

A Review: Construction of Statistical Distributions

Kai-Tai Fang ^{1,2} , Yu-Xuan Lin ^{1,3}  and Yu-Hui Deng ^{1,3,*}

¹ Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science (IRADS), Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, Tangjiawan, Zhuhai 519087, China; ktfang@bnu.edu.cn (K.-T.F.); yuxuanlin@bnu.edu.cn (Y.-X.L.)

² The Key Lab of Random Complex Structures and Data Analysis, The Chinese Academy of Sciences, Beijing 100045, China

³ Faculty of Science and Technology, Beijing Normal-Hong Kong Baptist University, 2000 Jintong Road, Tangjiawan, Zhuhai 519087, China

* Correspondence: ivandeng@bnu.edu.cn; Tel.: +86-0756-362-0618

Abstract

Statistical modeling is fundamentally based on probability distributions, which can be discrete or continuous and univariate or multivariate. This review focuses on the methods used to construct these distributions, covering both traditional and newly developed approaches. We first examine classic distributions such as the normal, exponential, gamma, and beta for univariate data, and the multivariate normal, elliptical, and Dirichlet for multidimensional data. We then address how, in recent decades, the demand for more flexible modeling tools has led to the creation of complex meta-distributions built using copula theory.

Keywords: construction of statistical distribution; distribution family; entropy; function of random variables; mean square error; quasi-Monte Carlo; representative points; statistical distribution

1. Introduction

Statistical distributions form the foundation of probability theory, statistical inference, and predictive modeling. They provide a mathematical framework for describing the randomness, variability, and uncertainty inherent in data. Different types of data—univariate or multivariate, discrete or continuous—require different distributions, including mixture models for complex patterns.

Numerous books have offered comprehensive introductions to useful distributions. The seminal four-volume work by Johnson et al. [1–4] provides an exceptionally detailed encyclopedia of distributions and their interrelations, though its sheer scope can make it challenging for readers to locate specific information efficiently. Other key texts include Fang and Xu’s book [5] (in Chinese, which limits its accessibility), Fang et al. [6] focusing treatment on multivariate symmetric distributions, Johnson [7] reviewing statistical simulation for multivariate distributions, and the new work by Fang et al. [8] offering a comprehensive study on representative points of distributions and their applications.

The purpose of this review is to synthesize various methods for constructing statistical distributions. By presenting a unified overview of both traditional and modern techniques, we aim to equip readers with the knowledge to understand existing distributions and develop new ones to meet evolving data modeling needs.

This paper is structured as follows. Section 2 introduces construction methods for univariate distributions, encompassing both traditional and newer models. Section 3



Academic Editors: Nikolai Leonenko and Boris Ryabko

Received: 7 October 2025

Revised: 6 November 2025

Accepted: 20 November 2025

Published: 23 November 2025

Citation: Fang, K.-T.; Lin, Y.-X.; Deng, Y.-H. A Review: Construction of Statistical Distributions. *Entropy* **2025**, *27*, 1188. <https://doi.org/10.3390/e27121188>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

briefly reviews methods for constructing discrete distributions, including both univariate and multivariate frameworks. Section 4.4 then details construction techniques for multivariate statistical distributions, specifically the families of elliptical distributions and meta-distributions created via copula methods. The paper concludes with final remarks in Section 5.

2. Construction of Univariate Statistical Distributions

This section introduces several approaches for constructing univariate continuous distributions:

- (1) Distribution families, including location-scale families, the Pearson system, and exponential families;
- (2) Functions of random variables via stochastic representations;
- (3) Parameter expansion generalizes existing families by adding parameters. For example, normal $\rightarrow$ skew normal; exponential $\rightarrow$ Weibull.
- (4) Generalized families via transformation of the cdf. A powerful and prolific approach for generating new families of distributions is through the functional transformation of an existing baseline cdf, $F(x)$. These transformations introduce new parameters, thereby creating more flexible distributions capable of modeling complex data behaviors.
- (5) Combination methods, such as mixture distributions, compound distributions, convolutions, and copula-based techniques.

2.1. Distribution Families

A distribution family is defined by a specific structural property. Among many such families, the following three are particularly well established in the literature.

2.1.1. Location-Scale Distribution Family

A distribution $F(x; \mu, \sigma)$ belongs to the location-scale family if its probability density function (pdf) can be expressed as

$$f(x) = \frac{1}{\sigma} f_0\left(\frac{x - \mu}{\sigma}\right), \quad (1)$$

where μ is the location parameter, $\sigma > 0$ is the scale parameter, and f_0 is the standard density of the distribution $Z \sim F(x; 0, 1)$. This implies that for any $X \sim F(x; \mu, \sigma)$, the standardized variable $Z = \frac{(X - \mu)}{\sigma}$ follows the standard distribution $F(x; 0, 1)$.

A key property of this family is its invariance under affine transformations, implying that if $X \sim F(x; \mu, \sigma)$, then $a + bX \sim F(x; a + b\mu, |b|\sigma)$. The normal distribution is a classic example of the location-scale distribution. Other prominent examples include the Cauchy and logistic distributions.

- Cauchy distribution: Its pdf is given by

$$f(x; \theta, \lambda) = \frac{1}{\pi \lambda \left[1 + \left(\frac{x - \theta}{\lambda} \right)^2 \right]},$$

where θ is the location parameter and λ is the scale parameter. Note that the Cauchy distribution does not have a defined mean or variance.

- Logistic distribution: Named for its relationship to the logistic (sigmoid) function, the logistic distribution resembles the normal distribution in shape but possesses heavier

tails. It finds applications in logistic regression, growth models, and economics. Its cumulative distribution function (cdf) is given by

$$F(x; \alpha, \beta) = \frac{1}{1 + \exp(-(x - \alpha)/\beta)}, \quad x, \alpha \in \mathbb{R}, \quad \beta > 0.$$

There are more other distributions in this category. For further reading, see [9,10].

2.1.2. Pearson Distribution Family

The Pearson family is a system of continuous probability distributions, introduced by Karl Pearson [11], designed to model a wide range of empirical data characterized by properties such as skewness and kurtosis that deviate from the normal distribution. The system is defined by the Pearson differential equation:

$$\frac{df(x)}{dx} = f(x) \frac{x - \lambda}{ax^2 + bx + c}, \quad (2)$$

where $f(x)$ is a probability density function and the parameters a, b, c, λ determine the specific distribution type.

Roots of the denominator ($ax^2 + bx + c$) classify distributions into eight or more types listed in Table 1, where type 0 is the normal distribution; type I distribution is the beta distribution; type II distribution curve is a symmetrical U -shaped curve; type III distribution curve is the gamma distribution; type IV distribution curve is skewed heavy-tailed; type V is the shift inverse gamma distribution; type VI is an inverse beta distribution; type VII is the t -distribution. Some references consider more types, like types VIII and IX, which are special cases of the power function distribution, and type X, which is the exponential distribution. More details can be found in Chapter 12 of [1,12].

Table 1. Classification of Pearson distributions.

Type	Condition	Distribution	PDF Example
0	$a = 0, b = 0$	Normal distribution	$f(x) \propto e^{-(x-\mu)^2/2\sigma^2}$
I	$b^2 - 4ac > 0$	Beta distribution	$f(x) \propto (1+x)^{r_1}(1-x)^{r_2}, x \leq 1$
II	$b^2 - 4ac = 0,$ $\lambda = -b/2a$	Symmetric beta	$f(x) \propto (1-x^2)^r, x \leq 1$
III	$a = 0$	Gamma distribution	$f(x) \propto x^{k-1}e^{-x/\theta}, x > 0$
IV	$b^2 - 4ac < 0$	Skewed heavy-tailed	Complex form
V	$c = 0$	Inverse gamma	$f(x) \propto x^{k-1}e^{-\theta/x}, x > 0$
VI	$b^2 - 4ac > 0, \lambda \neq \text{root}$	Beta prime, F -distribution	$f(x) \propto x^{r_1}(1+x)^{-(r_1+r_2)}$
VII	$b = 0$	Student's t -distribution	$f(x) \propto (1 + x^2/\nu)^{-(\nu+1)/2}$

2.1.3. The Exponential Distribution Family

Consider it as a “family” in the biological sense, implying that though its members, such as the normal, exponential, gamma, beta, and Poisson distributions, may appear distinct characteristics, they share a common underlying mathematical structure. This shared “DNA” makes them exceptionally powerful and convenient to use in statistical modeling and machine learning.

Formally, the exponential family comprises distributions whose probability density (or mass) functions can be expressed in the following canonical form:

$$f(x; \theta) = h(x) \exp\{\theta^\top \mathbf{T}(x) - B(\theta)\}, \quad (3)$$

where $h(x)$ is a function as base measure that depends only on the data X , not the parameters $\theta = (\theta_1, \dots, \theta_k)^\top$. $\mathbf{T}(x) = (T_1(x), \dots, T_k(x))^\top$ is a sufficient statistic, and $B(\theta)$ is a log-partition function and a normalization constant that ensures that the function integrates/sums to 1. $B(\theta)$ also generates the moments (mean, variance) of the distribution.

This common structure leads to several powerful properties:

- (1) All members have a low-dimensional sufficient statistic $\mathbf{T}(x)$.
- (2) In Bayesian statistics, the exponential family always has a natural conjugate prior. This makes Bayesian updating mathematically neat and computationally tractable.
- (3) Distributions in the exponential family are the maximum entropy distributions given constraints on the expected values of the sufficient statistics. For example, the normal distribution has the maximum entropy for a given mean and variance.

Further details can be found in references such as [1,9]. Many additional families of univariate distributions have also been cataloged in [1].

2.2. Distributions Produced by a Function of Random Variables

When the distribution of a random variable X or a random vector $\mathbf{X} = (X_1, \dots, X_k)$ is known, applying a continuous function $Y = g(X)$ or $Y = h(\mathbf{X})$ in order to yield a new distribution is a prevalent distribution construction approach. In this context, the original distribution is termed the initial distribution, while the resulting distribution is called the target distribution. The inverse transformation method [13] is a widely used technique for statistical simulation, based on the following theorem.

Theorem 1. Let X have a continuous cdf $F(x)$ and a uniform random variable $U \sim U(0, 1)$. Then we have

- (a) $F^{-1}(U) \sim F(\cdot)$;
- (b) $F(X) \sim U(0, 1)$.

Following Theorem 1, we can leverage this connection between any continuous distribution and the uniform distribution to derive the distribution of a transformed random variable $Y = g(X)$. The key insight is to use the probability integral transform from part (b) of the theorem. We know that $F_X(X) \sim U(0, 1)$. Therefore, the cdf of the transformation $Y = g(X)$ can be found by relating the event $\{Y \leq y\}$ back to this uniform distribution. Specifically, the cdf of Y is

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y).$$

The solution to this probability depends on the nature of the function g . For monotonic and differentiable transformations, this approach leads directly to the well-known change of variables formula for the pdf of Y :

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|,$$

where $x = g^{-1}(y)$ is the inverse function. This formula can be derived by differentiating the cdf $F_Y(y)$, which itself is computed using the distribution of X that Theorem 1 helps us to understand and manipulate through the uniform distribution.

A systematic study and generalization of the inverse transformation method can be found in Li et al. [14]. Numerous distributions have been constructed using this approach. The following are some functions commonly employed in the literature for this purpose.

2.2.1. Transformation of Random Variables

Some transformation examples are presented later.

- Linear transformation $Y = a + bX$, for example, used in the location-scale distributions;
- Power transformation $Y = X^r$, for example, from the normal to the χ^2 distribution by $Y = X^2$;
- Exponential transformation $Y = e^X$, for example, from the normal to the lognormal distribution by $Y = e^X$;
- Inverse transformation $Y = 1/X$, for example, the inverse gamma distribution.
- Transformation $Y = g(X_1, \dots, X_m)$, for example, the F -distribution and beta distribution are produced by functions of two χ^2 distribution variables.

2.2.2. Order Statistics

These include fundamental objects in statistics, for example, the sample minimum, maximum, and median. They deal with the properties and applications of sorted random samples. Let $X_1, \dots, X_n$ be a random sample from a population with cdf F and pdf f . The order statistics, denoted by $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, represent the sorted values of the sample. The cdf of the k th order statistic, $X_{(k)}$, is given by

$$P(X_{(k)} \leq x) = \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}. \quad (4)$$

The density functions for specific order statistics are provided in Table 2. Order statistics have wide-ranging applications, including robust statistics, sample range and interquartile range, quantiles and Q–Q plots, reliability engineering, and non-parametric statistical tests (e.g., the Wilcoxon rank-sum test). A comprehensive introduction to order statistics can be found in [15].

Table 2. Distributions of the order statistics.

Distribution	Density $f(x)$
The general k -th order statistic ¹	$\binom{n}{k-1} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x)$
Minimum ($X_{(1)}$)	$n[1 - F(x)]^{n-1} f(x)$
Maximum ($X_{(n)}$)	$n[F(x)]^{n-1} f(x)$

¹ Note: The pdf for the median exists in a closed form only when n is odd ($n = 2m + 1$), given by $p_{\text{median}}(x) = \frac{n!}{(m!)^2} [F(x)]^m [1 - F(x)]^m f(x)$. For even n , the median is defined as $\frac{1}{2}(X_{(n/2)} + X_{(n/2+1)})$ and its distribution is derived from the joint distribution of these two order statistics.

2.2.3. Extreme Value Theory (EVT)

EVT is a statistical framework for assessing the probabilities of extreme events. The fundamental result in EVT is the extreme value theorem, stating that the distribution of properly normalized maxima (or minima) converges to one of three types of distributions, regardless of the underlying distribution of the original data. These three types are collectively known as the generalized extreme value (GEV) distribution. The cdf for the GEV is given by

$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \text{ if } 1 + \xi \frac{x - \mu}{\sigma} > 0, \quad (5)$$

where parameters are μ for location, $\sigma > 0$ for scale, and ξ for shape. The three special cases of the GEV distribution are

- Type I (Gumbel distribution) $Gu(\mu, \sigma)$ light-tailed, with $\xi = 0$. Its pdf is

$$\frac{1}{\sigma} \exp \left\{ -\frac{x - \mu}{\sigma} - \exp \left\{ -\frac{x - \mu}{\sigma} \right\} \right\}, \sigma > 0.$$

- Type II (Fréchet distribution) heavy-tailed, with $\xi > 0$.
- Type III (Weibull distribution) short-tailed and bounded with $\xi < 0$.

The extreme value distribution has been used in finance, environmental science, engineering, and reliability testing. For a rigorous treatment of the GEV distribution and extreme value theory, see [16].

2.3. Generalized Families Via Transformation of the cdf

Given a baseline cdf $F(x)$, the exponentiated (or power-transformed) family is defined by

$$G(x) = [F(x)]^\alpha, \quad \alpha > 0.$$

This construction, detailed in [17], generates a vast range of new distributions. If $X_1, X_2, \dots, X_n$ is a random sample from $F(x)$, then $G(x)$ is the cdf of the sample maximum, $\max(X_1, \dots, X_n)$, when $\alpha = n$ is an integer. Notable members include the exponentiated Weibull and exponentiated exponential distributions.

A dual transformation yields the following complementary exponentiated family:

$$G(x) = 1 - [1 - F(x)]^\beta, \quad \beta > 0.$$

This is the cdf of the sample minimum, $\min(X_1, \dots, X_n)$, when $\beta = n$. It is also known as the proportional odds model in survival analysis. Together, the exponentiated and complementary exponentiated families form the foundation of the broader class of maximum–minimum stable distributions.

