International Journal of Chemistry, Mathematics and Physics (IJCMP)
[Vol-9, Issue-4, Oct-Dec, 2025]
https://dx.doi.org/10.22161/ijcmp.9.4.3
ISSN: 2456-866X

Bivariate normal-geometric distribution with FGM copula: Properties and parameter estimation

Sung-Hyon Ri*, Hyon-A Kim, Kwang-Hyok Kim

Faculty of Applied Mathematics, Kim Chaek University of Technology, Pyongyang, Democratic People's Republic of Korea
lsh6579@star-co.net.kp (Sung-Hyon Ri), kha 05426@star-co.net.kp (Hyon-A Kim), kkh9511@star-co.net.kp (Kwang-Hyok Kim)
*Corresponding author

Received: 23 Sep 2025; Received in revised form: 19 Oct 2025; Accepted: 23 Oct 2025; Available online: 27 Oct 2025

***Abstract*— *It is of practical importance to construct new effective statistical model for discrete and continuous data arising in the real world. The bivariate distribution with FGM copula whose two marginal distributions are normal and geometric, respectively, is proposed. The explicit characteristics and the moment generating function of the distribution are derived. Also, the conditional distributions and their characteristics of the distribution are obtained. The moment and maximum likelihood (ML) estimates for the parameters are investigated and the numerical results for them are presented. Real data analysis that indicates the usefulness of the model is also shown.*

Keywords*— *FGM copula, moment generating function, conditional distribution, moment estimation, ML estimation

I. INTRODUCTION

In analyzing data from reliability, economics, health sciences and other advanced areas, probability distributions are effectively used. Many new distributions are created to describe real-world data more accurately.

Copulas are an important tool to express the relationship between random variables and they are widely used in constructing multidimensional distributions (see [8], [11], [12] and [20]). Using copula, the bivariate normal-Weibull distributions [23], bivariate generalized Rayleigh distributions [7], and bivariate Fréchet distributions [3] have been studied. In [5], [6] and [17], several bivariate distributions have been constructed using Marshall-Olkin copula.

The Farlie-Gumbel-Morgenstern (FGM) copula proposed by Morgenstern in 1956 is one of the most widely known parametric copula. The bivariate FGM copula is given as:

$$C(u_1,u_2)=u_1u_2[1+\alpha(1-u_1)(1-u_2)],\ -1\le\alpha\le 1 .$$

Let (X_1, X_2) have FGM copula, then the cumulative distribution function (cdf) of (X_1, X_2) is expressed as:

$$F(x_1,x_2)=F_1(x_1)F_2(x_2)[1+\alpha(1-F_1(x_1))(1-F_2(x_2))],\quad (x_1,x_2)\in R^2 \tag{1}$$

where $F_i(x)$ (i=1, 2) is the cdf of X_i.

Using FGM copula, the continuous bivariate distributions of different types, e. g., a bivariate exponential distribution with normal and exponential marginals [10], a bivariate generalized exponential [1], a bivariate Weibull distribution [2], a bivariate modified Weibull distribution [21], and a bivariate Teissier distribution [22] have been presented and their statistical properties were investigated.

For discrete bivariate distributions, we refer to [13]. The bivariate geometric and Poisson distributions [24] and a bivariate discrete generalized exponential distribution [18] have been proposed and their properties and parameter estimation have been studied. Several negatively correlated bivariate Poisson distributions have been discussed in [9] and a model with Poisson marginals that allows negative correlation have been developed in [16]. Nelsen [19] has constructed probability functions for bivariate discrete random variables with any correlation using convex linear combinations of probability functions for Fréchet marginals. A bivariate discrete Weibull distribution that is a discrete analogy of the Marshall-Olkin bivariate Weibull distribution have been introduced and its various properties have been studied in [15]. In [14], a bivariate continuous-discrete distribution with i.i.d. exponential sum and geometric marginal distributions has been constructed and used to analyze the finance data.

However, there are few papers on the results concerning the distribution with continuous and discrete marginal distributions, and to the best of our knowledge, the construction of a continuous-discrete model using a copula has not been discussed.

In the real world, as the grade level, classification index, etc. are represented by discrete variables, its observation and the continuous index will require various continuous-discrete model. The aim of this paper is to propose a bivariate normal-geometric distribution with FGM copula and show its properties.

The rest of the paper is organized as follows. In Section 2, the bivariate normal-geometric distribution with FGM copula is defined, and the probabilistic properties of the distribution are described. Also, the explicit expressions for the conditional distributions and the conditional characteristics of the distribution are derived. In Section 3, the moment and ML estimates for the parameters are investigated and the numerical results for them are presented. In Section 4, the analysis of the real data between the grade level and weight is described. Finally, Section 5 concludes this paper

II. BIVARIATE NORMAL-GEOMETRIC DISTRIBUTION AND THEIR PROPERTIES

Let X_1 and X_2 be two iid copies of a random variable X with a cdf $F(x)$, then the cdf of $X^{(2)}=\max\{X_1, X_2\}$ is $F^{(2)}(x) = F(x)^2$. Therefore, Eq. (1) can be expressed as:

$$F(x_1,x_2) = F_1(x_1)F_2(x_2) + \alpha[F_1(x_1) - F_1^{(2)}(x_1)][F_2(x_2) - F_2^{(2)}(x_2)]\,. \tag{2}$$

Let $p_k = P\{K = k\},\ k = 0, \pm 1, \pm 2, \cdots$ be the probability mass function (pmf) of a random variable K taking a nonnegative integer values, then the pmf of $K^{(2)} = \max\{K_1, K_2\}$ is expressed as:

$$p_k^{(2)} = p_k^2 + 2p_k \sum_{j<k} p_j,\ k = 0, \pm 1, \pm 2, \cdots.$$

Proposition 2.1. Let X be a continuous random variable with the probability density function (pdf) $f(x)$ and K a discrete random variable with the pmf $p_k=P\{K=k\}$, $k = 0, \pm 1, \pm 2, \cdots$, and (X, K) have FGM copula. Then the pdf of (X, K) is expressed as:

$$f(x,k) = \frac{d}{dx} P\{X \le x, K = k\} = f(x)p_k + \alpha(f(x) - f^{(2)}(x))(p_k - p_k^{(2)})\,, \tag{3}$$

where $f^{(2)}(x)$ is the pdf of $F^{(2)}(x)$.

