isid/ms/2008/05
July 18, 2008
http://www.isid.ac.in/~statmath/eprints

On Stein's Identity and Its Application

Sudheesh Kumar Kattumannil

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

ON STEIN'S IDENTITY AND ITS APPLICATION

Sudheesh Kumar Kattumannil[*,†]

* Indian Statistical Institute, Delhi Centre,

New Delhi, India.

Abstract. Stein's identity and its role in inference procedures have been discussed widely in literature. We extend the identity to a general framework using an absolutely continuous function $g(x)$ that characterizes the probability distributions. It is shown that some of the identities available in the literature are special cases of the proposed one. Further, we also discuss various applications of the proposed identity.

Keywords: Stein's identity, exponential family, Pearson family, Generalized Pearson family, Gini index, sample moments.

1. Introduction

Charles Stein (1973) introduced a natural identity for a random variable whose distribution belongs to an exponential family. As a special case, if $X \sim N(\mu, \sigma^2)$ and $c(x)$ is a differentiable function satisfying $E(c'(X)) < \infty$, then

$$E(c(X)(X - \mu)) = \sigma^2 E(c'(X)).$$

This has come to be known in literature as Stein's identity or Stein's lemma. And it has been widely exploited since; it is discussed with some members of certain families of distributions (see Arnold et al. (2001), Landsman (2006), Brown et al. (2006) and Landsman and Neslehova (2008)).

Inspired from Stein, Hudson (1978) obtained an identity for the exponential family of distributions and studied its uses in multi parameter estimation. Prakasa Rao (1979) characterized the exponential family of distributions using the identity given by Hudson (1978), and then

†*Author E-mail*: skkattu@yahoo.co.in.

used the characterization results to establish some limit theorems. Arnold et al. (2001) introduced a multivariate version of Stein's identity and then applied the results for deriving consistent moment based estimates of parameters. They also considered the estimation of parameters in bivariate settings where the conditional distributions belong to exponential family. Nicoleris and Sagris (2002) examined the prediction of a random function of a parameter by means of Stein's identity. Landsman (2006) generalized the Stein's identity to the case of elliptical class of distributions. The results were illustrated through multivariate generalized Student-t distribution. Brown et al. (2006) derived an expectation identity using the heat equation and then showed that the identity is equivalent to Stein's identity. They pointed out series of applications to the area of probability and statistics. The ongoing interest in dealing with Stein's identity and its applications in different dimensions inspired us to generalize it to the distributions belong to a wider class of continuous probability distributions satisfying specific conditions. Then we look for some applications that were not considered by the earlier researchers in this context.

The rest of the paper is organized as follows. In Section 2 we derive the generalized Stein's identity and then deduce the identity given by Hudson (1978) as a special case. Also we obtain the exact expression for the proposed identity for the distributions belong to Pearson family and generalized Pearson family. In Section 3 we discuss various applications of the proposed identity.

2. **Generalized Stein's Identity**

Associated with an absolutely continuous random variable X with support $-\infty \leq < a < X < b \leq \infty$, let $f(x)$, $F(x)$ be the distribution function and density function respectively. Also let $h(x)$ be a Borel measurable function of the random variable X such that $E(h^2(X)) < \infty$ and $E(h(X)) = E(X) = \mu$.

Defining $\mathcal{A}$ to be the class of all absolutely continues function with derivatives $c'(x)$ defined on the range of X we present the following theorem as a generalization of the Stein's identity.

Theorem 1. Let X be an absolutely continuous random variable with support $-\infty \leq a < X < b \leq \infty$. If the density function $f(x)$ satisfies the differential equation

$$f'(x)/f(x) = -g'(x)/g(x) + (\mu - h(x)/g(x), \tag{2.1}$$

for some function $g(x)$, then for any $c(x)$ in $\mathcal{A}$ satisfying $E|c(X)h(X)| < \infty$, $E(c^2(X)) < \infty$, $E|g(X)c'(X)| < \infty$, we have the following identity

$$E(c(X)(h(X) - \mu)) = E(c'(X)g(X)), \tag{2.2}$$

provided $lim_{x \to b} g(x)f(x) = 0$.

