Journal of Statistical Modelling: Theory and Applications
Vol. 6, No. 2, 2025, pp. 117-126
Yazd University Press 2025

JSMTA

Research Paper

A decomposition approach to the asymptotic distribution of the sample variance in discrete models

Narges Abbasi*, Masoud Yarmohammadi, Ali Shadrokh, Mahnaz Lashani
Department of Statistics, Payme Noor University, Tehran, Iran

Received: December 23, 2025/ Revised: July 05, 2026/ Accepted: July 16, 2026

Abstract: Understanding the asymptotic behavior of the sample variance is important in statistical theory and inference. While classical results provide chi-square limiting distributions for continuous populations, discrete random variables often exhibit non-classical behavior, complicating both theoretical analysis and practical applications. In this paper, we propose a decomposition of the sample variance for discrete random variables into two theoretically tractable components, enabling a finer characterization of their stochastic structure. We derive the asymptotic distributions of these components under a specific condition, revealing that certain components converge to chi-square laws - a phenomenon previously observed only in Bernoulli and binomial models with success probability one-half. Extensive Monte Carlo simulations confirm the accuracy of the theoretical approximations and demonstrate their relevance even for moderate sample sizes. These results provide new insights into the interplay between discreteness, variance decomposition, and asymptotic behavior, extending classical chi-square asymptotics to a broader class of discrete models.

Keywords: Binomial distribution; Chi-square distribution; Discrete random variable; Limiting distribution; Moment.
Mathematics Subject Classification (2010): 60F05, 60E05.

1 Introduction

Among fundamental summary statistics, the sample variance plays a central role as a measure of dispersion. Its exact distribution is well understood under normality, but for non-Gaussian populations the distributional behavior becomes considerably

*Corresponding author: `n_abbasi@pnu.ac.ir`

more complex. In such settings, higher order moments of the underlying distribution influence the asymptotic form of the statistic, and classical chi-square approximations may fail to hold (Kendall and Stuart, 1977; Lehmann and Casella, 1998). Consequently, a substantial literature has examined the asymptotic properties of the sample variance under broader distributional frameworks.

Discrete random variables appear in numerous statistical applications, including count data analysis, reliability studies, epidemiology, and categorical modeling. Their intrinsic lattice structure and lack of smoothness pose analytical challenges that distinguish them from continuous models. These features may invalidate standard asymptotic arguments or give rise to nonclassical limiting distributions for statistics commonly used in practice (Johnson et al., 2005).

Several studies have shown that discreteness can fundamentally alter limiting behavior. Ma et al. (2011) demonstrated that sample quantiles of discrete distributions may exhibit non Gaussian or even discrete limit distributions, contrasting sharply with classical asymptotic theory. Such results emphasize that traditional arguments cannot be directly transferred to discrete settings; structural characteristics of the distribution must be taken into account.

Nevertheless, classical limits may reappear in discrete models when particular symmetry conditions are met. Serfling (1980) established asymptotic normality for the sample variance of discrete random variables under mild moment assumptions. Moreover, as a consequence of the second order delta method, it has been shown that for Bernoulli random variables with success probability $p = 1/2$, an appropriately normalized sample variance converges to a chi square distribution; see Van der Vaart (2000). A comparable outcome holds for the binomial distribution under the same symmetry. These findings indicate that, despite the inherent discreteness of the underlying models, classical chi-square limits may arise when sufficient structural balance is present.

Motivated by these observations, the present paper studies a decomposition of the sample variance for discrete random variables into two analytically tractable components. This representation facilitates a more detailed investigation of the stochastic structure of the variance and serves as the foundation for deriving the limiting distributions of the resulting components under a specific structural condition. Importantly, the decomposition functions solely as a theoretical device and introduces no additional modeling assumptions.

To evaluate the practical implications of the theoretical results, several approaches exist for simulating dependent random variables, including joint-based techniques recently applied by Le and Vargas (2025) for generating temporal and spatial data in biological and environmental studies. In the present work, Monte Carlo simulations are performed to assess the finite sample performance of the proposed asymptotic approximation. The simulation results confirm the accuracy of the limiting distribution and demonstrate that the approximation performs well even for moderate sample sizes.

