Submitted version of article published in GARP Risk Review with additional section to prove high-order accuracy of Gauss-Hermite Quadrature

Gauss-Hermite Quadrature in Financial Risk Analysis

Joe Pimbley*

Introduction

Financial risk analysis often focuses on calculating the probability of loss or expected loss of a given risky transaction or portfolio of transactions. In structured finance applications, these calculations may include the presence of a loss buffer (otherwise known as "equity") when the calculations pertain to a risk position that is not the first loss. The analyst compiles the PDF (probability density function) for the loss process and integrates this PDF with the desired risk measure to give a loss probability or expected loss with or without the presence of an equity loss buffer.

While this process may seem straightforward, there are many, many practical obstacles. The analyst must make many simplifying assumptions to the point that the problem that is "solved" may bear little resemblance to the true problem. The greatest challenge, then, is to possess the judgment to interpret what meaning the solution to the approximate problem has for the true problem. Two lesser but significant problems are solution complexity (including calculation time) and difficulty of explaining results (and uncertainty of results) to colleagues.

We've stumbled across Gauss-Hermite Quadrature (GHQ) as a contribution to these two "lesser but significant problems" of financial analysis. GHQ permits us to use traditional "risk management stress tests" to determine approximations of the PDF integrations necessary for calculating the various default probabilities and expected losses of financial risk analysis. GHQ saves calculation time and also performs calculations that are not otherwise practical. Given the similarity to familiar stress tests, it will likely be easier to explain results to a wider audience.

We proceed here in several steps. First, Appendix I introduces Hermite polynomials by discussing their origin and several relevant properties. Next, Appendix II discusses the numerical integration procedure known as Gauss Quadrature. GHQ is simply one specific implementation of Gauss Quadrature. Finally, we apply GHQ to a few typical financial risk analyses. It is <u>not</u> necessary to read and understand Appendices I and II to use this GHQ technique. These appendices derive the equation (1) approximation of the next section and discuss its accuracy.

* Mr. Pimbley is Executive Vice President, Head of Institutional Risk, at ACA Capital in New York. He acknowledges gratefully the assistance of ACA colleague Eduardo Robinovich for the calculations of the structured finance example in this article.

Risk Analysis Examples

The material in the appendices is not new. In fact, it was all discovered more than 150 years ago.* The reader need only understand equation (A-8) which we reprise (and shorten) here as equation (1):

$$\int_{-\infty}^{+\infty} e^{-x^2}\ f(x)\,dx \approx \sum_{i=0}^{n} w_i\ f(x_i) \quad (1)\ .$$

As will become clear shortly, the expression on the left-hand side of (1) will represent a risk measure (such as "expected loss") while the right-hand side is an easily calculable approximation to the risk measure. The analyst chooses the value of *n* for the approximation (with higher values generally giving greater accuracy). The function $f(x)$ is known as are all the w_i (which sum to $\sqrt{\pi}$) and x_i.

Equation (1) looks simple given all the work that went into deriving it. In fact, there's a very simple interpretation of equation (1) that, while wrong, is "usefully wrong". Specifically, it's tempting to consider the x_i as "possible values of *x*" (which they are) and the w_i as associated "probabilities of *x* being equal to x_i". Of course, we'd need to normalize the w_i through division by $\sqrt{\pi}$ for these w_i to look like probabilities (so that they would add to unity). We give the normalized w_i the notation $\hat{w}_i$ with the definition $\hat{w}_i = w_i / \sqrt{\pi}$. With this interpretation, equation (1) tells us to "add up the function evaluations at the points x_i weighted with the 'probabilities' $\hat{w}_i$, and then multiply by $\sqrt{\pi}$, to get the integral result we seek".#

In financial risk analysis, these x_i will be what we normally call "stresses". As a typical example, we may choose five yield curve scenarios: "forward LIBOR"; "forward LIBOR plus or minus one standard deviation"; and "forward LIBOR plus or minus two standard deviations". These "stress scenarios" of the financial risk world are analogous to the five zeroes of the fifth-order Hermite polynomial $H_5(x)$ (which also include zero as the middle value and are otherwise symmetric about zero). We say "analogous" since the actual zeroes of $H_5(x)$ are <u>not</u> $-2, -1, 0, +1, +2$. Rather, they are $-2.02, -0.96, 0, +0.96, +2.02$ (rounded to the nearest one-hundredth). (We find it

---

* Useful references are http://mathworld.wolfram.com/Hermite-GaussQuadrature.html, http://www.sosmath.com/diffeq/series/series06/series06.html, and http://www.maths.ox.ac.uk/~murc/lecture_notes/b5die/index.html .

