The Statistician's Problem

Dr. Jann-Huei Jinn
Associate Professor
Department of Mathematics and Statistics
Grand Valley State University
Allendale, Michigan 49546

Abstract: We present the solution to a tricky combination problem challenged by the authors of " The Mississippi Problem" published in The American Statistician, February 1998, Vol. 52, No.1. The word **MISSISSIPPI** contains three equal neighbors **SS**, **SS**, and **PP**. The authors considered a random permutation of the eleven letters in the word and found the probability that the permutation has no equal neighbors. In this paper, We solved the probability that a random permutation of the word **STATISTICS** has no equal neighbors is 1/5.

1. The Solution

Choose an initial order of the letters in the word Statistics, for example, SSSTTTIIAC. We solve the problem in three steps: Start with SSSTTT, insert II, then insert A, and finally, insert C.

Let X_1 be the number of equal neighbors in a random permutation of the six letters SSSTTT. To obtain the solution we need the probability function $P(X_1 = k)$.
We first determine the probability function of the total number of runs (see the references.) Consider a random permutation of m 1's and n 0's. Denote by U the number of runs of 1's and by V the number of runs of 0's. We need the probability function of $P(U = r, V = s)$. (Note that when $|r - s| > 1$, $P(U = r, V = s)$=0.) The m 1's can be partitioned into r groups in $\binom{m-1}{r-1}$ ways. Similarly, the n 0's can be partitioned into s groups in $\binom{n-1}{s-1}$ ways. We know that the number of permutations with r 1-runs and s 0-runs is the product of these two binomial coefficients when $|r - s|$=1, and twice that product when $r = s$. Since the total number of permutations is $\binom{m+n}{m}$, we obtain

$$P(U = r, V = s) = \frac{\binom{m-1}{r-1}\binom{n-1}{s-1}}{\binom{m+n}{m}} \qquad (1.1)$$

for r =1,2,…,m and s =1,2,…,n such that $|r - s|$=1. If $r = s$, then

$$P(U = r, V = r) = \frac{2\binom{m-1}{r-1}\binom{n-1}{r-1}}{\binom{m+n}{m}} \qquad (1.2)$$

Step 1. Consider the permutation of SSSTTT, we know that m=n=3 and Table 1 gives the probabilities $P(U = r, V = s)$ for this case.

Table 1: m=n=3 Probability Function P(U=r,V=s)

r\s	1	2	3
1	2/20	2/20	0
2	2/20	8/20	2/20
3	0	2/20	2/20

The probability function $P(U+V=k)$ of the total number of runs is obtained from the previous distribution (1.1) by summation. The result when m=n=3 is given in Table 2.

Table 2: m=n=3 Probability Function P(U+V=k) of the Total Number of Runs

k	2	3	4	5	6
P(U+V=k)	2/20	4/20	8/20	4/20	2/20

Let W be the number of equal neighbors in the random permutation. The relation between runs and equal neighbors is W = m+n-U-V. Hence, when m=n=3, the probability of 4 equal neighbors is equal to the probability 2/20 of two runs, the probability of 3 equal neighbors is 4/20, and so on. Therefore, X_1=6-(U+V). The probability function of X_1 required for the solution of the Statistician problem is given in Table 3.

Table 3: Probability Function of X_1

i	0	1	2	3	4
$P(X_1=i)$	2/20	4/20	8/20	4/20	2/20

Step 2. Insert II

Let X_2 be the number of equal neighbors among the eight letters, SSSTTTII, obtained. Since $P(X_2=j)=\sum_i P(X_2=j|X_1=i)P(X_1=i)$. (1.3)

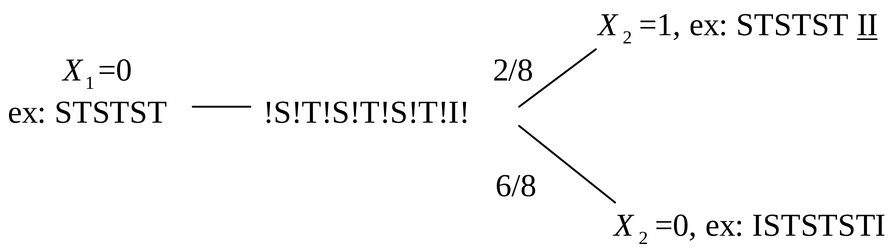
We only need to consider j=0,1,2. Because when j=3, for example, T SS TT S I I, we only have two letters A and C left and it is impossible to get no equal neighbors. The conditional probabilities required are given in Table 4.

