Part IA - Probability

Lectured by R. Weber

Lent 2015

Basic concepts
Classical probability, equally likely outcomes. Combinatorial analysis, permutations and combinations. Stirling's formula (asymptotics for $\log n!$ proved). [3]

Axiomatic approach
Axioms (countable case). Probability spaces. Inclusion-exclusion formula. Continuity and subadditivity of probability measures. Independence. Binomial, Poisson and geometric distributions. Relation between Poisson and binomial distributions. Conditional probability, Bayes's formula. Examples, including Simpson's paradox. [5]

Discrete random variables
Expectation. Functions of a random variable, indicator function, variance, standard deviation. Covariance, independence of random variables. Generating functions: sums of independent random variables, random sum formula, moments.

Conditional expectation. Random walks: gambler's ruin, recurrence relations. Difference equations and their solution. Mean time to absorption. Branching processes: generating functions and extinction probability. Combinatorial applications of generating functions. [7]

Continuous random variables
Distributions and density functions. Expectations; expectation of a function of a random variable. Uniform, normal and exponential random variables. Memoryless property of exponential distribution. Joint distributions: transformation of random variables (including Jacobians), examples. Simulation: generating continuous random variables, independent normal random variables. Geometrical probability: Bertrand's paradox, Buffon's needle. Correlation coefficient, bivariate normal random variables. [6]

Inequalities and limits
Markov's inequality, Chebyshev's inequality. Weak law of large numbers. Convexity: Jensen's inequality for general random variables, AM/GM inequality.

Moment generating functions and statement (no proof) of continuity theorem. Statement of central limit theorem and sketch of proof. Examples, including sampling. [3]

Contents

0 Introduction **3**

1 Classical probability **4**
- 1.1 Classical probability 4
- 1.2 Sample space and events 4

2 Combinatorial analysis **6**
- 2.1 Counting 6
- 2.2 Sampling with or without replacement 6
- 2.3 Sampling with or without regard to ordering 7
- 2.4 Four cases of enumerative combinatorics 7

3 Stirling's formula **8**
- 3.1 Multinomial coefficient 8
- 3.2 Stirling's formula 8

4 Axiomatic approach **12**
- 4.1 Boole's inequality 13
- 4.2 Inclusion-exclusion formula 14

0 Introduction

In every day life, we often encounter the use of the term probability, and they are used in many different ways. For example, we can hear people say:

(i) The probability that a fair coin will land heads is $1/2$.

(ii) The probability that a selection of 6 members wins the National Lottery Lotto jackpot is 1 in $\binom{19}{6} = 13983816$ or 7.15112×10^{-8}.

(iii) The probability that a drawing pin will land 'point up' is 0.62.

(iv) The probabilty that a large earthquake will occur on the San Andreas Fault in the next 30 years is about 21%

(v) The probability that humanity will be extinct by 2100 is about 50%

The first two cases are things derived from logic. We know that the coin either lands heads or tails. By definition, a fair coin is equally likely to land heads or tail. So the probability of either must be $1/2$.

The third is something probably derived from experiments. Perhaps we did 1000 experiments and 620 of the pins point up. The fourth and fifth examples belong to yet another category that talks about are beliefs and predictions.

We call the first kind "classical probability", the second kind "frequentist probability" and the last "subjective probability". In this course, we only consider classical probability.

1 Classical probability

1.1 Classical probability

Definition (Classical probability)**.** *Classical probability* applies in a situation when there are a finite number of equally likely outcome.

Consider the problem of points

A and B play a game in which they keep throwing coins. If a head lands, then A gets a point. Otherwise, B gets a point. The first person to get 10 points wins a prize.

Now suppose A has got 8 points and B has got 7. The game has to end because an earthquake struck. How should they divide the prize? We answer this by finding the probability of A winning. Someone must have won by the end of 19 rounds, i.e. after 4 more rounds. If A wins at least 2 of them, then A wins.

The number of ways this can happen is $\binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 11$. So A should get $11/16$ of the prize.

1.2 Sample space and events

Consider an experiment that has a random outcome.

Definition (Sample space)**.** The set of all possible outcomes is the *sample space*, Ω. We can lists the outcomes as $\omega_1, \omega_2, \cdots \in \Omega$. Each $\omega \in \Omega$ is an *outcome*.

Definition (Event)**.** A subset of Ω is called an *event*.

