Quant Interview Notes

Griffin Pinney

This document – first started in June of 2025 – is dedicated to my personal study for quantitative finance interviews, motivated by the realisation that writing down what one knows can be disproportionately useful. Though I'm making this document public in the hopes that it may aid others in their future interview preparation, I would like to stress that it was written purely with personal use in mind. As such, I've only included material relevant for my own study, and consequently, certain topics which I'm reasonably comfortable with (e.g. Markov chains) have been omitted, whereas other topics which many readers are likely comfortable with (e.g. Poisson processes) are discussed in detail. The document is structured as follows:

- **Some Content:** we begin by reviewing some basic content one needs to be familiar with to approach certain types of interview problems, so that the rest of the document may be better placed into context. (Again, we only review those topics which I happen to not already be confident with, so there will be large chunks of important material missing.)
- **Some Tricks:** we compile a collection of miscellaneous tricks and heuristics which are often useful in interview problems. This will be a mixture of the more 'classical' tricks which are widely taught, as well as some more 'personal' tricks I've identified through independent study. (Again, most of the tricks I feel I have already internalised will be omitted, but this collection of tricks will be more comprehensive than the collection of content from the previous section.)
- **The Classics:** we compile a collection of problems which seem to be repeatedly asked in interviews, or at least in various interview preparation materials. Though rote learning is not a good strategy in general, these questions are so common that their answers and methods of solution perhaps ought to be known by heart. When relevant, interesting variants of these classical problems are also explored.
- **Addressing Common (Personal) Mistakes:** the more one prepares, the more one realises that they are prone to repetitively making certain mistakes (unless one is already the perfect candidate). This section is devoted to addressing such mistakes which I personally tend to make, in the hopes that being acutely aware of them may help me avoid making them in the future. Essentially, every time during my preparation that I felt frustrated about getting an answer wrong or not having reached an answer in the optimal way, instead of beating myself up about it, I sought to understand why the mistake was made and how I should have thought about it differently, and wrote about it in this section. Naturally, this section is highly specific to me, but I have no doubt that some of my mistakes are shared by others who could benefit from reading these notes.

The problems discussed in this document will fall into one of four categories: probability/combinatorics, brainteasers, statistics, and other. To help keep them distinct, we enclose probability/combinatorics problems in blue boxes, brainteasers in red boxes, statistics problems in orange boxes, and other problems in grey boxes. Throughout the document, solutions are enclosed in green boxes.

Some Content

We dedicate this section to reviewing some basic content which, despite its simplicity, I have personally either struggled with or failed to internalise in the past. In each case, only the most important aspects of the given topic are outlined, with a more in-depth treatment to be sought elsewhere if need be. When relevant, we include example problems related to the material under discussion.

Discrete random variables

A discrete random variable is a random variable which takes on at most a countable number of distinct values, and its associated probability distribution (a *discrete distribution*) specifies the probability of each possible outcome. We list some common discrete distributions below together with their key properties, namely expectation and variance. One ought to be able to recall the defining characteristics of these distributions instantly, and derive their expectations and variances with ease. That being said, these quantities should also be known by heart, as this saves precious time when solving interview problems and demonstrates familiarity with the core material. It's my past failure to consistently recall some of the less memorable of these properties which has earned this topic a place in this document.

Uniform distribution

In a discrete uniform distribution, each outcome from within a finite set has an equal probability of occurring. That is, denoting the random variable by X, we have $\mathbb{P}(X = x) = 1/|S|$ for all $x \in S$, where S is some finite set. If $S = \{a, a+1, \cdots, b\}$, we write $X \sim \mathrm{Unif}\{a, \cdots, b\}$, and we have

$$\mathbb{E}[X] = \frac{b+a}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a+1)^2 - 1}{12}.$$

Bernoulli distribution

A Bernoulli distribution takes only the values 1 or 0, with probability p and $q := 1 - p$ respectively, where p is some parameter. That is, denoting the random variable by X, we have $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$, and we write $X \sim \mathrm{Bern}(p)$. For such a random variable, we have

$$\mathbb{E}[X] = p, \qquad \mathrm{Var}(X) = p(1-p) = pq.$$

Binomial distribution

A binomial distribution describes the probability of observing a given number of successes in n independent trials of an experiment which succeeds with probability p, where both n and p are parameters. Denoting the random variable by X, we write $X \sim \mathrm{B}(n, p)$ (or sometimes $X \sim \mathrm{Bin}(n, p)$), and we have

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

for each $k = 0, \cdots, n$. We note that there is a close connection between Bernoulli and binomial random variables: technically, Bernoulli random variables are the special case of binomial random variables in which $n = 1$, and conversely, $\mathrm{B}(n, p)$ may be described as the distribution of $X_1 + \cdots + X_n$ for i.i.d random variables $X_1, \cdots, X_n \sim \mathrm{Bern}(p)$. Recalling this makes the following properties easy to remember: for $X \sim \mathrm{B}(n, p)$, we have

$$\mathbb{E}[X] = np, \qquad \mathrm{Var}(X) = np(1-p) = npq.$$

Geometric distribution

A geometric distribution describes the number of independent trials needed to succeed in an experiment which succeeds with probability p and fails with probability $1-p$ (i.e. repeated samples from i.i.d $\text{Bern}(p)$ random variables), where p is some parameter. Denoting the random variable by X, we write $X \sim \text{Geom}(p)$, and we have

$$\mathbb{P}(X = k) = (1-p)^{k-1}p$$

for each $k = 1, 2, \cdots$, and

$$\mathbb{E}[X] = \frac{1}{p}, \qquad \text{Var}(X) = \frac{1-p}{p^2} = \frac{q}{p^2}.$$

The first of the above identities is used repeatedly, and is quite intuitive: if our experiment succeeds with probability p, we expect to have to repeat it $1/p$ times to see our first success. This can be derived from the definition of expectation by appealing to simple facts about power series, but it can also be proven more elegantly by appealing to recursion, via the law of total expectation: with probability p, we will see our first success after one trial, but with probability $1-p$, our first trial fails, in which case we are back where we started, and the number of additional trials we expect to perform until our first success is again described by $\mathbb{E}[X]$. It follows that $\mathbb{E}[X] = p + (1-p)(1+\mathbb{E}[X])$, which we solve to find $\mathbb{E}[X] = 1/p$. The formula for the variance of a geometric distribution is less intuitive, and one must resort to power series computations to derive it.

Negative binomial distribution

A negative binomial distribution describes the number of independent trials needed to achieve r successes in an experiment which succeeds with probability p and fails with probability $1-p$, where both r and p are parameters. Denoting the random variable by X, we write $X \sim \text{NB}(r, p)$, and we have

$$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r (1-p)^{r-k}$$

for each $k = r, r+1, \cdots$. We note that there is a close connection between negative binomial and geometric random variables: geometric random variables are the special case of negative binomial random variables in which $r = 1$, and conversely, $\text{NB}(r, p)$ may be described as the distribution of $X_1 + \cdots + X_r$ for i.i.d random variables $X_1, \cdots, X_r \sim \text{Geom}(p)$. Recalling this makes the following properties easy to remember: for $X \sim \text{NB}(r, p)$, we have

$$\mathbb{E}[X] = \frac{r}{p}, \qquad \text{Var}(X) = \frac{r(1-p)}{p^2} = \frac{rq}{p^2}.$$

Poisson distribution

A Poisson distribution describes the probability of a given number of events occurring in a fixed interval of time if these events occur with a known average rate λ, independently of the time since the last event. Denoting the random variable by X, we write $X \sim \text{Pois}(\lambda)$, and we have

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

for each $k = 0, 1, \cdots$, which can be derived as a limiting case of the binomial distribution as the number of trials goes to infinity and the expected number of successes remains fixed. As can be shown by simple power series manipulations, we have

$$\mathbb{E}[X] = \lambda, \qquad \text{Var}(X) = \lambda.$$

Continuous random variables

A continuous random variable is a random variable which takes on an uncountable number of distinct values, and its associated probability distribution (a *continuous distribution*) is described by its probability density function (PDF) and cumulative distribution function (CDF). Given a continuous random variable X, its PDF is a non-negative function f_X with the property that

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

for any interval $[a, b]$ (note: it's technically only *absolutely continuous* random variables which admit a PDF, so we are tacitly assuming that 'continuous random variable' really means 'absolutely continuous random variable', as seems to be common practice in the informal setting of interview problems); its CDF is defined by

$$F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t)\,dt,$$

from which it follows by the fundamental theorem of calculus that $f_X(x) = \frac{d}{dx}F_X(x)$. We list some common continuous distributions below together with their key properties, namely expectation and variance.

