The Use of Stochastic Collocation for Sampling from Expensive Distributions with Applications in Finance

Hilmarie Brand

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

May 24, 2016

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy at the University of the Cape Town. It has not been submitted before for any degree or examination in any other University.

_____

May 24, 2016

Abstract

The pricing of financial derivatives using numerical methods often requires sampling from expensive distributions. These are distributions with inverse cumulative distribution functions that are difficult to evaluate, thus requiring significant computation time. To mitigate this, Grzelak *et al.* (2015) introduced the stochastic collocation Monte Carlo sampler. This sampling method is based on a generalisation of the stochastic collocation method of Mathelin and Hussaini (Mathelin and Hussaini, 2003) which was introduced in the context of solving stochastic partial differential equations (Babuška *et al.*, 2007; Loeven *et al.*, 2007).

The stochastic collocation Monte Carlo sampling method entails sampling from a cheaper distribution and then transforming the samples to obtain realisations from the expensive distribution. The function that transforms the quantiles of the cheap distribution to the corresponding quantiles of the expensive distribution is approximated using an interpolating polynomial of a prespecified degree. The points at which the interpolating polynomial is constructed to exactly match the true quantile-to-quantile transformation function are known as collocation points. Any number of realisations from the expensive distribution may be read off using the interpolating polynomial, leading to a significant reduction in computation time when compared to methods like the inverse transform method.

This dissertation provides an overview of the stochastic collocation method, using distributions and models frequently encountered in finance as examples. Where possible, goodness of fit tests are performed. The major contribution of the dissertation is the investigation of the roots of Chebyshev polynomials of the first kind as collocation points, as opposed to Gaussian quadrature points used by Babuška *et al.* (2007), Loeven *et al.* (2007) and Grzelak *et al.* (2015). The roots of the Chebyshev polynomials are constrained to lie in a specified closed interval and hence are convenient to use when the statistic to be estimated does not depend on the entire distribution of interest, e.g. option prices or conditional expectations like expected shortfall.

Acknowledgements

I would like to extend my gratitude to my supervisors, Thomas McWalter and Jörg Kienitz, for their guidance, time and for sharing their knowledge with me. I would also like to thank my parents, Piet Brand and Estelle Brand, and my friend Violet Williams, for their never-ending support and encouragement.

Soli Deo Gloria.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The pricing of financial derivatives using numerical methods often requires the generation of realisations from distributions which are computationally 'expensive' to sample from. These are distributions with inverse cumulative distribution functions (inverse CDF) that are difficult to evaluate, thus requiring significant computation time. Samples from the non-central chi-squared distribution are, for instance, required for simulation of the Cox-Ingersoll-Ross (CIR) model, which describes the dynamics of the integrated variance process under the Heston model. Sampling from such expensive distributions using traditional methods such as the rejection method or the popular inverse transform method is slow. Over the last decade more efficient sampling algorithms have been proposed. Many of these new algorithms are based on the inverse transform method, using a polynomial approximation of either the probability density function (PDF) or the inverse cumulative distribution function of the distribution of interest.

In a recent paper by Grzelak *et al.* (2015) a sampling method based on a generalisation of the stochastic collocation method of Mathelin and Hussaini (2003) was proposed for sampling from computationally expensive distributions. The paper focusses on distributions and processes often encountered in finance, in particular the non-central chi-squared distribution, CIR process, Heston model and SABR model. The method was initially introduced as an efficient method to propagate uncertainty in numerical simulation in the context of solving stochastic partial differential equations (Tatang *et al.*, 1997; Webster *et al.*, 1996). At that time stochastic collocation was proposed as an alternative to Monte Carlo methods and polynomial chaos expansions approaches to uncertainty quantification. The stochastic collocation method enjoys both the ease of implementation of Monte Carlo methods and the fast convergence rates associated with polynomial chaos expansions.

1.1 Literature review

1.1.1 Polynomial chaos expansion

The polynomial chaos expansion approach has its origin in the work of Wiener (1938) and, like stochastic collocation, was introduced for the purpose of propagating uncertainty in the context of solving partial differential equations with random inputs (i.e. stochastic partial differential equations).

When solving stochastic partial differential equations using polynomial chaos expansions, the random inputs and solution are expressed as truncated expansions in orthogonal polynomials in the underlying random variables driving the uncertainty in the inputs. The optimal type of polynomials for use in the expansion is determined by the distribution of the random input variables. Given a particular probability distribution, the set of orthogonal polynomials that is optimal to use in the chaos expansion is the set of polynomials that are orthogonal with respect to the probability density function of that distribution. Hermite polynomials are, for example, optimal to use when the underlying random variables are normally distributed, while Laguerre polynomials are optimal when the underlying random variables are exponentially distributed. Using the standard Galerkin approach to polynomial chaos expansion (which involves the projection of the governing equation onto the subset of basis functions used in the truncated expansion), solving the stochastic partial differential equation requires the solution of a large coupled system of equations. Although typically this method is much less computationally expensive than the Monte Carlo approach to uncertainty quantification and also has much faster convergence rates, it has the drawback that its implementation becomes very difficult in the case of a non-polynomial non-linear relationship between the solution of the stochastic partial differential equation and the underlying random variable(s) (Mathelin and Hussaini, 2003).

In a recent paper by Olver and Townsend (2013) a computationally efficient algorithm based on polynomial chaos expansion and the inverse transform method was proposed for sampling from a broad class of smooth distributions. The algorithm entails approximating the probability density function by a truncated expansion in Chebyshev polynomials which is then integrated to obtain an approximate cumulative distribution function. Following this, pseudo random samples from the uniform distribution (with domain $[0, 1]$) are used together with this approximate cumulative distribution function to obtain the corresponding samples from the distribution defined by the approximate probability density function by means of the bisection method. They also suggested an extension of the algorithm to bivariate distributions whereby the probability density function is approximated by a

function of lower rank. The proposed algorithm was tested on four univariate and four bivariate distributions. It was found that the proposed algorithm substantially outperformed a MATLAB implementation of slice sampling for the univariate distributions that were considered. For two of the four bivariate distributions the computation time required by the algorithm was substantially less than that required by the slice sampling algorithm. In the paper it was specifically pointed out that the proposed algorithm would be very useful in the context of sampling from expensive distributions since the original distribution can be discarded once the approximation to the probability density function has been determined.

1.1.2 The stochastic collocation method

The stochastic collocation method was introduced in 2003 in a paper by Mathelin and Hussaini (2003) as an efficient method for propagating uncertainty in numerical simulation within the context of solving SPDEs. In contrast to the solution of an SPDE using polynomial chaos expansions, solving an SPDE using stochastic collocation requires the solution of a decoupled system of equations. The resulting (approximate) solution to the SPDE satisfies the corresponding deterministic PDE at the boundary conditions as well as at chosen interpolation points, which are referred to as collocation points in this context.

In 2007 two generalisations of the stochastic collocation method of Mathelin and Hussaini (2003) were proposed (Babuška *et al.*, 2007; Loeven *et al.*, 2007) Unlike the method proposed by Loeven *et al.* (2007), the method proposed by Babuška *et al.* (2007) allows random input variables to be correlated. When the input variables are independent the two methods are identical. A different approach to the application of stochastic collocation in the context of correlated input variables was later proposed by Navarro *et al.* (2015). Both the method proposed by Babuška *et al.* (2007) and that proposed by Loeven *et al.* (2007) uses Gaussian quadrature points as collocation points. Specifically, when the random input variables are independent, the collocation points relating to each random variable are taken to be the roots of polynomials that are orthogonal with respect to the PDF of that random variable. Loeven *et al.* (2007) referred to this generalisation as probabilistic collocation but more often than not the method is referred to as stochastic collocation in literature and will be referred to as such in this dissertation. Henceforth the term 'stochastic collocation' will be used to refer to the generalisation of the method of Mathelin and Hussaini (2003) whereby the roots of polynomials that are orthogonal with respect to the PDF of the random input variable are used as collocation points unless explicitly stated otherwise. The two generalisations proposed by Babuška *et al.* (2007) and Loeven *et al.* (2007) share a lot of similarities with a collocation method pro-

posed by Tatang (1995), which he referred to as probabilistic collocation, to solve black box type models.

Babuška *et al.* (2007) proved that his method attains exponential convergence with respect to the number of collocation points. For the univariate case, when the random input variable is uniformly distributed over $[-1, 1]$, then the methods proposed by Mathelin and Hussaini (2003), Babuška *et al.* (2007) and Loeven *et al.* (2007) are equivalent. Like the method of Mathelin and Hussaini (2003), this method results in a decoupled system of (deterministic) equations; one equation for every collocation point.

The stochastic collocation method was first applied in the field of finance and specifically derivative pricing by Pizzi (2012). In Pizzi (2012) a European call option was priced using Monte Carlo as well as the stochastic collocation technique applied to the Black-Scholes PDE with stochastic volatility. Results showed that much faster convergence rates are achieved with the stochastic collocation method.

More recently the stochastic collocation method was proposed as a more efficient approach to sample from computationally expensive distributions (Grzelak *et al.*, 2015). The authors refer to this sampling method as the stochastic collocation Monte Carlo sampler. The method entails sampling from a 'cheaper' distribution and then transforming the realisations of the cheap random variable to obtain realisations of the expensive random variable. The function that transforms the quantiles of the cheap distribution to the corresponding quantiles of the expensive distribution is approximated by an interpolating polynomial of some prespecified order. In the sampling context, the cheap distribution plays the same role as the input random variables that drive the uncertainty in the context where the method is used to solve SPDEs.

The polynomial used to approximate the quantile-to-quantile transformation function is constructed so that it coincides with the exact evaluation of that function at each of a set of collocation points. Once the approximating polynomial has been determined, it can be used to obtain any number of samples from the expensive distribution. This naturally leads to a significant reduction in computation time compared to a method such as the inverse transform method where the number of expensive evaluations required is the same as the total sample size. Again, Gaussian quadrature points are used as collocation points. Specifically, it is the roots of polynomials that are orthogonal with respect to the probability density of the cheap random variable that are used as collocation points.

The paper by Grzelak *et al.* (2015) demonstrates that the exact simulation of the Heston model, the simulation of the Stochastic Alpha Beta Rho (SABR) model as well as sampling from the squared Bessel process can be performed efficiently

using the stochastic collocation sampling method.

1.1.3 Other interpolating polynomial methods

Derflinger *et al.* (2009) proposed a numerical inversion algorithm, which according to the authors, was the fastest numerical inversion method available at the time. The algorithm entails dividing the domain of the inverse cumulative distribution function into subintervals and then approximating the function by an interpolating polynomial using Chebyshev interpolation points for each subinterval separately. The algorithm requires as inputs the density function, a typical point in the domain, which is not far from the mode, as well as the maximum allowable size of the difference between the true and approximate inverse cumulative distribution function.

The performance of the algorithm was tested on a few standard distributions like the normal, Cauchy, exponential, gamma, beta and t-distributions as well as a non-central chi-squared, hyperbolic, generalized hyperbolic and α-stable distribution. The results indicated that the newly proposed inversion method was much faster than the built in quantile functions in R — for the normal, exponential and Cauchy distributions only approximately three times as fast, but for a beta distribution with scale parameter greater than one between 80 and 120 times as fast. Generating one million samples from the generalised hyperbolic distribution and non-central chi-squared distribution using the proposed algorithm was respectively 1000 and 10000 times faster than the built in functions in R.

1.2 Dissertation outline

In Chapter 2 the stochastic collocation sampling method is described in detail. The chapter focuses solely on the application of the method to sample from univariate distributions, however attention is given to both marginal and conditional distributions. The latter requires a multivariate application of the stochastic collocation sampling method. Information regarding two types of collocation points that will be investigated in this dissertation namely Gaussian quadrature points and roots of Chebyshev polynomials of the first kind is also provided. The chapter is concluded with some guidelines regarding the choice of collocation points as well as some other practical issues. Chapter 3 illustrates the application of the method to sample from computationally expensive distributions such as the non-central chi-squared distribution as well as that of the integrated variance under the Heston model. Particular attention is given to distributions of non-negative random variables with

an atom at zero. In Chapter 3 various statistics of interest are approximated using the samples obtained from the stochastic collocation sampling method. It is shown that accurate estimates are attainable even when the null hypothesis that the method yields true random realisations of the expensive distribution is rejected by goodness-of-fit tests like the Kolmogorov Smirnov and Chi-squared tests. Concluding remarks are made in Chapter 4.

Chapter 2

The stochastic collocation sampling method

The stochastic collocation sampling method presented in Grzelak *et al.* (2015), which the authors refer to as the stochastic collocation Monte Carlo sampler, aims to reduce the computation time required to generate realisations from distributions that are computationally expensive to sample from. This is done by generating realisations of some distribution that is easy to sample from and then transforming those realisations to obtain approximate realisations from the computationally expensive distribution. Specifically, the samples are transformed using an interpolating polynomial that approximates the true quantile-to-quantile transformation function.

Not only are polynomials easy to evaluate but according to the Weierstrass Approximation Theorem there exists, for every continuous function defined on a closed interval, a polynomial that approximates that function as closely as desired over that interval, (Jeffreys and Jeffreys, 1988).

2.1 The method

2.1.1 Sampling from univariate marginal distributions

Consider a continuous random variable X that is distributed according to the cumulative distribution function, F_X. The random variable $F_X(X)$ is known to be uniformly distributed over the interval $(0, 1)$. By the inverse transform method it follows that a continuous random variable Y is identical in distribution to the random variable $F_Y^{-1}(F_X(X))$, i.e.

$$Y \overset{d}{=} g(X) := F_Y^{-1}(F_X(X))$$

where

$$F_Y^{-1}(u) = \inf \{y : F_Y(y) = u\} . \tag{2.1}$$

That is, realisations of Y can be obtained from realisations of any other continuous random variable X by evaluating the quantile-to-quantile transformation function $g = F_Y^{-1} \circ F_X$ for each realisation of X. However, if F_Y^{-1} is difficult to evaluate, generating a large number of realisations of Y in this way will be computationally expensive, even if F_X is easy to sample from. Note that the inverse CDF as defined in (2.1) will be well-defined unless the PDF of Y has an atom i.e. unless there exists a value, y, which is such that $\mathbb{P}[Y = y] > 0$.

The stochastic collocation sampling method involves the approximation of the quantile-to-quantile transformation function g by a global interpolating polynomial. This interpolating polynomial is constructed to coincide with g for each of a prespecified set of N values in the support of f_X, the PDF of X. Specifically, g is approximated by the polynomial with the smallest order that coincides with each of the points in the set,

$$\{(x_i^\star, y_i^\star)\}_{i=1}^N,$$

where $x_i^\star$ denotes the i^{th} collocation point and

$$y_i^\star := g(x_i^\star),$$

for $i = 1, ..., N$. This interpolating polynomial of smallest order is unique (Ryaben'kii and Tsynkov, 2006), has a maximum order of $(N-1)$ and will be denoted by g_N. Given g_N, which represents the approximate relationship between the two random variables X an Y, an approximate realisation of Y, $\hat{y}$, can be read off the interpolating polynomial for any realisation of X, say x, i.e.

$$\hat{y} = g_N(x).$$

The total number of evaluations of the computationally expensive function g will therefore be equal to the number of collocation points, N, irrespective of the number of realisations of Y that is required. Consequently, the computation time required to obtain a large number of realisations of Y using the stochastic collocation sampling method should be significantly less than that required using the inverse transform method.

It should be noted that although the quantile-to-quantile transformation function g is non-decreasing, there is no guarantee that the interpolation polynomial g_N will be non-decreasing. Also, it is only when g_N is non-decreasing that it is guaranteed that,

$$\mathbb{P}[X \leq x_i^\star] = \mathbb{P}[g_N(X) \leq g_N(x_i^\star)] = \mathbb{P}[\hat{Y} \leq y_i^\star].$$

2.1.1.1 The interpolating polynomial

The Lagrangian form of the unique interpolating polynomial of smallest order that coincides with each of the N points in the set, $\{(x_i^\star, y_i^\star)\}_{i=1}^N$, is given by

$$\hat{Y} = g_N(X) = \sum_{i=1}^{N} y_i^\star l_i(X),$$

where l_i denotes the i^{th} Lagrange basis function,

$$l_i(X) = \prod_{j=1, j\neq i}^{N} \frac{X - x_j^\star}{x_i^\star - x_j^\star}.$$

An alternative form of g_N that requires less calculations than the Lagrangian form is the first form of the so-called Barycentric form of the interpolating polynomial. This is given by

$$g_N(X) = \sum_{i=1}^{N} y_i^\star \frac{\lambda_i l(X)}{X - x_i^\star},$$

where

$$\lambda_i = \frac{1}{\prod_{j=1, j\neq i}^{N}(x_i^\star - x_j^\star)} \quad \text{and} \quad l(X) = \prod_{j=1}^{N}(X - x_j^\star).$$

A further simplification yields the second or true form of the Barycentric interpolation formula. The simplification involves dividing g_N by the Barycentric interpolation of the constant function $h(X) = 1$, i.e.

$$h(X) = l(X) \sum_{i=1}^{N} \frac{\lambda_i}{(X - x_i^\star)}.$$

Doing so yields

$$g_N(X) = \frac{\sum_{i=1}^{N} y_i^\star \frac{\lambda_i}{(X - x_i^\star)}}{\sum_{i=1}^{N} \frac{\lambda_i}{(X - x_i^\star)}}.$$

2.1.1.2 Interpolation error

When approximating the function g with the interpolating polynomial, g_N, an interpolation error is incurred for every x in the support of f_X. If g is N times continuously differentiable on (a, b), then for any $x \in (a, b)$ there exists an $\xi \in (a, b)$ which

is such that the interpolation error is given by

$$g(x) - g_N(x) = \frac{g^{(N)}(\xi)}{(N)!} \prod_{i=1}^{N} (x - x_i^\star) \leq \frac{\max|g^{(N)}(\xi)|}{(N)!} (b-a)^N \tag{2.2}$$

(Ryaben'kii and Tsynkov, 2006). It is clear from (2.2) that the magnitude of the interpolation error is determined by the magnitude of

$$\frac{g^{(N)}(\xi)}{(N)!},$$

which depends on the exact shape of g, the number of collocation points and the value of ξ, which is dependent on the set of interpolation points, as well as the magnitude of the product,

$$\prod_{i=1}^{N} (x - x_i^\star). \tag{2.3}$$

When the roots of Chebyshev polynomials of the first kind are used as collocation points (see Section 2.2.2), the magnitude of the product in (2.3) is minimised.

2.1.2 Sampling from univariate conditional distributions

Sampling from a univariate conditional distribution using the stochastic collocation sampling method requires the implementation of a multivariate polynomial interpolation scheme, which in turn requires a grid of collocation points. The dimension of this grid will depend on the number of variables conditioned on. In general, when considering a conditional distribution of a random variable Y where the number of variables conditioned on is M, an $(M+1)$-dimensional grid of collocation points will be required - one dimension for the cheap variable and one for each of the M variables conditioned on. The approximate conditional realisations of Y will be read off the surface that coincides with the quantile-to-quantile transformation function at each of the collocation points in the grid.

The two main types of grid that can be used are tensor product grids and sparse grids. A tensor product grid constructed from $(M+1)$ sets of univariate collocation points consists of every possible $(M+1)$-tuple of these collocation points. If the number of collocation points chosen for the cheap variable is N and the number of collocation points chosen for the j^{th} variable conditioned on is N_j then the total number of $(M+1)$-dimensional collocation points contained in the tensor product grid will be $N \prod_{j=1}^{M} N_j$. The number of evaluations of the expensive quantile-to-quantile transformation function that is required when using this tensor grid of collocation points is therefore also $N \prod_{j=1}^{M} N_j$. Sparse grids contain less points

than tensor grids and will therefore require less computation time. However, since the approximate quantile-to-quantile transformation function will coincide with the true quantile-to-quantile transformation function at fewer points, sparse grids also usually lead to less accurate results compared to those obtained using tensor product grids.

It should be noted that since the number of points on a tensor product grid increases exponentially with the number of variables, using such a grid of collocation points may become computationally expensive when the number of variables on which the distribution of Y is conditioned, is large. For such cases some sparse grid of collocation points, e.g. the grid proposed by Smolyak (Smolyak, 1963), can be used. The Smolyak grid and other sparse grids have been implemented for the stochastic collocation method in the context of solving SPDEs (Xiu and Hesthaven, 2005; Pizzi, 2012; Nobile *et al.*, 2008) but is beyond the scope of this dissertation. In the remainder of this dissertation only tensor product grids of collocation points are considered.

2.1.2.1 Sampling from univariate conditional distributions using tensor product grids

In order to sample from the distribution of Y conditional on the value of one other variable, V, using the stochastic collocation method requires the implementation of a two-dimensional polynomial interpolation. This in turn requires a two-dimensional grid of collocation points. The interpolating surface is constructed to coincide with the quantile-to-quantile transformation function,

$$g(x|V=v) = F^{-1}_{Y|V=v}(F_X(x)),$$

for each of the collocation points in the grid. If the number of collocation points chosen for the cheap variable, X, is N and the number of collocation points chosen for V is N_1, then the two dimensional tensor product grid constructed from these collocation points will contain $(N \times N_1)$ points.

