No. 8

Kohei Marumo* and Rodney Wolff†

Density Approximation with Orthogonal Expansions Under Constraints

March 2015

*marumo@mail.saitama-u.ac.jp. WH Bryan Mining and Geology Research Centre, The University of Queensland, Brisbane QLD 4072, AUSTRALIA and Faculty of Economics, Saitama University, JAPAN.

†rodney.wolff@uq.edu.au. WH Bryan Mining and Geology Research Centre, The University of Queensland Australia

Abstract

We discuss the use of the orthogonal expansions for the approximation of probability density functions under constraints. For given independent and identically distributed samples, it has been shown by Marumo and Wolff (2013) that smooth functions that approximate the empirical distribution function can be constructed using orthogonal polynomial expansion. The approximation of the density function is given as the derivative of this smooth function.

In this paper, we show that the approximations under constraints on the moments and risk measures such as Value at Risk and Expected Shortfall can be constructed using orthogonal expansions.

The biggest advantage of our method over the typical existing methods is that our method has an explicit formula for the constrained approximations, whereas others often require intensive numerical calculations.
Keywords: Hermite polynomials; Orthogonal expansion; Quadratic programming; Smoothing methods.

1 Introduction

In this paper, we discuss the approximation of a density under constraints using orthogonal expansions, as introduced by Marumo and Wolff (2013). Given a set of independent and identically distributed (iid) observations, we have a variety of choices for methods for approximating or estimating the underlying density function: the use of kernels, splines, mixture distributions, orthogonal polynomials and wavelets are among the most popular methods (see Silverman (1986) for density estimation methods). As pointed out by Hall and Presnell (1999), we sometimes wish densities to have particular characteristics. For example, densities with a mean and variance that are identical to the sample mean and variance can be useful for simulation. Such desire can be the motivation for density approximation under constraints.

Let $X_1, \ldots, X_N$, be iid observations from density f. Hall and Presnell (1999) use the weighted kernel estimator,

$$\tilde{f}(x|\boldsymbol{p}) = \frac{1}{h}\sum_{i=1}^{N} p_i K\left(\frac{x - X_i}{h}\right),$$

where K, h and $\boldsymbol{p} = (p_1, \ldots, p_N)$ are the kernel, its bandwidth, and the weights that sum up to the unity. Their aim is to find the $\boldsymbol{p}$ that minimises the power-divergence distance between $\boldsymbol{p}$ and the uniform weights $(1/N, \ldots, 1/N)$, given by

$$D_\rho(\boldsymbol{p}) = \frac{N - \sum_{i=1}^{N}(Np_i)^\rho}{\rho(1-\rho)},$$

where $\rho \in \mathbb{R}$ is a parameter, under the constraints. They show that, if the constraints are of the form

$$T_j(\tilde{f}(\cdot|\boldsymbol{p})) = t_j,\ j = 0, \ldots, m, \tag{1}$$

where T_j and t_j are linear functionals and constants, respectively, then Lagrange multipliers can be applied to this optimisation problem, and that the optimal $\boldsymbol{p}$ can be obtained by solving a system of non-linear equations determined by the constraints. They pointed out that constraints on moments and quantiles can be expressed in the form of Equation (1). They further consider the constraints on the entropy and by adjusting the entropy and bandwidth they obtain a unimodal density.

Eloyan and Ghosh (2011) use the mixture distribution of the form

$$\tilde{f}_n(x|\boldsymbol{p}_n) = \sum_{k=1}^{n} p_{k,n} f_{k,n}(x),$$

where $n \geq 1$ is the number of mixture components, and $f_{k,n}$ are some known densities. They use an iteration procedure called the Expectation Maximisation Algorithm in order to obtain the weights $\boldsymbol{p}_n$ that minimise the empirical Kullback-Leibler divergence,

$$\frac{1}{N} \sum_{i=1}^{N} \log \frac{f(X_i)}{\tilde{f}_n(X_i|\boldsymbol{p}_n)},$$

or equivalently, maximise the log-likelihood, under the constraints on the moments.

Musso and Oudjane (2009) use a kernel that can take negative values, in order to reduce the bias in the mean integrated squared error, given by

$$\mathrm{E}\left(\int_{-\infty}^{\infty} \{\tilde{f}(x) - f(x)\}^2 \mathrm{d}x\right).$$

Their estimator takes the form

$$\tilde{f}(x|\alpha_1, \ldots, \alpha_m) = \left(\frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right) - \sum_{r=1}^{m} \alpha_r f_r(x)\right)_+,$$

where f_r are real functions determined by the constraints, and $(\cdot)_+$ denotes $\max(\cdot, 0)$. They suggest that we estimate $\alpha_1, \ldots, \alpha_m$ by numerical integration, or by Monte Carlo, while they show that analytical solutions are available in some special cases.

Marumo and Wolff (2013) use orthogonal expansions to approximate the empirical distribution function. They show that the empirical distribution function can be approximated by a smooth function using the orthogonal expansions, and the associated density function can be obtained from the approximation. In the remainder of this paper, we discuss the application of the orthogonal expansion to approximating density functions under constraints on the moments and risk measures. Section 2 reviews the orthogonal expansion method. Section 3 introduce the approximation under constraints, and Section 4 shows the examples. We conclude in Section 5.

2 Orthogonal expansion method

We briefly review the method introduced by Marumo and Wolff (2013). For simplicity, let us assume that the samples $X_1, \ldots, X_N$ are standardised[1], that is, $\bar{X} = \sum_{i=1}^{n} X_i/N = 0$ and $s_X^2 = \sum_{i=1}^{N}(X_i - \bar{X})^2/N = 1$.

2.1 Orthogonal expansion

Let us consider the orthogonal expansion of the form,

$$\tilde{f}_n(x|\boldsymbol{c}_n) = \phi(x)\sum_{k=0}^{n} c_k \mathrm{He}_k(x), \tag{2}$$

where $n \geq 0$ is the degree of expansion, $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ is the density function of the Standard Normal distribution, and

$$\mathrm{He}_k(x) = \frac{1}{\sqrt{k!}}\frac{1}{\phi(x)}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\phi(x),\ \ k = 0, \ldots, n, \tag{3}$$

are the Hermite polynomials. Our aim here is to determine the coefficients $\boldsymbol{c}_n = (c_0, \ldots . c_n)'$ that make $\tilde{f}_n(\cdot|\boldsymbol{c}_n)$ a 'good' approximation in respect of the density function that generated the samples. Note that from the condition $\int_{-\infty}^{\infty} \tilde{f}_n(x|\boldsymbol{c}_n)\mathrm{d}x = 1$, we always have $c_0 \equiv 1$.