Further generalizations have been proposed to introduce additional flexibility and skewness. The Alpha Power Family (or T-X family), defined by

$$G(x) = \frac{\alpha^{F(x)} - 1}{\alpha - 1}, \quad \alpha > 0, \alpha \neq 1,$$

effectively introduces skewness to the baseline distribution [18]. A highly flexible generalization is the Kumaraswamy-G family [19], given by

$$G(x) = 1 - [1 - F(x)^a]^b, \quad a > 0, b > 0.$$

This family, which contains the exponentiated family as a special case ($b = 1$), enjoys the advantage of having a quantile function in a simple closed form.

These methods, along with other systematic frameworks like the T-X family [20], demonstrate that simple transformations of a baseline cdf can serve as an engine for generating numerous new and useful distributions, greatly expanding the toolbox for statistical modeling.

2.4. Combination Methods

There are several approaches to constructing a distribution through combining the existing distributions.

2.4.1. Types of Mixtures

- Finite mixture $f(x) = \sum_{i=1}^k w_i f_i(x)$, which has various applications in statistical inference with flexibility in density fitting [21], for example, a mixture of two normals [22] or two Weibull distributions [23].
- Continuous mixture $f(x) = \int f(x; \theta) g(\theta) d\theta$. One example is the scale mixture. A primary method to create such continuous mixtures is through compounding, which we detail in the next section.

The mixing distribution $g(\theta)$ in an infinite mixture is not required to be continuous. When $g(\theta)$ is a discrete distribution, the resulting mixture, often called a countable mixture or discrete mixture, integrates over a countable set of parameters. This approach is technically distinct from the continuous mixture defined above but remains a powerful tool for deriving flexible distributions. Prominent examples of distributions derived from a discrete mixing mechanism include the following:

- The negative binomial distribution, which can be derived as a Poisson-gamma mixture where the Poisson rate parameter is governed by a gamma distribution. While the gamma is continuous, a discrete analogue exists (e.g., a Poisson mixture with a discrete log-gamma distribution would yield a similar over-dispersed count distribution).
- The Student's t -distribution, which is a scale mixture of normal distributions where the precision (inverse variance) parameter follows a gamma distribution.
- The beta-binomial distribution, which arises when the success probability p of a binomial distribution follows a beta distribution.
- The compound Poisson distribution and related models, where the intensity or other parameters are driven by a latent discrete process.

2.4.2. Compounding

It is a fundamental concept in probability theory that a parameter of one distribution is itself a random variable following another distribution (see [1]). This approach generates a more flexible distribution that incorporates uncertainty in the parameter. It is important to note that the marginal distribution resulting from compounding is mathematically equivalent to a continuous mixture distribution, as defined in Section 2.4.1. The compounding framework provides a hierarchical data-generation perspective for the same model. Formally, if $X|\theta \sim F_1(\theta)$ and $\theta \sim F_2(\lambda)$, then the distribution of X is the marginal distribution of a compound distribution. Examples are the beta-binomial, gamma-Poisson (negative binomial), normal-normal, Dirichlet-multinomial, compound Poisson process models. The normal-normal model is as follows:

- First stage (conditional distribution): $X|\mu \sim \mathcal{N}(\mu, \sigma^2)$. The observation X is normally distributed with mean μ and a known variance σ^2 ;
- Second stage (parameter distribution): $\mu \sim \mathcal{N}(\mu_0, \tau^2)$. The mean μ itself is assumed to be normally distributed with prior mean μ_0 and variance τ^2 ;
- Compound distribution: The unconditional distribution of $X \sim \mathcal{N}(\mu_0, \sigma^2 + \tau^2)$.

2.4.3. Convolution

The random variable $Z = X + Y$ is termed the convolution of X and Y , and it has a density function given by

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z-x) dx,$$

where $f_{X,Y}(\cdot, \cdot)$ is the joint density of (X, Y) . If X and Y are independent, this simplifies to $f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx$.

A distribution of significant practical importance arises when X follows a normal distribution $N(\mu, \sigma^2)$ and Y follows an exponential distribution $\text{Exp}(\lambda)$, with X and Y being independent. This specific convolution forms the cornerstone of the normal–exponential stochastic frontier model [24], widely used in econometrics to measure the efficiency of firms and other decision-making units. In this context, the normal random variable represents the random noise (e.g., measurement error, luck), while the exponential random variable captures one-sided inefficiency.

Then the density function of $Z = X + Y$ is given by

$$f_Z(z; \mu, \sigma, \lambda) = \frac{\lambda}{2} e^{\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2z)} \cdot \text{erfc}\left(\frac{\mu + \lambda\sigma^2 - z}{\sqrt{2}\sigma}\right),$$

where $\text{erfc}(z)$ represents the complementary error function defined as

$$\text{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt.$$

2.5. Some Key Univariate Distributions and Related Distributions

The first continuous statistical distribution to be formally studied was the uniform distribution, dating back to Huygens [25]. However, the most historically significant early continuous distribution is the normal distribution, also referred to as the Gaussian distribution. Although the uniform distribution provided the simplest continuous model, it was the normal distribution that first underwent extensive analytical study. The development of the normal distribution emerged later through the work of de Moivre [26] and Gauss [27] during the 18th and 19th centuries. Its fundamental role in error theory and the central limit theorem established it as the cornerstone of continuous probability theory.

2.5.1. Distributions Related to the Normal Distribution

The probability density function of a normal distribution $X \sim N(\mu, \sigma^2)$ is given by

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, \quad (6)$$

where parameters μ and σ^2 are the mean and variance of X , respectively. When $\mu = 0$ and $\sigma^2 = 1$, the distribution is referred to as the standard normal distribution, denoted by $Z \sim N(0, 1)$. Several distributions derived from the normal distribution are listed in Table 3 and can be constructed as follows:

- (1) Lognormal distribution: $X \sim LN(\mu, \sigma^2)$ if $\ln(X) \sim N(\mu, \sigma^2)$;
- (2) Skew-normal distribution: proposed by Azzalini [28] with the pdf

$$f(z; \alpha) = 2\phi(z)\Phi(\alpha z), z \in \mathbb{R}, \alpha \in \mathbb{R}, \quad (7)$$

where $\phi(z)$ is the standard normal pdf, $\Phi(\alpha z)$ is the standard normal cdf evaluated at αz and α is a shape parameter. For more general cases, refer to Table 3. Skew-normal distribution is one of the skew-elliptical distributions, and there are more such distributions including skew- t , skew-Cauchy, skew-logistic, and skew-Laplace distributions; see [29,30]. There is more discussion about this topic in Section 4.6.

- (3) Chi-square distribution: If $X_1, \dots, X_\nu$ are independent and identically distributed (*i.i.d.*) random variables of $N(0, 1)$, then $X = X_1^2 + \dots + X_\nu^2 \sim \chi^2(\nu)$, the χ^2 -distribution with degrees of freedom ν ;
- (4) Chi distribution: If $X_1, \dots, X_\nu$ are *i.i.d.* $N(0, 1)$, then $X = \sqrt{X_1^2 + \dots + X_\nu^2} \sim \chi(\nu)$ the χ -distribution with degrees of freedom ν ;

- (5) Beta distribution (special case): If $Y \sim \chi^2(2m)$ and $Z \sim \chi^2(2n)$ are independent, then $X = \frac{Y}{Y+Z} \sim Be(m, n)$, referred to Table 3;
- (6) Beta prime distribution: If $X \sim Be(\alpha, \beta)$, then $\frac{X}{1-X} \sim BP(\alpha, \beta)$;
- (7) Arcsine distribution: It is defined as $Be(\frac{1}{2}, \frac{1}{2})$;
- (8) F -distribution: If $Y \sim \chi^2(m)$ and $Z \sim \chi^2(n)$ are independent, then $X = \frac{nY}{mZ} \sim F(m, n)$;
- (9) t -distribution: If $Y \sim \chi^2(\nu)$ and $Z \sim N(0, 1)$ are independent, then $T = \sqrt{\nu} \frac{Z}{\sqrt{Y}} \sim t_\nu$.

Table 3. Some continuous distributions constructed by or closely related to $N(\mu, \sigma^2)$.

Distribution	Notation	Density $f(x)$
Normal	$N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, x, \mu \in \mathbb{R}, \sigma > 0$
Lognormal	$LN(\mu, \sigma^2)$	$\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), x > 0, \mu \in \mathbb{R}, \sigma > 0$
Skew-normal	$SN(\mu, \sigma, \alpha)$	$f(x; \mu, \sigma, \alpha) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\alpha \left(\frac{x-\mu}{\sigma}\right)\right), \sigma > 0, x, \mu, \alpha \in \mathbb{R}$
$\chi^2(\nu)$	$\chi^2(\nu)$	$[2^{\nu/2} \Gamma(\frac{\nu}{2})]^{-1} x^{\nu/2-1} e^{-x/2}, x > 0, \nu > 0$
$\chi(\nu)$	χ_ν	$[2^{\nu/2-1} \Gamma(\frac{\nu}{2})]^{-1} x^{\nu-1} e^{-x^2/2}, x > 0, \nu > 0$
Beta	$Be(a, b)$	$\frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, 0 \leq x \leq 1, a, b > 0$
Beta prime	$BP(\alpha, \beta)$	$\frac{1}{B(\alpha, \beta)} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}}, x > 0, \alpha, \beta > 0$
Arcsine	arcsine	$\frac{1}{\pi \sqrt{x(1-x)}}, 0 < x < 1$
F	$F(m, n)$	$\frac{1}{B(\frac{m}{2}, \frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, x > 0, m, n > 0$
Student- t	t_ν	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \nu > 0, x \in \mathbb{R}$

Note: The title “constructed by” implies a direct derivation from the normal distribution. For distributions like χ^2 , χ , F , and t , this direct construction (e.g., as a sum of squares of independent standard normal variables) is most straightforward when their shape/degrees of freedom parameters (ν, m, n) are positive integers. The density functions listed here represent the analytic continuations of these distributions to positive real-valued parameters, which are their standard definitions. The beta and arcsine distributions are included due to their relationship with normal random variables via the transformation of ratios or functions thereof.

2.5.2. Distributions Related to the Exponential Distribution

The exponential distribution, defined on the positive real number space, was extensively studied in the early 19th century due to its valuable properties, such as memorylessness and its connection to the Poisson process for modeling rare events (see [1]). Several distributions derived from the exponential distribution are listed in Table 4.

- (1) Laplace distribution: Also known as the double exponential distribution, the Laplace distribution extends the exponential distribution symmetrically to the entire real number space, $\mathbb{R}$.
- (2) Weibull distribution: The Weibull distribution is formed by introducing location δ and sharp α parameters to the density of the exponential density. It constitutes a flexible family of distributions widely used for modeling time-to-failure, lifespan, and other positive-valued measurements.
- (3) Rayleigh distribution: Proposed by Rayleigh [31] for modeling wave amplitudes, the Rayleigh distribution is a special case of the Weibull distribution with $Wei(2, \beta, 0)$. It also arises from the normal distribution, $N(0, \sigma^2)$, by $X = \sqrt{Y^2 + Z^2} \sim Ray(\sigma)$, where Y and Z are *i.i.d.* $N(0, \sigma^2)$ and $\beta = 2\sigma^2$.

- (4) Topp–Leone exponential distribution (TLED): Introduced by [32], the TLED adds a shape parameter α to the exponential distribution, yielding the cumulative distribution function $F(x) = (1 - e^{-2x/\beta})^\alpha$. Its density is provided in Table 4. The TLED generalizes certain distributions. For instance, TLED includes the Burr type as $TLED(2, \beta)$ and Frèchet distribution as $TLED(1/2, \beta)$.

Table 4. Some continuous distributions constructed by the exponential distribution.

Distribution	Notation	Density $p(x)$
Exponential	$Exp(\beta)$	$\frac{1}{\beta}e^{-x/\beta}, x \geq 0, \beta > 0$
Laplace	$Laplace(\mu, \beta)$	$\frac{1}{2\beta}e^{-\frac{ x-\mu }{\beta}}, x \in \mathbb{R}, \mu \in \mathbb{R}, \beta > 0$
Weibull	$Wei(\alpha, \beta, \delta)$	$\frac{\alpha}{\beta} \left(\frac{x-\delta}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x-\delta}{\beta}\right)^\alpha}, x \geq \delta, \delta \in \mathbb{R}, \alpha, \beta > 0$
Rayleigh	$Ray(\beta)$ or $Ray(\sigma)$	$\frac{2x}{\beta}e^{-x^2/\beta}, x \geq 0, \beta = 2\sigma^2$
TLED	$TLED(\alpha, \beta)$	$\frac{2\alpha}{\beta}e^{-2x/\beta}(1 - e^{-2x/\beta})^{\alpha-1}, x > 0, \alpha, \beta > 0$

It is worth mentioning that the exponential distribution and the Poisson distribution share a fundamental symbiotic relationship within the context of a Poisson process. If the number of events occurring in a fixed time interval follows a Poisson distribution with rate parameter λ , then the waiting time between consecutive events follows an exponential distribution with the same rate parameter $\lambda = \frac{1}{\beta}$. Thus, the former models event counts, and the latter models inter-arrival times.

2.5.3. Distributions Related to the Gamma Distribution

The gamma distribution, $Gamma(\alpha, \beta)$, was formally defined in the 20th century and can be considered as an extension of the exponential distribution. It retains several desirable properties of the exponential distribution, such as additivity: the sum of independently and identically distributed variables $X_i \sim Gamma(\alpha_i, \beta)$ for $i = 1, \dots, m$ is distributed as $Gamma(\sum_{i=1}^m \alpha_i, \beta)$. The gamma distribution is foundational, as the chi-square distribution χ^2_ν is a special case, $Gamma(\nu/2, 1/2)$, and the beta distribution $Be(a, b)$ can be defined by $\frac{Y}{Y+Z}$, where $Y \sim Gamma(a, 1)$ and $Z \sim Gamma(b, 1)$ are independent. Table 5 lists several important extensions of the gamma distribution.

- (1) Location-shift gamma: Adding a location parameter δ forms the three-parameter distribution $Gamma(\alpha, \beta, \delta)$.
- (2) Generalized gamma: This distribution exists in several forms. The version shown with shape a , scale d , and power p parameters was introduced by Stacy [33]. It includes several common distributions as special cases: the standard gamma ($p = 1$), the Weibull ($a = 1$), the Rayleigh ($a = p = 2$), and the exponential ($a = p = 1$).
- (3) Inverse gamma: Defined as $Y = 1/X$ for $X \sim Gamma(\alpha, \beta)$, this two-parameter distribution is positive-skewed. It is primarily used in Bayesian statistics as the conjugate prior for the variance of a normal distribution.
- (4) Beta prime distribution: The ratio X/Y of two independent variables, $X \sim Gamma(\alpha, \theta)$ and $Y \sim Gamma(\beta, \theta)$, follows a beta prime distribution $BP(\alpha, \beta)$.
- (5) General beta prime distribution: This is a further extension of the beta prime distribution, incorporating two additional parameters.

Table 5. Some continuous distributions constructed by the gamma distribution.

