

ADA 170 057

AD-A170 057

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY NA			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA			5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR. 88-0442		
4. PERFORMING ORGANIZATION REPORT NUMBER(S) FSU Statistics Report No. M727			7a. NAME OF MONITORING ORGANIZATION AFOSR-NM		
6a. NAME OF PERFORMING ORGANIZATION Florida State University		6b. OFFICE SYMBOL (If applicable)	7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling Air Force Base, DC 20332-6448		
6c. ADDRESS (City, State and ZIP Code) Department of Statistics Tallahassee, FL 32306-3033			9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0007		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		8b. OFFICE SYMBOL (If applicable)	10. SOURCE OF FUNDING NOS.	PROGRAM ELEMENT NO. 6.1102F	PROJECT NO. 2304
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448			TASK NO.	WORK UNIT NO.	
11. TITLE (Include Security Classification) Easily-Stated But Hard Statistical Problems					
12. PERSONAL AUTHOR(S) Myles Hollander					
13a. TYPE OF REPORT Technical		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Yr., Mo., Day) May, 1986	15. PAGE COUNT 17
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary; use block number)		
FIELD	GROUP	SUB. GR.			
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This is an expository paper written in response to an invitation to give a non-technical talk in a session devoted to "Practical Aspects of Statistics" at the 1986 Annual Statistical Meetings. The goal of the session is to encourage interest in statistics among non-statisticians. ^{The original case} I have chosen to describe three problems of current research interest. The problems have the feature that they can be stated in a relatively easy fashion. The solutions however are difficult. References to partial solutions are given; all three problems have aspects that remain unsolved and are currently under study. The problem of Section 2 deals with survivorship data and concerns estimation of average remaining life. Section 3 considers a problem that pertains to assessing the degree of					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL Frank Proschan/Myles Hollander			22b. TELEPHONE NUMBER (Include Area Code) (904)644-3218	22c. OFFICE SYMBOL AFOSR/NM	

DTIC ELECTE
S JUL 23 1986 D

DTIC FILE COPY

similarity between species' presence or absence on islands. Section 4 presents a problem in geometrical probability. To conform to the spirit of the session, I have chosen to describe the problems in words, de-emphasizing symbols and mathematics and aiming for the non-statistician.

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution / _____	
Availability Codes	
Dist	Avail. and/or Special
A-1	



Easily-Stated But Hard Statistical Problems

by

Myles Hollander¹

FSU Statistics Report No. M727
AFOSR Technical Report No. 86-188

May, 1986

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

¹Research supported by the Air Force Office of Scientific Research under Grant AFOSR F49620-85-C-0007 to the Florida State University.

Approved for public release
distribution unlimited.

86 7 23 16 3

Easily-Stated But Hard Statistical Problems

Myles Hollander¹

The Florida State University

1. INTRODUCTION

This paper was written in response to an invitation to deliver a non-technical talk at the 1986 Annual Statistical Meetings. Dr. Robert L. Mason, the session organizer, charged the speakers with encouraging "interest in statistics among non-statisticians." I have chosen to describe three problems of current research interest. The problems have the feature that they can be stated in a relatively easy fashion. The solutions however are difficult. References to partial solutions are given; all three problems have aspects that remain unsolved and are currently under study. The problem of Section 2 deals with survivorship data and concerns estimation of average remaining life. Section 3 considers a problem that pertains to assessing the degree of similarity between species' presence or absence on islands. Section 4 presents a problem in geometrical probability.

To conform to the spirit of the session, I have chosen to describe the problems in words, de-emphasizing symbols and mathematics and aiming for the non-statistician.

2. HOW MUCH TIME IS LEFT?

Table 1 gives estimated values of the average remaining lifetime for the female population of the United States corresponding to the 1969 - 1971 era.

For example, the entry 56.59 corresponding to the age interval 20 - 21 is the

¹Research supported by the Air Force Office of Scientific Research under Grant AFOSR F49620-85-C-0007 to Florida State University.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
OFFICE OF OPERATIONAL RESEARCH (AFOSR/OOR)
This technical report has been reviewed and is
approved for publication under AFOSR 190-12.
DISTRIBUTION IS UNLIMITED.
MATTHEW J. KERFER
Chief, Technical Staff

remaining number of years of life a female may expect to live on the average as she celebrates her twentieth birthday.

Table 1. Estimated average remaining lifetimes for females, United States, 1967 - 1971.

