STAT:5100 (22S:193) Statistical Inference I

Week 2

Luke Tierney

University of Iowa

Fall 2015

Recap

- Basic definition of probability
- Consequences of the definition
- Finite sample spaces
- Equally likely outcomes
- Basic counting principles

Example (at least one six in four rolls, continued)

- The number of outcomes in the sample space

$$S = \{(x_1, x_2, x_3, x_4) : x_i \in \{1, 2, 3, 4, 5, 6\}\}$$

  is $6 \times 6 \times 6 \times 6 = 6^4 = 1296$.
- We can compute the number of outcomes with at least one six a number of ways.
- The easiest approach is to look at the complementary event of rolling no sixes.
- The number of outcomes containing no sixes is $5 \times 5 \times 5 \times 5 = 5^4 = 625$.
- So the probability of rolling at least one six is

$$P(\text{at least one six}) = 1 - P(\text{no sixes}) = 1 - \left(\frac{5}{6}\right)^4 = 1 - \frac{625}{1296} \approx 0.517.$$

Example (at least one six in four rolls, continued)

A check by simulation:

```
> R <- 4
> N <- 10000
> K <- 6
> rolls <- matrix(sample(1 : K,  R * N, replace = TRUE), N, R)
> phat <- mean(rowSums(rolls == K) > 0)
> phat
[1] 0.5153
```

Simulation standard error:

```
> se <- sqrt(phat * (1 - phat) / N)
> se
[1] 0.004997659
```

Interval estimate:

```
> c(phat - 2 * se, phat + 2 * se)
[1] 0.5053047 0.5252953
```

Ordered Sampling with Replacement

Previously we looked at computing the probability of at least one six in four rolls of a fair die. Another experiment that can be represented by the same sample space:

- A box contains balls numbered $1, 2, \dots, 6$.
- We select a ball at random (which means each ball is equally likely to be selected), record its number, and replace it in the box.
- Repeat three more times.
- This is an example of *ordered sampling with replacement*.

General Case

We can choose an ordered set of r out of n items *with replacement* in

$$\underbrace{n \times n \times \cdots \times n}_{r \text{ terms}} = n^r$$

ways.

Example (roll at least one six, continued)

- Suppose we roll a die r times. The probability of rolling at least one six is then

$$P(\text{at least one six in } r \text{ rolls}) = 1 - \left(\frac{5}{6}\right)^r$$

- What happens as r becomes very large?

$$P(\text{at least one six in } r \text{ rolls}) = 1 - \left(\frac{5}{6}\right)^r \to 1$$

  as $r \to \infty$.

- How many rolls would be need to make this probability at least 0.9?
- We need to find the smallest r such that

$$1 - \left(\frac{5}{6}\right)^r \geq 0.9.$$

- We have equality for

$$r = \frac{\log(1 - 0.9)}{\log(5/6)} \approx 12.629$$

  so the smallest integer satisfying the inequality is $r = 13$.

Example

- There are 35 students registered for a class
- 19 students have last name beginning with letters M–Z.
- Suppose I write student names on pieces of paper, place these in a box, then, one at a time, select and record three names, but do *not* replace the names in the box.
- This is an example of *ordered sampling without replacement*.
- What is the probability that at least one of the selected names starts with M–Z?

Example (continued)

- A possible sample space is all triples of names, with no repetitions.
- The number of outcomes in the sample space is

$$35 \times 34 \times 33 = 39270$$

- Again it is easiest to consider the complementary event that the selection contains no one with a last name starting with M–Z.
- The number of students with names last names that do not start with M–Z is $35 - 19 = 16$.
- So the number of outcomes with no names starting with M–Z is

$$16 \times 15 \times 14 = 3360$$

- The probability that at least one of the three selected students has a last name that starts with M–Z is therefore

$$1 - \frac{3360}{39270} \approx 0.914$$

Permutations: Ordered Sampling Without Replacement

- Suppose we want to choose an ordered list of r out of n items *without replacement*.
- We can do this

$$P_{r,n} = \underbrace{n \times (n-1) \times \cdots \times (n-r+1)}_{r \text{ terms}}$$

ways.

- $P_{r,n}$ is the number of *permutations* of n things taken r at a time.

