Draft: Paper presented at the Second International Conference on "Human Dimensions in the Information Age", New Delhi, 8–9 Jan 2001. Won the "Best Paper Award"

Supercomputing in Finance

C. K. Raju

1 Macro-Economic Background

The key point of liberalization and globalization is to permit the unhindered movement of capital across national boundaries. While cross-border movement of people is still subject to controls, so far as movement of capital is concerned, the nation-state has virtually ceased to exist.

There was a time when movement of capital across long distances presented also physical barriers. Funds had to be converted into some internationally acceptable *de facto* standard like gold, the gold was stowed into ships, and the ships set sail for far off lands. This was a time consuming and risky procedure, and many a treasure hunter today still searches for underwater treasures and bullion in sunken ships. However, information technology has changed all that: there are no longer any serious physical barriers to the movement of capital. Funds can now be transferred electronically. Information technology has made long-distance fund transfer secure and rapid. At the beginning of this century, travel round the world in eighty days was an astonishing matter of Wellsian science fiction. Today, with information technology, and international banking, large volumes of funds can travel round the world in eighty milliseconds.

Recent amendments to our laws have made such electronic fund transfers legal together with the digital signatures needed to initiate them.

To summarise, liberalization and globalization, combined with developments in information technology, have made the movement of capital across national boundaries free, rapid, and legal. Large volumes of capital can legally move in and out of the country at extremely short notice of a few seconds.

1.1 Capital mobility

Liberalization, IT, and the accompanying legal changes, all concern the 'how' of capital movements. One can also ask: why do capital movements take place? Capital comes in because it lacks productive avenues of investment at home. The aim is to maximise profit. In search of a high-rate of profit, the capital often flows into speculative activities. In India at least, this is mostly what has been happening. The bulk of the incoming capital has gone into high-profit, high-risk, speculative activities like the stock-market and real estate, and into money-lending against real estate and white-good security.

Since the incoming capital does not get tied to any productive physical assets, it tends to remain in a liquid or semi-liquid state. That is, there are also no serious physical hurdles to the quick conversion of the incoming capital. Accordingly, large volumes of capital can, in fact, move in and out of the country at very short notice of a few days. There are not even any motivational inhibitions: for the aim is not the good of this or that nation or group of people, but solely to maximise profit.

The high mobility of capital, motivated solely by the desire to maximise profit, has made stock prices very volatile.[1] This high volatility of stock prices is a qualitatively new feature. This qualitatively new feature puts time-tested investment strategies to high risk.

To summarise, liberalization, IT, and the recent legal changes have made capital very mobile, and this increased mobility of capital has created a qualitatively new situation by making stock prices very volatile.

1.2 Plain language description

This qualitatively new situation may be described in plain language as follows. In the short term, the inflow of large volumes of funds makes stock prices buoyant, and pushes them up. This makes all investors richer. Short-term stability boosts investor confidence, and whets their desire for more profit, so that more funds are sucked into speculative activities. However, high volatility of stock prices means that this happy situation does not last for very long. In a Markovian steady-state kind of model, the expected time period in which an event occurs is inversely proportional to its probability, so the higher the risk, the sooner a crisis is likely to occur. In any case, when the capital recedes, most people are left poorer than they were to start with. The short term prosperity arising from footloose capital risks longer-term bankruptcy. South East Asia was merely an example. Thus, the short-term prosperity due to capital inflow seems almost like a con trick which serves to attract funds that are swept away by the receding capital.

As is clear from the last remark, the larger investors are not only better able to protect themselves from such increased volatility, they may even profit from it. The number of billionaires in Mexico increased after the Mexican crisis.

What if anything can one do about this? There are two possible responses. The first is that one can try and reverse the liberalization process, and re-impose restrictions on

---

[1] For the definition of volatility, see J. C. Hull, *Options, Futures, and Other Derivatives*, 3rd ed, Prentice Hall, New Delhi, 1997.

the movement of capital, as Malayasia has recently successfully demonstrated. At the level of the individual investor, this is not a feasible proposition, and the investor will typically proceed by accepting the existing situation as a *fait accompli*. Can anything be done at this level?

New models and methods are required to deal with the emerging high-risk situation. In particular, derivatives and hedges are often used to manage risk. An example is provided by the well-known Black-Scholes option-pricing formula used in hedging strategies. The difficulty is that the new methods of managing risk require

(a) mathematical sophistication, and

(b) high-performance computation.

In addition, I will argue that in situations involving very high risk, one also needs a fundamental paradigm shift in both statistics and mathematics.

Irrespective of the paradigm shift that is needed, the typical investor lacks even the combination of skills (a) and (b). In fact, Taiwanese bankers have refused to handle hedge funds for this reason. However, a struthious refusal to handle hedge funds is hardly an insurance against macro-economic risk!

What will our Indian bankers do? Will Indian bankers invest in the development of appropriate new models and software to manage high-risk? Can the Indian financial managers understand the implications well enough to cope with them? Unfortunately, our management system is hierarchical, and the managers at the top typically tend to lack technical skills, or have become de-skilled to the point of being technologically illiterate. Such a system of management is unsuited to meet the emerging challenges, for the decision-making people at the top are not only unable to assess the technical issues, they are frequently unable to assess the technical competence of those who advise them on technical issues! The difficulty that technically illiterate people have in selecting the right advisors is, in fact, exactly analogous to the difficulty that ordinarily illiterate people face in selecting a good doctor or a competent lawyer.

1.3 An example

An example might make matters clearer. The RBI holds USD 30 billion in foreign exchange. At least 5% of these funds are discretionary. Thus, even a 1% improvement in the method of investing these discretionary funds means a gain of over Rs 50 crores annually. An improved method of handling investments is an implicit consequence of an improved method of handling risk, for this implies the possibility of riskless arbitrage. A small fraction of the expected annual gain could suffice to pay for the development costs of the requisite new models and software. Despite this logic, what actually transpired was something else.

In fact, in 1994–95, in C-DAC, I wrote a parallel code for a modified Black-Scholes model.[2] After I left C-DAC, no one followed this up. Presumably people thought that this was some exotic mathematical stuff of no practical relevance. Interest in this was rekindled when Scholes was awarded the Nobel prize in Economics in 1997: Indian managers are always as willing to imitate and follow precedents as they are unwilling to try new things. Accordingly, a concrete proposal for an improved method of handling risk was made to RBI in 1998. However, the inevitable response was: “Why aren’t others doing this?” Clearly, in the Indian managerial world the only safe thing is to follow precedents, so that all new ideas can only come from abroad. The idea of gaining a technological edge through new ideas is absent from the world of Indian managers. Accordingly RBI continues to rely on J. P. Morgan’s RiskMetrics software, and does not care to explore more appropriate methods of handling high risk situations.

There is a structural reason for this tendency of Indian managers towards mimesis. It is impossible in the Indian managerial system to give credit at the lower levels: all credit for any organizational achievement is inevitably absorbed by the top managers. All discredit for any organizational failure is inevitably transferred down the line. Accordingly, no rewards accrue to the manager from a good decision, while punishments are likely to follow any wrong decision. A decision which departs from precedents is likely to be seen as bad. This naturally encourages stasis and total avoidance of the slightest risk. It is a moot point whether even a crisis will bring out the unsuitability of the existing managerial system to meet the emerging stresses generated by liberalization.

However, for those individual investors who are still seeking a method of safeguarding themselves against the emerging high-risk situation, the following is a non-technical summary of what can be done. The technical details are in the appendix.

