TABLES FOR THE DISTRIBUTION OF THE NUMBER OF EXCEEDANCES[1]

By Benjamin Epstein

Wayne University

1. Introduction and Summary. Consider a random sample of size n taken from a continuous distribution $f(x)$. Let another random sample, independent of the first sample and also of size n, be drawn from the same population. Let U_r^n be the random variable associated with the number of values in the second sample which exceed the rth smallest value in the first sample. Similarly let V_s^n be the random variable associated with the number of values in the second sample which exceed the sth largest value in the first sample. Due to the fact that the rth smallest value in a sample of size n is at the same time the sth largest value in the sample with $s = n - r + 1$, it follows that

$$\Pr(U_r^n = x) \equiv \Pr(V_s^n = x), \tag{1}$$

$$s = n - r + 1; \qquad r = 1, 2, \cdots, n; \qquad x = 0, 1, 2, \cdots, n.$$

The probability distribution of U_r^n (and hence of V_s^n) is given by:

$$\Pr(U_r^n = x) = \binom{n-x+r-1}{r-1}\binom{n-r+x}{x}\Big/\binom{2n}{n} = \tfrac{1}{2}P_{n-x+r-1,r-1}\, P_{n-r+x,x}/P_{2n,n}\,, \tag{2}$$

$$x = 0, 1, 2, \cdots, n.$$

Formula (2) can be proved by combinatorial methods; details are omitted. An alternative formula, derived in another way [3], is

$$\Pr(U_r^n = x) = \tfrac{1}{2}\binom{n-1}{r-1}\binom{n}{x}\Big/\binom{2n-1}{n-r+x} = \tfrac{1}{2}P_{n-1,r-1}P_{n,x}/P_{2n-1,n-r+x}\,. \tag{2a}$$

In formulae (2) and (2a), $P_{n,x} = (\frac{1}{2})^n\binom{n}{x}$. Formulae in terms of $P_{n,x}$ are particularly convenient for hand computation, since one can use the extensive tables of the binomial probability distribution published by the National Bureau of Standards.

If the values of $\Pr(U_r^n \leqq x)$, for $x = 0, 1, 2, \cdots, n - 1, r = 1, 2, \cdots, n$ are written (for fixed n) in matrix form, one notes certain useful symmetries, which can be expressed by the identities

$$\Pr(U_r^n \leqq x) = \Pr(U_{x+1}^n \leqq r - 1), \tag{3}$$

$$\Pr(U_r^n \leqq x) + \Pr(U_{n-r+1}^n \leqq n - x - 1) = 1. \tag{4}$$

If one takes $x = n - r$ in (4) and uses the relation (3), it is readily verified that

$$\Pr(U_r^n \leqq n - r) = \tfrac{1}{2}. \tag{5}$$

* Received 7/20/53, revised 9/18/53.

[1] Work sponsored by the Office of Naval Research.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

TABLE 1

$\Pr(U_r^n \leq x)$

n	r	$x = 0$	1	2	3	4	5	6	7	8	9	10	11	12	13
2	1	.1667													
3	1	.0500	.2000												
4	1	.0143	.0714	.2143											
	2		.2429												
5	1	$.0^2397$	.0238	.0833	.2222										
	2		.1032	.2619											
6	1	$.0^2108$	$.0^2758$	.0303	.0909	.2273									
	2		.0400	.1212	.2727										
	3			.2835											
7	1	$.0^3291$	$.0^2233$	.0105	.0350	.0962	.2308								
	2		.0146	.0513	.1329	.2797									
	3			.1431	.2960										
8	1	$.0^4777$	$.0^3699$	$.0^2350$	.0128	.0385	.1000	.2333							
	2		$.0^2505$	.0203	.0594	.1410	.2846								
	3			.0660	.1573	.3042									
	4				.3096										
9	1	$.0^4206$	$.0^3206$	$.0^2113$	$.0^2452$	.0147	.0412	.1029	.2353						
	2		$.0^2169$	$.0^2761$	.0249	.0656	.1471	.2882							
	3			.0283	.0767	.1674	.3100								
	4				.1735	.3186									
10	1	$.0^5541$	$.0^4595$	$.0^3357$	$.0^2155$	$.0^2542$	.0163	.0433	.1053	.2368					
	2		$.0^3547$	$.0^2274$	$.0^2988$	.0286	.0704	.1517	.2910						
	3			.0115	.0349	.0849	.1749	.3142							
	4				.0894	.1849	.3250								
	5					.3281									

