TWO NOTES ON FINANCIAL MATHEMATICS

Daniel Dufresne
Centre for Actuarial Studies
University of Melbourne

The two notes which follow are aimed at students of financial mathematics or actuarial science. The first one, "A note on correlation in mean variance portfolio theory" is an elementary discussion of how negative correlation coefficients can be, given any set of n random variables.

The second one, "Two essential formulas in actuarial science and financial mathematics," gives a more or less self-contained introduction to Jensen's inequality and to another formula, which I have decided to call the "survival function expectation formula." The latter allows one to write the expectation of a function of a random variable as an integral involving the complement of its cumulative distribution function.

A NOTE ON CORRELATION IN MEAN VARIANCE PORTFOLIO THEORY

Daniel Dufresne
Centre for Actuarial Studies
University of Melbourne

Abstract

This note is aimed at students and others interested in mean variance portfolio theory. Negative correlations are desirable in portfolio selection, as they decrease risk. It is shown that there is a mathematical limit to how negative correlations can be among a given number of securities. In particular, in an "average correlation model" (where the correlation coefficient between different securities is constant) the correlation has to be at least as large as $-1/(n-1)$, n being the number of securities.

Keywords: Mean-variance portfolio theory; average correlation models

1. INTRODUCTION

One of the more technical points encountered when teaching the basics of mean variance portfolio theory is the restrictions which apply on the covariance matrix of the securities' returns. It is not usually possible to spend much time on this topic in class, which is in part why I am writing this note. The main result is Theorem 2, which may be known to specialists, though I have not found it in the literature (comments are welcome).

Negative correlations betwen security returns are desirable, if one aims at reducing the variance of portfolio returns. One question which came from a student was "In a single-index model, is it possible for all securities to be negatively correlated to the market?" The answer is obviously no, at least if what is called the market return is indeed a weighted average of the all the returns of all securities in the model (observe that this is not the usual definition of the "market," and this is a small difficulty with the single-index model, see footnote 6, p. 138 in Elton *et al*. (2003)).

A similar question is "in an average correlation model, can all correlations between distinct securities be negative?" An average correlation model has all off-diagonal correlations equal to the same number ρ. (Such models would be more accurately called "constant correlation models.") This is the question which gave rise to this note.

Intuitively, the correlation coefficient is a measure of the linear dependence between two random variables. Strictly speaking, there is linear dependence between X and Y only when $\rho = \pm 1$, since it is known that this condition is equivalent to the existence of constants a, b such that

$$X = aY + b. \tag{1.1}$$

This result has a simple proof. To begin with, it is easy to see that, if we define

$$\tilde{X} \;=\; X - \mathsf{E}X, \quad \tilde{Y} \;=\; Y - \mathsf{E}Y,$$

then, in all cases,

$$\mathsf{E}\Big(\tilde{X} - \rho\frac{\sigma_X}{\sigma_Y}\tilde{Y}\Big)^2 \;=\; \sigma_X^2 + \rho^2\frac{\sigma_X^2}{\sigma_Y^2}\sigma_Y^2 - 2\rho\frac{\sigma_X}{\sigma_Y}\mathsf{Cov}(X,Y) \;=\; \sigma_X^2(1-\rho^2). \tag{1.2}$$

Next, temporarily assume that (1.1) holds. One can then find the expression of a in terms of the moments of X and Y. Taking expectations on both sides of (1.1) yields $\mathsf{E}X \;=\; a\mathsf{E}Y + b$, and so

$$\tilde{X} \;=\; X - \mathsf{E}X \;=\; a(Y - \mathsf{E}\,Y) \;=\; a\tilde{Y}.$$

Multiplying the last identity by $\tilde{Y}$ and taking expectations again then gives

$$\mathsf{Cov}(X,Y) \;=\; a\sigma_Y^2 \quad\Longrightarrow\quad a \;=\; \rho\frac{\sigma_X}{\sigma_Y}.$$

Now, suppose that the correlation coefficient of X and Y is either $+1$ or -1. From (1.2), we get

$$\mathsf{E}\Big(\tilde{X} - \rho\frac{\sigma_X}{\sigma_Y}\tilde{Y}\Big)^2 \;=\; 0.$$

The only distribution which has a second moment equal to 0 is the one which assigns probability one to the point $x = 0$. Hence, $\rho = \pm 1$ implies

$$X - \mathsf{E}X \;=\; \rho\frac{\sigma_X}{\sigma_Y}(Y - \mathsf{E}\,Y),$$

which implies (1.1).