Proof: Consider

$$E(c(X)(h(X) - \mu)) = \int_a^b c(x)(h(x) - \mu)f(x)d(x)$$

$$= \int_a^b (h(x) - \mu)(c(x) - c(a))f(x)d(x)$$

$$= \int_a^b (h(x) - \mu)(\int_a^x c'(t)d(t))f(x)d(x).$$

Applying Fubini's Theorem

$$E(c(X)(h(X) - \mu)) = \int_a^b c'(t)(\int_t^b (h(x) - \mu)f(x)d(x))d(t). \tag{2.3}$$

Now, the equation (2.1) can be written as

$$g(x)f'(x) = -g'(x)f(x) + (\mu - h(x)f(x).$$

Integrating with respect to x from t to ∞ and basis of the assumption $lim_{x \to b} g(x)f(x) = 0$

$$-g(t)f(t) - \int_t^b g'(x)f(x)d(x) = -\int_t^b g'(x)f(x)d(x) + \int_t^b (\mu - h(x)f(x)d(x),$$

which simplifies to

$$\int_t^b (h(x) - \mu)f(x)d(x) = g(t)f(t). \tag{2.4}$$

Substituting (2.4) in (2.3)

$$E(c(X)(h(X)-\mu)) = \int_a^b c'(t)g(t)f(t)d(t).$$

$$= E(c'(X)g(X)).$$

Hence the proof is completed.

Remark 1. The differential equation (2.1) can be written as

$$d\log f(x)/dx + d\log g(x)/d(x) = (\mu - h(x))/g(x).$$

Integrating with respect to x from t to b and assuming $lim_{x\to b} g(x)f(x) = 0$

$$\log f(t) + \log g(t) = \int_t^b ((\mu - h(x)/g(x))d(x),$$

which gives

$$f(t) = [g(t)]^{-1}\exp(\int_t^b ((\mu - h(x)/g(x))d(x)),$$

Hence, for a given $h(x)$ the value of $g(x)$ uniquely determines the distribution of X.

Remark 2. When the distribution of X belongs to exponential family, the identity (2.2) reduces to the moment identity given by Hudson (1978).

The statement made in the Remark 2.2 is reflected upon in the following theorem.

Theorem 2. Suppose that the distribution of X belongs to the exponential family with probability density

$$f(x) = \exp\{\theta x - \varphi(\theta)\}k(x), -\infty \le x \le \infty$$

Let

$$t(x) = -k'(x)/k(x),$$

then for any absolutely continuous function $c(x)$ on R

$$E(c(X)(t(X)-\theta)) = E(c'(X)). \tag{2.5}$$

Proof: Consider

$$d \log f(x)/dx = \theta + k'(x)/k(x),$$

which can be written as

$$f'(x) = (\theta - t(x))f(x).$$

Integrating with respect to x from $-\infty$ to t and assuming $lim_{x \to -\infty} f(x) = 0$

$$f(t) = \int_{-\infty}^{\infty} (\theta - t(x))f(x)d(x). \tag{2.6}$$

By simple algebra, we can show that the differential equation (2.1) takes the form

$$\int_{-\infty}^{x} (\mu - h(x))f(x)d(x) = f(t)g(t),$$

and comparing with (2.6) we get

$$h(x) = t(x) - \theta + \mu \text{ and } g(x) = 1.$$

For the above choice of $h(x)$ and $g(x)$ the identity (2.2) reduces to (2.5).

For modelling and inference purposes, considering families of distributions is more desirable since it enables the results to be deduced for individual distributions. Thus we derive the exact expression for the identity (2.2) for some well-known families like Pearson family and generalized Pearson family. And then give some examples to check the validity of the results.

Theorem 3. Suppose that the distribution of X belongs to the Pearson family specified by

$$f'(x)/f(x) = -(x + d)/(a_0 + a_1 x + a_2 x^2), -\infty \leq x \leq \infty,$$

then

$$E(c(X)(X - \mu)) = E(c'(X)(b_0 + b_1 X + b_2 X^2)). \tag{2.7}$$

with $b_i = a_i/(1 - 2a_2), a_2 \neq 1/2, a_i \in R, i = 0, 1, 2.$

Proof: Nair and Sankaran (1991) showed that the random variable X belongs to the Pearson family if and only if

$$E(X|X > x) = \mu + (b_0 + b_1 x + b_2 x^2)k(x).$$

where $k(x) = f(x)/(1 - F(x))$ and b_i's are as stated in the theorem. The above identity can be written as

$$\int_x^\infty (x - \mu) f(x) d(x) = (b_0 + b_1 x + b_2 x^2) f(x).$$

Choosing $h(x) = x$ and comparing with (2.4) we get $g(x) = (b_0 + b_1 x + b_2 x^2)$. Then the identity (2.2) deduces to (2.7).

