g)'(\nabla h)] \\
&= E_{\underset{\sim}{\mu},t}[(\nabla g)'(\nabla h)] \\
&\geq 0, \text{ if } (\nabla g)'(\nabla h) \geq 0 \ \forall \underset{\sim}{x},
\end{aligned}$$

and hence $\text{Cov}_{\underset{\sim}{\mu},t}(g(\underset{\sim}{x}), h(\underset{\sim}{x})) \geq 0 \ \forall \underset{\sim}{\mu}, t$ as can be seen on letting $t \to 0$. The only if part is proved similarly.

8.4. Covariance Formulas for Polyharmonic Functions

In the preceding section, we proved that if $\underset{\sim}{X} \sim N(\underset{\sim}{0}, tI)$, then the covariance between $||\underset{\sim}{X}||^2$ and any harmonic function $g(\underset{\sim}{X})$ is zero. We show in this section that in fact for any polyharmonic function $g(\underset{\sim}{X})$, the covariance between $||\underset{\sim}{X}||^{2m}$ and $g(\underset{\sim}{X})$ admits a very clean general formular that depends only on the local behavior of g at $\underset{\sim}{0}$. The derivation of this general covariance formula uses what is known to potential theorists as the finite Almansi expansion of a polyharmonic function. We first provide these requisite technical background as well as the necessary definitions.

Definition 1. Let $g : \mathbb{R}^p \to \mathbb{R}^1$ be a $\mathcal{C}^{(2n)}$ function. g is said to be polyharmonic of degree n if $\Delta^n g = 0$, where Δ^n is the nth iterate of the Laplacian operator Δ.

The following representation of polyharmonic functions will be shortly used by us and is called its finite Almansi expansion; see Aronszajn, Creese and Lipkin (1983).

Lemma 3. Let g be polyharmonic of degree n. Then there exist unique harmonic functions $g_0, g_1, \ldots, g_{n-1}$ such that $g(\underset{\sim}{x}) = \sum\limits_{k=0}^{n-1} ||\underset{\sim}{x}||^{2k} g_k(\underset{\sim}{x})$.

Another fact about the functions g_k we will need in our covariance derivation is the following serio-integral representation; see pp 135 in Aronszajn, Creese and Lipkin (1983).

Lemma 4.

a. The harmonic function $g_0(\underset{\sim}{x})$ of Lemma 3 admits the representation

$$g_0(\underset{\sim}{x}) = g(\underset{\sim}{x}) + \sum_{i=1}^{\infty} \frac{(-1)^i ||\underset{\sim}{x}||^{2i}}{2^{2i}(i-1)!i!} \int_0^1 \tau^{\frac{p}{2}} - 1\tau^{i-1}(1-\tau)^{i-1}(\Delta^i g)(\tau \underset{\sim}{x}) d\tau; \tag{93}$$

b. For $k \geq 1$,

$$g_k(\underset{\sim}{x}) = \frac{1}{2^{2k}k!} \sum_{i=0}^{\infty} \frac{(-1)^i ||\underset{\sim}{x}||^{2i}}{2^{2i} i!(k+i)!} \int_0^1 \tau^{\frac{p}{2}-1}\tau^{i-1}(1-\tau)^{k+i-1}(i+k\tau)(\Delta^{k+i} g)(\tau \underset{\sim}{x}) d\tau. \tag{94}$$

We now state our general covariance formula.

Theorem 12. Let $g : \mathbb{R}^p \to \mathbb{R}^1$ be polyharmonic of degree n for some $n \geq 1$ and let $m \geq 1$ be any fixed integer. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{0}, tI)$. Then

$$\mathrm{Cov}_t(||\underset{\sim}{x}||^{2m}, g(\underset{\sim}{x})) = \frac{(2t)^m}{\Gamma(\frac{p}{2})} \sum_{k=1}^{n-1} \frac{\{\Gamma(\frac{p}{2})\Gamma(\frac{p}{2}+m+k) - \Gamma(\frac{p}{2}+m)\Gamma(\frac{p}{2}+k)\}}{k!\Gamma(\frac{p}{2}+k)2^k} (\Delta^k g)(\underset{\sim}{0}) t^k; \tag{95}$$

in particular, if $n = 1$ (i.e. if g is harmonic), then for any $m \geq 1, ||\underset{\sim}{X}||^{2m}$ and $g(\underset{\sim}{X})$ are uncorrelated (which is an extension of the correlation zero result in Proposition 13 for $m = 1$).

Proof:

Step 1. From part a. of Proposition 13, if h is a $\mathcal{C}_2$ harmonic function, $E_t(||\underset{\sim}{x}||^2 h(\underset{\sim}{x})) = E_t(||\underset{\sim}{x}||^2) E_t(h(\underset{\sim}{X})) = ph(\underset{\sim}{0})t$.

