*The Annals of Statistics*
1991, Vol. 19, No. 2, 1106–1108

# A NOTE ON BLACKWELL AND HODGES (1957) AND DIACONIS AND GRAHAM (1981)

BY MICHAEL PROSCHAN

*National Heart, Lung, and Blood Institute*

The papers of Blackwell and Hodges (1957) and Diaconis and Graham (1981) contain an error concerning the asymptotic distribution of the number of returns to the origin of a constrained random walk. The correct distribution is given.

Blackwell and Hodges (1957) evaluate different designs to control selection bias in experiments. Suppose there are two different treatments and patients are entered sequentially into the experiment. One method of ensuring that there are an equal number of patients in the two treatments is to use random allocation: Half of the $2n$ patients in the study are randomly selected to receive treatment A, and the remaining $n$ patients receive treatment B. If the experimenter has control over who is allowed to enter the experiment, and if he or she has some idea as to which group the next patient will be assigned, the experimenter may attempt to bias the results by choosing very ill or very healthy patients (this is called *selection bias*). Blackwell and Hodges point out that the optimal policy in terms of largest expected number of correct guesses for an experimenter is, at each stage of the experiment, to guess that the next patient will be assigned to the treatment having occurred least frequently thus far. If there have been an equal number of patients assigned to A and B thus far, the experimenter may pick either treatment or flip a coin to decide. This strategy is also optimal in terms of making the number of guesses stochastically largest. This is pointed out by Diaconis and Graham (1981), who add that it is easily proved by backward induction. Let $B(t)$ be the number of successes in $t$ independent Bernoulli trials with parameter $\frac{1}{2}$. Blackwell and Hodges show that the number of correct guesses of the experimenter is $G = n + B(T)$, where $T$ is the number of returns to the origin of a symmetric random walk $S_i$, conditioned on the event that $S_{2n} = 0$. They assert that

(1) "$\ldots T/\sqrt{n}$ has asymptotically the distribution of an absolute normal deviate."

Diaconis and Graham use this to conclude that [(2.4) of their Theorem 1]

$$\text{(2)} \qquad \Pr\left(\frac{G-n}{\sqrt{n/4}} \le x\right) \to 2\Phi(x) - 1 \quad \text{for } x \ge 0.$$

Received December 1989; revised April 1990.
*AMS* 1980 *subject classifications*. 60C05, 62P10.
*Key words and phrases*. Random allocation, returns to the origin of a random walk, selection bias.

If (2) were correct, then using the optimal strategy, $\Pr(G \geq n + 1.645\sqrt{n/2})$ would be approximately 0.02 for large $n$. But this is smaller than the probability 0.05 associated with flipping a coin to guess the next treatment assignment! Since (2) does follow from (1), the problem must be with (1). It is known [Feller (1968)] that if $R$ is the number of returns to the origin of an unconditioned, symmetric random walk, then $R/\sqrt{2n}$ has asymptotically the distribution $\Phi^+$ of an absolute normal deviate. It is not difficult to establish that $T$ is stochastically larger than $R$, so it seemed that (1) must have had a typographical error and that it should have begun with $T/\sqrt{2n}$ instead of $T/\sqrt{n}$. Even making this correction, as can be seen from calculations in Blackwell and Hodges, the mean of $T/\sqrt{2n}$ converges to $\sqrt{\pi/2}$, and the mean of $\Phi^+$ is $\sqrt{2/\pi}$. Because $E(T/\sqrt{2n})^2 \to 4$ as $n \to \infty$ [this follows from calculations in Wei (1978)], $T/\sqrt{2n}$ is uniformly integrable. Thus the discrepancy in limiting means rules out $\Phi^+$ as its limiting distribution.