Age Interval	Average Number of Years of Life Remaining at Beginning of Age Interval	Age Interval	Average Number of Years of Life Remaining at Beginning of Age Interval
Days		Years	
0-1	74.64	25-26	51.80
1-7	75.21	26-27	50.84
7-28	75.50	27-28	49.88
28-365	75.54	28-29	48.92
		29-30	47.97
Years		30-31	47.01
0-1	74.64	31-32	46.06
1-2	74.97	32-33	45.11
2-3	74.05	33-34	44.16
3-4	73.11	34-35	43.22
4-5	72.16	35-36	42.28
5-6	71.19	36-37	41.34
6-7	70.22	37-38	40.41
7-8	69.25	38-39	39.48
8-9	68.27	39-40	38.56
9-10	67.29	40-41	37.64
10-11	66.31	41-42	36.73
11-12	65.33	42-43	35.82
12-13	64.35	43-44	34.91
13-14	63.36	44-45	34.02
14-15	62.38	45-46	33.13
15-16	61.41	46-47	32.24
16-17	60.44	47-48	31.37
17-18	59.47	48-49	30.49
18-19	58.51	49-50	29.63
19-20	57.55	50-51	28.77
20-21	56.59	51-52	27.92
21-22	55.63	52-53	27.08
22-23	54.67	53-54	26.24
23-24	53.71	54-55	25.41
24-25	52.75		

Table 1 (continued).

Age Interval	Average Number of Years of Life Remaining at Beginning of Age Interval	Age Interval	Average Number of Years of Life Remaining at Beginning of Age Interval
Years		Years	
55-56	24.59	85-86	5.63
56-57	23.77	86-87	5.28
57-58	22.97	87-88	4.96
58-59	22.17	88-89	4.67
59-60	21.38	89-90	4.40
60-61	20.60	90-91	4.14
61-62	19.83	91-92	3.90
62-63	19.06	92-93	3.69
63-64	18.31	93-94	3.50
64-65	17.56	94-95	3.33
65-66	16.83	95-96	3.18
66-67	16.11	96-97	3.06
67-68	15.40	97-98	2.95
68-69	14.70	98-99	2.85
69-70	14.02	99-100	2.77
70-71	13.35	100-101	2.69
71-72	12.70	101-102	2.62
72-73	12.06	102-103	2.56
73-74	11.44	103-104	2.51
74-75	10.84	104-105	2.46
75-76	10.26	105-106	2.42
76-77	9.70	106-107	2.38
77-78	9.16	107-108	2.34
78-79	8.64	108-109	2.30
79-80	8.15	109-110	2.27
80-81	7.68		
81-82	7.22		
82-83	6.80		
83-84	6.39		
84-85	6.00		

Insurance companies use estimates of average remaining lifetimes to determine the premium to be charged corresponding to the age of a new purchaser of life insurance. Actually, companies base their rates on much more detailed information about the individual. Not only would age and sex be relevant, but also other variables including occupation, health history, marital status, and so on. Other groups interested in average remaining life include pension planners, governmental planners, industrial market specialists, economists, and a variety of other analysts.

How do statisticians derive estimates of the average remaining lifetime? To illustrate a standard method, we use the following data of Bjerkedal (1960). These data have also been analyzed by Hall and Wellner (1981). Bjerkedal studied the lifelengths of guinea pigs after injection with different amounts of tubercle bacilli. Guinea pigs are known to have a high susceptibility to human tuberculosis, and that is one reason for choosing this species. Table 2 gives estimated average remaining lifetimes for study "M" in which animals in a single cage are under the same regimen. The regimen number is the common log of the number of bacillary units in 0.5 ml of the challenge solution. Here we focus on regimen 5.5.

Table 2. Estimated average remaining life in days at the unique times of death for the 72 guinea pigs under regimen 5.5.

<u>Number of Deaths</u>	<u>Time of Death</u>	<u>Estimated Average Remaining Life</u>
0	0	141.85*
1	43	100.24
1	45	99.64
1	53	92.97
2	56	92.66
1	57	93.05
1	58	93.46
1	66	86.80
1	67	87.16
1	73	82.47

Table 2 (continued).

