Math 3338: Probability (Fall 2006)

Jiwen He

Section Number: 10853

`http://math.uh.edu/~jiwenhe/math3338fall06.html`

2.3 Counting Techniques (I) - Ordered Samples

Elements of combinatorial analysis

- **Proposition (The product rule of ordered pairs):** Given n_1 elements $a_1, \ldots, a_{n_1}$ and n_2 elements $b_1, \ldots, b_{n_2}$, there are precisely $n_1 \times n_2$ distinct ordered pairs (a_i, b_j) containing one element of each kind.

  Proof: Arrange the pairs in a rectangular array in the form of a multiplication table with n_1 rows and n_2 columns so that (a_i, b_j) stands at the interection of the ith row and jth column. Then each pair appears once and only once, and the assertion becomes obvious.

- **Proposition (General product rule of ordered multiplets):** Given n_1 elements $a_1, \ldots, a_{n_1}$, n_2 elements $b_1, \ldots, b_{n_2}$, etc, up to n_r elements $x_1, \ldots, x_{n_r}$, there are precisely $n_1 \times n_2 \times \cdots \times n_r$ distinct ordered r-tuples $(a_{i_1}, b_{i_2}, \ldots, x_{i_r})$ containing one element of each kind.

  Proof (by induction): If $r = 2$ the assertion reduces to the product rule of ordered pairs. Suppose it holds for $r - 1$, so that in particular there are precisely $n_2 \cdots n_r$ $(r-1)$-tuples $(b_{i_2}, \ldots, x_{i_r})$ containing one element of each kind. Then, regarding the $(r-1)$-tuples as elements of a new kind, we note that each r-tuple $(a_{i_1}, b_{i_2}, \ldots, x_{i_r})$ can be regarded as made up of a $(r-1)$-tuple $(b_{i_2}, \ldots, x_{i_r})$ and an element a_{i_1}. Hence, by the product rule of ordered pairs, there are precisely $n_1(n_2 \cdots n_r) = n_1 n_2 \cdots n_r$ r-tuples containing one element of each kind.

Examples

- **2.17:** Given $n_1 = 12$ plumbers $P_1, \ldots, P_{12}$ and $n_2 = 9$ electricians $Q_1, \ldots, Q_9$, there are $N = n_1 n_2 = (12)(9) = 108$ possible way of choosing the two types of contractors (P_i, Q_j).
- **2.18:** There are $n_1 = 4$ obstetricians, $Q_1, \ldots, Q_4$, and for each Q_i there are $n_2 = 3$ choices of pediatricians P_j for which O_i and P_j are associated with the same clinic. Applying the product rule gives $N = n_1 n_2 = (4)(3) = 12$ possible choices.
- **2.19 (2.17 continued):** Given $n_3 = 5$ appliance dealers. There are $N = n_1 n_2 n_3 = (12)(9)(5) = 540$ way to choose the three types of contractors (P_i, Q_j, D_k).
- **2.20 (2.18 continued):** If each clinic has both $n_3 = 3$ specialists in internal medicine and $n_4 = 2$ general surgeons, applying the product rule gives $N = n_1 n_2 n_3 n_4 = (4)(3)(3)(2) = 72$ possible ways to select one doctor of each type such that all doctors practice at the same clinic.

Ideal Experiment

- **Important facts:**
  - The sample space (together with probability distribution defined on it) defines the *ideal experiment*.
  - The nature of the sample points is irrelevant for our theory.
  - The same sample space can admit of a great variety of different interpretation. A number of situations in which the intuitive background varies, all are however, abstractly equivalent to the same sample space, in the sense that the outcome differ only in their verbal description.
- **Combinatorial product rule:** Let

$$A_1 = \{ a_1, \ldots, a_{n_1} \}, \quad B_2 = \{ b_1, \ldots, b_{n_2} \}, \cdots, X_r = \{ x_1, \ldots, x_{n_r} \}.$$

We define

$$A_1 \times B_2 \times \cdots \times X_r = \{ (a_{i_1}, b_{i_2}, \ldots, x_{i_r}), i_i = 1, \cdots, n_1, \cdots, i_r = 1, \cdots, n_r \}.$$

We have then

$$|A_1 \times B_2 \times \cdots \times X_r| = |A_1| \times |B_2| \times \cdots \times |X_r|.$$

Ordered Samples

- **Sampling with replacement and with ordering:** Suppose that we choose r objects in succession from a set or "population" of n distinct objects $a_1, \ldots, a_n$, in such a way that after choosing each object and recording the choice, we return the object to the population before making the next choice. This gives an *ordered sample of size* r $(a_{i_1}, a_{i_2}, \ldots, a_{i_r})$. Setting $n_1 = n_2 = \cdots = n_r = n$ in the product rule, we find that there are precisely

$$N = n^r$$

distinct ordered samples of the size r.