TABLE 1—*Concluded*

11	1	$.0^{5}142$	$.0^{4}170$	$.0^{3}111$	$.0^{3}516$	$.0^{2}193$	$.0^{2}619$	.0175	.0451	.1071	.2381				
	2		$.0^{3}173$	$.0^{3}953$	$.0^{2}376$	.0119	.0317	.0743	.1554	.2932					
	3			$.0^{2}446$	.0150	.0402	.0913	.1807	.3176						
	4				.0431	.0992	.1935	.3297							
	5					.1974	.3350								
12	1	$.0^{6}370$	$.0^{5}481$	$.0^{4}337$	$.0^{3}168$	$.0^{3}673$	$.0^{2}229$	$.0^{2}686$	.0186	.0466	.1087	.2391			
	2		$.0^{4}536$	$.0^{3}322$	$.0^{2}138$	$.0^{2}471$	.0136	.0343	.0775	.1584	.2950				
	3			$.0^{2}166$	$.0^{2}614$	.0180	.0447	.0965	.1854	.3202					
	4				.0196	.0498	.1069	.2002	.3334						
	5					.1102	.2068	.3401							
	6						.3421								
13	1	$.0^{7}961$	$.0^{5}135$	$.0^{4}101$	$.0^{4}538$	$.0^{3}229$	$.0^{3}824$	$.0^{2}261$	$.0^{2}745$	.0196	.0478	.1100	.2400		
	2		$.0^{4}163$	$.0^{3}106$	$.0^{3}491$	$.0^{2}180$	$.0^{2}558$	.0151	.0365	.0801	.1609	.2965			
	3			$.0^{3}601$	$.0^{2}242$	$.0^{2}771$	.0207	.0484	.1008	.1891	.3224				
	4				$.0^{2}847$	.0236	.0554	.1131	.2055	.3364					
	5					.0576	.1189	.2142	.3441						
	6						.2169	.3475							
14	1	$.0^{7}249$	$.0^{6}374$	$.0^{5}299$	$.0^{4}170$	$.0^{4}763$	$.0^{3}290$	$.0^{3}966$	$.0^{2}290$	$.0^{2}797$	.0204	.0489	.1111	.2407	
	2		$.0^{5}491$	$.0^{4}344$	$.0^{3}171$	$.0^{3}670$	$.0^{2}221$	$.0^{2}638$	.0164	.0384	.0824	.1630	.2978		
	3			$.0^{3}211$	$.0^{3}919$	$.0^{2}316$	$.0^{2}915$	.0230	.0516	.1043	.1923	.3242			
	4				$.0^{2}351$	.0107	.0271	.0601	.1182	.2099	.3388				
	5					.0285	.0642	.1259	.2200	.3473					
	6						.1284	.2247	.3518						
	7							.3532							
15	1	$.0^{8}645$	$.0^{6}103$	$.0^{6}887$	$.0^{5}526$	$.0^{4}250$	$.0^{3}100$	$.0^{3}350$	$.0^{2}110$	$.0^{2}316$	$.0^{2}843$	.0211	.0498	.1121	.2414
	2		$.0^{5}146$	$.0^{4}109$	$.0^{4}579$	$.0^{3}242$	$.0^{3}850$	$.0^{2}260$	$.0^{2}710$	.0176	.0400	.0843	.1648	.2989	
	3			$.0^{4}725$	$.0^{3}339$	$.0^{2}125$	$.0^{2}389$	.0105	.0251	.0543	.1074	.1949	.3257		
	4				$.0^{2}141$	$.0^{2}461$	.0127	.0302	.0641	.1225	.2135	.3408			
	5					.0134	.0328	.0697	.1318	.2249	.3499				
	6						.0716	.1362	.2311	.3552					
	7							.2331	.3576						

		$x = 0$	1	2	3	4	5	6	7	8	9
20	1	$.0^{11}725$	$.0^{9}152$	$.0^{8}168$	$.0^{7}128$	$.0^{7}771$	$.0^{6}385$	$.0^{5}167$	$.0^{5}644$	$.0^{4}225$	$.0^{4}727$
	2		$.0^{8}291$	$.0^{7}291$	$.0^{6}203$	$.0^{5}110$	$.0^{5}501$	$.0^{4}197$	$.0^{4}685$	$.0^{3}216$	$.0^{3}624$
	3			$.0^{6}265$	$.0^{5}168$	$.0^{5}832$	$.0^{4}343$	$.0^{3}122$	$.0^{3}386$	$.0^{2}110$	$.0^{2}287$
	4				$.0^{5}969$	$.0^{4}438$	$.0^{3}164$	$.0^{3}532$	$.0^{2}153$	$.0^{2}396$	$.0^{2}935$
	5					$.0^{3}180$	$.0^{3}616$	$.0^{2}182$	$.0^{2}477$	.0112	.0242
	6						$.0^{2}192$	$.0^{2}519$	.0124	.0268	.0527
	7							.0128	.0281	.0555	.1001
	8								.0564	.1025	.1703
	9									.1715	.2636
	10										.3762