Definition (Set notations)**.** Given any two events $A, B \subseteq \Omega$,

- The *complement* of A is $A^C = A' = \bar{A} = \omega \setminus A$.
- "A or B" is the set $A \cup B$.
- "A and B" is the set $A \cap B$.
- A and B are *mutually exclusive* or *disjoint* if $A \cap B = \emptyset$.
- $A \subseteq B$ means $A \Rightarrow B$.

Definition (Probability)**.** Suppose $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$. Let $A \subseteq \Omega$ be an event. Then the *probability* of A is

$$P(A) = \frac{\text{Number of outcomes in } A}{\text{Number of outcomes in } \omega} = \frac{|A|}{N}$$

Example. Suppose r digits are drawn at random from a table of random digits from 0 to 9. What is the probability of

(i) No digit exceeds k

(ii) The largest digit drawn is k

The sample space is $\Omega = \{(a_1, a_2, \cdots, a_r) : 0 \leq a_i \leq 9\}$. Then $|\Omega| = 10^r$.

Let A_k = [no digit exceeds k] $= \{(a_1, \cdots, a_r) : 0 \leq a_i \leq k\}$. Then $|A_k| = (k+1)^r$. So $P(A_k) = (k+1)^r/10^r$.

Now let B_k = [largest digit drawn is k]. We can find this by finding all outcomes in which no digits exceed k, and subtract it by the number of outcomes in which no digit exceeds $k-1$. So $|B_k| = |A_k| - |A_{k-1}|$ and $P(B_k) = [(k+1)^r - k^r]/10^r$.

2 Combinatorial analysis

2.1 Counting

Example. A menu has 6 starters, 7 mains and 6 desserts. How many meals are there? Clearly $6 \times 7 \times 6 = 252$.

This is the fundamental rule of counting:

Theorem (Fundamental rule of counting)**.** Suppose we have to make r multiple choices in sequence. There are m_1 possibilities for the first choice, m_2 possibilities for the second etc. Then the total number of choices is $m_1 \times m_2 \times \cdots m_r$.

Example. How many ways can $1, 2, \cdots, n$ be ordered? The first choice has n possibilities, the second has $n-1$ possibilities etc. So there are $n \times (n-1) \times \cdots \times 1 = n!$.

2.2 Sampling with or without replacement

Let $N = \{1, 2, \cdots, n\}$ be a list. Let $X = \{1, 2, \cdots, x\}$ be the items. Let $f : N \to X$ with $f(i) =$ item at the ith position.

Definition (1. Sampling with replacement)**.** When we sample with replacement, after choosing at item, it is put back and can be chosen again. Then any sampling function f satisfies sampling with replacement.

Definition (2. Sampling without replacement)**.** After choosing an item, we burn it and cannot choose it again. Then f must be an injective function, and clearly we must have $X \geq n$.

We can also have sampling with replacement, but we require each item to be chosen at least once. In this case, f must be surjective.

Example. Suppose $N = \{a, b, c\}$ and $X = \{p, q, r, s\}$. How many injective functions are there $N \to X$?

When we choose $f(a)$, we have 4 options. When we choose $f(b)$, we have 3 left. When we choose $f(c)$, we have 2 choices left. So there are 24 possible choices.

Example. I have n keys in my pocket. We select one at random once and try to unlock. What is the possibility that I succeed at the rth trial?

Suppose we do it with replacement. We have to fail the first $r-1$ trials and succeed in the rth. So The probability is

$$\frac{(n-1)(n-1)\cdots(n-1)(1)}{n^r} = \frac{(n-1)^{r-1}}{n^r}.$$

Now suppose we are smarter and try without replacement. Then the probability is

$$\frac{(n-1)(n-2)\cdots(n-r+1)(1)}{n(n-1)\cdots(n-r+1)} = \frac{1}{n}.$$

Example (Birthday problem). How many people are needed in a room for there to be a probability of at least a half that two people have the same birthday?

Suppose $f(r)$ is the probability that, in a room of r people, there is a birthday match.

We solve this by finding the probability of no match, $1 - f(r)$. The total number of possibilities of birthday combinations is 365^r. For nobody to have the same birthday, the first person can have any birthday. The second has 364 else to choose, etc.

$$P(\text{no match}) = \frac{365 \cdot 364 \cdot 363 \cdots (366 - r)}{365 \cdot 365 \cdot 365 \cdots 365}$$

If we calculate this with a computer, we find that $f(22) = 0.475695$ and $f(23) = 0.507297$.