The unique two-dimensional interpolating polynomial of smallest order that coincides with the quantile-to-quantile transformation function for each of the $(N \times N_1)$ points on the tensor grid constructed from the set of N collocation points corresponding to X, $\{x_i^\star\}_{i=1}^{N}$, and the set of N_1 collocation points corresponding to V, $\{v_j^\star\}_{i=1}^{N_1}$, is given by

$$g_N(x|V=v) = \sum_{i=1}^{N}\sum_{j=1}^{N_1} F^{-1}_{Y|V=v}(F_X(x_i^\star))l_i^X(x)l_j^V(v),$$

where

$$l_i^X(x) = \prod_{j=1, j\neq i}^{N} \frac{x - x_j^\star}{x_i^\star - x_j^\star} = \frac{\lambda_i^X l^X(x)}{x - x_i^\star},$$

with

$$\lambda_i^X = \frac{1}{\prod_{j=1, j\neq i}^{N}(x_i^\star - x_j^\star)} \quad \text{and} \quad l^X(X) = \prod_{i=1}^{N}(x - x_i^\star)$$

and similarly

$$l_j^V(v) = \prod_{i=1, i\neq j}^{N_1} \frac{v - v_i^\star}{v_j^\star - v_i^\star} = \frac{\lambda_j^V l^V(v)}{v - v_j^\star},$$

with

$$\lambda_j^V = \frac{1}{\prod_{i=1, i\neq j}^{N_1}(v_j^\star - v_i^\star)} \quad \text{and} \quad l^V(v) = \prod_{j=1}^{N_1}\left(v - v_j^\star\right).$$

Similarly, in order to sample from the distribution of Y conditional on the values of two variables, V and W, a three-dimensional interpolating polynomial is required. The unique three-dimensional interpolating polynomial of least order that passes through each of the $(N \times N_1 \times N_2)$ points,

$$\left\{\left(x_i^\star, v_j^\star, w_k^\star, g\left(x_i^\star | V = v_j^\star, W = w_k^\star\right)\right)\right\}_{i,j,k},$$

is given by

$$\begin{aligned} y \approx g_N(x|V = v, W = w) &= \sum_{i=1}^{N}\sum_{j=1}^{N_1}\sum_{k=1}^{N_2} F^{-1}_{Y|V=v,W=w}(F_X(x_i^\star)) l_i^X(x) l_j^V(v) l_k^W(w) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N_1}\sum_{k=1}^{N_2} F^{-1}_{Y|V=v,W=w}(F_X(x_i^\star)) \frac{\lambda_i^X \lambda_j^V \lambda_k^W l^X(x) l^V(v) l^W(w)}{(x - x_i^\star)(v - v_j^\star)(w - w_k^\star)}. \end{aligned}$$

2.2 Collocation points

Two types of collocation points will be considered in this work. The first is Gaussian quadrature points and the second is the roots of Chebyshev polynomials of the

first kind, also referred to as Chebyshev nodes or Chebyshev-Gauss points. These two types of nodes have different properties which make them suited for different applications. The definitions of these nodes and their properties will be discussed in the sections to follow.

2.2.1 Gaussian quadrature collocation points

For the integral

$$\int_a^b p(x)w(x)dx,$$

where p is a polynomial of order less or equal to $(2N-1)$ and w is a non-negative weighting function with support $[a,b]$ with $a,b \in [-\infty,\infty]$, the set of N Gaussian-quadrature points are those points, $\{x_i\}_{i=1}^N$, that are such that

$$\int_a^b p(x)w(x)dx = \sum_{i=1}^N p(x_i)w_i, \tag{2.4}$$

where $\{w_i\}$ are non-negative weights. The definition of the weights is provided in Section 2.2.1.3. The points, $\{x_i\}_{i=1}^N$, are known to be the N roots of the N^{th} order polynomial, p_N, from the family of polynomials that is orthogonal with respect to $w(x)$, that is, the polynomials $\{p_i\}$ that satisfy

$$\mathbb{E}[p_i(X)p_j(X)] = \int_a^b w(x)p_i(x)p_j(x)dx = \delta_{ij}\mathbb{E}[p_i^2(X)] \quad i,j = 0,...,N, \tag{2.5}$$

p_i is a polynomial of order i, δ_{ij} denotes the Kronecker delta and $f_X(x)$ is the PDF of X. Note that by the linearity of the integral operator, the integral of any linear function of p, say $h(p(\cdot))$ can also be evaluated deterministically by

$$\int_a^b h(p(x))w(x)dx = \sum_{i=1}^N h(p(x_i^\star))w_i.$$

The set of N Gaussian quadrature collocation points is the set of N roots of the N^{th} order polynomial p_N. If the quantile-to-quantile transformation function, g, is a polynomial of order $(2N-1)$ or less, the expected value of $g_N(X)$, which can be evaluated exactly as

$$\mathbb{E}[g_N(X)] = \sum_{i=1}^N g(x_i^\star)w_i,$$

will be equal to $\mathbb{E}[g(X)]$. If g is a polynomial of order $(N-1)$ or less, g_N will be identical to g. It follows that for a polynomial quantile-to-quantile transformation function, g_N will converge to g pointwise and $\mathbb{E}[g_N(X)]$ will converge to $\mathbb{E}[g(X)]$

as N tends to infinity. For an arbitrary continuous non-polynomial function, g, very specific conditions regarding g, the weighting function as well as the type of collocation points used need to be satisfied in order to guarantee convergence of g_N to g as well as convergence of $\mathbb{E}[g_N(X)]$ to $\mathbb{E}[g(X)]$. Some convergence results related to Gaussian quadrature collocation points are discussed in Sections 2.2.1.1 and 2.2.1.2 respectively. These two sections consider the case of bounded and unbounded weighting functions respectively.

2.2.1.1 Convergence results related to Gaussian quadrature points for weighting functions of bounded support

Erdős and Turán (1937) proved that given a continuous function g and a positive weighting function, w, with support $[-1, 1]$ that is such that

$$\int_{-1}^{1} w(x)dx < \infty,$$

the Lagrangian interpolating polynomial, g_N that matches g exactly at the N roots of the N^{th} order polynomial belonging to the family of polynomials that are orthogonal with respect to w, satisfies

$$\int_{-1}^{1} |g_N(x) - g(x)|^2 \, w(x)dx \to 0. \tag{2.6}$$

Since for every $x \in [-1, 1]$, $w(x) \geq M > 0$, where M is a positive constant, it then follows that

$$0 \leq \int_{-1}^{1} |g_N(x) - g(x)|^2 \, dx \leq \frac{1}{M} \int_{-1}^{1} |g_N(x) - g(x)|^2 \, w(x)dx,$$

which by the squeeze theorem implies that

$$\int_{-1}^{1} |g_N(x) - g(x)|^2 \, dx \to 0.$$

Erdős and Turán (1937) further proved that

$$\int_{-1}^{1} |g_N(x) - g(x)| \, dx \to 0$$

for all bounded and Riemann integrable functions g.

Note that if (2.6) is satisfied, then so is

$$\left(\int_{-1}^{1} |g_N(x) - g(x)|^2 \, w(x)dx \right)^{0.5} \to 0, \tag{2.7}$$

i.e. $g_N(x)$ converges to $g(x)$ with respect to the L_2 norm. This in turn implies that $g_N(x)$ converges to $g(x)$ in the L_1 norm, that is,

$$\left(\int_{-1}^{1} |g_N(x) - g(x)|\, w(x)dx\right) \to 0.$$

Now, since

$$\left(\int_{-1}^{1} |g_N(x) - g(x)|\, w(x)dx\right) \geq \left|\left(\int_{-1}^{1} (g_N(x) - g(x))\, w(x)dx\right)\right| \geq 0$$

it follows from the squeeze theorem that

$$\left|\int_{-1}^{1} (g_N(x) - g(x))\, w(x)dx\right| \to 0.$$

Application of the reverse triangle inequality gives

$$\left|\left|\int_{-1}^{1} g_N(x)w(x)dx\right| - \left|\int_{-1}^{1} g(x)w(x)dx\right|\right| \leq \left|\int_{-1}^{1} (g_N(x) - g(x))\, w(x)dx\right|.$$

Now, since the left hand side of the above inequality is greater or equal to zero, it follows from the squeeze theorem that

$$\left|\left|\int_{-1}^{1} g_N(x)w(x)dx\right| - \left|\int_{-1}^{1} g(x)w(x)dx\right|\right| \to 0.$$

If $w(x) = f_X(x)$, where $f_X(x)$ is the PDF of the random variable X, then this means that

$$|\mathbb{E}[g_N(X)]| \to |\mathbb{E}[g(X)]|.$$

For positive weighting functions, w with support $[a, b]$,

$$\int_a^b |g_N(x) - g(x)|^2\, w(x)dx \to 0 \tag{2.8}$$

$$\Rightarrow \quad \left(\int_a^b |g_N(x) - g(x)|^2\, w(x)dx\right)^{0.5} \to 0 \tag{2.9}$$

$$\Rightarrow \quad \int_a^b |g_N(x) - g(x)|\, w(x)dx \to 0 \tag{2.10}$$

$$\Rightarrow \quad \left|\int_a^b (g_N(x) - g(x))\, w(x)dx\right| \to 0. \tag{2.11}$$

As before, if $w(x) = f_X(x)$, then this means that

$$\therefore \quad |\mathbb{E}[g_N(X)]| \to |\mathbb{E}[g(X)]|.$$

If $\mathbb{E}[g_N(X)] > 0 \quad \forall N$ and $\mathbb{E}[g\,(X)] > 0$ or if $g_N(x) > g(x)$ or $g_N(x) < g(x)$ for all $x \in [a, b]$ then $\mathbb{E}[g_N(X)] \to \mathbb{E}[g\,(X)]$.

It follows from the above results that if the cheap distribution is taken to be a uniform distribution, then g_N constructed on the Gaussian quadrature points corresponding to that uniform distribution will converge to $g(x)$ in L_2 and L_1. The Gaussian quadrature points associated with the uniform distribution are however not located as far into the two tails of the distribution as say the Gaussian quadrature points corresponding to the standard normal distribution (see Section 2.2.1.8). In Section 2.5 a motivation will be provided for using a set of collocation points that are stretched far into the tails of the cheap distribution when the statistic of interest depends on the entire distribution or when it is not possible to determine the exact part of the cheap distribution that is of relevance for the statistic to be estimated.

It is important to note that (2.8)-(2.11) does not imply that

$$\int_c^d |g_N(x) - g(x)|^2 \, w(x)dx \to 0,$$
$$\left(\int_c^d |g_N(x) - g(x)|^2 \, w(x)dx\right)^{0.5} \to 0,$$
$$\int_c^d |g_N(x) - g(x)| \, w(x)dx \to 0,$$

and

$$\int_c^d |g_N(x) - g(x)| \, w(x)dx \to 0$$

respectively, for any $[c, d] \subset [a, b]$. This implies that when a statistic that depends on only a subset of the expensive distribution is estimated using realisations generated using the stochastic collocation sampling method with Gaussian quadrature points as collocation points, the estimate of that statistic may not converge to the true value as N tends to infinity. This means that statistics that depend on only a portion of the expensive distribution may be poorly approximated even for large N, especially when that part of the distribution contains very few collocation points.

If $g(x) > 0$ for all x in the support of f_X and

$$g_N^*(x) := \begin{cases} g_N(x) & \text{if } g_N(x) \geq 0 \\ 0 & \text{if } g_N(x) < 0 \end{cases}$$

then

$$|g_N^*(x) - g(x)| \begin{cases} = |g_N(x) - g(x)| & \text{if } g_N(x) \geq 0 \\ < |g_N(x) - g(x)| & \text{if } g_N(x) < 0. \end{cases}$$

From this it follows that, for every N,

$$0 \leq \sup\left\{|g_N^*(x) - g(x)|\right\} \leq \sup\left\{|g_N(x) - g(x)|\right\},$$
$$0 \leq \int_a^b |g_N^*(x) - g(x)|\, dx \leq \int_a^b |g_N(x) - g(x)|\, dx,$$
$$0 \leq \int_a^b |g_N^*(x) - g(x)|\, f_X(x)dx \leq \int_a^b |g_N(x) - g(x)|\, f_X(x)dx$$

and

$$0 \leq \left(\int_a^b |g_N^*(x) - g(x)|^2 f_X(x)dx\right)^{0.5} \leq \left(\int_a^b |g_N(x) - g(x)|^2 f_X(x)dx\right)^{0.5}. \tag{2.12}$$

Hence, if g_N converges to $g(x)$ in the L_2 norm, then so will $g_N^*(x)$. In addition to this, (2.12) implies that the rate at which g_N^* will converge to g in L_2 will be at least as fast as the rate at which g_N converges to g in L_2.

Note that for a given N, it is not necessarily the case that

$$|\mathbb{E}[g_N^*(X)] - \mathbb{E}[g\,(X)]| = \left|\int_a^b g_N^*(x) f_X(x)dx - \int_a^b g(x) f_X(x)dx\right|$$

will be smaller or equal to

$$|\mathbb{E}[g_N(X)] - \mathbb{E}[g(X)]| \leq \left|\int_a^b g_N(x) f_X(x)dx - \int_a^b g(x) f_X(x)dx\right|.$$

The same holds for the absolute error incurred by the estimate of any integral statistic $\int_c^d k(g(x)) f_X(x)dx$ where k is an arbitrary real function and $[c, d] \subset [a, b]$.

2.2.1.2 Convergence results related to Gaussian quadrature points for weighting functions of unbounded support

The properties of the Lagrangian interpolating polynomial constructed using the roots of polynomials that are orthogonal with respect to a weighting function with unbounded support as interpolation points were considered in Nevai (1980). One of the results proved in Nevai (1980) that is of particular importance here, is that if $g(x)$ is continuous and $|g(x)| \exp\left(-x^2/2\right)$ converges to zero at a rate faster than $1/|x|$, i.e. if

$$|g(x)| \exp(-x^2/2) = \mathrm{o}\left(\frac{1}{|x|}\right)$$

then

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} \left[|L_N(x) - g(x)| \exp(-0.5x^2)\right]^p dx = 0$$

for $p > 1$, where L_N is the interpolating polynomial constructed on the N roots of the N^{th} order polynomial that is orthogonal with respect to the weighting function

$$w(x) = \exp(-0.5x^2).$$

Note that since the density of the standard normal distribution, $\frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$, is a scalar multiple of $\exp(-x^2/2)$, it can be substituted for $\exp(-x^2/2)$ in the above result. It follows that when the quantile-to-quantile transformation function g is continuous, Nevai's result holds for the interpolating polynomial, g_N, constructed from the N roots of the N^{th} order polynomial p_N that satisfies (2.5) for $w(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$.

As before, the above result does not imply that the integral over a subset of the support of the weighting function will converge to zero, i.e. that

$$\lim_{N\to\infty} \int_c^d \left[|g_N(x) - g(x)| \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2) \right]^p dx = 0$$

for some $[c, d] \subset (-\infty, \infty)$. As before, this implies that statistics that depend on only a portion of the distribution of the expensive random variable may be poorly approximated even for large N. This will be illustrated with an example in Section 3.2.

For g_N^* defined earlier it is true that

$$0 \leq \int_{-\infty}^{\infty} \left[|g_N^*(x) - g(x)| \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2) \right]^p dx \leq \int_{-\infty}^{\infty} \left[|g_N(x) - g(x)| \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2) \right]^p dx$$

for any N and for all $x \in (-\infty, \infty)$ and so by the squeeze theorem it follows that

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} \left[|g_N^*(x) - g(x)| \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2) \right]^p dx = 0$$

for $p > 1$. The rate of convergence of the integral is also at least as fast as that of the integral containing g_N.

Note that as before, for a given N, it is not necessarily the case that

$$|\mathbb{E}[g_N^*(X)] - \mathbb{E}[g(X)]| \leq |\mathbb{E}[g_N(X)] - \mathbb{E}[g(X)]|,$$

nor is it necessarily the case that

$$\left| \int_c^d k(g_N^*(x)) f_X(x) dx - \int_c^d k(g(x)) f_X(x) dx \right| \leq \left| \int_c^d k(g_N(x)) f_X(x) dx - \int_c^d k(g(x)) f_X(x) dx \right|$$

for an arbitrary real function k and $[c, d] \subset [a, b]$.

Given time and space limitations the topic of the rate of convergence will not be considered in this dissertation. Further research on this topic is needed and would add value to the work contained in this dissertation.

2.2.1.3 Calculation of the Gaussian quadrature points

Theorem 2.1. *For any density function f_X, a unique sequence of monic orthogonal polynomials, $\{p_n\}$, where the order of p_n is exactly n, exists and can be constructed using the three-term recurrence relation (Favard, 1935),*

$$p_{j+1}(X) = (X - \alpha_{j+1})\, p_j(X) - \beta_j p_{j-1}(X) \text{ for } j = 0, ..., N-1$$

where

$$p_{-1}(X) \equiv 0, \quad p_0(X) \equiv 1,$$
$$\alpha_{j+1} = \frac{\mathbb{E}[X p_j^2(X)]}{\mathbb{E}[p_j^2(X)]}$$

and

$$\beta_j = \frac{\mathbb{E}[p_j^2(X)]}{\mathbb{E}[p_{j-1}^2(X)]} \quad \text{with} \quad \beta_0 = 0.$$

For the corresponding set of orthonormal polynomials, the recurrence relation becomes

$$\sqrt{\beta_{j+1}} p_{j+1}(X) = (X - \alpha_{j+1})\, p_j(X) - \sqrt{\beta_j} p_{j-1}(X)$$
$$\Rightarrow \quad X p_j(X) = \sqrt{\beta_j} p_{j-1}(X) + \alpha_{j+1} p_j(X) + \sqrt{\beta_{j+1}} p_{j+1}(X),$$

for $j = 0, ..., N-1$. In matrix notation this relation is given by

$$X \mathbf{p}(X) = \hat{J} \mathbf{p}(X) + \sqrt{\beta_N} p_N(X)\, \mathbf{e}_N,$$

where $\mathbf{e}_N$ denotes a unit vector of length N with N^{th} element equal to 1 and all other elements equal to 0 and

$$\hat{J} = \begin{bmatrix} \alpha_1 & \sqrt{\beta_1} & 0 & 0 & 0 \\ \sqrt{\beta_1} & \alpha_2 & \sqrt{\beta_2} & 0 & 0 \\ 0 & \sqrt{\beta_2} & \alpha_3 & \sqrt{\beta_3} & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \sqrt{\beta_{N-2}} & \alpha_{N-1} & \sqrt{\beta_{N-1}} \\ 0 & 0 & 0 & \sqrt{\beta_{N-1}} & \alpha_N \end{bmatrix}.$$

The eigenvalues of the symmetric tridiagonal matrix $\hat{J}$ are the roots of the orthogonal polynomial $p_N(X)$ (Golub and Welsch, 1969).

The i^{th}th weight, w_i, in the Gaussian quadrature formula given in (2.4) is given by

$$w_i = (v_{i1})^2,$$

where v_{i1} is the first component of the orthonormal eigenvector of $\hat{J}$ that correspond to the i^{th} largest eigenvalue of $\hat{J}$.

The recursion coefficients, $\{\alpha_j\}_{j=1}^{N}$ and $\{\beta_j\}_{j=1}^{N-1}$ can also be determined from the Cholesky decomposition of the Gram matrix,

$$M\left[i,j\right] = \mathbb{E}[X^{i+j-2}].$$

Specifically,

$$\alpha_j = \frac{r_{j,j+1}}{r_{j,j}} - \frac{r_{j-1,j}}{r_{j-1,j-1}}, \quad j = 1, ..., N$$
$$\beta_j = \left(\frac{r_{j+1,j+1}}{r_{j,j}}\right)^2, \quad j = 1, ..., N-1,$$

where $r_{i,j}$ denotes the element in the i^{th} row and j^{th} column of the upper triangular matrix R, where

$$M = R'R$$

(Golub and Welsch, 1969). The following subsections specify how to obtain the Gram matrix for the standard normal, gamma, non-central chi-squared and log-normal random variables.

2.2.1.4 Gram matrix for the standard normal random variable

The k^{th} raw moment of the standard normal variable is given by,

$$\mathbb{E}[X^k] = \begin{cases} 0 & \text{if } k \text{ is odd} \\ k!! & \text{if } k \text{ is even} . \end{cases}$$

2.2.1.5 Gram matrix for a gamma random variable

A gamma random variable with a shape parameter α and scale parameter β has k^{th} raw moment,

$$\mathbb{E}[X^k] = \frac{\Gamma\left(k+\alpha\right)\beta^k}{\Gamma\left(\alpha\right)}.$$

2.2.1.6 Gram matrix for the non-central chi-squared random variable

The raw moments of a non-central chi-squared random variable with d degrees of freedom and a non-centrality parameter λ are given by the recursive equation,

$$\mathbb{E}[X^k] = 2^{k-1}\left(k-1\right)!\left(d+k\lambda\right) + \sum_{i=1}^{k-1}\frac{(k-1)!2^{i-1}}{(k-i)!}\left(d+i\lambda\right)\mathbb{E}[X^{k-1}].$$

The collocation points determined by the moments of the non-central chi-squared distribution will be used in Section 3.5 in the simulation of the integrated variance process under the Heston stochastic volatility model.

2.2.1.7 Gram matrix for a lognormal random variable

The k^{th} raw moment of a lognormal random variable, X, with parameters μ and σ such that $\log(X) \sim N(\mu, \sigma)$ [1], is given by

$$\begin{aligned}\mathbb{E}[X^k] &= e^{\mu k + 0.5k^2\sigma^2} \\ &= m_{\log X}(k),\end{aligned}$$

where $m_{\log X}$ denotes the moment generating function of $\log(X)$. Collocation points determined by these raw moments are required, for example, for the simulation of the SABR model (Grzelak *et al.*, 2015).

2.2.1.8 Gaussian quadrature points for various cheap distributions

The greater the degree of non-linearity in the relationship between X and Y, the more collocation points are typically required to attain a certain goodness of fit. The number of collocation points required can become particularly large when the non-linearity cannot be described by a polynomial. It therefore seems natural to expect that using a cheap distribution that is more similar in shape to that of the expensive distribution should yield more accurate results.

