

## STEIN'S IDENTITY FOR DISCRETE DISTRIBUTIONS

K. Teerapabolarn

Department of Mathematics

Faculty of Science

Burapha University

Chonburi, 20131, THAILAND

**Abstract:** In this article, we propose a generalization of Stein's identity for discrete distributions, which is an extension of the identity in Sudheesh Kumar [13]. For applications, we give two results of the Poisson and binomial approximations via Stein's method and the two corresponding Stein's identities.

**AMS Subject Classification:** 62E10, 62E17

**Key Words:** binomial approximation, Poisson approximation, Stein's identity, Stein's method

### 1. Introduction

Let  $X$  be the normal random variable with mean  $\mu$  and variance  $\sigma^2$  and let  $c$  be a differentiable function with derivative  $c'$  such that  $\mathbb{E}|c'(X)| < \infty$ . It follows from [12] that

$$\mathbb{E}[(X - \mu)c(X)] = \sigma^2\mathbb{E}[c'(X)].$$

This has come to be known as Stein's identity. Similar types of identities for other distributions and their applications can be found in [1], [3] and [7, 8, 9, 10]. In the recent article, Sudheesh Kumar [13] proposed a generalization of the identity related to a wider class of continuous probability distributions.

For this article, in view of the interest in generalized Stein's identity of [13], we propose a generalization of the identity for discrete probability distributions, which is an extension of the identity in Theorem 2.1 of [13], and the details are in Section 2. In Section 3, for applications, we give two results of the Poisson and binomial approximations via Stein's method together with the two corresponding identities.

## 2. Result (Generalized Stein's Identity for Discrete Distributions)

Consider continuous distributions, we let  $X$  be a continuous random variable with support an interval  $(a, b)$ , mean  $\mathbb{E}(X) = \mu$ , variance  $\text{Var}(X) = \sigma^2$  and density  $f$ . Let  $h$  be a real valued function such that  $\mathbb{E}[h(X)] = \mu$  and  $\mathbb{E}[h^2(X)] < \infty$ . Suppose  $f$  is differentiable function with derivative  $f'$  and there exists a non-vanishing function  $z$  such that

$$\frac{f'(x)}{f(x)} = \frac{-z'(x)}{z(x)} + \frac{\mu - h(x)}{z(x)}, \quad x \in (a, b). \quad (2.1)$$

Sudheesh Kumar [13] established the generalized Stein's identity

$$\mathbb{E}\{[h(X) - \mu]c(X)\} = \mathbb{E}[z(X)c'(X)], \quad (2.2)$$

for  $c$  is an absolutely continuous function satisfying  $E|z(X)c'(X)| < \infty$  and

$$z(x) = \frac{1}{f(x)} \int_x^b [h(t) - \mu]f(t)dt \quad (2.3)$$

$$= \frac{1}{f(x)} \int_a^x [\mu - h(t)]f(t)dt, \quad (2.4)$$

where (2.3) is obtained by the relation (2.1) and the condition  $\lim_{x \rightarrow b} z(x)f(x) = 0$ .

In the case of discrete distributions, we can use the same arguments as in the continuous case to establish a generalization of Stein's identity for this case as follows.

Let  $X$  be a non-negative integer-valued random variable with support  $\mathcal{S}(X)$ , mean  $\mu$ , variance  $\sigma^2$  and probability function  $p(x)$ . Let  $h$  be a real valued function such that  $\mathbb{E}[h(X)] = \mu$  and  $\mathbb{E}[h^2(X)] < \infty$ , then, replacing integrals by sums and  $c'(x)$  by  $\Delta g(x) = g(x+1) - g(x)$  in the proof of Theorem 2.1 of [13], we arrive at the discrete analogue of the generalized stein's identity in [13].

