Mathematics Letters
2024; 10(1): 7-11
http://www.sciencepublishinggroup.com/j/ml
doi: 10.11648/ml.20241001.12
ISSN: 2575-503X (Print); ISSN: 2575-5056 (Online)

Statistical Properties of Points Between Two Random Points

Reza Habibi

Banking Department, Iran Banking Institute, Central Bank of Iran, Tehran, Iran

Email address:
Habibi1356@yahoo.com

To cite this article:
Reza Habibi. (2024). Statistical Properties of Points Between Two Random Points. *Mathematics Letters*, *10*(1), 7-11.
https://doi.org/10.11648/ml.20241001.12

Received: December 17, 2023; **Accepted**: January 24, 2024; **Published**: February 5, 2024

Abstract: Important inferences in statistics, economics and finance such as mixture distribution fitting in portfolio management are closely related to finding statistical properties of points between two random points. This problem is studied in the literature; however, accurate and fast approximations and Monte Carlo simulations are not well studied. This paper is concerned to finding these properties such as distribution function and moment generating function of points between two random points are derived. To this end, the random linear transformation technique plays important role. Also, the moment generating function is represented as expectation of random variable indexed by a Poisson variable. This note is useful to propose the Monte Carlo simulation of generating function. Two applications in mixture distribution fitting and properties of weighted averages are given. These two applications have been used in the literature for Bayesian bootstrap, change point analysis, DNA segmentations, where all theoretical results may be applied in these fields, directly. Finally, conclusions are presented.

Keywords: Linear Transformation, Mixture Distribution, Moment Generating Function, Monte Carlo, Random Points

1. Introduction

The application of probability theory in Euclidean geometry has a long history. One of the pioneers in this topic is [15], who proposed geometric probabilities and their applications. In recent years, [13] calculated the average distance between two random points in a n dimensional space. The distribution of this distance in a box is studied by [2].

Let X, Y be two independent random variables from common continuous distribution and density functions F and f with the moment generating function $M(t)$. A point between $(\min(X, Y), \max(X, Y))$ is represented by

$$Z = UX + (1 - U)Y$$

where $U \in (0,1)$ is assumed to be independent of X, Y and has beta distribution with parameters α, β. The assumption that X, Y have same distribution is simplifying assumption and their distributions can be different, without loss of generality. One can see that the k-th non-central moment of Z is given by, see [7]:

$$E(Z^k) = \sum_{j=0}^{k} \binom{k}{j} \frac{beta(j+\alpha, \beta+k-j)}{beta(\alpha, \beta)} \mu_j^* \mu_{k-j}^*,$$

where $\mu_j^* = E(X^j), j = 1,2, \ldots$ is the j- th non-central moment of X and $beta(\alpha, \beta)$ is the beta function. As soon as, U has uniform distribution with $\alpha = \beta = 1$, then

$$E(Z^k) = \frac{1}{k+1} \sum_{j=0}^{k} \mu_j^* \mu_{k-j}^*.$$

The standard proof of the last equation is to use the binomial expansion, however, the other method is to use the moment generating function as follows:

$$M_z(t) = E\left(E\left(e^{UtX} e^{(1-U)tY} \middle| U\right)\right) = \int_0^1 M(tu)\, M\big(t(1-u)\big) du.$$

Notice that using the Taylor expansion, see [3] and we have

$$M(tu) = \sum_{j=0}^{\infty} \mu_j^* u^j \frac{t^j}{j!} \text{ and}$$

$$M(t(1-u)) = \sum_{j=0}^{\infty} \mu_j^* (1-u)^j \frac{t^j}{j!}.$$

Therefore,

$$M(tu)M\big(t(1-u)\big)=\sum_{j=0}^{\infty}c_j^*(u)\frac{t^j}{j!},$$

at which, we have

$$c_j^*(u)=\sum_{r=0}^{j}\binom{r}{j}u^j(1-u)^{r-j}\mu_j^*\mu_{r-j}^*.$$

One can notice that

$$\int_0^1 M_X(tu)\,M_X\big(t(1-u)\big)du=\sum_{j=0}^{\infty}\int_0^1 c_j^*(u)du\frac{t^j}{j!}$$

and it is understood that

$$\int_0^1 c_j^*(u)du=\frac{1}{k+1}\sum_{j=0}^{k}\mu_j^*\mu_{k-j}^*\coloneqq d_k^*$$

This completes the proof.