Distribution	Notation	Density $f(x)$
Gamma	$\text{Gamma}(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0, \alpha, \beta > 0$
Location-shift gamma	$\text{Gamma}(\alpha, \beta, \delta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} (x - \delta)^{\alpha-1} e^{-\beta(x-\delta)}, x \geq \delta, \delta \in \mathbb{R}, \alpha, \beta > 0$
Generalized gamma	$G - \text{gamma}(a, d, p)$	$\frac{p}{d^a \Gamma(a/p)} x^{a-1} e^{-(x/d)^p}, x > 0, a, d, p > 0$
Inverse gamma	$\text{Inv} - \text{Gamma}(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x}, x > 0, \alpha, \beta > 0$
Beta prime	$BP(\alpha, \beta)$	$\frac{1}{B(\alpha, \beta)} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}}, x > 0, \alpha, \beta > 0$
General beta prime	$GBP(\alpha, \beta, p, q)$	$\frac{ \alpha (1+(x/\beta)^\alpha)^{-p-q}(x/\beta)^{p\alpha-1}}{\beta B(p, q)}, \alpha \in \mathbb{R}, \alpha \neq 0, p, q, \beta > 0, x > 0.$

2.6. The Generalized Hyperbolic Distribution Family

Schrödinger [34] described the probability distribution of the first passage time in Brownian motion. Thirty years later, Tweedie [35] gave the inverse relationship between the cumulant generating function of the first passage time distribution and that of the normal distribution, and named it the inverse Gaussian (IG) distribution. This new distribution was proposed in the top journal *Nature*, which shows its importance. The IG distribution, denoted as $IG(\mu, \lambda)$, is defined for $x > 0, \mu > 0$, and $\lambda > 0$ by the probability density function:

$$f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\frac{\lambda(x - \mu)^2}{2\mu^2 x}\right\}. \quad (8)$$

The IG distribution models the first passage time of a Brownian motion, making it highly useful for analyzing lifetime data and the frequency of event occurrences in various fields. For a comprehensive reference, see [36]. The IG distribution has many extensions, among which the generalized hyperbolic distribution (GH) is a more general one.

The GH distribution is a continuous probability distribution that represents a highly flexible family of models for describing data with skewness and heavy tails. Its key strength lies in its ability to nest several other important distributions as special or limiting cases, providing a unified framework for statistical modeling. It was introduced by [37] to model the grain size distributions of wind-blown sand. However, its remarkable fit to financial data catapulted it to prominence in financial econometrics and risk management (see [38]).

A random variable X is said to follow a GH distribution $GH(\lambda, \alpha, \beta, \delta, \mu)$ if its pdf is

$$f(x) = C(\lambda, \alpha, \beta, \delta) (\delta^2 + (x - \mu)^2)^{(\lambda-0.5)/2} K_{\lambda-0.5} \left(\alpha \sqrt{\delta^2 + (x - \mu)^2} \right) \exp(\beta(x - \mu)), \quad (9)$$

where $x \in \mathbb{R}$ and

$$C(\lambda, \alpha, \beta, \delta) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi} \alpha^{\lambda-0.5} \delta^\lambda K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})}, \alpha > 0, |\beta| < \alpha, \delta > 0, \mu, \lambda \in \mathbb{R}.$$

The parameter λ influences the tail behavior and the subfamily classification; α controls the heaviness of the tails and the steepness; β controls the asymmetry; δ is the scale parameter

related to the dispersion; and μ is the location parameter. $K_\lambda(\cdot)$ is the modified Bessel function of the second kind of order λ defined by

$$K_\lambda(x) = \frac{1}{2} \int_0^\infty t^{\lambda-1} \exp\left(-\frac{x}{2}\left(t + \frac{1}{t}\right)\right) dt.$$

The GH distribution involves many distributions as its special cases: hyperbolic distribution ($\lambda = 1$), normal inverse Gaussian distribution ($\lambda = -0.5$), variance gamma distribution ($\lambda > 0, \delta = 0$), etc. A normal variance-mean mixture distribution, termed the normal inverse Gaussian distribution proposed by [39], is used to construct stochastic processes that appear of interest for statistical modeling purposes, particularly in turbulence and finance. Its pdf is defined as

$$f(x; \alpha, \beta, \delta, \mu) = C(\alpha, \beta, \delta) q\left(\frac{x - \mu}{\delta}\right)^{-1} K_1\left(\alpha \delta q\left(\frac{x - \mu}{\delta}\right)\right) \exp(\beta(x - \mu)), \quad x \in \mathbb{R},$$

where

$$C(\alpha, \beta, \delta) = \frac{\alpha}{\pi} \exp\left(\delta \sqrt{\alpha^2 - \beta^2}\right), \quad q(t) = \sqrt{1 + t^2},$$

and K_1 is the modified Bessel function of the second kind of order 1. The parameters have the following role: $\mu \in \mathbb{R}$ (location) and $\delta > 0$ (scale); $\alpha > 0$ (tail heaviness) controls the steepness of the distribution. A larger α leads to lighter tails and $\alpha > |\beta|$, where β (asymmetry/skewness) controls the skewness. $\beta = 0$ gives a symmetric distribution, $\beta > 0$ implies positive skewness, and $\beta < 0$ implies negative skewness.

The GH distribution is a normal variance-mean mixture of the generalized inverse Gaussian (GIG) distributions. This construction is the source of its flexibility and mathematical tractability. The GIG distribution is a continuous probability distribution that generalizes several important distributions and plays a critical role as a mixing distribution in normal variance-mean mixtures. The GIG distribution denoted by $GIG(\lambda, \chi, \psi)$ is defined by the given pdf

$$f(x; \lambda, \chi, \psi) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right)\right), \quad x > 0, \chi, \psi \geq 0, \quad (10)$$

where parameter λ determines the shape of the distribution and is deeply connected to the index of the Bessel function. χ is the scale parameter often related to the “left-hand” behavior, with the constraint: $\lambda > 0$ whenever $\chi = 0$. ψ is another scale parameter that is often related to the “right-hand” tail with the constraint: $\lambda < 0$ whenever $\psi = 0$, and $\psi = \chi = 0$ is not allowed. The GIG involves inverse Gaussian distribution ($\lambda = -0.5$), reciprocal inverse Gaussian distribution ($\lambda = 0.5$), and gamma distribution ($\chi = 0, \lambda > 0$).

A random variable Y follows a GH distribution $GH(\lambda, \alpha, \beta, \delta, \mu)$ if it has the following stochastic representation:

$$Y = \mu + \beta X + \sigma \sqrt{X} Z,$$

where $Z \sim N(0, 1)$, and $X \sim GIG(\lambda, \chi, \psi)$ is independent of Z . The parameters of the resulting $GH(\lambda, \alpha, \beta, \delta, \mu)$ distribution are identified as follows: $\alpha = \sqrt{\psi\sigma^2 + \beta^2}$, $\delta = \sqrt{\chi}$. By choosing different GIG subfamilies as the mixer, we obtain the different GH subfamilies:

- (1) Mixing with the inverse Gaussian ($\lambda = -0.5$) distribution results in the normal inverse Gaussian distribution.

- (2) Mixing with the gamma ($\chi = 0, \lambda > 0$) distribution results in the variance gamma distribution.
- (3) Mixing with the GIG ($\lambda = 1$) distribution results in the hyperbolic distribution.

2.7. Beta-Generated Distributions

A significant research area is the creation of flexible univariate distribution classes that extend classical models. The beta-generated (BG) distributions by Arnold et al. [40,41] achieve this by merging the beta distribution with a chosen baseline distribution.

Definition 1. For a baseline cdf $F(x)$ with pdf $f(x)$ on a domain D , the beta-generated (BG) distributions have the following pdf:

$$g_F(x; a, b) = \frac{1}{B(a, b)} F(x)^{a-1} [1 - F(x)]^{b-1} f(x), \quad x \in D, \quad a > 0, b > 0. \quad (11)$$

We denote this by $X \sim BG(a, b; F)$. The corresponding cdf is derived via the substitution $y = F(t)$:

$$G_F(x; a, b) = \frac{1}{B(a, b)} \int_0^{F(x)} y^{a-1} (1 - y)^{b-1} dy = I_{F(x)}(a, b),$$

which is the incomplete beta function ratio.

A key property of the $BG(a, b; F)$ distribution is its stochastic representation:

$$X \stackrel{d}{=} F^{-1}(B), \quad \text{for } B \sim Be(a, b). \quad (12)$$

This representation provides a foundation for applying the generalized inverse transformation method in statistical simulation using representative points. It offers a direct method for simulating from the density (11) and for constructing multivariate extensions. The moments of X are consequently given by

$$E(X^r) = E([F^{-1}(B)]^r), \quad r > 0. \quad (13)$$

The beta-generated family encompasses a wide range of distributions depending on the choice of the baseline F . To demonstrate the flexibility and utility of this class, we select two specific examples. For the generalization of this generating mechanism, see [42].

2.7.1. The Beta-Normal Distribution

If $F(x)$ is the cdf of $N(\mu, \sigma^2)$, the corresponding BG distribution is the beta-normal distribution, denoted as $X \sim BN(a, b; \mu, \sigma)$. As studied by Eugene et al. [43], its cdf is

$$G(x) = \frac{1}{B(a, b)} \int_0^{\Phi(\frac{x-\mu}{\sigma})} t^{a-1} (1 - t)^{b-1} dt, \quad (14)$$

where $\Phi(\cdot)$ is the standard normal cdf. For the standard case $BN(a, b; 0, 1)$, the pdf is

$$g(x) = \frac{1}{B(a, b)} \Phi(x)^{a-1} [1 - \Phi(x)]^{b-1} \phi(x), \quad x \in \mathbb{R}. \quad (15)$$

Its moments are given by the formula in (13) (Gupta and Nadarajah [44]).

This four-parameter distribution defines location, scale, and shape, and can be unimodal or bimodal. Its fitting performance has been compared to that of normal mixture models in simulation studies. Note that the beta distribution is a special case of the Dirichlet distribution (cf. Section 4.4.2).

2.7.2. The Beta-Weibull Distribution

The beta-Weibull distribution, denoted by $BW(a, b; c, \gamma)$, has the following pdf:

$$f_{BW}(x) = \frac{1}{B(a, b)} \frac{c}{\gamma} \left(\frac{x}{\gamma}\right)^{c-1} \left[1 - \exp\left\{-\left(\frac{x}{\gamma}\right)^c\right\}\right]^{a-1} \exp\left\{-b\left(\frac{x}{\gamma}\right)^c\right\}, \quad x > 0. \quad (16)$$

A system study on this distribution refer to Famege et al. [45] and Mdziniso [46]. The beta-Weibull distribution includes many useful distributions, including the Weibull distribution ($a = b = 1$), beta-exponential distribution ($c = 1$), Rayleigh distribution ($a = b = 1, c = 2$), two-parameter Burr-type distribution ($b = 1, c = 2$), etc.

The beta-Weibull distribution has many good properties.

- (1) The limit of beta-Weibull density (16) is

$$\lim_{x \rightarrow \infty} f_{BW}(x) = \begin{cases} \infty, & ac < 1, \\ \frac{c\Gamma(a+b)}{\gamma\Gamma(a)\Gamma(b)}, & ac = 1, \\ 0, & ac > 1. \end{cases}$$

- (2) Let $Y \sim Be(a, b)$, then the random variable $X = \gamma[-\log(1 - Y)]^{1/c}$ follows $BW(a, b, c, \gamma)$, where $c, \gamma > 0$. It provides a way to generate a random sample from the BW distribution.
- (3) Let Y follow a beta-exponential distribution with parameters (a, b, θ) , then random variable $X = \theta(Y/\theta)^{1/c}$ follows a beta-Weibull distribution.
- (4) The beta-Weibull distribution is unimodal. The mode is at $x_0 = 1$ whenever $ac < 1$ or $ac = 1$ and $b \geq (c - 1)/2c$. For other cases, x_0 is the solution of the equation

$$bc - (c - 1)(x_0/\gamma)^{-c} = c(a - 1)[\exp(x_0/\gamma)^c]^{-1}.$$

- (5) The r th moments of $BW(a, b, c, \gamma)$ is given by

$$E\left[\left(\frac{X}{\gamma}\right)^r\right] = \frac{1}{B(a, b)} \int_0^\infty t^{r/c} (1 - e^{-t})^{a-1} e^{-bt} dt,$$

by the transformation $t = (x/\gamma)^c$.

3. Construction of Univariate Discrete Distributions and an Approximation

A discrete distribution is characterized by its support points and their probabilities:

$$\begin{array}{c|cccc} X & x_1 & x_2 & \cdots & x_k \\ \hline p_i & p_1 & p_2 & \cdots & p_k \end{array} \quad (17)$$

where the random variable X has support $x_1, \dots, x_k$ and $P(X = x_i) = p_i, i = 1, \dots, k$.

3.1. Classical Univariate Discrete Distributions

Standard discrete distribution model outcomes from fundamental random processes/trials:

1. Bernoulli(p)/Binomial(n, p): The Bernoulli distribution, Bernoulli(p), models a single trial, and Binomial(n, p) counts successes in n independent trials with sample success probability p .
2. Geometric(p): It models the number of Bernoulli trials needed to obtain the first success, representing the waiting time for a success. It is a special case of the negative binomial (with $r = 1$).

3. Negative Binomial(r, p): As a generalization of the geometric distribution, it models the number of Bernoulli trials needed to achieve r successes. The negative binomial distribution presented here in its classic form (modeling the number of trials) is equivalent to the parameterization modeling the number of failures before the r -th success (with support $y = 0, 1, 2, \dots$), which is prevalent in many modern software packages and generalized linear models.
4. Poisson(λ): It models the number of events occurring in a fixed interval of time or space, given a constant average rate (λ) and independence between events. It can be derived as a limiting case of the binomial distribution when $n \rightarrow \infty$ and $p \rightarrow 0$ such that $np \rightarrow \lambda$.
5. Hypergeometric(s, n, N, M): It represents the number s of successes in n draws without replacement from a population with N size with M success items, where $\max(0, n - (N - M)) \leq s \leq \min(M, n)$. This differs from the binomial distribution, where trials are independent (i.e., with replacement).

See Johnson and Kotz [47] for an extensive catalog. A guide to selecting common discrete probability distributions is illustrated in Figure 1, with probability mass functions summarized in Table 6.

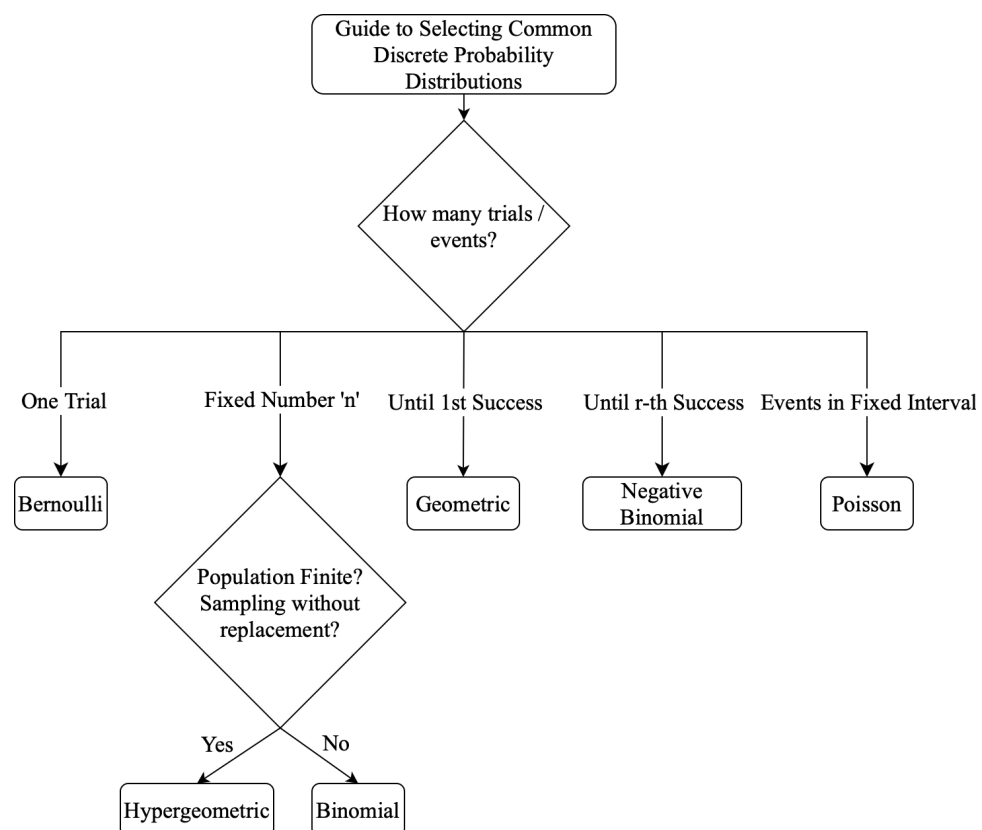


Figure 1. A guide to selecting common discrete probability distributions.