2.2 Naïve expansion

From Equations (2) and (3), the distribution function can be expressed as,

$$\tilde{F}_n(x|\boldsymbol{c}_n) = \int_{-\infty}^{x} \tilde{f}_n(u|\boldsymbol{c}_n)\mathrm{d}u = \Phi(x) + \phi(x)\sum_{k=1}^{n}\frac{c_k}{\sqrt{k}}\mathrm{He}_{k-1}(x), \tag{4}$$

where $\Phi(x) = \int_{-\infty}^{x}\phi(u)\mathrm{d}u$ is the distribution function of the Standard Normal distribution.

Define the empirical distribution function as

$$F^N(x) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{X_i \leq x\}},$$

and consider the weighted integrated squared difference (WISD) defined by

$$\mathrm{WISD} = \int_{-\infty}^{\infty}\frac{\{\tilde{F}_n(x|\boldsymbol{c}_n) - F^N(x)\}^2}{\phi(x)}\mathrm{d}x = \sum_{k=1}^{n}\frac{(c_k - \hat{c}_k)^2}{k} + \sum_{k=n+1}^{\infty}\frac{\hat{c}_k^2}{k},$$

where the secnd equality is given by considering the Parseval identity (see Marumo and Wolff (2013)), and $\hat{c}_k$ are the naïve estimators defined by,

$$\hat{c}_k = \frac{1}{N}\sum_{i=1}^{N}\mathrm{He}_k(X_i),\ \ k = 0, \ldots, n.$$

[1] See Marumo and Wolff (2013) for this standardisation.

Thus the WISD is minimised when $\boldsymbol{c}_n = \hat{\boldsymbol{c}}_n$, where the components of $\hat{\boldsymbol{c}}_n$ are the naïve estimators. Note that if the samples are standardised as suggested above, we have $\hat{c}_1 = \hat{c}_2 = 0$.

They show that $\tilde{F}_n(\cdot|\hat{\boldsymbol{c}}_n) \to F^N$ as $n \to \infty$, whereas that $\tilde{f}_n(\cdot|\hat{\boldsymbol{c}}_n)$ is divergent. Let us call $\tilde{F}_n(\cdot|\hat{\boldsymbol{c}}_n)$ and $\tilde{f}_n(\cdot|\hat{\boldsymbol{c}}_n)$ the naïve expansions of the distribution and density functions, respectively.

2.3 Smoothed expansion

From Equations (3) and (4), the second derivative of the distribution function is expressed as

$$\tilde{F}_n''(x|\boldsymbol{c}_n) = \phi(x) \sum_{k=0}^{n} \sqrt{k+1} c_k \mathrm{He}_{k+1}(x).$$

Define the weighted integrated squared curvature (WISC) by $\int_{-\infty}^{\infty} \{\tilde{F}_n''(x|\boldsymbol{c}_n)\}^2/\phi(x)\,\mathrm{d}x$, which is calculated to be

$$\mathrm{WISC} = \int_{-\infty}^{\infty} \frac{\{\tilde{F}_n''(x|\boldsymbol{c}_n)\}^2}{\phi(x)} \mathrm{d}x = \sum_{k=0}^{n} (k+1) c_k^2,$$

and consider the average of the two integrals weighted by $0 \le q \le 1$, as

$$\begin{aligned}(1-q)\mathrm{WISD} + q\mathrm{WISC} \\ = \sum_{k=1}^{n} \left(\frac{1-q+qk(k+1)}{k} c_k^2 - 2\frac{(1-q)\hat{c}_k}{k} c_k \right) + (1-q) \sum_{k=1}^{\infty} \frac{\hat{c}_k^2}{k} + q.\end{aligned} \tag{5}$$

This is minimised when

$$c_k = \hat{c}_k^S(q) = \frac{1-q}{1-q+qk(k+1)} \hat{c}_k, \ k = 0, \ldots, n.$$

Let us call the expansions that use the coefficients $\hat{\boldsymbol{c}}_n^S(q) = (\hat{c}_0^S(q), \ldots, \hat{c}_n^S(q))'$ the smoothed expansions. Marumo and Wolff (2013) show that $\tilde{F}_n(\cdot|\hat{\boldsymbol{c}}_n^S(q))$ and $\tilde{f}_n(\cdot|\hat{\boldsymbol{c}}_n^S(q))$ are convergent for $q > 0$ as $n \to \infty$.

Note that $q = 0$ corresponds to the naïve expansion, and $q = 1$ to the approximation by the Standard Normal distribution.

2.4 Choice of n and q

In the smoothed expansion, n and q are arbitrary parameters. Choosing appropriate values for these parameters are somewhat similar to the problem of choosing appropriate kernel and bandwidth in kernel estimation and unfortunately both problems do not seem to have an unambiguous or definitive solution.

Choice of n

One idea is to choose n large enough so that $\tilde{F}_n$ is close to $\tilde{F}_\infty$. We can measure the difference between the two distribution function by the WISD as,

$$\int_{-\infty}^{\infty} \frac{\{\tilde{F}_n(x|\hat{\boldsymbol{c}}_n^S(q)) - \tilde{F}_\infty(x|\hat{\boldsymbol{c}}_\infty^S(q))\}^2}{\phi(x)} \mathrm{d}x = \sum_{k=n+1}^{\infty} \frac{\hat{c}_k^S(q)^2}{k}.$$

We normally do not know the value of this quantity, but observing the value of $\sum_{k=1}^{n} \hat{c}_k^S(q)^2/k$ for $n = 1, \ldots$ can be useful. We can stop increasing n when the growth of the sum is small enough.

Alternatively, if we wish a simple formula, a small n such as $n \leq 8$ may be of great use.

Choice of q

From the definition, the estimate is close to the empirical distribution when q is small. In this sense, smaller q is more desirable. However, from the fact that the density function is not convergent when $q = 0$, we can assume that the estimate can be erratic for small q; that is, we are more likely to encounter the negative density. One idea is that we choose the smallest q that does not suffer form the effect of negative density.

3 Approximation Under Constraints

Now, we consider the density approximation under constraints; that is, we firstly find $\boldsymbol{c}_n$ that let the associated distribution satisfy the constraints and be close to the empirical distribution. Then the density is given by $\tilde{f}(\cdot|\boldsymbol{c}_n)$.

As already mentioned, the constraint $c_0 = 1$ is always imposed and we regard it as constant. Hence, hereafter, we use the notation $\boldsymbol{c}_n = (c_1, \ldots, c_n)'$, from which c_0 is removed.