But let us be clear that the interpretation is wrong. A specific $\hat{w}_i$ does not represent a probability that *x* will have the value x_i. But we'll get a very good approximation for the integral calculation if we make this assumption. There are other examples in modeling of such "usefully wrong" assumptions (*e.g.*, thinking of an electron as a small, negatively charged sphere). There should be a special word to describe them … perhaps "Potemkin probabilities" in this case?

to be a fascinating coincidence that these actual five zeroes are so close to the industry's standard stresses.)

In broad outline, then, our "plan" is to run our familiar stress cases on any risk situation of interest. We'll just adjust the stress cases somewhat (as with the zeroes of $H_5(x)$ above). Then, by assigning the pseudo-probabilities to the stresses (which we'll now call "GHQ stresses"), we'll be able to compute approximate integral risk measures with equation (1). With essentially no additional work, we get more information and we bridge the gap of "stress testing versus probabilistic outcomes".

In the discussion of this section, however, we've forgotten to make one key point. The integral of equation (1) has the term $\exp(-x^2)$. We wanted it there! That's why we're working with Hermite polynomials. This exponential belongs in our analysis as long as the underlying stochastic process is normal (Gaussian) or log-normal. In both of these cases we can re-write our integral risk measures to have the form of equation (1). The great majority of market variables (equities, currencies, interest rates, default swap spreads, *et cetera*) are expressible as normal or log-normal variables. Two exceptions are stochastic processes for bond prices or for default recoveries of bonds and loans. Further, there do exist alternate processes for the variables above that are <u>not</u> normal or log-normal (*e.g.*, Cox-Ingersoll-Ross model for the yield curve). Hence, this project assumes that the analyst is comfortable with normal or log-normal treatments. Our earlier discussion of five yield curve stresses presumes that we will employ a one-factor, log-normal model for movements in the yield curve.

The PDF for a random variable *X* that obeys a normal distribution $g(x)$ is

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(x-\mu)^2/2\sigma^2\right] \quad (2)$$

where μ is expected value of *X* (also written as $E\{X\}$) and σ is the standard deviation of *X*. (Note that we follow the standard convention to use upper case to refer to a random variable and its properties and lower case to refer to specific values as they appear in equation (2).)

As we've said, the random variable *X* obeys a normal distribution. If we're studying a random variable that is log-normal, it's easier to keep using *X* (the normal random variable). For example, a stock price *S* follows a log-normal distribution. We can simply write $S = \mathrm{S}_0 \exp(X)$ and *S* will behave "correctly" while X obeys equation (2) as long as we set appropriate values for μ and σ.

Finally, as a last step before we consider some real problems, let's show how the GHQ approximation will appear. Think of $f(x)$ as the function we will integrate. It may be "loss" or "expected positive value" or something similar. The function $f(x)$ will generally be fairly complex since it will include the value of a derivative or bond or equity which may require a *Monte Carlo* simulation or an *Intex* call. Further, as we've said, the "*x*" is normally distributed. The function $f(x)$ will also convert this normal random variable to a log-normal variant when necessary. With all of this, we find

$$I = \int_{-\infty}^{+\infty} g(x)\, f(x)\, dx \approx \sum_{i=0}^{n} \hat{w}_i \; f\left(\mu + \sqrt{2}\sigma x_i\right) \quad (3)$$

where the summation is over all the zeroes of the desired order of Gauss-Hermite quadrature. Note that we've written equation (3) in terms of the normalized GHQ weights $\hat{w}_i$.