Remark 1. When X comes from a gamma distribution with parameters α,θ, then $\mu_j^*=\frac{\Gamma(\alpha+j)}{\Gamma(\alpha)}\theta^j$. Therefore,

$$d_k^*=\frac{\sum_{j=0}^{k}\Gamma(\alpha+j)\Gamma(\alpha+k-j)}{(k+1)\Gamma^2(\alpha)}\theta^k,$$

see [5], where $\Gamma(\alpha)$ is the value of gamma function in the α. For the case of X has normal standard distribution, then $\frac{1}{k+1}\sum_{j=0}^{k}\mu_j^*\mu_{k-j}^*$ is zero if k is an odd number.

The next section derives alternative formulas for $M_z(t)$ which is useful for Monte Carlo simulation of $M_z(t)$.

2. Alternative Formulas

Let $\mu=E(X)=E(Y)$ then $\mu=E(Z)$. Then,

$$M_{z-\mu}(t)=e^{-\mu t}M_z(t)=\sum_{j=0}^{\infty}d_j\frac{t^j}{j!}$$

where $d_j=E(Z-\mu)^j$ is the $j-th$ central moment of Z. One can see that

$$Z-\mu=U(X-\mu)+(1-U)(Y-\mu).$$

By applying the above arguments, it is seen that, see [6]:

$$M_{z-\mu}(t)=\sum_{j=0}^{\infty}\frac{\sum_{r=0}^{j}\mu_r\mu_{j-r}}{j+1}\frac{t^j}{j!}.$$

where μ_r is $r-th$ central moment of X. Thus

$$d_j=\frac{1}{j+1}\sum_{r=0}^{j}\mu_r\mu_{j-r}.$$

Hence,

$$e^{-t}M_z(t)=e^{\mu t}\sum_{j=0}^{\infty}d_je^{-t}\frac{t^j}{j!}=e^{\mu t}E(d_N),$$

where N has Poisson distribution with parameter t and independent of Z. The last equation is changed to

$$M_z(t)=e^{(\mu-1)t}E(d_N),$$

which is useful for Monte Carlo simulation of $M_z(t)$ by simulating $E(d_N)$. The following proposition summarizes the above discussion.

Function $\mathrm{erfi}(x)$ has an series approximation up to third term as $\mathrm{erfi}(x)=\pi^{-0.5}x\left\{2+\frac{2x^2}{3}+\frac{x^4}{5}\right\}$, see [4].

Proposition 1. The moment generating functions of $Z=UX+(1-U)Y$ is given by

$$M_z(t)=e^{(\mu-1)t}E(d_N),$$

where $d_j=\frac{1}{j+1}\sum_{r=0}^{j}\mu_r\mu_{j-r}$ at which $\mu=E(X)$, μ_r is $r-th$ central moment of X and where N has Poisson distribution with parameter t and independent of Z. Function $M_z(t)$ is estimated using the Monte Carlo method.

Remark 2. It is easy to see that

$$\frac{1}{j+1}\sum_{r=0}^{j}\mu_r\mu_{j-r}=\sum_{r=0}^{j}E(X^j)\,E(Y^{r-j})=$$
$$\sum_{r=0}^{j}E(X^j)\,E(Y^{r-j})=\sum_{r=0}^{j}E(X^jY^{r-j})=$$
$$E\left\{\sum_{r=0}^{j}X^jY^{r-j}\right\}=E\{\frac{X^{k+1}-Y^{k+1}}{X-Y}\}.$$

This relation is useful in running Monte Carlo simulation of $M_z(t)$, easily.

