On a Misuse of Mathematics in Optimal Stopping and Finance Models, and a Misuse of Statistics in Data Analysis

Larry Shepp, U Penn; emeritus at Rutgers U.

Prague, St. Petersberg, and Karelia, 2010

I will first discuss a retirement problem. The problem has implications for optimal stopping (pun intended) theory and also for the downturn of 2008. Joint with Dean Foster, U Penn.

I will then discuss a probability problem which arises in a misapplication of statistics in which independent processes can be made to appear correlated. Joint with Abraham Wyner, U Penn

Part 1. How to Retire Early

Abstract. We pose an optimal control problem arising in a perhaps new model for retirement investing. We seek an investment strategy in which we invest an amount $f(X(t))dt$ if our current fortune is $X(t)$ at any time t. Suppose we have a steady income of dt in each time interval, dt, that we need a fortune of M dollars to retire, and that we choose to model our fortune stream, $X(t), t \geq 0$, by the Ito process, depending on the choice of the function, f,

$$dX(t) = (1 + f(X(t))dt + f(X(t))dW(t), \qquad X(0) = x,$$

where W is standard Brownian.

The problem is to choose $f = f(x)$ so that $V(x; f) = E_x(\tau_M^f)$, is as small as possible, where τ_M^f is the first time, t, that $X(t) = X^f(t) = M$, starting at $X(0) = x$, his initial fortune. We show how to choose an optimal $f = f_0$ and find an explicit formula for $V(x) = V(x; f_0)$, and show that the choice of f_0 is optimal among all nonanticipative investment strategies, not just among Markovian ones. We also consider a more general case where the diffusion coefficient, f, is replaced by Af^{α}. The general case reveals that the optimal investment strategy as well as the expected time until retirement are very strongly dependent in interesting and insightful ways on the particular model used.

The problem is to find a nonanticipating process, $f(t), t \geq 0$, so that if τ_M^f is the hitting time of M of the Ito process $X(t) = X^f(t), t \geq 0$, with $X(0) = x$, and

$$dX(t) = (1 + f(t))dt + f(t)dW(t),$$

then $V(x; f) = E_x \tau_M^f$ is a minimum over all such allowable f. Note that we are assuming that the state space for this optimal control problem is the right half line so that we do not allow negative values of $X(t)$. If $X(t) < 0$ for some $t < \tau_M^f$, which is possible if $f(t)$ is bounded away from zero and also bounded, then there is a need to define what happensif the investor is in debt; we assume the game is over in this case and then $\tau_M^f = \infty$ so that with our definition, we do not even achieve afinite expectation, much less a minimum. Other definitions, for example, some rule for borrowing additional capital, are possible to consider, but our model assumes that we are extremely adverse to being in debt.

We will show that under this assumption, the optimum control, f_0, exists and is unique. Any reasonable person would guess that $f(t) = f(X(t))$, i.e., that the optimal f is "Markovian", i.e., the optimal strategy depends only on the present fortune. But even if we guess that f should be Markovian, how do we learn which particular f is best? There is a nice way, involving a lot of nice guessing. Once one guesses f the proof that it is optimal is routine crank-turning, by martingale theory as we will see.

To get lower bounds on $V(x) = \inf_f V(x; f)$, one needs to find, in the usual way, a function, $\bar{V}(x)$, with $\bar{V}(M) = 0$, for which, for any f, the process, $Y(t) = t + \bar{V}(X^f(t))$ is a submartingale. If this is the case, then, we have from optional sampling, $E_x Y(\tau_M) \geq Y(0)$. This gives that for any f and $0 \leq x \leq M$,

$E_x \tau_M^f = EY(\tau_M^f) \geq Y(0) = \bar{V}(x)$,

and since this holds for any f and $0 \leq x \leq M$, we get that $V(x) \geq \bar{V}(x)$.

Equality will hold for all x, for the greatest lower bound, $\bar{V}$. The class of all such $\bar{V}$'s is a convex class determined by the Ito inequalities defining a submartingale, which are that $\bar{V} \geq 0$, and that for all x and all f,
$E[dY(t)|\mathcal{F}_t] = \bar{V}'(x)(1+f)dt + \frac{f^2}{2}\bar{V}''(x)dt + dt \geq 0.$
Since this must hold for all choices of f and all choices of $x = X(t)$ in $[0, \infty)$, and since this is quadratic in the real variable f (if this seems somewhat aggressive with respect to logic, recall that we are just using this reasoning for **guessing** the right $\bar{V}$).
For any such $\bar{V}$, we have that for any f, $V(x; f) \geq \bar{V}(x)$, which gives us the lower bound, $\bar{V}(x)$ on $V(x)$. Which $\bar{V}$'s satisfy the above submartingale condition?

Setting the derivative wrt. f equal to zero we see that we must have for each x, $\bar{V}''(x) > 0$, and then the minimum occurs at $f = f(x) = -\frac{\bar{V}'(x)}{\bar{V}''(x)}$. Putting this f back into the submartingale inequality we need, cancelling a term, that

$1 + \bar{V}'(x) - \frac{1}{2}\frac{(\bar{V}')^2(x)}{\bar{V}''(x)} \geq 0$.

For the best f, we need equality to hold everywhere in the string of inequalities above so that we would choose $\bar{V}$ to satisfy the last inequality with equality throughout. If we set

$g(x) = -\bar{V}'(x) \geq 0$,

then we seek g to satisfy

$g'(x)(\frac{1}{g(x)} - \frac{1}{g^2(x)}) \equiv -\frac{1}{2}$.

Integrating, we have for some integration constant, c,

$\frac{1}{g(x)} + \log g(x) = \frac{x+c}{2}$.

