

---

What Time do the Baboons Come down from the Trees? (An Estimation Problem)

Author(s): Stephen S. Wagner and Stuart A. Altmann

Reviewed work(s):

Source: *Biometrics*, Vol. 29, No. 4 (Dec., 1973), pp. 623-635

Published by: [International Biometric Society](#)

Stable URL: <http://www.jstor.org/stable/2529130>

Accessed: 19/10/2012 13:36

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at  
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*International Biometric Society* is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*.

<http://www.jstor.org>

## WHAT TIME DO THE BABOONS COME DOWN FROM THE TREES? (AN ESTIMATION PROBLEM)

STEPHEN S. WAGNER AND STUART A. ALTMANN

*Allee Laboratory of Animal Behavior, University of Chicago, Chicago, Illinois 60637, U.S.A.*

### SUMMARY

We consider a system that can be in either of two observable states, that begins each day in the first state, and exactly once each day makes a transition to the second state. An observer records the time of transition on a number of days, but on certain other days arrives too late to witness the transition. On those days he can only say that the transition occurred sometime earlier than his arrival time. We show how to use all the available data to obtain an unbiased estimate of the distribution of transition times.

### 1. PROBLEM

In the course of analyzing data on the ecology of baboons in East Africa, the following problem in statistical estimation arose. Consider a system that can be in one of two observable states, that always begins the day in the first state and exactly once each day makes a transition to the second state. The observer's task is to estimate the probability distribution of the transition times. However, the time at which observations begin each day is variable, so that on some days the observer arrives in time to witness the transition, whereas on other days he arrives too late. How can such an observer utilize the data that are available to obtain an unbiased estimate of the probability that transitions will take place during any specified interval?

The biological problem that gave rise to this estimation problem is this. The Amboseli Reserve, Kenya, consists primarily of open savannah, punctuated here and there by a grove of trees or a water hole. Baboons in this area sleep in trees every night, but descend from the trees in the morning and move out across the savannah where they spend the day foraging, often quite far from their sleeping trees. One of the baboon groups in Amboseli was studied in detail. In this group the time—hereafter called the “descent time”—at which the median member of the group descended to the ground was recorded whenever we arrived in time to witness it. On a number of other days, however, we arrived too late to see this median descent and could only record the fact that more than half (usually all) of the members of the group were on the ground. On each day, the time at which observation on this group began was recorded. We would like to know the distribution of descent times, since it is an important variable in determining exactly

how the baboons allocate the hours of the day among the various activities or states necessary for their survival.

In this paper we shall present a technique for utilizing such data to obtain an estimate of the probability distribution. Although we describe the technique in terms of baboon descents, it will be clear that it applies as well to any other comparable data.

The African field research was carried out by the junior author in collaboration with Jeanne Altmann (see Altmann and Altmann [1970]). The estimation technique was developed by the authors of this paper.

## 2. METHOD

One might suppose that he could simply assign to each time-segment  $(t, t + \Delta)$  the proportion of observed descents falling in that segment. Unfortunately that procedure is biased: late descents were more likely to be witnessed than early ones, and hence that procedure would bias the results in favor of late descents (displace the distribution "to the right").

There are two ways in which the true probability distribution can be approximated. First, we could restrict ourselves to those days on which we actually witnessed the median descent and recorded its time. Such data can be used without bias in the following manner. For any time-segment  $(t, t + \Delta)$  use only data from days on which we arrived at the study group and began observing by time  $t$ ; assuming independence between arrival time and descent time, these data will be unbiased. Divide them into "successes", i.e. descents during the interval  $(t, t + \Delta)$ , and "failures", i.e. descents before or after  $(t, t + \Delta)$ . In the usual way, then, the number of successes divided by the number of (successes + failures) is the maximum likelihood (ML) estimate of the probability of descent during  $(t, t + \Delta)$ .

However, this technique makes no use of the times when we began observations part way through the interval in which we saw the baboons descend, data which could not be included without introducing bias, and it makes only incomplete use of the times when we arrived before the interval in which they descended. Furthermore, it ignores those times when we arrived and found the baboons on the ground. But the fact is that all three kinds of data contain some information about descent time and should not be discarded.