Remark 3. Using the above equation, it is seen that

$$M_z(t)=E\left\{\frac{\sum_{j=0}^{\infty}\frac{t^j}{j!}\frac{\left(X^{j+1}-Y^{j+1}\right)}{(j+1)}}{X-Y}\right\}=\frac{1}{t}E\left\{\frac{e^{tX}-e^{tY}}{X-Y}\right\}.$$

This relation gives another representation for Monte Carlo simulation of $M_z(t)$. Monte Carlo simulation lets you see all possible outcomes of your decisions, including the actual probabilities each will occur, by running simulations with random variables thousands of times. These variables are described by them probability distribution which can be estimated with historical data or defined using expert opinion. Then, one can run sensitivity analysis to identify which variables have the largest impact on the outcome. This method lets you quantitatively assess the impact of risk, allowing for more accurate forecasting and, ultimately, better decision-making under uncertainty, see [10].

Remark 4. Another method to use the Monte Carlo simulation is to apply the relation

$$M_z(t)=e^{\frac{t^2}{4}}\left\{1+\frac{t^2}{12}+\frac{t^4}{160}\right\}.$$

$$M_z(t)=E\{M(tU)-M\big(t(1-U)\big)\}=$$
$$\int_0^1 M(tu)M\big(t(1-u)\big)du.$$

However, this relation is useful to obtain a closed form for the $M_z(t)$, in some special cases. For example, let X be standard normal random variable. Then,

$$M_z(t) = \int_0^1 e^{\frac{t^2u^2}{2}} e^{\frac{t^2(1-u)^2}{2}} du = e^{\frac{t^2}{4}} \int_0^1 e^{t^2(0.5-u)^2} du = 2e^{\frac{t^2}{4}} \int_0^{0.5} e^{t^2u^2} du = 2t^{-1} e^{\frac{t^2}{4}} \int_0^{\frac{t}{2}} e^{u^2} du.$$

Notice that $\int_0^{\frac{t}{2}} e^{u^2} du = \frac{\sqrt{\pi}}{2}\mathrm{erfi}(\frac{t}{2})$, where $\mathrm{erfi}(x)$ is the imaginary error function with $\mathrm{erfi}(0) = 0$. Thus,

$$M_z(t) = \frac{\sqrt{\pi}}{t} e^{\frac{t^2}{4}} \mathrm{erfi}(\frac{t}{2}).$$

As soon as X has normal distribution with mean μ and variance σ^2, then using $Z = UX + (1-U)Y$, it can be seen that yet μ and σ are the location and scale parameter of distribution of Z. Thus,

$$M_z(t) = \frac{\sqrt{\pi}}{\sigma t} e^{\mu t + \frac{\sigma^2 t^2}{4}} \mathrm{erfi}(\frac{\sigma t}{2}).$$

About the density function, notice that

$$f_Z(z) = \int_0^1 f_{uX+(1-u)Y}(z) du.$$

Using the transformation method, one can see that

$$f_{uX+(1-u)Y}(z) = \int_{-\infty}^{\infty} \frac{f(x)}{1-u} f(\frac{z-ux}{1-u}) dx.$$

Therefore, it is seen that

$$f_Z(z) = \int_0^1 \int_{-\infty}^{\infty} \frac{f(x)}{1-u} f(\frac{z-ux}{1-u}) dx\, du = E\left\{\frac{1}{1-U} f\left(\frac{z-UX}{1-U}\right)\right\}.$$

According to the [9] and [12], equations of $F_Z(z)$ and $f_Z(z)$ are useful relations to simulated these quantities using the Monte Carlo method. The following proposition summarizes the above discussion.