Uniform distribution

In a continuous uniform distribution, each outcome from within a fixed interval $[a, b]$ is equally likely to occur. Denoting the random variable by X, we write $X \sim U(a, b)$, which has the PDF and CDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(x) = \begin{cases} 0 & \text{for } x < a, \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b, \\ 1 & \text{for } x > b. \end{cases}$$

From this, one may derive

$$\mathbb{E}[X] = \frac{b+a}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Exponential distribution

An exponential distribution describes the waiting time until a single event occurs, under the assumption that events occur with a known average rate λ, independently of the time since the last event (note that this is the same setting under which the Poisson distribution is defined). Denoting the random variable by X, we write $X \sim \mathrm{Exp}(\lambda)$, which has the PDF and CDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

From this, one may derive

$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

It's occasionally useful to note that as a consequence of the above CDF, one has $\mathbb{P}(X > x) = e^{-\lambda x}$ for all $x \geq 0$, from which one may easily derive the *memorylessness property* of exponential random variables: $\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$ for all $s, t \geq 0$. That is, conditioned on a failure to observe an event during some initial period of time s, the distribution of the remaining waiting time is the same as the original unconditional distribution.

Normal distribution

A normal distribution (or sometimes *Gaussian distribution*) describes an important and particularly natural (in a sense made more precise by the central limit theorem) distribution with mean μ and variance σ^2, where both μ and σ^2 are parameters. Denoting the random variable by X, we write $X \sim \mathcal{N}(\mu, \sigma^2)$, which has the PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(we omit the CDF since it has no closed-form expression). That the PDF integrates to 1 over the real line may be proven by the classical trick of considering two i.i.d normal random variables and integrating their joint PDF over the plane using the polar coordinates formula, or by appealing to contour integration from complex analysis. As mentioned, we have the identities

$$\mathbb{E}[X] = \mu, \qquad \mathrm{Var}(X) = \sigma^2,$$

each of which may be derived by integrating. A key property of normal distributions is that any finite linear combination of independent normally distributed random variables is itself normally distributed (with easily determinable mean and variance), as may be proven by appealing to characteristic functions or by simply taking the convolution of the PDFs (and wading through some algebra). Moreover, for $X \sim \mathcal{N}(\mu, \sigma^2)$, we have $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$ for any constants $a, b \in \mathbb{R}$. A special case of note is that $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ – this latter distribution often being referred to as the *standard normal distribution*. Being a standardised form of the normal distribution, one may commit to memory the values of certain probabilities involving this distribution and use them to compute probabilities for the original distribution after applying the transformation $x \mapsto \frac{x-\mu}{\sigma}$. Some particularly well-known cases are that if $X \sim \mathcal{N}(0, 1)$, we have $\mathbb{P}(|X| < 1) \approx 0.68$, $\mathbb{P}(|X| < 2) \approx 0.95$, and $\mathbb{P}(|X| < 3) \approx 0.997$, which is referred to as the *68-95-99.7 rule*. Some examples of its use are as follows:

Problem

Suppose $X \sim \mathcal{N}(50, 4)$. Compute $\mathbb{P}(X > 52)$ to the nearest hundredth.

Solution

We have $\mu = 50$, and $\sigma^2 = 4$ hence $\sigma = 2$. Letting $Z = \frac{X-\mu}{\sigma}$, it follows by applying the transformation $x \mapsto \frac{x-\mu}{\sigma}$ that $\mathbb{P}(X > 52) = \mathbb{P}(Z > 1) = \frac{1}{2}(1 - \mathbb{P}(|Z| < 1))$, where the last equality is by the symmetry of the PDF of $Z \sim \mathcal{N}(0, 1)$. By the 68-95-99.7 rule, we have $\mathbb{P}(|Z| < 1) \approx 0.68$, and the answer is therefore $\frac{1}{2}(1 - 0.68) = \mathbf{0.16}$ to the nearest hundredth.

Problem

Suppose $X \sim \mathcal{N}(100, 36)$. Compute $\mathbb{P}(94 \leq X \leq 112)$ to the nearest hundredth.

Solution

We have $\mu = 100$, and $\sigma^2 = 36$ hence $\sigma = 6$. Letting $Z = \frac{X-\mu}{\sigma}$, it follows by applying the transformation $x \mapsto \frac{x-\mu}{\sigma}$ that $\mathbb{P}(94 \leq X \leq 112) = \mathbb{P}(-1 \leq Z \leq 2) = \frac{1}{2}\mathbb{P}(|Z| < 1) + \frac{1}{2}\mathbb{P}(|Z| < 2)$, where the last equality is by the symmetry of the PDF of $Z \sim \mathcal{N}(0, 1)$. By the 68-95-99.7 rule, we have $\mathbb{P}(|Z| < 1) \approx 0.68$ and $\mathbb{P}(|Z| < 2) \approx 0.95$, and the answer is therefore $\frac{0.68+0.95}{2} = \mathbf{0.82}$ to the nearest hundredth.

If x is sampled from $X \sim \mathcal{N}(\mu, \sigma^2)$, it's so common to use the value $\frac{x-\mu}{\sigma}$ that this is often given the name *z-score*, and can be thought of as the (signed) number of standard deviations which x is away from the mean.

Gamma distribution

A gamma distribution is parameterised by $\alpha, \lambda > 0$, called the *shape parameter* and *rate parameter* respectively. Denoting the random variable by X, we write $X \sim \mathrm{Gamma}(\alpha, \lambda)$, which has the PDF

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

(we omit the CDF since it has no closed-form expression). Here Γ, refers to the *gamma function* defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx,$$

noting that due to a change of variables in the integral, the division by $\Gamma(\alpha)$ normalises the PDF so that it integrates to 1.

In the case that the shape parameter α is a positive integer, $\mathrm{Gamma}(\alpha, \lambda)$ may be described as the distribution of $X_1 + \cdots + X_\alpha$ for i.i.d random variables $X_1, \cdots, X_\alpha \sim \mathrm{Exp}(\lambda)$, and therefore describes the waiting time until α events occur under the assumption that events occur with the known average rate λ, independently of the time since the last event. In particular, the distribution $\Gamma(1, \lambda)$ is the same as $\mathrm{Exp}(\lambda)$, which helps one remember the exact expression for $f_X(x)$ if all one can recall is that it has an exponential factor $e^{-\lambda x}$ together with some powers of λ and x. The relationship between the gamma distribution and exponential distribution also makes the following properties easier to remember: for $X \sim \mathrm{Gamma}(\alpha, \lambda)$ we have

$$\mathbb{E}[X] = \frac{\alpha}{\lambda}, \qquad \mathrm{Var}(X) = \frac{\alpha}{\lambda^2},$$

which holds even for non-integral α. The following is an example of some simple computations involving gamma distributions:

Problem

Let $X \sim \mathrm{Gamma}(\alpha, \lambda)$. Compute the *moment-generating function* $M_X(t) = \mathbb{E}[e^{tX}]$.

