COUNTING

Jerzy Szulga

Department of Mathematics and Statistics
Auburn University

MATH 5670-6670 FALL 2019

PROBABILITY I

August 23, 2019

Outline

The most general discrete

Elementary outcomes are non-divisible (a.k.a. "atoms").

Probabilities - a sequence of nonnegative numbers summing up to 1:

atoms	a_1	a_2	$\ldots$	a_n	$\cdots$
probabilities	p_1	p_2	$\ldots$	p_n	$\cdots$

Physical modeling:

- a tape partitioned in proportion to p_i's; a random point chooses an interval (easy on computers, hard in real word);
- a spinner (akin to "Wheel of Fortune") spun and stopped at random; angles proportional to p_i's;
- an urn, if S is finite and $p_i = \frac{n_i}{n}$ are rational, with n_i balls marked by "i"; a ball is selected at random.

Uniform discrete probability distribution

The principal example of a **uniform** probability model involves

- a finite sample space Ω (often denoted by S),
- equally likely elementary outcomes (a.k.a. atoms),
- events being sets of some atoms.

Thus, denoting by $|A|$ the count of elements in $A \subset \Omega$:

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|}. \tag{1}$$

Warning: The elementary outcomes **must be** equally likely!

If they are not, the formula (1) is false.

Multiplication Rule

If an event is described by a number of **independent** steps, then

$$\Big(\text{Total count}\Big) = \Big(\text{count in Step 1}\Big) \times \Big(\text{count in Step 2}\Big) \times \cdots$$

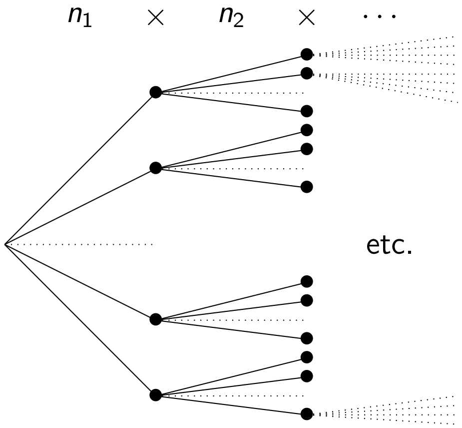
independent - the principles of count are not affected by the previous steps although the counts themselves might

Example. Draw two cards without replacement from a deck.

Step 1: by uniformity, there are 52 equally likely choices;

Step 2: by uniformity (not affected), there are 51 (change) equally likely choices;

Multiplication Rule - the tree



Useful only for small counts $n_1, n_2, \ldots$ or just for few steps.

SAMPLING

Merriam-Webster Dictionary

sample - a finite part of a statistical population whose properties are studied to gain information about the whole

sampling - specifically : the act, process, or technique of selecting a representative part of a population for the purpose of determining parameters or characteristics of the whole population

Well... what is "statistical population", "a part" - ordered or unordered? What if that part is not studied, or studied just for itself, or for fun, or to manipulate the minds of people? - not a "sample"?

Oxford English Dictionary

sample - A relatively small quantity of material, or an individual object, from which the quality of the mass, group, species, etc. which it represents may be inferred; a specimen.

Only next different but related meanings are given for various areas such as commerce, education, psychology, science, and so on, including statistics.

Statistics. A portion drawn from a population, the study of which is intended to lead to statistical estimates of the attributes of the whole population.

OED, a.k.a. **Lost in Words** - population?

From Latin *populus* - people

Genetics. A group of animals, plants, or humans, within which breeding occurs.

Physics. The (number of) atoms or subatomic particles that occupy any particular energy state.

Astronomy. Any of several groups, originally two in number, into which stars and other celestial objects are categorized on the basis of where in the galaxy they were formed.

Statistics. A (real or hypothetical) totality of objects or individuals under consideration, of which the statistical attributes may be estimated by the study of a sample or samples drawn from it.