3.1 Methodology

Weighted average

We consider minimising the weighted average in Equation (5), under constraints. We define

$$\boldsymbol{c}_n^* = D_n \boldsymbol{c}_n,$$

where D_n is a diagonal matrix,

$$D_n = \mathrm{diag}\left(\sqrt{\frac{1 - q + qk(k+1)}{k}};\ k = 1, \ldots, n\right).$$

Then the weighted average can be rewritten as

$$\sum_{k=1}^{n}\left(c_k^* - \frac{(1-q)\hat{c}_k}{\sqrt{(1-q)k+qk^2(k+1)}}\right)^2 - \sum_{k=1}^{n}\frac{(1-q)^2\hat{c}_k^2}{(1-q)k+qk^2(k+1)} + (1-q)\sum_{k=1}^{\infty}\frac{\hat{c}_k^2}{k} + q,$$

where c_k^* are the components of $\boldsymbol{c}_n^*$.

Since only the first term is relevant to $\boldsymbol{c}_n^*$, we can define our objective function as

$$I_n(\boldsymbol{c}_n^*) = \sum_{k=1}^{n}\left(c_k^* - \frac{(1-q)\hat{c}_k}{\sqrt{(1-q)k+qk^2(k+1)}}\right)^2.$$

Now, $I_n(\boldsymbol{c}_n^*)$ can be considered as the squared radius of a sphere in n dimensional space, centred at

$$\boldsymbol{u}_n^* = \begin{pmatrix} \frac{(1-q)\hat{c}_1}{\sqrt{1+q}} \\ \vdots \\ \frac{(1-q)\hat{c}_n}{\sqrt{(1-q)n+qn^2(n+1)}} \end{pmatrix}.$$

Linear constraints

As we review later, many constraints including those on the moments, quantiles and Expected Shortfalls can be expressed as linear equations in terms of $c_1, \ldots, c_n$. Let us call such constraints linear constraints. Assume that we have a set of m linear constraints expressed as $A_n\boldsymbol{c}_n = \boldsymbol{a}$, where A_n is an $m \times n$ rank m matrix with $m \leq n$ and $\boldsymbol{a}$ is a constant vector. This set of constraints can be rewritten as $A_n^*\boldsymbol{c}_n^* = \boldsymbol{a}$, where $A_n^* = A_n D_n^{-1}$. Hence, $\boldsymbol{c}_n^*$ under the constraints forms a linear subspace Π_n^*, which is orthogonal to all of the row vectors of A_n^*, and is shifted by $\boldsymbol{a}$.

Optimisation

Under the above settings, $\boldsymbol{c}_n^* \in \Pi_n^*$ that minimises $I_n(\boldsymbol{c}_n^*)$ is given by the orthogonal projection of $\boldsymbol{u}^*$ onto Π_n^*. That is, the optimal $\boldsymbol{c}_n^*$ is given by

$$\boldsymbol{c}_n^{*\dagger} = (I - A_n^{*\prime}(A_n^*A_n^{*\prime})^{-1}A_n^*)\boldsymbol{u}_n^* + A_n^{*\prime}(A_n^*A_n^{*\prime})^{-1}\boldsymbol{a}.$$

Thus, the optimal $\boldsymbol{c}_n$ is given by

$$\boldsymbol{c}_n^\dagger = D_n^{-1}\boldsymbol{c}_n^{*\dagger} = (I - D_n^{-2}A_n'(A_nD_n^{-2}A_n')^{-1}A_n)\boldsymbol{u}_n + D_n^{-2}A_n'(A_nD_n^{-2}A_n')^{-1}\boldsymbol{a}, \tag{6}$$

where

$$\boldsymbol{u}_n = D_n^{-1}\boldsymbol{u}_n^* = \begin{pmatrix} \frac{1-q}{1+q}\hat{c}_1 \\ \vdots \\ \frac{1-q}{1-q+qn(n+1)}\hat{c}_n \end{pmatrix} = \begin{pmatrix} \hat{c}_1^S(q) \\ \vdots \\ \hat{c}_n^S(q) \end{pmatrix},$$

and

$$D_n^{-2} = (D_n^{-1})^2 = \mathrm{diag}\left(\frac{k}{1-q+qk(k+1)};\ k=1,\ldots,n\right).$$

Note that this is among the simplest cases of the quadratic programming problem. If we do not cling on the explicit formula in Equation (6), we can generalise the constraints to inequality constraints, where the constraints are expressed as a half-subspace.

Convergence

The convergence of expansions using $\boldsymbol{c}_n^\dagger$ as $n \to \infty$ can be verified as follows. For $n > m$, we have

$$\begin{aligned} |\boldsymbol{u}_{n+1}^* - \boldsymbol{c}_n^{*\dagger}|^2 &= |\boldsymbol{u}_{n+1}^* - \boldsymbol{u}_n^*|^2 + |\boldsymbol{u}_n^* - \boldsymbol{c}_n^{*\dagger}|^2 \\ &= u_{n+1}^{*2} + I_n^\dagger, \end{aligned}$$

where we consider the n-dimensional vectors as being in the $\mathbb{R}^n$-subspace in $\mathbb{R}^{n+1}$, that is, we consider $\boldsymbol{c}_n^{*\dagger} = (c_{n,1}^{*\dagger}, \ldots, c_{n,n}^{*\dagger}, 0)'$, and so on, and $I_n^\dagger = I_n(\boldsymbol{c}_n^{*\dagger})$. See Figure 1. By considering the facts, $\boldsymbol{c}_n^{*\dagger} \in \Pi_n^* \subseteq \Pi_{n+1}^*$ and $\boldsymbol{c}_{n+1}^{*\dagger} \in \Pi_{n+1}^*$, we have

$$\begin{aligned} |\boldsymbol{u}_{n+1}^* - \boldsymbol{c}_{n+1}^{*\dagger}|^2 &\leq |\boldsymbol{u}_{n+1}^* - \boldsymbol{c}_n^{*\dagger}|^2 \\ I_{n+1}^\dagger &\leq u_{n+1}^{*2} + I_n^\dagger, \end{aligned}$$

which leads to

$$I_{n+1}^\dagger - I_n^\dagger \leq u_{n+1}^{*2}, \tag{7}$$

$$0 \leq I_n^\dagger \leq I_m^\dagger + \sum_{k=m+1}^{n} u_k^{*2}. \tag{8}$$

Using $\lim_{n\to\infty} \hat{c}_n/\sqrt{n} = 0$, and $0 \leq \sum_{k=1}^\infty \hat{c}_n^2/n < \infty$ shown by Marumo and Wolff (2013), and $|u_n^*| < |\hat{c}_n/\sqrt{n}|$ for large n, we have $\lim_{n\to\infty} u_n^{*2} = 0$ and $0 \leq \sum_{k=m+1}^\infty u_k^{*2} < \infty$. By considering the limit $n \to \infty$ in Inequalities (7) and (8), we find that $I_n^\dagger$ is convergent. Since $I_n^\dagger = |\boldsymbol{u}_n^* - \boldsymbol{c}_n^{*\dagger}|^2$ and $\lim_{n\to\infty} |\boldsymbol{u}_n^*|^2 < \infty$, we have $\sum_{k=1}^\infty c_k^{*\dagger 2} < \infty$. From definition, we have $c_k^{*\dagger 2} > c_k^{\dagger 2}$ for large k, and hence $\sum_{k=1}^\infty c_k^{\dagger 2} < \infty$. This suggests that $\tilde{f}_n(\cdot|\boldsymbol{c}_n^\dagger)$ is convergent.