2 What can be done: non-technical summary

Liberalization and developments in IT have increased stock- price volatility. A paradigm shift in statistics is needed to handle this increased volatility. The normal distribution does not apply to such a high-risk situation. However, existing models of option pricing and hedging, such as the Black-Scholes model, estimate stock-price volatility using a normal distribution.

The normal distribution does not apply to a high-risk situation since a high-risk situation leads to failure of the central limit theorem, or the law of large numbers. To avoid a complete shift from the existing paradigm of statistics, built almost entirely around the central limit theorem, it is possible to quibble that there is no actual failure

---

[2] C. K. Raju“ Stochode: Software for Simulation of Stochastic Differential Equations Driven by Levy Motion”, Centre for Development of Advanced Computing, unpublished internal report, 1996.

of the central limit theorem, but that there is ultra-slow convergence to the normal. Such quibbles are of no value from a practical point of view: irrespective of how one chooses to describe the situation, practically the fact remains that the use of existing statistical formulae will give inaccurate results in high-risk situations.

An appropriate generalisation which replaces the central limit theorem and the normal distribution is the notion of a stable distribution: more specifically, the α-stable Lévy distributions.

The use of alpha-stable Levy distributions in place of the normal distribution necessitates computation. Large-scale use requires high-performance computing or supercomputing. How secure are the resulting computations? To assess this, a paradigm shift is required from formal proof-oriented mathematics to a new kind of computational mathematics. This is a very subtle matter, with which I have dealt more fully elsewhere.[3]

The technical details are in the appendix. The corresponding software for stock-price simulation, using α-stable Lévy distributions is available on the author's website, and may be downloaded and used after getting a password from the author.

3 Technical Appendix

Stock market prices cannot be modeled deterministically. Bachelier first suggested that they should be modeled as stochastic processes.

3.1 Random variables and stochastic processes

A real-valued *random variable* ξ is specified by specifying the probability Pr $\{ \xi < x \}$, for every real number x. Thus, if (Ω, B, P) is a standard Borel probability space, and $\xi : \Omega \to R$, then ξ must be a Borel-measurable function,[4] since for the above probability to be defined, we must have $\xi^{-1}(-\infty, x) \varepsilon \Omega$, for each $x \varepsilon R$.

A (real-valued) *stochastic process* or a random function $\xi(t)$ is a map $\xi : \Omega \times R \to R$, such that

---

[3] C. K. Raju, "Computers, Mathematics Education, and the Alternative Epistemology of the Calculus in the YuktiBhâsâ", Invited Plenary Talk, 8th East-West Conference, Hawai'i, Jan 2000, (to appear) in: *Philosophy East and West*, July 2001.

[4] For the definition of a random variable, see, P. R. Halmos, *Measure Theor*y, D. Van Nostrand Co., 1950, Affiliated East-West Press, New Delhi, 1964.

(1) for each $t \varepsilon R$, $\xi(t)$ is a random variable, and if
$F_{t_1,t_2,t_3,\ldots,t_n}(x_1, x_2, x_3,\ldots, x_n) =$
$Pr\{ \xi(t_1)<x_1, \xi(t_2)<x_2, ..., \xi(t_n)<x_n\}$
is the distribution function of $(\xi(t_1), \xi(t_2), \ldots,\xi(t_n))$, then

(2) (symmetry)
$F_{t_{i_1},t_{i_2},t_{i_3},\ldots,t_{i_n}}(x_{i_1}, x_{i_2}, x_{i_3},\ldots, x_{i_n}) =$
$F_{t_1,t_2,t_3,\ldots,t_n}(x_1, x_2, x_3,\ldots, x_n)$
for every permutation $(i_1, i_2, \ldots,i_n)$ of $(1, 2,..., n)$, and

(3) (compatibility)
$F_{t_1,t_2,\ldots,t_m,t_{m+1},\ldots,t_n}(x_1, x_2,\ldots, x_m,\infty,\infty,...,\infty) =$
$F_{t_1,t_2,\ldots,t_m}(x_1, x_2,\ldots, x_m)$
for all $t_1,t_2,\ldots,t_m$.

A stochastic process $X(t)$ generalises the idea of a sequence of identically distributed random variables (i.d.r.v's) to a continuous variable t. For each fixed t , $X(t) \equiv X(\omega)$ is a random variable, and for any $t_1, t_2, \ldots, t_n$, the random variables $X(t_1), X(t_2), \ldots,X(t_n)$ are i.d.r.v.'s

Four types of stochastic processes will be important for us.

- Brownian motion (Wiener Process)
- Geometric Brownian motion
- Lévy motion
- Geometric Lévy motion

3.2 Standard Brownian motion

A stochastic process $z(t) \equiv z(t, \omega)$ is called standard Brownian motion[5] if

1. It is a process with independent increments, i.e.,

$\Delta_1 z = z(t_2) - z(t_1)$, and $\Delta_2 z = z(t_3) - z(t_2)$

---

[5] For the definition of stochastic process, and standard Brownian motion, see H. P. McKean Jr, *Stochastic Integrals*, Academic Press, New York, 1969.

are statistically independent.

2. The increments are normally distributed with variance equal to the time increment Δt:

$\Delta z \sim N(0, \sqrt{\Delta t})$.

3. $z(0) = 0$ almost everywhere, i.e. with probability 1.

4. Almost all sample paths are continuous: i. e., for each fixed ω,
$z(t)$ is continuous almost surely, i.e. with probability 1.

However, the sample paths of Brownian motion are almost everywhere non-differentiable.

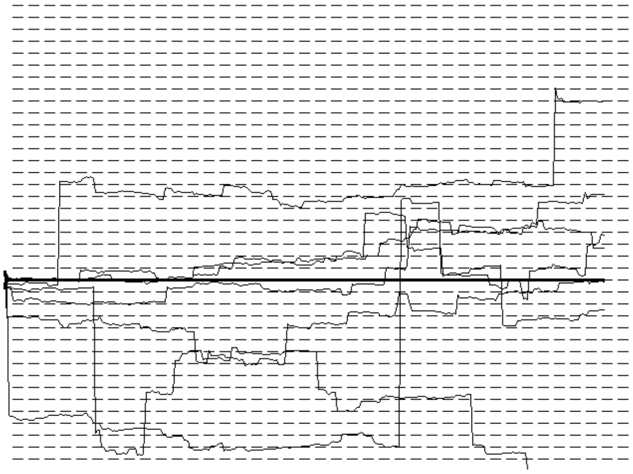

Fig: 1: Sample paths of standard Brownian motion

3.3 Stochastic differential equation

Langevin et al tried to model the dynamics of Brownian motion through an equation of the type

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\xi_t\, dt \qquad (1)$$

This models the process through an average or drift term + random perturbation.

However, if X_t is the Brownian motion, then the required ξ_t must be "Gaussian white noise", i.e., $E(\xi_s \xi_{s+t}) = \delta(t)$, where $\delta(t)$ is the Dirac delta function (Schwartz distribution).