**Theorem 2.1.** Suppose  $p(x) > 0$  for every  $x \in \mathcal{S}(X)$  and there exists a function  $z$  such that

$$z(x) = \frac{1}{p(x)} \sum_{k=0}^x [\mu - h(k)]p(k), \quad x \in \mathcal{S}(X). \tag{2.5}$$

Then, for any function  $g : \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}$  satisfying  $E|z(X)\Delta g(X)| < \infty$ , we have

$$\mathbb{E}\{[h(X) - \mu]g(X)\} = \mathbb{E}[z(X)\Delta g(X)]. \tag{2.6}$$

**Remark 2.1.** It is observed that  $z(x)$  in (2.5) can be expressed in the form

$$z(x) = \frac{1}{p(x)} \sum_{k \in \mathcal{S}(X) \setminus \{0, \dots, x\}} [h(k) - \mu]p(k), \quad x \in \mathcal{S}(X), \tag{2.7}$$

and by using (2.5), we have the following recurrence relation

$$z(0) = \mu - h(0), \quad z(x) = z(x - 1) \frac{p(x - 1)}{p(x)} + \mu - h(x), \quad x \in \mathcal{S}(X) \setminus \{0\}. \tag{2.8}$$

**Remark 2.2.** For a given  $h(x)$  the function  $z(x)$  uniquely determines the distribution of  $X$ , denoted by  $\mathcal{L}(X)$ . Thus, for  $h(x) = x$ , the identity (2.6) becomes

$$\mathbb{E}\{(X - \mu)g(X)\} = \mathbb{E}[z(X)\Delta g(X)], \tag{2.9}$$

and the random variable  $X$  has the distribution  $\mathcal{L}(X)$  if and only if

$$z(0) = \mu, \quad z(x) = z(x - 1) \frac{p(x - 1)}{p(x)} + \mu - x, \quad x \in \mathcal{S}(X) \setminus \{0\}. \tag{2.10}$$

**Remark 2.3.** Consider the identity (2.9), we have  $\mathbb{E}[z(X)] = \sigma^2$ .

The following examples are provided to illustrate the identity (2.6).

**Example 2.1.** Suppose that the distribution of  $X$  belongs to a discrete exponential family with the probability function

$$p(x) = k(x) \exp\{x\theta - \varphi(\theta)\}, \quad x = 0, 1, \dots, \theta > 0.$$

Let

$$t(x) = \begin{cases} 0 & \text{if } x = 0, \\ \frac{k(x - 1)}{k(x)} & \text{if } x = 1, 2, \dots, \end{cases}$$

then we have the identity

$$\mathbb{E} \left\{ \left[ t(X) - e^\theta \right] g(X) \right\} = e^\theta \mathbb{E}[\Delta g(X)], \quad \theta > 0. \quad (2.11)$$

For  $x \in \mathbb{N}$ , it follows that

$$\frac{p(x) - p(x-1)}{p(x)} = 1 - \frac{p(x-1)}{p(x)} = 1 - \frac{k(x-1)}{k(x)} e^{-\theta},$$

which gives  $p(x) - p(x-1) = \{1 - t(x)e^{-\theta}\} p(x)$ .

Summation with respect to  $x$  from  $t+1$  to  $\infty$  and using the fact that  $\lim_{x \rightarrow \infty} p(x) = 0$

$$e^\theta p(t) = \sum_{x=t+1}^{\infty} \{t(x) - e^\theta\} p(x),$$

and by using (2.7), yields

$$z(t)p(t) = \sum_{x=t+1}^{\infty} [h(x) - \mu] p(x).$$

By comparing these equations, we obtain

$$h(x) = t(x) - e^\theta + \mu \text{ and } z(x) = e^\theta,$$

thus the identity (2.6) reduces to the identity (2.11).

**Remark 2.4.** The Poisson distribution with mean  $\lambda$  belongs to the discrete exponential family with  $e^\theta = \lambda$ . Hence the identity (2.11) reduces to the Stein-Chen identity

$$\mathbb{E}[Xg(X)] = \lambda \mathbb{E}[g(X+1)], \quad (2.12)$$

which was pursued by Chen [5].

**Example 2.2.** Let  $X$  be distributed as the binomial distribution with parameters  $n \in \mathbb{N}$  and  $p \in (0, 1)$ . Then, replacing  $\mu$  by  $np$ , the identity (2.9) takes the form

$$\mathbb{E}[(1-p)Xg(X)] = \mathbb{E}[(n-X)pg(X+1)], \quad (2.13)$$

this can be found in Barbour et al. [2].

