RYAN G. MCCLARREN

ORTHOGONAL CHAOS

A Tutorial for Estimating Parametric Uncertainties using Polynomial Chaos Methods

ORION SCIENTIFIC PRESS

Contents

Introduction 5

Hermite Expansions for Normally-Distributed Parameters 9

Generalized Polynomial Chaos 19

Example from a PDE: Poisson Equation with an uncertain source 37

Mutli-dimensional Polynomial Chaos Expansions 41

Sparse Quadrature 49

Estimating Expansions Using Regularized Regression 59

Introduction

In computational science it is often the case that there are parameters in the underlying models that are only known up to some uncertainty. These uncertain parameters could be reaction rate coefficients (e.g., cross-sections in a particle transport calculation), constants in an equation-of-state model, the strength and spatial distribution of source terms, etc. As our models and numerical solutions to the models become more accurate, the uncertainty introduced in our calculations due to parameter uncertainty becomes more important.

The question that an engineer or scientist is now faced with is how much do these uncertainties in the inputs affect some important quantity for the system being analyzed. This is in contrast to previous eras where the simulations could only be reliable for order-of-magnitude indication or scaling relations. In other words, computational science has been successful enough to change the way simulations are consumed.

The term "Uncertainty Quantification" is used to refer to the process of determining how uncertainties in a calculation, including parameter uncertainty, model error, and numerical error, affect the result of a simulation. One way to pose the problem is to consider the solution $u(x,t)$ to a system of PDEs where the operator L is a function of some set of random variables Θ, and source term q that is also a function of Θ:

$$L(\Theta)u = q(\Theta). \tag{1}$$

Therefore, the solution $u(x,y)$ will be implicitly dependent on Θ, i.e., $u(x,t;\Theta)$. The use of the semicolon implies that the variables to the right are random variables that describe parameters that are uncertain. Therefore, we have expanded the phase space by the dimension of Θ, potentially making the problem much harder to solve. Indeed, we are now solving a stochastic PDE. Given knowledge of the distribution of the components of Θ, we could discretize and solve the enlarged problem to determine $u(x,t;\Theta)$. In general, however, this would involve different numerical methods (and, more importantly,

a different code). Additionally, the end user is not likely to be interested in $u(x,t;\Theta)$, rather some summary of the solution is how a system is often judged.

Most applications are interested in a small number of single scalar outputs in a calculation. We call these Quantities of interest (QoI). Some examples

- Will the fuel of the nuclear reactor remain less than the melting point during an accident scenario (QoI $= T_{\text{melt}}$)?
- Given uncertainty in the inputs to climate models, what is the predicted temperature rise (QoI $= \Delta T$)?
- How much of the heat shield on a space vehicle will be ablated during re-entry (QoI $= m_{\text{loss}}$)?

Most of the time a QoI is a point-value or integral of the solution $u(x,t;\Theta)$. The engineer wants to know information about the distribution of the QoI, such as the mean, median, variance, etc. This is a much easier problem to solve, and we could brute force the solution using a Monte Carlo procedure.

Going back to the original PDE system, assume that we can generate a realization (or sample) of the random variables Θ that we call Θ_i. We could then get a realization of $u(x,t;\Theta_i)$ by forming the operator L and source q as

$$L(\Theta = \Theta_i)u(x,t;\Theta_i) = q(\Theta = \Theta_i), \tag{2}$$

and then computing the QoI

$$\text{QoI}_i = \text{QoI}\left[u(x,t;\Theta_i)\right]. \tag{3}$$

If we repeat this process N times, then we will have N samples from the distribution of the quantity of interest. From these samples we can compute statistics such as the mean, variance, and empirical distribution functions. Such an approach to estimating the properties of the distribution of the QoI is called a Monte Carlo simulation.

The benefit of this approach is that solving Eq. (2) requires solving the original PDE system with the uncertain parameters N times. Therefore, if we have an existing code to solve the PDE system, we just have to run it many times. Additionally, since the realizations are independent, we can run this process in parallel easily. The downside of the Monte Carlo approach is that it converges slowly. The statistical error in a Monte Carlo estimated quantity (for instance the mean of the distribution of the QoI) decreases proportionally to $1/\sqrt{N}$. To cut the error in half, we must increase the number of

realizations by a factor of four. Given that each sample involves solving a system of PDEs, and that solution could take hours, or even weeks, to compute, we want to be as parsimonious as possible with the number of realizations (i.e., code runs) we need to compute.

An alternative approach, and the one that we will consider in some detail here, is to write the quantity of interest as an expansion in orthogonal polynomials. In particular we will pick the orthogonal polynomials so that the weighting function in the orthogonality condition "matches" the distribution of the parameters. To compute the integrals in the expansion we will use a collocation procedure and Gauss quadrature. In the process we will encounter many classic approximation techniques and have to review a host of statistics, special functions, and quadrature techniques.

Input Distribution	Orthogonal Polynomial	Support
Normal	Hermite	$(-\infty, \infty)$
Uniform	Legendre	$[a, b]$
Beta	Jacobi	$[a, b]$
Gamma	Laguerre	$[0, \infty)$

Table 1: The orthogonal polynomials and support corresponding to the different families of input random variables.

This approach is known as stochastic spectral projection, but the expansions we use are called polynomial chaos expansions. If the quantity of interest is a smooth function of the random variables, then we expect the expansion to be accurate with only a few terms. The benefit of spectral projection is, like Monte Carlo, it is an non-intrusive method: existing codes and methods can be applied out of the box. The approach does suffer from the curse of dimensionality in that the number of terms in the expansion explodes as the dimension of the random variable space increases. Later we will discuss approaches to mitigate this, using sparse grids and compressed sensing techniques.

Notation

Throughout this work we will use capital italic letters to denote a random variable, e.g., X, and lower case italics to denote a realization or single value of that random variable x. Additionally, the tilde will be used to indicate how a random variable is distributed, and calligraphic letters to denote a specific type of distribution. For example, shortly we will write $X \sim \mathcal{N}(\mu, \sigma^2)$ to indicate that the random variable X is a normal (or Gaussian) random variable with mean μ and variance σ^2.

Hermite Expansions for Normally-Distributed Parameters

We consider a continuous random variable, X, that is distributed as a normal, also called Gaussian, random variable. A real, continuous random variable is defined by its probability density function, $f(x)$, which is defined so that

$$f(x)dx = \text{The probability that the random variable } X \text{ takes a value in } dx \text{ about } x. \tag{4}$$

By this definition the following normalization is natural,

$$\int_{-\infty}^{\infty} f(x)\, dx = 1,$$

because it implies that the random variable takes on a value somewhere on the real line.

The probability density function (PDF) for a normal random variable is given by

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{5}$$

where the parameters of the distribution are

- μ, the mean of the distribution,
- σ, the standard deviation of the distribution and its square, σ^2, which is called the variance.

This PDF integrates to one, as can easily be checked.

Related to the probability distribution is the cumulative density function (CDF), $F(x)$, which is defined as

$$F(x) = \text{The probability that the random variable } X \text{ takes a value less than or equal to } x. \tag{6}$$

There are two relationships between the CDF and the PDF from basic calculus

$$F(x) = \int_{-\infty}^{x} f(x')\,dx', \qquad f(x) = \frac{dF}{dx}. \tag{7}$$

The CDF is a non-decreasing and right continuous function. It also has the property that the difference in two values of the CDF is the probability that the random variable is in a range:

$$F(b) - F(a) = \int_{a}^{b} f(x)\,dx = P(a < X \leq b).$$

The CDF for a normal random variable can be written in terms of the error function

$$F(x \mid \mu, \sigma^2) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]. \tag{8}$$

A normal random variable, with a given μ and σ^2 is often written as

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Expectation Value

An important concept with random variables is that of the expectation value. The expectation value of a function of a random variable is the integral of the function times the PDF. In particular, for a function $g(x)$ the expectation is written as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx. \tag{9}$$

There are a few special expectations that are useful to define. The mean of a random variable is the expectation of the variable itself. The mean is sometimes written as $\bar{x}$:

$$\bar{x} = E[X] = \int_{-\infty}^{\infty} x f(x)\,dx. \tag{10}$$

The variance, is the difference of the expectation of x^2 and the square of the mean,

$$\operatorname{Var}(X) = E[X^2] - E[X]^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \left(\int_{-\infty}^{\infty} x f(x)\,dx\right)^2. \tag{11}$$

For the normal distribution, we can see that μ is the mean by computing the integral

$$\bar{x} = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu.$$

Similarly, it can be shown than for a normal variable, $\text{Var}(X) = \sigma^2$.

Standard Normal Distribution

There is a special case of the normal distribution, called the standard normal. This is a normal distribution with zero mean, and unit variance, i.e., $Z \sim \mathcal{N}(0,1)$. In this case, we give the PDF a special symbol, $\phi(z)$:

$$\phi(z) \equiv f(z \mid \mu = 0, \sigma^2 = 1) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \tag{12}$$

Similarly, the CDF for the standard normal is written as $\Phi(x)$ and is defined as

$$\Phi(z) = F(z \mid \mu = 0, \sigma^2 = 1) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right]. \tag{13}$$

The standard normal is important because we can transform any normal random variable, $X \sim \mathcal{N}(\mu, \sigma^2)$, into a standard normal, $Z \sim \mathcal{N}(0,1)$, via the transform:

$$z = \frac{x - \mu}{\sigma}. \tag{14}$$

This relation shows that z is a measure of how many standard deviations from the mean a given value is. The inverse of this transform is

$$x = \mu + \sigma z. \tag{15}$$

Using these relations we can relate the expectation of a function of X to a function of Z as

$$E[g(X)] = E[g(\mu + \sigma Z)]. \tag{16}$$

We can check this in the formula for the mean

$$E[X] = E[\mu + \sigma Z] = \mu + \sigma \cancel{E[Z]}.$$

Hermite Polynomials

The Hermite polynomials, $He_n(x)$, are a set of orthogonal polynomials that form a basis for square-integrable functions on the real line with weight,

$$w(x) = e^{-x^2/2},$$

and inner product

$$\langle g(x), h(x) \rangle = \int_{-\infty}^{\infty} g(x)h(x)\, e^{-\frac{x^2}{2}}\, \mathrm{d}x,$$

i.e., the polynomials form an orthogonal basis for $L^2(\mathbb{R}, w(x)\, dx)$. The Hermite polynomials are defined as

$$He_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}. \tag{17}$$

There are two definitions of the Hermite polynomials that are scalings of each other. We use the "probabilist" version of the functions because of similarities with the standard normal distribution in the weighting function. The "physicist" version of the polynomials is slightly different and form a natural expression of the quantum harmonic oscillator.

The first few Hermite polynomials are

$$\begin{aligned}
He_0(x) &= 1, \\
He_1(x) &= x, \\
He_2(x) &= x^2 - 1, \\
He_3(x) &= x^3 - 3x, \\
He_4(x) &= x^4 - 6x^2 + 3, \\
He_5(x) &= x^5 - 10x^3 + 15x.
\end{aligned}$$

The orthogonality relation for the Hermite polynomials is

$$\int_{-\infty}^{\infty} He_m(x) He_n(x)\, e^{-\frac{x^2}{2}}\, \mathrm{d}x = \sqrt{2\pi} n! \delta_{nm}. \tag{18}$$

The expansion of a function in terms of Hermite polynomials is written as

$$g(x) = \sum_{n=0}^{\infty} c_n He_n(x), \tag{19}$$

where the expansion constants are given by

$$c_n = \frac{\langle g(x), He_n(x) \rangle}{\sqrt{2\pi} n!}. \tag{20}$$

Hermite Expansion of a function of a standard normal random variable

Consider a function $g(X)$ where $X \sim \mathcal{N}(0,1)$. The value of the function is also a random variable that we will call $G \sim g(X)$. If we

compute the zeroth order constant in the Hermite expansion of this function, we get

$$c_0 = \int_{-\infty}^{\infty} \frac{g(x)}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = E[G] = \bar{g}. \tag{21}$$

In other words, the constant c_0 in the expansion is the mean of the random variable G.

Recall that the variance of G is given by $E[G^2] - E[G]^2$, which is equal to

$$\begin{aligned} \mathrm{Var}(G) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\sum_{n=0}^{\infty} c_n He_n(x) \right)^2 e^{-\frac{x^2}{2}}\, dx - c_0^2 \\ &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} c_n^2 \langle He_n(x), He_n(x) \rangle - c_0^2 \\ &= \sum_{n=1}^{\infty} n! c_n^2. \end{aligned} \tag{22}$$

Here we have used the orthogonality of the Hermite polynomials to get the second equation, followed by the value of the integral in Eq. ((18)) to get the final result.

As an example, let us consider the function $g(X) = \cos(x)$. In this case we can directly compute the expansion coefficients:

$$c_n = \frac{1}{\sqrt{2\pi} n!} \int_{-\infty}^{\infty} \cos(x) He_n(x) e^{-x^2/2}\, dx = \begin{cases} 0 & n \text{ odd} \\ (-1)^{\frac{n}{2}} \frac{e^{-1/2}}{n!} & n \text{ even} \end{cases}. \tag{23}$$

This makes the approximation to the function

$$\cos(X) = e^{-\frac{1}{2}} \sum_{n \text{ even}} (-1)^{\frac{n}{2}} \frac{He_n(x)}{n!}, \qquad X \sim \mathcal{N}(0,1). \tag{24}$$

This implies that the mean of $g(x)$ is $e^{-1/2}$ and that the variance is

$$\mathrm{Var}(G) = e^{-1} \sum_{n \text{ even}, n>1} \frac{1}{n!} = e^{-1}(\cosh(1) - 1) \approx 0.19978820.$$

We can get a baseline for comparison between the expansion and the actual distribution of G by sampling a value for X from a standard normal and then evaluating $g(x)$. The resulting distribution is a Monte Carlo approximation to the true distribution of G. We then can compare that to the values obtained by sampling X and then evaluating the expansion in Eq. (24) with different orders of expansion. These results are shown in Figure 1.

In these results we see that improvement obtained as we go to higher order expansions. The zeroth-order expansion only gives a

For a function that is expensive to evaluate, we may not be able to estimate the distribution using Monte Carlo. Nevertheless, sampling X and then evaluating a polynomial of X is basically free.

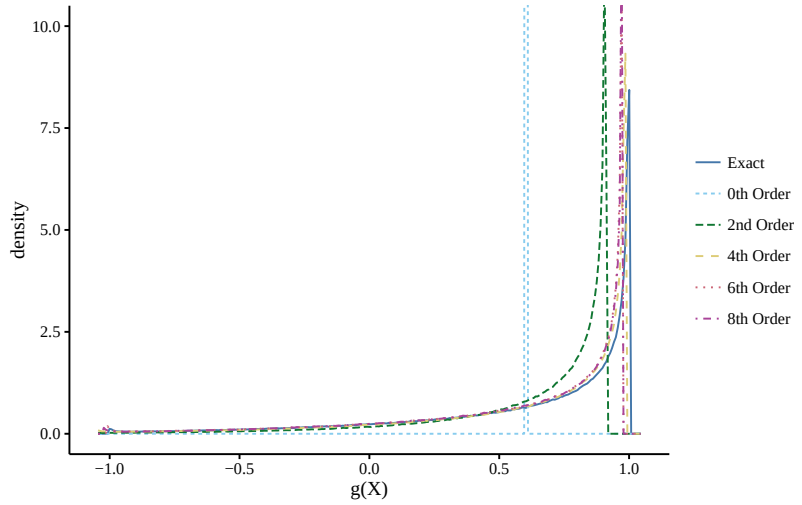


Figure 1: PDF of the random variable $g(X) = \cos(x)$, where $X \sim \mathcal{N}(0, 1)$, and various approximations. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

value of the mean, and there is a large improvement in going to the second-order expansion. There is a noticeable difference between the fourth- and second-order expansions, though beyond that, there is little difference on in the figure. We can track improvement in the higher-order expansions by looking at the convergence of the variance. In Table 2 we show that adding more terms to the expansion does improve the estimate of the variance, though modestly beyond the second-order expansion.

order	variance
0	0
2	0.183939721
4	0.199268031
6	0.199778974
8	0.199788098
∞	0.199788200

Table 2: The convergence of $\mathrm{Var}(G)$ for $g(X) = \cos(x)$, where $X \sim \mathcal{N}(0, 1)$.

Hermite Expansion of a function of a general normal random variable

If the random variable is normal, but not standard normal, then we need to change the procedure a bit. Let's say that $g(X)$ is a function of the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. In this case we will change variables to express the function as $g(Z)$ where Z and X are related by Eq. (15). Therefore, in this case

$$c_n = \frac{\langle g(\mu + \sigma z), He_n(z) \rangle}{\sqrt{2\pi} n!}. \tag{25}$$

The bounds of the inner product's integration are not affected because they are infinite, this may not be the case when we have different random variables.