(1) may be reinterpreted as

$$X_t(\omega) = X_0(\omega) + \int_0^t a(s, X_s(\omega))ds + \int_0^t b(s, X_r(\omega))\xi_r(\omega)dr \tag{2}$$

The mathematics of ξ_t was not then clear. But Wiener proved the existence of a process (Brownian motion, Wiener process) such that

$$dz_t = \xi_t \, dt.$$

Ito showed that (1) and (2) could be rewritten

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dz \tag{1'}$$

$$X_t(\omega) = X_0(\omega) + \int_0^t a(s, X_s(\omega))ds + \int_0^t b(s, X_r(\omega))dz(\omega) \tag{2'}$$

3.4 Stochastic Integral

Let (Ω, B, P) be a standard Borel probability space. Let z_t be the standard Brownian motion, and let F_t, $t \geq 0$, be an increasing family of σ-algebras generated by z_t. Consider a step function $f(t,\omega) \, \varepsilon \, L_2$, such that

$$f = \sum_{j=0}^{n} f_j(\omega) \chi_{[t_j, t_{j+1})}(t)$$

where each $f_j \, \varepsilon \, L_2$ is F_{t_j} measurable.

Define

$$I(f)(\omega) = \sum_{j=0}^{n} f_j(\omega)\{ z_{t_{j+1}} - z_{t_j} \}$$

Then

$$E(\ I(f)^2\ ) = \sum E f_j^2\ E\Big(\ (z_{t_{j+1}} - z_{t_j})\ |\ \boldsymbol{F_{t_j}}\ \Big)$$

$$= \sum E(f_j^2)\Big(t_{j+1} - t_j\Big)$$

For a general f, s.t. $f\,(t,\cdot)$ is F_t measurable for each t, and Ef_t^2 is continuous in t, $I(f)$ is the L_2 limit of an approximating sequence $f^{(n)}$ of step functions.[6]

3.5 Geometric Brownian motion

Stock/asset prices follow geometric Brownian motion. If S is the asset price

$$dS = \mu\ S\ dt + \sigma\ S\ dz$$

where
z = standard Brownian motion
μ = drift rate = rate of return on S
σ = volatility.

since, the rate of return μ is independent of stock price, and the uncertainty σ independent of stock price.[7]

3.6 Ito’s lemma

If X satisfies

$$dX = a(X,\ t)dt + b(X,\ t)dz$$

with *drift coefficient a*, and *dispersion coefficient b*, and $G \equiv G(X,\ t)$, then G satisfies[8]

---

[6] For a ‘proper’ definition of the stochastic integral see McKean, cited above. See also Raju, cited above, for the epistemological and practical worth of such mathematical propriety.

[7] For a detailed economic justification of this principle, see Hull, cited above.

[8] For a proof of Ito’s lemma, see McKean cited above.

$$dG = \left(\frac{\partial G}{\partial x} a + \frac{\partial G}{\partial t} + ½ \frac{\partial^2 G}{\partial x^2} b^2 \right) dt + \frac{\partial G}{\partial x} b dz.$$

Given that the stock price S follows geometric Brownian motion

$$dS = \mu\, S\, dt + \sigma\, S\, dz,$$

if $G \equiv \ln S$, $\frac{\partial G}{\partial S} = \frac{1}{S}$, so that G satisfies

$$dG = (\mu - \frac{\sigma^2}{2})\, dt + \sigma\, dz,$$

which is a generalised Wiener process. Hence, $G \equiv \ln S_T$ is normally distributed with mean $(\mu - \frac{\sigma^2}{2})T$, and variance $\sigma^2 T$.

Hence, S_T is lognormally distributed[9] with

$$E(S_T) = Se^{\mu T} \text{ and } \mathrm{Var}(S_T) = S^2 e^{2\mu T} \left[e^{\sigma^2 T} - 1 \right]$$

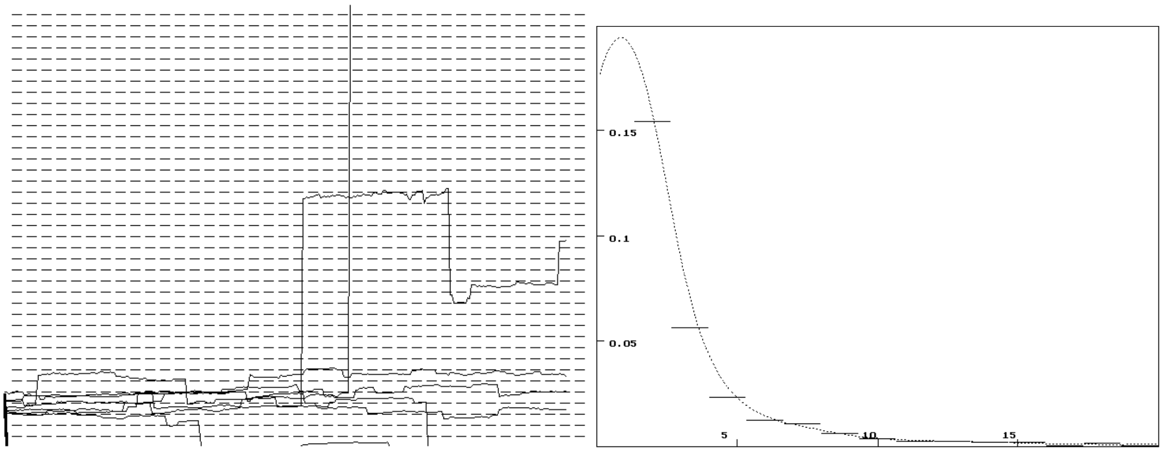

Fig. 2:
Sample paths of geometric Brownian motion (left) and the associated lognormal distribution of stock prices (right)

---

[9] For more details on the lognormal distribution, and its significance in economics, see J. Aitchison and J. A. C. Brown, *The Lognormal Distribution, with special reference to its uses in economics*, Cambridge University Press, 1957.

3.7 The Black-Scholes Option Pricing Formula

A (European) *call option* is the *right* to buy an asset on the expiry of a maturity period T, for a predetermined "strike price" K. If the asset price is S, the option will be exercised only if $S_T > K$, hence

- Payoff at time T is:

$$\max(S_T - K, 0). \tag{1}$$

- Expected payoff at time T is:

$$E \max(S_T - K, 0) \tag{2}$$

- Present value of expected payoff is:

$$c = e^{-r(T-t)} E \max(S_T - K, 0). \tag{3}$$

(t denotes the present time, and r is the riskless [bank] interest rate with continuous compounding).

Using the previous lognormal distribution for S_T this can be calculated to yield the Black-Scholes formula[10] for the option price:

$$c = S(t) \cdot N(d_1) - Ke^{-r(T-t)} \cdot N(d_2)$$

where

$$d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)(T-t)}{\sigma \cdot \sqrt{(T-t)}}$$

and

$$d_2 = d_1 - \sigma \cdot \sqrt{(T-t)}$$

where $N(x)$ is the cumulative distribution function for a normal variate.

3.8 Normal vs α-stable Lévy distributions

The Black-Scholes formula needs correction since stock prices are not lognormally distributed. They exhibit "fat tails", self-similarity under scaling, etc. CORRECTION MEANS THE POSSIBILITY OF ON-THE-AVERAGE RISKLESS ARBITRAGE.

---

[10] F. Black and M. Scholes," The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, **81** (1973) 637–59. See also, B. Lauterbach and P. Schultz, "Pricing Warrants: An Empirical Study of the Black-Scholes Model and Its Alternatives" *Journal of Finance*, **45** (1990) 1181–1209.

Basically, the normal distribution is not applicable. Applicability of the normal distribution rests on the following.

