Probability and Statistics
MANUSCRIPT

William F. Barnes [1]

April 12, 2024

Contents

Chapter 1

Probability and Statistics

1 Events and Probability

Probability theory is a branch of mathematics for studying systems with inherent randomness or uncertainty. It works closely with *statistics*, another branch of mathematics concerned with the organization and interpretation of data, along with *combinatorics*, a formal method of counting.

Systems that exhibit random or pattern-less behavior contain a *stochastic* component. Typical stochastic processes may include flipping a coin, drawing a card from a deck, rolling a dice, or playing darts while blindfolded. A *stochastic event* is any data generated by a stochastic process, and the set of all possible stochastic events is called the system's *sample space*.

1.1 Events

Elementary Events

Events that are considered *elementary* carry one 'unit' of information, loosely speaking. A coin landing on 'heads', or a dice landing on 3 qualify as elementary events.

Compound Events

Simple events that occur in groups are called *compound events*. Drawing a Queen of Hearts from a deck of cards carries two units information, and may be interpreted in several ways: 'draw a Queen AND a heart', or 'draw a Queen OR a Heart', or perhaps 'draw NOT a Diamond'. Such events are compound for this reason.

Compound Event Notation

Borrowing the familiar symbols from elementary logic, we denote the word 'AND' with the 'cap' symbol $\cap$, equivalent to multiplication ($\cdot$). Meanwhile, the word 'OR' uses the 'cup' symbol $\cup$, or sometimes just a plus sign (+). The 'NOT' operator is abbreviated by a dash above the symbol, as in 'NOT' $A = \bar{A}$. Any event that is infinitely improbable, impossible, or undefined is denoted by the 'Empty set' symbol, $\emptyset$. In summary:

$$\begin{aligned} A \text{ AND } B &= A \cdot B = A \cap B \\ A \text{ OR } B &= A + B = A \cup B \\ \text{NOT } A &= \bar{A} \\ \text{Empty set} &= \emptyset \end{aligned}$$

The logic of probabilistic analysis is the same as 'ordinary' logic. For instance, the philosophical axiom 'nothing can be and not be simultaneously' is contained in the statement:

$$A \cap \bar{A} = \emptyset$$

State

The *state* of a system, loosely defined, is any particular configuration of the variables used to describe that system. For instance, a snapshot of a chessboard contains the present state of the game. Any event taking place in a system usually changes its state. If the system is to evolve in time, as would a game of chess, then future states evolves from the present state according to some rules or model of evolution.

1.2 Probability

Statistical Probability

A stochastic process that iterates over a very large or infinite number of trials will produce data points randomly distributed among the space of all possible data points for that process. For all events of type A, the ratio of occurrences N_A over all N events is called the *statistical probability* of event A, defined as:

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N} \tag{1.1}$$

$P(A)$ strictly has values between 0 and 1, inclusive.

Normalization Conditions

All other events B, C, etc., are represented by the symbol $\bar{A}$ ('NOT' A), and obey:

$$N_A + N_{\bar{A}} = N \tag{1.2}$$

$$P\left(\bar{A}\right) + P(A) = 1 \tag{1.3}$$

Classical Probability

A definition that skirts around the invocation of $N \to \infty$ is called the *classical probability.* For instance, it does not require an infinite number of rolls on a six-sided dice to know the chances of landing on 3 are one out of six, as this quality is built into the dice itself. Classical systems like dice or playing cards are most succinctly analyzed using classical probability.

Counting States

In probability and statistics, it's often necessary to know the total number of states available to a system, sometimes requiring rigorous combinatorial consideration.

Example 1

The last four digits of a phone number have the format ABCD, where each letter represents any integer from 0 to 9, inclusive. What is the probability of randomly guessing the number 7766?

Right away, we know how to list all possible states of the password, starting from 0000 and ending at 9999 in numerical order. With $N = 1000$ passwords, the probability of randomly choosing the correct password $N_A = 7766$ is:

$$P(7766) = \frac{1}{N} = \frac{1}{10000}$$

Example 2

A bank account password has format ABCD, where each letter represents any integer from 0 to 3, inclusive. What is the probability of randomly guessing the password?

All two-digit arrangements solved by AB are contained in:

$$\begin{aligned} \omega = &00, 01, 02, 03, \\ &10, 11, 12, 13, \\ &20, 21, 22, 23, \\ &30, 31, 32, 33 \end{aligned}$$

From here, observe that all four-digit arrangements are contained on an $\omega \times \omega$ grid having $N = 16^2 = 256$ total members, or

$$P(N_A) = \frac{1}{N} = \frac{1}{256}.$$

1.3 Mutually Exclusive Events

A pair of *mutually exclusive events* A and B are those that cannot occur simultaneously. Their coincidence can only belong to the *empty set* as

$$A \cap B = \emptyset \,.$$

If two events are mutually exclusive, the probability of either event occurring is the sum of the individual probabilities:

$$P\left(A \text{ or } B\right) = P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) \qquad (1.4)$$

Example 3

Calculate the probability of rolling a 3 or a 4 on a six-sided dice. As mutually exclusive events, we simply have

$$P = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \,.$$

Example 4

Calculate the probability that a random three-card hand drawn from a 52-card deck contains the Queen of Hearts.

$$P = \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{3}{52}$$

1.4 Non-Exclusivity

Non-mutually exclusive events are those that cause 'double counting' in $P\left(A \cup B\right)$, and are adjusted by subtracting the probability that both occur:

$$P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right) \qquad (1.5)$$

Or, in street terms, the above reads:

$$P\left(A \text{ or } B\right) = P\left(A\right) + P\left(B\right) - P\left(A \text{ and } B\right)$$

Example 5

From a 52-card deck, calculate the probability of drawing a Heart or a Face card, or one that is both.

$$P = \frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{22}{52}$$

Example 6

A class of 30 students is in session. 16 are studying French, and 21 are studying Spanish. Choosing a student at random, find the probability that they:

- study French
- study Spanish
- study French and Spanish
- study only French
- study only Spanish
- study French or Spanish

Denote F for French and S for Spanish. Then the easy ones can be listed off:

$$P\left(F\right) = 16/30$$
$$P\left(S\right) = 21/30$$

If the number of 'multilingual' students studying both French and Spanish is denoted M, then

$$\left(16 - M\right) + M + \left(21 - M\right) = 30$$

must hold, telling us $M = 7$, or

$$P\left(F \cap S\right) = 7/30 \,.$$

With M known, the the number of students studying just one subject can be written:

$$P\left(\text{French only}\right) = P\left(F\right) - P\left(F \cap S\right) = 9/30$$
$$P\left(\text{Spanish only}\right) = P\left(S\right) - P\left(F \cap S\right) = 14/30$$

Finally, the number of students studying French or Spanish should equal the total, which is the sum of those studying French only, Spanish only, or both. The probability should equal one:

$$P\left(F \cup S\right) = \frac{9}{30} + \frac{14}{30} + \frac{7}{30} = 1$$

1.5 Independent Events

Two events A and B that occur simultaneously as the compound event $A \cap B$ are *independent* if not causally connected.

In general, the statistical probability for the compound event $A \cap B$ reads

$$P\left(A \cap B\right) = \lim_{N \to \infty} \frac{1}{N} N_{A \cap B} \,,$$

where in the $N \to \infty$, limit the quantity $N_{A \cap B}$ becomes $N_A \cdot P\left(B\right)$. We deduce that, for independent events, the compound probability is the product of the individual probabilities:

$$P\left(A \cap B\right) = P\left(A\right) P\left(B\right) \qquad (1.6)$$

Example 7

Calculate the probability of two fair coin tosses each landing on 'tails'.

$$P(T \cap T) = P\left(T\right) \cdot P\left(T\right) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Example 8

From a 52-card deck, what is the probability of randomly drawing (i) a Queen, (ii) a Heart, (iii) the Queen of Hearts?

$$P(Q) = 1/4$$
$$P(H) = 1/13$$
$$P(Q \cap H) = P(Q) \cdot P(H) = \frac{1}{4} \cdot \frac{1}{13} = \frac{1}{52}$$

Coworker Problem

Two people are neighbors and travel to the same job. Person A owns car A, which has a 70% chance of starting in the morning, and a 30% chance of stalling (not starting). Person B owns car B with an 80% chance of starting. If one or both cars start, both people arrive at work. If neither car starts, they both miss work. In a span of 100 workdays, how many days are missed?

