Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2006-18

More on the Distribution of the Sum of Uniform Random Variables

Heinrich Potuschak and Werner G. Müller

November 2006

Abstract

The paper provides a simplified derivation of the density of the sum of independent non-identically distributed uniform random variables via an inverse Fourier transform. We also provide examples illustrating the quality of the Normal approximation and corresponding `MATHEMATICA` code.

1 Introduction

The sum of independent albeit not necessarily identical uniformly distributed random variables arises naturally in the aggregation of scaled values with differing numbers of significant figures. Its distribution was first established by Olds (1951) via an inductive proof. Sadooghi *et al.* (2007) find this distribution by employing a Laplace transform, also seemingly utilizing beforehand knowledge of the result.

In contrast, Bradley and Gupta (2002) derive an explicit formula in all its generality through characteristic function inversion. However, their proof, although elegant, is quite involved. In the following we present a considerably simplified version, that should be comprehensible at the level of a first course in mathematical statistics, such as Casella and Berger (2002).

2 Another Derivation of the Density of S_n

Let S_n be the sum of n independent random variables X_k, which are uniformly distributed in $(-a_k, a_k)$. We require this setting for symmetry purposes, but it is quite evident that generality is not restricted, since by simple shifts any intervals may be considered. Eventually we will give a corresponding example in section 3.

2.1 The derivation of the density using the characteristic function

The characteristic function of X_k is $E(\exp[itX_k]) = \frac{e^{it\cdot a_k} - e^{-it\cdot a_k}}{2it\cdot a_k} = \frac{\sin t\cdot a_k}{t\cdot a_k}$. Due to independence the characteristic function of S_n is the product $\varphi_n(t) = \prod_{k=1}^{n} \frac{\sin t\cdot a_k}{t\cdot a_k}$. The integral $\int_{-\infty}^{\infty} |\varphi_n(t)| dt$ exists and so we determine the density from $\varphi_n(t)$ as the inverse Fourier transform

$$f_n(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-its} \cdot \varphi_n(t) dt.$$

Remark: To see that $f_n(s)$ is a real-valued function, we make use of the fact that $e^{-its} = \cos(ts) - i \cdot \sin(ts)$, and due to the odd function $\sin(ts) \cdot \varphi_n(t)$ the imaginary part vanishes from the integral. Now, using $\sin x = (e^{ix} - e^{-ix})/2i$ gives

$$f_n(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-its}}{(2i)^n \cdot t^n \cdot \prod a_k} \cdot \prod_{k=1}^{n} \left(e^{it\cdot a_k} - e^{-it\cdot a_k}\right) dt.$$

Multiplying out the product results in a sum of 2^n terms of the form $\sigma_j \cdot \exp[it \cdot \underline{a}.\underline{\epsilon}_j]$. The n-dimensional row vector $\underline{\epsilon}_j$ is one of the 2^n possible length n sequences of arranging the signs +1 and -1. The signs σ_j result from the product over the entries in $\underline{\epsilon}_j$ and are positive in the case of an even number of -1 in $\underline{\epsilon}_j$ and negative otherwise. Multiplying this sum by e^{-its} gives $\sum_{j=1}^{2^n} \sigma_j \cdot \exp[it \cdot (\underline{a}.\underline{\epsilon}_j - s)] = \sum_{j=1}^{2^n} \sigma_j \cdot e^{it\cdot \underline{b}_j}$, where the row vector $\underline{b}_j$ denotes $\underline{a}.\underline{\epsilon}_j - s$. Reversing the order of integration and summation gives

$$f_n(s) = \frac{1}{2^{n+1}\pi \cdot \prod a_k} \sum_{j=1}^{2^n} \sigma_j \cdot \int_{-\infty}^{\infty} \frac{e^{i\cdot t\cdot \underline{b}_j}}{(it)^n} dt$$

Let us consider the summands $I_j = \int_{-\infty}^{\infty} \frac{e^{i\cdot t\cdot \underline{b}_j}}{(it)^n} dt$ and apply ν times the recursion $\int \frac{e^{itb}}{(it)^n} dt = \frac{1}{n-1} \left(\frac{-e^{itb}}{i(it)^{n-1}} + b \cdot \int \frac{e^{itb}}{(it)^{n-1}} dt \right)$:

$$\begin{aligned} I_j =& \frac{1}{n-1} \lim_{A\to\infty} \left. \frac{-e^{itb}}{i(it)^{n-1}} \right|_{-A}^{A} + \cdots + \frac{(n-\nu-1)!}{(n-1)!} \lim_{A\to\infty} \left. \frac{-b^{\nu-1} \cdot e^{itb}}{i(it)^{n-\nu}} \right|_{-A}^{A} + \\ &+ \frac{b^{\nu}(n-\nu-1)!}{(n-1)!} \int_{-\infty}^{\infty} \frac{e^{itb}}{(it)^{n-\nu}} dt \end{aligned}$$

