East Asian Math. J.
Vol. 39 (2023), No. 5, pp. 537–547
http://dx.doi.org/10.7858/eamj.2023.040

OPTIMAL APPROXIMATION BY ONE GAUSSIAN FUNCTION TO PROBABILITY DENSITY FUNCTIONS

Gwang Il Kim, Seung Yeon Cho, and Doobae Jun*

Abstract. In this paper, we introduce the optimal approximation by a Gaussian function for a probability density function. We show that the approximation can be obtained by solving a non-linear system of parameters of Gaussian function. Then, to understand the non-normality of the empirical distributions observed in financial markets, we consider the nearly Gaussian function that consists of an optimally approximated Gaussian function and a small periodically oscillating density function. We show that, depending on the parameters of the oscillation, the nearly Gaussian functions can have fairly thick heavy tails.

1. Introduction

It is well known that the empirical distributions of major targets analyzed in financial markets, such as daily stock returns, are non-normal/non-Gaussian [1]. Particularly, the non-normal distributional peculiarities in the thickness of the tail, skewness, and kurtosis of their distributions have attracted the attention of many researchers, which have been analyzed in various ways, according to several distributions and methods suggested as alternatives: In [2, 3], targeting major countries such as the US, Canada and Japan, the non-normality of the distribution of daily returns of financial assets was analyzed. Also, in [4], by using more flexible distributions like the exponential generalized beta (EGB) and skewed generalized t (SGT) distribution, it was shown that the distribution of daily stock returns of a large emerging European stock market, invested in the Istanbul Stock Exchange, has highly non-normal characteristics. In [5], authors show that the distribution of market returns is not only leptokurtic but also skewed, and in [6], by adopting the Pearson type families of distributions,

Received August 4, 2023; Revised September 25, 2023; Accepted September 30, 2023.
2010 *Mathematics Subject Classification.* 62G30.
Key words and phrases. Optimal approximation, non-normal distribution, periodically oscillating distribution, nearly-Gaussian function.
This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. NRF-2021R1I1A3048350, RS-2022-00166144, NRF-2021R1A2C1014001).
*Corresponding Author.

(pISSN 1226-6973, eISSN 2287-2833)

the estimation of the parameters of the stochastic differential equation (SDE) related to the distributions was performed for the stock returns in the US and Japanese markets. In addition, in [7, 8], several alternatives to the distributions of market returns of diverse countries were proposed and their ability as alternatives were investigated.

In this paper, we propose another perspective for understanding the non-normality of the empirically observed distributions in financial markets, according to the following two steps: First, we introduce the optimal approximation by a Gaussian function to a probability density function and show when it can be achieved. Second, we consider nearly Gaussian function represented as the weighted sum of an optimal Gaussian approximation and a small periodically oscillating density function. And, we show how the tail of the nearly Gaussian function can be thicker than that of the given optimal Gaussian function.

The outline of this paper is given as follows: In Section 2, we introduce the optimal approximation by a Gaussian function. Then, in Section 3, we analyze the behaviors of the mean and variance of the nearly Gaussian function with a small periodically oscillating density. Finally, in Section 4, we summarize our results and propose some directions for future work.

2. Optimal approximation by a Gaussian function

Definition 1. For arbitrary functions $f, g \in \mathrm{L}^2(\mathbb{R})$, we define $\langle f, g\rangle$ by

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)dx.$$

As well-known, using the inner product $\langle\ ,\ \rangle$, we can naturally introduce a vector norm $\|\cdot\|$ in $\mathrm{L}^2(\mathbb{R})$: $\|f\| = \langle f, f\rangle^{1/2}$, $\forall f \in \mathrm{L}^2(\mathbb{R})$, with which we can also measure the difference between two functions $f, g \in \mathrm{L}^2(\mathbb{R})$ as $\|f - g\|$.

Definition 2. In $\mathrm{L}^2(\mathbb{R})$, a normalized Gaussian function φ with mean μ and variation σ^2 is given by

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{\sigma\sqrt{\pi}}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{1}$$

Now, we consider the Gaussian function approximation to an arbitrary probability density function in $\mathrm{L}^2(\mathbb{R})$.

Definition 3. For a probability density function $\rho \in \mathrm{L}^2(\mathbb{R})$, the normalized Gaussian function φ with mean μ and variation σ^2 is said to be *optimal* with weight k in approximating ρ, if $\|\rho - k\varphi_{\mu,\sigma}\|$ is minimized. In this case, we call the function $k\varphi_{\mu,\sigma}$ the Gaussian function approximation of ρ.