Example: Gain of an Equity Investment

Here are two questions we can answer analytically. What is the probability that an equity will gain value over a given time period (with known expected appreciation rate and volatility)? What is the expected gain for this same equity?

In the notation of equation (2), we can derive

$$\text{Probability of Gain } = 1 - \Phi(-\mu/\sigma) \quad \text{(4a)}$$

$$\text{Expected Gain } = e^{\mu+\sigma^2/2}\left[1 - \Phi\left(\frac{-\mu}{\sigma} - \sigma\right)\right] - \left[1 - \Phi(-\mu/\sigma)\right] \quad \text{(4b)}$$

In equation (4b), we've written the expected gain as a ratio to the original stock price. The expression $\Phi(x)$ is the cumulative normal distribution function. Since we can derive (4a) and (4b), there's no need for an approximation such as GHQ. But it's always helpful to test an approximation method against exact results (when you can find them).

In applying equation (3), the function $f(x)$ is $H(x)$ for the "probability of gain" and $H(x)\left(e^x - 1\right)$ for the "expected gain". This function $H(x)$ is the Heaviside function (also known as the step function) which is zero for *x* less than zero and unity for *x* greater than zero. Hence, we need only use these function evaluations in (3) to get GHQ approximations to (4a) and (4b). Numerical results are interesting! See the table below:

Method		Probability of Gain		Expected Gain
GHQ – 2		0.500		0.114
GHQ – 3		0.833		0.074
GHQ – 4		0.500		0.103
GHQ – 5		0.767		0.082
Exact		0.510		0.094

Table I

In Table I, the method "GHQ – 2" means the Gauss-Hermite quadrature with two zeroes (values of x_i). Technically, that means we use first-order expansion in Hermite polynomials since the number of zeroes we use is always one greater than the order of the polynomial. Similarly, the other methods have the number of zeroes shown. The "Exact" result at the bottom comes from equations (4a) and (4b) where we've assumed three months for the stock movement, volatility of 40% per annum and expected appreciation rate of 10% per annum.

At first glance, the results are not all that impressive. Looking at probability of gain, the exact result is 0.510 (which means there's a 51% chance the stock value will rise over the three-month period). The approximations GHQ-2 and GHQ-4 are close, but the odd-numbered approximations are not. As we increase the order beyond what is shown in Table I, the agreement will improve (and, we expect, will "converge" to the exact result). Even though this problem is simple, we've deliberately chosen a difficult case for Gauss-Hermite quadrature. The challenge is that the "probability of gain" function $f(x) = H(x)$ is discontinuous very close to one of the zeroes (when there are an odd number of zeroes). The discontinuity slows the convergence.

Agreement in the "Expected Gain" column is much better. The exact result shows that the average gain over the three-month period will be 9.4% (counting any losses as "zero gain"). The pattern is that even-numbered zero approximations are too high while odd-numbered zero approximations are too low. It's likely that the next two approximations would be much closer to the 9.4% exact result. In this case, the "expected gain" function $f(x) = H(x)\left(e^x - 1\right)$ is continuous everywhere (though not differentiable near one of the zeroes of the odd-numbered approximations).

Example: The Cosine and Square Cosine

The previous example in which we calculated the GHQ approximation to an exactly solvable problem is a bit disappointing as we mentioned. So, mostly for fun, let's try another solvable problem in which the integrand is smooth (continuously differentiable). Forgetting finance for the moment, let's consider the two integrals in which we show the exact solutions:

$$\int_{-\infty}^{+\infty} dx\, e^{-x^2} \cos x = \sqrt{\pi}\, e^{-1/4} \approx 1.380388$$

$$\int_{-\infty}^{+\infty} dx\, e^{-x^2} \cos^2 x = \frac{\sqrt{\pi}}{2}\left(1+e^{-1}\right) \approx 1.212252$$

Again, the idea here is that we know the exact values of these two integrals. Let's apply GHQ to the integrals as if we *didn't* know the exact values and see how close the GHQ approximations are:

Method		**Cosine**		**Cosine Square**
GHQ – 2		**1.347498**		**1.024428**
GHQ – 3		**1.382040**		**1.249614**
GHQ – 4		**1.380329**		**1.206943**
GHQ – 5		**1.380390**		**1.212252**
Exact		**1.380388**		**1.212252**

Table II

With only three zeroes (GHQ-3), the approximation is accurate to better than 0.2% for the cosine integrand and to roughly 3% for the cosine square integrand. Further, we see rapid convergence. This behavior is typical of Gauss quadrature for smooth functions.