The remainder of the paper is organized as follows. Section 2 introduces the variance decomposition and the underlying assumptions. Section 3 presents the main asymptotic results for limiting distribution of sample variance. Section 4 reports simulation studies, and Section 5 concludes with a discussion and directions for future research.

2 Structure for discrete random variables

Let discrete random variable Y with probability model $P(Y = a_1) = s_1$ and $P(Y = a_2) = s_2$; $s_1 + s_2 = 1$, a_1 and $a_2 \neq 0$. We rewrite this variable as follows

$$Y = a_1 X_1 + a_2(1 - X_1),$$

where $X_1 \sim B(1, p_1)$ and $s_1 = p_1$, $s_2 = 1 - p_1 = q_1$.

In the case in which Y takes values a_1, a_2, and a_3 with probabilities s_1, s_2, and s_3, respectively, it can be represented in terms of two independent Bernoulli random variables:

$$Y = a_1 X_1 + a_2(1 - X_1)X_2 + a_3(1 - X_1)(1 - X_2),$$

where $X_1 \sim B(1, p_1)$ and $X_2 \sim B(1, p_2)$ are independent and $s_1 = p_1$, $s_2 = q_1 p_2$, and $s_3 = q_1 q_2$. It is observed that the expressions formed as a sum of random variables are not independent variables; for example, $(1 - X_1)X_2$ are not independent of $(1 - X_1)(1 - X_2)$ and for a binomial random variable $B(n = 2, p = 1/2)$, equals to the discrete random variable Y when $a_1 = 1$, $a_2 = 0$, $a_3 = 2$ and $s_1 = 1/2$, $s_2 = s_3 = 1/4$; to have a different representation: $Y = X_1 + 2(1 - X_1)(1 - X_2)$. Although it becomes troublesome in the calculations of the characteristics of this variable, the behavior of the limit distribution of central moments, give a general result. Let Y be a discrete random variable with the probability model

$$P(Y = a_i) = s_i, \quad i = 1, 2, \ldots, k,$$

with $\sum_{i=1}^{k} s_i = 1$, $s_1 \geq s_2 \geq \cdots \geq s_k > 0$. The i-th central moment is defined by

$$\mu_i = E[(Y - \mu)^i] = \sum_{j=1}^{k} s_j (a_j - \mu)^i, \quad i = 1, 2, \ldots.$$

Let $X_i \sim B(1, p_i)$, $i = 1, \ldots, k - 1$, be independent Bernoulli random variables with $q_i = 1 - p_i$. The random variable Y could be represented as follows

$$Y = a_1 X_1 + \sum_{i=2}^{k-1} \{a_i X_i \prod_{j=1}^{i-1} (1 - X_j)\} + a_k \prod_{j=1}^{k-1} (1 - X_j). \tag{1}$$

The corresponding probabilities are

$$\begin{aligned}
P(Y = a_1) &= P(X_1 = 1) = s_1 = p_1, \\
P(Y = a_i) &= P(X_1 = \cdots = X_{i-1} = 0,\ X_i = 1) = s_i = p_i \prod_{j=1}^{i-1} q_j; \quad i = 2, \ldots, k - 1, \\
P(Y = a_k) &= P(X_1 = \cdots = X_{k-1} = 0) = s_k = \prod_{j=1}^{k-1} q_j.
\end{aligned}$$

From solving the above equations, we have $p_1 = s_1$, and $p_i = s_i / (1 - \sum_{j=1}^{i-1} s_j)$, $i = 1, 2, \ldots, k - 1$. The mean of Y equals to

$$\mu_Y = E(Y) = a_1 p_1 + \sum_{i=2}^{k-1} \left[a_i p_i \prod_{j=1}^{i-1} q_j \right] + a_k \prod_{j=1}^{k-1} q_j,$$

and the central moments are expressed as

$$\begin{aligned}\mu_i &= E(Y-\mu_Y)^i \\ &= (a_1-\mu_Y)^i p_1 + \sum_{j=2}^{k-1}\left[(a_j-\mu_Y)^i p_j \prod_{\ell=1}^{j-1} q_\ell\right] + (a_k-\mu)^i \prod_{j=1}^{k-1} q_j; \quad i=1,2,\ldots.\end{aligned}$$

As an alternative representation of an arbitrary discrete random variable, define $B_1 = X_1$, $B_i = X_i \prod_{j=1}^{i-1}(1-X_j)$, $i = 2,\ldots,k-1$, and $B_k = \prod_{j=1}^{k-1}(1-X_j)$. Then Y could be expressed as $Y = \sum_{i=1}^{k} a_i B_i$, where the B_i's are Bernoulli random variables, although they are not independent. We do not adopt this representation because the corresponding moment calculations become complicated. Instead, we can use an alternative structure, presented as follows.