It turns out not to be difficult to compute explicitly the limit of the survival function $S(t)$ of $T$. Again let $R$ denote the number of returns to the origin of an unconditioned symmetric random walk $S_i$ by time $2n$, and let $\tau_j$ be the time of the $j$th return to the origin of $S_i$. Used in the following derivation is the fact that $\Pr(S_{2n} = 0) = \Pr(S_{2i} \neq 0,\ i = 1, \ldots, n)$. This fact, proven in Feller (1968) by a reflection argument, is used in the third line of the following:

$$\begin{aligned}
\Pr(T > j) &= \frac{\Pr(\tau_j < 2n,\ S_{2n} = 0)}{\Pr(S_{2n} = 0)} \\
&= \frac{2^{2n}}{\binom{2n}{n}} \sum_{i=1}^{n-1} \Pr(\tau_j = 2i,\ S_{2n} = 0) \\
&= \frac{2^{2n}}{\binom{2n}{n}} \sum_{i=1}^{n-1} \Pr(\tau_j = 2i,\ \tau_{j+1} > 2n) \\
&= \frac{2^{2n}}{\binom{2n}{n}} \left(\Pr(R = j) - \Pr(\tau_j = 2n)\right) \\
&= \frac{2^{2n}}{\binom{2n}{n}} \left[\frac{1}{2^{2n-j}}\binom{2n-j}{n} - \frac{j}{(2n-j)2^{2n-j}}\binom{2n-j}{n}\right] \\
&= \frac{\binom{2n-j}{n}(n-j)2^{j+1}}{\binom{2n}{n}(2n-j)}.
\end{aligned}$$

The next-to-last equality follows from Feller (1968). Suppose that $j$ and $n$ approach $\infty$ in such a way that $n - j$ and $n/j$ also tend to $\infty$. Applying

Stirling's formula, one obtains

$$\Pr(T > j) \sim \left(\frac{n - j/2}{n}\right)^n \left(\frac{n - j/2}{n - j}\right)^{n-j} \left(\frac{2n - j}{2(n - j)}\right)^{1/2}. \tag{3}$$

Now let $j = \lambda\sqrt{2n}$, where $\lambda$ is a constant. Then (3) is asymptotic to

$$\left(1 - \frac{\lambda}{\sqrt{2n}}\right)^n \left(1 + \frac{\lambda\sqrt{2n}}{2(n - \lambda\sqrt{2n})}\right)^{n - \lambda\sqrt{2n}}$$

It follows that

$$\begin{aligned} \ln\left(\Pr(T > \lambda\sqrt{2n})\right) &= n\left(\frac{-\lambda}{\sqrt{2n}} - \frac{\lambda^2}{4n} + o\left(\frac{1}{n}\right)\right) + (n - \lambda\sqrt{2n}) \\ &\quad \times \left(\frac{\lambda\sqrt{2n}}{2(n - \lambda\sqrt{2n})} - \frac{2n\lambda^2}{8(n - \lambda\sqrt{2n})^2} + o\left(\frac{1}{n}\right)\right) \\ &\to -\frac{\lambda^2}{2}. \end{aligned}$$

Therefore,

$$\Pr\left(\frac{T}{\sqrt{2n}} > \lambda\right) \to \exp\left(-\frac{\lambda^2}{2}\right),$$

a Weibull survival function with scale parameter $\sqrt{2}$ and shape parameter 2. Thus (1) and (2) should be replaced by

$$\frac{T}{\sqrt{2n}} \text{ is asymptotically Weibull}(\sqrt{2}, 2), \tag{1'}$$

$$\Pr\left(\frac{G - n}{\sqrt{n/2}} \le x\right) \to 1 - \exp\left(-\frac{x^2}{2}\right) \quad \text{for } x \ge 0. \tag{2'}$$

The probability of guessing more than $n + 1.645\sqrt{n/2}$ assignments correctly is roughly 0.26 for large $n$, as opposed to 0.05 associated with using coin flips to guess.

## REFERENCES


BLACKWELL, D. and HODGES, J. L. (1957). Design for the control of selection bias. *Ann. Math. Statist.* **28** 449–460.

DIACONIS, P. and GRAHAM, R. (1981). The analysis of sequential experiments with feedback to subjects. *Ann. Statist.* **9** 3–23.

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

WEI, L. J. (1978). On the random allocation design for the control of selection bias in sequential experiments. *Biometrika* **65** 79–84.



421 SNURE ROAD
SILVER SPRING, MARYLAND 20901