OKU Fen Bilimleri Enstitüsü Dergisi
7(2): 581-591, 2024

OKU Journal of The Institute of Science and Technology, 7(2): 581-591, 2024

Osmaniye Korkut Ata Üniversitesi
Fen Bilimleri Enstitüsü
Dergisi

Osmaniye Korkut Ata University
Journal of The Institute of Science
and Technology

# İstatistikte Dağılım Fonksiyonu ile Fourier Dönüşümü

**Seçil YALAZ**[1*]

[1]Dicle Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 21280, Diyarbakır

[1]https://orcid.org/0000-0001-7283-9225
*Sorumlu yazar: syalaz@dicle.edu.tr

**Araştırma Makalesi**

***Makale Tarihçesi:***
Geliş tarihi: 25.01.2023
Kabul tarihi:27.10.2023
Online Yayınlanma: 11.03.2024

***Anahtar Kelimeler:***
Fourier dönüşümü
İntegrallenebilme
Yoğunluk fonksiyonu
Dağılım fonksiyonu
Ölçüm hatalı değişkenler
Asimptotik özellikler

**ÖZ**

Karmaşık integral denklemlerini basit cebirsel denklemlere dönüştürme yeteneğine sahip olan ve hem matematikte hem de istatistikte sıkça kullanılan en önemli yöntemlerden biri Fourier dönüşümüdür. Fourier dönüşümü, matematikte belli koşullar altında her fonksiyon için geçerli olmasına rağmen istatistiğin matematikten çok farklı olması nedeniyle bu durum istatistikte daha karmaşık hale gelebilmektedir. İstatistikte her durum için farklı gözlem değerleri yani farklı $x$ler söz konusu iken matematikte her $x$ için bir fonksiyon tanımlanır. İstatistikte fonksiyonlardan ziyade rasgele değişkenlerle ilgilenilmektedir ve ilgilenilen gözlem değerlerinin yoğunluk fonksiyonları da bilinmelidir. Asimptotik özelliklerin incelendiği parametrik olmayan modellerde Fourier dönüşümü kullanıldığı görülmektedir. Hem dağılım hem de yoğunluk fonksiyonu kullanılarak gerçekleştirilebilen Fourier dönüşümünde bilinmeyen veya integrallenebilir olmayan yoğunluk fonksiyonları ya da çok yavaş yakınsama oranı söz konusu olduğunda (asimptotik özellikler düşünüldüğünde) yoğunluk fonksiyonunun kullanılması mümkün olamamaktadır. Böyle durumlarda Fourier dönüşümünün dağılım fonksiyonu ile gerçekleştirilmesi daha uygun olacaktır. Bu çalışmada, Fourier dönüşümünün hangi koşullarda dağılım fonksiyonu ile gerçekleştirilmesinin daha uygun olacağı üzerine öneriler sunulmaktadır.

# Fourier Transform by Distribution Function in Statistics

**Research Article**

***Article History:***
Received: 25.01.2023
Accepted: 27.10.2023
Published online: 11.03.2024

***Keywords:***
Fourier transform
Integrability
Density function
Distribution function
Errors in variables
Asymptotic properties

**ABSTRACT**

The Fourier transform is one of the most important methods, which has the ability to transform complex integral equations into simple algebraic equations and is frequently used in both mathematics and statistics. Although the Fourier transform is valid for every function in mathematics under certain conditions, this situation can become more complicated in statistics because of the fact that statistics is very different from mathematics. While in statistics, different observation values, that is, different $x$ values, are considered for each situation, in mathematics for each $x$ a function is defined. In statistics, random variables are concerned rather than functions, and the density functions of the observed values of interest should also be known. It is seen that the Fourier transform is used in non-parametric models in which asymptotic properties are examined. In the Fourier transform, which can be performed using both distribution and density functions, it is not possible to use the density function when there are unknown or non-integrable density functions or very slow convergence rate (considering asymptotic properties). In such cases, it would be more appropriate to perform the Fourier transform with the distribution function. In this study, suggestions are presented on under which conditions it would be more appropriate to perform the Fourier transform with the distribution function.