Each morning, one of four things happen:

$$P_1 = \text{A starts, B starts}$$
$$P_2 = \text{A starts, B stalls}$$
$$P_3 = \text{A stalls, B starts}$$
$$P_4 = \text{A stalls, B stalls}$$

As independent events, we further have

$$P_1 = (0.7)(0.8) = 0.56$$
$$P_2 = (0.7)(1 - 0.8) = 0.14$$
$$P_3 = (1 - 0.7)(0.8) = 0.24$$
$$P_4 = (1 - 0.7)(1 - 0.8) = 0.06 ,$$

which passes the sanity check

$$\sum_{j=1}^{4} P_j = 1 .$$

The answer to the question is the combined probability of at least one car starting. For this, we simply have

$$P = P_1 + P_2 + P_3 = 0.56 + 0.14 + 0.24 = 0.94 ,$$

or 94%. Six days are missed of every hundred.

1.6 Conditional Probability

In contrast to independent events, systems may bear a notion of 'dependent events', meaning that event B can occur only if event A occurs. This is called a *conditional probability*, denoted $P(B|A)$, enunciated 'B given A'. By definition, the probability of event B occurring given condition A is

$$P(B|A) = \lim_{N\to\infty} \frac{1}{N_A} N_{A\cap B} .$$

The term $N_{A\cap B}$ is the number of events B that occur given event A, which shows up again in the equation for $P(A \cap B)$:

$$P(A \cap B) = \lim_{N\to\infty} \frac{N_{A\cap B}}{N} .$$

Divide the two equations and simplify to derive the statement of conditional probability:

$$P(A \cap B) = P(B|A)\, P(A) \tag{1.7}$$

Note that the above generalizes the case of independent events, for if events A and B are independent, this result reduces to $P(A \cap B) = P(A)\, P(B)$ again.

Example 9

In a 52-card deck, calculate the probability that the first three cards are Kings.

$$P(KKK) = P(K)\, P(K|K)\, P(K|(K|K))$$
$$= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \approx 0.000181$$

Example 10

In a 52-card deck, calculate the probability that the first three cards are KQJ, in that order, with mixed suits allowed.

$$P(KQJ) = P(K)\, P(Q|K)\, P(J|(Q|K))$$
$$= \frac{4}{52} \cdot \frac{4}{51} \cdot \frac{4}{50} \approx 0.000483$$

Example 11

Suppose a pair of six-sided dice are rolled, landing on faces X and Y, respectively. What is the probability that $X = 2$ given that $X + Y \le 5$?

Of the 36 possible outcomes for a pair of dice, 10 of them obey $X+Y \le 5$. Listing these, find that only $(2, 1)$, $(2, 2)$, $(2, 3)$ satisfy $X = 2$, a total of three outcomes. The final answer is the ratio of these counts:

$$P\left(X = 2 \middle| X + Y \le 5\right) = \frac{3}{10}$$

Inversion Trick

Making use of the normalization condition

$$P\left(\bar{A}\right) + P(A) = 1 ,$$

it often helps in problem solving to use the negated event $\bar{A}$ as the working variable.

Example 12

Calculate the probability that a random three-card hand drawn from a 52-card deck contains the

Queen of Hearts. For this, define the event $\bar{A}$ of *not* drawing the Queen of Hearts and use the inversion trick as follows:

$$\bar{A} = \bar{A}_1 \cdot \bar{A}_2 \cdot \bar{A}_3$$

$$\begin{aligned} P\left(\bar{A}_1 \cap \bar{A}_2\right) &= P\left(\bar{A}_1\right) P\left(\bar{A}_2 | \bar{A}_1\right) \\ &= \frac{51}{52} \cdot \frac{50}{51} = \frac{50}{52} \end{aligned}$$

$$\begin{aligned} P\left(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3\right) &= P\left(\bar{A}_1 \cap \bar{A}_2\right) P\left(\bar{A}_3 | \bar{A}_1 \cap \bar{A}_2\right) \\ &= \frac{50}{52} \cdot \frac{49}{50} = \frac{49}{52} \end{aligned}$$

$$P(A) = 1 - P\left(\bar{A}\right) = 1 - \frac{49}{52} = \frac{3}{52}$$

Radioactive Decay

An unstable atom is one that expels energy by ejecting a subatomic particle or photon. Having no internal time-keeping mechanism, an unstable atom is entirely 'unaware' of its absolute age, and the its decay occurs at a random moment after becoming unstable.

Supposing the observation of an unstable atom begins at $t = 0$, the conditional probability of the atom decaying in a small time window Δt after time $t > 0$ is

$$P^{\text{decay}}_{\Delta t/t} = \tau^{-1} \Delta t \,,$$

where τ^{-1} is a constant related to (but not precisely equal to) the statistical half-life of the element, defined such that $\Delta t \ll \tau$.

The probability of the atom being 'still alive' in the interval Δt is

$$P^{\text{alive}}_{\Delta t/t} = 1 - \Delta t / \tau \,.$$

Decompose the entire 'alive' state into a product of conditional probabilities by slicing the time t into n identical copies of the short interval Δt as:

$$P^{\text{alive}}(t) = P^{\text{alive}}_{\Delta t/t_1} \cdot P^{\text{alive}}_{\Delta t/t_2} \cdots P^{\text{alive}}_{\Delta t/t_n} = \left(1 - \frac{t}{n\tau}\right)^n$$

Letting $n \to \infty$ permits use of the identity

$$\lim_{n\to\infty} \left(1 + \frac{A}{n}\right)^n = e^A \,.$$

It follows that the probability that a single unstable atom will still be 'alive' obeys

$$P(t) = e^{-t/\tau} \,. \tag{1.8}$$

One can work out the so-called half life $\tau_{1/2}$ of the atom by inquiring when $P(t)$ reduces to $1/2$.

Missing Face Problem

A six-sided dice is rigged to keep rolling if it lands on 2. Prove that the statistical probability of rolling a 3 is 1/5.

Denote B as the event 2, and denote A as event 3. With a single roll, the probabilities of A or B occurring are easy to write down:

$$\begin{aligned} p_1(A) &= 1/6 \\ p_1(B) &= 1/6 \end{aligned}$$

Of course, event B is unstable and induces a re-roll, which has a 1/6 chance of generating event A again, and the same chance for event B:

$$\begin{aligned} p_2(A|B) &= (1/6)(1/6) = (1/6)^2 \\ p_2(B|B) &= (1/6)(1/6) = (1/6)^2 \end{aligned}$$

With event $B|B$ comes another re-roll, and we stack on the conditional probabilities as

$$\begin{aligned} p_3(A|B|B) &= (1/6)^3 \\ p_3(B|B|B) &= (1/6)^3 \,, \end{aligned}$$

and the pattern is clear.

It follows that A could occur after any number of rolls, or potentially occur after an infinite string of B-events, and the total probability for A occurring is

$$\begin{aligned} P(A) = &\, p_1(A) + p_2(A|B) \\ &+ p_3(A|B|B) + p_4(A|B|B|B) + \cdots \,, \end{aligned}$$

simplifying to:

$$P(A) = \frac{1}{6} \cdot \left(1 + \left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)^2 + \left(\frac{1}{6}\right)^3 + \cdots\right)$$

In the infinite limit, the geometric series in parentheses converges to 6/5. The probability of event A occurring is therefore:

$$P(A) = \frac{1}{6} \cdot \frac{6}{5} = \frac{1}{5}$$

1.7 Bayes' Theorem

For two events A and B, recall that the statement of conditional probability reads

$$P(A \cap B) = P(B|A) P(A) \,,$$

which gives the likelihood events A and B simultaneously occurring. It's equally valid to write the statement with A and B swapped, giving a complimentary statement for 'A given B':

$$P(B \cap A) = P(A|B) P(B)$$

Now, since $A \cap B$ is logically equivalent to $B \cap A$, we immediately know $P(A \cap B) = P(B \cap A)$, allowing the two conditional equations to be combined to arrive at *Bayes' theorem*:

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} \tag{1.9}$$

Laptop Repair Shop

Source: CMPSCI240 UMass Amherst 2013

Problem 1

You work in a laptop repair shop. 80% of laptops brought in have been dropped, 15% of laptops have had a drink spilled on them, and 5% of laptops have a variety of other problems. A customer drops off a laptop and doesn't tell you what happened to it. You notice the laptop is emitting a slight coffee-like smell. Based on your knowledge of broken laptops, you estimate that 20% of dropped laptops have a slight coffee-like smell, 65% of laptops that have had something spilled on them have a slight coffee-like smell, and 5% of laptops that have some other problem have a slight coffee-like smell.

Provide labels for the events described in the problem. Find the probability that the laptop had something spilled on it given that is has a slight coffee-like smell.

Part 2:

After closer inspection, you note that the laptop has no cracks on the case. Based on your knowledge of broken laptops, you estimate that 80% of dropped laptops have cracked cases, 11% of laptops that have had something spilled on them have cracked cases, and 9% of laptops that have some other problem have cracked cases.