**Example 2.3.** For the negative binomial distribution with probability function

$$p(x) = \frac{\Gamma(n+x)}{\Gamma(n)x!} p^n (1-p)^x, \quad x = 0, 1, \dots,$$

and it follows from (2.9) that

$$\mathbb{E}[Xg(X)] = \mathbb{E}[(1-p)(n+X)g(X+1)] \quad (2.14)$$

is Stein's identity for the negative binomial distribution with parameters  $n$  and  $p$ , which is the same identity in Brown and Phillips [4].

### 3. Applications

In this section, we use Stein's method and Stein's identities to give two results in approximating the distribution of non-negative integer-valued random variable by the Poisson and binomial distributions. Let  $\mathcal{P}(\lambda)$  and  $\mathcal{B}(n, p)$  denote the Poisson and binomial distributions. The two total variation distances of  $\mathcal{L}(X)$  and  $\mathcal{P}(\lambda)$  and  $\mathcal{L}(X)$  and  $\mathcal{B}(n, p)$  are defined by

$$d_{TV}(\mathcal{L}(X), \mathcal{P}(\lambda)) = \sup_{A_1} |\mathcal{L}(X)\{A_1\} - \mathcal{P}(\lambda)\{A_1\}| \quad (3.1)$$

and

$$d_{TV}(\mathcal{L}(X), \mathcal{B}(n, p)) = \sup_{A_2} |\mathcal{L}(X)\{A_2\} - \mathcal{B}(n, p)\{A_2\}|, \quad (3.2)$$

where  $A_1$  is a subset of  $\mathbb{N} \cup \{0\}$  and  $A_2$  a subset of  $\{0, \dots, n\}$ .

#### 3.1. Application to the Poisson Approximation

The Stein's method was first introduced by Stein [11] and the version appropriate for the Poisson case was first developed by Chen [5]. Stein's identity for the Poisson distribution with mean  $\lambda > 0$ , every subset  $A_1$  of  $\mathbb{N} \cup \{0\}$  and for any bounded real valued function  $g = g_{A_1} : \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}$  is of the form

$$\mathcal{L}(X)\{A_1\} - \mathcal{P}(\lambda)\{A_1\} = \mathbb{E}[\lambda g(X+1) - Xg(X)]. \quad (3.3)$$

For any subset  $A_1$  of  $\mathbb{N} \cup \{0\}$ , Barbour et al. [2] showed that

$$\sup_{k, A_1} |\Delta g(k)| \leq \lambda^{-1}(1 - e^{-\lambda}). \quad (3.4)$$

**Theorem 3.1.** *If  $\lambda = \mu$ , then we have the following:*

$$d_{TV}(\mathcal{L}(X), \mathcal{P}(\lambda)) \leq (1 - e^{-\lambda}) \mathbb{E} \left| \frac{z(X)}{\lambda} - 1 \right|. \tag{3.5}$$

*Proof.* In view of (3.1) and (3.3), for every subset  $A_1$  of  $\mathbb{N} \cup \{0\}$ , we have

$$\begin{aligned} d_{TV}(\mathcal{L}(X), \mathcal{P}(\mu)) &= |\lambda \mathbb{E}[g(X + 1)] - \mathbb{E}[Xg(X)]| \\ &= |\lambda \mathbb{E}[\Delta g(X)] - \mathbb{E}[(X - \mu)g(X)]|. \end{aligned}$$

Since  $\mathbb{E}|z(X)\Delta g(X)| \leq \lambda^{-1}(1 - e^{-\lambda})\sigma^2 < \infty$ , by Theorem 2.1, we get

$$\begin{aligned} d_{TV}(\mathcal{L}(X), \mathcal{P}(\mu)) &= |\lambda \mathbb{E}[\Delta g(X)] - \mathbb{E}[z(X)\Delta g(X)]| \\ &\leq \mathbb{E} [|\lambda - z(X)| |\Delta g(X)|]. \end{aligned}$$

Using (3.4), finishes the proof of theorem. □

**Corollary 3.1.** *If  $z(x) \geq / < \lambda$  for every  $x \in \mathcal{S}(X)$ , then*

$$d_{TV}(\mathcal{L}(X), \mathcal{P}(\lambda)) \leq (1 - e^{-\lambda}) \left| \frac{\sigma^2}{\lambda} - 1 \right|. \tag{3.6}$$

