



Contaminated Normal Distribution with Discrete Random Variable

Sudarno Sudarno¹, Deby Fakhriyana²

^{1,2}Department of Statistics, Diponegoro University, Semarang, Indonesia

ABSTRACT: The contaminated normal distribution is the distribution of the normal random variable mixed with other random variables. The combination of the other random variables can be either the discrete random variable or the continuous random variable. The normal random variable is the main random variable, while the other random variable is follower random variables. The purpose of this study is to determine the cumulative distribution function and the probability density function of the contaminated standard normal random variable. In addition, it also seeks the properties of the contaminated standard normal random variable. The contaminated normal discusses random variables constructed by the standard normal random variable and the discrete random variable. The contaminated normal is expanded in the form of the linear contaminated normal. The mean of the contaminated normal is zero, while the variance of the contaminated normal is a constant. If the contaminated normal random variable is a linear function, then the mean of the linear contaminated normal is equal intercept, but the variance of the linear contaminated normal is a function of the square of the slope. The properties of the contaminated standard normal random variable are the center point of the data at the mean, the variance is a constant, the shape of the curve is bell-shaped, the data distribution is symmetric, and the axis of symmetry is at the mean of the contaminated standard normal random variable.

KEY WORDS: Contaminated Normal Distribution, Mixture Distribution, Bell-shaped, Symmetric

1. INTRODUCTION

The continuous random variable distribution frequently used in statistics is the normal or Gaussian distribution. The normal distribution is a probability function that shows the symmetrical distribution of a variable around its mean. This function is visualized as a bell-shaped graph. In this graph, data clusters around the mean and gradually decreases towards both sides, creating a balanced and symmetrical pattern [1-2].

The contaminated normal distribution is the distribution of a normal random variable augmented by another random variable. The additional random variable can be a discrete random variable or a continuous random variable. In this case, the normal random variable is the primary random variable, while the other random variables are follower random variables [1]. Several definitions of the normal mixture model. A normal mixture model is a probabilistic model that represents data as a combination of several normal distributions, each with its own mean and variance, weighted by a mixing coefficient. The normal mixture models are commonly used for clustering and density estimation, since they can capture complex, multimodal distributions where data points may naturally group around different centers rather than a single mean [4-5]. The normal mixture model is defined as a statistical method that represents a mixture of multiple Gaussian distributions, utilizing a weighted sum of their probability density functions. It involves an iterative procedure based on the Expectation-Maximization algorithm to estimate the parameters and determine the optimal number of Gaussian components [7]. Scientifically, an expansion of the contaminated normal distribution is performed. The random variable expansion takes the form of the linear contaminated normal.

Existing research on the contaminated normal distribution is that the use of Gaussian mixtures models can be compromised in the presence of mild outliers and groups with heavy tails. A mixture of contaminated normal distribution represents a mixture of two normal components with a common location parameter and one scale parameter being the multiple of the other one. The component with inflated covariance can help model potentially heavy distribution tails. Furthermore, the use of contaminated normal components is to detect mild outliers [3]. The purpose of this article is to determine the cumulative distribution function and the probability density function of the random variable of the contaminated standard normal. It also seeks the properties of the distribution of the contaminated standard normal random variable.

Cite the Article: Sudarno, S., Fakhriyana, D. (2026). Contaminated Normal Distribution with Discrete Random Variable. Current Science Research Bulletin, 3(10), 221-227. <https://doi.org/10.55677/csr/b/01-V03I10Y2026>

Publication Date: October 12, 2026

Corresponding Author: Sudarno Sudarno

221

License: This is an open access article under the CC BY 4.0 license: <https://creativecommons.org/licenses/by/4.0/>

II. LITERATURE REVIEW

The literature underlying this article consists of the random variable and the normal distribution. The random variable consists of the discrete random variable and the continuous random variable [1-2]. Meanwhile, the normal distribution consists of the standard normal random variable and the normal random variable [1,6].

2.1. The Random Variable

The Discrete Random Variable

The random variable is a discrete random variable if its space D is either finite or countable. Let X be a discrete random variable with space D . The probability mass function of the random variable X is given by

$$p_X(x) = P(X = x), \text{ for } x \in D.$$

Let Ω be a sample space and $E \subseteq \Omega$ be an event. The indicator function of E , denoted by I_E , is the random variable defined as

$$I_{E(\omega)} = \begin{cases} 1, & \text{if } \omega \in E \\ 0, & \text{if } \omega \notin E \end{cases}$$

The indicator function is a discrete random variable. Because space $D = \{0,1\}$ is finite.