Proposition 2.1. *For a probability density function $\rho \in L^2(\mathbb{R})$, the Gaussian function approximation optimized at $\boldsymbol{a} = (\mu_0, \sigma_0)$ is given by*

$$\phi(x) = \frac{k_0}{\sqrt{\sigma_0\sqrt{\pi}}} \exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right), \tag{2}$$

where $k_0 = \langle \rho, \varphi_{\mu_0,\sigma_0} \rangle$ *and* $\boldsymbol{a}$ *is the solution to the system of equations*

$$\left\langle \rho, \frac{\partial}{\partial \mu} \varphi_{\mu,\sigma} \right\rangle = 0, \ \left\langle \rho, \frac{\partial}{\partial \sigma} \varphi_{\mu,\sigma} \right\rangle = 0. \tag{3}$$

Proof. First, let $\sum = \{(k, \mu, \sigma) \mid k, \mu \in \mathbb{R}, \text{ and}, \ \sigma \in \mathbb{R}^+\}$ and $\Psi(k, \mu, \sigma) = \|\rho - k\varphi_{\mu,\sigma}\|^2$. Then $\Psi(k, \mu, \sigma)$ is a C^1 function on $\sum$. Hence, if the Gaussian function approximation $\phi(x)$ is optimized with a special weight k_0 at $\mathbf{a} = (\mu_0, \sigma_0)$, i.e. $\|\rho - k\varphi_{\mu,\sigma}\|$ is minimized at (k_0, μ_0, σ_0) (equivalently, so is $\|\rho - k\varphi_{\mu,\sigma}\|^2$), the following equations hold:

$$\frac{\partial}{\partial k}\Psi(k_0, \mu_0, \sigma_0) = 0, \frac{\partial}{\partial \mu}\Psi(k_0, \mu_0, \sigma_0) = 0, \frac{\partial}{\partial \sigma}\Psi(k_0, \mu_0, \sigma_0) = 0. \tag{4}$$

Here, since $\langle \varphi_{\mu,\sigma}, \varphi_{\mu,\sigma} \rangle = 1$, we have that

$$\begin{aligned} \Psi(k, \mu, \sigma) &= \|\rho - k\varphi_{\mu,\sigma}\|^2 \\ &= \langle \rho, \rho \rangle - 2k\langle \rho, \varphi_{\mu,\sigma} \rangle + k^2. \end{aligned} \tag{5}$$

Therefore, by Eq. (4), we obtain

$$-2\langle \rho, \varphi_{\mu_0,\sigma_0} \rangle + 2k_0 = 0, \quad -2k_0 \left\langle \rho, \frac{\partial}{\partial \mu}\varphi_{\mu,\sigma} \right\rangle \Big|_{\mathbf{a}} = 0,$$
$$-2k_0 \left\langle \rho, \frac{\partial}{\partial \sigma}\varphi_{\mu,\sigma} \right\rangle \Big|_{\mathbf{a}} = 0.$$

This completes the proof. □

Proposition 4 provides the necessary condition that the optimal Gaussian approximation should satisfy. In the following theorem, we show that minimizing $\Psi(k, \mu, \sigma)$ is equivalent to maximizing $\langle \rho, \varphi_{\mu,\sigma} \rangle^2$.

Theorem 2.2. *Let* $\Omega = \{(\mu, \sigma) \mid \mu \in \mathbb{R}, \text{ and}, \ \sigma \in \mathbb{R}^+\}$ *and let* $\rho \in L^2(\mathbb{R})$ *be a probability density function. Then the Gaussian function approximation* $k\varphi_{\mu,\sigma}$ *is optimized at* $\boldsymbol{a} = (\mu_0, \sigma_0) \in \Omega$ *if* $\langle \rho, \varphi_{\mu,\sigma} \rangle^2$ *has its global maximum at* $\boldsymbol{a}$ *and* $k = \langle \rho, \varphi_{\mu_0,\sigma_0} \rangle$.