Table 4: Some Conditional Probabilities $P(X_2=j|X_1=i)$

i\j	0	1	2
0	42/56	14/56	0
1	12/56	32/56	12/56
2	2/56	20/56	24/56
3	0	6/56	24/56
4	0	0	12/56

We calculated all the probabilities given in Table 4. Suppose that X_1=0, for example, STSTST, we will insert the first I then second I. It does not matter to insert the first I at any place, for example, if we insert the first I to the right end of STSTST, i.e., STSTSTI, then we obtain !S!T!S!T!S!T!I! (where "!" represent the space to insert the second I). It is convenient to use a tree diagram to do illustration; see Figure 1.

Figure 1: Tree diagram for calculation of the conditional probability $p(X_2 = i | X_1 = 0)$

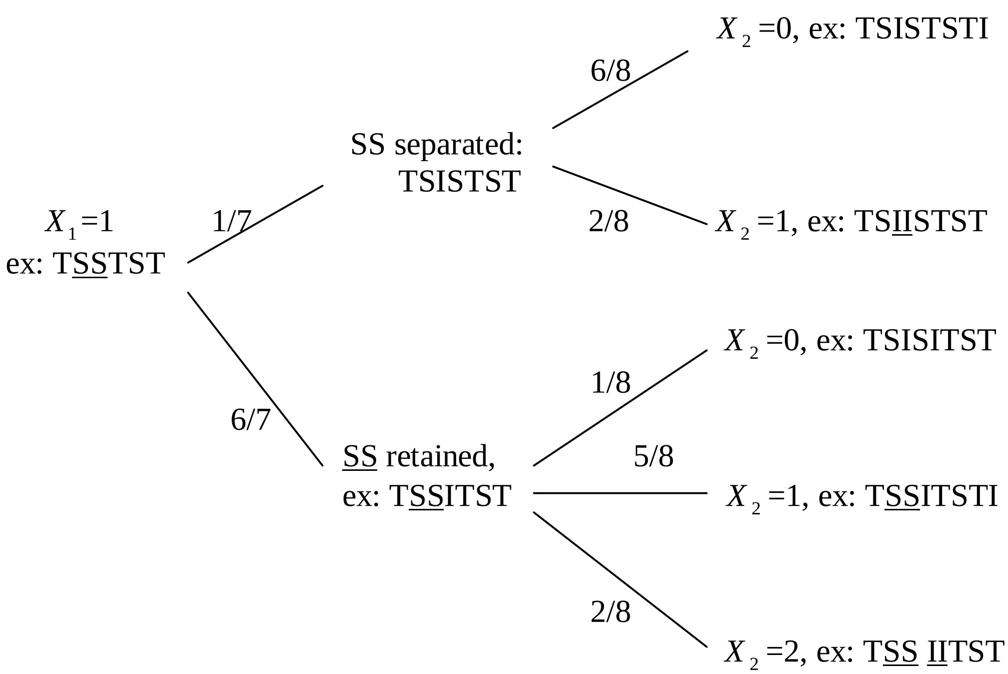


Therefore, $P(X_2 = 1 | X_1 = 0)$ =2/8=14/56 and $P(X_2 = 0 | X_1 = 0)$ =6/8=42/56.