By (1), since our analysis is centered on the component corresponding to the symmetry point where the success probability equals one half, we express the structure of the discrete random variable in the form

$$W = (1-X_1)\{a_2X_2 + a_3X_3(1-X_2) + \cdots + a_{k-1}X_{k-1}\prod_{j=2}^{k-2}(1-X_j) + a_k\prod_{j=2}^{k-1}(1-X_j)\}.$$

where W is structured as the product of two independent random variables.

To simplify the notation for the remainder of this derivation, we denote the sequence of Bernoulli variables and coefficients as X_1, a_1 and p_1, respectively, using X, a and p where the indices are clear from the context. We define $Y = aX + W$, which is formulated as the sum of two components, both of which involve the random variable X. Notably, W vanishes when $X = 1$, which occurs with probability p. Furthermore, $E[XW] = 0$ because XW is identically to zero; $P(XW = 0) = 1$. The mean and variance of Y are given by $\mu_Y = ap + \mu_W$ and

$$\sigma_Y^2 = a^2p(1-p) + \sigma_W^2 + 2a\sigma_{X,W}, \tag{2}$$

where $E(W) = \mu_W$, $\sigma_Y^2 = Var(Y)$, $\sigma_W^2 = Var(W)$ and $\sigma_{X,W} = Cov(X,W)$ is given by

$$\sigma_{X,W} = E(XW) - E(X)E(W) = -p\mu_W.$$

The formula for the variance of W is $\sigma_W^2 = \sum_{i=2}^{k} a_i^2 s_i - \mu_W^2$ and with a some modification, it can be written

$$\sigma_W^2 = \sigma_Y^2 - a^2p(1+p) + 2ap\mu_Y.$$

For $p = \frac{1}{2}$ and $a = 1$, we have $\sigma_W^2 = \sigma_Y^2 + \mu_Y - \frac{3}{4}$ and $\sigma_{X,W} = -\frac{1}{2}\mu_W$. We employ the following notation for the product moment of the two variables X and W, which will be used in the following sections for notational convenience:

$$\mu_{ij} = E\left[\left(X-p)^i(W-\mu_W)^j\right)\right]; \quad i,j = 0,1,2,\ldots.$$

3 Sample variance

With the following assumptions, we consider the variable Y in equation (1) as two components $Y = aX + W$. Now $Y_1, \ldots, Y_n$ be identically independent random sample from Y, and let the pairs $(X_1, W_1), \ldots, (X_n, W_n)$ represent the partitioned components associated with these samples. Let $S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, $S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, and $S_W^2 = \frac{1}{n}\sum_{i=1}^{n}(W_i - \bar{W})^2$ denote the sample variances of Y, X, and W, respectively, and $S_{X,W} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(W_i - \bar{W})$ the sample covariance between X and W. It can be easily shown that

$$S_Y^2 = a^2 S_X^2 + S_W^2 + 2aS_{X,W} \tag{3}$$

Now, from (2) and (3) we have

$$\begin{aligned}\sqrt{n}\left(S_Y^2 - \sigma_Y^2\right) &= \frac{a^2}{\sqrt{n}}\left[n\left(S_X^2 - p(1-p)\right)\right] \\ &\quad + \sqrt{n}\left(S_W^2 - \sigma_W^2\right) + 2a\sqrt{n}\left(S_{X,W} - \sigma_{X,W}\right).\end{aligned} \tag{4}$$

Note that none of the three expressions given in (4) are independent of each other. In the next section, we analyze the limiting distributions for the individual components on the right-hand side of the sample variance decomposition. We derive their means and variances, as well as the covariances between these components, providing a complete characterization of their asymptotic behavior.

