The first divisible sum

Barbour, A D ; Grübel, R

DOI: https://doi.org/10.1007/BF02213453

Originally published at:
Barbour, A D; Grübel, R (1995). The first divisible sum. Journal of Theoretical Probability, 8(1):39-47.
DOI: https://doi.org/10.1007/BF02213453

The first divisible sum

A.D. Barbour and R. Grübel

Universität Zürich and Technische Universiteit Delft

now: "Universität-Gesamthochschule-Paderborn"

Abstract. We consider the distribution of the first sum of a sequence of iid random variables which is divisible by d. This is known to converge, when divided by d, to a geometric distribution as $d \to \infty$. We obtain results on the rate of convergence, using two contrasting approaches. In the first, Stein's method is adapted to geometric limit distributions. The second method is based on the theory of Banach algebras. Each method is shown to have its merits.

Keywords: discrete renewal theory, Stein's method, Banach algebras.

AMS Classification: 60K05, 60F05

Postal addresses: Institut für Angewandte Mathematik, Rämistrasse 74, CH–8001 ZÜRICH and Department of Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, NL–2628CD DELFT.

now: "Fachbereich 17 (Mathematik-Informatik), Universität-Gesamthochschule-Pad erborn, 33095 Paderborn, Germany"

1. Introduction.

> There is a mistake in the very first sentence that had a result similar to that of a missing semicolon in certain computer programs - it fouled the whole thing up. For me, $\mathbb{N}_0$ ('the $\mathbb{N}$ with the na ught') has always been $\{0,1,2,\ldots\}$, but of course, it should then be $\mathbb{N}$ ($= \{1,2,\ldots\}$) here. I suggest that we insert something like "We will write $\mathbb{N}$ for the set $\{1,2,3,\ldots\}$ of strictly positive integers and $\mathbb{Z}_+$ for the set $\{0,1,2,\ldots\}$ of non-negative integers." Many of the referee's comments wil l become vacuous once $\mathbb{N}$ and $\mathbb{N}_0$ are interpreted as above.

Let $(X_i : i \in \mathbb{N}_0)$ be a sequence of independent and identically distributed random variables with values in $\mathbb{N}$ and let $p_n = P(X_1 = n)$. We assume that $\mu = EX_1$ is finite and that the sequence $p = (p_n : n \in \mathbb{N}_0)$ is aperiodic, i.e. $\gcd\{n \in \mathbb{N}_0 : p_n > 0\} = 1$. We are interested in the distribution of Y_d, the (normalized) first sum that is divisible by d,

$$Y_d = \frac{1}{d} S_{\tau_d}, \quad \tau_d = \inf\{n \in \mathbb{N} : d \mid S_n\}, \quad S_n = \sum_{i=1}^{n} X_i.$$

> Maybe we should indeed add something about exceptional sets where Y_d is not d efined. I suggest something like: "Here we take the infimum of an empty set to be infini te; on $\tau_d = \infty$ we set $Y_d = 0$. Under our assumptions τ_d will be fi nite with probability 1, which means that this is irrelevant for our results as these refer t o distributions." This might also be the right place to add some 'motivation', e.g. "The variable $d \cdot Y_d$ arises as the time between observed renewals in a periodically insp ected self-renewing aggregate."

In Grübel (1985) it was shown that Y_d converges in distribution to $\mathrm{Ge}(1/\mu)$. Here $\mathrm{Ge}(\theta)$ (or Ge_θ) denotes the geometric distribution on $\mathbb{N}$ with parameter θ, i.e. $Y \sim \mathrm{Ge}(\theta)$ means $P(Y = n) = (1-\theta)^{n-1}\theta$ for all $n \in \mathbb{N}$. In the present note we investigate the associated rate of convergence and we establish some related upper bounds. We use two rather different techniques: Stein's method (used here for the first time in connection with the geometric distribution) in Section 2, and an analytic approach in Section 3. The latter is based on results from the theory of commutative Banach algebras, see Gelfand, Raikov and Shilov (1964); a general reference for Stein's method is Stein (1986). In Section 4 we compare the two approaches.

