J. Appl. Prob. **39,** 650–656 (2002)
Printed in Israel

FROM MATCHBOX TO BOTTLE: A STORAGE PROBLEM

LUDWIG BARINGHAUS * AND
RUDOLF GRÜBEL,* ** *Universität Hannover*

Abstract

We generalize Banach's matchbox problem: demands of random size are made on one of two containers, both initially with content t, where the container is selected at random in the successive steps. Let Z_t be the content of the other container at the moment when the selected container is found to be insufficient. We obtain the asymptotic distribution of Z_t as $t \to \infty$ under quite general conditions. The case of exponentially distributed demands is considered in more detail.

Keywords: Asymptotic normality; Bessel functions; convergence in distribution; exponential distribution; Knuth's old sum; renewal theory

AMS 2000 Subject Classification: Primary 60K30
Secondary 60F05; 60K05

Dedicated to Professor D. Morgenstern on the occasion of his seventy-eighth birthday

At times $n = 1, 2, \dots$, a demand of random size $W_n > 0$ needs to be satisfied. We have two containers, A and B, initially filled to capacity t. At each time one of these is chosen at random. We are interested in the distribution of the content Z_t of the other container once the content of the chosen container is found to be insufficient. Formally, we have a sequence (H_i, W_i), $i \in \mathbb{N}$, of independent and identically distributed two-dimensional random vectors, with independent components and $\mathrm{P}(H_i = 0) = \mathrm{P}(H_i = 1) = \frac{1}{2}$ (here $H_i = 1$ means that we choose container A in the ith step). Let μ be the distribution of $W := W_1$.

With μ concentrated at the single value 1, this is Banach's matchbox problem; see Feller (1968). Replacing the smoking habit by a drinking habit we may imagine a person carrying two bottles of size t in his right and left pocket, repeatedly choosing one of them at random, and then taking a swig of random size.

Over the years, Banach's matchbox problem and its generalizations have been analyzed by various authors. Holst (1989) gave an approach based on Poisson process embedding. The problem also arises in connection with paired comparisons; see Uppuluri and Blot (1974) and the references given therein for statistical applications. Cacoullos (1967) considered a generalization with more than two boxes; again, the problem is put into a statistical context. Berghahn (1966) considered the above set-up with μ an exponential distribution, and he derived various explicit series representations. Another generalization, which lets the choice depend on the content of the containers, is known as the toilet paper problem or the transparent matchbox problem; see Knuth (1984) and Stirzaker (1988).

Received 4 July 2002.
* Postal address: Institut für Mathematische Stochastik, Universität Hannover, Postfach 60 09, D-30060 Hannover, Germany.
** Email address: rgrubel@stochastik.uni-hannover.de

It is helpful to regard the overall procedure as a two-dimensional random walk $(S_n)_{n\in\mathbb{N}_0}$, where

$$S_0 \equiv \begin{pmatrix}0\\0\end{pmatrix} \quad \text{and} \quad S_n := \begin{pmatrix}S_{A,n}\\S_{B,n}\end{pmatrix} \quad \text{for } n \in \mathbb{N},$$

with

$$S_{A,n} := \sum_{i=1}^{n} H_i W_i \quad \text{and} \quad S_{B,n} := \sum_{i=1}^{n} (1 - H_i) W_i;$$

the latter two terms denote the total demand made up to time n on A and B respectively. If notationally convenient, we will write $S_A(n)$ and $S_B(n)$ instead of $S_{A,n}$ and $S_{B,n}$. The random walk is stopped as soon as it leaves the square $[0, t] \times [0, t]$. Formally, $\tau_t := \tau_{A,t} \wedge \tau_{B,t}$ with

$$\tau_{A,t} := \inf\{n \in \mathbb{N} : S_{A,n} > t\}, \qquad \tau_{B,t} := \inf\{n \in \mathbb{N} : S_{B,n} > t\}.$$

As the support of the step distribution of the random walk is contained in the set $\{0\} \times [0, \infty) \cup [0, \infty) \times \{0\}$, either $\tau_t = \tau_{A,t}$ or $\tau_t = \tau_{B,t}$. Then

$$Z_t := \begin{cases} t - S_B(\tau_{A,t}) & \text{if } \tau_t = \tau_{A,t}, \\ t - S_A(\tau_{B,t}) & \text{if } \tau_t = \tau_{B,t}, \end{cases}$$

and

$$R_t := \begin{cases} S_A(\tau_{A,t}) - t & \text{if } \tau_t = \tau_{A,t}, \\ S_B(\tau_{B,t}) - t & \text{if } \tau_t = \tau_{B,t}, \end{cases}$$

denote, respectively, the content of the other container and the part of the current demand that remains to be satisfied, both at the moment that the amount in the chosen container is found to be insufficient.