We then conclude that (1) if $\rho = +1$, then $X - \mathsf{E}X$ and $Y - \mathsf{E}\,Y$ have the same sign, and (2) if $\rho = -1$, then $X - \mathsf{E}X$ and $Y - \mathsf{E}\,Y$ have opposite signs. It is then impossible for three random variables Y_1, Y_2, Y_3 to have pairwise correlation coefficients all equal to -1: if $\tilde{Y}_1 > 0$, then $\tilde{Y}_2 < 0$, which in turn implies $\tilde{Y}_3 > 0$, and the last inequality implies $\tilde{Y}_1 < 0$, a contradiction.

Let us try to make the same reasoning when $-1 < \rho < 0$. Formula (1.2) says that if $|\rho|$ is close to one, then the variance of

$$\tilde{X} - \rho\frac{\sigma_X}{\sigma_Y}\tilde{Y} \;=\; X - \mathsf{E}X - \rho\frac{\sigma_X}{\sigma_Y}(Y - \mathsf{E}Y)$$

is small. In other words, if ρ is sufficiently negative, then there is a high probability that the signs of $\tilde{X}$ and $\tilde{Y}$ are the same.

We can thus say that "if $\tilde{Y}_1 > 0$, then there is a high probability that $\tilde{Y}_2 < 0$, which means that there is a high probability that $\tilde{Y}_3 > 0$, and, if the last inequality holds, then there is a high probability that $\tilde{Y}_1 < 0$." This is not a contradiction, but it gives a hint that some correlation matrices may not be possible, if the correlation coefficients are "too negative." Whether a correlation matrix is possible can be determined by finding out if it is positive semidefinite. This is explained in Section 2. For instance, if three securities have pairwise correlation coefficients all equal to ρ, then the correlation matrix is

$$C = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix},$$

which is positive semidefinite if, and only if, $\rho \geq -\frac{1}{2}$. Thus, it is not possible to have all correlations equal to, say, $-.75$. This fact is extended to an arbitrary number of securities in Section 3.

A financial interpretation of the results of Sections 2 and 3 is that, given any set of n securities, it is mathematically impossible for the average correlation to be less than $-1/(n-1)$. This is a little disappointing for portfolio selection, but may be a partial explanation for the well-known fact that it is difficult to find negative correlations among actual securities.

All the random variables in this note are supposed to have finite second moments.

2. A GENERAL NECESSARY CONDITION

A symmetric $n \times n$ matrix C is said to be *positive definite* if, for all column vectors $x \in \mathbb{R}^n$, $x \neq 0$,

$$x^T C x > 0.$$

The matrix is said to be *positive semidefinite* (or *semipositive definite*) if $x^T C x \geq 0$ for all $x \in \mathbb{R}^n$.

Now, suppose C is the correlation matrix of a vector $Y = [Y_1, \ldots, Y_n]^T$, with variances σ_i^2, $i = 1, \ldots, n$. For $x \in \mathbb{R}^n$,

$$\mathsf{Var}\left(\sum_{i=1}^{n} x_i \frac{Y_i}{\sigma_i}\right) = x^T C x.$$

Since the variance is never negative, we see that a correlation matrix is always positive semidefinite. Conversely, any positive semidefinite matrix is the correlation matrix of some random vector (this is because a positive semidefinite matrix A has a "square root," that is, another positive definite matrix B such that $B^2 = A$).

For instance, in dimension two we find

$$x^T C x = [x_1, x_2] \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 2 r x_1 x_2 + x_2^2 = (x_1 + r x_2)^2 + (1 - r^2) x_2^2.$$

If $-1 \leq r \leq 1$, then this cannot be negative; if $|r| > 1$, then one can find x such that the expression is negative. Hence, the matrix is positive semidefinite if, and only if, $1 \leq r \leq 1$. This is not surprising. For higher dimensions it takes more effort to determine whether a matrix is positive semidefinite; a necessary and sufficient condition is given in Section 3.

We thus see that not all matrices with "1's" on the diagonal can be correlation matrices. The following necessary condition is a weaker form of the result of Section 3.

Theorem 1. *For $n \geq 2$ let $C = [\rho_{ij}]$ be any $n \times n$ correlation matrix, and define the average (off-diagonal) correlation coefficient as*

$$\overline{\rho} \;=\; \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \rho_{ij}.$$

Then $\overline{\rho} \;\geq\; -\dfrac{1}{n-1}$.

To prove this, let $x = [1, \ldots, 1]^T \in \mathbb{R}^n$. Then, since C is positive semidefinite,

$$0 \;\leq\; x^T C x \;=\; [1, \ldots, 1] \begin{bmatrix} \sum_{j=1}^n \rho_{1j} \\ \vdots \\ \sum_{j=1}^n \rho_{nj} \end{bmatrix} \;=\; \sum_{i=1}^n \sum_{j=1}^n \rho_{ij} \;=\; n + n(n-1)\overline{\rho},$$

which proves the theorem.