Example: A Falling Bond Price

We noted earlier that a bond price does not obey a normal or log-normal density function. Yet we can postulate a one-factor yield curve stochastic process that is log-normal. We then compute a bond price based on the changing yield curve.

Consider a simple fixed-coupon with bullet maturity and coupon C. Today's value is par which implies that the discount factor weighted value of today's forward curve is also C. We seek to approximate both the expected loss and the expected value of this bond value at one year in the future. Again, a gain on the bond is considered "zero loss" for purposes of "expected loss". Further, let's add that the remaining maturity in one year will be T and that the forward T-year rate is F_0.

While this problem is "simple", the concepts and algebra soon run deep. As a function of the normally distributed variable X, the bond value in one year is

$$\text{Par} + \text{Par}\left(C_0 - F_0 e^x\right) WAD(x) \quad (5)$$

where $WAD(x)$ is the remaining weighted average duration. The ubiquitous "duration approximation" treats this duration as independent of x (the future yield curve). We will retain the yield curve dependence.

Applying equation (3) to this relationship between bond price and the normal variable X, we find the following Gauss-Hermite approximations for the bond expected value and expected loss:

Method		**Expected Value**		**Expected Loss**
GHQ – 2		**0.99233**		**0.0263**
GHQ – 3		**0.99234**		**0.0183**
GHQ – 4		**0.99234**		**0.0238**
GHQ – 5		**0.99234**		**0.0200**
GHQ – 18		**0.99234**		**0.0222**
GHQ – 19		**0.99234**		**0.0216**

Table III

To generate these results, we assigned these values: coupon C is 5.0%; forward T-year rate F_0 is 5.2%; yield curve volatility is 20%; and the remaining time T is 5 years. We've also simplified the $WAD(x)$ functional form. We apply a continuous time framework with the discount factor equal to $F_0 e^x$. In "real life", we solve this problem more appropriately with the true payment dates (in lieu of a continuous time assumption) and build the necessary discount factors at these payment dates with the common bootstrapping technique.

The expected loss converges slowly while the expected bond price reaches the precision we show with only three zeroes (GHQ-3). The expected value is less than par since we've stipulated that F_0 is greater than the coupon C (which is typical).

Example: Structured Transaction with Simulated Cash Flows

Consider typical RMBS (residential mortgage-backed securities) investments. RMBS are a subset of ABS (asset-backed securities) which form yet a further subset of "structured finance". In RMBS, a pool of assets consists of several thousand mortgage loans to homeowners. Investors in RMBS provide the cash that ultimately goes to the homeowners for house purchases. These investors choose a desired "tranche" within the capital structure from the most senior (low risk and low return) to the most junior (high risk and high potential return).

Modeling the performance of these investments is notoriously complex. The primary parameters that impact investment outcome are level of interest rates (*i.e.*, the yield curve), default losses of the mortgage loans, and prepayment rates of the mortgage loans. The manner in which any of these three variables affects tranche performance is convoluted. Depending on further critical details, for example, rising interest rates can help or hurt the investment tranches. Analysts must run obscure cash flow models to determine expected tranche performance. These models are invariably "scenario models" in that the user specifies the future yield curve, default loss, and prepayment assumptions.

One persistent problem with scenario models, as we discussed early in this article, is that the user gets no information on the probability that any particular scenario occurs. Further, we have three critical variables which means the user must specify scenarios for three variables simultaneously. Hence, this complex problem is a candidate for GHQ analysis. We must say immediately, though, that GHQ can only alleviate the yield curve contribution. Plausible stochastic processes for mortgage default losses and prepayments are not yet available while the one-factor, log-normal process for the yield curve fits GHQ well. The application of GHQ, then, reduces the analysis to that of specifying scenarios for just the two remaining variables (default losses and prepayments) and using apparent yield curve stresses to give integrated performance data (such as expected return or tranche default probability) for the scenarios of default losses and prepayments.