2. Stein's method. Let X_0 be an additional random variable, defined on the same probability space as and independent of the X_i's. The process $(\xi_n : n \in \mathbb{N}_0)$

of *forward recurrence times* associated with the delayed renewal process $(X_0+S_n: n \in \mathbb{N}_0)$ is then defined by

$$\xi_n \;=\; \min\{X_0+S_i : i \in \mathbb{N}_0,\, X_0+S_i \geq n\} \;-\; n.$$

This process is a Markov chain with state space $\mathbb{N}_0$, and Y_d defined as $d^{-1}\inf\{n \in d\mathbb{N}_0 : \ \xi_n = 0\}$ is the same random variable as in Section 1, when $X_0 = 0$ a.s.

> I would change this sentence into e.g.: "This process is a Markov chain with sta te space $\mathbb{Z}_+$. Let $Y_d = d^{-1}\inf\{n \in d\mathbb{N} : \ \xi_n = 0\}$; this generalizes the setup introduced in the introduction where $X_0 = 0$."

If instead $X_0 \sim \pi$, where π is the probability distribution on $\mathbb{N}_0$ defined by $\pi_k = P(X_1 > k)/\mu$, (ξ_n) is stationary, and all ξ_n's have distribution π. We will write P_π, E_π for the corresponding probability and expectation respectively and P_i, E_i for probability and expectation conditional on $X_0 = i$.

> Maybe the last sentence should have a John the Baptist type introduction: "Follo wing the custom in Markov chain theory, ...".

We are interested in the distribution of Y_d under $P = P_0$ or, more precisely, in the total variation distance between this distribution and the geometric distribution with mean μ,

$$\|P_0^{Y_d} - \mathrm{Ge}(1/\mu)\|_{\mathrm{TV}} \;=\; \sup_{A\subset\mathbb{N}} \bigl|P_0(Y_d \in A) - \mathrm{Ge}_{1/\mu}(A)\bigr|.$$

Our upper bound will be a multiple of the tail $\mathbb{P}(\tau_{\pi,0} > d)$ of the coupling time of any pair of ξ-processes, one of which is stationary and the other starting with $X_0 = 0$; the coupling could be chosen to make $\mathbb{P}(\tau_{\pi,0} > d) = \|P_0^{\xi_d} - \pi\|_{\mathrm{TV}}$, or the processes could run independently until coupling.

> Shall we change this? I offer: "Our upper bound will be a multiple of the tail of the coupling time of any pair of ξ-processes, one of which is stationary and the other starting with $X_0 = 0$. By this we mean the following: if $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space, if $\xi^{(1)}$ and $\xi^{(2)}$ are two processes on this space such that the distribution of $xi^{(1)}$ under $\mathbb{P}$ is the same as the distribution of ξ under P_π and the distribution of $\xi^{(2)}$ equals the distribution of ξ under P_0, a nd if further τ denotes the first index k with $\xi_k^{(1)} = \xi_k^{(2)}$ ($\tau = \inf ty$ if no 'coupling' takes place), then our upper bound will involve the tail pr obabilities of τ under $\mathbb{P}$. To make the dependence on the

two different in itial distrib utions of $\xi^{(1)}$ and $\xi^{(2)}$ clear, we will write $\mathbb{P}(\tau_{\pi,0} > n)$ for these probabilities. There are, of course, many such co nstructions. It is known that $(\Omega, \mathcal{F}, \mathbb{P})$, $\xi^{(1)}$ and $\xi^{(2)}$ c an be chosen such that $\mathbb{P}(\tau_{\pi,0} > d) = \|P_0^{\xi_d} - \pi\|_{\mathrm{TV}}$ (see e.g . ?? and ??). Our upper bound does not depend on any specific construction, we could use, for example, an independent coupling where $\mathbb{P}$ is the product of P_π and P_0."

Because

$$\|P_\pi^{Y_d} - P_0^{Y_d}\|_{\mathrm{TV}} \leq \mathbb{P}(\tau_{\pi,0} > d), \tag{1}$$

it is enough, if only the order of approximation is of interest, to consider the stationary process.