If the probability that a laptop smells like coffee and the probability that a laptop has a cracked case are conditionally independent of each other given the cause of the damage (drop, spill, or other), what is the probability that the laptop had something spilled on it if it has a slight coffee-like smell and no cracks in the case?

Solution 1

Denote D for 'drop', S for 'spill', and O for 'other'. Let the letter F denote 'slight coffee-like smell'. The information provided in the problem may be written:

$$\begin{aligned} P(D) &= .80 \\ P(S) &= .15 \\ P(O) &= .05 \end{aligned}$$

$$\begin{aligned} P(F|D) &= .20 \\ P(F|S) &= .65 \\ P(F|O) &= .05 \end{aligned}$$

We further deduce:

$$\begin{aligned} P(F) = P(F|D)\,P(D) + P(F|S)\,P(S) \\ + P(F|O)\,P(O) = .26 \end{aligned}$$

To answer the question, we need to compute $P(S|F)$, which is the inversion of $P(F|S)$. Applying Bayes' theorem, we easily find:

$$P(S|F) = \frac{P(S)\,P(F|S)}{P(F)} = .375$$

Part 2:

The problem asks us to evaluate $P\left(S|F \cap Z^C\right)$, where Z denotes 'crack' and Z^C denotes 'no crack,' using the information

$$\begin{aligned} P(Z|D) &= .80 \\ P(Z|S) &= .11 \\ P(Z|O) &= .09\,, \end{aligned}$$

or equivalently:

$$\begin{aligned} P\left(Z^C|D\right) &= 1 - .80 = .20 \\ P\left(Z^C|S\right) &= 1 - .11 = .89 \\ P\left(Z^C|O\right) &= 1 - .09 = .91 \end{aligned}$$

Proceed by applying Bayes' theorem directly to write

$$P\left(S|F \cap Z^C\right) = \frac{P\left(F \cap Z^C|S\right)P(S)}{P\left(F \cap Z^C\right)}\,.$$

Due to the independence between F and Z^C, the term $P\left(F \cap Z^C|S\right)$ decouples into $P(F|S)\,P\left(Z^C|S\right)$.

The denominator term $P\left(F \cap Z^C\right)$ can be recast as a sum that factors in a similar way:

$$\begin{aligned} P\left(F \cap Z^C\right) &= \sum_{i=D,S,O} P\left(F \cap Z^C|X_i\right)P(X_i) \\ &= \sum_{i=D,S,O} P(F|X_i)\,P\left(Z^C|X_i\right)P(X_i) \end{aligned}$$

Evaluating the final answer is now straightforward:

$$P\left(S|F \cap Z^C\right) = .7169$$

1.8 Copernicus Method

Source: Futility Closet
`https://futilitycloset.com/`

Princeton astrophysicist J. Richard Gott was visiting the Berlin Wall in 1969 *when a curious thought occurred to him.* His visit occurred at a random moment t years after the wall was created. Dividing the total lifespan T of the wall into four equal intervals, Gott reasoned there is a 50% chance that t lands within the middle two quarters of the wall's lifespan. Built in 1961, the wall was $t = 8$ years old at the time of his visit.

With this setup, we see at minimum that t is one quarter of T. At the other extreme, t could be three quarters of T. We therefore write

$$T_{\max} = 4t \qquad T_{\min} = t \times \frac{4}{3}$$

to establish upper and lower estimates of the wall's lifespan. Inserting $t = 8$ years, we find

$$T_{\max} = 32\,\text{yr} \qquad T_{\min} \approx 10.67\,\text{yr}\ ,$$

which, when these are added to 1961, produces the pair of results

$$\text{Year}_{\max} \approx 1993 \qquad \text{Year}_{\min} \approx 1972\ .$$

That is, Gott found a 50% chance that the Berlin Wall would fall between the years 1972 and 1993. The wall came down in 1989.

Generalization

Now generalize the above method using N intervals instead of four. Doing so, we begin with

$$T_{max} = Nt \qquad T_{min} = \frac{Nt}{N-1}\ .$$

Of course, the window defined by $T_{max} - T_{min}$ no longer corresponds to a probability of 50%, but must be adjusted to

$$p\left(N\right) = \frac{N-2}{N}\ ,$$

or

$$N\left(p\right) = \frac{2}{1-p}\ ,$$

which must be an integer.

An interesting exercise is one that calculates $\Delta T = T_{max} - T_{min}$ and expresses the result all in terms of $p\left(N\right)$. Figuring this out, one finds

$$\Delta T = \frac{4tp}{1-p^2}\ ,$$

which can be inverted via the quadratic formula:

$$p\left(\Delta T\right) = \frac{-2t}{\Delta T} + \sqrt{\left(\frac{2t}{\Delta T}\right)^2 + 1}$$

<u>Example 13</u>

Suppose you encounter a man for the first time in the 42^{nd} year of his life. Determine the upper and lower bounds of the interval in which he has a 33% chance of expiring (in total years). Repeat for 66% and 75%.

$$T_{max(33\%)} = 3 \times 42 = 126$$
$$T_{min(33\%)} = \frac{3}{2} \times 42 = 63$$

$$T_{max(66\%)} = 6 \times 42 = 252$$
$$T_{min(66\%)} = \frac{6}{5} \times 42 = 50.4$$

$$T_{max(75\%)} = 8 \times 42 = 336$$
$$T_{min(75\%)} = \frac{8}{7} \times 42 = 48$$

1.9 Dice Stacking

Suppose you are given a pair of distinguishable three-sided dice, abbreviated $2d3$, meant to be rolled simultaneously or in sequence. Denoting the outcomes of either given dice as 1, 2, 3, we can list the possible outcomes for one roll of such a pair of dice:

$$\omega = 11, 12, 21, 13, 22, 31, 23, 32, 33$$

The concatenation of outcomes from $2d3$ is equivalent to a single roll of an *effective* nine-sided dice, abbreviated $d9$.

Effective d5?

One may wonder if other effective dice can be derived from the roll $2d3$. For instance, replacing the concatenation of outcomes with the sum (inserting a plus sign between each digit in the above ω-list), one finds

$$\sigma = 2, 3, 3, 4, 4, 4, 5, 5, 6\ .$$

In a more visual notation, the above is equivalent to:

$\sigma_2 =$ ⚀⚀
$\sigma_3 =$ ⚁⚀, ⚀⚁
$\sigma_4 =$ ⚀⚂, ⚁⚁, ⚂⚀
$\sigma_5 =$ ⚁⚂, ⚂⚁
$\sigma_6 =$ ⚂⚂

In total, there are five possible outcomes as sums ranging from 2 to 6. The distribution of outcomes, however, is nonuniform. For instance, one sees that σ_4 can be reached three ways, whereas all other σ_j are less common.

Density Inversion

From the information contained in σ, we can work toward an effective $d5$ so long as the nonuniform distribution problem can be dealt with. Proceed by listing the allowed outcomes divided by the respective density, i.e.

$$\gamma = \frac{1}{1}(2), \frac{1}{2}(3), \frac{1}{3}(4), \frac{1}{2}(5), \frac{1}{1}(6) \; .$$

Multiply by $3 \cdot 2$ to get rid of all denominators:

$$\gamma = 6\,(2)\,, 3\,(3)\,, 2\,(4)\,, 3\,(5)\,, 6\,(6)$$

Explicitly, γ is a list with twenty items:

$$\gamma = 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6, 6, 6, 6$$

Qualitatively, we see outcomes that are under-represented in σ are over-represented in γ, and vice-versa.

The d5 Gamma Ray

The item γ is the 'gamma-array', or 'gamma ray' for short. Taking an gamma ray γ, assign the items in γ to an index via $\gamma(k)$, where k is an integer between 1 and 20, inclusive.

In the above example, one can see there is nothing particularly special about the arrangement of items 2, 3, 4, etc. in the array. These are listed by group in ascending order for mere convenience.

Since the intent is to build a uniformly-behaving $d5$ dice, it's prudent to conceive of the set $\{\gamma\}$ of *all* possible reshuffles of the items of γ. Of course, one wouldn't attempt writing down the full set $\{\gamma\}$, or stepping through its members in any systematic way. It suffices instead to have a function that shuffles an existing gamma ray to produce another.