The Continuous Random Variable

The random variable is a continuous random variable if its cumulative distribution function $F_X(x)$ is a continuous function for all $x \in R$. The cumulative distribution function $F_X(x)$ is

$$F_X(x) = \int_{-\infty}^x f_X(t) dt,$$

where $f_X(t)$ is the probability density function of the random variable X . If $f_X(x)$ is continuous, then implies that

$$\frac{d}{dx} F_X(x) = f_X(x).$$

2.2. The Normal Distribution

The standard normal random variable

A random variable Z has a standard normal distribution if its pdf is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad -\infty < z < \infty.$$

The moment generating function of the random variable Z is

$$M_Z(t) = \exp\left(\frac{1}{2}t^2\right), \quad -\infty < t < \infty.$$

The first and second derivatives of the moment generating function of the random variable Z with respect to t are, respectively,

$$M'_Z(t) = t \exp\left(\frac{1}{2}t^2\right), \quad -\infty < t < \infty. \quad (1)$$

and

$$M''_Z(t) = \exp\left(\frac{1}{2}t^2\right) + t^2 \exp\left(\frac{1}{2}t^2\right), \quad -\infty < t < \infty. \quad (2)$$

Therefore, based on Equations (1) and (2) we can determine the mean and the variance of the random variable Z , that is

The mean of Z is

$$\mu_Z = E(Z) = M'_Z(0) = t \exp\left(\frac{1}{2}t^2\right) \Big|_{t=0} = 0. \quad (3)$$

The variance of Z is

$$\sigma_Z^2 = Var(Z) = M''_Z(0) - (M'_Z(0))^2 = \exp\left(\frac{1}{2}t^2\right) + t^2 \exp\left(\frac{1}{2}t^2\right) \Big|_{t=0} = 1. \quad (4)$$

To make the distribution of the random variable Z easier to remember, it is abbreviated with the notation $Z \sim N(0,1)$. The letter N indicates the normal distribution, the value 0 indicates the mean, and the value 1 represents the variance of the random variable Z .

The Normal Random Variable

If we define the new random variable X , that is $X = bZ + a$, where a, b be constants and $b > 0$. This is a one-to-one transformation. The inverse of the transformation function is $Z = (X - a) / b$ and the Jacobian is $J = b^{-1}$. Therefore, the probability density function of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}b} \exp\left\{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right\}, \quad -\infty < x < \infty.$$

Based on Equations (3) and (4) the random variable X will have the mean and the variance are, respectively,

$$\mu_X = E(X) = E(bZ + a) = a$$

and

$$\sigma_X^2 = \text{Var}(X) = \text{Var}(bZ + a) = b^2.$$

The shorthand for the distribution of the random variable X is written with the notation $X \sim N(a, b^2)$. The letter N indicates the normal distribution, the value a indicates the mean, while the value b^2 is the variance of the random variable X .

The following will calculate the mean and variance values for several constants a and b as in Table 1.

Table 1. The mean and variance of the random variable $X \sim N(a, b^2)$

a	b	$E(X)$	$\text{Var}(X)$	$N(a, b^2)$
0	1	0	1	$N(0,1)$
5	2	5	4	$N(5,4)$
10	3	10	9	$N(10,9)$
15	4	15	16	$N(15,16)$
20	5	20	25	$N(20,25)$

In Table 1, it can be seen that the mean of the random variable X is equal to constant a , while the variance of the random variable X is equal to b^2 .

If we draw for $X \sim N(10,9)$, then it will look like Figure 1 as follow.

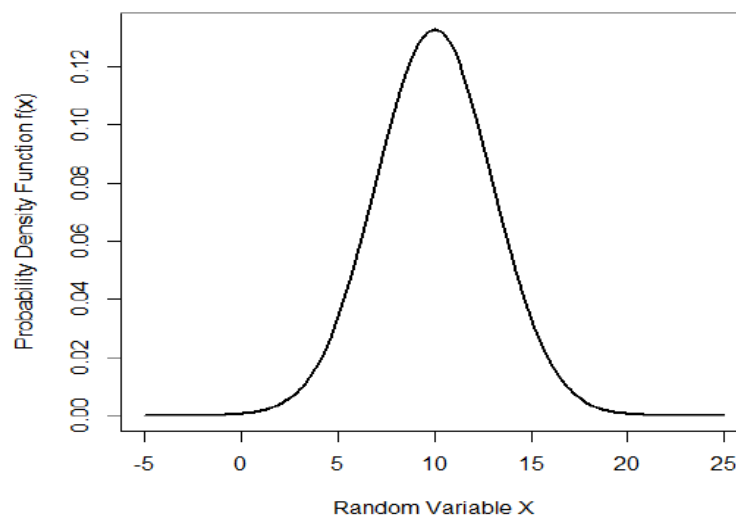


Figure 1. The standard normal density of $X \sim N(10,9)$

The graph $X \sim N(10,9)$ has a center at 10, and the standard deviation is 3. The shape of the graph is symmetrical like a bell shape.