Due to stationarity, the process $(\xi_{d+n} : n \in \mathbb{N}_0)$ has the same distribution as the original $(\xi_n : n \in \mathbb{N}_0)$. Let Y_d' be the normalized first divisible sum for the shifted process; Y_d and Y_d' have the same conditional distributions, in the sense that, for all k, $\mathcal{L}(Y_d|\xi_0 = k) = \mathcal{L}(Y_d'|\xi_d = k)$. On $Y_d > 1$ we have $Y_d' = Y_d - 1$, and a decomposition with respect to the value of ξ_d gives the first of the following equalities, for f any bounded function on $\mathbb{N}$:

$$\begin{aligned} E_\pi f(Y_d) &= \pi_0 f(1) + \sum_{j\geq 1} \pi_j E_j f(Y_d + 1) \\ &= \pi_0 f(1) + (1-\pi_0) E_\pi f(Y_d+1) + \pi_0 \big(E_\pi f(Y_d+1) - E_0 f(Y_d+1)\big), \end{aligned}$$

or

$$E_\pi\{(1-\pi_0) f(Y_d+1) + \pi_0 f(1) - f(Y_d)\} = \pi_0\{E_0 f(Y_d+1) - E_\pi f(Y_d+1)\}.$$

From (1) it follows that

$$\big|E_\pi f(Y_d+1) - E_0 f(Y_d+1)\big| \leq \mathbb{P}(\tau_{\pi,0} > d) \sup_{j,k\in\mathbb{N}} |f(j) - f(k)|.$$

Hence, if we define an operator $\mathcal{A}$ acting on functions $f : \mathbb{N} \to \mathbb{R}$ by

$$\big(\mathcal{A}f\big)(j) = (1-\theta) f(j+1) + \theta f(1) - f(j) \qquad \text{for all } j \in \mathbb{N}$$

with $\theta = \pi_0 = 1/\mu$, then the above yields

$$\big|E_\pi \mathcal{A} f(Y_d)\big| \leq \frac{1}{\mu} \mathbb{P}(\tau_{\pi,0} > d) \sup_{j,k\in\mathbb{N}} |f(j) - f(k)|.$$

What is the referee complaining about? Maybe we should replace "the above yields " by "this upper bound, together with the `previous equality, gives". Alternative ly, we could number the formulas. Maybe I overlook something.`

This bound will be very useful if, for each $A \subset \mathbb{N}$, we can find a bounded function $f = f_A$ such that

$$(1-\theta)f(j+1) + \theta f(1) - f(j) \;=\; 1_A(j) - \mathrm{Ge}_\theta(A), \tag{2}$$

since with this f we have $E_\pi \mathcal{A} f(Y_d) = P_\pi(Y_d \in A) - \mathrm{Ge}_\theta(A)$. Equation (2) can be considered as a Stein equation for geometric limit distributions; a heuristic explanation is given at the end of the section.

Let $h_A(j) = 1_A(j) - \mathrm{Ge}_\theta(A)$. It is easy to check that $\mathcal{A}f = h$ is solved by

$$f(j) = f(1) + (1-\theta)^{-j} \sum_{i=1}^{j-1} (1-\theta)^i h(i) \quad \text{for all } j > 1,$$

with any choice of $f(1)$. For h_A as above it holds that $\sum_{i=1}^{\infty}(1-\theta)^i h(i) = 0$, so that the corresponding solution can equally be written as

$$f_A(j) \;=\; f_A(1) \;-\; \sum_{i=0}^{\infty} (1-\theta)^i h_A(j+i),$$

from which the bound $\sup_{j,k\in\mathbb{N}} |f(j) - f(k)| \le 1/\theta = \mu$ for all $A \subset \mathbb{N}$ is immediate, so that

$$\|P_\pi^{Y_d} - \mathrm{Ge}(\theta)\|_{\mathrm{TV}} \le \; \mathbb{P}(\tau_{\pi,0} > d).$$

Combining the above considerations gives us the following result.

Theorem 1 *With $\tau_{\pi,0}$ any coupling time as above, we have*

$$\|P^{Y_d} - \mathrm{Ge}(1/\mu)\|_{\mathrm{TV}} \;\le\; 2\mathbb{P}(\tau_{\pi,0} > d).$$

The literature contains numerous results on the tails of coupling times. Pitman (1974) shows that $EX_1 < \infty$ implies that $\mathbb{P}(\tau_{\pi,0} < \infty) = 1$, and Lindvall (1979) has results showing that $EX_1^{1+\gamma} < \infty$ for some $\gamma > 0$ implies $\mathbb{P}(\tau_{\pi,0} > d) = o(d^{-\gamma})$ for the independent coupling. Combining this with the above theorem we obtain the following corollary.