SAMPLING

r **items are sampled from a population of** n **distinct objects** (e.g., a city of 1000 people, a deck of 52 cards):

1. **without replacement** (necessarily $r \leq n$)

   (e.g., 5 people for a city council, 5 cards for a poker hand).

   with order or **without order**

2. **with replacement** (no restriction on r vs. n)

   (e.g., electing the sheriff each year for 5 years, selecting a card, shuffling, and again 5 times)

Sequences

A **sequence** *is function from a subset of natural numbers into some set $X \neq \emptyset$.*

We denote finite sequences by $\mathbf{x} = (x_1, x_2, \ldots, x_r)$, where $r = x = |\mathbf{x}|$ is called the sequence's **length**. We often skip brackets and commas, e.g., in MISSISSIPPI, 2019, 01001001, passwords, VIN...

Elements of a sequence <u>may repeat</u> and are automatically <u>ordered</u>. For example, $(1, 2, 1)$ is a correct sequence, and $(1, 2, 1) \neq (1, 1, 2)$. We use the **ordinals** such as 1^{st}, 2^{nd}, 3^{rd}, 4^{th} element, etc.

A sequence with no repetitions is called a **permutation**.

Counting sequences

How many 3-digit numbers can you form using $1, 2, 3, 4, 5$, when repetitions are allowed?

Answer: By the Multiplication Rule, $5 \cdot 5 \cdot 5 = 5^3 = 125$.

In general, the count of **sequences of length r with elements selected from a set of n objects** is

$$n^r = \underbrace{n \cdot n \cdots n}_{r \text{ factors}}$$

Warning: Don't confuse the power with 3^5 or r^n (a common error)!

Counting permutations

How many 3-digit numbers can you form using $1, 2, 3, 4, 5$ without repetition?

Answer: By the Multiplication Rule, $5 \cdot 4 \cdot 3 = 60$.

The count of **permutations of r objects selected from a set of n objects**:

$$(n)_r = P^n_r = \frac{n!}{(n-r)!} = \underbrace{n \cdot (n-1) \cdot \cdots \cdot (n-k+1)}_{r \text{ factors}}$$

For example,

$$(9)_4 = P^9_4 = \frac{9!}{5!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot \cancel{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}}{\cancel{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}} = 9 \cdot 8 \cdot 7 \cdot 6.$$

Sequences: crisp notation

Recall the **Cartesian product** of two or more sets:

$$A \times B = \{ (a, b) : a \in A, b \in B \},$$

$$A_1 \times \cdots \times A_r = \{ (x_1, \ldots, x_r) : x_i \in A_i,\ i = 1, ..., r \}.$$

When $A_1 = A_2 = \cdots A_r = A$, we simply write

$$A^r = \underbrace{A \times \cdots \times A}_{r}.$$

For the set of permutations we may write $A^{(r)}$.

Combinations

A **combination** *is a finite subset of X.*

A combination (i.e., a set) can be specified by using curly brackets:

$$\{ x_1, \ldots, x_r \} .$$

Now, r is called the set's **size** or **count**.

A more fancy word is "**cardinality**" or "**cardinal number**".

Elements of a set must not repeat and there is no order.

For example, there is no 1st nor 2nd element, and

$\{ 1, 2, 3 \} = \{ 3, 2, 1 \}$. The symbol $\{ 1, 2, 1 \}$ is incorrect.

Counting combinations

In a permutation of $r = 4$ from $n = 9$ the order is important, e.g. $(1, 7, 4, 5) \neq (4, 1, 5, 7)$. Now, disregard the order.

There are $24 = 4!$ permutations of such sequence. Therefore, we must divide the number of permutations by that factorial: $\frac{P_4^9}{4!}$.

The number of combinations is

$$\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{(n)_r}{r!} = \frac{n!}{(n-r)!r!}$$

a.k.a. the **binomial coefficient**.

Which symbols to use?

P_r^n or C_r^n are archaic (but appear on fancier calculators);
$(n)_r$ or $\binom{n}{r}$ are modern.

