TR1509-1: The Mode Augmented Boxcar (MAB) Distribution

George Rebane, v31dec15

Refs : [1] PD101 report: Rebane, G.J., TR0405-1: New Approaches to Securities Modeling (CatMAB) and Simplified Portfolio Design (PD101), 2004.
[2] BudgetRisk report: Rebane, G.J., TR1508-1: BudgetRisk, v30dec15

1 Introduction

The Mode Augmented Boxcar (MAB) distribution was developed as an accessible planning tool for capturing, expressing, and computing subjective probability assessments for investment portfolio design (Ref 1). Since its inception as a financial engineering tool, MAB applications distributions have been expanded into the areas of generating risk conversant budgets and cost estimates (Ref 2).

MAB is defined by four parameters that are easily understood and generated by anyone involved in the quantitative estimation of future revenues, costs, returns, etc that by their very nature have non-deterministic values that are, at best, known only to within a probability distribution. This makes it clear that such line item values are really random variables that will be realized from a distribution that represents the 'domain expert's' (DE's) current assessment of future uncertainty and, therefore, risk.

A MAB is fully specified by four parameters consisting of the line item's potential low and high boundary values, its mode or most likely value, and a modal confidence measure expressing the DE's belief that the realized value will be at or close to the mode. A moment's reflection confirms that every DE considers such factors in the traditional exercise of selecting the single scalar that summarizes his uncertain knowledge that surrounds each line item in the budget or income/cost estimate. The MAB's defining 4-tuple described above formalizes and captures this thought process.

The following development of the MAB probability density function is based on the work reported in [1]. The analytical expression of the MAB cumulative distribution was motivated by the subsequent work reported in [2].

2 Mathematical Development

From statistics we recall that a probability density function (pdf) describes the behavior of a random variable in terms of which values of that r.v. we may expect to see. In Figure 1, if we let x be the values of a specific line item, then $p(x)$ is its pdf over a given range of anticipated values. Suppose all we are certain of is that the r.v. representing the future lies somewhere between a low value and a high value as shown in Figure 1, and that we have no other knowledge/belief about how it will be realized. In short, we are totally ignorant as to where the line item will turn out between the low and the high. Such a state of knowledge is represented by the well-known uniform or 'boxcar' pdf which has the absolute simplest structure/shape (and therefore minimum information content) between the stated limits. Very often such a boxcar distribution is literally the 'shape of our knowledge' about many realworld things.[1]

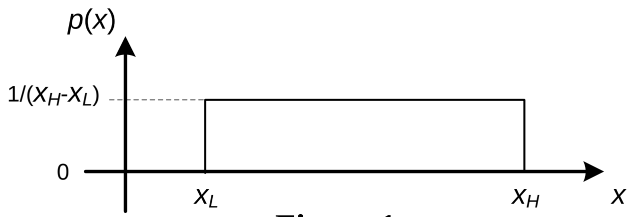

Figure 1

Probability density functions allow us to compute the probability that the value of the represented r.v. will be between two different values (say, \$5,000 and \$5,700) no matter how close together they lie. The actual probability is equal to the area under the $p(x)$ curve between the specified values. Since all probabilities lie between zero (the impossible event) and one (the certain event), and since we are certain that our line item will be realized between x_L and x_H, we know that the area under the total $p(x)$ function in Figure 1 must be unity. Therefore, since the boxcar is a rectangle, its height must be $1/(x_H - x_L)$. This gives us the simple expression for the pdf of a boxcar, $p(x) = [U(x_L) - U(x_H)]/(x_H - x_L)$ where $U(\ )$ is the well-known unit step function.[2]

However, in many cases the DE also believes that the line item's value may achieve a most likely price, cost or even a percent change/discount. This most likely value x_M represents the mode, or highest value, of $p(x)$ and is the basis for the naming such distributions MAB. In such cases the MAB pdf allows the DE to also specify x_M and the strength of his/her confidence of the location and prominence of this most likely value