In our first theorem, we obtain the distributional asymptotics of Z_t as $t \to \infty$. For Banach's matchbox problem, this has already been done by Uppuluri and Blot (1974); see also Holst (1989). Throughout, we assume that μ has finite second moment. We write '$\xrightarrow{\mathrm{D}}$' for convergence in distribution and '$\overset{\mathrm{D}}{=}$' for equality in distribution. Also, $Z \sim N(a, b)$ means that Z is a normally distributed random variable with mean a and variance b.

Theorem 1. *With $\sigma^2 := 2\,\mathrm{E}(W^2)/\mathrm{E}(W)$ we have*

$$t^{-1/2} Z_t \xrightarrow{\mathrm{D}} |Z| \quad \text{as } t \to \infty, \text{ with } Z \sim N(0, \sigma^2).$$

Proof. Let $\eta_0 := 0$ and $\eta_n := \inf\{i > \eta_{n-1} : H_i = 1\}$ for $n \in \mathbb{N}$, and let

$$X_n := W_{\eta_n}, \qquad Y_n := \sum_{k=\eta_{n-1}+1}^{\eta_n - 1} W_k$$

for $n \in \mathbb{N}$. Thus, $(X_n)_{n\in\mathbb{N}}$ is the sequence of demands made from container A or, in the random walk interpretation, of the moves towards the right, while $(Y_n)_{n\in\mathbb{N}}$ is the sequence of successive total demands from container B between two choices of container A or, in the random walk interpretation, the sums of the steps in the upward direction between two single steps towards the right. As usual, we take an empty sum to have the value 0, which arises in the present context if container A is chosen first and whenever there are two successive steps towards the right.

The random variables $X_1, X_2, X_3, \dots, Y_1, Y_2, Y_3, \dots$ are independent, and X_n has distribution μ for all $n \in \mathbb{N}$. The common distribution of the Y-variables is that of a random sum of terms with distribution μ, where the number of terms has the same distribution as $\eta_1 - 1$, which is the geometric distribution with parameter $\frac{1}{2}$. In particular,

$$\mathrm{E}(Y_n) = \mathrm{E}(\eta_1)\,\mathrm{E}(W) = \mathrm{E}(W), \qquad \mathrm{var}(Y_n) = \mathrm{E}(W^2) + (\mathrm{E}(W))^2.$$

Now let

$$T_n := \sum_{k=1}^{n} (Y_k - X_k) \quad \text{for } n \in \mathbb{N}.$$

This results in yet another sequence of independent and identically distributed random variables, with mean and variance given by

$$\mathrm{E}(Y_k - X_k) = 0, \qquad \mathrm{var}(Y_k - X_k) = 2\,\mathrm{E}(W^2).$$

Hence the central limit theorem can be applied, leading to

$$n^{-1/2} T_n \xrightarrow{\mathrm{D}} T \quad \text{as } n \to \infty, \text{ with } T \sim N(0, 2\,\mathrm{E}(W^2)).$$

Now let $(N_t)_{t \ge 0}$ be the renewal process associated with the X-variables, i.e.

$$N_t := \sup\left\{ n \in \mathbb{N}_0 : \sum_{k=1}^{n} X_k \le t \right\}.$$

It is known that N_t/t converges to the constant $1/\mathrm{E}(X_1)$ in probability as $t \to \infty$. This result, together with other renewal theoretic facts that we need below, can be found in Chapter XI of Feller (1971). We thus have

$$\frac{T_{N_t+1}}{\sqrt{t}} = \sqrt{\frac{N_t + 1}{t}}\,\frac{T_{N_t+1}}{\sqrt{N_t + 1}} \xrightarrow{\mathrm{D}} \frac{T}{\sqrt{\mathrm{E}(W)}} \quad \text{as } t \to \infty$$

by Anscombe's theorem; see for example Chung (1974, p. 216). Let $V_t := S_B(\tau_{A,t})$ be the y-coordinate of the two-dimensional random walk once the x-coordinate has crossed the level t. Then we have

$$\frac{1}{\sqrt{t}}(V_t - t) = \frac{1}{\sqrt{t}}\left(\sum_{k=1}^{N_t+1} Y_k - t \right) = \frac{1}{\sqrt{t}} T_{N_t+1} - \frac{1}{\sqrt{t}} L_t,$$

where

$$L_t := \sum_{k=1}^{N_t+1} X_k - t = S_A(\tau_{A,t}) - t$$

denotes the residual waiting time at time t associated with the renewal process $(N_t)_{t \ge 0}$. It is known that L_t converges in distribution as $t \to \infty$ in the nonlattice case; in the lattice case, t would need to be restricted to the multiples of the lattice width. In either case, $(L_t)_{t \ge 0}$ is stochastically bounded, which implies that $L_t/\sqrt{t} \to 0$ in probability as $t \to \infty$.