Central Limit Theorem: Let $X_1, X_2, \ldots, X_n$ be i.i.d.r.v.'s (i.e., random sample of size n). Let

$$S_n = X_1 + X_2 + \ldots + X_n$$
$$EX_i = \mu,\ EX_i^2 = \sigma^2, \text{ then}$$

$$\frac{S_n - ES_n}{\sqrt{ES_n^2}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(\mu, \sigma).$$

Weak Law of Large Numbers:
If the variance does not exist, but the mean does,

$$\Pr\left\{ \left|\frac{S_n}{n} - \mu\right| > \varepsilon \right\} \to 0.$$

If neither mean nor variance exists, the limiting distribution must be an α-stable distribution (infinitely divisible distribution).[11]

Hence, CLASSICAL STATISTICS FAILS, FORCING COMPUTATION.

3.9 The α-stable Lévy distributions

The characteristic function of a random variable X is Ee^{itX}.

The characteristic functions of the stable distributions are related to the function $e^{-|t|^\alpha}$, where $\alpha \varepsilon (0,2]$ is called the index of stability.

The characteristic function ϕ of the most general stable distribution must have the form:

$$\log \phi(t) = -\sigma^\alpha |t|^\alpha \left\{ 1 - i\beta \operatorname{sgn}(t) \tan\frac{\pi\alpha}{2} \right\} + i\mu t$$

We say that $X \sim S_\alpha(\sigma, \beta, \mu)$, where
α = stability parameter, $0<\alpha\leq 2$
β = skewness parameter, $-1\leq\beta\leq 1$
σ = scale parameter, $\sigma>0$
(if $\alpha<2$ the variance does not exist, hence σ cannot be called variance)
μ = location parameter

---

[11] For a compact introduction to stable distributions, see W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, 1966, Wiley Eastern, New Delhi, 1969, pp 165–73, and 540–9. Feller's treatment of stable distributions, however, involves a small error, which invalidates some of his formulae.

(if $\alpha<1$ the mean does not exist, hence μ cannot be called the mean).

An alternative form, with $\sigma = 1$, $\mu = 0$, $\overline{\alpha} = \min(\alpha, 2-\alpha)$, is

$$\log \phi(t) = \begin{cases} -|t|^{\alpha} e^{-\frac{i\pi}{2}\cdot\,\overline{\alpha}\cdot\delta\,\cdot \mathrm{sgn}(t)}, & \alpha \neq 1 \\ -|t|\left(1 + i\delta\,\frac{2}{\pi}\,\mathrm{sgn}(t)\log|t|\right), & \alpha = 1 \end{cases}$$

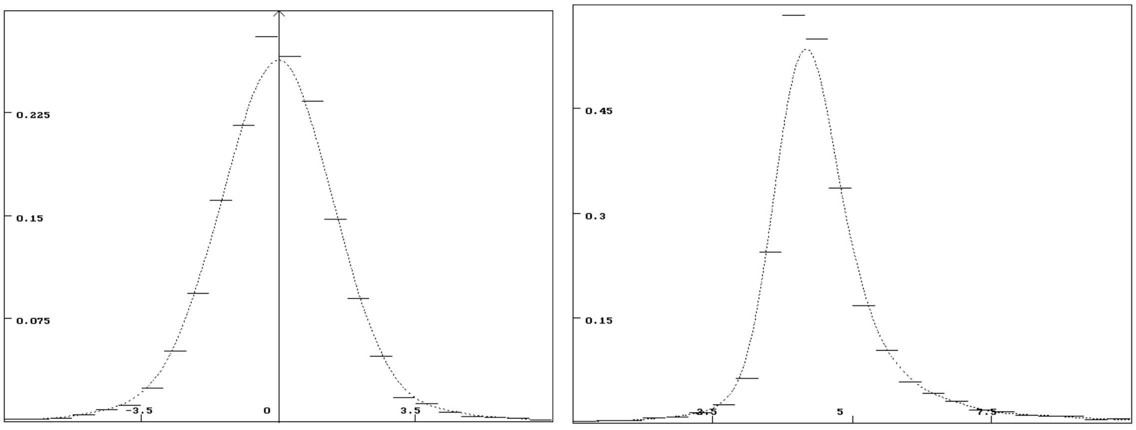

Fig. 3: Example Lévy densities

(Left) Symmetric case $\alpha = 1.7$, $\delta = 0$, $\mu = 0.0$, $\sigma = 1.0$, and general case, (right) $\alpha = 1.2$, skewness $\delta = 0.8$, $\mu = -1.0$, $\sigma = 5.0$,

3.10 Special cases

Let $X \sim S_\alpha(\beta, \mu, \sigma)$. We can try to understand the nature of the distribution by looking at various special cases.

$\alpha = 2$, $\beta = 0$: $S_2(0, \mu, \sigma) = N(\mu, \sigma)$, the normal distribution.

$\alpha = 1$, $\beta = 0$: $S_1(0, \mu, \sigma) =$ Cauchy, with density

$$f(x) = \frac{2\sigma}{\pi((x-\mu)^2 + 4\sigma^2)}.$$

$\alpha = ½$, $\beta = 1$: $S_{½}(1, \mu, \sigma) =$ classical Lévy with density

$$f(x) = \left(\frac{\sigma}{2\pi}\right)^{½} (x-\mu)^{-3/2} \exp\left\{-\frac{\sigma}{2(x-\mu)}\right\}$$

Tail probabilities: If $X \sim S_\alpha(\beta, \mu, \sigma)$ has density f_α, then asymptotically

$$\Pr\{\, X > \lambda \,\} \approx \frac{D_\alpha \beta \sigma^\alpha}{\lambda^\alpha}, \qquad f_\alpha(x) \approx O\left(\frac{1}{\lambda^{1+\alpha}}\right)$$

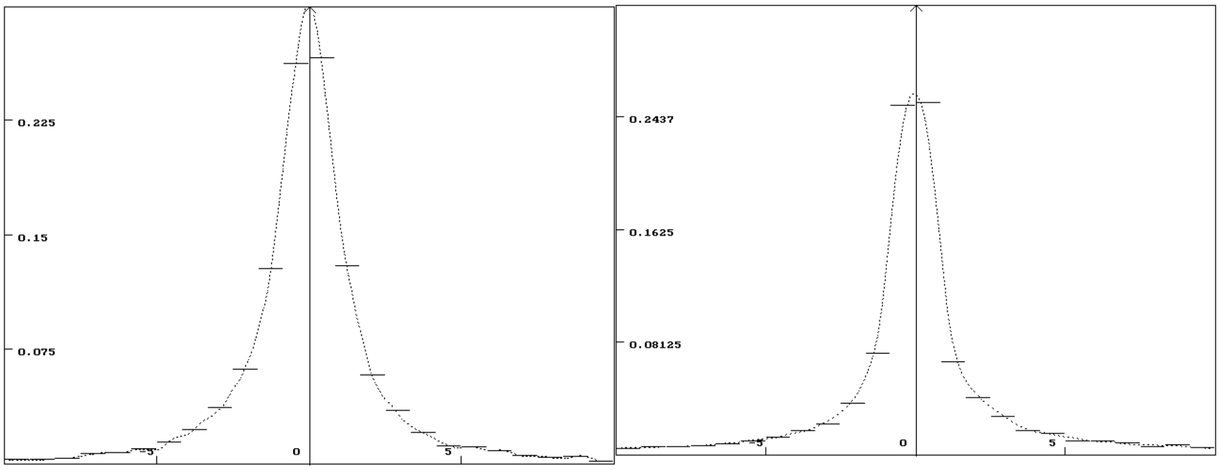

Fig. 4: Special cases of stable distributions
Left: Density function of Cauchy distribution. Right: Density function of classical Lévy distribution.