Going back to our example from before where $g(X) = \cos(x)$, we now say that $X \sim \mathcal{N}(\mu = 0.5, \sigma^2 = 4)$. Performing the integrals for the coefficients in Eq. (25) gives the following expansion, to fifth

order,

$$\cos(X) \approx e^{-2}\left(1 - 2He_2(z) + \frac{2}{3}He_4(z)\right)\cos\left(\frac{1}{2}\right) + e^{-2}\left(2He_1(z) + \frac{4}{3}He_3(z) - \frac{4}{15}He_5(z)\right)\sin\left(\frac{1}{2}\right) \quad (26)$$

The mean is

$$\bar{g} = e^{-2}\cos\left(\frac{1}{2}\right) \approx 0.1187678845769458,$$

and the variance is

$$\mathrm{Var}(G) = \frac{\left(e^4 - 1\right)\left(e^4 - \cos(1)\right)}{2e^8} \approx 0.48598481520881144144.$$

The distributions produced by various approximations to G are shown in Figure 2. The exact distribution is more difficult to capture with a polynomial expansion, partly because of the non-smoothness in the solution at ± 1. By the sixth-order expansion, the overall shape of the distribution is correct, though the peaks are not in the correct place. Note that in all of these curves the mean is the same. Also, notice that even though the minimum value of $g(X)$ is -1, the expansion can give a nonzero probability of getting a value less than -1.

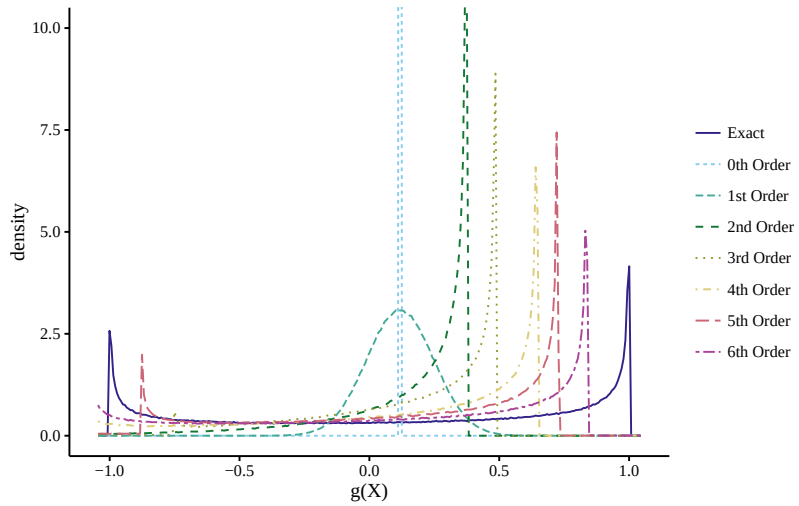


Figure 2: PDF of the random variable $g(X) = \cos(x)$, where $X \sim \mathcal{N}(\mu = 0.5, \sigma^2 = 4)$, and various approximations. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

In this example the variance also takes longer to converge. In Table 3, we see that even the sixth-order expansion only has 1 digit correct.

order	variance
0	0
1	0.016807404
2	0.128990805
3	0.173419006
4	0.329091747
5	0.380416942
6	0.458346473
∞	0.485984815

Table 3: The convergence of $\mathrm{Var}(G)$ for $g(X) = \cos(x)$, where $X \sim \mathcal{N}(\mu = 0.5, \sigma^2 = 4)$.

Gauss-Hermite Quadrature

Recall that our ultimate goal is to use polynomial expansions to provide information about the distribution of output quantities from

a computer simulation. To that end we will need to estimate the coefficients in the Hermite expansion. If we use a quadrature rule to estimate the integrals in these coefficients, then we would like a quadrature rule to require as few evaluations of the integrand as possible, because each evaluation requires running a new simulation at a different point in input space.

The most common way to approximate the required integrals is to use Gauss-Hermite quadrature, which is a Gauss quadrature rule for computing integrals of the form

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}\,dx \approx \sum_{i=1}^{n} w_i f(x_i), \tag{27}$$

where the abscissas, x_i, are given by the n roots of $He_n(x)$, and the weights are given by

$$w_i = \frac{\sqrt{\pi}n!}{n^2\left(He_{n-1}\left(\sqrt{2}x_i\right)\right)^2}. \tag{28}$$

The values for the weights and abscissas are given up to $n = 6$ are given in Table 4. Note that the points are symmetric about the origin, so we only give the magnitude of the abscissas.

n	$\lvert x_i\rvert$	w_i
1	0	$\sqrt{\pi}$
2	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}\sqrt{\pi}$
3	0	$\frac{2}{3}\sqrt{\pi}$
	$\frac{1}{2}\sqrt{6}$	$\frac{1}{6}\sqrt{\pi}$
4	0.524647623275290	0.804914090005514
	1.65060123885785	0.0813552017779922
	0	0.945308720482942
5	0.958572464613819	0.3936193231522404
	2.020182704560 86	0.019953268807 48209
	0.436077411927617	0.7246295952243919
6	1.335849074013697	0.1570673203228565
	2.350604973674492	0.004530009905508858

Table 4: The abscissas and weights for Gauss-Hermite quadrature up to order 6.

This quadrature set has the standard features of Gauss quadrature. The rule will be exact when $f(x)$ is a polynomial of degree $2n-1$ or less.

There is a slight issue in Gauss-Hermite quadrature in that it uses the a weight function of $\exp\left(-x^2\right)$, rather than $\exp\left(-x^2/2\right)$ we used in our inner product definition. Therefore, we need to make the change of variable $x \to x'/\sqrt{2}$. This makes the approximation to the inner product

$$\langle g(x), He_m(x) \rangle \approx \sqrt{2} \sum_{i=1}^{n} w_i g(\sqrt{2} x_i) \tag{29}$$

We can use our previous example, of $g(X) = \cos(x)$, where $X \sim \mathcal{N}(\mu = 0.5, \sigma^2 = 4)$, as a test of estimating the inner-products using Gauss-Hermite quadrature rules. In Figure 3, the distribution, as approximated by a fifth-order Hermite expansion, is computed using Gauss-Hermite quadratures of different values of n. For this expansion, we need at least 8 quadrature points to get an accurate estimate of the coefficients. We can see the convergence in the coefficients with the number of quadrature points in Table 5. Here we see that to estimate the mean, c_0, with two-digits of accuracy we need $n = 6$, whereas the c_5 term needs $n = 9$ to get that many digits of accuracy.

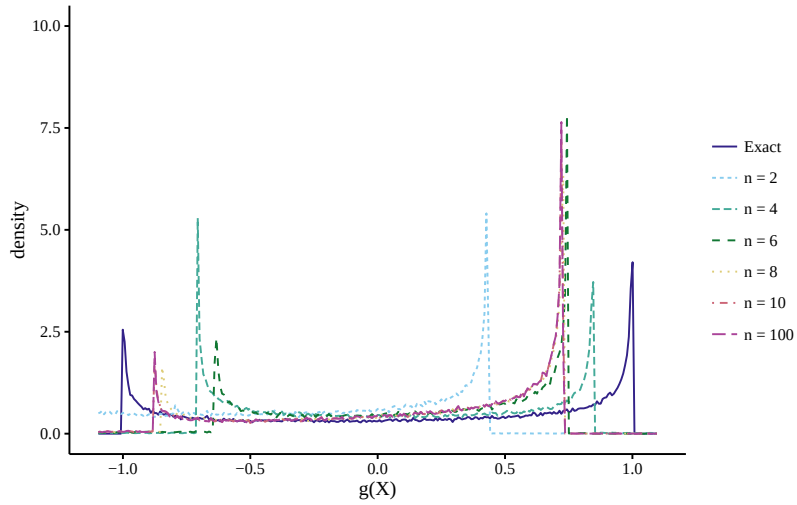


Figure 3: PDF of the random variable $g(X) = \cos(x)$, where $X \sim \mathcal{N}(\mu = 0.5, \sigma^2 = 4)$ using a fifth-order Hermite expansion with various Gauss-Hermite quadrature rules to approximate the coefficients. This figure was generated from 10^5 samples of X that were used to evaluate $g(X)$ and the various approximations.

Exercises

1. A beam of radiation that strikes a slab of material will have the intensity decreased by a factor $t = \exp(-kx)$ where x is the thickness of the slab and k is the extinction coefficient, sometimes called a macroscopic cross-section. If $K \sim \mathcal{N}(\mu = 5, \sigma^2 = 1)$ and $x = 1$. Compute the mean and variance of the $t(K)$. Plot the distribution of $t(K)$ as well.

n	c_0	c_1	c_2	c_3	c_4	c_5
2	-0.365203	-0.435940	-0.000000	0.145313	0.030434	-0.021797
3	0.307609	0.087730	-0.569973	-0.000000	0.142493	-0.004386
4	0.065646	-0.219271	-0.023343	0.173281	0.000000	-0.034656
5	0.130446	-0.103803	-0.322800	0.037629	0.141446	0.000000
6	0.116662	-0.135589	-0.213171	0.104748	0.048382	-0.028531
7	0.119090	-0.128702	-0.242956	0.081489	0.089843	-0.012370
8	0.118725	-0.129931	-0.236549	0.087602	0.076377	-0.018886
9	0.118773	-0.129744	-0.237688	0.086315	0.079768	-0.016907
10	0.118767	-0.129769	-0.237515	0.086541	0.079075	-0.017382
100	0.118768	-0.129766	-0.237536	0.086511	0.079179	-0.017302

Table 5: The convergence of the first six coefficients in the Hermite polynomial expansion $g(X) = \cos(x)$, where $X \sim \mathcal{N}(\mu = 0.5, \sigma^2 = 4)$ as estimated by different Gauss-Hermite quadrature rules.

2. Repeat the exercise using $K \sim \mathcal{N}(\mu = 2, \sigma^2 = 1)$.
3. Consider a stochastic medium where the distribution of thicknesses of two different materials is unknown. In this case the beam transmission will be given by

$$t = \exp(-k_1 x_1 - k_2(x - x_1)).$$

If $k_1 = 5$, $k_2 = 0.2$, $x = 1$ and $X_1 \sim \mathcal{N}(\mu = 0.5, \sigma^2 = 0.1)$, compute the mean and variance of the $t(x_1)$ and plot the distribution. Is there a value of $\bar{k}$ that you can define so that

$$\exp(-\bar{k}x) = E[\exp(-k_1 x_1 - k_2(x - x_1)]?$$

Generalized Polynomial Chaos

When the input parameter is not normally distributed, we need a different polynomial expansion to approximate the mapping from input parameter to output random variable. We will cover three such cases, as enumerated in Table 1. First, we tackle uniform random variables.

Uniform Random Variables: Legendre Polynomials

Consider a random variable X that is uniformly distributed in the range $[a, b]$. In this case we write $X \sim \mathcal{U}[a, b]$. Additionally, the PDF of X is

$\mathcal{U}(a, b)$ denotes a uniform distribution between a and b.

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}. \tag{30}$$

The mean of a uniform distribution is $(b - a)/2$ and the variance is $(b - a)^2/12$.

As with normal random variables, it is useful to convert general uniform random variables to a standardized random variable. In this case, we map the interval $[a, b]$ to $[-1, 1]$ to correspond with the support with the standard definition of Legendre polynomials. In particular, if $Z \sim \mathcal{U}[-1, 1]$, then

For statisticians it is more common to think of a standard uniform random variable as having the range $[0, 1]$. However, defining the standard to be symmetric about the origin makes for easier algebra down the road. This will also be the case with beta-distributed random variables later.

$$x = \frac{b-a}{2} z + \frac{a+b}{2}, \tag{31}$$

and

$$z = \frac{a + b - 2x}{a - b}. \tag{32}$$

Therefore, the expectation operator on a uniform random variable transforms to

$$E[g(X)] = \frac{1}{b-a} \int_a^b g(x)\, dx = \frac{1}{2} \int_{-1}^{1} g\left(\frac{b-a}{2} z + \frac{a+b}{2} \right) dz. \tag{33}$$

For a function on the range $[-1, 1]$ the Legendre polynomials form an orthogonal basis. The Legendre polynomials are defined as

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left[(x^2 - 1)^n \right]. \tag{34}$$

The first ten of these polynomials are given in Table 6.

Table 6: The first ten Legendre polynomials.

n	$P_n(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2 - 1)$
3	$\frac{1}{2}(5x^3 - 3x)$
4	$\frac{1}{8}(35x^4 - 30x^2 + 3)$
5	$\frac{1}{8}(63x^5 - 70x^3 + 15x)$
6	$\frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$
7	$\frac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x)$
8	$\frac{1}{128}(6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35)$
9	$\frac{1}{128}(12155x^9 - 25740x^7 + 18018x^5 - 4620x^3 + 315x)$
10	$\frac{1}{256}(46189x^{10} - 109395x^8 + 90090x^6 - 30030x^4 + 3465x^2 - 63)$

The orthogonality relation for Legendre polynomials is written as

$$\int_{-1}^{1} P_n(x) P_{n'}(x)\, dx = \frac{2}{2n+1} \delta_{nn'}. \tag{35}$$

The expansion of a square-integrable function on the interval $[a, b]$ in Legendre polynomials is then

$$g(x) = \sum_{n=0}^{\infty} c_n P_n \left(\frac{a + b - 2x}{a - b} \right), \qquad x \in [a, b], \tag{36}$$

where c_n is defined by

$$c_n = \frac{2n+1}{2} \int_{-1}^{1} g\left(\frac{b-a}{2} z + \frac{a+b}{2} \right) P_n(z)\, dz. \tag{37}$$

As before, c_0 will be the mean of the random variable $G \sim g(X)$:

$$\begin{aligned} c_0 &= \frac{1}{2} \int_{-1}^{1} g\left(\frac{b-a}{2} z + \frac{a+b}{2} \right) dz \\ &= \frac{1}{b-a} \int_{a}^{b} g(x)\, dx \\ &= E[G]. \end{aligned} \tag{38}$$

Additionally, the variance of the G is equivalent to the sum of the squares of the coefficients with $n \geq 1$:

$$\begin{aligned}\mathrm{Var}(G) &= \frac{1}{2}\int_{-1}^{1}\left(\sum_{n=0}^{\infty} c_n P_n(z)\right)^2 dz - c_0^2 \\ &= \sum_{n=1}^{\infty}\frac{c_n^2}{2n+1}.\end{aligned} \tag{39}$$

As a demonstration of the Legendre expansion, we will once again turn to the function $g(X) = \cos(x)$. This time, however, $X \sim \mathcal{U}(0, 2\pi)$. In this case we get

$$c_n = \frac{2n+1}{2}\int_{-1}^{1}\cos(\pi z + \pi)\, P_n(z)\, dz = -\frac{2n+1}{2}\int_{-1}^{1}\cos(\pi z)\, P_n(z)\, dz. \tag{40}$$

This makes the expansion, through sixth-order

$$\cos(X) \approx \frac{15}{\pi^2}P_2(z) + \frac{45\left(4\pi^2-42\right)}{2\pi^4}P_4(z) + \frac{273\left(7920-960\pi^2+16\pi^4\right)}{16\pi^6}P_6(z) \quad X \sim \mathcal{U}(0,2\pi), \tag{41}$$

and z is related to x via Eq. (32). The variance of this function is given by

$$\mathrm{Var}(G) = \frac{1}{2\pi}\int_{0}^{2\pi}\cos^2(x)\, dx = \frac{1}{2}. \tag{42}$$

The convergence of the variance estimate is given in Table 7.

order	variance
0	0
2	0.461969
4	0.499663
6	0.499999
8	0.500000
∞	0.500000

Table 7: The convergence of $\mathrm{Var}(G)$ for $g(X) = \cos(x)$, where $X \sim \mathcal{U}(0, 2\pi)$.

The convergence of the approximation to G as a function of the order of the Legendre expansion is shown in Figure 4. In this case, the approximation converges rather quickly: the eighth-order expansion is indistinguishable from the exact distribution.