3.2 Comparison with other methods

Here, we compare the proposed method with others already reviewed above.

The biggest advantage of the proposed method is that it has the explicit formula given by Equation (6) for the optimised estimators. Optimisation is achieved using elementary linear algebra. This makes a contrast to other methods: Hall and Presnell (1999) require solving numerically a system of non-linear equations, Eloyan and Ghosh (2011) use an iteration algorithm,

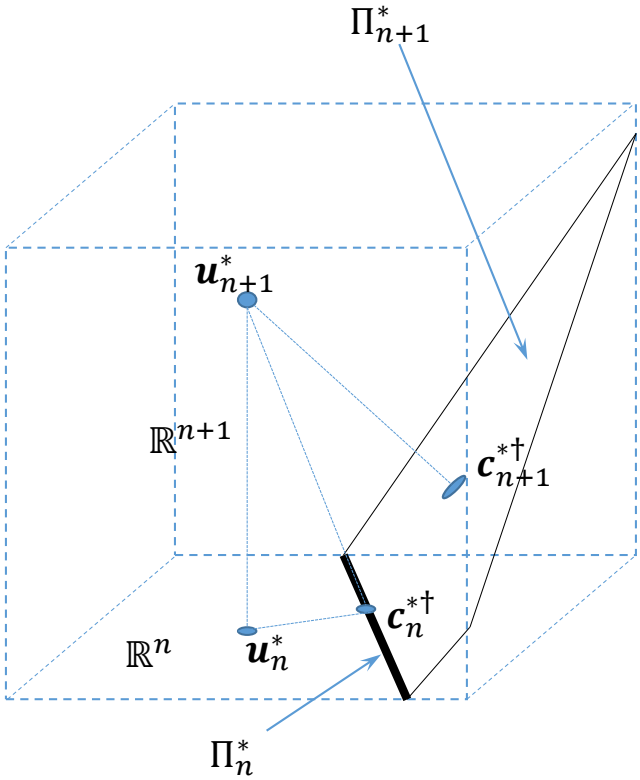

Figure 1: Illustrative diagram (for $n = 2$). $\boldsymbol{u}^*_n$, $\boldsymbol{c}^{*\dagger}_n$ and Π^*_n are in the $\mathbb{R}^n$-subspace, while others are in the $\mathbb{R}^{n+1}$-space. We use the fact $(\boldsymbol{u}^*_{n+1} - \boldsymbol{u}^*_n) \perp (\boldsymbol{c}^{*\dagger}_n - \boldsymbol{u}^*_n)$.

and Musso and Oudjane (2009) use numerical integration, in order to obtain the estimates.

While arbitrariness can sometimes provide extra degree of freedom, it is often considered as a problem. The methods using the kernel have arbitrariness in choosing the kernels and the bandwidth. Guidelines for making these choices exist, as pointed out by Musso and Oudjane (2009), however, the choice of the bandwidth is often made by many trials, as done by Hall and Presnell (1999). The arbitrariness of the proposed method is those of choosing n and q. We might use the idea mentioned in Subsection 2.4.

The biggest disadvantage of the proposed method is the effect of possible negative density. As shown later, we find from numerical examples that we can mitigate the effect by choosing n and q appropriately, although it is difficult to eliminate it completely.

4 Examples

We apply the method explained in the previous Section to some examples.

4.1 Constraints on the moments

Assume that the random variable $\tilde{X}$ has the density function of the form in Equation (2). It can be easily shown that

$$\mathrm{E}(\mathrm{He}_k(\tilde{X})) = \int_{-\infty}^{\infty} \mathrm{He}_k(x)\tilde{f}_n(x|\boldsymbol{c}_n)\mathrm{d}x = c_k,$$

for $k = 0, \ldots, n$. On the other hand, we have

$$\mathrm{E}(\mathrm{He}_k(\tilde{X})) = h_{k,0}\tilde{\mu}_0 + h_{k,1}\tilde{\mu}_1 + \cdots + h_{k,k}\tilde{\mu}_k,$$

where $h_{k,r}$ are the rth coefficient of kth Hermite polynomial and $\tilde{\mu}_r$ is the rth moment of $\tilde{X}$. Hence, the moments and the coefficients have a linear relationship of the form,

$$\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} h_{0,0} & & & \\ h_{1,0} & h_{1,1} & & 0 \\ \vdots & & \ddots & \\ h_{n,0} & \cdots & \cdots & h_{n,n} \end{pmatrix} \begin{pmatrix} \tilde{\mu}_0 \\ \tilde{\mu}_1 \\ \vdots \\ \tilde{\mu}_n \end{pmatrix}. \tag{9}$$

Since we have $c_0 = h_{0,0} = \tilde{\mu}_0 = 1$, Equation (9) can be rewritten as

$$\boldsymbol{c}_n = \boldsymbol{h}_0 + H_n\tilde{\boldsymbol{\mu}}_n, \tag{10}$$

where $\boldsymbol{h}_0 = (h_{1,0}, \ldots, h_{n,0})'$, $\tilde{\boldsymbol{\mu}}_n = (\tilde{\mu}_1, \cdots, \tilde{\mu}_n)'$ and $H_n = (h_{k,r};\ k = 1, \ldots, n,\ r = 1, \ldots, n)$. It is obvious that the matrix H_n is non-singular.

Setting the lower moments to certain values

Sometimes we wish the approximation to have the same mean, variance, skewness and kurtosis as those of the samples. We learn from Equation (9) that this constraint can be easily imposed by letting $\boldsymbol{c}_n = (\hat{c}_1, \ldots, \hat{c}_4, \hat{c}_5^S(q), \ldots, \hat{c}_n^S(q))'$.

Other linear moment constraints

A set of linear constraints on the moments can be written in a form, $B\tilde{\boldsymbol{\mu}}_n = \boldsymbol{b}$, where B and $\boldsymbol{b}$ are a constant matrix and vector, respectively. Using the relationship in Equation (10), this can be written in terms of that of $\boldsymbol{c}_n$ as $BH_n^{-1}\boldsymbol{c}_n = \boldsymbol{b} + BH_n^{-1}\boldsymbol{h}_0$. The optimal $\boldsymbol{c}_n$ is given by Equation (6) with $A = BH_n^{-1}$ and $\boldsymbol{a} = \boldsymbol{b} + BH_n^{-1}\boldsymbol{h}_0$.