- **Sampling without replacement and with ordering:** Suppose that we choose r objects in succession from a set or "population" of n distinct objects $a_1, \ldots, a_n$, in such a way that an object once chosen is removed from the population. This gives a permutation of size r, which is again an *ordered sample of size* r $(a_{i_1}, a_{i_2}, \ldots, a_{i_r})$, but now $a_1, \ldots, a_n$ are distinct, and there are $n-1$ objects left after the first choice, $n-2$ objects left after the second choice, and so on. Cleary this corresponds to setting $n_1 = n$, $n_2 = n-1, \ldots, n_r = n-r+1$ in the product rule. Hence, instead of n^r distinct samples as in the case of sampling with replacement, there are now only

$$N = n(n-1)\cdots(n-r+1) =: P_{r,n}$$

distinct *permutations* of size r. Denote $P_{n,n}$ by $n! := n(n-1)\cdots 1$. We have

$$P_{r,n} = \frac{n!}{(n-r)!}.$$

Distribution of r balls in n cells

- Suppose we place r distinguishable balls into n different cells. "Placing balls into cells" amounts to choosing one cell for each ball. Numbering both the balls and the cells, let i_1 be the number of the cell into which the first ball is placed, i_2 be the number of the cell into which the second ball is placed, and so on. Then the arrangement of the balls in the cells is described by an ordered r-tuple $(i_1, i_2, \dots, i_r)$. This is equivalent to sampling with ordering.
  - **Cell allowed to contain more than one ball:** With r balls we have r independent choices. This is equivalent to sampling with replacement and with ordering. Therefore r balls can be placed into n cells in n^r different ways.
  - **No cell allowed to contain more than one ball ($r \leq n$):** Clearly, there are $n_1 = n$ empty cells originally, $n_2 = n - 1$ emply cells after one cell has been occupied, and so on. Then the arrangement of the balls in the cells is described by a permutations of size r. Hence the total number of distinct arrangements of the balls in the cells is $P_{r,n}$.
- A great variety of conceptual experiments are abstractly equivalent to that of placing balls into cells.

Distribution of r balls in n cells (cont.)

1. **Birthdays:** The possible configurations of the birthdays of r people correspond to the different arrangements of r balls in $n = 365$ cells (assuming the year to have 365 days).
2. **Dice:** The possible outcomes of the experiment of throwing a die r times (or throwing r dice) correspond to placing r balls into $n = 6$ cells.
3. **Coin:** When *tossing a coin*, we are in effect dealing with only $n = 2$ cells.
4. **Sampling:** Let a group of r people be classified according to age,or profession. The categories play the role of our cells, the people that of balls.
5. **Ramdom digits:** The possible orderings of a sequence of r digits correspond to the distribution of r balls (=places) into ten cells called $0, 1, \ldots, 9$.
6. **Accidents:** Classifying r accidents according to the weekdays when they occurred is equivalent to placing r balls into $n = 7$ cells.
7. **Elevator:** An elevator starts with r passengers and stops at n floors. The different arrangements of discharging the passengers are replicas of the different distribution of r balls in n cells.
8. **Sex distribution:** The sex distribution of r people. Here we have $n = 2$ cells and r balls.

Example - Birthdays

- **Question:** There are r people gathered in a room.
  1. What is the probability that two (at least) have the same birthday?
  2. What is the probability that at least one has the same birthday as you?
- **Answer:** Denote the sample space by $\Omega = \{(i_1, i_2, \ldots, i_r)\}$, where, for $k = 1, \ldots, r$, $i_k \in \{1, \ldots, n = 365\}$ is the birthday of the kth person. Then $|\Omega| = 365^r$.
  1. Denote the complementary event

     $$A^c = \{\text{all birthdays are different}\},$$

     which is equivalent to sampling without replacement and with ordering. The product rule gives $|A^c| = P_{r,365}$. So the probability that all birthdays are distinct is $P(A^c) = \frac{P_{r,365}}{365^r}$ and that two or more people have the same birthday is $P(A) = 1 - P(A^c) = 1 - \frac{P_{r,365}}{365^r}$. For $r = 22$, $P(A) \approx 0.4927$; for $r = 23$, $P(A) \approx 0.5243$.
  2. Denote the complementary event

     $$A^c = \{\text{The birthdays of others are different than my birthday}\},$$

     which is equivalent to sampling with replacement and with ordering. The product rule gives $|A^c| = (365 - 1)^r$. So the probability that at least one has the same birthday as me is $P(A) = 1 - P(A^c) = 1 - \frac{364^r}{365^r}$. If we want $P(A) \approx 1/2$, we have $r \approx -1/log_2(364/365) \approx 252.61$.

