7-1-2006

The Heat Equation and Stein's Identity: Connections, Applications

Lawrence D. Brown
University of Pennsylvania

Anirban DasGupta

Leonard R. Haff

William E. Strawderman

Recommended Citation

Brown, L. D., DasGupta, A., Haff, L. R., & Strawderman, W. E. (2006). The Heat Equation and Stein's Identity: Connections, Applications. *Journal of Statistical Planning and Inference, 136* (7), 2254-2278. http://dx.doi.org/10.1016/j.jspi.2005.12.001

The Heat Equation and Stein's Identity: Connections, Applications

Abstract
This article presents two expectation identities and a series of applications. One of the identities uses the heat equation, and we show that in some families of distributions the identity characterizes the normal distribution. We also show that it is essentially equivalent to Stein's identity. The applications we have presented are of a broad range. They include exact formulas and bounds for moments, an improvement and a reversal of Jensen's inequality, linking unbiased estimation to elliptic partial differential equations, applications to decision theory and Bayesian statistics, and an application to counting matchings in graph theory. Some examples are also given.

Keywords
Bayes risk, harmonic, heat equation, inadmissibility, matching polynomial, Stein's identity, unbiased

Disciplines
Applied Mathematics | Statistics and Probability

THE HEAT EQUATION AND STEIN'S IDENTITY : CONNECTIONS,APPLICATIONS

by

L. Brown
University of Pennsylvania

A. DasGupta
Purdue University
and University of California,
San Diego

L. R. Haff
University of California,
San Diego

and

W. E. Strawderman
Rutgers University

Technical Report #00-04

Department of Statistics
Purdue University
West Lafayette, IN USA

November 2002

⊛

THE HEAT EQUATION AND STEIN'S IDENTITY : CONNECTIONS,APPLICATIONS

by

L. Brown
University of Pennsylvania

A. DasGupta
Purdue University
and University of California,
San Diego

L. R. Haff
University of California,
San Diego

and

W. E. Strawderman
Rutgers University

Abstract

This article presents two expectation identities and a series of applications. One of the identities uses the heat equation, and we show that in some families of distributions the identity characterizes the normal distribution. We also show that it is essentially equivalent to Stein's identity. The applications we have presented are of a broad range. They include exact formulas and bounds for moments, an improvement and a reversal of Jensen's inequality, linking unbiased estimation to elliptic partial differential equations, applications to decision theory and Bayesian statistics, and an application to counting matchings in graph theory. Some examples are also given.

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 two expectation identities, and present a series of applications. One of the identities is derived by using the heat equation. We show that this identity is essentially equivalent to Stein's identity. It is thus not very surprising that it seems to have a broad range of potential applications, as we indicate in this article. The other identity is applicable to problems in time series, and to normally distributed data when both the mean and variance are unknown. However, it is the first identity derived from the heat equation that we have

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

analyzed and applied in more detail.

We have indicated a broad range of applications. They can be roughly classified into the following areas:

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

b) an improvement as well as a reversal of Jensen's inequality;

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 to its Bayes risk;

f) applications in graph theory, and specifically, establishing a connection to counting perfect matchings in graphs.

These applications and other illustrative examples are presented in sections 2 through 8.

We have not developed all the applications to their full potential in this article. We think that there is an excellent potential for additional applications. In fact, that seems to be the best feature of the results we have presented. These results developed over a period of several years. We are indebted to Persi Diaconis, Joe Eaton and Malay Ghosh for valuable suggestions, and for graciously reading through the article's previous drafts.

2. NOTATION AND IDENTITIES

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 $\frac{\partial^2}{\partial t^2} g, \nabla_x g$ 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. $||y||$ 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. $E_{\underset{\sim}{\mu}, t}$ will denote expectation and often will be written as just E; similarly, var$(\cdot)$ will stand for variance and cov for covariance.

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

g. $R(\underset{\sim}{\mu}, t, \delta)$: 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 $r(\pi, t) = r(\pi, t, \delta)$ the Bayes risk of an estimate δ.

h. $\phi(\cdot)$ 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.

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

j. $N_p(\underset{\sim}{\mu}, t\Sigma_0)$ 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 chi-square 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. Heat Equation Identity

Theorem 1.

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 $\mathrm{O}(e^{c||\underset{\sim}{x}||})$ for some $0 \leq c < \infty$. Then

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

Identity (1) will be referred to as the **Heat Equation Identity**.