Gauss-Legendre Quadrature

For estimating the coefficients in a Legendre expansion, Gauss-Legendre quadrature is a natural choice. Gauss-Legendre quadrature

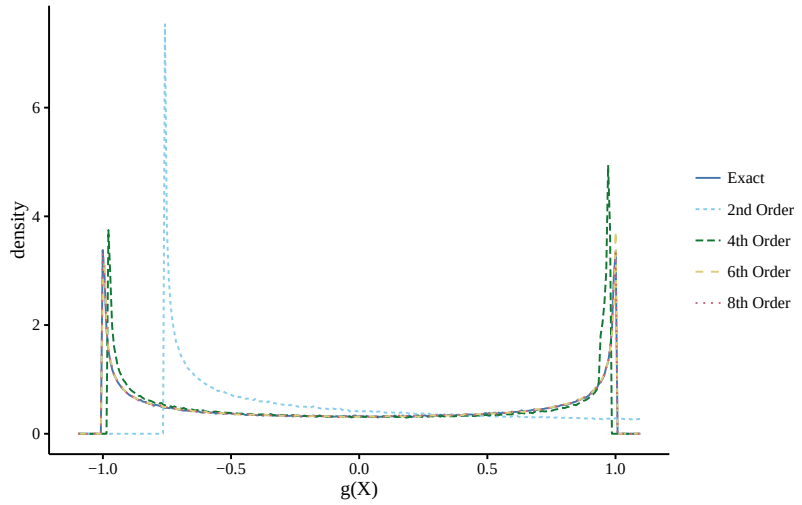


Figure 4: PDF of the random variable $g(X) = \cos(x)$, where $X \sim \mathcal{U}(0, 2\pi)$, and various approximations. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

approximately integrates functions on the range $[-1, 1]$ as

$$\int_{-1}^{1} f(z)\, dz \approx \sum_{i=1}^{n} w_i f(z_i), \tag{43}$$

where the z_i are the roots of P_n, and the weights are given by

$$w_i = \frac{2}{\left(1 - z_i^2\right)\left[P_n'(z_i)\right]^2}. \tag{44}$$

Gauss-Legendre quadrature integrates polynomials of degree $2n - 1$ exactly.

We can use our previous example, of $g(X) = \cos(x)$, where $X \sim \mathcal{U}(0, 2\pi)$, as a test of estimating the inner-products using Gauss-Legendre quadrature rules. In Figure 5, the distribution, as approximated by a fifth-order Legendre expansion, is computed using Gauss-Legendre quadratures of different values of n. We can see the convergence in the coefficients with the number of quadrature points in Table 9. Here we see that to estimate the mean, c_0, with two-digits of accuracy we need $n = 4$, whereas the c_4 term needs $n = 7$ to get that many digits of accuracy.

Beta Random Variables: Jacobi Polynomials

A random variable that takes on a value in the range, $[-1, 1]$, can often be described by a beta distribution[1]. A random variable Z that is beta-distributed is written as $Z \sim \mathcal{B}(\alpha, \beta)$, where $\alpha > -1$ and

[1] The definition of the beta distribution used here is not the typical statistician's distribution. That distribution has support on $[0, 1]$ and uses parameters α' and β' that are equal to $\alpha' = \alpha + 1$, and $\beta' = \beta + 1$. As we will see the definition in Eq. (45) is well-suited to expansion in Jacobi polynomials.

n	$\lvert x_i \rvert$	w_i
1	0	2
2	$\frac{1}{\sqrt{3}}$	1
3	0	$\frac{8}{9}$
	$\sqrt{\frac{3}{5}}$	$\frac{5}{9}$
4	0.3399810436	0.652145155
	0.8611363116	0.347854845
5	0	0.568888889
	0.5384693101	0.47862867
	0.9061798459	0.2369268851
6	0.2386191860	0.467913935
	0.6612093865	0.360761573
	0.9324695142	0.171324492

Table 8: The abscissas and weights for Gauss-Legendre quadrature up to order 6.

n	c_0	c_1	c_2	c_3	c_4	c_5
2	0.240619	0.000000	0.000000	0.000000	-0.842165	0.000000
3	-0.022454	0.000000	1.955092	0.000000	-2.639374	0.000000
4	0.001068	0.000000	1.478399	0.000000	-0.000000	0.000000
5	-0.000031	0.000000	1.521801	0.000000	-0.637516	0.000000
6	0.000001	0.000000	1.519760	0.000000	-0.579819	0.000000
7	0.000000	0.000000	1.519819	0.000000	-0.582523	0.000000
8	0.000000	0.000000	1.519818	0.000000	-0.582445	0.000000
9	0.000000	0.000000	1.519818	0.000000	-0.582447	0.000000
10	0.000000	0.000000	1.519818	0.000000	-0.582447	0.000000
100	0.000000	0.000000	1.519818	0.000000	-0.582447	0.000000

Table 9: The convergence of the first six coefficients in the Legendre polynomial expansion $g(X) = \cos(x)$, where $X \sim \mathcal{U}(0, 2\pi)$ as estimated by Gauss-Legendre quadrature rules using different values of n.

$\beta > -1$ are parameters. The PDF for Z is given by

$$f(z) = \frac{2^{-(\alpha+\beta+1)}}{\alpha+\beta+1}\frac{\Gamma(\alpha+1)+\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)}(1+z)^{\beta}(1-z)^{\alpha} \qquad z \in [-1,1]. \tag{45}$$

The reason that is sometimes called a beta distribution is that the PDF can be expressed in terms of the beta function, $\mathrm{B}(\alpha, \beta)$

$$\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \tag{46}$$

as

$$f(z) = \frac{2^{-(\alpha+\beta+1)}}{\mathrm{B}(\alpha+1,\beta+1)}(1+z)^{\beta}(1-z)^{\alpha} \qquad z \in [-1,1]. \tag{47}$$

There is some subtlety regarding the support of z. If α or β is less than 0 then one or both of the endpoints is excluded due to

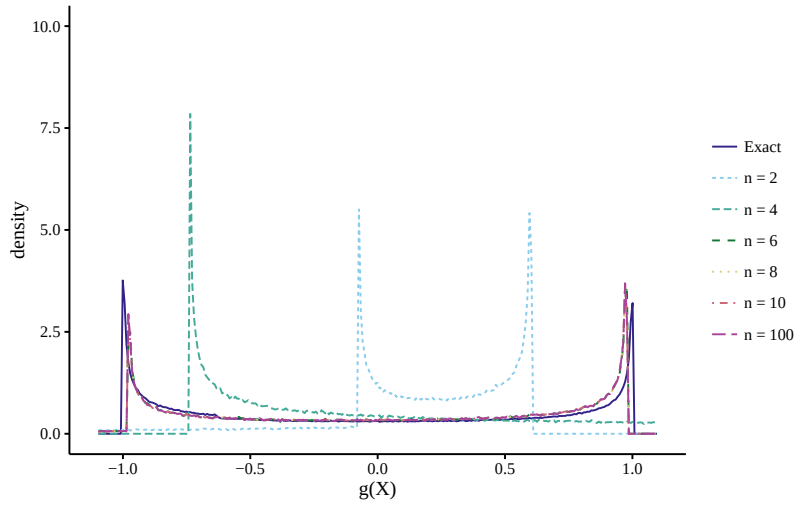


Figure 5: PDF of the random variable $g(X) = \cos(x)$, where $X \sim \mathcal{U}(0, 2\pi)$ using a fifth-order Legendre expansion with various Gauss-Legendre quadrature rules to approximate the coefficients. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

a singularity. The PDF for various values of α and β are shown in Figure 6.

As before, we can scale the distribution to a general range $X \in [a, b]$ using Eqs. (31) and (32). The expectation operator in this case is given by

$$E[g(X)] = \int_{-1}^{1} g\left(\frac{b-a}{2}z + \frac{a+b}{2}\right) \frac{2^{-(\alpha+\beta+1)}(1+z)^{\beta}(1-z)^{\alpha}}{\mathrm{B}(\alpha+1, \beta+1)}\, dz. \quad (48)$$

From this we get following for a beta distribution on the range $[a, b]$:

$$\bar{x} = \frac{(\alpha+1)a + (\beta+1)b}{\alpha+\beta+2}, \qquad \mathrm{Var}(X) = \frac{(\alpha+1)(\beta+1)(a-b)^2}{(\alpha+\beta+2)^2(\alpha+\beta+3)}. \quad (49)$$

The Jacobi polynomials, $P_n^{(\alpha,\beta)}(z)$ are orthogonal polynomials under the weight $(1-z)^{\alpha}(1+z)^{\beta}$ for the interval $z \in [-1, 1]$. These polynomials can be defined in several ways, including Rodrigues' formula:

$$P_n^{(\alpha,\beta)}(z) = \frac{(-1)^n}{2^n n!}(1-z)^{-\alpha}(1+z)^{-\beta}\frac{d^n}{dz^n}\left\{(1-z)^{\alpha}(1+z)^{\beta}\left(1-z^2\right)^n\right\}. \quad (50)$$

The general form of these polynomials is given up to order 3 in Table 10. Note that when $\alpha = \beta = 0$ these polynomials are the Legendre polynomials.

These polynomials have the, somewhat ugly, orthogonality relation

$$\langle P_m^{(\alpha,\beta)}(z) P_n^{(\alpha,\beta)}(z)\rangle = \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)n!}\delta_{nm}, \qquad \alpha,\ \beta > -1. \quad (51)$$

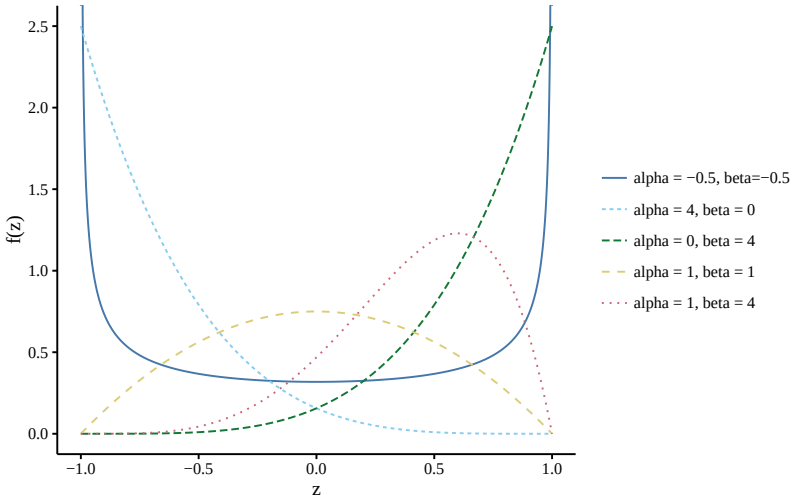


Figure 6: PDF $Z \sim \mathcal{B}(\alpha, \beta)$ for several values of α and β. Note that when $\alpha = \beta$ the distribution is symmetric about $\frac{1}{2}$, and swapping α and β creates mirror images.

where

$$\langle g(z), h(z) \rangle = \int_{-1}^{1} (1-z)^{\alpha}(1+z)^{\beta} g(z) h(z)\, dz. \tag{52}$$

Note that if $n = 0$ then we can use the identity $\Gamma(z+1) = z\Gamma(z)$ to get the normalization constant used in the PDF for the beta distribution:

$$\frac{2^{\alpha+\beta+1}}{\alpha+\beta+1} \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+1)} = 2^{\alpha+\beta+1} \mathrm{B}(\alpha+1, \beta+1).$$

Table 10: The first three Jacobi polynomials.

n	$P_n^{(\alpha,\beta)}(z)$
0	1
1	$\frac{1}{2}(\alpha - \beta + z(\alpha+\beta+2))$
2	$\frac{1}{2}(\alpha+1)(\alpha+2) + \frac{1}{8}(z-1)^2(\alpha+\beta+3)(\alpha+\beta+4) + \frac{1}{2}(z-1)(\alpha+2)(\alpha+\beta+3)$
3	$\frac{1}{6}(\alpha+1)(\alpha+2)(\alpha+3) + \frac{1}{48}(z-1)^3(\alpha+\beta+4)(\alpha+\beta+5)(\alpha+\beta+6) +$ $+\frac{1}{8}(z-1)^2(\alpha+3)(\alpha+\beta+4)(\alpha+\beta+5) + \frac{1}{4}(z-1)(\alpha+2)(\alpha+3)(\alpha+\beta+4)$

A function that is square-integrable with respect to the inner product in Eq. (52) can be written as

$$g(x) = \sum_{n=0}^{\infty} c_n P_n^{(\alpha,\beta)}\left(\frac{a+b-2x}{a-b}\right), \qquad x \in [a, b], \tag{53}$$

where the constant is defined as

$$c_n = \langle P_n^{(\alpha,\beta)}(z) P_n^{(\alpha,\beta)}(z) \rangle^{-1} \int_{-1}^{1} g\left(\frac{b-a}{2}z + \frac{a+b}{2}\right) P_n^{(\alpha,\beta)}(z)(1-z)^{\alpha}(1+z)^{\beta}\, dz. \tag{54}$$

It is worthwhile to look at c_0. This formula gives us that, as before c_0 is the mean (expected value) of $G \sim g(X)$:

$$c_0 = \frac{2^{-(\alpha+\beta+1)}}{\mathrm{B}(\alpha+1,\beta+1)} \int_{-1}^{1} g\left(\frac{b-a}{2}z + \frac{a+b}{2}\right)(1-z)^{\alpha}(1+z)^{\beta}\,dz = E[g(X)]. \tag{55}$$

Also, by construction the variance in $g(X)$ is the sum of the squares of the c_n for $n > 0$:

$$\mathrm{Var}(G) = E[g^2(X)] - (E[g(X)])^2 = \frac{2^{-(\alpha+\beta+1)}}{\mathrm{B}(\alpha+1,\beta+1)} \sum_{n=1}^{\infty} c_n^2 \langle P_n^{(\alpha,\beta)}(z) P_n^{(\alpha,\beta)}(z)\rangle. \tag{56}$$

As a test of this expansion, we will consider $g(X) = \cos(x)$, where $X \in [0, 2\pi]$ and X is derived from a standard beta random variable $Z \sim \mathcal{B}(4,1)$. The density plot from 10^6 samples of this distribution is shown in Fig. 7.

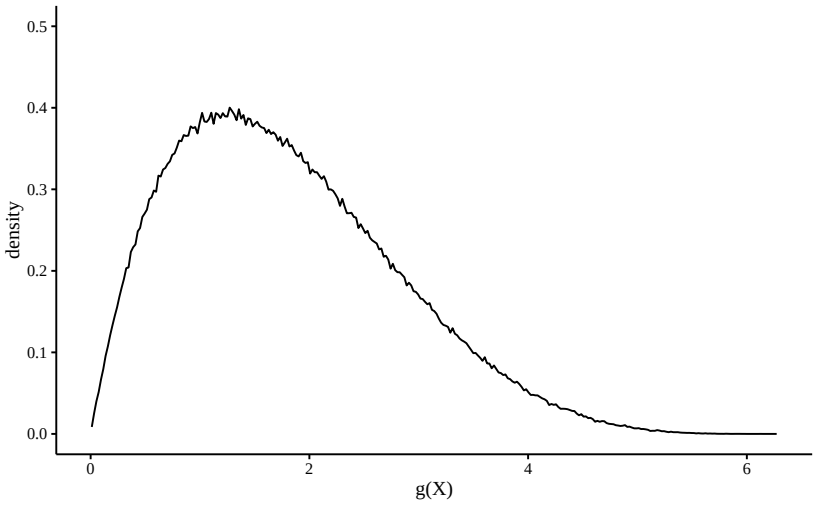


Figure 7: Density plot of 10^6 samples of $\pi z + \pi$ where $Z \sim \mathcal{B}(4,1)$. These samples were used to generate the results in Figure 8.

In this case we get

$$c_n = \langle P_n^{(\alpha,\beta)}(z) P_n^{(\alpha,\beta)}(z)\rangle^{-1} \int_{-1}^{1} \cos(\pi z + \pi)\, P_n^{(4,1)}(z)\,dz. \tag{57}$$

There is not a tidy formula for the coefficients, but we can calculate them (with the help of Mathematica). The mean value of $G \sim \cos(X)$ is

$$c_0 = -\frac{15\left(\pi^2 - 9\right)}{2\pi^4} \approx -0.0669551. \tag{58}$$

The expansion, through third-order is

$$\cos(X) \approx -\frac{15\left(\pi^2-9\right)}{2\pi^4} + \frac{6\left(315-60\pi^2+2\pi^4\right)}{\pi^6} P_1^{(4,1)}(z) - \frac{35\left(630-75\pi^2+\pi^4\right)}{2\pi^6} P_2^{(4,1)}(z)$$
$$+ \frac{12\left(-51975+8190\pi^2-315\pi^4+2\pi^6\right)}{\pi^8} P_3^{(4,1)}(z) \quad Z \sim \mathcal{B}(4,1), \quad (59)$$

and z is related to x via $x = \pi z + \pi$. For completeness, the values of the Jacobi polynomials in this equation are given in Table 11.