3.1 Variances and covariances

For large sample sizes, the properly normalized second central moment converges in distribution to a normal distribution with variance τ^2, where $\tau^2 = \mathrm{Var}\big((X-\mu)^2\big) = \mathbb{E}\big[(X-\mu)^4\big] - \sigma^4$ represents the asymptotic variance of the sample variance. For finite sample sizes, the second central moment has mean σ^2 and variance $\mathrm{Var}(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$, where $\mu_4 = \mathbb{E}\big[(X-\mu)^4\big]$ denotes the fourth central moment (Kendall and Stuart, 1977). Similar to the finite sample case, it can be demonstrated that

$$\begin{aligned}\mathrm{Var}(S_W^2) &= \frac{1}{n}\left[E\big((W-\mu_W)^4\big) - \frac{n-3}{n-1}\sigma_W^4\right], \\ \mathrm{Var}(S_X^2) &= \frac{1}{n}\left[E\big((X-p)^4\big) - \frac{n-3}{n-1}p^2(1-p)^2\right], \\ \mathrm{Var}(S_{X,W}) &= \frac{1}{n}\left[E\big((X-p)^2(W-\mu_W)^2\big) - \frac{n-3}{n-1}\sigma_{X,W}^2\big)\right].\end{aligned}$$

The exact expressions include higher-order terms, such as $o(1/n)$ and $O(1/n^2)$, which vanish as $n \to \infty$ and are thus omitted for clarity. The direct computation of the variances and covariances among the components of equation (4) is highly time consuming. Instead, by leveraging the known limiting variance of the overall sample variance and examining the structural form and the influence of the coefficient a, we can effectively infer the variances and covariances of the individual components, providing a practical and accurate approximation for their asymptotic behavior. Although the factor

$(n-3)/(n-1)$ approaches to one in the limit, performing the decomposition as in (4) allows us to directly determine the variances and covariances of the component random variables in a transparent manner. To this end, we first derive the fourth central moment of Y:

$$\begin{aligned}\mu_4 &= E\left[(Y-\mu_Y)^4\right] \\ &= a^4E\left[(X-p)^4\right]+4a^3E\left[(X-p)^3(W-\mu_W)\right] \\ &\quad +6a^2E\left[(X-p)^2(W-\mu_W)^2\right]+4aE\left[(X-p)(W-\mu_W)^3\right]+E\left[(W-\mu_W)^4\right].\end{aligned}$$

The squared variance of Y is

$$\begin{aligned}\sigma_Y^4 &= \left(a^2\sigma_X^2+\sigma_W^2+2a\sigma_{X,W}\right)^2 \\ &= a^4\sigma_X^4+4a^3\sigma_X^2\sigma_{X,W}+a^2\left(4\sigma_{X,W}^2+2\sigma_X^2\sigma_W^2\right)+4a\sigma_{X,W}\sigma_W^2+\sigma_W^4.\end{aligned}$$

After substituting the fourth central moment and the squared variance into (5), a straightforward algebraic simplification yields the variances and pairwise covariances of the statistics.

$$\begin{aligned}V &= \operatorname{Var}\left(\sqrt{n}\left(S_Y^2-\sigma_Y^2\right)\right)=\mu_4-\frac{n-3}{n-1}\sigma_Y^4 \\ &= a^4A+4a^3B+4a^2C+2a^2D+4aE+F, \qquad (5)\end{aligned}$$

where

$$\begin{aligned}A &= \operatorname{Var}\left(\sqrt{n}\left(S_X^2-\sigma_X^2\right)\right)=E\left[(X-p)^4\right]-\frac{n-3}{n-1}\sigma_X^4, \\ B &= \operatorname{Cov}\left(\sqrt{n}\left(S_X^2-\sigma_X^2\right),\sqrt{n}\left(S_{X,W}-\sigma_{X,W}\right)\right)=\mu_{31}-\frac{n-3}{n-1}\sigma_X^2\sigma_{X,W}, \qquad (6) \\ C &= \operatorname{Var}\left(\sqrt{n}\left(S_{X,W}-\sigma_{X,W}\right)\right)=\mu_{22}-\frac{n-3}{n-1}\sigma_{X,W}^2, \\ D &= \operatorname{Cov}\left(\sqrt{n}\left(S_X^2-\sigma_X^2\right),\sqrt{n}\left(S_W^2-\sigma_W^2\right)\right)=\mu_{22}-\frac{n-3}{n-1}\sigma_X^2\sigma_W^2, \qquad (7) \\ E &= \operatorname{Cov}\left(\sqrt{n}\left(S_{X,W}^2-\sigma_{X,W}\right),\sqrt{n}\left(S_W^2-\sigma_W^2\right)\right)=\mu_{13}-\frac{n-3}{n-1}\sigma_{X,W}\sigma_W^2, \qquad (8) \\ F &= \operatorname{Var}\left(\sqrt{n}\left(S_W^2-\sigma_W^2\right)\right)=E\left[(W-\mu_W)^4\right]-\frac{n-3}{n-1}\sigma_W^4.\end{aligned}$$

