Siegel's formula via Stein's identities

Jun S. Liu
Department of Statistics
Harvard University

Abstract
Inspired by a surprising formula in Siegel (1993), we find it convenient to compute covariances, even for order statistics, by using Stein (1972)'s identities. Generalizations of Siegel’s formula to other order statistics as well as other distributions are obtained along this line.

KEY WORDS: Exponential family; multivariate Stein’s identity; covariance with order statistics.

1 Siegel’s result and Stein’s identities

In a recent paper of Siegel (1993), a remarkable formula is found for the covariance of $X_1$ with the minimum of the multivariate normal vector $(X_1, \ldots, X_n)$. The theorem is stated as follows.

**Theorem [Siegel]** Let $(X_1, \ldots, X_n)$ be multivariate normal, such that the variables are distinct, with arbitrary mean and variance structure. Then
\[
\text{cov}[X_1, \min(X_1, \ldots, X_n)] = \sum_{i=1}^{n} \text{cov}(X_1, X_i) \Pr[X_i = \min(X_1, \ldots, X_n)].
\]

On the other hand, Stein (1972) observes that for $Z \sim N(\mu, \sigma^2)$ and any function $f$ such that $E[f'(Z)] < \infty$,
\[
E[(Z - \mu)f(Z)] = \sigma^2 E[f'(Z)].
\]

Similar identities hold for many other distributions. For example, if $Z \sim \text{Poisson}(\lambda)$, then
\[
E[(Z - \lambda)f(Z)] = \lambda E[f(Z + 1) - f(Z)].
\]
If \( Z \sim t_\nu(0, \sigma^2) \), i.e., \( Z = Y / \sqrt{W / \nu} \), where \( Y \sim N(0, \sigma^2) \) and \( W \sim \chi_\nu^2 \), then

\[
E[Zf(Z)] = \sqrt{\nu / (\nu - 2)} \sigma^2 E[f'(Z^*)],
\]

where \( Z^* \sim t_{\nu-2}(\mu, \sigma^2) \). The above identities are directly proved as an application of integration by parts. To illustrate, we show the last identity when \( \sigma = 1 \):

\[
E[Zf(Z)] = \int_{-\infty}^{\infty} zf(z) \frac{\Gamma((\nu + 1)/2)}{\nu \pi \Gamma(\nu/2)} \left( 1 + \frac{z^2}{\nu} \right)^{-(\nu + 1)/2} dz
\]

\[
= \frac{\Gamma((\nu + 1)/2)}{\nu \pi \Gamma(\nu/2)} \left( \frac{\nu}{\nu - 2} \right) \int_{-\infty}^{\infty} \left( 1 + \frac{z^2}{\nu} \right)^{-(\nu - 1)/2} f(z) dz
\]

\[
= \frac{\nu}{\nu - 2} \sqrt{\nu/ (\nu - 2)} \frac{\Gamma((\nu - 1)/2)}{\Gamma(\nu/2 - 1)} \left( 1 + \frac{z^2}{\nu} \right)^{-(\nu - 1)/2} \int_{-\infty}^{\infty} f(z) dz
\]

\[
= \frac{\nu}{\nu - 2} E[f'(Z^*)].
\]

There is also a multivariate version of Stein’s identity as follows.

**Lemma 1** Let \( X = (X_1, \ldots, X_n) \) be multivariate normally distributed with arbitrary mean vector \( \mu \) and covariance matrix \( \Sigma \). For any function \( h(x_1, \ldots, x_n) \) such that \( \partial h / \partial x_i \) exists almost everywhere and \( E|\partial h / \partial x_i| < \infty \), \( i = 1, \ldots, n \), we write \( \nabla h(X) = (\partial h(X) / \partial x_1, \ldots, \partial h(X) / \partial x_n)^T \). Then the following identity is true:

\[
\text{cov}[X, h(X)] = \Sigma \ E[\nabla h(X)].
\]

Specifically,

\[
\text{cov}[X_1, h(X_1, \ldots, X_n)] = \sum_{i=1}^{n} \text{cov}(X_1, X_i) E[\frac{\partial}{\partial x_i} h(X_1, \ldots, X_n)].
\]

**Proof:** Let \( Z = (Z_1, \ldots, Z_n) \), where the \( Z_i \) are i.i.d. \( N(0, 1) \) random variables. It is straightforward to show by (1) that for any \( g(X) \), differentiable almost everywhere, \( \text{cov}[Z_i, g(Z)] = E[\partial g(Z) / \partial z_i] \). Hence

\[
\text{cov}[Z, g(Z)] = E[\nabla g(Z)].
\]