If we expand the definitions of the Jacobi polynomials and make

Table 11: The Jacobi polynomials $P_n^{(4,1)}$ through order 3.

n	$P_n^{(4,1)}(z)$
0	1
1	$\frac{1}{2}(7z+3)$
2	$9(z-1)^2+24(z-1)+15$
3	$\frac{165}{8}(z-1)^3+\frac{315}{4}(z-1)^2+\frac{189(z-1)}{2}+35$

numerical approximations of the coefficients we can get

$$\cos(X) \approx 2.50342z^3 + 4.14706z^2 - 0.536325z - 1.00484 \quad Z \sim \mathcal{B}(4,1).$$

The variance of G is given by

$$\mathrm{Var}(G) = \frac{2^{-(\alpha+\beta+1)}}{\mathrm{B}(\alpha+1,\beta+1)} \int_{-1}^{1} \cos^2(\pi z+\pi)(1-z)^\alpha(1+z)^\beta\, dz - \left(\frac{15\left(\pi^2-9\right)}{2\pi^4}\right)^2 \quad (60)$$
$$= \frac{1}{64}\left(\frac{135}{\pi^4}+32-\frac{60}{\pi^2}\right) - \frac{225\left(\pi^2-9\right)^2}{4\pi^8} \approx 0.4221832.$$

The convergence of the variance estimate is given in Table 12. Notice that at fourth-order the estimate is correct to three digits.

Table 12: The convergence of $\mathrm{Var}(G)$ for $g(X) = \cos(x)$, where $x = \pi z + \pi$ and $Z \sim \mathcal{B}(4,1)$.

order	variance
1	0.3302376
2	0.4001581
4	0.4220198
6	0.4221829
8	0.4221832
∞	0.4221832

The convergence of the approximation to G as a function of the order of the Jacobi expansion is shown in Figure 8. The "exact" distribution is determined by evaluating $g(x)$ at the 10^6 points shown in Figure 7. By the fourth-order expansion the overall character of the true distribution is captured. The eighth-order expansion is indistinguishable from the exact distribution.

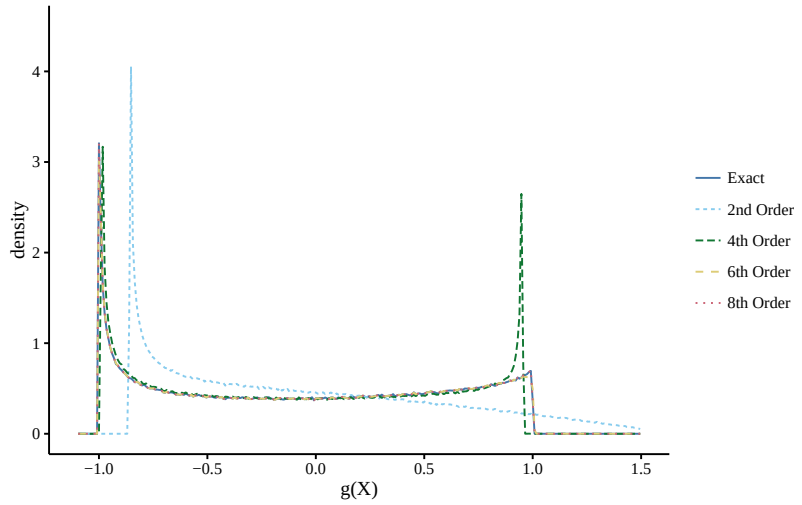


Figure 8: PDF of the random variable $g(X) = \cos(x)$, where $x = \pi z + \pi$ and $Z \sim \mathcal{B}(4,1)$, and various approximations. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

Gauss-Jacobi Quadrature

To estimate the integrals required to compute a Jacobi expansion of a function of a beta-distributed random variables, we turn to Gauss-Jacobi quadrature. As in Gauss-Legendre quadrature (recall that Legendre polynomials are a special case of Jacobi polynomials), the quadrature rule looks like

$$\int_{-1}^{1} f(z)(1-z)^{\alpha}(1+z)^{\beta}\,dz \approx \sum_{i=1}^{n} w_i f(z_i). \tag{61}$$

The abscissas, z_i, for the quadrature rule are the n roots of $P_n^{(\alpha,\beta)}(z)$, and the weights are given by

$$w_i = \frac{2n+\alpha+\beta+2}{n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)(n+1)!}\frac{2^{\alpha+\beta}}{P_n'^{(\alpha,\beta)}(z_i)P_{n+1}^{(\alpha,\beta)}(z_i)}, \tag{62}$$

Here, unlike in Gauss-Legendre quadrature, the weights and abscissas depend on the choice of α and β. Therefore, we will not give an extensive table of coefficients because the generality makes the formulas lengthy. The first-order quadrature ($n = 1$) is

$$x_1 = \frac{b-a}{a+b+2}, \qquad w_1 = \frac{2^{a+b+1}\Gamma(a+2)\Gamma(b+2)}{(a+1)(b+1)\Gamma(a+b+2)}. \tag{63}$$

Beyond $n = 1$ the formulas for the weights and abscissas will not fit on a page, so they do not appear here.

For our example from above, where $Z \sim \mathcal{B}(4,1)$, the quadrature rules are given in Table 13. Notice that unlike Gauss-Legendre

quadrature rules, these rules are not symmetric about the origin. Moreover, the weights sum to the integral of the weight function over the domain:

$$\sum_{i=1}^{n} w_i = \int_{-1}^{1} (1-z)^4(1+z)\,dz = \frac{32}{15}. \tag{64}$$

n	z_i	w_i
1	$-\frac{3}{7}$	$\frac{32}{15}$
2	0	$\frac{16}{21}$
	$-\frac{2}{3}$	$\frac{48}{35}$
3	0.273378	0.213558
	-0.313373	1.121472
	-0.778187	0.798303
4	0.451910	0.062182
	-0.037021	0.545298
	-0.497091	1.049649
	-0.840875	0.476204
5	0.573288	0.019805
	0.169240	0.233970
	-0.247188	0.732908
	-0.615377	0.850154
	-0.879964	0.296496

Table 13: The abscissas and weights for Gauss-Jacobi quadrature up to order 5 with $\alpha = 4$ and $\beta = 1$.

We can use our previous example, of $g(X) = \cos(x)$, where $x = \pi z + \pi$ and $Z \sim \mathcal{B}(4,1)$, as a test of estimating the coefficients using Gauss-Jacobi quadrature rules. In Figure 9, the distribution, as approximated by a sixth-order Jacobi expansion, is computed using Gauss-Jacobi quadratures of different values of n. The distribution at about $n = 6$ the approximation is fairly accurate. This means that we only need six function evaluations to estimate the coefficients. We can see the convergence in the coefficients with the number of quadrature points in Table 14. This table bears out the observation that $n = 6$ is an adequate level of approximation.

Gamma Random Variables: Laguerre Polynomials

The final class of random variable we will consider are gamma random variables. These random variables have support on $(0, \infty)$ and if X is a gamma-distributed random variable we will write

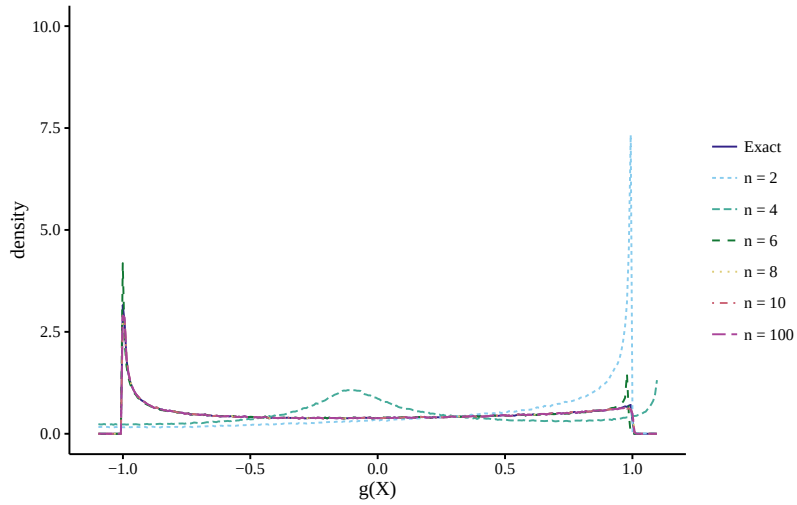


Figure 9: PDF of the random variable $g(X) = \cos(x)$, where $x = \pi z + \pi$ and $Z \sim \mathcal{B}(4,1)$ using a sixth-order Jacobi expansion with various Gauss-Jacobi quadrature rules to approximate the coefficients. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

n	c_0	c_1	c_2	c_3	c_4	c_5
2	-0.035714	-0.642857	0.000000	0.589286	-0.157292	-0.259369
3	-0.069292	-0.503277	0.282089	0.000000	-0.280037	0.478186
4	-0.066861	-0.514456	0.229440	0.132105	-0.000000	-0.135492
5	-0.066957	-0.513982	0.233355	0.120895	-0.058189	0.000000
6	-0.066955	-0.513994	0.233197	0.121391	-0.053616	-0.011632
7	-0.066955	-0.513994	0.233201	0.121378	-0.053807	-0.011110
8	-0.066955	-0.513994	0.233201	0.121378	-0.053802	-0.011124
9	-0.066955	-0.513994	0.233201	0.121378	-0.053802	-0.011124
10	-0.066955	-0.513994	0.233201	0.121378	-0.053802	-0.011124
100	-0.066955	-0.513994	0.233201	0.121378	-0.053802	-0.011124

Table 14: The convergence of the first six coefficients in the Jacobi polynomial expansion $g(X) = \cos(x)$, where $x = \pi z + \pi$ and $Z \sim \mathcal{B}(4,1)$ as estimated by Gauss-Jacobi quadrature rules using different values of n.

$X \sim \mathcal{G}(\alpha, \beta)$ where the PDF of the random variable is[2]

$$f(X) = \frac{\beta^{(\alpha+1)} x^{\alpha} e^{-\beta x}}{\Gamma(\alpha+1)}, \qquad x \in (0,\infty), \quad \alpha > -1, \ \beta > 0. \tag{65}$$

The distribution gets its name from the appearance of the gamma function in the PDF.

As in other variables, it will be useful to have a standardized gamma random variable. In this case we define a $Z \sim \mathcal{G}(\alpha, 1)$, so that Z has the PDF

$$f(z) = \frac{z^{\alpha} e^{-z}}{\Gamma(\alpha+1)}, \qquad z \in (0,\infty), \quad \alpha > -1. \tag{66}$$

We can change from Z to X using a simple scaling

$$z = \beta x. \tag{67}$$

The PDF for a gamma random variable with several different values for the α and β parameters is shown in Figure 10. Here we see

[2] There are several definitions of gamma random variables. One common definition has a different parameter $\alpha' = \alpha + 1$, but the same parameter β.

that α moves the peak of the distribution and that β, as we mentioned above, scales the distribution.

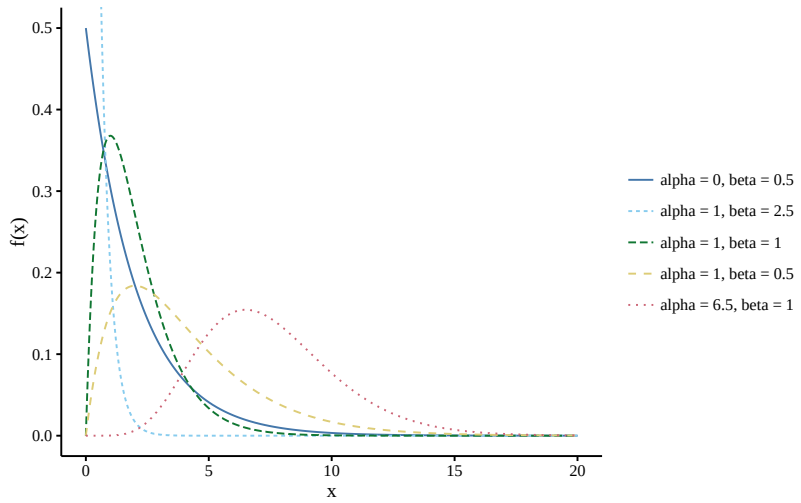


Figure 10: PDF $X \sim \mathcal{G}(\alpha, \beta)$ for several values of α and β. Note that adjusting α moves the peak of the distribution and β scales the distribution along x.

The expectation operator for a gamma random variable can be written as

$$E[g(X)] = \int_0^\infty g(x) \frac{\beta^{(\alpha+1)} x^\alpha e^{-\beta x}}{\Gamma(\alpha+1)}\, dx = \int_0^\infty g\left(\frac{z}{\beta}\right) \frac{z^\alpha e^{-z}}{\Gamma(\alpha+1)}\, dz. \tag{68}$$

Additionally, the mean and variance are given by

$$\bar{x} = \frac{\alpha+1}{\beta}, \qquad \text{Var}(X) = \frac{\alpha+1}{\beta}. \tag{69}$$

The orthogonal polynomials that we will use with functions of a gamma random variable are generalized Laguerre polynomials. Rodrigues' formula for these polynomials is

$$L_n^{(\alpha)}(x) = \frac{x^{-\alpha} e^x}{n!} \frac{d^n}{dx^n} \left(e^{-x} x^{n+\alpha}\right). \tag{70}$$

Some low-order generalized Laguerre polynomials are given in Table 15.

The generalized Laguerre polynomials have the following orthogonality condition

$$\int_0^\infty x^\alpha e^{-x} L_n^{(\alpha)}(x) L_m^{(\alpha)}(x)\, dx = \frac{\Gamma(n+\alpha+1)}{n!} \delta_{n,m}. \tag{71}$$

The generalized Laguerre polynomials form a basis for functions on $(0, \infty)$ that are square integrable with the inner product

$$\langle g(z), h(z) \rangle = \int_0^\infty z^\alpha e^{-z} g(z) h(z)\, dz. \tag{72}$$

Table 15: The first three generalized Laguerre polynomials.

n	$L_n^{(\alpha)}(z)$
0	1
1	$\alpha - x + 1$
2	$\frac{1}{2}\left(\alpha^2 + 3\alpha + x^2 - 2\alpha x - 4x + 2\right)$
3	$\frac{1}{6}\left(\alpha^3 + 6\alpha^2 + 11\alpha - x^3 + 3\alpha x^2 + 9x^2 - 3\alpha^2 x - 15\alpha x - 18x + 6\right)$

Therefore, we can write a function $g(X)$ where $X \sim \mathcal{G}(\alpha, \beta)$ using the following expansion

$$g(x) = \sum_{n=0}^{\infty} c_n L_n^{(\alpha)}(\beta x), \tag{73}$$

where the expansion coefficients are

$$c_n = \frac{n!}{\Gamma(n+\alpha+1)} \int_0^{\infty} g\left(\frac{z}{\beta}\right) z^{\alpha} e^{-z} L_n^{(\alpha)}(z)\, dz. \tag{74}$$

The value of c_0 is once again the mean of $G \sim g(X)$ where $X \sim \mathcal{G}(\alpha, \beta)$:

$$c_0 = \int_0^{\infty} g\left(\frac{z}{\beta}\right) \frac{z^{\alpha} e^{-z}}{\Gamma(\alpha+1)}\, dz = E[g(X)]. \tag{75}$$

The variance of G is related to the sum of the squares of the expansion coefficients:

$$\begin{aligned} \mathrm{Var}(G) &= \int_0^{\infty} \left(\sum_{n=0}^{\infty} c_n L_n^{(\alpha)}(z)\right)^2 \frac{z^{\alpha} e^{-z}}{\Gamma(\alpha+1)}\, dz - c_0^2 \\ &= \sum_{n=1}^{\infty} \frac{\Gamma(n+\alpha+1)}{\Gamma(\alpha+1)n!} c_n^2. \end{aligned} \tag{76}$$

As an example we will examine $G \sim g(X)$ where $g(x) = \cos x$ and $X \sim \mathcal{G}(1, 2)$. For the generalized Laguerre expansion of this function we have expansion coefficients given by

$$c_n = \frac{n!}{\Gamma(n+2)} \int_0^{\infty} \cos\left(\frac{z}{2}\right) z e^{-z} L_n^{(1)}(z)\, dz. \tag{77}$$

The expected value of G is

$$c_0 = \int_0^{\infty} \cos\left(\frac{z}{2}\right) z e^{-z}\, dz = \frac{12}{25}. \tag{78}$$

The expansion to third-order is

$$\cos(X) \approx \frac{12}{25} + \frac{44}{125}(2 - 2x) + \frac{28}{625}\left(2x^2 - 6x + 3\right) + \frac{656}{9375}\left(x^3 - 6x^2 + 9x - 3\right) \quad X \sim \mathcal{G}(1, 2). \tag{79}$$

The variance of G is given by

$$\mathrm{Var}(G) = \sum_{n=1}^{\infty} \frac{\Gamma(n+2)}{\Gamma(2)n!} c_n^2 = \frac{337}{1250} = 0.2696. \tag{80}$$

The convergence of the variance estimate is given in Table 16. As we saw previously, the variance is well-estimated by the fourth-order expansion. We will also see that the fourth-order expansion is also a good estimate of the distribution of G.

order	variance
1	0.2478080
2	0.2538291
4	0.2693313
6	0.2695484
8	0.2695967
∞	0.2696000

Table 16: The convergence of $\mathrm{Var}(G)$ for $g(X) = \cos(x)$, where $X \sim \mathcal{G}(1,2)$.