Remark 3. The normal random variable with mean μ and variance σ^2 belongs to the Pearson family with $a_0 = \sigma^2$,$a_1 = a_2 = 0$, hence the identity (2.7) reduces to

$$E(c(X)(X - \mu)) = \sigma^2 E(c'(X)),$$

which is the well known Stein's identity.

Remark 4. By successive application of (2.2) we can arrive at the following results

(1) $E(c(X)(h(X) - \mu)^2) = E(c''(X)g^2(X) + c'(X)g'(X)g(X) + c(X)h'(X)g(X))$

(2) $E(c(X)(h(X) - \mu)^4) = E(c^{iv}(X)g^4(X) + 6c'''(X)g'(X)g^3(X) + 6c''(X)h'(X)g^3(X)$
$+4c''(X)g''(X)g^3(X) + 7c''(X)(g'(X))^2 g^2(X) + 8c'(X)h''(X)g^3(X)$
$+14c'(X)h'(X)g'(X)g^2(X) + c'(X)g'''(X)g^3(X) + 4c'(X)g''(X)g'(X)g^2(X)$
$+c'(X)(g'(X))^3 g(X) + 3c(X)h'''(X)g^3(X) + 9c(X)h''(X)g'(X)g^2(X)$
$+3c(X)(h'(X))^2 g^2(X) + 3c(X)h'(X)g''(X)g^2(X) + 3c(X)h'(X)(g'(X))^2 g(X))$

When $X \sim N(\mu, 1)$ we have $g(x) = 1$ with $h(x) = x$ so that $h'(x) = 1$, $h''(x) = 0$, $g'''(x) = g''(x) = g'(x) = 0$ and the above results reduces to the Lemma 4 of Stien (1981).

In an effort to improve the richness in members of the Pearson family and there by extend the domain of application, Sindhu (2003) has replaced the linear term in the differential equation stated in Theorem 3 by a quadratic, obtaining

$$f'(x)/f(x) = (b_0 + b_1 x + b_2 x^2)/(a_0 + a_1 x + a_2 x^2). \tag{2.8}$$

Besides containing all the members of the Pearson family (corresponding to $b_2 = 0$),this extended Pearson system consists of many new members like the Inverse Gaussian, Random Walk, Maxwell and Rayleigh distributions. And the next theorem is concerned about the generalized Pearson family.

Theorem 4. Let X be the random variable belongs to the generalized Pearson family specified by (2.8), then

$$E(c(X)(pX^2 + qX + r)) = -E(c'(X)(a_2X^2 + a_1X + a_0)). \tag{2.9}$$

with $p = b_2$, $q = b_1 + 2a_2$ and $r = b_0 + a_1$.

Proof: The generalized Pearson family is characterized by the property (See Sankaran et al. (2003))

$$E((b_2X^2 + (b_1 + 2a_2)X + b_0 + a_1 + \mu)|X > x) = \mu - (a_2x^2 + a_1x + a_0)k(x),$$

this gives

$$\int_x^\infty (b_2x^2 + (b_1 + 2a_2)x + b_0 + a_1)f(x)d(x) = -(a_2x^2 + a_1x + a_0)f(x).$$

Taking $h(x) = px^2 + qx + r + \mu$, and comparing with the identity (2.4) we get $g(x) = -(a_2x^2 + a_1x + a_1)$. Substituting the values of $h(x)$ and $g(x)$ in (2.2) the identity (2.9) follows.

Example 1. The Rayleigh distribution with probability density function

$$f(x) = 2\lambda x \exp(-\lambda x^2),$$

belongs to the family specified by (2.8) and we find $g(x) = -x$ with $h(x) = -2\lambda x^2 + 2 + \mu$. And the identity (2.9) takes the form

$$2E(c(X)(1 - \lambda X^2)) = E(Xc'(X))$$

Example 2. For the inverse Gaussian distribution with probability density function

$$f(x) = \sqrt{\lambda/2\pi x^3}\exp(-\lambda(x-\alpha)^2/2x\alpha^2)$$

we have $g(x) = -2\alpha^2x^2$ with $h(x) = -\lambda x^2 + \alpha^2 x + \lambda\alpha^2 + \mu$. And the moment identity (2.9) becomes

$$E(c(X)(\lambda\alpha^2 + \alpha^2X - \lambda X^2)) = 2\alpha^2E(X^2c'(X)).$$

So far we have specialized Theorem 1 to some families of distributions that include many of the continuous distributions used in common. However there are some important distributions like the Weibull, Burr that are not members of these families, and at the same time quite useful in lifetime analysis. Our general framework in the Theorem 1 permits us to include them also with appropriate choice of $h(x) = x$ in each case as illustrated in the following example.