**To Cite:** Yalaz S. Fourier Transform by Distribution Function in Statistics. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2024; 7(2): 581-591.

## 1. Introduction

Fourier transform is one of the most important methods, which has the ability to transform complex integral equations into simple algebraic equations and is frequently used in both mathematics and statistics. The lemma developed by German mathematician Georg Frederick Bernhard Riemann, who lived between 1826 and 1866 and put forward by the French mathematician Henri Lebesque, who lived between 1875 and 1941, and also known as the Riemann - Lebesque lemma, said that the Fourier or Laplace transform of a function on $L^1$ goes to infinity.

According to the lemma if $f$ is one dimensional Lebesque metric, $L^1$, integrable on the real coordinate space of dimension $d$, $\mathbb{R}^d$, which means that Lebesque integral of $|f|$ is finite, the Fourier transform of $f$ gives the following notation (Bochner and Chandrasekharan, 1949; URL 2, 2022),

$$\hat{f}(z) := \int_{\mathbb{R}^d} f(x)\, exp(-izx)\, dx \to 0, \qquad |z| \to \infty.$$

Although this idea is known to have been assserted by Riemann, it is known that the idea was asserted for the first time by the French mathematicians Cauchy and Poisson between 1810 and 1840 without the condition that $f$ is $L^1$ integrable on $\mathbb{R}^d$. Riemann and Lebesque developed this idea, since it would not be possible for all the transformations to be correct without this condition.

Hence, $f \epsilon L^1(-\infty, \infty)$, the forward Fourier transform is $\psi_f(s) = \int e^{isx} f(x)\, dx$. According to Riemann's definition, this notation is seen as $\psi_f \epsilon L^1(-\infty, \infty)$, the inverse Fourier transform is $f(x) = \frac{1}{2\pi} \int e^{-isx} \psi_f(s)\, ds$.

This is true for every function in mathematics. However, statistics is more concerned with random variables rather than functions. In statistics, this representation can be made as $\psi_F(s) = \int e^{isx} f(x)\, dx$ if the density function $f(x)$ is known.

When the distribution function is used, it can be done as $X \sim F \;\Rightarrow\; \psi_F(s) = \int e^{isx} F(dx)$. The aim here is to obtain $\psi_F$ from $F$ and thus to obtain the experimental function $\psi_{F_n}$ and return to $F$ again obtaining $F_n$ from $\psi_{F_n}$. This cycle is given in Figure 1.

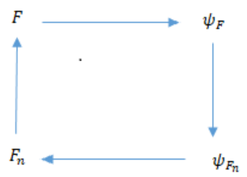


**Figure 1.** Fourier transform with the distribution function

It is also true if $f(x)$ is known (See Figure 2).

$$
\begin{tikzcd}
f \arrow[r] & \psi_f \arrow[d] \\
f_n \arrow[u] & \psi_{f_n} \arrow[l]
\end{tikzcd}
$$

**Figure 2.** Fourier transform with density function

However, it is not suggested to use $f(x)$. $f$ function may not be $L^1$ integrable on $\mathbb{R}^d$ or not be exactly known. Also, considering the asymptotic features, the convergence rate may be much slower.