3. AVERAGE CORRELATION MODELS

As said before, in an average correlation model all off-diagonal correlation coefficients are the same number, say $\rho = r$.

Define

$$C^n = C^n(r) \;=\; \begin{bmatrix} 1 & r & \cdots & r & r \\ r & 1 & \cdots & r & r \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r & r & \cdots & 1 & r \\ r & r & \cdots & r & 1 \end{bmatrix}_{n \times n}$$

and $d_n = d_n(r) = \det(C^n)$. One immediately notes that d_n is 0 if $r = -1/(n-1)$, since the sum of the rows is then the 0 vector. The determinant of $C^n(r)$ is also 0 if $r = 1$.

Theorem 2. *For $n \geq 2$, $C^n(r)$ is positive semidefinite if, and only if, $-1/(n-1) \leq r \leq 1$, and*

$$d_n(r) \;=\; (1-r)^{n-1}[1 + (n-1)r].$$

The proof of this result will now be given. For an $n \times n$ matrix A, let A_i be the "leading principal submatrix" of A obtained by deleting rows and columns $i+1, \ldots, n$. Then a symmetric matrix $A_{n\times n}$ is positive semidefinite if, and only if, $\det A_i \geq 0$ for $i = 1, \ldots, n$; likewise, A is positive definite if, and only if, $\det A_i > 0$ for $i = 1, \ldots, n$ (for a proof of these facts, see Horn & Johnson

(1985), pp.403-404). The leading principal submatrices of $C^n(r)$ are precisely, $C^1(r), \ldots, C^{n-1}$, and so $C^n(r)$ is positive semidefinite if, and only if, $\det C^k(r) \geq 0$ for all $k = 1, \ldots, n$.

The determinant of $C^n(r)$ may be calculated directly for low values of n:

$$d_1(r) \;=\; 1, \quad d_2(r) \;=\; 1 - r^2, \quad d_3(r) \;=\; 1 - 3r^2 + 2r^3, \quad d_4(r) \;=\; 1 - 6r^2 + 8r^3 - 3r^4,$$

and so on. These all agree with Theorem 2, but a general proof has to be found for arbitrary n.

Consider the matrix

$$\widetilde{C}^n \;=\; \begin{bmatrix} r & r & \cdots & r & r \\ r & 1 & \cdots & r & r \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r & r & \cdots & 1 & r \\ r & r & \cdots & r & 1 \end{bmatrix}_{n\times n} .$$

The matrices C^n and $\widetilde{C}^n$ are the same except for the first element of the first row. By expanding their determinants along the first row, it can be seen that the difference between them is

$$\det C^n - \det \widetilde{C}^n \;=\; (1-r)\det C^{n-1}.$$

The determinant of $\widetilde{C}^n$ is easy to calculate, because determinants are unaffected when one row is multiplied by a constant and then added to another row. Hence

$$\det \widetilde{C}^n \;=\; \det \begin{bmatrix} r & r & \cdots & r & r \\ 0 & 1-r & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1-r & 0 \\ 0 & 0 & \cdots & 0 & 1-r \end{bmatrix}_{n\times n} \;=\; r(1-r)^{n-1}.$$

Hence, d_n satisfies the recurrence equation

$$d_n \;=\; (1-r)d_{n-1} + r(1-r)^{n-1}, \quad n \geq 2.$$

The solution of this equation is especially easy to find if one makes the change of variable

$$d_n \;=\; (1-r)^{n-1}\bar{d}_n \quad \Longrightarrow \quad \bar{d}_n \;=\; \bar{d}_{n-1} + r.$$

By iterating the last equation, we get

$$\bar{d}_n \;=\; \bar{d}_1 + (n-1)r \;=\; 1 + (n-1)r \quad \Longrightarrow \quad d_n \;=\; (1-r)^{n-1}[1 + (n-1)r].$$

We conclude that $C^n(r)$ is positive semidefinite if, and only if, $-1/(k-1) \leq r \leq 1$ for $k = 2, \ldots, n$, which is the same as the condition $-1/(n-1) \leq r \leq 1$. Similarly, $C^n(r)$ is positive definite if, and only if, $-1/(n-1) < r < 1$. This ends the proof.

REFERENCES

Elton, E., Gruber, M.J., Brown, S.J., and Goetzmann, W.N. (2003). *Modern Portfolio Theory and Investment Analysis (6th Ed.)*. John Wiley & Sons.