Since the left side is of the form $\frac{1}{y} + \log y \geq 1$ for all $y > 0$, it is tempting to choose $c = 2$ since this makes the right side greater than or equal to one for $x \geq 0$. N.B. It is our privilege to do this, since we are just guessing. We have almost arrived at a guess for the best $\bar{V}$, namely we have to solve the last equation for $g(x) = -\bar{V}'(x)$, and then $\bar{V}$ is determined because we have $\bar{V}(M) = 0$.

A plot of y vs. $\frac{1}{y} + \log y$ is given in Figure 1 which shows that the inverse function defining $g(x)$ by

$\frac{1}{g(x)} + \log(g(x)) = 1 + \frac{x}{2}$,

is not unique since the inverse is not one-one. Which one do we use, the left side branch or the right side branch to define $g(x)$ for each $x \in [0, M]$? Recall that we must have $g'(x) = -\bar{V}''(x) > 0$, so we guess to use the left side to determine $g(x)$. There is then clearly a unique solution, $g(x)$, and we declare this as our guess at $g(x) = -\bar{V}'(x)$.

Using the condition $\hat{V}(M) = 0$, we have
$\hat{V}(x) = -\int_x^M \hat{V}'(u)du = \int_x^M g(u)du$.
We have already set up the proof that this $\hat{V}(x) \equiv V(x)$. We have also seen that any optimal choice of $f = f_0$ must satisfy
$f(x) = -\frac{\bar{V}'(x)}{\bar{V}''(x)}$,
which we can express in terms of the $g(x)$ we have already defined because we have seen that V' can be expressed in terms of g, and so we get that
$f_0(x) = \frac{g(x)}{-g'(x)} = 2\frac{1-g(x)}{g(x)}$.
We note that near $x = 0$, we have $g(x) \sim 1 - \sqrt{x}$, and so it follows that $f_0(x) = 2(g(x) - 1) \sim 2\sqrt{x}$. We note that near $x = \infty$, $g(x) \sim \frac{2}{1+x+\log\frac{x}{2}}$, so that $f_0(x) \sim x + \log\frac{x}{2} - 1$.

To complete that proof that this $f = f_0$ is optimal, with this choice of $f = f_0$, and $\bar{V}$, the inequalities now hold for every other choice of f that $Y^f(t) = t + \bar{V}^f(X^f(t))$ is a submartingale. It follows that $V(x) \geq \bar{V}(x)$ for all $x > 0$. Also equality holds for $f = f_0$ given above because in this case the submartingale is a local martingale. We need to show the equality $EY(\tau_M) = Y(0)$ holds, where Y is the process, $Y(t) = t + \bar{V}(X_0^f(t))$. It is enough to prove that $\tau_M < \infty$ w.p. 1. The difficulty is that the process, $X(t) = X^{f_0}(t)$ hits zero uncountably many times with positive probability starting from any $0 \leq x < M$.

How do we know that X cannot take negative values? When $X(t) = 0$, then the unit drift moves it to the right, but how do we know that the term $f_0(X(t))dW(t)$ does not cause the process to reach the negative half-line? Each time the process hits zero, imagine that $f(x)$ is turned off, so that $f(x) = 0$ for $0 \leq x \leq \epsilon$. There is still a unit drift present so that the process takes time ϵ to reach the point $x = \epsilon$. The probability starting at ϵ that the process hits M before it reaches zero again is easily seen to be $1 - c\epsilon$. It follows from this that the expected time to reach M starting from any x is finite and the conclusion follows.

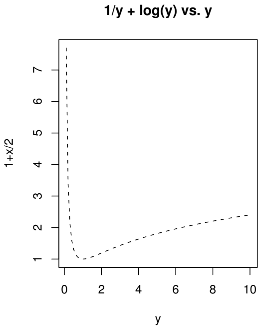
It seems remarkable that the process X^{f_0} behaves as if there is a reflecting barrier at zero. It does not pass through zero to the negative half-axis because $f_0(0) = 0$. This means that it slows down as it gets near zero, but, unlike the Black-Scholes process, it actually hits zero. The drift, $1 + f(0) = 1$, so that it then moves away from zero but it hits zero uncountably many times (if it hits it once), just as the reflecting Brownian process, $|W(t)|$, does, because the set of zeros of $X(t)$ is a perfect set. It is interesting that X reflects off zero even though there is no local time in its Ito representation. This completes the proof that for the optimal investment strategy, $f = f_0$, the fortune of the young man reaches M in a finite time with minimum expected value. It is remarkable that the young man goes broke repeatedly with positive probability before achieving his goal.

Remark. It is often remarked of some rich people that because they were "aggressive, they went into bankruptcy several times before making it". Somehow, mathematics seems to have already been aware of this common observation! Note that it is always true that $V(x) \leq M - x$ since an investor can always choose $f \equiv 0$ and "save the way to retirement".

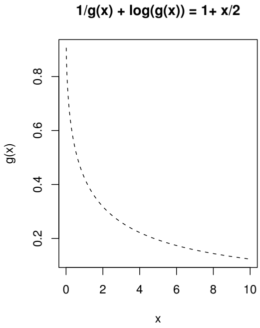
Corollaries of the solution to the problem The solution is not so trivial and illustrates the theme that "one must guess the answer" in optimal stochastic control problems which are convex optimization problems and one needs to find the appropriate extreme point of a certain convex set. Some people deny this and believe instead that the solution, can be found systematically, **in a crank-turning way**, by using the Bellman equation directly, without any guessing. I claim this problem, among others, provides a convincing counterexample. Some people (maybe in this very audience!) also think that in some such problems one can avoid the use of martingales and semimartingale ideas. I claim that for any interesting optimal control problem it is impossible to avoid martingale ideas because somewhere in the proof of optimality use must be made that the strategy is actually **nonanticipating** - since with anticipating strategies one can do strictly better in every problem of any interest.