Proof. From Eq. (2), we have

$$P\{X \le x, K = k\} = F(x)p_k + \alpha(F(x) - F^{(2)}(x))(p_k - p_k^{(2)}) .$$

Differentiating the above expression by x yields the result of the proposition. □

2.1. Definition and characteristics

Let X~$N(0, 1)$, then the cdf (pdf) of $X^{(2)}$ is $\Phi(x)^2$ ($\varphi^{(2)}(x) = 2\varphi(x)\Phi(x)$), where $\varphi(x)$ and $\Phi(x)$ are the pdf and cdf of the standard normal distribution, respectively [4]. This is the skew normal distribution with slant parameter 1. The mathematical expectation and variance of $X^{(2)}$, respectively, are

$$EX^{(2)} = \frac{1}{\sqrt{\pi}} , \ VarX^{(2)} = 1 - \frac{1}{\pi} . \tag{4}$$

Let K follows geometric distribution $Ge(q)$, i. e. $P\{K = k\} = (1-q)q^{k-1} = q^{k-1} - q^k , k = 1,2,\cdots$,

then the pmf of $K^{(2)}$ is as:

$$P\{K^{(2)} = k\} = 2A(k, q) - A(k, q^2) , \quad k = 1,2,\cdots,$$

where $A(k,q) = q^{k-1} - q^k$.

Definition 2.1. Let the bivariate random variable (X, K) have the FGM copula, and $X \sim N(\mu, \sigma^2)$ and $K \sim G(q)$. Then, we say that (X, K) follows the bivariate normal-geometric distribution with FGM copula and denote by $(X, K) \sim NGF(\mu, \sigma^2, q; \alpha)$.

From (3), the pdf of $(X, K) \sim NGF(\mu, \sigma^2, q; \alpha)$ is expressed as:

$$f(x,k) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})\left[A(k,q) + \alpha\left(1 - 2\Phi(\frac{x-\mu}{\sigma})\right)(A(k,q^2) - A(k,q))\right]. \tag{5}$$

First, we find the moment generating function of $NGF(\mu, \sigma^2, q; \alpha)$.

Proposition 2.2. The moment generating function of $(X, K) \sim NGF(\mu, \sigma^2, q; \alpha)$ is given by

$$M(t,z) = Ee^{tX+zK} = e^{\mu t + \frac{\sigma^2 t^2}{2}}\left[\frac{(1-q)e^z}{1-qe^z} + \alpha\left(1 - 2\Phi(\frac{\sigma t}{\sqrt{2}})\right)\left(\frac{(1-q)e^z}{1-qe^z} + \frac{(1-q^2)e^z}{1-q^2e^z}\right)\right].$$

Proof. Let $X_0 \sim N(0,1)$ and $K \sim G(q)$ then the moment generating functions of $X_0, K, X_0^{(2)}$ and $K^{(2)}$ are respectively expressed as [4]:

$$M_{X_0}(t) = Ee^{tX_0} = e^{\frac{t^2}{2}} ,$$

$$M_{X_0}^{(2)}(t) = Ee^{tX_0^{(2)}} = 2e^{\frac{t^2}{2}}\Phi(\frac{t}{\sqrt{2}}) ,$$

$$M_K(z) = Ee^{zK} = \frac{(1-q)e^z}{1-qe^z} ,$$

$$M_K^{(2)} = Ee^{zK^{(2)}} = \frac{2(1-q)e^z}{1-qe^z} - \frac{(1-q^2)e^z}{1-q^2e^z}.$$

From Eq. (3), the moment generating function of $(X_0, K) \sim NGF(0,1,q;\alpha)$ can be obtained as

$$M_0(t,z) = M_{X_0}(t)M_K(z) + \alpha(M_{X_0}(t) - M_{X_0}^{(2)}(t))(M_K(z) - M_K^{(2)}(z))$$

$$= e^{\frac{t^2}{2}}\left[\frac{(1-q)e^z}{1-qe^z} + \alpha\left(1 - 2\Phi(\frac{t}{\sqrt{2}})\right)\left(\frac{(1-q)e^z}{1-qe^z} + \frac{(1-q^2)e^z}{1-q^2e^z}\right)\right].$$

While, let $X = \mu + \sigma X_0$, then $(X,K) \sim NGF(\mu, \sigma^2, q; \alpha)$, so the moment generating function of (X, K) is

$$M(t,z) = Ee^{tX+zK} = e^{\mu t}M_0(\sigma t, z).$$

From the above, the result of the proposition is obtained. □

Next, the mathematical expectation and the covariance matrix of $(X,K) \sim NGF(\mu, \sigma^2, q; \alpha)$ are derived, which are used for the moment estimation of the parameter considered in Section 3.

Proposition 2.3. Let $(X,K) \sim NGF(\mu, \sigma^2, q; \alpha)$, then

$$EX = \mu,\ EK = 1/(1-q),\ VarX = \sigma^2,\ VarK = q/(1-q)^2,$$

and the correlation coefficient between X and Y is

$$r(X,K) = \frac{\alpha\sqrt{q}}{\sqrt{\pi}(1+q)}.$$

Therefore, the covariance matrix of $(X, K) \sim NGF(\mu, \sigma^2, q; \alpha)$ is expressed as:

$$Cov(X,K) = \begin{pmatrix} \sigma^2 & \frac{\alpha\sigma q}{\sqrt{\pi}p(1+q)} \\ \frac{\alpha\sigma q}{\sqrt{\pi}(1-q)(1+q)} & \frac{q}{(1-q)^2} \end{pmatrix}. \quad (6)$$

Proof. Let $X_0 = (X-\mu)/\sigma$, then from Eq. (3)

$$E(X_0K) = EX_0EK + \alpha(EX_0 - EX_0^{(2)})(EK - EK^{(2)}),$$

While, from Eq. (4),

$$EX_0^{(2)} = \frac{1}{\sqrt{\pi}},\quad EK = \frac{1}{1-q},\ EK^{(2)} = \frac{1+2q}{(1-q)(1+q)}.$$

Therefore, we have

$$Cov(X_0,K) = \frac{\alpha q}{\sqrt{\pi}(1-q)(1+q)}.$$

The correlation coefficient of X and K is

$$r(X,K)=\frac{\alpha\sqrt{q}}{\sqrt{\pi}(1+q)}.$$

The rest is obvious. □

From the above proposition, we can see that $|r(X,K)|\le(2\sqrt{\pi})^{-1}=0.2821$. This shows that $NGF(\mu,\sigma^2,q;\alpha)$ can only be used as a model for weakly related discrete and continuous random variables, but not for strongly correlated cases.