Definition (Factorials)

For an integer $n > 0$ define

$$n! = n \times (n-1) \times \cdots \times 2 \times 1 = n \times (n-1)!.$$

For $n = 0$ define $0! = 1$.

Some properties:

- $n!$ is the number of ways to choose an order for n items.
- Using factorials we can write $P_{r,n}$ as

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

- This is usually *not* a good computing formula:
  - it is too much work
  - it will have numerical problems
- Computation with factorials is usually done on the logarithm scale using the Gamma function.

Example (names, continued)

- Suppose I do not record the order in which the names are selected
  - For example, I might record them in alphabetical order
- Alternatively, suppose I select three names at once.
- Does this change the probability that at least one last name in the sample starts with M–Z?
  - How I record names clearly does not.
  - Selecting an ordered list can be done by selecting a group and then selecting a random order.
- Either way,
  - our probability would be the same;
  - all possible unordered samples would be equally likely.

Example (continued)

- How many different unordered samples (combinations, subsets) of 3 out of 35 students can be formed?

$$\begin{aligned}\#(\text{ordered samples}) &= \#(\text{unordered samples}) \times \#(\text{orderings})\\ P_{3,35} &= \#(\text{unordered samples}) \times 3!\\ 39270 &= \#(\text{unordered samples}) \times 6\end{aligned}$$

- So

$$\#(\text{unordered samples}) = \frac{P_{3,35}}{3!} = \frac{39270}{6} = 6545.$$

- Similarly,

$$\#(\text{unordered samples with no M–Z names}) = \frac{P_{3,16}}{3!} = \frac{3360}{6} = 560$$

- Our probability is therefore

$$1 - \frac{560}{6545} \approx 0.914$$

as before.

Combinations: Unordered Sampling Without Replacement

- The number of possible unordered samples without replacement (number of subsets, number of *combinations*) of r out of n items is

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

- Note that

$$C_{r,n} = C_{n-r,n}$$
$$\binom{n}{r} = \binom{n}{n-r}$$

- Why?
  - Use algebra, or:
  - Use a counting argument:
    - Each subset of size r has a complement of size $n-r$.
    - So the number of subsets of size r equals the number of subsets of size $n-r$

Example (names, continued)

- Suppose we want to compute the probability that our sample of 3 out of 35 students contains exactly one person with a last name starting with M–Z.
- We can do this based on ordered or unordered sampling.
- The answer will be the same; using unordered sampling is easier.
- Creating a subset of size 3 with exactly one person with last name starting with M–Z can be done in two stages:
  - Choose the person with last name starting with M–Z;
    - there are 19 choices.
  - Choose the set of two people with last names starting with A–L
    - there are $\binom{16}{2} = \frac{16 \times 15}{2} = 120$ choices.
- There are $19 \times 120 = 2280$ such subsets.
- So the probability that our sample of 3 out of 35 students contains exactly one person with a last name starting with M–Z is

$$\frac{19\binom{16}{2}}{\binom{35}{3}} = \frac{2280}{6545} \approx 0.348.$$

A common setting

- Suppose we have a *population* of N items.
- M of these items have property X
- r items are selected at random, without replacement.
- What is the probability that k of the items selected have property X?
- There are $\binom{N}{r}$ possible unordered samples.
- The number of samples containing k items with property X is $\binom{M}{k}\binom{N-M}{r-k}$.
- Assuming all possible samples are equally likely,

$$P(\text{exactly } k \text{ have property X}) = \frac{\binom{M}{k}\binom{N-M}{r-k}}{\binom{N}{r}}$$

- There are many applications, such as
  - survey sampling
  - quality control sampling
  - capture/recapture methods

Example

- Suppose a fair coin is tossed 10 times. What is the probability that four flips are heads (and six flips are tails)?
- One possible sample space is

$$S = \{(x_1 x_2 \dots x_{10}) : x_i \in \{H, T\}\}.$$

- Since the coin is *fair* we can assume that each of the $2^{10} = 1024$ patterns of 10 H and T symbols is equally likely.
- How many of these patters contain four H and six T symbols?