One may read in short:

$\binom{n}{r} = C_r^n$ - "*the number of combinations of r from n*",
- "*n choose r*"

$(n)_r = P_r^n$ - "*the number of permutations of r from n*",
- "*n choose r with order*"

Combinations: easy to confuse or abuse

Here, the "combination" is a precisely defined technical term.

Its meaning differs from the common usage.

For example, "*a combination to the lock*" is not our combination, because the order matters.

In general, "combination" means a process of putting, using, or mixing things together, or an outcome of such process. Therefore it may cover sequences and permutations as well.

Advice: Avoid this general meaning! Use some synonyms instead:

arrangements, configurations, patterns, schemes, options, ways, ...

Non uniform discrete models

Every discrete probability problem must begin with a sample space Ω. The elements of Ω are called **elementary outcomes** or **atoms**.

The ratio of counts as the probability is valid if and only if atoms are equally likely. A statement to this effect must be present in a solution of every problem prior to using that ratio.

Exercise. A room has 4 tables. 3 people enter and take a seat at a table at random. What is the probability that they sit alone? Consider and discuss (many) possible scenarios.

Summary: sampling

Sampling with or without replacement from a set of distinct objects.

This notion covers

- **sequences** - ordered samples with replacement
- **permutations** - ordered samples without replacement
- **combinations** - unordered samples (no repetitions)

The sampling from a set of identical objects does not fit the above classification. However...

PLACING

One dirty exam problem from the past

[10 points] In both cases describe the outcome space S **prior** to computing the probability.

Case 1 [3 points]. Pixie and Dixie, the cute mice, were chased by Mr. Jinks, the cat. Two holes, A and B, were of their rescue, and they succeeded. What is the probability they end up in one hole if each mouse chose a hole at random?

Case 2 [7 points]. Two cute gray mice were chased by a cat. Two holes, A and B, were of their rescue, and they succeeded. What is the probability they end up in one hole if mice chose holes at random?

Case 1. Placing distinct items in cells

n distinct items are placed in r distinct cells

So, both items and cells can be ordered, marked, or enumerated.

- **permutations** - the exclusion rule: no 2 items in one cell;

  the count $(n)_r$, $r \leq n$.

- **sequences** - unlimited supply of each item, no exclusion;

  the count n^r.

Case 2. Placing identical items in cells

Note the reversal of r and n.

r identical items are placed in n distinct cells

Consequently, the items cannot be marked or ordered.

- **combinations** - the exclusion rule applies: no 2 items in one cell; the count $\binom{n}{r}$, $r \leq n$;
- **free combinations** - no exclusion rule: the count $\binom{r+n-1}{r}$, or $\binom{r+n-1}{n-1}$ by symmetry.

The latter formula - why?

A free combination is an ordered partition of an integer r into n summands:

$$(r_1, \ldots, r_n): \qquad r = r_1 + \cdots + r_n, \quad 0 \leq n_i \leq r.$$

For small r and n it's easy to list all partitions but the listing soon becomes very tedious. E.g., for $r = 4$, $n = 3$ the count is 15:

$$(4,0,0),\ (0,4,0),\ (0,0,4),$$
$$(3,1,0),\ (1,3,0),\ (3,0,1),\ (0,3,1),\ (0,1,3),\ (1,0,3),$$
$$(2,2,0),\ (2,0,2),\ (0,2,2),$$
$$(2,1,1),\ (1,2,1),\ (1,1,2).$$

Any idea for a quick count in general?

Recall combinations through **Sampling**

Sample 3 items from a set of distinct 7 items, say from
$X = [1, 7] = \{ 1, 2, 3, 4, 5, 6, 7 \}$.

first choose a permutation in $(7)_3$ ways,
then disregard the order $\binom{7}{3} = \frac{(7)_3}{3!}$

Equivalent to "choose left behind", $\binom{7}{4} = \binom{7}{3}$.