Putting pieces together we see that

$$\frac{V_t - t}{\sqrt{t}} \xrightarrow{\mathrm{D}} Z \quad \text{as } t \to \infty,$$

with $Z \sim N(0, \sigma^2)$, and σ^2 as in the statement of Theorem 1. It remains to relate Z_t to V_t. We have $Z_t = t - V_t$ on $\{\tau_t = \tau_{A,t}\}$, where $\tau_t = \tau_{A,t}$ is equivalent to $V_t \le t$. The symmetry of the step distribution of the two-dimensional random walk $(S_n)_{n\in\mathbb{N}_0}$,

$$\begin{pmatrix} H_i W_i \\ (1-H_i)W_i \end{pmatrix} \overset{\mathrm{D}}{=} \begin{pmatrix} (1-H_i)W_i \\ H_i W_i \end{pmatrix},$$

implies that

$$\mathrm{P}(Z_t \le z, \tau_t = \tau_{A,t}) = \mathrm{P}(Z_t \le z, \tau_t = \tau_{B,t}),$$

so that

$$\mathrm{P}(Z_t \le z) = 2\,\mathrm{P}(Z_t \le z, \tau_t = \tau_{A,t}) = 2\,\mathrm{P}(t - z \le V_t \le t)$$

for $0 \le z \le t$.

We now consider the probability that the remaining part of the demand from the empty container cannot be satisfied from the other container. Using the notation introduced above, we can write this as

$$\psi(t) := \mathrm{P}(Z_t < R_t).$$

As a corollary to Theorem 1 we obtain $\lim_{t\to\infty} \psi(t) = 0$ since, with the notation introduced in its proof,

$$\begin{aligned}
\psi(t) &= \mathrm{P}(Z_t < R_t, \tau_t = \tau_{A,t}) + \mathrm{P}(Z_t < R_t, \tau_t = \tau_{B,t}) \\
&= 2\,\mathrm{P}(V_t \le t < V_t + L_t) \\
&= 2\,\mathrm{P}(V_t \le t) - 2\,\mathrm{P}(V_t + L_t \le t) \\
&= 2\,\mathrm{P}\big(t^{-1/2}(V_t - t) \le 0\big) - 2\,\mathrm{P}\big(t^{-1/2}(V_t - t) - t^{-1/2}L_t \le 0\big) \\
&\to \tfrac{1}{2} - \tfrac{1}{2} = 0 \quad \text{as } t \to \infty,
\end{aligned}$$

where we have used symmetry and $R_t = L_t$ on $\{\tau_t = \tau_{A,t}\}$ in the second step and $L_t/\sqrt{t} \to 0$ in probability in the last step.

Our second theorem gives an explicit representation for ψ in the case of exponentially distributed demands. We write $\mathrm{Exp}(\lambda)$ for the exponential distribution with parameter λ.

Theorem 2. *If* $\mu = \mathrm{Exp}(\lambda)$*, then*

$$\psi(t) = \mathrm{e}^{-\lambda t} \sum_{n=0}^{\infty} \frac{1}{n!\,n!} \left(\frac{\lambda t}{2}\right)^{2n}.$$

Proof. Rescaling if necessary, we may assume that $\lambda = 1$. We continue to use the notation introduced in the proof of Theorem 1. In the case $\mu = \mathrm{Exp}(1)$ the residual waiting time L_t has distribution $\mathrm{Exp}(1)$ and is independent of N_t, which in turn has a Poisson distribution with parameter t. Given $N_t = n$, we can write V_t as the sum of $n+1$ independent random variables that are in turn random sums of independent random variables with distribution $\mathrm{Exp}(1)$, where the number of terms has a geometric distribution with parameter $\frac{1}{2}$. Hence V_t is then a random sum, where the number K_t of terms has a negative binomial distribution with parameters $n+1$ and $\frac{1}{2}$. In particular,

$$\mathrm{P}(N_t = n, K_t = k) = \mathrm{e}^{-t} \frac{t^n}{n!} \binom{k+n}{n} \frac{1}{2^{n+k+1}}.$$