In Grzelak *et al.* (2015) only the standard normal distribution was investigated as cheap distribution. For cases where the expensive distribution differs a lot from the standard normal distribution in terms of for example skewness or kurtosis, Grzelak *et al.* (2015) proposed using moment matching to determine a better cheap distribution. In Section 3.1 distributions other than the standard normal will used as cheap distributions for sampling from a gamma distribution. Surprisingly the results indicate that using a cheap distribution that is more similar in shape to that of the expensive distribution does not necessary result in more accurate results. The rest of this section contains some general remarks about Gaussian quadrature points.

The relative positions of Gaussian quadrature points are unaffected by parameters that do not affect the shape of the distribution, e.g. parameters that do not affect the skewness and/or kurtosis of the distribution. For example, $\{F_X(x_i^\star)\}_i$ is unaffected by the scale parameter when X is gamma distributed and by μ when X is lognormally distributed with $\mathbb{E}[\log(X)] = \mu$. Similarly, the relative positions of

[1] In this dissertation $\log(x)$ will refer to the natural logarithm of x.

$U\,(0,1)$	$x^\star$	0.05	0.23	0.50	0.77	0.95
$U\,(0,1)$	$F_X\,(x^\star)$	0.05	0.23	0.50	0.77	0.95
$N\,(0,1)$	$x^\star$	-2.86	-1.36	-0.00	1.36	2.86
$N\,(0,1)$	$F_X\,(x^\star)$	0.00	0.09	0.50	0.91	1.00
$\text{logN}\,(0,0.1)$	$x^\star$	0.79	0.91	1.05	1.20	1.39
$\text{logN}\,(0,0.1)$	$F_X\,(x^\star)$	0.01	0.18	0.67	0.96	1.00
$\text{logN}\,(0,0.5)$	$x^\star$	0.62	1.44	3.08	6.59	15.27
$\text{logN}\,(0,0.5)$	$F_X\,(x^\star)$	0.17	0.77	0.99	1.00	1.00
$\text{logN}\,(0,1)$	$x^\star$	1.18	11.05	90.02	733.51	6886.91
$\text{logN}\,(0,1)$	$F_X\,(x^\star)$	0.56	0.99	1.00	1.00	1.00
$\Gamma\,(5,2)$	$x^\star$	3.97	8.68	15.26	24.38	37.70
$\Gamma\,(5,2)$	$F_X\,(x^\star)$	0.05	0.44	0.88	0.99	1.00
$\Gamma\,(2,2)$	$x^\star$	1.23	4.23	9.22	16.80	28.52
$\Gamma\,(2,2)$	$F_X\,(x^\star)$	0.13	0.62	0.94	1.00	1.00
$\Gamma\,(0.5,2)$	$x^\star$	0.24	2.15	6.17	12.83	23.61
$\Gamma\,(0.5,2)$	$F_X\,(x^\star)$	0.37	0.86	0.99	1.00	1.00

Tab. 2.1: *The set of five Gaussian quadrature points for various different distributions*

the Gaussian quadrature points are unaffected by both μ and σ when X is normally distributed and unaffected by both a and b when $X \sim U(a,b)$.

The Gaussian quadrature points corresponding to distributions that are skewed to the right, i.e. positively skewed, do not stretch far into the lower tails of those distributions. The greater the extent to which a distribution is skewed to the right, the greater the extent to which the Gaussian quadrature points are clustered in the upper 50%, and in particular the extreme right tail, of the distribution. This is evident upon comparison of the relative positions of the Gaussian quadrature points corresponding to the standard normal distribution, which has zero skewness, to those corresponding to distributions like the gamma and lognormal distributions which are skewed to the right. It is also evident upon comparison of the relative positions of the Gaussian quadrature points for different gamma and lognormal distributions. Upon inspection of the five Gaussian quadrature points corresponding to the three different gamma distributions considered in Table 2.1, it is clear that given a fixed scale parameter, the extent to which the Gaussian quadrature points are clustered in the right tail of the distribution increases as the shape parameter decreases and hence the skewness of the distribution increases. Similarly, for a fixed value of μ (which is zero in Table 2.1), the greater the value of σ and hence the

$\mathbf{N=2}$	$x^\star$	-1.00	1.00	–	–	–	–	–	–
$\mathbf{N=2}$	$F_X(x^\star)$	0.16	0.84	–	–	–	–	–	–
$\mathbf{N=3}$	$x^\star$	-1.73	-0.00	1.73	–	–	–	–	–
$\mathbf{N=3}$	$F_X(x^\star)$	0.04	0.50	0.96	–	–	–	–	–
$\mathbf{N=4}$	$x^\star$	-2.33	-0.74	0.74	2.33	–	–	–	–
$\mathbf{N=4}$	$F_X(x^\star)$	0.01	0.23	0.77	0.99	–	–	–	–
$\mathbf{N=5}$	$x^\star$	-2.86	-1.36	-0.00	1.36	2.86	–	–	–
$\mathbf{N=5}$	$F_X(x^\star)$	0.00	0.09	0.50	0.91	1.00	–	–	–
$\mathbf{N=6}$	$x^\star$	-3.32	-1.89	-0.62	0.62	1.89	3.32	–	–
$\mathbf{N=6}$	$F_X(x^\star)$	0.00	0.03	0.27	0.73	0.97	1.00	–	–
$\mathbf{N=7}$	$x^\star$	-3.75	-2.37	-1.15	0.00	1.15	2.37	3.75	–
$\mathbf{N=7}$	$F_X(x^\star)$	0.00	0.01	0.12	0.50	0.88	0.99	1.00	–
$\mathbf{N=8}$	$x^\star$	-4.14	-2.80	-1.64	-0.54	0.54	1.64	2.80	4.14
$\mathbf{N=8}$	$F_X(x^\star)$	0.00	0.00	0.05	0.29	0.71	0.95	1.00	1.00

Tab. 2.2: *Gaussian quadrature points for the standard normal distribution*

greater the skewness of the lognormal distribution, the more the Gaussian quadrature points are clustered in the right tail of the distribution. Upon comparison the relative positions of the Gaussian quadrature points corresponding to the standard normal and uniform distributions, it is clear that the Gaussian quadrature points of the uniform distribution, which has zero skewness and negative excess kurtosis, are pulled more towards the center of the distribution compared to those of the standard normal distribution.

Table 2.2 contains the Gaussian quadrature points corresponding to the standard normal distribution for $N = 2, ..., 8$. It is evident that for $N \geq 5$, the Gaussian quadrature points are distributed over the entire distribution, stretching far into both the left and right tails. It will be explained in Section 2.5 that having a set of collocation points that stretch far into both the left and right tails of a distribution is desirable when the statistic to be estimated depends on the entire distribution of $g(X)$ or when it is not possible to determine the subset of the support of f_X that the statistic depends on.

2.2.2 Zeros of Chebyshev polynomials as collocation points

Recall from Section 2.1.1.2 that the interpolation error incurred by approximating a function g that is continuously differentiable on (a, b) with a Lagrangian polyno-

mial based on N interpolation points, can be expressed as

$$g\left(x\right)-g_N(x)=\frac{g^{(N)}\left(\xi\right)}{(N)!}\prod_{i=1}^{N}\left(x-x_i^{\star}\right),$$

where $\xi\in(a,b)$. The maximum absolute value of the product,

$$\prod_{i=1}^{N}\left(x-x_i^{\star}\right), \tag{2.13}$$

over all $x\in[a,b]$ is minimised by the taking the collocation points, $\{x_i^{\star}\}_{i=1}^{N}$, to be the roots of the (translated) N^{th} order Chebyshev polynomial of the first kind that has all N of its roots in $[a,b]$. As the magnitude of both $\frac{g^{(N)}(\xi)}{(N)!}$ and the product in (2.13) depend on the set of interpolation points, minimisation of the product does not necessarily imply minimisation of the magnitude of the interpolation error. What is known is that the upper bound of of the interpolation error,

$$g(x)-g_N(x),$$

over the interval $[a,b]$, is reduced from

$$\frac{\max_{a\leq\xi\leq b}\left|g^{(N)}\left(\xi\right)\right|}{(N)!}\left(b-a\right)^N$$

to

$$\frac{\max_{a\leq\xi\leq b}\left|g^{(N)}(\xi)\right|}{(N)!}\left(\frac{b-a}{2}\right)^N\frac{1}{2^{N-1}}.$$

when using the Chebyshev nodes in $[a,b]$ are used as collocation points. Using the Chebyshev nodes as collocation points also minimises the effect of the Runge phenomenon, which refers to the tendency of interpolating polynomials constructed with equidistant interpolation points to oscillate near the edges of the interval over which the interpolation is done (Runge, 1901).

The N^{th} order Chebyshev polynomial of the first kind has N roots in the interval $[-1,1]$. These are

$$\cos\left(\frac{(2k-1)\pi}{2N}\right),\quad k=1,2,...,N. \tag{2.14}$$

The roots of the N^{th} order translated Chebyshev polynomial of the first kind for which all N roots are contained in the interval $[a,b]$ can be obtained by a simple linear transformation of (2.14). Specifically, the N roots are given by,

$$0.5(a+b)+0.5(b-a)\cos\left(\frac{(2k-1)\pi}{2N}\right),\quad k=1,2,...,N. \tag{2.15}$$

2.2.2.1 Convergence results related to Chebyshev nodes

When Chebyshev nodes are used to construct the interpolating polynomial g_N, then, when the $(r-1)^{\text{th}}$ order derivative of g, $g^{(r-1)}$, is Lipschitz-continuous on the interval $[a,b]$, g_N converges to g uniformly on $[a,b]$ at the rate

$$\mathrm{O}\left(\frac{1}{(N-1)^r}\log(N)\right),$$

as N tends to infinity (Ryaben'kii and Tsynkov, 2006). The interpolating polynomial constructed on the Chebyshev nodes attains an error that is not much greater than the unavoidable error that is incurred when approximating an r times differentiable function using N interpolation points on a uniform grid, which is $\mathrm{O}\left(\frac{1}{N^r}\right)$ (Ryaben'kii and Tsynkov, 2006).

When $g_N(x)$ converges to $g(x)$ uniformly, $g_N(x)$ must be bounded and hence, by the dominated convergence theorem the previous result implies that

$$\int_a^b g_N(x) f_X(x) dx \to \int_a^b g(x) f_X(x) dx.$$

For any uniformly continuous function, h, it then follows that

$$\int_a^b h\left(g_N(x)\right) f_X(x) dx \to \int_a^b h\left(g(x)\right) f_X(x) dx.$$

Note that all continuous linear functions are uniformly continuous.

When both F_Y^{-1} and F_X are Lipschitz continuous, the quantile-to-quantile transformation function, $g = F_Y^{-1} \circ F_X$, will be Lipschitz continuous, as the composite of two Lipschitz continuous functions, is Lipschitz continuous. Note that the PDF of the standard normal distribution and every derivative thereof are (globally) Lipschitz continuous.

It is important to note that since there is no information regarding whether or not g_N constructed from Chebyshev nodes in $[a,b]$ converges to g over the entire support of f_X, nothing can be stated about whether or not the expected value $\mathbb{E}[g_N(X)]$ converges and if it does, to what value. What is known is that interpolating polynomials tend to display wild behaviour in the region to the left of the smallest interpolation point and the region to the right of the largest interpolation point. Hence it is likely that g_N will poorly approximate g in these regions. Depending on the probability weight associated with these subsets of the support of f_X, poor approximation of g in those regions could have a great adverse effect on the accuracy of approximations to statistics that depend on those parts of the distribution of X.

Note that if $g(x) > 0$ for all x in the support of f_X, then if g_N converges to g uniformly on $[a, b]$, so will g_N^* defined in Section 2.2.1.1. Also, the rate at which g_N^* converges to g will be at least as fast as the rate at which g_N converges to g since

$$|g_N^*(x) - g(x)| \leq |g_N(x) - g(x)| \tag{2.16}$$

for all $x \in [a, b]$. By the dominated convergence, the uniform convergence of g_N^* to g implies that

$$\int_a^b g_N^*(x) f_X(x) dx \to \int_a^b g(x) f_X(x) dx.$$

and that

$$\int_a^b h\left(g_N^*(x)\right) f_X(x) dx \to \int_a^b h\left(g(x)\right) f_X(x) dx.$$

for any uniformly continuous function h.

The inequality in (2.16) implies that for every N,

$$\begin{aligned}
\sup\left\{|g_N^*(x) - g(x)|\right\} &\leq \sup\left\{|g_N(x) - g(x)|\right\}, \\
\int_a^b |g_N^*(x) - g(x)|\, dx &\leq \int_a^b |g_N(x) - g(x)|\, dx
\end{aligned}$$

and

$$\int_a^b |g_N^*(x) - g(x)|\, f_X(x) dx \leq \int_a^b |g_N(x) - g(x)|\, f_X(x) dx.$$

It is, however, not necessarily the case that

$$\int_c^d |g_N^*(x) - g(x)|\, dx \leq \int_c^d |g_N(x) - g(x)|\, dx$$

and

$$\int_c^d |g_N^*(x) - g(x)|\, f_X(x) dx \leq \int_c^d |g_N(x) - g(x)|\, f_X(x) dx,$$

for $[c, d] \subset [a, b]$. It is also not necessarily the case that, for a given N

$$\left| \int_a^b g_N^*(x) f_X(x) dx - \int_a^b g(x) f_X(x) dx \right| \leq \left| \int_a^b g_N(x) f_X(x) dx - \int_a^b g(x) f_X(x) dx \right|.$$

Similarly it does not imply that

$$\left| \int_c^d k(g_N^*(x)) f_X(x) dx - \int_c^d k(g(x)) f_X(x) dx \right| \leq \left| \int_c^d k(g_N(x)) f_X(x) dx - \int_c^d k(g(x)) f_X(x) dx \right|,$$

for $[c, d] \subset [a, b]$.

Note that since for every N and $x \in [a, b]$,

$$\int_a^b |g_N^*(x) - g(x)| \, f_X(x)dx \leq \int_a^b |g_N(x) - g(x)| \, f_X(x)dx$$

it follows that

$$\left(\int_a^b |g_N^*(x) - g(x)|^p \, f_X(x)dx \right)^{1/p} \leq \left(\int_a^b |g_N(x) - g(x)|^p \, f_X(x)dx \right)^{1/p},$$

for all $p \geq 0$. Now, since uniform convergence implies almost sure convergence as well as convergence in L_p for $p \geq 1$, the above inequality implies that the rate at which g_N^* converges to g_N in L_p must be at least as fast as the rate at which g_N converges to g. Since convergence in L_p for $p \geq 1$ implies convergence in probability, which in turn implies convergence in distribution, it follows that when $g(x)$ is Lipschitz continuous, $g_N(X)$ as well as $g_N^*(X)$ will converge to $g\,(X)$ in distribution.

The fact that Chebyshev nodes can be calculated for any closed interval $[a, b]$ is very useful. When the statistic that is to be approximated only depends on some part of the distribution of $g(X)$, say $\{g(x) : a_y \leq g(x) \leq b_y\}$, the collocation points can be constrained to lie in the corresponding subset of the support of f_X, say $[a_x, b_x]$, by using the roots of the appropriately translated Chebyshev polynomials of the first kind. Since g_N approximates g most accurately in the vicinity of collocation points, such a set of collocation points is expected to yield a more accurate estimate of the statistic, than a set of collocation points that are not restricted to $[a_x, b_x]$ conditional on the value of $g_N(x)$ being ignored for every $x \notin [a_x, b_x]$ when calculating the estimate. When using the set of Chebyshev nodes in $[a_x, b_x]$ as collocation points, $g(x)$ may be very poorly approximated by $g_N(x)$ for $\{x : x < a_x\}$ and $\{x : x > b_x\}$ since there will be no collocation points in these regions. Consequently, allowing the values $\{g_N(x) : x < a_x \text{ or } x > b_x\}$ to contribute to the estimation of the statistic may have very large adverse effects of the accuracy of the estimate.

2.2.3 Adjusting the set of collocation points to avoid numerical instability

When the number of collocation points becomes large, the collocation points bundle up in the far ends of one or both of the tails of the distribution of X. Depending on the shape of the distribution of Y, numerical instability may arise when calculating $F_Y^{-1}(F_X(x_i^\star))$ for collocation points, $x_i^\star$, for which $F_X(x_i^\star)$ is close to zero or one. When this happens, a method called grid stretching proposed by Grzelak *et al.* (2015) can be used to pull the collocation points towards the middle of the distribution of X so as to avoid this numerical instability. In order to achieve this, the col-

location points are scaled by a factor of $\frac{1}{\sigma}$ where σ is such that $F_X\left(\frac{\max_{1\leq i\leq N}\{x_i^\star\}}{\sigma}\right)$ is equal to some prespecified value close to one, e.g. 0.995.

An alternative to the grid stretching approach to deal with numerical instabilities is to use the set of N Chebyshev nodes in $[a, b]$ where a and b are such that $F_X(a)$ and $F_X(b)$ are close to zero and one respectively, but also far enough away to avoid numerical instabilities.

2.3 Stochastic collocation sampling for distributions with an atom at zero

When a non-negative random variable Y has an atom at zero, i.e.

$$F_Y(0) > 0,$$

$F_Y^{-1}(z)$ does not exist whenever $z < F_Y(0)$. Consequently, the quantile-to-quantile transformation function, g does not exist for $x \in \{x : F_X(x) < F_Y(0)\}$.

One very naive solution to the aforementioned problem may be to set $g(x)$ equal to zero for every $x \in \{x : F_X(x) < F_Y(0)\}$ and then obtain the interpolating polynomial of the resulting function. A more extreme approach may be to ignore the region of $g(x)$ corresponding to $\{x : F_X(x) < F_Y(0)\}$ altogether. If the set

$$\{x : F_X(x) < F_Y(0)\}$$

includes one of more Gaussian quadrature points, this approach will involve ignoring those collocation points.

An alternative approach proposed by Grzelak *et al.* (2015), is to linearly extrapolate the CDF of Y to the left of $F_Y(0)$ until it hits zero at some value, y_0, and then use this extended CDF to determine the interpolating polynomial g_N. The resulting interpolating polynomial is likely to provide a more accurate approximation to $g(x)$ over the support of f_X than the two naive approaches mentioned earlier. Approximate realisations of Y may then be read off g_N^* which was defined in Section 2.2.1.8.

The extended CDF used in Grzelak *et al.* (2015) corresponds to a hypothetical variable Y^+ and is defined by

$$F_{Y^+}(y) = \begin{cases} F_Y(y) & \text{if } y > 0 \\ f(y) & \text{if } y \leq 0, \end{cases}$$

where $f(y)$ is a straight line stretching from $F_Y(0)$ to $f(y_0)$ where $y_0 < 0$ and $f(y_0) = 0$. Note that

$$F_{Y^+}^{-1}\left(F_X(x_i^\star)\right) = F_Y^{-1}(F_X(x)) \quad \text{if} \quad F_X(x^\star) \geq F_Y(0).$$

The interpolating polynomial, g_N, is then constructed to pass through each of the points in the set,

$$\left\{x_i^\star, F_{Y^+}^{-1}\left(x_i^\star\right)\right\}_{i=1}^{N}.$$

A collocation point, $x^\star$, for which $F_X(x^\star) < F_Y(0)$, is referred to by Grzelak *et al.* (2015) as a virtual collocation point.

For each value in the support of f_X, the corresponding value of Y^+ can be read off g_N. The value of Y corresponding to an arbitrary value of X, x, is determined by

$$\hat{y} := g_N^*(x) := \max\left(\sum_{i=1}^{N}\left(F_Y^+\right)^{-1}\left(F_X\left(x_i^\star\right)\right) l_i\left(x\right), 0\right).$$

This approach will be applied in Section 3.4 to generate realisations from a CEV process.

As an alternative to the approach proposed by Grzelak *et al.* (2015), the roots of the N^{th} order (translated) Chebyshev polynomial for which all N roots are contained in

$$\left[F_X^{-1}(F_Y(0)), F_X^{-1}(p)\right],$$

with p close to one, can be used as collocation points. To ensure that

$$\mathbb{P}[\hat{Y} = 0] \approx \mathbb{P}[Y = 0],$$

the value of $g_N(x)$ should be set to zero whenever $F_X(x) \leq F_Y(0)$. The resulting function aims to approximate as well as possible, the true relationship between X and the expensive variable of interest, Y, over the support of f_X, unlike the approximating function proposed by Grzelak *et al.* (2015), which aims to approximate the true relationship between X and the artificial random variable, Y^+, over the support of f_X.

This approach will also be illustrated in Section 3.4 and will be shown to outperform the approach suggested by Grzelak *et al.* (2015) for the example considered.

2.4 Goodness of fit

The chi-squared and Kolmogorov-Smirnov tests will be used to measure how well the CDF of $g(X)$ is approximated by the empirical CDF associated with the realisations generated using the stochastic collocation sampling method.

Typically interest does not lie with the accuracy of the approximation to the CDF of $g(X)$ over the entire support of f_X but rather with the accuracy of approximations to statistics that depend on only a portion of the support of f_X. When

the true value of the statistic of interest is known, the absolute difference between the true value and the approximation as well as the maximum absolute difference between g and the approximation to g over the subset of the support of f_X that is of interest, will be considered to assess the goodness of fit. The number of collocation points required to achieve an accurate estimate of a statistic depends on the statistic itself as well as the shape of the quantile-to-quantile transformation function, g.

2.5 The choice of collocation points and other practical guidelines

As mentioned earlier, there are two potential problems with the interpolating polynomial g_N used to approximate the quantile-to-quantile transformation function, g. The first is that g_N is not guaranteed to be non-decreasing; the second is that $g_N(x)$ is not guaranteed to be non-negative for non-negative random variables.