Suppose that X_1=1, an example where this occurs is TSSTST. We obtain X_2=0 by separating the pair SS either with the first or second I inserted, but not both. It is convenient to use a tree diagram; see Figure 2. We obtain

$P(X_2 = 0 | X_1 = 1) = \frac{1}{7} \times \frac{6}{8} + \frac{6}{7} \times \frac{1}{8} = \frac{12}{56}$, $P(X_2 = 1 | X_1 = 1) = \frac{1}{7} \times \frac{2}{8} + \frac{6}{7} \times \frac{5}{8} = \frac{32}{56}$ and $P(X_2 = 2 | X_1 = 1) = \frac{6}{7} \times \frac{2}{8} = \frac{12}{56}$.

Figure 2: Tree diagram for calculation of the conditional probability $P(X_2 = i | X_1 = 1)$

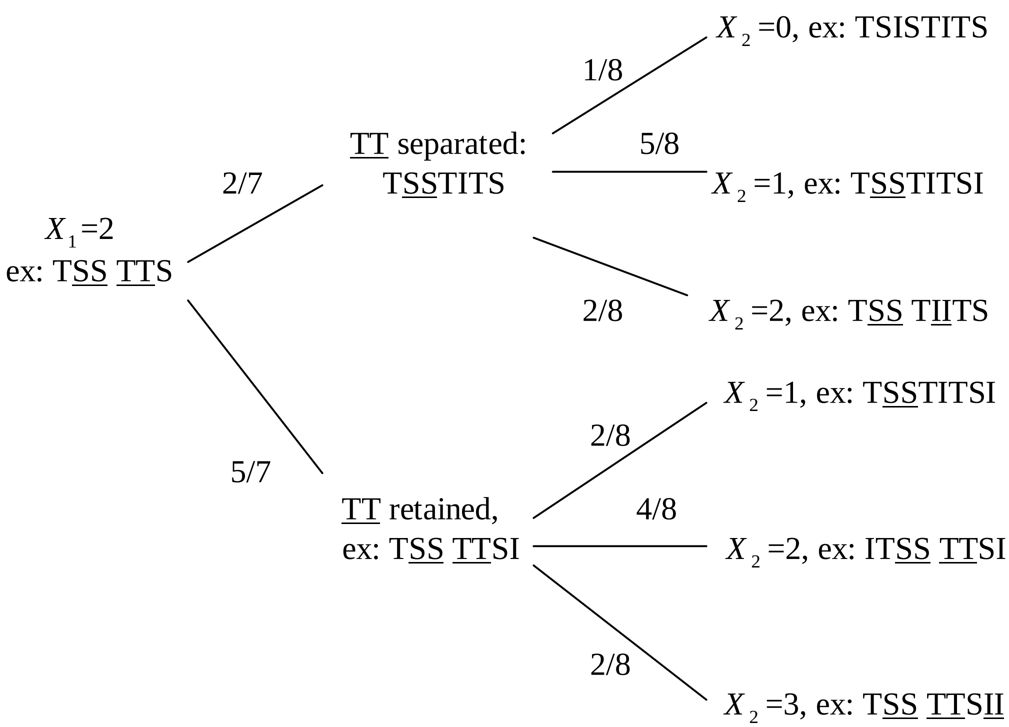


Suppose that X_1=2, an example is TSS TTS. Refer to Figure 3 and we obtain

$P(X_2 = 0 | X_1 = 2) = \frac{2}{7} \times \frac{1}{8} = \frac{2}{56}$, $P(X_2 = 1 | X_1 = 2) = \frac{2}{7} \times \frac{5}{8} + \frac{5}{7} \times \frac{2}{8} = \frac{20}{56}$, and

$P(X_2 = 2 | X_1 = 2) = \frac{2}{7} \times \frac{2}{8} + \frac{5}{7} \times \frac{4}{8} = \frac{24}{56}$.