The convergence of the approximation to G as a function of the order of the Laguerre expansion is shown in Figure 11. The "exact" distribution is determined by evaluating $g(x)$ at the 10^6 samples from $X \sim \mathcal{G}(1,2)$. By the fourth-order expansion the overall character of the true distribution is captured.

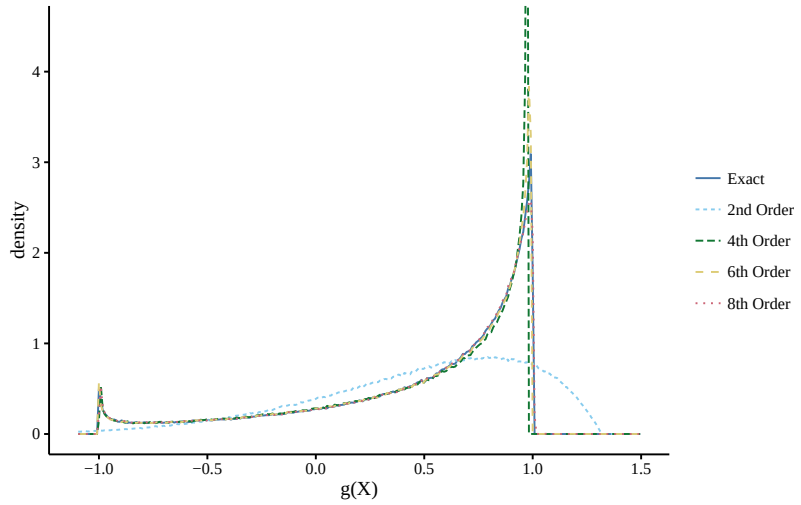


Figure 11: PDF of the random variable $g(X) = \cos(x)$, where $X \sim \mathcal{G}(1,2)$, and various approximations. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

Gauss-Laguerre Quadrature

To estimate the integrals required to compute a generalized Laguerre expansion of a function of a gamma-distributed random variables, we turn to generalized Gauss-Laguerre quadrature. The quadrature rule

has the form

$$\int_0^\infty f(z) z^\alpha e^{-z}\, dz \approx \sum_{i=1}^{n} w_i f(z_i). \tag{81}$$

The abscissas, z_i, for the quadrature rule are the n roots of $L_n^{(\alpha)}(z)$, and the weights are given by

$$w_i = \frac{\Gamma(n+\alpha) z_i}{n!(n+\alpha)(L_{n-1}^{\alpha}(z_i))^2}. \tag{82}$$

The first-order quadrature ($n = 1$) is

$$x_1 = 1 + \alpha, \qquad w_1 = \frac{(\alpha+1)\Gamma(a+1)}{a+1}. \tag{83}$$

For $n = 2$ we have

$$x_{1,2} = \alpha \pm \sqrt{\alpha+2}, \qquad w_{1,2} = \frac{\left(3 \pm \sqrt{3}\right)\Gamma(a+2)}{2(a+2)\left(a+1-\left(3\pm\sqrt{3}\right)\right)^2}. \tag{84}$$

Beyond second-order the quadratures are too lengthy to write for a general value of α. Note that if $\alpha = 0$, then the quadrature rule reduces to simple Gauss-Laguerre quadrature.

To use the generalized Gauss-Laguerre quadratures to compute the inner products for the Laguerre expansion of a gamma-distributed random variable, $X \sim \mathcal{G}(\alpha, \beta)$ as

$$\int_0^\infty f\left(\frac{z}{\beta}\right) z^\alpha e^{-z}\, dz \approx \sum_{i=1}^{n} w_i f\left(\frac{z_i}{\beta}\right). \tag{85}$$

For our example from above, where $X \sim \mathcal{G}(1,2)$, the quadrature rules are given in Table 17. In this case the weights sum to the integral of the weight function over the domain:

$$\sum_{i=1}^{n} w_i = \int_0^\infty z e^{-z}\, dz = 2. \tag{86}$$

We can use our previous example, of $g(X) = \cos(x)$, where $X \sim \mathcal{G}(1,2)$ as a test of estimating the coefficients using generalized Gauss-Laguerre quadrature rules. In Figure 12, the distribution, as approximated by a fifth-order Laguerre expansion, is computed using generalized Gauss-Laguerrequadratures of different values of n. The distribution at about $n = 8$ the approximation is fairly accurate. We can see the convergence in the coefficients with the number of quadrature points in Table 14. This table bears out the observation that $n = 8$ is an adequate level of approximation.

n	z_i	w_i
1	2	1
2	$3 \pm \sqrt{3}$	$\frac{3 \pm \sqrt{3}}{3\left(2-\left(3 \pm \sqrt{3}\right)\right)^2}$
3	7.758770	0.020102
	3.305407	0.391216
	0.935822	0.588681
4	10.953894	0.001316
	5.731179	0.074178
	2.571635	0.477636
	0.743292	0.446871
5	14.260103	0.000069
	8.399067	0.008720
	4.610833	0.140916
	2.112966	0.502281
	0.617031	0.348015

Table 17: The abscissas and weights for generalized Gauss-Laguerre quadrature up to order 5 with $\alpha = 1$.

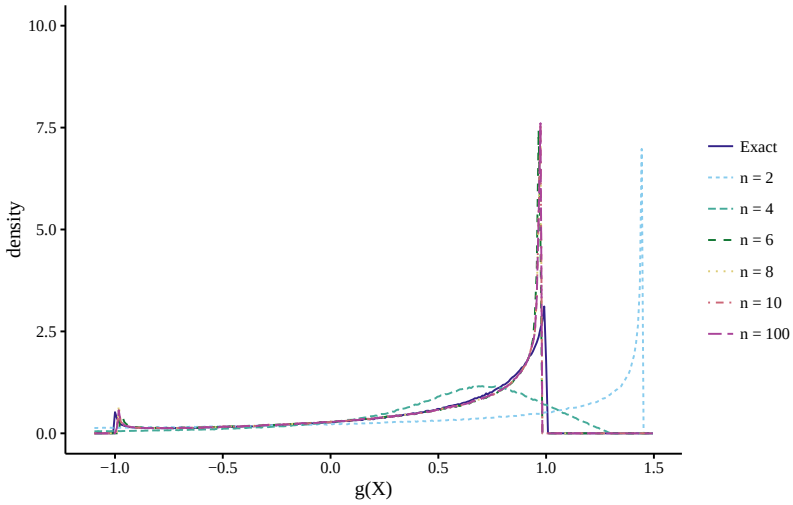


Figure 12: PDF of the random variable $g(X) = \cos(x)$, where $x = \pi z + \pi$ and $Z \sim \mathcal{B}(4, 1)$ using a sixth-order Jacobi expansion with various Gauss-Jacobi quadrature rules to approximate the coefficients. This figure was generated from 10^6 samples of X that were used to evaluate $g(X)$ and the various approximations.

n	c_0	c_1	c_2	c_3	c_4	c_5
2	0.484528	0.438701	0.000000	-0.219350	-0.223933	-0.140776
3	0.478523	0.343285	0.077209	-0.000000	-0.046325	-0.099540
4	0.480185	0.352313	0.038293	-0.054229	-0.000000	0.036153
5	0.479984	0.352043	0.045559	-0.053931	-0.036908	-0.000000
6	0.480001	0.351990	0.044746	-0.052110	-0.029267	-0.004078
7	0.480000	0.352001	0.044801	-0.052532	-0.029939	-0.000867
8	0.480000	0.352000	0.044800	-0.052475	-0.029968	-0.001564
9	0.480000	0.352000	0.044800	-0.052480	-0.029949	-0.001480
10	0.480000	0.352000	0.044800	-0.052480	-0.029952	-0.001484
100	0.480000	0.352000	0.044800	-0.052480	-0.029952	-0.001485

Table 18: The convergence of the first six coefficients in the generalized Laguerre polynomial expansion $g(X) = \cos(x)$, where $X \sim \mathcal{G}(1, 2)$ as estimated by generalized Gauss-Laguerre quadrature rules using different values of n.

Example from a PDE: Poisson Equation with an uncertain source

The examples we have seen so far have been functions that have been simple to evaluate. In such examples, there is no benefit to minimizing the number of function evaluations. For a more realistic example where the function evaluations are expensive, but not too expensive, we will look at a a quantity related to the solution of the 2-D Poisson equation with Dirichlet boundary conditions:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u(x,y) = -q(x,y). \tag{87}$$

$$u(1,y) = u(x,1) = u(-1,y) = u(x,-1) = 0. \tag{88}$$

The source q will be a Gaussian in space with an uncertain center in y:

$$q(x,y) = \exp\left[-x^2 - (y-\omega)^2\right]. \tag{89}$$

The center of the Gaussian in the y-coördinate will be a uniform random variable in the range $[-0.25, 0.25]$ (i.e., $\omega \sim \mathcal{U}(-0.25, 0.25)$). We are interested in the integral over a quarter of the domain. Our quantity of interest is therefore

$$g(\omega) = \int_0^1 dx \int_0^1 dy\, u(x,y;\omega). \tag{90}$$

The notation $u(x,y;\omega)$ denotes that the solution depends on the center of Gaussian ω.

Because ω is a uniform random variable, we will use a Legendre expansion to compute an approximation to $G \sim g(c)$. From Eq. (37) we are interested in computing the integral

$$c_n = \frac{2n+1}{2} \int_{-1}^{1} g\left(\frac{z}{4}\right) P_n(z)\, dz. \tag{91}$$

We will estimate the Legendre expansion coefficients using Gauss-Legendre quadrature. For example, using an $n = 2$ quadrature rule

we would estimate the coefficients as

$$c_n \approx \frac{2n+1}{2}\left(g\left(-\frac{1}{4\sqrt{3}}\right)P_n\left(-\frac{1}{4\sqrt{3}}\right)+g\left(\frac{1}{4\sqrt{3}}\right)P_n\left(-\frac{1}{4\sqrt{3}}\right)\right). \tag{92}$$

Note, to compute the c_n in this case will require solving Poisson's equation twice, each time with different sources, and computing the integral in Eq. (90). There are, at least, a countably infinite number of ways to estimate the solution to Poisson's equation. Here we will use Mathematica's `NDSolve` function. Solving Poisson's equation with these two values of ω gives

$$g\left(-\frac{1}{4\sqrt{3}}\right) = 0.381378, \qquad g\left(\frac{1}{4\sqrt{3}}\right) = 0.381378.$$

Therefore, for example, c_0 will be

$$c_0 \approx \frac{1}{2}\left[g\left(-\frac{1}{4\sqrt{3}}\right)+g\left(\frac{1}{4\sqrt{3}}\right)\right] = 0.381378. \tag{93}$$

In Table 19 estimates for the expansion coefficients up to c_n are shown. Note that in the best case, we could only hope for a quadrature rule with n points to integrate up to c_{2n-1} accurately, and that this would only be the case if g we are a constant function. From this table it seems that the integrals are accurate (though not exact) up to c_n for an n point quadrature rule once $n > 2$.

n	c_0	c_1	c_2	c_3	c_4	c_5
1	0.386712	0.000000	-0.966780	0.000000	1.305153	0.000000
2	0.381378	0.000000	-0.000000	-0.000000	-1.334823	-0.000000
3	0.381406	-0.000000	-0.010613	-0.000000	0.014327	0.000000
4	0.381406	-0.000000	-0.010559	0.000000	-0.000000	0.000000
5	0.381406	0.000000	-0.010559	0.000000	0.000071	-0.000000
6	0.381406	-0.000000	-0.010559	0.000000	0.000071	-0.000000
7	0.381406	-0.000000	-0.010559	0.000000	0.000071	-0.000000
8	0.381409	0.000000	-0.010567	-0.000000	0.000079	0.000000
9	0.381406	0.000000	-0.010559	-0.000000	0.000071	-0.000000
10	0.381406	0.000000	-0.010559	-0.000000	0.000071	-0.000000

Table 19: The convergence of the first six coefficients in the 2-D Poisson equation example as a function of the number of Gauss-Legendre quadrature points used.

Using the results from Table 19 we can create empirical PDF of G for different quadrature rules. For a given polynomial expansion, generating 10^6 samples requires only evaluating that many polynomials. In Figure 13 these PDFs are shown for quadrature rules using $2, 4, 6$, and 10 points as well as the PDF from 3000 Monte Carlo samples of G by randomly selecting ω's. Note that with only 6 function evaluations using the $n = 6$ quadrature rule we get a better representation of G than thousands of Monte Carlo samples, and a savings of about 0.75 hours on my laptop.

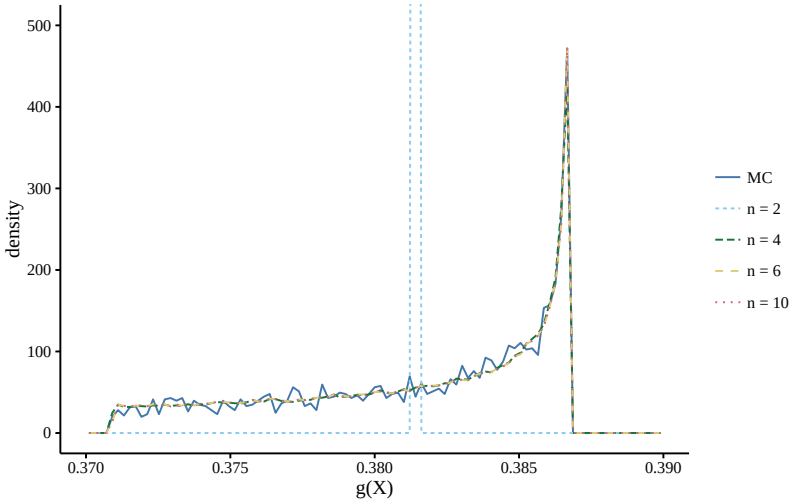


Figure 13: PDF of the random variable $g(\omega) = \int_0^1 dx \int_0^1 dy\, u(x, y; \omega)$, where $\omega \sim \mathcal{U}(-0.25, 0.25)$ and u is the solution to Eq. (87), using several different Gauss-Legendre quadrature rules and a Monte Carlo simulation using 3×10^3 numerical solutions of Poisson's equation.

Mutli-dimensional Polynomial Chaos Expansions

It is likely that in a realistic problem there will be several sources of uncertainty and several uncertain parameters. It may also be possible that the different parameters may have different types of distributions. Let us consider a generic function of d random variables, θ_i, with an expansion given by

$$g(\theta_1, \ldots, \theta_d) = \sum_{l_1=0}^{\infty} \cdots \sum_{l_d=0}^{\infty} c_{l_1,\ldots,l_d} \mathfrak{P}_{l_1,\ldots,l_d}(\theta_1, \ldots, \theta_d). \tag{94}$$

Here $\mathfrak{P}_{l_1,\ldots,l_d}(\theta_1, \ldots, \theta_d)$ is a product of the d orthogonal polynomials,

$$\mathfrak{P}_{l_1,\ldots,l_d}(\theta_1, \ldots, \theta_d) = \prod_{i=1}^{d} P_{l_i}(\theta_i), \tag{95}$$

and the expansion coefficients are

$$c_{l_1,\ldots,l_d} = \int_{D_1} d\theta_1 \cdots \int_{D_d} d\theta_d \, g(\theta_1, \ldots, \theta_d) \mathfrak{P}_{l_1,\ldots,l_d}(\theta_1, \ldots, \theta_d) \mathfrak{w}(\theta_1, \ldots, \theta_d), \tag{96}$$

and $\mathfrak{w}(\theta_1, \ldots, \theta_d)$ is the product of the weight functions for the d bases. If the sum is truncated at degree N polynomials then there will be $(1+N)^d$ terms in the expansion.

