Outcomes, Events, and Sample Spaces
Counting Methods

Statistics 110

Summer 2006

Outcomes, Events, and Sample Spaces

When dealing with probability, we need to determine what we want to assign probabilities to

- Elementary Outcome: A complete result of the experiment under consideration. Also known as an outcome, simple event, or sample point.

- Sample Space: The set of all possible outcomes of the experiment

  Examples:

  1. Rolling a single die example: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
  2. Radioactive decay: $\Omega = \{0, 1, 2, \ldots\}$.

3. Mendel's peas:

$$\Omega = \left\{ (x_1, x_2, x_3, x_4) : x_i \geq 0 \text{ \& } \sum x_i = 560 \right\}$$
$$= \{(560, 0, 0, 0), (559, 1, 0, 0), (559, 0, 1, 0), \ldots\}.$$

   There happen to be $\binom{563}{3} = 29583961$ different outcomes in Ω.

4. SST anomaly forecasts (at a single location): $\Omega = (-\infty, \infty)$.
5. Flip a coin three times: $\Omega = \{$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$\}$

Note that sometimes its easier to include events may not be possible. For example, the SST temperature anomalies really can't get outside a fairly small range. As we will see later on, adding outcomes with zero probability isn't a problem.

- Event: An outcome or a set of outcomes. A subset of the sample space. A statement about the outcome of the experiment.

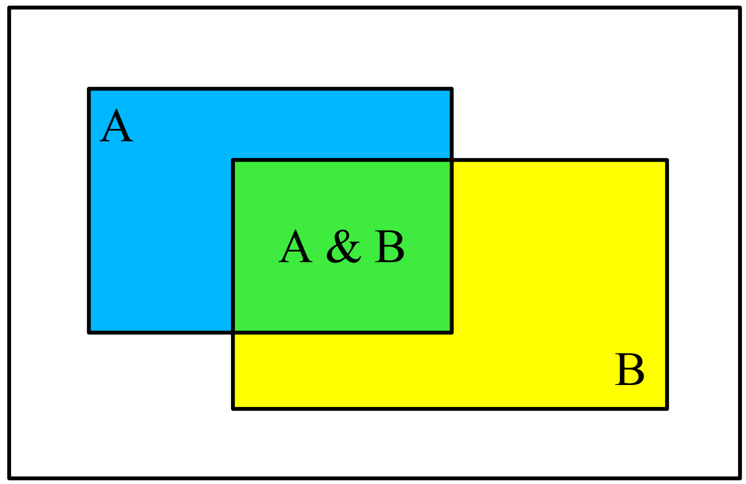
  Note that events are usually denoted by capital letters.

  Examples:

  1. Rolling a single die:
     (a) Roll is even — $A = \{2, 4, 6\}$
     (b) Roll is greater than 4 — $B = \{5, 6\}$
  2. Flip a coin three times:
     (a) All flips are the same — $C = \{$HHH, TTT$\}$
     (b) At least two heads — $D = \{$HHT, HTH, THH, HHH$\}$
  3. Flip a coin twice and roll a die once:
     (a) At least one tail and roll is 6 — $E = \{$HT6, TH6, TT6$\}$
     (b) Two tails — $F = \{$TT1, TT2, TT3, TT4, TT5, TT6$\}$

Combining Events

- Intersection (and): The set of all outcomes that occur in all of the sets



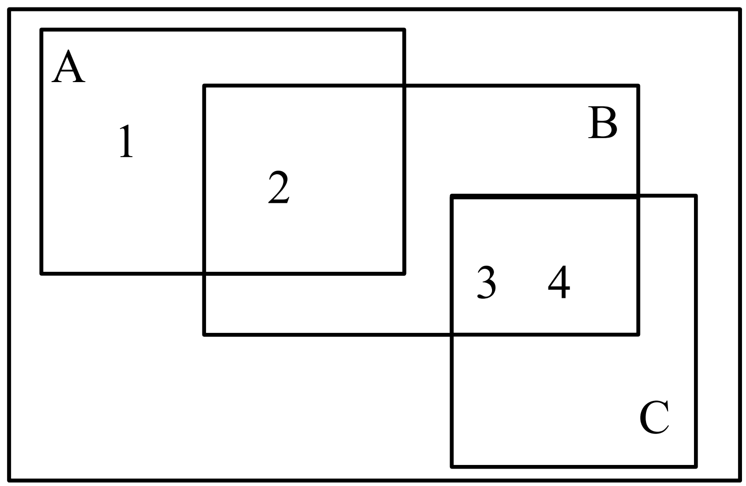
$$A \cap B = B \cap A \quad \text{(Commutative)}$$