4.2 Constraints on risk measures

In applications to financial risk management, we are often interested in risk measures associated with the profit and loss distribution. Among a number of risk measures, the Value at Risk (VaR) and Expected Shortfall (ES) are the most commonly discussed. The VaR is defined as the quantile of the profit and loss distribution, and the ES, as the expected loss exceeding the VaR. [2]

[2] See McNeil et al. (2005) and Basel Committee on Banking Supervision (2013).

Density approximations whose risk measures are the same as those of the empirical distribution obtained from the markets can be useful, for instance, for simulating possible future profit and loss. Hall and Presnell (1999) consider the density estimations under constraints on quantile, mainly from the statistical interest.

Constraints on Value at Risk

Suppose that X is a random variable that denotes the profit and loss of the portfolio. Then the VaR with a confidence level α is defined as

$$\mathrm{VaR}_\alpha = -\sup\{x | P(X \le x) \le 1-\alpha\}.$$

Let $\tilde{F}_n(\cdot|\boldsymbol{c}_n)$ be our approximation for the distribution function of X. Then the constraint that sets the VaR a certain value $-x_\alpha$ can be expressed as $\tilde{F}_n(x_\alpha|\boldsymbol{c}_n) = 1-\alpha$. This can be written as a linear constraint of $\boldsymbol{c}_n$ as

$$\sum_{k=1}^{n} \frac{\mathrm{He}_{k-1}(x_\alpha)}{\sqrt{k}} c_k = \frac{1-\alpha-\Phi(x_\alpha)}{\phi(x_\alpha)}.$$

Therefore it is trivial to apply the method explained above to this case.

Constraints on Expected Shortfall

Given $\mathrm{VaR}_\alpha = -x_\alpha$, the ES with the confidence level α is defined as

$$\mathrm{ES}_\alpha = -\mathrm{E}(X|X \le x_\alpha).$$

The ES can be expressed using an orthogonal expansion as

$$\begin{aligned}
\mathrm{ES}_\alpha &= -\frac{\int_{-\infty}^{x_\alpha} u \mathrm{d}\tilde{F}_n(u|\boldsymbol{c}_n)}{1-\alpha} \\
&= -\frac{1}{1-\alpha}\Bigg[\phi(x_\alpha) + \{\phi(x_\alpha)x_\alpha - \Phi(x_\alpha)\}c_1 \\
&\qquad + \phi(x_\alpha)\sum_{k=2}^{n}\left\{\frac{x_\alpha \mathrm{He}_{k-1}(x_\alpha)}{\sqrt{k}} - \frac{\mathrm{He}_{k-2}(x_\alpha)}{\sqrt{k(k-1)}}\right\}c_k\Bigg].
\end{aligned}$$

Hence, the constraint $\mathrm{ES}_\alpha = -\eta_\alpha$, for some constant η_α, can be expressed as a linear constraint of $\boldsymbol{c}_n$ as

$$\left\{x_\alpha - \frac{\Phi(x_\alpha)}{\phi(x_\alpha)}\right\}c_1 + \sum_{k=2}^{n}\left\{\frac{x_\alpha \mathrm{He}_{k-1}(x_\alpha)}{\sqrt{k}} - \frac{\mathrm{He}_{k-2}(x_\alpha)}{\sqrt{k(k-1)}}\right\}c_k = -1 - \frac{1-\alpha}{\phi(x_\alpha)}\eta_\alpha.$$

4.3 Numerical examples

Let us review numerical examples of the proposed method.

Data sets

We use four data sets from the four markets shown in Table 1. We take daily returns of the equity indices and treat them as iid samples.[3]

	S&P 500	FT 100	Nikkei 225	AS 51
Start Date	2 Oct 2012	8 Oct 2012	13 Sep 2012	8 Oct 2012
End Date	30 Sep 2014	30 Sep 2014	30 Sep 2014	30 Sep 2014
Number of Samples	500	500	500	500
Mean ($\times 10^{-4}$)	6.211	2.510	11.73	3.326
Standard Deviation ($\times 10^{-3}$)	6.931	6.887	14.46	6.996
Skewness	-0.487	-0.137	-0.645	-0.232
Kurtosis	4.340	4.681	5.608	3.855
Max ($\times 10^{-2}$)	2.509	3.032	4.826	2.593
Min ($\times 10^{-2}$)	-2.533	-3.027	-7.597	-2.354

Table 1: Summary statistics of the samples

Constraints

We consider the following two constraints:

Constraint 1 The mean, variance, skewness and kurtosis are identical to those of the samples.

Constraint 2 In addition to Constraint 1, 97.5% VaR and ES are identical to those of the samples.

Choice of n

We calculate $\sum_{k=1}^{n} c_k^2/k$ for up to $n = 100$, for $q = 0.001, 0.01$ and 0.05 (upper plots in Figures 2 to 5). From these plots, we find that $n = 60$ can be close to the convergence in all cases.

Choice of q

We firstly define the 'range of approximation' as $\pm 6\sigma$, where σ is the sample standard deviation. We then numerically searched for the smallest q with

	S&P 500	FT 100	Nikkei 225	AS 51
Unconstrained	0.0441	0.0132	0.3801	0.0558
Constraint 1	0.0075	0.0109	0.0163	0.0105
Constraint 2	0.0049	0.0095	0.0200	0.0068

Table 2: The smallest q that realise non-negative density within the range of approximation ($\pm 6\sigma$).

which the density function is non-negative within the range of approximation. The value of q obtained is shown in Table 2.

The range of approximation $\pm 6\sigma$ is large enough to contain all of the samples with margins, and we use the values in Table 2 in the lower plots in Figures 2 to 5.[4] However, we find that widening the range of approximation can lead to a sharp rise in minimum q, or difficulty in finding appropriate value of q. See Table 3 for the example with the range of approximation $\pm 10\sigma$.

	S&P 500	FT 100	Nikkei 225	AS 51
Unconstrained	1.0000	1.0000	1.0000	1.0000
Constraint 1	0.9873	0.9969	0.9997	0.9950
Constraint 2	-	-	-	-

Table 3: The smallest q that realise non-negative density within the range of approximation ($\pm 10\sigma$). The '-' symbol denotes that the value of q that realise non-negative density is not found by our numerical search.

Density and distribution functions

The lower plots in Figures 2 to 5 show the approximations for the density function and distribution function. The plots for the distribution function show that Constraint 1, constraints on up to the fourth moments, can make the approximation closer to the empirical distributions. Further, the approximations with Constraint 2 are closest to the empirical distributions, as expected.

Impact of negative density

By construction, the approximations may exhibit negative density in the outside of the range of approximation. In our numerical examples, the size of negative density is around 10^{-8} at most, and we may consider it negligible compared to the densities in the range of approximation. See Figures 6 to 9 and Table 4.