As a simple example, consider the function $g = \cos(\theta_1)\cos(\theta_2)$ with $\theta_i \sim \mathcal{U}(0, 2\pi)$. A second-order expansion would have the form

$$\begin{aligned} g(\theta_1, \theta_2) &= c_{0,0} + c_{1,0}P_1(\pi\theta_1 + \pi) + c_{0,1}P_1(\pi\theta_2 + \pi) + c_{2,0}P_2(\pi\theta_1 + \pi) \\ &+ c_{0,2}P_2(\pi\theta_2 + \pi) + c_{1,1}P_1(\pi\theta_1 + \pi)P_1(\pi\theta_2 + \pi) + c_{2,1}P_2(\pi\theta_1 + \pi)P_1(\pi\theta_2 + \pi) \\ &+ c_{1,2}P_1(\pi\theta_1 + \pi)P_2(\pi\theta_2 + \pi) + c_{2,2}P_2(\pi\theta_1 + \pi)P_2(\pi\theta_2 + \pi). \end{aligned} \tag{97}$$

To compute the expansion coefficients we can use what is known as a tensor-product quadrature set. Here we take a 1-D quadrature rule with n points and weights given by $\{w_i, x_i\}$ for $i = 1 \ldots n$, that we denote as Q_n so that

$$Q_n f(x) = \sum_{l=1}^{n} w_l f(x_l), \tag{98}$$

and apply it over all dimensions as

$$Q_n^{(d)} g(\theta_1, \ldots, \theta_d) = \sum_{l_1=1}^{n} \cdots \sum_{l_d=1}^{n} w_{l_1} \cdots w_{l_d} g(\theta_{1l_1}, \ldots, \theta_{dl_d}), \tag{99}$$

where θ_{i,l_j} is the ith input evaluated at its jth point in the quadrature set. It is sometimes convenient to write $Q^{(d)}$ as a tensor product of 1-D quadrature rules. We define a tensor product of two quadrature rules as

$$Q_n \otimes Q_m = \{\{w_i w_j, (x_i, x_j)\} : i = 1 \ldots n, j = 1 \ldots m\}. \tag{100}$$

Therefore, we can write a tensor-product quadrature comprised of n point quadratures as

$$Q_n^{(d)} g(\theta_1, \ldots, \theta_d) = (Q_n^{(1)} \otimes \cdots \otimes Q_n^{(d)}) g. \tag{101}$$

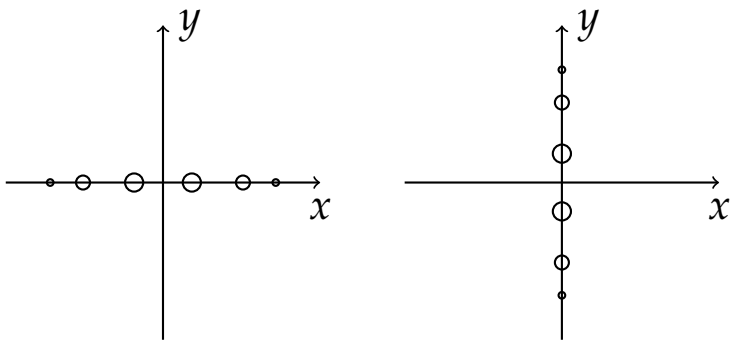
We could in principle have each dimension have a different number of quadrature points, and in many cases this will make the calculation more efficient.

The number of quadrature points scales geometrically with d. This is the so-called curse of dimensionality because the number of function evaluations needed explodes as d gets larger. For example, using a two-point quadrature rule, when $d = 26$, requires one simulation for every person in Germany. Even worse $d = 26$ requires a mole (6×10^{23}) of calculations for the $n = 8$ rule. In a full-scale engineering system, 78 uncertain parameters is not out of the question.

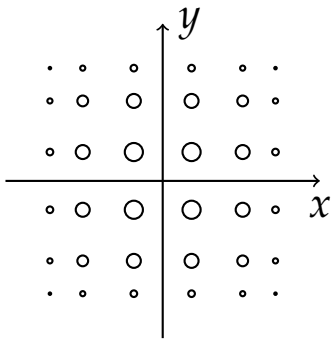
As an example, the tensor product quadrature for $d = 2$ comprised of 1-D, six-point Gauss-Legendre quadrature rules is shown in Figure 14. Two things are evident in this figure: the weights are much larger in the middle of domain, and the points are more densely packed near the corners. These effects are even more pronounced as the number points in the quadrature set goes up.

Example Three-Dimensional Expansion: Black-Scholes Pricing Model

For an example of a polynomial chaos expansion in multiple dimensions we will look at the solution to the Black-Scholes partial differential equation for the value of a call option. A call option gives the holder the ability to purchase a stock at a given price, called the strike price, at a given future date. The value of the option is a function of the current price of the stock (S), the strike price (K), the time to expiration in years (T), the risk-free interest rate (r), the dividend rate the stock pays q, and the volatility of the stock (σ). Three of these, r, q, and σ are uncertain parameters.



1-D quadrature in x 1-D quadrature in y



Tensor product quadrature

Figure 14: Illustration of the 2-D tensor-product quadrature derived from the six-point Gauss-Legendre quadrature set. The size of a point is proportional to its weight.

The Black-Scholes model is based on assuming that the stock price follows geometric Brownian motion. The solution for the price of an option from the Black-Scholes model can be given by

$$p = e^{-rT}\left(F\Phi(v_1) - K\Phi(v_2)\right), \tag{102}$$

where

$$F = Se^{(r-q)T}, \tag{103}$$

$$v_1 = \frac{\log\frac{S}{K} + (r - q + \frac{1}{2}\sigma^2)T}{T\sqrt{T}}, \qquad v_2 = v_1 - \sigma\sqrt{T}, \tag{104}$$

and $\Phi(z)$ is the standard normal CDF function, c.f. Eq. (13).

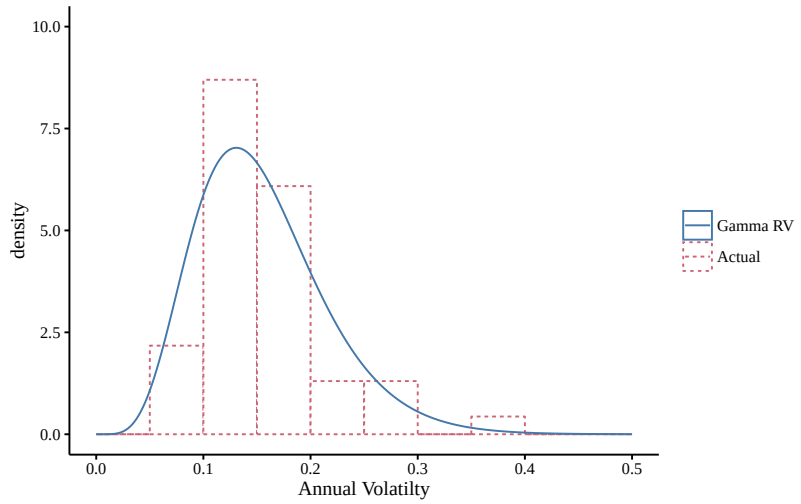


Figure 15: The empirical distribution and a fitted Gamma distribution in the annual percentage change in Coca-Cola stock for each year between 1970 and 2015. The distribution has a mean of 0.154083 and variance of 0.0036984. This corresponds to a Gamma distribution of $\Sigma \sim \mathcal{G}(5.46636, 41.8142)$.

We are interested in calculating the current value of a call option for stock in the Coca-Cola company, ticker symbol KO. On 15 August 2016, KO was trading at \$44.15. We will consider a call with strike price of \$44. The option expiration is 158 days away ($T = 0.432877$). This option is trading at \$1.46. We need to estimate the distribution of random variables, r, q, and σ. The distribution that we will use for the volatility σ will be based on the actual annual standard deviations of daily returns for the years from 1970 to 2015. The histogram of these 45 volatilities is shown in Figure 15, along with a Gamma distribution that matches the mean and variance of the observations, $\Sigma \sim \mathcal{G}(5.46636, 41.8142)$. This distribution is found by solving Eq. (69) to match the observed mean and variance in the data. Note that the distribution indicates that 10% if the time the volatility will be greater than 23.6%. For the interest rate, r, we will use the benchmark LIBOR 30-day interest rate with a Gamma distribution $r = 0.0048x$ with $X \sim \mathcal{G}(0, 1)$. This distribution had a mean of the current rate

(0.48%). For the dividend rate we will use a uniform distribution so that $Q \sim \mathcal{U}(0.025, 0.045)$.

The expansion of $p(X, D, \Sigma)$ will have the form

$$p(X,Q,\Sigma) = \sum_{l_x=0}^{\infty}\sum_{l_d=0}^{\infty}\sum_{l_\sigma=0}^{\infty} c_{l_x l_d l_\sigma} L_{l_x}^{(0)}(x) P_{l_d}\left(\frac{2d-0.7}{0.2}\right) L_{l_\sigma}^{(5.46636)}(41.8142\sigma). \tag{105}$$

From this equation, we can compute the mean of the distribution, c_{000} as

$$\bar{p} = c_{000} = \int_0^\infty dx \int_{0.025}^{0.045} dq \int_0^\infty dz\, p\left(x, q, \frac{z}{41.8142}\right) \frac{z^{5.46636}}{\Gamma(6.46636)} e^{-x-z} \left(\frac{1}{0.02}\right) \approx 1.56662. \tag{106}$$

Note that this is slightly higher than the price the option is trading at, \$1.46.

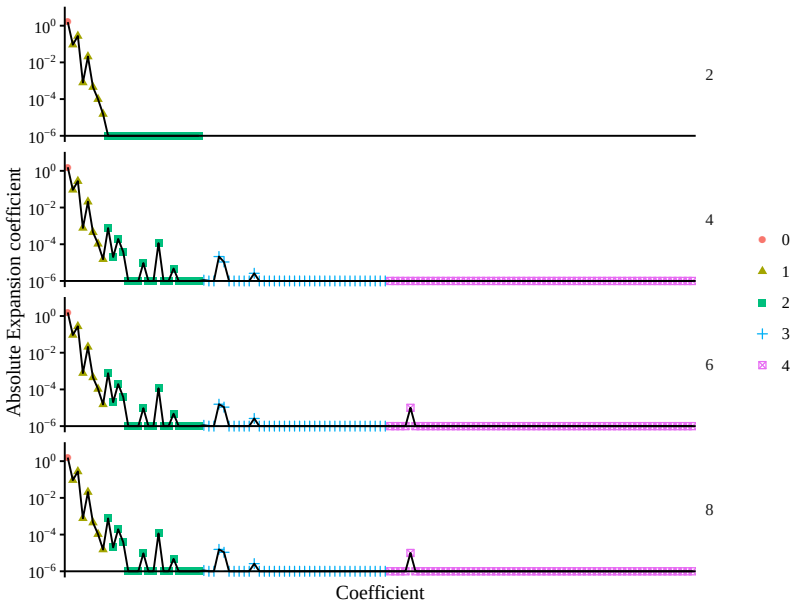


Figure 16: The magnitude of the coefficients in the expansion of the value of a call option as a function of three uncertain parameters, $r = 0.0048x$ with $X \sim \mathcal{G}(0,1)$, $Q \sim \mathcal{U}(0.025, 0.045)$, and $\Sigma \sim \mathcal{G}(5.46636, 41.8142)$. The color and shape of the points indicates the maximum polynomial degree that the coefficient responds to, e.g., c_{011} would be a "1" in the figure. The different panels on the figure indicate the number, n, of Gauss quadrature points in each dimension. Those points with a maximum polynomial degree greater than n are not shown, and the coefficients are "floored" to a minimum of 10^{-6}.

Because the price of the option is a well-behaved function, we will expand p with polynomial degree up to order four:

$$p(X,Q,\Sigma) = \sum_{l_x=0}^{4}\sum_{l_d=0}^{4}\sum_{l_\sigma=0}^{4} c_{l_x l_d l_\sigma} L_{l_x}^{(0)}(x) P_{l_d}\left(\frac{2d-0.7}{0.2}\right) L_{l_\sigma}^{(5.46636)}(41.8142\sigma). \tag{107}$$

Such an expansion will have $5^3 = 125$ terms. Using tensor-product Gauss quadrature—Gauss-Laguerre in x and σ, Gauss-Legendre in q—we can estimate these coefficients. The results from these calculations with various numbers of points in the 1-D quadrature rules that comprise the tensor-product quadrature are shown Figure

16. This figure indicates the maximum single polynomial degree in each point using a color/shape. Here we only show coefficients with a magnitude larger than 10^{-6}, for the $n = 2$ rules we do not show any coefficients corresponding to degree three or four polynomials.

From Figure 16 we can see that the $n = 2$ quadrature rule does a good job of estimating the low-order, large-magnitude coefficients. This indicates that most of the variation in the distribution can be captured using only $2^3 = 8$ evaluations of the function. The higher-order coefficients have a smaller magnitude and can be captured using $n = 4$ rules, and the largest, significant coefficient c_{004} or the coefficient for a quartic in volatility can be captured using $n = 6$ or a total of $6^3 = 216$ function evaluations.

One way to compare the quadrature rules is to look at the convergence of the variance. The variance in the price is

$$\text{Var}(P) = \sum_{l_x=1}^{\infty} \sum_{l_d=1}^{\infty} \sum_{l_\sigma=1}^{\infty} \frac{\Gamma(l_x+1)\Gamma(l_\sigma+6.46636)}{l_x! l_\sigma! \Gamma(6.46636)(2l_d+1)} c^2_{l_x l_d l_\sigma}. \tag{108}$$

The results for this calculation using the expansion coefficients in Figure 16 are shown in Table . This table indicates that the $n = 2$ coefficients estimate the variance to three digits of accuracy.

n	$\text{Var}(P)$
2	0.486085
4	0.486321
6	0.486321
8	0.486321

Table 20: The convergence of the variance in the option price as a function of the quadrature rule used.

We compare random 10^6 samples from the Black-Scholes solution, to the same number of samples to the expansions as estimated with the various quadrature rules, in Figure 17. This figure indicates that because of the smoothness of an expansion with only the a few terms is accurate.

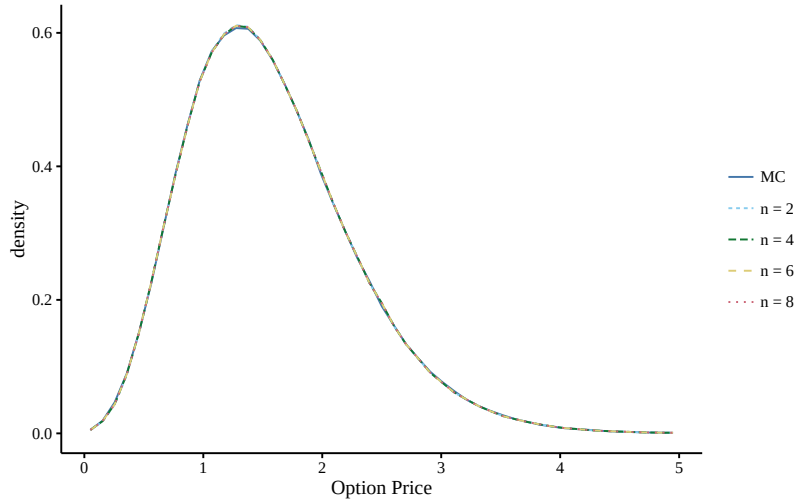


Figure 17: The distribution of the price of an option with a strike price of \$44, a stock price of \$44.15, and days to expiration of 158. The risk-free interest rate is $r = 0.0048x$ with $X \sim \mathcal{G}(0,1)$, the dividend rate is $Q \sim \mathcal{U}(0.025, 0.045)$, and the volatility of the stock is $\Sigma \sim \mathcal{G}(5.46636, 41.8142)$. We compare the polynomial chaos expansion as computed using tensor-products of quadrature rules with $n = 2, 4, 6, 8$ and compare these distributions to a Monte Carlo distribution with 10^6 samples.

From this example, several things are evident. With a smoothly varying function, the expansion order required to estimate the distribution of the quantity of interest, and the number of function

evaluations needed are small. The results also indicate that of the many coefficients possible in a high-order expansion will be negligible. In next chapters we will investigate how to take advantage of this structure.

Sparse Quadrature

The explosion of terms in multi-dimensional expansions comes, in part, from the cross-terms that appear in the expansion. For example, in a fourth-order expansion we end up with 4^d degree polynomials because the highest order terms in the series are a product of four d-degree polynomials. The tensor-product Gauss quadratures that we use to estimate the expansion can accurately integrate these polynomials. Nevertheless, it is often the case that these high-degree interactions (that is the product of several high-degree polynomials) are unnecessary in the expansion (as we saw in the previous case).

In such a scenario it can be useful to change the way that we expand the output in orthogonal polynomials. Instead of including combinations of polynomials up to a given degree, we look to include only polynomials up to a maximum degree. In other words

$$g(\theta_1 \ldots, \theta_d) \approx \sum_{l_1+\cdots+l_d<N} c_{l_1,\ldots,l_d} \mathfrak{P}_{l_1,\ldots,l_d}(\theta_1, \ldots, \theta_d), \tag{109}$$

With this expansion, we no longer need to integrate any polynomials of degree higher than N. Therefore, our tensor-product quadrature rule integrates higher-degree polynomials than we need.

For this situation we can use Smolyak sparse quadrature sets. These rules construct quadrature points that do not grow as fast as product quadrature grids. To accomplish this, we combine quadrature rules to ensure that a polynomial of a given degree in any single dimension, but not products of polynomials of that degree, is exactly integrated.