Remark. The only way to make use of nonanticipatingness is to use the martingale optional sampling theorem; how else can one do it? This is a metamathematical argument but it seems unimpeachable. I will also discuss Ito modeling for pricing derivatives, widely used, which some people, including me, believe played a leading role in the 2008 worldwide economic disaster. There are people in this audience who may dispute this which is of course why I am bringing it up.

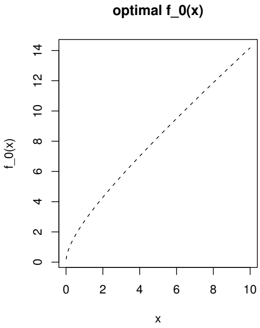
A graph of $g(x) = -V'(x)$ is given in Figure 2, a graph of the optimal payoff, $V(x)$, is given in Figure 3, and a graph of the optimal investment strategy, f_0, is given in Figure 4.



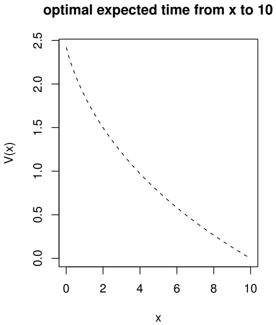
Plot of $\frac{1}{y} + \log y$ vs. y



Plot of $g = g(x)$, $\frac{1}{g} + \log g = 1 + \frac{x}{2}$.



Plot of the optimal $f = f_0$, $f(x) = \frac{g(x)}{g'(x)}$.



Plot of the optimal payoff $V_M^f(x), M = 10, f = f_0$.

Generalization of the problem

A more general model for retirement than the model used so far, namely,

$dX(t) = (1 + f(X(t)))dt + f(X(t))dW(t), X(0) = x,$

would allow the diffusion term to be any fixed function of $f(X(t))$ rather than simply $f(X(t))$ itself. The most natural choice was made above because this is used in the Black-Scholes-Samuelson model for stock prices which was arrived at under the argument that doubling an investment empirically seems to double the volatility, but this is a crude argument and other possibilities seem to be worth exploring. We propose considering the more general model:

$dX(t) = (1 + f(X(t))dt + \phi(f(X(t)))dW(t), X(0) = x,$

where $\phi(u)$ is any increasing function. For tractibility, we will restrict the discussion to the particular forms $\phi(u) = Au^{\alpha}$, where $A > 0$, and $\alpha > 0$ are parameters.

The same method of proof shows that for $\alpha = 1$, as before, but using general A, we have
$g(x; A; \alpha = 1) = -V'(x; A; \alpha = 1) = g(\frac{x}{A^2}; A = 1; \alpha = 1)$.
It follows that
$V(x; A; \alpha = 1) = A^2 V(\frac{x}{A^2}; A = 1; M = \frac{M}{A^2})$.
This is as expected; the original model used f as both the drift and the diffusion parameter because one could scale time to make the diffusion equal to one in appropriate time units. However if one wants to *compare* models, then the parameter A must be retained. If one does this, one sees that
$g(x; A; \alpha = 1) \to g(0; A = 1; \alpha = 1) \equiv 1$
so that $\lim_{A\to\infty} V(x; A; \alpha = 1) = \int_x^M du = M - x$.

The conclusion is that if the investor has the choice of investing in a risky market or instead to be conservative by saving his salary without investing, then the conservative strategy is asymptotically (as $A \to \infty$) superior even though the resulting time to retirement is $M - x$, which is the maximum delay among all models since there is no advantage to investment. The conservative investment advice to avoid risk is usually given to older investors; our modelling assumptions conclusions bear this out, even for young investors in the limit as risk gets very large.

We next consider the case $A = 1$, and $\frac{1}{2} \leq \alpha < 1$. Since the diffusion speed is larger for $\alpha < 1$ than for $\alpha = 1$, one would think that as α decreases the expected time would increase and investing would be disadvantageous, but this is surprisingly not the case, as we see below. Moreover we will show that for $A = 1$, and $0 < \alpha < \frac{1}{2}$, one can find investment strategies that allow retirement in time ϵ, arbitrarily small. Another surprise is that there is a sharp discontinuity in $V(x, \alpha)$ as $\alpha \uparrow \frac{1}{2}$. We show that $V(x, \alpha) = 0$ for $\alpha < \frac{1}{2}$ but as we see below $V(x; \alpha = \frac{1}{2}) = \frac{1}{2}[e^{-2x} - e^{-2M}]$.

Requiring that Y be a martingale for the best choice of $f = f_0$ above gives an ode for $g(x) = g(x, A, \alpha) = -\bar{V}'(x, A, \alpha)$. After a calculation, very similar to the one above for $A = \alpha = 1$, the ode is:
$\frac{-g'(x)(1-g(x))^{2\alpha-1}}{g^{2\alpha}(x)} = \frac{(1-\frac{1}{2\alpha})^{2\alpha-1}}{A^2\alpha}$.
Integrating gives
$\int_{g(x)}^{1} \frac{(1-u)^{2\alpha-1}}{u^{2\alpha}} du = x\frac{(1-\frac{1}{2\alpha})^{2\alpha-1}}{A^2\alpha} + c$.
Again we guess that $c = 0$ and we can solve for $g(x) \in [0, 1]$ for any $x \geq 0$. We see that so long as $\alpha > .5$, there is no trouble. We can write (since $V(M) = 0$),
$V(x) = \int_x^M -V'(u)du = \int_x^M g(u)du = \int_x^M \frac{g(u)}{g'(u)} g'(u)du =$
$b(\alpha) \int_0^{g(x)} (\frac{1-u}{u})^{2\alpha-1} du$,
where $b(\alpha) = \frac{(1-\frac{1}{2\alpha})^{2\alpha-1}}{\alpha A^2}$. Finally we have that $f_0 = f_0(x, A, \alpha)$ is given by
$f_0(x) = (\frac{1}{g(x)} - 1)\frac{1}{1-\frac{1}{2\alpha}}$.