As an example, consider a senior tranche that pays a floating coupon of L + 40 bps pa. The best outcome is that the investor will receive this full coupon and all principal as the tranche pays down. In this case, the DM ("discount margin") is 40 bps pa. When we choose "base" case stress values for default losses and prepayments, our cash flow model finds that the senior tranche loses no principal and pays the full coupon under six of the nine GHQ yield curve stress points when we choose the ninth-order Hermite polynomial approximation. The remaining three stress points correspond to rising interest rates. The x_i values for the less-than-full-repayment scenarios are approximately 1.5, 2.3, and 3.2. As equation (3) shows, we multiply these values by $\sqrt{2}$ to determine the yield curve stress (so that the respective stresses are 2.1, 3.2, and 4.5 multiplied by the yield curve volatility).

With the base case stresses for default losses and prepayments, the GHQ analysis shows the expected DM to be 38 bps pa. When we choose an "extreme" prepayment stress of half the base case (*i.e.*, prepayments occur at half the expected rate), the expected DM falls to negative 14 bps pa. The point is that GHQ permits us to remove yield curve assumptions from our scenario-based cash flow analysis.

Summary

There are two distinct "ways to think" in risk management. The risk analyst can apply "stress tests" to a firm's risk positions and impose limits on the variability of portfolio value under these stresses. Or the analyst can employ a stochastic methodology (*e.g.*, *Monte Carlo* simulation) to monitor probability of loss (or expected loss).

In many cases, the ancient mathematical integration technique known as Gauss-Hermite Quadrature (GHQ) blends these two risk management viewpoints. GHQ transforms an integral risk measure of a stochastic distribution into (what appears to be) a conventional stress test. At first recognition, the result is incredible! This article provides the derivation (primarily in the appendices) and gives several examples of the application.

Appendix I: Hermite Polynomials

Consider the linear, second-order ordinary differential equation:

$$y'' - 2xy' + 2ny = 0 \quad \text{(A-1)}$$

where n is a non-negative integer and y' and y'' represent the first and second derivatives, respectively, of y with respect to x. By searching for series solutions for y as a function of x, one finds a solution for each n that is a polynomial of order n.[#] For example, if $n = 1$, we see by inspection that $y = 2x$ solves equation (A-1). Of course, $y = 2x$ is a polynomial of order 1. More generally, let's call the solution for each non-negative integer n $y = H_n(x)$. We've already seen that $H_1(x) = 2x$. We list the first few Hermite polynomials below:

$$H_0(x) = 1$$

$$H_1(x) = 2x$$

$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$

$$H_4(x) = 16x^4 - 48x^2 + 12 \quad \text{(A-2)} .$$

The reader may notice that all the polynomials here would remain solutions of (A-1) after multiplication by any constant. The list we provide is the common convention for choosing the constants to satisfy recurrence relations that facilitate analysis. The leading term of $H_n(x)$ will always be $2^n x^n$.

In (A-2) we listed just the first five Hermite polynomials. Think of the infinite sequence of Hermite polynomials as we let n go from zero to infinity. The functions have the critically important properties that they are orthogonal and complete on the infinite interval $(-\infty, +\infty)$ with the weight function $\exp(-x^2)$. This orthogonality means that

$$\int_{-\infty}^{+\infty} e^{-x^2} H_m(x) H_n(x)\, dx = 0, \; m \neq n \quad \text{(A-3)} .$$

[#] Equation (1) has two linearly independent solutions. We are interested only in the solution that we can represent as a finite series. The solution that we ignore remains an infinite series.

As a sketch of the proof of orthogonality, we would re-write (1) as

$$\left(e^{-x^2}\ y'\right)' + 2n\,e^{-x^2}\ y\ =\ 0 \quad \text{(A-4)}$$

and then evaluate the integral

$$\int_{-\infty}^{+\infty}\left(e^{-x^2}\ y_n'\right)'\ y_m\ dx$$

twice by parts. With the integration-by-parts result and equation (A-4), we prove (A-3).

