Ma 3/103
Introduction to Probability and Statistics

KC Border
Winter 2017

Lecture 3: Learning to count; Binomial Distribution

Relevant textbook passages:

Pitman [10]: Sections 1.5–1.6, pp. 47–77; Appendix 1, pp. 507–514.

Larsen–Marx [9]: Sections 2.4, 2.5, 2.6, 2.7, pp. 67–101. Sections

3.1 Examples of counting and probability

To calculate the probability of the event E, when the experimental outcomes are all equally likely, simply count the number of outcomes that belong to E and divide by the total number of outcomes in the outcome space S.

3.1.1 How many different outcomes are there for the experiment of tossing a coin n times?

$$2^n$$

3.1.2 Binomial probabilities

What is the probability of getting k heads in n independent tosses of a fair coin?

Let's do this carefully. The sample space S is the set of sequences of length n where each term s_i in the sequence is H or T. For each point $s \in S$, let $A_s = \{i : s_i = H\}$. Since there are only two outcomes, if you know A_s, you know s and vice versa. Now let E be any subset of S that has exactly k elements. There is exactly one point $s \in S$ such that $A_s = E$. Thus the number of elements of S such that $|A_s| = k$ is precisely the same as the number of subsets of S of size k, namely $\binom{n}{k}$. Thus

$$\text{Prob(exactly } k \text{ Heads)} = \frac{\big|\{s \in S : |A_s| = k\}\big|}{|S|} = \frac{\binom{n}{k}}{2^n} = \frac{n!}{k!(n-k)!2^n}.$$

Here is an example with $n = 3$:

s	A_s
HHH	$\{1,2,3\}$
HHT	$\{1,2\}$
HTH	$\{1,3\}$
HTT	$\{1\}$
THH	$\{2,3\}$
THT	$\{2\}$
TTH	$\{3\}$
TTT	$\varnothing$

For $k = 2$, the set of points $s \in S$ with exactly two heads is the set $\{HHT, HTH, THH\}$, which has $3 = \binom{3}{2}$ elements, and probability $3/8$.

We can use Pascal's Triangle to write down these probabilities.

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & \frac{1}{2} & & \frac{1}{2} & & & \\ & & \frac{1}{4} & & \frac{2}{4} & & \frac{1}{4} & & \\ & \frac{1}{8} & & \frac{3}{8} & & \frac{3}{8} & & \frac{1}{8} & \\ \frac{1}{16} & & \frac{4}{16} & & \frac{6}{16} & & \frac{4}{16} & & \frac{1}{16} \\ & & & & \text{etc.} & & & & \end{array}$$

(Prob of 0 Heads in 0 tosses)
(Prob of 0, 1 Heads in 1 toss)
(Prob of 0, 1, 2 Heads in 2 tosses)
etc.

3.1.3 How many ways can a standard deck of 52 cards be arranged?

Here the order matters, so we want to the number of lists, which is

$$52! \approx 8.06582 \times 10^{67}$$

or more precisely:

80,658,175,170,943,878,571,660,636,856,403,766,975,289,505,440,883,277,824,000,000,000,000.

This is an astronomically large number. In fact, since the universe is about 10–20 billion years old and there are about 7.28 billion people (according to Siri), if every person on the planet were set to work arranging a deck in a given order, and could do so in one second, it would take about 2×10^{40} lifetimes (to date) of the universe to go through all the possible arrangements of the deck. For this course we make the usual assumption that after shuffling the deck a few times all possible arrangements are equally likely. This is ludicrous. But the assumption may actually give reasonably good results for typical questions we ask about card games, such as those that follow. (For more about the distribution of cards after shuffling, see the papers by Bayer and Diaconis [3] and Assaf, Diaconis, and Soundararajan [2]. A rule of thumb is that it takes at least seven riffle shuffles for the deck to sufficiently mixed up to be able to use the model that all orders of the cards are equally likely—provided what yo want to do to calculate the probabilities of events typically associated with card games.)

3.1.4 How many different five-card draw poker hands are there?

In standard five-card draw poker, each player is dealt five cards before any betting occurs. The order in which you receive your cards does not matter for how you will bet, only the set of cards in your hand. In various forms of stud poker, there is betting before you receive all their cards, so the order in which you receive cards may influence your bets. There are

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2,598,960$$

possible five-card hands.