Example (continued)

- Each pattern corresponds uniquely to a subset of four integers out of $1, \ldots, 10$ representing the positions of the H symbols.
- So there are as many patterns containing four H and six T symbols as there are subsets of size four out of a set of size 10:

$$\binom{10}{4} = 210$$

- So the probability of 10 tosses resulting in four heads and six tails is

$$\frac{\binom{10}{4}}{2^{10}} = \frac{210}{1024} \approx 0.2051$$

Recap

- Basic counting principles
- Examples
- Ordered sampling with replacement
- Permutations, ordered sampling without replacement
- Combinations, subsets, unordered sampling without replacement

Example

- Suppose we roll a fair die 10 times. What is the probability that at least three rolls are sixes?
- We can compute this as

$$P(\text{at least three sixes}) = \sum_{k=3}^{10} P(k \text{ sixes}) = 1 - \sum_{k=0}^{2} P(k \text{ sixes})$$

- The sample space has $6^{10} = 60,466,176$ equally likely outcomes.
- An outcome with exactly k sixes can be constructed by
  - choosing the positions of the sixes
  - choosing values from $1, 2, \ldots, 5$ for the $10 - k$ other spots
- So the number of outcomes with exactly k sixes is

$$\binom{10}{k} 5^{10-k}$$

Example

- So the probability of exactly k sixes in 10 rolls is

$$P(k \text{ sixes}) = \binom{10}{k}\frac{5^{10-k}}{6^{10}} = \binom{10}{k}\left(\frac{1}{6}\right)^{k}\left(\frac{5}{6}\right)^{10-k}$$

- This is an example of a *binomial probability*
- The probability of at last three sixes in 10 rolls is therefore

$$1 - \left(\frac{5}{6}\right)^{10} - 10\frac{1}{6}\left(\frac{5}{6}\right)^{9} - 45\left(\frac{1}{6}\right)^{2}\left(\frac{5}{6}\right)^{8} \approx 0.2248$$

Examples

- How many partial derivatives of order r does a function of n variables have?

  For $n = 2$ and $r = 3$ we have four cases:

$$\frac{\partial^3}{\partial x^3}, \frac{\partial^3}{\partial y^3}, \frac{\partial^3}{\partial x^2 \partial y}, \frac{\partial^3}{\partial x \partial y^2}$$

- How many multinomials of degree r in n variables are there?

  For $n = 3$ and $r = 2$ there are six cases:

$$x^2, y^2, z^2, xy, xz, yz$$

Examples

- How many sequences $k_1, \dots, k_n$ of non-negative integers are there such that $k_1 + \cdots + k_n = r$? These are called *weak compositions*.

  For $r = 3$ and $n = 2$ there are four:

  $$(0, 3), (1, 2), (2, 1), (3, 0)$$

- How many sequences $k_1, \dots, k_n$ of positive integers are there such that $k_1 + \cdots + k_n = r$? These are called *compositions*.

  For $r = 5$ and $n = 2$ there are four:

  $$(1, 4), (2, 3), (3, 2), (4, 1)$$

One way to do this:

$$\overbrace{XX}^{1} \mid \overbrace{X}^{2} \mid \cdots \mid \overbrace{XXX}^{n} = 1, 1, 2, \ldots, n, n, n$$

- We need r X's and $n-1$ $|$'s,
- Each pattern corresponds to a unique weak composition.
- For example, for $n = 4$ and $r = 7$

$$\begin{array}{ccccccccccc} 1 & 1 & 1 & & 2 & 2 & & & 4 & 4 \\ X & X & X & \mid & X & X & \mid & \mid & X & X \end{array}$$

- The number of patterns is the number of ways to choose r spots for the X's (or $n-1$ spots for the $|$'s) out of $r+n-1$ positions:

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$$

Examples

- Partial derivatives: For a function of two variables ($n = 2$) there are
  - $r = 2$: $\binom{2+2-1}{2} = \binom{3}{2} = 3$ 2nd order partials.
  - $r = 3$: $\binom{2+3-1}{3} = \binom{4}{3} = 4$ 3rd order partials.
- Multinomials: For a second degree $r = 2$ there are
  - $n = 3$: $\binom{3+2-1}{2} = \binom{4}{2} = 6$ multinomials in 3 variables.
  - $n = 4$: $\binom{4+2-1}{2} = \binom{5}{2} = 10$ multinomials in 4 variables.

