Probability

Part 1 - Basic Counting Principles

1. References

(1) R. Durrett, *The Essentials of Probability*, Duxbury.

(2) L.L. Helms, *Probability Theory with Contemporary Applications*, Freeman.

(3) J.J. Higgins and S. Keller-McNulty, *Concepts in Probability and Stochastic Modeling*, Duxbury.

(4) R.V. Hogg and E.A. Tanis, *Probability and Statistical Inference*, Prentice-Hall.

2. Solved Problems

Problem 1. An experiment consists of choosing a sequence of 6 digits. Describe the sample space of this experiment. How many possible outcomes of the experiment are there? Let A denote the event that the sequence begins with two consecutive digits in natural order. How many outcomes belong to the event A?

Solution: The sample space S of an experiment is defined to be the set of all possible outcomes of the experiment. Here it is the set of all sequences $(x_1, x_2, x_3, x_4, x_5, x_6)$, with each $x_i \in \{0, 1, 2, \cdots, 9\}$. Since repetitions are not specifically prohibited, we will assume that they may occur. There are 10 possibilities for x_1, 10 possibilities for x_2, $\cdots$, and 10 possibilities for x_6. By the multiplication principle, the total number of outcomes is $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^6$. This is the number of functions from a set of size 6 to a set of size 10. In general, the number of functions from a set of size n to a set of size m is m^n.

Let B_1 denote the set of all sequences that begin with $(0,1)$, B_2 those that begin with $(1,2)$, $\cdots$, and finally let B_9 denote the set of sequences that begin with $(8,9)$. The number of outcomes in each B_i is 10^4, the B_i are pairwise disjoint, and $A = \cup_{i=1}^{9} B_i$. Therefore, A has 9×10^4 outcomes.

Problem 2. Consider the same experiment and event A described in the preceding problem, but assume that the terms of the sequence must be distinct. How many outcomes are in the sample space and the event A?

Solution: An outcome is now a non-repeating sequence of digits $(x_1, \cdots, x_6)$. It may be helpful to think of the digits as being chosen sequentially in time. There are 10 possibilities for x_1. Once x_1 has been selected its value is excluded for future selections. Thus there are 9 possibilities for x_2, 8 possibilities for x_3, and so on. Finally, there are 5=10-6+1 possibilities for x_6. By the multiplication principle, there are $10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151200$ outcomes in all. In general, the number of nonrepeating sequences of length n from a set of m elements $(m \geq n)$ is $m(m-1)\cdots(m-n+1) = m!/(m-n)!$. This is also the number of one-to-one functions from a set of size n to a set of size m, or the number of *permutations* of length n from a set of m objects.

Define events B_1, B_2, etc. as in Problem 1. For B_1, the initial terms are 0 and 1 and the last four terms of an outcome in B_1 must be a nonrepeating sequence from the remaining eight digits. Thus B_1 has $8 \times 7 \times 6 \times 5 = 1680$ outcomes. The other B_i have the same number. Therefore, the number of outcomes in A is $9 \times 1680 = 15120$.

Problem 3. Assume that the experiment in Problem 1 has *equally likely* outcomes. What is the probability of the event A? What is the probability of A in Problem 2?

Solution: The phrase "equally likely outcomes" applied to an experiment with a finite number N of possible outcomes means that each individual outcome, or singleton event, has probability $1/N$. Equivalently, the probability of any event A is $\#A/N$, where $\#A$ denotes the cardinality of A. Thus, in Problem 1, each individual outcome has probability 10^{-6} and the probability of A is $9 \times 10^4/10^6 = 0.09$.

In Problem 2, if outcomes are equally likely, the probability of event A is $P(A) = 15120/151200 = 0.10$.

Problem 4. An experiment consists of selecting a subset (not a sequence) of six of the ten digits. How many outcomes does this experiment have? Define A to be the event that the chosen subset contains two consecutive digits. How many outcomes belong to A? If all outcomes are equally likely, what is the probability of A?

Solution: The number of subsets of size n from a set of size m is given by the binomial coefficient $\binom{m}{n} = \frac{m!}{n!(m-n)!}$. Thus, the number of outcomes for this experiment is $\binom{10}{6} = 210$. To find the number of outcomes containing a pair of consecutive digits, it is easier to count the ones that don't contain a consecutive pair. Suppose there is such an outcome and its members are arranged in increasing order. Between the 6 members of the outcome there are 5 gaps where other integers between 0 and 9 belong. This makes at least 11 integers between 0 and 9, which is clearly impossible. Therefore, every outcome belongs to A. If all outcomes of the experiment are equally likely, the probability of A is $210/210 = 1$.

Problem 5 (The birthday problem): What is the probability that in a set of n people, at least two will have the same birthday?

Solution: In problems like this one, where no information to the contrary is given, you may assume that outcomes are equally likely. The difficulty is usually in deciding exactly what the outcomes are. In this case, an outcome is an assignment of one of the 365 birthdays to each person in the set. In other words, it is a function from a set of size n to a set of size 365. Therefore, there are 365^n equally likely outcomes. If an outcome belongs to the event "no two people with the same birthday", then the function is one-to-one and there are $365!/(365-n)!$ of these. Hence, the probability that no two people have the same birthday is

$$\frac{365 \times \cdots \times (365-n+1)}{365^n} = (1-\frac{1}{365})(1-\frac{2}{365})\cdots(1-\frac{n-1}{365})$$

The probability that two or more people do have the same birthday is 1 minus this quantity. It is an increasing function of n up to $n = 365$. For $n = 22$ it is greater than 1/2.

Definition 2.1. *Suppose $\mathcal{P}$ is a finite population. An ordered sample* with replacement *of size n from $\mathcal{P}$ is a possibly repeating sequence of length n from $\mathcal{P}$. An ordered sample* without replacement *of size n from $\mathcal{P}$ is a nonrepeating sequence of length n of members of $\mathcal{P}$. An* unordered sample *of size n from $\mathcal{P}$ is a subset of $\mathcal{P}$ having n members.*

If the sampled population has N members, the numbers of samples of each of the three types are, respectively, N^n, $N(N-1)\cdots(N-n+1)$ and $\binom{N}{n}$.

3. Unsolved Problems

Problem 1: Three students are seated in a row of 10 chairs for an exam. Assuming all outcomes are equally likely, what is the probability that they are in three adjacent chairs? What is the probability that no two adjacent chairs are occupied?

Problem 2: A six-sided die is thrown 5 times. Assuming all outcomes are equally likely, what is the probability of getting three sixes in a row and no more? What is the probability of getting exactly three sixes in the 5 throws?

Problem 3: What is the probability that among 3 friends, at least two were born in the same month? Assume all outcomes are equally likely.

Problem 4: A population has N members. An ordered sample without replacement of size 3 is chosen in such a way that after each element of the sample is selected, all the remaining elements of the population are equally likely to be chosen next. Show that if K is any three-member subset of the population, the probability that K is the result of the sequential selection procedure is $1/\binom{N}{3}$. Generalize to ordered samples of size n. This shows that an unordered sample without replacement can be chosen sequentially. That is the way it is usually done.

Problem 5: Suppose a population $\mathcal{P}$ has N members and that an ordered sample of size n is chosen from it with replacement. Assume all outcomes are equally likely. What is the probability that no member of the population is included more than once in the sample? (Hint: This is a restatement and generalization of the birthday problem.)

Problem 6 (For those who like calculus and inequalities): In problem number 5, suppose N and n are both allowed to vary. Show that the probability of no repetitions in the sample approaches 1 if and only if $\frac{n^2}{N}$ goes to 0.