While this might sound odd since 23 is small, we shouldn't be counting the number of people, but the number of pairs. With 23 people, we have $23 \times 22/2 = 253$ pairs, which is quite large compared to 365.

2.3 Sampling with or without regard to ordering

Do the labels on the list positions or items matter?

Recall we have the function $f : N \to X$ and they map $f(1), f(2), \cdots, f(n)$. After choosing the numbers, we can

- Leave the list alone
- Sort the list ascending. i.e. we might get $(2, 5, 4)$ and $(4, 2, 5)$. If we don't care about list positions, these are just equivalent to $(2, 4, 5)$.
- Re-number each item by the number of the draw on which it was first seen. For example, we can rename $(2, 5, 2)$ and $(5, 4, 5)$ both as $(1, 2, 1)$. This happens if the labelling of items doesn't matter.
- Both of above. So we can rename $(2, 5, 2)$ and $(8, 5, 5)$ both as $(1, 1, 2)$.

Combining these four possibilities with whether we have replacement or not, we have 12 possible cases of counting.

2.4 Four cases of enumerative combinatorics

- Replacement + with ordering: the number of ways is x^n.
- Without replacement + with ordering: the number of ways is $x_{(n)} = x^{\underline{n}} = x(x-1)\cdots(x-n+1)$.
- Without replacement + without order: we only care which items get selected. The number of ways is $\binom{x}{n} = C_n^x = x_{(n)}/n!$.
- Replacement + without ordering: we only care how many times the item got chosen. We want to find ways to partition $n = n_1 + n_2 + \cdots n_x$. Say $n = 6$ and $x = 3$. We can write a particular partition as

  $$* * \mid * \mid * **$$

  So we have $n + x - 1$ symbols and $x - 1$ of them are bars. So the number of ways is $\binom{n+x-1}{x-1}$.

3 Stirling's formula

3.1 Multinomial coefficient

Suppose we fill successive positions in a list of length n, with replacement, from a set of x items. How many ways can this be done so that item $i \in X$ is used n_i times? (of course we must have $n_1 + n_2 + \cdots + n_x = n$.

In the case of $x = 2$, we have the binomial coefficient. In general, we have

Definition (Multinomial coefficient)**.** A *multinomial coefficient* is

$$\binom{n}{n_1, n_2, \cdots, n_x} = \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1\cdots-n_{x-1}}{n_x} = \frac{n!}{n_1!n_2!\cdots n_x!}.$$

It is the number of ways to distribute x items into n positions with replacement, in which the ith position has n_i items.

Example. We know that

$$(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \cdots + y^n.$$

If we have a trinomial, then

$$(x+y+z)^n = \sum_{n_1,n_2,n_3}\binom{n}{n_1, n_2, n_3}x^{n_1}y^{n_2}y^{n_3}.$$

Example. How many ways can we deal 52 cards to 4 player, each with a hand of 13? There are

$$\binom{52}{13, 13, 13, 13} = \frac{52!}{(13!)^4} = 53644737765488792839237440000 = 5.36 \times 10^{28}.$$

While computers are still capable of calculating that, what if we have more cards? Suppose each person has n cards. Then the number of ways is

$$\frac{(4n)!}{(n!)^4} = \frac{2^{8n}}{\sqrt{2}(\pi n)^{3/2}}.$$

We can use Stirling's Formula to approximate it:

3.2 Stirling's formula

Before we state and prove Stirling's formula, we prove a weaker (but examinable) version:

Proposition. $\log n! \sim n\log n$

Proof. Note that

$$e^n = 1 + n + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!} + \cdots.$$

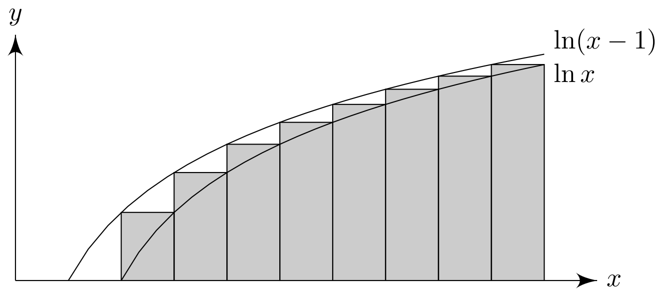
So $1 \le n^n/n! \le e^n$. Note that

$$\log n! = \sum_{k=1}^n \log k.$$

Now we claim that

$$\int_1^n \log x \, \mathrm{d}x \leq \sum_1^n \log k \leq \int_1^{n+1} \log x \, \mathrm{d}x.$$