Example

- Compositions: If $(k_1, \ldots, k_n)$ is a composition of r, then $(k_1 - 1, \ldots, k_n - 1)$ is a weak composition of $r - n$, and conversely.
- So the number of compositions of size $n = 2$ of the integer 5 is identical to the number of weak compositions of size $n = 2$ of the integer 3.

Unordered sampling with replacement

- The problem of counting the number of partial derivatives or multinomials is sometimes called *unordered sampling with replacement*
- There is no reasonable sampling mechanism that makes these outcomes equally likely.
- If you find yourself using this result in a probability problem you are most likely doing something wrong.

Example (Matching Problem)

- At the beginning of an evening n people place their hats in a cloak room.
- At the end of the evening hats are randomly assigned so all assignments are equally likely.
- What is the chance that no one receives their own hat?
- What is the chance exactly k people receive their own hat?
- Other variations:
  - matching cards
  - guessing in psychic experiments
  - home finding in animal studies

Example (Matching problem, continued)

Simple explorations:

```
> idx <- 1 : 10
> sample(idx)
 [1]  7  5  6  3 10  9  1  4  2  8
> sample(idx) == idx
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
> sum(sample(idx) == idx)
[1] 1
> replicate(20, sum(sample(idx) == idx))
 [1] 2 1 0 1 2 0 3 1 1 0 1 0 1 2 1 1 1 1 1 2
```

Example (Matching problem, continued)

A function to sample the number of matches:

```
rmatch <- function(R, N) {
   idx <- 1 : N
   replicate(R, sum(sample(idx) == idx))
}
```

Exploring the probability of no matches:

```
R <- 10000
n <- 1 : 10
p <- sapply(n, function(N) mean(rmatch(R, N) == 0))
plot(n, p)
se <- sqrt(p * (1 - p) / R)
segments(n, p - 2 * se, n, p + 2 * se)
```

The code for this example is available at

http://www.stat.uiowa.edu/~luke/classes/193/matching.R.

Example (Matching problem, continued)

- Let's focus on computing the probability of no matches.
- One possibility is a direct approach: calculate

$$d_n = \text{number of assignments of } n \text{ hats with no matches}$$

- d_n is the number of *derangements* of $1, \ldots, n$.
- Clearly $d_1 = 0$ and $d_2 = 1$.

Example (Matching problem, continued)

- We can relate d_{n+1} to smaller problems:
  - The first hat chosen can be any of the n hats numbered $2, \ldots, n+1$; suppose it is hat k.
  - Then hat 1 can be in position k or not.
  - If hat 1 is in position k, then there are d_{n-1} assignments of the remaining $n-1$ hats that do not contain any matches.
  - If hat 1 is not in position k then there are d_n assignments of the remaining n hats, including hat 1, that do not contain a match or put hat 1 in position k.
  - This means that d_{n+1} satisfies

$$d_{n+1} = n(d_{n-1} + d_n) = nd_{n-1} + nd_n$$

  - This is a *difference equation*.

Example (Matching problem, continued)

- We can solve this difference equation directly, or first convert to probabilities.
- Let $q_n = P(\text{no matches with } n \text{ hats})$.
- Then $q_n = d_n/n!$ and

$$q_{n+1} = \frac{1}{n+1} q_{n-1} + \frac{n}{n+1} q_n$$

- Subtracting q_n from both sides gives

$$q_{n+1} - q_n = \frac{1}{n+1} q_{n-1} - \frac{1}{n+1} q_n$$

or

$$q_{n+1} - q_n = \frac{-1}{n+1}(q_n - q_{n-1}).$$

Example (Matching problem, continued)

- Repeated application of this identity, together with the initial conditions $q_2 = \frac{1}{2}$ and $q_1 = 0$ gives

$$\begin{aligned}
q_2 - q_1 &= \frac{1}{2} - 0 = \frac{1}{2}\\
q_3 - q_2 &= -\frac{1}{3}(q_2 - q_1) = -\frac{1}{3 \times 2}\\
q_4 - q_3 &= -\frac{1}{4}(q_3 - q_2) = +\frac{1}{4 \times 3 \times 2}\\
q_5 - q_4 &= -\frac{1}{5}(q_4 - q_3) = -\frac{1}{5 \times 4 \times 3 \times 2}
\end{aligned}$$

- This suggests that

$$q_{n+1} - q_n = \frac{(-1)^{n+1}}{(n+1)!}$$

- You can formally verify this using induction.