If, for example, we are interested in the time period (0845, 0900), and on a certain day we located the group at 0903 and they were *still in the trees*, we are justified in using this day as an instance of their being in the trees throughout the interval 0845–0900, and indeed, during all earlier intervals that day, on the assumption that the group does not descend more than once per morning. (Notice that, even if the group ascended and descended several times before moving off to begin foraging—something which it has never been observed to do—we would simply define it to be in state one until the last descent and in state two thereafter.) Or suppose that we are considering the time period (0845, 0900), and on some days we located the

group after 0900 and they were already on the ground. While the descent time for such "late" days is not known, the *distribution* of descent times for such days among the periods preceding our arrival is likely to be the same as the observed distribution for days on which we arrived before the interval; thus, data from "late" days might be apportioned among each preceding time interval on the basis of the probability of descending during that interval, as estimated from all descents during the interval that were observed on days when we arrived before the interval began. Or better yet, since data added from days on which we arrived after the onset of the interval can be used to increase the accuracy of the probability estimates, the apportionment for each "late" datum should also be based on this additional information.

The technique that we have developed makes use of all the data and does so in an unbiased way. It can be explained most easily through the use of the following notation: for any time-segments  $(a, b)$ ,  $(s, t)$ , and  $(u, v)$ , let  $D_i^s(a, b)$  denote the number of days on which observations began between  $a$  and  $b$  o'clock and the baboons descended between  $s$  and  $t$  o'clock, and let  $P(d_i^s | d_v^u)$  denote the probability that the baboons descend between  $s$  and  $t$  o'clock given that they descend between  $u$  and  $v$  o'clock. (For the sake of comparison, we remark that, in Altmann and Altmann [1970], the notation  ${}_v^a D_i^s$  was used for  $D_i^s(a, b)$ .)

We make the following three assumptions. First, we assume a stationary source, i.e. that the probabilities do not change from day to day. Second, we assume the characteristic mentioned earlier, namely, that the animals always begin the day in state one (in the trees) and make exactly one transition to state two (on the ground). Third, we assume that the time of transition from state one to state two is independent of the time at which observations begin, or, symbolically, that

$$\frac{D_i^s(a, b)}{D_v^u(a, b)} \simeq \frac{D_i^s(a', b')}{D_v^u(a', b')} \quad (1)$$

for any time-segments  $(a, b)$ ,  $(a', b')$ ,  $(s, t)$ , and  $(u, v)$  with  $(s, t) \subseteq (u, v)$ —where the equivalence symbol  $\simeq$  indicates that both sides are unbiased estimates of the probability  $P(d_i^s | d_v^u)$ .

Two remarks are in order concerning equation (1). (i) If the segment  $(a, b)$  is larger than the segment  $(a', b')$ , the left-hand side will typically be a better estimate of  $P(d_i^s | d_v^u)$  than the right-hand side, being based on a larger sample. Consequently, we shall always take  $a = 0$ , and  $b$  as large as possible. (ii) In general, however, we cannot take  $b$  larger (later) than  $u$  or the value of  $D_v^u(0, b)$  will be unknown. (For, if  $b > u$ , a missed descent after  $u$  cannot be distinguished from one before  $u$ .) An important exception is the case where  $u = 0$ : so long as  $b \leq v$ , the value of  $D_v^0(0, b)$  will be known.

We will estimate the probability of descent during any interval from the relative frequency of descents, both observed and estimated, during that interval. The frequency for each interval will be obtained from the sum of four terms,  $A$ ,  $B$ ,  $C$ , and  $D$ , which represent the number of descents during  $(t, t + \Delta)$  for each of the four possibilities of arrival time: either we arrived

before  $t$  (and therefore saw the median descent), or we arrived during  $(t, t + \Delta)$  but in time to see the descent, or we arrived during the interval but too late to see the descent, or finally, we arrived after  $t + \Delta$ , and thus found the baboons on the ground.

We make the following abbreviations. Let  $N$  denote the number of days of observation, including the days on which we arrived after the baboons' descent, and let  $\omega$  denote the latest time of observation. Let  $D_{t+\Delta}^t (= D_{t+\Delta}^t(0, \omega))$  denote the (usually unknown) number of days on which the baboons descended between time  $t$  and  $t + \Delta$ , and let  $D_{t+\Delta}^t(x)$  denote the number of days (typically 0 or 1, occasionally 2 or more because of the rounding-off of  $x$ ; unknown if either  $x > t$  and  $t \neq 0$  or  $x > t + \Delta$ ) on which observations began at  $x$  o'clock and the baboons descended between  $t$  and  $t + \Delta$  o'clock. Then

$$P(d_{t+\Delta}^t) \simeq \frac{D_{t+\Delta}^t}{N} = \frac{1}{N} (A + B + C + D),$$

where

$$A = D_{t+\Delta}^t(0, t) \quad \text{and} \quad B = \sum_{t < y < t+\Delta} D_{t+\Delta}^y(y)$$

are known through observation and

$$C = \sum_{t < y < t+\Delta} D_y^t(y) \quad \text{and} \quad D = D_{t+\Delta}^t(t + \Delta, \omega)$$

have to be estimated.