Proof of Theorem 1:

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

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

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

$$\int\limits_{B(\underset{\sim}{0}, r)} (g\Delta_x p - p\Delta_x g) d\underset{\sim}{x} = \int\limits_{\partial B(\underset{\sim}{0}, r)} (g\nabla p - p\nabla g)' n d\sigma, \tag{3}$$

where n denotes the unit outer normal.

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

$$\begin{aligned}
&\int\limits_{\partial B(\underset{\sim}{0}, r)} (g\nabla p - p\nabla g)' n d\sigma \\
&\leq \int\limits_{\partial B(\underset{\sim}{0}, r)} |g| \; ||\nabla p|| d\sigma + \int\limits_{\partial B(\underset{\sim}{0}, r)} p||\nabla g|| d\sigma \\
&\leq A e^{cr} \int\limits_{\partial B(\underset{\sim}{0}, r)} ||\nabla p|| d\sigma + B e^{cr} \int\limits_{\partial B(\underset{\sim}{0}, r)} p d\sigma,
\end{aligned} \tag{4}$$

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

Step 4. From (2), write $\frac{\partial}{\partial t}E(g(\underset{\sim}{X},\underset{\sim}{\mu}))$ as

$$\frac{\partial}{\partial t}E(g(\underset{\sim}{X},\underset{\sim}{\mu})) = \frac{1}{2}\int(p\Delta_x g)d\underset{\sim}{x} + \frac{1}{2}\int(g\Delta_x p - p\Delta_x g)d\underset{\sim}{x}. \tag{5}$$

Step 5. Finally,

$$\begin{aligned}&\int(g\Delta_x p - p\Delta_x g)d\underset{\sim}{x}\\ =&\lim_{r\to\infty}\int_{B(\underset{\sim}{O},r)}(g\Delta_x p - p\Delta_x g)d\underset{\sim}{x}\\ =&0 \text{ by (3) and (4).}\end{aligned}$$

2.3. Canonical Normal Identity

Two statistically important cases not covered by the preceding theorem 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.

Theorem 2. Let $c, m \geq 0$. Suppose $X \sim N(\mu, ct), Y \sim t\chi^2(m)$, and suppose $X,\ Y$ 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 $O(e^{a|x|}y^k)$ for some $0 \leq a, k < \infty$. Then

$$\frac{\partial}{\partial t}E(g(X,Y,\mu)) = \frac{c}{2}E(g_{xx}(X,Y,\mu)) + \frac{1}{t}E(Yg_y(X,Y,\mu)). \tag{6}$$

Identity (6) 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,

$$\frac{\partial}{\partial t}p_1 = \frac{c}{2}\frac{\partial^2}{\partial x^2}p_1, \tag{7}$$

$$\text{and} \quad \frac{\partial}{\partial t}p_2 = \frac{y}{2t^2}p_2 - \frac{m}{2t}p_2 = -\frac{y}{t}\frac{\partial}{\partial y}p_2 - \frac{p_2}{t}, \tag{8}$$

on some calculations.

Step 2. Since X, Y are independent, the joint density is given by

$$p(x, y, \mu, t) = p_1(x, \mu, t)p_2(y, t). \tag{9}$$

Multiplying both sides by $g(x, y, \mu)$, (6) 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

The identities in Section 2 have some 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 some of the applications in Theorem 3 are reexpressions of known facts in the area of partial differential equations.

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.

Theorem 3. 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 $\lim_{t\to 0+} h(\underset{\sim}{\mu}, t) = g(\underset{\sim}{\mu})$ exists, then $g(\underset{\sim}{X})$ is an unbiased estimate of h provided $E(g(\underset{\sim}{X}))$ exists.

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 3, 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 A = max (a, b), we have, respectively, $|h(\underset{\sim}{\mu})| \leq \mathrm{A}(1 + ||\underset{\sim}{\mu}||^i), i = 1, 2$, 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 3,

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

$$\Rightarrow \quad cf'(\mu + c\sqrt{t}) = 2\sqrt{t} f''(\mu + c\sqrt{t}). \tag{10}$$

Consider now (μ, t) lying on the one-dimensional curve $\mu = -c\sqrt{t}$. Then, from (10), $cf'(0) = 2\sqrt{t} f''(0)$. 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.

Remark: 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 3,

$$\begin{aligned} \frac{\partial}{\partial t} h &= \frac{k}{2} c_0 t^{\frac{k}{2}-1} + c_1 \frac{k-1}{2} t^{\frac{k-3}{2}} \mu + \sum_{i=2}^{k} c_i \mu^i \frac{k-i}{2} (\sqrt{t})^{k-i-2} \\ &= \frac{1}{2} \frac{\partial^2}{\partial \mu^2} h \\ &= \frac{1}{2} \sum_{i=0}^{k-2} (i+1)(i+2) c_{i+2} \mu^i (\sqrt{t})^{k-i-2}. \end{aligned} \tag{11}$$

By comparing coefficients of the powers of μ on the two sides of (11), 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 3 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.