The d5 Algorithm

An effective $d5$ is achieved with the following steps:

1. Choose a random gamma ray from $\{\gamma\}$ and let $k = 1$.

2. Let x equal the random sum of a $2d3$ roll:

$$x = \texttt{dice}(3) + \texttt{dice}(3)$$

3. If $x = \gamma(k)$, record $x - 1$ as a valid result.

4. Let $k \to k + 1$.

5. If $k > 20$ then Goto 1.

6. Goto 2.

2 Combinatorics

2.1 Arrangements

The number A_n of all *arrangements* of n distinguishable (non-repeated) elements is equal to the factorial of the total number of elements:

$$A_n = n!$$

If the set of n elements contains any number m identical members, the number of arrangements overcounts by a factor of m-factorial, which must be divided out:

$$A_n^m = \frac{n!}{m!} \tag{1.10}$$

<u>Example 1</u>

Consider the set of twelve elements $ACEFGILMNTUX$. What is the probability that a random arrangement of the elements will spell out $MAGNETICFLUX$?

$$P\,(MAGNETICFLUX) = \frac{1}{12!}$$

<u>Example 2</u>

Consider the word *FLUXELECTRIC*. What is the probability that a random arrangement of the letters will spell out *ELECTRICFLUX*?

$$P\,(ELECTRICFLUX) = \frac{2!2!2!}{12!}$$

2.2 Permutations

For a set of width n, partition each of the $n!$ arrangements into two bins such that one bin contains the first m elements in the arrangement, and the other bin contains the remaining $n - m$ elements. For each of the m elements in the first bin, the unused elements in the second bin are subject to $(n - m)!$ arrangements. Dividing out this factor yields the *permutation* number:

$$P_n^m = \frac{n!}{(n-m)!} \tag{1.11}$$

Qualitatively, the permutation number tells how many ways there are to choose m unique elements from a set of n total elements.

<u>Example 3</u>

A door keypad is unlocked by a code of four different integers between 0 to 9, inclusive. The same integer cannot be used twice. How any possible passwords are there?

For a $k = 4$ digit password drawing (and consuming) from $N = 10$ integers, observe that N of them are available for the first digit A. $N - 1$ of the digits are available for the second digit B, and so on, with the k^{th} digit selecting from $N-k+1$ unused integers. In general, we can intuitively write

$$P_N^k = N\left(N-1\right)\ldots\left(N-k+1\right) = \frac{N!}{\left(N-k\right)!} \, ,$$

which builds the permutation formula:

$$P_{10}^4 = \frac{10!}{\left(10-4\right)!} = \frac{10!}{6!} = 5040$$

Example 4

In a 52-card deck, calculate the probability that the first three cards are Kings. (This is a repeat of an earlier problem.)

The total number of ways to draw any three cards from 52 is

$$P_{52}^3 = \frac{52!}{\left(52-3\right)!} = 52 \cdot 52 \cdot 50 \, .$$

Meanwhile, the number of ways to draw any three Kings from four total Kings is

$$P_4^3 = \frac{4!}{\left(4-3\right)!} = 4 \cdot 3 \cdot 2 \, .$$

The ratio of these is the answer:

$$P\left(KKK\right) = \frac{P_4^3}{P_{52}^3} = \frac{4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50} \approx 0.000181$$

Example 5

In a 52-card deck, calculate the probability that the first three cards are KQJ, in that order, with mixed suits allowed. (This is a repeat of an earlier problem.)

Consider the three cards KQJ in that order. Listing off all ways this could occur, we see there are 4^3 possibilities in total, as each card has four suits to choose from. The total number of ways to draw any three cards from 52 is

$$P_{52}^3 = \frac{52!}{\left(52-3\right)!} = 52 \cdot 52 \cdot 50 \, .$$

The ratio of these is the answer:

$$P\left(KQJ\right) = \frac{4^3}{52 \cdot 51 \cdot 50} \approx 0.000483$$

2.3 Birthday Problem

Consider a room populated by N people. What is the probability that any two people were born on the same day? (Ignore leap year.)

Conditional Probability Analysis

Begin with the trivial case $N = 2$, in where there is a $1/365$ chance of a common birthday:

$$P\left(2\right) = \frac{1}{365} = 1 - \frac{364}{365}$$

The result is written in the form $1 - X$ so we may focus on X, the probability of *no* common birthday.

A third person entering the system, making $N = 3$, has $365-2 = 363$ available days to avoid a common birthday. The probability becomes

$$P\left(3\right) = 1 - \frac{364}{365} \cdot \frac{363}{365} \, ,$$

and the pattern becomes clear. For total population N, the probability that some pair of people share a birthday ought to be:

$$\begin{aligned} P\left(N\right) &= 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{\left(365-N+1\right)}{365} \\ &= 1 - \frac{365!}{365^N \left(365-N\right)!} \end{aligned}$$

On the right, note that X has been expressed as a recursion of conditional probabilities:

$$X\left(n|n-1\right) = \frac{365-\left(n-1\right)}{365}$$

$$X\left(N\right) = \prod_{n=2}^{N} X\left(n|n-1\right)$$

Permutation Analysis

The same result can be written directly by the permutation formula. Choosing N people of 365, we have

$$P_{365}^N = \frac{365!}{\left(365-N\right)!} \, .$$

Meanwhile, the number of ways to assign birthdays to N people is 365^N without worrying about sharing. The ratio of these is the same $X\left(N\right)$ calculated above:

$$X\left(N\right) = \frac{1}{365^N} P_{365}^N = \frac{365!}{365^N \left(365-N\right)!} \, .$$

Following is a list of various populations N with their corresponding $P(N)$:

N	P(N)
5	2.71%
10	11.7%
20	41.1%
23	50.7%
30	70.6%
50	97.0%

Remarkably, the population need only be 23 in order for there to be a 50% chance that any two people share a birthday.

2.4 Combinations

Extending the derivation of the permutation formula, it may happen that the precise order of elements in the 'm' bin does not matter, meaning the list of permutations is overpopulated by a factor of $m!$. Dividing out this factor, we attain the number of *combinations* in the system:

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

The numbers C_n^m are none other than the binomial coefficients.

<u>Example 6</u>

From a 52-card deck, a five-card hand is drawn at random. How many five-card hands are possible?

$$C_{52}^5 = \frac{52!}{5!\,(52-5)!}$$

<u>Example 7</u>

From a 52-card deck, calculate the probability of drawing a royal flush (A-K-Q-J-10) in any order in any one suit.

$$P\left(RF\right) = \frac{4}{C_{52}^5} = \frac{1}{649740} \approx 0.00000154$$

Lottery Game

In a lottery game, the winning numbers are five non-repeating integers between 1 and 75, inclusive, along with one bonus integer between 1 and 15, inclusive. Guessing the five winning numbers at random, let us calculate the probability $P = (n, b)$ that n of the guessed numbers match the winning numbers, with or without the bonus b also being correctly guessed.

As an application of combinatoric analysis, it follows that there are $C_{75}^5 = 17,259,390$ ways to guess the winning five numbers, and $C_{15}^1 = 15$ choices for the bonus number. Following are the probabilities of guessing partial winning numbers, with and without the bonus.

$$\begin{aligned}
P\left(5,1\right) &= \frac{1}{C_{75}^5 \cdot C_{15}^1} = \frac{1}{258,890,850} \\
P\left(4,1\right) &= \frac{C_5^4 \cdot C_{70}^1}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{739,688} \\
P\left(3,1\right) &= \frac{C_5^3 \cdot C_{70}^2}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{10,720}
\end{aligned}$$

$$\begin{aligned}
P\left(2,1\right) &= \frac{C_5^2 \cdot C_{70}^3}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{473} \\
P\left(1,1\right) &= \frac{C_5^1 \cdot C_{70}^4}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{56} \\
P\left(0,1\right) &= \frac{C_5^0 \cdot C_{70}^5}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{21}
\end{aligned}$$

$$\begin{aligned}
P\left(5,0\right) &= \frac{C_5^5 \cdot C_{70}^0 \cdot C_{14}^1}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{18,492,204} \\
P\left(4,0\right) &= \frac{C_5^4 \cdot C_{70}^1 \cdot C_{14}^1}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{52,835} \\
P\left(3,0\right) &= \frac{C_5^3 \cdot C_{70}^2 \cdot C_{14}^1}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{766}
\end{aligned}$$

$$\begin{aligned}
P\left(2,0\right) &= \frac{C_5^2 \cdot C_{70}^3 \cdot C_{14}^1}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{34} \\
P\left(1,0\right) &= \frac{C_5^1 \cdot C_{70}^4 \cdot C_{14}^1}{C_{75}^5 \cdot C_{15}^1} \approx \frac{1}{4} \\
P\left(0,0\right) &= \frac{C_5^0 \cdot C_{70}^5 \cdot C_{14}^1}{C_{75}^5 \cdot C_{15}^1} \approx \frac{2}{3}
\end{aligned}$$

3 Variables and Expectations

3.1 Normalization

For an event A, the probability $P(A)$ of the event occurring has a trivial yet important relationship to $P\left(\overline{A}\right)$, via the normalization condition

$$1 = P\left(A\right) + P\left(\overline{A}\right) .$$

In words, normalization means there is a 100% chance that event A either occurs or does not occur.