Based on the relationships established between X and W, and $p=1/2$, we have $\sigma_{X,W}=\frac{-1}{2}\mu_W$, $\mu_{31}=-\frac{1}{8}\mu_W$, $\mu_{22}=\frac{1}{4}\sigma_W^2$, and $\mu_{13}=-\frac{1}{2}\left(\mu_{03}+\mu_W^3\right)$. Therefore,

$$\begin{aligned}A &= \operatorname{Var}\left(\sqrt{n}\left(S_X^2-\frac{1}{4}\right)\right)=\frac{1}{8(n-1)}, \\ B &= \operatorname{Cov}\left(\sqrt{n}\left(S_X^2-\frac{1}{4}\right),\sqrt{n}\left(S_{X,W}-\mu_{X,W}\right)\right)=-\frac{1}{4}\frac{n-2}{n-1}\mu_W, \\ C &= \operatorname{Var}\left(\sqrt{n}\left(S_{X,W}-\sigma_{X,W}\right)\right)=\frac{1}{4}\left(\sigma_W^2-\frac{n-3}{n-1}\mu_W^2\right), \\ D &= \operatorname{Cov}\left(\sqrt{n}\left(S_X^2-\frac{1}{4}\right),\sqrt{n}\left(S_W^2-\sigma_W^2\right)\right)=\frac{1}{2}\frac{1}{(n-1)}\sigma_W^2,\end{aligned}$$

$$E = \operatorname{Cov}\left(\sqrt{n}\left(S^2_{X,W} - \sigma_{X,W}\right), \sqrt{n}\left(S^2_W - \sigma^2_W\right)\right) = \frac{-1}{2}\left(\mu_{03} + \mu^3_W - \frac{n-3}{n-1}\mu_W \sigma^2_W\right),$$
$$F = \operatorname{Var}\left(\sqrt{n}\left(S^2_W - \sigma^2_W\right)\right) = E\left[(W - \mu_W)^4\right] - \frac{n-3}{n-1}\sigma^4_W.$$

The three statistics $S^2_W, S_{X,W}$, and S^2_X are centered by subtracting their respective means and then multiplied by $\sqrt{n}$. The resulting random vector has the covariance matrix

$$\begin{pmatrix} F & E & D \\ E & C & B \\ D & B & A \end{pmatrix}.$$

We are now ready to present the limiting distributions of the three components in the structure of the sample variance given in (4).

3.2 Limiting distribution of second central moment

The asymptotic distribution of the sample variance for Bernoulli and binomial distributions was derived using a very straightforward (Abbasi et al., 2024). The work highlighted the sensitivity of the limiting distributions when the success probability is one half, providing valuable insight into the emergence of classical chi-square type behavior in these symmetric scenarios.

$$a^2 n\left(S^2_X - \frac{1}{4}\right) \xrightarrow{D} a^2\frac{-1}{4}\chi^2_1.$$

The second component from (4), $Z_1 = \sqrt{n}\left(S^2_W - \sigma^2_W\right)$ has normal distribution $N(0, F)$ and the third component, $Z_2 = \sqrt{n}\left(S_{X,W} - \sigma_{X,W}\right)$ follows normal distribution $N(0, C)$. The covariance structure among these components is determined according to B, D, and E in (6), (7) and (8), respectively. From all of the above, we can write

$$\sqrt{n}\left(S^2_Y - \sigma^2_Y\right) \sim \frac{a^2}{\sqrt{n}}\left(\frac{-1}{4}\chi^2_1\right) + N(0, F) + 2aN(0, C).$$

Two points are worth emphasizing regarding the chi-square component in the limiting model:
1-Sample-size effect: As n becomes large, the influence of the chi-square component diminishes (since its coefficient is of order $1/\sqrt{n}$). Consequently, we recover the result commonly presented in standard texts for the limiting distribution of sample moments: a normally distributed random variable with fixed coefficients.
2-Effect of the parameter a: An important issue is whether the parameter a affects the limiting model. A trivial case occurs when $a = 0$, in which case the chi-square component vanishes entirely. Another case arises when a is sufficiently large and corresponds to a mode located at the maximum (or minimum) point of the sample space. Under this configuration, the variance of Y increases substantially. As this variance becomes dominant relative to the variances of the other components, the contribution of the chi-square term to the limiting distribution becomes negligible and may effectively disappear.