III. RESULT AND DISCUSSION

The contaminated normal discusses random variables constructed from the standard normal random variable and the discrete random variable. It also extends to the linear contaminated normal. We want to find the mean, variance, probability density function, and cumulative distribution function of this random variable.

3.1. Contaminated Normal

A contaminated normal is a random variable whose distribution is a mixture of normal distributions. The mixture normal distribution can be mixed with the distribution of discrete or continuous random variables. The following discusses the contaminated normal between the random variable with a standard normal distribution and the discrete random variable.

Let $Z \sim N(0,1)$ and $I_{1-\varepsilon}$ be a discrete random variable defined by

$$I_{1-\varepsilon} = \begin{cases} 1 & \text{with probability } 1-\varepsilon \\ 0 & \text{with probability } \varepsilon, \end{cases}$$

The random variable $I_{1-\varepsilon}$ has a probability mass function of discrete type, with the value is $\{0,1\}$. Assumed that Z and $I_{1-\varepsilon}$ are independent. Let us define the random variable $W = ZI_{1-\varepsilon} + \sigma_c Z(1-I_{1-\varepsilon})$, where $Z \sim N(0,1)$, σ_c , and ε , $0 \leq \varepsilon \leq 1$ be constants. We'll determine the cumulative distribution function and the probability density function of the random variable W . Since Z and $I_{1-\varepsilon}$ are independent imply that the cumulative distribution function of the random variable W is

$$\begin{aligned} F_W(w) &= P(W \leq w) = P(W \leq w, I_{1-\varepsilon} = 1) + P(W \leq w, I_{1-\varepsilon} = 0) \\ &= P(W \leq w / I_{1-\varepsilon} = 1)P(I_{1-\varepsilon} = 1) + P(W \leq w / I_{1-\varepsilon} = 0)P(I_{1-\varepsilon} = 0) \\ &= P(Z \leq w)(1-\varepsilon) + P(Z \leq w / \sigma_c)\varepsilon = \Phi(w)(1-\varepsilon) + \Phi(w / \sigma_c)\varepsilon \end{aligned} \quad (5)$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Therefore, the distribution of the random variable W is a mixture of normal distribution. Next, the probability density function of the random variable W will be determined. To get the probability density function of the random variable W by differentiating the cumulative distribution function $F_W(w) = \Phi(w)(1-\varepsilon) + \Phi(w / \sigma_c)\varepsilon$ with respect to w . The probability density function of the random variable W is

$$f_W(w) = F'_W(w) = \phi(w)(1-\varepsilon) + \phi(w / \sigma_c)\frac{\varepsilon}{\sigma_c}$$

where $\phi(\cdot)$ is the probability density function of a standard normal distribution.

Next, we will look for the mean and variance of the random variable W . Because $W = ZI_{1-\varepsilon} + \sigma_c Z(1-I_{1-\varepsilon})$, then

$$E(W) = E(ZI_{1-\varepsilon} + \sigma_c Z(1-I_{1-\varepsilon})) = 0. \quad (6)$$

Since based on Equation 3 that $E(Z) = 0$.

$$Var(W) = Var(ZI_{1-\varepsilon} + \sigma_c Z(1-I_{1-\varepsilon})) = 1 + \varepsilon(\sigma_c^2 - 1). \quad (7)$$

Since based on Equation 4 that $Var(Z) = 1$ and Z and $I_{1-\varepsilon}$ are independent so $Cov(Z, I_{1-\varepsilon}) = 0$. Therefore, the mean of the random variable W is located at the origin, whereas the variance of the random variable W is influenced by the values of ε and σ_c^2 .

To determine the distribution of the random variable W to the probability density function and the cumulative distribution function, simulations were carried out using several values W . In this case, the values $\varepsilon = 0.20$ and $\sigma_c^2 = 225$ are selected as in Table 2 below.