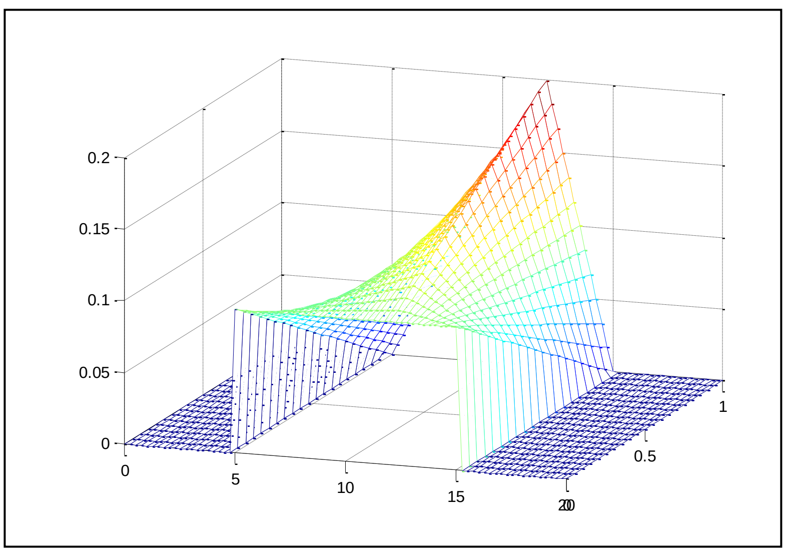

Figure 2

[1] Using such a distribution between limits allows us to say that we are totally ignorant of anything more than the r.v. will certainly fall somewhere within those limits.

[2] $U(x_0) = 0$ for $x < x_0$, and $U(x_0) = 1$ for $x \geq x_0$.

by supplying a positive confidence factor C that ranges from zero to unity – i.e. $0 < C \le 1$. Setting $C = 0$ is the same as having no confidence in the value of x_M which, of course, is the same as specifying the flat boxcar distribution. Alternatively, setting $C = 1$ shows the highest support for the selected mode and turns the shape of the MAB into a triangle distribution with its apex at x_M. Figure 2 shows how a typical MAB density changes shape as C is varied. In this figure $x_L = 5$, $x_H = 15$, and $x_M = 12$. C varies from 0 to 1. A detailed description of the MAB distribution, along with its mathematical formalities, follows.

2.1 The MAB pdf and moments

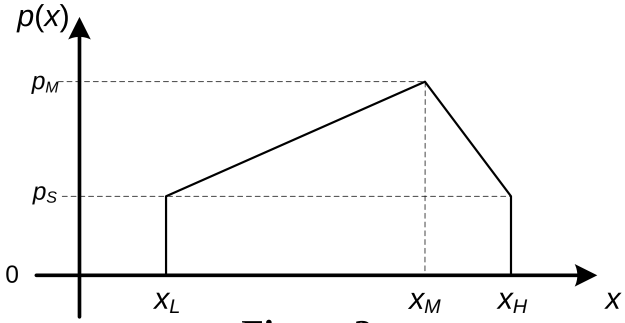

Figure 3

Figure 3 presents the generalized MAB probability density function (pdf). We note that the boxcar rectangular shape now takes on the silhouette of a house with a peak (mode) at height p_M and 'shoulders' with height p_S. The pdf is expressed mathematically as

$$p(x) = [m_L x + b_L][U(x_L) - U(x_M)] + [m_H x + b_H][U(x_M) - U(x_H)]. \tag{1}$$

The pdf is the sum of two straight lines only one of which is operative for any value of x. The integral of $p(x)$ over its non-zero range gives the area under the curve that must equal unity. This integral is easy to evaluate geometrically in terms of its mode and shoulder heights.

$$\begin{aligned} \int_{x_L}^{x_H} p(x)\,dx &= (x_H - x_L)\,p_S + \frac{(x_H - x_L)(p_M - p_S)}{2} \\ &= \frac{(x_H - x_L)(p_M + p_S)}{2} = 1 \end{aligned} \tag{2}$$

We note that as $p_M \to p_S$ the mode disappears and the pdf assumes the shape of the standard boxcar shown in Figure 1. From (2) we can derive a number of useful relationships.

$$p_M = \frac{2}{x_H - x_L} - p_S \tag{3}$$

It is apparent that the mode assumes a maximum when the shoulder height goes to zero, and alternately takes on its minimum value at the shoulder (which gives us the modeless boxcar).

$$p_{M,\max} = \frac{2}{x_H - x_L}, \qquad p_{M,\min} = \frac{1}{x_H - x_L} = p_S \tag{4}$$

Equations (4) suggest the existence of a natural modal confidence measure $C \in [0,1]$ in terms of which the generalized modal height can be expressed as

$$p_M = \frac{1+C}{x_H - x_L}\,. \tag{5}$$

Substituting this into (3) and solving for p_S gives the equivalent formula for the generalized shoulder height,

$$p_S = \frac{1-C}{x_H - x_L}\,. \tag{6}$$

These symmetrical relations demonstrate how the MAB distribution can capture the wide latitude of DE beliefs in easy to visualize shapes that range from pure (right to isosceles) triangles, through shouldered triangles, to the humble boxcar by the specification of the four parameters x_L, x_H, x_M, and C. Armed with this concise description of the future we may now proceed to compute the statistics needed for its use in various applications, financial and otherwise.