For n repeated events, also called trials, A is replaced by A_k, where the index k tracks the event number such that $1 \leq k \leq n$. Using this notation, we lump $\overline{A}$ and all subsequent $\overline{A}_k$ into the coefficients A_k to write a more general normalization condition:

$$1 = \sum_{k=1}^{n} P\left(A_k\right) \tag{1.12}$$

3.2 Statistical Average

Expanding out the normalization condition above, we have a sequence with n terms on the right:

$$1 = P(A_1) + P(A_2) + \cdots + P(A_n)$$

By multiplying A_k into each respective $P(A_k)$ term, the equation becomes the *statistical average*, or *weighted average* $\langle A \rangle$ of the events A_k. To denote this, we write:

$$\langle A \rangle = A_1 \cdot P(A_1) + A_2 \cdot P(A_2) + \cdots + A_n \cdot P(A_n)$$

In summation notation, the above result reads:

$$\langle A \rangle = \sum_{k=1}^{n} A_k \cdot P(A_k) \tag{1.13}$$

3.3 Expectation Value

A function f that depends on any event A_k can also be averaged using this apparatus. Generalizing the above, we can easily write an equation for the *expectation value* of f:

$$\langle f \rangle = \sum_{k=1}^{n} f(A_k) \cdot P(A_k) \tag{1.14}$$

With the above, we may also calculate $\langle f^2 \rangle$ without hesitation:

$$\langle f^2 \rangle = \sum_{k=1}^{n} (f(A_k))^2 \cdot P(A_k) \tag{1.15}$$

Example 1

A six-sided dice that chooses a random number $1 \leq A_k \leq 6$ is tossed in succession to produce $n \gg 1$ events. Calculate the average outcome.

$$\langle A \rangle = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{21}{6} = 3.5$$

Example 2

A six-sided dice that is missing the 2-face but has an extra 4-face is tossed in succession to produce $n \gg 1$ events. Calculate the average outcome.

$$\langle A \rangle = \frac{1}{6} + \frac{0}{6} + \frac{3}{6} + \frac{4 \cdot 2}{6} + \frac{5}{6} + \frac{6}{6} = \frac{23}{6} = 3.83$$

3.4 Standard Deviation

Further insight into f can be gained by inserting $(f(A_k) - \langle f \rangle)^2$ as the argument in Equation (1.19). By doing so, and then taking the square root of the entire result, we arrive at the *standard deviation* in the system:

$$\sigma_f = \sqrt{\sum_{k=1}^{n} (f(A_k) - \langle f \rangle)^2 \cdot P(A_k)} \tag{1.16}$$

Using only the definitions above, it's easy to show that the standard deviation is equivalent to

$$\sigma_f = \sqrt{\langle f^2 \rangle - 2 \langle f \rangle \langle f \rangle + \langle f \rangle^2}$$
$$\sigma_f = \sqrt{\langle f^2 \rangle - \langle f \rangle^2} \; . \tag{1.17}$$

3.5 Continuous Distributions

For a stochastic process that produces events A_k in a continuous range instead of a discrete set, the normalization condition

$$\sum_{k=1}^{n} P(A_k) = 1$$

becomes an infinite sum. When confronted with this, the sum becomes an integral according to

$$\sum_{k=1}^{n} P(A_k) \to \int dP(A_k) = 1 \; .$$

Probability Distribution Function

At this point, we abbreviate $A_k \to k$, and then use the chain rule to write

$$1 = \int_n \frac{dP(k)}{dk} \, dk = \int_n w(k) \, dk \; .$$

The continuous function $w(k)$ is called the *probability density*, or *probability distribution function* (although 'w' stands for *weight*). Specifically, $\omega(k)$ is the probability of an event occurring within a window $[k, k + dk]$.

In the continuous limit, the equations for the statistical average, general expectation value, and standard deviation generalize to:

$$\langle k \rangle = \int_n k \cdot w(k) \, dk \tag{1.18}$$

$$\langle f \rangle = \int_n f(k) \, w(k) \, dk \tag{1.19}$$

$$\sigma_f = \sqrt{\int_n (f(k) - \langle f \rangle)^2 \; w(k) \, dk} \tag{1.20}$$

Note that Equation (1.17) still holds in the continuous distribution.

Example 3

What is the expected area of a right triangle with a hypotenuse of k whose non-right angles are uniformly distributed over the interval $(0, \pi/2)$?

$$\begin{aligned}\langle A \rangle &= \frac{k^2/2}{\pi/2} \int_0^{\pi/2} \cos(\theta) \sin(\theta) \, d\theta \\ &= \frac{k^2/2}{\pi/2} \int_0^1 x \, dx = \frac{k^2}{2\pi}\end{aligned}$$

Example 4

Divide a given line segment into two other line segments. Then, cut each of these new line segments into two more line segments. What is the probability that the resulting four line segments are the sides of a quadrilateral?

Let the total length be L, and require that no one side be longer than $L/2$. After the initial cut, let the longer segment have length x, and the shorter segment $L - x$. Diving the longer segment at point z (from the start of x), it is required that $z < L/2$ and simultaneously $x - z < L/2$. Therefore, the window of allowed z has width $L/2 - (x - L/2) = L - x$. The normalized probability of an allowed z along x is:

$$\begin{aligned}P &= N \int_{L/2}^{L} \frac{L - x}{x} \, dx \\ &= \frac{(L \ln x - x)\big|_{L/2}^{L}}{L/2} = 2 \ln 2 - 1 \approx 38.6\%\end{aligned}$$

3.6 Random Variables

Consider a set $\{A_k\}$ of random (not necessarily independent) variables.

Sum of Random Variables

Suppose that the sum of random variables comes to A:

$$A = \sum_{k=1}^{n} A_k$$

In the continuous large-n limit, the average value of A can be written as an n-dimensional integral

$$\langle A \rangle = \int A \cdot w(A_1, \ldots, A_n) \, dA_1 \ldots dA_n \, .$$

Replace A in the above with its sum representation:

$$\langle A \rangle = \sum_{k=1}^{n} \int A_k \cdot w(A_1, \ldots, A_n) \, dA_1 \ldots dA_n \, ,$$

where the 'sum' symbol has been harmlessly pulled outside all n of the integrals.

Simplifying the above is a straightforward exercise, with the majority of integrals satisfying the normalization condition and resolving to one. After the dust settles, one finds

$$\langle A \rangle = \langle A_1 \rangle + \langle A_2 \rangle + \cdots + \langle A_n \rangle \, ,$$

which, strictly translated, means *the average of the sum is the sum of the averages*:

$$\langle A \rangle = \sum_{k=1}^{n} \langle A_k \rangle \tag{1.21}$$

Independent Random Variables

More can be said about the weight function $w(k)$ in the regime of independent random variables. In the same sense that $P(A \cap B) = P(A) P(B)$ applies to independent events, we write

$$w(A_1, A_n, \ldots, A_n) = w(A_1) \, w(A_2) \cdots w(A_n)$$

when all probability distribution values $w(A_k)$ are independent.

Product of Independent Random Variables

Suppose that the product of random variables $\{B_k\}$ of n comes to

$$B = \prod_{k=1}^{n} B_k = B_1 \cdot B_2 \cdots B_n \, ,$$

and let us calculate the average value $\langle B \rangle$. Going by definition, this amounts to

$$\langle B \rangle = \prod_{k=1}^{n} \int B_k \cdot w(B_1) \cdots w(B_n) \, dB_1 \ldots dB_n \, ,$$

the 'product' symbol has been pulled outside all n of the integrals, and the probability distribution is factored to accommodate independent B_k.

From this, we see the right side is the product of n independent integrals, and conclude

$$\langle B \rangle = \langle B_1 \rangle \cdot \langle B_2 \rangle \cdots \langle B_n \rangle \, ,$$

which, strictly translated, means *the average of the product is the product of the averages*:

$$\langle B \rangle = \prod_{k=1}^{n} \langle B_k \rangle \tag{1.22}$$

3.7 Variance

Starting from the sum

$$A = \sum_{k=1}^{n} A_k ,$$

square both sides and convince yourself that

$$A^2 = \left(\sum_{i=1}^{n} A_i\right)\left(\sum_{j=1}^{n} A_j\right) = \sum_{k=1}^{n} A_k^2 + \sum_{i\neq j} c_{ij} A_i A_j ,$$

where c_{ij} are the binomial coefficients to represent all cross terms.