3.11 Standard Lévy motion

A stochastic process $L_\alpha(t) \equiv L_\alpha(t, \omega)$ is called standard Lévy motion if

1. It is a process with independent increments, i.e.,

$\Delta_1 L_\alpha = L_\alpha(t_2) - L_\alpha(t_1)$, and $\Delta_2 L_\alpha = L_\alpha(t_3) - L_\alpha(t_2)$
are statistically independent.

2. The increments are α-stable Lévy distributed:

$$\Delta L_\alpha \sim S_\alpha\,(\beta, 0, (\Delta t)^{1/\alpha})$$

3. $L_\alpha(0) = 0$ almost surely.

There is no requirement of continuity. The process may have jumps.

The stochastic integral with respect to standard Lévy motion is defined similarly except that $f \in L^{\alpha}$, and we take the L^{α} limit of the stochastic integral of simple functions, $I(f^{(n)})$, to obtain $I(f)$.

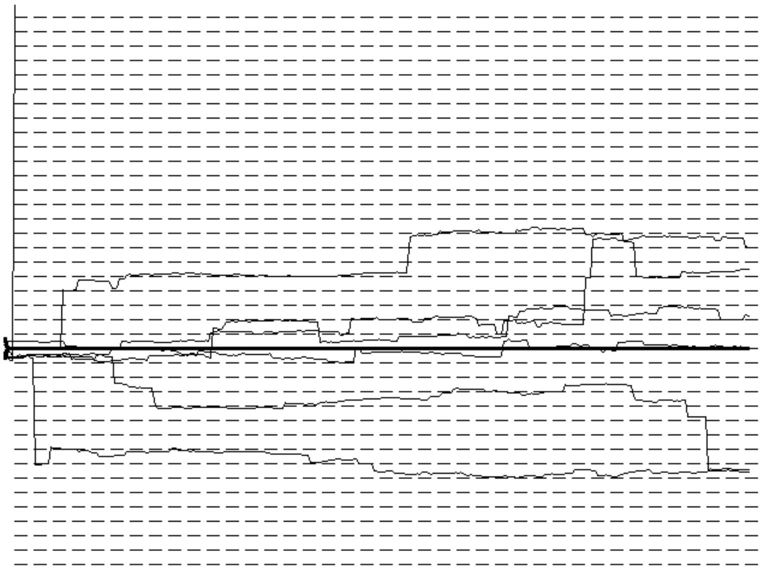

Fig. 5: Sample paths of standard Lévy motion

3.12 Discretisation of the Stochastic Integral

To solve

$dX = a(X, t)dt + b(X, t)dL_{\alpha}$, with $X(0) = X_0$, an α-stable r.v.,

we need to use an appropriate discretisation scheme.[12]

In an explicit scheme:

$\{t_i = i\tau, i{=}0, 1, 2..., I\}$, $\tau = T/I$ is a fixed mesh on $[0, T]$.

$X(t_i)$ are r.v.'s approximating $\{X(t)\}$.

$$\Delta M^{\tau}_{\alpha,i} = M_{\alpha}([t_{i-1}, t_i)) = L_{\alpha}(t_i) - L_{\alpha}(t_{i-1}) \quad \sim S_{\alpha}(\tau^{1/\alpha}, \beta, 0)$$

approximates the stochastic measure.

$X(t_i)$ is further approximated by a sample $\{X_{t_i}(n)\}_{n=1}^{N}$.

This corresponds to an algorithm based on the Euler method.

---

[12] For some of the ways to discretize the stochastic integral, see P. E. Kloeden, E. Platen, H. Schurz, *Numerical Solution of SDE Through Computer Experiments*, Springer Verlag, Berlin, 1994.

$X_0^\tau = X_0$

For $i = 1, 2, \ldots, I$

$$X_{t_i}^\tau = X_{t_{i-1}}^\tau + Y_{t_i}^\tau$$

$$Y_{t_i}^\tau = a(t_{i-1}, X(t_{i-1}))\ \tau + \quad + \ b(t_{i-1}, X(t_{i-1}))\Delta M_{\alpha,i}^\tau$$

Replace $X_0^\tau \to \left\{X_0^\tau(n)\right\}_{n=1}^{N}$, etc. by random samples. Note that, unlike the case of Brownian motion, convergence cannot actually be proved for a number of simple schemes.[13] Hence, something further is needed to assess the reliability of the result.

4 Simulation of random variables

The actual execution of the above plan requires that random variables be readily available.[14] Assume that we can generate a uniform random variate: $X \sim U(0, 1)$.

4.1 Inversion method

Consider a r.v. ξ with a probability density function $f(x)$, and for $x \varepsilon [a, b)$ let

$$F(x) = \int_a^x f(y)dy$$

If F^{-1} exists and $\gamma \sim U(0, 1)$ then $F^{-1}(\gamma) \sim F(x)$.

Example. For the **exponential distribution** with parameter λ, $f(x) = \lambda e^{-\lambda x}$, $F(x) = 1 - e^{-\lambda x}$, $x >= 0$. If $\gamma \sim U(0, 1)$, then $\tau = -\frac{1}{\lambda} \log (1-\gamma) \sim F(x)$.

---

[13] For more material specific to alpha-stable distributions, see A. Janicki and A. Weron, *Simulation and Chaotic Behavior of Alpha Stable Stochastic Processes*, Marcel Dekker, New York, 1994. The book, however, seeks to hide as much as it seeks to communicate, for obvious reasons.

[14] A good introduction to general Monte Carlo Simulation is: M. H. Kalos, and P. A. Whitlock, *Monte Carlo Methods*, Vol. 1: *Basics*, Wiley-Interscience, New York, 1986. More general is: D. E. Knuth, *The Art of Computer Programming 2: Semi-Numerical Algorithms*, Addison-Wesley, Reading, Mass, 2nd ed, 1981.

4.2 Normal distribution.

By the Central Limit Theorem, if $\xi_1, \xi_2, \ldots, \xi_n$ are i.i.d.r.v.'s with $E(\xi_n) = \mu$ and $D^2(\xi_n) = \sigma^2$ then

$$\lim_{n\to\infty} P\left\{\frac{\xi_1+\xi_2+\ldots+\xi_n - n\mu}{\sigma\sqrt{n}}\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

If $\gamma_1, \gamma_2, \ldots, \gamma_n \sim U(0, 1)$ $(\mu = 1/2, \sigma^2 = 1/12)$ then

$$P\{\gamma_1+\gamma_2+\ldots+\gamma_{12} - 6 < x\} \approx \int_{-\infty}^{-x} e^{-u^2/2}\, du$$

Alternatively, one can use the Box-Muller method: if $\gamma_1, \gamma_2 \sim U(0, 1)$ then

$$\eta_1 = \sqrt{-2\ln(\gamma_1)}\, \cos(2\pi\gamma_2)$$
$$\eta_2 = \sqrt{-2\ln(\gamma_1)}\, \sin(2\pi\gamma_2)$$

are i.i.d. $N(0,1)$.

4.3 Simulation of α-stable distributions

The following steps are used.[15]

Generate a r.v. $V \sim U(-\pi/2, \pi/2)$.

Generate an exponential r.v. W with mean 1.