When the entire distribution of $g(X)$ is of interest, the effect of the first issue can usually be minimised by using a set of collocation points that stretches far into both tails of the distribution. Then, the region to the left of the smallest collocation point and the region to the right of the largest collocation point will be associated with very little probability weight. Consequently, if g_N is decreasing in one or both of these two tail regions, the effect of the poor approximation of $g(x)$ in these regions on the accuracy of the approximation of a statistic should be relatively small. The set of Gaussian quadrature nodes associated with the standard normal distribution works well in this regard, especially for $N \geq 5$ (see Table 2.2). The Gaussian quadrature points associated with the uniform distribution do not lie far enough into either one of the distribution's tails, while the Gaussian quadrature points associated with positively skewed distributions like the lognormal and gamma distributions do not lie far enough into the lower tails of the distributions (see Table 2.1). As an alternative to using the Gaussian quadrature points associated with the standard normal distribution, the Chebyshev nodes in the interval $[Z_\alpha, Z_\beta]$ with α close to zero and β close to one, can be used as collocation points. It is of course possible for g_N to be decreasing in the region between the smallest and largest collocation points although this behaviour is not as typical in this region as in the two tail regions. That is, unless $g(x)$ is almost flat for a reasonably large subset of the support of f_X. For such cases the interpolating polynomial may oscillate around zero, especially when a large number of collocation points is used. This behaviour of the interpolating polynomial will be illustrated in Section 3.3.

The type of collocation points that should be used depends on the statistic to be estimated. When an estimate of the unconditional expectation of $g(X)$ is re-

quired, it is beneficial to use a set of Gaussian quadrature points since the expected value of $g_N(X)$ can be evaluated exactly using the Gaussian quadrature formula $\sum_{i=1}^{N} g_N(x_i^\star)w_i$ where $\{x_i^\star\}$ is the set of Gaussian quadrature points. Note that the sample mean, $\frac{1}{n}\sum_{i=1}^{n} g_N(x_i)$, contains a sampling error and is only an estimate of $\mathbb{E}[g_N(X)]$. When a set of Chebyshev nodes are used as collocation points, the expected value of $g(X)$ cannot be evaluated exactly. Consequently, when using a set of Chebyshev nodes as collocation points, the expected value of $g(X)$ needs to be approximated using a set of realisations of $g(X)$ generated using the stochastic collocation sampling method. The computation time required to compute the sample mean, $\frac{1}{n}\sum_{i=1}^{n} g_N(x_i)$, will be longer than that required to compute $\mathbb{E}[g_N(X)]$ exactly using the Gaussian quadrature formula. Note that by the linearity of the integral operator, the unconditional expected value of any linear function of $g(X)$ can also be evaluated exactly using the Gaussian quadrature formula.

When the statistic of interest depends on a portion of the distribution of $g(X)$ only, say $\{g(x) : a_y \leq g(x) \leq b_y\}$, and it is possible to determine the corresponding subset of the support of f_X, i.e.

$$\{x : a_x \leq x \leq b_x\} = \{x : a_y \leq g(x) \leq b_y\},$$

then using N collocation points that are all located in this subset, will usually yield more accurate estimates than a set of Gaussian quadrature points, which is not restricted to that particular subset. Recall from Section 2.2.2 that using the N roots of the N^{th} order (translated) Chebyshev polynomial of the first kind that has all N its roots contained in $[a_x, b_x]$ as collocation points minimises the upper bound on the size of the interpolation error incurred for $x \in [a_x, b_x]$ and that the interpolating polynomial constructed with these Chebyshev nodes as collocation points, converges uniformly to g over the interval $[a_x, b_x]$ under certain continuity conditions. This makes Chebyshev nodes particularly useful when estimating statistics that depend on only a subset of the support of f_X. Examples considered in Sections 3.3 and 3.4 will clearly illustrate this point.

One example of a statistic that depends on only a portion of the distribution of $g(X)$ is an option price. In order to estimate the price of a call option on an underlying with price Y and strike K, the roots of the N^{th} order translated Chebyshev polynomial with roots in $\left[F_X^{-1}(F_Y(K)), F_X^{-1}(p_2)\right]$ with p_2 is close to one should be used as collocation points. After determining the interpolating polynomial g_N, its value should be set to zero for every $x \notin \left[F_X^{-1}(F_Y(K)), F_X^{-1}(p_2)\right]$ to ensure that the behaviour of $g_N(x)$ in that region does not negatively affect the estimate of the call option price. When a put-call parity relationship exists, a more accurate and precise estimate of the price of a call or put option on an underlying with price Y and

strike K will most likely be obtained by using as collocation points the roots of the N^{th} order translated Chebyshev polynomial with roots in the smaller of the the two intervals, $\left[F_X^{-1}(p_1), F_X^{-1}(F_Y(K))\right]$ with p_1 close to zero and $\left[F_X^{-1}(F_Y(K)), F_X^{-1}(p_2)\right]$ with p_2 is close to one. Note that the interval $\left[F_X^{-1}(p_1), F_X^{-1}(F_Y(K))\right]$ corresponds to the part of the distribution of Y for which the put option is in the money, while the interval $\left[F_X^{-1}(F_Y(K)), F_X^{-1}(p_2)\right]$ corresponds to the part of the distribution of Y for which the call option is in the money. If $\left[F_X^{-1}(p_1), F_X^{-1}(F_Y(K))\right]$ is the smaller of the two intervals, the stochastic collocation sampling method should be used to calculate the price of the put option with strike K using the Chebyshev nodes in that interval. The price of the call option with strike K can then be obtained using the put-call parity relationship. This approach is illustrated in Section 3.4.

When the statistic to be estimated is a percentile of the expensive distribution, it only makes sense to use samples generated with the stochastic collocation sampling method to estimate that percentile if it is not possible to determine the corresponding percentile of the cheap distribution. Otherwise the percentile of the expensive distribution, p_y, can be obtained by evaluating $F_Y^{-1}(F_X(p_x))$, where p_x is the corresponding percentile of the cheap distribution. This only entails one evaluation of the computationally expensive function g compared to the N evaluations that will be required to estimate p_y with the sample percentile obtained from the stochastic collocation realisations. If it is not possible to determine the corresponding percentile of X but it is possible to determine a range of X values that contains with certainty all x values that correspond to the bottom $(p_y \times 100)\%$ of the expensive distribution, then the set of N Chebyshev nodes in that region can be used as collocation points. All x values that fall outside of that range should be discarded. This should usually yield more accurate results than using a set of collocation points distributed over the entire distribution of X since the approximation to $g(x)$ is most accurate in the vicinity of the collocation points.

When it is not possible to determine the subset of the support of f_X that corresponds to the part of the expensive distribution that contributes to the calculation of the statistic of interest, or at least to determine some larger subset that definitely contains within it the subset of the support of f_X that is of interest, the collocation points should be spread out well over the entire support of f_X. It is important that the collocation points lie far into both the lower and upper tail so that poor approximation of $g(x)$ in the tail regions have minimal effect on the approximations of statistics of interest. The Gaussian quadrature nodes associated with the standard normal distribution exhibit this property, especially for $N \geq 5$. (see Section 2.2.1.8)

For a non-negative random variable, $g(X)$, the convergence properties of g_N^*, are at least as good as those of the unadjusted interpolating polynomial, g_N. For

a non-negative random variable, $g(X)$ negative values of $g_N(x)$ should therefore always be set to 0 even if the individual samples are not of interest. However, for a given N, the absolute error incurred when approximating an integral statistic using g_N^* is not guaranteed to be smaller than the absolute error incurred when using g_N (see Section 2.2.1.8).

Note that if $g_N(x)$ is negative for one or more values of x, the distribution of $g_N^*(X)$ will have an atom at zero, even when the distribution of $Y = g(X)$ does not. This may in turn result in a goodness-of-fit test like the Kolmogorov-Smirnov test, which considers the largest absolute difference between the true CDF of $g(X)$ and the empirical CDF of $g_N(X)$, to reject the null hypothesis that the generated samples are random realisations of the expensive distribution of interest, even if the rest of the distribution is accurately estimated.

It is important to note that while the expected value of $g_N(X)$ can be evaluated exactly using the Gaussian quadrature formula, the expected value of $g_N^*(X)$ cannot. The estimate of the expected value of $\mathbb{E}[g_N^*(X)]$ determined by the samples generated using the stochastic collocation sampling method, i.e. $\frac{1}{n}\sum_{i=1}^{n} g_N^*(x_i)$, contains a sampling error. Given this as well as the greater computation time that is required for calculating the sample mean, $\frac{1}{n}\sum_{i=1}^{n} g_N^*(x_i)$, it may be beneficial to rather use $\mathbb{E}[g_N(X)]$ as approximation of $\mathbb{E}[g(X)]$. Note however that it is not necessarily the case that the Monte Carlo estimate of $\mathbb{E}[g_N(X)]$, i.e. $\frac{1}{n}\sum_{i=1}^{n} g_N(x_i)$, is a worse estimate of $\mathbb{E}[g(X)]$ than is $\mathbb{E}[g_N(X)]$.

Chapter 3

Applications and results

In this chapter the stochastic collocation sampling method and variations thereof that were discussed in Chapter 2 are applied to examples frequently encountered in finance. Both Gaussian quadrature points, which were used as collocation points in Grzelak *et al.* (2015), as well as Chebyshev nodes, which have to date not been used in combination with the stochastic collocation sampling method, will be used to generate samples.

The chapter commences with the generation of samples from a gamma distribution using various cheap distributions. In Sections 3.2 and 3.3 the generation of realisations from a non-central chi-squared distribution without significant mass near zero and a non-central chi-squared distribution with significant mass near zero, is considered respectively. In Section 3.4 a CEV process with an atom at zero is simulated using the two approaches described in Section 2.3. The application of the stochastic collocation sampling method to sampling from a conditional distribution is illustrated in Section 3.5. In this section the sampling method is used to generate samples from the integrated variance process of an asset whose dynamics are described by a Heston model.

For each of the expensive distributions that is considered, the samples generated using the stochastic collocation sampling method will be used to approximate some statistic that depends on only a portion of the distribution. The approximation of these integral statistics will highlight the benefits of using Chebyshev nodes as collocation points. The examples considered in this chapter, especially those in Section 3.3, will illustrate that accurate estimates of statistics are possible even when the CDF of the expensive distribution is poorly approximated according to the chi-squared and Kolmogorov-Smirnov (K-S) goodness of fit tests.

3.1 Sampling from a gamma distribution

In this section approximate realisations from a gamma distribution with a shape parameter of 5 and a scale parameter of 2 will be generated using the stochastic collocation sampling method with various different cheap distributions.

The 'cheap' random variables that will be used to generate realisations of the expensive random variable, $Y \sim \Gamma(5, 2)$, are,

1. $X_1 \sim U(0, 1)$,
2. $X_2 \sim N(0, 1)$,
3. $X_3 \sim \text{logN}(2.696, 0.284^2)$ and
4. $X_4 \sim \Gamma(4, 2)$.

The distribution of X_4, which is very similar to that of Y, is not a computationally cheap distribution as such but has been included to illustrate how the results obtained with cheap distributions like the standard normal distribution, which has a shape very dissimilar to that of the distribution of Y, compare with results obtained with a distribution that is very similar in shape to that of Y. The parameters of the lognormal distribution were chosen so that the variance and skewness of the distribution match those of the distribution of Y. The entries of the Table 3.1 are the values, $\left\{F_{X_j}(x_i^{\star j})\right\}$, with $x_i^{\star j}$ denoting the i^{th} collocation point corresponding to X_j. These values reflect the relative positions of the collocation points for each of the 'cheap' distributions.

$\boldsymbol{X_1 \sim U(0,1)}$	0.04691	0.23077	0.50000	0.76923	0.95309
$\boldsymbol{X_2 \sim N(0,1)}$	0.00214	0.08761	0.50000	0.91239	0.99786
$\boldsymbol{X_3 \sim \text{logN}(2.696, 0.284^2)}$	0.04584	0.44871	0.89938	0.99637	0.99999
$\boldsymbol{X_4 \sim \Gamma(4,2)}$	0.06446	0.48081	0.89660	0.99487	0.99997

Tab. 3.1: *The relative positions of the Gaussian quadrature collocation points of the five 'cheap' distributions for $N = 5$.*

Note that the set of collocation points corresponding to the standard normal distribution, which is very dissimilar to the distribution of Y, is the only set for which there are collocation points located in the extreme lower and upper tails. The collocation points of the distribution of X_1, which is also very dissimilar to that of Y, do not stretch as far into either the lower or upper tail of the distribution. The collocation points of the distributions of X_3 and X_4, which are similar to the distribution of Y in shape, are asymmetric and do not stretch far into the lower tails of those distributions.

One hundred thousand (approximate) realisations of Y were generated using

the stochastic collocation sampling method, using each of the four 'cheap' distributions in turn. The null hypothesis that the generated realisations are random samples from the distribution of Y, were tested using chi-squared and Kolmogorov-Smirnov goodness of fit tests with a significance level of 5%. The observed test statistic values are provided in Table 3.2. The critical value for the one-sided chi-squared test is 30.14 while critical value corresponding to the Kolmogorov-Smirnov test is 1.36. The goodness of fit test statistic values indicate that the samples gen-

		$N=2$	$N=3$	$N=4$	$N=5$
Chi-squared	$X_1 \sim U(0,1)$	25155.49	21927.98	2268.36	1312.74
	$X_2 \sim N(0,1)$	7269.98	56.14	23.85	21.96
	$X_3 \sim logN(2.696, 0.284^2)$	189.77	185.36	53.26	26.61
	$X_4 \sim \Gamma(4,2)$	292.44	55.67	30.47	26.55
K-S	$X_1 \sim U(0,1)$	33.7728	19.2659	10.5113	7.2459
	$X_2 \sim N(0,1)$	19.1291	1.5014	0.6930	0.6424
	$X_3 \sim logN(2.696, 0.284^2)$	2.8122	2.4625	1.8372	1.0643
	$X_4 \sim \Gamma(4,2)$	3.5294	1.2706	0.8380	1.0338

Tab. 3.2: *Chi-squared and Kolmogorov-Smirnov test statistic values based on* 100000 *generated samples for each of* $N = 2, 3, 4$ *and* 5.

erated using the stochastic collocation sampling method with X_2 and $N = 4$ or $N = 5$ as well as those generated with X_2, X_3 and X_4 and $N = 5$, are sufficiently close to true random realisations of that distribution to be used as such. Note that according to the goodness of fit test statistic values, the samples generated using the standard normal distribution, which is very dissimilar from the distribution of Y, are more similar to true random realisations of Y than those generated using the two distributions that are similar to that of Y in shape, namely the distributions of X_3 and X_4, when four or five collocation points are used. The samples generated using the lognormal distribution, only yield better goodness of fit test statistic values than the samples generated using the standard normal distribution when two collocation points are used. The samples generated using the gamma distribution, only yield better goodness of fit test statistic values than the samples generated using the standard normal distribution when two or three collocation points are used, with the test statistic values only negligibly smaller for $N = 3$.

Table 3.3 contains the absolute errors incurred when approximating $\mathbb{E}[Y] = 10$ with the expected value of $g_N(X)$ for each of the four 'cheap' distributions and $N = 2, ..., 5$. The absolute errors indicate that the expected value of $g_N(X_2)$ is the

most accurate approximation of $\mathbb{E}[Y]$ for $N = 2, ..., 5$, while the expected value of $g_N(X_1)$ is the least accurate approximation of $\mathbb{E}[g(X)]$ for $N = 2, ..., 5$.

	$N = 2$	$N = 3$	$N = 4$	$N = 5$
$X_1 \sim U(0, 1)$	0.232685	0.118900	0.072139	0.048417
$X_2 \sim N(0, 1)$	0.003062	0.000043	0.000002	0.000000
$X_3 \sim logN(2.696, 0.284^2)$	0.021874	0.002795	0.000356	0.000031
$X_4 \sim \Gamma(4, 2)$	0.010237	0.002577	0.000900	0.000383

Tab. 3.3: *Absolute errors incurred when approximating the expected value of Y by the expected value of $g_N(X)$.*

Even though the shape of the standard normal distribution is very dissimilar to that of Y, using this distribution as the cheap distribution yielded the most desirable results in terms of the approximation of the expected value of Y and goodness of fit test statistic values. This can most likely be explained by the fact that the Gaussian quadrature points of the standard normal distribution is stretched further into both tails of the distribution than any of the other four 'cheap' distributions (see Section 2.5).

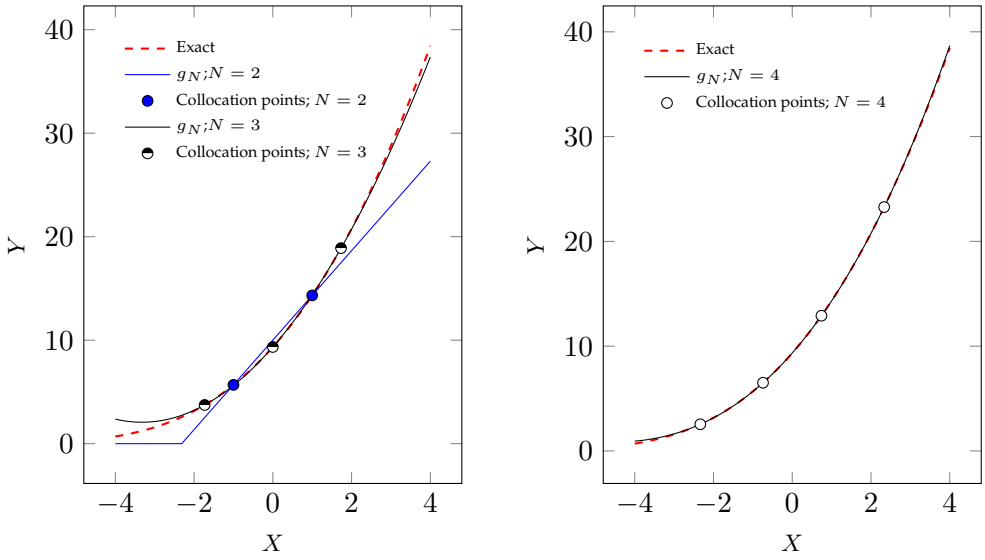
Unfortunately the time and space limitations placed on this dissertation did not allow for a more in depth investigation of the choice of the cheap distribution. Given these limitations and in light of the results above, only the standard normal distribution will be considered as cheap distribution in the remainder of this dissertation. The standard normal random variable will be denoted by X. Further research is needed to confirm whether the choice of the standard normal distribution as cheap distribution will always be appropriate.

The true relationship between $X \sim N(0, 1)$ and $g(X) = Y \sim \Gamma(5, 2)$ as well as the approximate relationships given by $g_N^*(x)$ constructed using $N = 2$, $N = 3$ and $N = 4$ Gaussian quadrature points as collocation points, are illustrated in Figure 3.1.

Recall from Section 2.5 that approximate realisations of non-negative random variables should not be read off from the interpolation polynomial, $g_N(x)$, but rather from the adjusted interpolating polynomial, $g_N^*(x)$, defined as

$$g_N^*(x) = \begin{cases} g_N(x) & \text{if } g_N(x) \geq 0 \\ 0 & \text{if } g_N(x) < 0 \end{cases}.$$

However, it was suggested that the expected value of $g_N(X)$ should be used as approximation to $\mathbb{E}[g(X)]$ since it can be evaluated exactly using the Gaussian quadrature formula unlike the expected value of $\mathbb{E}[g_N^*(X)]$, which can only be approxi-



(a) **(b)**

Fig. 3.1: The exact relationship between $X \sim N(0,1)$ and $g(X) \sim \Gamma(5,2)$ together with the approximate relationships obtained with $N = 2$ and $N = 3$ Gaussian quadrature collocation points respectively in (a) and $N = 4$ Gaussian quadrature collocation points in (b).

mated using a set of generated samples, thus also requiring greater computation time.

Table 3.4 contains the absolute errors that were incurred when $\mathbb{E}[g(X)]$ was approximated by $\mathbb{E}[g_N(X)]$ and each of the sample means,

$$\frac{1}{n}\sum_{i=1}^{n} g_N(x_i), \frac{1}{n}\sum_{i=1}^{n} g_N^*(x_i) \text{ and } \frac{1}{n}\sum_{i=1}^{n} g_N^{*Cheb}(x_i),$$

where g_N^{*Cheb} is the adjusted interpolating polynomial constructed using the set of N Chebyshev nodes contained in the interval

$$\left[F_X^{-1}(0.00001), F_X^{-1}(0.99999)\right]$$

as collocation points. The sample means were calculated using the hundred thousand samples generated using the stochastic collocation method with the standard normal distribution as cheap distribution.

Note that for $N = 2$ the sample mean, $\frac{1}{n}\sum_{i=1}^{n} g_N(x_i)$, is a more accurate estimate of $\mathbb{E}[g(X)]$ than is $\mathbb{E}[g_N(X)]$. This illustrates a point made in Section 2.5 namely that, even though the sample mean $\frac{1}{n}\sum_{i=1}^{n} g_N(x_i)$ is only an approximation of $\mathbb{E}[g_N(X)]$, it is not necessarily the case that it will be a worse estimate of

	$N = 2$	$N = 3$	$N = 4$	$N = 5$
$\mathbb{E}[g_N(X)]$	0.003062	0.000043	0.000002	0.000000
$\frac{1}{n}\sum_{i=1}^{n} g_N(x_i)$	0.000384	0.007654	0.007472	0.007454
$\frac{1}{n}\sum_{i=1}^{n} g_N^*(x_i)$	0.013505	0.007654	0.007472	0.007454
$\frac{1}{n}\sum_{i=1}^{n} g_N^{*Cheb}(x_i)$	5.241249	0.023293	0.030430	0.007849

Tab. 3.4: *Absolute errors incurred when approximating the expected value of $g(X)$.*

$\mathbb{E}[g(X)]$ than is $\mathbb{E}[g_N(X)]$. Note that the absolute error incurred when approximating the expected value of $g(X)$ using the sample mean, $\frac{1}{n}\sum_{i=1}^{n} g_N^{*Cheb}(x_i)$, is greater than the absolute error incurred when using $\mathbb{E}[g_N(X)]$.