2.2 Conditional distributions and conditional characteristics

In this subsection, we show the conditional distributions of the bivariate normal-geometric distribution with FGM copula and explicit expressions for their characteristics. Then, we describe the sampling method from $NGF(\mu,\sigma^2,q;\alpha)$

Let $(X,K)\sim NGF(\mu,\sigma^2,q;\alpha)$. Using Eq. (5) and Eq. (6), the conditional pdf of X given $K=k$ is is derived as follows:

$$f_{X|K=k}(x)=\frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})[1+\alpha(1-2\Phi(\frac{x-\mu}{\sigma}))(q^{k-1}+q^k-1)] \tag{7}$$

The conditional expectation and variance of X given $K=k$ are respectively obtained as follows:

$$E(X|K=k)=\mu-\frac{\alpha\sigma}{\sqrt{\pi}}(q^{k-1}+q^k-1),$$

$$Var(X|K=k)=\sigma^2\left(1-\frac{\alpha^2(q^{k-1}+q^k-1)^2}{\pi}\right)$$

Also, the conditional pmf of K given $X=x$ is derived as:

$$f_{K|X=x}(k)=\left[1-\alpha(1-2\Phi(\frac{x-\mu}{\sigma}))\right](q^{k-1}-q^k)+\alpha\left[1-2\Phi(\frac{x-\mu}{\sigma})\right](q^{2k-2}-q^{2k}).$$

The conditional expectation and variance of K given $X=x$ are respectively obtained as follows:

$$E(K|X=x)=\frac{1}{1-q}-\alpha\left[1-2\Phi(\frac{x-\mu}{\sigma})\right]\frac{q}{1-q^2},$$

$$Var(K|X=x)=\frac{q}{(1-q)^2}-\alpha\left[1-2\Phi(\frac{x-\mu}{\sigma})\right]\frac{q(1+q^2)}{(1-q^2)^2}+\alpha^2\left[1-2\Phi(\frac{x-\mu}{\sigma})\right]^2\frac{q^2}{(1-q^2)^2}.$$

Next, we consider the sampling method from $NGF(\mu,\sigma^2,q;\alpha)$. From Eq. (7), the conditional cdf of X given $K=k$ is respectively obtained a

$$F_{X|K=k}(x)=(1+D(k))\Phi(\frac{x-\mu}{\sigma})-D(k)\Phi(\frac{x-\mu}{\sigma})^2,$$

where $D(k)=\alpha(q^{k-1}+q^k-1)$. Then, sample x from $F_{X|K=k}(x)$ is generated as follows:

$$x=\mu+\sigma\Phi^{-1}\left(\frac{1+D(k)\pm\sqrt{(1+D(k))^2-4D(k)u}}{2D(k)}\right), \tag{8}$$

where u is a random number of unit uniform distribution $U[0, 1]$ and Φ^{-1} is the inverse function of Φ. And, among the two signs $\pm$, it is selected that the value in brackets enters the interval (0, 1).

Algorithm 1. (Sampling generation algorithm from $NGF(\mu, \sigma^2, q; \alpha)$).

Step 1. Generate k from $G(q)$.

Step 2 Generate u from $U[0, 1]$. Find x from Eq. (8).

III. ESTIMATION OF PARAMETERS

In this section, we consider the moment and the ML estimations of the parameters of the bivariate normal-geometric distribution. Let $(x_i, k_i), i = 1, 2, \cdots, n$ be n iid samples from $NGF(\mu, \sigma^2, q; \alpha)$.

3.1. Moment estimation

By Proposition 2.3, the moment estimators of the parameters μ, σ^2, q and α can be respectively constructed as:

$$\tilde{\mu} = \bar{x} = n^{-1}\sum_{i=1}^{n} x_i ,$$

$$\tilde{\sigma}^2 = s^2 = n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 ,$$

$$\tilde{q} = 1 - \bar{k}^{-1}, \quad \bar{k} = n^{-1}\sum_{i=1}^{n} k_i ,$$

$$\tilde{\alpha} = \frac{\sqrt{\pi}(1-\tilde{q}^2)}{\tilde{\sigma}\tilde{q}}\left(n^{-1}\sum_{i=1}^{n} x_i k_i - \bar{x}\bar{k}\right).$$

3.2. ML estimation

From Eq. (5), the likelihood function is written as follows:

$$L(\mu, \sigma, q, \alpha) = \prod_{i=1}^{n}\frac{1}{\sigma}\varphi(z_i)[A(k_i, q) + \alpha(1 - 2\Phi(z_i))(A(k_i, q^2) - A(k_i, q))],$$

where $z_i = \dfrac{x_i - \mu}{\sigma}$. Then the log-likelihood function becomes:

$$l = \ln L(\mu, \sigma, q, \alpha) = -n\ln\sigma + \sum_{i=1}^{n}\ln\varphi(z_i) + \sum_{i=1}^{n}\ln D(\mu, \sigma, q, \alpha, z_i, k_i) ,$$

where $D(\mu, \sigma, q, \alpha, z, k) = (1-\alpha)A(k, q) + \alpha A(k, q^2) - 2\alpha\Phi(z)(A(k, q^2) - A(k, q))$.

The ML estimator is obtained by numerical methods, using the Newton-Raphson method. The initial approximation can be put into the moment estimate as usual.