Consider a random variable X following the discrete distribution defined in (17) and a continuous function $g(\cdot)$. The transformed variable $Y = g(X)$ follows a distribution with the following structure:

$$\begin{array}{c|cccc} Y & g(x_1) & g(x_2) & \cdots & g(x_k) \\ \hline p_i & p_1 & p_2 & \cdots & p_k \end{array}. \quad (18)$$

In other words, the law of Y is determined by mapping the support points of X via g while preserving the associated probabilities. This elegant property facilitates the analysis

of transformed discrete variables. Li et al. [14] utilized this property to compute nearly representative points for the distribution of $g(X)$.

Table 6. Famous discrete distributions.

Distribution	Support Points	Probability Mass Function $p_i(x)$
Bernoulli(p)	0, 1	$p_0 = 1 - p, p_1 = p$
Binomial(n, p)	0, 1, 2, ..., n	$p_i = \binom{n}{i} p^i (1 - p)^{n-i}, i = 0, 1, 2, \dots, n$
Geometric(p)	1, 2, ...	$p_i = p(1 - p)^{i-1}, i = 1, 2, \dots$ (Number of trials until first success)
Negative Binomial(r, p)	$r, r + 1, \dots$	$p_i = \binom{i-1}{r-1} p^r (1 - p)^{i-r}, i = r, r + 1, \dots$ (Number of trials until r -th success)
	0, 1, 2, ...	$p_i = \binom{i+r-1}{i} p^r (1 - p)^i, i = 0, 1, 2, \dots$ (Number of failures before r -th success)
Poisson(λ)	0, 1, 2, ...	$p_i = e^{-\lambda} \frac{\lambda^i}{i!}, i = 0, 1, 2, \dots$
Hypergeometric(s, n, N, M)	$\max(0, n - (N - M)), \dots, \min(M, n)$	$p_i = \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$

Note on alternative derivations and parameterizations: The distributions listed above can often be derived through multiple probabilistic mechanisms, which leads to different (but equivalent) parameterizations and interpretations. These alternative derivations highlight the interconnectedness of discrete probability distributions and provide flexibility in choosing the most appropriate model for a given data-generating process.

3.2. Distributions Generated by Mixtures

A powerful method for constructing flexible probability models is through mixture distributions, where the parameter of one distribution is itself governed by another distribution. This approach naturally incorporates heterogeneity and over-dispersion, commonly observed in real-world data.

3.2.1. Poisson Mixtures

When the rate parameter λ of a Poisson distribution is not fixed but follows a continuous distribution $G(\lambda)$ with support on $\mathbb{R}^+$, the resulting marginal distribution is a Poisson mixture. Its probability mass function is given by

$$P(X = k) = \int_0^\infty \frac{e^{-\lambda} \lambda^k}{k!} dG(\lambda), \quad k = 0, 1, 2, \dots$$

The most prominent example is the negative binomial distribution, which arises as a Poisson-gamma mixture. Specifically, if

$$X | \lambda \sim \text{Poisson}(\lambda) \quad \text{and} \quad \lambda \sim \text{Gamma}(r, \beta = (1 - p)/p),$$

then the marginal distribution of X is negative binomial with pmf:

$$P(X = k) = \frac{\Gamma(r + k)}{k! \Gamma(r)} p^r (1 - p)^k, \quad k = 0, 1, 2, \dots$$

This model is widely used for count data where the variance exceeds the mean.

3.2.2. Binomial Mixtures

Similarly, allowing the success probability p of a binomial distribution to vary according to a continuous distribution $F(p)$ on $[0, 1]$ leads to a binomial mixture. The marginal pmf is

$$P(X = k) = \int_0^1 \binom{n}{k} p^k (1 - p)^{n-k} dF(p), \quad k = 0, 1, \dots, n$$

The beta-binomial distribution is the canonical example, where $p \sim \text{Beta}(\alpha, \beta)$. Its pmf is

$$P(X = k) = \binom{n}{k} \frac{B(k + \alpha, n - k + \beta)}{B(\alpha, \beta)}$$

where $B(\cdot, \cdot)$ is the beta function. This distribution effectively models counts with extra-binomial variation.

Other notable mixtures include the Poisson-inverse Gaussian and the negative binomial-beta distributions [48].

3.3. Distributions Generated by Random Sums

Distributions generated by random sums, also known as compound distributions, provide a framework for modeling the aggregate outcome of a random number of events. Let N be a non-negative integer-valued random variable with pmf $p_N(n)$, and let $\{X_i\}$ be a sequence of *i.i.d.* random variables, independent of N . The random sum is defined as

$$S = X_1 + X_2 + \cdots + X_N, \quad (\text{with the convention that } S = 0 \text{ if } N = 0).$$

The distribution of S is called the compound distribution of N and X_i . Its probability generating function (pgf) is given by $G_S(t) = G_N(G_X(t))$, from which moments and probabilities can be derived.

Prominent examples include the following:

- Compound Poisson distribution: If $N \sim \text{Poisson}(\lambda)$, then S is compound Poisson. This is a highly flexible family used in insurance (to model total claim amounts) and queueing theory. Special cases include the Poisson-exponential and Poisson-gamma (which is equivalent to the negative binomial under a specific parameterization) distributions.
- Geometric sum: If $N \sim \text{Geometric}(p)$, and $X_i \sim \text{Exp}(\theta)$, then $S \sim \text{Exp}(\theta p)$. This result is foundational in renewal theory.

The family of phase-type distributions can also be constructed using random sums of exponential random variables, demonstrating the versatility of this method for generating distributions with specific tail behaviors [49].

3.4. Approximation to Univariate Continuous Distributions by Representative Points

Consider a random variable $X \sim F(x)$, where $F(x)$ is a univariate cumulative distribution function. The empirical distribution function $F_k(x)$ is defined as

$$F_k(x) = \frac{1}{k} \sum_{i=1}^k I\{x_i \leq x\}, \quad (19)$$

where I_A is the indicator function of A . The empirical distribution $F_k(x)$ is a discrete distribution with support points $x_1, \dots, x_k$, each assigned an equal probability mass of $1/k$. It serves as a consistent approximation of $F(x)$, meaning $F_k(x) \rightarrow F(x)$ in distribution as $k \rightarrow \infty$. This distribution can be represented as

$$\begin{array}{c|cccc} Y_{MC} & x_1 & x_2 & \cdots & x_k \\ \hline p_i & \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{array}. \quad (20)$$

Alternatively, one can construct other approximations using the so-called representative points (RPs) as support points. For a given distribution $F(x)$ in $\mathbb{R}$, several types of RPs exist, including

- Monte Carlo RPs (MC-RPs): the empirical sample points.
- Quasi-Monte Carlo RPs (QMC-RPs): deterministic points based on low-discrepancy sequences.
- Mean square error RPs (MSE-RPs): points minimizing a mean squared error criterion.

For a univariate continuous distribution $F(x)$, the QMC-RP approximation is given by the following distribution:

$$\frac{Y_{QMC} \mid F^{-1}(\frac{1}{2k}) \quad F^{-1}(\frac{3}{2k}) \quad \dots \quad F^{-1}(\frac{2k-1}{2k})}{p_i \mid \frac{1}{k} \quad \frac{1}{k} \quad \dots \quad \frac{1}{k}}. \quad (21)$$

MSE-RPs and related approximation distribution is

$$\frac{Y_{MSE} \mid \xi_1 \quad \xi_2 \quad \dots \quad \xi_k}{p_i \mid p_1 \quad p_2 \quad \dots \quad p_k}. \quad (22)$$

where the support points $\{\xi_1, \dots, \xi_k\}$ are determined by

$$\text{MSE}(\mathcal{Y}) = \int_{-\infty}^{+\infty} \min_i (x - \xi_i^{(k)})^2 p(x) dx = \sum_{j=1}^k \int_{I_j} (x - \xi_i^{(k)})^2 p(x) dx, \quad (23)$$

where

$$I_1 = (a_1, a_2), \quad I_2 = (a_2, a_3), \quad \dots, \quad I_k = (a_k, a_{k+1}), \quad (24)$$

$$a_1 = -\infty, \quad a_i = (\xi_{i-1}^{(k)} + \xi_i^{(k)})/2, \quad i = 2, \dots, k, \quad a_{k+1} = \infty.$$

Fang and Pan [50] and Fang et al. [8] provided a comprehensive review of the theory, generating algorithms, and applications of representative points in statistical inference.

3.4.1. Difference Between Two Distributions

The Kullback–Leibler (KL) divergence, or relative entropy, of two probability distributions with probability density functions $p(x)$ and $q(x)$ on $\mathcal{X} \subset \mathbb{R}^p$ is defined as the expected logarithmic ratio of densities concerning $p(x)$:

$$D_{KL}(p(x)|q(x)) = \int_{\mathcal{X}} \log\left(\frac{p(x)}{q(x)}\right) p(x) dx = E_{p(x)}\left[\log\left(\frac{p(x)}{q(x)}\right)\right]. \quad (25)$$

The KL divergence is non-negative and is equal to 0 if and only if the two distributions are identical.

Now, consider the univariate case. For any discrete distribution, we can find a corresponding density estimator, denoted by $\hat{p}(x)$. Let $F(x)$ be a cdf with pmf $p(x)$, and let $\hat{F}(x)$ with pmf $\hat{p}(x)$ be an approximating distribution. Several criteria in the literature measure the distance between F and $\hat{F}$:

$$\begin{aligned} \text{KL divergence (KL)} : \text{KL}(p \parallel \hat{p}) &= \sum_x \log\left(\frac{p(x)}{\hat{p}(x)}\right) p(x); \\ L^2\text{-norm of pmfs (} L^2 \text{ pdf)} : \|p, \hat{p}\| &= \left(\sum_x [p(x) - \hat{p}(x)]^2 \right)^{1/2}; \\ L^2\text{-norm of cdfs (} L^2 \text{ cdf)} : \|F, \hat{F}\| &= \left(\int_{\mathbb{R}} [F(x) - \hat{F}(x)]^2 dx \right)^{1/2}. \end{aligned}$$

Another metric for the distance between $F(x)$ and $\hat{F}(x)$ is the absolute bias index (ABI), which evaluates the overall bias in parameter estimation. For a population distribution $F(x; \theta)$ with parameter vector $\theta = (\theta_1, \dots, \theta_m)$ and its estimate $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)$ under $\hat{F}(x)$, the ABI is defined as follows:

$$ABI = \frac{1}{m} \sum_{i=1}^m \left| \frac{\theta_i - \hat{\theta}_i}{\theta_i} \right|. \quad (26)$$

3.4.2. Discrete Analogues

Discrete analogs aim to create new discrete distributions that preserve one or more defining characteristics of an original continuous distribution. An approximation $\hat{F}(x)$ is considered successful if it maintains as much information from $F(x)$ as possible, whether measured by the distance between the distributions, the similarity of their moments, or the properties of simulated samples.

As a continuous random variable X can be characterized by its pdf, moments, generating functions, or other features, the construction of a discrete analog Y (with pmf) from X (with density $f(x)$) focuses on matching these properties. The primary methods for this construction, summarized by [51], are classified as follows.

- (1) Discrete analogs of the Pearson differential equation. This approach constructs discrete probability distributions by employing difference equations that mirror the structure of the Pearson differential equation, which defines the continuous Pearson family [52,53].
- (2) The probability mass function (pmf) of Y retains the form of the pdf of X , and the support of Y is determined from the full range of X . The pmf is defined as

$$P(Y = k) = \frac{f(k)}{\sum_{j=-\infty}^{\infty} f(j)}, \quad k = 0, \pm 1, \pm 2, \dots$$

This approach has been used to create discrete analogs of various continuous distributions, including the gamma, general Dirichlet, normal, lognormal, exponential, Laplace, generalized exponential, and gamma distributions [54–56].

- (3) The survival function (SF) of Y retains the form of the survival function of X , and the support of Y is determined from the full range of X . The discrete survival function is defined as $S_Y(k) = P(Y \geq k)$, and its corresponding cdf is $F_Y(k) = P(Y \leq k)$, related by $S_Y(k) = 1 - F_Y(k - 1)$. This technique has generated numerous discrete distributions, such as the discrete exponential, Weibull, geometric Weibull, normal, Rayleigh, Maxwell, extended exponential, and Lindley distributions [57,58].
- (4) The hazard (failure) rate function of Y retains the form of the hazard rate function of X . This method preserves the hazard rate function. For a continuous random variable X with survival function $S_X(x) = P(X \geq x)$ and hazard rate function $\lambda_X(x) = f_X(x)/S_X(x)$, the survival function of the discrete analog Y is given by

$$P(Y \geq k) = \prod_{i=1}^{k-1} (1 - \lambda_X(i)), \quad k = 1, \dots, m,$$

and its pmf is

$$P(Y = k) = \lambda_X(k) \prod_{i=1}^{k-1} (1 - \lambda_X(i)).$$

This approach has been used to propose new discrete distributions, such as the discrete Lomax, Weibull, and Rayleigh [59].

- (5) The moments of Y and X coincide up to a certain order. This method requires that Y and X share the same finite r^{th} moment for $r = 0, 1, \dots, 2N - 1$, and that their cdfs coincide at least at $M + 1$ points. The support of the discrete random variable Y is $\{y_1, \dots, y_{MN}\}$. It has been used to define new discrete uniform, normal, gamma, and beta distributions [60].

There are more other construction methods, for instance, an extensive review on discrete distributions defined for integer data is provided in [61].

4. Multivariate Distributions

The commonly used discrete distributions can be extended to multivariate cases, including multinomial and multivariate negative binomial, hypergeometric, and Poisson distributions.

4.1. Multinomial Distribution

An extension of the binomial distribution is the so-called multinomial distribution. It is the fundamental model for counting events that fall into multiple categories, making it essential in fields like statistics, genetics, machine learning (e.g., topic modeling with naive Bayes), and survey analysis. Considering an experiment consists of n identical, independent trials, each trial results in one of k mutually exclusive and exhaustive categories with probabilities $p_i, i = 1, \dots, k; p_1 + \dots + p_k = 1$. Let random vector $\mathbf{X} = (X_1, \dots, X_k)$ follow a multinomial distribution $Multinomial(n; p_1, \dots, p_k)$. Its probability mass function is

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1! \dots n_k!}, n_1, \dots, n_k \in \mathbb{N}_0, n_1 + \dots + n_k = n. \quad (27)$$

The multinomial distribution has many nice properties. One is related to Poisson distribution, that lets $X_1, \dots, X_k$ be independent and $X_i \sim \text{Poisson}(\lambda_i), i = 1, \dots, k$. Then the conditional distribution of $(X_1, \dots, X_k | X_1 + \dots + X_k = n)$ follows $Multinomial(n; p_1, \dots, p_k)$ where $p_i = \lambda_i / (\lambda_1 + \dots + \lambda_k), i = 1, \dots, k$. The multinomial distribution has been used in many statistical applications. One is in the Pearson's chi-square test for goodness-of-fit, using the fact that the test statistic, calculated from observed multinomial counts and expected counts (np_i) , follows a chi-square distribution asymptotically (as n grows large).