3.1.5 How many different deals?

How many distinct *deals* of five-card draw poker hands for seven players are there? (The order of hands matters to the betting, but the order of cards within hands does not.)

The number of distinct deals is

$$\underbrace{\binom{52}{5}\binom{47}{5}\binom{42}{5}\binom{37}{5}\binom{32}{5}\binom{27}{5}\binom{22}{5}}_{\text{7 terms}} \approx 6.3 \times 10^{38}.$$

Each succeeding hand has five fewer cards to choose from, the others being used by the earlier hands.

3.1.6 How many five-card poker hands are flushes?

To get a *flush* all five cards must be of the same suit. There are thirteen ranks in each suit, so there $\binom{13}{5}$ distinct flushes from a given suit. There are four suits, so there

$$4\binom{13}{5} = 5148 \text{ possible flushes.}$$

(This includes straight flushes.)

So what is the **probability** of a flush?

$$\frac{4\binom{13}{5}}{\binom{52}{5}} = \frac{5148}{2,598,960} \approx .002$$

3.1.7 Deals in bridge

In Contract Bridge, all fifty-two cards are dealt out to four players, so each has thirteen. The first player can have any one of $\binom{52}{13}$ hands, so the second may have any of $\binom{39}{13}$ hands, the third may have any of $\binom{26}{13}$ hands, and the last player is stuck with the $\binom{13}{13} = 1$ hand left over.

Thus there are

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} \approx 5.36447 \times 10^{28}.$$

distinct *deals* in bridge.

3.1.8 Splits in bridge

Suppose your opponents have n Clubs between them. What is the probability that they are split k–$(n-k)$ between West and East?

This is the probability that West (the player on your left) has k of the n. East will have the remaining $n-k$.

There are $\binom{26}{13} = 10,400,600$ possible hands for West. In order for West's hand to have k Clubs, they[1] must have one of the $\binom{n}{k}$ subsets of size k from the n Clubs. The remaining $13-k$ must be made up from the $26-n$ non-Clubs. There are $\binom{26-n}{13-k}$ possibilities. Thus there are

$$\binom{n}{k}\binom{26-n}{13-k}$$

hands in which West has k clubs, so the probability is

$$\frac{\binom{n}{k}\binom{26-n}{13-k}}{\binom{26}{13}}$$

that West has k clubs.

For the case $n = 3$ this is $11/100$ for $k = 0, 3$, and $39/100$ for $k = 1, 2$.

3.2 Bernoulli Trials

A **Bernoulli trial** is a random experiment with two possible outcomes, traditionally labeled "success" and "failure." The probability of success is traditionally denoted p. The probability of failure $(1-p)$ is often denoted q. A **Bernoulli random variable** is simply the indicator of success in a Bernoulli trial. That is,

Pitman [10]: p. 27

$$X = \begin{cases} 1 & \text{if the trial is a success} \\ 0 & \text{if the trial is a failure.} \end{cases}$$

3.3 The Binomial Distribution

If there are n stochastically independent Bernoulli trials with the same probability p of success, the probability distribution of the number of successes is called the **Binomial distribution**. A **Binomial random variable** is simply the count of the number of successes in n trials. To get exactly k successes, there must be $n-k$ failures. There are

Pitman [10]: § 2.1

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

such outcomes (where by convention $0! = 1$), and by independence each has probability $p^k(1-p)^{n-k}$. **N.B. In this case, since success and failure are not equally likely, so the points in the sample space are not equally likely. Simple counting is not going to be adequate.** Thus

$$P\,(k \text{ successes in } n \text{ independent Bernoulli trials}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Another way to write this is in terms of the binomial random variable X that counts success in n trials:

$$P\,(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Note that the Binomial random variable is simply the sum of the Bernoulli random variables for each trial. Compare this to the analysis in Subsection 3.1.2, and note that it agrees because $1/2^n = (1/2)^k (1/2)^{n-k}$.

Since $p + (1-p) = 1$ and $1^n = 1$, the Binomial Theorem assures us that the binomial distribution is a probability distribution.

3.3.1 Example (The probability of n heads in $2n$ coin flips) For a fair coin the probability of n heads in $2n$ coin flips is

$$\binom{2n}{n} \left(\frac{1}{2}\right)^{2n}.$$

We can see what happens to this for large n by using Stirling's approximation:

3.3.2 Proposition (Stirling's formula)

$$n! = e^{-n} n^n \sqrt{2\pi n}(1 + \varepsilon_n)$$

where $\varepsilon_n \to 0$ as $n \to \infty$.