The first term,  $A$ , is an unbiased sample of descents observed during the interval; for this we must, as indicated above, use the number of just those days on which we began observing on or before the onset of the interval, and saw the baboons descend during the interval.

The second term,  $B$ , is the total number of descents in the interval on the days when we arrived during the interval but in time to witness the descent.

The third term is the number of days on which the descent occurred in the interval out of all the days on which we arrived during the interval and the group had, by then, already descended. The expected value of this number is

$$C = \sum_y D[y] P(d_{t+\Delta}^t | d_y^0) \quad (t < y < t + \Delta),$$

where  $D[y]$  abbreviates  $D_y^0(y)$ , the number of days on which we arrived at time  $y$  and found the baboons on the ground. On such days the median descent occurred at some time before  $y$ . The probability  $P(d_{t+\Delta}^t | d_y^0)$  can be estimated from

$$\frac{D_y^t(0, t)}{D_t^0(0, t) + D_y^t(0, t)} = D_y^t(0, t) / D_y^0(0, t),$$

so we have the estimate

$$C = \sum_y D[y] \frac{D_y^t(0, t)}{D_y^0(0, t)}$$

where the summation is carried out over all  $y$  in  $(t, t + \Delta)$  for which  $D[y]$  is not zero.

The fourth term is the number of descents expected in the interval out of those days not included so far, namely, those on which we began observations after the end of the interval and the group had already descended. The expected value for these days is

$$D = \sum_z D[z] P(d_{t+\Delta}^t \mid d_z^0) \quad (z > t + \Delta),$$

where, once again,  $D[z]$  abbreviates  $D_z^0(z)$ , the number of late arrivals at time  $z$ . The probability factors here could be estimated in an unbiased way as  $D_{t+\Delta}^t(0, t)/D_z^0(0, t)$  by the same reasoning as in the calculation of  $C$ . However, the technique that follows uses more of the available data and is therefore to be preferred.

We note that because  $0 < t < t + \Delta$ , we have  $d_{t+\Delta}^t \equiv d_{t+\Delta}^0 \cap d_{t+\Delta}^t$ , and thus

$$\begin{aligned} D &= \sum_{z>t+\Delta} D[z] \cdot P(d_{t+\Delta}^0 \cap d_{t+\Delta}^t \mid d_z^0) \\ &= \sum_{z>t+\Delta} D[z] \cdot P(d_{t+\Delta}^t \mid d_{t+\Delta}^0 \cap d_z^0) \cdot P(d_{t+\Delta}^0 \mid d_z^0), \end{aligned}$$

where the extra factor has been introduced by recalling that for any events  $E, F$ , and  $G$ ,  $P(E \cap F \mid G) = P(F \mid E \cap G) \cdot P(E \mid G)$ . But  $d_{t+\Delta}^0 \cap d_z^0 = d_{t+\Delta}^0$ , so we have

$$D = \sum_{z>t+\Delta} D[z] \cdot P(d_{t+\Delta}^t \mid d_{t+\Delta}^0) \cdot P(d_{t+\Delta}^0 \mid d_z^0).$$

This can be estimated as follows

$$\begin{aligned} D &= \sum_{z>t+\Delta} D[z] \frac{D_{t+\Delta}^t(0, t)}{D_t^0(0, t)} \cdot \frac{D_{t+\Delta}^0(0, t + \Delta)}{D_z^0(0, t + \Delta)} \\ &= \frac{D_{t+\Delta}^t(0, t)}{D_{t+\Delta}^0(0, t)} \cdot \frac{D_{t+\Delta}^0(0, t + \Delta)}{D_z^0(0, t + \Delta)} \sum_{z>t+\Delta} \frac{D[z]}{D_z^0(0, t + \Delta)}, \end{aligned}$$

where the summation is carried out over all  $z$  greater than  $t + \Delta$  for which  $D[z]$  is not zero.

The final formula is therefore

$$\begin{aligned} P(d_{t+\Delta}^t) &= \frac{1}{N} \left[ D_{t+\Delta}^t(0, t) + \sum_y D_{t+\Delta}^y(y) + \sum_y D[y] \frac{D_y^t(0, t)}{D_y^0(0, t)} \right. \\ &\quad \left. + \frac{D_{t+\Delta}^t(0, t)}{D_{t+\Delta}^0(0, t)} \frac{D_{t+\Delta}^0(0, t + \Delta)}{D_z^0(0, t + \Delta)} \sum_z \frac{D[z]}{D_z^0(0, t + \Delta)} \right], \quad (2) \end{aligned}$$

where  $t < y < t + \Delta$  and  $z > t + \Delta$  and, as before,  $D[y]$  and  $D[z]$  represent the number of late arrivals at times  $y$  and  $z$  respectively.