Fourier transform is used in non-parametric models in which asymptotic properties are examined in statistics. In the convolution theorem, which is used especially in nonparametric equations with measurement errors, the nonparametric regression function can be estimated by the Fourier transform of a nonparametric estimator of the density of the flawed measurement, such as a Kernel estimator (Carroll et al., 1995). In the deconvolution technique, in order to eliminate the effect of measurement error, it is recommended to use a Kernel whose Fourier transform has a certain basis in the estimation of the non-parametric regression function (Fan, 1991). However, the rate of convergence depends on the variable of nonparametric function. The smoother the density of the variable, the faster the Fourier transform of the function deteriorates as the frequency approaches infinity, and the smaller the kernel bandwidth, the faster the bias decreases. In the literature focusing on the deconvolution technique, the smoothness of a density is typically defined in terms of the asymptotic decay rate of the Fourier transform as the frequency goes to infinity. The basis for such an explanation is that the number of derivatives of a continuous density is directly related to the asymptotic behavior of the Fourier transform as the frequency goes to infinity. This leads to the traditional distinction between " ordinary smooth" functions (which take a finite number of continuous derivatives and whose Fourier transform degrades under certain conditions) and "super smooth" functions (which take an infinite number of continuous derivatives and whose Fourier transform degrades under certain conditions). The use of the deconvolution technique in the estimation of the nonparametric function depends on knowing the distribution of the variable. In the absence of information about the distribution, when there are two consecutive measurement errors, the estimator is converted to simple factors with the Fourier transform process and the solution is obtained (Schennach, 2004; Yalaz and Tez, 2019).

In Fourier transform, which is the basis of all these important studies, it is important to decide which function will be used and when it will be used. To get rid of the stated handicaps may be the idea, instead of using the unknown density function $f$ using the distribution function $F$, can be adopted. Mentioned idea forms the basis of our research.

## 2. Material and Method

### *2.1. Probability Function and Distribution Function*

There are two types of random variables in statistics; discrete random variables and continuous random variables. The probability functions of discrete random variables ($X$) are probability values that take a finite number of values $x_1, x_2, \cdots, x_n$,

$$f(x_i) = P(X = x_i), i = 1, 2, \cdots, n.$$

The distribution function is the probability that $X$ is less than or equal to $x$,

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i).$$

Probability density function of continuous random variables ($X$) are probability values that take the values defined in the range $(-\infty, \infty)$ and satisfy the conditions $\int_{-\infty}^{\infty} f(x)dx = 1$, $f(x) \geq 0$; $-\infty < x < \infty$. In addition to this, the probability of finding the continuous random variable $X$ between $c$ and $d$ is

$$P(c < X < d) = \int_{c}^{d} f(x)dx,$$

which shows the area bounded by $f(x)$ curve, $x$-axis and $x = c$, $x = d$ lines. The distribution function is defined as (Akdeniz, 2002),

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds.$$

Distributions produced from discrete random variables are called discrete probability distributions, and distributions produced from continuous random variables are called continuous probability distributions.

Using $f(x)$ in Fourier transform is inconvenient. In order to use $f$, it must be $L^1$ integrable on $\mathbb{R}^d$, which means that $f \epsilon L^1(-\infty, \infty)$. Because the continuity condition cannot be met, discrete distributions are often ignored in the Fourier transform. Discrete probability distributions, which are well known and frequently used in statistics are Poisson, Bernoulli, Binomial, Geometric, Negative Binomial and Discrete Uniform distributions. It is not meaningful to ignore these important distributions and to deal with other distributions.

The maximum likelihood estimation method, used for the estimation of parameters in parametric statistical methods, proposed in years between 1912 and 1922 by Ronald Aymler Fisher, a British statistician, biologist and geneticist, is also valid for the independent sample with the same distributions. Because the joint probability density function can be written in these conditions as,

$$f(x_1, x_2, \ldots, x_n|\lambda) = f(x_1|\lambda) \times f(x_2|\lambda) \times \ldots \times f(x_n|\lambda).$$

From here,

$$L(\lambda; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\lambda) = \prod_{i=1}^{n} f(x_i|\lambda)$$

can be written. (The likelihood function $L(\lambda|X) = P(X|\lambda)$ can be used when the observed $x_1, x_2, \ldots, x_n$variables are considered as the constant parameters of this function as the variables of the function $\lambda$) (URL 3, 2022). Taking the natural logarithm of the likelihood function is a widely used method in practice:

$$\ln L(\lambda; x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i|\lambda).$$

However, as mentioned earlier, statistics is very different from mathematics. While a function is defined for every $x$ in mathematics, different $x$ values for each situation or different observation values are appeared in statistics. Therefore, it is necessary to know exactly what distribution the functions (in the language of statistics, density functions) of the observation values we are interested in have. When it is not known from which distribution the functions come from, parametric statistical methods, such as the maximum likelihood estimation method, cannot be used, so non-parametric statistical methods will be more compatible. Because of the fact that, non-parametric methods are based on the idea that the distributions of random variables should be obtained from the data, not the assumption that the distributions are known beforehand.