If $A = \{1, 2\}, B = \{2, 3, 4\}$ and $C = \{3, 4\}$, then

$$A \cap B = \{2\}$$

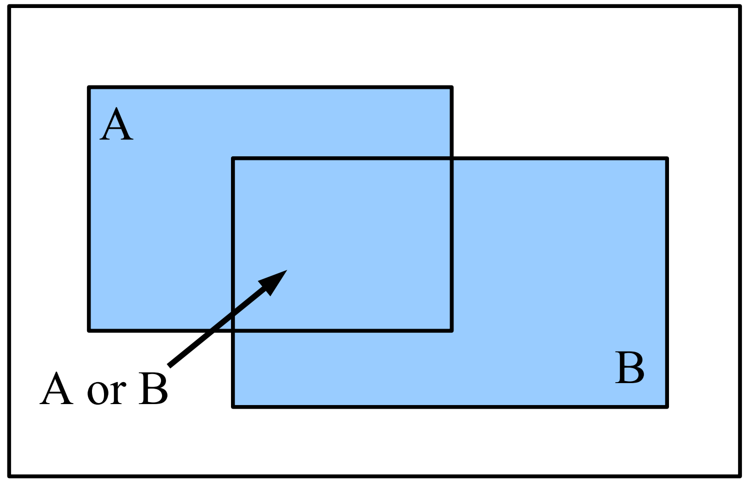
$$A \cap C = \phi \qquad\qquad \text{(Empty Set)}$$

$$B \cap C = \{3, 4\}$$



If $A \cap B = \phi$, the events A and B are said to be **disjoint**.

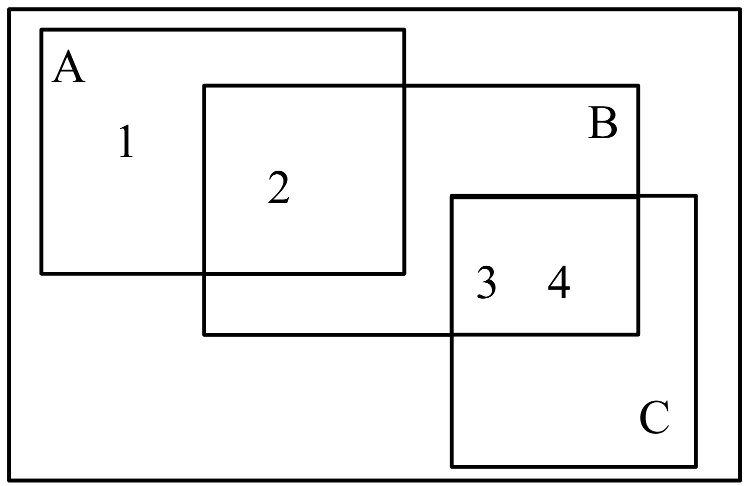
- Union (or): The set of all outcomes that occur in at least one of the sets