Table 2. The probability density function and cumulative distribution function of random variable W

w	$f_W(w)$	$F_W(w)$
-5	0.005032966	0.0738885
-3	0.008759381	0.08522798
-1	0.198884	0.2216089
0	0.3244731	0.5
1	0.198884	0.7783911
3	0.008759381	0.914772
5	0.005032966	0.9261115

In Table 2, by taking value w symmetrical to the center at $w = 0$, the probability density function is also symmetrical, while the cumulative distribution function is a monotonic increasing sequence. This is displayed in the form of a picture as in Figure 2 as follows.

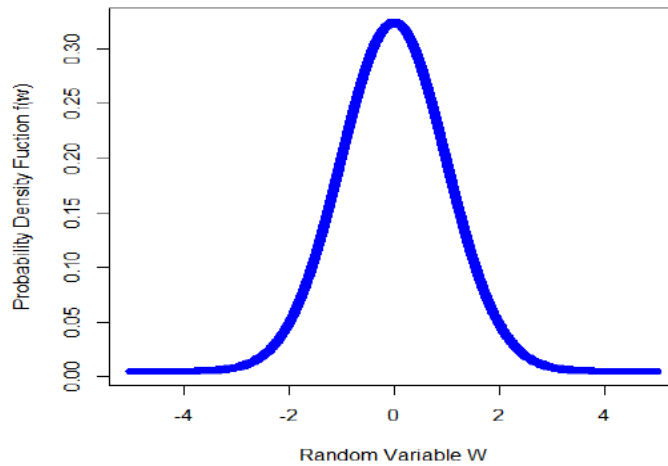


Figure 2. The standard normal density of the random variable W

The probability density function of the random variable W is a bell shape, symmetric with center at $w=0$. The highest point is $(0, 0.3244731)$ and its cumulative distribution function is $\Phi(0) = P(W \leq 0) = 0.5$. This is a half of the whole area, which is one.

3.2. The Linear Contaminated Normal

If we form the new random variable, that is linear contaminated normal $Y = bW + a$, where a, b be constants, and $b > 0$. We want to find the cumulative distribution function and the probability density function of the random variable Y . Based on Equation 5, then the cumulative distribution function of the random variable Y is

$$F_Y(y) = \Phi\left(\frac{y-a}{b}\right)(1-\varepsilon) + \Phi\left(\frac{y-a}{b\sigma_c}\right)\varepsilon. \quad (8)$$

This cumulative distribution function is a mixture of normal cumulative distribution functions. Meanwhile, to obtain the probability density function of the random variable Y , Equation 8 needs to be differentiated with respect to the random variable Y . We be

known that $F_Y(y) = \Phi\left(\frac{y-a}{b}\right)(1-\varepsilon) + \Phi\left(\frac{y-a}{b\sigma_c}\right)\varepsilon$. Then the probability density function of the random variable Y is

$$f_Y(y) = F_Y'(y) = \phi\left(\frac{y-a}{b}\right)(1-\varepsilon)/b + \phi\left(\frac{y-a}{b\sigma_c}\right)\varepsilon/(b\sigma_c).$$

If we use transformation method, the probability density function of the random variable W is

$$f_W(w) = F_W'(w) = \phi(w)(1-\varepsilon) + \phi(w/\sigma_c)\frac{\varepsilon}{\sigma_c}.$$

The random variable $Y = bW + a$ has the transformation function $w = (y-a)/b$ and Jacobian $J = b^{-1}$. Therefore, the probability density function of the random variable Y is

$$f_Y(y) = F_Y'(y) = b^{-1} \left(\phi\left(\frac{y-a}{b}\right)(1-\varepsilon) + \phi\left(\frac{y-a}{b\sigma_c}\right)(\varepsilon/\sigma_c) \right) = \phi\left(\frac{y-a}{b}\right)(1-\varepsilon)/b + \phi\left(\frac{y-a}{b\sigma_c}\right)\varepsilon/(b\sigma_c)$$

where $\phi(\cdot) = \Phi'(\cdot)$ is the probability density function of the standard normal random variable.