Example 3. For the Weibull distribution with probability density function

$$f(x) = (\alpha/\beta)(\alpha/\beta)^{\alpha-1}\exp(-(x/\beta)^{\alpha}), x > 0, \alpha, \beta > 0$$

we have $h(x) = x^{\alpha} + \mu - \beta^{\alpha}$ and $g(x) = \alpha^{-1}x\beta^{\alpha}$. And the identity (2.2) will takes the form

$$E(c(X)(X^{\alpha} - \beta^{\alpha})) = \alpha^{-1}\beta^{\alpha}E(Xc'(X)).$$

3. **Applications**

Most of the researchers discussed the Stein's identity in connection with estimation problems in the classical as well as in the Bayesian setup (see Stein (1981), Arnold et al. (2001), Nicoleris and Sagris (2002), Landsman (2006), Brown et al (2006), Landsman and Neslehova (2008)). Their views can be easily extended to the case where the probability distributions satisfy the differential equation (2.1). In this section we discuss the application of the identity (2.2) in economics problem.

(a) Lorenz curve and Gini index

The Lorenz curve and Gini index is extensively used in studying income inequalities and the related characterization problems. We suggest an alternate form for the same using the identity (2.2). The proposed form is very useful when finding these quantities for specified probability distributions. We also deduce the alternate form given by Tziafetas (1989). Here we confined the support of X to $(0, \infty)$.

Definition 1. Lorenz curve for an absolutely continuous positive random variable X is defined as the graph of the ratio

$$L(F(x)) = E(X|X \leq x)F(x)/E(X) \tag{3.1}$$

to $F(x)$. If X represents the annual income $L(p)(p = F(x))$ is the proportion of total income that accrues to individual having the lowest income.

Definition 2. Gini index denoted by $c(x)$ is defined as (Tse (2006))

$$c(x) = 2\int_0^1 (p - L(p))d(p) = 1 - 2\int_0^1 L(p)d(p) \tag{3.2}$$

The following theorem suggests an alternate form for Lorenz curve and Gini index.

Theorem 5. Suppose that the random variable X belongs to the class of distributions specified by the differential equation (2.1), then

(i) Lorenz curve is given by

$$L(F(x)) = F(x) - \mu^{-1} f(x) g(x)$$

(ii) Gini index is given by

$$c(x) = 2E(f(X)g(X))/E(X) \tag{3.3}$$

Proof:(i) Choosing $h(x) = x$, the identity (2.4) reduces to the form

$$E(X|X \leq x)F(x) = \mu F(x) - f(x)g(x)$$

Substituting in equation (3.1) we have the alternate form given in the theorem.

(ii) By definition

$$c(x) = 1 - 2\int_0^1 L(p)d(p), p = F(x)$$

Consider

$$\int_0^1 L(F(x))d(F(x)) = \int_0^1 (E(X|X \leq x)F(x)/E(X))d(F(x))$$

$$= (1/\mu)\int_0^1 (\int_0^x tf(t)d(t))d(F(x)).$$

Changing Riemann- Stieltjes integral to Riemann integral

$$\int_0^\infty L(F(x))d(F(x)) = (1/\mu)\int_0^\infty (\int_0^x tf(t)d(t))f(x)d(x).$$

Applying Fubini's Theorem

$$\int_0^1 L(F(x))d(F(x)) = (1/\mu)\int_0^\infty (\int_t^\infty f(x)d(x))tf(t)d(t).$$

$$= (1/\mu)\int_0^\infty (1 - F(t))tf(t)d(t)$$

$$= 1 - (1/\mu)E(XF(X))$$

$$= 1/2 - (1/\mu)E((X - \mu)F(X)),$$

since $E(F(X)) = 1/2$. Using identity (2.2) the above equation can be written as

$$\int_0^1 L(F(x))d(F(x)) = 1/2 - (1/\mu)E(f(X)g(X)).$$

Substituting in equation (3.2) we get the alternate form (3.3).

Example 4. Consider the beta distribution with probability density function

$$f(x) = (1/B(p,q))x^{p-1}(1-x)^{q-1}, 0 < x < 1.$$

Here $g(x) = x(1-x)/(p+q)$. Consider

$$E(f(X)g(X)) = (1/B^2(p,q)(p+q)) \int_0^1 x^{p-1}(1-x)^{q-1}x^{p-1}(1-x)^{q-1}x(1-x)d(x).$$

$$= B(2p, 2q)/B^2(p,q)(p+q).$$

Hence the Gini index is given by

$$c(x) = 2B(2p, 2q)/B^2(p,q)p.$$

The expression (3.3) can be used even when the support of X is $(-\infty, \infty)$.