Also see Stein (1981) for a more elaborate proof. Now for the random vector \( X \), we can write \( X = \Sigma^{1/2} Z + \mu \), and \( g(Z) = h(\Sigma^{1/2} Z + \mu) \). Hence the left hand side of (2) is

\[
\text{cov}[X, h(X)] = \text{cov}[\Sigma^{1/2} Z, g(Z)] = \Sigma^{1/2} E[\nabla g(Z)].
\]

Since \( \nabla g(Z) = \Sigma^{1/2} \nabla h(X) \), the identity (2) follows immediately. \( \square \)
REMARK: A version of this lemma appears in Stein (1981) with \( \Sigma \) being the identity matrix. However, the search of an explicit formula as (2) is not successful, although it is believed that the form is well-known in the fields of sliced inverse regression and decision theory. Siegel (1993)’s method can also be applied to prove the identity (2) but is less direct.

We find that Siegel’s theorem is closely related to Stein’s identities and can be generalized to covariances of \( X_1 \) with other order statistics, other functionals, and for other distributions.

2 A simple proof of the generalized Siegel’s formula

Let \( f_u(z) = \min(z, u) \). Then \( f_u \) is piecewise linear in \( z \). Hence for \( Z \sim N(\mu, \sigma^2) \), Stein’s identity can be applied to get that

\[
\text{cov}[Z, f_u(Z)] = \sigma^2 E[f_u'(Z)] = \sigma^2 E[I_{\{Z<u\}}] = \sigma^2 \Pr[Z = \min(Z, u)],
\]

which easily leads to Lemmas 2 and 3 of Siegel (1993). Then some more properties of the normal distribution are used in Siegel (1993) to obtain the proof of his theorem. We find, however, that it is convenient to use a multivariate version of Stein’s identity directly to prove the following generalized Siegel’s formula.

**Theorem 1** Let \((X_1, \ldots, X_n)\) be multivariate normal, such that the variables are distinct, with arbitrary mean and variance structure. Let \(X_{(i)}\) be the \(i\)th largest among \(X_1, \ldots, X_n\), then

\[
\text{cov}[X_1, X_{(i)}] = \sum_{j=1}^{n} \text{cov}(X_1, X_j) \Pr(X_j = X_{(i)}).
\]

**Proof:** Define \( h(x_1, \ldots, x_n) = x_{(i)} \). Then \( h \) is piecewise linear and therefore almost everywhere differentiable. Also it is noticed that \( \partial h(x_1, \ldots, x_n)/\partial x_j \) is one if \( x_j = x_{(i)} \) and zero if \( x_j \neq x_{(i)} \). Hence by applying identity (3) we obtain the result. \( \square \)

REMARK: When \( X_1, \ldots, X_n \) are i.i.d., Carl Morris and I found the following ancillary argument to prove the above result. Since \( X_{(i)} - \bar{X} \) is independent of \( \bar{X} \), \( \text{cov}(\bar{X}, X_{(i)}) = \text{cov}(\bar{X}, \bar{X}) = \text{var}(X_1)/n \). On the other hand, since \( \text{cov}(X_j, X_{(i)}) = \text{cov}(X_k, X_{(i)}) \) for any \( j \) and \( k \) because of the i.i.d. assumption, we obtain that \( \text{cov}(X_1, X_{(i)}) = \text{var}(X_1)/n \).

For a multivariate normal random vector, the covariance of \( X_1 \) with \( \min(|X_1|, \ldots, |X_n|) \) can also be computed similarly. Define \( h(x_1, \ldots, x_n) = \min(|x_1|, \ldots, |x_n|) \), then \( \partial h(x)/\partial x_j \) is equal to
1 if $0 < x_j < u$; $-1$ if $-u < x_j < 0$; and $0$ if $|x_j| > u$, where $u = \min(|X_k|, k \neq j)$. Hence Stein’s identity (3) leads to the following corollary.

**Corollary 1** Let $Y = \min(|X_1|, \ldots, |X_n|)$, where $(X_1, \ldots, X_n)$ is multivariate normal with arbitrary mean and covariance structure. Then

$$
cov(X_1, Y) = \sum_{i=1}^{n} cov(X_1, X_i)[Pr(|X_i| = Y, X_i > 0) - Pr(|X_i| = Y, X_i < 0)].
$$

When $X_1, \ldots, X_n$ have mean zero, the above covariance equals zero.