The mean and the variance of the random variable Y are, respectively,

$$E(Y) = E(bW + a) = bE(W) + a = b(0) + a = a.$$

Since in Equation 6, $E(W) = 0$.

$$\text{Var}(Y) = \text{Var}(bW + a) = b^2 \text{Var}(W) = b^2(1 + \varepsilon(\sigma_c^2 - 1)).$$

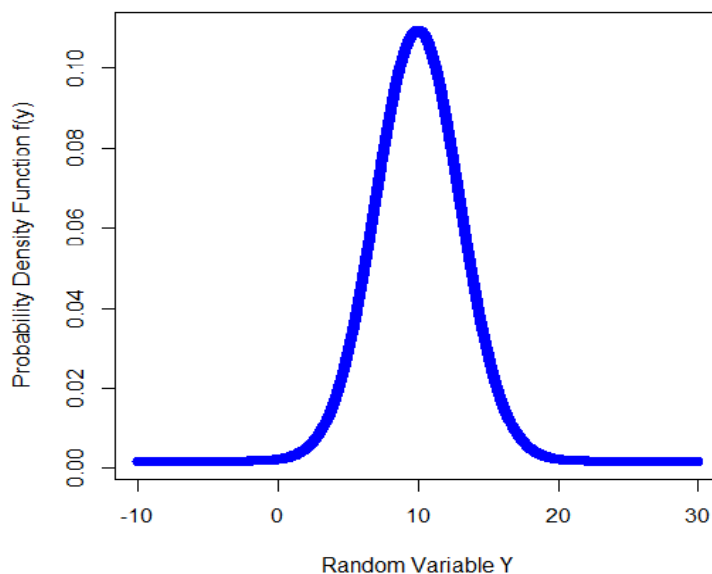
Because in Equation 7, $\text{Var}(W) = 1 + \varepsilon(\sigma_c^2 - 1)$.

To find out the distribution of random variable Y in the probability density function and the cumulative distribution function, it is necessary to carry out simulations with several values of the random variable W as in Table 3.

Table 3. The probability density function and cumulative distribution function of random variable Y

a	b	w	$f_Y(y)$	$F_Y(y)$
0	1	-10	0.004259307	0.05049851
5	2	-5	0.002516483	0.0738885
10	3	0	0.1081577	0.5
15	4	5	0.001258241	0.9261115
20	5	10	0.0008518613	0.9495015

In Table 3, for arbitrary values of a and b , assuming the value w is symmetric to the data center at $w=0$, the probability density function is also symmetric, while the cumulative distribution function is a monotonic increasing sequence. If we want to understand the graph of $f_Y(y)$, for $Y=3W+10$. We can simulate it by assuming $\varepsilon=0.20$ and $\sigma_c=15$. The graph is shown in Figure 3 below.

Figure 3. The standard normal density of the random variable Y

The probability density function of the random variable Y is bell-shaped, symmetric with the data center at $y=10$. The highest point is $(10, 0.1081577)$ and the cumulative distribution function $\Phi(10) = P(W \leq 10) = 0.5$. Because the shape of the curve is bell-shaped and symmetric, the axis of symmetry lies on the line $y=10$.

IV. CONCLUSION

The contaminated normal distribution is the distribution of the standard normal random variable mixed with the distribution of another random variable. The other random variable can be a discrete random variable or a continuous random variable. Because the primary random variable is a standard normal random variable, the properties of the standard normal random variable influence the new random variable. The properties of the contaminated standard normal random variable are: a. The center point of the data is at the mean. b. The variance is constant. c. The shape of the curve is bell-shaped. d. The data distribution is symmetric. e. The axis of symmetry is at $y=\mu$, where μ is the mean of the contaminated standard normal random variable.

REFERENCES

1. Hogg, R.V., McKean, J.W., and Craig, A.T., 2019. Introduction to Mathematical Statistics, Eighth Edition, Pearson Education, Inc., USA.
2. Larsen, R.J., and Marx, M.L., 2018. *An Introduction to Mathematical Statistics and Its Applications*, Sixth Edition, Pearson Education, Inc., USA.
3. Melnykov, Y., 2026. *On the Use of Contaminated Normal Distributions for Modeling Data Groups with Heavy Tails and Outliers*, Journal of Classification. Volume 43, pages 66–85. doi.org/10.1007/s00357-025-09518-1.
4. Noble, J., 2025. Gaussian Mixture Model. <https://www.ibm.com/think/topics/gaussian-mixture-model>.

5. Reynolds, D., 2018. Gaussian Mixture Models. Published in Encyclopedia of Biometrics. DOI:[10.1007/978-0-387-73003-5_196](https://doi.org/10.1007/978-0-387-73003-5_196).
6. Sugiyama, M., 2016. *Introduction to Statistical Machine Learning*, Elsevier, Inc., London.
7. Sciencedirect, 2026. Gaussian Mixture Model. <https://www.sciencedirect.com/topics/engineering/gaussian-mixture-model>.