[3] We are aware of possible existence of serial dependence structures in the samples; however, we ignore it here for simplicity. Marumo and Wolff (2008) consider approximating the distribution functions where the samples have serial dependence structures.

[4] For the Normal distribution, for example, -6σ corresponds to the 9.8659×10^{-10} quantile.

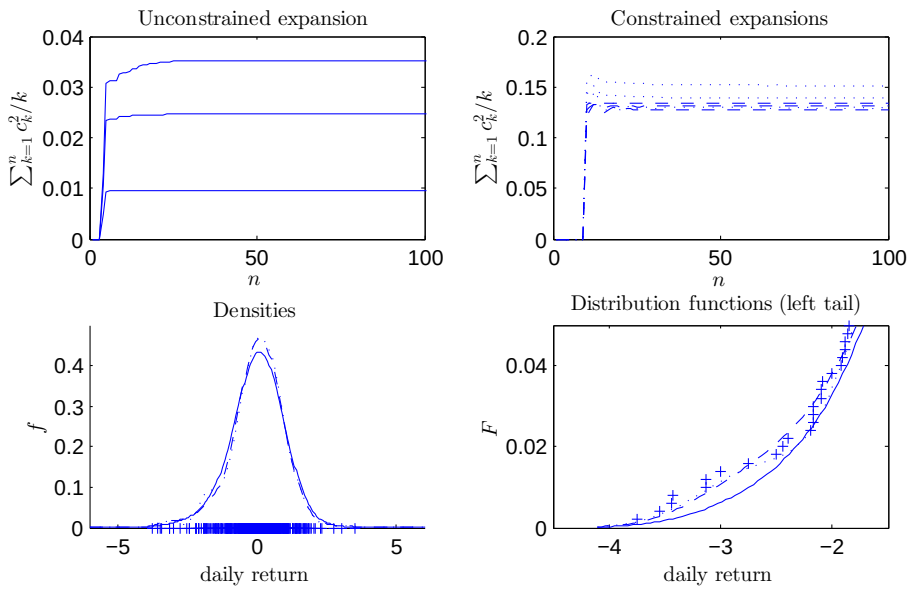

Figure 2: Results for S&P 500 daily returns (standardised). Upper plots show $\sum_{k=1}^{n} c_k^2/k$ against n. The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraints 1 and 2, respectively. The value of q is set to 0.001, 0.01 and 0.05, from the top to bottom curves. The lower left plots show the density functions with $n = 60$. For unconstrained approximation (solid) and approximations with Constraints 1 (dashed) and 2 (dotted), the value of q is set to 0.0441, 0.0075 and 0.0049, respectively. The '+' marks on the horizontal axis correspond to the samples. The lower right plots show the left tails of the distribution functions. The '+' marks correspond to the empirical distribution function.

5 Discussion

In this paper, we propose a method for approximating density function under constraint using orthogonal expansions, given an iid data set. We show that the explicit formula for the approximation is available, as long as the constraints are expressed as a set of linear equations. This is the advantage over other methods reviewed above.

The biggest disadvantage of the method is possible existence of negative density. The numerical examples demonstrate that we can mitigate its effect, while it is difficult to eliminate it completely. Strictly speaking, in order to ensure that the density is non-negative, we need to investigate the condition on $c_1, \ldots, c_n$ with which the polynomial $\sum_{k=0}^{n} c_n \mathrm{He}_k(x)$ does not have roots. However, it is a tricky task to derive the necessary and sufficient condition when n is large. Instead, finding some sufficient condition or focusing on small n may be useful. This is our future task.

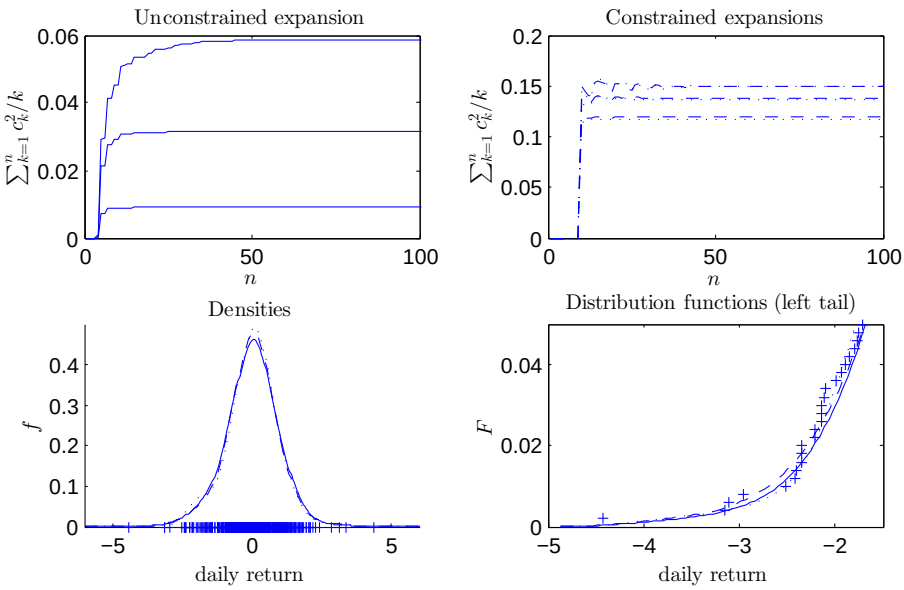

Figure 3: Results for FT 100 daily returns (standardised). Upper plots show $\sum_{k=1}^{n} c_k^2/k$ against n. The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraints 1 and 2, respectively. The value of q is set to 0.001, 0.01 and 0.05, from the top to bottom curves. The lower left plots show the density functions with $n = 60$. For unconstrained approximation (solid) and approximations with Constraints 1 (dashed) and 2 (dotted), the value of q is set to 0.0132, 0.0109 and 0.0095, respectively. The '+' marks on the horizontal axis correspond to the samples. The lower right plots show the left tails of the distribution functions. The '+' marks correspond to the empirical distribution function.

($\times 10^{-8}$)	S&P 500	FT 100	Nikkei 225	AS 51
Unconstrained	-0.0394	-1.0594	-0.0385	-0.0755
Constraint 1	-0.1578	-1.1524	-1.2519	-0.1110
Constraint 2	-0.7928	-1.3945	-0.9081	-0.2204

Table 4: The largest negative densities. See Figures 6 to 9.