Figure 3: Tree diagram for calculation of the conditional probability $P(X_2 = i | X_1 = 2)$

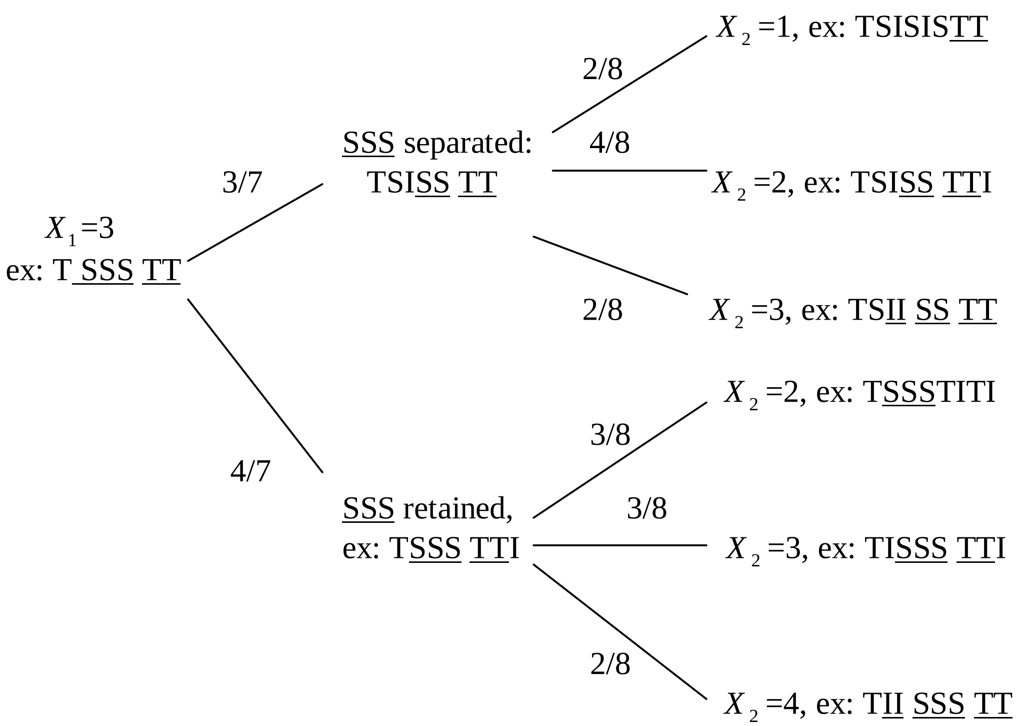


Suppose that X_1=3, an example is T SSS TT. Refer to Figure 4 and we obtain

$P(X_2 = 0 | X_1 = 3) = 0$, $P(X_2 = 1 | X_1 = 3) = \frac{3}{7} \times \frac{2}{8} = \frac{6}{56}$, and

$P(X_2 = 2 | X_1 = 3) = \frac{3}{7} \times \frac{4}{8} + \frac{4}{7} \times \frac{3}{8} = \frac{24}{56}$.

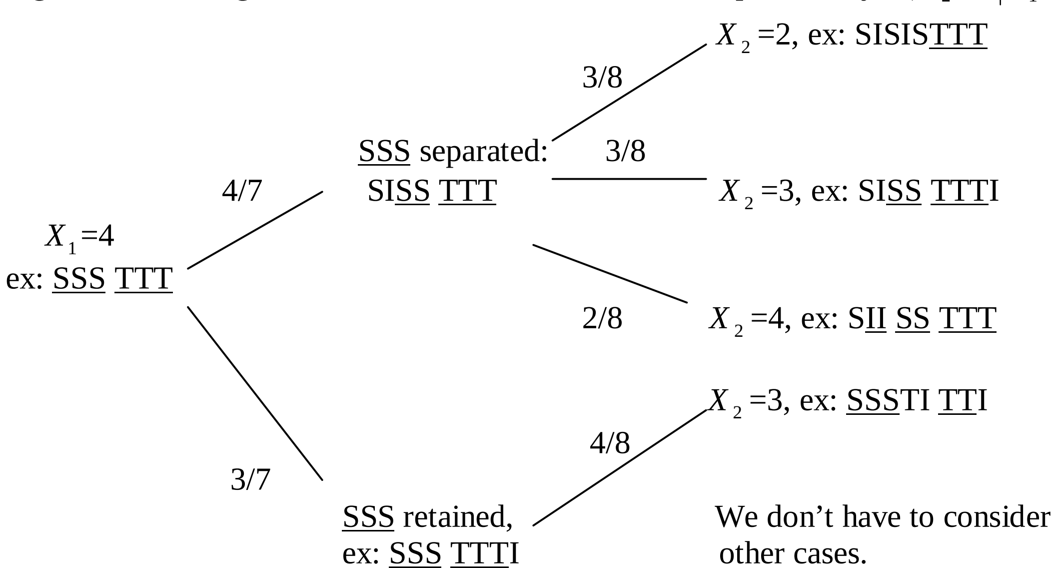
Figure 4: Tree diagram for calculation of the conditional probability $P(X_2 = i | X_1 = 3)$