$$A \cup B = B \cup A \quad \text{(Commutative)}$$

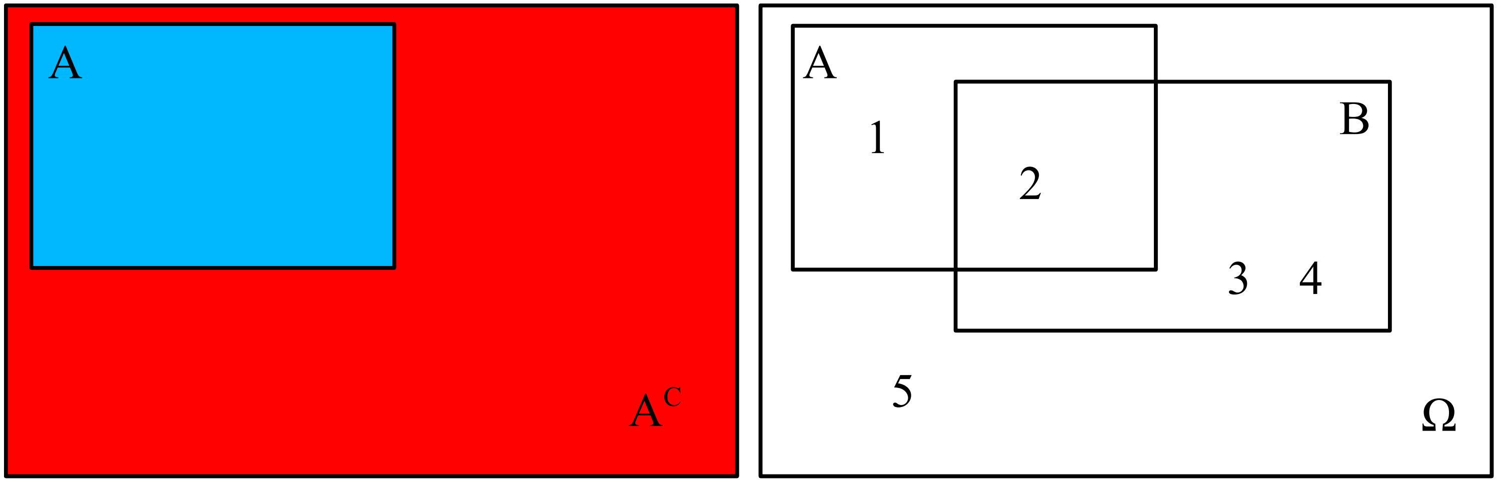
If $A = \{1, 2\}, B = \{2, 3, 4\}$ and $C = \{3, 4\}$, then

$$A \cup B = \{1, 2, 3, 4\}$$
$$A \cup C = \{1, 2, 3, 4\}$$
$$B \cup C = \{2, 3, 4\}$$



Note that "or" means one, or the other, or both (not exclusive or).