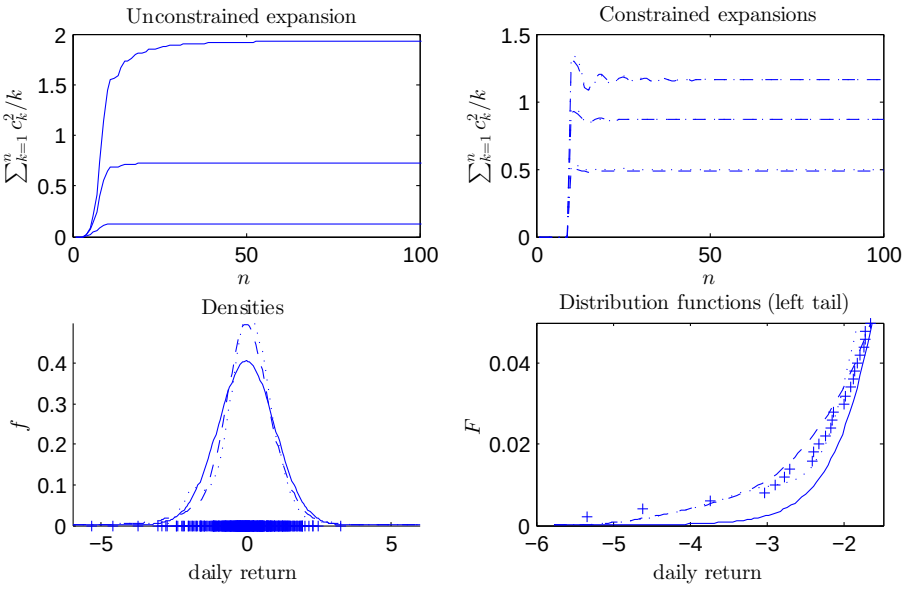

Figure 4: Results for Nikkei 225 daily returns (standardised). Upper plots show $\sum_{k=1}^{n} c_k^2/k$ against n. The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraints 1 and 2, respectively. The value of q is set to 0.001, 0.01 and 0.05, from the top to bottom curves. The lower left plots show the density functions with $n = 60$. For unconstrained approximation (solid) and approximations with Constraints 1 (dashed) and 2 (dotted), the value of q is set to 0.3801, 0.0163 and 0.0200, respectively. The '+' marks on the horizontal axis correspond to the samples. The lower right plots show the left tails of the distribution functions. The '+' marks correspond to the empirical distribution function.

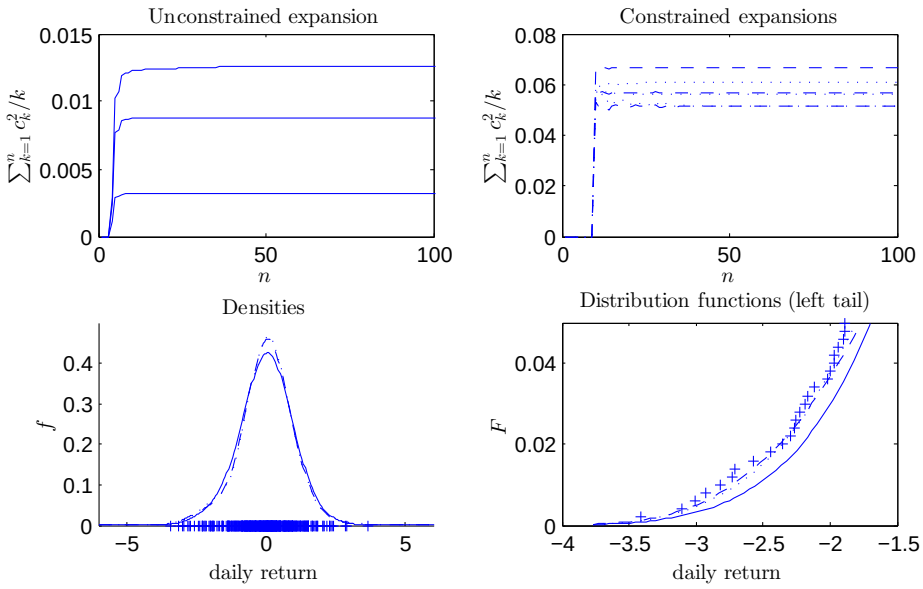

Figure 5: Results for AS 51 daily returns (standardised). Upper plots show $\sum_{k=1}^{n} c_k^2/k$ against n. The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraints 1 and 2, respectively. The value of q is set to 0.001, 0.01 and 0.05, from the top to bottom curves. The lower left plots show the density functions with $n = 60$. For unconstrained approximation (solid) and approximations with Constraints 1 (dashed) and 2 (dotted), the value of q is set to 0.0558, 0.0105 and 0.0068, respectively. The '+' marks on the horizontal axis correspond to the samples. The lower right plots show the left tails of the distribution functions. The '+' marks correspond to the empirical distribution function.

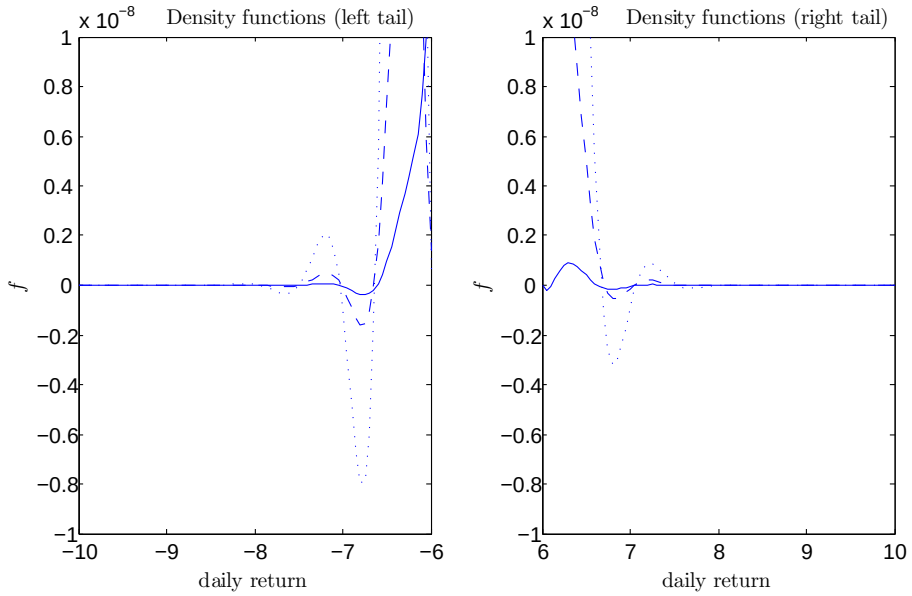

Figure 6: Far tails of the density functions for S&P 500 daily returns (standardised). The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraint 1 and 2, respectively. The value for n and p are the same as those in Figure 2.