Solution

We have

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \int_{-\infty}^\infty e^{tx} f_X(x)\, dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda - t)x}\, dx \\
&= \frac{\lambda^\alpha}{(\lambda - t)^\alpha} \cdot \frac{(\lambda - t)^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x}\, dx \\
&= \left(\frac{\lambda}{\lambda - t}\right)^\alpha,
\end{aligned}$$

where the last line follows since $\frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x}$ is the PDF of a $\mathrm{Gamma}(\alpha, \lambda - t)$ random variable, and therefore integrates to 1 over the given domain. The answer is therefore $M_X(t) = \boldsymbol{\left(\frac{\lambda}{\lambda - t}\right)^\alpha}$.

We note that technically, the above computations are only valid for $t < \lambda$.

Beta distribution

A beta distribution is used for modeling random variables that take values in the interval $[0, 1]$, especially when the variable represents some unknown probability. Beta distributions are parameterised by $\alpha, \beta > 0$, and denoting the random variable by X, we write $X \sim \text{Beta}(\alpha, \beta)$, which has PDF given by

$$f_X(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

(we omit the CDF since it has no closed-form expression). Here B, refers to the *beta function* defined by

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx,$$

noting that the division by $B(\alpha, \beta)$ normalises the PDF so that it integrates to 1. A useful property of the beta function is that $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ which, together with the useful identity $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, allows one to easily derive the identities

$$\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta}, \qquad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

To be added later:

The central limit theorem and law of large numbers

Joint and marginal probability distributions

Covariance and correlation

Poisson processes

Martingales and random walks

Betting games

Linear regression

Some Tricks

This section is dedicated to outlining a collection of miscellaneous tricks which may occasionally be useful. Each trick is accompanied by some examples and, when relevant, some suggestions on identifying when it may be applicable.

Cancelling complicated events

If, in the process of computing a certain probability, one needs to calculate the probability of some other complicated event, it is sometimes the case that this complicated probability will cancel in the computation, and therefore need not be evaluated. For example:

Problem

If 150 coins are tossed, what is the probability that the number of heads is odd?

Solution

Let p denote the probability that the number of heads in the first 149 tosses is odd. Then, with probability p, we need the last toss to show tails, and with probability $1-p$, we need the last toss to show heads. It follows by the law of total probability that the answer is $\frac{1}{2}p + \frac{1}{2}(1-p) = \mathbf{\frac{1}{2}}$.

In the above example, the probability p was not known a priori, but its computation could be avoided due to cancellation. (Note: a posteriori, we also have $p = \frac{1}{2}$.)

The following is a somewhat transparent variant of the same problem::

Problem

There are n coins, one of which is fair, and the other $n-1$ of which show heads with probability $0 < \lambda < 1$. If all coins are tossed, what is the probability that the number of heads is odd?

Solution

Let $p(n, \lambda)$ denote the probability that the number of heads among the $n-1$ biased coins is odd. Then, with probability $p(n, \lambda)$, we need the fair coin to show tails, and with probability $1 - p(n, \lambda)$, we need the fair coin to show heads. It follows by the law of total probability that the answer is $\frac{1}{2}p(n, \lambda) + \frac{1}{2}(1 - p(n, \lambda)) = \mathbf{\frac{1}{2}}$.

Perhaps the above trick could be better described as 'collapsing' an event (in this case, the tossing of n coins) down to one *determining* sub-event (the tossing of the final coin). To spot applications of this trick, one should keep a sharp eye for steps in a sequential process which entirely determine the overall outcome. Here's one more example:

Problem

Emma has two 6-sided dice with values 1 - 6. One of the dice is fair, while the other is loaded such that each side appears in proportion to its value. If Emma rolls both dice, find the probability that the sum of the outcomes will be odd.

Solution

The outcome of rolling the biased die determines the required parity of the outcome of rolling the unbiased die. Whatever this parity is, it occurs with probability $\mathbf{\frac{1}{2}}$.

Seek a recurrence relation

Even when a problem fixes a certain number of objects (such as coins or dice) and asks for the probability of a particular event happening with those objects (or the number of ways in which a particular event can happen), it is sometimes the case that computing the analogous probability (or number of possibilities) in the generality of n objects is easier, especially when the relevant events for n and $n-1$ (and perhaps $n-2$, and so on...) objects can be nicely related to each other. In such cases, one should look for a recurrence relation expressing the probability of (or number of possibilities for) the relevant event for n objects in terms of the probability of (or number of possibilities for) the relevant event for $n-1$ objects, and hope that this recurrence relation either has a nice closed-form solution, or can simply be computed up to the desired n. If all other approaches have failed on a given problem, this can often be the crux move that leads to a solution.

Problem

A fair coin is flipped 10 times. Given that no heads appear consecutively in the sequence, find the probability that the first flip was a heads.

Solution

In order for no consecutive heads to appear in a sequence of 10 coin flips, the sequence must begin either T or HT, and then be followed by a sequence of either 9 or 8 flips respectively in which there are also no consecutive heads. This is recursive behaviour, so we seek to find for general n the number of ways in which a coin can be flipped n times with no consecutive heads appearing. Denoting this quantity by $f(n)$, the same logic as above gives $f(n) = f(n-1) + f(n-2)$, and since $f(1) = 2$ and $f(2) = 3$ we are led to $f(n) = F_{n+2}$, where F_n denotes the Fibonacci sequence beginning $F_0 = 0$, $F_1 = 1$. So, there are $F_{12} = 144$ ways in which a fair coin can be flipped 10 times with no consecutive heads, and among these, those for which the first flip is heads must begin with HT, followed by 8 flips with no consecutive heads, which can happen in $F_{10} = 55$ ways. It follows that the answer is $\mathbf{\frac{55}{144}}$.

We also have the following similar example:

Problem

Suppose a coin is flipped 10 times and the outcomes are recorded. Find the probability that any tails occur only in consecutive pairs. For example, with four flips, TTHH, TTTT, and HHHH are all valid, but HTHH and TTTH are not.

Solution

In order for tails to occur in consecutive pairs in a sequence of 10 coin flips, the sequence must begin either H or TT, and then be followed by a sequence of either 9 or 8 flips respectively in which any tails occur in consecutive pairs. This is recursive behaviour, so we seek to find for general n the number of ways in which a coin can be flipped n times in such a way that any tails occur in consecutive pairs. Denoting this quantity by $f(n)$, the same logic as above gives $f(n) = f(n-1)+f(n-2)$, and since $f(1) = 1$ and $f(2) = 2$, we are led to $f(n) = F_{n+1}$, where F_n denotes the Fibonacci sequence beginning $F_0 = 0$, $F_1 = 1$. So, there are $F_{11} = 89$ ways in which a fair coin can be flipped 10 times with any tails occurring in consecutive pairs. Since there are $2^{10} = 1024$ possible outcomes for the 10 coin flips, it follows that the answer is $\mathbf{\frac{89}{1024}}$.

Count the complement

This is an obvious strategy, but is so simple that it can sometimes be overlooked. If one ever finds oneself in the situation that a certain subset of a set is difficult to enumerate, it's worth considering whether the complement of that subset is easier to enumerate.

Problem

A dot is marked at each vertex of a triangle ABC. Then 2, 3, and 7 more dots are marked on the sides AB, BC, and CA respectively. How many triangles have their vertices at these dots?

Solution

There are $2 + 3 + 7 + 3 = 15$ dots in total. Clearly, there are $\binom{15}{3} = 455$ ways to choose three dots, but determining which of these choices result in non-degenerate triangles is difficult. Instead, we count those choices which lead to *degenerate* triangles, i.e. we count the complement. A degenerate triangle must have all three points collinear, and for three points in the given configuration to be collinear, they must be contained in a single side of the triangle. Including the vertices, side AB contains 4 dots, side BC contains 5 dots, and side CA contains 9 dots. Thus, there are $\binom{4}{3} + \binom{5}{3} + \binom{9}{3} = 98$ ways to choose dots collinearly, meaning the answer is $455 - 98 = \mathbf{357}$.