Step 2. By direct verification, $\Delta(uv) = u\Delta v + v\Delta u + 2(\nabla u)'\nabla v$. Hence, by Step 1, induction, and some algebra, for $r \geq 1$,

$$E_t(||\underset{\sim}{x}||^{2r}h(\underset{\sim}{x})) = p(p+2)\dots(p+2(r-1))h(\underset{\sim}{0})t^r$$
$$= 2^r \frac{\Gamma(\frac{p}{2}+r)}{\Gamma(\frac{p}{2})} h(\underset{\sim}{0})t^r. \quad (96)$$

Step 3. By Lemma 3,

$$\text{Cov}_t(||\underset{\sim}{x}||^{2m}, g(\underset{\sim}{x})) = \sum_{k=0}^{n-1} \text{Cov}_t(||\underset{\sim}{x}||^{2m}, ||\underset{\sim}{x}||^{2k} g_k(\underset{\sim}{x}))$$
$$= \sum_{k=0}^{n-1} \{E_t(||\underset{\sim}{x}||^{2(m+k)} g_k(\underset{\sim}{x})) - E_t(||\underset{\sim}{x}||^{2m}) E_t(||\underset{\sim}{x}||)^{2k} g_k(\underset{\sim}{x})\}$$
$$= \sum_{k=0}^{n-1} \{2^{m+k} \frac{\Gamma(\frac{p}{2}+m+k)}{\Gamma(\frac{p}{2})} g_k(\underset{\sim}{0}) t^{m+k} - 2^m t^m \frac{\Gamma(\frac{p}{2}+m)}{\Gamma(\frac{p}{2})} \times 2^k \frac{\Gamma(\frac{p}{2}+k)}{\Gamma(\frac{p}{2})} g_k(\underset{\sim}{0}) t^k\}, \quad (97)$$

the last line using Step 2 and the formula for the mth moment of a $\chi^2(p)$ distribution.

Step 4. In (97), the term corresponding to $k = 0$ cancels. For $k \geq 1$, by Lemma 4,

$$g_k(\underset{\sim}{0}) = \frac{1}{2^{2k} k!(k-1)!} (\Delta^k g)(\underset{\sim}{0}) \int_0^1 \tau^{\frac{p}{2}-1}(1-\tau)^{k-1} d\tau$$
$$= \frac{\Gamma(\frac{p}{2})\Gamma(k)}{2^{2k} k!(k-1)!\Gamma(\frac{p}{2}+k)} (\Delta^k g)(\underset{\sim}{0}). \quad (98)$$

Step 5. Substituting (98) into (97) yields the formula stated in (95) on algebra.

9. DIFFERENTIAL EQUATIONS DRIVING A MOMENT SEQUENCE AND APPLICATIONS

We will now show that it follows from the heat equation identity (3) that if $x \sim N(\mu, t)$, then for any $n \geq 1, E_{\mu,t}((x-\mu)^{2n} g(x))$ satisfies a 'universal' nth order linear differential equation. Precisely, if $h(t) \triangleq E_{\mu,t}(g(x))$, then there exists a fixed triangular array of constants $\{a_{i,n}\}$ such that $a_{0,n}t^n h(t) + a_{1,n}t^{n+1}h'(t) + \dots + a_{n,n}t^{2n}h^{(n)}(t) - E((x-\mu)^n g(x)) \equiv 0$. A similar equation holds for $E_{\mu,t}((x-\mu)^{2n+1} g(x))$. It is interesting that such a universal differential equation should hold at all; it is also interesting that the results say that by knowing **only** $E(g(x))$ and $E(g'(x))$, as a function of t, one can find $E((x-\mu)^n g(x))$ for any $n \geq 1$. In section 9.2, we will use this development to outline a general formal approximation for $E(g(|\hat{\theta}_n - \theta|)$ for an estimate $\hat{\theta}_n$ of a parameter of interest θ. The formal approximation is then applied to a variety of cases, with quite nice accuracy in each case. First we present the universal differential equation.

9.1. A Universal Differential Equation

For simplicity, we present the derivation for the case $\underset{\sim}{\mu} = 0$. First we state a lemma that would be used in the derivation. This lemma is the reason that the constants $\{a_{i,n}\}$ in the differential equation can be written explicitly, which adds to the utility of the equation.

Lemma 3. Let $H_n(x)$ be the nth Hermite polynomial defined as

$\frac{d^n}{dx^n}(e^{\frac{-x^2}{2}}) = (-1)^n H_n(x) e^{-\frac{x^2}{2}}$. Then $x^n = \sum_{i-0}^{[\frac{n}{2}]} \frac{n!}{i!(n-2i)!2^i} H_{n-2i}(x)$.

In particular,

$$x^{2n} = \sum_{i=0}^{n} \frac{(2n)!2^i}{(n-i)!(2i)!2^n} H_{2i}(x), \tag{99}$$

and

$$x^{2n+1} = \sum_{i=0}^{n} \frac{(2n+1)!2^i}{(n-i)!(2i+1)!2^n} H_{2i+1}(x). \tag{100}$$

Proof: This representation of the powers x^n in terms of Hermite polynomials may be derived from the identity given in Problem 77 in pp 389 in Szego (1975).

The differential equation is given next.