The next example will illustrate the usefulness of Chebyshev nodes when approximating statistics that depend on only a subset of the support of f_X. Specifically, the estimation of the integral

$$\int_{-\infty}^{\infty} g(x)\mathbb{I}_{\{K_1 \leq g(x) \leq K_2\}} f(x)dx, \tag{3.1}$$

with $K_1 = F_Y^{-1}(0.1)$ and $K_2 = F_Y^{-1}(0.3)$, is considered. Note that the statistic looks similar to the price of a double barrier option. Using quadrature, the 'exact' value of the integral was determined to be 1.229. The estimates obtained using g_N^* constructed from Gaussian quadrature points and those obtained using g_N^* constructed from Chebyshev nodes in the interval $\left[F_X^{-1}(0.1), F_X^{-1}(0.3)\right]$ are compared for sample sizes of 10000 to 1000000 in steps of 10000. For convenience, let g_N^{*Gauss} and g_N^{*Cheb} denote the adjusted interpolating polynomials constructed from Gaussian quadrature points and Chebyshev nodes respectively.

Figure 3.2 presents the exact relationship between X and Y, i.e. $g(x)$ together with g_N^{*Gauss} and g_N^{*Cheb} in Figures 3.2a and 3.2b respectively. Note that for $N = 2$ and $N = 4$ there is only one Gaussian quadrature point in the interval

$$\left[F_X^{-1}(0.1), F_X^{-1}(0.3)\right],$$

while for $N = 3$ and $N = 4$ there are none (see Table 2.2). Table 3.5 contains the maximum absolute error, $|g(x) - g_N^*(x)|$, over the interval $x \in \left[F_X^{-1}(0.1), F_X^{-1}(0.3)\right]$ using one hundred thousand samples. For every N the maximum absolute error incurred using g_N^{*Cheb} is smaller than that incurred using g_N^{Gauss}. For both g_N^{Gauss} and g_N^{*Cheb} the maximum absolute error incurred decreases as N becomes larger.

The estimates of the integral obtained with $N = 3$ collocation points are provided in Figure 3.3. Even though overall g_N^{*Cheb} is a poorer approximation to $g(x)$ than g_N^{*Gauss}, it yields more accurate estimates of the integral statistic because all

	$N = 2$	$N = 3$	$N = 4$	$N = 5$
g_N^{*Gauss}	0.4691474248	0.0556123853	0.0065125922	0.0001647612
g_N^{*Cheb}	0.0427026626	0.0004409501	0.0000042321	0.0000000114

Tab. 3.5: *Maximum absolute error incurred over the interval* $\left[F_X^{-1}(0.1), F_X^{-1}(0.3)\right]$.

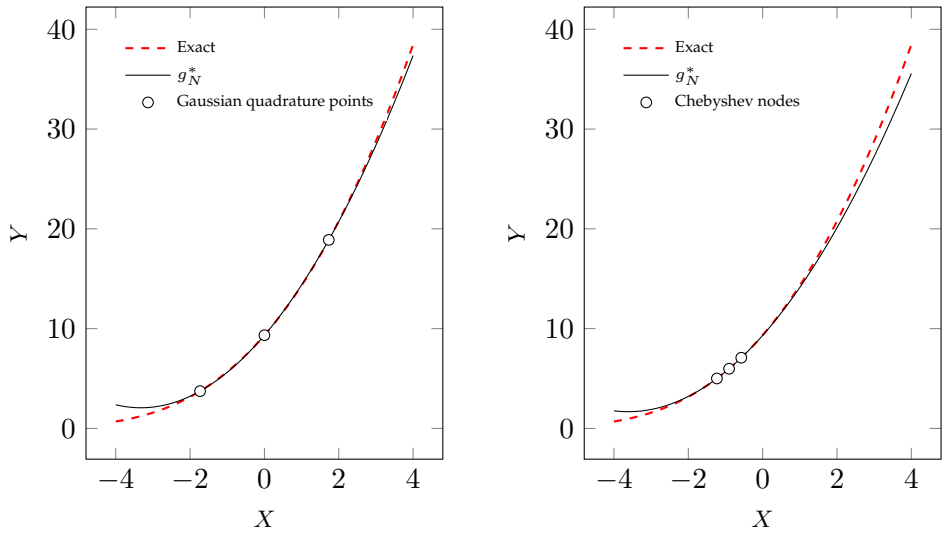


Fig. 3.2: The interpolating polynomial constructed using (a) $N = 3$ Gaussian quadrature points and (b) $N = 3$ Chebyshev nodes.

three collocation points are in the region of interest. This illustrates how choosing a set of collocation points located in the region of interest can improve the accuracy of an estimate.

3.2 Sampling from a non-central chi-squared distribution

In this section approximate realisations from a non-central chi-squared distribution with $d = 4$ degrees of freedom and a non-centrality parameter of $\lambda = 3$ are generated using the stochastic collocation sampling method. The standard normal distribution will be considered as cheap distribution. The overall fit of the CDF as measured by goodness of fit tests, the accuracy of approximations to the unconditional expected value as well as the accuracy of approximations to a tail integral statistic will be considered.

The true relationship between $X \sim N(0,1)$ and $Y \sim \chi_4'^2(3)$ as well as the approximate relationships given by $g_N^*(x)$ constructed using $N = 2$, 3, 4 and 5

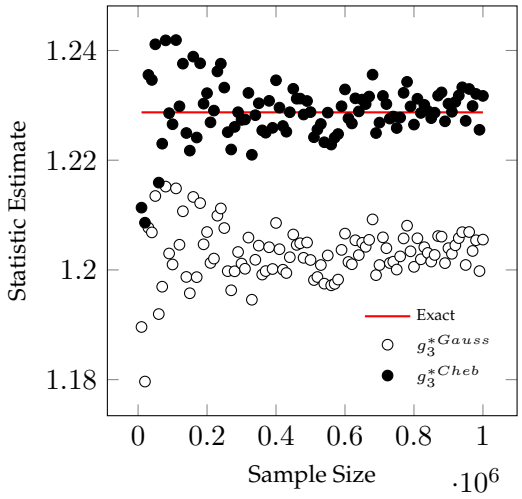


Fig. 3.3: Estimates of the integral statistic in (3.1).

Gaussian quadrature points as collocation points, are presented in Figure 3.4.

Ten sets of 100000 samples each were generated using the stochastic collocation sampling method with $N = 2, 3, 4, 5$ and 6 Gaussian quadrature collocation points respectively. The mean computation times are presented in Table 3.6. The mean time required to obtain 100000 realisations using the inverse transform method was 4.95 seconds while Matlab's built in function, ncx2rnd, required a mere 0.027 seconds. The results show that up to $N = 5$ collocation points, it is faster, on average, to obtain 100000 approximate samples from the non-central chi-squared distribution using the stochastic collocation sampling method than it is to obtain the same number of samples using Matlab's built in function. When six collocation points are used, the computation times are approximately the same.

$\boldsymbol{N=2}$	$\boldsymbol{N=3}$	$\boldsymbol{N=4}$	$\boldsymbol{N=5}$	$\boldsymbol{N=6}$
0.0191	0.0185	0.0210	0.0234	0.0266

Tab. 3.6: *Mean computation times for generating* 100000 *approximate realisations from* $\chi_4'^2(3)$

The null hypothesis that the samples generated using the stochastic collocation sampling method are indeed random realisations from the non-central chi-squared distribution of interest was tested with chi-squared and Kolmogorov-Smirnov tests. Each of the values in Table 3.7 is the mean of the 10 realisations of the goodness of fit test statistic obtained from the ten sets of generated samples. In each cell the value in brackets is the proportion of realisations for which the null hypothesis

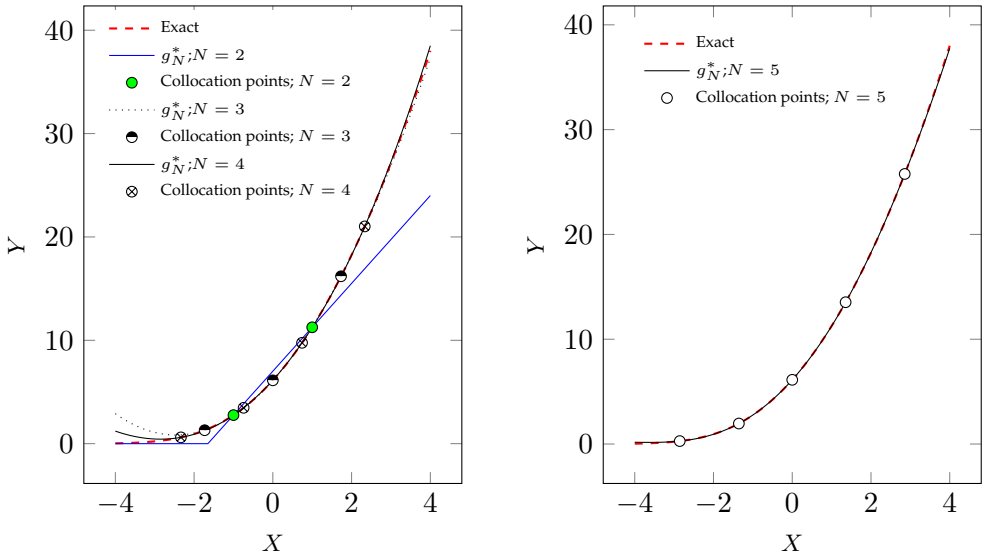


Fig. 3.4: Exact and approximate relationships between $X \sim N(0,1)$ and $Y \sim \chi_4'^2(3)$.

	$N=2$	$N=3$	$N=4$	$N=5$	$N=6$
Kolmogorov-Smirnov	25.75(1)	5.71(1)	1.64(1)	0.94(0)	0.83(0)
Chi-squared	13353.25(1)	71.07(1)	37.23(0.8)	20.38(0)	20.02(0)

Tab. 3.7: *Goodness of fit test statistic values.*

was rejected.

The critical values for the one-sided Kolmogorov-Smirnov and chi-squared tests performed at a significance level of 5%, are 1.36 and 30.14 respectively. The results show that at a 5% significance level the aforementioned null hypothesis was rejected for every realisation of the Kolmogorov Smirnoff test when the samples were generated using 4 or less collocation points. However, when using five or six collocation points, the null hypothesis could not be rejected for any of the ten sets of samples. Based on the chi-squared test the null hypothesis was rejected for all ten sets of samples generated using $N = 2$ and $N = 3$ collocation points and rejected for eight of the ten sets generated with $N = 4$ collocation points. As for the Kolmogorov-Smirnov test statistic, the null hypothesis could not be rejected for any of the sample sets generated using $N = 5$ or $N = 6$ collocation points. The goodness of fit results therefore indicate that, when $N = 5$ and $N = 6$ Gaussian quadrature points are used as collocation points, the samples generated with

the stochastic collocation sampling method are indistinguishable from true random realisations from the non-central chi-squared distribution of interest. In the next section sampling from a more extreme non-central chi-squared distribution with significant mass near zero will be considered. Obtaining sufficiently accurate realisations from that distribution will prove much more difficult.

Table 3.8 contains the absolute error incurred when approximating the expected value of Y, which is known to be 7, with the expected value of $g_N(X)$ as a function of the number of Gaussian quadrature collocation points, N.

$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
0.0106	0.0003	0.0002	0.0000	0.0000

Tab. 3.8: *Absolute errors incurred when the expected value of* $\chi_4'^2(3)$ *is approximated by* $\mathbb{E}[g_N(X)]$.

As an example of a statistic that depends on only a subset of the support of f_X, consider the integral,

$$\int_{F_X^{-1}(0.0001)}^{F_X^{-1}(0.05)} g\left(x\right) f_X\left(x\right) dx. \tag{3.2}$$

The estimate of this statistic obtained from samples generated using $N = 3$ Gaussian quadrature points as collocation points will be compared to the estimate obtained from samples generated using the three roots of the translated Chebyshev polynomial of the first kind with roots in $\left[F_X^{-1}(0.0001), F_X^{-1}(0.05)\right]$ as collocation points. As before, let g_N^{*Gauss} and g_N^{*Cheb} denote the adjusted interpolating polynomials constructed from Gaussian quadrature points and Chebyshev nodes respectively. Figures 3.5a and 3.5b present the exact relationship between X and Y, i.e. g(x), together with g_N^{*Gauss} and g_N^{*Cheb} respectively.

The maximum absolute errors incurred over the interval

$$\left[F_X^{-1}(0.0001), F_X^{-1}(0.05)\right]$$

when approximating $g(x)$ by g_N^{*Gauss} and g_N^{*Cheb} respectively, are presented in Table 3.9. Note that not only are the maximum absolute errors incurred when using g_N^{*Cheb} to approximate $g(x)$ much smaller than those incurred when using g_N^{*Gauss} to approximate $g(x)$, but $\left|g(x) - g_N^{*Cheb}\right|$, unlike $\left|g(x) - g_N^{*Gauss}\right|$, is a strictly decreasing function of N.

Note that the rejection of the null hypothesis that the samples generated using $N = 3$ Gaussian quadrature points as collocation points are random realisations

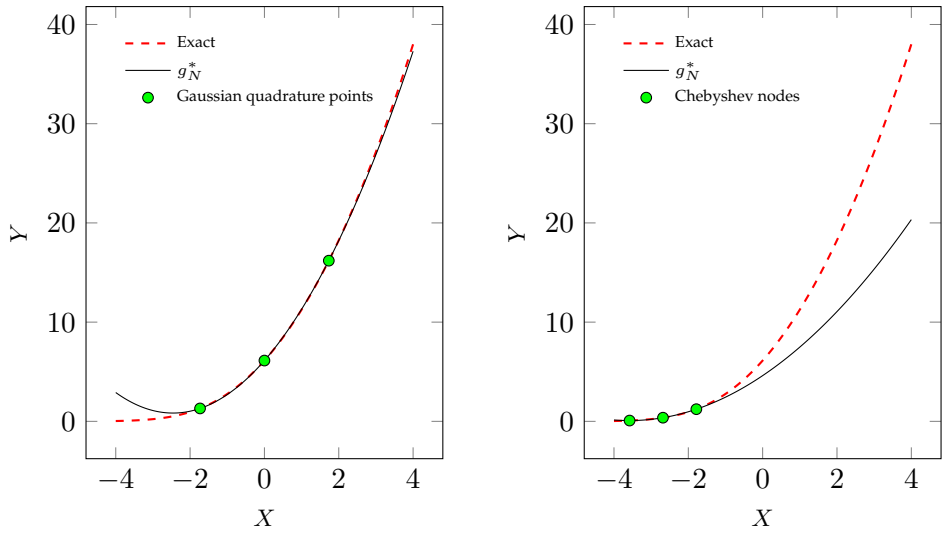


Fig. 3.5: (a) The interpolating polynomial constructed using $N = 3$ Gaussian quadrature points as collocation points. (b) The interpolating polynomial constructed using $N = 3$ Chebyshev nodes in $\left[F_X^{-1}(0.0001), F_X^{-1}(0.05)\right]$ as collocation points.

	$\boldsymbol{N=2}$	$\boldsymbol{N=3}$	$\boldsymbol{N=4}$	$\boldsymbol{N=5}$	$\boldsymbol{N=6}$
Gaussian quadrature	1.426791	2.153179	0.825066	0.096969	0.053247
Chebyshev	0.244005	0.027633	0.001198	0.000283	0.000023

Tab. 3.9: *Maximum absolute errors incurred over the interval* $\left[F_X^{-1}(0.0001), F_X^{-1}(0.05)\right]$.

from the non-central chi-squared distribution of interest does not guarantee that those samples will yield a poor estimate of a statistic that depends on only a subset of the support of f_X. Notice however that a large portion of the subset of interest for the estimation of (3.2) lies to the left of the smallest Gaussian quadrature point and contains only that one Gaussian quadrature point. Since interpolating polynomials often exhibit wild behaviour in the regions to the left and right of the smallest and largest interpolation points respectively, the integral statistic is likely to be poorly estimated by those samples. Figure 3.6 presents the estimates of the statistic in (3.2) obtained using g_3^{*Gauss} and g_3^{*Cheb} respectively, for sample sizes from 10000 to one million.

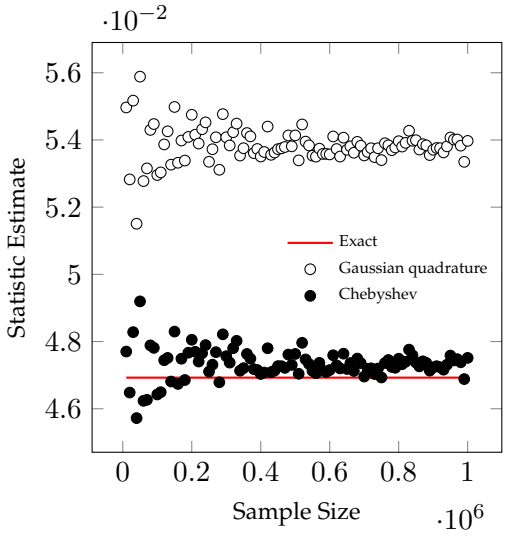


Fig. 3.6: The exact value of the tail integral statistic in (3.2) together with estimates based on samples generated using $N = 3$ (a) Gaussian quadrature points and (b) Chebyshev nodes as collocation points. The value of g_3^{*Gauss} and g_3^{*Cheb} was set to zero for every $x \notin \left[F_X^{-1}(0.0001), F_X^{-1}(0.05)\right]$.

3.3 Sampling from a non-central chi-squared distribution with significant mass near zero

In this section the stochastic collocation sampling method will be used to generate approximate realisations from a non-central chi-squared distribution with $d = 0.08$ degrees of freedom and a non-centrality parameter of $\lambda = 0.00054$. This is a very extreme non-central chi-squared distribution that has significant probability mass near zero. Such extreme distributions do however arise in financial applications. In fact, the unconditional distribution of the variance, $V(T)$, of an asset price, $S(T)$, with dynamics described by a Heston stochastic volatility model, i.e.

$$\begin{aligned} dS(t) &= rS(t)\,dt + \sqrt{V(t)}S(t)\,dW_S(t), \quad S(t_0) = S_0 \\ dV(t) &= \kappa\left(\bar{V} - V(t)\right)dt + \omega\sqrt{V(t)}dW_V(t), \quad V(t_0) = V_0, \end{aligned}$$

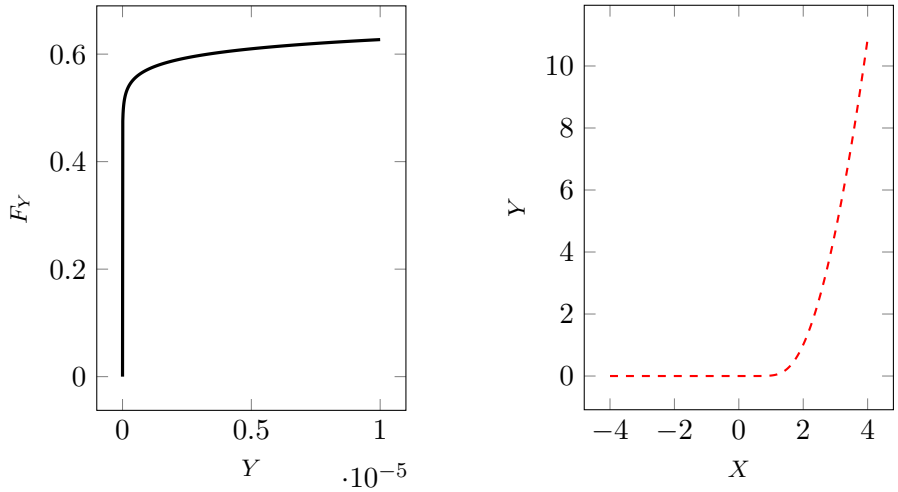
where W_S and W_V are two correlated Brownian motions, with

$$\begin{aligned} S(0) = 100, \quad r = 0, \quad \omega = 1, \quad \kappa = 0.5, \quad \bar{V} = V_0 = 0.04, \\ \text{corr}\left(W_S, W_V\right) = -0.9 \text{ and } \quad T = 10, \end{aligned}$$

is proportional to this extreme non-central chi-squared distribution (Grzelak *et al.*, 2015) [1]. See equations (3.6) and (3.7) for the relationship between the parameters of

[1] See the November 2014 version of this paper.

the Heston model and that of the non-central chi-squared unconditional variance.



(a) **(b)**

Fig. 3.7: (a) The CDF of Y. (b) The exact relationship between X and Y.

The CDF of this particular non-central chi-squared distribution is presented in Figure 3.7a. Note the very steep gradient of the CDF in the lower tail of the distribution, indicating significant probability mass near zero. In fact, approximately 70% of the distribution's probability weight corresponds to $0 < Y \leq 0.000156$. This translates into $g(x)$ exhibiting a large seemingly flat region, before it starts increasing visibly. This is reflected in Figure 3.7b, which illustrates the true relationship between $X \sim N(0,1)$ and $Y \sim \chi'^2_{0.08}(0.00054)$, i.e. $g(x)$.