Kernel density estimation in non-parametric statistical methods provides very good estimates for the density function (Fan and Truong, 1993):

$$\widehat{f_h}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right).$$

where $K$, a non-negative function whose integral is equal to one, $h > 0$ is the smoothing parameter called the bandwidth.

Kernel density estimation, frequently used for measurement error models, is lean on the investigation of asymptotic features (Toprak, 2015; Yalaz, 2019). However, considering the asymptotic properties, it is known that $h$ has the same asymptotic ratio with $n^{-1/5}$. In this case, $\frac{1}{nh} = \frac{1}{nn^{-1/5}} = n^{-4/5}$. That is, the rate of convergence is much slower, $n^{-4/5}$, than the parametric methods convergence rate, $n^{-1}$, because

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} P(X_i \leq x).$$

The graph for $F(y) = \int_{-\infty}^{y} f(x)\,dx$ is indicated in Figure 3.

**Figure 3.** Distribution function graph

When smoothing methods are used, this returns to $F^s(y) = \int_{-\infty}^{y} f_n(x)\,dx$ and the graph changes to the Figure 4.

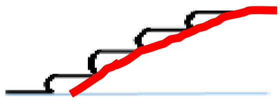

**Figure 4.** Flattened distribution function graph

As a result of smoothing, the piecewise function disappears and becomes a smooth function. This function can be used to find the unknown density function, but it should not be ignored that the convergence rate is very slow.

Since smoothing methods suggest the estimation of an unknown $f(x)$ function, although it is a preferred method, the convergence rate of the obtained function is very slow. The best way to avoid this handicap may be not to use the unknown probability function $f$ and to use the distribution function $F$ instead of it.

## 3. Findings and Discussion

### *3.1. Playing Around with a General Expression, Random Variables and Demonstrating the Expected Value with Fourier Transform*

In this study, instead of limiting the study and dealing with any function, we have proceeded through a method that aims to deal with relations called functional and then find functions that provide these equations. The most important aspect of working with functionals is not to make any special assumptions. For example, if it is not given that the desired function is differentiable or if it is not calculated in a way that can be differentiable, the derivative and related properties should not be used in the solution (URL 1, 2022).

Let $T(X)$ be a functional for $X \sim F$. Then,

$$E[T(X)] = \int T(X)F(dx).$$

Due to the drawbacks of the mentioned flattening methods, instead of using flattening methods here, representations have been made by playing with random variables a little (disturbing the random variables) as similar with these methods. So, let's take $X + hN$ instead of taking $X$. Here, $X$ and $N$ are independent random variables and $N$ has a standard normal distribution.

If $N \sim N(0,1)$, the characteristic function of the random variable $N$ is known as

$$\psi_N(s) = E\left[e^{isN}\right] = \int e^{isx}\theta_x(dx)$$

$$= \int e^{isx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int e^{-\frac{2isx-x^2}{2}}dx$$

$$= e^{-\frac{s^2}{2}}$$

This means that, $\theta(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi}}\psi_N(x)$.