Proof. We first note that

$$\begin{aligned} \Psi(k, \mu, \sigma) &= \|\rho - k\varphi_{\mu,\sigma}\|^2 \\ &= \langle \rho, \rho \rangle - 2k \langle \rho, \varphi_{\mu,\sigma} \rangle + k^2 \\ &= \langle \rho, \rho \rangle + (k - \langle \rho, \varphi_{\mu,\sigma} \rangle)^2 - \langle \rho, \varphi_{\mu,\sigma} \rangle^2 . \end{aligned} \tag{6}$$

Since ρ is a given function, if $\langle \rho, \varphi_{\mu,\sigma} \rangle^2$ is maximized at $\mathbf{a}$ and $k = \langle \rho, \varphi_{\mu_0,\sigma_0} \rangle$, it is trivial that $\Psi(k, \mu, \sigma)$ is optimized at $\mathbf{a}$. □

Example 2.3. Consider a probability density function $\rho(x)$ as follows:

$$\rho(x) = \begin{cases} 0, & x < -2; \\ 0.1, & -2 \leq x < 0; \\ 0.3, & 0 \leq x < 2; \\ 0.2, & 2 \leq x < 3; \\ 0, & 3 \leq x. \end{cases}$$

According to Theorem 5, in order to find the optimal Gaussian approximation of a probability density function ρ, we need to maximize $\langle \rho, \varphi_{\mu,\sigma} \rangle^2$. For this, we solve a non-linear system:

$$\langle \rho, \partial_\mu \varphi_{\mu,\sigma} \rangle = 0, \quad \langle \rho, \partial_\sigma \varphi_{\mu,\sigma} \rangle = 0, \tag{7}$$

and use the resultants for comupting $k = \langle \rho, \varphi_{\mu,\sigma} \rangle$. In this example, we are informed N values of distribution function ρ on grid points $x_i = x_{min} + i\Delta x$, $i = 0, 1, ..., N-1$, where the number N is sufficiently large to resolve the shape of the distribution function ρ. Now, we treat (7) in a discrete manner, i.e., we aim to find $p = (\mu, \sigma)$ such that

$$F_1(p) = \sum_i \rho_i \partial_\mu \varphi_{\mu,\sigma}(x_i) = 0, \quad F_2(p) = \sum_i \rho_i \partial_\sigma \varphi_{\mu,\sigma}(x_i) = 0. \tag{8}$$

Recalling the form of $\varphi_{\mu,\sigma}(x)$ in (1), the partial derivatives are given by

$$\partial_\mu \varphi_{\mu,\sigma}(x) = \frac{x-\mu}{\sigma^2} \varphi_{\mu,\sigma}(x), \quad \partial_\sigma \varphi_{\mu,\sigma}(x) = \left(-\frac{1}{2\sigma} + \frac{(x-\mu)^2}{\sigma^3} \right) \varphi_{\mu,\sigma}(x).$$

In order to solve (8), we use Newton's method:

$$p_{k+1} = p_k - J_F(p_k)^{-1} F(p_k),$$
$$F(p_k) = \begin{bmatrix} F_1(p_k) \\ F_2(p_k) \end{bmatrix} \quad p_k = \begin{bmatrix} \mu_k \\ \sigma_k \end{bmatrix},$$

where the Jacobian matrix $J_F(p)$ is given by

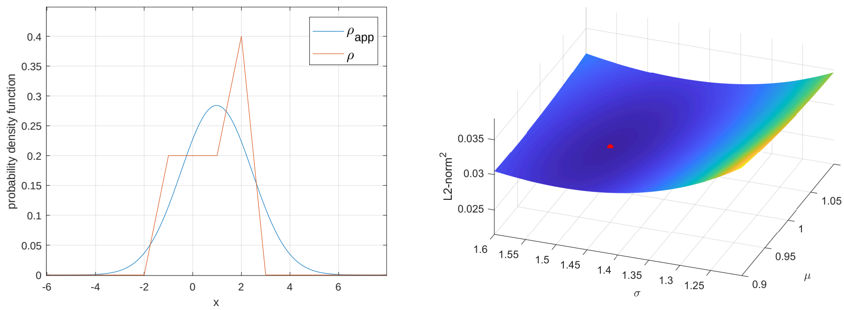
$$J_F(p) = \begin{bmatrix} \partial_\mu F_1(p) & \partial_\sigma F_1(p) \\ \partial_\mu F_2(p) & \partial_\sigma F_2(p) \end{bmatrix} = \begin{bmatrix} \sum_i \rho_i \partial_\mu^2 \varphi_{\mu,\sigma}(x_i) & \sum_i \rho_i \partial_\sigma \partial_\mu \varphi_{\mu,\sigma}(x_i) \\ \sum_i \rho_i \partial_\mu \partial_\sigma \varphi_{\mu,\sigma}(x_i) & \sum_i \rho_i \partial_\sigma^2 \varphi_{\mu,\sigma}(x_i) \end{bmatrix}.$$