The case $\alpha < \frac{1}{2}$ is especially interesting. The argument given for guessing the optimal $\hat{V}$ breaks down and says there is no $\hat{V}$ that will make $\hat{V}(X^f(t)) + t$ a submartingale for all choices of $f = f(x)$ except the trivial case $\hat{V}(x) \equiv 0$. This is the best lower bound that submartingale theory can provide which lead one to suspect that $V(x;\alpha) \equiv 0$ for $\alpha < \frac{1}{2}$. To prove this we need to find an f that makes the expected time to reach M arbitrarily small. Consider the investment strategy

$f(x) = 0, 0 \le x \le \epsilon, \qquad f(x) = c, \epsilon < x < M.$

If we can find a function, $g = g(x) = g_{\epsilon,c}(x), 0 \le x \le M$ for which $Y(t) = g(X_t) + t$ is a martingale and $g(M) = 0$, then the martingale theorem gives that $g(x) = Y(0) = E_x Y(\tau_M) = E_x \tau_M$. Ito calculus gives that g must be, for appropriate constants, $A', B, D,$

$g(x) = D - x, 0 \le x \le \epsilon,$

$g(x) = \frac{M-x}{1+c} + A'(e^{-\frac{2(1+c)}{c^2\alpha}x} - e^{-\frac{2(1+c)}{c^2\alpha}M}), \epsilon \le x \le M.$

Finally, set $\epsilon = \frac{1}{1+c}$, and let $c \to \infty$ and verify that A' and D tend to zero as $c \to \infty$ and we have shown that $V(x;\alpha) \equiv 0$ for $\alpha < \frac{1}{2}$. Under this model one can retire in arbitrarily small expected time. This should probably be interpreted that the model with $\alpha < \frac{1}{2}$ does not represent the real world; models should be looked at carefully and rejected if they do not conform to reality, unless of course, *reality* is incorrect.

Part 2. How not to do statistics.

The problem is to find the density, $f_\theta(x), -1 \leq x \leq 1$, of the *empirical correlation coefficient*:

$$\theta = \frac{\int_0^1 W_1(t)W_2(t)dt - (\int_0^1 W_1(s)ds)(\int_0^1 W_2(t)dt)}{\sqrt{\int_0^1 W_1^2(t)dt - (\int_0^1 W_1(t)dt)^2}\sqrt{\int_0^1 W_2^2(t)dt - (\int_0^1 W_2(t)dt)^2}},$$

of two *independent* standard Wiener processes, W_1, W_2. A simple Monte Carlo computation allows one to graph the histogram of θ. Our main motivation is to compare what probability theory can contribute to this study vs. simple Monte Carlo. We believed we would be able to calculate $f_\theta(x)$ *to arbitrary precision* and thereby increase the accuracy of the Monte Carlo calculation, however we did not succeed despite much effort. We obtain (with much effort) an explicit formula for the variance of θ. We could derive similar formulas for the higher moments of θ with still more effort. It is frustrating that we cannot obtain a formula for the density of θ, from which one can get arbitrary accuracy.

We can, with a hard calculation, give a Fredholm integral equation for $f_\theta(x), -1 < x < 1$, which has a provably unique solution (because it determines the moments of f_θ, and the moment problem has a unique solution because the support of f_θ is compact). However, the integral equation is of the first, or difficult Fredholm type, and it is not at all clear how to get anything useful from it such as, for example, the asymptotics of $f_\theta(x)$ as $x \to \pm 1$. The Monte Carlo computation indicates that $f_\theta(\pm 1) = 0$, but this may not be true. Thus, our purpose is to draw attention to this problem in the hope that others will be able to salvage something useful from our technique which we believe is new.

Let $W_i(t), 0 \leq t \leq 1, i = 1, 2$ denote two independent standard Brownian motions on $[0, 1]$ and define the empirical correlation between them,

$$\theta = \frac{\int_0^1 W_1(t)W_2(t)dt - (\int_0^1 W_1(s)ds)(\int_0^1 W_2(t)dt)}{\sqrt{\int_0^1 W_1^2(t)dt - (\int_0^1 W_1(t)dt)^2}\sqrt{\int_0^1 W_2^2(t)dt - (\int_0^1 W_2(t)dt)^2}}.$$

One might naively expect that θ would be small because the processes, W_i are independent. However, Brownian motion is *self-correlated*, so a spurious correlation is induced between the independent Brownian motions. It is a frequent mistake to use the correlation of the partial sums of time series as a proxy for the correlation of the time series themselves. This bad practice gives rise to spurious correlation. We seem to be the first to attempt to obtain the actual distribution of the spurious correlation.