Horn, R.A, and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge University Press.

TWO ESSENTIAL FORMULAS IN ACTUARIAL SCIENCE AND FINANCIAL MATHEMATICS

Daniel Dufresne
Centre for Actuarial Studies
University of Melbourne

Abstract

The aim of this note is to give students a unified view of two formulas that are used in various contexts. Textbooks often give statements or proofs that are quite restrictive, and it may be useful to have a broader look at their proofs and applicability.

The first part of the note discusses Jensen's inequality, which allows one to compare a "function of the mean with the mean of a function," if the function is convex. Applications are given: geometric mean rate of return, life annuities, American options.

The second formula does not have a name, as far as I know, which is why I coined a rather ugly new one: the "survival function expectation formula." This formula allows one to write the expected value of a function of a random variable as the integral of the derivative of the function times one minus the cumulative distribution function. The applications given are: stop-loss premiums ("censored expectations"), stochastic dominance, life annuities.

1. INTRODUCTION

The Fundamental Theorem of Calculus implies that a differentiable function is the integral of its derivative:

$$\phi(b) \;=\; \phi(a) + \int_a^b \phi'(x)\,dx. \tag{1.1}$$

If $\phi''(x)$ exists for all x, then we may apply the same treatment to ϕ', to obtain

$$\begin{aligned}\phi(b) \;&=\; \phi(a) + \int_a^b \left[\phi'(a) + \int_a^x \phi'(y)\,dy\right]\,dx \\ &=\; \phi(a) + (b-a)\phi'(a) + \int_a^b \int_a^x \phi''(y)\,dy\,dx \\ &=\; \phi(a) + (b-a)\phi'(a) + R(b),\end{aligned} \tag{1.2}$$

where $R(b) = \int_a^b \int_a^x \phi''(y)\,dy\,dx$ is the remainder, or error term, for the first-order Taylor expansion of ϕ about a.

There are functions ϕ which satisfy a formula similar to (1.1), but that do not have a derivative at every point in the interval (a, b). More precisely, a function ϕ is said to be *absolutely continuous* if there exists ψ such that

$$\phi(b) \;=\; \phi(a) + \int_a^b \psi(x)\,dx, \quad a \le x \le b \tag{1.3}$$

(see Rudin, 1987, pp.145-146). Every differentiable function is absolutely continuous, but the converse is not true. For instance, the function $|x|$ satisfies (1.3) with

$$\psi(x) \;=\; \mathbf{1}_{\{x>0\}} - \mathbf{1}_{\{x<0\}}.$$

If (1.3) holds, then it can be shown that $\phi'(x)$ exists "almost everywhere" and that it is equal to $\psi(x)$ also "almost everywhere" (for the meaning of the last expression, see Rudin (1987, p.27), or any text on measure theory).

2. JENSEN'S INEQUALITY

Definition. *A function f defined on an interval $I = (a, b)$, $-\infty \leq a < b \leq \infty$, is convex if, for all $a < x, y < b$, and all $0 \leq \lambda \leq 1$,*

$$f(\lambda x + (1-\lambda)y) \;\leq\; \lambda f(x) + (1-\lambda)f(y).$$

Graphically, this says that, if $x < z < y$, then the point $(z, f(z))$ is below or on the line segment joining the points $(x, f(x))$ and $(y, f(y))$.

A *concave* function is defined in precisely the same way, except that the inequality is in the opposite direction. Hence, a function f is concave if, and only if, $-f$ is convex.

A question that arises from time to time is: which of $f(\mathsf{E}(Z))$ or $\mathsf{E}[f(Z)]$ is larger? There is a simple answer in the case of convex functions.

Theorem 1. **(Jensen's Inequality)** *Suppose f is a convex function over an open interval I, and that Z is a random variable with values in I. Then*

$$f(\mathsf{E}(Z)) \;\leq\; \mathsf{E}[f(Z)].$$

A trick to remember the direction of the inequality is to recall that the function $f(x) = x^2$ is convex, and that $\mathsf{Var}(Z) \geq 0$ actually means

$$f(\mathsf{E}Z) \;=\; [\mathsf{E}(Z)]^2 \;\leq\; \mathsf{E}(Z^2) \;=\; \mathsf{E}[f(Z)].$$

A twice differentiable function f is convex if, and only if, $f''(x) \geq 0$ for all $a < x < b$. This is the case for the functions x^2 and e^x. For these cases a first proof of Jensen's inequality will be given, based on a first-order Taylor expansion (this is similar to the proof often found in textbooks). Other functions, such as $(x-c)_+$ (if c is a constant) are not twice differentiable everywhere. A more general second proof will be given for those cases. A third proof is also given.