For convenience, let denote the parameter by $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\mu, \sigma, q, \alpha)$. The gradient and Hess matrix of the log-likelihood function, respectively, are

$$\nabla l = (g_1, g_2, g_3, g_4)^T, \quad H = (h_{ij})_{4\times 4},$$

where $g_i = \dfrac{\partial l}{\partial \theta_i}$, $h_{ij} = \dfrac{\partial^2 l}{\partial \theta_i \partial \theta_j}$. Using

$$dz_i / d\mu = -1/\sigma, \; dz_i / d\sigma = -z_i / \sigma, \; d\varphi(x)/dx = -x\varphi(x),$$

the components of the gradient can be calculated as follows:

$$g_1 = \frac{dl}{d\mu} = \frac{1}{\sigma}\left(\sum_{i=1}^n z_i + 2\alpha \sum_{i=1}^n \frac{\varphi(z_i)(A(k_i, q^2) - A(k_i, q))}{D(\mu, \sigma, q, \alpha, z_i, k_i)} \right),$$

$$g_2 = \frac{dl}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\left(\sum_{i=1}^n z_i^2 + 2\alpha \sum_{i=1}^n \frac{z_i\varphi(z_i)(A(k_i, q^2) - A(k_i, q))}{D(\mu, \sigma, q, \alpha, z_i, k_i)} \right),$$

$$g_3 = \frac{dl}{dq} = \frac{1}{q}\sum_{i=1}^n \frac{k_i A(k_i, q) - q^{k_i - 1} + \alpha(1 - 2\Phi(z_i))[k_i(2A(k_i, q^2) - A(k_i, q)) + q^{k_i - 1}(1 - 2q^{k_i - 1})]}{D(\mu, \sigma, q, \alpha, z_i, k_i)},$$

$$g_4 = \frac{dl}{d\alpha} = \sum_{i=1}^n \frac{-A(k_i, q) + A(2k_i, q) - 2\Phi(z_i)(A(k_i, q^2) - A(k_i, q))}{D(\mu, \sigma, q, \alpha, z_i, k_i)}.$$

Also, the components of the Hess matrix can be calculated as follows:

$$h_{11} = \frac{\partial^2 l}{\partial \mu^2} = \frac{1}{\sigma^2}\left(-n + 2\alpha \sum_{i=1}^n \frac{L_{\mu\mu}(\mu, \sigma, q, \alpha, z_i, k_i)}{D(\mu, \sigma, q, \alpha, z_i, k_i)^2} \right),$$

$$h_{12} = \frac{\partial^2 l}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^2}\left(\sum_{i=1}^n z_i - \alpha \sum_{i=1}^n \frac{L_{\mu\sigma}(\mu, \sigma, q, \alpha, z_i, k_i)}{D(\mu, \sigma, q, \alpha, z_i, k_i)^2} \right),$$

$$h_{13} = \frac{\partial^2 l}{\partial \mu \partial q} = \frac{2\alpha}{\sigma q}\sum_{i=1}^n \frac{L_{\mu q}(\mu, \sigma, q, \alpha, z_i, k_i)}{D(\mu, \sigma, q, \alpha, z_i, k_i)^2},$$

$$h_{14} = \frac{\partial^2 l}{\partial \mu \partial \alpha} = \frac{2}{\sigma}\sum_{i=1}^n \frac{L_{\mu\alpha}(\mu, \sigma, q, \alpha, z_i, k_i)}{D(\mu, \sigma, q, \alpha, z_i, k_i)^2},$$

$$h_{22} = \frac{\partial^2 l}{\partial \sigma^2} = \frac{1}{\sigma^2}\left(n - 3\sum_{i=1}^n z_i^2 + 2\alpha \sum_{i=1}^n \frac{L_{\sigma\sigma}(\mu, \sigma, q, \alpha, z_i, k_i)}{D(\mu, \sigma, q, \alpha, z_i, k_i)^2} \right),$$

$$h_{23} = \frac{\partial^2 l}{\partial \sigma \partial q} = \frac{2\alpha}{\sigma q}\sum_{i=1}^n \frac{L_{\sigma q}(\mu, \sigma, q, \alpha, z_i, k_i)}{D(\mu, \sigma, q, \alpha, z_i, k_i)^2},$$

$$h_{24}=\frac{\partial^2 l}{\partial\sigma\partial\alpha}=\frac{2}{\sigma}\sum_{i=1}^{n}\frac{L_{\sigma\alpha}(\mu,\sigma,q,\alpha,z_i,k_i)}{D(\mu,\sigma,q,\alpha,z_i,k_i)^2},$$

$$h_{33}=\frac{\partial^2 l}{\partial q^2}=\frac{1}{q^2}\sum_{i=1}^{n}\frac{L_{qq}(\mu,\sigma,q,\alpha,z_i,k_i)}{D(\mu,\sigma,q,\alpha,z_i,k_i)^2},$$

$$h_{34}=\frac{\partial^2 l}{\partial q\partial\alpha}=\frac{1}{q}\sum_{i=1}^{n}\frac{L_{q\alpha}(\mu,\sigma,q,\alpha,z_i,k_i)}{D(\mu,\sigma,q,\alpha,z_i,k_i)^2},$$

$$h_{44}=\frac{\partial^2 l}{\partial\alpha^2}=\sum_{i=1}^{n}\frac{L_{\alpha\alpha}(\mu,\sigma,q,\alpha,z_i,k_i)}{D(\mu,\sigma,q,\alpha,z_i,k_i)^2},$$

where

$$\begin{aligned}L_{\mu\mu}&=\varphi(z_i)(A(k_i,q^2)-A(k_i,q))\times\\&\times[z_i((1-\alpha)A(k_i,q)+\alpha A(k_i,q^2)-2\alpha\Phi(z_i)(A(k_i,q^2)-A(k_i,q)))-\\&-2\alpha\varphi(z_i)(A(k_i,q^2)-A(k_i,q))]\end{aligned}$$

$$\begin{aligned}L_{\mu\sigma}&=\varphi(z_i)(A(k_i,q^2)-A(k_i,q))\times\\&\times[(z_i^2-1)((1-\alpha)A(k_i,q)+\alpha A(k_i,q^2)-2\alpha\Phi(z_i)(A(k_i,q^2)-A(k_i,q)))-\\&-2\alpha z_i\varphi(z_i)(A(k_i,q^2)-A(k_i,q))]\end{aligned}$$

$$L_{\mu q}=\varphi(z_i)[k_iA(k_i,q)A(k_i,q^2)-2q^{2k_i-2}A(k_i,q)+q^{k_i-1}A(k_i,q^2)]$$