The next result uses identity (6).

Theorem 4. Let $c, m > 0$ and suppose $X \sim N(\mu, ct)$, $\quad Y \sim t\chi^2(m)$, 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 (6) as

$$\frac{\partial}{\partial t}E(g(X,Y) = \frac{c}{2}\frac{\partial^2}{\partial \mu^2}E(g(X,Y)) + \frac{1}{t}E(Y \frac{\partial}{\partial Y} g(X,Y)). \tag{12}$$

a. This is an immediate consequence of the above identity (12).

b. Again, by (12),

$$0 = \frac{c}{2}h''(\mu) + \frac{1}{t}E(Yg_2'(Y))$$
$$\Rightarrow \frac{1}{t}E(Yg_2'(Y)) = -\frac{c}{2}h''(\mu). \tag{13}$$

The RHS of (13) 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 (13) 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 $\alpha, \beta, \gamma, \eta, \delta$,

$$h(\mu) = \alpha\mu^2 + \beta\mu + \gamma = E(g_1(X)) + E(\delta Y + \eta), \tag{14}$$

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 (1) of Section 2 says 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 (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 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 5. 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 $\mathcal{C}^1(\mathbb{R})$, 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,

$$2\frac{d}{dt}Eg(X) = -2\int g(x)\left\{\frac{1}{2t^{3/2}}p\left(\frac{x}{\sqrt{t}}\right) + \frac{x}{2t^{3/2}}\frac{1}{\sqrt{t}}p'\left(\frac{x}{\sqrt{t}}\right)\right\}dx$$

$$= Eg''(X)$$

$$= \int g''(x)\frac{1}{\sqrt{t}}p\left(\frac{x}{\sqrt{t}}\right)dx \qquad \forall t > 0$$

Step 2. Therefore,

$$-\int g(x)\left\{\frac{1}{t}p\left(\frac{x}{\sqrt{t}}\right) + \frac{x}{t^{3/2}}p'\left(\frac{x}{\sqrt{t}}\right)\right\}dx$$

$$= \int g''(x)p\left(\frac{x}{\sqrt{t}}\right)dx$$

$$= \int g(x)\frac{1}{t}p''\left(\frac{x}{\sqrt{t}}\right)dx \text{ (on integration by parts) } \forall t > 0. \tag{15}$$

Hence, by using $t = 1$,

$$\int g(x)\{p''(x) + xp'(x) + p(x)\} = 0 \text{ for all } g \text{ in } \mathcal{C}_0^\infty(\mathbb{R})$$

$$\Rightarrow\; p''(x) + xp'(x) + p(x) = a.e. \text{ (and hence everywhere)} \tag{16}$$

Step 3. One solution of (16) 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 (16) 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,

$$h(x)(T'(x))^2 - tT''(x)h(x) - 2tT'(x)h'(x) + t^2h''(x)$$

$$= 2h(x)T(x) + 2t^2\frac{\beta'(t)}{\beta(t)}h(x) \qquad \forall t, \forall x. \tag{17}$$

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

$$(T'(x))^2 - 2T(x) = 0, \tag{18}$$

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

Step 3. Substituting $T(x) = \frac{(x-\mu)^2}{2}$ in (17) and setting $x = \mu$, one now gets:

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

$$\Rightarrow 2t\frac{\beta'(t)}{\beta(t)} = t\frac{h''(\mu)}{h(\mu)} - 1 \qquad \forall t > 0. \tag{19}$$

Step 4. From (19), 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 now show that the heat equation identity (1) 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 6.

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

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

c. For $p > 1$, Identity (1) $\Rightarrow$ 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 (1), define $h(x, \mu) = g_x(x, \mu)$. Then by Stein's identity,

$$tE(g_{xx}(X, \mu)) = tE(h_x(X, \mu)) = E((X - \mu)h(X, \mu)). \tag{20}$$

Step 2. However,

$$
\begin{aligned}
&E((X-\mu)h(X,\mu)) \\
&= \int (x-\mu)\frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2t}(x-\mu)^2} g_x(x,\mu)dx \\
&= \sqrt{t}\int \frac{1}{\sqrt{2\pi}} z e^{-\frac{1}{2}z^2} g_x(\mu + z\sqrt{t},\mu)dz, \qquad (21)
\end{aligned}
$$

by a change of variable.