First Proof of Theorem 1. It has already been said that if a convex function f has a second derivative f'' then necessarily $f''(x) \geq 0$ for all x. Apply formula (1.2), with $\phi = f$, $a = \mathsf{E}(Z) = \mu$, $b = Z$:

$$f(Z) \;=\; f(\mu) + (Z-\mu)f'(\mu) + \int_{\mu}^{Z}\int_{\mu}^{x} f''(y)\,dy\,dx.$$

The last term (the remainder in Taylor's formula) is never negative, whatever the value of Z. Hence, taking expectations on both sides, the expectation of the remainder is non-negative, and

$$\mathsf{E}[f(Z)] \;\geq\; f(\mu) + \mathsf{E}[(Z-\mu)]f'(\mu) \;=\; f(\mu).$$

Second Proof of Theorem 1. If f is defined on an open interval I and is convex, then it is necessarily continuous (try to picture why). Moreover, it can be shown that such a function is also absolutely continuous, so that it is the integral of some function g:

$$f(b) \;=\; f(a) + \int_a^b g(y)\,dy.$$

Just as in cases where the derivative f' exists, the function g is non-decreasing (more correctly: it is always possible to find g which satisfies the equation above and is non-decreasing). We thus have

$$\begin{aligned} f(Z) \;&=\; f(\mu) + \int_a^Z g(y)\,dy \\ &=\; f(\mu) + (Z-\mu)g(\mu) + \int_\mu^Z [g(y) - g(\mu)]dy. \end{aligned}$$

Now the integral term in the last equation cannot be negative: if $Z > \mu$, then the integrand is non-negative, and if $Z < \mu$, then the integrand is non-positive, which makes the integral non-negative. By taking expectations on both sides, we see that Jensen's inequality holds.

Third Proof of Theorem 1. This is a standard proof, see for instance Shreve (2004, p.30) for more details. The proof rests on the following fact: if ϕ is convex, then, for every x,

$$\phi(x) \;=\; \max\{\ell(x) : \ell \text{ is a linear function with } \ell(y) \leq \phi(y)\ \forall y \in \mathbb{R}\}. \tag{2.1}$$

If ℓ is a linear function which lies below ϕ, then $\mathsf{E}[\phi(x)] \geq \mathsf{E}[\ell(x)] = \ell(\mathsf{E}\,X)$. Using (2.1), the maximum over all ℓ which lie below ϕ of $\ell(\mathsf{E}\,X)$ is $\phi(\mathsf{E}\,X)$, which ends the proof. □

It is often desirable to know whether the inequality in Theorem 1 is strict. If the function g in the second proof is strictly increasing, and if Z is not degenerate (that is, if Z takes more than a single value), then the inequality is strict. This is the case if f is twice differentiable and $f''(x) > 0$ for all x.

(*N.B*. To learn more about the function g above, refer to Rudin (1987, pp.61-62) and Rockafellar (1970). It can be either the left-hand or the right-hand derivative of f, both of which always exist if f is convex.)

Applications

Example 2.1. (*Geometric vs arithmetic mean rate of return*) If the arithmetic rate of return is R, the *geometric mean rate of return* is defined as

$$g_R \;=\; \exp\{\mathsf{E}[\log(1+R)]\} - 1.$$

In order to find which of $\mathsf{E}(R)$ and g_R is largest, apply Jensen's inequality, with $Z = \log(1+R)$ and $f(x) = e^x$, to see that

$$g_R \;\le\; \mathsf{E}(R). \tag{2.2}$$

If the distribution of R is not degenerate then the inequality is strict, since the second derivative of the exponential function is strictly greater than 0.

This is related to a more widely-known fact, that the geometric mean of n positive numbers $a_1, \ldots, a_n$ is never larger than their arithmetic average. To prove this, imagine that R has distribution

$$\mathsf{P}(R = a_i - 1) \;=\; \frac{1}{n}, \quad i = 1, \ldots, n.$$

Then

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \;=\; \exp\{\mathsf{E}[\log(1+R)]\}$$

and

$$\mathsf{E}(1+R) \;=\; \frac{1}{n}\sum_{i=1}^{n} a_i.$$

Inequality (2.2) is then rewritten as

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \;\le\; \frac{1}{n}\sum_{i=1}^{n} a_i \quad \forall\, a_1, \ldots, a_n > 0.$$

<u>*Example 2.2*</u>. (*Life annuities*) The present value, at geometric rate of interest δ, of a continuously paid t-year annuity certain is

$$\bar{a}_{\overline{t}|} \;=\; \int_0^t e^{-\delta s}\, ds \;=\; \frac{1 - e^{-\delta t}}{\delta}.$$

