## TR1908-1: The Extended MAB (EMAB) Distribution

George Rebane, 20 August 2019 (V20nov19)

Refs: 1. Rebane, G.J., TR1509-1: The Mode Augmented Boxcar (MAB) Distribution, September 2015 v26aug18.

2. Rebane, G.J., TR1601-1: The Impulse Extended MAB (IMAB) Distribution, 7 January 2016.

3. Rebane, G.J., 'EMAB – Extended MAB', 15 July 2019 in 'Investment Analytics Notes'

# 1 Introduction

The ability to quantitatively express uncertainty based on subjective and less rigorous estimates of random variables was addressed in the development of the Mode Augmented Boxcar (MAB) distribution described in [1]. The MAB allows its user to specify the distribution with four easily understood parameters ($x_L$, $x_H$, $x_M$, $C$) that augment the uniform or 'boxcar' distribution by adding an emphasized mode within [$x_L$, $x_H$], its low/high range. The $C$ parameter expresses the user's confidence that the r.v. $x$ will fall in the vicinity of the mode at $x_M$. Given that $x$ may actually fall outside [$x_L$, $x_H$], the user is advised to set these limits into the narrowest range so that the resulting MAB contains, say, 95% or more of the probability mass.

Among other applications, the MAB has been used to effectively express the future performance of securities, uncertainties in upcoming revenues and costs, and the development of more risk revealing budgets in general. Its uses have supplied much needed information for assessing risk inherent in subjective inputs provided by 'domain experts' who collaborate on projects in which it is required to understand the probability distributions of aggregated sums of money. The IMAB [2] was developed specifically to extend subjective risk assessments into situations that require so-called mixture models of probability – e.g. the IMAB captures the uncertainty in the amount of a future order, given that it is placed, along with ability to also include the probability that the order will not be received.

Here we extend MAB so that it will contain the 'complete' probability mass of unity starting with the four MAB parameters, and asking the estimator or domain expert to include the probability $P$ that the r.v. lies within [$x_L$, $x_H$]. Given this additional data we form the Extended MAB (EMAB) p.d.f. by appending exponentially diminishing tails to MAB's shoulders so that its integral from [-∞,∞] equals unity. The resulting EMAB is shown in Figure 1 wherein the location of its mean is indicated with the dark blue line and the standard deviation from the mean is shown with the turquoise lines.

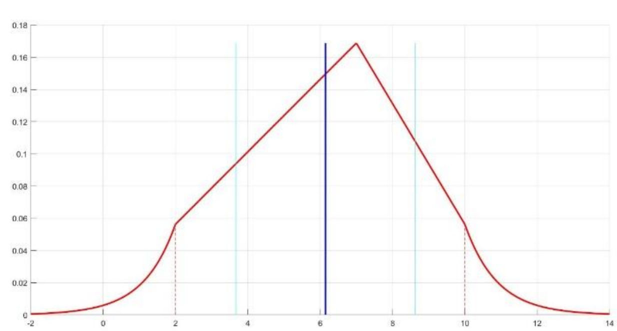


Figure 1

## 2 Mathematical Development

### 2.1 EMAB PDF and Moments

In this report we develop the expressions for the the EMAB's density and cumulative distributions, its mean and variance, along with a programmable model that generates EMAB samples. The derivations will be abbreviated with more detail provided in [1] and [3].

Figure 2 is the EMAB annotated with terms that fully illustrate and parametrize its p.d.f. Here we note that the extended tails each contain one half of the residual probability mass not collected under the MAB itself. The exponential decay constant *a* is chosen so that the tails meet the MAB at its shoulder height as derived in [1].