### 3.2. Application to the Binomial Approximation

Similar to that of the Poisson approximation, Stein’s identity for the binomial with parameters  $n \geq 1$  and  $p = q \in (0, 1)$ , every subset  $A_2$  of  $\{0, \dots, n\}$  and for any bounded real valued function  $g = g_{A_2} : \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}$  is of the form

$$\mathcal{L}(X)\{A_2\} - \mathcal{B}(n, p)\{A_2\} = \mathbb{E}[(n - X)p\lambda g(X + 1) - qXg(X)]. \tag{3.7}$$

For any subset  $A_2$  of  $\{0, \dots, n\}$ , Ehm [6] proved that

$$\sup_{k, A_2} |\Delta g(k)| \leq \frac{1 - p^{n+1} - q^{n+1}}{(n + 1)pq}. \tag{3.8}$$

**Theorem 3.2.** *If  $np = \mu$ , then we have the following:*

$$d_{TV}(\mathcal{L}(X), \mathcal{B}(n, p)) \leq \frac{1 - p^{n+1} - q^{n+1}}{(n + 1)pq} \mathbb{E}|\mu - pX - z(X)|. \tag{3.9}$$

*Proof.* In view of (3.2) and (3.7) and using the fact that  $\mathbb{E}|z(X)\Delta g(X)| < \infty$ , we have

$$d_{TV}(\mathcal{L}(X), \mathcal{B}(n, p)) = |\mathbb{E}[npg(X + 1)] - p\mathbb{E}[X\Delta g(X)] - \mathbb{E}[Xg(X)]|$$

$$\begin{aligned}
 &= |\mathbb{E}[\mu\Delta g(X)] - p\mathbb{E}[X\Delta g(X)] - \mathbb{E}[(X - \mu)g(X)]| \\
 &= |\mathbb{E}[\mu\Delta g(X)] - p\mathbb{E}[X\Delta g(X)] - E[z(X)\Delta g(X)]| \\
 &= |\mathbb{E}\{[\mu - pX - z(X)]\Delta g(X)\}| \\
 &\leq \mathbb{E}[|\mu - pX - z(X)||\Delta g(X)|].
 \end{aligned}$$

Using (3.8), (3.9) is obtained. □

**Corollary 3.2.** *If  $\mu - px - z(x) \geq / < 0$  for every  $x \in \mathcal{S}(X)$ , then*

$$d_{TV}(\mathcal{L}(X), \mathcal{B}(n, p)) \leq \frac{1 - p^{n+1} - q^{n+1}}{(n + 1)pq} |q\mu - \sigma^2|. \tag{3.10}$$

**Example 3.1.** The hypergeometric random variable  $X$  has the probability function

$$p(x) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}, \quad x = 0, \dots, \min\{m, n\}.$$

Then its mean and variance are  $\mu = \frac{mn}{N}$  and  $\sigma^2 = \frac{mn(N-m)(N-n)}{N^2(N-1)}$ , respectively.

Using (2.10), we have  $z(x) = \frac{(m-x)(n-x)}{N}$ . Note that  $\lambda = \frac{mn}{N} \geq z(x)$  for all  $0 \leq x \leq \min\{m, n\}$  and, for  $p = \frac{m}{N}$ , we also have  $\mu - px - z(x) = \frac{(n-x)x}{N} \geq 0$  for all  $0 \leq x \leq \min\{m, n\}$ . Therefore, applying corollaries 3.1 and 3.2, the results in approximating the hypergeometric distribution with parameters  $N$ ,  $m$  and  $n$ , denoted by  $\mathcal{H}(N, m, n)$ , by the Poisson and binomial distributions are as follows:

$$d_{TV}(\mathcal{H}(N, m, n), \mathcal{P}(mn/N)) \leq (1 - e^{-mn/N}) \frac{N(n - 1) + m(N - n)}{N(N - 1)}$$

and

$$d_{TV}(\mathcal{H}(N, m, n), \mathcal{B}(n, p)) \leq (1 - p^{n+1} - q^{n+1}) \frac{n(n - 1)}{(n + 1)(N - 1)}.$$