If $\delta > 0$, then the function above is concave, and, for any non-negative random variable T,

$$\mathsf{E}[\bar{a}_{\overline{T}|}] \;\le\; \bar{a}_{\overline{\mathsf{E}T}|}.$$

When $T = T_x$ represents the future lifetime of an individual currently aged x, the left-hand side is denoted "$\bar{a}_x$." The inequality is strict if T_x is not degenerate. Jensen's inequality thus shows that the present value of a life annuity is never larger than an annuity certain for a term equal to the life expectancy. The same result holds for annuities for a fixed term, just replace T_x with $T_x \wedge n$ in the proof above:

$$\bar{a}_{x:\overline{n}|} \;=\; \mathsf{E}[\bar{a}_{\overline{T_x \wedge n}|}] \;\le\; \bar{a}_{\overline{\mathsf{E}(T_x \wedge n)}|}.$$

(The life expectancy $\mathsf{E}T_x$ is denoted $\mathring{e}_x$, while the partial life expectancy $\mathsf{E}(T_x \wedge n)$ is denoted $\mathring{e}_{x:\overline{n}|}$.)

Exercise. Can you find a general inequality between

$$\bar{A}_x \;=\; \mathsf{E}e^{-\delta T_x} \text{ and } e^{-\delta \mathsf{E} T_x}?$$

<u>*Example 2.3*</u>. (*American options*) Consider a no-arbitrage market with a risk-free geometric rate of interest $r > 0$ (see Shreve (2004) and references therein). It is well-known that an American call option on some security S which does not pay dividends before maturity is worth more alive than exercised. Letting T be the maturity of the American option, one way to prove this is that the payoff of the corresponding European call is

$$(S_T - K)_+ \;\geq\; S_T - K,$$

and that the time-0 price of this payoff is $S_0 - Ke^{-rT}$, which is strictly larger than the intrinsic value of the American call at time 0, if the strike price $K > 0$.

Shreve (2004, pp.111-112) proves a more general version of this result, in a discrete-time framework. Consider an American option with intrinsic value $g(s)$ when the current price of the underlying security is s, the function g being convex and, moreover, satisfying $g(0) = 0$. Then it can be proved that the price of this American option, denoted V_0^A, is the same as the price of an otherwise identical European option, denoted V_0^E; in other words, it is never a profitable to exercise the American option before maturity.

As Shreve (2004) explains, if g takes negative values we can replace it with

$$g^+(s) \;=\; \max\{g(s), 0\}.$$

This function is also convex, and $g^+(0) = 0$. In the sequel, we write g for g^+.

Shreve's proof is based on the fact that the process $\{e^{-rt}g(S_t)\}$ is a submartingale (r is the risk-free rate of interest). We offer a different, "model-free" proof of this result, using the idea in the third proof of Jensen's inequality. It is assumed that there is no arbitrage, that the underlying security does not pay dividends, and that $r \geq 0$.

Consider a linear function ℓ which lies below g. Then

$$\ell(x) \;=\; ax + b$$

for some constants a and b. Since $g(0) = 0$, b must be smaller than or equal to 0. Hence, the payoff of the European option then satisfies

$$g(S_T) \;\geq\; \ell(S_T) \;=\; aS_T + b.$$

Because the model is arbitrage-free, and also because the security S does not pay dividends, the time-0 price of the payoff on the right is

$$aS_0 + be^{-rT} \;=\; aS_0 + b + b(e^{-rT} - 1) \;\geq\; \ell(S_0).$$

Hence, for each linear function ℓ which lies below g,

$$\text{time-0 price of payoff } g(S_T) \;=\; V_0^E \;\geq\; \ell(S_0).$$

This implies $V_0^E \;\geq\; g(S_0)$. Hence, there is no advantage in exercising the American option at time 0. We can replace "time 0" with any other time before maturity.