We have all the pieces we need to define a Smolyak sparse quadrature set. For a given value of ℓ, in d dimensions the quadrature rule is defined as

$$S_\ell^{(d)} f = \sum_{q=\ell-d}^{\ell-1} (-1)^{\ell-1-q} \binom{d-1}{\ell-1-q} \sum_{\|\vec{k}\|_1=q+d} Q_{2^{k_1}-1} \otimes \cdots \otimes Q_{2^{k_d}-1} f, \tag{110}$$

where $\|\vec{k}\|_1 = \sum_{i=1}^{d} |k_i|$.

Looking at this formula we see that the tensor products where the sum of the number of points in each dimension equals a constant are included. Note that the quadrature rule can have negative weights.

To demonstrate how these rules work we will look at the quadrature rule with $\ell = 3$ and Gauss-Legendre quadrature. In this case we should have the a quadrature rule with up to $2^3 - 1$ points:

$$\begin{aligned} S_3^{(2)} f &= \sum_{q=1}^{2} (-1)^{2-q} \binom{1}{2-q} \sum_{\|\vec{k}\|_1 = q+2} Q_{2^{k_1}-1} \otimes Q_{2^{k_2}-1} f \\ &= - \sum_{\|\vec{k}\|_1 = 3} Q_{2^{k_1}-1} \otimes Q_{2^{k_2}-1} f + \sum_{\|\vec{k}\|_1 = 4} Q_{2^{k_1}-1} \otimes Q_{2^{k_2}-1} f \\ &= - (Q_1 \otimes Q_3) f - (Q_3 \otimes Q_1) f + (Q_3 \otimes Q_3) f + (Q_1 \otimes Q_7) f + (Q_7 \otimes Q_1) f \end{aligned}$$

Counting up the total number of points in this rule there are 21 compared to 49 for the tensor product quadrature rule for $Q_7 \otimes Q_7$.

The $(Q_1 \otimes Q_3)$ and $(Q_3 \otimes Q_1)$ rules are completely redundant with $(Q_3 \otimes Q_3)$ and the $(Q_1 \otimes Q_7)$ and $(Q_3 \otimes Q_1)$ rules share the origin with $(Q_3 \otimes Q_3)$,

We show the points for $S_3^{(2)}$ based on Gauss-Legendre quadrature in Figure 18 as well as the comparable tensor-product quadrature rule, $Q_7 \otimes Q_7$,

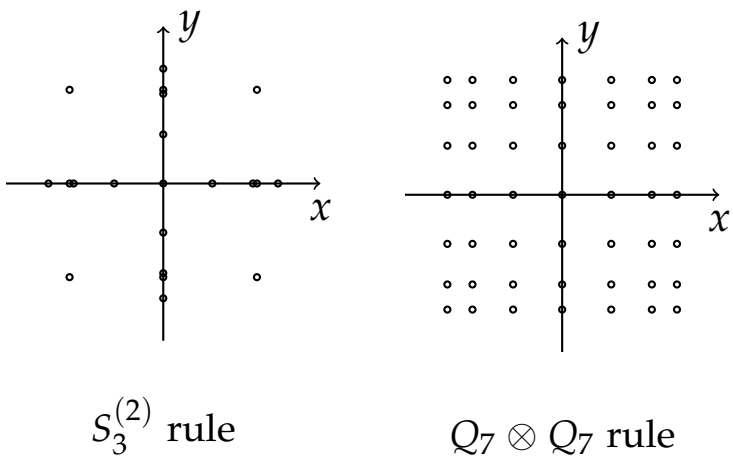


Figure 18: Comparison of the Smolyak sparse quadrature rule of level $\ell = 3$ and the tensor-product rule comprised of 7-point Gauss-Legendre quadrature rules.

Another way to show the construction of a 2-D Smolyak quadrature rule is to write all the quadrature rules up to order $2^\ell - 1$ in a tableau of tensor-product quadratures where the number of points in the x-direction increases from left to right, and the number of points in the y-direction increases from bottom to top. The Smolyak quadrature rule will be a linear combination of the tensor-product rules from the diagonal and below. This construction is shown in Figure 19.

Now that we have seen how the sparse grids work, we will discuss why they are constructed in the form that they are. As we have said, a product quadrature rule comprised of n points in 1-D, will

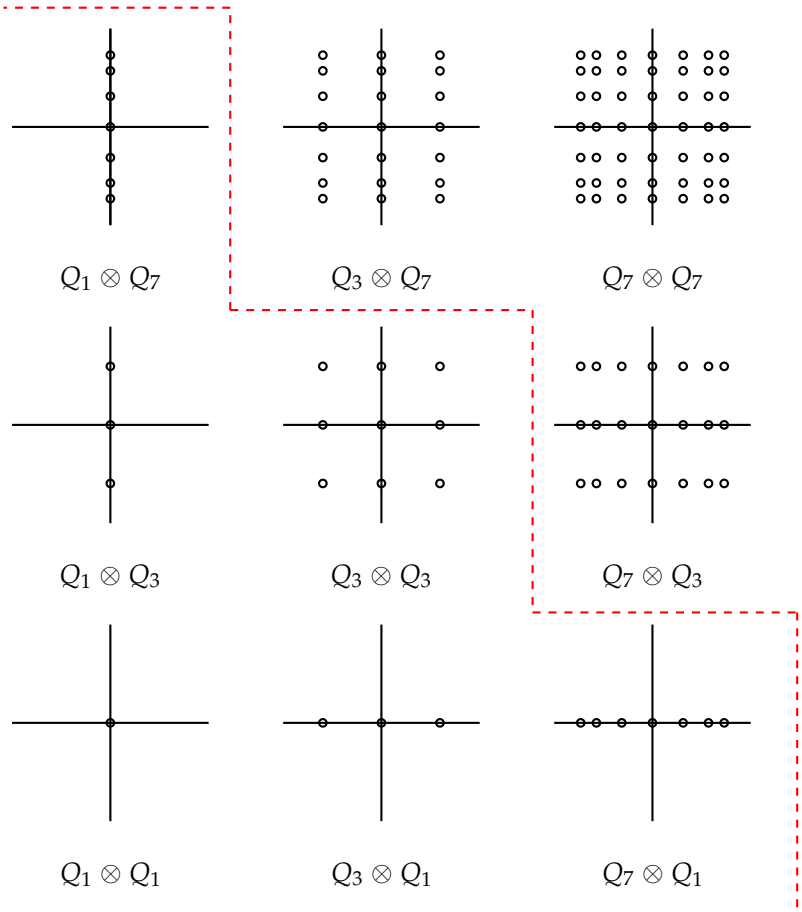


Figure 19: Demonstration of the construction of the Smolyak quadrature rule with $\ell = 3$ in two dimensions comprised of Gauss-Legendre quadrature rules. The Smolyak quadrature is a linear combination of the points below the dashed line.

integrate d-dimensional polynomials where *any single* component polynomial has degree less than or equal to $(2n-1)$. The Smolyak construction is designed to integrate polynomials with a total degree of equal to $(2n-1)$. This is shown in Figure 20 for $n=2$. Indeed, it can be shown that the Smolyak sparse grid that is exact on polynomials of N in the one-dimensional quadrature rules will be exact on polynomials of total degree N for the multi-dimensional integral.

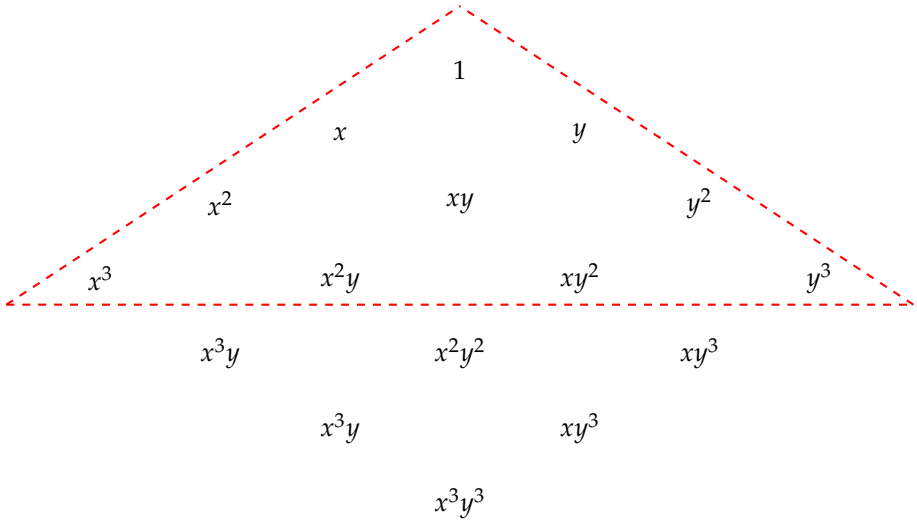


Figure 20: The monic polynomials that can be integrated exactly by a two-dimensional tensor-product Gauss quadrature rule comprised of two-point rules. The dashed line encloses the polynomials that the sparse grid will integrate.

The construction of the quadrature set will illuminate the origin of Eq. (110), in particular why there needs to be negatively-weighted points. Looking at Figure 20, to integrate the polynomials in the triangle we can think about it in terms of "adding" quadrature rules:

$$\begin{pmatrix} & 1 & \\ x & & y \\ x^2 & & y^2 \end{pmatrix} = \begin{pmatrix} & 1 & \\ x & & \\ x^2 & & \end{pmatrix} + \begin{pmatrix} & 1 & \\ & y & \\ & & y^2 \end{pmatrix} + \begin{pmatrix} & 1 & \\ x & & y \\ & xy & \end{pmatrix} - \begin{pmatrix} & 1 & \\ x & & \\ & & \end{pmatrix} - \begin{pmatrix} & 1 & \\ & y & \\ & & \end{pmatrix} \tag{111}$$

Here we see the reason for the appearance of the term $(-1)^{\ell-1-q}$ term in Eq. 110. It is worth mentioning that there is an alternate form of the Smolyak rule. For this will also need to define a difference in quadratures as

$$\Delta_{2^l-}f = Q_{2^l-}f - Q_{2^{l-1}-1}f, \tag{112}$$

and $Q_0 = \emptyset$. Using these the Smolyak rule can be written as

$$S_\ell^{(d)} f = \sum_{q=0}^{\ell-1} \sum_{\|\vec{k}\|_1 = q+d} \Delta_{2^{k_1}-1} \otimes \cdots \otimes \Delta_{2^{k_d}-1} f. \tag{113}$$

Black-Scholes Example Redux

Turning back to our Black-Scholes example from before, we will construct a Smolyak sparse grid for this 3-D expansion. As we saw in before, a tensor-product quadrature comprised of six-point quadrature rules was able to capture the most important coefficients in the expansion. This rule has $6^3 = 216$ function evaluations. In this case, we will use the $\ell = 3$ Smolyak sparse grid to compute the coefficients in the expansion. This quadrature rule can be calculated from

$$\begin{aligned}
S_3^{(3)} f = &\sum_{q=0}^{2} (-1)^{2-q} \binom{2}{2-q} \sum_{\|\vec{k}\|_1 = q+3} Q^{(\sigma)}_{2^{k_1}-1} \otimes Q^{(x)}_{2^{k_2}-1} \otimes Q^{(z)}_{2^{k_3}-1} f \\
= &Q_1^{(\sigma)} \otimes Q_1^{(x)} \otimes Q_1^{(z)} f \\
&- 2\left[Q_3^{(\sigma)} \otimes Q_1^{(x)} \otimes Q_1^{(z)} f + Q_1^{(\sigma)} \otimes Q_3^{(x)} \otimes Q_1^{(z)} f + Q_1^{(\sigma)} \otimes Q_1^{(x)} \otimes Q_3^{(z)} f \right] \\
&+ Q_7^{(\sigma)} \otimes Q_1^{(x)} \otimes Q_1^{(z)} f + Q_1^{(\sigma)} \otimes Q_7^{(x)} \otimes Q_1^{(z)} f + Q_1^{(\sigma)} \otimes Q_1^{(x)} \otimes Q_7^{(z)} f + \\
&+ Q_3^{(\sigma)} \otimes Q_3^{(x)} \otimes Q_1^{(z)} f + Q_3^{(\sigma)} \otimes Q_1^{(x)} \otimes Q_3^{(z)} f + Q_1^{(\sigma)} \otimes Q_3^{(x)} \otimes Q_3^{(z)} f
\end{aligned}$$

The component 1-D rules in $S_3^{(3)}$ are shown in Table 21. Note that only the z points are nested at all (notice the repeated 0).

	$\beta\sigma$	w_σ	x	w_x	z	w_z
Q_1	6.466360	271.060701	1.000000	1.000000	0.000000	2.000000
	13.811184	13.236834	6.289945	0.010389	0.774597	0.555556
Q_3	7.787369	148.010162	2.294280	0.278518	0.000000	0.888889
	3.800528	109.813705	0.415775	0.711093	-0.774597	0.555556
	28.226889	0.000454	19.395728	0.000000	0.949108	0.129485
	20.399826	0.129138	12.734180	0.000016	0.741531	0.279705
	14.769642	4.663395	8.182153	0.001074	0.405845	0.381830
Q_7	10.417345	42.165053	4.900353	0.020634	0.000000	0.417959
	6.984121	116.015439	2.567877	0.147126	-0.405845	0.381830
	4.281556	93.279531	1.026665	0.421831	-0.741531	0.279705
	2.185142	14.807693	0.193044	0.409319	-0.949108	0.129485

Table 21: The 1-D quadrature rules that comprise the sparse rule $S_3^{(3)}$.

The nesting of points in the z direction leads to 7 redundant points and a total of 50 unique points in the $S_3^{(3)}$ set. The points in the set are shown in in Figure 21.

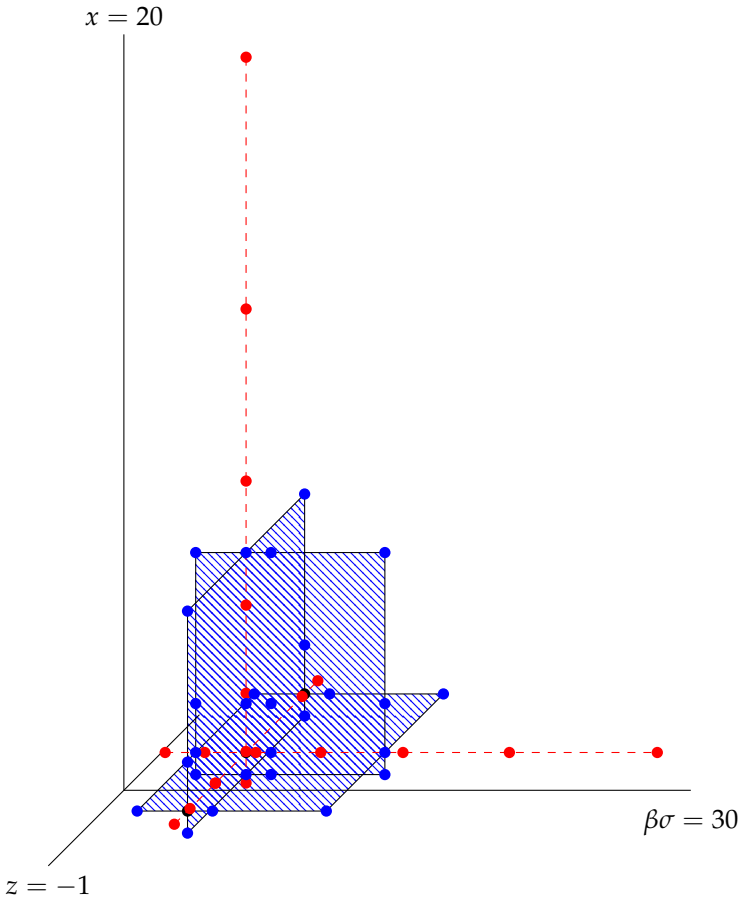


Figure 21: Depiction of the points for the $S_3^{(3)}$ quadrature set comprised of the rules in Table 21. The red points are the Q_7 rules in each dimension, and the points and planes in blue are those from the three permutations of the rule $Q_3 \otimes Q_3 \otimes Q_1$. The black points are the two non-redundant points from permutations of the $Q_3 \otimes Q_1 \otimes Q_1$ rules.