Theorem 13. Let $x \sim N(0,t)$.

a. Let $n \geq 1$ and suppose $g^{(2j)}$ satisfies the heat equation identity (3) for $j = 0, 1, \ldots,$ $n-1$. Then

$$E_t(x^{2n} g(x)) = \sum_{i=0}^{n} a_{i,n} t^{n+i} \frac{d^i}{dt^i}(E_t g(x)), \tag{101}$$

$$\text{where } a_{i,n} = \frac{(2n)!2^{2i}}{2^n(2i)!(n-i)!}, 0 \leq i \leq n;$$

b. Let $n \geq 1$ and suppose $g^{(2j+1)}$ satisfies the heat equation identity (3) for $j = 0, 1, \ldots, n-1$. Then

$$E_t(x^{2n+1}g(x)) = \sum_{i=0}^{n} c_{i,n} t^{n+i+1} \frac{d^i}{dt^i}(E_t g'(x)), \tag{102}$$

$$\text{where } c_{i,n} = \frac{(2n+1)!2^{2i}}{2^n(2i+1)!(n-i)!}, 0 \leq i \leq n.$$

Proof: We will only prove part a) here as part b) is similar. Towards this end,

$$\begin{aligned}
&E_t(x^{2n}g(x)) \\
&= t^n E_{t=1}(x^{2n}g(x\sqrt{t})) \\
&= t^n \sum_{i=0}^{n} \frac{(2n)!2^i}{(n-i)!(2i)!2^n} E_{t=1}(H_{2i}(x)g(x\sqrt{t})) \text{ (Lemma 3)} \\
&= t^n \sum_{i=0}^{n} \frac{(2n)!2^i}{(n-i)!(2i)!2^n} t^i E_t(g^{(2i)}(x)) \text{ (integration by parts)} \\
&= t^n \sum_{i=0}^{n} \frac{(2n)!2^i}{(n-i)!(2i)!2^n} t^i 2^i \frac{d^i}{dt^i}(E_t g(x)) \text{ (heat equation identity)} \\
&= \sum_{i=0}^{n} \frac{(2n)!2^{2i}}{(n-i)!(2i)!2^n} t^{n+i} \frac{d^i}{dt^i}(E_t g(x)),
\end{aligned}$$

as claimed.

The coefficients $a_{i,n}$ and $c_{i,n}$ as given in Theorem 12, are tabulated below for $n \leq 5$ for the user's convenience.

Table 2

i \ n	$a_{i,n}$: 1	2	3	4	5	$c_{i,n}$: 1	2	3	4	5
0	1	3	15	405	945	3	15	105	945	10395
1	2	12	90	840	9450	2	20	210	2520	34650
2	0	4	60	840	12600	0	4	84	1512	27720
3	0	0	8	224	5040	0	0	8	288	7920
4	0	0	0	16	720	0	0	0	16	880
5	0	0	0	0	32	0	0	0	0	32

Is Theorem 13 of any demonstrable practical use? We think it is; in the next section, we will give a few examples and we think the potential for further practical applications is high. But first we give below an interesting consequence of Theorem 13.

Corollary 6. Let $X \sim N(0, t)$.

a. Let $g(x) = h(x^2)$ be an even function of x. If X^2 and $g(X)$ are uncorrelated for each $t > 0$, then g must be a constant.

Proof:

a. From part a. of Theorem 13, one readily gets $\text{Cov}_t(x^2, g(x)) = 2t^2(\frac{d}{dt}E_t g(x))$. Thus, $\text{Cov}_t(x^2, g(x)) = 0 \;\; \forall t \Rightarrow E_t g(x) \equiv c$ (a constant). Since $g(x)$ is a function of x^2, by the completeness of x^2, it follows that g itself is a constant.

9.2. Formal Gram-Charlier Expansions and Approximate Computation of Expectations

The material in this section is our most prominent attempt at using theorem 12 for practically useful applications. The goal is to supply a general recipe by using Theorem 12 for approximate computation of $E(g(Z))$ when Z does not have exactly a normal distribution, but an "approximately" normal distribution. For instance, Z may be $\frac{\sqrt{n}(\overline{X} - E_F(X_1))}{\sqrt{\text{Var}_F(X_1)}}$ for $X_1, \ldots, X_n$ iid from some F; other types are considered in section 9.4. We expect, by virtue of the central limit theorem, that Z would be approximately normal. Our recipe below for approximation of $E(g(Z))$then has a good chance of producing an adequately accurate result. The recipe, as presented below, is a formalism; we did not give error bounds, although we think rigorous error bounds can in fact be given. The formal recipe for approximation of $E(g(Z))$ for a nonnormal Z presented below uses two main ingredients: Theorem 13 above and a formal Gram-Charlier density expansion for Z. The density expansion is used **only** as a tool to produce an approximation. But in concrete applications, the approximation may be used even if Z is discrete. In fact, in our illustrative demonstrations, we will try our approximation on a Poisson and a Binomial example. Now let us see the formal approximation.