This formula is applicable to all intervals except the first—the so-called “left tail”, say  $(0, t)$ . For that we have the simpler formula

$$P(d_i^0) = \frac{1}{N} \left[ D_i^0(0, t) + D_i^0(0, t) \sum_{z>t} \frac{D[z]}{D_z^0(0, t)} \right]. \quad (3)$$

For both these formulas the time  $t$  (and the number  $\Delta$ ) must be chosen large enough that, out of the days we arrived by time  $t$ , at least one descent occurred by the earlier of  $t + \Delta$  and the time, say  $y$ , of the earliest arrival after  $t$  with a missed descent. This is necessary and sufficient to make both  $D_y^0(0, t)$  and  $D_{t+\Delta}^0(0, t)$  nonzero and hence to ensure that no fraction in either formula has a zero denominator.

### 3. COMPUTATION PROCEDURE

Choice of interval size is an important initial consideration in these calculations, since the numbers obtained via equation (2) for adjacent intervals cannot be lumped. Our choice of  $\Delta$  as 15 minutes was based on values of  $D_{t+\Delta}^0(0, t)$ —the first term of equation (2)—since they are easy to obtain and indicate the minimal sample size per interval. We tried to keep  $\Delta$  small, so that the number of intervals (and thus the number of degrees of freedom) would be kept large.

In deciding which intervals should be grouped together to form the left and right tails, we lumped until each tail included at least one observed descent. A less stringent, but equally satisfactory requirement would be that each tail contain one *known* descent (observed or deducible). An even less stringent requirement—but one which *must* be satisfied to avoid a zero denominator—is this: if the first interval (the left tail) contains no known descent, then it must contain the arrival time corresponding to the earliest observed descent, *and* that descent must occur in the second interval and earlier than any late arrival. If this condition is satisfied, no zero denominators will be encountered in the calculations for any interval.

The estimation technique that we have presented in equations (2) and (3) leads to a rather formidable counting task. Worse, it will have to be done by hand, since, in any situation that we can imagine, to program the procedure and prepare the data for a computer would require even more time. For that reason, we have devised a graphic technique that greatly facilitates the counting, and makes it practicable even with moderately large samples. The technique is as follows:

(1) Prepare a list of observed descent times, indicating also the time that observations began on each of these days (see Table 1). Prepare a separate list of the times of day that observations began for those days on which you arrived after the descent of the median number of the group (see Table 2).

(2) Label a large sheet of graph paper with time of day across the top, allowing one graph line for each minute—or for whatever was the smallest time unit used in the observations. (We used graph paper that was ruled

TABLE 1  
DESCENT TIME DATA FOR DAYS ON WHICH ARRIVAL WAS BEFORE MEDIAN DESCENT

	Date	Arrival time	Descent time		Date	Arrival time	Descent time
1.	25-11-63	0655	0656	30.	5-6-64	0825	0844
2.	29-10-63	0657	0659	31.	17-7-64	0842	0845
3.	5-11-63	0658	0720	32.	12-6-64	0815	0846
4.	12-2-64	0715	0721	33.	28-2-64	0730	0848
5.	29-3-64	0634	0743	34.	14-5-64	0830	0850
6.	14-2-64	0738	0747	35.	7-7-64	0831	0855
7.	18-2-64	0729	0750	36.	6-7-64	0822	0858
8.	1-4-64	0727	0751	37.	2-7-64	0837	0858
9.	8-2-64	0732	0754	38.	17-3-64	0803	0859
10.	26-5-64	0758	0758	39.	10-6-64	0848	0859
11.	19-2-64	0731	0805	40.	11-3-64	0830	0900
12.	7-6-64	0758	0808	41.	23-7-64	0807	0904
13.	22-6-64	0753	0810	42.	27-2-64	0723	0905
14.	24-5-64	0753	0811	43.	31-3-64	0750	0905
15.	21-2-64	0750	0815	44.	10-4-64	0824	0907
16.	13-2-64	0734	0815	45.	22-4-64	0833	0908
17.	11-6-64	0805	0820	46.	7-3-64	0832	0910
18.	21-6-64	0756	0820	47.	29-2-64	0815	0910
19.	13-3-64	0820	0825	48.	13-5-64	0758	0915
20.	12-7-64	0817	0827	49.	20-4-64	0830	0920
21.	30-6-64	0758	0828	50.	27-4-64	0801	0930
22.	5-5-64	0823	0831	51.	28-4-64	0835	0930
23.	12-5-64	0817	0832	52.	23-4-64	0900	0932
24.	25-4-64	0831	0832	53.	4-3-64	0845	0935
25.	26-3-64	0810	0833	54.	6-5-64	0840	0935
26.	18-3-64	0813	0836	55.	26-6-64	0815	0945
27.	15-3-64	0711	0840	56.	25-3-64	0722	0948
28.	6-3-64	0755	0842	57.	8-7-64	0821	0952
29.	11-5-64	0817	0844	58.	21-4-64	0810	1027