If $K_t = k$, then the distribution of V_t is gamma with parameters k and 1, hence $V_t + L_t$ has a gamma distribution with parameters $k + 1$ and 1. Conditionally on $N_t = n$ and $K_t = k$, the event that $V_t \le t < V_t + L_t$ can therefore be seen as the event that there are exactly k points in the time interval $[0, t]$ in a unit rate Poisson process, and hence

$$\mathrm{P}(V_t \le t < V_t + L_t \mid N_t = n, K_t = k) = \mathrm{e}^{-t}\frac{t^k}{k!}.$$

This yields

$$\begin{aligned}
\psi(t) &= 2\,\mathrm{P}(V_t \le t < V_t + L_t) \\
&= 2\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \mathrm{P}(V_t \le t < V_t + L_t \mid N_t = n, K_t = k)\,\mathrm{P}(N_t = n, K_t = k) \\
&= 2\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \mathrm{e}^{-t}\frac{t^k}{k!}\mathrm{e}^{-t}\frac{t^n}{n!}\binom{k+n}{n}\frac{1}{2^{n+k+1}} \\
&= \mathrm{e}^{-2t}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\frac{t^{n+k}}{2^{n+k}}\frac{(k+n)!}{n!\,n!\,k!\,k!} \\
&= \mathrm{e}^{-2t}\sum_{n=0}^{\infty}\frac{1}{n!\,n!}\sum_{k=n}^{\infty}\frac{t^k}{2^k}\frac{k!}{(k-n)!\,(k-n)!} \\
&= \mathrm{e}^{-2t}\sum_{k=0}^{\infty}\frac{t^k}{2^k}\frac{1}{k!}\sum_{n=0}^{k}\binom{k}{n}^2 \\
&= \mathrm{e}^{-2t}\sum_{k=0}^{\infty}\frac{(2t)^k}{k!}\frac{1}{2^{2k}}\binom{2k}{k}.
\end{aligned}$$

Note that

$$\frac{1}{2^{2k}}\binom{2k}{k} = \mathrm{E}(U^k),$$

where the distribution of the random variable U is beta with parameters $p = q = \frac{1}{2}$. Thus, if M is a random variable which has the Poisson distribution with parameter $2t$ and is independent of U, then

$$\psi(t) = \mathrm{E}(U^M) = \mathrm{E}(\mathrm{e}^{2t(U-1)}) = \mathrm{e}^{-t}\,\mathrm{E}(\mathrm{e}^{-t(1-2U)}).$$

Due to the symmetry of the distribution of $1 - 2U$ about 0, the odd moments of $1 - 2U$ vanish. Using the fact that $(1 - 2U)^2 \stackrel{\mathrm{D}}{=} U$, we obtain

$$\begin{aligned}
\psi(t) &= \mathrm{e}^{-t}\sum_{n=0}^{\infty}\frac{t^{2n}}{(2n)!}\,\mathrm{E}(U^n) \\
&= \mathrm{e}^{-t}\sum_{n=0}^{\infty}\frac{t^{2n}}{(2n)!}\frac{1}{2^{2n}}\binom{2n}{n} \\
&= \mathrm{e}^{-t}\sum_{n=0}^{\infty}\frac{1}{n!\,n!}\left(\frac{t}{2}\right)^{2n}.
\end{aligned}$$

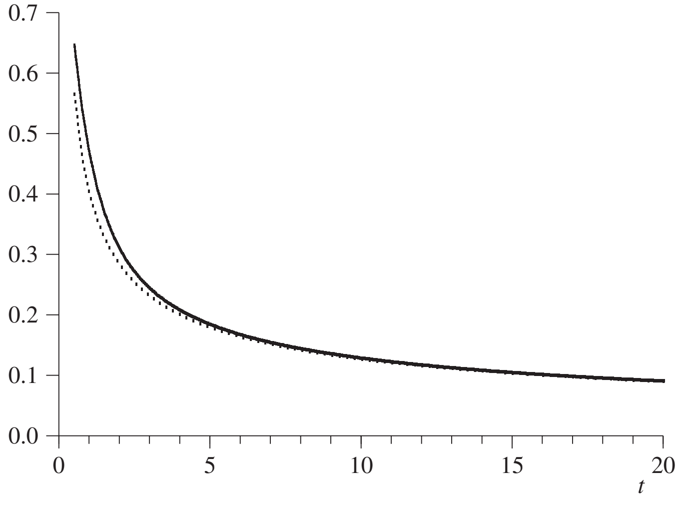

FIGURE 1: The probability ψ (solid line) and its approximation (dotted line).