This is true by considering the diagram:



Perform the integral to obtain

$$n \log n - n + 1 \leq \log n! \leq (n+1)\log(n+1) - n;$$

Divide both sides by $n \log n$ and let $n \to \infty$. Both sides tend to 1. So

$$\frac{\log n!}{n \log n} \to 1.$$

$\square$

Now we prove Stirling's Formula:

Theorem (Stirling's formula)**.** As $n \to \infty$,

$$\log\left(\frac{n!e^n}{n^{n+\frac{1}{2}}}\right) = \log\sqrt{2\pi} + O\left(\frac{1}{n}\right)$$

Corollary.

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

Proof. (non-examinable) Define

$$d_n = \log\left(\frac{n!e^n}{n^{n+1/2}}\right) = \log n! - (n+1/2)\log n + n$$

Then

$$d_n - d_{n+1} = (n+1/2)\log\left(\frac{n+1}{n}\right)$$

Write $t = 1/(2n+1)$. Then

$$d_n - d_{n+1} = \frac{1}{2t}\log\left(\frac{1+t}{1-t}\right).$$

We can simplifying by noting that

$$\log(1+t) - t = -\frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4 + \cdots$$
$$\log(1-t) + t = -\frac{1}{2}t^2 - \frac{1}{3}t^3 - \frac{1}{4}t^4 - \cdots$$

Then if we subtract the equations and divide by $2t$, we obtain

$$\begin{aligned}
d_n - d_{n+1} &= \frac{1}{3}t^2 + \frac{1}{5}t^4 + \frac{1}{7}t^6 \\
&\leq \frac{1}{3}t^2 + \frac{1}{3}t^4 + \frac{1}{3}t^6 = \cdots \\
&= \frac{1}{3}\frac{t^2}{1-t^2} \\
&= \frac{1}{3}\frac{1}{(2n+1)^2 - 1} \\
&= \frac{1}{12}\left(\frac{1}{n} - \frac{1}{n+1}\right)
\end{aligned}$$

By summing these bounds, we know that

$$d_1 - d_n \leq \frac{1}{12}\left(1 - \frac{1}{n}\right)$$

Then we know that d_n is bounded below by d_1+ something, and is decreasing since $d_n - d_{n+1}$ is positive. So it converges to a limit A.

Suppose $m > n$. Then $d_n - d_m < \left(\frac{1}{n} - \frac{1}{m}\right)\frac{1}{12} - \frac{2}{15}\frac{1}{(2n+1)^4}$ by adding back the term we removed when changing $t^4/5$ to $t^4/3$. So taking the limit as $m \to \infty$, $A < d_n < A + 1/(12n)$, with strict inequalities due to the second term.

To find A, we have a small detour to prove a formula:

Take $I_n = \int_0^{\pi/2} \sin^n \theta \, \mathrm{d}\theta$. This is decreasing as n increases as $\sin^n \theta$ gets smaller. We also know that

$$\begin{aligned}
I_n &= \int_0^{\pi/2} \sin^n \theta \, \mathrm{d}\theta \\
&= -\cos\theta \sin^{n-1}\theta\Big|_0^{\pi/2} + \int_0^{\pi/2} (n-1)\cos^2\theta \sin^{n-2}\theta \, \mathrm{d}\theta \\
&= (n-1)(I_n - 2 - I_n)
\end{aligned}$$

So

$$I_n = \frac{n-1}{n} I_{n-2}.$$

We can directly evaluate the integral to obtain $I_0 = \pi/2$, $I_1 = 1$. Then

$$\begin{aligned}
I_{2n} &= \frac{1}{2}\cdot\frac{3}{4}\cdots\frac{2n-1}{2n}\pi/2 = \frac{(2n)!}{(2^n n!)^2}\frac{\pi}{2} \\
I_{2n+1} &= \frac{2}{3}\cdot\frac{4}{5}\cdots\frac{2n}{2n+1} = \frac{(2^n n!)^2}{(2n+1)!}
\end{aligned}$$