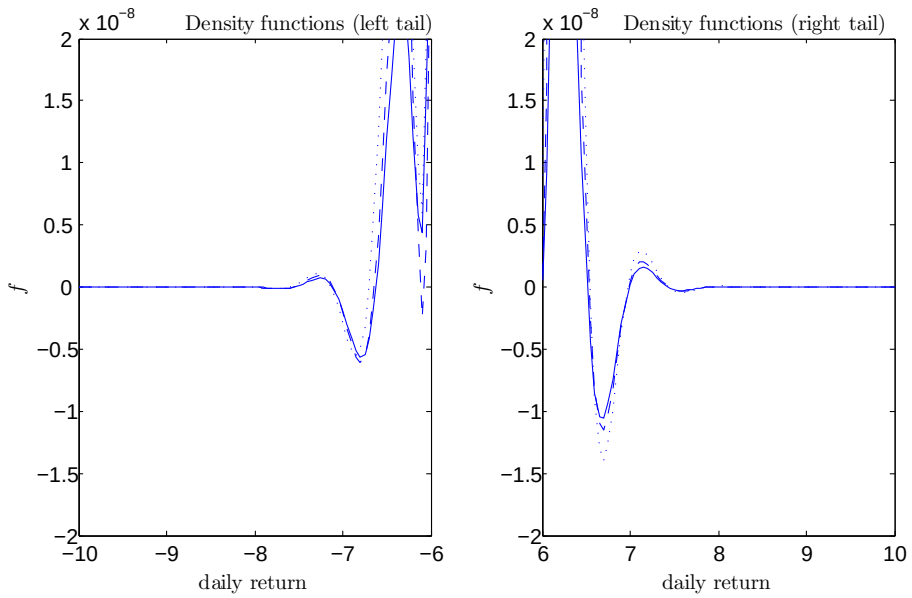

Figure 7: Far tails of the density functions for FT 100 daily returns (standardised). The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraint 1 and 2, respectively. The value for n and p are the same as those in Figure 3.

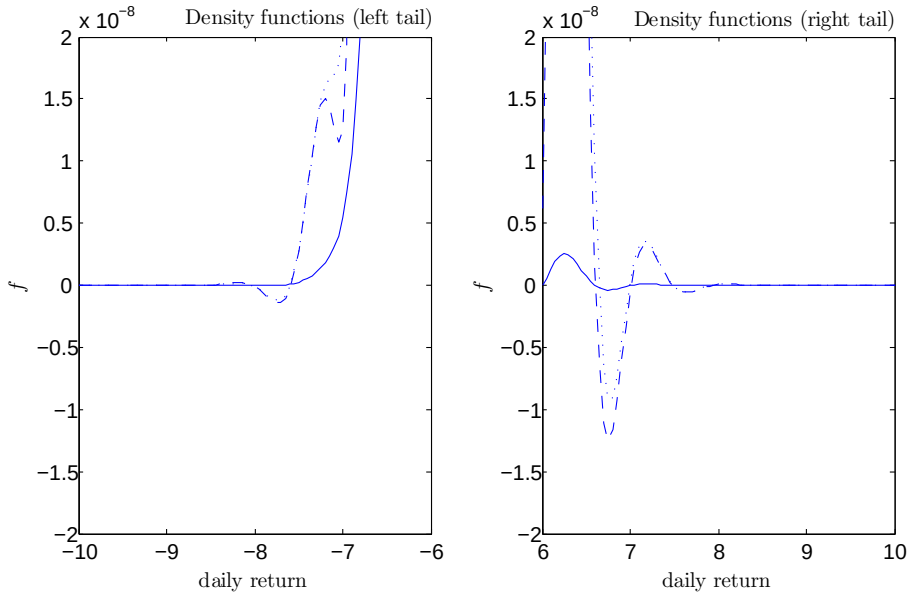

Figure 8: Far tails of the density functions for Nikkei 225 daily returns (standardised). The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraint 1 and 2, respectively. The value for n and p are the same as those in Figure 4.

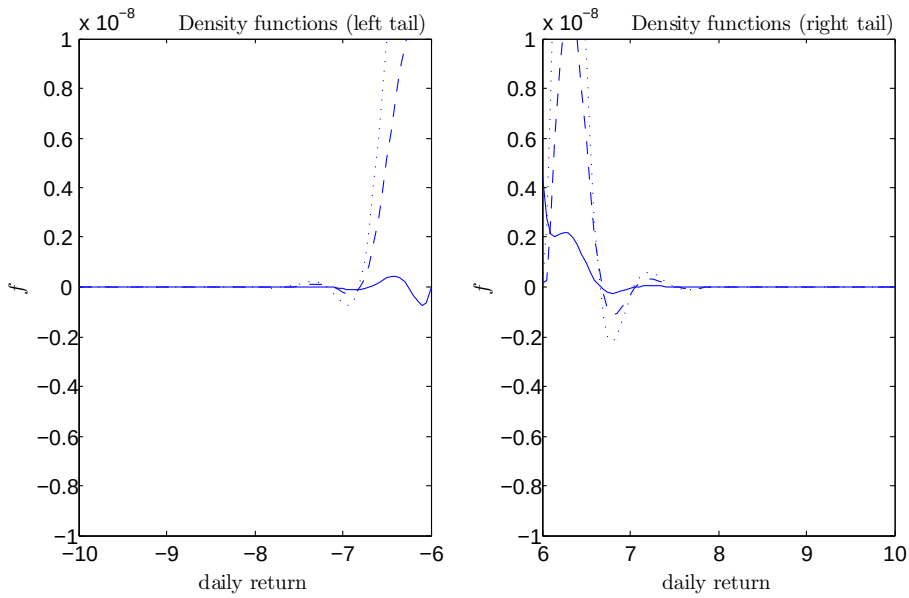

Figure 9: Far tails of the density functions for AS 51 daily returns (standardised). The solid, dashed and dotted curves correspond to unconstrained approximation, approximations with Constraint 1 and 2, respectively. The value for n and p are the same as those in Figure 5.

References

Basel Committee on Banking Supervision. Fundamental review of the trading book : A revised market risk framework. Consultative Document, Bank for International Settlements, Oct. 2013.

A. Eloyan and S. K. Ghosh. Smooth density estimation with moment constraints using mixture distributions. *Journal of Nonparametric Statistics*, 23(2):513–531, 2011.

P. Hall and B. Presnell. Density estimation under constraints. *Journal of Computational and Graphical Statistics*, 8(2):259–277, 1999.

K. Marumo and R. C. Wolff. Expansion methods applied to asset return distributions. *Journal of Risk*, 10(2):3–24, 2008.

K. Marumo and R. C. Wolff. A non-parametric method for approximating. *Faculty of Economics Saitama University Working Paper*, (4), February 2013.

A. J. McNeil, R. Frey, and Embrechts. *Quantitative Risk Management*. Princeton, 2005.

C. Musso and N. Oudjane. l^2-density estimation under constraints. In *European Signal Processing Conference*, pages 2499–2503, 2009.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall London, 1986.