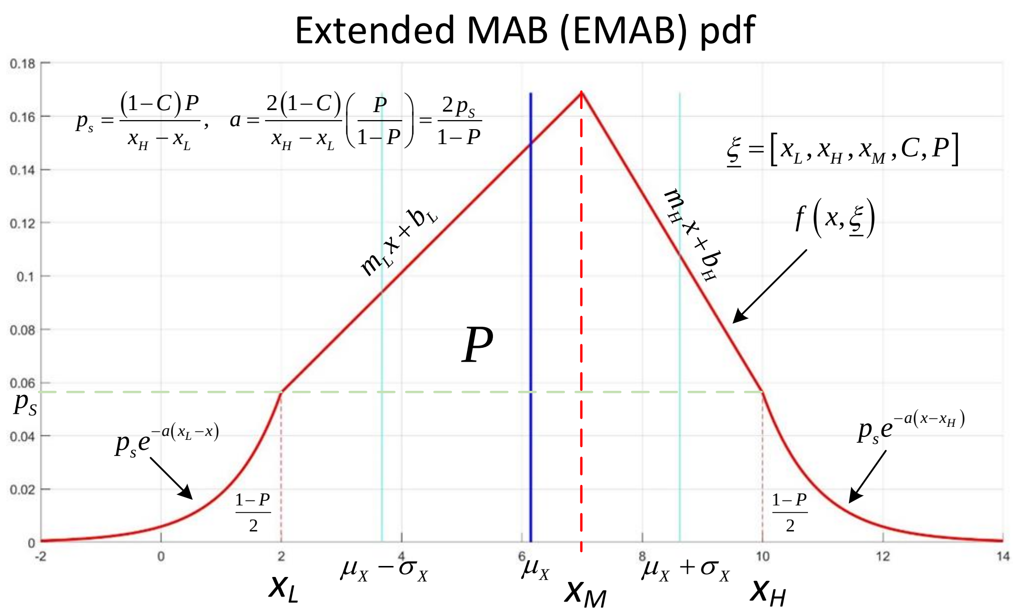


**Figure 2**

As seen from the figure, the entire EMAB is defined by specifying its parameter vector $\underline{\xi}$. Writing out the relations for the MAB slopes and intercepts allows the rest of the EMAB parameters to be developed. The piecewise p.d.f. definition is developed from Figure 2.

$$
\begin{aligned}
f(x, \underline{\xi}) &= \left\{p_S e^{-a(x_L - x)}\right\}(-\infty \le x \le x_L)\ldots \\
&+ \{m_L x + b_L\}(x_L < x \le x_M)\ldots \\
&+ \{m_H x + b_H\}(x_M < x \le x_H)\ldots \\
&+ \left\{p_S e^{-a(x - x_H)}\right\}(x_H < x \le \infty)
\end{aligned}
\tag{1}
$$

where

$$\Delta x_L = x_M - x_L, \quad \Delta x_H = x_H - x_M, \quad \Delta x = x_H - x_L = \Delta x_L + \Delta x_H$$

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

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

The EMAB mean is computed from the usual

$$\mu_{EMAB} = \int_{-\infty}^{\infty} x f\left(x, \underline{\xi}\right) dx \ . \tag{3}$$

Using $\int x e^{ax} dx = \frac{e^{ax}}{a^2}\left(ax - 1\right)$ , gives us

$$\mu_{EMAB} = \left\{-\frac{p_S}{a^2} e^{ax_L}\left(ax_L + 1\right)\right\} + \left\{m_L \frac{\left(x_M^2 - x_L^2\right)}{2} + b_L\left(x_M - x_L\right)\right\} + \ldots$$
$$\left\{m_H \frac{\left(x_H^2 - x_M^2\right)}{2} + b_H\left(x_H - x_M\right)\right\} + \left\{-\frac{p_S}{a^2} e^{ax_H}\left(ax_H + 1\right)\right\} \tag{4}$$

which compacts to

$$\boxed{\mu_{EMAB} = \frac{x_L + x_H}{2} + \frac{PC}{6}\left(\Delta x_L - \Delta x_H\right)} \tag{5}$$

Note the similarity of the EMAB mean to that of the MAB given in [1]. Here the shift of the mean away from the standard boxcar's mean is weighted by the probability *P* assigned to the core MAB in the EMAB distribution.