Step 3. Write (21) as

$$
2t\int \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\frac{\partial}{\partial t}g(\mu+z\sqrt{t},\mu)dz = 2t + \frac{\partial}{\partial t}\int \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}g(\mu+z\sqrt{t},\mu)dz \qquad (22)
$$

Step 4. Now, make the change the change of variable back to $x = \mu + z\sqrt{t}$, yielding (22)$= 2t\frac{\partial}{\partial t}E(g(X,\mu))$, hence establishing identity (1).

b. Step 1. Given h as in Stein's identity, define $g(x,\mu) = \int_0^x h(u,\mu)du$.

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

$$
\begin{aligned}
&= \int_0^\infty\int_0^x \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dudx - \int_{-\infty}^0\int_x^0 \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dudx \\
&= \int_0^\infty\int_u^\infty \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dxdu - \int_{-\infty}^0\int_{-\infty}^u \frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-\mu)^2}{2t}}h(u)dxdu \\
&= \int_0^\infty \left\{1-\Phi\left(\frac{u-\mu}{\sqrt{t}}\right)\right\}h(u)du - \int_{-\infty}^0 \Phi\left(\frac{u-\mu}{\sqrt{t}}\right)h(u)du. \qquad (23)
\end{aligned}
$$

Step 3. By the heat equation identity,

$$
\begin{aligned}
\frac{d}{dt}E(g(X)) &= \frac{1}{2}E(g_{xx}(X)) \\
&= \frac{1}{2}E(h_x(X)). \qquad (24)
\end{aligned}
$$

Step 4. Therefore, by (23),

$$\frac{1}{2}E(h_x(X)) = \frac{d}{dt}\Big[\int_0^{\infty}\Big\{1-\Phi\Big(\frac{u-\mu}{\sqrt{t}}\Big)\Big\}h(u)du - \int_{-\infty}^{0}\Phi\Big(\frac{u-\mu}{\sqrt{t}}\Big)h(u)du\Big]$$

$$= \int_{-\infty}^{\infty}\frac{u-\mu}{2t^{3/2}}\phi\Big(\frac{u-\mu}{\sqrt{t}}\Big)h(u)du, \tag{25}$$

on differentiation.

But (25) $= \frac{1}{2t}E((X-\mu)h(X))$, 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 provide a few applications of identity (1). First we present a 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 1. 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 (1) and in addition assume that $E(|\Delta_x g(\underset{\sim}{X}, \underset{\sim}{\mu})|) < \infty$. Then

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

$$\text{where } \; e(\underset{\sim}{\mu}, t) = e_g(\underset{\sim}{\mu}, t) = \frac{1}{4\pi^{\frac{p}{2}}}\int (\Delta_x g)||\underset{\sim}{x} - \underset{\sim}{\mu}||^{2-p}\Gamma(\frac{p}{2}-1, \frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{2t})d\underset{\sim}{x}. \tag{26}$$

Proof:

Step 1. By identity (1), 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:

$$
\begin{aligned}
& E_{\underset{\sim}{\mu},t}(g(\underset{\sim}{X}, \underset{\sim}{\mu})) - g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) \\
&= \int_0^t \frac{\partial}{\partial s} E_{\underset{\sim}{\mu},s}(g(\underset{\sim}{X}, \underset{\sim}{\mu}))ds \\
&= \frac{1}{2}\int_0^t E_{\underset{\sim}{\mu},s}(\Delta_x g)ds. \qquad (27)
\end{aligned}
$$

Step 2. Now,

$$
\begin{aligned}
& \int_0^t E_{\underset{\sim}{\mu},s}(\Delta_x g)ds \\
&= \int_0^t \int \frac{1}{(2\pi s)^{\frac{p}{2}}} e^{-\frac{1}{2s}||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}(\Delta_x g)d\underset{\sim}{x}ds \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \int (\Delta_x g) \int_0^t \frac{e^{-\frac{1}{2s}||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}}{s^{p/2}} ds d\underset{\sim}{x}. \qquad (28)
\end{aligned}
$$

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 (26) follows after a few lines of algebra.