[1] Some pedants will claim that the use of *they* or *their* as an ungendered singular pronoun is a grammatical error. There is a convincing argument that those pedants are wrong. See, for instance, Huddleston and Pullum [8, pp. 103–105]. Moreover there is a great need for an ungendered singular pronoun, so I will use *they* in that role.

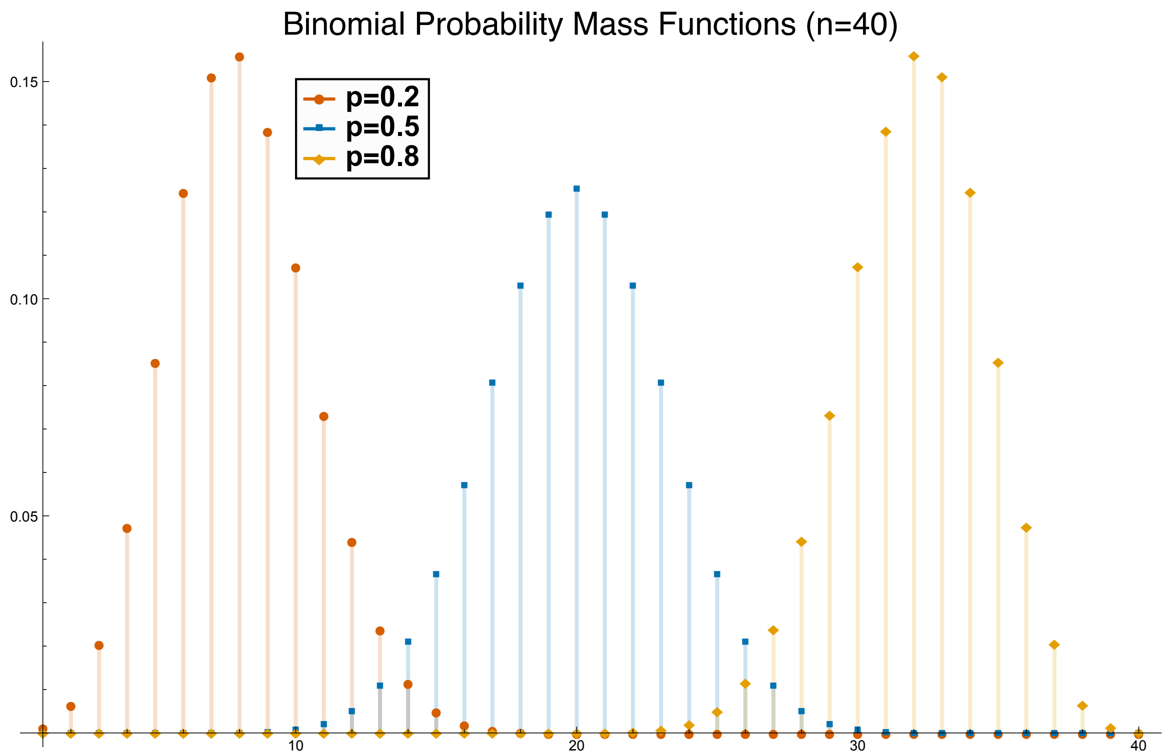

Figure 3.1. Binomial Probabilities

For a proof, see, e.g., Robbins [11], Feller [7, p. 52] or [5, 6] or Ash [1, pp.43–45], or Diaconis and Freedman [4], or the exercises in Pitman [10, p. 136].

Thus we may write

$$\frac{(2n)!}{n!\,n!} = \frac{e^{-2n}(2n)^{2n}\sqrt{4\pi n}}{e^{-n}e^{-n}n^n n^n\sqrt{2\pi n}\sqrt{2\pi n}}(1+\delta_n) = \frac{2^{2n}}{\sqrt{\pi n}}(1+\delta_n),$$

where $\delta_n \to 0$ as $n \to \infty$.

So the probability of n heads in $2n$ attempts is

$$\begin{aligned} &\frac{2^{2n}}{\sqrt{\pi n}}2^{-2n}(1+\delta_n) \\ =\ &\frac{1}{\sqrt{\pi n}}(1+\delta_n) \\ \longrightarrow\ &0 \end{aligned}$$

as $n \to \infty$.