We will show that the density of θ satisfies the integral equation

$\int_0^1 dx f_\theta(x) x^2 K(xz) = g(z), 0 \leq z \leq 1,$

where the kernel, $K(x,z) = K(xz)$, and $g = g(z)$ are given by
$K(u) = \sum_{n=1}^{\infty} u^{2(n-1)} \frac{n!(n-1)!2^{2n}}{(2n)!} = 2\,{}_2F_1(1,1;\frac{3}{2},u^2) = 2\frac{\sin^{-1}(u)}{u\sqrt{1-u^2}}$,
and with $v = v(t,z) = \sqrt{t^2 + z^2(1-t^2)}$,
$g(z) = \int_0^1 dt \int_0^\infty du S(u\sqrt{1+v}) S(u\sqrt{1-v}) \frac{u}{v}[T(u\sqrt{1+v}) - T(u\sqrt{1-v})]$,
where $S(u) = \sqrt{\frac{u}{\sinh u}}$ and $T(u) = \frac{1}{u}\frac{S'(u)}{S(u)}$.

This is a Fredholm equation of the "bad" kind, but it is easy to show that it uniquely determines f_θ because K and g are known analytic functions in the unit disk, and so the moments of f_θ can in principle be determined by equating coefficients of powers of z; the moment problem on a bounded interval determines the density. We obtain the following expression for the variance of θ by comparing coefficients of z^2,

$$E\theta^2 = \int_0^\infty du_1 \int_0^\infty du_2 S(u_1)S(u_2)\frac{u_1 u_2}{u_1+u_2}\left(\frac{T(u_1)-T(u_2)}{u_1-u_2}\right).$$

It does not seem possible to do this double integral exactly in elementary terms; but it has removable singularities and converges very nicely at all points where any of u_1, u_2, or $u_1 - u_2$ vanishes. The numerical value is $\mu_2 = .240522\ldots$, according to our calculation. Thus the standard deviation of θ is nearly .5 which is hard to explain if one alleges that the correlation is zero.

There are formulas for all the moments of θ, so we can compute the density, $f_\theta(x)$, in principle, but we have to admit that despite much effort we were unable to get a formula or an algorithm which would allow the calculation of the density of θ to high precision. The spurious correlation is induced because each Wiener process is "self-correlated" in time since a Wiener process is an integral of pure noise and so its values at different time points are correlated. This phenomenon is well-known to probabilists because it is involved in the arc-sine law. Some practitioners of statistics, unaware of the arc-sine laws, overlook the spurious correlation between independent *sum* processes, $S_i, S_i', i = 1, \ldots, n$, which are sums of independent random variables, and erroneously believe that the so-called empirical running correlation,

$$\theta_n = \frac{\sum_{i=1}^n S_i S_i' - \frac{1}{n}(\sum_{i=1}^n S_i)(\sum_{i=1}^n S_i')}{\sqrt{\sum_{i=1}^n S_i^2 - \frac{1}{n}(\sum_{i=1}^n S_i)^2}\sqrt{\sum_{i=1}^n (S_i')^2 - \frac{1}{n}(\sum_{i=1}^n S_i')^2}},$$

ought to be small. The first few moments of θ, the limit in law of θ_n, (approximate S_i by $W(\frac{i}{n})\sqrt{n}$ by the central limit theorem) show that θ_n is non-negligible for large n.

It is therefore clear the distribution of θ need not be accurately computed because one should not be using the partial sums as a proxy for the actual time series. Nevertheless this is an nice problem for developing probabilistic methodology.

Getting the distribution of θ.
Jan R. Magnus gave the moments of the ratio of a *pair of quadratic forms* in normal variables by a technique similar to ours. Our paper gives a method for the correlation coefficient which is a ratio involving *three* quadratic forms of normal variables as well as square roots. It seems clear that the "three-form" problem solved here requires a new method.

The empirical correlation is,
$\theta = \frac{\int_0^1 W_1(t)W_2(t)dt - (\int_0^1 W_1(s)ds)(\int_0^1 W_2(t)dt)}{\sqrt{\int_0^1 W_1^2(t)dt - (\int_0^1 W_1(t)dt)^2}\sqrt{\int_0^1 W_2^2(t)dt - (\int_0^1 W_2(t)dt)^2}}$.
Noting that $m_i = \int_0^1 W_i(t)dt, i = 1, 2$ are the empirical mean values,we can also write $\theta = \frac{\int_0^1 (W_1(t) - m_1)(W_2(t) - m_2)dt}{\sqrt{\int_0^1 (W_1(t) - m_1)^2 dt}\sqrt{\int_0^1 (W_2(t) - m_2)^2 dt}}$,
and note that $-1 \leq \theta \leq 1$ by Schwarz's inequality. We will use a special method to obtain the distribution of θ somewhat similar to earlier uses made of a related idea. First note that
$\theta = \frac{X_{1,2}}{\sqrt{X_{1,1}X_{2,2}}}$, where
$X_{1,2} = X_{2,1} = \int_0^1 W_1(t)W_2(t)dt - (\int_0^1 W_1(s)ds)(\int_0^1 W_2(t)dt)$
$X_{i,i} = \int_0^1 W_i^2(t)dt - (\int_0^1 W_i(t)dt)^2, i = 1, 2.$

Note we can write for $i, j \in \{1, 2\}$,
$X_{i,j} = \int_0^1 dW_i(s_1) \int_0^1 dW_j(s_2)[\min(s_1, s_2) - s_1 s_2]$,
because the right side is, interchanging order of integrations,
$\int_0^1 dW_i(s_1) \int_0^1 dW_j(s_2)[\int_0^{\min(s_1,s_2)} dt - (\int_0^{s_1} ds)(\int_0^{s_2} dt)] =$
$= \int_0^1 dt(\int_t^1 dW_i(s_1))(\int_t^1 dW_j(s_2) -$
$(\int_0^1 ds \int_s^1 dW_i(s_1))(\int_0^1 dt \int_t^1 dW_j(s_2))) =$
$= \int_0^1 dt(W_i(1) - W_i(t))(W_j(1) - W_j(t)) - (\int_0^1 (W_i(1) -$
$W_i(s))ds)(\int_0^1 dt(W_j(1) - W_j(t))dt) =$
$= \int_0^1 W_i(t)W_j(t)dt - (\int_0^1 W_i(s)ds)(\int_0^1 W_j(t)dt) = X_{i,j}$.