Compute the probability of the complement

The probabilistic version of the above trick is to compute the probability of the complement of an event if one finds the original probability difficult to calculate. This is worth considering any time it's either unclear how to approach the computation of the original probability, or if the computation feels messy with several cases to consider.

Problem

Suppose that $X \sim \text{Geom}(\frac{1}{4})$. Compute $\mathbb{P}(X \leq 8 | X \geq 5)$.

Solution

Recall that the distribution $\text{Geom}(p)$ describes the probability that it takes n trials to succeed in an experiment which succeeds with probability p. Thus, the information $X \geq 5$ says that this experiment failed in each of the first 4 trials, and we want the probability that the experiment succeeds within 8 trials. That is, we want the probability that the experiment succeeds within the next 4 trials. This can be partitioned into cases according to whether the experiment succeeds on the next trial, the trial after that, or the trial after that, and so on, but computing the probabilities of each of these events is more work than is necessary. We instead compute the probability of the complement, i.e. the probability that the next 4 trials of the experiment fail, which is $(\frac{3}{4})^4 = \frac{81}{256}$. The final answer is therefore $1 - \frac{81}{256} = \mathbf{\frac{175}{256}}$.

Computing probabilities for maxima of independent random variables

Suppose one has independent random variables $X_1, \cdots, X_n$. Note that it's easy to compute the probability that the maximum $M := \max(X_1, \cdots, X_n)$ is *at most* a given value, since $M \leq x$ if and only if $X_i \leq x$ for all $i = 1, \cdots, n$, hence $\mathbb{P}(M \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x)$. If one wishes to compute other probabilities involving the maximum, it is therefore useful to cast them in terms of probabilities of the form $\mathbb{P}(M \leq x)$ when possible, usually by taking the difference of two such terms. This can be useful in both the discrete and continuous settings. For example:

Problem

You roll two fair dice. What is the probability that the highest value rolled is a four?

Solution

Let X_1 and X_2 denote the outcomes of the first and second die rolls respectively. Then, $\mathbb{P}(\max(X_1, X_2) = 4) = \mathbb{P}(\max(X_1, X_2) \leq 4) - \mathbb{P}(\max(X_1, x_2) \leq 3) = \left(\frac{4}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = \mathbf{\frac{17}{36}}$.

Problem

Let $X_1, X_2, X_3, X_4 \sim \text{Unif}(0, 4)$ be i.i.d. Find the probability that the maximum of these 4 random variables is in the interval $(2, 3]$.

Solution

Let $M = \max(X_1, X_2, X_3, X_4)$. We have $\mathbb{P}(2 < M \leq 3) = \mathbb{P}(M \leq 3) - \mathbb{P}(M \leq 2) = \left(\frac{3}{4}\right)^4 - \left(\frac{2}{4}\right)^4 = \mathbf{\frac{65}{256}}$.

Recall basic facts from circle geometry

There are several problems which have the setup of some number of points being chosen uniformly at random from the circumference of a circle. Depending on what is being asked, it's sometimes useful to recall some basic facts from circle geometry in solving such problems. The most important of these are as follows:

- The angle subtended by an arc of a circle at its centre is twice the angle it subtends anywhere on the circle's circumference.
- As a consequence of the previous result, three points on the circumference of a circle form an obtuse triangle if and only if they all lie in a semicircle.
- Related to the previous point, three points on the circumference of a circle form a triangle which contains the centre of the circle if and only if they don't all lie in a semicircle.
- If N points on the circumference of a circle are to lie in a semicircle, then exactly one of the N points must be the 'clockwise-most' in the sense that the other $N - 1$ points lie on the counterclockwise semicircular arc starting at that point. That such a point exists is clear; that it is unique follows since the counterclockwise arc joining the last of the N points back to the first (ordered according to the existing counterclockwise semicircular arc) must be at least half the circumference long.

Problem

Given N points drawn uniformly at random on the circumference of a circle, what is the probability that they are all within a semicircle?

Solution

Label the points $p_1, \cdots, p_N$, and let A_i denote the event that all other points lie in the counterclockwise semicircular arc starting at the point p_i. Then, denoting the event that the N points are all within a semicircle by A, we have by the above facts from circle geometry that A is the disjoint union $A = \bigcup_{i=1}^N A_i$. It follows that $\mathbb{P}(A) = \sum_{i=1}^N \mathbb{P}(A_i) = N\mathbb{P}(A_1)$, where the last equality is a consequence of symmetry. It's clear that $\mathbb{P}(A_1) = \frac{1}{2^{N-1}}$, since each of the other $N - 1$ points must lie in a particular half of the circle, so the answer is $\mathbf{\frac{N}{2^{N-1}}}$.

Problem

What is the probability that the centre of a circle is contained within the triangle formed by choosing three points uniformly at random on its circumference?

Solution

Letting A denote the event that the 3 points are all within a semicircle, we have by the previous example that $\mathbb{P}(A) = 3/4$. By the above facts from circle geometry, it follows that the answer is $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = \mathbf{\frac{1}{4}}$.

Problem

2025 points are evenly distributed on the circumference of a circle. If a set of three of these points is chosen uniformly at random, find the probability that the three points form the vertices of an obtuse triangle.

Solution

There is a total of $\binom{2025}{3}$ ways to choose the three points; we must determine how many of these choices result in an obtuse triangle. By the above facts from circle geometry, this will be the case if and only if the three points lie in a semicircle, in which case exactly one of the three points will be the 'clockwise-most' in the sense that the other 2 points lie on the counterclockwise semicircular arc starting at that point. This point can be chosen in 2025 ways, and having chosen this, the remaining 2 points can be chosen in $\binom{1012}{2}$ ways. The answer is therefore

$$\frac{2025 \cdot \binom{1012}{2}}{\binom{2025}{3}} = \mathbf{\frac{3033}{4046}}.$$

If order is all that matters, focus only on order

Given i.i.d continuous random variables $X_1, \cdots, X_n$, each of their $n!$ orderings is equally likely. As such, if we wish to compute, say, $\mathbb{P}(X_1 < X_2 < \cdots < X_n)$, we need not launch into some highly complicated computation involving iterated integrals and the probability density function of the random variables X_i; instead, we simply observe that by symmetry, this particular ordering is just as likely as any other, hence the probability is $\frac{1}{n!}$ (to be more precise, we are using the *exchangeability* of this sequence of random variables).

More generally, one may use symmetry/intuition to help compute the probabilities of other events involving the relative positions of the random variables X_i. For example, in order to compute $\mathbb{P}(X_1 < X_2, X_1 < X_3, \cdots, X_1 < X_{10})$, we note that this event is equivalent to X_1 being the minimum of the 10 random variables $X_1, \cdots, X_{10}$. But by symmetry, each random variable is equally likely to be the minimum, hence the probability is $\frac{1}{10}$.

Problem

Let $X_1, \cdots, X_{20}$ be i.i.d $\text{Unif}(0,1)$ random variables. Compute $\mathbb{P}(X_1 > \max\{X_2, X_{10}, X_{15}\})$.

Solution

The event $X_1 > \max\{X_2, X_{10}, X_{15}\}$ is equivalent to X_1 being the maximum of the four random variables X_1, X_2, X_{10}, X_{15}. Of these four random variables, each is equally likely to be the maximum, and the answer is therefore $\mathbf{\frac{1}{4}}$.