Compute

$$X = \frac{\sin(\alpha V)}{\cos^{1/\alpha}(V)} \times \left\{\frac{\cos(V-\alpha V)}{W}\right\}^{(1-\alpha)/\alpha}$$

For skewed α-stable r.v. $Y \sim S_\alpha(1, \beta, 0)$ we use the formula

---

[15] For the simulation of alpha-stable and other distributions, see L. Devroye, *Non-Uniform Random Variate Generation*, Springer, Berlin, 1986. The original reference for alpha-stable distributions is J. M. Chambers, C. L. Mallows, and B. W. Stuck, “A Method for Simulating Stable Random Variable”, *Journal of the American Statistical Association*, **71** (1976) 340–44.

$$Y = D_{\alpha,\beta} \times \frac{\sin(\alpha(V+C_{\alpha,\beta}))}{\cos^{1/\alpha}(V)} \times \left\{ \frac{\cos(V-\alpha(V+C_{\alpha,\beta}))}{W} \right\}^{(1-\alpha)/\alpha}$$

$$\text{where } C_{\alpha,\beta} = \frac{\arctan(\beta \tan(\pi\alpha/2))}{1 - |1-\alpha|}$$

$$D_{\alpha,\beta} = [\arctan(\beta \tan(\pi\alpha/2))]^{-1/\alpha}$$

5 Pseudo Random Number Generation

An $X \sim U(0, 1)$ can be obtained using a prn generator.

True random numbers cannot be used for high-speed computation since

(a) Natural phenomena cannot produce these numbers as fast as, say, 1 per ms, which is too slow even for a machine at 1 MFlop/s.

(b) Table look-up will require compilation, storage, verification, and access of enormously large tables (min. 100 quadrillion or 10^{18} entries) .

Definition: prn’s are sequences of numbers that are

(a) easy to generate on the computer
(b) satisfy some [not all] statistical tests of randomness.

It is important to have good Pseudo Random Number Generators (PRNG’s).

5.1 Cost of correlation

In the case of the Black-Scholes model, modified to be driven by Lévy motion, we saw that simulation is essential. There is no other way to value an option.

Parallel computation is desirable because it reduces cost/increases speed. However, a poorly thought out parallelisation strategy means a bad sequence of prn’s. And, in financial analysis, a bad (correlated) sequence of prn’s means

- Sample size is systematically over-estimated.
- Hence, risk is systematically under-estimated.
- Hence, wrong decisions may be frequently taken.

Hence, bad parallel PRNG = bad simulation = big financial loss.
How does one get a good PRNG?

5.2 The eight-fold way for a sequential PRNG

A good sequential pseudo random number generator should generate a stream of numbers that

- are uniformly distributed
- are statistically independent
- never repeat
- are reproducible (for debugging)
- are portable (the same on any computer)
- can be adjusted using an initial “seed” value
- can be easily split into independent subsequences
- can be generated rapidly using limited computer resources (memory/storage).

5.3 The four-fold way for a parallel PRNG

A good parallel PRNG should be

- a good sequential PRNG on each processor
- scalable (case N=1 good for debugging)
- independent also across processors
- efficient (preferably no data movement between processors after initialization)

5.4 Common PRNG’s

Some common PRNG’s are the following:

- von Neumann’s generator (obsolete)
- Linear Congruential Generators
- Lagged Fibonacci Generators

- Shift Register (Tauseworth) Generators (obsolete, since architecture-specific))
- Combination generators

von Neumann's generator:

Take a number x_n with $2a$ digits. Square it and discard the a most significant and the a least significant digits. Form x_{n+1} from the remaining digits.

$$x_{n+1} = \left[\frac{x_N^2}{b^a}\right] - \left[\frac{s_n^2}{b^{3a}}\right] \cdot b^{2a}$$

e.g. $x_n = 40$, $b = 10$, $a = 1$, then $x_{n+1} = 60$.

This is now obsolete since it has a very short period

Linear Congruential Generators:

$$x_{n+1} \equiv a_0 x_n + a_1 x_{n-1} + \cdots + a_j x_{n-j} + b \pmod{M}$$

If $a_j = 0$ for $j \geq 1$ one has the common special case

$$x_{n+1} = \lambda \cdot x_n + b \pmod{M}$$

The max period of this generator is M ($M-1$ if $b = 0$).

If $M >$ machine precision, then slower multi-precision arithmetic must be used.

If $M <$ machine precision, period may be too small. E.g. on a 32 bit machine max period would be $2^{32} \sim 10^9$, and a 10 MFlop/s machine can run through this in a couple of minutes.

It is most convenient to have M a power of 2, e.g. 2^{31}, (as in IBM RANDU) but then the lowest order bits tend to be correlated. The scatter plot of n-tuples shows a lattice structure. Hence M must be chosen to be a prime P.

Statistical properties are mostly decided by λ and P. Possible values are
$P = 2^{31}-1$, $\lambda = 16{,}807 = 7^5$.

Lagged Fibonacci Generators:

$$X_n = X_{n-p} \otimes X_{n-q}$$

where

p, q are lags, $p>q$, say,

$\otimes$ denotes a binary operation, such as

+ addition mod M (or mod 1)
- subtraction mod M
* multiplication mod M
$\oplus$ XOR, bitwise exclusive or.

If $M = 2^b$, maximum period for optimal lags p and q is

$2^p - 1$	for $\oplus$ (XOR)
$(2^p - 1)2^{b-1}$	for + and -
$(2^p - 1) \cdot 2^{b-3}$	for *

The period can be increased by increasing p and q. Suggested values are

$$p = 1279, \quad q = 1063$$

(C and Unix generator RANDOM has a default lag of 31, not recommended for serious computing.)

Shift Register (Tauseworth) Generators

First generate q random integers by the LCG

$$x_{n+1} \equiv a_0 x_n + a_1 x_{n-1} + \cdots + a_j x_{n-j} + b \pmod P$$

where

$$1 + a_0 y + a_1 y^2 + \cdots + a_{n-1} y^n$$

should be a primitive polynomial over the field { 0, 1} to give the maximum period of

$$2^{n+1} - 1$$

Then perform a lagged bitwise XOR to generate further random integers

$$x_k = x_{k-q+p} \oplus x_{k-q}$$

Suggested values are the trinomial

$$1 + y^q + y^p$$

with $p = 103$, $q = 250$

These do not perform so well on statistical tests. But they were popular because of speed. However, they are now obsolete since + performs almost as fast as ⊕.

Combination Generators

Currently P. L'Ecure's combination generator is used at CERN.

Consider J LCG's

$$x_{j,\,n} \equiv a_j\, x_{j,\,n-1} \bmod m_j, \quad 1 \le j \le J$$

where m_i's are distinct primes, and j th generator has (maximal) length

$$\rho_j = m_j - 1.$$

Let $\delta_1, \delta_2, \ldots, \delta_J$ be arbitrary integers, s.t.

$$(\delta_j, m_j) = 1, \; 1 \le j \le J$$

Form

$$z_n \equiv \left(\sum_{j=1}^{J} \delta_j\, x_{j,\,n} \right) \bmod m_{1,} \qquad u_n = z_n / m_1$$

and

$$w_n = \left(\sum_{j=1}^{J} \frac{\delta_j\, x_{j,\,n}}{m_j} \right) \bmod 1$$

thenu_n and w_n both have period of length

$$\rho = \mathrm{lcm}\,(\rho_1, \rho_2, \ldots, \rho_J)$$

with a maximum value

$$\rho = (m_1 - 1)\cdot(m_2 - 1)\cdots(m_J - 1)/2^{J-1}$$

Further, w_n obeys the recurrence

$$x_n \equiv a x_{n-1} \bmod m, \quad w_n = x_n / m$$

where

$$m = \prod_{j=1}^{J} m_j \qquad .$$

Suggested values are those of L’Ecure:[16]

j	m_j	a_j
1	2147483647	45991
2	2147483543	207707
3	2147483423	138556
4	2147483323	49689

$$m = 21267641435849934371830464348413044909$$

$$a = 5494569482908719143153333426731027229$$

$$\rho = (2^{31}-2)(2^{31}-106)(2^{31}-226)(2^{31}-326)/2^3 \approx 2^{121}$$

5.5 Parallelisation strategies for PRNG

Simulation parallelisable to the extent that PRNG is parallelisable. Intuitively this is 100% parallelisable Three possible strategies for parallel PRNG are

- Centralized: Use one PRNG and distribute its output among processors
- Replicated: Replicate the same PRNG on each processor, possibly with some minor differences
- Distributed: Use a single PRNG and partition its tasks to multiple processors.