10 lines to the half inch, both vertically and horizontally, but somewhat wider rulings would be easier to read.) This axis will represent descent times. For each day on which you arrive in time, the observed descent time will be plotted along this axis. For any day on which you arrive too late, the arrival time will be plotted; this is the latest possible descent time for that day. In the left margin, at any convenient spacing, mark off and label vertical segments to represent the time intervals into which the observer's arrival times will be grouped, with the first (earliest) time period beginning about one inch below the top of the graph. "Convenient" means one or two inches per interval—somewhat more if very large samples were obtained. Draw a horizontal "base line" at the top of the graph, but below the time marks (see Figure 1); the use of this line will be explained below.

(3) Draw a "staircase" as follows. On a horizontal graph line near the lower end of each arrival-time segment, draw a horizontal "tread" line under the corresponding column segment. Join the ends of adjacent horizontal treads with vertical lines ("risers"), as in Figure 1.

(4) Plot each observed descent as a red dot, positioned horizontally according to the time of median descent and vertically according to the time segment in which observations began that day. Place the dots on the vertical graph lines and just above the segment treads. If a descent was recorded on the boundary between two time segments, divide it in half



TABLE 2  
DESCENT TIME DATA FOR DAYS ON WHICH ARRIVAL WAS AFTER MEDIAN DESCENT

Date	Arrival time	Date	Arrival time	Date	Arrival time
1.	1-12-63	0705	32.	13-10-63	0840
2.	6-11-63	0710	33.	4-7-64	0845
3.	24-10-63	0715	34.	3-5-64	0850
4.	26-11-63	0720	35.	25-5-64	0851
5.	18-10-63	0720	36.	24-11-63	0853
6.	7-5-64	0730	37.	15-7-64	0855
7.	7-11-63	0740	38.	16-2-64	0856
8.	23-11-63	0750	39.	10-3-64	0857
9.	28-11-63	0750	40.	28-7-64	0858
10.	27-11-63	0753	41.	18-6-64	0858
11.	28-5-64	0755	42.	20-2-64	0858
12.	5-7-64	0757	43.	2-8-64	0859
13.	28-3-64	0800	44.	27-5-64	0900
14.	23-3-64	0805	45.	28-10-64	0905
15.	26-10-63	0805	46.	15-5-64	0907
16.	11-7-64	0805	47.	10-5-64	0908
17.	27-7-64	0807	48.	27-6-64	0915
18.	9-6-64	0810	49.	11-10-63	0915
19.	24-6-64	0812	50.	17-2-64	0920
20.	16-10-63	0812	51.	22-10-63	0920
21.	25-2-64	0813	52.	10-7-64	0925
22.	6-6-64	0814	53.	14-7-64	0926
23.	22-11-63	0815	54.	11-4-64	0931
24.	10-10-63	0815	55.	23-5-64	0933
25.	2-11-63	0815	56.	30-7-64	0943
26.	23-6-64	0817	57.	18-7-64	0945
27.	24-4-64	0823	58.	29-7-64	0946
28.	3-7-64	0830	59.	16-7-64	0950
29.	29-4-64	0831	60.	22-7-64	0955
30.	4-8-64	0838	61.	15-10-63	0955
31.	7-10-63	0840	62.	19-10-63	1005
63.	2-5-64	1012			
64.	1-3-64	1018			
65.	17-10-63	1020			
66.	23-10-63	1020			
67.	25-7-64	1020			
68.	13-7-64	1031			
69.	8-6-64	1050			
70.	9-3-64	1050			
71.	26-4-64	1100			
72.	14-10-63	1205			
73.	18-11-63	1245			
74.	2-3-64	1250			
75.	8-5-64	1405			
76.	1-7-64	1407			
77.	12-10-63	1500			
78.	31-7-64	1531			
79.	6-10-63	1535			
80.	19-6-64	1556			
81.	29-6-64	1603			
82.	9-5-64	1605			
83.	9-10-63	1625			
84.	8-3-64	1625			
85.	11-2-64	1653			
86.	30-5-64	1705			
87.	5-3-64	1708			
88.	26-2-64	1722			
89.	4-5-64	1728			
90.	12-3-64	1730			
91.	25-10-63	1730			
92.	29-11-63	1750			
93.	22-2-64	1801			
94.	22-3-64	1829			