- Complement: The set of outcomes that don't occur in the event

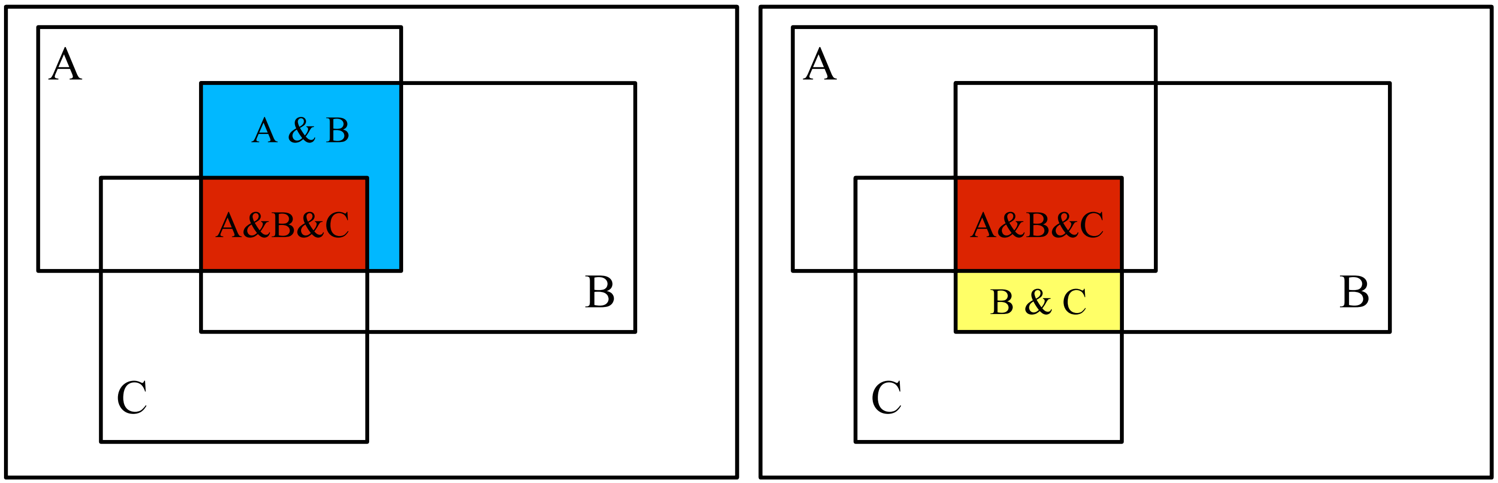


If $A = \{1, 2\}, B = \{2, 3, 4\}$ and $\Omega = \{1, 2, 3, 4, 5\}$, then

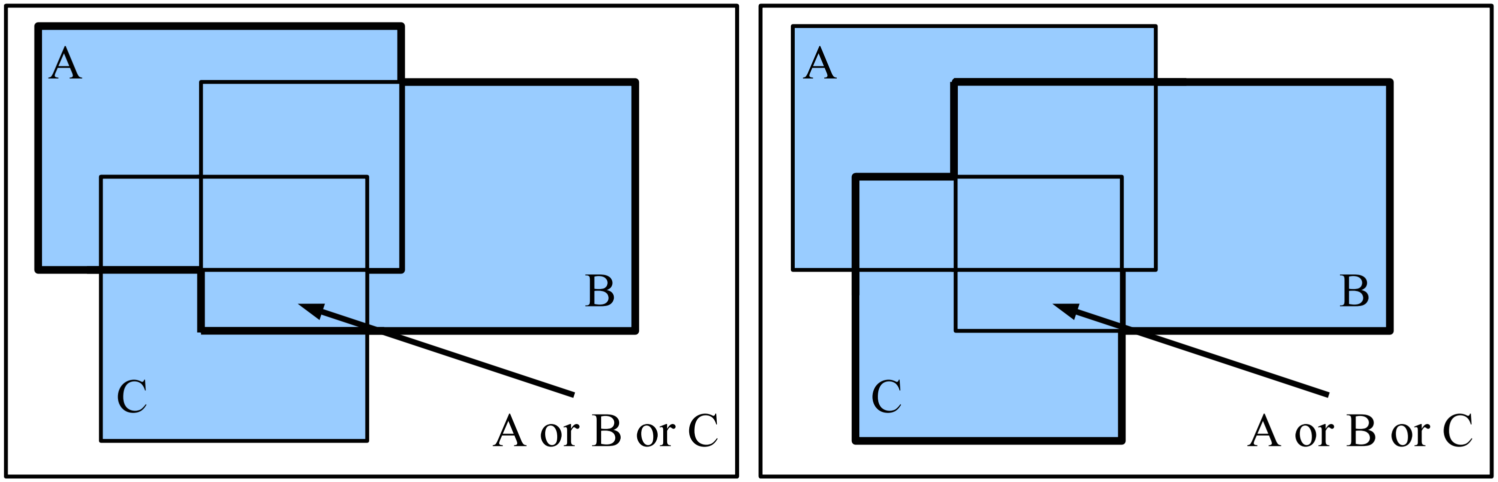
$$A^c = \{3, 4, 5\}$$
$$B^c = \{1, 5\}$$
$$\Omega^c = \phi$$