4.2. Multivariate Hypergeometric Distribution

The multivariate hypergeometric distribution is the generalization of the hypergeometric distribution to more than two categories. It describes the probability of drawing specific numbers of items from each category without replacement. Considering a population with size N , of which $M_1, \dots, M_k$ are the numbers of items in the k mutually exclusive and exhaustive categories with $\sum_{i=1}^k M_i = N$, suppose that the n random samples are sampling without replacement containing $n_1, \dots, n_k$ items of each corresponding category. The joint distribution of $\mathbf{X} = (X_1, \dots, X_k)$ has pmf

$$P(X_1 = n_1, \dots, n_k) = \frac{\prod_{i=1}^k \binom{M_i}{n_i}}{\binom{N}{n}}, \sum_{i=1}^k n_i = n, 0 \leq n_i \leq M_i, i = 1, \dots, k. \quad (28)$$

The distribution denoted as $Mult. Hypg.(n; M_1, \dots, M_k)$ is the so-called multivariate hypergeometric distribution. Without requiring the parameters in (28) to be non-negative, multivariate inverse hypergeometric, negative hypergeometric, and negative inverse hypergeometric distributions are then defined; see [62]. One property is that if $\mathbf{X}$ and $\mathbf{Y}$ are independent and each distributed as $Multinomial(n; p_1, \dots, p_k)$, then $\mathbf{X} | (\mathbf{X} + \mathbf{Y} = \mathbf{N})$ follows $Mult. Hypg.(n, N)$.

4.3. Multivariate Poisson Distribution

If the parameters $p_1, \dots, p_{k-1}$ in the multinomial distribution are allowed to tend to 0, this implies that p_k tends 1 as $n \rightarrow \infty$, such that $np_i \rightarrow \lambda_i$, $i = 1, 2, \dots, k-1$. Hence, the probability of obtaining the event $\cap_{i=1}^{k-1} (X_i = n_i)$ tends to

$$P(X_1 = n_1, \dots, X_{k-1} = n_{k-1}) = \prod_{i=1}^{k-1} \{e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!}\}. \quad (29)$$

This is the joint distribution of $k-1$ mutually independent Poisson random variables with means $\lambda_1, \dots, \lambda_{k-1}$. The distribution can also be obtained as a limiting form of negative multinomial distributions.

The simplest case, as derived from the multinomial limit, is one of independence, but more flexible distributions allowing for correlation are of greater practical interest. A more general and widely used approach to constructing a multivariate Poisson distribution with a covariance structure is to use a common mixture or latent variable model. A standard construction introduces a set of independent latent Poisson variables and defines the observed variables as their sums.

A classical bivariate Poisson distribution for random variables (X, Y) can be defined as follows:

$$X = Z_1 + Z_0, \quad Y = Z_2 + Z_0,$$

where Z_0, Z_1, Z_2 are independent Poisson random variables with rates $\lambda_0, \lambda_1, \lambda_2 \geq 0$, respectively. The joint probability mass function of (X, Y) is given by

$$P(X = x, Y = y) = \exp(-(\lambda_1 + \lambda_2 + \lambda_0)) \sum_{k=0}^{\min(x,y)} \frac{\lambda_1^{x-k}}{(x-k)!} \frac{\lambda_2^{y-k}}{(y-k)!} \frac{\lambda_0^k}{k!}.$$

The marginal distributions are $X \sim \text{Poisson}(\lambda_1 + \lambda_0)$ and $Y \sim \text{Poisson}(\lambda_2 + \lambda_0)$, and the covariance between X and Y is $\text{Cov}(X, Y) = \lambda_0 \geq 0$. This model can only account for non-negative correlation, which is often sufficient for many applications, such as modeling the number of insurance claims across different policy types or goals scored by two teams in a match.

This construction can be generalized to higher dimensions by introducing more common components, leading to a richer covariance structure [3,63,64]. These correlated multivariate Poisson models are essential tools in modern multivariate count data analysis.

4.4. Continuous Multivariate Distributions

Johnson et al. [4] provided a comprehensive compilation of many multivariate distributions. Johnson [7] offered an in-depth introduction to the generation of multivariate statistical simulations. Palmitesta and Provasi [65] reviewed a range of multivariate distributions and their computer generation methods. Their survey encompassed the Dirichlet distribution, the multivariate Johnson distribution system, multivariate beta and gamma distributions, the multivariate extreme value distribution, the inverse Gaussian multivariate distribution, the Cook–Johnson family of multivariate uniform distributions, and elliptically contoured distributions. They also examined techniques for generating random vectors with specified marginal distributions and correlation matrices. Fang, Kotz, and Ng [6] conducted a comprehensive study on symmetric multivariate distributions. Due to space limitations, this section focuses on selected multivariate distributions from the literature and meta-distributions constructed via copulas, see [66]. For more construction methods for multivariate distribution, see [67].

4.4.1. Stochastic Representation

The traditional definition of the multivariate normal distribution relies on its density function. This approach, however, has several weaknesses:

- (i) It assumes that the explicit form of the density function is known.
- (ii) It requires the analytical evaluation of high-dimensional integrals to compute probabilities, moments, and marginal distributions.

The stochastic representation operator, denoted by $\stackrel{d}{=}$, is a powerful tool for constructing new distributions. If two random vectors $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution, we write $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$.

Examples

$U \stackrel{d}{=} 1 - U$, where $U \sim U(0, 1)$;

$X \stackrel{d}{=} -X$, where $X \sim N(0, \sigma^2)$;

$X \stackrel{d}{=} \mu + \sigma Y$, where $X \sim N(\mu, \sigma^2)$ and $Y \sim N(0, 1)$.

Properties

- (1) $\mathbf{X} \stackrel{d}{=} \mathbf{Y} \Rightarrow (f_1(\mathbf{X}), \dots, f_m(\mathbf{X})) \stackrel{d}{=} (f_1(\mathbf{Y}), \dots, f_m(\mathbf{Y}))$, where f_j are measurable functions. Examples are $\mathbf{X}^\top \mathbf{X} \stackrel{d}{=} \mathbf{Y}^\top \mathbf{Y}$ and $\mathbf{X}/\|\mathbf{X}\| \stackrel{d}{=} \mathbf{Y}/\|\mathbf{Y}\|$.
- (2) X, Y , and Z are random variables,
 $X \stackrel{d}{=} Y$ can imply $X + Z \stackrel{d}{=} Y + Z$, if Z is independent of (X, Y) .
- (3) If $X + Z \stackrel{d}{=} Y + Z$, and Z is independent of (X, Y) , it is not necessary for $X \stackrel{d}{=} Y$.
- (4) If Z is independent of (X, Y) , then
 - a. $X \stackrel{d}{=} Y$ implies $XZ \stackrel{d}{=} YZ$;
 - b. $XZ \stackrel{d}{=} YZ$ implies $X \stackrel{d}{=} Y$ under some conditions.
- (5) $X \stackrel{d}{=} Y$ if and only if $X^+ \stackrel{d}{=} Y^+$ and $X^- \stackrel{d}{=} Y^-$.

Example 1. The multivariate normal distribution, denoted by $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, has the following stochastic representation:

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}^\top \mathbf{Z}, \quad (30)$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top$, $\mathbf{A}^\top \mathbf{A} = \boldsymbol{\Sigma}$, and $\mathbf{Z} = (Z_1, \dots, Z_p)^\top$ with i.i.d. elements $Z_i \sim N(0, 1)$. It is easy to find density function, moments, and marginal distributions of $\mathbf{Z}$ and then transfer these to $\mathbf{X}$. For example, from the density function of $\mathbf{Z}$, $\frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}$, it is straightforward to find the density of $\mathbf{X}$ as

$$f(\mathbf{x}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top (\mathbf{x} - \boldsymbol{\mu}) \right\}. \quad (31)$$

As the mean and covariance of $\mathbf{Z}$ are, respectively, $\mathbf{0}$ and $\mathbf{I}_p$, one can immediately find $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$, respectively. It is also easy to find that all the marginal and affine transformations of $\mathbf{X}$ are normally distributed. For more detailed discussion and applications, see [68].

Many efforts to extend constructing methods from univariate to multivariate, for example, Pearson system, exponential family, and many others, can be found in [4].

4.4.2. Dirichlet Distribution

Let $X_1, \dots, X_{m-1}$ and X_m be independent random variables, and $X_i \sim \text{Gamma}(\alpha_i)$, $i = 1, \dots, m$. Set

$$Y = X_1 + \dots + X_m, \quad Y_j = \frac{X_j}{Y}, j = 1, \dots, m.$$

The distribution of $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$ is called the Dirichlet distribution with parameters $\alpha_1, \dots, \alpha_m$ and is written as $(Y_1, \dots, Y_m) \sim D_m(\alpha_1, \dots, \alpha_m)$.

The beta distribution $Be(\alpha, \beta)$ is a special case of Dirichlet distribution. As $Y_1 + \dots + Y_m = 1$, the random vector $(Y_1, \dots, Y_m)$ does not have a density on $\mathbb{R}^m$. However, the density of the first $m - 1$ components (with respect to Lebesgue measure on the simplex) is given by

$$f(y_1, \dots, y_{m-1}) = \frac{1}{B_m(\boldsymbol{\alpha})} \prod_{i=1}^m y_i^{\alpha_i-1}, \quad \sum_{i=1}^m y_i = 1, \quad 0 \leq y_i \leq 1, i = 1, \dots, m,$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$ and

$$B_m(\alpha_1, \dots, \alpha_m) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_m)}{\Gamma(\alpha_1 + \dots + \alpha_m)}.$$

Moreover, Y is independent of $\{Y_j, j = 1, \dots, m - 1\}$.

There is a fundamental relationship between the Dirichlet distribution and the uniform distribution on the unit sphere $S^p = \{\mathbf{x} : \mathbf{x}^\top \mathbf{x} = 1\}$ [68].

Theorem 2. Suppose that $\mathbf{U} \sim U(S^p)$ in $\mathbb{R}^p$. Then $\mathbf{U}$ has a stochastic representation

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_m \end{bmatrix} \stackrel{d}{=} \begin{bmatrix} D_1 \mathbf{U}^{(1)} \\ \vdots \\ D_m \mathbf{U}^{(m)} \end{bmatrix}, \quad (32)$$

where $\mathbf{U}^{(j)} \in \mathbb{R}^{t_j}$, $t_1 + \dots + t_m = p$, and

- (i) $\mathbf{U}^{(j)} \sim U(S^{t_j})$, the uniform distribution on S^{t_j} , $j = 1, \dots, m$;
- (ii) $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(m)}$ and $(D_1, \dots, D_m)$ are mutually independent;
- (iii) $D_j > 0$, $j = 1, \dots, m$, and $(D_1^2, \dots, D_m^2) \sim D_m(t_1/2, \dots, t_m/2)$.

Based on the Dirichlet distribution, the *multivariate Liouville distribution* is defined on $\mathbb{R}_+^m$ by the stochastic representation $\mathbf{X} \stackrel{d}{=} R\mathbf{Y}$, where $R > 0$ is a random variable independent of $\mathbf{Y}$ and $\mathbf{Y} \sim D_m(\alpha_1, \dots, \alpha_m)$. For a comprehensive study of the Liouville distribution, see Chapter 6 of [6].

4.4.3. Spherical and Elliptical Distributions

There are a number of equivalent ways to define the spherical distribution. Let $\mathbf{X}$ be a $p \times 1$ random vector. The following statements are equivalent [6]:

- (i) $\mathbf{X} \stackrel{d}{=} P\mathbf{X}$ for each $P \in O(p)$, where $O(p)$ is the set of orthogonal matrices of order p ;
- (ii) The characteristic function (c.f.) of $\mathbf{X}$ is of the form $\psi(\mathbf{t}^\top \mathbf{t})$ for some $\psi \in \Psi_p$ where

$$\Psi_p = \{\psi(\cdot) : \psi(t_1^2 + \dots + t_p^2) \text{ is a } p\text{-dimensional c.f.}\}; \quad (33)$$

- (iii) $\mathbf{X}$ has a stochastic representation $\mathbf{X} \stackrel{d}{=} R\mathbf{U}^{(p)}$, where random variable $R \geq 0$ is independent of $\mathbf{U}^{(p)}$, $\mathbf{U}^{(p)}$ is uniformly distributed on the unit sphere S^p and does not have a density;
- (iv) For any $\mathbf{a} \in S^p$ and $\mathbf{b} \in S^p$, we have $\mathbf{a}^\top \mathbf{X} \stackrel{d}{=} \mathbf{b}^\top \mathbf{X}$.

Theorem 3. Let $\mathbf{X} \stackrel{d}{=} R\mathbf{U}^{(p)}$ and $P(\mathbf{X} = \mathbf{0}) = 0$. Then

$$\|\mathbf{X}\| \stackrel{d}{=} R, \quad \mathbf{X}/\|\mathbf{X}\| \stackrel{d}{=} \mathbf{U}^{(p)},$$

and they are independent.

Example 2. The theorem has two immediate applications. First, it provides a method to generate a random sample from the uniform distribution on the sphere, $U(S^p)$: simply sample $\mathbf{X}$ from $N_p(\mathbf{0}, \mathbf{I}_p)$ and normalize it to obtain $\frac{\mathbf{X}}{\|\mathbf{X}\|}$. Second, it allows the mixed moments of $\mathbf{U}^{(p)}$ to be easily calculated from the known mixed moments of the standard normal vector $\mathbf{X}$ and the moments of its radius $R = \|\mathbf{X}\|$. A detailed discussion can be found in Section 3.1.2 of [6].

Theorem 4. Let $\mathbf{X} \sim S_p(\psi)$ denote a spherically symmetric distribution in $\mathbb{R}^p$, characterized by its characteristic function $\psi(\mathbf{t}^\top \mathbf{t})$ as (33). It is not necessary that $\mathbf{X}$ has a density. The random vector $\mathbf{X} \stackrel{d}{=} R\mathbf{U}^{(p)}$ has a density which must be of the form $g(\mathbf{x}^\top \mathbf{x})$, if and only if the radial variable R has a density, denoted by $f(r)$. The relationship between g and f is given by

$$f(r) = \frac{2\pi^{p/2}}{\Gamma(p/2)} r^{p-1} g(r^2),$$

where $g(\cdot)$ is called the density generator. We write $\mathbf{X} \sim S_p(g)$.

The necessary condition that $\mathbf{X} \sim S_p(\psi)$ has a density is

$$\int_0^\infty y^{p/2-1} g(y) dy < \infty$$

as

$$\int g(\mathbf{x}^\top \mathbf{x}) d\mathbf{x} = \frac{\pi^{p/2}}{\Gamma(p/2)} \int_0^\infty y^{p/2-1} g(y) dy = 1.$$

Table 7 gives many useful spherical distributions from [6].

Table 7. Some subclasses of p -dimensional spherical distributions.

