Derivation: How much data do I need?

A rederivation for @nntaleb's Technical Incerto Chapter 8.

Please help me to debug my derivation! I get an extra term which alters the conclusions about sample size.

Equation Definition

- Let *MAD(n)* be mean absolute deviation of *n* summed variables.
- Let X_{g1}, X_{g2}, X_{g3}, … be gaussian variables with mean μ and scale *M* (mean absolute deviation)
- Let X_{v1}, X_{v2}, X_{v3}, … be **non-gaussian** variables with mean μ and scale *M*
- The convergence speed is $\kappa_{1,n}$. It is defined as the value which solves the following:

$$\mathrm{MAD}(n) = \mathrm{MAD}(1)\, n^{\left(\frac{1}{2-\kappa_{1,n}}\right)} \tag{1}$$

Equation 2 tells you what sample size n_v you need. To be precise, n_v is the number of samples you need so that your sample average has the same mean absolute deviation as if you sampled n_g gaussian variables.

$$n_v = n_g^{-\frac{1}{\kappa_{1,n_v}-1}} \tag{2}$$

Derivation Attempt

Both the gaussian and non-gaussian distributions are normalized to the same scale, M.

We know that $\kappa_{1,n} = 0$ for the Gaussian. Inserting this to Equation (1) gives:

$$\mathrm{MAD}(n_g) = \sqrt{n_g}\, \mathrm{MAD}(1) \tag{3}$$

We are looking for the value n_v which makes the MAD of the sample averages match, i.e.:

$$\frac{\mathrm{MAD}(n_v)}{n_v} = \frac{\mathrm{MAD}(n_g)}{n_g} \tag{4}$$

Inserting Equations (1) and (3) into Equation (4) gives:

$$\frac{\mathrm{MAD}(1)\, n_v^{\left(\frac{1}{2-\kappa_{1,n_v}}\right)}}{n_v} = \frac{\sqrt{n_g}\, \mathrm{MAD}(1)}{n_g} \tag{5}$$

This Simplifies to:

$$n_v^{\left(\frac{1}{2-\kappa_{1,n_v}}-\frac{2-\kappa_{1,n_v}}{2-\kappa_{1,n_v}}\right)} = n_g^{-\frac{1}{2}} \tag{6}$$

Then to:

$$n_v^{\left(\frac{\kappa_{1,n_v}-1}{2-\kappa_{1,n_v}}\right)} = n_g^{-\frac{1}{2}} \tag{7}$$

And then to:

$$n_v = n_g^{-\frac{1}{2}\left(\frac{2-\kappa_{1,n_v}}{\kappa_{1,n_v}-1}\right)} \tag{8}$$

Finally to:

$$n_v = n_g^{-\left(\frac{1-0.5\kappa_{1,n_v}}{\kappa_{1,n_v}-1}\right)} \tag{9}$$

Conclusion

There is an extra $-0.5\kappa_{1,n_v}$ term compared to Equation (2). Including such a term would reduce the required sample size.

Please let me know if you understand why my derivation doesn't match the book!