		$x = 10$	11	12	13	14	15	16	17	18	
20	1	$.0^{3}218$	$.0^{3}614$	$.0^{2}164$	$.0^{2}416$	.0101	.0236	.0530	.1154	.2436	
	2	$.0^{2}167$	$.0^{2}418$	$.0^{2}983$	.0218	.0457	.0909	.1708	.3025		
	3	$.0^{2}691$	.0155	.0324	.0637	.1176	.2037	.3307			
	4	.0204	.0412	.0776	.1367	.2253	.3474				
	5	.0479	.0880	.1504	.2401	.3582					
	6	.0954	.1601	.2503	.3655						
	7	.1666	.2572	.3705							
	8	.2616	.3738								
	9	.3756									
	10										

TABLE 2

Values of $\Pr(U_r^5 \leqq x)$

r	$x = 0$	1	2	3	4
1	.0²397	.0238	.0833	.2222	.5000
2	.0238	.1032	.2619	.5000	.7778
3	.0833	.2619	.5000	.7381	.9167
4	.2222	.5000	.7381	.8968	.9762
5	.5000	.7778	.9167	.9762	.9²603

Proofs[2] of (3), (4), and (5) can be obtained by using the results of pages 257–258 of [3]. Because of these symmetries, the complete matrix (for any fixed r) can be constructed if one knows only the quantities, $\Pr(U_r^n \leqq x)$, $r = 1(1)[n/2]$, $x = r-1, r, r+1, \cdots, n-r-1$. In Table 1 these values are given[3] for $n = 2(1)15(5)20$. To see how the complete matrix is obtained from Table 1, it is interesting to verify, using (3), (4), and (5), that the complete matrix, in the special case $n = 5$, is given by Table 2.

A somewhat different, but related, exceedance problem is to take two random samples of size n from a continuous distribution $f(x)$. Let us for convenience attach the letter x to one of the samples and the letter y to the other sample. Further let $x_{r,n}$ and $y_{r,n}$ be respectively the rth smallest observations in each of the samples. Let us define $z_{r,n} = \max(x_{r,n}, y_{r,n})$. If $z_{r,n} = x_{r,n}$, count the number of y's which are $\geqq x_{r,n}$; if $z_{r,n} = y_{r,n}$, count the number of x's which are $\geqq y_{r,n}$. Denoting the number of exceedances as W_r^n, it is readily seen from (1) that the probability distribution of W_r^n is given by

$$(6)\quad \Pr(W_r^n = x) = 2\binom{n-x+r-1}{r-1}\binom{n-r+x}{x} \Big/ \binom{2n}{n}, \qquad x = 0, 1, 2, \cdots, n-r.$$

It is evident from the definition that,

$$(7)\quad \Pr(W_r^n \leqq x) = 1, \qquad x \geqq n-r.$$

Clearly one can find the values of $\Pr(W_r^n \leqq x)$ by using Table 1. Thus, for example, in the special case $n = 5$ one obtains Table 3.

2. Applications of exceedance theory. There are three principal uses of exceedance theory. These are:

(a) *Floods and droughts.* This theory was used by H. A. Thomas, Jr. [6] in making predictions about the recurrences of floods and droughts in the future on the basis of what is known from past data. In recent papers by Chow [1], [2], the interested reader will find further work in this direction.

---

[2] We wish to acknowledge with thanks a communication from Dr. E. J. Gumbel on this point.

[3] In Wayne University Technical Report No. 6 (July 1953) values were given for $n = 2(1)20(5)50$. We have also considered the practically important case where the two samples may be of unequal size. Tables for selected pairs of unequal values of the sample size will be available in the near future.

TABLE 3
$\Pr(W_r^5 \leqq x)$

r	$x = 0$	1	2	3	4
1	$.0^2794$	.0476	.1667	.4444	1.0000
2	.0476	.2064	.5238	1.0000	
3	.1667	.5238	1.0000		
4	.4444	1.0000			
5	1.0000				

(b) *Non-parametric tests for slippage.* The functions U_r^n, V_s^n, and W_r^n can be used to give two-sample nonparametric tests for slippage of the mean. There are close connections between the results in this paper and recent tests for slippage by Mosteller and Tukey [4] and [5].