5.6 Advantages and Disadvantages

- Centralized
  - computation and communication cannot be balanced: random numbers are consumed faster than they are produced.
  - In practice, use is restricted to distributed computing.
- Replicated
  - Using identical code on N processors is completely pointless: “efficiency of parallelisation” is 100% but nothing is gained for the application.
  - Using the same rand () function with different seeds does very little more, since this produces highly correlated sequences of prn’s.

---

[16] P. L’Ecuyer,“Efficient and Portable Combined Random Number Generators”, *Communications of the ACM*, **31** (1988) 742–749, and 774; “Random Numbers for Simulation”, *Communications of the ACM*, **3**3 (1990) 85–97. Also P. L’Ecuyer, “Random Number Generation” in J. Banks (ed) *Handbook on Simulation*, Wiley, 1997.

— Using e.g. LCG’s with a fixed sequence of (λ, P) pairs creates problems w.r.t. scalability.

— Using LCG’s with a random sequence of (λ, P) pairs creates a problem w.r.t. reproducibility, and debugging.

Hence, the method of choice is still that of Perclus and Kalos:[17] Use the LCG’s

$$x_{n,j} \equiv a\, x_{n,j} + b_i \mod m$$

where

$$b_j \equiv d^{\,j} \bmod m, \qquad d = m^{1/2} + 1$$

or

$$b_j = p_j \quad 1 \le j \le J$$

and

$$p_i\text{'s are the largest primes less than } \sqrt{m/2}.$$

The resulting prn streams are independent.[18]

- Distributed

Leapfrog method

Distribute prn’s among processors like a deck of cards. Thus, if there are N processors, the *p*th processor “leapfrogs" the original sequence by N to get the sequence:

$$x_p,\ x_{p+N},\ x_{p+2N},\ \dots$$

However each processor must generate its own subsequence.

For LCG’s replace λ and b by Λ and B:

$$\Lambda \equiv \lambda^n \bmod m \text{ and } B \equiv \frac{b(a^N - 1)}{a - 1} \bmod m$$

This method has been known to give spurious results.

---

[17] D. E. Perclus and M. Kalos,“Random Number Generators for MIMD Parallel Processors”, *Journal of Parallel and Distributed Computation*, **6** (1989) 477–497

[18] The parallel PRNG of Perclus and Kalos was tried out by some students as a project carried out under my supervision at C-DAC: see, V. Godbole and Samir Dani,“Monte Carlo Methods”, pp 304–312 in V. P. Bhatkar et al, *Advanced Computing 91*, Tata McGraw Hill, New Delhi, 1991.

Lehmer tree

Starting with root of the tree x_0, each element x has exactly one left successor x_L and one right successor x_R:

$$x_L \equiv (\lambda_L x + b_L) \bmod P$$
$$x_R \equiv (\lambda_R x + b_R) \bmod P$$

However, method tends to have statistical properties similar to that of single PRNG with random seeds.

6 Construction of density estimators

In the software provided, the solution of the stochastic differential equation is represented graphically in two ways: through

- sample paths, and
- using non-parametric density estimators.

The idea is to pass from a discrete histogram to a continuous density estimator.

A bar graph f_n approximates the graph of the unknown function f. Joining the middles of the bars by straight lines, one obtains a continuous piecewise linear approximation to the density f.

This may be formalized as follows.

Let $\xi_1, \xi_2, \ldots, \xi_n$ be i.i.d.r.v.'s $\sim f = f(x)$.

Let $\Xi_n = \{\xi_1, \xi_2, \ldots, \xi_n\}$.

Realizations of Ξ_n, i.e., its statistical sample, are denoted by
$\Xi^{(n)}(w) = \{\xi^{(1)}(w), \xi^{(2)}(w), \ldots, \xi^{(n)}(w)\}$.

A value of the density estimator, for any fixed $x \in \mathbf{R}$ is denoted by $f_n(x)$ or by $f_n(x, w)$, depending upon whether we regard it as an approximation to the density based on a sample or as a r.v.

6.1 The histogram

The histogram provides a non-parametric density estimator.

Let

$[a, b]$ = interval on which the density is to be estimated;
fix $n \,\varepsilon\, \mathbf{N}$ and a mesh size $h = {}^{(b-a)}\!/_{N}$.
$x_i = a+ih$ for $i = 0,1,\ldots,N$.

A histogram is a function $f_n = f_n(x)$, continuous on each interval $[x_i, x_{i+1})$, defined by

$$f_n(x) = \frac{K_n(x_i, x_{i+1})}{nh}$$

where

$$K_n(x_i, x_{i+1}) = \#\{k \mid \xi^{(k)} \,\varepsilon\, [x_i, x_{i+1})\}$$

counts the number of data values $\xi^{(k)}$ in each subinterval.

6.2 Kernel density estimators

More generally, one uses a continuous kernel function $K = K(u)$ on $\mathbf{R}$ satisfying

$$\int_{-\infty}^{\infty} K(u)\, du = 1, \quad K(u) \geq 0 \text{ for } u \,\varepsilon\, \mathbf{R}\,.$$

Commonly used kernel functions are the following.

- rectangular: $K(u) = \begin{cases} \frac{1}{2} & \text{if } |u| \leq 1 \\ 0 & \text{if } |u| > 1 \end{cases}$
- triangular: $K(u) = \begin{cases} \frac{1}{\sqrt{6}} - \frac{|u|}{6} & \text{if } |u| \leq \sqrt{6} \\ 0 & \text{if } |u| > \sqrt{6} \end{cases}$
- Gaussian: $K(u) = (2\pi)^{-1/2}\, e^{-u^2/2}$
- optimal: $K(u) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right) & \text{if } |u| \leq \sqrt{5} \\ 0 & \text{if } |u| > \sqrt{5} \end{cases}$

To pass from the histogram to a continuous density, estimated using a kernel density estimator, we further need a sequence $\{b_n\}$ such that

$$\lim_{n\to\infty} b_n = 0,$$

and

$$\lim_{n\to\infty} nb_n = \infty.$$

Usually, for practical calculations,

$$b_n = cn^{-1/5},$$

with c an appropriate constant.

- Rozenblatt-Parzen density estimator

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{b_n}K\left(\frac{x-\xi^{(i)}}{b_n}\right)$$

- Wolverton-Wagner-Yamato density estimator

$$f_n(x) = \frac{n-1}{n}f_{n-1}(x) + \frac{1}{nb_n}K\left(\frac{x-\xi^{(n)}}{b_n}\right)$$

$$f_0(x) = 0.$$

6.3 Description of the software

The software consists of three different executables, currently compiled as DOS-console applications.