4 Simulations

The limiting distribution obtained in this paper differs from the commonly found in the literature. For clarity, we refer to our result as Model II, while the conventional model is referred to as Model I.
Model I : $\sqrt{n}\left(S_Y^2 - \sigma_Y^2\right) \sim N(0, V)$,
Model II : $\sqrt{n}\left(S_Y^2 - \sigma_Y^2\right) \sim \frac{a^2}{\sqrt{n}}\left(\frac{-1}{4}\chi_1^2\right) + N(0, F) + 2aN(0, C)$. To evaluate the finite-sample performance of the two asymptotic models, we conducted a simulation studies under four discrete distributions. These distributions were selected because they represent different support sizes and different levels of dispersion and asymmetry. The four distributions used in the simulation are listed in Table 1. From distributions D1–D4, we computed the corresponding means, variances of Y and W.

For each distribution, we considered sample sizes n ranging from 20 to 300 to investigate how the performance of the two models changes as the sample size increases. These values were chosen to reflect small, moderate, and relatively large samples, so that the transition and convergence behavior of the asymptotic approximations could be assessed clearly.

Table 1: The four probability distributions employed in the simulation.

Distribution	Support Y	Probabilities	σ_Y^2	μ_4	σ_W^2
D1	-2, 1, 2	$\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$	$\frac{9}{4}$	$\frac{177}{16}$	2
D2	-2, -1, 1, 3	$\frac{1}{6}, \frac{1}{6}, \frac{1}{2}, \frac{1}{6}$	$\frac{31}{12}$	$\frac{667}{48}$	$\frac{7}{3}$
D3	-3, -2, 1, 2, 3	$\frac{1}{8}, \frac{1}{8}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}$	$\frac{7}{2}$	$\frac{467}{16}$	$\frac{13}{4}$
D4	-4, -3, -2, 1, 2, 3, 4	$\frac{1}{24}, \frac{2}{24}, \frac{3}{24}, \frac{1}{2}, \frac{3}{24}, \frac{2}{24}, \frac{1}{24}$	$\frac{49}{12}$	$\frac{2143}{48}$	$\frac{23}{6}$

To compare the two models, we used the Kolmogorov–Smirnov test. For Model I, a random sample of size n was drawn from the specified distribution, and the sample variance together with $T = \sqrt{n}\left(S_Y^2 - \sigma_Y^2\right)$ was computed. Based on repeated resampling, $m = 1000$ realizations of this statistic are generated. The one-sample Kolmogorov–Smirnov test is used and the corresponding p-value is calculated. This procedure was repeated $L = 2000$ times, and the mean p-value is reported.

For Model II, the simulation was based on a linear combination of two normal variables and one chi-square variable. We generated $m = 1000$ observations from this model and compared them with $m = 1000$ observations of T from the corresponding target distribution using the two-sample Kolmogorov–Smirnov test. The resulting p-value is calculated, and the procedure was again repeated $L = 2000$ times and the mean p-values are reported.

Figure 1 presents the mean p-values obtained for a range of sample sizes n under four discrete distributions D1-D4. For small sample sizes, some distributions yield unsatisfactory results. As n increases, the influence of the chi-square component in Model II diminishes, because its coefficient $1/\sqrt{n}$ converges to zero. Consequently, the outcomes of the two models become increasingly similar. Overall, the results clearly indicate that Model II outperforms Model I, as it yields consistently larger p-values. Note that for simulation studies the MASS library of R software is used.