Meanwhile, the square of the average $\langle A \rangle$ comes out to

$$\langle A \rangle^2 = \sum_k \langle A_k \rangle^2 + \sum_{i\neq j} c_{ij} \langle A_i \rangle \langle A_j \rangle .$$

We can also calculate $\langle A^2 \rangle$ by exploiting the the independence among A_k, resulting in

$$\langle A^2 \rangle = \sum_{k=1}^{n} \langle A_k^2 \rangle + \sum_{i\neq j} c_{ij} \langle A_i \rangle \langle A_j \rangle .$$

Taking the difference $\langle A^2 \rangle - \langle A \rangle^2$, the cross terms cancel and we arrive at a simple relation connecting A to its members:

$$\begin{aligned} \langle A^2 \rangle - \langle A \rangle^2 &= \sum_{k=1}^{n} \langle A_k^2 \rangle - \langle A_k \rangle^2 \\ &+ \cancel{\sum_{i\neq j} c_{ij} \langle A_i \rangle \langle A_j \rangle} - \cancel{\sum_{i\neq j} c_{ij} \langle A_i \rangle \langle A_j \rangle} \end{aligned}$$

The square root of $\langle A^2 \rangle - \langle A \rangle^2$ is defined as the *variance* in A:

$$\mathrm{Var}\,(A) = \sqrt{\langle A^2 \rangle - \langle A \rangle^2} \tag{1.23}$$

As we've built it, the variance has some more handy expressions:

$$\mathrm{Var}\,(A) = \sqrt{\sum_{k=1}^{n} \langle A_k^2 \rangle - \langle A_k \rangle^2} = \sqrt{\sum_{k=1}^{n} \left(\mathrm{Var}\,(A_k)\right)^2}$$

3.8 Dispersion

A variation in the sum A of independent variables, denoted ΔA, is also known as *dispersion*, defined as:

$$\Delta A = A - \langle A \rangle = \sum_{k=1}^{n} \left(A_k - \langle A_k \rangle\right) \tag{1.24}$$

From this, it's easy to show that the average dispersion is zero:

$$\langle \Delta A \rangle = \langle A \rangle - \langle A \rangle = 0$$

The expectation value $\langle \Delta A^2 \rangle$, however, is more telling. By brute force, first write

$$\begin{aligned} \Delta A^2 &= \left(\left(A_1 - \langle A_1 \rangle\right) + \left(A_2 - \langle A_2 \rangle\right) + \cdots\right)^2 \\ &= \sum_{k=1}^{n} \left(A_k - \langle A_k \rangle\right)^2 + \sum_{i\neq j} c_{ij} \Delta A_i \Delta Aj , \end{aligned}$$

so then:

$$\langle \Delta A^2 \rangle = \sum_{k=1}^{n} \left\langle \left(A_k - \langle A_k \rangle\right)^2 \right\rangle + \cancel{\sum_{i\neq j} c_{ij} \langle \Delta A_i \rangle \langle \Delta Aj \rangle}$$

This is result is perhaps not surprising, telling us the total ΔA^2 is the sum of its constituents:

$$\langle \Delta A^2 \rangle = \sum_{k=1}^{n} \langle \Delta A_k^2 \rangle \tag{1.25}$$

In the large-n limit, the average $\langle A \rangle$ scales with n, and meanwhile we see $\langle \Delta A^2 \rangle$ also scales with n. The ratio of the RMS dispersion over the average thus tends to zero, as

$$\frac{\sqrt{\langle \Delta A^2 \rangle}}{\langle A \rangle} \approx \frac{1}{\sqrt{n}} \to 0 ,$$

telling us that fluctuations in A become negligibly small.

3.9 Random Product Problem

Consider the real numbers in the interval $(0:2)$. Let $\tilde{x}_1$ be a random number chosen from this interval, let $\tilde{x}_2$ be a second random number, and so on up to $\tilde{x}_n$. (Repeats are allowed but unlikely.)

Expectation

With this setup, suppose we are interested in the product of numbers in the list:

$$X_n = \prod_{j=1}^{n} \tilde{x}_j = \tilde{x}_1 \cdot \tilde{x}_2 \cdot \tilde{x}_3 \cdots \tilde{x}_n$$

Sampling from $(0:2)$, it is true that the average random value is one:

$$\langle \tilde{x}_j \rangle = 1 .$$

This should mean right away that the average product is also one:

$$\langle X_n \rangle = 1 \cdot 1 \cdot 1 \cdots = 1$$

Disaster

All seems well until we try to verify $\langle X_n \rangle = 1$ on a calculator. To illustrate, take the contrived list with five members

$$\{\tilde{x}_j\} = \{0.8, .9, 1.0, 1.1, 1.2\} \ ,$$

so the product is

$$X_5 = (0.8)\,(0.9)\,(1.0)\,(1.1)\,(1.2) = 0.9504 \ ,$$

which is less than one.

The effect gets worse for increasing n, for if we continue the pattern so the list includes 0.7, 1.3, the product is

$$X_7 \approx 0.8648 \ .$$

The members $\tilde{x}_j < 1$ weigh down the product X_n more than members $\tilde{x}_j > 1$ weigh the product up. After many trials, the net result is $X_n \to 0$, in contradiction to $\langle X_n \rangle = 1$.

You're encouraged to verify this on a computer with a variety of $\tilde{x}_j$ and a variety of n-values to see there is clearly something wrong with the way X_n is expected to behave. It seems that X_n reliably *decreases* for increasing n, so we inevitably conclude $X_n \to 0$.

Modified Interval

Going back to the beginning, adjust the interval to $(0 : 2.5)$ so that

$$\langle \tilde{x}_j \rangle = 1.25 \ ,$$

and run similar experiments. Now we're multiplying a list of numbers who average is clearly greater than one. However, much like the previous setup, the product X_n still goes to zero.

Adjust the interval once more to $(0 : 3)$ and start over. This time, we have

$$\langle \tilde{x}_j \rangle = 1.5 \ ,$$

and pattern finally breaks. One can check that product X_n tends to grow for increasing n, and for large n, the trend $X_n \to \infty$ occurs.

Tuning the Interval

Given the evidence on hand, there should be some interval $(0 : p)$, where p is some number between 2.5 and 3 such that X_n does not tend to zero and does not tend to infinity:

$$X_n \propto \langle X_n \rangle$$

To estimate p, one may write a simple trial-and-error program that allows p to vary:

1. Choose an initial value for p.
2. Choose a sufficiently large sample of n values from the interval $(0 : p)$ and calculate the corresponding X_n.
3. If X_n goes to zero, increase p.
4. If X_n goes to infinity, decrease p.
5. Goto step 2.

Doing this, one finds, after many trials:

$$p \approx 2.718\ldots$$

This answer is tantalizingly close to Euler's constant. Who saw that coming?

Proper Analysis

To reconcile the random product problem, begin with the natural logarithm of the product X_n:

$$\ln(X_n) = \ln(\tilde{x}_1) + \ln(\tilde{x}_2) + \ln(\tilde{x}_3) + \cdots$$

In the limit $n \to \infty$, it stands to reason that *every* real number in the interval $(0 : p)$ is represented by some $\tilde{x}_j$ or another. Rearranging the sum to write these in order, we have

$$\ln(X) = \lim_{n\to\infty} \sum_{j=1}^{n} \ln\left(p\frac{j}{n}\right) \ .$$

The total interval $(0 : p)$ can be made from n copies of a small interval Δx, which means $p/n = \Delta x$. Also substituting $x = j/n$, the above becomes

$$\ln(X) = \lim_{\Delta x \to 0} \frac{1}{p} \sum_{x>0}^{1} \ln(px)\,\Delta x \ .$$

The sum becomes an integral in the continuous limit

$$\ln(X) = \frac{1}{p}\int_0^1 (\ln(p) + \ln(x))\,dx \ ,$$

and the solution is straightforward:

$$p\ln(X) = (x\ln(p) + x\ln(x) - x)\Big|_0^1$$
$$p\ln(X) = \ln(p) - 1$$

Now comes the final argument. By avoiding $\ln(0) \to -\infty$ and also $\ln(\infty) \to \infty$, we're asking for X to be a finite number. In the infinite limit, it can only be that $X \to 1$:

$$p\ln(1) = 0 = \ln(p) - 1$$

The only solution to $\ln(p) = 1$ is $p = e$ and we're done.

Problem 1

Consider the real numbers in the interval $(0:1)$, and let $\tilde{x}_1$, $\tilde{x}_2$, $\tilde{x}_3$, etc. represent random samples from this interval. How many times n must a random $\tilde{x}_j$ be multiplied into a very large number $A \gg 1$ until the product is approximately one? In other words, solve for n in the following:

$$1 \approx A \cdot \tilde{x}_1 \cdot \tilde{x}_2 \cdots \tilde{x}_n$$

Hint:

$$0 \approx \ln(A) + \sum_{j=1}^{n} (\ln(x) + 1) - \sum_{j=1}^{n} (1)$$

The answer is $n \approx \ln(A)$.