Calculating $F(\beta_1, \beta_2, a) = Ee^{a\beta_1\beta_2 X_{1,2} - \frac{\beta_1^2}{2} X_{1,1} - \frac{\beta_2^2}{2} X_{2,2}}$

.

Define for $|a| \leq 1, \beta_i \geq 0, i = 1, 2$ the integral
$F(\beta_1, \beta_2, a) = Ee^{a\beta_1\beta_2 X_{1,2} - \frac{\beta_1^2}{2} X_{1,1} - \frac{\beta_2^2}{2} X_{2,2}}$.

We first prove that the expectation is finite under the given conditions. This is because $X_{1,2} = \theta\sqrt{X_{1,1}X_{2,2}}$ and $|\theta| \leq 1$ so the exponent is at most
$-\frac{1}{2}(\beta_1\sqrt{X_{1,1}} - \beta_2\sqrt{X_{2,2}})^2 \leq 0.$
Thus the expectand is bounded by unity, and so for this range, $F(\beta_1, \beta_2, a) \leq 1$. To calculate it, define the kernel, $K_{i_1,i_2}(s_1, s_2) = K(i_1, s_1; i_2, s_2); i_1, i_2 \in \{1, 2\}, s_1, s_2 \in [0, 1]$, and note that $F(\beta_1, \beta_2, a) = Ee^{\frac{1}{2}\sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\int_0^1\int_0^1 K_{i_1,i_2}(s_1,s_2)dW_{i_1}(s_1)dW_{i_2}(s_2)}$,
where $K_{i,i}(s_1, s_2) = -\beta_i^2 M(s_1, s_2), i = 1, 2; s_1, s_2 \in [0, 1]$,
$K_{1,2}(s_1, s_2) = K_{2,1}(s_1, s_2) = a\beta_1\beta_2 M(s_1, s_2), s_1, s_2 \in [0, 1]$, and
$M(s_1, s_2) = \min(s_1, s_2) - s_1 s_2$, for $s_1, s_2 \in [0, 1]$.

The expectation, $F(\beta_1, \beta_2, a)$, is a Gaussian integral and we calculate it in terms of a Fredholm determinant. This determinant is the product of the eigenvalues of $I - K_{i_1,i_2}(s_1, s_2)$, i.e. $det(I - K) = \prod(1 - \alpha)$, where α runs through all the eigenvaluesof the kernel K, and then $F(\beta_1, \beta_2, a) = \frac{1}{\sqrt{det(I-K)}}$. To see this note that if α_n are the eigenvalues of $K_{i_1,i_2}(s_1, s_2)$, and if $\phi_n(i_1, s_1)$ are the corresponding orthonormalized eigenfuctions, then $\sum_{i_1=1}^{2} K_{i_1,i_2}(s_1, s_2)\phi_n(i_1, s_1)ds_1 = \alpha_n\phi_n(i_2, s_2), i_2 \in \{1, 2\}, s_2 \in [0, 1]$,
's theorem states that we may represent K by a series in the complete set of orthonormal eigenfunctions, $K_{i_1,i_2}(s_1, s_2) = \sum_{n=1}^{\infty} \alpha_n\phi_n(i_1, s_1)\phi_n(i_2, s_2); i_1, i_2 \in \{1, 2\}, s_1, s_2 \in [0, 1]$.
There are a discrete set of values α for which there is a nonzero function $\phi(\cdot, \cdot)$ satisfying the above eigenequation. Because of the form of the kernel, K, we guess that the eigenfunctions are of separableform:
$\phi(i, s) = \xi_i\phi(s); i \in \{1, 2\}, s \in [0, 1]$,
where $\phi(s)$ is an eigenfunction of the kernel,
$M = M(s_1, s_2) = \min(s_1, s_2) - s_1s_2$ which is the covariance of
pinned Brownian motion on [0, 1]. The eigenequation for M is

Mercer's theorem states that we may represent K by a series in the complete set of orthonormal eigenfunctions,
$K_{i_1,i_2}(s_1, s_2) = \sum_{n=1}^{\infty} \alpha_n \phi_n(i_1, s_1)\phi_n(i_2, s_2); i_1, i_2 \in \{1, 2\}, s_1, s_2 \in [0, 1]$.
There are a discrete set of values α for which there is a nonzero function $\phi(\cdot, \cdot)$ satisfying the above eigenequation. Because of the form of the kernel, K, we guess that the eigenfunctions are of separableform:
$\phi(i, s) = \xi_i \phi(s); i \in \{1, 2\}, s \in [0, 1]$,
where $\phi(s)$ is an eigenfunction of the kernel,
$M = M(s_1, s_2) = \min(s_1, s_2) - s_1 s_2$ which is the covariance of pinned Brownian motion on $[0, 1]$. The eigenequation for M is
$\int_0^1 M(s_1, s_2)\phi(s_1) ds_1 = \lambda \phi(s_2), s_2 \in [0, 1]$,
and it is easy to verify that the eigenfunctions and eigenvalues of M are $\phi_n(t) = \sqrt{2} \sin(\pi n t); n \geq 1, t \in [0, 1], \lambda_n = \frac{1}{\pi^2 n^2}; n \geq 1$.