$$L_{\mu\alpha}=\varphi(z_i)A(k_i,q)[A(2k_i,q)-A(k_i,q)]$$

$$\begin{aligned}L_{\sigma\sigma}&=z_i\varphi(z_i)(A(k_i,q^2)-A(k_i,q))\times\\&\times[(z_i^2-2)((1-\alpha)A(k_i,q)+\alpha A(k_i,q^2)-2\alpha\Phi(z_i)(A(k_i,q^2)-A(k_i,q)))-\\&-2\alpha z_i\varphi(z_i)(A(k_i,q^2)-A(k_i,q))]\end{aligned}$$

$$L_{\sigma q}=z_i\varphi(z_i)[k_iA(k_i,q)A(k_i,q^2)-2q^{2k_i-2}A(k_i,q)+q^{k_i-1}A(k_i,q^2)]$$

$$L_{\sigma\alpha}=z_i\varphi(z_i)A(k_i,q)[A(k_i,q^2)-A(k_i,q)]$$

$$\begin{aligned}L_{qq}&=\alpha^2(1-2\Phi(z_i))^2[k_iA(k_i,q)A(k_i,q^2)(3-k_i)-k_i(A(k_i,q)^2+2A(k_i,q^2)^2)-\\&-q^{2k_i-2}(1-2q^{k_i-1})^2+2q^{2k_i-2}(2k_iA(k_i,q)-3A(k_i,q)+3A(k_i,q^2))+\\&+2q^{k_i-1}(A(k_i,q)-A(k_i,q^2)-k_iA(k_i,q^2))]+\\&+\alpha(1-2\Phi(z_i))[k_iA(k_i,q)(2A(k_i,q)+k_iA(k_i,q^2)-3A(k_i,q^2))-\\&-2A(k_i,q)q^{2k_i-2}(2k_i-3)+2q^{k_i-1}(k_iA(k_i,q^2)-2A(k_i,q)+A(k_i,q^2))+2q^{2k_i-2}(1-2q^{k_i-1})]+\\&+2A(k_i,q)q^{k_i-1}-k_iA(k_i,q)^2-q^{2k_i-2}\end{aligned}$$

$$L_{q\alpha}=(1-2\Phi(z_i))[k_iA(k_i,q)A(k_i,q^2)-2q^{2k_i-2}A(k_i,q)+q^{k_i-1}A(k_i,q^2)]$$

$$L_{\alpha\alpha}=-(1-2\Phi(z_i))^2[A(k_i,q^2)-A(k_i,q)]^2.$$

3.3. Numerical results

We present the results of numerical calculations for the moment estimates and the ML estimates for the parameters μ, σ^2, q and α. The ML estimates are calculated according to the following algorithm.

Algorithm 2.

Step 1. The initial value of parameter θ, $\theta^{(0)}$, is set to the moment estimate.

Step 2. Set j=j+1 and update as follows.

$$\theta^{(j+1)} = \theta^{(j)} - H^{-1}(\theta^{(j)})\nabla l(\theta^{(j)}).$$

For the pre-fixed $\varepsilon > 0$, if $\left|\theta_i^{(j+1)} - \theta_i^{(j)}\right| < \varepsilon,\ i = 1, 2, 3, 4$, then the computation is stopped and the ML estimate is found to be $\hat{\theta} = \theta^{(j+1)}$, otherwise go to step 2.

The bias and the standard error of the (moment or ML) estimator $\tilde{\theta}$ of the parameter θ are calculated with n estimates $\tilde{\theta}_i,\ i = 1, \cdots, n$, as follows, respectively.

$$BIAS = \frac{1}{n}\sum_{t=1}^{n}\tilde{\theta}_t - \theta,\quad SE = \frac{SD}{\sqrt{n}}\quad \text{(where } SD^2 = \frac{1}{n-1}\sum_{t=1}^{n}(\tilde{\theta}_i - \theta)^2\text{)}.$$

The results of the calculations for $\mu = 0,\ \sigma = 1,\ q = 0.7$, $\alpha = \pm 0.2, \pm 0.5, \pm 0.7$ and $n = 10^4$ are presented in the tables below.

Table 1. Biases and standard errors in brackets for the moment and ML estimates of parameters with samples of size n = 25, 50, 100 and 150 from $NGF(\mu, \sigma^2, q; \alpha)$ *in the case when* $\mu = 0, \sigma = 1, q = 0.7, \alpha = 0.2$ *and* $n = 10^4$.

	Moment				ML			
n	μ	σ	q	α	μ	σ	q	α
25	0.0046	-0.0301	-0.0086	-0.0242	0.0024	-0.0257	-0.0073	0.0174
	(0.2022)	(0.1452)	(0.0526)	(0.5410)	(0.2034)	(0.1455)	(0.0526)	(0.5253)
50	0.0018	-0.0156	-0.0045	0.0124	-0.0015	-0.0152	-0.0045	0.0077
	(0.1423)	(0.1001)	(0.0365)	(0.4353)	(0.1405)	(0.0998)	(0.0365)	(0.4132)
100	0.0005	-0.0087	-0.0021	-0.0068	-0.0003	-0.0087	-0.0021	-0.0036
	(0.0992)	(0.0710)	(0.0253)	(0.3536)	(0.0992)	(0.0710)	(0.0253)	(0.3086)
150	0.0002	-0.0049	-0.0012	-0.0028	0.0001	-0.0049	-0.0012	-0.0020
	(0.0820)	(0.0584)	(0.0210)	(0.2946)	(0.0820)	(0.0584)	(0.0210)	(0.2522)

Table 2. Biases and standard errors in brackets for the moment and ML estimates of parameters with samples of size n = 25, 50, 100 and 150 from $NGF(\mu, \sigma^2, q; \alpha)$ *in the case when* $\mu = 0$, $\sigma = 1$, $q = 0.7$, $\alpha = -0.2$ *and n = 10*^{4}.