Let $X_1, X_2, \ldots, X_n$ be iid real valued random variables with $E(X_1^4) < \infty$; if the characteristic function of X_1 is in $L^j(\mathbb{R})$ for some j ≥ 1, then $Z = Z_n = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma}$ has a density for all $n \geq j$ and the density, say $f_n(z)$, admits the Gram-Charlier expansion

$$f_n(z) = \phi(z)[1 + \frac{\mu_3}{6\sigma^3\sqrt{n}}(z^3 - 3z) + \frac{\mu_4 - 3\sigma^4}{24\sigma^4 n}(z^4 - 6z^2 + 3)]$$
$$+ \frac{\mu_3^2}{72\sigma^6 n}(z^3 - 3z)] + 0(n^{-1}), \tag{103}$$

uniformly in z; here μ_i is the central moment $E(x-\mu)^i$ and $\sigma^2 = \mu_2$. See pp 508 in Feller (1966). Of course, the expansion (103) cannot be integrated for a general function $g(z)$ to obtain an approximation to $E(g(Z_n))$ with an error of $0(n^{-1})$ (or any error bound in fact). But interestingly, if we integrate $g(z)$ with respect to the two term expansion in (103), then the terms involve exactly $E_{N(0,1)}(Z^m g(Z))$, for which we have certain representations in Theorem 13. Use of that representation leads to a simple but formal recipe for approximate computation of $E(g(Z_n))$. We must reemphasize that the approximation is a formalism for we are not supplying error bounds. We shall see in the example that the formalism can work very well even for very small n.

To proceed with the formal approximation, denote

$$M_g(t) = E_{N(0,t)}(g(x)), M_g'(t) = \frac{d}{dt}M_g(t), \text{ and } M_g''(t) = \frac{d^2}{dt^2}M_g(t).$$

We present our formal approximation for an even function $g(z)$ only to simplify the steps; all the steps in the formalism go through for a general $g(z)$ at the expense of slightly more notational complexity.

By using part a. of Theorem 13, one gets

$$\left.\begin{array}{l} E_{N(0,1)}(x^2 g(x)) = M_g(1) + 2M_g'(1) \\ E_{N(0,1)}(x^4 g(x)) = 3M_g(1) + 12M_g'(1) + 4M_g''(1) \end{array}\right\} \tag{104}$$

Thus a formal approximation to $E(g(Z_n)) = \int_{-\infty}^{\infty} g(z) f_n(z) dz$ for an even function $g(z)$ is

$$E(g(Z_n)) \approx M_g(1) + \frac{\mu_4 - 3\sigma^4}{24\sigma^4 n}(3M_g(1) + 12M_g'(1) + 4Mg''(1) - 6M_g(1) - 12M_g'(1) + 3M_g(1))$$
$$= M_g(1) + \frac{\mu_4 - 3\sigma^4}{6\sigma^4 n}M_g''(1). \tag{105}$$

(105), really, is just a term-by-term integration; the point is that computation of $E_{N(0,1)}(x^{2i}g(x))$ for $i = 1, 2$ is replaced by computation of the single number $M_g''(1)$. Another point is that presumably the formal approximation will generally get more accurate by using more terms in the density expansion (103) involving higher moments. Let us see two examples of the use of the formal approximation (105).

Examples

Example 5. Suppose we are interested in writing an approximation to $E\left|\sum_{i=1}^{n} x_i - n\lambda\right|$ where $x_i \overset{iid}{\sim} Poi(\lambda)$.

Then in the notation of section 9.2, we want to approximate $E(g(Z_n))\sqrt{n\lambda}$ where $g(z) = |z|$. The approximation is accurate unless λ is small although Z_n does not even have a density at all. The quantities needed for execution of the formal approximation (105) are: $M_g(1) = \sqrt{\frac{2}{\pi}}, M_g''(1) = \frac{-1}{2\sqrt{2\pi}}, \sigma^2 = \lambda, \mu_4 = \lambda + 3\lambda^2$. Thus, the formal approximation is

$$E\left(\left|\sum_{i=1}^{n} x_i - n\lambda\right|\right) = \sqrt{n\lambda}E(g(Z_n)) \approx \sqrt{\frac{2n\lambda}{\pi}}(1 - \frac{1}{24n\lambda}). \tag{106}$$

Actually, $E(|\sum_{i=1}^{n} x_i - n\lambda|)$ also has an exact formula so the accuracy of the very simple approximation (106) can be checked. The exact formula is $E(|\sum_{i=1}^{n} x_i - n\lambda|) = \frac{2n\lambda e^{-n\lambda}(n\lambda)^{[n\lambda]}}{[n\lambda]!}$ (see Diaconis and Zabell (1991)). The following table reports the percentage error of approximation (106) when $\lambda = 1$; the percentage errors are very small.

Table 3

n	**Exact**	**Approximation (106)**	**% Error**
6	1.9275	1.9407	.68%
10	1.9383	1.9464	.42%
15	1.9437	1.9491	.28%
25	1.9478	1.9513	.18%

Example 6. This example illustrates use of the formal approximation (105) in a decision theory problem. Suppose $X \sim Bin(n, p)$ and we wish to estimate p using the loss function.

$$\left.\begin{aligned} L(p,a) &= 0 \text{ if } |p-a| \le \epsilon \\ &= |p-a| \text{ if } |p-a| > \epsilon \end{aligned}\right\} \tag{107}$$

The problem is to write a simple approximation to the risk function of the usual estimate $\hat{p} = \frac{X}{n}$. Thus, in the notation of Section 9.2, we want to approximate $E(g(Z_n))\sqrt{\frac{pq}{n}}$ where $q = 1 - p$, $Z_n = \frac{\sqrt{n}(\hat{p}-p)}{\sqrt{pq}}$, and $g(z) = |z| I_{|z| > \frac{\epsilon\sqrt{n}}{\sqrt{pq}}}$.