The EMAB variance is again computed from $\sigma_x^2 = \mathrm{E}(x^2) - \mu_x^2$, and requires the use of

$$\int x^2 e^{bx} dx = \frac{1}{b^3}\left[e^{bx}\left(b^2 x^2 - 2bx + 2\right)\right] \ . \tag{6}$$

Then

$$E\left(x^2\right) = \int_{-\infty}^{\infty} x^2 f\left(x, \underline{\xi}\right) dx$$
$$= \underbrace{\frac{p_S}{a^3}\left[a^2\left(x_H^2 + x_L^2\right) + 2a\left(x_H - x_L\right) + 4\right]}_{\text{Extended tails contribution}} + \ldots \tag{7}$$
$$\frac{1}{4}\left[m_L\left(x_M^4 - x_L^4\right) + m_H\left(x_H^4 - x_M^4\right)\right] + \frac{1}{3}\left[b_L\left(x_M^3 - x_L^3\right) + b_H\left(x_H^3 - x_M^3\right)\right]$$

Now substituting for the constants, and noting that variance should be a function of only the 'spread' or Δ-terms, and not the location of the p.d.f., we therefore set $x_L = 0$ which then makes $x_M = \Delta x_L$ and $x_H = \Delta x$. Now using these definitions, we can write (7) as

$$\begin{aligned} E\left(x^2\right) &= \frac{p_S}{a^3}\left[a^2\Delta x^2 + 2a\Delta x + 4\right] + \ldots \\ &\frac{1}{4}\left[m_L \Delta x^4 + m_H\left(\Delta x^4 - \Delta x_L^4\right)\right] + \frac{1}{3}\left[b_L \Delta x_L^3 + b_H\left(\Delta x^3 - \Delta x_L^3\right)\right] \end{aligned} \tag{8}$$

And from (5), after some rearrangement, we have

$$\mu_{EMAB}^2 = \frac{\Delta x^2}{4} + \frac{PC}{6}\left(\Delta x_L - \Delta x_H\right)\left[\Delta x + \frac{PC}{6}\left(\Delta x_L - \Delta x_H\right)\right] \tag{9}$$

This yields the desired EMAB variance

$$\boxed{\begin{aligned} \sigma_{EMAB}^2 &= E\left(x^2\right) - \mu_{EMAB}^2 \\ &= \frac{p_S}{a^3}\left[a^2\Delta x^2 + 2a\Delta x + 4\right] + \frac{1}{4}\left[\left(m_L + m_H\right)\Delta x^4 - m_H \Delta x_L^4\right] + \ldots \\ &\frac{1}{3}\left[\left(b_L - b_H\right)\Delta x_L^3 + b_H \Delta x^3\right] - \left\{\frac{\Delta x^2}{4} + \frac{PC}{6}\left(\Delta x_L - \Delta x_H\right)\left[\Delta x + \frac{PC}{6}\left(\Delta x_L - \Delta x_H\right)\right]\right\} \end{aligned}} \tag{10}$$

To demonstrate the correctness of (10), we would expect that, as most of the EMAB probability mass accumulates within the MAB interval $[x_L, x_H]$, we should then have $\sigma_{EMAB}$ approach the value of $\sigma_{MAB}$ where its variance from [1] is given below. At $P = 1$, when the probability mass in the tails disappears, the EMAB and MAB variances would be equal. To demonstrate this, it is not enough to just substitute $P = 1$ into (10), care must be taken to correctly handle the tails as $P \to 1$ in the limit. In this limit the tails' exponential decay constant $a \to \infty$, which makes the tails' contribution shown in (7) disappear. The remaining $E(x^2)$ formula is then identical to that obtained for calculating the MAB variance.

$$\sigma_{MAB}^2 = \frac{\Delta x^2}{12} - \frac{C}{6}\left[\Delta x_L \Delta x_H + \frac{C}{6}\left(\Delta x_H - \Delta x_L\right)^2\right] \tag{11}$$

This development is confirmed in Figure 3 plot below which illustrates $\sigma_{EMAB} \xrightarrow{P\to 1} \sigma_{MAB}$ .