Associative Laws

- $(A \cap B) \cap C = A \cap (B \cap C)$

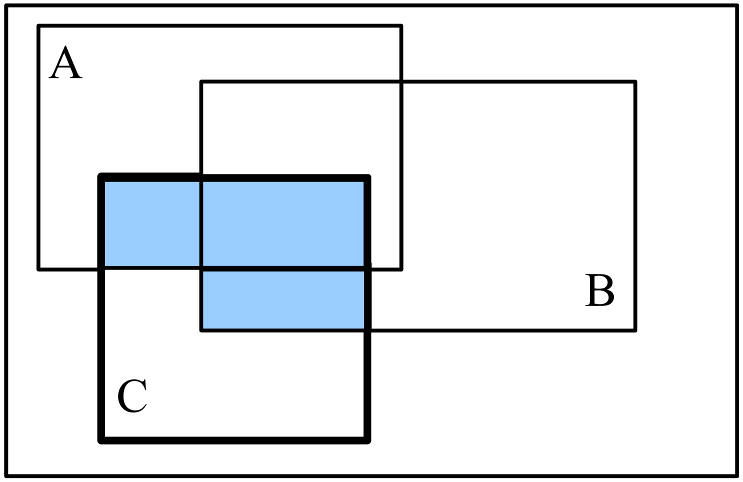


- $(A \cup B) \cup C = A \cup (B \cup C)$

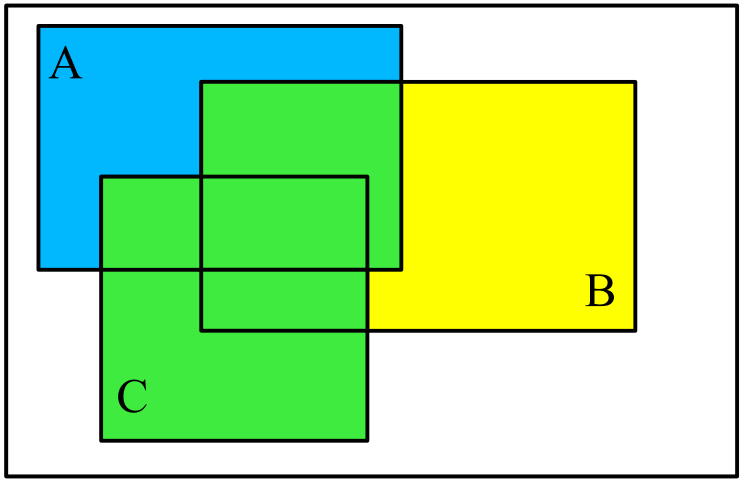


Distributive Laws

- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

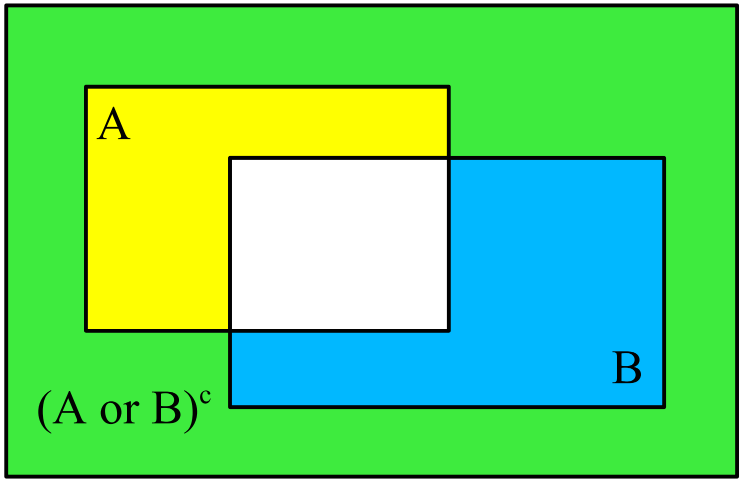


- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

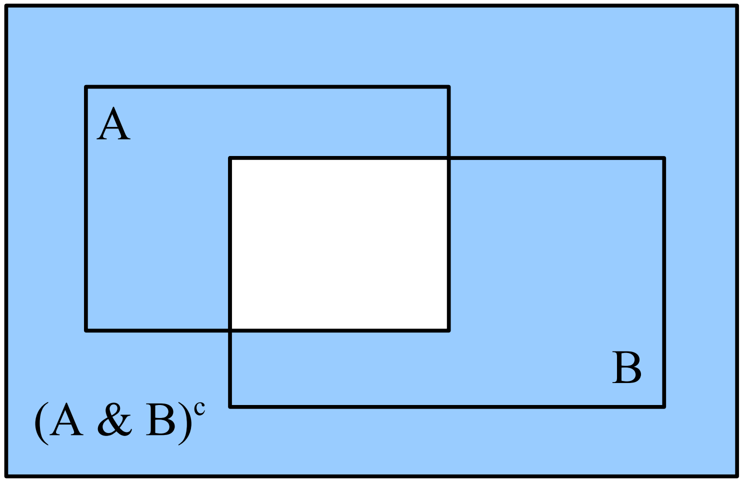


DeMorgan's Laws

- $(A \cup B)^c = A^c \cap B^c$



- $(A \cap B)^c = A^c \cup B^c$



Counting Methods

In some problems (such as rolling a fair die), each of the outcomes is equally likely. Thus, if there are N possible outcomes in the the sample space, each outcome has probability

$$p_0 = \frac{1}{N}$$

For example when placing 3 labelled balls in 3 boxes, there are 27 different possible outcomes ($= 3 \times 3 \times 3$), so

$$p_0 = \frac{1}{27}$$

In such equally likely cases, the probability of an event A, denoted by $P[A]$ satisfies

$$P[A] = (\text{Number points in } A) \times p_0$$
$$= \frac{\text{Number points in } A}{\text{Number points in } \Omega}$$

So for $A = \{\text{All balls end up in same box}\}$ and $B = \{\text{There is exactly 1 empty box}\}$, the probabilities are

$$P[A] = \frac{3}{27} = \frac{1}{9}; \quad P[B] = \frac{18}{27} = \frac{2}{3}$$

So for many problems, finding a probability reduces to figuring out how many possible outcomes satisfy the condition of interest.