We find that
$\phi_n(i_1, s_1) = \xi_{i_1}\phi_n(s_1)$ is an eigenvalue of K with eigenvalue α_n if and only if
$-\beta_2^2\xi_1 + a\beta_1\beta_2\xi_2 = \frac{\alpha_n}{\lambda_n}\xi_1$
$a\beta_1\beta_2\xi_1 - \beta_1^2\xi_2 = \frac{\alpha_n}{\lambda_n}\xi_2$.
There are thus two eigenvalues, $\alpha_n^{\pm}$ for each$\lambda_n = \frac{1}{\pi^2 n^2}$; $n \geq 1$ and ϕ_n. These eigenvalues are $\alpha_n^{\pm} = \lambda_n\frac{-(\beta_1^2+\beta_2^2)\pm\sqrt{(\beta_1^2-\beta_2^2)^2+4a^2\beta_1^2\beta_2^2}}{2}$.
We thus obtain that $det(I-K) = \prod_{\pm}\prod_{n=1}^{\infty}(1-\alpha_n^{\pm}) = \prod_{n=1}^{\infty}(1-\frac{(z^+)^2}{\pi^2 n^2})\prod_{n=1}^{\infty}(1-\frac{(z^-)^2}{\pi^2 n^2}) = \frac{\sin(z^+)}{z^+}\frac{\sin(z^-)}{z^-}$,
$z^{\pm} = \sqrt{\frac{-(\beta_1^2+\beta_2^2)\pm\sqrt{(\beta_1^2-\beta_2^2)^2+4a^2\beta_1^2\beta_2^2}}{2}}$,
where we used the product formula,
$\frac{\sin z}{z} = \prod_{n=1}^{\infty}(1-\frac{z^2}{\pi^2 n^2})$.

Noting that $z^{\pm}$ are purely imaginary, we write $z_{\pm} = ic^{\pm}$, and so $\frac{\sin z}{z} = \frac{\sinh iz}{iz}$, we see that the determinant is simply $det(I - K) = \frac{\sinh c^{+}}{c^{+}} \frac{\sinh c^{-}}{c^{-}}$, and
$c^{\pm} = c^{\pm}(\beta_1, \beta_2, a) = \sqrt{\frac{(\beta_1^2+\beta_2^2) \pm \sqrt{(\beta_1^2-\beta_2^2)^2+4a^2\beta_1^2\beta_2^2}}{2}}$.
We finally obtain $F(\beta_1, \beta_2, a)$, using Mercer's theorem

$$F(\beta_1, \beta_2, a) = \prod_n E e^{\frac{1}{2}[\sum_{i_1=1}^{2} \alpha_n \int_0^1 \phi_n(i_1, s_1) dW(s_{i_1})]^2} = \prod_n (1-\alpha_n)^{-\frac{1}{2}} = \frac{\quad}{\sqrt{\frac{\sinh}{c^{+}}}} \tag{1}$$

since $\sum_{i_1=1}^{2} \xi_{i_1} \phi_n(i_1, s_1) dW_{i_1}(s_1)$ are standard normal and independent.

The special trick to get the integral equation for f_θ

.

If $\gamma_i \geq 1, \beta_i \geq 0, i = 1, 2$, and $|a| < 1$, the quantities,
$F(\sqrt{\gamma_1}\beta_1, \beta_2, \frac{a}{\sqrt{\gamma_1}})$, $F(\beta_1, \sqrt{\gamma_2}\beta_2, \frac{a}{\sqrt{\gamma_2}})$, $F(\sqrt{\gamma_1}\beta_1, \sqrt{\gamma_2}\beta_2, \frac{a}{\sqrt{\gamma_1\gamma_2}})$,
are finite and well-defined, and we may define the quantity,
$G = G(\gamma_1, \gamma_2, a)$, by
$G = \int_0^\infty \frac{d\beta_1}{\beta_1} \int_0^\infty \frac{d\beta_2}{\beta_2}[F(\beta_1, \beta_2, a) - F(\sqrt{\gamma_1}\beta_1, \beta_2, \frac{a}{\sqrt{\gamma_1}}) - F(\beta_1, \sqrt{\gamma_2}\beta_2, \frac{a}{\sqrt{\gamma_2}}) + F(\sqrt{\gamma_1}\beta_1, \sqrt{\gamma_2}\beta_2, \frac{a}{\sqrt{\gamma_1\gamma_2}})]$.
By the above argument, this is equal to
$G(\gamma_1, \gamma_2, a) = \int_0^\infty \frac{d\beta_1}{\beta_1} \int_0^\infty \frac{d\beta_2}{\beta_2} E e^{a\beta_1\beta_2 X_{1,2}} (e^{-\frac{\beta_1^2 X_{1,1}}{2}} - e^{-\frac{\gamma_1\beta_1^2 X_{1,1}}{2}})(e^{-\frac{\beta_2^2 X_{2,2}}{2}} - e^{-\frac{\gamma_2\beta_2^2 X_{2,2}}{2}})$.