Proposition 2. $F_Z(z)$ and $f_Z(z)$ are given by

$$F_Z(z) = E\left\{F\left(\frac{z-UX}{1-U}\right)\right\}$$

$$f_Z(z) = E\left\{\frac{1}{1-U} f\left(\frac{z-UX}{1-U}\right)\right\}.$$

The inverse inference about U, based on observation $\mathrm{Z = z}$, is done by Bayesian posterior density $\mathrm{f(u|z)}$. A posterior probability, in Bayesian statistics, is the revised or updated probability of an event occurring after taking into consideration new information. The posterior probability is calculated by updating the prior probability using Bayes' theorem. In statistical terms, the posterior probability is the probability of event A occurring given that event B has occurred.

Let the marginal density $\mathrm{f(u) = 1}$ be the uninformative prior. Therefore, $\mathrm{f(u,z)}$ is proportional to $\mathrm{f(z|u)}$. One can see that

$$\mathrm{f(z|u)} = \frac{\mathrm{g(u)}}{\int_0^1 \mathrm{g(u)du}}$$

where

$$\mathrm{g(u)} = \frac{1}{1-\mathrm{u}} \mathrm{E_X}\left\{\mathrm{f_Y}\left(\frac{\mathrm{z-uX}}{1-\mathrm{u}}\right)\right\}.$$

If X has exponential distribution with rate θ, and assuming $\theta t < 1$, then according to the [17], then

$$M_z(t) = \int_0^1 \frac{du}{(1-\theta ut)(1-\theta(1-u)t)} = \frac{1}{(2-\theta t)}\left\{\int_0^1 \frac{du}{(1-\theta ut)} + \int_0^1 \frac{du}{(1-\theta(1-u)t)}\right\} = \frac{-\log(1-\theta t)}{\theta t(1-\theta t)(2-\theta t)}.$$

Regarding the distribution and density functions of Z, notice that

$$F_Z(z) = P(Z \le z) = \int_0^1 P(uX + (1-u)Y \le z) du = E\left\{F\left(\frac{z-UX}{1-U}\right)\right\}.$$

3. Simulations

Here, some simulation results are proposed.

Example 1. Let X be standard normal distribution. The following plot compares the Monte Carlo estimate of

$$M_z(t) = E\{M(tU) - M(t(1-U))\}$$

$$M_z(t) = e^{\frac{t^2}{4}}\left\{1 + \frac{t^2}{12} + \frac{t^4}{160}\right\}.$$

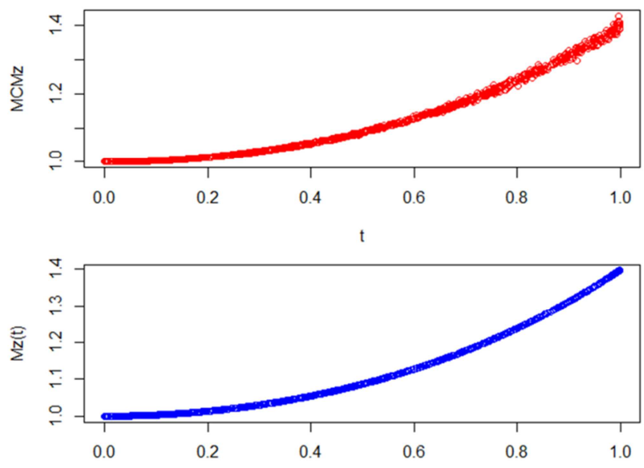

Figure 1. *Exact and Monte Carlo $M_z(t)$: Normal case.*

Example 2. Exponential distribution has many applications in modeling severities in risk banking operational risk management. Suppose that X has exponential distribution with unit rate. Therefore,

$$M_z(t) = \frac{-\log(1-t)}{t(1-t)(2-t)}, 0 < t < 1.$$

Example 3. Gamma distribution is useful distribution for making decision about life time in reliability field. Let X be a gamma random variable with parameters 2,1. Therefore,

$$d_k^* = \frac{\sum_{j=0}^{k} \Gamma(2+j)\Gamma(2+k-j)}{(k+1)}.$$

Some selected values are $d_0^* = 1$, $d_1^* = 2$, $d_2^* = \frac{10}{3}$, …. The following table gives $M_z(t)$ for $t = 0.5(0.25)2$, using relation $M_z(t) = e^{-t}E(d_N^*)$, and Monte Carlo simulation at which N has Poisson distribution with parameter t.