Since $c \cdot \lim_{A\to\infty} \frac{e^{iAb} - e^{-iAb}}{A^{n-\nu}} = c \cdot \lim_{A\to\infty} \frac{2\cos(A\cdot b)}{A^{n-\nu}} = 0$ the first ν terms vanish while $\nu < n$. Repeating $n-1$ times the recursion gives

$$I_j = \frac{\underline{b}_j^{n-1}}{(n-1)!} \cdot \int_{-\infty}^{\infty} \frac{e^{i\cdot t\cdot \underline{b}_j}}{i \cdot t} dt.$$

Substituting $y = t \cdot b$ yields

$$\int_{-b\cdot\infty}^{b\cdot\infty} \frac{e^{i\cdot y}}{i \cdot y} dy = \int_{-b\cdot\infty}^{b\cdot\infty} \frac{\cos y + i \cdot \sin y}{i \cdot y} dy = \int_{-b\cdot\infty}^{b\cdot\infty} \frac{\sin y}{y} dy = \operatorname{sign}(b) \cdot \pi,$$

due to the odd function $\frac{\cos y}{y}$ and the sine integral. Inserting I_j in the sum above finally leads to

$$f_n(s) = \frac{1}{2^{n+1}(n-1)! \cdot \prod a_k} \cdot \sum_{j=1}^{2^n} \sigma_j \cdot \operatorname{sign}(\underline{a}.\underline{\epsilon}_j - s) \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-1},$$

where $s \in I\!R$. This is the same formula as given in Theorem 1 in Bradley and Gupta (2002).

2.2 Simplification of the density

The above formula is symmetric about s, but has the disadvantage to sum up over all 2^n indices j. In Corollary 1 Bradley and Gupta (2002) proof the sufficiency of summing up only over $j : \underline{a}_j.\underline{\epsilon}_j + s > 0$. This proves to be correct, but does not follow as easily as they claim. We can prove that $\{\underline{a}.\underline{\epsilon}_j : \underline{a}.\underline{\epsilon} > s\}$ is a sufficient subset.

2.2.1

The first step is to show that $S_- := \sum_{j=1}^{2^n} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-1} = 0$. This is not evident since terms do not cancel in pairs. Expanding the binomial to a sum of n terms and changing the order of summation gives

$$S_- = \sum_{\nu=0}^{n-1} \binom{n-1}{\nu} \cdot s^{n-1-\nu} \cdot \sum_{j=1}^{2^n} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j)^\nu.$$

We consider the inner sum and use the fact that each row vector $\underline{\epsilon}_j$ has exactly one counterpart $\underline{\epsilon}_{j'} = -\underline{\epsilon}_j$. Using $\sigma_j = \prod_{k=1}^n \epsilon_{jk}$, term j' of the sum is $\prod_{k=1}^n (-\epsilon_{jk}) \cdot (-\underline{a}.\underline{\epsilon}_j)^{n-1}$ and is equal to minus term j. So the inner sum vanishes in pairs and $S_- = 0$ is proved.

In the Appendix an alternative proof is given. It is made by induction and shows, that S_- vanishes for all powers $0 < p < n$ of the binomial inside the sum.

2.2.2

Separating the index set $J = \{j : 1 \le j \le 2^n\}$ into $J_- = \{j : \underline{a}.\underline{\epsilon}_j < s\}$ and $J_+ = \{j : \underline{a}.\underline{\epsilon}_j > s\}$ yields

$$\begin{aligned} f_n(s) &= f_n(s) + S_- = \\ &= \frac{1}{2^{n+1}(n-1)! \prod a_k} \left(\sum_{j \in J} \sigma_j \operatorname{sgn}(\underline{a}.\underline{\epsilon}_j - s)\, (\underline{a}.\underline{\epsilon}_j - s)^{n-1} + \sum_{j \in J} \sigma_j\, (\underline{a}.\underline{\epsilon}_j - s)^{n-1} \right) = \\ &= \text{const} \cdot \left(- \sum_{j \in J-} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-1} + \sum_{j \in J+} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-1} \right) = \\ &= \frac{1}{2^n (n-1)! \prod a_k} \cdot \sum_{j \in J+} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-1}. \end{aligned}$$