Corollary 2 *Let $\gamma \ge 0$. If $EX_1^{1+\gamma} < \infty$ then*

$$\|P^{Y_d} - \mathrm{Ge}(1/\mu)\|_{\mathrm{TV}} \;=\; o(d^{-\gamma}).$$

The following is a heuristic explanation for the success of $\mathcal{A}$ in the present context: we can think of a sequence of successive Y_d's being embedded in the renewal process $(S_n : n \in \mathbb{N})$. The value of the current Y_d either increases by one if a multiple of d is passed, or a hit occurs in which case a new run for the next Y_d begins. This structure resembles that of a 'birth-catastrophe' process, by which we mean a Markov chain on $\mathbb{N}$ with transition probabilities

$$p_{i,i+1} = 1 - \theta, \quad p_{i1} = \theta \qquad \text{for all } i \in \mathbb{N}.$$

The stationary distribution of such a process is Ge_θ; further, it is easy to check that $\mathcal{A}$ is the associated generator. The latter fact can also be used to solve (2); see Barbour (1988) for more on the connection between Stein equations and generators of Markov processes.

3. An analytic approach. The quantity of interest is the distribution of Y_d and this distribution depends solely on the distribution of X_1. Identifying distributions on $\mathbb{N}_0$ with the associated sequences of point masses leads us to regard this dependence as a mapping from and to ℓ_1, the space of all summable sequences with index set $\mathbb{N}_0$. We will analyze the functional $\ell_1 \ni p \to p^{(d)} \in \ell_1$, where $p_k^{(d)} = P(Y_d = k)$, by decomposing it into several simpler functionals. A key idea is the relation to renewal sequences which has also been used in Grübel (1985). We assume that $EX_1^{1+\gamma} < \infty$ for some fixed $\gamma \geq 0$.

The renewal sequence $(u_n : n \in \mathbb{N}_0)$ associated with a distribution p can be defined recursively by $u_0 = 0$, $u_n = \sum_{i=1}^{n} p_i u_{n-i}$; u_n is the probability that one of the partial sums of the X-sequence lands in n. For a geometric distribution with parameter θ, symbolized by the sequence $p^{[\theta]}$, the associated renewal sequence $u^{[\theta]}$ has $u_n^{[\theta]} = \theta$ for all $n \in \mathbb{N}$. Renewal sequences are useful in the present context because of the following: $u^{(d)}$, the renewal sequence associated with $p^{(d)}$, is related to u by $u_n^{(d)} = u_{dn}$. Hence, a first decomposition of the functional consists of three steps: pass to the renewal sequence of the distribution of X_1, take every d^{th} element, then find the distribution associated with this renewal sequence.

For this to work we need a thorough understanding of the relationship between distributions and the associated renewal sequences. Here we can build upon a sizeable literature, beginning with one of the original proofs of the discrete renewal theorem given by Erdős, Feller and Pollard (1949). Again, a decomposition into simpler steps is crucial: let Σp and Δu be the sequences given by $(\Sigma p)_n = \sum_{k>n} p_k$ for all $n \in \mathbb{N}_0$, and $(\Delta u)_0 = 0$, $(\Delta u)_n = u_n - u_{n-1}$ for all $n \in \mathbb{N}$, respectively. For a finite mean, aperiodic p these sequences are both summable and are convolution inverse, i.e. $\Sigma p * \Delta u = \delta$ where $\delta_0 = 1$, $\delta_n = 0$ for all $n \in \mathbb{N}$.

```
I suggest: replace full stop by semicolon and continue with
"aperiodicity, for e xample, is used to show that the value 0
is not in the range of the Fourier transform of Σp, see Erdős et
al. (1949) for details."
```

Now

$$p \;\to\; \Sigma p \;\to \Delta u \;\to\; u \;\to\; u^{(d)} \;\to\; \Delta u^{(d)} \;\to\; \Sigma p^{(d)} \;\to\; p^{(d)}$$

summarizes all the necessary steps in one diagram.