3.10 Random Sums Problem

Accumulating random values $0 < r_k < 1$ in a sum, how many iterations $\langle n \rangle$ until the total is greater than one, on average?

Geometric Analysis

Begin by interpreting each interval $0 \leq r_k \leq 1$ as an independent 'number line' for each of the n variables needed. For $n = 2$, r_1, r_2 lie on orthogonal axes of a two-dimensional plane. For $n = 3$, r_1, r_2, r_3 lie on orthogonal axes of a three-dimensional volume, and so on.

Geometrically, the criteria

$$\sum_{j=1}^{n} r_j > 1$$

thus defines a triangular area in two dimensions, a pyramid-like volume in three dimensions, a hypervolume in four-dimensions, and so on. The space enclosed by each 'volume' is defined by

$$\sum_{j=1}^{n} r_j \leq 1 \; .$$

For convenience, let us label $r_1 \rightarrow z$, $r_2 \rightarrow y$, $r_3 \rightarrow x$, $r_4 \rightarrow t$, $r_5 \rightarrow u$.

Examining $n = 2$, the line $z + y = 1$ encloses half of the unit square, formally shown via

$$\begin{aligned} V_2 &= \int_0^1 \int_0^{1-z} dy \, dz \\ &= \int_0^1 (1-z) \, dz = \left(z - \frac{z^2}{2} \right) \Big|_0^1 = \frac{1}{2} \; . \end{aligned}$$

For $n = 3$, the plane $z + y + x = 1$ encloses one sixth of the unit cube:

$$V_3 = \int_0^1 \int_0^{1-z} \int_0^{1-z-y} dx \, dy \, dz = \frac{1}{6}$$

Jumping to $n = 4$ is impossible to visualize, however the required integral is easy enough to write and solve:

$$V_4 = \int_0^1 \int_0^{1-z} \int_0^{1-z-y} \int_0^{1-z-y-x} dt \, dx \, dy \, dz = \frac{1}{24}$$

Evidently, the enclosed volume is always the inverse of the factorial of the number of dimensions,

$$V_n = \frac{1}{n!} \; .$$

Probabilistic Calculation

We ultimately seek the expectation value $\langle n \rangle$, given by

$$\langle n \rangle = \sum_{n=2}^{\infty} n \cdot P(n) \; ,$$

where $P(n)$ is the probability of satisfying

$$\sum_{j=1}^{n} r_j < 1 \; .$$

By the geometric analysis, observe that $P(n)$ corresponds to the 'window' of volume bounded between V_n and V_{n-1}:

$$P(n) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}$$

With this, we can calculate the expectation value

$$\langle n \rangle = \sum_{n=2}^{\infty} n \cdot \frac{n-1}{n!} = \sum_{n=2}^{\infty} \frac{1}{(n-2)!} = \sum_{0}^{\infty} \frac{1}{n!} \; ,$$

which indeed converges to Euler's constant:

$$e = \sum_{0}^{\infty} \frac{1}{n!}$$

Amazingly, we conclude:

$$\langle n \rangle = e$$

4 Systems and Distributions

4.1 Two-State System

Consider a balanced coin that is tossed to generate n random events resulting in either H(eads) or T(ails). If we are interested in the portion m 'heads' events that occur without the order of events being important, the combination number

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

summarizes the system. Said another way, the multiplicity of the system Ω is 'n choose m':

$$C_n^m = \Omega\,(m,n) = \binom{m}{n}$$

The sum of all C_n^m across the whole range of m, namely from 0 to n, must resolve to the total multiplicity of events, namely 2^n for a coin tossing game:

$$2^n = \sum_{m=0}^{n} \frac{n!}{m!\,(n-m)!}$$

Normalized Probability Distribution

Knowing the total states available to the two-state system, we can write the probability of attaining m events among n trials in any two-state system as:

$$P\,(m,n) = \frac{1}{2^n}\frac{n!}{m!\,(n-m)!} \tag{1.26}$$

In the above definition, we divide by the factor 2^n so that the sum of all probabilities - accounting for all outcomes - sums to one.

The combination number C_n^m can be interpreted nicely by spotting the pattern that emerges in trivial cases. A single toss can result in T or H, which we denote

$$\omega_1 = (T,H) \ .$$

Denoting m as the number of H-events, we write

$$C_1^0 = 1 \qquad C_1^1 = 1$$

For a game of $n = 2$ tosses, the list of possible events is

$$\omega_2 = (TT,TH,HT,HH) \ .$$

Again denoting m as the number of H-events, we write

$$C_2^0 = 1 \qquad C_2^1 = 2 \qquad C_2^2 = 1$$

Similarly, a game of three tosses has

$$\omega_3 = (TTT,TTH,THT,THH,HTT,HTH,HHH)$$

with combinations

$$C_3^0 = 1 \qquad C_3^1 = 3 \qquad C_3^2 = 3 \qquad C_3^3 = 1\,.$$

The pattern in C_n^m (stand back and look at the page) matches the rows of Pascal's triangle.

Heuristic Derivation

There is a (perhaps) intuitive way to derive C_n^m. Take n coins and lay them all down showing T, represented by $C_n^0 = 1$. Turn any one of the coins to H and find $C_n^1 = n$. Turn any two of the coins to H and find

$$C_n^2 = n\frac{(n-1)}{2}\ ,$$

and for three,

$$C_n^3 = n\frac{(n-1)}{2}\frac{(n-2)}{3}\ ,$$

and so on. Building this up for m total H-faces, we find

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

the familiar combination number.

4.2 Binomial Distribution

Consider an *unbalanced* coin having inherent probability p to land on H(eads), and correspondingly $1-p$ to land on T(ails). As a generalized two-state system, a game of n tosses generates the same potential outcomes:

$$\begin{aligned}
\Omega_1 &= T,H \\
\Omega_2 &= TT,TH,HT,HH \\
\Omega_3 &= TTT,TTH,THT,THH,HTT,HTH,HHH
\end{aligned}$$

Of course, the probability P of generating m Heads-events requires an extra argument to account for the imbalance p. Denoting the modified combination symbol $P_n^m\,(p)$, the two-state analysis generalizes by:

$$\begin{aligned}
P_1^0\,(p) &= 1-p \\
P_1^1\,(p) &= p
\end{aligned}$$

$$\begin{aligned}
P_2^0\,(p) &= (1-p)^2 \\
P_2^1\,(p) &= 2\cdot p\,(1-p) \\
P_2^2\,(p) &= p^2
\end{aligned}$$

$$P_3^0(p) = (1-p)^3$$
$$P_3^1(p) = 3 \cdot p(1-p)^2$$
$$P_3^2(p) = 3 \cdot p^2(1-p)$$
$$P_3^3(p) = p^3$$

Evidently, the factors of p and $1-p$ compound into the terms p^m and $(1-p)^{n-m}$, but otherwise this analysis traces that of the two-state system exactly. Scanning for a pattern in the above, we evidently have

$$P_n^m(p) = \binom{n}{m}(1-p)^{n-m} p^m \, .$$

This result is known as the *binomial distribution*, and gives the probability of attaining, in general, m events of weight p among n trials:

$$P(m,n,p) = \frac{n!}{m!\,(n-m)!}(1-p)^{n-m} p^m \qquad (1.27)$$

Note there is no need to divide by 2^n. The binomial distribution as written is unit-normalized already.

Analysis

Define a random variable z_k that is equal to one if the event H with weight p occurs in the k-th trial, and is equal to zero otherwise. The average value of z_k is then

$$\begin{aligned}\langle z_k\rangle &= P(H)\,z(H) + P(T)\,z(T) \\ &= p \cdot 1 + (1-p)\cdot 0 = p \, ,\end{aligned}$$

and, simply enough, the average of z_k^2 reads

$$\left\langle z_k^2\right\rangle = p \cdot 1^2 + (1-p)\cdot 0^2 = p \, .$$

The standard deviation in z, denoted σ_z, is evidently

$$\sigma_z = \sqrt{\left\langle z_k^2\right\rangle - \langle z_k\rangle^2} = \sqrt{p-p^2} = \sqrt{p(1-p)} \, .$$

Next, note that the number m of H-events among the n independent trials is the sum

$$m = \sum_{k=1}^{n} z_k \, ,$$

implying

$$\left\langle \Delta m^2\right\rangle = \sum_{k=1}^{n}\left\langle \Delta z_k^2\right\rangle = n\left\langle \Delta z^2\right\rangle \, ,$$

or, in tighter notation for large-n systems,

$$\sigma_m = \sqrt{n\sigma_z^2} = \sqrt{np(1-p)} \, .$$

<u>Example 1</u>

Monique is practicing netball. She knows from past experience that the probability of her making any one shot is 70%. Her coach has asked her to keep practicing until she scores 50 goals. How many shots would she need to attempt to ensure that the probability of making at least 50 shots is more than 99%?