For both convenience and insight we define the following positive intervals for the MAB.

$$\Delta x_L = x_M - x_L, \quad \Delta x_H = x_H - x_M, \quad \Delta x = x_H - x_L = \Delta x_L + \Delta x_H \tag{7}$$

This lets us write the slopes and intercepts needed in (1) in terms of the input parameters.

$$m_L = \frac{p_M - p_S}{\Delta x_L} = \frac{2C}{\Delta x \Delta x_L}, \quad b_L = p_S - m_L x_L = \frac{1}{\Delta x}\left[1 - C\left(1 + \frac{2x_L}{\Delta x_L}\right)\right] \tag{8}$$

$$m_H = \frac{p_S - p_M}{\Delta x_H} = \frac{-2C}{\Delta x \Delta x_H}, \quad b_H = p_M - m_H x_M = \frac{1}{\Delta x}\left[1 + C\left(1 + \frac{2x_M}{\Delta x_H}\right)\right] \tag{9}$$

And the probability that the random line item variable will be less/more than or equal to the modal return is shown to be simply the ratios of the above defined intervals for all C.

$$P\left(x_L \le x \le x_M\right) = p_S \Delta x_L + \frac{\left(p_M - p_S\right)\Delta x_L}{2} = \frac{\Delta x_L}{\Delta x}, \;\; P\left(x_M \le x \le x_H\right) = 1 - \frac{\Delta x_L}{\Delta x} = \frac{\Delta x_H}{\Delta x} \tag{10}$$

We can now compute the mean μ_x of the MAB by using the above parameters in

$$\mu_x = \int_{x_L}^{x_H} x p(x)\,dx = \int_{x_L}^{x_M} x\left(m_L x + b_L\right)dx + \int_{x_M}^{x_H} x\left(m_H x + b_H\right)dx \tag{11}$$

to obtain

$$\boxed{\mu_x = \frac{x_L + x_H}{2} + \frac{C}{6}\left(\Delta x_L - \Delta x_H\right)} \tag{12}$$

Here the first r.h.s term is the familiar boxcar mean. The second term appropriately biases the mean toward the direction of the mode. Note that the mean is unchanged if the mode, no matter how pronounced, is located exactly at the center (i.e. when $\Delta x_L = \Delta x_H$).

To compute the variance σ_x^2 of the MAB we will use the form $\sigma_x^2 = \mathrm{E}(x^2) - \mu_x^2$. This requires us to compute the expectation

$$E\left(x^2\right) = \int_{x_L}^{x_H} x^2 p\left(x\right) dx = \int_{x_L}^{x_M} x^2 \left(m_L x + b_L\right) dx + \int_{x_M}^{x_H} x^2 \left(m_H x + b_H\right) dx \,. \tag{13}$$

Solving the integrals and substituting from (7)-(9) gives

$$E\left(x^2\right) = \frac{x_H^3 - x_L^3}{3\Delta x} + \frac{C}{6}\left[\left(\Delta x_L - \Delta x_H\right)\left(x_L + x_H\right) - \Delta x_L \Delta x_H\right]. \tag{14}$$

Finally, subtracting the square of the mean (12) and performing the necessary algebra gives us the desired compact form for variance.

$$\boxed{\sigma_x^2 = \frac{\Delta x^2}{12} - \frac{C}{6}\left[\Delta x_L \Delta x_H + \frac{C}{6}\left(\Delta x_H - \Delta x_L\right)^2\right]} \tag{15}$$

We again note that the first r.h.s. term is the familiar boxcar variance. And since the second term is never negative, any positive C value that allows the formation of a modal peak immediately begins to 'bunch' the probability mass and, therefore, diminish the variance as expected. It is clear that variance is a minimum for all values of C when the mode is at the center (i.e. $\Delta x_L = \Delta x_H$). It is also easy to show that the maximum variance contraction is $\sigma_{x,min}/\sigma_x = 0.707$ as C varies from 0 to 1. Similarly, the maximum possible displacement of the mean from center as C varies from 0 to 1 is $\Delta x/6$.