Then,

$$E[T(X + hN)] = \int\int T(x + hy)\theta(y)F(dx)\,dy = \frac{1}{\sqrt{2\pi}}\int\int T(x + hy)\psi_N(y)dyF(dx) \tag{1}$$

Let us take $x + hy = z$. In this case

$$hy = z - x \Rightarrow y = \frac{z-x}{h} \Rightarrow dy = \frac{1}{h}dz.$$

Substituting this in equation (1), one can get

$$E[T(X+hN)] = \frac{1}{\sqrt{2\pi}}\frac{1}{h}\int\int T(z)\psi_N\left(\frac{z-x}{h}\right)dzF(dx)$$
$$= \frac{1}{\sqrt{2\pi}}\frac{1}{h}\int\int\int T(z)e^{i\left(\frac{z-x}{h}\right)w}\theta(w)dwdzF(dx)$$
$$= \frac{1}{\sqrt{2\pi}}\frac{1}{h}\int\int\int T(z)e^{iz\frac{w}{h}}e^{-ix\frac{w}{h}}\theta(w)dwdzF(dx). \quad (2)$$

Taking $\frac{w}{h} = \omega, \frac{dw}{h} = d\omega$ is possible. If these are used in equation (2), we can write

$$\frac{1}{\sqrt{2\pi}}\frac{1}{h}\int\int\int T(z)e^{iz\frac{w}{h}}e^{-ix\frac{w}{h}}\theta(w)dwdzF(dx) = \frac{1}{\sqrt{2\pi}}\int\int\int T(z)e^{iz\omega}e^{-ix\omega}\theta(h\omega)d\omega dzF(dx) \quad (3)$$

where $\theta(h\omega) = \frac{1}{\sqrt{2\pi}}e^{-\frac{h^2\omega^2}{2}}$. Substitute this in equation (3), the presentation will be

$$\frac{1}{\sqrt{2\pi}}\int\int\int T(z)e^{iz\omega}e^{-ix\omega}\theta(h\omega)d\omega dzF(dx) = \frac{1}{2\pi}\int\int\int T(z)e^{iz\omega}e^{-ix\omega}e^{-\frac{h^2\omega^2}{2}}d\omega dzF(dx)$$
$$= \frac{1}{2\pi}\int\int T(z)e^{iz\omega}\int e^{-ix\omega}F(dx)dze^{-\frac{h^2\omega^2}{2}}d\omega. \quad (4)$$

Substituting $\int e^{-ixs}F(dx) = \psi_F(-s)$ in equation (4), it can be seen that

$$\frac{1}{2\pi}\int\int T(z)e^{iz\omega}\int e^{-ix\omega}F(dx)dze^{-\frac{h^2\omega^2}{2}}d\omega = \frac{1}{2\pi}\int\int T(z)e^{izs}\psi_F(-s)dze^{-\frac{h^2s^2}{2}}ds. \quad (5)$$

Since $\int e^{izs}T(z)dz = \psi_T(s)$ equation (5) will be

$$\frac{1}{2\pi}\int\int T(z)e^{izs}\psi_F(-s)dze^{-\frac{h^2s^2}{2}}ds = \frac{1}{2\pi}\int \psi_F(-s)\psi_T(s)e^{-\frac{h^2s^2}{2}}ds. \quad (6)$$

However, since $T(X+hN)$ was obtained while dealing with the functional $T(X)$ at the beginning, it should be returned to the situation where random variables have not been manipulated.

*3.2. Returning to the State of Not Playing with Random Variables*

The main goal was to get $E[T(X)]$. When it is desired to return $E[T(X)]$ again from equation (6), it is necessary to approximate $h \to 0$. In this case $e^{-\frac{h^2 s^2}{2}} \to 1$ happens. However, we should be very careful to the remaining expression

$$\frac{1}{2\pi}\int \psi_F(-s)\psi_T(s)ds. \tag{7}$$

At first glance, although approximating $h \to 0$ seems to make our job easier, the aforementioned troubles are also valid here. In this case, considering the equation (7), $\psi_F(-s)$ or $\psi_T(s)$ must be integrable. $\psi_F(-s)$ is certainly not integrable because $F$ distribution contains jumps (see Figure 3).