So using the fact that I_n is decreasing, we know that

$$1 \leq \frac{I_{2n}}{I_{2n+1}} \leq \frac{I_{2n-1}}{I_{2n+1}} = 1 + \frac{1}{2n} \to 1.$$

Using the approximation $n! \sim n^{n+1/2}e^{-n+A}$, where A is the limit we want to find, we can approximate

$$\frac{I_{2n}}{I_{2n+1}} = \pi(2n+1)\left[\frac{((2n)!)^2}{2^{4n+1}(n!)^4}\right] \sim \pi(2n+1)\frac{1}{ne^{2A}} \to \frac{2\pi}{e^{2A}}.$$

Since the last expression is equal to 1, we know that $A = \log\sqrt{2\pi}$ $\square$

Example. Suppose we toss a coin $2n$ times. What is the probability of equal number of heads and tails? The probability is

$$\frac{\binom{2n}{n}}{2^{2n}} = \frac{(2n)!}{(n!)^2 2^{2n}} \sim \frac{1}{\sqrt{n\pi}}$$

This approximation can be improved:

Proposition ((non-examinable))**.** We use the $1/12n$ term from the proof above to get a better approximation:

$$\sqrt{2\pi}n^{n+1/2}e^{-n} + \frac{1}{12n+1} \leq n! \leq \sqrt{2\pi}n^{n+1/2}e^{-n} + \frac{1}{12n}.$$

Example. Suppose we draw 26 cards from 52. What is the probability of getting 13 reds and 13 blacks? The probability is

$$\frac{\binom{26}{13}\binom{26}{13}}{\binom{52}{26}} = 0.2181.$$

4 Axiomatic approach

Definition (Probability space)**.** A *probability space* is a triple $(\Omega, \mathcal{F}, P)$. Ω is the *sample space*, $\mathcal{F}$ is a collection of subsets of Ω. $P : \mathcal{F} \to [0, 1]$ is the *probability measure*. $\mathcal{F}$ has to satisfy the following axioms:

(i) $\emptyset, \Omega \in \mathbf{F}$.

(ii) $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$.

(iii) $A_1, A_2, \cdots, \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} \in \mathcal{F}$.

And P has to satisfy the following *Kolmogorov axioms*:

(i) $o \leq P(A) \leq 1$ for all $A \in \mathcal{F}$

(ii) $P(\Omega) = 1$

(iii) For any countable collection of events $A_1, A_2, \cdots$ which are disjoint, i.e. $A_i \cap A_j = \emptyset$ for all i, j, then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i).$$

We say $P(A)$ is the probability of the event A.

Definition (Probability distribution)**.** Let $\Omega = \{\omega_1, \omega_2, \cdots\}$. Choose $\{p_1, p_2, \cdots, \}$ such that $\sum_i^{\infty} = 1$. Let $p(\omega_i) = p_i$. Then define

$$P(A) = \sum_{\omega_i \in A} p(\omega_i).$$

This $P(A)$ satisfies the above axioms, and $p_1, p_2, \cdots$ is the *probability distribution*

Theorem.

(i) $P(\emptyset) = 0$

(ii) $P(A^C) + 1 - P(A)$

(iii) $A \subseteq B \Rightarrow P(A) \leq P(B)$

(iv) $P(A \subseteq B) = P(A) + P(B) - P(A \cap B)$.

(v) Let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$. Then

$$P\left(\bigcup_1^{\infty} A_i\right) = \lim_{n \to \infty} P(A_n).$$

This states that P is a continuous set function.

Proof.

(i) Ω and $\emptyset$ are disjoint. So $P(\Omega) + P(\emptyset) = P(\Omega \cup \emptyset) = P(\Omega)$. So $P(\emptyset) = 0$.

(ii) $P(\Omega) = 1 = P(A) + P(A^C)$ since A and A^C are disjoint.

(iii) Write $B = A \cup (B \cap A^C)$. Then $P(B) = P(A) + P(B \cap A^C) \geq P(A)$.

(iv) $P(A \cup B) = P(A) + P(B \cap A^C)$. We also know that $P(B) = P(A \cap B) + P(B \cap A^C)$. Then the result follows.

(v) c.f. Lecture 7

□

From above, we know that $P(A \cup B) \leq P(A) + P(B)$. So we say that P is a *subadditive* function. We also know that $P(A \cap B) + P(A \cup B) \leq P(A) + P(B)$ (in fact they are equal!). We say P is *submodular*.