<u>Number of Deaths</u>	<u>Time of Death</u>	<u>Estimated Average Remaining Life</u>
1	74	82.80
1	79	79.10
2	80	80.79
3	81	84.15
1	82	84.69
2	83	86.90
1	84	87.59
1	88	85.26
1	89	85.98
2	91	87.55
2	92	90.40
1	97	87.34
2	99	89.40
2	100	92.83
1	101	94.18
3	102	100.94
1	103	102.80
1	104	104.79
1	107	104.88
1	108	107.13
1	109	109.55
1	113	109.07
1	114	111.79
1	118	111.64
1	121	112.67
1	123	114.92
1	126	116.40
1	128	119.17
1	137	114.96
1	138	119.14
1	139	123.76
1	144	124.70
1	145	130.21
1	147	135.33
1	156	133.76
1	162	135.75
1	174	132.00
1	178	137.14
1	179	146.62
1	184	153.42
1	191	159.73
1	198	168.00
1	211	172.22
1	214	190.38
1	243	184.43

Table 2. (continued)

<u>Number of Deaths</u>	<u>Time of Death</u>	<u>Estimated Average Remaining Life</u>
1	249	208.17
1	329	153.80
1	380	128.50
1	403	140.67
1	511	49.00
1	522	76.00
1	598	0.00

*Although 0 is not a time of death, we have included the estimated mean residual life at time 0.

To illustrate how the values in column 3 of Table 2 are computed, consider the specific time of death 403. The remaining lifetimes of the three guinea pigs still alive right after the death at time 403 are $511 - 403 = 108$, $522 - 403 = 119$ and $598 - 403 = 195$. Thus the estimated average remaining life of a hypothetical guinea pig who has survived regimen 5.5 for 403 days is

$$\frac{108 + 119 + 195}{3} = 140.67.$$

Similarly the estimated remaining life of a hypothetical guinea pig who has survived regimen 5.5 for 511 days is

$$\frac{(522 - 511) + (598 - 511)}{2} = 49.$$

We can use the sample of 72 lifetimes to estimate average remaining life at any time t , not just at those times corresponding to times of death in the sample. For example, at time 440, the estimated average remaining life would be

$$\frac{(511 - 440) + (522 - 440) + (598 - 440)}{3} = 103.67.$$

For all times greater than 598 (the time of death of the longest surviving guinea pig in the sample), the estimated average remaining life is taken to be 0.

Note that the estimated values of column 3 tend to decrease up to time 90 days, and then tend to increase up to about time 249, and then begin to decrease again. Even before conducting this experiment, it is not unreasonable to conjecture that the injection of tubercle bacilli would cause an adverse stage of aging where average remaining life decreases and then after the hardier guinea pigs have survived this adverse stage, the guinea pigs' natural systems recoup to yield a beneficial stage where (for a while) average remaining life increases.

Keep in mind that the Table 2, column 3 values computed from the sample of 72 guinea pigs are *estimates* of the *true average remaining lifetimes* of a hypothetical population of guinea pigs that could be subjected to regimen 5.5. This raises two related questions.

- A. How can the sample be used to "test" whether there is a trend change in the true average remaining life?
- B. Suppose it is known *a priori* that there is a change in trend (either a decreasing trend changing to an increasing trend or an increasing trend changing to a decreasing trend) in the true average remaining life. How can that prior information be utilized to yield better estimates of the average remaining lifetimes?

There are many situations where the type of prior information mentioned in B above would be available. They include:

- (i) *Length of time employees stay with certain companies:* An employee with a company for four years has more time and career invested in the company than an employee of only two months. The average remaining life of a four-year employee is likely to be longer than the average remaining life of a two-month employee. After this initial increasing trend (this is called "inertia" by social scientists), the processes of aging and retirement yield a decreasing period.

- (ii) *Length of wars*: In the initial stages as negotiations deteriorate and conflict escalates, we expect the war to be longer as time goes by. Eventually, a decreasing trend will be applicable as resources and lives are expended.
- (iii) *Life of certain television shows*: Many shows will initially be cancelled. The longer a show lasts the longer we expect it to continue. After this increasing period of average remaining life, it is reasonable to postulate a decreasing trend for the waning period.
- (iv) *Life lengths of humans*: High infant mortality explains the early interval of increasing average remaining life. (Note the first four rows of Table 1.) Deterioration and aging explain the later decreasing stage.

Guess, Hollander, and Proschan (1986) provide methods pertaining to Question A. Question B is harder, still open, and leads to our first easily-stated but hard problem, namely,

- I. Determine "optimal" estimates of true average remaining lifetimes when it is known that these lifetimes exhibit a reversal of trend.