Suppose that X_1=4, an example is SSS TTT. Refer to Figure 5 and we obtain

$$P(X_2 = 2 | X_1 = 4) = \frac{4}{7} \times \frac{3}{8} = \frac{12}{56}.$$

Figure 5: Tree diagram for calculation of the conditional probability $P(X_2 = i | X_1 = 4)$



Based on (1.3) and combining Table 3 and Table 4, we obtain

$$P(X_2 = 0) = \sum_{i=0}^{4} P(X_2 = 0 | X_1 = i)\ P(X_1 = i) = \frac{42}{56} \times \frac{2}{20} + \frac{12}{56} \times \frac{4}{20} + \frac{2}{56} \times \frac{8}{20} = \frac{148}{1120}.$$

Similarly,

$$P(X_2 = 1) = \frac{14}{56} \times \frac{2}{20} + \frac{32}{56} \times \frac{4}{20} + \frac{20}{56} \times \frac{8}{20} + \frac{6}{56} \times \frac{4}{20} = \frac{340}{1120} \text{ and}$$

$$P(X_2 = 2) = \frac{12}{56} \times \frac{2}{20} + \frac{24}{56} \times \frac{4}{20} + \frac{24}{56} \times \frac{8}{20} + \frac{12}{56} \times \frac{4}{20} = \frac{360}{1120}.$$

Table 5: Part of the Probability Function of X_2

j	0	1	2
$P(X_2 = j)$	148/1120	340/1120	360/1120

Step 3. Insert A then C

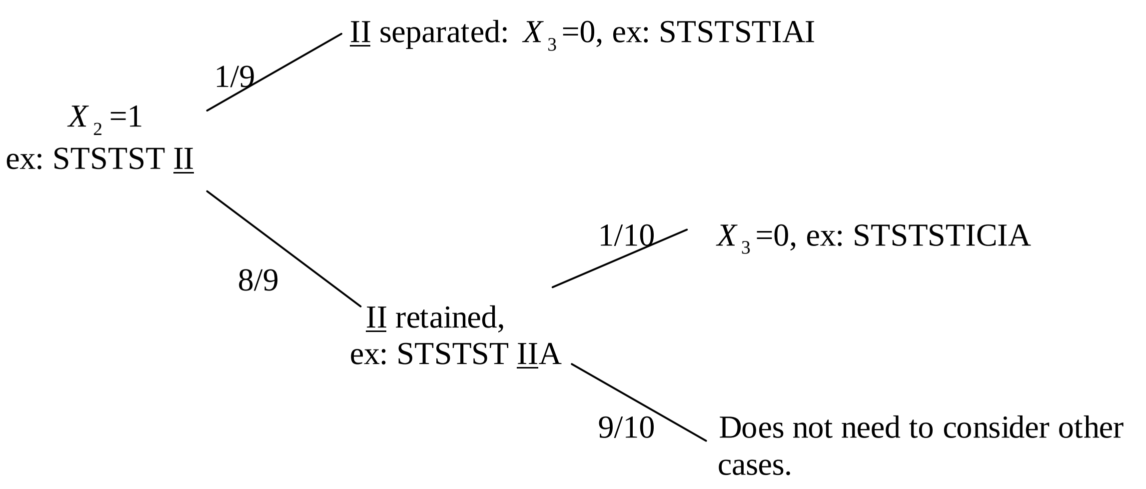
Let X_3 be the number of equal neighbors among the ten letters , SSSTTTIIAC, obtained.