Table 1. *Monte Carlo values of $M_z(t)$: Gamma case.*

t	0.5	0.75	1	1.25	1.5	2
$M_z(t)$	1.84	1.95	2.08	2.13	2.32	2.57

4. Two Applications

The proposed method has two applications, namely in weighted averages and mixture distribution fitting.

(a) Mixture distribution. A different insight to this problem is defining it as a randomized mixture distribution. A mixture distribution is a mixture of two or more probability distributions, see [1]. Random variables are drawn from more than one parent population to create a new distribution. The parent populations can be multivariate, although the mixed distributions should have the same dimensionality. In addition, they should either be all discrete probability distributions or all continuous probability distributions, see [8] and references therein.

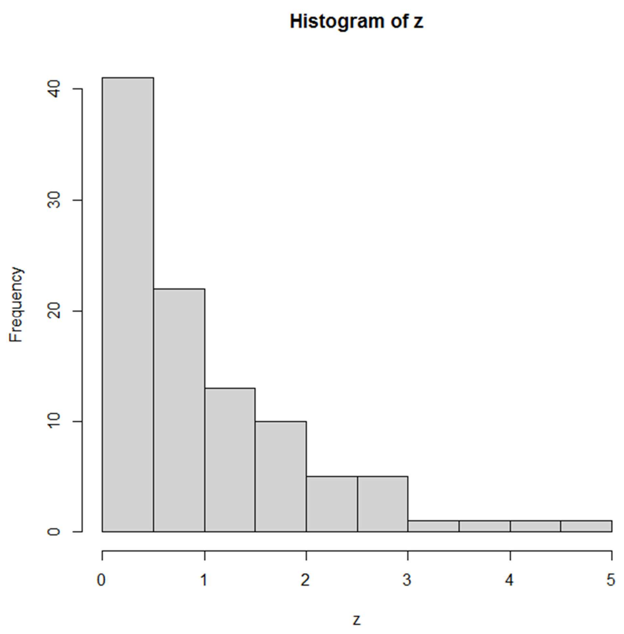

Figure 2. *Histogram of Z.*

Compounding or mixture distributions provide a rich class of models for applications ranging from models of heterogeneity, measurement error, distribution of stock returns and income to models of unemployment duration. Some very general mixtures will be considered which include many new mixture models and also provide a unified method of organizing and comparing previously considered models as well as a test of heterogeneity see [14]. To this end, replace U with I_A which is one if an event A is occurred and zero otherwise. Therefore, Z has a mixture distribution of distributions of X and Y with mixing portion P(A).

Consider a weighted average of stock returns of *Google* and *Amazon* during 15 December 2022 to 15 December 2023. Let A be the event that return of *Google* has increased as 0.01. The following histogram gives the mixture distribution of these returns.

b) Weighted average. The problem also can be viewed as random weighting problem. A weighted average is a calculation that takes into account the varying degrees of importance of the numbers in a data set. In calculating a weighted average, each number in the data set is multiplied by a predetermined weight before the final calculation is made. A weighted average can be more accurate than a simple average in which all numbers in a data set are assigned an identical weight, see [11].

In calculating a simple average, or arithmetic mean, all numbers are treated equally and assigned equal weight. But a weighted average assigns weights that determine in advance the relative importance of each data point. A weighted average is most often computed to equalize the frequency of the values in a data set. However, values in a data set may be weighted for other reasons than the frequency of occurrence. For example, if students in a dance class are graded on skill, attendance, and manners, the grade for skill may be given greater weight than the other factors, see [16].