	Moment				ML			
n	μ	σ	q	α	μ	σ	q	α
25	-0.0026	-0.0311	-0.0088	-0.0258	0.0018	-0.0313	-0.0089	0.0175
	(0.2000)	(0.1431)	(0.0534)	(0.6839)	(0.1938)	(0.1430)	(0.0534)	(0.6531)
50	-0.0017	-0.0171	-0.0044	0.0132	-0.0014	-0.0171	-0.0043	-0.0050
	(0.1417)	(0.1020)	(0.0365)	(0.4738)	(0.1421)	(0.1020)	(0.0365)	(0.4294)
100	0.0006	-0.0072	-0.0017	0.0077	-0.0004	-0.0072	-0.0017	-0.0026
	(0.0995)	(0.0713)	(0.0254)	(0.3506)	(0.0996)	(0.0713)	(0.0254)	(0.3087)
150	0.0003	-0.0057	-0.0012	0.0017	0.0001	-0.0057	-0.0012	0.0001
	(0.0804)	(0.0580)	(0.0205)	(0.2896)	(0.0805)	(0.0580)	(0.0205)	(0.2502)

Table 3. Biases and standard errors in brackets for the moment and ML estimates of parameters with samples of size n = 25, 50, 100 and 150 from $NGF(\mu, \sigma^2, q; \alpha)$ *in the case when* $\mu = 0$, $\sigma = 1$, $q = 0.7$, $\alpha = 0.5$ *and n = 10*^{4}.

	Moment				ML			
n	μ	σ	q	α	μ	σ	q	α
25	0.0031	-0.0299	-0.0086	-0.0888	-0.0021	-0.0320	-0.0089	-0.0554
	(0.2024)	(0.1448)	(0.0526)	(0.5464)	(0.2024)	(0.1450)	(0.0528)	(0.5239)
50	0.0026	-0.0138	-0.0045	-0.0470	0.0022	-0.0155	-0.0046	-0.0219
	(0.1421)	(0.1016)	(0.0366)	(0.5066)	(0.1435)	(0.102)	(0.0369)	(0.4974)
100	0.0008	-0.0072	-0.0017	0.0175	-0.0005	-0.0072	-0.0017	-0.0087
	(0.0998)	(0.0716)	(0.0236)	(0.3506)	(0.0994)	(0.0716)	(0.0236)	(0.3087)
150	0.0005	-0.0056	-0.0014	-0.0029	0.0001	-0.0056	-0.0014	-0.0018
	(0.0816)	(0.0585)	(0.0208)	(0.2735)	(0.0812)	(0.0582)	(0.0207)	(0.2406)

Table 4. Biases and standard errors in brackets for the moment and ML estimates of parameters with samples of size n = 25, 50, 100 and 150 from $NGF(\mu, \sigma^2, q; \alpha)$ *in the case when* $\mu = 0$, $\sigma = 1$, $q = 0.7$, $\alpha = -0.5$ *and n = 10*^{4}.

	Moment				ML			
n	μ	σ	q	α	μ	σ	q	α
25	0.0042	-0.0306	-0.0080	-0.0919	0.0036	-0.0329	-0.0083	0.0597
	(0.1977)	(0.1436)	(0.0533)	(0.5491)	(0.1980)	(0.1437)	(0.0534)	(0.5271)

50	-0.0035	-0.0159	-0.0045	0.0156	-0.0019	-0.0162	-0.0046	0.0136
	(0.1413)	(0.1015)	(0.0367)	(0.4755)	(0.1415)	(0.1020)	(0.0367)	(0.4463)
100	0.0005	-0.0087	-0.0021	-0.0124	0.0002	-0.0089	-0.0021	-0.0051
	(0.0991)	(0.0709)	(0.0253)	(0.3268)	(0.0988)	(0.0708)	(0.0253)	(0.2899)
150	0.0004	-0.0048	-0.0012	0.094	0.0002	-0.0048	-0.0012	0.0035
	(0.0816)	(0.0580)	(0.0205)	(0.2747)	(0.0812)	(0.0580)	(0.0205)	(0.2408)

Table 5. Biases and standard errors in brackets for the moment and ML estimates of parameters with samples of size n = 25, 50, 100 and 150 from $NGF(\mu, \sigma^2, q; \alpha)$ *in the case when* $\mu = 0$, $\sigma = 1$, $q = 0.7$, $\alpha = 0.7$ *and* $n = 10^4$.

	Moment				ML			
n	μ	σ	q	α	μ	σ	q	α
25	-0.0032	-0.0321	-0.0081	-0.1529	-0.0021	-0.0357	-0.0086	-0.1092
	(0.2011)	(0.1434)	(0.0526)	(0.5148)	(0.2000)	(0.1434)	(0.0526)	(0.4848)
50	0.0013	-0.0135	-0.0035	-0.0818	-0.0004	-0.0154	-0.0038	-0.0488
	(0.1423)	(0.1001)	(0.0362)	(0.3776)	(0.1412)	(0.0999)	(0.0362)	(0.3449)
100	0.0006	-0.0087	-0.0022	-0.0370	-0.0002	-0.0094	-0.0022	-0.0178
	(0.0992)	(0.0708)	(0.0253)	(0.2871)	(0.0983)	(0.0707)	(0.0252)	(0.257)
150	-0.0004	-0.0053	-0.0011	-0.0265	-0.0001	-0.0056	-0.0011	-0.0119
	(0.0819)	(0.0582)	(0.0206)	(0.2440)	(0.0810)	(0.0582)	(0.0205)	(0.2181)

Table 6. Biases and standard errors in brackets for the moment and ML estimates of parameters with samples of size n = 25, 50, 100 and 150 from $NGF(\mu, \sigma^2, q; \alpha)$ *in the case when* $\mu = 0$, $\sigma = 1$, $q = 0.7$, $\alpha = -0.7$ *and* $n = 10^4$.

