C. R. Math. Rep. Acad. Sci. Canada Vol. **42** (2) 2020, pp. 21–24

Generalizing Stein's Lemma

Moawia Alghalith

Presented by Pierre Milman, FRSC

Abstract. We generalize Stein's lemma. That is, we do not assume a specific probability distribution. We also provide new additional results for the covariance and the variance.

Résumé. Nous généralisons le lemme de Stein. Autrement dit, nous ne supposons pas une distribution de probabilité spécifique. Nous fournissons également de nouveaux résultats supplémentaires pour la covariance et la variance.

1. **Introduction** Stein's lemma is a very useful result that is widely used in many applications (see, for example, Alghalith (2017) and Shushi (2018). In 1973 Charles Stein presented the following identity, known as Stein's Lemma, under the assumption of the normality of X

$$Ef(X)(X-\mu_X)=\sigma_X^2 Ef'(X),$$

where μ_X is the mean of X and σ_X is its standard deviation (assuming $Ef'(X)$ exists).

There were some generalizations of the lemma. For example, Hudson (1978) obtained this identity for exponential distributions. Landsman (2006) and Landsman et al. (2015) generalized the identity to elliptical distributions. Adcock (2014) generalized it to the multivariate extended skew-t distribution. Shushi (2018) assumed truncated elliptical random vectors. Adcock et al. (2019) introduced a lemma for generalized skew-elliptical random vectors.

In this note, we generalize and extend Stein's lemma. That is, we do not assume any probability distribution. Furthermore, we provide new additional results for the covariance and the variance. These results are as important and useful as Stein's lemma.

2. The Method

Theorem. *For a differentiable function $f(x)$ and two random variables X*

Received by the editors on February 19, 2020; revised April 28, 2020.
AMS Subject Classification: 62E15.
Keywords: Stein's lemma, probability distribution, exact Taylor expansion, variance, covariance.

and Y *(where* $\mu \neq 0$[1] *is the arithmetic mean as a simple average of* $f'(X)$ *defined below; we also assume all the moments exist)*

(i) $\mathrm{Var}(f(X)) = \mu^2 \mathrm{Var}(X)$
(ii) $\mathrm{Cov}(f(X), Y) = \mu \mathrm{Cov}(X, Y)$
(iii) $\mathrm{Cov}(f(X), X) = \mu \mathrm{Var}(X)$
(iv) $Ef(X)(X - EX) = \mu \mathrm{Var}(X)$.

PROOF. First, we obtain a Taylor's expansion of the derivative of the function to be evaluated

$$f'(x_i) = f'(c) + R(x_i, c), \tag{1}$$

where R is the remainder (error), and x_i is an outcome. Thus

$$R(x_i, c) = f'(x_i) - f'(c). \tag{2}$$

There is a constant c^* that minimizes the sum of the errors, as follows

$$\underset{c}{\mathrm{Min}} \sum\nolimits_i R^2(x_i, c).$$

We can show that the second order condition holds. The first-order condition is

$$-2f''(c^*) \sum\nolimits_i R(x_i, c^*) = 0 \Longrightarrow \sum\nolimits_i R(x_i, c^*) = 0. \tag{3}$$

The first-order condition implies that

$$f'(c^*) = \frac{\sum_i f'(x_i)}{n} \equiv \mu, \tag{4}$$

where n is the number of outcomes; so that, for a large n, μ is the arithmetic mean (simple average) of $f'(x_i)$ (μ is a constant). For a continuous random variable, the simple average is the midpoint of the interval of the outcomes $[x_k, x_m]$, thus the simple average is $\frac{x_m - x_k}{2}$. We note that for asymmetrical distributions the arithmetic mean and the expected value are not generally equal. For symmetric distributions, both the simple average and expected value are the midpoint of the interval of outcomes (graphically the midpoint of the horizontal axis). For example, under a normal distribution, both the expected value and the simple average is the midpoint of the interval $(-\infty, +\infty)$ which is equal to zero. However, clearly, for asymmetrical distributions such as the log-normal distribution, the expected value is not the midpoint (simple average) of the interval $(0, +\infty)$.

Now consider this Taylor's expansion at the point of approximation c^*

$$f'(x_i) = f'(c^*) + R(x_i, c^*) \equiv \mu + R(x_i, c^*). \tag{5}$$

[1] There is no loss of generality from assuming $\mu \neq 0$, since the results can be achieved by a simple transformation of the function, such as $g'(X) = f'(X) + c$, $c \neq 0$. Then, the same results will be obtained for $f(X)$.