### 3 Generalizations to other distributions

Stein’s identities for other distributions can also be used to obtain similar results.

**Corollary 2** Let $X_1 \sim \text{Poisson}(\lambda_1)$ be independent of $X_2, \ldots, X_n$, then

$$
cov[X_1, \min(X_1, \ldots, X_n)] = \lambda_1 Pr[X_1 < X_j, \text{ for all } j \neq 1].
$$

**Proof:** Define $f_u(x)$ the same way as before, then by Stein’s identity for a Poisson distribution:

$$
cov[X_1, f_u(X_1)] = \lambda_1 E[f_u(X_1 + 1) - f_u(X_1)] = \lambda_1 Pr(X_1 < u).
$$

Hence conditionally

$$
cov[X_1, f_u(X_1) \mid X_2, \ldots, X_n] = \lambda_1 Pr[X_1 < \min(X_2, \ldots, X_n) \mid X_2, \ldots, X_n].
$$

Since $cov(U, V) = E[cov(U, V \mid W)] + cov[E(U \mid W), E(V \mid W)]$, and $X_1$ is independent of $X_j, j \neq 1$, the conclusion follows. □

**Corollary 3** Let $X_1 \sim \text{t}_{\nu}(0, \sigma^2)$ be independent of $X_2, \ldots, X_n$, then

$$
cov[X_1, \min(X_1, \ldots, X_n)] = \sqrt{\frac{\nu}{\nu - 2} \sigma^2 Pr[X_1^* < \min(X_2, \ldots, X_n)]},
$$

where $X_1^* \sim \text{t}_{\nu-2}(0, \sigma^2)$. Let $X_{(i)}$ be the $i$th order statistics, then more generally,

$$
cov[X_1, X_{(i)}] = \sqrt{\frac{\nu}{\nu - 2} \sigma^2 Pr[X_{(i-1)} < X_1^* < X_{(i+1)}]].
$$

**Proof:** Use Stein’s identity for $t$-distribution. □
More generally, Morris (1983) provides Stein-type identities (theorem 5.2) for natural exponential families with quadratic variance functions (NEF-QVF), which can be used to generalize the above arguments to a broader class of distributions (NEF-QVF) without much difficulty.

To illustrate, we compute $\text{cov}[Y_1, \min(Y_1, \ldots, Y_n)]$ where $Y_i \sim \text{Exponential}(\theta_i)$ are independent, $E(Y_i) = 1/\theta_i$, and $\text{var}(Y_i) = 1/\theta_i^2$. Stein’s identity takes the form

$$\text{cov}[Y_1, f(Y_1)] = \frac{1}{\theta_1} E[Y_1 f'(Y_1)].$$

Let $f_u(x) = \min(x, u)$, then $\text{cov}[Y_1, f_u(Y_1)] = (1/\theta_1) E(Y_1 I_{Y_1 < u})$. Since $U = \min(Y_2, \ldots, Y_n) \sim \text{Exponential}(\theta_2 + \cdots + \theta_n)$,

$$\text{cov}[Y_1, \min(Y_1, \ldots, Y_n)] = \frac{1}{\theta_1} E[E(Y_1 I_{Y_1 < U} | U)] = \frac{1}{\theta_1} E[Y_1 \Pr(U > Y_1 | Y_1)] = \frac{1}{(\theta_1 + \cdots + \theta_n)^2}.$$

REMARK: Firstly, it is noted that this covariance does not depend on $\theta_1$ at all, and a more general version is $\text{cov}[Y_i, \min(Y_1, \ldots, Y_n)] = \text{var}[\min(Y_1, \ldots, Y_n)]$. Secondly, Professor Carl Morris provides another ancillary argument for the result: consider a location shift family of exponential distributions with known scale parameters $\theta_1, \ldots, \theta_n$. Then $Y_{(1)} = \min(Y_1, \ldots, Y_n)$ is complete sufficient for the location parameter. Hence by Basu's theorem, $Y_1 - Y_{(1)}$ is independent of $Y_{(1)}$, and therefore the result holds.

Acknowledgement: The author thanks Professors Hal Stern and Carl Morris for bringing up the problem and providing very stimulating discussions. Professor Morris also suggested the computations for the exponential distribution.

REFERENCES