Note that $J_F(p)$ is just the linear combination of partial derivatives:

$$\partial_\mu^2 \varphi_{\mu,\sigma}(x) = \left(-\frac{1}{\sigma^2} + \frac{(x-\mu)^2}{\sigma^4} \right) \varphi_{\mu,\sigma}(x),$$
$$\partial_\sigma \partial_\mu \varphi_{\mu,\sigma}(x) = \partial_\mu \partial_\sigma \varphi_{\mu,\sigma}(x) = \left(\frac{x-\mu}{\sigma^2} \right) \left(-\frac{5}{2\sigma} + \frac{(x-\mu)^2}{\sigma^3} \right) \varphi_{\mu,\sigma}(x),$$
$$\partial_\sigma^2 \varphi_{\mu,\sigma}(x) = \left[\left(-\frac{1}{2\sigma} + \frac{(x-\mu)^2}{\sigma^3} \right)^2 + \left(\frac{1}{2\sigma^2} - \frac{3(x-\mu)^2}{\sigma^4} \right) \right] \varphi_{\mu,\sigma}(x).$$

Since the convergence of Newton's method depends on the initial guess of μ and σ, as an suitable initial guess p_0, we use the mean and variance of the N samples ρ_i, $i = 0, ...N-1$. In this example, Newton's method with tolerance

$tol = 10^{-8}$ gives an optimal parameter $(\mu_{opt}, \sigma_{opt}) \approx (0.9796, 1.4715)$ and $k_{opt} \approx 0.4588$ within only five iterations. In the left panel of Figure 1, we compare the two functions ρ and $\rho_{app} := k_{opt}\varphi_{\mu_{opt},\sigma_{opt}}$. Also, in the right panel of Figure 1, we show that the error $E(\mu, \sigma) := \|\rho - k\varphi_{\mu,\sigma}\|_2$ is minimized around the red point $p_{opt} = (0.9796, 1.4715)$.



(a) (b)

FIGURE 1. Optimal approximation of the density function ρ by the Gaussian function. (a) Shapes of ρ and ρ_{app}. (b) Error surface $E(\mu, \sigma)$.

Remark 1. On the other hand, in the case of probability density function with the shape similar to a Gaussian function, we can obtain a more precise approximation of the density function. However, if not as in the previous example, to obtain such a precise optimal approximation, we need to control the norm of its remainder density function (for example, performing additional approximations on the remainder density function several times) or using multiple Gaussian functions from the beginning.

In this paper, we focus on the case where the norm of the remainder density function can be well controlled, as shown in the next section. Here, for the sake of the following discussion, let us introduce *nearly Gaussian* functions, as follows: Let ψ_{μ_0,σ_0} be the normalized density function of $\phi = k_0\varphi_{\mu_0,\sigma_0}$ so that the integral of ψ_{μ_0,σ_0} on $\mathbb{R}$ is 1, i.e.

$$\psi_{\mu_0,\sigma_0}(x) = \frac{1}{k_0\sqrt{2\sqrt{\pi}\sigma_0}}\phi(x), \;\; x \in \mathbb{R}.$$

Then, we can we write the density function ρ as follows:

$$\begin{aligned}\rho &= k_0\varphi_{\mu_0,\sigma_0} + (\rho - k_0\varphi_{\mu_0,\sigma_0})\\ &= \psi_{\mu_0,\sigma_0} + k_0\varphi_{\mu_0,\sigma_0} - \psi_{\mu_0,\sigma_0} + (\rho - k_0\varphi_{\mu_0,\sigma_0})\\ &= \psi_{\mu_0,\sigma_0} + (k_0\sqrt{2\sqrt{\pi}\sigma_0} - 1)\psi_{\mu_0,\sigma_0} + (\rho - k_0\varphi_{\mu_0,\sigma_0}).\end{aligned}$$

Here, we note that, if $\|\rho - k_0\varphi_{\mu_0,\sigma_0}\|$ is sufficiently small and $k_0\sqrt{2\sqrt{\pi}\sigma_0} \simeq 1$, the density function ρ is close to Gaussian, i.e., $\rho \simeq \psi_{\mu_0,\sigma_0}$. In this case, we call the density function ρ nearly Gaussian function.

In the next section, we focus on the case where a small perturbation part of nearly Gaussian function is given by a harmonic oscillation.