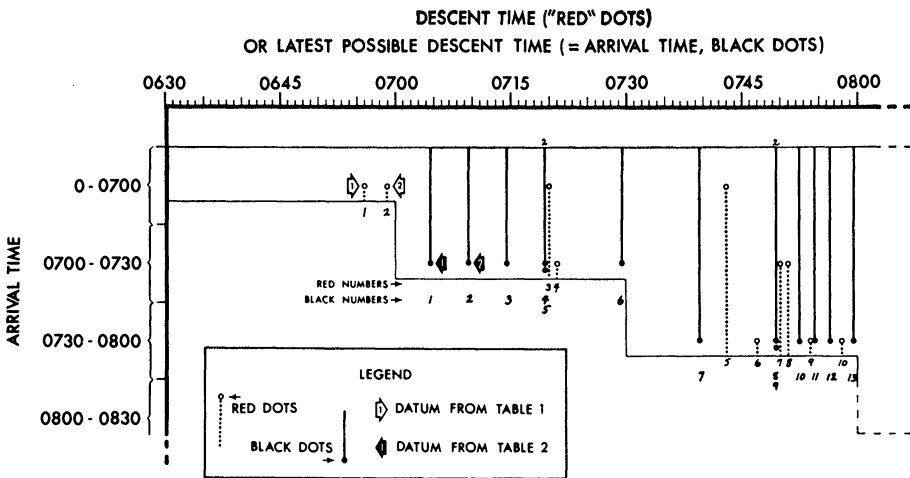


FIGURE 1  
GRAPH USED TO FACILITATE COUNTING. FOR EXPLANATION, SEE TEXT.

between the two; make a notation to indicate that these are half dots. If an arrival time falls on a time segment boundary, place the dot in the earlier of the two segments. Whenever more than one dot falls in the same position, place the additional dots immediately above those already plotted.

These procedures for handling boundary data are entailed by assumptions not heretofore described. With a finite number of observations and a finite number of segments, there is, in theory, a zero probability that any arrival time or descent time will fall exactly on the boundary between two segments. In practice, of course, one cannot record a time with an infinite number of significant figures, so an event may very well appear to fall on a boundary and hence outside of any segment. It is no solution to use intervals (i.e. inclusive of end-points) instead of segments, for then the categories would not be mutually exclusive. It is also no solution to use categories of the form  $[a, b)$ —i.e. including the left end-point but not the right—for that would tend to inflate the later categories and deflate the earlier. One way to handle the problem, and the way we have chosen, is to treat any descent that was recorded on the boundary between two time segments as half a descent in the one segment and half a descent in the other. A justification for this procedure might be found in the consideration that the probability is one-half that the true time of descent is within the first segment and one-half that it is in the second. On the other hand, whenever our *arrival* was recorded on a time boundary, we considered that the true arrival time was strictly before the boundary time. If one uses the above procedures, the number of events falling in the interval  $[a, b)$  is the same as the number falling in the segment  $(a, b)$ .

(5) Plot a black dot for each day on which observations began after the descent of the median animal, aligning the dot horizontally in the *space* immediately preceding the vertical line for the minute in which the arrival took place, [cf. (4)] and vertically in the corresponding time segment.

(6) Draw a vertical red line from each red dot down to the tread below it. For half dots, draw dashed lines. Draw a vertical black line from each black dot up to the base line (see step 2). If a black or red line represents more than one dot, indicate the number above the base line (black dots) or below the tread (red dots).

(7) Number the red dots from left to right, placing red numbers immediately below each tread. Similarly number the black dots, again beginning with 1 for the left-most dot.

This completes the graph. With it, the counts specified by the components of equation (2) for each time-segment  $(t, t + \Delta)$  become routine.

In order, they are found as follows:  $N$  is the total number of days in the sample. The number  $D_{t+\Delta}^r(0, t)$  will be the number of red dots directly (vertically) above the tread of time-segment  $(t, t + \Delta)$  that are at least as high as some higher tread. The number  $\sum_y D_{t+\Delta}^r(y)$  is the number of all remaining red dots directly above the tread of segment  $(t, t + \Delta)$ . Thus, the sum  $D_{t+\Delta}^r(0, t) + \sum_y D_{t+\Delta}^r(y)$  is simply the total number of red dots above the tread of  $(t, t + \Delta)$ .

The next three numbers are obtained for each minute  $y$  in the time segment, or, practically speaking, for each minute  $y$  such that  $D[y]$  is non-zero. The number  $\sum_v D[y]$  is the number of black dots in the space just to the left of time-line  $y$  (minute-line  $y$  in our study), yet still directly above the tread of  $(t, t + \Delta)$ . The number  $D_v^t(0, t)$  is the number of red dots that are anywhere left of time-line  $y$  and which lie at least one time-segment above the tread of  $(t, t + \Delta)$ . The number  $D_v^0(0, t)$  is the number of red and black dots that are anywhere to the left of time-line  $y$  and that are at least as high as the tread of  $(t, t + \Delta)$ .