Example (Matching problem, continued)

- This equation can be written as

$$q_{n+1} = q_n + \frac{(-1)^{n+1}}{(n+1)!}$$

- Repeated application of this, together with $q_1 = 0$ gives

$$q_n = \sum_{k=2}^{n} \frac{(-1)^k}{k!} = 1 + \sum_{k=1}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

- As n increases to infinity $q_n \to \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1} \approx 0.3678794$.
- Convergence is very fast:

```
> cumsum((-1)^(0:10)/factorial(0:10))
 [1] 1.0000000 0.0000000 0.5000000 0.3333333 0.3750000 0.3666667 0.3680556
 [8] 0.3678571 0.3678819 0.3678792 0.3678795
```

Recap

- Weak compositions
- Matching problem

Example (Matching problem, continued)

- An alternative approach that does not require solving difference equations works with the complementary event.
- Let B be the event that no one gets their own hat and let

$$A_i = \text{event that person } i \text{ gets their own hat}$$

- Then

$$\bigcup_{i=1}^{n} A_i = B^c.$$

- The probability we want is

$$q_n = P(B) = 1 - P(B^c) = 1 - P\left(\bigcup_{i=1}^{n} A_i\right).$$

- We can use the *union-intersection formula*.

Union-Intersection Formula

- For any n events $A_1, \ldots, A_n$ let

$$
\begin{aligned}
P_1 &= \sum_{1 \le i \le n} P(A_i) \\
P_2 &= \sum_{1 \le i < j \le n} P(A_i \cap A_j) \\
P_3 &= \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) \\
&\cdots \\
P_n &= P(A_1 \cap A_2 \cap \cdots \cap A_n)
\end{aligned}
$$

- Then

$$
P\left(\bigcup_{i=1}^{n} A_i\right) = P_1 - P_2 + P_3 - \cdots \pm P_n = \sum_{k=1}^{n} (-1)^{k+1} P_k.
$$

- This is sometimes called the *inclusion-exclusion formula*.

Some Notes

- This can be proved by induction.
- The union probability can be bounded as

$$P_1 \geq P\left(\bigcup_{i=1}^{n} A_i\right) \geq P_1 - P_2$$

$$P_1 - P_2 + P_3 \geq P\left(\bigcup_{i=1}^{n} A_i\right) \geq P_1 - P_2 + P_3 - P_4$$

$$\vdots$$

- This is very useful if $P_1, \ldots, P_n$ are easy to compute.

Some Notes

- The union-intersection, or inclusion-exclusion, formula also applies to counts.
- Let $A_1, \dots, A_n$ be subsets of a finite set S, let $\#(B)$ denote the number of elements in a set B, and let

$$\begin{aligned}
N_1 &= \sum_{1 \le i \le n} \#(A_i) \\
N_2 &= \sum_{1 \le i < j \le n} \#(A_i \cap A_j) \\
N_3 &= \sum_{1 \le i < j < k \le n} \#(A_i \cap A_j \cap A_k) \\
&\dots \\
N_n &= \#(A_1 \cap A_2 \cap \cdots \cap A_n)
\end{aligned}$$

- Then

$$\#\left(\bigcup_{i=1}^{n} A_i\right) = N_1 - N_2 + N_3 - \cdots \pm N_n = \sum_{k=1}^{n} (-1)^{k+1} N_k.$$

Example (Matching problem, continued)

- For a particular set $i_1, \ldots, i_k$ of k individuals

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \frac{(n-k)!}{n!} = \frac{1}{P_{k,n}}$$

- There are $\binom{n}{k}$ sets of k individuals.
- So for a particular k

$$P_k = \sum_{i_1,\ldots,i_k} P(A_{i_1} \cap \cdots \cap A_{i_k}) = \binom{n}{k}\frac{(n-k)!}{n!} = \frac{1}{k!}$$