3. Nearly Gaussian function with harmonic oscillation

In the financial market, the behavior of the return of an asset is described by the stochastic process, which means that its corresponding probability density function is represented by the function of x and t. To understand the non-normality of the probability density function, we assume that $\rho_\epsilon(x,t)$, the approximation of $\rho(x,t)$, consists of the optimally approximated Gaussian function and an very small remainder density term in simple harmonic oscillation, which is given as follows; for $0 < \epsilon \ll 1$,

$$\rho_\epsilon(x,t) = (1-\epsilon)\frac{1}{k_0(t)\sqrt{2\sqrt{\pi}\sigma_0(t)}}\phi(x,t) + \epsilon\eta(x - A\sin\omega t), \tag{9}$$

where η is a probability density function such that $0 \le \eta(x) \in \mathrm{L}^2(\mathbb{R})$ with $\int_{\mathbb{R}} \eta(x)dx = 1$. Here the function $\phi(x,t)$ can be thought as the optimal approximation of $\rho(x,t)$ obtained by Theorem 2.2 at each time t and $k_0(t)$, $\sigma_0(t)$ are its corresponding optimal parameters. Note that the first term on the right side of Eq. (9) can be simplified as a specific Gaussian function;

$$\frac{1}{k_0(t)\sqrt{2\sqrt{\pi}\sigma_0(t)}}\phi(x,t) = \frac{1}{\sqrt{2\pi}\sigma_0(t)}\exp\left(-\frac{(x-\mu_0(t))^2}{2\sigma_0(t)^2}\right),$$

and we immediately obtain that $\int_{\mathbb{R}} \frac{1}{k_0(t)\sqrt{2\sqrt{\pi}\sigma_0(t)}}\phi(x,t)dx = 1$. Although $\rho_\epsilon(x,t)$ in Eq. (9) is written as the weighted sum of a Gaussian function and a small oscillating function, it can also be understood as the sum of a Gaussian function and a small perturbation periodic in time:

$$\rho_\epsilon(x,t) = \frac{1}{k_0(t)\sqrt{2\sqrt{\pi}\sigma_0(t)}}\phi(x,t) + \epsilon\left(\eta(x - A\sin\omega t) - \frac{1}{k_0(t)\sqrt{2\sqrt{\pi}\sigma_0(t)}}\phi(x,t)\right). \tag{10}$$

Then from now on, we analyze how the mean and variance the distribution function changes according to the form of a Gaussian function and a small perturbation. Before proceeding, we first calculate the mean and variance of the accompanied distribution $\eta(x - A\sin\omega t)$ at time $t \ge 0$.

Lemma 3.1. *Assume that a distribution with the density function $\eta(x)$ has mean μ_1 and variance σ_1^2. Then the mean $\mu(t)$ and the variance $\sigma(t)^2$ of the distribution with the modified density function $\eta(x - A\sin\omega t)$ is given by*

$$\mu(t) = \mu_1 + A\sin\omega t, \quad \sigma(t)^2 = \sigma_1^2. \tag{11}$$

Proof. First, we find the mean $\mu(t)$ of the distribution with the density function $\eta(x - A\sin\omega t)$. Since $\int_{\mathbf{R}} \eta(x)dx = 1$, we can easily obtain

$$\begin{aligned}\mu(t) &= \int_{\mathbb{R}} x\eta(x - A\sin\omega t)dx \\ &= \int_{\mathbb{R}} x\eta(x)dx + A\sin\omega t \int_{\mathbb{R}} \eta(x)dx \\ &= \mu_1 + A\sin\omega t.\end{aligned}$$

Next, we use $\mu(t)$ to find the variance $\sigma(t)^2$ of the distribution $\eta(x - A\sin\omega t)$ as follows:

$$\begin{aligned}\sigma(t)^2 &= \int_{\mathbb{R}} (x - \mu(t))^2\eta(x - A\sin\omega t)dx \\ &= \int_{\mathbb{R}} (x + A\sin\omega t)^2\eta(x)dx - 2\mu(t)\int_{\mathbb{R}} (x + A\sin\omega t)\eta(x)dx \\ &\quad + \mu(t)^2 \int_{\mathbb{R}} \eta(x)dx \\ &= \int_{\mathbb{R}} x^2\eta(x)dx + 2A\sin\omega t\int_{\mathbb{R}} x\eta(x)dx + (A\sin\omega t)^2\int_{\mathbb{R}} \eta(x)dx \\ &\quad - 2\mu(t)\int_{\mathbb{R}} x\eta(x)dx + \left(-2\mu(t)A\sin\omega t + \mu(t)^2\right)\int_{\mathbb{R}} \eta(x)dx.\end{aligned}$$