2. SIMILARITY OF SPECIES' PRESENCE ON ISLANDS

Table 3 contains presence-absence data of the six species of ground finches in genus Geospiza on 23 Galápagos islands. The data are taken from Meeter (1986) who cites D. Simberloff (personal communication). Simberloff compiled the data from Abbott, Abbott, and Grant (1977), Grant (personal communication) and Harris (1973).

Table 3. Ground finches, genus Geospiza, present on 23 Galápagos islands

difficilis	1	1	1	0	1	0	0	0	0	0	0	0	0	0	6					
conirostris	1	0	0	0	1	0	0	0	1	0	0	0	0	0	4					
scandens	0	0	1	1	0	1	1	1	1	1	1	0	1	0	14					
foliginosa	0	0	1	1	0	1	1	1	1	1	1	1	1	1	19					
fortis	0	0	1	1	0	1	1	1	1	1	1	0	1	0	16					
magnirostris	0	1	1	1	1	1	0	1	1	0	0	0	0	0	8					
Column Totals	2	2	5	4	3	5	3	1	2	4	3	5	3	3	3	2	2	3	1	
Darwin																				
Wolf																				
Pinta																				
Marchena																				
Genovesa																				
San Salvador																				
Rabida																				
Bartolomé																				
Daphne Major																				
Seymour																				
Baltra																				
Santa Cruz																				
Plaza Sur																				
Santa Fe																				
San Cristóbal																				
Gardner nr. E.																				
Española																				
Santa María																				
Isabela																				
Fernandina																				
Los Hermanos																				
Pinzon																				
La Tortuga																				
Row Totals																				

In Table 3, a "1" entry denotes presence, a "0" entry denotes absence. For example, the "1" in the first row, second column of the table means *difficilis* is present on Wolf island whereas the "0" in the second row, second column signifies *conirostris* is not present on Wolf. One method of assessing the similarity of two species is to count the number of "1 - 1" matches between their respective rows in Table 3, and then determine if this number is significantly higher (or lower) than would be dictated by chance. As we shall soon see, the calculation of the relevant chance depends on the probability model used. A significantly low number of "1 - 1" matches between two species could indicate species competition, whereas a significantly high number could indicate that the colonization patterns of these two species are related (Simberloff and Connor, 1979).

REMARK 1: Simberloff and Connor (1981) and others also point out, however, that low numbers of "1 - 1" matches may be a consequence of different species having different habitats — rather than a consequence of direct competition—and high numbers could signify the presence of habitats favoring both species — rather than mutualism.

Consider the comparison between *difficilis* and *foliginosa*. There are three "1 - 1" matches in this comparison as Table 3 shows that both species are found on Pinta, San Salvador, and Santa Cruz islands. Is this value of three matches significantly small? That is, what is the probability of having three or less matches if the entries in Table 3 are filled at random? One must be precise about the term "at random" because the probability in question depends on the model chosen.

Meeter (1986) shows how to calculate probabilities for the number of matches in various models including the model where the row totals corresponding to the two species under consideration are fixed. Fixing the row totals

corresponds to conditioning on the relative rarity of the species. In other words, after allowing for the fact that some species are more "successful" than others, is there any evidence that they are "avoiding coexistence"? Note from Table 3 that the row totals for *difficilis* and *foliginosa* are 6 and 19, respectively. Meeter points out that row matches are more likely in rows having higher row totals and hence probabilistic models should reflect this. One model he considers which does reflect this property is the model in which the rows totals are fixed at 6 and 19. In this model, methods of Mosimann (1968) can be used to obtain

Probability of exactly 3 matches between *difficilis* and *foliginosa*

$$= \frac{\binom{6}{3} \times \binom{17}{16}}{\binom{23}{19}} = \frac{20 \times 17}{8,855} = .0384 . \quad (1)$$

(The symbol $\binom{6}{3}$, for example, denotes the number of different choices of 3 distinct objects chosen from among 6 distinct objects. It is one of a general class of such symbols more formally called binomial coefficients.) The probability is obtained as follows. The elements in the row corresponding to *difficilis* are regarded as fixed. The number of ways to get exactly three matches is equal to the "number of ways $\left[\binom{6}{3} \right]$ of placing 3 ones in the *foliginosa* row among the 6 columns containing ones in the *difficilis* row" multiplied by the "number of ways $\left[\binom{17}{16} \right]$ of placing the $19 - 3 = 16$ remaining ones in the remaining $23 - 6 = 17$ columns." To find the desired probability we divide by $\binom{23}{19}$, the total number of ways of putting 19 ones in the *foliginosa* row among 23 available positions. Readers familiar with probability calculations will recognize Display (1) as a probability obtained from the hypergeometric distribution. Similarly,

$$= \frac{\binom{6}{2} \times \binom{17}{17}}{\binom{23}{19}} = \frac{15 \times 1}{8,855} = .0016. \quad (2)$$