This general idea of focusing only on order and using indirect symmetric arguments to avoid tedious computations extends to situations other than that described above in which one has a sequence of i.i.d continuous random variables. For example, if one is concerned with the relative ordering of the outcomes of certain discrete experiments such as die rolls, it helps to make observations along the lines of certain outcomes being equally likely due to symmetry (for example, if you're repeatedly rolling a die, you're just as likely to roll a 1 before a 2 as you are to roll a 2 before a 1). A good concrete example is demonstrated in the next problem, which also falls under the umbrella of the following trick.

Extend a possibly terminating process past the end

If one is performing a sequential process which may terminate under certain conditions, and one wishes to compute some probability or expectation regarding the point of termination, it is often useful to extend the process past the point of termination until all outcomes have been observed.

Problem

You roll a die until you observe a 5. What is the expected minimum number rolled?

Solution

In any instance of this experiment, we can imagine continuing to roll the die until all five outcomes other than six have been observed. These five outcomes must be observed in some order, so we may define for each $i = 1, \cdots, 5$ a random variable X_i which describes the placement of the outcome i in this ordering. Then, the minimum number rolled before 5 will be 1 if and only if $X_1 < X_5$; but by symmetry, this happens with probability $\frac{1}{2}$. Similarly, the minimum number rolled before 5 will be 2 if and only if $X_2 < X_5 < X_1$, which happens with probability $\frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$ (since X_2 is the minimum of these three random variables with probability $\frac{1}{3}$, and then X_5 is the minimum of the remaining two random variables with probability $\frac{1}{2}$). Extending this pattern, we see that the answer is

$$1 \cdot \frac{1}{2} \cdot \frac{1}{1} + 2 \cdot \frac{1}{3} \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} \cdot \frac{1}{3} + 4 \cdot \frac{1}{5} \cdot \frac{1}{4} + 5 \cdot \frac{1}{5} = \mathbf{\frac{137}{60}}.$$

Use conditional probability for one event occurring before another

Sometimes, conditional probability may be used to simplify the computation of the probability that one event occurs before another from a collection of possible events. Indeed, suppose an experiment can give one of three outcomes A, B, or C, and we are interested in the probability that event A occurs before event B if the experiment is repeated indefinitely. In this case, any trial of the experiment which gives event C can be ignored, and it's clear that the probability that A occurs before B is then given by $\mathbb{P}(A|A \cup B)$.

Problem

Suppose you are rolling a pair of 6-sided dice. If you roll a total of 7 you win, and if you roll a total of 11, you lose. If you get any other total, you continue rolling. What is the probability of winning?

Solution

Using the above trick, we compute the probability of rolling a total of 7 given that a total of either 7 or 11 was rolled. Each pair of dice rolls is equally likely, and there is a total of 6 ways to roll a 7 and a total of 2 ways to roll an 11. The answer is therefore $\frac{6}{6+2} = \mathbf{\frac{3}{4}}$.

We note that the above problem may also be solved by standard recursive techniques. For example, on any given roll, one wins with probability $\frac{6}{36}$ and loses with probability $\frac{2}{36}$, hence the game continues with probability $1-(\frac{6}{36}+\frac{2}{36})=\frac{28}{36}$. If we let p denote the probability of winning, it follows that $p=\frac{6}{36}+\frac{28}{36}p$, and we can solve to find $p=\frac{3}{4}$. More generally, any of this class of problems can be solved using Markov chain techniques, but the conditional probability approach sometimes simplifies matters, as is more evident in the following example:

Problem

In a tennis match starting at 30-30, with Alice having a probability of 0.6 of winning each serve independently, what is the probability that Alice wins the game (given standard tennis scoring rules)?

Solution

We note that if the score is tied (whether it be 30-30 or deuce), a player wins if they score the next two points in a row; otherwise, the game returns to a tie. Letting A denote the event that Alice wins two points in a row starting from a tie, and B the event that Alice's opponent wins two points in a row starting from a tie, it follows that Alice wins if and only if A occurs before B, which happens with probability $\mathbb{P}(A|A\cup B)=\frac{0.6^2}{0.6^2+0.4^2}=\mathbf{\frac{9}{13}}$.

The method of reduced sample space

The method of reduced sample space is a technique that simplifies problems by focusing only on the outcomes relevant to the event in question. The sample space is reduced to only include outcomes that contribute to the result, thereby simplifying calculations.

Problem

Alice rolls a fair 6-sided die until she obtains her first 6. What is the probability that Alice obtains the value 5 exactly four times before she stops rolling?

Solution

The only relevant outcomes of any given roll are 5 and 6, so we ignore all rolls which give any other value. In this case, each roll is equally likely to be either a 5 or a 6, so the probability of observing the sequence 55556 is $\left(\frac{1}{2}\right)^5=\mathbf{\frac{1}{32}}$.

Problem

Alice rolls a fair 6-sided die until she obtains her first 6. What is the probability that Alice sees exactly two 2's and two 3's before she stops rolling?

Solution

The only relevant outcomes of any given roll are 2, 3, and 6, so we ignore all rolls which give any other value. In this case, each roll is equally likely to be either a 2, 3, or 6, and there are $\binom{4}{2}=6$ ways to arrange the two 2's and two $3's$ occurring before the 6, so the answer is $6\cdot\left(\frac{1}{3}\right)^5=\mathbf{\frac{2}{81}}$.

We note that in general, we are replacing the probability of each relevant event with the probability that that particular event has occurred given that *some* relevant event has occurred.

The tail-sum formula

The tail-sum formula is an elegant trick for computing expectations of **non-negative** random variables. Traditionally, the formula is applied to discrete random variables and takes the form

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$$

(hence the name *tail-sum formula*), but an analogous identity also holds for non-negative continuous random variables, namely

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > t)\, dt.$$

These identities are consequences of the classic layer-cake decomposition from analysis, and are useful whenever the probabilities $\mathbb{P}(X > n)$ are easier to compute than the probabilities $\mathbb{P}(X = n)$ (which is often the case when X is realised as the minimum of some family of random random variables, for instance). If one ever finds that computing expectations according to the traditional formula $\sum_{n=0}^{\infty} n\mathbb{P}(X = n)$ (or its continuous analogue) is overcomplicated, it's worth trying the tail-sum formula.

The discrete tail-sum formula is often written in the equivalent form $\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n)$, though one must be careful to note that the sum begins at index 1 in this case. Clearly, the continuous version may also be written $\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X \geq t)\, dt$.

Problem

You play a game in which you choose a number each round, the winning number on the nth round being selected uniformly at random from the set $\{1, 2, \cdots, n+2\}$. Let W be the random variable representing how many rounds you play before you win. What is $\mathbb{E}[W]$?

Solution

Computing the expectation according to the traditional formula leads to

$$\mathbb{E}[W] = \sum_{n=0}^{\infty} n\mathbb{P}(W = n) = \sum_{n=0}^{\infty} \frac{n}{n+2} \left(\prod_{j=2}^{n} \frac{j}{j+1} \right) = \sum_{n=0}^{\infty} \frac{2n}{(n+1)(n+2)}.$$

Though it's clear by elementary convergence tests that this expectation is infinite, the result and intermediate computations are much clearer if we use the tail-sum formula. Indeed, given $n \geq 1$, we have $W > n$ if and only if we do not win on rounds $1, 2, \cdots, n$, which occurs with probability $\mathbb{P}(W > n) = \prod_{j=1}^{n} \frac{j+1}{j+2} = \frac{2}{n+2}$. Hence, $\mathbb{E}[W] = 1 + \sum_{n=1}^{\infty} \frac{2}{n+2}$ **is infinite**.

Problem

You randomly throw a dart at a dartboard of radius 1. What is the expected distance of the dart from the centre of the board?

Solution

Let R denote the distance the dart lands from the centre of the dartboard. One could compute the PDF of R by first computing the CDF and differentiating, and then use the PDF to compute the expectation. However, this takes more steps than is necessary. We note that $\mathbb{P}(R > x) = \pi(1 - x^2)/\pi = 1 - x^2$, hence $\mathbb{E}[R] = \int_0^1 (1 - x^2)\, dx = \mathbf{\frac{2}{3}}$, by the continuous tail-sum formula.