5.2. Reduction to Useful Forms

We will now show that the general moment formula (26) reduces to more useful forms. The reduction presented below is as follows:

a. for p =1,2, and 3, formula (26) 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 1. Among the immediate applications of these bounds are an improvement as well as a reversal of Jensen's inequality for convex functions;

b. for $p \geq 4$, formula (26) will in fact be reduced to a considerably more useful exact form if p is even. This latter exact formula (formula (38) below) for $E(g(\underset{\sim}{X}, \underset{\sim}{\mu}))$ is a surprising reduction and leads to lower and upper bounds again. 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 $p \leq 3, \Gamma\left(\frac{p}{2} - 1, z\right) \geq \frac{e^{-z} z^{\frac{p}{2}-1}}{z+2-\frac{p}{2}}$; (29)

b. For $p = 1, \Gamma\left(\frac{p}{2} - 1, z\right) \leq \frac{2e^{-z}}{\sqrt{z}}$; (30)

For $p = 2, \Gamma\left(\frac{p}{2} - 1, z\right) \leq e^{-z}(1 + |\log z|)$; (31)

For $p = 3, \Gamma\left(\frac{p}{2} - 1, z\right) \leq \frac{e^{-z}}{\sqrt{z}}$; (32)

c. If $p \geq 4$ and is even, $\Gamma\left(\frac{p}{2} - 1, z\right) = \left(\frac{p}{2} - 2\right)! e^{-z} \sum_{i=0}^{\frac{p}{2}-2} \frac{z^i}{i!}$. (33)

Proof:

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

$$\Gamma(\alpha, z) = \frac{e^{-z} z^{\alpha}}{\Gamma(1-\alpha)} \int_0^{\infty} \frac{e^{-t} t^{-\alpha}}{z+t} dt \tag{34}$$

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

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

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

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 definitely 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.

Proposition 2. 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 1.

a. If $p = 1$,

$$E(g(X, \mu)) \leq g(\mu, \mu) + t \cdot E(|g_{xx}(X, \mu)|). \tag{35}$$

If $p = 2$,

$$E(g(\underset{\sim}{X}, \underset{\sim}{\mu})) \leq g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + \frac{t}{2} \cdot E\left(\left|\Delta_x g\left(\underset{\sim}{X}, \underset{\sim}{\mu}\right)\right|\left(1 + \left|\log \frac{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}{2t}\right|\right)\right). \tag{36}$$

If $p = 3$

$$E(g(\underset{\sim}{X}, \underset{\sim}{\mu})) \leq g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + t \cdot E\left(\frac{|\Delta_x g(\underset{\sim}{X}, \underset{\sim}{\mu})|}{\frac{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}{t}} \right). \tag{37}$$

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

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

$$\left\{ \sum_{j=1}^{\frac{p}{2}-1} \frac{2^{j-1}}{(\frac{p}{2} - j - 1)!} E\left(\frac{\Delta_x g(\underset{\sim}{X}, \underset{\sim}{\mu})}{\left(\frac{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}{t} \right)^j} \right) \right\} \tag{38}$$

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

$$E(g(X, \mu)) \geq g(\mu, \mu) + t \cdot E\left(\frac{g_{xx}(X, \mu)}{\frac{(X-\mu)^2}{t} + 3} \right). \tag{39}$$

If $p = 2$ and g is subharmonic,

$$E(g(\underset{\sim}{X}, \underset{\sim}{\mu})) \geq g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + t \cdot E\left(\frac{\Delta_x g(\underset{\sim}{X}, \underset{\sim}{\mu})}{\frac{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}{t} + 2} \right). \tag{40}$$

If $p = 3$ and g is subharmonic,

$$E(g(\underset{\sim}{X}, \underset{\sim}{\mu})) \geq g(\underset{\sim}{\mu}, \underset{\sim}{\mu}) + t \cdot E\left(\frac{\Delta_x g(\underset{\sim}{X}, \underset{\sim}{\mu})}{\frac{||\underset{\sim}{X} - \underset{\sim}{\mu}||^2}{t} + 1} \right). \tag{41}$$

Proof: The bounds for $p = 1, 2, 3$ and the equality of $p \geq 4$ all follow on combining the basic moment formula (26) with Lemma 1; we therefore 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 Proposition 2.

Example 1. 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

$$\pi(\underset{\sim}{\mu}) = \frac{\Gamma(\frac{\alpha+p}{2})}{(\alpha\pi)^{p/2}\Gamma(\frac{\alpha}{2})} \frac{1}{(1 + \frac{\underset{\sim}{\mu}'\underset{\sim}{\mu}}{\alpha})^{\frac{\alpha+p}{2}}}; \tag{42}$$

see Berger (1986).

The marginal density of $\underset{\sim}{X}$ is

$$m(\underset{\sim}{x}) = \int \frac{1}{(2\pi t)^{\frac{p}{2}}} e^{-\frac{||\underset{\sim}{x}-\underset{\sim}{\mu}||^2}{2t}} \pi(\underset{\sim}{\mu}) d\underset{\sim}{\mu}, \tag{43}$$

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 Proposition 2 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 (37), then we have, uniformly in $\underset{\sim}{x}, m(\underset{\sim}{x}) \le \pi(\underset{\sim}{x}) + \frac{12t}{\pi^2}$, a simple bound (of the correct order).