Due to the fact that $g(x)$ is almost flat for at least the bottom 70% of the support of f_X, most of the collocation points are located in that region. Consequently, there are relatively few collocation points in the region where $g(x)$ is increasing more rapidly, especially when a small number of collocation points, like $N = 4$, is used. This results in poor approximation of $g(x)$ in the upper tail of the distribution of X. This is illustrated in Figure 3.8a, which presents the exact relationship between X and Y, i.e. $g(x)$, for $x \in [0.5, 4]$, together with the adjusted interpolating polynomial, $g^*_N(x)$, constructed from four Gaussian quadrature collocation points. Increasing N, and hence the order of the interpolating polynomial, on the other hand, yields a polynomial that oscillates around zero in the region where $g(x)$ is very close to zero. This is illustrated in Figure 3.8b, which shows $g(x)$ together with the (unadjusted) interpolating polynomial constructed with six Gaussian quadrature collocation points, $g^*_6(x)$. A variation of the stochastic collocation method that is based on piecewise polynomial interpolation instead of global polynomial interpolation may yield better results for a distribution like this. Such investigation is

however beyond the scope of this dissertation.

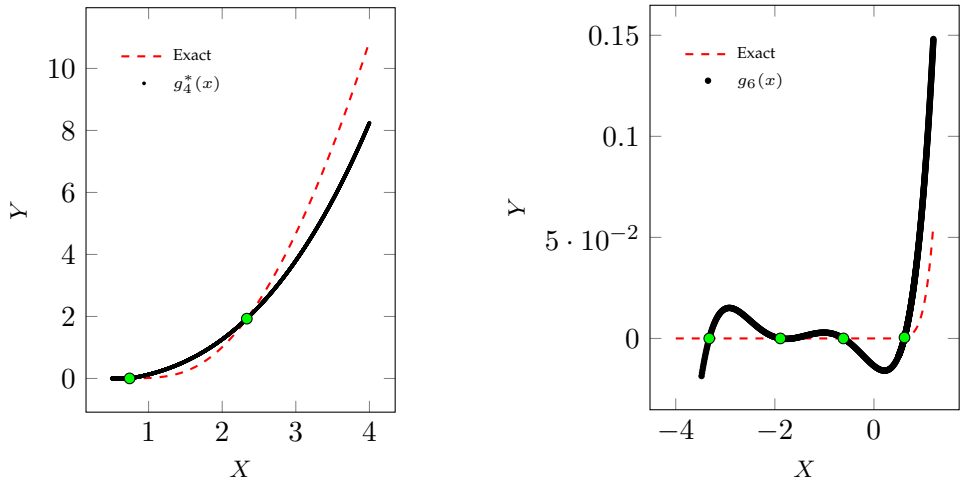


Fig. 3.8: The exact relationship between $X \sim N(0,1)$ and $Y \sim \chi'^2_{0.08}(0.00054)$ together with $g_4^*(x)$ in (a) and $g_6(x)$ in (b).

Figure 3.9 presents the true CDF of Y, F_Y, together with the empirical CDF based on 100000 samples generated using the stochastic collocation sampling method with six Gaussian quadrature collocation points. Note that the distribution of $g_6^*(X)$, unlike that of $g(X)$, has an atom at zero as a result of setting all negative values of the interpolating polynomial equal to zero. Consequently, the true and approximate CDFs differ greatly at $x = 0$. The poor approximation of $g(x)$ over the support of f_X results in the rejection of the null hypothesis that the generated samples are random realisations from the non-central chi-squared distribution of interest (see Table 3.10).

Ten sets of 100000 samples each were generated using the stochastic collocation sampling method with $N = 4$, 6, 8 and 10 Gaussian quadrature collocation points. Table 3.10 contains the mean Kolmogorov-Smirnov and chi-squared test statistics for $N = 4$, 6, 8 and 10. The values in brackets are the proportions of the ten sets for which the null hypothesis was rejected. Both goodness of fit tests rejected the null hypothesis for all ten sample sets generated with ten collocation points. In fact, even when using $N = 30$ Gaussian quadrature collocation points the null hypothesis was rejected by both tests for all ten sets of generated samples.

Table 3.11 contains the mean computation times for $N = 2$, 4, 6, 8 and 10. The mean time required to generate 100000 samples with the inverse transform method was 10 minutes 88 seconds while Matlab's built in function, *ncx2rnd*, required a mere 0.0288 seconds. The mean time required to obtain 100000 realisations using

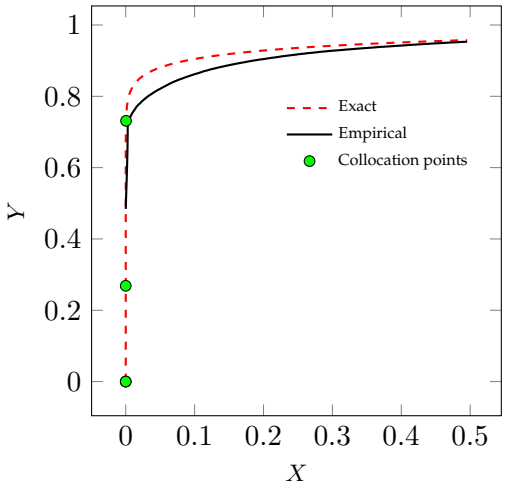


Fig. 3.9: True CDF and empirical CDF of $Y \sim \chi'^2_{0.08}(0.00054)$ determined by 100000 stochastic collocation samples obtained using $N = 6$ Gaussian quadrature collocation points.

	$\boldsymbol{N = 4}$	$\boldsymbol{N = 6}$	$\boldsymbol{N = 8}$	$\boldsymbol{N = 10}$
Kolmogorov-Smirnov	116.72(1)	109.77(1)	111.44(1)	96.91(1)
Chi-squared	625092.20(1)	485231.79(1)	389097.28(1)	495872.50(1)

Tab. 3.10: *Mean goodness of fit test statistics for ten sets of* 100000 *generated samples.*

the inverse transform method was 10 minutes 88 seconds while Matlab's built in function, *ncx2rnd*, required a mere 0.0288 seconds. The reason for the very large computation time that was required when using the inverse transform method is that the root finding function, *fzero*, and not the function *ncx2inv*, was used to determine the set of values, $\left\{F_Y^{-1}(F_X(x_i^\star))\right\}_{i=1}^N$, as well as the exact relationship between X and Y. The function *fzero* was used because the function *ncx2inv* did not provide the correct values in the region where the gradient of the CDF of Y is very steep. The mean time required to generate 100000 samples with the stochastic collocation sampling method when $N = 10$ collocation points were used, was more than three

$\boldsymbol{N = 4}$	$\boldsymbol{N = 6}$	$\boldsymbol{N = 8}$	$\boldsymbol{N = 10}$
0.0506	0.0713	0.0834	0.0999

Tab. 3.11: *Mean computation times for stochastic collocation sampling of* 100000 *realisations of* $\chi'^2_{0.08}(0.00054)$.

times the computation time required by Matlab's built in function, *ncx2rnd*.

Table 3.12 contains the absolute errors incurred when the expected value of Y, which is known to be 0.08054, is approximated by $\mathbb{E}[g_N(X)]$. The errors indicate that reasonably accurate estimates of the unconditional expected value is obtained using as few as $N = 8$ Gaussian quadrature collocation points.

$N = 4$	$N = 6$	$N = 8$	$N = 10$
0.008762	0.004934	0.000968	0.000337

Tab. 3.12: *Absolute errors incurred when approximating the expected value of* $\chi'^2_{0.08}(0.00054)$ *with* $\mathbb{E}[g_N(X)]$.

Next, the approximation of the integral,

$$\int_{F_X^{-1}(0.5)}^{F_X^{-1}(0.95)} g\left(x\right) f_X\left(x\right) dx, \tag{3.3}$$

is considered. The approximation of the integral based on samples generated with $N = 9$ Gaussian quadrature points and $N = 9$ Chebyshev nodes in the interval

$$\left[F_X^{-1}(0.5), F_X^{-1}(0.95)\right]$$

respectively, is calculated using sample sizes of 10000 up to one million in steps of 10000. Figure 3.10 presents the true relationship between X and Y together with the approximate relationships based on each of the two sets of collocation points, respectively. Denote the adjusted interpolating polynomials constructed from the set of Gaussian quadrature points and Chebyshev nodes by g_9^{*Gauss} and g_9^{*Cheb} respectively.

The estimates of the integral statistic in (3.3) are presented in Figure 3.11. When the integral statistic was estimated, all values of $g_N^*(x)$ for which $x < F_X^{-1}(0.5)$ or $x > F_X^{-1}(0.95)$ were set to zero to prevent poor approximation of $g(x)$ in these regions from affecting the accuracy of the estimates negatively. Figure 3.12a contains the approximations of the integral statistic using $N = 9$ Chebyshev nodes in $\left[F_X^{-1}(0.5), F_X^{-1}(0.95)\right]$ as collocation points without setting the value of $g_N^*(x)$ to zero for $x \notin \left[F_X^{-1}(0.5), F_X^{-1}(0.95)\right]$. This example again illustrates how beneficial it is to choose a set of collocation points that lie in the region of interest and that in doing so, accurate estimates can be obtained even when the distribution as a whole is poorly approximated.

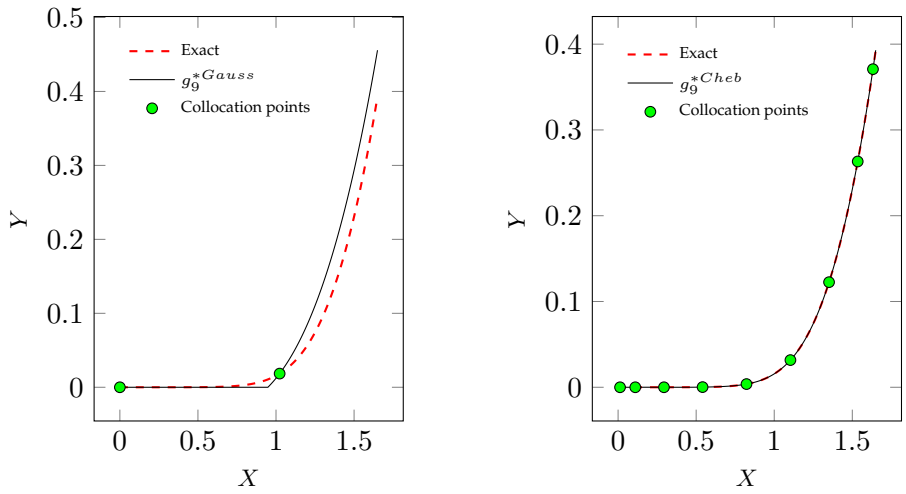


Fig. 3.10: The exact relationship between X and Y together with g_9^{*Gauss} in (a) and g_9^{*Cheb} in (b).

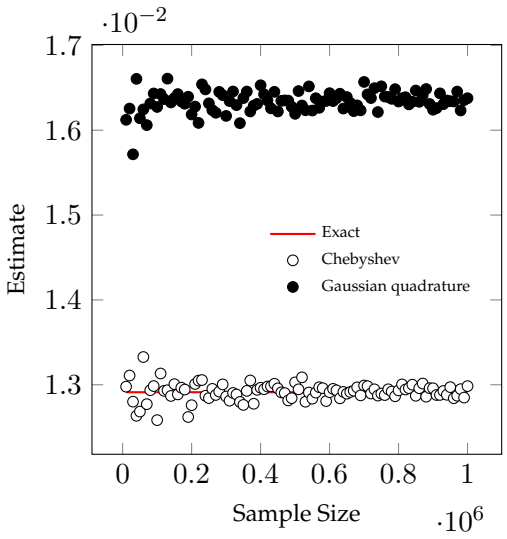


Fig. 3.11: Estimates of the integral statistic in (3.3) obtained from 100000 samples generated using $N = 9$ Gaussian quadrature points and $N = 9$ Chebyshev nodes as collocation points respectively. The value of g_9^{*Gauss} and g_9^{*Cheb} was set to zero for every $x \notin \left[F_X^{-1}(0.5), F_X^{-1}(0.95)\right]$.

3.4 Simulating a CEV process with an atom at zero

Consider an asset $S(t)$ with price process described by the CEV process,

$$dS(t) = \sigma S^{\beta}(t) dW(t),$$

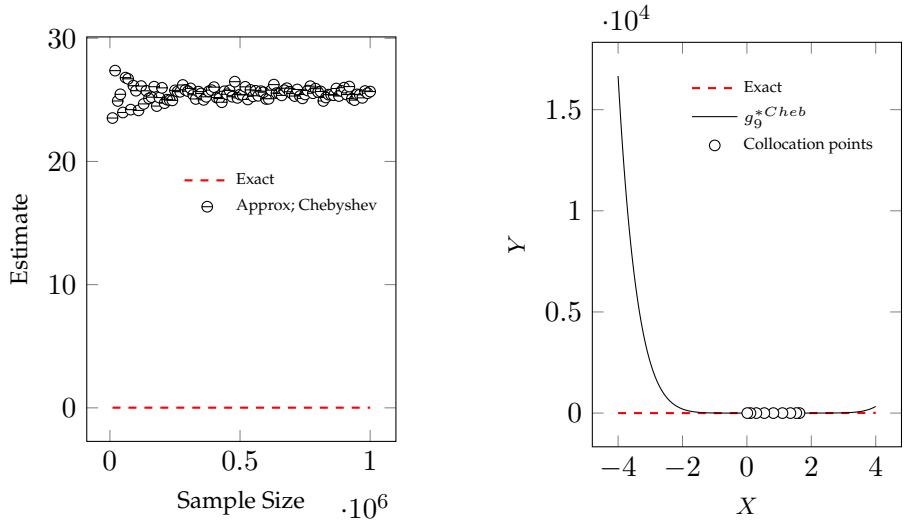


Fig. 3.12: (a) Estimates of the integral statistic in (3.3) obtained from 100000 samples generated using $N = 9$ Chebyshev nodes as collocation points. The values of g_9^{*Cheb} for $x \notin \left[F_X^{-1}(0.5), F_X^{-1}(0.95)\right]$ were not changed to zero.; (b) The exact relationship between X and Y together with g_9^{*Cheb}.

with $\sigma = 0.4$, $\beta = 0.5$ and $S_0 = 0.07$. The distribution of $S(2)$ has an atom at zero. Specifically, $\mathbb{P}[S(2) = 0] = 0.65$. The stochastic collocation sampling method will be used to generate realisations from the distribution of $S(2)$, using the two approaches described in Section 2.3. Both approaches are applied using five collocation points.

Schroder (1989) proved that the (risk neutral) CDF of $S(T)$ can be expressed in terms of the CDF of a non-central chi-squared distribution in the following way:

$$P\left[S(T) \leq y | S(0)\right] = 1 - F_{\chi_b'^2(c(y))}(a)$$

where

$$a = \frac{S_0^{2(1-\beta)}}{(1-\beta)^2 \sigma^2 T}, \quad b = \frac{1}{1-\beta}, \quad c\left(y\right) = \frac{y^{2(1-\beta)}}{(1-\beta)^2 \sigma^2 T} \quad .$$

3.4.1 Approach proposed by Grzelak *et al.* (2015)

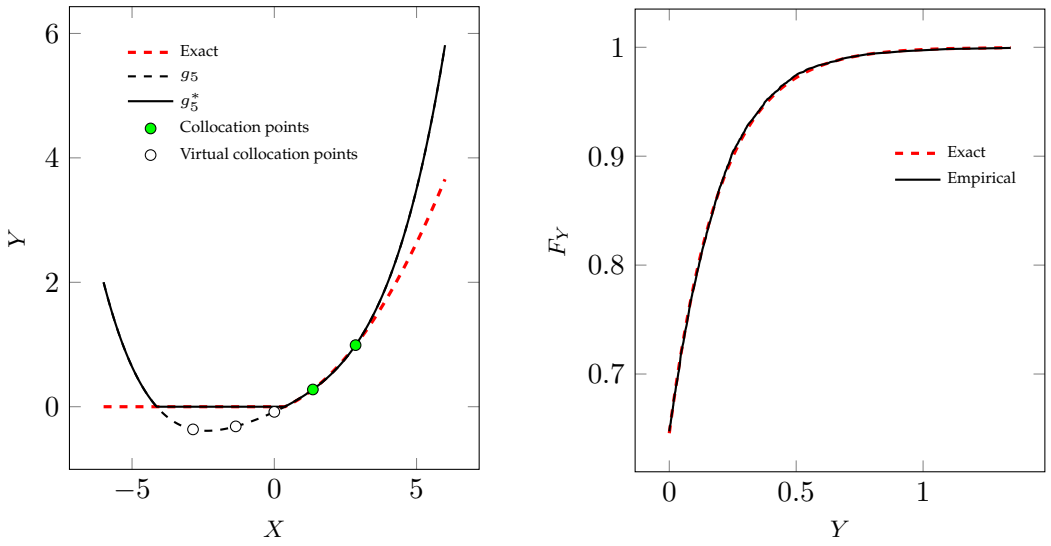
For a given realisation, x, of $X \sim N(0,1)$, the corresponding (approximate) realisation of $S(2)$ is given by

$$s = \max\left(\sum_{i=1}^{N} F_{S(2)^+}^{-1}\left(F_X\left(x_i^\star\right)\right) l_i\left(x\right), 0\right).$$

	$\boldsymbol{x_1^\star}$	$\boldsymbol{x_2^\star}$	$\boldsymbol{x_3^\star}$	$\boldsymbol{x_4^\star}$	$\boldsymbol{x_5^\star}$
$x_i^\star$	-2.86	-1.36	-0.00	1.36	2.86
$F_{N(0,1)}(x_i^\star)$	0.00	0.09	0.50	0.91	1.00
$F^{-1}_{S(2)}(F_X(x_i^\star))$	–	–	–	0.28	0.99
$F^{-1}_{S(2)^+}(F_X(x_i^\star))$	-0.36	-0.32	-0.08	0.28	0.99

Tab. 3.13: *Collocation points and corresponding values of* $S(2)$.

Table 3.13 contains the five Gaussian quadrature collocation points as well as the exact inverse of the CDF of $S(2)$, F_S, and the inverse of the extended CDF at each of the collocation points. Note that there are three collocation points for which $F_X(x^\star) < F_S(0) = 0.65$, i.e. virtual collocation points. The inverse CDF of $S(2)$ does not exist for these three collocation points. Figure 3.13a presents the exact and approximate relationship between the variables X and $S(2)$ while Figure 3.13b presents the exact and approximate (empirical) CDFs.



(a) **(b)**

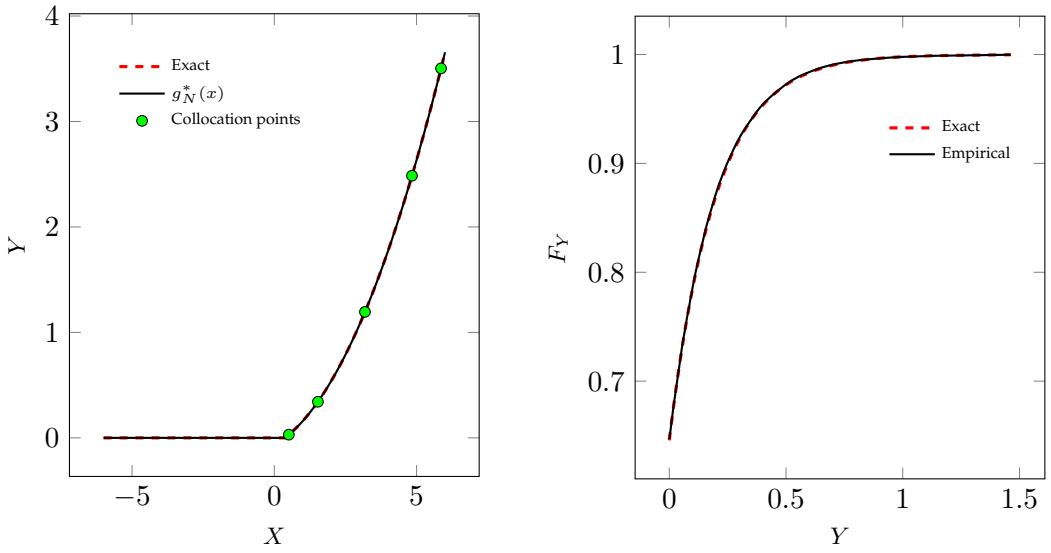
Fig. 3.13: The exact relationship between the standard normal random variable, X, and $S(2)$ together with the approximate relationship obtained by using the five Gaussian quadrature points as collocation points. (b) The CDF of $S(2)$ together with the empirical CDF obtained from 100000 generated samples using the collocation points in (a).

3.4.2 The Chebyshev nodes approach

Following the approach in Section 2.3, the roots of the 5^{th} order translated Chebyshev polynomial of the first kind that has all its roots in the interval,

$$\left[F_X^{-1}(F_S(0)), F_X^{-1}(0.9999)\right],$$

are used as collocation points when approximating the relationship between X and $S(2)$. After constructing the interpolating polynomial with these collocation points, the value of the polynomial is set to zero for every $x \leq F_X^{-1}(F_Y(0))$. This will ensure that $\mathbb{P}[\hat{S(2)} = 0] \approx F_Y(0)$ for sufficiently large sample sizes. The resulting function is used to approximate the exact relationship between X and Y. Figure 3.14a presents the exact and approximate relationships between X and Y. The empirical CDF corresponding to 100000 approximate realisations of $S(2)$ obtained using this approach is represented in Figure 3.14b together with the true CDF of $S(2)$. Upon comparison of Figures 3.13 and 3.14, it is clear that the approach implemented here, using Chebyshev nodes as collocation points, yielded a more accurate approximation of $g(x)$ than did the approach proposed by Grzelak *et al.* (2015).



(a) **(b)**

Fig. 3.14: (a) The exact relationship between the standard normal random variable, X, and $S(2)$ together with the approximate relationship obtained by using the $N = 5$ Chebyshev nodes in $\left[F_X^{-1}(F_S(0)), F_X^{-1}(0.9999)\right]$ as collocation points. (b) The CDF of $S(2)$ together with the empirical CDF obtained from 100000 generated samples using the collocation points in (a).