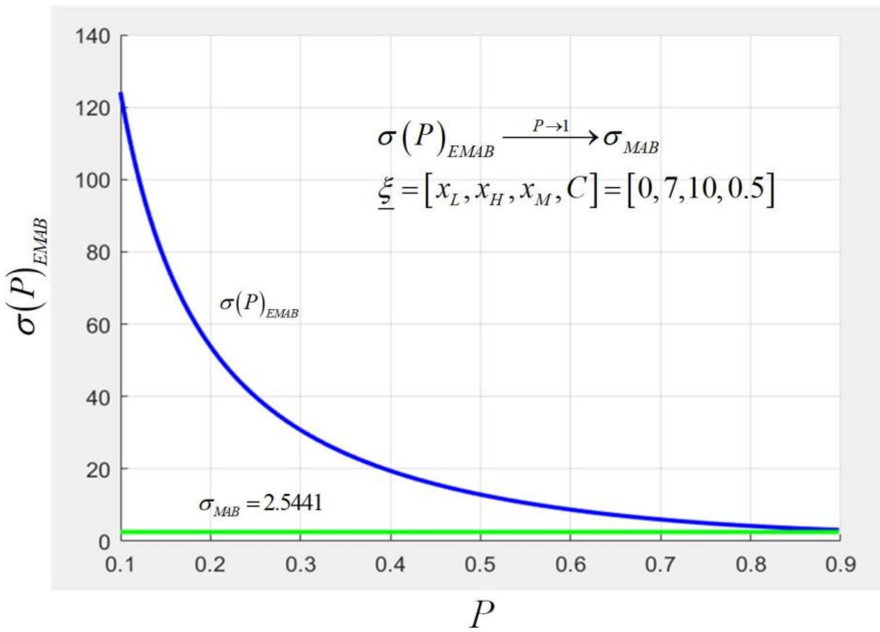


**Figure 3**

### *2.2 The EMAB Cumulative Distribution*

The cumulative distribution function, c.d.f., is defined in the standard way as $F(x) = \int_{-\infty}^{x} f(\eta) d\eta$ where the EMAB p.d.f. is given in (1). Since $f$ is defined piecewise, $F(x)$ is also calculated piecewise using the appropriate programmable logicals to 'activate' only the needed pieces.

$$
\begin{aligned}
F(x) = \int_{-\infty}^{x} f(\eta) d\eta = \int_{-\infty}^{x} p_S e^{-a(x_L - x)} d\eta \left(-\infty \le x < x_L\right) + \ldots \\
\left\{ \frac{1-P}{2} + \int_{x_L}^{x} (m_L \eta + b_L) d\eta \right\} \left(x_L \le x < x_M\right) + \ldots \\
\left\{ \frac{1-P}{2} + \frac{\Delta x_L}{\Delta x} P + \int_{x_M}^{x} (m_H \eta + b_H) d\eta \right\} \left(x_M \le x < x_H\right) + \ldots \\
\left\{ \frac{1+P}{2} + \int_{x_H}^{x} p_S e^{-a(x - x_H)} d\eta \right\} \left(x_L \le x \le \infty\right)
\end{aligned}
\quad (12)
$$

The integrals of the tails contributions are

$$
\int_{-\infty}^{x} p_S e^{-a(x_L - \eta)} d\eta = \left(\frac{1-P}{2}\right) e^{a(x - x_L)}, \quad \int_{x_H}^{x} p_S e^{-a(\eta - x_H)} d\eta = \left(\frac{1-P}{2}\right) e^{-a(x - x_H)} \quad (13)
$$

The central integrals evaluate identically to those of the MAB c.d.f. in [1], save for the addition of the $P$ factor. This yields the desired EMAB c.d.f.

$$
\begin{aligned}
F_{EMAB}(x) = & \left(\frac{1-P}{2}\right) e^{a(x-x_L)} \left(-\infty \le x < x_L\right) + \ldots \\
& \left\{\frac{1-P}{2} + P\left(\frac{x-x_L}{\Delta x}\right)\left[C\left(\frac{x-x_M}{\Delta x_L}\right)+1\right]\right\}\left(x_L \le x < x_M\right) + \ldots \\
& \left\{\frac{1-P}{2} + P\left[\frac{\Delta x_L}{\Delta x} + \frac{x-x_M}{\Delta x}\left[C\left(\frac{x_H-x}{\Delta x_H}\right)+1\right]\right]\right\}\left(x_M \le x < x_H\right) + \ldots \\
& \left\{\frac{1+P}{2} + \frac{1-P}{2} e^{a(x_H-x)}\right\}\left(x_H \le x \le \infty\right)
\end{aligned}
\qquad (14)
$$

## 2.3 Sampling the EMAB Distribution

One of the most important uses of the MAB and EMAB expressions of subjective uncertainty is to employ these distributions in appropriate parts of stochastic systems simulation models. Such models frequently have components with operating parameters that involve random variables whose distributions are not known, but can be captured from the knowledge of domain experts in the form of MAB and/or EMAB distributions. To simulate the performance of such systems would require the sampling of all modeled probability distributions. The MAB sampling model was presented in [1], and here we develop the programmable algorithm that samples the EMAB.