1. Distrib.exe
2. sde.exe
3. Option.exe

6.4 Software for generating distributions

To illustrate the various distributions involved, the software also provides histograms, and kernel density sketches of normal, lognormal, exponential, Lévy, log-Lévy etc.

distributions. The `distrib.exe` simply generates a random variate distributed according to the following distributions, and constructs a kernel density estimate.

1. Exponential
2. Normal
3. Cauchy
4. Pareto
5. Lognormal
6. Classical Lévy
7. Symmetric Lévy
8. General Lévy
9. Log-Lévy

The various parameters of these distributions need to be specified, along with the sample size. For example, for the general Lévy distribution one needs to specify the following parameters.

- the index of stability $\alpha \varepsilon [0,2]$
- the location parameter μ
- the scale parameter $\sigma>0$, and the
- the skewness parameter $\delta \varepsilon [-1,1]$.

Some parameters are needed for sketching the histogram

- the interval on which the density is to be estimated
- the scale on the y-axis
- the class interval

For the kernel density estimator, the user must specify

- the kernel density function to use: rectangular, optimal, Gaussian, or triangular,
- the value of $b_n = cn^{1/5}$ by specifying the value of the constant c,
- the step size used to plot the density on the screen.

6.5 Software for solving stochastic differential equations

The sde.exe solves a general stochastic differential equation

$$dX(t) = a(X, t)dt + b(X, t)dL_\alpha(t)$$

where

$a(X,t) =$ drift coefficient (deterministic part)

$b(X,t) =$ dispersion coefficient

$L_\alpha(t) =$ standard Lévy motion, with index of stability α.

The solution is exhibited graphically using

- a vector field for the deterministic part
- sample paths for the full equation including the stochastic part.

The statistical characteristics of the solution $X(T)$are also exhibited graphically, using a

- histogram, and
- kernel density estimator.

The user must specify the drift and dispersion coefficients, the index of stability α, and the sample size, and various parameters needed for the histogram and the kernel density estimator.

Specifically, the user needs to specify the following.

- Coefficients of SDE

1. drift function `a(t, x) =`
2. dispersion function `b(t, x) =`

These functions may be entered symbolically, and are parsed internally.

- Specification of statistical characteristics of initial data

3. Index of stability for X(0), `alpha =`
4. Scale parameter for X(0), `sigma =`
5. Shift parameter for X(0), `mu =`

- Specification of statistical characteristics of driving perturbation

6. Index of stability for dL(t), `alpha =`
7. Scale parameter dL(t), `sigma =`

- Interval of integration, step size, sample size

8. Interval of integration [0, T], `T =`
9. Step size: The max number of time steps in [0, T], `ntMax =`
10. Sample size =

- Data for graphics

11. See deterministic vector field? Esc/Other key
12. Trajectory window [0, T] x [c, d]. `c =`
13. Trajectory window, `d =`
14. No. of trajectories to display, `nTr =`
15. No. of quantiles, and defn of quantile from (0.005 to 0.5)

- Data for histogram of solution X(T)

16. Estimation interval, `Low =`
17. Estimation interval, `High =`
18. Scale on the y-axis, `yMax =`
19. No. of class intervals for the histogram, `hiMax` =
    - Data for kernel density estimators
20. Parameters for displaying the density estimate, integer `xstep =`
21. The parameter *c* for the sequence `bn =`

6.6 Software for estimating option prices

For the particular case of a modified Black-Scholes option pricing model, using a log-Lévy distribution of stock prices, rather than a lognormal distribution assumed in the original Black-Scholes model, we need to solve the stochastic differential equation

$$dS = \mu\, S\, dt + \sigma\, S\, dL_\alpha$$

where
L_α = standard Lévy motion
μ = drift rate = rate of return on S
σ = volatility.

(The original Black-Scholes model is included by using the case $\alpha = 2$.) Thus, to specify the drift and dispersion coefficients, it is only necessary to specify two numbers—the drift rate and the volatility—both of which must be determined empirically. The drift rate is just the riskless interest rate or bank rate of interest (with continuous compounding) which is usually known easily. The key problem is to estimate the volatility. From

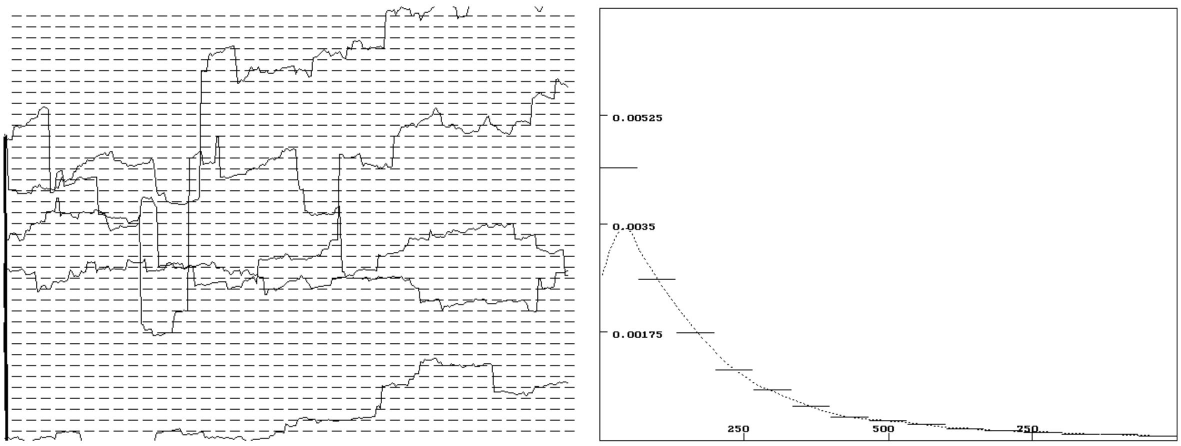

Fig. 6: Geometric Lévy motion and the log-Lévy density
Shown above are some sample paths (possible scenarios) of realistic stock price fluctuations. According to the above theory, the fluctuations follow geometric Lévy motion rather than geometric Brownian motion. Hence the associated distribution of stock prices follows a log-Lévy distribution (right) rather than the lognormal distribution used by the classical Black-Scholes model.

a knowledge of the solution, the price of an option may be estimated by using the standard formula (3.7.3). The software does this explicitly.

Thus, the user primarily needs to specify only the volatility, from the empirical data on stock prices. As we have pointed out, the estimation of volatility must start by assuming an underlying Lévy distribution. Various empirical studies have arrived at somewhat varying values of α to characterize the empirical distribution of stock prices.

For demonstration purposes, various input parameters have been provided. These are subject to the overall restrictions put on the sample size for the demonstration executable. Using the sample values provided, the software can be executed simply by invoking the accompanying batch files. All figures shown above were obtained in this way.

About the Author

C. K. Raju holds a Ph.D. from the Indian Statistical Institute. He taught mathematics and statistics at the University of Poona for many years before playing a lead role in the C-DAC team which built the first Param supercomputer. He has been a Fellow of the Indian Institute of Advanced Study. He has put forward a new theory of physics based on mixed-type functional differential equations (*Time: Towards a Consistent Theory*, Kluwer, 1994), and has been listed among the top 2000 scientists of the 20th century. He is also widely known for his work on the foundations of mathematics.