If $s > 0$ this formula reduces the number of terms to sum up from 2^n to $2^n/4$ on average and also avoids subtracting similarly large terms if n is a great number. As a by-product we see $f_n(s)$ vanishing for $|s| > \sum a_k$, because no $\underline{a}.\underline{\epsilon}_j$ is greater than $\sum a_k$ and $f_n(s)$ is symmetric about 0. Thus the most parsimonious way to calculate or plot $f_n(s)$ for any s is to replace it by $|s|$.

3 Some illustrative examples calculated with MATHEMATICA

3.1 Plotting the density $f_n(s)$

For the general case, that $Y_k \sim U(g_k, h_k)$, its distribution can be regarded as $U(-a_k + c_k, a_k + c_k)$. Because $\sum_{k=1}^{n} Y_k = S_n + \sum c_k$ the density of the sum is symmetrical with respect to $\sum c_k$ and can thus be calculated directly by making use of the formula above where s is replaced by $|s - \sum c_k|$.

The implementation of the distribution in MATHEMATICA required a few amendments, which are described in the comments to the code below, which plots $f_n(s)$ together with the density of an approximating normal distribution.

In1:= **<< Statistics`**

The package <<Statistics' is only required for the evaluation of the normal pdf.

In2:= $\boldsymbol{n = 3; a = \{0.5, 1.2, 1.5\}; c = \{-1, 3, 4\};}$

We have rather arbitrarily chosen $n = 3$, $a = \{0.5, 1.2, 1.5\}$, and $c = \{-1, 3, 4\}$ as an example.

In3:= $\mathbf{Sa = Apply[Plus, \mathit{a}]; Sa2 = Apply\,[Plus, \mathit{a}^2]\,; Sc = Apply[Plus, \mathit{c}];}$
$\mathbf{cnst = 2^{\mathit{n}} * (\mathit{n} - 1)! * Apply[Times, \mathit{a}];}$
$\mathbf{\mathit{e} = Table\,[IntegerDigits[\mathit{i} - 1, 2, \mathit{n}], \{\mathit{i}, 2^{\mathit{n}}\}] * 2 - 1;}$

The rows of the $(2^n, n)$-matrix ε are calculated as vectors of the dual numbers of 0 up to $2^n - 1$, filled up with zeros from the left and with zeros replaced by -1.

In4:= $\mathbf{ea = \mathit{e}.\mathit{a}; \sigma = Apply[Times, Transpose[\mathit{e}]];}$

The 2^n-dimensional vector $\underline{a}.\underline{\epsilon}$ is calculated as product **e.a** and denoted by **ea**. The sign σ_j counts the number of -1 in row j of ϵ and is calculated as $\prod \epsilon_{jk}$.

In5:= $\mathbf{fn[s\_] := \sum_{\mathit{j}=1}^{2^{\mathit{n}}} \sigma_{[[\mathit{j}]]} * Max\left[ea_{[[\mathit{j}]]} - Abs[\mathit{s} - Sc], 0\right]^{\mathit{n}-1} \Big/ cnst}$

The range of summation $J_+ = \{j : \underline{a}.\underline{\epsilon}_j < s\}$ holds also if $\max[\underline{a}.\underline{\epsilon}_j - s, 0]$ is used for $j \in [1, 2^n]$.

In6:= $\mathbf{ll = -Sa + Sc - 1; uu = Sa + Sc + 1;}$

The density of the sum is positive in $(-\sum a_k + \sum c_k, \sum a_k + \sum c_k)$ and is plotted for the domain $(-\sum a_k + \sum c_k - 1, \sum a_k + \sum c_k + 1)$ for better visibility.

In7:= $\mathbf{ndist = NormalDistribution\left[Sc, \sqrt{Sa2/3}\right];}$

The mean of the corresponding normal approximation is $\sum c_k$ and its variance is $\Sigma(2a_k)^2/12$.