We recall that to sample a probability distribution, we first generate a uniformly distributed random number $u \in [0,1]$, and then set that equal to the c.d.f. at the desired sample value $x_S$. By solving the c.d.f. for $x_S$ as a function of $u$ yields the desired formula. Since the EMAB c.d.f. is composed over four domains as shown in (14), we must solve for $x_S$. in those domains as a function of $u$, given into which of the four probability ranges $u$ falls. Note that the formula below contains the form where the c.d.f. is shifted to $x_L = 0$. This faster form is given in (15) and allows recovery of the actual $x_S$ by adding back $x_L$ to the shifted sampled value.

The validity of the EMAB sampling formula is demonstrated in Figure 4 which shows the theoretical EMAB p.d.f. in red for $x \in [-\infty, \infty]$ along with the related histogram of its 100,000 samples computed from (15). In the figure we have calculated and plotted the theoretical and sample means and standard deviations (vertical lines in red and black). As seen, the plotted line pairs are indistinguishable because their values are almost identical.

$$\text{For } u \in \left[0, \frac{1-P}{2}\right), \quad F_{EMAB}\left(x_S\right) = \left[\frac{1-P}{2}\right] e^{a x_S} = u$$

$$x_S\left(x_L = 0\right) = \frac{1-P}{P}\left[\frac{\Delta x}{2\left(1-C\right)}\right]\ln\left(\frac{2u}{1-P}\right),$$

$$\text{for } u \in \left[\frac{1-P}{2}, \frac{1-P}{2} + \frac{\Delta x_L}{\Delta x}P\right), \quad F_{EMAB}\left(x_S\right) = \frac{1-P}{2} + P\left(\frac{x_S}{\Delta x_L}\right)\left[C\left(\frac{x_S}{\Delta x_L} - 1\right) + 1\right] = u$$

$$x_S\left(x_L = 0\right) = \frac{\Delta x_L}{2C}\left\{1 - \left[1 - \frac{4C\Delta x}{P\Delta x_L}\left(u - \frac{1-P}{2}\right)\right]^{\frac{1}{2}}\right\},$$

$$\text{for } u \in \left[\frac{1-P}{2} + \frac{\Delta x_L}{\Delta x}P, \frac{1+P}{2}\right), \quad F_{EMAB}\left(x_S\right) = \frac{1-P}{2} + P\left\{\frac{\Delta x_L}{\Delta x} + \left(\frac{x_S - \Delta x_L}{\Delta x}\right)\left[C\left(\frac{\Delta x - x_S}{\Delta x_H}\right) + 1\right]\right\} = u$$

$$x_S\left(x_L = 0\right) = \frac{\Delta x_H}{2C}\left\{1 + C\left(\frac{\Delta x_L + \Delta x}{\Delta x_H}\right) - \left[\begin{array}{l}\left[1 + C\left(\frac{\Delta x_L + \Delta x}{\Delta x_H}\right)\right]^2 \ldots \\ -\frac{4C\Delta x}{\Delta x_H}\left[C\frac{\Delta x_L}{\Delta x_H} + \frac{1}{P}\left(u - \frac{1-P}{2}\right)\right]\end{array}\right]^{\frac{1}{2}}\right\},$$

$$\text{for } u \in \left[\frac{1+P}{2}, \infty\right), \quad F_{EMAB}\left(x_S\right) = \frac{1+P}{2} + \frac{1-P}{2}\left[1 - e^{a\left(\Delta x - x_S\right)}\right] = u$$

$$x_S\left(x_L = 0\right) = \frac{1-P}{P}\left[\frac{\Delta x}{2\left(1-C\right)}\right]\ln\left[2\left(\frac{1-u}{1-P}\right)\right] - \Delta x. \qquad (15)$$

**Figure 4 - EMABsampleTest02.m**