Appeal to parity

Sometimes, one can make a simple appeal to parity to demonstrate that certain events of interest are impossible, or to otherwise shed light on the possible outcomes of a given experiment. This is worth considering whenever considerations of parity are already built into a problem, or if a problem requires a finite collection of numbers to be chosen or arranged to satisfy a certain condition.

Problem

Consider the sequence $1 \_ 2 \_ 3 \_ 4 \_ 5 \_ 6 \_ 7 \_ 8 \_ 9$, where each blank can be filled with either a $+$ or a $-$ with equal probability. What is the probability that the final result, after performing all operations, equals 0?

Solution

We note that there are four even terms and five odd terms. The equation will therefore always equal 1 when reduced mod 2, which is to say the result is always odd regardless of how the symbols are inserted. In particular, the result can never be 0, so the probability of achieving this final result is **0**.

Problem

Hannah has 12 blocks, 2 each of 6 distinct colors. She randomly arranges the blocks in a straight line. What is the probability that there are an even number of blocks between every identically coloured pair of blocks?

Solution

Picture 12 slots in which blocks may be placed, numbered from 1 to 12. We note that there is an even number of spaces between two blocks if and only if their corresponding slots have numbers of opposite parity. It follows that there will be an even number of blocks between every identically coloured pair if and only if there is one block of each colour in the odd-numbered slots and one block of each colour in the even-numbered slots. This can be achieved in $(6!)^2$ ways, and since there's a total of $12!/(2!)^6$ ways to arrange the blocks with no restrictions, the answer is $\frac{(6!)^2}{12!/(2!)^6} = \mathbf{\frac{16}{231}}$.

The hockey-stick identity

The hockey-stick identity is the simple combinatorial identity

$$\sum_{i=r}^{n} \binom{i}{r} = \binom{n+1}{r+1},$$

which holds for all natural numbers n and r with $n \geq r$. This identity derives its name from its appearance when the relevant binomial coefficients are highlighted on Pascal's triangle: the terms on the left-hand side form a diagonal line stemming from the right-hand edge of the triangle, with the term on the right-hand side jutting out from the end of this line like the head of a hockey stick. The visualisation itself makes the form of the identity easy to remember, but one must take care to note that the 'hockey stick' must originate from the edge of the triangle, i.e. the first term in the sum must be of the form $\binom{r}{r}$.

The hockey-stick identity is likely to be useful whenever one must evaluate a sum of binomial coefficients in which the term on the bottom remains constant across all summands. For example:

Problem

There are 20 balls in an urn labeled from 1 through 20. You pick 10 balls out of this urn. What is the expected maximum value of the 10 balls you picked out?

Solution

The maximum value can be any of $k = 10, 11, \cdots, 20$. For the maximum value to be k, the 9 other balls must have been drawn from the collection of balls labeled $1, \cdots, k-1$, and there are therefore $\binom{k-1}{9}$ ways for this to happen. Since there are $\binom{20}{10}$ ways for the balls to be chosen without any restrictions, each of which is equally likely, it follows that the expected value of the maximum ball drawn is

$$\begin{aligned}\frac{1}{\binom{20}{10}} \sum_{k=10}^{20} k\binom{k-1}{9} &= \frac{10}{\binom{20}{10}} \sum_{k=10}^{20} \binom{k}{10} \\ &= \frac{10}{\binom{20}{10}} \cdot \binom{21}{11} = \mathbf{\frac{210}{11}},\end{aligned}$$

where the last line follows by the hockey-stick identity.

The sum of squares formula

While everybody is familiar with the formula $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$, the formula for the sum of consecutive squares is slightly harder to remember:

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Despite the above identity being less widely known, one is occasionally expected to compute a sum of squares, in which case knowing the formula saves considerable time and demonstrates good breadth of knowledge.

Problem

Example coming!

Solution

Solution coming!

Keeping room here for the above problem and example...

To be added later:

- Symmetric strategies in two-player games
- Recall basic facts about order statistics
- Tiling a grid
- Forming words with no adjacent characters the same
- Using Catalan numbers for certain counting problems
- A martingale betting trick
- Stirling's approximation

The Classics

This section is dedicated to a series of problems that could be considered 'classical' in the sense that they seem to have been asked repetitively, and therefore ought to be known by heart. The extent to which such questions are *still* asked is unclear, but it would be wise to internalise these problems nonetheless so as to never be caught off guard. The questions are not given in any particular order, and those which happen to have been covered in the previous section are not repeated.

$n+1$ coins vs n coins

Problem

Alice has $n+1$ fair coins and Bob has n fair coins. What is the probability that Alice will flip more heads than Bob if both flip all their coins?

Solution

Let E_1 denote the event that Alice flips the same number of heads as Bob in her first n flips, and let E_2 denote the event that Alice flips more heads than Bob in her first n flips. By the law of total probability, the answer is then $\frac{1}{2}\mathbb{P}(E_1) + \mathbb{P}(E_2) = \frac{1}{2}(\mathbb{P}(E_1) + 2\mathbb{P}(E_2))$.
Now, by symmetry, we have $\mathbb{P}(E_2) = \mathbb{P}(E_3)$, where E_3 denotes the event that Alice flips fewer heads than Bob in her first n flips. Since $\mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) = 1$, it follows that $\mathbb{P}(E_1) + 2\mathbb{P}(E_2) = 1$, from which we conclude that the answer is $\mathbf{\frac{1}{2}}$.

Though by no means classical, it is worth mentioning the following considerably more difficult (or perhaps just more annoying) variant of the above problem:

Problem

Alice has $n+2$ fair coins and Bob has n fair coins. What is the probability that Alice will flip more heads than Bob if both flip all their coins?

Solution

We let E_1, E_2, E_3 denote the same events as in the previous solution, further subdividing E_3 into the disjoint events E_3' and E_3'', corresponding to Alice tossing exactly one less heads than Bob or more than one fewer heads than Bob respectively, in her first n tosses. In the case of event E_1, Alice must flip at least one heads in her remaining two tosses to win, and in the case of event E_2, Alice has already won. However, unlike the previous problem in which case Alice cannot win in the case of event E_3, this time Alice can win in the case of event E_3' by flipping two heads on her remaining two tosses. It follows that the answer is $\frac{3}{4}\mathbb{P}(E_1) + \mathbb{P}(E_2) + \frac{1}{4}\mathbb{P}(E_3') = \frac{1}{2} + \frac{1}{4}\mathbb{P}(E_1) + \frac{1}{4}\mathbb{P}(E_3')$ (where we have used the fact that $\mathbb{P}(E_1) + 2\mathbb{P}(E_2) = 1$, as in the previous solution), and we have no choice but to get our hands dirty and compute. First, we have:

$$\mathbb{P}(E_1) = \sum_{k=0}^{n} \binom{n}{k}^2 \left(\frac{1}{2}\right)^{2n} = \left(\frac{1}{2}\right)^{2n} \sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k} = \left(\frac{1}{2}\right)^{2n} \binom{2n}{n},$$

where we have used Vandermonde's identity. Next, we have:

$$\mathbb{P}(E_3') = \sum_{k=1}^{n} \binom{n}{k}\binom{n}{k-1} \left(\frac{1}{2}\right)^{2n} = \left(\frac{1}{2}\right)^{2n} \sum_{k=1}^{n} \binom{n}{n-k}\binom{n}{k-1} = \left(\frac{1}{2}\right)^{2n} \binom{2n}{n-1},$$

where we have again used Vandermonde's identity (which is more evident after reindexing the sum). It then follows by Pascal's formula that the answer is $\mathbf{\frac{1}{2} + \left(\frac{1}{2}\right)^{2n+2} \binom{2n+1}{n}}$.