The reader is asked to convince him/herself that, given 19 ones in the foliginosa row and 6 ones in the difficilis row, the probability of exactly 1 match in the 23 columns is 0 and the probability of exactly no matches is also 0. The probability of having three or less matches is then $.0384 + .0016 = .040$, obtained by adding the values given by Displays (1) and (2).

Meeter considers other models. He notes that the probability model might also be selected to reflect the fact that matches between two rows are more likely in columns with high totals. Compare San Salvador with La Tortuga! Thus Meeter considers the model where column totals are fixed. Fixing the column totals can be viewed as conditioning on the species richness of islands. He shows that the chance of three or less matches between difficilis and foliginosa when the column totals are fixed (at those column totals given in Table 3) is .156. Both the "row totals fixed" and the "column totals fixed" models seem to support evidence of competition between difficilis and foliginosa, with the "row totals fixed" model showing stronger support. (But recall Remark 1!)

Another model considered by Meeter is the one in which all column totals and two row totals are fixed. For this model, Meeter calculates the probability of three or less matches between difficilis and foliginosa to be .042. Meeter also mentions the model in which all column totals and all row totals are fixed. This model is being studied by A. Zaman and D. Simberloff, and is the basis for our second easily-stated but hard problem.

- II. Determine the probability distribution of the number of "1 - 1" matches between two rows when all column totals and row totals are fixed.

Except for small tables (i.e., when the number of rows and columns is small), Problem II is unsolved.

For more information about the pairwise comparisons of species in Table 3, see Meeter (1986). Meeter's results are described in terms of an ecological problem, but, as he points out, his results are applicable in a variety of situations in which "individuals" are scored as to the presence or absence of certain "characteristics", and it is of interest to assess the degree of similarity between pairs of individuals.

3. THE CHANCE OF COVERING A CIRCLE BY RANDOMLY PLACED ARCS

We begin with a simplified version of the problem. Consider a circle whose circumference is of length 1. Consider also four arcs, with each arc of length .35. The four arcs are thrown *independently and uniformly on the circumference*. The preceding italicized phrase can be interpreted as follows. Imagine a dial which is flicked and comes to rest at some point on the circumference. Put the midpoint of the first arc at the point where the dial stops. Now spin the dial again. Use enough energy so that it is reasonable to assume that the dial's starting point does not affect its ending point, and place the midpoint of the second arc at the position on the circumference where the dial comes to rest. Repeat this process two more times, thus placing all four arcs on the circumference. What is the chance that the circumference is completely covered by these four arcs?

The previous question has been answered for the case of any number of equal-length arcs by Stevens (1939) who gave an explicit formula for calculating the desired chance. For simplicity here we took the number of arcs to be four and the common length to be .35. In this case, Stevens' formula shows the chance is .0635. The interpretation of this probability is as follows. Suppose the process consisting of four spins—with an arc placement after each spin—was

repeated a large number of times. In about 6 percent of the replications of the process, the circumference would be completely covered and in about 94 percent the circumference would not be covered. The non-statistician reader will naturally ask: What do you mean by "a large number of times?" Do you mean 100 replications, or 1,000 or 10,000, or just how many replications? Probability theory provides a precise answer. Here we will crudely state that roughly speaking, the greater number of replications, the more likely it is that the percentage of times the circumference will be completely covered by the four arcs is close to 6.35%.

We started with the case where the arcs had the same length. An explicit solution for the case of general number of arcs and general arc lengths is at present unavailable. Huffer and Shepp (1987) [hereafter denoted as HS(1987)] state "It seems hopeless to give a simple formula for the case of general arc lengths." Thus our third easily-stated but hard statistical problem is:

III. For the case of general number of arcs and general arc lengths, provide a simple formula for the probability of complete coverage.