Example 2. Improving on Jensen's Inequality. Of course, in general, for a convex function g, one only can assert that $E(g(X)) \ge g(E(X))$. (39) 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 $f'(t) \ge a \ge 0$ and $f''(t) \ge b \ge 0$, then $g''(z) \ge 2(a + 2bz^2)$ and so, if we apply (39), then we get

$$\begin{aligned} E(g(Z)) \ge& g(0) + 2E\frac{a+2bZ^2}{Z^2+3} \\ &= g(0) + 4(b + (\phi(\sqrt{3}) - 1)e^{\frac{3}{2}}\sqrt{\frac{\pi}{6}}(66 - a)) \\ &= g(0) + .54a + .766, \end{aligned} \tag{44}$$

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

6. APPLICATION TO DECISION THEORY

The heat equation identity (1) and the canonical normal identity (6) are now used to provide some applications to decision theory.

i) We 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.

ii) We show that the heat equation identity produces a totally new proof of a Bayes risk lower bound previously derived using altogether different means.

6.1. 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 (6) is the key. We shall also see how an upwardly biased estimate of risk and thence certain inadmissibility results follow from this inequality. It should be remarked that inequality (46) below can be derived from Stein's identity itself.

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 $\underset{\sim}{h}: \quad \mathbb{R}^p \to \mathbb{R}^p$ has the following properties:

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

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

$\underline{\text{a}}$ Then, $E(\underset{\sim}{T}_0'\underset{\sim}{h}(\underset{\sim}{T}_o)) \leq \frac{m}{m+p-2}E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)).$ (45)

$\underline{\text{b}}$ 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,

$$E((\underset{\sim}{T} - \underset{\sim}{\mu})'\underset{\sim}{h}(\underset{\sim}{T})) \leq \frac{m}{m+p-2}E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T})). \tag{46}$$

Proof: The proof uses the multidimensional version of the canonical normal identity (6). 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

$$\frac{\partial}{\partial t}E(g(\underset{\sim}{Z}, Y)) = \frac{1}{2n}E(\Delta_z g(\underset{\sim}{Z}, Y)) + \frac{1}{t}E((Y)g_y(\underset{\sim}{Z}, Y)). \tag{47}$$

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 (47) 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,

$$
\begin{aligned}
\Delta_z g(\underset{\sim}{z}, y) &= \frac{mn}{y} \nabla \cdot \underset{\sim}{h}(\underset{\sim}{t}_0), \\
\text{and} \quad g_y(\underset{\sim}{z}, y) &= -\frac{1}{2y} \sum_{i=1}^{p} t_{0i} h_i(\underset{\sim}{t}_0),
\end{aligned}
\tag{48}
$$

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

Step 4. Therefore, by identity (47),

$$
E\left(\sum_{i=1}^{p} T_{0i} h_i(\underset{\sim}{T}_0)\right) = mt \cdot E\left(\frac{\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)}{Y}\right). \tag{49}
$$

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,

$$
\begin{aligned}
E\left(\sum_{i=1}^{p} T_{0i} h_i(\underset{\sim}{T}_0)\right) &= mt \cdot E\left(\frac{\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)}{Y}\right) \\
&\leq mt \cdot E\left(\frac{\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)}{Y + n\underset{\sim}{Z}'\underset{\sim}{Z}}\right) \\
&\quad \text{(since by assumption } \nabla \cdot \underset{\sim}{h} \leq 0) \\
&= mt \cdot E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)) \cdot E\left(\frac{1}{Y + n\underset{\sim}{Z}'\underset{\sim}{Z}}\right) \\
&= \frac{m}{m+p-2} E(\nabla \cdot \underset{\sim}{h}(\underset{\sim}{T}_0)),
\end{aligned}
\tag{50}
$$

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

(50) proves part $\underline{\text{a}}$ of Lemma 2. Part $\underline{\text{b}}$ follows from part $\underline{\text{a}}$.

Corollary 3. Let $f:\ \mathbb{R}^p \to \mathbb{R}^1$ be superharmonic and define $h = \nabla f$. Then

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

Corollary 3 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 $\Delta f \leq 0$.