What about the probability of between $n-k$ and $n+k$ heads in $2n$ tosses? Well the probability of getting j heads in $2n$ tosses is $\binom{2n}{j}(1/2)^{2n}$, and this is maximized at $j = n$ (See, e.g., Pitman [10, p. 86].) So we can use this as an upper bound. Thus for $k \geqslant 1$

$$\begin{aligned} P\,(\text{between } n-k \text{ and } n+k \text{ heads}) &< \frac{2k+1}{\sqrt{\pi n}}(1+\delta_n) \\ &\longrightarrow 0 \end{aligned}$$

as $n \to \infty$.

So any reasonable "law of averages" will have to let k grow with n. We will come to this in a few more lectures.

□

3.4 The Multinomial Distribution

Larsen–Marx [9]: Section 10.2, pp. 494–499
Pitman [10]: p. 155

The **multinomial distribution** generalizes the binomial distribution to random experiments with more than two "types" of outcomes. If there are m possible outcome types and the i^{th} type has probability p_i, then in n independent trials, if $k_1 + \cdots + k_m = n$,

$$P\left(k_i \text{ outcomes of type } i,\ i = 1, \ldots, m\right) = \frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_m!}\, p_1^{k_1} \cdot p_2^{k_2} \cdot \cdots \cdot p_m^{k_m}.$$

3.4.1 Remark If you find the above claim puzzling, this may help. Recall that in Subsection 3.1.2 we looked at the number of sets of size k and showed that there was a one-to-one correspondence between sets of size k and points in the sample space with exactly k heads. The same sort of reasoning shows that there is a one-to-one correspondence between partitions of the set of trials, $\{1, \ldots, n\}$, into m sets $E_1, \ldots, E_m$ with $|E_i| = k_i$ for each i and the set of points s in the sample space where there are k_i outcomes of type i for each $i = 1, \ldots, m$. Each such sample point has probability $p_1^{k_1} p_1^{k_2} \cdots p_m^{k^m}$. How many are there?

Well there are $\binom{n}{n-k_1}$ sets of trials of size k_1. But now we have to chose a set of size k_2 from the remaining $n - k_1$ trials, so there are $\binom{n-k_1}{k_2}$ ways to do this for each of the $\binom{n}{n-k_1}$ choices we made earlier. Now we have to choose a set of k_3 trials from the remaining $n - k_1 - k_2$ trials, etc. The total number of possible partitions of the set of trials is thus

$$\binom{n}{k_1} \times \binom{n-k_1}{k_2} \times \binom{n-k_1-k_2}{k_3} \times \cdots \times \binom{n-k_1-k_2-\cdots-k_{m-1}}{k_m}.$$

Expanding this gives

$$\frac{n!}{k_1!(n-k_1)!} \times \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \times \frac{(n-k_1-k_2)!}{k_3!(n-k_1-k_2-k_3)!} \times \cdots \times \frac{(n-k_1-k_2-\cdots-k_{m-1})!}{k_m!\underbrace{(n-k_1-k_2-\cdots-k_{m-1}-k_m)}_{=0}!}.$$

Now observe that the second term in each denominator cancels the numerator in the next fraction, and (recalling that $0! = 1$) we are left with

$$\frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_m!}$$

points $s \in S$, each of which has probability $p_1^{k_1} \cdot p_2^{k_2} \cdot \cdots \cdot p_m^{k_m}$.

We can use random vectors to describe what is happening. For each type $i = 1, \ldots, m$, let X_i denote the number of outcomes of type i. Then the random vector $\boldsymbol{X} = (X_1, \ldots, X_m)$ has a distribution given by

$$P\left(\boldsymbol{X} = (k_1, \ldots, k_m)\right) = \frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_m!}\, p_1^{k_1} \cdot p_2^{k_2} \cdot \cdots \cdot p_m^{k_m}.$$

3.4.2 Example Suppose you roll 9 dice. What is the probability of getting 3 aces (ones) and 6 boxcars (sixes)?

$$\frac{9!}{3!\,0!\,0!\,0!\,0!\,6!}\left(\frac{1}{6}\right)^9 = 94\frac{1}{10,077,696} \approx 0.0000083.$$

(Recall that $0! = 1$.) □

Figure 3.2. The archetypal urn.