The completeness of the $H_n(x)$ means that any well-behaved function* $f(x)$ can be expressed as a linear combination of the $H_n(x)$. That is, there exist coefficients C_n such that

$$f(x) = \sum_{n=0}^{\infty} C_n\ H_n(x)\ \ . \quad \text{(A-5)}$$

We've now exhausted what we need to know about Hermite polynomials. But let's ponder equation (A-5) before we leave. This equation gives us an exact representation of $f(x)$ as an infinite series of Hermite polynomials. We will use this representation later but will find that we can't use all the (infinite number of) terms. Rather, we may use only the first five or ten or twenty. These <u>finite</u> series will be approximations of $f(x)$.

* In this discussion, "well behaved" means that $f(x)$ is continuous everywhere except at isolated points where it is permitted to have jump discontinuities.

Appendix II: Gauss Quadrature

Forget Hermite polynomials for a few minutes. The word “quadrature” is a synonym for “numerical integration”. If we need to know the integral from zero to four of $f(x) = x^2$, we simply apply the anti-derivative $x^3/3$ and evaluate this entity at the two endpoints to give $64/3$. But if we need to integrate the more difficult function $f(x) = \exp\left(1/_{x+1}\right)$, we must estimate the result numerically. Please remark the word “estimate”. All numerical integrations are approximations. The goal of numerical integration is to achieve a desired accuracy with a good estimate of this accuracy for the least amount of computation.

Possibly the simplest form of numerical integration is to look at the mid-point of the interval ($x = 2$ in this example where we integrate over the interval $x \in (0,4)$) and approximate the integral as four (the length of the interval) times $f(2)$ (the function evaluated at the mid-point). (This result is 16 and is not a good approximation for $64/3$.) The next simplest method is the Trapezoidal Rule which evaluates the function $f(x)$ at the two endpoints (rather than at just the mid-point). (This Trapezoidal Rule result is 32 and is also a bad approximation.) With computer programs it is elementary to divide any given interval (such as $x \in (0,4)$) into many smaller intervals and apply a method such as the Trapezoidal Rule on the smaller intervals to get better accuracy.

We must say, though, that writing a program to apply the Trapezoidal Rule to ever smaller intervals is highly inelegant. It’s crude and barbaric. You’ll have many brilliant minds of the past three hundreds years spinning in their graves. Before computers, research and discovery of improved numerical integration methods was highly important. Hence, there are many known techniques superior to the Trapezoidal Rule and we will jump to the method that may be the most brilliant.

Gauss Quadrature begins with the specification of an infinite, complete, orthogonal sequence of polynomials in which the domain of the polynomials matches the desired integration interval and the orthogonality condition is consistent with the desired integrand. We choose the Hermite polynomials since they suit our financial risk applications well in terms of both the integration interval and desired integrand. (As the reader may have guessed, Gauss Quadrature with Hermite polynomials is “Gauss-Hermite Quadrature”.) More specifically, we will want to approximate integrals of the form

$$\int_{-\infty}^{+\infty} e^{-x^2} \, f(x)\, dx$$

in which the e^{-x^2} term denotes either the normal or log-normal probability density functions.

Let's call this integral I and use equations (A-5) and (A-3) of this memorandum:

$$I = \int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx$$

$$= \int_{-\infty}^{+\infty} e^{-x^2} \sum_{n=0}^{\infty} C_n H_n(x)\,dx$$

$$= \sum_{n=0}^{\infty} C_n \int_{-\infty}^{+\infty} e^{-x^2} H_n(x)\,dx$$

$$= C_0 \int_{-\infty}^{+\infty} e^{-x^2}\,dx = \sqrt{\pi}\,C_0 \quad \text{(A-6)}\,.$$

This is outstanding! The end result is simple. The desired integral I is just the square root of π multiplied by the coefficient of the zero-th order Hermite polynomial. Hence, all we need to do is get the expansion of the function $f(x)$ in terms of the Hermite polynomials and then use just the zero-th order coefficient C_0. If we can accomplish this task, then equation (A-6) is an <u>exact</u> result. In practice, though, we will expand $f(x)$ into a <u>finite</u> series of Hermite polynomials as an approximation. The approximation improves as we increase the number of terms.