Basketball practice

Problem

Frank is shooting free throws. He makes his first free throw and misses his second. For $n \geq 3$, the probability of making the nth free throw is equal to the proportion of free throws he made during his first $n-1$ attempts. What is the probability that Frank makes exactly 50 free throws in 100 attempts?

Solution

Simple calculations reveal that after three attempts, Frank makes a total of either one or two free throws each with a probability of $1/2$. After investigating a couple more small cases, we are led to the hypothesis that the distribution of the number of free throws made up to and including the nth attempt is uniform over the set of all possibilities (i.e. 1 through $n-1$). Indeed, assuming that this is true for n, it follows that after $n+1$ attempts, only 1 free throw is made if first only 1 free throw is made from the first $n-1$ attempts, and then the next free throw misses; by induction, this occurs with probability $\frac{1}{n-1}\frac{n-1}{n} = \frac{1}{n}$. Similarly, after $n+1$ attempts, n free throws are made if first $n-1$ of the first n free throws are made, and then the next free throw is made; by induction, this occurs with probability $\frac{1}{n-1}\frac{n-1}{n} = \frac{1}{n}$. For $1 < k < n$, we see that k free throws are made if either k are made from the first n attempts and then the next free throw is missed, or $k-1$ are made from the first n attempts and then the next free throw is made. Again by induction, this occurs with probability $\frac{1}{n-1}\frac{n-k}{n} + \frac{1}{n-1}\frac{k-1}{n} = \frac{1}{n}$. This proves our hypothesis, and we conclude that Frank makes exactly 50 out of 100 free throws with probability $\mathbf{1/99}$.

We also have the following simple variant:

Problem

Frank is shooting free throws. He makes his first free throw and misses his second. For $n \geq 3$, the probability of making the nth free throw is equal to the proportion of free throws he made during his first $n-1$ attempts. What is the expected number of free throws that Frank makes in 100 attempts?

Solution

Using the results of the previous solution, we simply take the expectation of a uniformly distributed discrete random variable, and it is immediate that the answer is 50. Supposing one does not realise the distribution is uniform as proven above, there is also a more direct way to solve this version of the problem by deriving a recurrence relation. Indeed, let X_n denote the number of free throws made after n attempts. Then, $X_n = X_{n-1} + (X_n - X_{n-1})$, and we have by the law of total expectation that

$$\mathbb{E}[X_n] = \sum_{k=1}^{n-2} \mathbb{P}(X_{n-1} = k)\mathbb{E}[X_{n-1} + (X_n - X_{n-1})|X_{n-1} = k].$$

Clearly, $\mathbb{E}[X_{n-1}|X_{n-1} = k] = k$, and if $X_{n-1} = k$, $X_n - X_{n-1}$ is a Bernoulli random variable with probability of success $\frac{k}{n-1}$, hence $\mathbb{E}[X_n - X_{n-1}|X_{n-1} = k] = \frac{k}{n-1}$. It follows that

$$\mathbb{E}[X_n] = \sum_{k=1}^{n-2} \mathbb{P}(X_{n-1} = k)\left(k + \frac{k}{n-1}\right) = \mathbb{E}[X_{n-1}] + \frac{\mathbb{E}[X_{n-1}]}{n-1} = \frac{n}{n-1}\mathbb{E}[X_{n-1}].$$

Noting that $X_2 \equiv 1$, it follows by a telescoping product argument that $\mathbb{E}[X_n] = \frac{n}{2}\mathbb{E}[X_2] = \frac{n}{2}$. In particular, $\mathbb{E}[X_{100}] = \mathbf{50}$.

First ace

Problem

On average, how many cards in a normal deck of 52 playing cards do you need to flip over to observe your first ace?

Solution

The placement of the four aces partitions the remaining 48 cards into five groups of size $X_1, \cdots, X_5$. We have $\sum_{i=1}^{5} X_i = 48$, and by symmetry, $\mathbb{E}[X_1] = \cdots = \mathbb{E}[X_5]$. It follows by the linearity of expectation that $\mathbb{E}[X_1] = \frac{48}{5}$, hence the expected number of cards we must flip to observe the first ace is $\frac{48}{5} + 1 = \frac{53}{5} = \mathbf{10.6}$

Cats and dogs

Problem

Six dogs and six cats are sitting at a circular table uniformly at random. Find the probability that there are at least four dogs in a row somewhere in the circle.

Solution

There will be four dogs in a row somewhere if and only if, reading clockwise around the table, there is some starting position from which the sequence of cats and dogs begins CDDDD. Letting A_i denote the event that this sequence begins at seat i, the answer is given by

$$\mathbb{P}\left(\bigcup_{i=1}^{12} A_i\right).$$

Now, the events A_i are pairwise disjoint (if there is overlap between the strings of five seats beginning from seats $i \neq j$, then $A_i \cap A_j \neq \emptyset$ would imply that some seat has to have both a dog and a cat in it, which is impossible; if there is no overlap, $A_i \cap A_j \neq \emptyset$ would imply there is at least eight dogs, which is also impossible). Since the events A_i are all equally likely by symmetry, it follows that the answer is equal to $12\mathbb{P}(A_1)$. But it is clear that $\mathbb{P}(A_1) = \frac{6}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} \cdot \frac{4}{9} \cdot \frac{3}{8}$, since the animals have been seated uniformly at random, hence the animal in seat 1 is a cat with probability $\frac{6}{12}$, after which the animal in seat 2 is a dog with probability $\frac{6}{11}$, and so on. Multiplying by 12, we see that the answer is $\mathbf{\frac{3}{11}}$.

The birthday paradox

Problem

Assuming 365 days in a year, how many people do we need in a class to make the probability that at least two people have the same birthday more than $\frac{1}{2}$?

Solution

Given n people, the probability that they all have distinct birthdays is $\prod_{k=1}^{n-1} \frac{365-k}{365}$, hence the probability that at least two people share a birthday is $f(n) = 1 - \prod_{k=1}^{n-1} \frac{365-k}{365}$. The answer is the first n for which this is at least $\frac{1}{2}$, which turns out to be **23**.

Bubble sort

Problem

Suppose you run one iteration of bubble sort on a random permutation of $(1, 2, \cdots, 100)$. What is the probability that the list is now sorted?

Solution

At each step in an iteration of bubble sort, one considers positions i and $i+1$ and decides to either swap them or not. In our case, this is done at positions $i = 1, \cdots, 99$, and since there are 2 choices at each position, we conclude there are 2^{99} permutations of $(1, 2, \cdots, 100)$ which will be sorted after one iteration. The answer is therefore $\mathbf{\frac{2^{99}}{100!}}$.

To be more precise, one can exhibit a bijection between $\{0,1\}^{99}$ and the favourable permutations of $(1, 2, \cdots, 100)$ as follows: given any $(x_1, \cdots, x_{99}) \in \{0,1\}^{99}$ and starting with the fully sorted list $(1, 2, \cdots, 100)$, we iterate in i from 99 to 1, at each step either swapping elements i and $i+1$ or doing nothing, according to whether $x_i = 1$ or $x_i = 0$ respectively. It's clear that this map is injective and maps surjectively to the set of favourable permutations of $(1, 2, \cdots, 100)$.

Alternative Solution

The above solution – like most in this document – is not my own. However, I would like to record the solution I found when I first saw this problem, which is of a similar spirit (i.e. one finds a clever bijection to a set that we can count).