Although HS(1987) do not solve Problem III, they present results which yield inequalities concerning the probability of complete coverage. To give an indication of the nature of the HS(1987) results, we return for simplicity to the case of four arcs. We initially discussed the case where the four arcs were each of length .35; let us call that configuration 1. To illustrate the Huffer-Shepp results we will introduce two other arc configurations, where the arc lengths are not equal, but where the total length of the four arcs is 1.4 as it is for the equal-lengths configuration 1.

The concept of majorization is critical to the HS(1987) development. For two configurations of four arcs each, configuration 2 is said to majorize configuration 1 if the following four conditions hold:

- (i) The length of the longest arc in configuration 2 is greater than or equal to the length of the longest arc in configuration 1.
- (ii) The sum of the lengths of the longest and second-longest arcs in configuration 2 is greater than or equal to the sum of the lengths of the longest and second-longest arcs in configuration 1.
- (iii) The sum of the lengths of the three longest arcs in configuration 2 is greater than or equal to the sum of the lengths of the three longest arcs in configuration 1.
- (iv) The sum of the lengths of the four arcs in configuration 2 is equal to the sum of the lengths of the four arcs in configuration 1.

Now consider configurations 1, 2, and 3 below.

Configuration 1: .35, .35, .35, .35

Configuration 2: .40, .35, .35, .30

Configuration 3: .40, .38, .32, .30

From the definition of majorization, one can check that (a) configuration 2 majorizes configuration 1, and also (b) configuration 3 majorizes configuration 2.

HS(1987) show that the probability of complete coverage preserves the "partial ordering of majorization." This means that if a configuration of arcs, say configuration A, majorizes another configuration of arcs, say configuration B, then the probability of complete coverage using the configuration-A arcs is greater than or equal to the probability of complete coverage using the configuration-B arcs. For our little example with three configurations, the results of HS(1987) show that the probability of complete coverage using configuration 3 is greater than or equal to the probability of complete

coverage using configuration 2 which in turn is greater than or equal to the probability of complete coverage using configuration 1. Recall the latter probability is known to be .0635.

More generally, the results of HS(1987) provide many inequalities. For configurations ordered by majorization, HS(1987) yields a comparison of the respective coverage probabilities without necessitating actual calculation of the probabilities. Furthermore, since every configuration of n arcs having a specified total arc length L , say, majorizes the configuration of n arcs with equal arc lengths having total length L , HS(1987) yields, for any configuration, a lower bound to the probability of complete coverage. The lower bound is obtained from Stevens' formula.

Acknowledgments

I am grateful to Fred Huffer, Duane Meeter, Edsel Peña and Arif Zaman for helpful discussions.

REFERENCES

- Abbott, I., Abbott, L. K., and Grant, P. R. (1977). Comparative ecology of Galapagos ground finches (Geospiza Gould): Evaluation of the importance of floristic diversity and interspecific competition. Ecological Monographs 47, 151-184.
- Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. Amer. J. Hygiene 72, 130-148.
- Guess, F., Hollander, M. and Proschan, F. (1986). Testing exponentiality versus a trend change in mean residual life. Ann. Statist. 14. (To appear.)
- Hall, W. J. and Wellner, J. A. (1981). Mean residual life. Statistics and Related Topics (Csörgö, M., Dawson, D. A., Rao, J. N. K. and Saleh, A. K. Md.E., eds.). North-Holland, Amsterdam, 169-184.
- Harris, M. P. (1973). The Galapagos avifauna. Condor 75, 265-278.
- Huffer, F. and Shepp, L. A. (1987). On the probability of covering the circle by random arcs. App. Prob. (To appear.)
- Meeter, D. (1986). Row similarity in binary-valued matrices. Florida State University, Department of Statistics Technical Report No. M723.
- Moisimann, J. E. (1968). Elementary Probability for the Biological Sciences, New York: Appleton-Century-Crofts.
- Simberloff, D. S. and Connor, E. F. (1979). Q-mode and R-mode analyses of biogeographic distributions: Null hypotheses based on random colonization, in Contemporary Quantitative Ecology and Related Ecometrics, G.P. Patil and M. L. Rozenzweig (eds.), Fairland, MD: International Cooperative Publishing House, 123-138.
- Simberloff, D. S. and Connor, E. F. (1981). Missing species combinations, American Naturalist 118, 215-239.
- Stevens, W. L. (1939). Solution to a geometrical problem in probability. Ann. Eugenics 9, 315-320.
- U.S. Life Tables: 1969-71, Vol. 1, No. 1, U.S. National Center for Health Statistics.