	Moment				ML			
n	μ	σ	q	α	μ	σ	q	α
25	-0.0044	-0.0301	-0.0083	0.1283	-0.0036	-0.0327	-0.0086	-0.1026
	(0.2011)	(0.1443)	(0.0526)	(0.5345)	(0.2001)	(0.1444)	(0.0526)	(0.4738)
50	-0.0016	-0.0138	-0.0035	0.0823	-0.0014	-0.0145	-0.0035	0.0483
	(0.1423)	(0.1002)	(0.0363)	(0.3735)	(0.1402)	(0.0999)	(0.0362)	(0.3324)
100	-0.0009	-0.0076	-0.0019	-0.0367	-0.0002	-0.0080	-0.0018	0.0174
	(0.0983)	(0.0706)	(0.0253)	(0.2838)	(0.0973)	(0.0704)	(0.0253)	(0.2550)
150	0.0008	-0.0049	-0.0011	0.0134	0.0003	-0.0052	-0.0011	0.0020
	(0.0804)	(0.0580)	(0.0207)	(0.2408)	(0.0798)	(0.0579)	(0.0207)	(0.2131)

IV. DATA ANALYSIS

As an application of the proposed distribution, we analyze data on the entrance grade level and weight (unit: kg) of 1250 male students who entered KUT in 2023. The 1250 male students were divided into 10 groups according to the low to high order entrance grades. The number of individuals in each group is set to 125. The individual included in the ith (i=1, 2,⋯, 10) group is assigned a grade level i.

From the geometric distribution with parameter q=0.7, 50 samples are generated to find the individual frequencies according to the grade levels. The frequencies for more than 10 grade levels are all included in the 10 grade level. The frequencies obtained for grade levels 1 to 10 are shown in Table 7. Fortunately, in the generation run, the frequencies for more than 9 grade levels are all zero.

Table 7. Individual frequencies according to the grade levels obtained based on 50 samples generated from a geometric distribution with a parameter of q=0.7.

i (Grade level)	1	2	3	4	5	6	7	8	9	10	Sum
k_i (Frequency)	16	11	6	7	1	5	2	2	0	0	50

50 data of grade level and weight are constructed by randomly selecting k_i individuals from the ith group for i=1, 2,⋯, 10, which are presented in Table 8.

Table 8. 50 data of grade level (k) and weight (x) constructed by randomly selecting k_i individuals from the ith group for i = 1,…, 10.

No	k	x	No	k	x	No	k	x	No	k	x	No	k	x
1	1	66.4	11	1	68.9	21	2	57.7	31	3	79.3	41	5	66.1
2	1	63.7	12	1	68.8	22	2	81.4	32	3	61.2	42	6	62.5
3	1	71.5	13	1	71.6	23	2	69.9	33	3	64	43	6	65.5
4	1	63	14	1	65.9	24	2	63.5	34	4	64.6	44	6	70.3
5	1	66.4	15	1	71	25	2	75.9	35	4	65.4	45	6	70.2
6	1	64.9	16	1	64.7	26	2	64.5	36	4	66.2	46	6	67.7
7	1	64.5	17	2	71.4	27	2	69.9	37	4	70.2	47	7	63.6
8	1	67.8	18	2	70.8	28	3	60.8	38	4	73.2	48	7	68.8
9	1	65	19	2	63.8	29	3	64	39	4	59.4	49	8	62.6
10	1	61.7	20	2	72.5	30	3	68.7	40	4	66.6	50	8	65.8

The model for the data in Table 8 is assumed to be $NGF(\mu, \sigma^2, p; \alpha)$ and the parameters of the distribution are estimated. The moment estimates of the parameters are calculated as follows:

$$\tilde{\mu} = 67.0737,\ \tilde{\sigma} = 4.6089,\ \tilde{q} = 0.6644,\ \tilde{\alpha} = -0.3089$$

Also, the ML estimates are obtained as follows:

$$\hat{\mu} = 67.0868,\ \hat{\sigma} = 4.6108,\ \hat{q} = 0.664,\ \hat{\alpha} = -0.2982$$

According to Proposition 2.3, plug-in estimators of the correlation coefficients corresponding to the moment and the ML estimators are calculated, respectively, as follows:

$$\tilde{r} = -0.08535 \ , \ \ \hat{r} = -0.08239 \ .$$

Fig. 1 shows the conditional mathematical expectations (solid line) and the conditional upper and lower standard deviation limits (dashed lines) of weight X with respect to the grade level K based on moment estimates (a) and maximum likelihood estimates (b), respectively. Fig. 2 shows the conditional mathematical expectations (solid line) and the conditional upper and lower standard deviation limits (dashed lines) of grade level K with respect to the weight X based on moment estimates (a) and maximum likelihood estimates (b), respectively.

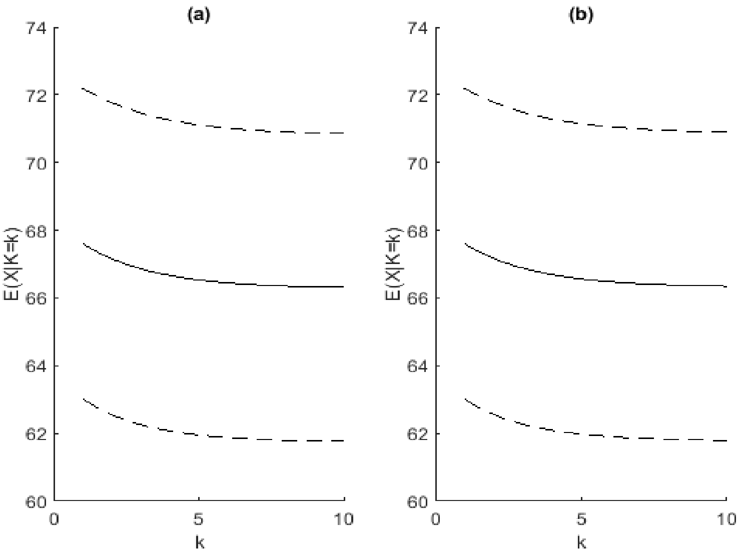

Fig 1. The conditional mathematical expectations (solid line) and the conditional upper and lower standard deviation limits (dashed lines) of weight X with respect to the grade level K based on moment estimates (a) and maximum likelihood estimates (b)