4.1 Boole's inequality

This is also known as the "union bound".

Theorem (Boole's inequality)**.** For any $A_1, A_2, \cdots$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i).$$

Proof. The axiom states a similar formula that only holds for disjoint sets. So we need a clever trick to make them disjoint. We define

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_2 \setminus B_1 \\ B_i &= A_i \setminus \bigcup_{k=1}^{i-1} A_k. \end{aligned}$$

So we know that

$$\bigcup B_i = \bigcup A_i.$$

But the B_i are disjoint. So our Axiom (iii) gives

$$P(\bigcup_i A_i) = P(\bigcup_i B_i) = \sum_i P(B_i) \leq \sum_i P(A_i).$$

Where the last inequality follows from (iii) of the theorem above. □

Example. Suppose we have countably infinite number of biased coins. Let $A_k = [k\text{th toss head}]$ and $P(A_k) = p_k$. Suppose $\sum_1^\infty p_k < \infty$. What is the probability that there are infinitely many heads?

The event "there is at least one more head after the ith coin toss" is $\bigcup_{k=i}^\infty A_k$. There are infinitely many heads if and only if there are unboundedly many coin tosses, i.e. no matter how high i is, there is still at least more more head after the ith toss.

So the probability required is

$$P\left(\bigcap_{i=1}^{\infty}\bigcup_{k=i}^{\infty} A_k\right) \leq \lim_{i\to\infty} P\left(\bigcup_{k=i}^{\infty} A_k\right) = \lim_{i\to\infty} \sum_{k=i}^{\infty} p_k \to 0$$

Therefore $P(\text{infinite number of heads}) = 0$.

Example (Erdos 1947)**.** Is it possible to colour a complete n-graph (i.e. a graph of n vertices with edges between every pair of vertices) red and black such that no k-vertex complete subgraph with monochrome edges?

Erdös said this is possible if

$$\binom{n}{k}2^{1-\binom{k}{2}} < 1.$$

We colour edges randomly, and let $A_i = [i\text{th subgraph has monochrome edges}]$. Then the probability that at least one subgraph has monochrome edges is

$$P\left(\bigcup A_i\right) \leq \sum_i P(A_i) = \binom{n}{k} 2\cdot 2^{-\binom{k}{2}}.$$

The last expression is obtained since there are $\binom{n}{k}$ ways to choose a subgraph; a monochrome subgraph can be either red or black, thus the multiple of 2; and the probability of getting all red (or black) is $2^{-\binom{k}{2}}$.

If this probability is less than 1, then there must be a way to colour them in which it is impossible to find a monochrome subgraph, or else the probability is 1.

4.2 Inclusion-exclusion formula

Theorem (Inclusion-exclusion formula)**.**

$$\begin{aligned}P\left(\bigcup_i^n A_i\right) &= \sum_1^n P(A_i) - \sum_{i_1<i_2} P(A_{i_1}\cap A_{j_2}) + \sum_{i_1<i_2<i_3} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}) - \cdots\\ &\quad + (-1)^{n-1}P(A_1\cap\cdots\cap A_n).\end{aligned}$$

Proof. Perform induction on n. $n = 2$ is proven above.

Then

$$P(A_1\cup A_2\cup\cdots A_n) = P(A_1) = P(A_2\cup\cdots\cup A_n) - P\left(\bigcup_{i=2}^n (A_1\cap A_i)\right).$$

and we can apply the induction hypothesis for $n-1$. □

Example. Let $1, 2, \cdots, n$ be randomly permuted to $\pi(1), \pi(2), \cdots, \pi(n)$. If $i \neq \pi(i)$ for all i, we say we have a *derangement*.

Let $A_i = [i = \pi(i)]$.

Then

$$\begin{aligned}P\left(\bigcup_{i=1}^n A_i\right) &= \sum_k P(A_k) - \sum_{k_1<k_2} P(A_{k_1}\cap A_{k_2}) + \cdots\\ &= n\cdot\frac{1}{n} - \binom{n}{2}\frac{1}{n}\frac{1}{n-1} + \binom{n}{3}\frac{1}{n}\frac{1}{n-1}\frac{1}{n-2}\\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}\\ &\to e^{-1}\end{aligned}$$

So the probability of derangement is $1 - P(\bigcup A_k) = 1 - e^{-1} \approx 0.632$.