Let us take $h \to 0$ for a moment. In this case, the expression

$$E[T(X+hN)] = \frac{1}{2\pi}\int \psi_F(-s)\psi_T(s)e^{-\frac{h^2 s^2}{2}}ds$$

will return to

$$E[T(X)] = \frac{1}{2\pi}\int \psi_F(-s)\psi_T(s)ds.$$

It is known that $\psi_F(-s)$ is not integrable. So, in order to $E[T(X)]$ be valid, $T$ should be integrable for $|T| \epsilon L^1(-\infty,\infty)$, and for taking inverse process $\psi_T(s) = \int e^{isx}T(x)dx$ should be integrable for $\psi_T \epsilon L^1(-\infty,\infty)$. Situations that satisfy both conditions can be achieved through a multitude of assumptions known as restrictions. In this case, $\psi_T(s)$ needs to be defined more specifically and be integrable. Initially our aim has kept $\psi_T(s)$ as any functional. Therefore, taking $h \to 0$ would not be correct.

The recommendation here is taking $h \to 0^-$ and $h \to 0^+$ rather than taking $h \to 0$. Then,

$$E[T(X+hN)] = \frac{1}{2}E[T(X^+)] + \frac{1}{2}E[T(X^-)].$$

If we take $T(x) = 1_{a<x\leq b}$, because of

$T(x^+) = \lim_{y\downarrow x} T(y)$ and $T(x^-) = \lim_{y\uparrow x} T(y)$, then

$$T(x) = \frac{1}{2}T(x^+) + \frac{1}{2}T(x^-).$$

However, it is necessary to pay attention to the idea of taking $h \to 0^-$ and $h \to 0^+$. Lebesque measurement theory does not allow us to integrate limits approaching from the right and left. In this case, we may encounter more limitations.

## 4. Results

Kernel density estimation, which is frequently used for measurement error in variables, is based on the logic of smoothing non-smooth probability density functions $(f)$. For this reason, while studying the subject of measurement error in variables, it has also been applied to investigate the asymptotic properties in the literature. However, the convergence rate is much slower under weak assumptions than the convergence rate in parametric methods. Instead of using unknown probability function $f$, using the distribution function $F$ is the best way to avoid this handicap. By paying attention to the integrability condition in the smoothing process using the distribution function, it became possible to achieve the desired result by assuming that the smoothing parameter approaches zero from the right and left. However, it is necessary to pay attention to the idea that limits used in the functions and equations cannot be included in the integral.

## Statement of Conflict of Interest

The author has declared no conflict of interest.

## Author's Contributions

The contribution of the author is equal.

## References

Akdeniz F. Olasılık ve istatistik. Adana, Baki Kitabevi, 2002.

Fan J. On the optimal rates of convergence for nonparametric deconvolution problems. Annals of Statistics 1991; 19: 1257-1272.

Fan J., Truong YK. Nonparametric regression with errors in variables. Annals of Statistics 1993; 21: 1900-1925.

Matematik Dünyası, URL 1: http://www.matematikdunya si.org/arsiv/PDF_eskisayilar/92_4_22_25_FONKSIYONEL.pdf (Erişim zamanı; Eylül, 15, 2022).

Carroll RJ., Ruppert D., Stefanski LA. Measurement error in nonlinear models. Chapman and Hall, 1995.

Bochner S., Chandrasekharan K. Fourier transform. New Jersey, Princeton University Press, 1949.

Schennach SM. Estimation of nonlinear models with measurement error. Econometrica, 2004; 72: 33-75.

Toprak S. Semiparametric regression models with errors in variables. Dicle University Institute of Science, PHd Thesis, Diyarbakır, Turkey, 2015.

Yalaz S. Multivariate partially linear regression in the presence of measurement error. AStA Advances in Statistical Analysis, 2019; 103(1): 123–135.

Yalaz S., Tez M. Semiparametric EIV regression model with unknown errors in all variables. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 2019; 8(4): 1177–1183.

Wikipedia, URL 2:
https://en.m.wikipedia.org/wiki/ Riemann–Lebesgue_lemma
(Erişim zamanı; Eylül, 15, 2022).

Wikipedia, URL 3:
https://en.m.wikipedia.org/wiki /Maximum_likelihood_estimation
(Erişim zamanı; Eylül, 15, 2022).