In8:= **Plot[{fn[*s*], PDF[ndist, *s*]}, {*s*, ll, uu},**
AxesLabel → {"s", "fn(s)"}, PlotStyle → {{}, Dashing[{0.01}]}];

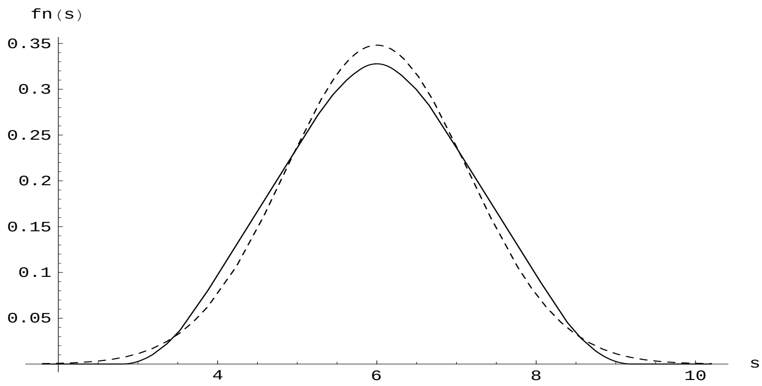

Figure 1: The density $f_n(s)$ for an example (solid line) and the corresponding normal approximation (dashed line).

3.2 The quality of the normal approximation

The normal approximation in Figure 1 seems reasonable. However to better judge its quality it may be advantageous to compare the respective cumulative distribution functions. Calculating $F_n(s)$ in MATHEMATICA is given by

In9: $\mathbf{Fn[s\_]:=\left(h = \sum_{j=1}^{2^n} \sigma_{[[j]]} * Max\left[ea_{[[j]]} - Abs[s - Sc], 0\right]^n / (n * cnst);\right.}$
$\mathbf{\left. If[s < Sc, h, 1 - h]\right)}$

This, since the integral of sums $\int_{-\infty}^{s} f_n(x)dx$ is equal to a sum of integrals and $\int_{-\infty}^{s}(\underline{a}.\underline{\epsilon}_j - x)^{n-1}dx = -(\underline{a}.\underline{\epsilon}_j - s)^n/n$.

We now illustrate the deviations from the normal by plotting the differences between the cdf's and pdf's respectively in Figure 2. In this example the numbers of variables were chosen as series $n = \{3, 6, 9, 12\}$, the endpoints as $\underline{a} = 1 + k/10$, and the translations as $\underline{c} = \underline{0}$.

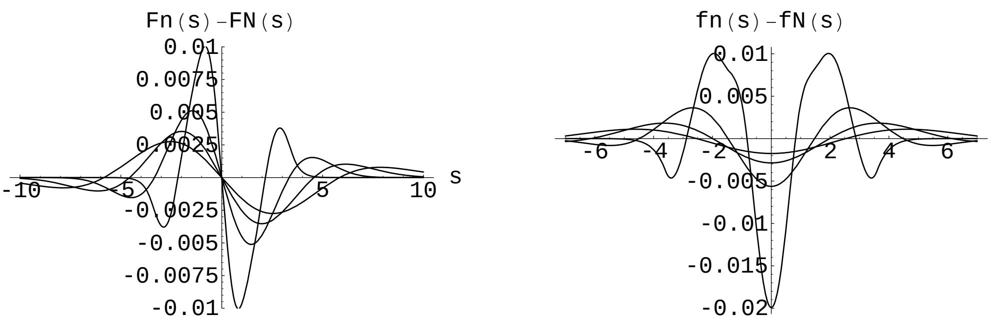

Figure 2: Differences of cdf's and pdf's of the exact distribution and the approximation for various n.

4 Another way to compute the density

Instead of summing up over $j \in J_+$ there is an equivalent way by using sums of k-tuples of $\underline{a}$: Adding $\underline{a}.\underline{\epsilon}_j + \sum a_k - \sum a_k$ cancels each $\underline{a}.\underline{\epsilon}_j < 0$ and a doubled sum of k-tuples minus $\sum a_k$ remains.

We denote A as matrix of k-tuples A_{kl} with row index $0 \leq k \leq n$ and column index $1 \leq l \leq \binom{n}{k}$ and define $A_{01} = 0$. The number of -1 in $\underline{\epsilon}_j$ agrees with the index $n-k$ of the row occupied by $\underline{a}.\underline{\epsilon}_j$ and so

$$f_n(s) = \text{const} \cdot \sum_{k=0}^{n} (-1)^{n-k} \cdot \sum_{l \in Lk+} (2 \cdot A_{kl} - \sum a_k - s)^{n-1}$$

with the indexing set $L_{k+} = \{l : 2A_{kl} - \sum a_k > s\}$.