2.2 The MAB Cumulative Distribution

Understanding and/or vetting a DE's MAB for a line item requires that the distribution can be queried to discover the probabilities that the r.v. will assume values in certain ranges. This can most readily be done through the standardized use of the MAB's cumulative distribution which we develop here. Formally the MAB's cumulative distribution is defined as

$$F(x)=\int_{x_L}^{x} p(\xi)d\xi=\int_{x_L}^{x}\{(\xi\le x_M)[m_L\xi+b_L]+(\xi>x_M)[m_H\xi+b_H]\}d\xi$$
$$=(\xi\le x_M)\int_{x_L}^{x}[m_L\xi+b_L]d\xi+(\xi>x_M)\left\{F(x_M)+\int_{x_M}^{x}[m_H\xi+b_H]d\xi\right\} . \tag{16}$$

Here we have rewritten the MAB pdf from (1) in a more compact form using logicals that evaluate to 0 or 1 for the variable of integration since we always will require that $x\in[x_L,x_H]$. This allows (16) to be expressed as the sum of two integrals which will be evaluated separately.

$$F(x\le x_M)=\int_{x_L}^{x}[m_L\xi+b_L]d\xi=\left[\frac{m_L}{2}\xi^2+b_L\xi\right]\Bigg|_{x_L}^{x}$$
$$=\frac{1}{\Delta x}\left\{\frac{C}{\Delta x_L}\left(x^2-x_L^2\right)+\left[1-C\left(1+\frac{2x_L}{\Delta x_L}\right)\right](x-x_L)\right\} \tag{17}$$

After some algebraic acrobatics using previously defined terms, the above can be expressed compactly as

$$\boxed{F(x\le x_M)=\frac{x-x_L}{\Delta x}\left[C\left(\frac{x-x_M}{\Delta x_L}\right)+1\right]} \tag{18}$$

From (18) we derive the needed $F(x_M)$ in the second part of (16) by setting $x = x_M$. This gives

$$F(x_M)=\frac{x_M-x_L}{\Delta x}\left[C\left(\frac{x_M-x_M}{\Delta x_L}\right)+1\right]=\frac{\Delta x_L}{\Delta x} . \tag{19}$$

We finish the derivation by evaluating the second part of the r.h.s. of (16) that is valid for when $x\in(x_M,x_H]$.

$$F(x>x_M)=F(x_M)+\int_{x_M}^{x}[m_H\xi+b_H]d\xi=F(x_M)+\left[\frac{m_H}{2}\xi^2+b_H\xi\right]\Bigg|_{x_M}^{x}$$
$$=F(x_M)+\frac{1}{\Delta x}\left\{\frac{C}{\Delta x_H}\left(x^2-x_M^2\right)+\left[1+C\left(1+\frac{2x_M}{\Delta x_H}\right)\right](x-x_M)\right\} \tag{20}$$

Substituting from (19) and performing more algebra yields the desired compact form

$$\boxed{F(x>x_M)=\frac{1}{\Delta x}\left\{\Delta x_L+(x-x_M)\left[C\left(\frac{x_H-x}{\Delta x_H}\right)+1\right]\right\}} \tag{21}$$

Again we may check the correctness of (21) by letting x assume the various boundary values which yield the familiar and known cumulative probabilities. We conclude by collecting (18) and (21) into an easily programmable expression that is also suitable for incorporating into spreadsheet formats.

$$\boxed{\begin{aligned} F(x) &= \frac{(x \le x_M)}{\Delta x}\left\{(x - x_L)\left[C\left(\frac{x - x_M}{\Delta x_L}\right) + 1\right]\right\} \\ &\quad + \frac{(x > x_M)}{\Delta x}\left\{\Delta x_L + (x - x_M)\left[C\left(\frac{x_H - x}{\Delta x_H}\right) + 1\right]\right\} \end{aligned}} \qquad (22)$$

We conclude with Figure 4 showing three example plots of the MAB cumulative function for parameters $x_L = 5$, $x_H = 15$, and $x_M = 12$ for $C = 0$ (boxcar), 0.5, 1.0. Note that all three intersect at their common inflection point at x_M where $F(x_M) = \Delta x_L / \Delta x = 7/10 = 0.7$ as required by (22).