Recalling the assumptions:

$$\int_{\mathbf{R}} \eta(x)dx = 1, \quad \int_{\mathbf{R}} x\eta(x)dx = \mu_1, \quad \int_{\mathbf{R}} (x - \mu_1)^2\eta(x)dx = \sigma_1^2$$

we have

$$\begin{aligned}\sigma(t)^2 &= \sigma_1^2 + \mu_1^2 + 2A\sin\omega t\mu_1 + (A\sin\omega t)^2 - 2\mu_1\mu(t) - 2\mu(t)A\sin\omega t + \mu(t)^2 \\ &= \sigma_1^2 + (\mu_1 - \mu(t))^2 + 2A\sin\omega t(\mu_1 - \mu(t)) + (A\sin\omega t)^2 \\ &= \sigma_1^2 + (\mu_1 - \mu(t) + A\sin\omega t)^2 \\ &= \sigma_1^2.\end{aligned}$$

This completes the proof. □

By Lemma 3.1, we see that any modification by a simple harmonic oscillation on the density function of a probability distribution never changes the variance of the probability distribution, even though the mean of the probability distribution shows the same simple harmonic oscillation as in the modification.

In the following proposition, we show that the small harmonic oscillation $\eta(x - A\sin\omega t)$, included as part of the distribution function $\rho_\epsilon(x,t)$ in Eq. (9), can play important roles in determining the mean and variance of the distribution function $\rho_\epsilon(x,t)$ at time t.

Proposition 3.2. *Suppose that a probability density function $\rho_\epsilon(x,t)$ is given as in Eq. (9). Then the mean $\mu_\epsilon(t)$ and the variance $\sigma_\epsilon(t)^2$ of the distribution with the density function $\rho_\epsilon(x,t)$ is given by*

$$\mu_\epsilon(t) = (1-\epsilon)\mu_0(t) + \epsilon(\mu_1 + A\sin\omega t), \tag{12}$$

$$\sigma_\epsilon(t)^2 = (1-\epsilon)(\sigma_0(t)^2 + \epsilon^2 h(t)^2) + \epsilon(\sigma_1^2 + (1-\epsilon)^2 h(t)^2), \tag{13}$$

where μ_1 and σ_1^2 are respectively the mean and the variance of the distribution with the density function $\eta(x)$, and $h(t) = \mu_0(t) - \mu_1 - A\sin\omega t$.

Proof. For brevity, we omit the t in μ_0 and σ_0 in the proof. We begin by finding the mean $\mu_\epsilon(t)$ of the distribution function $\rho_\epsilon(x,t)$.

$$\begin{aligned}
\mu_\epsilon(t) &= (1-\epsilon)\int_{\mathbb{R}} x\frac{1}{k_0\sqrt{2\sqrt{\pi}\sigma_0}}\phi(x)dx + \epsilon\int_{\mathbb{R}} x\eta(x - A\sin\omega t)dx \\
&= (1-\epsilon)\int_{\mathbb{R}} x\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right)dx + \epsilon\int_{\mathbb{R}}(x + A\sin\omega t)\eta(x)dx \\
&= (1-\epsilon)\mu_0 + \epsilon(\mu_1 + A\sin\omega t).
\end{aligned}$$

Using this, we obtain the variance $\sigma_\epsilon(t)^2$ as follows:

$$\begin{aligned}
\sigma_\epsilon(t)^2 = (1-\epsilon)&\int_{\mathbb{R}}(x^2 - 2\mu_\epsilon(t)x + \mu_\epsilon(t)^2)\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right)dx \\
&+ \epsilon\int_{\mathbb{R}}(x^2 - 2\mu_\epsilon(t)x + \mu_\epsilon(t)^2)\eta(x - A\sin\omega t)dx \\
= I_1 + I_2.&
\end{aligned}$$

Since we know

$$\int_{\mathbf{R}}\eta(x)dx = 1, \quad \int_{\mathbf{R}} x\eta(x)dx = \mu_1, \quad \int_{\mathbf{R}}(x-\mu_1)^2\eta(x)dx = \sigma_1^2,$$

we can compute I_1 as

$$\begin{aligned}
I_1 &= (1-\epsilon)\int_{\mathbb{R}} x^2\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right)dx \\
&\quad - 2(1-\epsilon)\mu_\epsilon(t)\int_{\mathbb{R}} x\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right)dx \\
&\quad + (1-\epsilon)\mu_\epsilon(t)^2\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right)dx \\
&= (1-\epsilon)(\sigma_0^2 + \mu_0^2) - 2(1-\epsilon)\mu_\epsilon(t)\mu_0 + (1-\epsilon)\mu_\epsilon(t)^2 \\
&= (1-\epsilon)\left(\sigma_0^2 + (\mu_0 - \mu_\epsilon(t))^2\right).
\end{aligned}$$