3.4.3 Option pricing

In this section, approximate realisations of $S(T)$ generated using the stochastic collocation sampling method will be used to approximate vanilla European call and put options with strike, $K = 0.08$ and time to maturity, $T = 2$. Closed form solutions for vanilla European call and put options on an asset with a price process described by the CEV process,

$$dS(t) = rStdt + \sigma S^{\beta}(t)dW(t),$$

were derived in Schroder (1989) and are given by,

$$c(S_0, K, \sigma, \beta, T) = S_0 \left[1 - F_{\chi'^2_d(\lambda)}(y)\right] - Ke^{-rT} F_{\chi^2_{(d-2,y)}}(\lambda)$$

and

$$p(S_0, K, \sigma, \beta, T) = -S_0 F_{\chi'^2_d(\lambda)}(y) + Ke^{-rT}\left[1 - F_{\chi'^2_{d-2}(y)}(\lambda)\right]$$

respectively, where

$$\lambda = \kappa S_0^{2(1-\beta)} e^{2r(1-\beta)T}, \quad y = \kappa K^{2(1-\beta)}, \quad d = 2 + \frac{1}{1-\beta}$$

and

$$\kappa = \frac{2r}{\sigma^2 (1-\beta)\left(e^{2r(1-\beta)} - 1\right)}.$$

When the risk free interest rate, r, is zero as in the example considered in Section 3.4, the call and put pricing formulas become,

$$c(S_0, K, \sigma, \beta, T) = S_0 \left[1 - F_{\chi'^2_d(\lambda)}(y)\right] - KF_{\chi'^2_{d-2}(y)}(\lambda)$$

and

$$p(S_0, K, \sigma, \beta, T) = -S_0 F_{\chi'^2_d(\lambda)}(y) + K\left[1 - F_{\chi'^2_{d-2}(y)}(\lambda)\right]$$

respectively, with

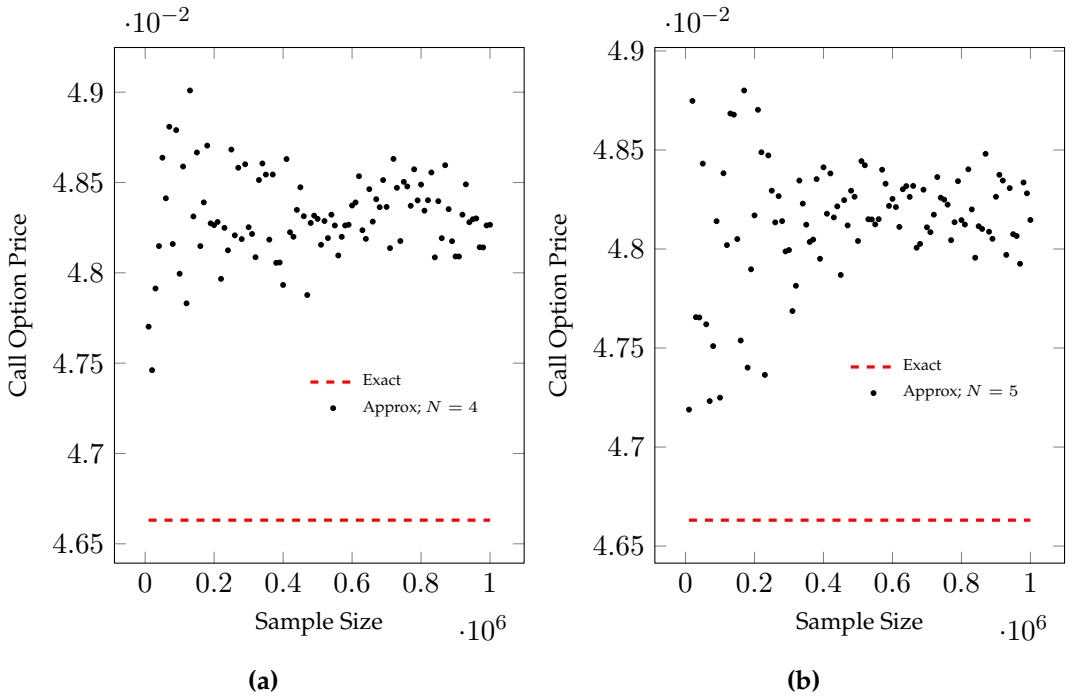
$$\lambda = \kappa S_0^{2(1-\beta)}, \quad y = \kappa K^{2(1-\beta)}, \quad d = 2 + \frac{1}{1-\beta}$$

and

$$\kappa = \frac{1}{\sigma^2 (1-\beta)^2 T}.$$

When calculating an estimate of an option price using samples generated using the stochastic collocation sampling method, the values of the interpolating polynomial corresponding to x-values in the subset(s) of f_X for which the option is out-of-the-money, should be ignored, irrespective of the set of collocation points used. That is, when estimating the price of a call option with strike K, the set of values, $\{g_N^*(x) : x < F_S^{-1}(K)\}$, should be ignored. Similarly, the set of values, $\{g_N^*(x) : x > F_S^{-1}(K)\}$, should be ignored when estimating the price of a put option with strike K.

The estimates of the call option with strike, $K = 0.08$ and time to maturity, $T = 2$, calculated from approximate realisations of $S(2)$ generated using the stochastic collocation sampling method with $N = 4$ and $N = 5$ Gaussian quadrature collocation points respectively are presented in Figures 3.15a and 3.15b, for sample sizes starting at 10000 up to one million.



(a) **(b)**

Fig. 3.15: The exact call option price together with the approximate call option prices obtained from samples generated using the set of $N = 4$ Gaussian quadrature collocation points in (a) and $N = 5$ Gaussian quadrature collocation points in (b).

Next, the call option price is estimated using Chebyshev nodes contained in the interval

$$\left[F_X^{-1}(F_S(0.08)), F_X^{-1}(0.9999)\right],$$

as collocation points.

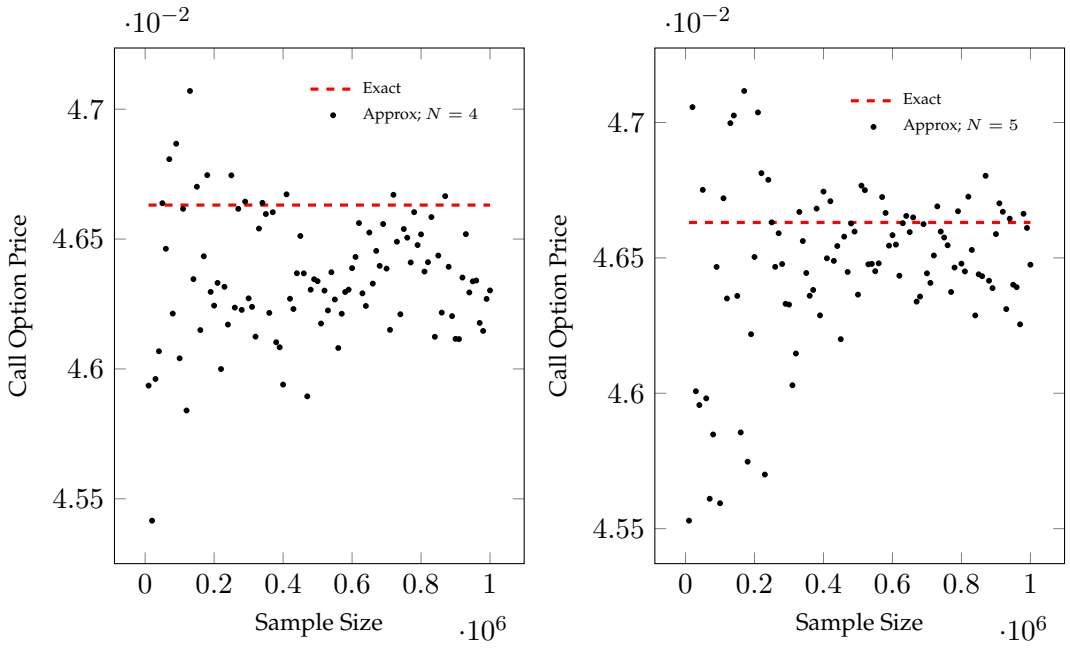


Fig. 3.16: The exact call option price together with the approximate call option prices obtained from samples generated using the set of $N = 4$ Chebyshev nodes in the interval $\left[F_X^{-1}(F_S(0.08)), F_X^{-1}(0.9999)\right]$ as collocation points in (a) and the set of $N = 5$ Chebyshev nodes in the interval $\left[F_X^{-1}(F_S(0.08)), F_X^{-1}(0.9999)\right]$ as collocation points in (b).

The put-call parity relationship,

$$c + Ke^{-rT} = p + S_0,$$

can be utilised to obtain more accurate and precise estimates of the call option price by constructing the interpolating polynomial from the Chebyshev nodes in the smaller of the two intervals,

$$\left[F_X^{-1}(F_S(K)), F_X^{-1}(0.9999)\right] \tag{3.4}$$

and

$$\left[F_X^{-1}(0.9999), F_X^{-1}(F_S(K))\right]. \tag{3.5}$$

The approximate price of the option which is in-the-money in that smaller interval should then be calculated using the set of samples generated from the interpolating

polynomial. If the interval in (3.5) is the smaller of the two intervals, that means, that the generated samples would be used to approximate the put option price. The corresponding approximate call option price can then be determined using the put call parity relationship. For $K = 0.08$, it is indeed the interval in (3.4) which is the smaller of the two intervals. Following the approach just described, the approximate call option prices were obtained and presented in Figure 3.17.

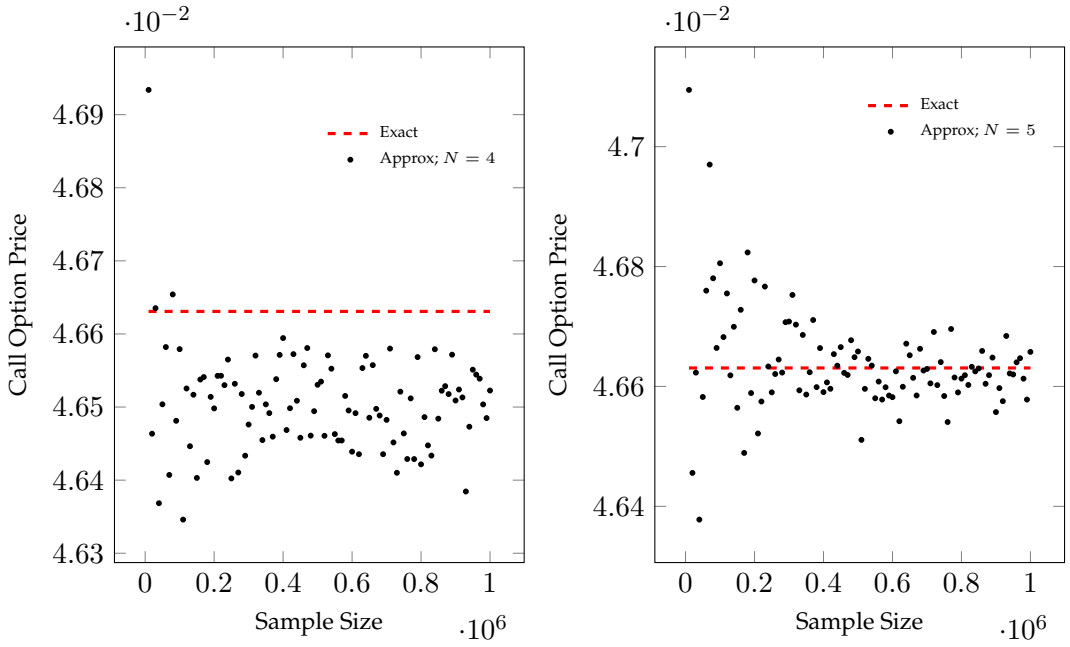


Fig. 3.17: Approximate call option prices obtained from samples generated using Chebyshev nodes in $\left[F_X^{-1}(F_S(0.08)), F_X^{-1}(0.9999)\right]$ as collocation points and utilising the put-call parity relationship.

3.5 The integrated variance process under the Heston model

Consider an asset $S(t)$ with price process described by the Heston stochastic volatility model, i.e.

$$dS(t) = rS(t)dt + \sqrt{V(t)}S(t)dW_S(t), \quad S(t_0) = S_0$$
$$dV(t) = \kappa\left(\bar{V} - V(t)\right)dt + \omega\sqrt{V(t)}dW_V(t), \quad V(t_0) = V_0.$$

When pricing path dependent options, realisations of the distribution of $S(T_2)$ conditional on $S(T_1)$ need to be generated. (Broadie and Kaya, 2006) proved that the

conditional distribution of the log asset price at time T_2, $\log(S(T_2))$ given the log asset price at time T_1, is normally distributed with

$$\mathbb{E}[\log(S(T_2))|\log(S(T_1))] = \log(S(T_1)) + \mu_S(T_1, T_2)$$

and

$$\mathbb{V}[\log(S(T_2))|\log(S(T_1))] = \sigma_S(T_1, T_2)^2$$

where

$$\begin{aligned}\mu_S(T_1, T_2) = r(T_2 - T_1) + \left[\frac{\kappa\rho}{\omega} - 0.5\right] Y(T_1, T_2) \\ + \frac{\rho}{\omega}\left[V(T_2|T_1) - V(T_1) - \kappa\bar{V}(T_2 - T_1)\right]\end{aligned}$$

and

$$\sigma_S^2(T_1, T_2) = \left(1 - \rho^2\right) Y(T_1, T_2), \quad Y(T_1, T_2) = \int_{T_1}^{T_2} V(s)ds$$

where $V(T_2|T_1)$ denotes the variance at T_2 conditional on the variance at T_1. It is clear that in order to sample from the distribution of $S(T_2)$ conditional on the value of $S(T_1)$, realisations of the random variables $V(T_1)$, $V(T_2)$ and $Y(T_1, T_2)$ are required.

The variance at time T_1 follows a distribution proportional to a non-central chi-squared distribution with

$$d = \frac{4\kappa\bar{V}}{\omega^2} \tag{3.6}$$

degrees of freedom and a non-centrality parameter of $\lambda(T_1, V_0)$ where

$$\lambda(t, V_0) = \frac{4\kappa V_0 e^{-\kappa t}}{\omega^2\left(1 - e^{-\kappa t}\right)}. \tag{3.7}$$

Specifically,

$$\begin{aligned}V(T_1) &\sim c(T_1)\chi_d'^2(\lambda(T_1, V(T_0)) \\ \text{where} \quad c(t) &= \frac{\omega^2}{4\kappa}\left(1 - e^{-\kappa t}\right).\end{aligned}$$

Similarly,

$$V(T_2) \sim c(T_2)\chi_d'^2(\lambda(T_2, V(T_0)).$$

The conditional distribution of $V(T_2)$ given the value of $V(T_1)$ is proportional to a non-central chi-squared distribution with d degrees of freedom and a non-centrality parameter of $\lambda(T_2 - T_1, V(T_1))$. Specifically,

$$V(T_2|T_1) \sim c\,(T_2 - T_1)\,\chi_d'^2(\lambda(T_2 - T_1, V(T_1))).$$

Realisations of $V(T_1)$, $V(T_2)$, and $V(T_2|T_1)$ can be obtained as explained in Section 2 and illustrated in Sections 3.2 and 3.3. The remainder of this section will focus on the generation of realisations of the integrated variance, $Y(T_1, T_2)$ given realisations of $V(T_1)$ and $V(T_2|T_1)$.

The characteristic function of the integrated variance, $Y(T_1, T_2)$, conditional on the variance at time T_1, $v(s)$, and the variance at time T_2, $v(t)$, was derived in Broadie and Kaya (2006) and is given by,

$$\begin{aligned}\phi(u) = &\frac{D(u)e^{-0.5(D(u)-\kappa)\tau}\left(1-e^{-\kappa\tau}\right)}{\kappa\left(1-e^{-D(u)\tau}\right)} \\ &\times \exp\left(\frac{v\left(s\right)+v\left(t\right)}{\omega^2}\left[\frac{\kappa\left(1+e^{-\kappa\tau}\right)}{1-e^{-\kappa\tau}} - \frac{D\left(u\right)\left(1+e^{-D(u)\tau}\right)}{1-e^{-D(u)\tau}}\right]\right) \\ &\times \frac{I_\nu\left(\sqrt{v\left(s\right)v\left(t\right)}\frac{4D(u)e^{-0.5D(u)\tau}}{\omega^2\left(1-e^{-D(u)\tau}\right)}\right)}{I_\nu\left(\sqrt{v\left(s\right)v\left(t\right)}\frac{4\kappa e^{-0.5\kappa\tau}}{\omega^2(1-e^{-\kappa\tau})}\right)}\end{aligned}$$

$$\text{where} \quad D\left(u\right) = \sqrt{\kappa^2 - 2\omega^2 iu} \quad \text{and} \quad \nu = \frac{2\kappa\theta}{\omega^2} - 1,$$

where I_ν denotes the modified Bessel function of the first kind. Unfortunately the characteristic function, ϕ, may exhibit some discontinuities. Lord and Kahl (2010) proposed evaluating the continuous algebraic equivalent of $\phi(u)$,

$$\varphi\left(u\right) = \phi\left(u\right)\frac{e^{\nu\log(z(u))}}{z\left(u\right)^\nu}$$

where

$$z\left(u\right) = \frac{D\left(u\right)e^{-0.5D(u)\tau}}{1-e^{-D(u)\tau}} \quad \text{and } f\left(u\right) = \frac{D\left(u\right)}{1-e^{-D(u)\tau}}$$

and $\log\left(z\left(u\right)\right)$ is evaluated as

$$\log\left(z\left(u\right)\right) = 0.5D\left(u\right)\tau + \log\left(f\left(u\right)\right).$$

The CDF corresponding to $\varphi(u)$ can be evaluated using the Gil-Pelaez theorem (Gil-Pelaez, 1951),

$$\mathbb{P}[Y \le x] = 1 - \left(0.5 + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-iux}\varphi_Y(u)}{iu}\right]du\right).$$

The above probability can be approximated using a finite sum,

$$\mathbb{P}[Y \leq x] \approx 1 - \left(0.5 + \frac{1}{\pi} \sum_{j=1}^{N} Re \left[\frac{e^{-iu_j x} \varphi_Y(u_j)}{iu_j} \right] \right). \tag{3.8}$$

The three-dimensional interpolating polynomial that will be used to generate realisations of $Y(T_1, T_2)$ can be constructed to coincide with the integrated variance at every possible three-tuple of collocation points, $(x_i^\star, v_j^\star, w_k^\star)$, i.e. each point on the three-dimensional tensor product grid defined by the collocation points of X, $\{x_i^\star\}$, the collocation points of $V(T_1)$, $\{v_j^\star\}$ and the collocation points of $V(T_2)$, $\{w_k^\star\}$. Alternatively, it can be constructed to coincide with every point on some sparse grid, however, this is beyond the scope of this work.

This section is concluded with the implementation of the stochastic collocation sampling method to generate realisations from the integrated variance process with the following set of parameters,

$$\omega = 0.2, \quad \kappa = 0.5, \quad \bar{V} = V_0 = 0.1, \quad T_1 = 5, \quad T_2 = 10.$$

For this particular set of parameters,

$$\begin{aligned} d = 5, \quad c(T_1) = 0.0184, \quad \lambda(T_1, V_0) = 0.4471, \\ c(T_2) = 0.0199, \quad \lambda(T_2, V_0) = 0.0339, \\ c(T_2 - T_1) = 0.0184, \quad \lambda(T_2 - T_1, V(T_1)) = \lambda(5, V(T_1)). \end{aligned}$$

For the construction of the interpolating polynomial, $N = 5$ collocation points are used for the cheap distribution, $N_1 = 2$ collocation points for the distribution of $V(T_1)$ and $N_2 = 3$ collocation points for the distribution of $V(T_2)$. These collocation points are provided in Table 3.14. The three-dimensional interpolating polynomial is constructed to coincide with the integrated variance for each point on the tensor product grid defined by the three sets of collocation points.

$\mathbf{x_i^\star}$	-2.857	-1.356	-0.000	1.356	2.857
$\mathbf{v_j^\star}$	0.065	0.214	–	–	–
$\mathbf{w_k^\star}$	0.049	0.152	0.339	–	–

Tab. 3.14: *Collocation points for sampling from the integrated variance process.*

Figure 3.18 presents the two-dimensional interpolating polynomial constructed to coincide with the quantile of $V(T_2|T_1)$ that corresponds to the values of X and $V(T_1)$ for every point on the two-dimensional tensor product grid constructed from

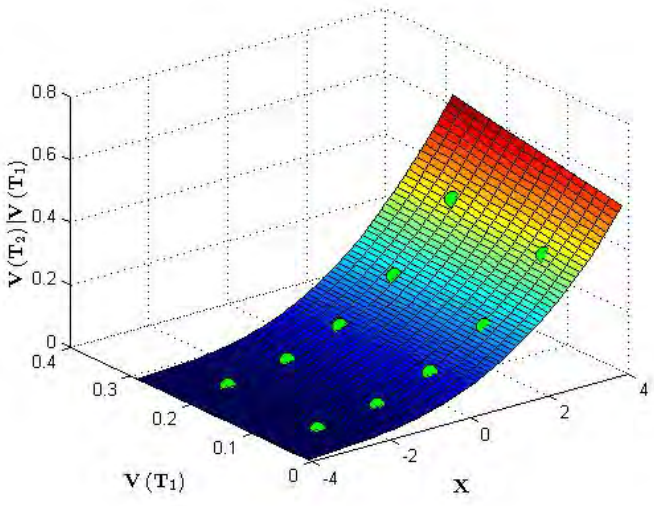


Fig. 3.18: The two-dimensional interpolating polynomial used to read off approximate realisations of $V(T_2|T_1)$ together with the two-dimensional tensor product constructed from the collocation points corresponding to X and $V(T_1)$.

the collocation points corresponding to X and $V(T_1)$. This interpolating polynomial is used to obtain approximate realisations of $V(T_2|T_1)$.