This problem is analogous to flipping a weighted coin with bias p. The multiplicity of scoring k shots in N tosses is

$$\Omega(k,N,p) = \frac{N!}{k!\,(N-k)!}(1-p)^{N-k} p^k \, ,$$

where summing over k gives the cumulative distribution:

$$99\% = \sum_{k=50}^{N} \frac{N!}{k!\,(N-k)!} .3^{N-k} .7^k$$

This is best solved by a computer, where one should find

$$N = 86 \, .$$

<u>Example 2</u>

Haldor the Viking has slain sixteen ooze creatures in the swamp. After a thorough forensic analysis, Haldor finds a single gold cup among the corpses. He remembers from swamp lore that a slain ooze has a 1/3 chance to drop a gold cup. What are the chances he found just one cup after slaying sixteen oozes? Repeat the calculation for finding two cups, three cups, etc., up to sixteen cups. Also account for zero cups.

Model a slain ooze as a weighted coin with a Heads probability of 1/3, and a Tails probability of 2/3, which calls for a straightforward application of the binomial distribution. For finding one gold cup, we have

$$\begin{aligned}P(16,1,1/3) &= \frac{16!}{1!\,(16-1)!}(2/3)^{16-1}(1/3)^1 \\ &= \frac{16}{3}\left(\frac{2}{3}\right)^{15} \approx 0.01218 \, ,\end{aligned}$$

and then for other numbers of gold cups:

$$P(16,2,1/3) \approx 0.04567$$
$$P(16,3,1/3) \approx 0.1066$$
$$P(16,4,1/3) \approx 0.1732$$
$$P(16,5,1/3) \approx 0.2078$$

$$P(16,6,1/3) \approx 0.1905$$
$$P(16,7,1/3) \approx 0.1361$$
$$P(16,8,1/3) \approx 0.07654$$
$$P(16,9,1/3) \approx 0.03402$$

$$
\begin{aligned}
P(16,10,1/3) &\approx 0.01191 \\
P(16,11,1/3) &\approx 0.003247 \\
P(16,12,1/3) &\approx 0.0006765 \\
P(16,13,1/3) &\approx 0.0001041
\end{aligned}
$$

$$
\begin{aligned}
P(16,14,1/3) &\approx 0.00001115 \\
P(16,15,1/3) &\approx 0.0000007434 \\
P(16,16,1/3) &\approx 0.00000002323 \\
P(16,0,1/3) &\approx 0.001522
\end{aligned}
$$

4.3 Multi-State System

A generalization of the two-state system is the *multi-state* system. Going for a modest example, consider a three-sided coin with faces A, B, C. Flipping such a coin to generate n total events, let:

- $n_A \rightarrow$ Number of outcomes A
- $n_B \rightarrow$ Number of outcomes B
- $n_C \rightarrow$ Number of outcomes C
- $p_A \rightarrow$ Probability of of outcome A
- $p_B \rightarrow$ Probability of of outcome B
- $p_C \rightarrow$ Probability of of outcome C

With this, we can write the probability of the three-state system exhibiting the state (n_A, n_B, n_C, n):

$$
P(n_A, n_B, n_C, n) = \frac{n!}{n_A! n_B! n_C!} \, p_A^{n_A} p_B^{n_B} p_C^{n_C}
$$

In the special case $C = 0$, the above reduces to the non-normalized binomial distribution.

4.4 Gaussian Distribution

Recall that the probability of generating k results among n total trials in a two-state system is given by

$$
P(k,n) = \frac{1}{2^n} \frac{n!}{k!\,(n-k)!} \; .
$$

Introduce the shift

$$
k \rightarrow k + \frac{n}{2} \; ,
$$

which modifies the above:

$$
P(k,n) = \frac{1}{2^n} \frac{n!}{\left(\frac{n}{2}+k\right)! \left(\frac{n}{2}-k\right)!}
$$

In the large-k limit, making k a continuous variable, it makes sense to describe the system solely in terms of expectation values and their deviations, a notion formally called the *central limiting theorem.* Here we develop this idea on a two-state system to derive a central equation in probability theory called the *Gaussian distribution.*

To proceed in the large n-limit, we deploy Stirling's approximation for large numbers

$$
\begin{aligned}
\ln(n!) &\approx n \ln(n) - n + \ln\left(\sqrt{2\pi n}\right) \\
n! &\approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \; ,
\end{aligned}
$$

and the probability density reduces to

$$
w(k) = e^{-2k^2/n} \sqrt{\frac{2}{\pi n}} \; . \tag{1.28}
$$

The result $w(k)$ is the famed normalized Gaussian distribution centered at $k = 0$. Introducing a nonzero shift of base-point value a, the generalized equation is

$$
w(k) = e^{-2(k-a)^2/n} \sqrt{\frac{2}{\pi n}} \; .
$$

Using Gaussian integrals, the average values and standard deviation are readily calculated:

$$
\begin{aligned}
\langle k \rangle &= \int_n k \cdot w(k) \; dk = a \\
\langle k^2 \rangle &= \int_n k^2 \cdot w(k) \; dk = \frac{n}{4} + a^2 \\
\sigma_k &= \sqrt{\langle k^2 \rangle - \langle k \rangle^2} = \sqrt{\frac{n}{4}}
\end{aligned}
$$

4.5 Poisson Distribution

Imagine trying to count the number of water molecules that pass a point in a river flowing at average speed v. Over time interval t, the average molecule count is directly proportional to vt. To reduce notation clutter, let us ignore the proportionality constant and take vt as a dimensionless quantity. Due to local random fluctuations in the river, an actual measurement would never precisely land on vt, but instead on an interval surrounding vt. Naturally we wonder, what is the time-varying probability $P_k(t)$ that k molecules are measured over the interval t?

To begin, partition the elapsed time t into n identical bins of width Δt such that $\Delta t \rightarrow 0$, and observe that each $P_k(\Delta t)$ relates to its $k-1$ and $k+1$ neighbors as:

$$
\lim_{\Delta t \rightarrow 0} P_0(\Delta t) \gg P_1(\Delta t) \gg P_2(\Delta t) \gg P_3(\Delta t) \gg \cdots
$$

This means it's more likely to measure few molecules in a small Δt-interval as opposed to many. We may proceed using weighted two-state analysis, wherein a Δt-interval may either be unfilled with zero molecules, or filled with one or more molecules. Borrowing the apparatus developed previously, we write

$$P\left(k, n, v\Delta t\right) = \frac{n!}{k!\left(n-k\right)!}\left(1 - v\Delta t\right)^{n-k}\left(v\Delta t\right)^{k} \ ,$$

where n and k are integers. Substituting $t = n\Delta t$, we have

$$P\left(k, n, vt\right) = \frac{\left(vt\right)^{k}}{k!}\left(\frac{n!}{\left(n-k\right)!}\frac{1}{n^{k}}\right)\left(1 - \frac{vt}{n}\right)^{n-k} \ .$$

In the large-n limit, the approximations

$$\begin{aligned} \frac{n!}{\left(n-k\right)!} &\approx n^{k} \\ \left(1 - \frac{vt}{n}\right)^{n-k} &\approx e^{-vt} \end{aligned}$$

are valid, and re-casting vt as a dimensionless variable q lands us at the anticipated *Poisson distribution*:

$$P_{k}\left(q\right) = \frac{q^{k}}{k!}e^{-q} \tag{1.29}$$

Summing over the variable k tells us $P_k\left(t\right)$ is already normalized:

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

With $P_k\left(t\right)$ on hand, we may calculate $\langle k\rangle$, $\left\langle k^2\right\rangle$, and the standard deviation:

$$\begin{aligned} \langle k\rangle &= \sum_{k=0}^{\infty} k\frac{q^{k}}{k!}e^{-q} = e^{-q}\sum_{k=1}^{\infty}\frac{q^{k}}{\left(k-1\right)!} \\ &= e^{-q}\sum_{p=0}^{\infty}\frac{q^{(p+1)}}{p!} = e^{-q}qe^{q} = q \\ \left\langle k^{2}\right\rangle &= \sum_{k=0}^{\infty} k^{2}\frac{q^{k}}{k!}e^{-q} = e^{-q}q\sum_{p=0}^{\infty}\left(p+1\right)\frac{q^{p}}{p!} \\ &= q + q^{2} \\ \sigma_{k} &= \sqrt{q^{2} + q - q^{2}} = \sqrt{q} \end{aligned}$$