Formally:

$$\begin{aligned} S = & \{ A \subset [1, 7] : |A| = 3 \} \\ = & \{ \{ x_1, x_2, x_3 \} : x_i \in [1, 7], i = 1, 2, 3 \} \end{aligned}$$

Alternative: combinations through **Placing**

Now the numbers $1, 2, 3, 4, 5, 6, 7$ play the roles of "cells":

1	2	3	4	5	6	7

Selecting numbers means placing a mark, say $*$, in a cell.

For a sample of 3 numbers, for example, $\{3, 5, 6\}$:

		$*$		$*$	$*$	
1	2	3	4	5	6	7

Use 1 instead of $*$, denote $\mathbf{d} = d_1 d_2, \ldots d_7$, $d = |\mathbf{d}| = \sum_i d_i = 3$:

$$S = \left\{ \mathbf{d} \in \{0, 1\}^7 : \, d = 3 \right\}, \qquad \text{e.g., } \{3, 5, 6\} \leftrightarrow 0010110.$$

Simple counting of free combinations

Say, $r = 3$ identical items $\ast \quad \ast \quad \ast$ are to be placed in $n = 4$ cells. The sample space S consists of sequences (r_1, r_2, r_3, r_4), where r_i is the number of items in the i^{th} cell. Clearly $0 \leq r_i \leq r$ and $r_1 + r_2 + r_3 + r_4 = r$. We record patterns as in the examples:

(0,1,1,1)	$\Big\vert \quad \vert \quad \ast \quad \vert \quad \ast \quad \vert \quad \ast \quad \Big\vert$	$r+n-1=6$,
(2,0,1,0)	$\Big\vert \quad \ast \quad \ast \quad \vert \quad \vert \quad \ast \quad \vert \quad \Big\vert$	positions,
(0,0,3,0)	$\Big\vert \quad \vert \quad \vert \quad \ast \quad \ast \quad \ast \quad \vert \quad \Big\vert$	or $r = 3$ items.

We choose places for items, or equivalently, for cells' inner walls. So the count is $\binom{6}{3}$. The general case works the same.

01-coding of free combinations

n cells and r indistinguishable items.

$r = r_1 + \cdots + r_n$, $0 \leq r_i \leq r$, the order of r_i's matters, $i = 1, \ldots, n$.

A placement pattern is coded as a 01 sequence of length $r + n - 1$.

The number r_i of items in the i^{th} cell is coded as the i^{th} sequence of 1's of length r_i. For example, for $n = 6$ cells and $r = 6$ items, e.g.:

$$(0, \mathbf{3}, 0, 0, \mathbf{1}, \mathbf{2}) \quad \longleftrightarrow \quad 0\mathbf{111}00\mathbf{1}, \mathbf{11} \longleftrightarrow 0\mathbf{111}00\mathbf{1}\ \mathbf{11}$$

(the comma, space, or other separator is necessary).

Total count is $\binom{11}{6} = \binom{11}{5} = 464.$

Summary

While placing items in cells, the count of the outcomes is:

	n items in r cells distinct, so orderable	r items in n cells identical, so no order
repetitions OK	n^r	$\binom{r+n-1}{r}$
no 2 in 1 cell	$(n)_r$	$\binom{n}{r}$

A physics connection

Elementary particles (electrons, protons, photons, quarks, etc.) are indistinguishable. They occupy certain quantum states - "cells" in our counting model.

- **Fermions** are subject to the **Pauli Exclusion Principle** - e.g., electrons or protons, yielding the **Fermi-Dirac** statistic $\binom{n}{r}$.
- **Bosons** follow no occupancy restriction - e.g. photons or mesons, the latest Higgs bosons, or speculative gravitons, yielding the **Boson-Einstein** statistic $\binom{r+n-1}{n-1}$.

How can we tell: distinct or identical?

This is a "Bayesian vs. frequentist" question.

Distinct is as distinct does.

Example. People as individuals are distinguishable. But in number they form a crowd. In a crowd they cease to be distinguishable.

How many people make a crowd? How many trees make a forest?

So, a Bayesian after speculation may claim one or the other. Yet, a repeated experiment and a frequency based statistics shall decide.