We now introduce the space $\ell_1(\gamma)$ of all sequences $(a_n : n \in \mathbb{N}_0)$ with the property that $\|a\|_\gamma < \infty$ where $\|a\|_\gamma = \sum_{n=0}^{\infty}(1+n)^\gamma |a_n|$. Endowed with this norm $\ell_1(\gamma)$ becomes a Banach algebra; in particular, the norm inequality $\|a \star b\|_\gamma \le \|a\|_\gamma \, \|b\|_\gamma$ holds for all $a, b \in \ell_1(\gamma)$.

Let $\mathcal{I}$ be the set of invertible elements in ℓ_1. A well-known result from the theory of Banach algebras (see e.g. Gelfand, Raikov and Shilov (1964), Section 19) states that, if $a \in \mathcal{I} \cap \ell_1(\gamma)$, then $a^{\star(-1)} \in \ell_1(\gamma)$: taking the inverse does not lead out of $\ell_1(\gamma)$. Hence: $EX_1^{1+\gamma} < \infty$ means $\Sigma p \in \ell_1(\gamma)$ which implies $\Delta u \in \ell_1(\gamma)$.

Using the relationship between u and $u^{(d)}$ we see that

$$\left|\left(\Delta u^{(d)} - \Delta u^{[\theta]}\right)_n\right| \;\le\; \begin{cases} 0, & \text{if } n = 0, \\ |u_d - \theta|, & \text{if } n = 1, \\ \sum_{k=1}^{d} \left|(\Delta u)_{(n-1)d+k}\right|, & \text{if } n > 1, \end{cases}$$

which gives, with $\theta = 1/\mu = \lim_{n\to\infty} u_n$,

$$\begin{aligned} \|\Delta u^{(d)} - \Delta u^{[\theta]}\|_\gamma \;&\le\; 2^\gamma |u_d - \theta| \;+\; \sum_{n=2}^{\infty}(1+n)^\gamma \sum_{k=1}^{d}\left|(\Delta u)_{(n-1)d+k}\right| \\ &\le\; 2^\gamma \sum_{k>d}\left|(\Delta u)_k\right| \;+\; 2^\gamma d^{-\gamma} \sum_{n=2}^{\infty}\sum_{k=1}^{d}(1+nd)^\gamma \left|(\Delta u)_{(n-1)d+k}\right| \\ &\le\; 2^\gamma d^{-\gamma}\left(\sum_{k>d}(1+k)^\gamma \left|(\Delta u)_k\right| \right. \\ &\qquad\qquad \left. +\; \left(\sup_{k>d}\frac{1+k+d}{1+k}\right)^\gamma \sum_{k>d}(1+k)^\gamma\left|(\Delta u)_k\right|\right) \\ &=\; o(d^{-\gamma}). \end{aligned}$$

From the norm inequality and simple algebra we obtain

$$\begin{aligned} \|\Sigma p^{(d)} - \Sigma p^{[\theta]}\|_\gamma \;&=\; \|\Sigma p^{(d)} \star \Sigma p^{[\theta]} \star (\Delta u^{(d)} - \Delta u^{[\theta]})\|_\gamma \\ &\le\; \|\Sigma p^{(d)}\|_\gamma \, \|\Sigma p^{[\theta]}\|_\gamma \, \|\Delta u^{(d)} - \Delta u^{[\theta]}\|_\gamma. \end{aligned}$$

The middle factor does not depend on d and the last factor is $o(d^{-\gamma})$. To bound the first factor we apply another piece of analysis: the set $\mathcal{I}\cap\ell_1(\gamma)$ is $\|\cdot\|_\gamma$-open in $\ell_1(\gamma)$, and $a \to a^{\star(-1)}$ is continuous on this set (Gelfand, Raikov and Shilov (1964), Section 2). We know that $\Delta u^{(d)}$ converges to $\Delta u^{[\theta]}$, which is in $\mathcal{I}\cap\ell_1(\gamma)$, with respect to $\|\cdot\|_\gamma$ (we even have a rate), hence, for any given $\epsilon > 0$ we can find a $d(\epsilon)$ such that

$$\|(\Delta u^{(d)})^{\star(-1)} - (\Delta u^{[\theta]})^{\star(-1)}\|_\gamma \;\le\; \epsilon \qquad \text{for all } d \ge d(\epsilon).$$

This implies $\sup_{d\in\mathbb{N}} \|\Sigma p^{(d)}\|_\gamma < \infty$, so that we have proved the following theorem.