The key idea in calculating $G(\gamma_1, \gamma_2, a)$ is to employ integrals of the form $\frac{d\beta_i}{\beta_i}$ because these allow replacing β_i by $\frac{\beta_i}{\sqrt{X_{i,i}}}$ without introducing a Jacobian. Indeed, making these transformations and putting the expectation back inside the integral, we arrive at the following equation for the moment generating function, $\phi_\theta(z) = Ee^{z\theta}$, of θ,

$$G(\gamma_1, \gamma_2, a) = \int_0^\infty \frac{d\beta_1}{\beta_1} \int_0^\infty \frac{d\beta_2}{\beta_2} \phi_\theta(a\beta_1\beta_2)(e^{-\frac{\beta_1^2}{2}} - e^{-\frac{\gamma_1\beta_1^2}{2}})(e^{-\frac{\beta_2^2}{2}} - e^{-\frac{\gamma_2\beta_2^2}{2}} \tag{2}$$

$= \int_0^\infty \frac{d\beta_1}{\beta_1} \int_0^\infty \frac{d\beta_2}{\beta_2} [F(\beta_1, \beta_2, a) - F(\sqrt{\gamma_1}\beta_1, \beta_2, \frac{a}{\sqrt{\gamma_1}}) - F(\beta_1, \sqrt{\gamma_2}\beta_2, \frac{a}{\sqrt{\gamma_2}}) + F(\sqrt{\gamma_1}\beta_1, \sqrt{\gamma_2}\beta_2, \frac{a}{\sqrt{\gamma_1\gamma_2}})].$

We think of the last identity as one determining the density, $f_\theta(x)$, of θ. We put the expression for F into (2) and differentiate both sides of (2) wrt. both γ_1 and γ_2. We obtain, relying on the corollary of Fubini's theorem that states that integration and differentiation wrt. a parameter can be interchanged if the integral of the differentiated integrand converges absolutely, which it does, on both sides, as we will see later, *for certain values of a*:

$$\frac{\partial^2 G(\gamma_1,\gamma_2,a)}{\partial\gamma_1\partial\gamma_2} = \int_0^\infty d\beta_1 \int_0^\infty d\beta_2 \frac{\beta_1\beta_2}{4} Ee^{a\beta_1\beta_2\theta} e^{-\gamma_1\frac{\beta_1^2}{2}} e^{-\gamma_2\frac{\beta_2^2}{2}} =$$

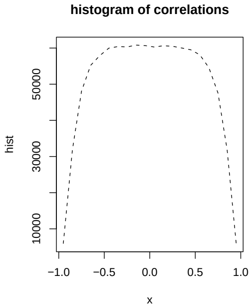
$$= \int_0^\infty \frac{d\beta_1}{\beta_1} \int_0^\infty \frac{d\beta_2}{\beta_2} \frac{\partial^2}{\partial\gamma_1\partial\gamma_2} F(\sqrt{\gamma_1}\beta_1, \sqrt{\gamma_2}\beta_2, \frac{a}{\sqrt{\gamma_1\gamma_2}}).$$

e may write this in the form of an integral equation
$\int_0^1 dxx^2 f_\theta(x)K(xz) = g(z), 0 \leq z \leq 1,$
where $g(z)$ is the right hand side and where
$K(xz) = \sum_{n=1}^{\infty} \frac{(xz)^{2(n-1)}}{2n} \frac{(n!)^2 2^{2n}}{(2n)!}.$
Unfortunately, this is an integral equation of the "bad", or "first" kind; the good kind is of the form $f - Kf = g$. Equations of the bad kind are not discussed much in the integral equation literature because there are severerestrictions on g, which must be smooth enough to be in the range of K. But there is another way to look at our particular one, namely as a set of inequalities:
$\int_0^1 f(x)K(x,z) \leq g(z), 0 \leq z \leq 1,$
if maximize the linear form, $\int_0^1 f(x)dx$ subject to the constraints,we have an infinite dimensional linear program whose answer ought to be$f(x) = f_\theta(x), 0 \leq x \leq 1$ since f_θ satisfies the inequalities and they hold with equality. Moreover, we have seen that any function f which satisfies the inequalities with equality has its moments uniqely determined by them and since we are on a finite interval, the moment problem is unique.

A graph and a conclusion

.

We are not the first to point out the spurious correlation created by using partial sums instead of the actual variates of time series. We seem to be the first to try to obtain the actual distribution of the correlation coefficient by a formula. While approximations can be obtained simple Monte Carlo, it might be useful to have an exact formula and accurate graph of the density of θ. This perhaps has its main interest in pure probability theory and to explore methods of computation. We give a graph of the histogram of one million samples of θ (Figure 1).



Histogram approximation to f_θ

The statistic which uses the *actual random variables*, X_k, X'_k instead of their partial sums, $\sum_{j=1}^{k} X_j, \sum_{j=1}^{k} X'_j$ does not produce spurious correlation. Indeed in the limit, $n \to \infty$,
$$\theta'_n = \frac{\frac{1}{n}\sum_{k=1}^{n} X_k X'_k - (\frac{1}{n}\sum_{k=1}^{n} X_k)(\frac{1}{n}\sum_{k=1}^{n} X'_k)}{(\sqrt{\frac{1}{n}\sum_{k=1}^{n} X_k^2 - (\frac{1}{n}\sum_{k=1}^{n} X_k)^2})(\sqrt{\frac{1}{n}\sum_{k=1}^{n} X_k^2 - (\frac{1}{n}\sum_{k=1}^{n} X_k)^2})},$$
is easily seen to tend to zero, by the law of large numbers, if the r.v's, $X_k, X'_k, k = 1, \ldots$ are iid sequences, independent of each other, with positive finite variances. This shows that the spurious correlation is a consequence of using the partial sums in place of the variables themselves.

The reason that the partial sums are self-correlated and thereby induce spurious correlation seems related to the arcsine law. The history of Sparre Andersen's major combinatorial contribution to the proof of the arcsine law, raises the question of whether a formula can be derived even for discrete sequences of partial sums by similar combinatorial methods employing cyclic permutations. This would be very elegant, but seems unlikely.

The question is why is it so easy to obtain a graph without using any analytical probability theory and so hard to get much further with analytical probability theory?