Or equivalently,

$$f'(x) = \mu + R(x, c^*). \tag{6}$$

Integrating yields

$$\int_{\alpha}^{x} f'(u)\,du = \int_{\alpha}^{x} \mu du + \int_{\alpha}^{x} R(x, c^*)\,du \tag{7}$$

since $\sum_i R(x_i, c^*) = 0$ (which implies that $\int_{\alpha}^{x} R(x, c^*)\,du = 0$); also $\int_{\alpha}^{x} f'(u)\,du = f(x) - f(\alpha)$ and $\int_{\alpha}^{x} \mu du = \mu(x - \alpha)$. Therefore

$$f(x) = f(\alpha) + \mu(x - \alpha). \tag{8}$$

Thus,

$$f(X) = f(\alpha) + \mu(X - \alpha), \tag{9}$$

since (8) holds for each outcome. Clearly, from (9)

$$\mathrm{Var}(f(X)) = \mu^2 \mathrm{Var}(X),$$

$$\mathrm{Cov}(f(X), X) = \mu \mathrm{Var}(X),$$
$$\mathrm{Cov}(f(X), Y) = \mu \mathrm{Cov}(X, Y),$$
$$Ef(X)(X - EX) = \mu \mathrm{Var}(X).$$

□

ALTERNATIVE PROOF. Consider this Taylor's expansion around an outcome x_i

$$f(dX + x_i) - f(x_i) = f'(x_i)\,dX, \tag{10}$$

Taking the summation yields

$$\sum\nolimits_i [(f(dX + x_i) - f(x_i)] = \sum\nolimits_i f'(x_i)\,dX. \tag{11}$$

We note that $\sum\limits_i f'(x_i)\,dX = \frac{X-c}{n}\sum\limits_i^n f'(x_i) = \left(\frac{\sum\limits_i^n f'(x_i)}{n}\right)(X - c) = \mu(X - c)$ as $n \to \infty$, where c is a constant, the arithmetic mean of f' is $\mu = \frac{\sum\limits_i^n f'(x_i)}{n}$ as

$n \to \infty$. Similarly, $\sum_i \left[f\left(dX + x_i\right) - f\left(x_i\right) \right] = \frac{f(X)-\alpha}{n} \sum_i^n 1 = f\left(X\right) - \alpha$, where α is a constant. Thus (11) can be rewritten as

$$f\left(X\right) = \alpha + \mu\left(X - c\right). \tag{12}$$

The results are directly derived from (12). □

In conclusion, the results are intuitive and they will be extremely useful in many disciplines, since they express an arbitrary function in a linear form. For example, the form of the value function (or the utility function) is usually unknown; however, by using our results, the expected utility function can be expressed as $EU\left(W\right) = \alpha + \mu EW$, where W is the random wealth, α is a constant, and μ is defined as before, and $\mathrm{Var}\left(U\right) = \mu^2 \mathrm{Var}\left(W\right)$. Our results also can be conveniently used in non-linear regressions. In many cases, the functional form is unknown or cumbersome. The regression equation can be conveniently linearized using our method. Furthermore, our method can be used in optimization, since it will enable us to obtain explicit optimal solutions by substituting our results in the first-order condition.

References

1. Alghalith, M. (2017). A note on the stochastic portfolio optimization. Economics Bulletin, 37, 1265–1266.
2. Adcock, C.J. (2014). Mean–variance–skewness efficient surfaces, Stein's lemma and the multivariate extended skew-Student distribution, European Journal of Operational Research, 234, 392–401.
3. Adcock, C., Landsman, Z., and Shushi, T. (2019). Stein's Lemma for generalized skew-elliptical random vectors. Communications in Statistics-Theory and Methods, 1–16.
4. Hudson, H.M. (1978). A natural identity for exponential families with application in a multiparameter estimation. Ann. Statist., 6, 473-484.
5. Landsman, Z, Vanduffel, S. and Yao, J. (2015). Some Stein-type inequalities for multivariate elliptical distributions and applications, Statist. Probab. Lett., 97, 54–62.
6. Landsman, Z. (2006). On the generalization of Stein's Lemma for the elliptical class of distributions, Statist. Probab. Lett., 76, 1012–1016.
7. Shushi, T. (2018). Stein's lemma for truncated elliptical random vectors. Statistics & Probability Letters, 137, 297–303.
8. Stein, C.M. (1973). Estimation of the mean of a multivariate normal distribution. Proceedings of Prague Symposium on Asymptotic Statistics, 345–381.

Department of Economics, University of the West Indies, St. Augustine, Trinidad
e-mail: malghalith@gmail.com