- Therefore

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}(-1)^{k+1}\frac{1}{k!}.$$

Example (Matching problem, continued)

- So

$$
\begin{aligned}
P(B) &= 1 - \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!} \\
&= 1 + \sum_{k=1}^{n} (-1)^{k} \frac{1}{k!} = \sum_{k=0}^{n} (-1)^{k} \frac{1}{k!}
\end{aligned}
$$

- As n increases to infinity,

$$
\begin{aligned}
q_n = P(B) &= \sum_{k=0}^{n} (-1)^{k} \frac{1}{k!} \\
&\to \sum_{k=0}^{\infty} (-1)^{k} \frac{1}{k!} \\
&= e^{-1}
\end{aligned}
$$

Example (Matching problem, continued)

- What is the probability that exactly r individuals get their own hat?
- For a particular set of r individuals $i_1, \dots, i_r$ the number of ways for those individuals and no others to get their own hats is

$$d_{n-r} = (n-r)! q_{n-r} = (n-r)! \left[\sum_{k=0}^{n-r} (-1)^k \frac{1}{k!} \right]$$

- There are $\binom{n}{r}$ sets of r individuals.

Example (continued)

- So the number of ways for exactly r individuals to get their own hats is

$$\binom{n}{r} d_{n-r} = \binom{n}{r}(n-r)!\left[\sum_{k=0}^{n-r}(-1)^k \frac{1}{k!}\right] = \frac{n!}{r!}\left[\sum_{k=0}^{n-r}(-1)^k \frac{1}{k!}\right]$$

- The probability of exactly r individuals getting their own hats is thus

$$\frac{\frac{n!}{r!}\left[\sum_{k=0}^{n-r}(-1)^k \frac{1}{k!}\right]}{n!} = \frac{1}{r!}\left[\sum_{k=0}^{n-r}(-1)^k \frac{1}{k!}\right]$$

- As n increases to infinity this converges to

$$\frac{1}{r!}\left[\sum_{k=0}^{\infty}(-1)^k \frac{1}{k!}\right] = \frac{1}{r!} e^{-1}$$

- This is a *Poisson* probability.

Example (Generalized Birthday Problem)

- An urn contains m balls numbered $1, \dots, m$.
- A sample of size n is drawn with replacement.
- What is the probability that all sampled balls are different?
- What is the probability that the sample contains k distinct balls?
- Equivalently, what is the probability that the sample omits $r = m - k$ balls?
- For $n \geq m$, what is the probability the sample includes every ball at least once?
- The classical birthday problem: what is the probability that no students in a class of n students share the same birthday?
- The general case has applications to *bootstrapping*, where $m = n$.

Example (Generalized Birthday Problem, continued)

- The probability that all sampled balls are distinct is

$$\frac{P_{n,m}}{m^n} = \frac{m!}{(m-n)!m^n}$$

for $n \leq m$ and zero otherwise.

- For a class of $n = 35$ and $m = 365$ the chance that no two students share a birthday is

$$\frac{365!}{(365-35)!365^{35}} \approx 0.1856.$$

Example (Generalized Birthday Problem, continued)

- To find the probability that every ball appears at least once, i.e. that no ball is omitted, let

$$A_i = \{\text{ball } i \text{ is not in the sample}\}$$

- The event that the sample omits no balls is

$$A_1^c \cap \cdots \cap A_m^c = \bigcap_{i=1}^{m} A_i^c = \left(\bigcup_{i=1}^{m} A_i\right)^c$$

- We may be able to compute the union probability with the union-intersection formula.

Example (Generalized Birthday Problem, continued)

- For a particular set of k balls $i_1, \ldots, i_k$

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \frac{(m-k)^n}{m^n}.$$

- There are $\binom{m}{k}$ possible sets of k balls to omit, so by the union-intersection formula the probability that the sample omits at least one ball is

$$P\left(\bigcup_{i=1}^m A_i\right) = \sum_{k=1}^m (-1)^{k+1} \binom{m}{k} \frac{(m-k)^n}{m^n}$$