Example 5. Consider the Normal distribution with probability density function

$$f(x) = (1/\sqrt{2\pi}\sigma)\exp(-(x-\mu)^2/2\sigma^2), -\infty < x < \infty.$$

Here $g(x) = \sigma^2$ and

$$c(x) = 2E(f(X)g(X))/E(X) = (2\sigma^2/\mu)E(f(X)) = \sigma/\mu\sqrt{\pi}$$

Remark 5. The Example 4 and Example 5 are discussed by Tziafetas (1989) and McDonald (1978) respectively and our proof is much easier.

Remark 6. The proof for the alternate form of gini index given by Tziafetas (1989)

$$c(x) = 2Cov(X, F(X))/E(X)$$

is straightforward when we use generalize Stein's identity. By definition

$$Cov(X, F(X)) = E((X-\mu)F(X))$$

$$= E(f(X)g(X))(\text{using}(2.2))$$

(b) Expression for covariance between sample mean and sample variance

The joint distribution of $(\overline{X}, S^2)$ is quiet complicated for many probability distributions and hence finding the covariance between sample mean $(\overline{X})$ and sample variance (S^2) is a tedious job. Here we obtain a simple expression for the same by means of the identity (2.2). Zhang (2007) discussed about an identity connecting covariance between sample mean and sample variance and third central moment and it is given by

$$Cov(\overline{X}, S^2) = E(X - \mu)^3/n$$

Applying the identity (2.2) successively we get

$$Cov(\overline{X}, S^2) = E(g(X)g'(X))/n$$

The above identity suggests an easy way to find the covariance between sample mean and sample variance for the probability distributions satisfying the differential equation (2.1).

Remark 7. For the normal random variable described in Remark 3 $g(x) = \sigma^2$ and $g'(x) = 0$, we obtain $Cov(\overline{X}, S^2) = 0$, a well known result available in the literature.

References

1. Arnold, B.C, Castillo, E and Sarabia, J.M, 2001. A multivariate version of Stein's identity with applications to moment calculations and estimation of conditionally specified distributions. Comm. Statist. Theory Methods, 30, 2517-2542.

2. Brown, L and Das Gupta, A, 2006. The heat equations and Stein's Identity; Connections and Applications, J. Statist. Plann. Inf., 136, 2254-2278.

3. Hudson, H.M, 1978. A natural identity for exponential families with application in a multiparameter estimation. Ann. Statist., 6, 473-484.

4. Landsman, Z, 2006. On the generalization of Stein's Lemma for the elliptical class of distributions, Statist. Probab. Lett., 76, 1012-1016.

5. Landsman, Z and Neslehova, J (2008). Stein's Lemma for elliptical random vectors, J. Multivariate. Anal., 99, 912 - 927

6. McDonald, J.B, 1978. The Beta as a model for the distribution income, Brigham Young University.

7. Nair, N.U and Sankaran, P.G, 1991. Characterization of Pearson family of distribution. IEEE Trans. Rel., 40, 75-77.

8. Nicoleris, T And Sagaris, A, 2002 Random function prediction and Stein's identity. Statist. Probab. Lett., 59, 293-305.

9. Prakasa Rao, B.L.S, 1979. Characterizations of distribution through some identities, J. Appl. Probab. 16, 903-909.

10. Sankaran, P.G, Nair, N.U and Sindhu, T.K, 2003. A generalized Pearson system useful in reliability analysis. Statist. Papers, 4(1), 125-130.

11. Sindhu, T.K, 2003. An extended Pearson system useful in reliability analysis, Unpublished Ph. D thesis, Cochin University of Science and Technology.

12. Stein, C.M, 1973 Estimation of the mean of a multivariate normal distribution. Proceedings of Prague Symposium on Asymptotic Statistics, 345-381.

13. Stein, C.M, 1981. Estimation of the mean of a multivariate normal distribution, Ann. Statist., 6, 1135-1151.

14. Tziafetas, G.N, 1989. A formula for the Gini coefficient and its decomposition, Biometrical Journal,. 31, no. 8, pp. 961-967.

15. Zhang, L, 2007. Sample mean and sample variance: Their covariance and their (in) dependence, Amer. Statist., 61, 159- 160.

16. Tse, S.M (2006), Lorenz curve for truncated and censored data, Ann. Instit. Stat. Math., 58, 675- 686.