More Examples

- **Generality:** We consider random samples of size r with replacement taken from a population of the n elements $a_1,\ldots, a_n$. We are interested in the event A that in such a sample $a_1,\ldots, a_n$ no element appears twice, that is, that our sample could have been obtained also by sampling without replacement. The product rule shows that there are n^r different samples in all, of which $P_{r,n}$ satisfy the stipulated condition. Assuming that all arrangements have equal probability, we conclude that *the probability of no repetition in our sample* is

$$P(A) = \frac{|A|}{|\Omega|} = \frac{P_{r,n}}{n^r} = \frac{n(n-1)\cdots(n-r+1)}{n^r}.$$

The following concrete interpretations of this formula will reveal surprising features:

1. **Throwing a die six times:** the probability that all faces $(n = 6)$ turn up is $\frac{n!}{n^n} = \frac{6!}{6^6} \approx 0.01543$, extremely improbable.
2. **Elevator:** An elevator starts with $r = 7$ passengers and stops at $n = 10$ floors. The probability that no two passengers leave at the same floor is $\frac{P_{7,10}}{10^7} \approx 0.06048$, quite improbable.
3. **Random sampling numbers:** Consider the number $e = 2.71828\ldots$. Every succession of five digites represents a sample of size $r = 5$ for a population consisting of the ten digits $0, 1, \ldots, 9$. The probability that five consecutive random digits are all different is $\frac{P_{5,10}}{10^5} = 0.3024$.

2.3 Counting Techniques (II) - Subpopulations

Subpopulation: sampling without ordering

- Any *set* of r elements chosen from a population of n elements, without replacement and without ordering is called a *subpopulation* of size r of the original population. The number of such subpopulations is given by
  - **Theorem:** A population of n elements has precisely
  $$C_r^n = \frac{n!}{r!(n-r)!}$$
  subpopulations of size $r \leq n$.
  - **Proof:** If order mattered, then the elements of each subpopulation could be arranged in $r!$ distinct ways. Hence there are $r!$ times more "ordered samples" of r elements than subpopulations of size r. But there are precisely $n(n-1)\cdots(n-r+1)$ such ordered samples and hence jsut
  $$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$
  subpopulations of size r.
  - **Remark:** An expression of the form C_r^n is called a *binomial coefficient*, often denoted by $\binom{n}{r}$ instead of C_r^n. The number C_r^n is sometimes called the *number of combinations of n things taken r at a time* (without regard of order).

Bridge and Poker

- **Example 2.22 - Bridge:** A bridge hand consists of any 13 cards selected from a 52-card deck without regard to order. There are $C^{52}_{13} \approx 635$ billion different bridge hands. Suppose a bridge hand is dealt from a well-shuffled deck (i.e., 13 cards are randomly selected from among 52 probabilities) and C^{52}_{13} possible outcomes are equally likely. Let

$$A = \{\text{the hand consists entirely of spades and clubs with both suits represented}\}$$

$$B = \{\text{the hand consists of exactly two suits}\}$$

  Since there are 13 cards in each suit, the number of hands consisting entirely of spades and/or clubs (i.e., no red cards) is $C^{26}_{13} \approx 10$ millions. One of these C^{26}_{13} consists entirely of spades, and one consists entirely of clubs, so $|A| = C^{26}_{13} - 2$. Since there are $C^4_2 = 6$ combinations consisting of two suits, of which spades and clubs is one such combination, using the product rule gives $|B| = C^4_2 \times |A|$. Then

$$P(A) = \frac{C^{26}_{13} - 2}{C^{52}_{13}} = 0.0000164, \quad P(B) = \frac{C^4_2 \times (C^{26}_{13} - 2)}{C^{52}_{13}} = 0.0000983 \approx \frac{1}{10,000}.$$

- **Poker:** There exisits $C^{52}_5 \approx 2.5$ millions hands at poker. Let

$$A = \{\text{the hand consists of five different faces values}\}$$

  These face values can be chosen in C^{13}_5 ways, and corresponds to each card we are free to choose one of the four suits. Using the product rule gives $|A| = 4^5 \times C^{13}_5$ and $P(A) = \frac{4^5 \times C^{13}_5}{C^{52}_5} = 0.5071.$

More examples

- **Placing r balls in n cells:** There are n^r possible arrangements. Let $k \le n$ and

$$A = \{\text{a specified cell contains exactly } k \text{ balls}\}$$

The k balls can be chosen in C_k^r ways, and the remaining $r-k$ balls can be placed into the remaining $n-1$ cells in $(n-1)^{r-k}$ ways; using the product rule gives $|A| = C_k^r \times (n-1)^{r-k}$. It follows that

$$P(A) = \frac{|A|}{n^r} = \frac{C_k^r \times (n-1)^{r-k}}{n^r} = C_k^r \times \frac{1}{n^k} \times \left(1 - \frac{1}{n}\right)^{r-k}.$$

The is a special case of the so-called *binomial distribution*.

- **Example 2.23:** A technician selects 6 of 25 printers, of which 10 are laser printers and 15 are inkject printers; there are $N = C_6^{25}$ ways of doing it. Let $r \le 6$ and

$$A_r = \{\text{exactly r of the 6 selected are inkjet printers}\}.$$

There are C_r^{15} ways of choosing the r inkjet printers and then C_{6-r}^{10} ways of choosing the $6-r$ laser printers; using the product rule gives $|A_r| = C_r^{15} \times C_{6-r}^{10}$. Then $P(A_3) = \frac{|A_3|}{N} = .3083$, and

$$P(A_3 \cup A_4 \cup A_5) = P(A_3) + P(A_4) + P(A_5) = \frac{|A_3|}{N} + \frac{|A_4|}{N} + \frac{|A_5|}{N} = .853.$$

2.3 Counting Techniques (III) - Partitions