Table 3.15 contains the value of the inverse CDF corresponding to each point on the three-dimensional tensor product grid of collocation points.

$\mathbf{v_j^\star}$	$\mathbf{w_k^\star}$	$\mathbf{x_1^\star}$	$\mathbf{x_2^\star}$	$\mathbf{x_3^\star}$	$\mathbf{x_4^\star}$	$\mathbf{x_5^\star}$
$\mathbf{v_1^\star}$	$\mathbf{w_1^\star}$	0.13	0.21	0.34	0.54	0.86
$\mathbf{v_1^\star}$	$\mathbf{w_2^\star}$	0.20	0.32	0.50	0.76	1.15
$\mathbf{v_1^\star}$	$\mathbf{w_3^\star}$	0.36	0.55	0.80	1.13	1.59
$\mathbf{v_2^\star}$	$\mathbf{w_1^\star}$	0.24	0.37	0.57	0.84	1.24
$\mathbf{v_2^\star}$	$\mathbf{w_2^\star}$	0.34	0.52	0.77	1.11	1.57
$\mathbf{v_2^\star}$	$\mathbf{w_3^\star}$	0.54	0.80	1.12	1.53	2.07

Tab. 3.15: *Inverse CDF of the integrated variance process for each point on the tensor grid of collocation points.*

One thousand approximate realisations of the integrated variance, $Y(T_1, T_2)$, were generated using the stochastic collocation sampling method. Figure 3.19 presents the first 100 approximate realisations of $Y(T_1, T_2)$ together with the corresponding exact realisations. The computation time that was required to generate the 100 real-

isations of the integrated variance with the stochastic collocation sampling method was 5.51 seconds. To generate the same number of realisations using the inverse transform method required 13.8 seconds. The total time required when Matlab's built in function, *ncx2rnd*, was used to sample from the non-central chi-squared distribution, was 13.77 seconds. This computation time is very similar to that required by the inverse transform method since the bulk of the computation time for both methods was required for evaluating the CDF and then finding the inverse using a root finding method.

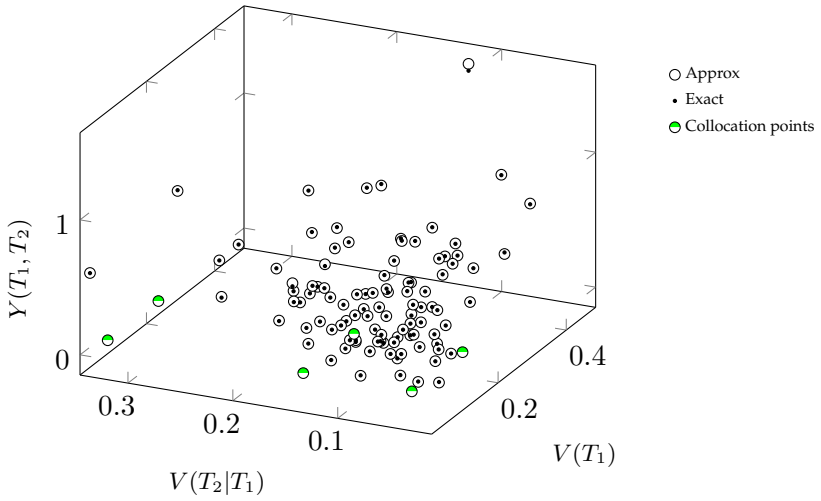


Fig. 3.19: Exact and approximate realisations of the integrated variance. The spheres represent collocation points.

Figure 3.20 contains the empirical CDF corresponding to 100000 approximate realisations of the integrated variance for $V(T_1) = 0.169$ and $V(T_2) = 0.08$ as well as the 'true' CDF obtained using the finite sum approximation in (3.8).

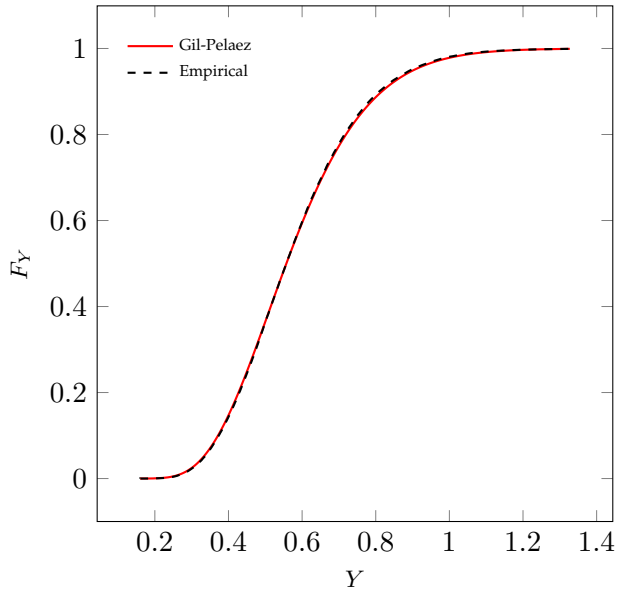


Fig. 3.20: The empirical CDF corresponding to 100000 generalisations of the integrated variance for $V(T_1) = 0.169$ and $V(T_2) = 0.08$ and the 'exact' CDF given by Gil-Pelaez.

Chapter 4

Summary and future research

This dissertation focused on the application of the stochastic collocation method to the sampling from computationally expensive distributions. The method originated in the area of uncertainty quantification (Tatang *et al.*, 1997; Webster *et al.*, 1996; Mathelin and Hussaini, 2003) and was first applied to sample from expensive distributions in a paper by Grzelak *et al.* (2015). In the context of sampling from expensive distributions the stochastic collocation method involves the transformation of realisations from a computationally cheap distribution to obtain approximate realisations of the expensive random variable, using an interpolating polynomial approximation to the true quantile-to-quantile transformation function that relates the two distributions. This dissertation builds on the work of Grzelak *et al.* (2015), in particular with regards to the type of collocation points used, the estimation of statistics using the approximate realisations obtained using the stochastic collocation sampling method, assessing the overall goodness of fit of the resulting empirical CDF and, to a lesser extent, the type of cheap distribution used.

4.1 Choice of cheap distribution

In Section 3.1 the stochastic collocation sampling method was used to generate samples from a gamma distribution using four different 'cheap' distributions, which included the standard normal distribution, a uniform distribution, a lognormal distribution with skewness and variance identical to that of the gamma distribution of interest and a different gamma distribution that is very similar in shape to the gamma distribution of interest. Results showed that using the Gaussian quadrature nodes corresponding with the standard normal distribution as collocation points resulted in the most accurate approximations of the unconditional expected value of the gamma distribution of interest for $N = 1, 2, 3, 4$, and 5. Based on the realised values of the goodness of fit test statistics, the Gaussian quadrature nodes associated with the standard normal distribution also produced the set of samples

most similar to a set of random samples from the gamma distribution of interest for $N = 4$ and $N = 5$, even though the distribution is less similar to that gamma distribution than the lognormal distribution and 'cheap' gamma distribution that were considered. Based on these results together with the time and space limits of this dissertation, it was decided to only consider the standard normal distribution as cheap distribution in the remainder of the dissertation. Further research is required to confirm whether or not the standard normal distribution is always the best, or at least always a good, choice for the cheap distribution.

4.2 The number of collocation points

Since the quantile-to-quantile transformation function is a monotonically increasing non-linear function, more than two collocation points should be used. That is, unless one is only interested in calculating the unconditional expected value of the expensive random variable, or a linear function thereof, and it is known that the quantile-to-quantile transformation function is a third order polynomial or very close to one. In that case two Gaussian quadrature points can be used since the Gaussian quadrature formula can exactly evaluate the unconditional expected value associated with any polynomial of order three or less. Unfortunately, increasing N does not guarantee improved fit of the CDF as measured by goodness-of-fit tests. However, increasing N when using Gaussian quadrature nodes will usually lead to an increase in the accuracy of the estimate of the unconditional expected value, while increasing N when using Chebyshev nodes will usually lead to a decrease in the maximum absolute interpolation error over the interval containing the Chebyshev nodes.

As the number of collocation points is increased, the collocation points tend to become more and more clustered in the tails of the distribution. This is not a problem unless numerical instability arises when computing the value of the quantile-to-quantile transformation function at collocation points lying in the very extreme tails of the distribution. For such cases, the grid stretching method proposed in Grzelak *et al.* (2015) can be used. Alternatively, the set of Chebyshev nodes constrained to lie in a closed interval with lower limit set to be the α^{th} percentile of the cheap distribution and the upper limit set to be the β^{th} percentile of the cheap distribution, can be used as collocation points. Here, α and β are close to zero and one respectively, but not so close as to result in numerical instability.

4.3 Different types of collocation points

In addition to Gaussian quadrature points, the only type of collocation points considered in Grzelak *et al.* (2015), the roots of Chebyshev polynomials of the first kind, also referred to as Chebyshev nodes, were used as collocation points in this dissertation. Gaussian quadrature points and Chebyshev nodes have different properties, which make them suited for different applications. Two important differences between the two types of collocation points are that a set of Gaussian quadrature nodes is associated with a particular distribution and cannot be constrained to a specific region of that distribution while a set of Chebyshev nodes is not associated with any particular distribution and can be constrained to lie in any closed interval on the real line. Gaussian quadrature points are particularly useful when the statistic to be estimated is the unconditional expected value of the expensive random variable, as it can be approximated by the expected value of the random variable defined by the (unadjusted) interpolating polynomial, which can be evaluated exactly using the Gaussian quadrature formula (see Section 2.2.1). The same can be said of the unconditional expected value of any linear function of the expensive random variable. Sometimes, reasonably accurate estimates of such unconditional expected values can be obtained even when the distribution as a whole is poorly approximated. This was demonstrated for the non-central chi-squared distribution considered in Section 3.3.

If the statistic to be estimated is the unconditional expected value of a non-linear function of the expensive random variable and the CDF associated with this function of the random variable is known, then the Gaussian quadrature nodes can be used to construct the interpolating polynomial that relates the cheap random variable to the random variable defined by that non-linear function. The expected value can then be calculated using the Gaussian quadrature formula. When, however, the distribution of the random variable defined by the non-linear function is not known, the expected value needs to be approximated by the corresponding sample mean calculated from a set of samples generated using the interpolating polynomial that relates the cheap random variable to the original expensive random variable.

The interpolating polynomial constructed using Chebyshev nodes constricted to a particular closed interval will converge uniformly to the quantile-to-quantile transformation function over that interval under certain continuity conditions (see Section 2.2.2). Also, for the closed interval containing the set of Chebyshev nodes, the upper bound on the absolute interpolation error is minimised by the interpolating polynomial constructed from the Chebyshev nodes. For the examples consid-

ered in Chapter 3 it was seen that the maximum absolute interpolation error over the interval typically decreases when the number of Chebyshev collocation points was increased. Chebyshev nodes have been used in combination with other sampling approaches like that proposed in Derflinger *et al.* (2009) but have not yet been used in combination with the stochastic collocation sampling method proposed in Grzelak *et al.* (2015). Using these nodes as collocation points substantially increases the usefulness of this sampling method as it will typically yield a set of samples that will produce more accurate estimates of statistics that depend on only a part of the expensive distribution, compared to the set of samples generated using Gaussian quadrature collocation points. This is especially true when the region of the cheap distribution that contributes to the value of the statistic to be estimated, contains very few Gaussian quadrature points.

Other types of collocation points, e.g. the extrema of the Chebyshev polynomials of the first kind (Ryaben'kii and Tsynkov, 2006), as well as other types of interpolation, e.g. monotone cubic Hermite polynomial interpolation, which would always yield a monotonically increasing interpolating polynomial between the smallest and largest collocation points (Grzelak *et al.*, 2015), should also be investigated.

4.4 Estimating statistics that depend on only a subset of the distribution

For some very extreme distributions, e.g. the non-central chi-squared distribution with significant mass near zero that was considered in Section 3.3, it may not be possible to generate a set of samples that is indistinguishable from a set of random samples drawn from the expensive distribution, even when a large number of collocation points is used. However, even for such extreme distributions, it may be possible to generate a set of samples that yield accurate estimates of statistics that depend on only a portion of the distribution. This can be done by using the roots of the translated Chebyshev polynomials of the first kind that has roots in the corresponding part of the cheap distribution, as collocation points. Since the interpolating polynomial approximates the quantile-to-quantile transformation function most accurately in the vicinity of collocation points, such a set of Chebyshev collocation points is expected to yield a more accurate approximation to the quantile-to-quantile transformation function for that particular part of the cheap distribution than a set of Gaussian quadrature points, which is not constrained to that region of interest. This will be especially true when the part of the cheap distribution that is of interest contains very few Gaussian quadrature points.

When a statistic that depends on only some portion of the cheap distribution

is estimated from samples generated using a set of Chebyshev nodes constrained to lie in that region, it is important to ignore all values of the interpolating polynomial that correspond to regions of the cheap distribution that are not of interest. This will prevent poor approximation of the true quantile-to-quantile transformation function in these regions from adversely affecting the accuracy of approximations. This approach can, however, only be used when it is possible to identify the regions of the cheap distribution that the statistic depends on. Similarly, for an expensive distribution with a known atom at zero, the approximate samples obtained using Chebyshev nodes (collocation points) constrained to lie in the region of the cheap distribution for which the quantile-to-quantile transformation function exists, should typically yield more accurate approximations than the approximate samples obtained using a set of Gaussian quadrature points as collocation points. For the generation of samples from a CEV process, the Chebyshev nodes approach was compared to an approach suggested by Grzelak *et al.* (2015) which entails the use of a set of Gaussian quadrature nodes as collocation points and a linearly extended version of the CDF of the expensive random variable. It makes sense that this approach should typically outperform the approach proposed by Grzelak *et al.* (2015) since the final set of collocation points are used to approximate the true quantile-to-quantile transformation function and not some adjusted version thereof.

Sometimes there may be some information regarding the regions of the cheap distribution that contribute to the value of the statistic even though it is not possible to determine exactly where those regions are. It may, for example, be known that the set of values of the cheap random variable that contributes to the statistic forms a subset of the upper 40% of the distribution. In that case the set of Chebyshev nodes in the upper 40% of the cheap distribution could be used as collocation points. On the other hand, if it is only known that the likelihood that a value in the bottom 60% of the distribution will contribute to the value of the statistic is very unlikely, then a combination of a relatively small number of Chebyshev nodes in the bottom 60% of the distribution and a relatively large number of Chebyshev nodes in the upper 40% of the distribution, could be used as collocation points.

When it is not possible to determine which part of the cheap distribution contributes to the value of a statistic to be estimated, a set of collocation points that is spread over the entire support of the cheap distribution should be used. The same is, of course, true when samples are required to approximate a statistic dependent on the entire cheap distribution. To ensure that the statistic of interest is approximated as well as possible, the collocation points need to be positioned such that no region of the cheap distribution that is associated with great probability

weight, contains zero collocation points. In particular, collocation points need to stretch far into both tails of the distribution in order to minimise the effect of any wild behaviour of the interpolating polynomial in the tails on the statistic to be estimated. The set of Gaussian quadrature nodes associated with the standard normal distribution exhibits these properties (see Section 2.2.1.8).

4.5 Computation time and sparse grids

The stochastic collocation sampling method proved to be an efficient approach to generate (approximate) realisations from the computationally expensive distributions considered in Chapter 3. The major advantage of this sampling method is the relatively short computation times required to generate approximate realisations from expensive distributions compared to methods like the inverse transform method. This is especially true when the CDF of the expensive random variable is not known analytically and needs to be determined from the characteristic function. This was illustrated in Section 3.5, where realisations from the integrated variance of an asset price with dynamics described by a Heston model, were generated. In addition to requiring shorter computation times, the method is also easy to implement.

When the stochastic collocation method is used to sample from a conditional distribution, a collocation grid is required. When the dimensionality of the collocation grid becomes large, the computation times required by the stochastic collocation method, although typically much smaller than that required by say the inverse transform method, may become impractical. This is especially true if tensor product collocation grids are used. For such high dimensional problems, the reduction in computation time resulting from using the stochastic collocation sampling method instead of the inverse transform method may be increased substantially by considering sparse grids of collocation points, which contain less points than tensor grids. These sparse grids will typically result in less accurate approximations than tensor grids, since they contain less points. The resulting reduction in computation time may, however, be more important than accuracy in a particular scenario — the key is to find the optimal trade-off between accuracy and computation time. For high dimensional problems the aim is to use the type of sparse grid that leads to the most accurate results with the smallest number of collocation points. Some sparse grids have been implemented for the stochastic collocation method in the context of solving SPDEs. In particular, the Smolyak grid has been implemented by Xiu and Hesthaven (2005), Nobile *et al.* (2008) and Pizzi (2012). The use of sparse grids in the context of sampling from expensive distributions is, however, still to

be investigated.

4.6 Conclusion

Unfortunately there is no 'one size fits all solution' for the choice of type and number of collocation points. The statistic to be estimated as well as the shape of the expensive distribution of interest substantially affect the choice of type and number of collocation points.

Relative to the inverse transform method, the stochastic collocation sampling method proved to be a very efficient sampling method for the expensive distributions that were considered in this dissertation. The use of Chebyshev nodes as collocation points proved particularly beneficial when the realisations obtained from the expensive distribution were used to approximate a statistic dependent on a known subset of the support of the cheap distribution's density function. The performance of the method still needs to be compared against that of other existing efficient sampling methods, like those proposed in Derflinger *et al.* (2009) and Olver and Townsend (2013).

A detailed investigation of the convergence properties, including the rate of convergence, of approximations to statistics calculated from samples generated using the two types of collocation points considered, will add substantial value to the work contained in this dissertation. Unfortunately time and space limitations did not allow for such investigations.

A potentially useful application of the stochastic collocation sampling method that was not investigated in this dissertation, is the approximation of computationally expensive copulas.

Bibliography

Babuška, I., Nobile, F. and Tempone, R. (2007). A stochastic collocation method for elliptic partial differential equations with random input data, *SIAM Journal on Numerical Analysis* **45**(3): 1005–1034.

Broadie, M. and Kaya, Ö. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes, *Operations Research* **54**(2): 217–231.

Derflinger, G., Hörmann, W., Leydold, J. and Sak, H. (2009). *Efficient numerical inversion for financial simulations*, Springer.

Erdős, P. and Turán, P. (1937). On interpolation I, *Annals of Math* **38**: 142–155.

Favard, J. (1935). Sur les polynomes de tchebicheff, *Comptes Rendus Hebdomadaires des Sances de lAcadmie des Sciences, Paris* **200**: 2052–2053.

Gil-Pelaez, J. (1951). Note on the inversion theorem, *Biometrika* **38**(3-4): 481–482.

Golub, G. H. and Welsch, J. H. (1969). Calculation of gauss quadrature rules, *Mathematics of computation* **23**(106): 221–230.

Grzelak, L. A., Witteveen, J., Suarez-Taboada, M. and Oosterlee, C. W. (2015). The stochastic collocation monte carlo sampler: Highly efficient sampling from'expensive'distributions, *Available at SSRN 2529691* .

Jeffreys, H. and Jeffreys, B. (1988). Weierstrasss theorem on approximation by polynomials, *Methods of Mathematical Physics* pp. 446–448.

Loeven, G., Witteveen, J. and Bijl, H. (2007). Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties, *Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit, AIAA paper*, Vol. 317.

Lord, R. and Kahl, C. (2010). Complex logarithms in heston-like models, *Mathematical Finance* **20**(4): 671–694.

Mathelin, L. and Hussaini, M. Y. (2003). A stochastic collocation algorithm for uncertainty analysis.

Navarro, M., Witteveen, J. and Blom, J. (2015). Stochastic collocation for correlated inputs, *Uncertainty quantification in computational sciences and engineering*.

Nevai, P. G. (1980). Mean convergence of Lagrange interpolation, II, *Journal of Approximation Theory* **30**(4): 263–276.

Nobile, F., Tempone, R. and Webster, C. G. (2008). A sparse grid stochastic collocation method for partial differential equations with random input data, *SIAM Journal on Numerical Analysis* **46**(5): 2309–2345.

Olver, S. and Townsend, A. (2013). Fast inverse transform sampling in one and two dimensions, *arXiv preprint arXiv:1307.1223* .

Pizzi, F. (2012). *Stochastic collocation and option pricing*, Master's thesis, Italy.

Runge, C. (1901). Über empirische funktionen und die interpolation zwischen äquidistanten ordinaten, *Zeitschrift für Mathematik und Physik* **46**(224-243): 20.

Ryaben'kii, V. S. and Tsynkov, S. V. (2006). *A Theoretical Introduction to Numerical Analysis*, Chapman and Hall/CRC.

Schroder, M. (1989). Computing the constant elasticity of variance option pricing formula, *Journal of Finance* pp. 211–219.

Smolyak, S. A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions, *Dokl. Akad. Nauk SSSR*, Vol. 4, p. 123.

Tatang, M. A. (1995). *Direct incorporation of uncertainty in chemical and environmental engineering systems*, PhD thesis, Massachusetts Institute of Technology.

Tatang, M. A., Pan, W., Prinn, R. G. and McRae, G. J. (1997). An efficient method for parametric uncertainty analysis of numerical geophysical models, *Journal of Geophysical Research: Atmospheres (1984–2012)* **102**(D18): 21925–21932.

Webster, M. D., Tatang, M. A. and McRae, G. J. (1996). Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model.

Wiener, N. (1938). The homogeneous chaos, *American Journal of Mathematics* pp. 897–936.

Xiu, D. and Hesthaven, J. S. (2005). High-order collocation methods for differential equations with random inputs, *SIAM Journal on Scientific Computing* **27**(3): 1118–1139.