3.5 Sampling with and without replacement

Suppose you have an urn holding N balls, of which B are black and the remaining $W = N - B$ are white. If the urn is sufficiently well churned, the probability of drawing a black ball is simply B/N. Now think of drawing a sample of size $n \leqslant N$ from this underlying population, and ask what the probability of the composition of the sample is.

3.5.1 Sampling without replacement

Sampling without replacement means that a ball is drawn from the urn and set aside. The next time a ball is drawn from the urn, the composition of the balls has changed, so the probabilities have changed as well.

For $b \leqslant n$, what is the probability that exactly b of the balls are black, and $w = n - b$? are white?

Let's dispose of some obvious cases. In order to have b black and w white balls, we must have

$$b \leqslant \min\{B, n\} \quad \text{and} \quad w \leqslant \min\{W, n\}.$$

There are $\binom{B}{b}$ sets of size b of black balls and $\binom{W}{w}$ sets of size w of white balls. Thus there are $\binom{B}{b}\binom{W}{w}$ possible ways to get exactly b black balls and w white balls in a sample of size $n = w + b$, out of $\binom{N}{n}$ possible samples of size n. Thus

$$P\left(b \text{ black } \& \ w \text{ white}\right) = \frac{\binom{B}{b}\binom{W}{w}}{\binom{N}{n}}.$$

Note that if $b > B$ or $w > W$, by convention $\binom{B}{b} = \binom{W}{w} = 0$ (there are no subsets of size b of a set of size $B < b$), so this formula works even in this case.

These probabilities are known as the **hypergeometric distribution**.

3.5.2 Sampling with replacement

Sampling with replacement means that after a ball is drawn from the urn, it is returned, and the balls are mixed well enough so that each is equally likely. Thus repeated draws are independent and the probabilities are the same for each draw.

What is the probability that sample consists of b black and w white balls? This is just the binomial probability

$$P\left(b \text{ black } \& \ w \text{ white}\right) = \binom{n}{b}\left(\frac{B}{N}\right)^b\left(\frac{W}{N}\right)^w.$$

3.5.3 Comparing the two sampling methods

Intuition here can be confusing, since without replacement every black ball drawn reduces the pool of black balls making it less likely to get another black ball relative sampling with replacement, but every white ball drawn makes more likely to get a black ball. On balance you might think that sampling without replacement favors a sample more like the underlying population.

To compare the probabilities of sampling without replacement to those with replacement, we can rewrite the hypergeometric probabilities to make them look more like the binomial probabilities as follows.

$$P\left(\text{exactly } b \text{ balls out of } n \text{ are black}\right) = \frac{\binom{B}{b}\binom{W}{w}}{\binom{N}{n}} = \frac{\frac{B!}{b!(B-b)!}\frac{W!}{w!(W-w)!}}{\frac{N!}{n!(N-n)!}} = \frac{\frac{B!}{(B-b)!}\frac{W!}{(W-w)!}}{\frac{N!}{(N-n)!}}\frac{n!}{b!w!},$$

or in terms of the "order notation" (Subsection 2.10.3) we have

$$\begin{aligned} P\left(b \text{ black } \& \ w \text{ white}\right) &= \binom{n}{b}\frac{(B)_b\,(W)_w}{(N_n)} \\ &= \binom{n}{b}\frac{\overbrace{B \times (B-1) \times \cdots \times (B-b+1)}^{b \text{ terms}} \times \overbrace{W \times (W-1) \times \cdots \times (W-w+1)}^{w \text{ terms}}}{\underbrace{N \times (N-1) \times \cdots \times (N-n+1)}_{n \text{ terms}}} \end{aligned}$$

for sampling without replacement versus

$$P\left(b \text{ black } \& \ w \text{ white}\right) = \binom{n}{b}\left(\frac{B}{N}\right)^b\left(\frac{W}{N}\right)^w = \binom{n}{b}\frac{\overbrace{B \times \ldots B}^{b \text{ terms}} \times \overbrace{W \times \ldots W}^{w \text{ terms}}}{\underbrace{N \times \cdots \times N}_{n = b+w \text{ terms}}}.$$

for sampling with replacement.

The ratio of the probability without replacement to the probability with replacement can be written as

$$\frac{B}{B} \times \frac{B-1}{B} \times \cdots \times \frac{B-b+1}{B} \times \frac{W}{W} \times \frac{W-1}{W} \times \cdots \times \frac{W-w+1}{W} \times \frac{N}{N} \times \frac{N}{N-1} \times \cdots \times \frac{N}{N-n+1}.$$

If $b = 0$, the terms involving B do not appear, and similarly for $w = 0$. Whether this ratio is greater or less than one is not obvious. But if we increase N keeping B/N (and hence W/N) constant, then holding the sample size n fixed, each term in this ratio converges to 1 for each b. Therefore the ratio converges to one.