By calculation the quantities needed for application of the formal approximation (105) are

$$\left.\begin{aligned} M_g(1) &= \sqrt{\tfrac{2}{\pi}} e^{-\frac{n\epsilon^2}{2pq}} \\ M_g''(1) &= \frac{e^{-\frac{n\epsilon^2}{2pq}} (2p^3 - p^4 - 2np\epsilon^2 + n^2\epsilon^4 - p^2(1-2n\epsilon^2))}{2\sqrt{2\pi} p^2 q^2} \\ \sigma^2 &= pq, \ \mu_4 = pq(p^3 + q^3); \end{aligned}\right\} \quad (108)$$

the approximation to the risk function is $R(p, \hat{p}) = \sqrt{\frac{pq}{n}}(M_g(1) + \frac{\mu_4 - 3\sigma^4}{6\sigma^4 n} M_g''(1))$. Curiously, approximately 135 years ago, Todhunter published a formula (Todhunter (1865)) for the binomial distribution which can be used to give a closed form formula, although frightening, for $R(p, \hat{p})$. This formula is as follows: let $b(k, n, p)$ denote the binomial *pmf* $P_{n,p}(X = k)$. Let [ ] and { } denote the floor and ceiling of a. nonnegative integer. Let also m denote the smallest integer $> np$. In the expression below, $\{n(p - \epsilon)\}$ is interpreted as 0 and $[n(p + \epsilon)]$ is interpreted as n when they are outside of the range 0 to n. Then,

$$R(p, \hat{p}) = \frac{\begin{array}{c} 2qm\, b(m, n, p) - (\{np\}q\, b(\{np\}, n, p) - (n - [n(p + \epsilon)])p \\ b([n(p + \epsilon)], n, p) - \{n(p - \epsilon)\}q\, b(\{n(p - \epsilon)\}, n, p) + (n - [np])p\, b([np], n, p)) \end{array}}{n}; \quad (109)$$

in comparison, (108) is much simpler. The following table reports the accuracy of the approximation (108) when $p = .25$ and $\epsilon = .01$. The accuracy seems to be high.

Table 4

n	**Exact Risk**	**Approximation (108)**
5	.1582	.1552
10	.1126	.1093
20	.0759	.0770
30	.0637	.0626
40	.0541	.0540

9.4. Bahadur Representation and Hájek Projections

So far we have illustrated our formal approximation formula (105) to approximate $E(g(Z_n))$ when Z_n is exactly $\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}$ for an iid sequence $x_1, x_2, \ldots$. But we can also use formula (105) when Z_n is not a standardized sample mean but only admits a linear representation

$$Z_n = \frac{1}{n}\sum_{i=1}^{n}\phi_F(x_i) + R_n, \tag{110}$$

where ϕ_F is an approximate function and the remainder R_n is of a "lower" order; often $R_n = o_p(\frac{1}{\sqrt{n}})$. There are two main ways to write such linear representations for nonlinear statistics; certain statistics are known to admit a Bahadur representation and for certain statistics one can compute the Hájek projection. Bahadur representations are available (under conditions) for sample percentiles, M estimates, Bayes estimates, etc. Hájek projections, in principle, are available for the attractive general class of U statistics (and certain statistics which are not U statistics, e.g. suitable linear combinations of order statistics). The formal approximation (105) for $E(g(Z_n))$ may be used for all such statistics; whether the approximation can be produced will depend on two factors: can we produce $\phi_F(X_i) = Y_i$ (say) in (110) explicitly, and can we compute the moments $\sigma^2(Y_1)$ and $\mu_4(Y_1)$ explicitly. In the cases where these can be done, the formal approximation formula (105) may be applied and may be a practically useful thing to have. We now present two additional examples to illustrate this idea.

Example 7. This example illustrates the use of our moment approximation formula (105) when the statistic has a known Bahadur representation.

Specifically, let $\hat{\xi}_{p,n}$ be the *pth* sample percentile and ξ_p the *pth* population percentile for a standard Cauchy distribution. $\hat{\xi}_{p,n}$ admits the Bahadur representation

$$\hat{\xi}_{p,n} - \xi_p = \overline{Z}_n + R_n, \tag{111}$$

where $Z_i = \frac{p - I(X_i \leq \xi_p)}{f(\xi_p)}$ and $R_n = 0(n^{\frac{-3}{4}}(\log \log \mathrm{n})^{\frac{3}{4}})$ (see Kiefer (1967)). To approximate $E|\hat{\xi}_{p,n} - \xi_p|$, we will use the approximation formula (105) with $g(z) = |z|$. The moments $\sigma^2(Z_1)$ and $\mu_4(Z_1)$ are equal to $\sigma^2(Z_1) = \frac{pq}{f^2(\xi_p)}$ and $\mu_4(Z_1) = \frac{pq^4 + p^4q}{f^4(\xi_p)} = \frac{pq(1-3pq)}{f^4(\xi_p)}$. This gives the simple approximation

$$E|\hat{\xi}_{p,n} - \xi_p| \approx \frac{\sqrt{pq}}{\sqrt{n}f(\xi_p)}\sqrt{\frac{2}{\pi}}(1 - \frac{1-6pq}{24pqn}). \tag{112}$$

No exact formulae for $E|\hat{\xi}_{p,n} - \xi_p|$ are possible and so we verify the accuracy of (112) by simulation; a simulation size of 10,000 was used. In the table below, $p = .5$; again the approximation is accurate.