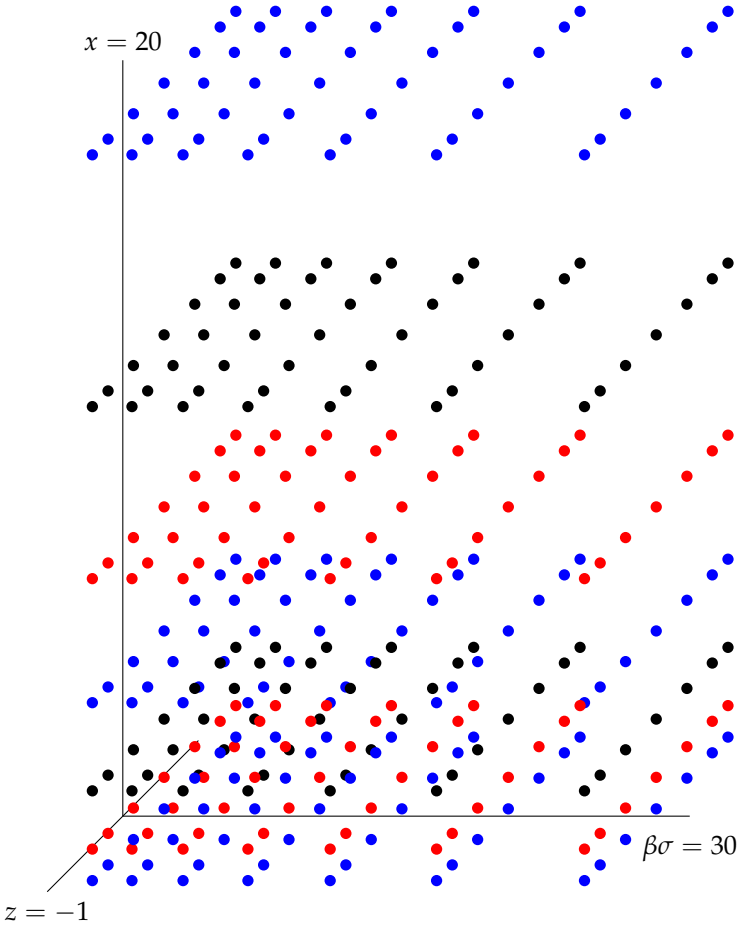


Figure 22: The points from the $Q_7 \otimes Q_7 \otimes Q_7$ tensor-product quadrature using the points from Table 21. The different x levels are colored to distinguish them in the 2-D projection.

Using the sparse quadrature we look at calculating the expansion coefficients for the Black-Scholes example in Figure 23. In these results we see that the $l = 2$ rule exactly integrates the 1-D polynomials up to order 2. The $l = 3$ rule is accurate for the monic polynomials up to degree 4. The mixed degree polynomials are less accurate at $l = 3$, as observed in the polynomials with maximum order 1-3. At $l = 4$ the coefficients are as accurate as the tensor-product quadrature set.

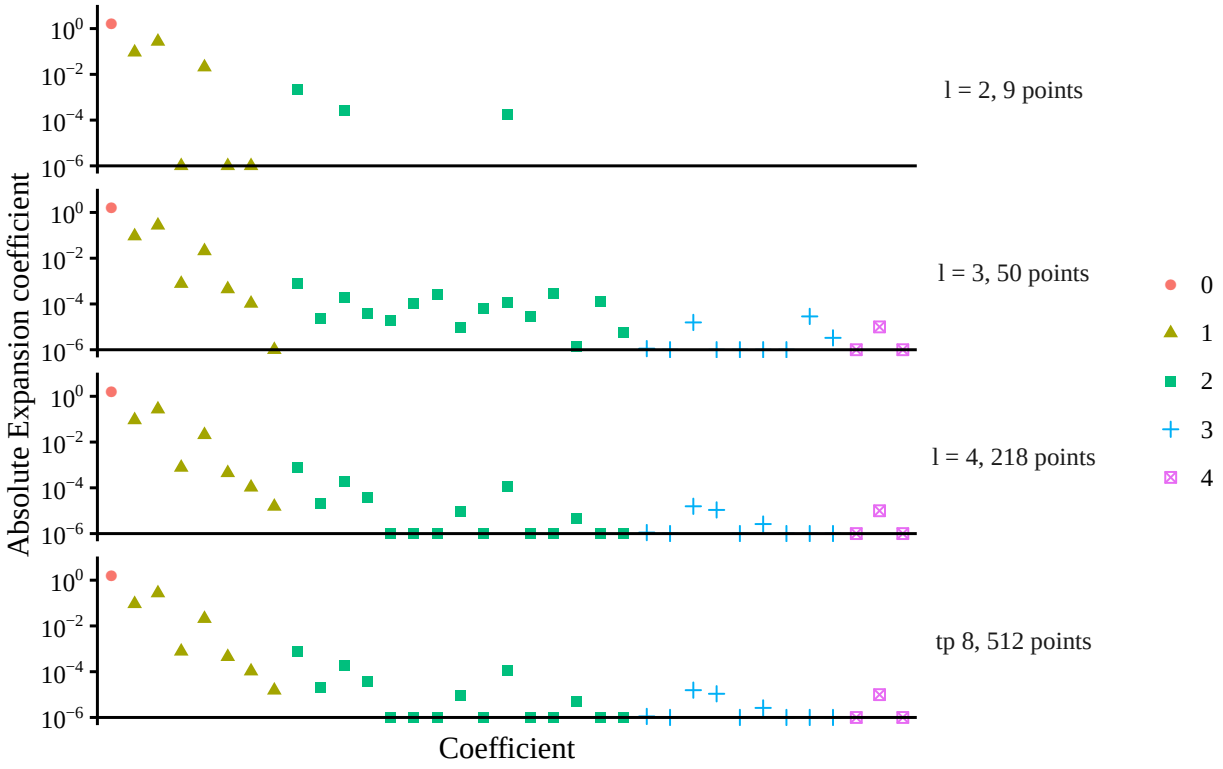


Figure 23: The magnitude of the coefficients in the expansion of the value of a call option as a function of three uncertain parameters, $r = 0.0048x$ with $X \sim \mathcal{G}(0,1)$, $Q \sim \mathcal{U}(0.025, 0.045)$, and $\Sigma \sim \mathcal{G}(5.46636, 41.8142)$. The color and shape of the points indicates the maximum polynomial degree that the coefficient responds to, e.g., c_{011} would be a "1" in the figure. The different panels on the figure indicate the level l, of the sparse Gauss quadrature followed by the total number of points in the quadrature. The bottom panel shows the coefficients calculated with a tensor-product quadrature rule with 8 points in the 1-D quadrature rules. Those coefficients with a total polynomial degree greater than $(2^l - 1)/2$ are not shown, and the coefficients are "floored" to a minimum of 10^{-6}.

Extensions to Sparse Quadratures

The Smolyak sparse quadrature addresses the problem of the number of quadrature points growing exponentially with the number of dimensions and leads to polynomial growth in the number of quadrature points. It does not, however, address the issue regarding how many poiints will be needed in any single dimension. In fact, our Black-Scholes example indicates that the volatility variable should require more points than the other two. One way to accomplish this is to use a weighted quadrature.

Anisotropic quadratures are a way to handle integrals that require more accuracy in a given dimension. A simple way of doing this is to introduce a weight into the selection of quadrature rules. This makes Eq. (**??**)

$$S^{(d)}_{\ell,\vec{a}} f = \sum_{q=\ell-d}^{\ell-1} (-1)^{\ell-1-q} \binom{d-1}{\ell-1-q} \sum_{q+d-1<\|\vec{k}\|_{\vec{a}} \leq q+d} Q_{2^{k_1}-1} \otimes \cdots \otimes Q_{2^{k_d}-1} f, \tag{114}$$

where $\vec{a}$ is a d-length vector of weights, and $\|\vec{k}\|_{\vec{a}} = \sum_{i=1}^{d} |a_i k_i|$.

As an example, if $\vec{a} = (1, 0.5)$, then the $\ell = 3$ quadrature rule with $d = 2$ would be

$$\begin{aligned} S^{(d)}_{\ell,(1,0.5)} f = \\ - Q_1 \otimes Q_7 f - Q_1 \otimes Q_{15} f - Q_3 \otimes Q_3 f - Q_3 \otimes Q_1 f + \\ Q_3 \otimes Q_{15} f + Q_3 \otimes Q_7 f + Q_1 \otimes Q_{31} f + Q_7 \otimes Q_3 f + Q_7 \otimes Q_1 f \end{aligned} \tag{115}$$

This rule has a maximum of 31 points in one direction and 7 in the other dimension.

Another possible extension is to make the quadrature adaptive in each dimension to try and automatically determine which direction to add more points in. In such a procedure we would compute a quadrature rule as

$$A^d f = \sum_{\vec{k} \in I} \Delta_{2^{k_1}-1} \otimes \cdots \otimes \Delta_{2^{k_d}-1} f, \tag{116}$$

where I is the set of all indices included in the rule. The adaptive algorithm starts with $I = \{(1, \cdots, 1)\}$. Then, we add a point to I with an additional level in the dimension with the largest value of the tensor product of Δ quadratures because the magnitude of a Δ quadrature indicates how much the integral changes when adding new points. Then an additional level in the direction of the level just added. The rule grows by considering those tensor products that are adjacent to terms already in the set.

Figure 24 demonstrates an example of an adaptive quadrature rule in two dimensions. We start with $\Delta_1 \otimes \Delta 1$ being the only member of I. Then we compute $\Delta_1 \otimes \Delta_2 f$ and $\Delta_2 \otimes \Delta_1 f$; these are the hatched blocks in part (a) of Figure 24. In the figure the magnitude of $\Delta_1 \otimes \Delta_2 f$ is larger, so it is added to I, and then $\Delta_1 \otimes \Delta_3 f$ is computed, c.f. part (b). Then $\Delta_2 \otimes \Delta_1 f > \Delta_1 \otimes \Delta_3 f$, so $\Delta_2 \otimes \Delta_1 f$ is added to the set in part (c). Now to proceed $\Delta_2 \otimes \Delta_2 f$ and $\Delta_3 \otimes \Delta_1 f$ are calculated. Finally, in part (d) $\Delta_3 \otimes \Delta_1 f$ is added to I, and $\Delta_4 \otimes \Delta_1 f$ is computed. This process will continue until some stopping criterion is reached, such as the maximum magnitude of the $\Delta_i \otimes \Delta_j f$ under consideration is

smaller than some threshold. Once this stopping criteria is reached, all the hatched blocks computed can be included in the set because the work of including them has already been done.

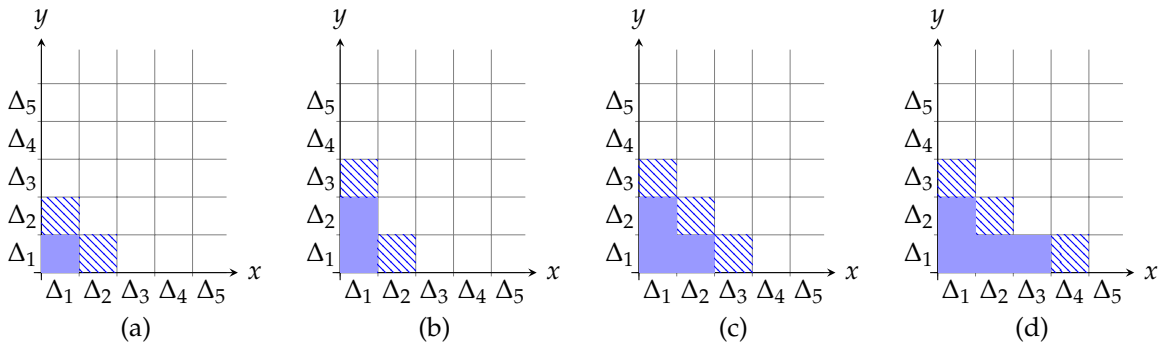


Figure 24: Demonstration of an adaptive sparse quadrature rule. The set is built from left to right and the solid blocks are the elements of the quadrature rule, and the hatched blocks are the new tensor products to be considered. The first step in the cycle is (a), and based on the results the quadrature rule grows to those seen in (b) through (d).

Estimating Expansions Using Regularized Regression

The approximation of a quantity using polynomial chaos can be constructed using other means than quadrature. One possibility is to think of the expansion as a function to be estimated via regression and use regularized regression techniques to minimize the number of function evaluations needed.

To set the stage for this approach let us consider an Nth order Hermite expansion of a function $g(x)$,

$$g(x) \approx \sum_{n=0}^{N} c_n He_n(x).$$

Now consider that we have evaluated $g(x)$ at M values of x. The resulting data gives us the following system of equations

$$\begin{aligned}
g(x_1) &= c_0 He_0(x_1) + c_1 He_1(x_1) + \cdots + c_n He_n(x_1) + \epsilon_1, \\
g(x_2) &= c_0 He_0(x_2) + c_1 He_1(x_2) + \cdots + c_n He_n(x_2) + \epsilon_2, \\
&\vdots \\
g(x_M) &= c_0 He_0(x_M) + c_1 He_1(x_M) + \cdots + c_n He_n(x_M) + \epsilon_M.
\end{aligned}$$

Here we have written the expansion error for each case as ϵ_i. This system is M equations for $N+1$ unknowns, the c_n coefficients, and therefore has no unique solution unless $M = N+1$. We can write this system using rectangular matrices as

$$\mathbf{y} = \mathbf{A}\mathbf{c},$$

where $\mathbf{y}$ is the vector of length M that contains the $g(x_i)$, $\mathbf{A}$ is the $M \times (N+1)$ matrix of the Hermite functions evaluated at x_i, and $\mathbf{c}$ is a vector of length $(N+1)$ for the unknown c_i coefficients.

One could use standard least squares regression to estimate the coefficients by multiplying the equation on both sides by $\mathbf{A}^\mathrm{t}$ and then solving to get

$$\mathbf{c}_\mathrm{ls} = (\mathbf{A}^\mathrm{t}\mathbf{A})^{-1}\mathbf{A}^\mathrm{t}\mathbf{y}. \tag{117}$$

That is, if the inverse of $(\mathbf{A}^\mathrm{t}\mathbf{A})$ exists, which it will not under many conditions, including when $M < N+1$. Therefore, using least squares

will not necessarily save us the growth in function evaluations as N increases.

The least squares solution is the minimizer of the sum of the squares of the residuals, ϵ_i. That is, least squares minimizes the function

$$J_{\text{ls}}(\mathbf{c}) = \frac{1}{2}\sum_{i=1}^{M}(y_i - \mathbf{a}_i\mathbf{c})^2, \tag{118}$$

where a_i is the ith row of $\mathbf{A}$. The minimizer of J_{ls} can be found by taking the derivative with respect to $\mathbf{c}$ and setting the result to zero, to get

$$\sum_{i=1}^{M}\mathbf{a}_i^{\text{t}}(y_i - \mathbf{a}_i\mathbf{c}) = 0.$$

This can be re-written in matrix form as,

$$\mathbf{A}^{\text{t}}\text{y} = \mathbf{A}^{\text{t}}\mathbf{A}\mathbf{c},$$

the solution of which is $\mathbf{c}_{\text{ls}}$.

The idea behind regularized regression is to modify the least-squares problem to make it possible to solve the minimization problem if the inverse of $(\mathbf{A}^{\text{t}}\mathbf{A})$ does not exist. In particular, we will be interested in regularizations where we minimize the magnitude of the coefficients c_i. This is because, as we have seen, that many of the coefficients can be small as in the Black-Scholes example or the coefficients decay to zero as $N \to \infty$. A common way to regularize the least squares minimization is to add a penalty term corresponding to the norm of the coefficients

$$J_{\text{el}} = \frac{1}{2}\sum_{i=0}^{M}(y_i - \mathbf{a}_i\mathbf{c})^2 + \lambda\sum_{n=1}^{N}|c_n| + \lambda\frac{(1-\alpha)}{2}\sum_{n=1}^{N}c_n^2.$$

The minimizer of this function is called the elastic net minimizer. The parameter $\lambda > 0$ controls the amount of penalty for large coefficients, and $\alpha \in [0, 1]$ determines how to balance the L_1 (i.e., the sum of absolute values) and the L_2 (i.e., the sum of squares) penalties. Choosing $\alpha = 1$ only considers the L_1 penalty and is called the lasso penalty, and $\alpha = 0$ is called the ridge penalty.

The goal of elastic net is to find estimates of the expansion parameters that both approximate the values of $g(x)$ in the data and constrict the coefficients in the expansion. This will work well if many of the correct coefficients are small. The balance between correctly matching the data and making the coefficients small is struck through the parameter λ. Ideally, we want λ to be as small as possible. To obtain λ we typically perform a cross-validation procedure: use a subset of the data (called training data) to construct an elastic net fit with a given value of λ and then test the fit on the data *not*

used to create the fit (called the testing data). The smallest value of λ that has an acceptable error on the test data is used for the fit.

To construct the elastic net fit, one can use the implementation for `R`, called `glmnet`, or using `Scikit-Learn` for `python`. `Matlab` has a function for lasso and ridge penalties.