With the approximation

$$f(x) \approx \sum_{i=0}^{n} C_i H_i(x) \quad , \text{(A-7)}$$

we can solve a linear system of equations to get the C_0 of equation (A-6) so that

$$I = \int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx = \sqrt{\pi}\,C_0 \approx \sum_{i=0}^{n} w_i f(x_i) \quad \text{(A-8)}$$

in which the x_i are the $n+1$ zeroes of $H_{n+1}(x)$ and the w_i are the coefficients we construct by solving the linear equations that force equation (A-7) to be exactly correct at the points x_i. Fortunately, both the x_i and the w_i are tabulated. We don't need to re-derive them. The sum of the w_i is $\sqrt{\pi}$. In arriving at (A-8) we have omitted a long discussion of why we choose zeroes of $H_{n+1}(x)$ for the x_i and why the approximation in (A-8) is much more accurate than one would imagine.

Additional Section Excluded from GARP Article to Explain High-Order Accuracy of Gauss-Hermite Quadrature

Let's take an example to make this real. We want to approximate

$$I = \int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx$$

with the first three Hermite polynomials: $H_0(x); H_1(x); H_2(x)$. (See equations (A-2).) To get the "best fit", I will choose three points on the abscissa (*x*-axis) and force the approximation

$$f(x) \approx C_0 H_0(x) + C_1 H_1(x) + C_2 H_2(x) \quad \text{(A-9)}$$

to be exact at these three points (which we'll call $x_0, x_1, \text{and } x_2$). We will choose these three values to be 0, $+\sqrt{3/2}$, and $-\sqrt{3/2}$. These may seem like arbitrary choices, but we'll circle back to this choice in just a minute. For now, note that applying equation (7) to each of the points $x_0, x_1, \text{and } x_2$ gives three linear equations in the three unknowns $C_0, C_1, \text{and } C_2$. It's quite easy to solve these equations since the $H_n(x)$ have even or odd symmetry depending on whether *n* is even or odd and our three points $x_0, x_1, \text{and } x_2$ are symmetric about zero. Further, we only need to get C_0. Of course, this C_0 depends on the function value $f(x)$ at all three points $x_0, x_1, \text{and } x_2$. With minimal algebra, we find

$$I = \int_{-\infty}^{+\infty} e^{-x^2} f(x)\,dx = \sqrt{\pi}\, C_0 = \frac{\sqrt{\pi}}{6}\left[f\left(-\sqrt{3/2}\right) + 4f(0) + f\left(+\sqrt{3/2}\right)\right]$$

Now let's get back to the pressing question of why we choose the $x_0, x_1, \text{and } x_2$ as we did. First, we need three points to find the three unknowns $C_0, C_1, \text{and } C_2$.[@] Second, it is helpful that these points be symmetric about zero. Most fascinating is that these $x_0, x_1, \text{and } x_2$ are the zeroes of the <u>next</u> Hermite polynomial $H_3(x)$. Why would we do that? Why does $H_3(x)$ matter? In short, the answer is we get the "right answer" for C_0 by doing the work for the equation (A-9) approximation with the first three Hermite polynomials $H_0(x); H_1(x); H_2(x)$ while, in reality, extending the approximation to $H_3(x)$. Stated differently, once we find C_0 with the approximation of the first three Hermite polynomials, we know immediately that we'd still get the same

[@] The reader might object that we only care about C_0. But we still count the other two unknowns for purposes of determining how many abscissa values are necessary.

C_0 if we improve the approximation to four Hermite polynomials. The accuracy of our approximation is better than we expect.