Table 5

n	**Simulated** $E\lvert\hat{\xi}_{p,n}-\xi_p\rvert$	**Approximation (112)**
20	.2897	.2814
30	.2342	.2295
40	.2005	.1986
60	.1638	.1620

Example 8. This example illustrates the use of formula (105) for a U statistic. For purposes of illustration, we take our statistic to be the Gini mean difference $T_n = \frac{1}{\binom{n}{2}} \sum\sum_{j<k} |X_j - X_k|$. T_n has the Hoeffding decomposition

$$T_n - E|X_1 - X_2| = \sum_{i=1}^{n} E(T_n|X_i) - nE|X_1 - X_2| + R_n, \tag{113}$$

where $R_n = 0(n^{-1}(\log \text{n})^{\frac{1}{2}+\delta})$ for any $\delta > 0$ if $E(X_1^2) < \infty$ (see Geertsema (1970) or Serfling (1980)).

For the ensuing calculation, denote $E|X_1 - X_2| = \theta$ and $E|X_1 - u| = m(u)$. Then the terms $E(T_n|X_i)$ in the Hájek projection can be calculated if we can calculate $E|X_j - X_k| \Big| X_i$ for all pairs $j < k$. For instance,

$$\begin{aligned}
&E(T_n|X_1) \\
&= \frac{1}{\binom{n}{2}} \sum\sum_{j<k} E|X_j - X_k| \Big| X_1 \\
&= \frac{1}{\binom{n}{2}} \left\{ \sum_{k=2}^{n} E|X_1 - X_k| \Big| X_1 + \sum_{j=2}^{n-1} \sum_{k=j+1}^{n} E|X_j - X_k| \right\} \\
&= \frac{1}{\binom{n}{2}} \left\{ (n-1)m(X_1) + \frac{(n-1)(n-2)}{2}\theta \right\}
\end{aligned} \tag{114}$$

Thus

$$\begin{aligned}
\sum_{i=1}^{n} E(T_n|X_i) &= \frac{1}{\binom{n}{2}} \left\{ (n-1) \sum_{i=1}^{n} m(X_i) + \frac{n(n-1)(n-2)}{2}\theta \right\} \\
&= \frac{2}{n} \sum_{i=1}^{n} m(X_i) + (n-2)\theta.
\end{aligned} \tag{115}$$

Hence, $T_n - \theta = 2\left(\frac{1}{n}\sum_{i=1}^{n} m(X_i) - \theta\right) + R_n$, with $R_n = 0(n^{-1}(\log n)^{\frac{1}{2}+\delta})$. If we denote $m(X_i) = Z_i$, then to approximate $E|T_n - \theta|$ we use formula (105) with $g(z) = |z|$, i.e.,

$$E|T_n - \theta| \approx \frac{2\sigma(Z_1)}{\sqrt{n}}\left(\sqrt{\frac{2}{\pi}} - \frac{1}{2\sqrt{2\pi}}\frac{\mu_4(z_1) - 3\sigma^4(Z_1)}{6\sigma^4(Z_1)n}\right). \tag{116}$$

If we take, for illustration, the population density to be $\frac{1}{\lambda}e^{\frac{-x}{\lambda}}$, then on calculation

$$\left.\begin{aligned} m(u) &= u - \lambda + 2\lambda e^{\frac{-u}{\lambda}} \\ \sigma^2(Z_1) &= \tfrac{\lambda^2}{3}, \mu_4(Z_1) = \tfrac{149}{45}\lambda^4 \end{aligned}\right\} (117)$$

Substitution into (116) yields, on algebra,

$$E|T_n - \theta| \approx \frac{2\lambda}{\sqrt{3n}}\sqrt{\frac{2}{\pi}}(1 - \frac{67}{60n}). \tag{118}$$

Again, an exact expression for $E|T_n - \theta|$ is not available and we check the accuracy of (118) by simulating $E|T_n - \theta|$; we have taken $\lambda = 1$ in the following table.

Table 6

n	**Simulated $\mathbf{E}|\mathbf{T_n} - \theta|$**	**Approximation (118)**
15	.2356	.2202
20	.2074	.1945
30	.1687	.1619
40	.1473	.1416

Compared to the previous examples, it takes a larger value of n for the approximation to get very accurate; it is probably due to the pronounced skewness in the original Exponential population.

10. APPLICATIONS IN GRAPH THEORY

The heat equation identity (3) leads to certain applications in graph theory. In the following, we will indicate its application in counting matchings in graphs. Roughly speaking, the heat equation identity gives a method to count perfect matchings in a graph by breaking it into graphs with successively smaller numbers of vertices. If the original graph has n vertices, then the reduced graphs have $n-2, n-4, n-6, \ldots$ vertices. This method may have some practical utility due to reduction to simpler graphs for which counting matchings may be physically easier. Matchings in graphs appear to have been independently reinvented a number of times in different branches of science. Important applications have been made in statistical physics and theoretical chemistry; Godsil (1993) describes the discovery that the properties of aromatic hydrocarbons depend on the number of matchings if a molecule is represented as a graph with the atoms as vertices and the bonds as edges. Farrell (1979) introduced matching polynomials in the combinatorics literature; a later exposition is Godsil and Gutman (1981). A common example of the use of perfect matchings is assignment of a set of tasks to a set of competent individuals so that none is assigned more that one task. First we present the definitions, notation and certain technical facts that we will use in our derivation.