As in the proof of Proposition 3.2, we compute I_2 as

$$\begin{aligned}
I_2 &= \epsilon \int_{\mathbb{R}} \left(x^2 - 2\mu_\epsilon(t)x + \mu_\epsilon(t)^2\right) \eta(x - A\sin\omega t)dx \\
&= \epsilon \int_{\mathbb{R}} \left((x + A\sin\omega t)^2 - 2\mu_\epsilon(t)(x + A\sin\omega t) + \mu_\epsilon(t)^2\right) \eta(x)dx \\
&= \epsilon(\sigma_1^2 + \mu_1^2) + 2\epsilon(A\sin\omega t - \mu_\epsilon(t))\mu_1 + \epsilon(A\sin\omega t - \mu_\epsilon(t))^2 \\
&= \epsilon\left(\sigma_1^2 + (\mu_1 + A\sin\omega t - \mu_\epsilon(t))^2\right).
\end{aligned}$$

Combining I_1 and I_2, we have

$$\sigma_\epsilon(t)^2 = (1-\epsilon)\left(\sigma_0^2 + (\mu_0 - \mu_\epsilon(t))^2\right) + \epsilon\left(\sigma_1^2 + (\mu_1 + A\sin\omega t - \mu_\epsilon(t))^2\right).$$

Letting $h(t) = \mu_0 - \mu_1 - A\sin\omega t$, we have

$$\begin{aligned}
&\mu_0 - \mu_\epsilon(t) = \epsilon(\mu_0 - \mu_1 - A\sin\omega t) = \epsilon h(t), \\
&\mu_1 + A\sin\omega t - \mu_\epsilon(t) = -(1-\epsilon)(\mu_0 - \mu_1 - A\sin\omega t) = -(1-\epsilon)h(t),
\end{aligned}$$

and hence $\sigma_\epsilon(t)^2$ can be simplified as

$$\sigma_\epsilon(t)^2 = (1-\epsilon)(\sigma_0^2 + \epsilon^2 h(t)^2) + \epsilon(\sigma_1^2 + (1-\epsilon)^2 h(t)^2).$$

□

Proposition 3.2 shows how the mean and the variance of the modified distribution change by the added small distribution with simple harmonic oscillation: They all depend strongly on the simple harmonic oscillation, despite the algebraic way in which the added distribution combines with the existing distribution. In particular, even though the algebraic scale ϵ is sufficiently small, when the amplitude of A and μ_1 is very large, the changes in all of the mean and the variance of the modified distribution may be so large that it cannot be ignored.

In the following theorem, we provide an sufficient condition for $\sigma_\epsilon(t)^2$ to be large as we want. The result implies that one may describe the fat tail of a probability density function with nearly Gaussian function if μ_1 is sufficiently far from μ_0.

Theorem 3.3. *Suppose that a probability function $\rho_\epsilon(x,t)$ is given as in Eq. (9). Given $\nu \geq 0$, if the parameters of Gaussian distribution function and ϵ satisfy*

$$|\mu_0(t) - \mu_1| > (\sqrt{\nu} + 1)\sigma_0/\sqrt{\epsilon}, \quad A \simeq \sigma_0(t)/\sqrt{\epsilon}, \tag{14}$$

then

$$\frac{\sigma_\epsilon(t)^2}{\sigma_0(t)^2} = \lambda + \mathcal{O}(\epsilon), \quad \lambda > \nu.$$

Proof. By the Proposition 3.2, we have

$$\frac{\sigma_\epsilon(t)^2}{\sigma_0^2} = (1-\epsilon)\left(1 + \epsilon^2 \frac{h(t)^2}{\sigma_0^2}\right) + \epsilon\left(\frac{\sigma_1^2}{\sigma_0^2} + (1-\epsilon)^2 \frac{h(t)^2}{\sigma_0^2}\right). \tag{15}$$