The most immediate method to prove this property is to consider the linear system of equations we solved to get $C_0, C_1, \text{and } C_2$. When we keep the three abscissa points $x_0, x_1, \text{and } x_2$ and then add a third point x_3 in order to compute C_3, the matrix of the linear system has three zeroes that tell us immediately that the prior $C_0, C_1, \text{and } C_2$ values are correct and we have just one remaining equation for C_3 in terms of the $C_0, C_1, \text{and } C_2$.

But the story gets even stranger. We just found that an approximation to order *n* is actually correct to order $n+1$ (due to our choice of the zeroes of $H_{n+1}(x)$ for the abscissa points in the order *n* approximation). It's even better than this. We find the correct integral (*i.e.*, C_0) for integrand functions $f(x)$ all the way to order *2n*!

I could not recall the proof of this amazing property nor did I find it in the references I consulted. Yet it forms a key justification for applying Gauss Quadrature methods. So I managed to derive the proof and find it to be both obscure and brilliant. That's a bad combination! Brilliant observations are usually simple in their final form. There's likely a more elegant proof than what I'll describe here.

In determining the weights w_i, we forced the equality

$$\sum_{i=-k}^{+k} w_i \, H_j(x_i) = 0 , \quad j = 1, \cdots, 2k \quad \text{(A-10)}$$

where we assume our approximation is to order *2k*. If this statement is <u>also</u> true for continued higher values of *j*, then the integral approximation will be accurate to these higher orders. When equation (A-10) is satisfied, the value of C_0 is unaffected when the true function $f(x)$ has components of these higher order Hermite polynomials.

But why should equation (A-10) be satisfied by higher values of *j*? Some references argue that the freedom to choose abscissa values (the x_i) provide this flexibility. But that's not a comforting explanation. Further, it seems wrong on its face. After all, we chose the zeroes of $H_{2k+1}(x)$ *en masse*. This doesn't feel like flexibility. First, here's an easy answer. Equation (A-10) is trivially satisfied for $j = 2k+1$ since $H_{2k+1}(x)$ is zero at all the x_i.

The next step, then, is to determine why equation (A-10) should work with $j = 2k+2$. There's an answer! A tremendously valuable recurrence relation for Hermite polynomials is:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x) \quad . \quad \text{(A-11)}$$

We see that $H_{2k+2}(x)$ will satisfy (A-10) if we set $n = 2k+1$ in (A-11), multiply this equation by w_i and perform the indicated summation since the first term on the right-hand side is identically zero and the second term already satisfies (A-11).

But this argument only works once. It doesn't get us to the next even-order Hermite polynomial. If I didn't know the answer (accuracy to twice the *2k*), I'd assume there is no further accuracy. But it exists and there's a proof. The "secret" is that there are generalizations of equation (A-10) that we can prove:

$$\sum_{i=-k}^{+k} w_i \, x_i H_j(x_i) = 0 \,, \quad j = 2, \cdots, 2k \quad ; \text{(A-12a)}$$

$$\sum_{i=-k}^{+k} w_i \, x_i^2 \, H_j(x_i) = 0 \,, \quad j = 3, \cdots, 2k \,. \quad \text{(A-12b)}$$

These and extrapolated relationships (involving higher powers of x_i) make it possible to prove (A-10) for the necessary higher orders.

Let's generalize and conclude. In equation (A-9) we posited an approximation to a function $f(x)$ to order 2 (*i.e.*, to $H_2(x)$). Extending this approximation to order *n* we write

$$f(x) \approx \sum_{i=0}^{n} C_i \, H_i(x) \quad . \quad \text{(A-13)}$$

We then derive the integral approximation

$$I = \int_{-\infty}^{+\infty} e^{-x^2} \, f(x) dx = \sqrt{\pi} \, C_0 = \sum_{i=0}^{n} w_i \, f(x_i) \quad \text{(A-14)}$$

in which the x_i are the $n+1$ zeroes of $H_{n+1}(x)$ and the w_i are the coefficients we construct by solving the linear equations that force equation (A-13) to be exactly correct at the points x_i. Fortunately, both the x_i and the w_i are tabulated. We don't need to re-derive them. The sum of the w_i is $\sqrt{\pi}$.