> That is, the difference between sampling with and without replacement holding the sample size constant becomes insignificant as N gets large, holding B/N and W/N fixed.

But how big is big enough? The only time sampling with replacement makes a difference is when the same ball is chosen more than once. The probability that all n balls are distinct is $\frac{N}{N} \times \frac{N-1}{N} \times \cdots \times \frac{N-n+1}{N}$, so the complementary probability (of a duplicate) is $1 - \prod_{k=0}^{n}\big(1-(1/N)\big)$.

Now us the Taylor series approximation that $\ln(1-x) \approx -x$, to get that $\log P\,(\text{duplicate}) \approx -\sum_{k=0}^{n} k/N = -n(n+1)/2N$. (The probability is less than one, so its logarithm is negative.) So as Pitman [10, p. 125] asserts, if $n \ll \sqrt{N}$, this probability is very small. With modern software, you can see for yourself how the two sampling methods compare. See Table 3.1 for a modest example of results calculated by Mathematica 11.

b	Without Replacement	With Replacement	Ratio
0	0.33048	0.34868	0.94780
1	0.40800	0.38742	1.0531
2	0.20151	0.19371	1.0403
3	0.051794	0.057396	0.90240
4	0.0075532	0.011160	0.67680
5	0.00063980	0.0014880	0.42997
6	0.000030998	0.00013778	0.22498
7	8.1440×10^{-7}	8.7480×10^{-6}	0.093096
8	1.0411×10^{-8}	3.6450×10^{-7}	0.028564
9	5.1992×10^{-11}	9.0000×10^{-9}	0.0057769
10	5.7769×10^{-14}	1.0000×10^{-10}	0.00057769

Probability of b black balls in a sample of size $n = 10$ for $N = 100$, $B = 10$, $W = 90$.

b	Without Replacement	With Replacement	Ratio
0	0.34850	0.34868	0.99950
1	0.38761	0.38742	1.0005
2	0.19379	0.19371	1.0004
3	0.057348	0.057396	0.99917
4	0.011125	0.011160	0.99683
5	0.0014782	0.0014880	0.99340
6	0.00013625	0.00013778	0.98887
7	8.6016×10^{-6}	8.7480×10^{-6}	0.98326
8	3.5597×10^{-7}	3.6450×10^{-7}	0.97660
9	8.7200×10^{-9}	9.0000×10^{-9}	0.96889
10	9.6017×10^{-11}	1.0000×10^{-10}	0.96017

Probability of b black balls in a sample of size $N = 10{,}000$, $B = 1000$, $W = 9000$.

Table 3.1. Sampling without replacement vs. sampling with replacement.

3.6 Matching

There are n consecutively numbered balls and n consecutively numbered bins. The balls are arranged in the bins (one ball per bin) at random (all arrangements are equally likely). What is the probability that at least one ball matches its bin? (See Exercise 28 on page 135 of Pitman [10].)

Intuition is not a lot of help here for understanding what happens for large n. When n is large, there is only a small chance that any given ball matches, but there are a lot of them, so one could imagine that the probability could converge to zero, or to one, or perhaps something in between.

Let A_i denote the event that Ball i is placed in Bin i. We want to compute the probability of $\bigcup_{i=1}^{n} A_i$. This looks like it might be a job for the Inclusion–Exclusion Principle, since these events are not disjoint. Recall that it asserts that

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) = & \sum_i p(A_i) \\ & - \sum_{i<j} P(A_i A_j) \\ & + \sum_{i<j<k} P(A_i A_j A_k) \\ & \vdots \\ & + (-1)^k \sum_{i_1<i_2<\cdots<i_k} P(A_{i_1} A_{i_2} \cdots A_{i_k}) \\ & \vdots \\ & + (-1)^{n+1} P(A_1 A_2 \ldots A_n). \end{aligned}$$

Consider the intersection $A_{i_1} A_{i_2} \cdots A_{i_k}$, where $i_1 < i_2 < \cdots < i_k$. In order for this event to occur, ball i_j must be in bin i_j for $j = 1, \ldots, k$. This leaves $n-k$ balls unrestricted, so there are $(n-k)!$ arrangements in this event. And there are $n!$ total arrangements. Thus