Type	Density Function $g(\mathbf{x}^\top \mathbf{x})$ or c.f. $\psi(\mathbf{t}^\top \mathbf{t})$
Multinormal	$g(\mathbf{x}^\top \mathbf{x}) = c \exp(-\frac{1}{2}\mathbf{x}^\top \mathbf{x})$
Kotz-type	$g(\mathbf{x}^\top \mathbf{x}) = c(\mathbf{x}^\top \mathbf{x})^{N-1} \exp[-r(\mathbf{x}^\top \mathbf{x})^s]$, $r, s > 0$, $2N + p > 2$
Pearson type VII	$g(\mathbf{x}^\top \mathbf{x}) = c(1 + \mathbf{x}^\top \mathbf{x}/s)^{-N}$, $N > p/2, s > 0$
Multivariate t	$g(\mathbf{x}^\top \mathbf{x}) = c(1 + \mathbf{x}^\top \mathbf{x}/v)^{-(p+v)/2}$, $v > 0$
Multivariate Cauchy	$g(\mathbf{x}^\top \mathbf{x}) = c(1 + \mathbf{x}^\top \mathbf{x})^{-(p+1)/2}$
Pearson Type II	$g(\mathbf{x}^\top \mathbf{x}) = c(1 - \mathbf{x}^\top \mathbf{x})^q$, $q > 0$ for $\mathbf{x}^\top \mathbf{x} < 1$
Logistics	$g(\mathbf{x}^\top \mathbf{x}) = c \exp(-\mathbf{x}^\top \mathbf{x}) / [1 + \exp(-\mathbf{x}^\top \mathbf{x})]^2$
Scale mixture	$g(\mathbf{x}^\top \mathbf{x}) = c \int_0^\infty t^{-p/2} \exp(-\mathbf{x}^\top \mathbf{x}/2t) dG(t)$, $G(t)$ a cdf.

Elliptical Distribution

A random vector $\mathbf{X}$ has an elliptical distribution, denoted as $\mathbf{X} \sim EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \mathbf{A}^\top \mathbf{A}$, if it can be expressed as

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}^\top \mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{R} \mathbf{A}^\top \mathbf{U}^{(p)}, \quad (34)$$

where $\mathbf{Y} \sim S_p(\psi)$ is spherical. This relationship generalizes the way a normal variable $X \sim N(\mu, \sigma^2)$ is constructed from a standard normal $Z \sim N(0, 1)$. Consequently, any subclass of spherical distributions in Table 7 defines a corresponding subclass of elliptical distributions.

4.4.4. Approximation to Multivariate Continuous Distributions by Representative Points

For multivariate continuous distributions, the QMC-RPs and MSE-RPs can be defined in a similar way to the univariate case. Suppose that a random vector $\mathbf{X} \sim F(\cdot)$ in $\mathbb{R}^n$ has a density function $f(\cdot)$ with a finite mean vector and covariance matrix. A set of points $\Xi = \{\boldsymbol{\xi}_j, j = 1, \dots, k\}$ is called mean squared error representative points if it minimizes the mean squared error, $\text{MSE}(\Xi) = \int_{\mathbb{R}^n} \min_{j=1, \dots, k} \|\mathbf{x} - \boldsymbol{\xi}_j\|^2 f(\mathbf{x}) d\mathbf{x}$, where $\|\cdot\|$ denotes the L^2 -norm. For information regarding the construction of approximation distributions of a continuous multivariate distribution by their QMC-RPs, one can refer to [8,69,70]. For information regarding the construction of approximation distributions of a continuous multivariate distribution by their MSE-RPs, one can refer to [8,70–73]. Due to the complexity of computation in high-dimensional multivariate distribution, the related research works about MSE-RPs of this kind of distribution are limited. Representative points also have many applications for discrete data, especially for discrete massive data. In Chapter 9 of [8], some cases where the sample size is hugely larger than the number of variables have been considered, and some model-based and model-free subsampling methods are included.

4.5. Basic Statistics Under the Multivariate Population

Let $\mathbf{X}$ follow a multivariate continuous distribution $F(\mathbf{x}; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ represents parameters. A random sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ is drawn from this population. The distributions of basic statistics, such as the mean vector, the covariance matrix, and various test statistics, are fundamental to statistical inference and its applications. For example, when the population distribution is multivariate normal, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the standard unbiased estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are sample mean and the sample covariance, given, respectively, by

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^\top (\mathbf{x}_i - \bar{\mathbf{x}}). \quad (35)$$

It is known that $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n-1} \boldsymbol{\Sigma})$. Other key related distributions are described below.

Wishart distribution: The Wishart distribution, denoted by $W_p(n, \boldsymbol{\Sigma})$, is the multivariate generalization of the chi-square distribution. It is defined by $\mathbf{W} = \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^\top$ and $\mathbf{Z}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ are *i.i.d.* The sample covariance matrix $\mathbf{S}$ follows Wishart distribution: $(n-1)\mathbf{S} \sim W_p(n-1, \boldsymbol{\Sigma})$. This distribution was developed by John Wishart [74], who, in his seminal 1928 paper “The Generalised Product Moment Distribution in samples from a normal multivariate population”, published in *Biometrika*, derived the distribution of the matrix of sums of squares and cross-products. A crucial result for normally distributed data is that the sample mean $\bar{\mathbf{x}}$ and the sample covariance matrix $\mathbf{S}$ are independent, generalizing the corresponding univariate property.

Distributions of the eigenvalues of $\mathbf{S}$: Unlike the sample covariance matrix itself, the distributions of its individual eigenvalues are complex and do not follow simple, universal forms like the normal or Wishart distributions.

Hotelling's T-squared distribution: It is the multivariate generalization of the Student's t -distribution, used primarily for hypothesis testing on the mean vector. For a random vector $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ independent of $\mathbf{S} \sim W_p(n, \boldsymbol{\Sigma})$, the Hotelling's T^2 statistic is denoted as

$$T^2 = n\mathbf{X}^\top \mathbf{S}^{-1} \mathbf{X}. \quad (36)$$

The resulting T^2 statistic, despite being derived from multivariate data, is a scalar. Therefore, its sampling distribution, denoted by $T^2(p, n, \boldsymbol{\mu})$, is a univariate distribution. When $\boldsymbol{\mu} = \mathbf{0}$, the distribution is central and denoted by $T^2(p, n)$. The T^2 statistic is related to the F -distribution through the following identity: $\frac{n-p+1}{np} T^2(p, n) \sim F(p, n-p+1)$.

Wilks' lambda distribution: This distribution generalizes the beta distribution and is used in multivariate analysis of variance and likelihood ratio tests. Let $A_1 \sim W_p(m, \boldsymbol{\Sigma})$ and $A_2 \sim W_p(n, \boldsymbol{\Sigma})$ be independent. The Wilks' lambda statistic, defined as $\Lambda = \frac{A_2}{A_1 + A_2}$, is a scalar. Thus, its sampling distribution, denoted by $\Lambda(p, m, n)$, is a univariate distribution. Wilks' lambda has close relationships with the T^2 distribution and simplifies to an F or beta distribution when $p = 1$.

When the population $\mathbf{X}$ follows an elliptical distribution (a broader class than the normal), deriving the distributions of the aforementioned statistics becomes significantly more challenging. A comprehensive discussion of this topic can be found in [68].

Multivariate L_1 -Norm Symmetric Distribution

The elliptical distribution can be regarded as an extension of the multivariate normal distribution. Similarly, the exponential distribution can be generalized to a multivariate setting. Let $X_1, \dots, X_p$ be *i.i.d.* random variables from the standard exponential distribution, $Exp(1)$ (see Table 4). Consider the random vector $\mathbf{X} = (X_1, \dots, X_p)^\top$ and its L_1 -norm defined by $\|\mathbf{X}\|_1 = |X_1| + \dots + |X_p|$. Let $\mathbf{U} = \frac{\mathbf{X}}{\|\mathbf{X}\|_1}$. Fang and Fang [75] defined a class of multivariate distributions via the following stochastic representation:

$$\mathbf{X} \stackrel{d}{=} R\mathbf{U}, \quad (37)$$

where the radial variable R is independent of the directional vector $\mathbf{U}$. This class of distributions was studied systematically in [75,76] and in Chapter 5 of [6].

4.6. Definition by Distribution, Density or Survival Functions

4.6.1. Meta-Skew Symmetric Distributions

A traditional approach to defining a probability distribution is through its distribution function, density function, or survival function. For instance, the multivariate normal distribution in Anderson's book [77] is defined by its density function. A general univariate form of an asymmetric distribution, known as the skew-symmetric family, can be represented as a function that combines a cumulative distribution function and a probability density function; see [78,79].

Definition 2. A random variable X , whose density function has the form

$$p(x) = 2f(x)G(w(x)), \quad (38)$$

is called skew symmetric, where $f(x)$ is a density function centrally symmetric about 0, $w(x)$ is an odd real-valued function such that $w(-x) = -w(x)$, and $G(w(\cdot))$ is a cumulative distribution function on $\mathbb{R}$ such that $g = G'$ is an even density function, providing G is differentiable.

4.6.2. Skew-Normal Distribution

The multivariate skew-normal (MSN) distribution, introduced in Table 3, is a fundamental and powerful generalization of both the multivariate normal and the univariate skew-normal distributions. It provides a flexible framework for modeling multivariate data that exhibits asymmetry (skewness).

A p -dimensional random vector $\mathbf{Y}$ follows a multivariate skew-normal distribution with location vector $\boldsymbol{\mu}$, positive definite scale matrix $\boldsymbol{\Sigma}$, and skewness vector $\boldsymbol{\alpha}$ if its probability density function is given by

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}) = 2\phi_p(\mathbf{y} - \boldsymbol{\mu}; \boldsymbol{\Sigma})\Phi(\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\mu})),$$

where $\phi_p(\mathbf{z}; \boldsymbol{\Sigma})$ is the pdf of p -variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$; $\Phi(\cdot)$ is the cdf of $N(0, 1)$; $\boldsymbol{\omega}$ is the diagonal matrix formed from the standard deviations of $\boldsymbol{\Sigma}$ (the square roots of the diagonal elements), and the vector $\boldsymbol{\alpha}$ is the shape or skewness parameter vector. It denotes $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$.

The MSN distribution admits an intuitive stochastic representation. Consider a $(p + 1)$ -dimensional normal random vector $(X_0, \mathbf{X}_1^\top)^\top$,

$$\begin{bmatrix} X_0 \\ \mathbf{X}_1 \end{bmatrix} \sim N_{1+p} \left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & \boldsymbol{\delta}^\top \\ \boldsymbol{\delta} & \boldsymbol{\Sigma} \end{pmatrix} \right]. \quad (39)$$

Here, X_0 is a scalar and $\mathbf{X}_1$ is a p -vector. The vector $\boldsymbol{\delta}$ controls the correlation between X_0 and each component of $\mathbf{X}_1$. Then

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\omega}(\mathbf{X}_1 | X_0 > 0) \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}),$$

where the skewness parameter $\boldsymbol{\alpha}$ relates to $\boldsymbol{\delta}$ through the following equation:

$$\boldsymbol{\alpha} = \frac{1}{\sqrt{1 - \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}.$$

For comprehensive details on the theory and applications of the multivariate skew-normal distribution, see [80,81].

4.6.3. Skew-Elliptical Distributions

Skew-elliptical distributions form a broad family that generalizes symmetric elliptical distributions by introducing skewness (asymmetry). This extension moves beyond the familiar world of symmetric, bell-shaped curves into the asymmetric realm, providing a principled and mathematically sound framework for modeling a wider range of real-world data patterns. Among these, the skew- t distribution is often a particularly useful starting point for applied work due to its flexibility in handling both skewness and heavy tails.

The general form of the pdf for a skew-elliptical distribution is

$$f(\mathbf{z}) = 2f_0(\mathbf{z})\Pi(\mathbf{z}), \quad (40)$$

where $f_0(\mathbf{z})$ is the pdf of a symmetric elliptical distribution (the “parent”), and $\Pi(\mathbf{z})$ is a skewing function that maps onto the interval $[0, 1]$ and satisfies $\Pi(\mathbf{z}) + \Pi(-\mathbf{z}) = 1$. This function controls the direction and amount of skewness. The skew-elliptical family is vast, but several members are particularly important: skew-normal, skew- t , skew-Cauchy, and skew-Laplace distributions. The skew-elliptical distributions are used anywhere that data deviates from symmetry and normality, such as in finance, environmental science, biology, psychometrics, and social sciences. For a comprehensive study, refer to [82].

4.6.4. Definition by Survival Function

The survival function of a random variable $X \sim F(x)$ is defined as $\bar{F}(x) = P(X > x)$. This function is particularly useful in the analysis of biometrical data. Within this domain, the exponential and Weibull distributions play an important role.

A foundational model for dependent failures is the Marshall–Olkin multivariate exponential distribution [83]. Its strength lies in an intuitive shock-based construction and analytical tractability. The model assumes that shocks to the system components arrive according to three independent Poisson processes with rates λ_1 , λ_2 , and λ_{12} , corresponding to shocks that fatally affect only component 1, only component 2, or both components simultaneously. The arrival times of these shocks, Z_1 , Z_2 , and Z_{12} , are, therefore, independent exponential random variables with parameters λ_1 , λ_2 , and λ_{12} , respectively.

We set $X = \min\{Z_1, Z_{12}\}$ and $Y = \min\{Z_2, Z_{12}\}$. The joint survival function of the random vector (X, Y) is defined as $\bar{H}(x, y) = P(X > x, Y > y)$. For any $x, y \geq 0$ the survival function is given by

$$\begin{aligned}\bar{H}(x, y) &= P(Z_1 > x)P(Z_2 > y)P(Z_{12} > \max(x, y)) \\ &= \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)].\end{aligned}$$

Its marginal survival functions are

$$\bar{F}(x) = \exp(-(\lambda_1 + \lambda_{12})x), \text{ and } \bar{G}(y) = \exp(-(\lambda_2 + \lambda_{12})y).$$

Hence, X and Y are exponential random variables with parameters $\lambda_1 + \lambda_{12}$ and $\lambda_2 + \lambda_{12}$, respectively. This elegant construction can be extended to higher dimensions.

4.7. Meta-Distributions via Copula Techniques

Various statistical distributions play a key role in modeling real-world data. The rapid development of data science has increased the demand for a wider variety of multivariate distributions. A common challenge is to construct a multivariate distribution from given one-dimensional marginal distributions while incorporating a known dependence structure. Such constructions arise not only from statistical theory but also from practical problems, particularly in risk and decision analysis for complex multivariate systems. The copula is a powerful technique designed for this purpose.

Copulas are mathematical functions that fully capture the dependence structure among random variables, offering great flexibility in building multivariate stochastic models. Since their introduction in the 1950s, copulas have gained significant popularity in several applied fields such as finance, insurance, and reliability theory. Nowadays, they are a well-recognized tool for market and credit modeling, risk aggregation, portfolio selection, and more.

The following definition is based on probability theory. For a more general treatment, see [66,84]. As most research works and applications are interested in the copulas with respect to continuous distributions, and it becomes complicated for discrete distributions, this section only covers the continuous case.

Definition 3. If $C(z_1, \dots, z_p)$ is the cumulative distribution function on the unit hypercube $[0, 1]^p$ with uniform $U(0, 1)$ margins, then C is called a copula.

The following Lemma demonstrates the existence of the copula function.

Lemma 1. Let $H(x_1, \dots, x_p)$ be a continuous distribution with p continuous margins F_i on the domain D_i for $i = 1, \dots, p$. Then there exists a unique copula C such that

$$C(z_1, \dots, z_p) = H(F_1^{-1}(z_1), \dots, F_p^{-1}(z_p)). \quad (41)$$

It is well known that for a given set of p marginal distributions $F_i(x_i)_{i=1}^p$, there exist infinitely many p -dimensional distributions with these margins. If the dependence structure of the joint distribution is known, one may find a unique or a small family of such joint distributions $H(x_1, \dots, x_p)$. The copula technique is a fundamental tool for this purpose.