Let us look at two special cases:
4.1 If X_k are independent and uniformly distributed in $(0, a_k)$ the density of $Y_n = \sum_{k=1}^{n} X_k$ is

$$\begin{aligned} f_{n0}(y) &= 2 \cdot f_n(2s - \textstyle\sum a_k) = \\ &= 2 \cdot \text{const} \cdot \textstyle\sum_{k=0}^{n} (-1)^{n-k} \cdot \sum_{l \in Lk+} (2 \cdot A_{kl} - 2y)^{n-1} = \\ &= \frac{1}{(n-1)! \prod a_k} \cdot \textstyle\sum_{k=0}^{n} (-1)^{n-k} \cdot \sum_{l : A(k,l) > s} (A_{kl} - y)^{n-1} \quad \text{for } 0 < y < \sum a_k \end{aligned}$$

Another way to notate the inner sum is to use $1 \leq l \leq \binom{n}{k}$ and to sum up powers of $(y - A_{kl})_+ := Max[0, y - A_{kl}]$, as was done in Bradley and Gupta (2002) and Sadooghi *et al.* (2007) and in the inductive proof in Olds (1951).

4.2 In the case of $\underline{a} = \underline{1}$ and with $[y]$ as least integer less than y the density reduces to $f_n(y) = \frac{1}{(n-1)!} \cdot \sum_{k=0}^{[y]} (-1)^k \cdot \binom{n}{k} \cdot (y-k)^{n-1}$. This formula also is given in Sadooghi *et al.* (2007).

5 Conclusions

Sadooghi *et al.* (2007) claim that their proof is more easily comprehensible. After our amendments to the proof of Bradley and Gupta (2002), we leave it to the reader to decide.

Appendix A: another proof by induction

The inductional assumption is that $\sum_{j=1}^{2^n} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - s)^p = 0$ for $0 \leq p < n$.

At the zeroth step take $n = 2$ and $+(-a_1 - a_2 - s)^p - (-a_1 + a_2 - s)^p - (+a_1 - a_2 - s)^p + (+a_1 + a_2 - s)^p = 0$ is true for $p = 0$ and $p = 1$ and $s \in I\!R$.

In the induction step we make use of the set $(\underline{a}.\underline{\epsilon}_j + a_{n+1}) \cup (\underline{a}.\underline{\epsilon}_j - a_{n+1})$. Its length is 2^{n+1} and the signs $\underline{\sigma}$ remain unchanged if a_{n+1} is added and change if a_{n+1} is subtracted:

$$S_- = \sum_{j=1}^{2^n} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j + a_{n+1} - s)^n - \sum_{j=1}^{2^n} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - a_{n+1} - s)^n$$

We expand each term $(\underline{a}.\underline{\epsilon}_j - s \pm a_{n+1})^n$ to

$$\binom{n}{0} \cdot (\underline{a}.\underline{\epsilon}_j - s)^n \cdot a_{n+1}^0 \pm \binom{n}{1} \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-1} \cdot a_{n+1}^1 + \binom{n}{2} \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-2} \cdot a_{n+1}^2 \pm \dots$$

If $n = 2l$ and even or $n = 2l - 1$ and odd, term j of the subtracted sum is $2 \cdot \sum_{\nu=1}^{l} \binom{n}{2\nu-1} \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-2\nu+1} \cdot a_{n+1}^{2\nu-1}$. Changing the order of summations gives $S_- = 2 \cdot \sum_{\nu=1}^{l} \binom{n}{2\nu-1} \cdot a_{n+1}^{2\nu-1} \cdot \sum_{j=1}^{2^n} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - s)^{n-2\nu+1} = 0$ due to the inductional assumption: $\sum_{j=1}^{2^n} \sigma_j \cdot (\underline{a}.\underline{\epsilon}_j - s)^p$ vanishes if $p < n$.

Acknowledgements

We are grateful to Professor Robert Hafner for initial hints in improving the derivation in Section 2.

References

[1] Bradley, D.M. and Gupta, C.R., On the distribution of the sum of n non-identically distributed uniform random variables. Ann. Inst. Statist. Math. **54**(3), (2002) 689-700.

[2] Casella, G. and Berger, R.L., *Statistical Inference, 2nd edition* (Duxbury Press, 2002)

[3] Olds, E.G., A note on the convolution of uniform distributions. Ann.Math.Statist.**23,** (1952) 282-285.

[4] Sadooghi-Alvandi, S.M. and Nematollahi, A.R. and Habibi, R. (2007) On the distribution of the sum of independent uniform random variables. Statistical Papers.