6.2. Proving Inadmissibility from Biased Estimates of Risk

Lemma 2 and Corollary 3 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 3. 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 (46). For $m > 2$,

$$\begin{aligned} &E(||\underset{\sim}{T} + \underset{\sim}{h}(\underset{\sim}{T}) - \underset{\sim}{\mu}||^2) - E(||\underset{\sim}{T} - \underset{\sim}{\mu}||^2) \\ &\leq E(||\underset{\sim}{h}||^2 + \frac{2m}{m+p-2} \nabla \cdot h). \end{aligned} \tag{51}$$

Proposition 3 is evident because $\underset{\sim}{h}$ is assumed to satisfy inequality (46). The domination result to follow from this is given next.

Proposition 4. 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 3 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.

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 4 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 4) for $p \geq 3$.

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

6.3. Application to Bayesian Statistics

The heat equation identity can be usefully exploited to study Bayes risks in point estimation problems. Specifically, (1) leads to identities and bounds for Bayes risks. 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.

6.3.1. A Bayes Risk Identity

Proposition 5. Let $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, t\Sigma), p \geq 1$, 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

$$\frac{d}{dt}r(t,G) = E_m[\sum_{k=1}^{p}(\nabla\delta_{G,k})'\Sigma(\nabla\delta_{G,k})], \tag{52}$$

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

Corollary 4. Let $X \sim N(\mu, t)$ and let $\mu \sim G$. Then the Bayes risk $r(t,G)$ satisfies

$$\frac{d}{dt}r(t,G) = E_m\delta_G'(X)^2. \tag{53}$$

Since the proof of (52) is essentially the same as that of (53), we will only prove Corollary 4.

Proof of Corollary 4: Consider the function $g(x,\mu,t) = (\delta_G(x,t) - \mu)^2$. By the expectation identity (1),

$$\begin{aligned}\frac{d}{dt}r(t,G) &= \frac{d}{dt}\int E_\mu(g(X,\mu,t))dG(\mu)\\ &= \frac{1}{2}E_G E_\mu(g_x''(X,\mu,t)) + 2\int\int(\delta_G(x,t)-\mu)\frac{\partial}{\partial t}\delta_G(x,t)p(x|\mu,t)dxdG(\mu)\end{aligned} \tag{54}$$

now note that the second term in (54) 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 (54), $\frac{d}{dt}r(t,G) = E_m(\delta_G'(X,t))^2$.

Remark: A minor but immediate consequence of Proposition 5 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 5 goes further by laying out explicitly what $\frac{d}{dt}r(t,G)$ equals, not just that it is > 0. This exact formula is of some interest.

6.3.2. Bayes Risk Bounds

We will now show how one can obtain lower bounds on $r(t, G)$ by using the identity of Proposition 5. The derivation of the bound, as we shall now see, may seem to be strange! Not only shall we use the apparently new identity (53), 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 (1990)), 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 6 below is attained when G is a normal prior.

Proposition 6 Let $X \sim N(\mu, t)$ and let $\mu \sim G$. Then

$$r(t, G) \geq \frac{t}{1 + tI(G)} \tag{55}$$

where $I(G)$ denotes the Fisher information of G. (Note that (55) is formally valid even if $I(G) = \infty$)

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

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

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

Step 2. By Step 1 and Corollary 4,

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

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,

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

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,

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

completing the proof.

Discussion

Bounds similar to the one in Proposition 6 are obtainable by our methods for the case $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, t\Sigma)$ by using Proposition 5 and the Brown identity for $r(t, G)$ and $\delta_G(x, t)$ for the $N_p(\underset{\sim}{\mu}, t\Sigma)$ case.

7. DIFFERENTIAL EQUATIONS DRIVING A MOMENT SEQUENCE

We will now show that it follows from the heat equation identity (1) 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 $a_{0,n} t^n h(t) + a_{1,n} t^{n+1} h'(t) + \ldots + a_{n,n} t^{2n} h^{(n)}(t) - E((X - \mu)^n g(X)) \equiv 0$. A similar equation holds for $E_{\mu,t}((X - \mu)^{2n+1} g(X))$. We find it interesting that such a universal differential equation should hold at all. We then provide some applications, in particular an application to counting perfect matchings in graph theory.

7.1. 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,

$$x^{2n} = \sum_{i=0}^{n} \frac{(2n)! 2^i}{(n-i)!(2i)! 2^n} H_{2i}(x), \tag{56}$$

and

$$x^{2n+1} = \sum_{i=0}^{n} \frac{(2n+1)!2^i}{(n-i)!(2i+1)!2^n} H_{2i+1}(x). \tag{57}$$

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

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

$$E_t(X^{2n}g(X)) = \sum_{i=0}^{n} a_{i,n} t^{n+i} \frac{d^i}{dt^i}(E_t g(X)), \tag{58}$$

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

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

$$E_t(X^{2n+1}g(X)) = \sum_{i=0}^{n} c_{i,n} t^{n+i+1} \frac{d^i}{dt^i}(E_t g'(X)), \tag{59}$$

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

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 7, are tabulated below for $n \leq 5$ for the user's convenience.