3. THE SURVIVAL FUNCTION EXPECTATION FORMULA

The survival function is one minus the distribution function. Suppose Z is a random variable which takes values in $[a, \infty)$, and also suppose that ϕ, ψ satisfy (1.3) for $b = Z$, for any value of Z. Then

$$\phi(Z) \;=\; \phi(a) + \int_a^Z \psi(x)\,dx \;=\; \phi(a) + \int_a^\infty \psi(x) 1_{\{Z>x\}}\,dx.$$

In the last integral, the indicator function $1_{\{Z>x\}}$ limits integration to $x \in (a, Z)$. An integral is the limit of a sum, and the expectation of a sum is the sum of the expectations. It is therefore easy to accept that

$$\mathsf{E}\int_a^\infty \psi(x)1_{\{Z>x\}}\,dx \;=\; \int_a^\infty \mathsf{E}\left[\psi(x)1_{\{Z>x\}}\right]\,dx \;=\; \int_a^\infty \psi(x)\mathsf{P}(Z>x)\,dx.$$

This formula is given in most actuarial textbooks, though often with unnecessarily restrictive assumptions. Observe that the identity holds whatever the type of distribution Z has (discrete, continuous, mixed, singular,...). (The justication for the last identity is Fubini's Theorem. There are technical restrictions for this to hold; in applications, they are virtually always satisfied, if $\mathsf{E}\,\phi(Z)$ exists. The theorem below gives one condition which guarantees that the formula holds, but the result is true in many other cases where this assumption does not hold.)

Theorem 2. **(The Survival Function Expectation Formula)** *Suppose that ϕ, ψ satisfy (1.3) for $b \in [a, \infty)$ and that $\psi(x)$ does not change sign. Let Z be a random variable which takes values in $[a, \infty)$ only. Then*

$$\mathsf{E}[\phi(Z)] \;=\; \phi(a) + \int_a^\infty \psi(x)\mathsf{P}(Z>x)\,dx.$$

In case where $\phi'(x)$ exists for all x, then $\psi = \phi'$ and this becomes

$$\mathsf{E}[\phi(Z)] \;=\; \phi(a) + \int_a^\infty \phi'(x)\mathsf{P}(Z>x)\,dx.$$

Applications

Example 3.1. (*Tail expectations*) Suppose X is any random variable. The new variable $Z = (X - c)_+ = \max(X-c, 0)$ is greater than or equal to 0, so let $a = 0$. Note that, for all $x > 0$,

$$\mathsf{P}((X-c)_+ > x) \;=\; \mathsf{P}(X > c + x).$$

Hence, using $\phi(x) = x$, we get

$$\mathsf{E}(X-c)_+ \;=\; \int_0^\infty \mathsf{P}(X > c+x)\,dx \;=\; \int_c^\infty \mathsf{P}(X > y)\,dy. \tag{3.1}$$

Observe that it is not required that c be greater than 0, and that X may take values in $(-\infty, \infty)$. Textbooks often give a proof of (3.1) which assumes that X has a pdf, but, as Theorem 2 shows, X DOES NOT HAVE TO BE OF CONTINUOUS TYPE FOR (3.1) TO HOLD. (Moreover, (3.1) holds even when either side of the equation is infinite.) In particular, if $X \geq 0$, then

$$\mathsf{E}X \;=\; \int_0^\infty \mathsf{P}(X > x)\,dx,$$

again irrespective of the type of distribution X has.

Another common application of (3.1) is, in cases where

$$\mathsf{P}(X \in (c, c')) \;=\; 0,$$

to write

$$\mathsf{E}(X-c)_+ - \mathsf{E}(X-c')_+ \;=\; \int_c^{c'} [1 - F_X(y)]\,dy \;=\; (c'-c)(1-F_X(c)).$$

In the case where X takes the values $n = 0, 1, \ldots$, only, we see that $F_X(y)$ is constant and equal to $F_X(n)$ for $y \in (n, n+1)$, and so

$$\int_n^{n+1} \mathsf{P}(X > y)\,dy \;=\; \mathsf{P}(X > n) \;=\; 1 - F_X(n). \tag{3.2}$$

This implies

$$\mathsf{E}\,X \;=\; \sum_{n=0}^{\infty} [1 - F_X(n)].$$

If X is integer-valued, then (3.2) implies that the following recursive equation holds:

$$\mathsf{E}(X-n-1)_+ \;=\; \mathsf{E}(X-n)_+ - [1 - F_X(n)], \quad n = 0, 1, \ldots$$

Exampe 3.2. (*Stochastic dominance*) There is *stochastic dominance of* X over Y if

$$F_X(z) \;\leq\; F_Y(z) \quad \forall z$$

and if there is some z_0 for which the inequality is strict. (A little reflection shows that if there is z_0 where $F_X(z_0) < F_Y(z_0)$, then necessarily the same holds for z in some interval.) We will use the notation

$$X \overset{st}{\geq} Y$$

to indicate that X dominates Y stochastically. The interpretation is that the distribution of X is concentrated on larger values than the distribution of Y.