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = \frac{(n-k)!}{n!}.$$

Note that this depends only on k. Now there are $\binom{n}{k}$ size-k sets of balls. Thus the k term in the formula above satisfies

$$\sum_{i_1<i_2<\cdots<i_k} P(A_{i_1} A_{i_2} \cdots A_{i_k}) = \binom{n}{k} \frac{(n-k)!}{n!}.$$

Therefore the Inclusion–Exclusion Principle reduces to

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

Here are the values for $n = 1, \ldots, 10$:

n:	Prob(match)
1:	1
2:	$\frac{1}{2} = 0.5$
3:	$\frac{2}{3} \approx 0.666667$
4:	$\frac{5}{8} = 0.625$
5:	$\frac{19}{30} \approx 0.633333$
6:	$\frac{91}{144} \approx 0.631944$
7:	$\frac{177}{280} \approx 0.632143$
8:	$\frac{3641}{5760} \approx 0.632118$
9:	$\frac{28673}{45360} \approx 0.632121$
10:	$\frac{28319}{44800} \approx 0.632121$

Notice that the results converge fairly rapidly, but to what? The answer is $\sum_{k=1}^{\infty}(-1)^{k+1}\frac{1}{k!}$, which you may recognize as $1-(1/e)$. (See the supplementary notes on series.)

3.7 Waiting: The Negative Binomial Distribution

Pitman [10]: p. 213
Larsen–Marx [9]: § 4.5

The **Negative Binomial Distribution** is the probability distribution of the number of independent trials need for a given number of heads. What is the probability that the r^{th} success occurs on trial t, for $t \geqslant r$?

For this to happen, there must be $t-r$ failures and $r-1$ successes in the first $t-1$ trials, with a success on trial t. By independence, this happens with the binomial probability for $r-1$ successes on $t-1$ trials times the probability p of success on trial t:

$$\text{NB}(t;r,p) = \binom{t-1}{r-1}p^{r-1}(1-p)^{t-1-(r-1)} \times p = \binom{t-1}{r-1}p^{r}(1-p)^{t-r} \qquad (t \geqslant r).$$

Of course, the probability is 0 for $t < r$. The special case $r=1$ (number of trials to the first success) is called the **Geometric Distribution**.

Warning: The definition of the negative binomial distribution here is the same as the one in Pitman [10, p. 213] and Larsen–Marx [9, p. 262]. Both Mathematica and R use a different definition. They define it to be the distribution of the number of failures that occurs before the r^{th} success. That is, Mathematica's `PDF[NegativeBinomialDistribution[r, p], t]` is our $\text{NB}(t+r;r,p)$. Mathematica and R's definition assigns positive probability to 0, ours does not.

Bibliography

[1] R. B. Ash. 2008. *Basic probability theory*. Mineola, New York: Dover. Reprint of the 1970 edition published by John Wiley and Sons.

[2] S. Assaf, P. Diaconis, and K. Soundararajan. 2011. A rule of thumb for riffle shuffling. *Annals of Applied Probability* 21(3):843–875. http://www.jstor.org/stable/23033357

[3] D. Bayer and P. Diaconis. 1992. Trailing the dovetail shuffle to its lair. *Annals of Applied Probability* 2(2):294–313. http://www.jstor.org/stable/2959752

[4] P. Diaconis and D. Freedman. 1986. An elementary proof of Stirling's formula. *American Mathematical Monthly* 93(2):123–125. http://www.jstor.org/stable/2322709

[5] W. Feller. 1967. A direct proof of Stirling's formula. *American Mathematical Monthly* 74(10):1223–1225. http://www.jstor.org/stable/2315671

[6] ———. 1968. Correction to "A direct proof of Stirling's formula". *American Mathematical Monthly* 75(5):518. http://www.jstor.org/stable/2314719

[7] ———. 1968. *An introduction to probability theory and its applications*, 3d. ed., volume 1. New York: Wiley.

[8] R. Huddleston and G. K. Pullum. 2005. *A student's introduction to English grammar*. Cambridge: Cambridge University Press.

[9] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[10] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[11] H. Robbins. 1955. A remark on Stirling's formula. *American Mathematical Monthly* 62(1):26–29. http://www.jstor.org/stable/2308012