- The probability that our sample does not omit any balls is

$$\begin{aligned} q_{m,n} = P\left(\bigcap_{i=1}^m A_i^c\right) = 1 - P\left(\bigcup_{i=1}^m A_i\right) &= 1 - \sum_{k=1}^m (-1)^{k+1} \binom{m}{k} \frac{(m-k)^n}{m^n} \\ &= 1 + \sum_{k=1}^m (-1)^{k} \binom{m}{k} \frac{(m-k)^n}{m^n} \\ &= \sum_{k=0}^m (-1)^{k} \binom{m}{k} \frac{(m-k)^n}{m^n} \end{aligned}$$

Example (Generalized Birthday Problem, continued)

- The number of points in the sample space that do not omit any balls is

$$v_{m,n} = m^n q_{m,n} = \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n$$

- The probability of excluding a particular set of r balls and no others is

$$\frac{v_{m-r,n}}{m^n} = \sum_{k=0}^{m-r} (-1)^k \binom{m-r}{k} \frac{(m-r-k)^n}{m^n} = \sum_{k=0}^{m-r} (-1)^k \binom{m-r}{k} \left(1 - \frac{r+k}{m}\right)^n$$

- The probability of excluding exactly r balls is

$$\binom{m}{r} \frac{v_{m-r,n}}{m^n} = \binom{m}{r} \sum_{k=0}^{m-r} (-1)^k \binom{m-r}{k} \left(1 - \frac{r+k}{m}\right)^n$$

Some Notes on Counting Problems

Things to be careful about

- under counting: missing an entire case
- double counting: two or more ways to get the same object

Counting usually involves

- breaking a large problem into smaller ones
- transforming a complex problem into a simpler but equivalent one
  - need to make sure the simpler one is really equivalent, i.e. the relation is one-to-one
  - the simpler ones often correspond to counting permutations or combinations
  - if you find yourself wanting to use "unordered sampling with replacement" you are probably doing something wrong

Countably Infinite Sample Spaces

Example

- Suppose two players, A and B, take turns rolling a die.
  - The first player who rolls a six wins the game.
  - Player A rolls first.
- Some questions:
  - What is the probability that player A wins?
  - What is the probability that player B wins?
  - Does the game end with a winner?
- We can use as a sample space

$$S = \{1, 2, 3, \dots\}$$

or

$$S = \{1, 2, 3, \dots, \infty\}.$$

Example (continued)

- We can start by working out the probability p_k that the first roll to produce a six is roll k:
  - Suppose a die is rolled k times; there are 6^k possible outcomes.
  - The number of outcomes with a six on roll k and no sixes on rolls $1, \ldots, k-1$ is $5^{k-1} \times 1$.
  - So the probability p_k is

$$p_k = \frac{5^{k-1}}{6^k} = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}.$$

- An alternative approach:
  - compute $q_k = P(\text{no sixes in first } k \text{ rolls}) = \left(\frac{5}{6}\right)^k$;
  - then $p_k = q_{k-1} - q_k = \left(\frac{5}{6}\right)^{k-1} - \left(\frac{5}{6}\right)^k = \left(1 - \frac{5}{6}\right)\left(\frac{5}{6}\right)^{k-1}$

Example (continued)

- Let E be the event that the game ends.
- Is it certain that the game will end, i.e. is $P(E) = 1$?
- The sum of the p_k is

$$
\begin{aligned}
\sum_{k=1}^{\infty} p_k &= \sum_{k=1}^{\infty} \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} \\
&= \frac{1}{6}\sum_{k=1}^{\infty}\left(\frac{5}{6}\right)^{k-1} \\
&= \frac{1}{6}\left(\frac{1}{1-5/6}\right) \\
&= 1.
\end{aligned}
$$

Example (continued)

- The finite additivity axiom alone does *not* allow us to conclude that

$$P(E) = P\left(\bigcup_{k=1}^{\infty}\{\text{first six is on roll } k\}\right) = \sum_{k=1}^{\infty} p_k.$$

- But since $\sum_{k=1}^{\infty} p_k = 1$ we can argue this way:
  - Let E_n be the event that the game ends in n or fewer rolls.
  - Then for all $n \geq 1$

$$1 \geq P(E) \geq P(E_n).$$

  - Furthermore,

$$P(E_n) = \sum_{k=1}^{n} p_k \to 1$$

  - This implies $P(E) \geq 1$ as well, and so $P(E) = 1$.