Write $P(X_3 = 0) = \sum_{j=0}^{2} P(X_3 = 0 | X_2 = j) P(X_2 = j)$ (1.4)

When X_2 =0, for example, STISTIST, it does not matter where to insert A and C, there are no equal neighbors. Therefore, $P(X_3 = 0 | X_2 = 0) = 1$.

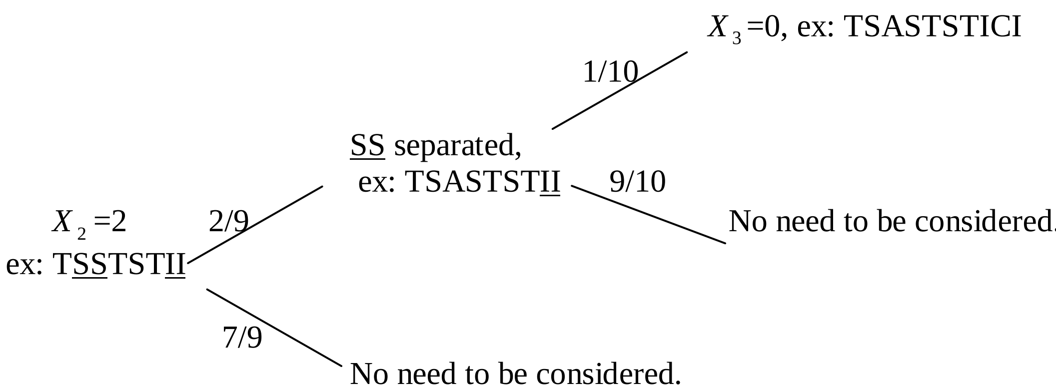
When X_2 =1, for example, STSTST<u>II</u>, $P(X_3 = 0 | X_2 = 1) = \frac{1}{9} + \frac{8}{9} \times \frac{1}{10} = \frac{18}{90}$ (see Figure 6.)

Figure 6: Tree diagram for calculation of the conditional probability $P(X_3 = 0 | X_2 = 1)$



When X_2=2, for example, TSSTSTII, $P(X_3=0|X_2=2)=\frac{2}{9}\times\frac{1}{10}=\frac{2}{90}$ (see Figure 7.)

Figure 7: Tree diagram for calculation of the conditional probability $P(X_3=0|X_2=2)$



Therefore,

$$P(X_3=0)=P(X_3=0|X_2=0)P(X_2=0)+P(X_3=0|X_2=1)P(X_2=1)+$$
$$P(X_3=0|X_2=2)P(X_2=2)=\left(\frac{148}{1120}\right)(1)+\left(\frac{18}{90}\right)\left(\frac{340}{1120}\right)+\left(\frac{2}{90}\right)\left(\frac{360}{1120}\right)=\frac{1}{5}.$$

This is the answer to the Statistics problem.

REFERENCES

Blom, G., Englund, J., and Sandell, D. (1998), *The Mississippi Problem,* The American Statistician, Vol. 52, 49-50.

Blom, G., Holst, L., and Sandell, D. (1994), *Problems and Snapshots from the World of Probability,* New York: Springer-Verlag.

Constantine, G.M. (1987), *Combinatorial Theory and Statistical Design*, New York: Willey.

David, F.N., and Barton, D.E. (1962), *Combinatorial Chance*, London: Griffin.

Mood, A.M. (1940), *The Distribution Theory of Runs,* The Annals of Mathematical Statistics, 11, 367-392.