Suppose we have p random variables $X_i \sim F_i(x_i)$ for $i = 1, \dots, p$, and we wish to construct a multivariate distribution H with these margins and a given dependence structure defined by a copula $C(z_1, \dots, z_p)$. This method is called the copula technique. The resulting distribution is termed a *meta-distribution* and is denoted by $F(C; F_1, \dots, F_p)$.

Theorem 5. Sklar's Theorem (Multivariate Case) Let $\mathbf{X}$ be a continuous random vector with joint distribution $H(x_1, \dots, x_p)$ and continuous marginal distributions F_i on domains D_i for $i = 1, \dots, p$. Then there exists a unique copula C such that

$$H(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p)). \quad (42)$$

Conversely, if C is a copula and $F_1, \dots, F_p$ are univariate distribution functions, then the function H defined by (42) is a p -dimensional distribution function with margins $F_1, \dots, F_p$.

If the margins F_i are continuous, then C is unique. The joint density function of $\mathbf{X}$ derived from (42) is given by

$$h(x_1, \dots, x_p) = c(F_1(x_1), \dots, F_p(x_p)) \prod_{i=1}^p f_i(x_i), \quad (43)$$

where f_i is the probability density function of X_i , and $c(\cdot)$ is the copula density, i.e., the p -th mixed partial derivative of C .

4.7.1. Some Examples of Copula

Many copula examples exist for lower dimensions. This subsection presents selected examples from Nelsen [66] to provide a tutorial for beginners to comprehend the ideas of copula. Examples 3 and 4 have different joint distributions but have the same copula.

Example 3. Adapted from Nelsen [66]

Consider the joint distribution

$$H(x, y) = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}, \quad (x, y) \in [-1, 1] \times [0, \infty).$$

Its marginal distributions are $X \sim U(-1, 1)$ with cdf $F(x) = \frac{x+1}{2}$ and $Y \sim \text{Exp}(1)$, the standard exponential distribution with cdf $G(y) = 1 - e^{-y}$, and copula

$$C(u, v) = \frac{uv}{u + v - uv}.$$

We now demonstrate how to derive the copula C from H , and, conversely, how to reconstruct H from C, F, G .

Deriving C from H :

Let $u = F(x)$ and $v = G(y)$, which implies $u = \frac{x+1}{2}$ and $v = 1 - e^{-y}$ or $e^y = \frac{1}{1-v}$.

$$\begin{aligned} C(u, v) &= H(F^{-1}(u), G^{-1}(v)) = \frac{(F^{-1}(u) + 1)(e^{G^{-1}(v)} - 1)}{F^{-1}(u) + 2e^{G^{-1}(v)} - 1} \\ &= \frac{[(2u - 1) + 1][e^{-\ln(1-v)} - 1]}{2u - 1 + 2e^{-\ln(1-v)} - 1} \\ &= \frac{2u(\frac{1}{1-v} - 1)}{2u - 1 + 2\frac{1}{1-v} - 1} = \frac{2u[1 - (1 - v)]}{(2u - 1)(1 - v) + 2 - (1 - v)} \\ &= \frac{uv}{u + v - uv}. \end{aligned}$$

Deriving H from $\{C, F, G\}$:

Note that $e^y = \frac{1}{1-v}$; we have

$$\begin{aligned} H(x, y) &= C(u, v) \\ &= \frac{\frac{x+1}{2}(1 - e^{-y})}{\frac{x+1}{2} + 1 - e^{-y} - \frac{x+1}{2}(1 - e^{-y})} \\ &= \frac{(x+1)(1 - e^{-y})}{x + 1 + 2(1 - e^{-y}) - (x+1)(1 - e^{-y})} \\ &= \frac{(x+1)(e^y - 1)}{(x+1) + 2e^y - 2} = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}. \end{aligned}$$

Example 4. Gumbel's bivariate logistic distribution [85].

$$H(x, y) = (1 + e^{-x} + e^{-y})^{-1}, x, y \geq 0.$$

Its two margins are

$$F(x) = (1 + e^{-x})^{-1}, G(y) = (1 + e^{-y})^{-1}.$$

Its copula is

$$C(u, v) = \frac{uv}{u + v - uv},$$

which is identical to the copula in Example 3, despite the marginal distributions being quite different.

For multivariate copulas, higher-dimensional margins are defined as per the following example.

Example 5. Consider the following three-dimensional distribution H :

$$H(x, y, z) = \frac{(x+1)(e^z - 1) \sin(z)}{x + 2e^y - 1}, \text{ on } [-1, 1] \times [0, \infty] \times [0, \frac{\pi}{2}].$$

Its three one-dimensional marginal distributions are

$$H_1(x) = \frac{x+1}{2}, H_2(y) = 1 - e^{-y}, H_3(z) = \sin(z).$$

The three two-dimensional marginal distributions are

$$\begin{aligned} H_{12}(x, y) &= \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}, \\ H_{23}(y, z) &= (1 - e^{-y}) \sin(z), \\ H_{13}(x, z) &= \frac{1}{2}(x+1) \sin(z). \end{aligned}$$

Example 6. Fang et al. [86] provided a comprehensive study of the copula defined for $\alpha > 0$ and $\beta \in [0, 1]$:

$$C(u, v; \alpha, \beta) = \frac{uv}{[1 - \beta(1 - u^{1/\alpha})(1 - v^{1/\alpha})]^\alpha}. \quad (44)$$

For different values of α and β , we have the following:

1. $\beta = 0$; U and V are independent.
2. $\beta = 1$; it reduces to the bivariate Clayton distribution [87].
3. $\alpha = 1$; it reduces to the Ali–Mikhail–Haq copula [88].

Its multivariate version is given by

$$C(u_1, \dots, u_n; \alpha, \beta) = \frac{\prod_{i=1}^n u_i}{[1 - \beta \prod_{i=1}^n (1 - u_i^{1/\alpha})]^\alpha}. \quad (45)$$

These examples demonstrate that the copula technique provides a powerful method for constructing numerous new distributions useful for statistical modeling.

4.7.2. Gaussian Copula and Meta-Gaussian Distribution

When $H(\cdot)$ is a multivariate normal distribution, the associated copula is called a Gaussian copula. When $H(\cdot)$ is an elliptical distribution, the corresponding copula is known as an elliptical copula. This family includes the Gaussian copula, the t -copula, and the Kotz-type copula, among others; for a comprehensive overview, see [89].

From the definition of a copula, the Gaussian copula for $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is identical to that of $\mathbf{Z} = (Z_1, \dots, Z_p) \sim N_p(\mathbf{0}, \mathbf{R})$, where $\mathbf{R}$ is the correlation matrix of $\mathbf{Z}$. Clearly, all marginal distributions of $\mathbf{Z}$ are standard normal $N(0, 1)$, yet its components have correlation matrix $\mathbf{R} = (r_{ij})$. Denote the cdf of $\mathbf{Z}$ by $G_{\text{normal}}(z_1, \dots, z_p \mid \mathbf{R})$ and define a function C as follows:

$$C_{\text{normal}}(u_1, \dots, u_p) = G_{\text{normal}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)), \quad (u_1, \dots, u_p)^\top \in [0, 1]^p, \quad (46)$$

where $\Phi(\cdot)$ is the cdf of $N(0, 1)$. This function C is the Gaussian copula. Kelly and Krzysztofowicz [90] studied the meta-Gaussian (or meta-normal) distribution. A more detailed analysis can be found in Section 2.6, “Multivariate Gaussian/Normal”, of [91].

Definition 4. The meta-Gaussian or meta-normal distribution with a set of given marginal cdfs $F_1, \dots, F_p$ is defined by its cdf

$$F_{X_1, \dots, X_p}(x_1, \dots, x_p) = G_{Z_1, \dots, Z_p}(\Phi^{-1}(F_1(x_1)), \dots, \Phi^{-1}(F_p(x_p))). \quad (47)$$

We denote this distribution as $\mathbf{X} \sim MN_p(\mathbf{0}, \mathbf{R}; F_1, \dots, F_p)$.

The pdf of $MN_p(\mathbf{0}, \mathbf{R}; F_1, \dots, F_p)$ is

$$h(x_1, \dots, x_p) = c_{\text{normal}}(F_1(x_1), \dots, F_p(x_p)) \prod_{i=1}^p f_i(x_i),$$

where c_{normal} is the density of the Gaussian copula and f_i is the pdf corresponding to F_i . In the case where the correlation matrix has all off-diagonal elements equal to ρ , Bouy'e [92] provided illustrative plots for various values of ρ .

Let $X \sim MN_2(\mathbf{0}, \mathbf{R}; F_1, F_2)$, where the correlation matrix with $\text{Corr}(X_1, X_2) = \rho$. Then the normal copula is given by

$$C_{normal}(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left\{-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right\} ds dt,$$

where $0 \leq u, v \leq 1$.

The correlation coefficient has been used in multivariate analysis for a long time. However, the copula and its related meta-distributions might have different correlation coefficients due to different marginal distributions. Therefore, the so-called “measure of concordance” is considered in the literature. It requests that the copula and its meta-distributions have the same values of the measure of concordance. The Kendall's τ and the Spearman's ρ are the most popular ones used in the applications. Their definitions can refer to [66]. Spearman's ρ for the binormal copula is given by $\frac{6}{\pi} \arcsin(\frac{\rho}{2})$ (see [93]). This result does not extend to other elliptical distributions, which are introduced in the next section.

The meta-normal distribution $X \sim MN_2(\mathbf{0}, \mathbf{R}; F_1, F_2)$ is given by

$$\begin{aligned} F(x_1, x_2) &= C_{normal}(F_1(x_1), F_2(x_2)) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(F_1(x_1))} \int_{-\infty}^{\Phi^{-1}(F_2(x_2))} \exp\left\{-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right\} ds dt. \end{aligned} \quad (48)$$

Mai and Scherer [94], in Example 1.7, discussed the Gaussian copula and meta-Gaussian distributions for the general case $X \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In this case, the marginal distributions of X_i ($i = 1, \dots, p$) are not necessarily identical if the variances $\text{Var}(X_i)$ differ. Now, consider the two-dimensional normal distribution $N_2(\mathbf{0}, \mathbf{R})$ with correlation matrix $\mathbf{R}$ and correlation coefficient ρ . Let $\xi_i = \Phi^{-1}(u_i)$ for $i = 1, 2$. The density of the Gaussian copula, as given in Chapter 7 of [95], is

$$c_{normal}(\boldsymbol{\xi}) = c_{normal}(\xi_1, \xi_2) = |\mathbf{R}|^{-1/2} \exp\left(-\frac{1}{2} \boldsymbol{\xi}^\top (\mathbf{R}^{-1} - \mathbf{I}) \boldsymbol{\xi}\right). \quad (49)$$

This expression can be factorized into two independent components

$$c_{normal}(\boldsymbol{\xi}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \boldsymbol{\xi}^\top \mathbf{R}^{-1} \boldsymbol{\xi}\right) [2\pi \cdot \exp(\frac{1}{2}(\xi_1^2 + \xi_2^2))].$$

The first factor is the density of $N_2(\mathbf{0}, \mathbf{R})$, while the second factor is the product of the densities of two independent standard normal variables (up to a constant),

$$\sqrt{2\pi} \exp(\frac{1}{2}\xi_1^2) \text{ and } \sqrt{2\pi} \exp(\frac{1}{2}\xi_2^2).$$

Recently, He et al. [96] derived the covariance matrix of the Gaussian copula and related properties.

4.7.3. Marshall–Olkin Copula

The Marshall–Olkin multivariate exponential distribution was introduced in Section 4.6.4. We now derive its corresponding survival copula, denoted by $\hat{C}_{MO}(u, v)$. We begin by expressing the joint survival function $\bar{H}(x, y)$ in terms of the marginal survival functions $u = \bar{F}(x)$ and $v = \bar{G}(y)$, recalling the relationship $\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y))$, where $\hat{C}$ is

the survival copula. Using the identity $\max(x, y) = x + y - \min(x, y)$, we can rewrite the survival function as follows:

$$\begin{aligned}\bar{H}(x, y) &= \hat{C}(\bar{F}(x), \bar{G}(y)) \\ &= \exp[-(\lambda_1 + \lambda_{12})x - (\lambda_2 + \lambda_{12})y + \lambda_{12} \min(x, y)] \\ &= \bar{F}(x)\bar{G}(y) \min\{\exp(\lambda_{12}x), \exp(\lambda_{12}y)\}.\end{aligned}$$

As $u^{\frac{-\lambda_{12}}{\lambda_1 + \lambda_{12}}} = \exp(\lambda_{12}x)$ and $v^{\frac{-\lambda_{12}}{\lambda_2 + \lambda_{12}}} = \exp(\lambda_{12}y)$, let $\alpha = \frac{\lambda_{12}}{\lambda_1 + \lambda_{12}}$ and $\beta = \frac{\lambda_{12}}{\lambda_2 + \lambda_{12}}$. Obviously, $0 < \alpha, \beta < 1$, and let

$$\begin{aligned}u &= \bar{F}(x) = e^{-(\lambda_1 + \lambda_{12})x} = e^{-\lambda_{12}x/\alpha}, \quad u^{-\alpha} = e^{\lambda_{12}x}, \\ v &= \bar{G}(y) = e^{-(\lambda_2 + \lambda_{12})y} = e^{-\lambda_{12}y/\beta}, \quad v^{-\beta} = e^{\lambda_{12}y}.\end{aligned}$$

The survival copula for the Marshall–Olkin distribution, known as the Marshall–Olkin copula, is given by

$$\begin{aligned}\hat{C}_{MO}(u, v) &= uv \min\left\{u^{\frac{-\lambda_{12}}{\lambda_1 + \lambda_{12}}}, v^{\frac{-\lambda_{12}}{\lambda_2 + \lambda_{12}}}\right\} = \min\left\{vu^{1-\frac{\lambda_{12}}{\lambda_1 + \lambda_{12}}}, uv^{1-\frac{\lambda_{12}}{\lambda_2 + \lambda_{12}}}\right\} \\ &= uv \min(u^{-\alpha}, v^{-\beta}) = \min(u^{1-\alpha}v, uv^{1-\beta}).\end{aligned}\quad (50)$$

Therefore, the survival copula for the Marshall–Olkin bivariate exponential distribution forms a two-parameter family of copulas given by

$$C_{\alpha, \beta}(u, v) = \begin{cases} u^{1-\alpha}v, & \text{if } u^\alpha \geq v^\beta, \\ uv^{1-\beta}, & \text{if } u^\alpha \leq v^\beta. \end{cases}\quad (51)$$

The application of copulas and meta-distributions is a cornerstone of modern quantitative finance, risk management, and many other fields that deal with multivariate dependence. They provide a powerful toolkit for modeling complex relationships beyond simple correlation.

5. Conclusions

The development of statistical distributions, driven by the need to model real data, is a cornerstone of statistical science. With a vast and well-established body of literature, a complete review of construction methods is beyond our scope. This article, instead, focuses on recent advancements in the field, particularly in simulation techniques and computational algorithms. Our aim is to offer readers an expanded perspective on the evolving landscape of statistical distributions.

Author Contributions: Conceptualization, investigation, writing—original draft preparation, resources, supervision, project administration: K.-T.F.; writing—review and editing, visualization: Y.-X.L.; writing—review and editing, funding acquisition: Y.-H.D. All authors have read and agreed to the published version of the manuscript.