Table 1

n / i	$a_{i,n}$ 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

8. 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.

8.1. 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

$$\mu_G(x) = \sum_{r=0}^{[\frac{n}{2}]} (-1)^r p(G, r) x^{n-2r}. \tag{60}$$

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.

8.2 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 3.

a. The matching polynomial of the disjoint union of any two graphs G and H satisfies
$$\mu_{G \cup H}(x) = \mu_G(x)\mu_H(x). \tag{61}$$

b. For any graph G, $\frac{d}{dx}\mu_G(x) = \sum_u \mu^{(x)}_{G-u}$. $\quad$ (62)

c. For any graph G, the total number of perfect matchings in $\overline{G}$ satisfies $\psi(\overline{G}) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \mu_G(x) e^{-\frac{x^2}{2}} dx.$ (63)

Remark. Formula (63) manifests the enjoyably surprising connection between matchings in graphs and the normal distribution. It will also be a key ingredient for our result.

8.3. 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 8. 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

$$\psi(\overline{G \cup H}) = \sum_{i=0}^{n} \frac{2^i}{(2i)!} \alpha_{i,n,H} (\Sigma_i \psi(\overline{G - u_1, u_2 \ldots u_{2i}})), \tag{64}$$

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

$$\alpha_{i,n,H} = \sum_{r=i}^{n} (-1)^{n-r} \frac{(2r)!}{2^r (r-i)!} p(H, n-r). \tag{65}$$

Proof: The principal tools for this identity on perfect matchings are Lemma 3 above and Theorem 7 in section 7.1. Towards this end,

$$\psi(\overline{G \cup H}) = \int \mu_{G \cup H}(x) \phi(x) dx \qquad \text{(equation (63))}$$

$$= \int \mu_G(x) \mu_H(x) \phi(x) dx \qquad \text{(equation (61))}$$

$$= \sum_{r=0}^{n} (-1)^r p(H, r) \int x^{2n-2r} \mu_G(x) \phi(x) dx \qquad \text{(equation (60))}$$

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

(equation (58))

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

(iteration of the heat equation identity (1))

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

(iteration of equation (62))

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

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

(equation (63) 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 5.

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

$$\psi(\overline{G\cup H}) = (1-p(H,1))\psi(\overline{G}) + \sum\sum_{v\neq u}\psi(\overline{G-uv}). \tag{66}$$

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

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

$$+\sum\sum\sum\sum_{z\neq w\neq v\neq u}\psi(\overline{G-uvwz}). \tag{67}$$

Proof: Each part follows on some algebra from Theorem 8.

REFERENCES

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

Basu, D., 1956, On Statistics Independent of a Complete Sufficient Statistic, *Sankhya*, **15**, 377–380.

Berger, J., 1986, Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Brandwein, A. C. and Strawderman, W. E., 1990, Stein Estimation: The spherically Symmetric Case, *Stat. Sc.*, **5**, 356–369.

Brown, L. D., 1971, Admissible Estimators, Recurrent Diffusions, and Insoluble Boundary Value Problems, *Ann. Math. Stat.*, **42**, 855–903.

Brown, L. D., 1986, Fundamentals of Exponential Families, IMS, Hayward, California.

Brown, L. D. and Gajek, L., 1990, Information Inequalities for the Bayes Risk, *Ann. Stat.*, 18, 1578–1594.

Cellier, D. and Fourdrinier, D., 1995, Shrinkage Estimators Under Spherical Symmetry for the General Linear Model, *J. Mult. Anal.*, **52**, 338–351.

Diaconis, P. and Zabell, S., 1991, Closed form Summation for Classical Distributions, *Stat. Sc.*, **6**, 284–302.

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

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

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

Gradsheyn, I. and Ryzhik, I., 1980, Table of Integrals, Series and Products, Academic Press, New York.

Lehmann, E. L. and Casella, G., 1998, Theory of Point Estimation, Chapman and Hall, New York.

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

Rudin, W., 1974, Real and Complex Analysis, McGraw Hill, New York.

Stein, C., 1981, Estimation of the Mean of a Multivariate Normal Distribution, *Ann. Stat.*, **9**, 1135–1151.

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