Theorem 3 *If, for some* $\gamma \ge 0$, $EX_1^{1+\gamma} < \infty$ *then*

$$\sum_{k=1}^{\infty} k^\gamma \left| P\big(Y_d \ge k\big) \;-\; \Big(1 - \frac{1}{\mu}\Big)^{k-1} \right| \;=\; o(d^{-\gamma}).$$

It is easy to see that this implies $o(d^{-\gamma})$-convergence of the total variation distance between the distribution of Y_d and the geometric distribution with mean μ, hence we reobtain Corollary 2. In fact, the theorem amplifies the total variation result in a manner discussed in Chapter 2.4 of Barbour, Holst and Janson (1992). It is also interesting to note that the weights k^γ are in some sense optimal; as $dY_d > X_1$ the left hand side of the equality in the theorem would become infinite for any $\gamma' > \gamma$ unless $EX_1^{1+\gamma'} < \infty$.

4. Synthesis. For mathematicians, the meaning of the word 'analytic' is twofold, and it may be a coincidence that both meanings can be attached to the approach of the previous section. We should perhaps point out that the dissection into simple parts, such as given in the diagram, also points the way towards an algorithm for obtaining the distribution of Y_d for a given distribution of X_1; see Grübel (1988) for the computation of renewal sequences and related problems.

A somewhat typical aspect of the approach in Section 3 is the fact that the switch to analysis, away from the stochastic model, occurs at a very early stage. From a probabilistic point of view Stein's method is more appealing as more use is made of the stochastic aspects of the problem. This is perhaps more than a matter of taste, especially if it comes to generalizing to non-independent models where Stein's method often has no direct competitors. For instance, if the X_i arise as the intervals between the points of a stationary, mixing point process ξ, the inequality in Theorem 1 can be proved in much the same way: P_0 is now the Palm distribution, and of the ξ processes used in the definition of $\tau_{\pi,0}$, one has the Palm distribution and the other the stationary distribution. On the other hand, for independent X_i's, the analytic approach gives a somewhat sharper result.

The contrast between the two approaches is not confined to this example. In proving bounds for the approximation in the central limit theorem, one has a similar choice between the operator technique of Trotter (1959) and Fourier methods on the one hand, and methods such as Stein's (1970) on the other. In the Poisson context, one has the operator technique of Deheuvels and Pfeifer (1986) and the complex analytic method of Uspensky (1931) as opposed to the Stein–Chen method (Chen, 1975). In these cases, too, Stein's method adapts more easily to dependent settings than the analytic methods.

References

Barbour, A.D. (1988) Stein's method and Poisson process convergence. *Journal of Applied Probability* **25A** 175–184.

Barbour, A.D., Holst, L. and Janson, S. *Poisson Approximation.* Clarendon Press, Oxford 1992.

Chen, L.H.Y. (1975) Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

Deheuvels, P. and Pfeifer, D. (1986) A semigroup approach to Poisson approximation. *Ann. Probab.* **14** 665–678.

Erdős, P., Feller, W. and Pollard, H. (1949) A property of power series with positive coefficients. *Bull. Am. Math. Soc.* **55** 201–204.

Gelfand, I.M., Raikov, D.A. and Shilov, G.E. *Commutative Normed Rings.* Moscow, 1960. Engl. Edition: Chelsea, New York, 1964.

Grübel, R. (1985) An application of the renewal theoretic selection principle: the first divisible sum. *Metrika* **32** 327–337.

Grübel, R. (1988) The fast Fourier transform in applied probability theory. *Nieuw Archief voor Wiskunde Ser. 4* **7** 289–300.

Lindvall, T. (1979) On coupling of discrete renewal processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **48** 57–70.

Pitman, J.W. (1974) Uniform rates of convergence for Markov chain transition probabilities. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **29** 193–227.

Stein, C. (1970) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 583–602.

Stein, C. *Approximate computation of expectations.* IMS Lecture Notes, Vol. 7, Hayward, California 1986.

Trotter, H.F. (1959) An elementary proof of the central limit theorem. *Archiv der Math.* **9** 226–234.

Uspensky, J.V. (1931) On Ch. Jordan's series for probability. *Ann. Math.* **32** 306–312.