There are two settings of interest that many counting problems fall into, ordered samples (e.g. a head followed by a tail is different from a tail followed by a head) and unordered sampling (e.g. all that matters is that there is a head and a tail).

- Ordered samples: Given a population of n elements $\{a_1, a_2, \ldots, a_n\}$ select an ordered sample or size r.

  1. Sampling with replacement:

     The sample space has $n \times n \times \ldots \times n = n^r$ outcomes.

     Two possible samples of size 3 (when $n = 7$) are $\{a_6, a_4, a_1\}$ and $\{a_3, a_5, a_3\}$.

     An example of this would be rerolling a die r times. For example rolling a die twice has 36 different possibilities (assuming you pay attention to the order).

2. Sampling without replacement:

The sample space has $(n)_r \stackrel{\text{def}}{=} n(n-1)(n-2)\ldots(n-r+1) = \frac{n!}{(n-r)!}$ outcomes.

A possible sample of size 3 (when $n = 7$) is $\{a_6, a_4, a_1\}$. The sample $\{a_3, a_5, a_3\}$ is not possible under this sampling scheme.

Special case: there are $n! \stackrel{\text{def}}{=} n(n-1)\ldots 2 \times 1 = (n)_n$ different orderings (permutations) of n elements.

So when sampling r items from a population of size n (with replacement), the probability of no repetition in our sample is

$$P[\text{No Repeats}] = \frac{(n)_r}{n^r}$$

Examples:

(a) An elevator starts with 5 passengers, with 7 floors where the passengers could get off. What's the chance that everybody gets off at a different floor? What the chance with 7 passengers? 8 passengers?

Set $n = 7$ and $r = 5$. Then

$$P[\text{No Repeats}] = \frac{(7)_5}{7^5} = \frac{2520}{16807} = 0.150$$

For $r = 7$,

$$P[\text{No Repeats}] = \frac{7!}{7^5} = 0.00612$$

For $r = 8$,

$$P[\text{No Repeats}] = 0$$

(b) What is the chance that all r students in a class have different birthdays? How big does r need to be for this probability to be less than 0.5? (An equivalent statement is how big does r need to be for the probability that at least two people have the same birthday is at least 0.5?)

Set $n = 365$ (We'll ignore leap day)

$$\begin{aligned} P[\text{Different Birthdays}] &= \frac{(365)_r}{365^r} \\ &= \frac{365}{365} \times \frac{364}{365} \times \ldots \times \frac{365 - r + 1}{365} \end{aligned}$$

For the second question, n needs to be at least 23 as

$$P[\text{Different Birthdays}|r = 22] = 0.524$$
$$P[\text{Different Birthdays}|r = 23] = 0.493$$

- Unordered samples (Subsets)

  Given a population of size n, the number of subsets of size $r (0 \leq r \leq n)$ is

$$\binom{n}{r} \stackrel{\text{def}}{=} \frac{(n)_r}{r!} = \frac{n!}{r!(n-r)!}$$

  Proof: Each unordered sample corresponds to $r!$ ordered samples.

  Example: What is the probability that a poker hand contains 5 different face values (e.g. 2, 6, 9, 10 ,K)

  – The sample space contains $\binom{52}{5}$ different hands
  – The number of different possible hands in our event is

$$\underbrace{\binom{13}{5}}_{\text{choice of 5 face cards}} \times \underbrace{4 \times 4 \times 4 \times 4 \times 4}_{\text{choice of suit for each card}}$$

– These give the probability

$$P = \frac{\binom{13}{5}4^5}{\binom{52}{5}} = 0.5071$$

A subset of size r is equivalent to partitioning the population into 2 parts, one with $r_1 = r$ points and the other with $r_2 = n - r$ points.

The number of partitions of n elements into k parts, with the 1st part containing r_1 elements, the 2nd with r_2 elements, the 3rd with r_3 elements, etc is

$$\binom{n}{r_1 \; r_2 \; \ldots \; r_k} \stackrel{\text{def}}{=} \frac{n!}{r_1! r_2! \ldots r_k!}$$

Note that $r_1 + r_2 + \ldots + r_k = n; r_i \geq 0$.