Weighted averages have applications in portfolio managements. Consider a weighted average of stock returns of *Apple co* and *Amazon* during 15 December 2022 to 15 December 2023. The following figure gives the probability of Z larger than 0.25 for various values of U's.

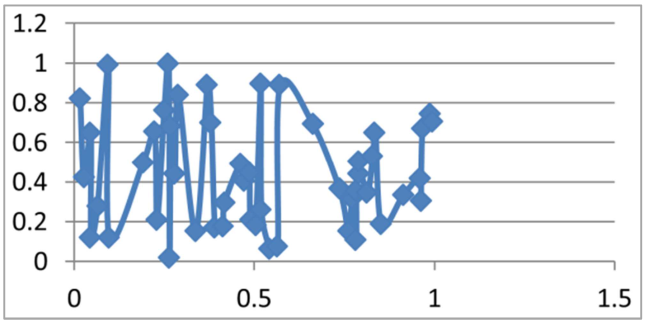

Figure 3. *Probability of larger than 0.25.*

5. Conclusions

In recent years, probability distribution of distance between two random points in a n dimensional space is well studied, see [2]. This note is devoted to this issue. The following conclusions are proposed:

1) Distributions of all random points between two random points are characterized.
2) Some approximate and exact formulas are proposed.
3) Simulation results are proposed.
4) Two applications in mixture distributions and portfolio management are considered.

ORCID

0000-0001-8268-0326

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Amendola, C.; Engstrom, A.; Haase, C. Maximum number of modes of Gaussian mixtures. Information and Inference: A Journal of the IMA 9, 2020, 587–600.

[2] Bailey, D. H., Borwein, J. M., and Crandall, R. E. Box integrals. *Journal of Computational and Applied Mathematics* 206; 2007, 196-208.

[3] Billingsley, P. *Convergence of probability measures*. 2010. Wiley. New York. USA.

[4] Bradley D. M., and Gupta, C. R. On the distribution of the sum of n non-identically distributed uniform random variables. *Annals of Mathematical Statistics* 54. 2002. 689–700.

[5] Csorgo, M. and Horvath, L. *Limit theorems in change point analysis*. 1997. Wiley, New York. USA.

[6] Fay, M. P. Designing Monte Carlo implementations of permutation or bootstrap hypothesis tests. *American Statistician* 56. 2002. 63-70.

[7] Ferguson, T. S. (1996). *A course in large sample theory*. Chapman and Hall. USA.

[8] Fruhwirth-Schnatter, S. *Finite mixture and Markov switching models*. 2006. Springer. USA.

[9] Glasserman, P. and Kim, K. K. (2008). Beta approximations for bridge sampling. *Proceedings of the 2008 Winter Simulation Conference*. USA.

[10] Hill, S. D. and Spall, J. C. Stationarity and Convergence of the Metropolis-Hastings algorithm. *IEEE Control Systems Magazine* 39, 2018. 56–67.

[11] James, F. *Statistical methods in experimental physics*. Singapore: World Scientific. 2006. Singapore.

[12] Lambert, B. *The posterior – the goal of Bayesian inference. A Student's Guide to Bayesian Statistics*. Sage Publication. 2018. USA.

[13] Philip, J. Calculation of expected distance on a unit cube. *www.math.kth.se/~johanph*, Jan 2007.

[14] Ray, R.; Lindsay, B. The topography of multivariate normal mixtures. The Annals of Statistics 33. 2005. 2042–2065.

[15] Solomon, H. *Geometric probability*. Society for Industrial and Applied Mathematics. 1978. USA.

[16] Strutz, T. *Data fitting and uncertainty (A practical introduction to weighted least squares and beyond)*. 2010. Springer. USA.

[17] Zivot, E. and Wang, J. Modeling financial time series with S-PLUS. 2018. Springer. USA.