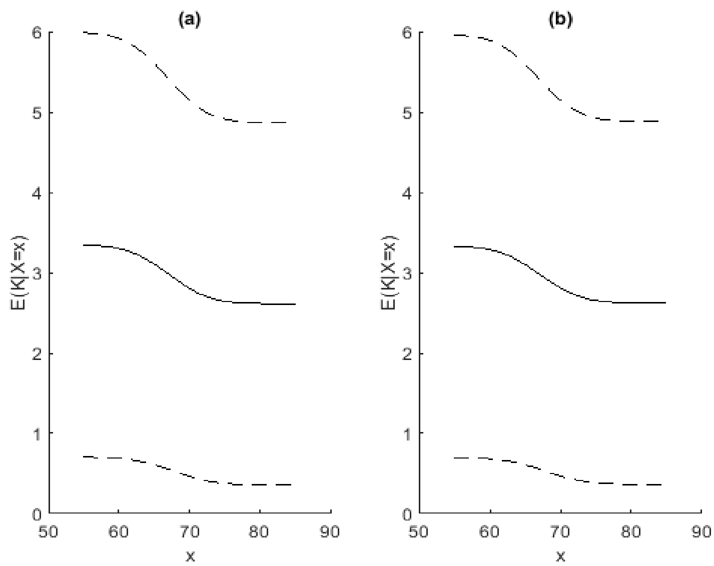

Fig.2. The conditional mathematical expectations (solid line) and the conditional upper and lower standard deviation limits (dashed lines) of grade level K with respect to the weight X based on moment estimates (a) and maximum likelihood estimates (b)

V. CONCLUSION

In this paper, the bivariate normal-geometric distribution with FGM copula has been proposed and its statistical properties have been investigated. Since the absolute correlation coefficient of the distribution is less than 0.2821, it is a model for continuous and discrete random variables with low correlation. The moment estimators for the parameters have been derived and the ML estimation performed by the Newton- Raphson method has been described, and the numerical results have been presented.

The distribution would be widely used as an efficient model for discrete-continuous data with low correlation. Replacing the FGM copula with other copula is expected to obtain a model with higher correlation. Finally, we believe that our approach could be extended to develop different multivariate continuous-discrete distributions.

ACKNOWLEDGEMENTS

The authors would like to thank Editor and anonymous referees for valuable comments and suggestions.

REFERENCES

[1] Abd Elaal M.K., Jarwan R.S. (2017). Inference of bivariate generalized exponential distribution based on copula functions. Appl Math Sci 11(24):1155–1186.

[2] Almetwally E.M., Muhammed H.Z., El-Sherpieny E.S.A. (2020). Bivariate Weibull distribution: properties and different methods of estimation. Ann Data Sci 7(1):163–193.

[3] Almetwally E.M., Muhammed H.Z. (2020). On a bivariate Fréchet distribution. J Stat Appl Probab 9(1):1– 21.

[4] Azzalini A., Capitanio A. (2014). The Skew-Normal and Related Families, Cambridge University Press.

[5] Eliwa M.S., El-Morshedy M. (2020). Bivariate odd Weibull-G family of distributions: properties, Bayesian and non-Bayesian estimation with bootstrap confidence intervals and application. J Taibah Univ Sci 14(1):331–345.

[6] El-Morshedy M., Alhussain Z.A., Atta D., Almetwally E.M., Eliwa M.S. (2020). Bivariate Burr X generator of distributions: properties and estimation methods with applications to complete and type-II censored samples. Mathematics 8(2):1–31.

[7] El-Sherpieny E.S., Almetwally E.M. (2019). Bivariate generalized Rayleigh distribution based on Clayton Copula. In: Proceedings of the annual conference on statistics (54rd), Computer Science and Operation Research, Faculty of Graduate Studies for Statistical Research, Cairo University, 1–19.

[8] Flores A.Q. (2009). Testing Copula functions as a method to derive bivariate Weibull distributions. APSA 2009 Toronto meeting paper. Available at SSRN: https://ssrn.com/abstract 1450064.

[9] Griffiths R.C., Milne R.K., Wood R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. Austral. J. Statist. 21, 238–255.

[10] Gumbel E.J. (1960). Bivariate exponential distributions. J Am Stat Assoc 55(292), 698–707.

[11] Jaworski P., Durante F., Hardle W., Rychlik T. (2010). Copula Theory and Its Applications, Springer-Verlag Berlin Heidelberg.

[12] Joe H. (2014) Dependence modeling with copulas. CRC Press, Cambridge.

[13] Kocherlakota S., Kocherlakota, K. (1992). Bivariate Discrete Distributions. Marcel Dekker, New York.

[14] Kozubowski T.J., Panorska A.K. (2005) A mixed bivariate distribution with exponential and geometric marginal, Journal of Statistical Planning and Inference 134, 501–520.

[15] Kundu D., Nekoukhou V. (2018). On bivariate discrete Weibull distribution, Communications in Statistics-Theory and Methods, V. 00, No. 0, 1–18.

[16] Lakshminarayana J., Pandit S., Rao K.S., (1999). On a bivariate Poisson distribution. Comm. Statist. Theory Methods 28, 267–276.

[17] Muhammed H.Z. (2016) Bivariate inverse Weibull distribution. J Stat Comput Simul 86(12), 2335–2345.

[18] Nekoukhou V., Kundu D. (2017). Bivariate discrete generalized exponential distribution. Statistics, 51, 1143–58.

[19] Nelsen R.B. (1987). Discrete bivariate distributions with given marginals and correlation. Comm. Statist. Simulation Comput. 16, 199–208.

[20] Nelsen R.B. (2007) An introduction to copulas, Springer.

[21] Peres M., Achcar J.A., Martinez E.Z. (2018) Bivariate modifed Weibull distribution derived from Farlie-Gumbel-Morgenstern copula: a simulation study. Electr J Appl Stat Anal 11(2), 463–488.

[22] Tyagi S. (2024). On bivariate Teissier model using Copula: dependence properties, and case studies, Int J Syst Assur Eng Manag (June 2024) 15(6), 2483–2499 https://doi.org/10.1007/s13198-024-02266-2.

[23] Verrill S.P., Evans J.W., Kretschmann D.E., Hatfield C.A. (2015). Asymptotically efficient estimation of a bivariate Gaussian–Weibull distribution and an introduction to the associated pseudo-truncatedWeibull. Commun Stat Theory Methods, 44(14):2957–2975.

[24] Piperigou V.E. (2009). Discrete distributions in the extended FGM family: The p.g.f. approach, Journal of Statistical Planning and Inference, 139, 3891 – 3899.