Funding: Our work was supported in part by the Guangdong Provincial Key Laboratory of IRADS (grant number 2022B1212010006), in part by National Natural Science Foundation of China (Key Program) (grant number 12231004), and in part by BNBU Research (grant number UICR0600048) at Beijing Normal-Hong Kong Baptist University, Zhuhai, PR China.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1994; Volume 1.
- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1995; Volume 2.
- Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Discrete Multivariate Distributions*; John Wiley & Sons: New York, NY, USA, 1997.
- Kotz, S.; Balakrishnan, N.; Johnson, N.L. *Continuous Multivariate Distributions, Volume 1: Models and Applications*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2000.
- Fang, K.T.; Xu, J.L. *Statistical Distribution*; China Higher Education Press: Beijing, China, 2016.
- Fang, K.T.; Kotz, S.; Ng, K.W. *Symmetric Multivariate and Related Distributions*; Chapman and Hall: London, UK, 1990.
- Johnson, M.E. *Multivariate Statistical Simulation*; John Wiley & Sons: New York, NY, USA, 1987.
- Fang, K.T.; Ye, H.; Zhou, Y. *Representative Points of Statistical Distributions and Their Applications in Statistical Inference*; CRC Press: Boca Raton, FL, USA, 2025.
- Lehmann, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.
- Bickel, P.J.; Doksum, K.A. *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd ed.; Chapman and Hall/CRC: New York, NY, USA, 2015; Volume 1.
- Pearson, K.X. Contributions to the mathematical theory of evolution—II. Skew variation in homogeneous material. *Philos. Trans. R. Soc. Lond. A* **1895**, 186, 343–414.
- Elderton, W.P.; Johnson, N.L. *Systems of Frequency Curves*; Cambridge University Press: Cambridge, UK, 1969.
- Hogg, R.V.; McKean, J.W.; Craig, A.T. *Introduction to Mathematical Statistics*, 6th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2005.
- Li, Y.; Sun, Z.H.; Fang, K.T. Generalized inverse transformation method via representative points in statistical simulation. *Commun. Stat.-Simul. Comput.* **2024**, 54, 4519–4545. [\[CrossRef\]](#)
- David, H.A.; Nagaraja, H.N. *Order Statistics*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2004.
- De Haan, L.; Ferreira, A. *Extreme Value Theory: An Introduction*; Springer: New York, NY, USA, 2006.
- Al-Hussaini, E.K.; Ahsanullah, M. Exponentiated distributions. *Atlantis Studies in Probability and Statistics*; Atlantis Press: Paris, France, 2015; Volume 21.
- Mahdavi, A.; Kundu, D. A new method for generating distributions with an application to exponential distribution. *Commun. Stat.-Theory Methods* **2017**, 46, 6543–6557. [\[CrossRef\]](#)
- Cordeiro, G.M.; De Castro, M. A new family of generalized distributions. *J. Stat. Comput. Simul.* **2011**, 81, 883–898. [\[CrossRef\]](#)
- Alzaatreh, A.; Lee, C.; Famoye, F. A new method for generating families of continuous distributions using the T-X family. *Metron* **2013**, 71, 6379. [\[CrossRef\]](#)
- Everitt, B.S.; Hand, D.J. *Finite Mixture Distributions*; Springer: Dordrecht, The Netherlands, 1981.
- Li, Y.; Fang, K.T.; He, P.; Peng, H. Representative points from a mixture of two normal distributions. *Mathematics* **2022**, 10, 3952. [\[CrossRef\]](#)
- Yan, T.; Fang, K.T.; Yin, H. A novel approach for parameter estimation of mixture of two Weibull distributions in failure data modeling. *Stat. Comput.* **2024**, 34, 221. [\[CrossRef\]](#)
- Aigner, D.; Lovell, C.K.; Schmidt, P. Formulation and estimation of stochastic frontier production function models. *J. Econom.* **1977**, 6, 21–37. [\[CrossRef\]](#)
- Huygens, C. De Ratiociniis in Ludo Aleae. In *Exercitationum Mathematicarum*; Schooten, F.V., Ed.; Elsevier: Leiden, The Netherlands, 1657; pp. 521–534.
- de Moivre, A. Approximatio ad Summam Terminorum Binomii $(a + b)^n$ in Seriem Expansi. *Misc. Anal.* **1733**, II, 1–7.
- Gauss, C.F. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum*; Perthes et Besser: Hamburg, Germany, 1809.
- Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, 12, 171–178.
- Martínez-Flórez, G.; Tovar-Falón, R.; Gómez, H. Bivariate Power-Skew-Elliptical Distribution. *Symmetry* **2020**, 12, 1327. [\[CrossRef\]](#)
- Zhou, Y.F.; Lin, Y.X.; Fang, K.T.; Yin, H. The Representative Points of Generalized Alpha Skew- t Distribution and Applications. *Entropy* **2024**, 26, 889. [\[CrossRef\]](#)
- Rayleigh, L. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philos. Mag.* **1880**, 10, 73–78. [\[CrossRef\]](#)
- Nadarajah, S.; Kotz, S. Moments of some J-shaped distributions. *J. Appl. Stat.* **2003**, 30, 311–317. [\[CrossRef\]](#)
- Stacy, E.W. A generalization of the gamma distribution. *Ann. Math. Stat.* **1962**, 33, 1187–1192. [\[CrossRef\]](#)
- Schrödinger, E. Zur theorie der fall-und steigversuche an teilchen mit brownischer bewegung. *Phys. Z.* **1915**, 16, 289–295.
- Tweedie, K.C.M. Inverse statistical variates. *Nature* **1945**, 155, 453. [\[CrossRef\]](#)
- Seshadri, V. *The Inverse Gaussian Distribution: Statistical Theory and Applications*; Springer: New York, NY, USA, 1999.

37. Barndorff-Nielsen, O.E. Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **1977**, *353*, 401–419.
38. Eberlein, E.; Keller, U. Hyperbolic distributions in finance. *Bernoulli* **1995**, *1*, 281–299. [\[CrossRef\]](#)
39. Barndorff-Nielsen, O.E. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* **1997**, *24*, 1–13. [\[CrossRef\]](#)
40. Arnold, B.C.; Castillo, E.; Sarabia, J.M. *Conditional Specification of Statistical Models*; Springer Series in Statistics; Springer: New York, NY, USA, 1999.
41. Arnold, B.C.; Castillo, E.; Sarabia, J.M. Conditionally specified distributions: An introduction (with comments and a rejoinder by the authors). *Stat. Sci.* **2001**, *16*, 249–274. [\[CrossRef\]](#)
42. Zografos, K.; Balakrishnan, N. On families of beta-and generalized gamma-generated distributions and associated inference. *Stat. Methodol.* **2009**, *6*, 344–362. [\[CrossRef\]](#)
43. Eugene, N.; Lee, C.; Famge, F. Beta-normal distribution and its applications. *Commun. Stat.-Theory Methods* **2002**, *31*, 497–512. [\[CrossRef\]](#)
44. Gupta, A.K.; Nadarajah, S. On the moments of the beta normal distribution. *Commun. Stat.-Theory Methods* **2005**, *33*, 1–13. [\[CrossRef\]](#)
45. Famge, F.; Lee, C.; Olumolade, O. The beta-Weibull distribution. *J. Stat. Theory Appl.* **2005**, *4*, 121–136.
46. Mdziniso, N.C. The Quotient of the Beta-Weibull Distribution. Master's Thesis, Marshall University, Huntington, WV, USA, 2012.
47. Johnson, N.L.; Kotz, S. *Urn Models and Their Applications*; John Wiley & Sons: New York, USA, NY, 1977.
48. Johnson, N.L.; Kemp, A.W.; Kotz, S. *Univariate Discrete Distributions*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 2005.
49. Ross, S.M. *Introduction to Probability Models*, 11th ed.; Academic Press: Cambridge, MA, USA, 2014.
50. Fang, K.T.; Pan, J. A review of representative points of statistical distributions and their applications. *Mathematics* **2023**, *11*, 2930. [\[CrossRef\]](#)
51. Chakraborty, S. Generating discrete analogues of continuous probability distributions—A survey of methods and constructions. *J. Stat. Distrib. Appl.* **2015**, *2*, 6. [\[CrossRef\]](#)
52. Kemp, A.W. A wide class of discrete distributions and the associated differential equations. *Sankhyā Indian J. Stat. Ser. A* **1968**, *30*, 401–410.
53. Sundt, B.; Jewell, W.S. Further results on recursive evaluation of compound distributions. *ASTIN Bull. J. IAA* **1981**, *12*, 27–39. [\[CrossRef\]](#)
54. Lisman, J.H.C.; Van Zuylen, M.C.A. Note on the generation of most probable frequency distributions. *Stat. Neerl.* **1972**, *26*, 19–23. [\[CrossRef\]](#)
55. Inusah, S.; Kozubowski, T.J. A discrete analogue of the Laplace distribution. *J. Stat. Plan. Inference* **2006**, *136*, 1090–1102. [\[CrossRef\]](#)
56. Kozubowski, T.J.; Inusah, S. A skew Laplace distribution on integers. *Ann. Inst. Stat. Math.* **2006**, *58*, 555–571. [\[CrossRef\]](#)
57. Roy, D. The discrete normal distribution. *Commun. Stat.-Theory Methods* **2003**, *32*, 1871–1883. [\[CrossRef\]](#)
58. Roy, D. Discrete Rayleigh distribution. *IEEE Trans. Reliab.* **2004**, *53*, 255–260. [\[CrossRef\]](#)
59. Krishna, H.; Pundir, P.S. Discrete Burr and discrete Pareto distributions. *Stat. Methodol.* **2009**, *6*, 177–188. [\[CrossRef\]](#)
60. Chakraborty, S.; Chakravarty, D. Discrete gamma distributions: Properties and parameter estimations. *Commun. Stat.-Theory Methods* **2012**, *41*, 3301–3324. [\[CrossRef\]](#)
61. Karlis, D.; Mamode Khan, N. Models for integer data. *Annu. Rev. Stat. Its Appl.* **2023**, *10*, 297–323. [\[CrossRef\]](#)
62. Janardan, K.G.; Patil, G.P. A unified approach for a class of multivariate hypergeometric models. *Sankhyā Indian J. Stat. Ser. A* **1972**, *34*, 363–376.
63. Kocherlakota, S.; Kocherlakota, K. *Bivariate Discrete Distributions*; Marcel Dekker: New York, USA, NY, 1992.
64. Karlis, D. An EM algorithm for multivariate Poisson distribution and related models. *J. Appl. Stat.* **2003**, *30*, 63–77. [\[CrossRef\]](#)
65. Palmitesta, P.; Provasi, C. *Computer Generation of Random Vectors from Continuous Multivariate Distributions*; Dipartimento di Metodi Quantitativi, Università degli Studi di Siena: Siena, Italy, 2001.
66. Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: New York, NY, USA, 2006.
67. Balakrishna, N.; Lai, C.D. Construction of bivariate distributions. In *Continuous Bivariate Distributions*, 2nd ed.; Springer: New York, NY, USA, 2009; pp. 179–228.
68. Fang, K.T.; Zhang, Y.T. *Generalized Multivariate Analysis*; Science Press: Beijing, China; Springer: Berlin/Heidelberg, Germany, 1990.
69. Fang, K.T.; Wang, Y. *Number-Theoretic Methods in Statistics*; Chapman & Hall: London, UK, 1994.
70. Pagès, G. *Numerical Probability: An Introduction with Applications to Finance*; Universitext Series; Springer: Berlin/Heidelberg, Germany, 2018.
71. Fang, K.T.; Yuan, K.H.; Benlter, P.M. Applications of number-theoretic methods to quantizers of elliptically contoured distributions, Multivariate Analysis and Its Applications. *IMS Lect. Notes-Monogr. Ser.* **1994**, *24*, 211–225.
72. Tarpey, T.; Flury, B. Self-consistency: A fundamental concept in statistics. *Stat. Sci.* **1996**, *11*, 229–243. [\[CrossRef\]](#)

73. Graf, S.; Luschgy, H. *Foundations of Quantization for Probability Distributions*; Springer: Berlin/Heidelberg, Germany, 2000.
74. Wishart, J. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **1928**, *20*, 32–52. [\[CrossRef\]](#)
75. Fang, K.T.; Fang, B.Q. Some families of multivariate symmetric distributions related to exponential distribution. *J. Multivar. Anal.* **1988**, *24*, 109–122. [\[CrossRef\]](#)
76. Fang, K.T.; Fang, B.Q. A characterization of multivariate l_1 -norm symmetric distributions. *Stat. Probab. Lett.* **1989**, *7*, 297–299. [\[CrossRef\]](#)
77. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*; John Wiley & Sons: New York, NY, USA, 1984.
78. Wang, J.; Boyer, J.; Genton, M.G. A skew-symmetric representation of multivariate distributions. *Stat. Sin.* **2004**, *14*, 1259–1270.
79. Arad, F.; Sheikhi, A. Asymmetric distributions based on the t -copula. In Proceedings of the 2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Bam, Iran, 2–4 March 2022; pp. 1–3.
80. Azzalini, A.; Valle, A.D. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [\[CrossRef\]](#)
81. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew-normal distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1999**, *61*, 579–602. [\[CrossRef\]](#)
82. Genton, M.G. (Ed.) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*; CRC Press: Boca Raton, FL, USA, 2004.
83. Marshall, A.W.; Olkin, I. A multivariate exponential distribution. *J. Am. Stat. Assoc.* **1967**, *62*, 30–44. [\[CrossRef\]](#)
84. Durante, F.; Sempi, C. Chapter 1: Copula theory: An introduction. In *Copula Theory and Its Applications, Proceedings of the Workshop Held in Warsaw, Poland, 25–26 September 2009*; Lecture Notes in Statistics—Proceedings, 198; Springer: Berlin/Heidelberg, Germany, 2010; pp. 3–32.
85. Gumbel, E.J. Multivariate logistic distributions. *Bull. Int. Stat. Inst.* **1961**, *38*, 1–6.
86. Fang, K.T.; Fang, H.B.; von Rosen, D. A family of bivariate distributions with non-elliptical contours. *Commun. Stat. Theory Methods* **2000**, *29*, 1885–1898. [\[CrossRef\]](#)
87. Clayton, U.G. A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **1978**, *65*, 141–151. [\[CrossRef\]](#)
88. Ali, M.M.; Mikhail, N.N.; Haq, M.S. A class of bivariate distributions including the bivariate logistic. *J. Multivar. Anal.* **1978**, *8*, 405–412. [\[CrossRef\]](#)
89. Fang, H.B.; Fang, K.T.; Kotz, S. The meta-elliptical distributions with given marginalis. *J. Multivar. Anal.* **2002**, *82*, 1–16. [\[CrossRef\]](#)
90. Kelly, K.S.; Krzysztofowicz, R. Bivariate meta-gaussian density for use in hydrology. *Stoch. Hydrol. Hydraul.* **1997**, *11*, 17–31. [\[CrossRef\]](#)
91. Joe, H. *Dependence Modeling with Copulas*; CRC Press: London, UK, 2015.
92. Bouyé, E.; Durrleman, V.; Nikeghbali, A.; Riboulet, G.; Roncalli, T. Copulas for Finance—A Reading Guide and Some Applications. Available online: <https://ssrn.com/abstract=1032533> (accessed on 1 October 2025).
93. Kruskal, W.H. Ordinal measures of association. *J. Am. Stat. Assoc.* **1958**, *53*, 814–861. [\[CrossRef\]](#)
94. Mai, J.F.; Scherer, M. *Simulating Copulas: Stochastic Models, Sampling Algorithms and Applications*, 2nd ed.; World Scientific: Hackensack, NJ, USA, 2017.
95. Zhang, L.; Singh, V.P. *Copulas and Their Applications in Water Resources Engineering*; Cambridge University Press: Cambridge, UK, 2019.
96. He, P.A.; Fang, K.T.; Ye, H.J. Gaussian copulas and meta Gaussian distributions: A theoretical study. *Stat. Comput.* **2025**, *in press*.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.