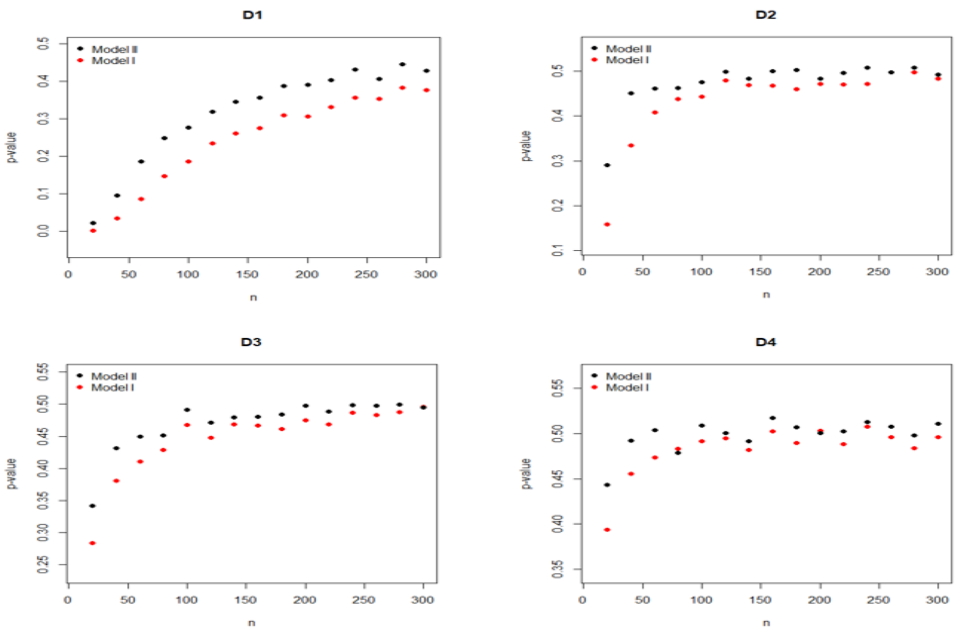

Figure 1: Mean p-values for Model I and Model II across different sample sizes under distributions D1-D4.

5 Discussion and conclusions

We investigated the asymptotic behavior of the sample variance for discrete populations by decomposing it into a collection of interpretable random components. By characterizing the limiting distributions of these components and their dependence structure, we obtained a new representation of the asymptotic distribution of the sample variance beyond the classical continuous setting. A key feature of our analysis is the explicit role played by the point at which the probability equals one half. In this case, the limiting behavior of the sample variance exhibits pronounced sensitivity, leading to the emergence of a chi-square distribution. This phenomenon, previously observed in special cases such as the Bernoulli and binomial distributions, is here embedded in a more general framework for discrete random variables. To support the theoretical findings, we conducted a simulation studies based on limiting distribution which provides an accurate approximation even for moderate sample sizes, and the goodness-of-fit results obtained from the Kolmogorov–Smirnov test further support the theoretical conclusions. The results presented in this approach extend the existing literature on the distribution of the sample variance by providing a unified asymptotic treatment tailored to discrete populations. This framework may be useful in practical applications involving discrete data, such as count data analysis, reliability studies, and categorical modeling. Possible directions for future research include extending the proposed approach to multivariate discrete settings and exploring its implications under alternative dependence structures.

Acknowledgement

The authors of this work are grateful to the reviewers of this article for their accuracy and the suggestions.

References

Abbasi, N., Keshavarz, N., Yarmohammadi, M. and Saadatmand, A. (2024). A simple method for determining the limiting distribution of sample central moments for two-point and Binomial distributions. *Journal of Statistical Modelling: Theory and Applications*, **4**(1):1–10.

Johnson, N.L., Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions.* 3rd edition, Hoboken, NJ: Wiley.

Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics, Vol. 1: Distribution Theory.* 4th edition, London: Charles Griffin.

Lehmann, E.L. and Casella, G.(1998). *Theory of Point Estimation.* 2nd edition, Springer, New York.

Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses.* 3rd edition, Springer, New York.

Le, V.H. and Vargas, R. (2025). Copula-based cosimulation for simulating temporal or spatial data in biogeosciences. *Journal of Geophysical Research: Biogeosciences*, **130**(10):e2025JG008802.

Ma, Y., Genton, M.G. and Parzen, E. (2011). Asymptotic properties of sample quantiles of discrete distributions. *The Annals of Statistics*, **39**(3):1380–1409.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

Van der Vaart, A.W. (2000). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, Series Number 3, Cambridge University Press.