10.1. Definitions and Notations

Definition.

a. An ***r*-matching** in a graph G with n vertices is a set of r edges no two of which have a vertex in common. It will be denoted as $p(G, r)$. By convention, $p(G, 0) = 1$.

b. The **matching polynomial** of a graph G with n vertices is the nth degree polynomial

$$\mu_G(x) = \sum_{r=0}^{[\frac{n}{2}]} (-1)^r p(G, r) x^{n-2r}. \tag{119}$$

c. A **perfect matching** in a graph G with n vertices is an r-matching with $r = \frac{n}{2}$ (thus, n has to be even). It will be denoted by $\psi(G)$.

d. The complement of a graph G is a graph with the same vertex set as G and two vertices sharing an edge if they did not share an edge in G. It will be denoted as $\overline{G}$.

e. For a graph G with n vertices, and $u_1, \ldots, u_t$ some $t \geq 1$ specified vertices, $G - u_1, \ldots u_t$ is the graph with $u_1, \ldots u_t$ deleted from the vertex set of G.

f. A graph G with n vertices having no edges is called an empty graph. It will be denoted as ϕ_n.

g. The complement of ϕ_n is called the complete graph on n vertices. It will be denoted as K_n.

10.2. Certain Known Technical Facts

Below we state a collection of facts on matchings and matching polynomials that we will use in the proof of our subsequent result. Significantly more similar facts are known but we do not state them here; many of these can be seen in Godsil and Gutman (1981) and Godsil (1993).

Lemma 6.

a. The matching polynomial of the disjoint union of any two graphs G and H satisfies $\mu_{G\cup H}(x) = \mu_G(x)\mu_H(x)$. (120)

b. For any graph G, $\frac{d}{dx}\mu_G(x) = \sum\limits_u \mu_{G-u}^{(x)}$. (121)

c. For any graph G, the total number of perfect matchings in $\overline{G}$ satisfies $\psi(\overline{G}) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \mu_G(x) e^{-\frac{x^2}{2}}\, dx.$ (122)

Remark. Formula (122) manifests the enjoyably surprising connection between matchings in graphs and the normal distribution. It will also be a key ingredient for our result.

10.3. Perfect Matchings and the Heat Equation

The result below gives an identity relating perfect matchings in the complement of the disjoint union of two graphs to matchings in appropriate subgraphs. Of course, there are various other ways to express the number of perfect matchings in complement of a union of graphs, but we will not mention them here.

Theorem 14. For some $m \geq n \geq 1$, let G be a graph on $2m$ vertices and H a graph on $2n$ vertices. Then the number of perfect matchings in the complement of $G \cup H$ equals

$$\psi(\overline{G \cup H}) = \sum_{i=0}^{n} \frac{2^i}{(2i)!} \alpha_{i,n,H} (\Sigma_i \psi(\overline{G - u_1, u_2 \ldots u_{2i}})), \tag{123}$$

where Σ_i denotes sum over all subsets of $2i$ vertices of G regarding identical collections with different orderings of $u_1, u_2, \ldots, u_{2i}$ as different, and

$$\alpha_{i,n,H} = \sum_{r=i}^{n} (-1)^{n-r} \frac{(2r)!}{2^r (r-i)!} p(H, n-r). \tag{124}$$

Proof: The principal tools for this identity on perfect matchings are Lemma 6 above and Theorem 13 in section 9.1. Towards this end,

$$\psi(\overline{G \cup H}) = \int \mu_{G\cup H}(x)\phi(x)dx \qquad \text{(equation (122))}$$

$$= \int \mu_G(x)\mu_H(x)\phi(x)dx \qquad \text{(equation (120))}$$

$$= \sum_{r=0}^{n}(-1)^r p(H,r) \int x^{2n-2r}\mu_G(x)\phi(x)dx \qquad \text{(equation (119))}$$

$$= \sum_{r=0}^{n}(-1)^r p(H,r)\{\sum_{i=0}^{n-r} \frac{(2n-2r)!2^{2i}}{2^{n-r}(2i)!(n-r-i)!}(\frac{d^i}{dt^i}E_{N(0,t)}(\mu_G(x)))_{t=1}\}$$

(equation (101))

$$= \sum_{r=0}^{n}(-1)^r p(H,r)\{\sum_{i=0}^{n-r} \frac{(2n-2r)!2^{2i}}{2^{n-r}(2i)!(n-r-i)!}\frac{1}{2^i}E_{N(0,1)}(\mu_G^{(2i)}(x))\}$$

(iteration of the heat equation identity (3))

$$= \sum_{r=0}^{n}(-1)^r p(H,r)\{\sum_{i=0}^{n-r} \frac{(2n-2r)!2^{i}}{2^{n-r}(n-r-i)!(2i)!}E_{N(0,1)}(\Sigma_i \mu(x)_{G-u_1,\ldots,u_{2i}}\}$$