Recalling the definition of the modified Bessel functions,

$$I_\nu(z) := \left(\frac{z}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{1}{\Gamma(\nu+k+1)k!}\left(\frac{z}{2}\right)^{2k},$$

we see that $\psi(t) = \mathrm{e}^{-\lambda t} I_0(\lambda t)$. It is known that $\mathrm{e}^{-t} I_\nu(t) \sim 1/\sqrt{2\pi t}$ as $t \to \infty$ (see for example Formula 9.7.1 in Abramowitz and Stegun (1964)), hence

$$\lim_{t\to\infty} \sqrt{t}\,\psi(t) = \frac{1}{\sqrt{2\pi\lambda}}.$$

Figure 1 shows the graph of ψ and the asymptotic approximation in the case $\lambda = 1$. It is interesting to note that the approximation is excellent even for moderate values of t and that the values of ψ are relatively large. If we accept a probability of 0.1 for the event that the current demand cannot be satisfied from the other container, then we need an initial capacity that is $16.1717\ldots$ times larger than the mean of the individual requests. If we accept a probability of 0.01, then we arrive at a factor of about 1592.

We close with three comments. First, the technique used in the proof of Theorem 1 for obtaining asymptotic normality of the y-position at the time that a level t is crossed in the x-direction as $t \to \infty$ easily generalizes to step distributions that do not satisfy the support constraints that are natural in the storage problem considered here and may therefore be of interest in its own right.

Second, the symmetry of the distribution of $1 - 2U$ about 0 and the equality in distribution $(1-2U)^2 \stackrel{\mathrm{D}}{=} U$ for random beta variables U with parameters $p = q = \frac{1}{2}$ used in the proof of Theorem 2 yield that the mth moment of $1 - 2U$ is

$$\sum_{k=0}^{m} \binom{m}{k} \frac{(-1)^k}{2^k} \binom{2k}{k} = \begin{cases} 0 & \text{if } m \text{ is odd,} \\ \dfrac{1}{2^m}\dbinom{m}{\frac{1}{2}m} & \text{if } m \text{ is even.} \end{cases}$$

This identity is known as 'Knuth's old sum'; see Prodinger (1994).

Third, the result of Theorem 2 may be written as $\psi(t) = \mathrm{P}(\tilde{N}_{A,t} = \tilde{N}_{B,t})$, where $\tilde{N}_A$ and $\tilde{N}_B$ denote the thinned Poisson processes in the A- and B-direction respectively that arise if we delete those points of the original Poisson processes that are not connected to a change in direction of the two-dimensional random walk. It would be interesting to have a probabilistic, 'noncomputational' proof of Theorem 2 that exploits this representation.

Acknowledgement

The problem discussed in this paper was brought to our attention by D. Morgenstern, founder of and Professor Emeritus at the Institut für Mathematische Stochastik, Universität Hannover; he also supervised Berghahn's diploma thesis.

References

ABRAMOWITZ, M. AND STEGUN, I. A. (1964). *Handbook of Mathematical Functions*. National Bureau of Standards, Washington, DC.

BERGHAHN, H.-H. (1966). Eine Verallgemeinerung des Banachschen 'Streichholzproblems'. Diploma Thesis, Universität Freiburg.

CACOULLOS, T. (1967). Asymptotic distribution for a generalized Banach match box problem. *J. Amer. Statist. Assoc.* **62,** 1252–1257.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd edn. Academic Press, New York.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edn. John Wiley, New York.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edn. John Wiley, New York.

HOLST, L. (1989). A note on Banach's match box problem. *Statist. Prob. Lett.* **8,** 441–443.

KNUTH, D. E. (1984). The toilet paper problem. *Amer. Math. Monthly* **91,** 465–470.

PRODINGER, H. (1994). Knuth's old sum—a survey. *Bull. EATCS* **54,** 232–245.

STIRZAKER, D. (1988). A generalization of the matchbox problem. *Math. Scientist* **13,** 104–114.

UPPULURI, V. R. R. AND BLOT, W. J. (1974). Asymptotic properties of the number of replications of a paired comparison. *J. Appl. Prob.* **11,** 43–52.