We also recall that

$$\frac{h(t)}{\sigma_0} = \frac{\mu_0 - \mu_1 - A\sin\omega t}{\sigma_0}.$$

Since the assumption implies $\dfrac{h(t)^2}{\sigma_0(t)} = \mathcal{O}\left(\dfrac{1}{\epsilon}\right)$, we can rewrite Eq. 15 as

$$\frac{\sigma_\epsilon(t)^2}{\sigma_0^2} = 1 + \epsilon\frac{h(t)^2}{\sigma_0^2} + \mathcal{O}(\epsilon).$$

Note that

$$\begin{aligned} 1 + \epsilon\frac{h(t)^2}{\sigma_0^2} &= 1 + \epsilon\left(\frac{(\mu_0 - \mu_1)^2 - 2(\mu_0 - \mu_1)A\sin\omega t + (A\sin\omega t)^2}{\sigma_0^2}\right) \\ &> 1 + \epsilon\left(\frac{(\mu_0 - \mu_1)^2 - 2|\mu_0 - \mu_1||A|}{\sigma_0^2}\right) \\ &\simeq 1 + \epsilon\left(\frac{(\mu_0 - \mu_1)^2 - 2|\mu_0 - \mu_1|\sigma_0/\sqrt{\epsilon}}{\sigma_0^2}\right). \end{aligned}$$

In the last approximation, we used $A \simeq \sigma_0/\sqrt{\epsilon}$. Then, we use that the following quadratic function g satisfies for $\nu \geq 0$

$$g(x) = x^2 - 2|x|\sigma_0/\sqrt{\epsilon} > g((\sqrt{\nu}+1)\sigma_0/\sqrt{\epsilon}), \quad |x| > (\sqrt{\nu}+1)\sigma_0/\sqrt{\epsilon}.$$

This completes the proof. □

4. Conclusions

In this paper, in contrast to other works, we have shown how the optimal approximation by a Gaussian function to a probability density function can be achieved. Then, we have shown that the nearly Gaussian function that consists of a Gaussian function and a small oscillating density function can be used for the description of the tail of a distribution by showing that its variance can be greater than that of the Gaussian one.

Finally, we propose two avenues for further research: First, it would be interesting to make this optimal approximation method more expansive so that we can approximate probability density functions with more complicated shapes, such as density functions with several peaks, unlike the one in this paper. We reason that it could be achieved by splitting the given distribution to several sub-distributions with only one peak so that the suitable linear combination of the sub-distributions can reproduce the original distribution, and by optimally approximating the sub-distributions by Gaussian functions. Second, it would be also interesting to consider near Gaussian function with various types of perturbations, and see how the changes affect the mean and variance of the distribution functions.

References

[1] S. T. Rachev (ed.), *H*andbook of heavy tailed distributions in finance, Elsevier Science B.V., New York, 2003.

[2] J. McDonald and Y. Xu, *A generalization of the beta distribution with applications*, Journal of Economics, **66**, (1995). 133-152.

[3] P. Theodossiou, *Financial data and the skewed generalized-t distribution*, Management Science, **44**, (1998). 1650-1661.

[4] R. D. Harris and C. Coskun Küçüközmen, *T*he empirical distribution of stock returns: evidence from an emerging European markets, Applied Economics Letters, 8:6, (2001). 367-371.

[5] T. Kariya, Y. Tsukuda, J. Matsue and K. Omaki, *A*n extensive analysis on the Japanese markets via S. Taylor's model, Financial Engineering snd the Japanese markets, **2**(1), (1995). 15-86.

[6] Y. Nagahara, *N*on-Gaussian distribution for stock returns and related stochastic differential equation, Financial Engineering and the Japanese markets, **3**, (1996). 121-149.

[7] A. Behr and U. Pötter, *A*lternatives to the normal model of stock returns: Gaussian mixture, generalised logF and generalised hyperbolic models, Annals of Finance, **5**, (2009). 49-68.

[8] C. G. Corlu, M. Metereliyoz and M. Tiniç, *E*mpirical distributions of daily equity index returns: A comparison, Expert Systems With Applications, **54**, (2016). 170-192.

Gwang Il Kim
Department of Mathematics, Gyeongsang National University, Jinju, 52828, Republic of Korea
Email address: `gikim@gnu.ac.kr`

Seung Yeon Cho
Department of Mathematics, Gyeongsang National University, Jinju, 52828, Republic of Korea
Email address: `chosy89@gnu.ac.kr`

Doobae Jun
Department of Mathematics, Gyeongsang National University, Jinju, 52828, Republic of Korea
Email address: `dbjun@gnu.ac.kr`