Suppose (1) $X \overset{st}{\geq} Y$, (2) $X, Y \geq a$, (3) u is a function with a positive derivative u'. Applying Theorem 2, we get

$$\mathsf{E}[u(X)] - \mathsf{E}[u(Y)] \;=\; \int_0^\infty u'(x)[F_Y(x) - F_X(x)]dx \;>\; 0. \tag{3.3}$$

When u is a utility function this is called the "First-Order Stochastic Dominance Theorem" (see Elton *et al*., 2003, Chapter 11).

Stochastic dominance is a weaker condition than $X \geq Y$. However, if $X \overset{st}{\geq} Y$, and one is comparing expectations of functions X only with expectations of functions of Y only (as in (3.3)), then one may proceed just as though the stronger condition $X \geq Y$ held, because of the following result:

Lemma. *If $X \overset{st}{\geq} Y$, then there exists a probability space on which are defined random variables X', Y' such that*

(i) the distribution of X' is the same as that of X;

(ii) the distribution of Y' is the same as that of Y;

(iii) $X' \geq Y'$ and $\mathsf{P}(X' > Y') > 0$.

Proof. Suppose **U**$\sim$ $(0, 1)$, and define

$$X' \;=\; F_X^{-1}(U), \quad Y' \;=\; F_Y^{-1}(U).$$

Here the inverse of a distribution function F is defined as

$$F^{-1}(u) \;=\; \inf\{x \,:\, F(x) > u\}.$$

The proof of (i) and (ii) is well-known (for example, in simulation). Part (iii) follows readily. $\square$

For instance, the First-Order Stochastic Dominance Theorem is an immediate consequence of this lemma, since

$$\mathsf{E}[u(X)] \;=\; \mathsf{E}[u(X')] \;>\; \mathsf{E}[u(Y')] \;=\; \mathsf{E}[u(Y)].$$

<u>*Example 3.3*</u>. (*Life annuities*) Suppose K_x is a random variable which takes the values $0, 1, \ldots$ only, representing the number of complete years lived by an individual currently aged x. An n-year annuity certain payable annually in arrears has present value

$$a_{\overline{n}|} \;=\; \frac{1 - v^n}{i}, \quad v = (1+i)^{-1}.$$

The actuarial present value of a life annuity (paid at the end of each year) from age x is defined as

$$a_x \;=\; \mathsf{E}[a_{\overline{K_x}|}].$$

The usual formula for the life annuity can be found by applying Theorem 2. If $\delta = \log(1+i)$, then

$$\begin{aligned}
\mathsf{E}[a_{\overline{K_x}|}] &= \int_0^\infty \frac{\delta}{i} v^t \mathsf{P}(K > t)\,dt \qquad (3.4)\\
&= \sum_{n=0}^\infty \frac{\delta}{i} \int_n^{n+1} v^t\,dt\,\mathsf{P}(K_x > n)\\
&= \sum_{n=0}^\infty v^{n+1}\mathsf{P}(K_x > n)\\
&= \sum_{n=0}^\infty v^{n+1}{}_{n+1}p_x.
\end{aligned}$$

This of course also implies

$$\ddot{a}_x \;=\; \sum_{k=0}^\infty v^k{}_n p_x.$$

A similar analysis applies to continuously paid annuities, yielding, for instance

$$\bar{a}_x \;=\; \mathsf{E}[\bar{a}_{\overline{T_x}|}] \;=\; \int_0^\infty v^t{}_t p_x\,dt.$$

Formulas for life annuities for a limited term n are found in the same way, by replacing K_x (resp. T_x) with $K_x \wedge n$ (resp. $T_x \wedge n$). Formulas for assurances may also be found. For instance, $\phi(y) = v^{y+1}$ yields

$$A_x \;=\; \mathsf{E}[v^{K_x+1}] \;=\; v - \delta\int_0^\infty v^{t+1}\mathsf{P}(K_x > t)\,dt.$$

From (3.4), this means that, if $d = iv$,

$$A_x \;=\; v - iva_x \;=\; v - iv(\ddot{a}_x + 1) \;=\; 1 - d\ddot{a}_x.$$

In the same way, one obtains

$$\bar{A}_x \;=\; 1 - \delta\bar{a}_x.$$

REFERENCES

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997). *Actuarial Mathematics* (Second Edition). Society of Actuaries, Itasca, Illinois.

Elton, E.J., Gruber, M.J., Brown, S.J., and Goetzmann, W.N. (2003). *Modern Portfolio Theory and Investment Analysis* (6th Edition). Wiley, New York.

Rockafellar, R.T., (1970). *Convex analysis*. Princeton University Press, 1970.

Rudin, W. (1987). *Real and Complex Analysis* (Third Edition). McGraw-Hill, New York.

Shreve, S. (2004). *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*. Springer, New York.