(iteration of equation (121))

$$= \sum_{r=0}^{n}(-1)^r p(H,r)\{\sum_{i=0}^{n-r} \frac{(2n-2r)!2^{i}}{2^{n-r}(n-r-i)!(2i)!}\Sigma_i E_{N(0,1)}(\mu(x)_{G-u_i\ldots u_{2i}})\}$$

$$= \sum_{r=0}^{n}(-1)^r p(H,r)\{\sum_{i=0}^{n-r} \frac{(2n-2r)!2^{i}}{2^{n-r}(n-r-i)!(2i)!}\Sigma_i \psi(\overline{G-u_1\ldots u_{2i}})\}$$

(equation (122) again)

$$= \sum_{i=0}^{n} \frac{2^i}{(2i)!}\{\sum_{r=0}^{n-i}(-1)^r p(H,r)\frac{(2n-2r)!}{2^{n-r}(n-r-i)!}\}\Sigma_i \psi(\overline{G-u_1\ldots u_{2i}})$$

(change in order of summation)

$$= \sum_{i=0}^{n} \frac{2^i}{(2i)!}\{\sum_{r=i}^{n}(-1)^{n-r} p(H,n-r)\frac{(2r)!}{2^{r}(r-i)!}\}\Sigma_i \psi(\overline{G-u_1\ldots u_{2i}}),$$

(change of variable)

completing the proof.

The following corollary is for the special case when H has 2 or 4 vertices.

Corollary 7.

a. If G has an even number of vertices and H has 2 vertices, then

$$\psi(\overline{G \cup H}) = (1 - p(H,1))\psi(\overline{G}) + \sum_{v \neq u}\sum \psi(\overline{G - uv}). \tag{125}$$

b. If G has 2m vertices for some $m \geq 2$ and H has 4 vertices, then

$$\psi(\overline{G \cup H}) = (3 - p(H,1) + p(H,2))\psi(\overline{G}) + (6 - p(H,1))\sum_{v \neq u}\sum \psi(\overline{G - uv})$$

$$+ \sum_{z \neq w \neq v \neq u}\sum\sum\sum \psi(\overline{G - uvwz}). \tag{126}$$

Proof: Each part follows on some algebra from Theorem 14.

REFERENCES

Aronszajn, N., Creese, T. M. and Lipkin, L. J., 1983, Clarendon Press, Oxford.

Axler, S., Bourdon, P. and Ramey, W., 1992, Springer-Verlag, New York.

Basu, D., 1956, *Sankhya*, **15**, 377–380.

Berger, J., 1986, Springer-Verlag, New York.

Brandwein, A. C. and Strawderman, W. E., 1990, *Stat. Sc.*, **5**, 356–369.

Brown, L. D., 1971, *Ann. Math. Stat.*, **42**, 855–903.

Brown, L. D., 1986, IMS, Hayward, California.

Brown and Hwang, 1982, Statistical Decision Theory and Related Topics, III, S. S. Gupta and J. Berger, Eds., Academic Press, New York.

Brown, L. D. and Gajek, L., 1990, *Ann. Stat.*, 18, 1578–1594.

Brown, L. D. and Liu, R. C., 1993, *Ann. Stat.*, **21**, 1–13.

Cellier, D. and Fourdrinier, D., 1995, *J. Mult. Anal.*, **52**, 338–351.

Delampady, M. et. al., 2000, Purdue University Technical Report, To appear.

Diaconis, P. and Zabell, S., 1991, *Stat. Sc.*, **6**, 284–302.

Doss, H. and Sethuraman, J., 1989, *Ann. Stat.*, **17**, 440–442.

Farrell, E. J., 1979, *J. Comb. Theory, Ser. B*, **27**, 75–86.

Feller, W., 1936, *Math. Ann.*, **113**, 113–160.

Feller, W., 1966, **Vol. 2**, John Wiley, New York.

Folland, G., 1992, Wadsworth and Brooks - Cole, Pacific Grove, California.

Geertsema, J. C., 1970, *Ann. Math. Stat.*, **41**, 1016–1026.

Godsil, C. J., 1993, Chapman and Hall, New York.

Godsil, C. J. and Gutman, I., 1981, *J. Graph Theory*, **5**, 137–144.

Gradshteyn, I. and Ryzhik, I., 1980, Academic Press, New York.

Ito, K., 1951, *Mem. Amer. Math. Soc.*, **4**.

Kiefer, J., 1967, *Ann. Math. Stat.*, **38**, 1323–1342.

Lehmann, E. L. and Casella, G., 1998, Chapman and Hall, New York.

Marsden, J. E. and Tromba, A. J., 1996, Freeman, New York.

Olver, F. W., 1997, A K Peters, Wellesley, Massachusetts.

Rudin, W., 1974, McGraw Hill, New York.

Serfling, R. J., 1980, John Wiley, New York.

Stein, C., 1981, *Ann. Stat.*, **9**, 1135–1151.

Szego, G., 1975, *Amer. Math. Soc.*, Providence, Rhode Island.

Todhunter, I., 1865, Macmillan, London.