(c) *Life testing.* It is a characteristic feature of life tests that data become available in order of size. Thus it becomes very natural to apply exceedance theory, which is based purely on order statistics. By so doing it is possible in many cases to shorten both the average time and average number of items destroyed in order to reach a decision as to whether or not the items in one population are in some sense superior to the items in another population.

3. Numerical examples.

EXAMPLE 1. What is the probability that the third largest flood during the past 20 years will be exceeded at least once during the next 20 years? *Answer.* The probability is

$$p = 1 - \Pr(V_3^{20} = 0) = 1 - \Pr(U_{18}^{20} = 0) = 1 - .1154 = .8846.$$

EXAMPLE 2. During a period of 20 years the lowest observed annual rainfall in a certain locality was 8.6 inches. What is the probability that in the next 20 years at least two of the years will have rainfall $\leqq$ 8.6 inches? *Answer.* The probability is $p = \Pr(U_1^{20} \leqq 18) = .2436$.

EXAMPLE 3. (one-sided test): We are now interested in making a choice between two lots A and B. In particular we are interested in some characteristic such as life or strength, where data become available in order of magnitude. Let it be known a priori that the probability density function associated with lot B is either the same as that of lot A or is displaced to the left (e.g., is inferior). Put in another way, we are thinking of a case where the only relevant parameter is some measure of slippage. We wish to test the hypothesis H_0 of no displacement against the alternative H_1 that B is displaced to the left of A. The Type I error is taken to be $\leqq .05$. Ten items are drawn from each of the lots and placed on life test. It is decided in advance that a decision will be based on how many failures occur in the sample from B before the second failure occurs in the sample from A. The two samples are put on test simultaneously and give the pattern $bbbabbb\cdots$, where a denotes a failure in the sample drawn from A, b denotes a failure in the sample drawn from B. The experiment is stopped at the seventh failure with rejection of the null hypothesis, because Pr $(U_2^{10} \leqq 4) = .0286 < .05$. If, however, we had obtained a pattern like $babba\cdots$, we would have stopped experimentation after the fifth failure with the acceptance of H_0.

EXAMPLE 4. (two-sided test): Given two lots A and B, we wish to test the null hypothesis that the life distributions of A and B are the same against the alternative that they are different. As in Example 3, let 10 items be drawn at random from each of the two lots and placed on life test. It is decided in advance that our decision will be based on the statistic W_2^{10}. If, for example, the failure pattern observed is $aaaaabaa\cdots$, the experiment will be terminated on the eighth trial with rejection of the null hypothesis (on the .05 level of significance). This is because Pr $(W_2^{10} \leqq 3) = .0198$. On the other hand a pattern like $babba\cdots$ would lead to acceptance of the null hypothesis on the fifth trial.

4. Discussion. Fairly extensive random sampling experiments have shown that the statistics W_1^{10}, W_2^{10}, and W_3^{10} are more effective than the run test, and somewhat less effective than the Wilcoxon rank test, for detecting slippage of the mean in the case where the underlying distributions are normal, all with the same variance. Since the improvement in power obtained by using W_2^{10} or W_3^{10} rather than W_1^{10} is minor in this case, there are sound practical reasons for preferring W_1^{10}. Decisions based on this statistic can be made at a great saving in average time to decision, as well as average number of items destroyed. It should be noted in Example 4 that if decisions were based on W_1^{10}, we would have truncated testing on the fifth trial with the rejection of H_0, since Pr $(W_1^{10} \leqq 5) = .0325$.

A detailed discussion of the points raised in the last paragraph will appear elsewhere.

5. Acknowledgement. I wish to thank John Lay for his work in computing the tables.

REFERENCES

[1] VEN TE CHOW, "A general formula for hydrologic frequency analysis," *Trans. Amer. Geophys. Union*, Vol. 32, No. 2 (1951), pp. 231–282.

[2] VEN TE CHOW, *Frequency Analysis of Hydrologic Data with Special Application to Rainfall Intensities*, University of Illinois Bulletin, Vol. 50, No. 81, July, 1953.

[3] E. J. GUMBEL AND H. VON SCHELLING, "The distribution of the number of exceedances," *Ann. Math. Stat.*, Vol. 21, (1952), pp. 247–262.

[4] F. MOSTELLER, "A k-sample slippage test for an extreme population," *Ann. Math. Stat.*, Vol. 19, (1948), pp. 58–65.

[5] F. MOSTELLER AND J. W. TUKEY, "Significance levels for a k-sample slippage test," *Ann. Math. Stat.*, Vol. 21, (1950), pp. 120–123.

[6] H. A. THOMAS, JR., "Frequency of minor floods," *J. Boston Soc. Civil Engineers*, Vol. 35, (1948), pp. 425–442.