The next number,  $D_{t+\Delta}^t(0, t)$ , has already been described. The number  $D_{t+\Delta}^0(0, t + \Delta)$  is the total number of dots (red or black) anywhere on or left of time-line  $t + \Delta$  but at least as high as the tread of  $(t, t + \Delta)$ .

The number  $D_{t+\Delta}^0(0, t)$  is the total number of dots (red or black) that are on or left of time-line  $t + \Delta$  and are at least as high as the next higher tread.

The next two numbers are obtained for each late arrival time  $z$  after the end of time-segment  $(t, t + \Delta)$ . The number  $D[z]$  is the number of black dots that are in the space immediately to the left of time-line  $z$ . Because  $z > t + \Delta$ , these dots are all above treads of later (lower) time segments. The number  $D_z^0(0, t + \Delta)$  is the number of red or black dots strictly to the left of time-line  $z$  and at least as high as the tread of  $(t, t + \Delta)$ . (Since we are dealing here with a descent that is known to have occurred before time  $z$ , the time  $z$  itself is not included.)

The graph also facilitates the counting that is specified by equation (3), for the left tail of the distribution.  $N$  is the total number of days in the sample.  $D_t^0(0, t)$  is the number of black and red dots directly above the left-most tread. The next two numbers are obtained for each minute  $z$  that is greater than  $t$  (i.e. anywhere to the right of the left-most riser) and for which  $D[z]$  is nonzero.  $D[z]$  is the number of black dots in the space immediately to the left of time-line  $z$ .  $D_z^0(0, t)$  is the number of black and red dots that are strictly to the left of minute  $z$  and above the level of the left-most tread. (As before, the time  $z$  is not included.)

#### 4. RESULTS

Data on descent of yellow baboons from groves of fever trees are given in Tables 1 and 2. These data were obtained in the Masai-Amboseli Game Reserve, Kenya, during 1963-64. The ecology of this baboon population has been described by Altmann and Altmann (1970). The probability distribution of descent times (Figure 2) was obtained from the data in Tables 1 and 2 by means of equations (2) and (3).

Since each day's sample is essentially a Bernoulli trial, binomial confidence limits may be placed on the estimate for each time period. For the purpose of calculating confidence limits for each time period  $(t, t + \Delta)$ , one must associate with that period a sample size. That size is certainly no less than the number  $N_t = D_w^0(0, t)$  of days on which we arrived by the beginning

of the interval. In fact, since some additional information is gained from the days on which we arrived after the beginning of the interval, the "true" sample size is no doubt larger than  $N_t$ —but probably not as large as  $N_t + D_{\omega}^0(t, \omega) = N$ . Those two values,  $N_t$  and  $N$ , determine an outer and an inner pair of confidence limits for the period  $(t, t + \Delta)$ , both of which are plotted in Figure 2.

The results, using 15-minute time periods, are shown in the bottom graph of Figure 2. For comparison, we have also carried out the calculations using 30-minute time periods (Figure 2, top graph). This gives smaller confidence intervals for the probability estimates because of the larger sample sizes, but of course the resulting probability statements have less temporal precision. (As we noted earlier, we cannot pool frequencies of adjacent 15-minute intervals to get the frequency distribution for 30-minute