Proof:

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-r_2-\ldots-r_{k-1}}{r_k}$$
$$=\frac{n!}{r_1!(n-r_1)!}\frac{(n-r_1)!}{r_2!(n-r_1-r_2)!}\cdots\frac{(n-r_1-r_2-\ldots-r_{k-1})!}{r_k!(n-r_1-r_2-\cdots-r_k)!}$$

Example: How many ways can we assign 12 programmers to three projects where Project A needs 3 people, Project B needs 2 people, Project C needs 4 people, and the remaining 3 people are held in reserve.

Partition 12 into 4 parts A, B, C, Reserve

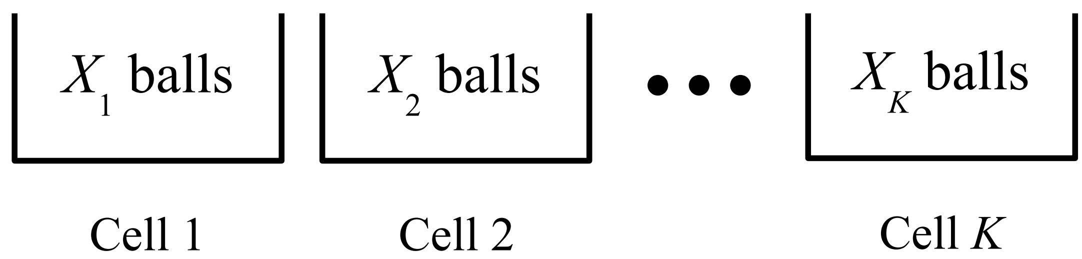
$$\#\ \text{Partitions} = \binom{12}{3\ 2\ 4\ 3} = 277200$$

- Indistinguishable Balls

  Consider the distribution of J balls into K cells (numbered 1 to K).

  – Distinguishable balls: sample space has $K \times K \times \ldots \times K = K^J$ outcomes.

  – Indistinguishable balls: any outcome looks like

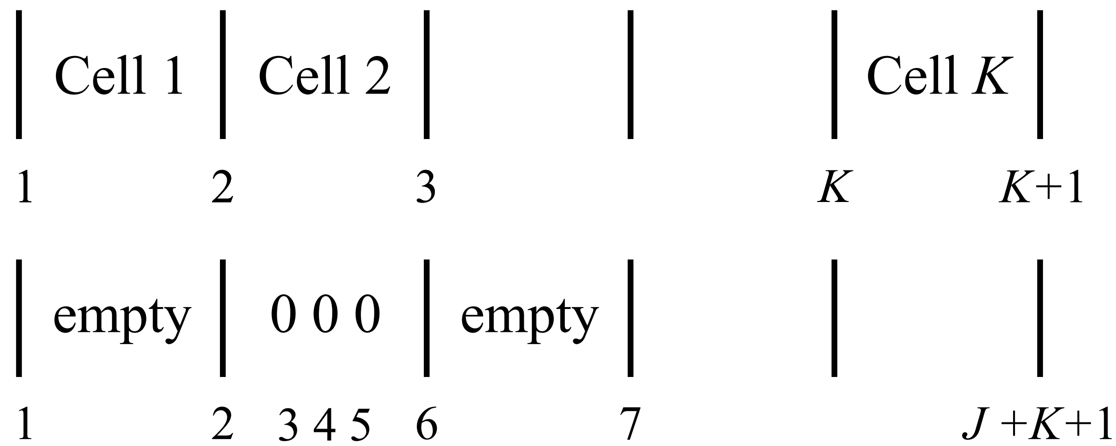


Theorem. *$\#$ of distributions of J indistinguishable balls into K cells is*

$$\#\left\{(x_1, x_2, \ldots, x_K) : x_i \geq 0 \ \&\ \sum x_i = J\right\}$$
$$= \binom{J+K-1}{J} = \binom{J+K-1}{K-1}.$$

Remark: This is the number of terms in the summation of the multinomial theorem.

Proof.



$J + K - 1$ variable positions
with J balls & $K - 1$ walls

Need to decide which of the $J + K - 1$ contain the J balls which is just $\binom{J+K-1}{J}$. $\square$

The number of different sample points in the Mendel example is given by this formula where $J = 560$ and $K = 4$.