**Example 3.2.** The Pólya random variable  $X$  has the probability function

$$p(x) = \frac{\binom{r/c+x-1}{x} \binom{(N-r)/c+n-x-1}{n-x}}{\binom{N/c+n-1}{n}}, \quad x = 0, \dots, n,$$

where parameters  $N, n, r$  and  $c$  are non-negative integer-valued. Then the mean and variance of  $X$  are  $\mu = \frac{nr}{N}$  and  $\sigma^2 = \frac{nr(N+cn)(N-r)}{N^2(N+c)}$ , respectively. Using (2.10), we have  $z(x) = \frac{(r+cx)(n-x)}{N}$ . Observe that, if  $r \geq c(n - 1)$ , then  $\lambda =$

$\frac{nr}{N} \geq z(x)$  for all  $0 \leq x \leq n$  and, for  $p = \frac{r}{N}$ , it follows that  $\mu - px - z(x) = \frac{-cx(n-x)}{N} \leq 0$  for all  $0 \leq x \leq n$ . Therefore, applying corollaries 3.1 and 3.2, the results in approximating the Pólya distribution with parameters  $N, n, r$  and  $c$ , denoted by  $\mathcal{PY}(N, n, r, c)$ , by the Poisson and binomial distributions are of the forms

$$d_{TV}(\mathcal{PY}(N, n, r, c), \mathcal{P}(nr/N)) \leq (1 - e^{-nr/N}) \frac{r(N + cn) - c(n-1)N}{N(N + c)},$$

for  $r \geq c(n-1)$ , and

$$d_{TV}(\mathcal{PY}(N, n, r, c), \mathcal{B}(n, p)) \leq (1 - p^{n+1} - q^{n+1}) \frac{cn(n-1)}{(n+1)(N+c)}.$$

## References

- [1] B.C. Arnold, E. Castillo, J.M. Sarabia, A multivariate version of Stein's identity with applications to moment calculations and estimation of conditionally specified distributions, *Commun. Stat. - Theory Methods*, **30** (2001), 2517-2542.
- [2] A.D. Barbour, L. Holst, S. Janson, *Poisson Approximation*, Oxford Studies in Probability 2, Clarendon Press, Oxford (1992).
- [3] L. Brown, A. DasGupta, L.R. Haff, W.E. Strawderman, The heat equations and Stein's identity: Connections and applications, *J. Stat. Plann. Inference*, **136** (2006), 2254-2278.
- [4] T.C. Brown, M.J. Phillips, Negative binomial approximation with Stein's method, *Meth. Comp. Appl. Probab.*, **1** (1999), 407-421.
- [5] L.H.Y. Chen, Poisson approximation for dependent trials, *Ann. Probab.*, **3** (1975) 534-545.
- [6] W. Ehm, Binomial approximation to the Poisson binomial distribution, *Statist. Probab. Lett.*, **11** (1991), 7-16.
- [7] H.M. Hudson, A natural identity for exponential families with application in a multi-parameter estimation, *Ann. Statist.*, **6** (1978), 473-484.
- [8] Z. Landsman, On the generalization of Stein's lemma for the elliptical class of distributions, *Statist. Probab. Lett.*, **76** (2006), 1012-1016.

- [9] Z. Landsman, J. Neslehova, Stein's lemma for elliptical random vectors, *J. Multivariate. Anal.*, **99** (2008), 912-927.
- [10] B.L.S. Prakasa Rao, Characterizations of distribution through some identities, *J. Appl. Probab.*, **16** (1979), 903-909.
- [11] C.M. Stein, A bound for the error in normal approximation to the distribution of a sum of dependent random variables, in *Proceedings of the sixth Berkeley Symposium on Mathematical Statistics and Probability*, **3** (1972), 583-602.
- [12] C.M. Stein, Estimation of the mean of a multivariate normal distribution, *In Proceedings of Prague Symposium on Asymptotic Statistics*, 345-381.
- [13] K. Sudheesh Kumar, On Stein's identity and its applications. *Statist. Probab. Lett.*, **79** (2009), 1444-1449.