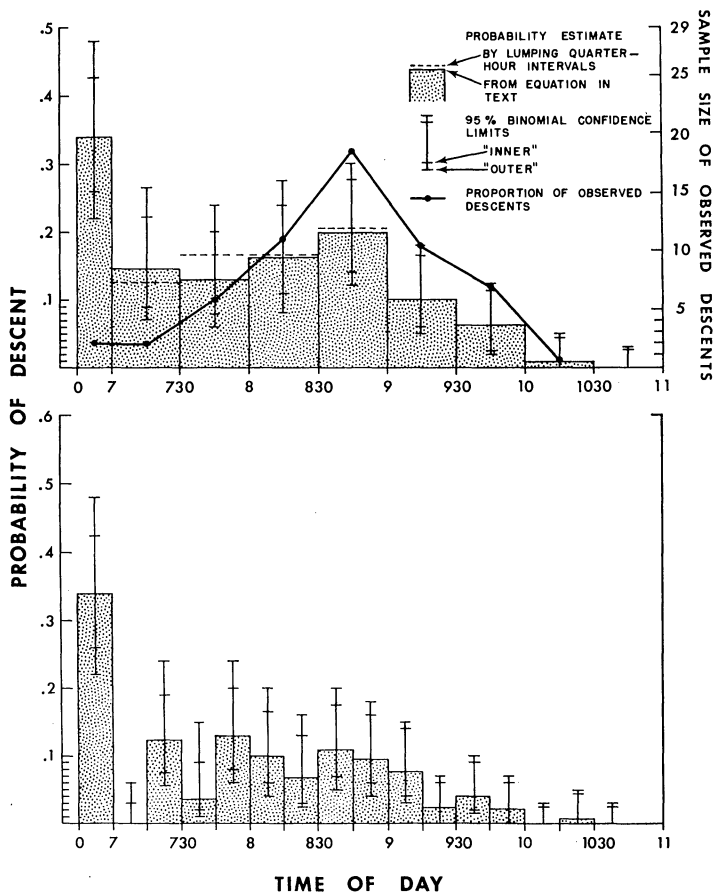


FIGURE 2

DISTRIBUTION OF DESCENT TIMES. FOR EXPLANATION, SEE TEXT. (REPRODUCED FROM ALTMANN AND ALTMANN [1970] BY PERMISSION OF S. KARGER AG).

intervals. Rather, for each new choice of interval size, we must carry out a new frequency count according to equations (2) and (3). The uncorrected distribution of observed descent times is shown in the upper graph of Figure 2; the bias of the data, particularly against early descents, is evident.

### 5. UNSOLVED PROBLEMS

Although unbiased estimates of relative frequencies were used to estimate the probabilities that are components of equations (2) and (3), we have not demonstrated that the resulting probabilities of transformation from state one to state two during each time interval are ML estimates, although it seems reasonable that they are. (A referee doubted this.)

It is not clear how to obtain unbiased estimates of the mean and variance of the population, since the distribution of some data is contingent upon the distribution of others.

The probability estimates obtained from equations (2) and (3) may not add up to 1, when summed over all time segments. This is so because the sample varies from one interval to the next. Any attempt to normalize the probabilities should take into account these differences in sample size, but the best way to do this is as yet unclear.

A referee has proposed an estimation procedure different from ours, namely that of first estimating the cumulative distribution function (cdf) and then differencing the cdf to construct the frequency histogram. Thus, the cdf being given by equation (3), the probability estimates would be given by

$$\begin{aligned} P(d_{t+\Delta}^1) &= P(d_{t+\Delta}^0) - P(d_t^0) \\ &= \frac{1}{N} D_{t+\Delta}^0(0, t + \Delta) \left[ 1 + \sum_{z>t+\Delta} \frac{D[z]}{D_z^0(0, t + \Delta)} \right] \\ &\quad - \frac{1}{N} D_t^0(0, t) \left[ 1 + \sum_{y>t} \frac{D[y]}{D_y^0(0, t)} \right]. \end{aligned}$$

This would have the advantage of being computationally simpler, and of giving a histogram already normalized. A disadvantage would be that the estimated cdf may fail to be nondecreasing, so that in certain cases some of the probability estimates would be negative. Also we suspect, but have been unable to establish, that this estimate will be subject to greater fluctuations than ours just because it is formed from differences rather than sums. With regard to other, perhaps more important, measures of superiority, such as a "fuller" use of the available data or greater "efficiency", the questions remain open.

### ACKNOWLEDGMENT

Field expedition supported by grants GB-683 from NSF, and MH-07336-01 from NIH. Data analysis supported by grants GB-2879, GB-4415, and GB-3903 from NSF, and by FR-00165 and MH-19617 from U.S.P.H.S.

A QUELLE HEURE LES BABOUINS DESCENDENT-ILS DES ARBRES?  
(UN PROBLEME D'ESTIMATION).

## RESUME

On considère un système qui peut être en l'un ou l'autre de deux états observables, qui commence chaque jour dans le premier état et, exactement une fois par jour, effectue une transition vers le second état. Un observateur note l'heure de transition un certain nombre de jours, mais certains autres jours il est arrivé trop tard pour être témoin de la transition. Ces jours là il peut seulement dire que la transition s'est passée antérieurement à son arrivée. Nous montrons comment utiliser toutes les données disponibles pour obtenir un estimateur sans biais de la distribution des temps de transition.

## REFERENCES

Altmann, S. A. and Altmann, J. [1970]. *Baboon Ecology: African Field Research*. University of Chicago Press, Chicago, and S. Karger, Basel (Bibliotheca Primatologica No. 12).

*Received May 1970, Revised March 1973*

*Key Words:* Censored data; Biased data; Transition times; Circadian rhythms; Diurnal cycles.