We note that in an iteration of bubble sort, each element can move at most one step backwards. Therefore, if one iteration of bubble sort is to fully sort the list, any element i which is ahead of its natural position can only be so by one step. One then sees without much difficulty that a favourable permutation must be comprised of successive blocks of the form $(i+n, i, i+1, \cdots, i+n-1)$, where the term $i+n$ is found at position i of the overall permutation. Such a block is entirely determined by its pair of endpoints, which establishes a bijection between the favourable permutations of $(1, 2, \cdots, 100)$ and subsets of $\{1, \cdots, 100\}$ which have an even number of elements. There are $\sum_{j=0}^{50} \binom{100}{2j} = 2^{99}$ such subsets, leading to the answer $\mathbf{\frac{2^{99}}{100!}}$.

Probability of two daughters given at least one daughter

Problem

Given that someone with two children has at least one daughter, what is the probability that they have two daughters?

Solution

A simple application of Bayes' rule reveals that the answer is $\mathbf{\frac{1}{3}}$. Alternatively, if someone has four children, the outcomes for the genders are uniformly distributed over the sample space $\{(b, g), (g, b), (b, b), (g, g)\}$. Conditioning on there being at least one daughter, the distribution remains uniform, but over the restricted sample space $\{(b, g), (g, b), (g, g)\}$. Having two daughters, which is to say the event (g, g), therefore occurs with probability $\mathbf{\frac{1}{3}}$.

Probabilistic strategy

Problem

Alice and Bob are placed in separate rooms. Alice will flip a coin, and Bob will also flip a coin. Alice must guess the outcome of Bob's coin toss, while Bob must guess the outcome of Alice's coin toss. They will win if both guesses are correct. Before the game, they're allowed to discuss and devise a strategy. What is the probability of their winning, assuming they use an optimal strategy?

Solution

Obviously, if Alice and Bob each agree to a fixed guess, they will win with probability $\frac{1}{4}$. We look for an outcome-dependent strategy which beats these odds, and the simple strategy of both Alice and Bob making the same guess as the outcome of their own toss comes to mind. Under this strategy, they win if their tosses come up either HH or TT, which happens with probability $\boldsymbol{\frac{1}{2}}$.

Though it's intuitively clear that the above is optimal, let's take some care to prove it. First, we note that any probabilistic strategy employed by both players is equivalent to them employing a deterministic strategy with some probability. It follows by the law of total probability that any upper bound we may find on the odds of winning across all deterministic strategies will also be an upper bound on the odds of winning across all probabilistic strategies. It therefore suffices to find a deterministic strategy which is at least as good as every other deterministic strategy, and we claim that ours is such a strategy. Indeed, if a deterministic strategy works in the case of some combination $(x, y) \in \{H, T\}^2$, then it can't work in either of the cases (x, x) or (y, y), and may or may not work in the case (y, x). That is, any deterministic strategy can work in at most two out of four cases, so our result of $\frac{1}{2}$ is indeed optimal.

Personally, I find the above problem somewhat tricky, mostly because the right strategy feels too 'stupid', and even when one arrives at this strategy, the argument for optimality is not immediately clear. However, having seen the above problem, the following simple variant becomes easy (noting that optimality is clear in this case):

Problem

Under the same setup as the previous problem with the new rule that Alice and Bob win if at least one of their guesses is correct, what is the probability of winning assuming Alice and Bob use an optimal strategy?

Solution

We again explore simple outcome-dependent deterministic strategies, expecting to beat our answer of $\frac{1}{2}$ to the previous problem since the win condition in this problem is less stringent. Without too much difficulty, we happen upon the following strategy: have Alice guess the same outcome as her own toss, and have Bob guess the opposite outcome of his toss. Whatever the outcome of Alice's toss, they will then win due to Alice's guess if Bob's toss matches hers, and they will win due to Bob's guess otherwise. They therefore win with probability $\mathbf{1}$ under this strategy, which is clearly optimal.

We consider a third (and considerably more difficult) problem in the same vein:

Problem

Three wizards are seated in a circular room. A magician will place one hat on each wizard's head, where each hat is either black or white, chosen uniformly at random. Each wizard cannot see their own hat but can see the hats of the other two wizards. Upon hearing a bell, all wizards must react simultaneously by either announcing the colour of their own hat or remaining silent. If at lease one wizard makes an announcement and all announcements are correct, the wizards win the game collectively (and if all wizards remain silent they lose). The wizards can confer and devise a strategy beforehand. What is the maximum probability of winning if they play optimally?

Solution

Each wizard will observe some element of the set $S = \{(B,B),(W,W),(B,W)\}$, so we must essentially decide on a function $f : S \to \{B,W,A\}$ for each of the three wizards, where A means 'abstain'. If we have each wizard adopt the function which maps $(B,B) \mapsto W$, $(W,W) \mapsto B$, and $(B,W) \mapsto A$, then the wizards will win in all six cases in which they're not all wearing the same colour (since the two of them who share a colour abstain from voting, and the one who wears a different colour makes the correct prediction). All wizards make the wrong prediction in the remaining two cases where all wizards wear the same colour, so the wizards win under this strategy with probability $\frac{6}{8} = \mathbf{\frac{3}{4}}$.

To see that this is optimal, we may restrict ourselves to symmetric strategies (i.e. each wizard choosese the same function f) due to the symmetry of the problem. If the function f maps (B,W) to either B or W, it's easy to see that this strategy fails in at least three out of six cases, and can therefore not beat the strategy found above. Hence, f must map (B,W) to A, and it remains to decide where f sends (B,B) and (W,W). There are only a handful of cases to consider, and examining the cases reveals that $(B,B) \mapsto W$ and $(W,W) \mapsto B$ is the best choice.

The Monty Hall problem

This classic probabilistic 'paradox' is so widely known that it is unlikely to be asked in any interview, but it is important to understand the misconception which leads to the apparent paradox so that it may be avoided in other contexts. First, we take the conditional probability of the event A given B to be axiomatically defined by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

In the event that the underlying probability space has finite cardinality and the probability measure is uniform, this formula simplifies to $\mathbb{P}(A|B) = \frac{|A \cap B|}{|B|}$, and in particular, each element of the set B is assigned probability $1/|B|$, so the conditional probability measure remains uniform. This observation is useful in various contexts; for example, suppose one wishes to know the probability of having rolled a 6 on a fair 6-sided die given that the value rolled was no less than 3. The outcomes of the die roll are uniformly distributed, and they remain so given the conditioning, so the answer is simply $\frac{1}{4}$. The key pitfall in apparent paradoxes like the Monty Hall problem is that this reasoning only applies to uniform distributions, and the possibilities of interest, when accounting for the extra information revealed in such problems, are not uniformly distributed despite it often appearing so. Adhering to the defining formula for conditional probability ought to remove any possibility for error, but even then one runs into pitfalls if one isn't careful, for it's easy to conflate the event that some object is in a certain state with the event that one has been told that some object is in a certain state (note that the latter event is actually a subset of the former in general). We illustrate these pitfalls and how to avoid them in the original Monty Hall problem below, together with a similar example.

Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say 1, and the host, who knows what's behind the doors, opens another door, say 3, which has a goat. He then says to you, "Do you want to pick door 2?" Is it to your advantage to switch your choice?

Solution

Given no additional information, the possible outcomes for what lies behind each door are CGG, GCG, and GGC, each outcome being equally likely. The erroneous reasoning is as follows: given that each of the above three outcomes are equally likely, and given that the information of a goat being behind door 3 eliminates the possibility GGC, each of the two remaining possibilities are equally likely, hence there is no advantage in switching. This appears to be supported by our definition of conditional probability: letting A denote the event that the car is behind door 1 and B the event that there is a goat behind door 3, we have

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/3}{2/3} = \frac{1}{2}.$$

The first line of reasoning may be corrected by observing that the full sample space is not just determined by what is behind which doors, but also which door the host chooses to reveal. Although it's true that each of the cases CGG, GCG, and GGC are equally likely, within the case CGG, it is equally likely that the host will choose to reveal the goats behind either door 2 or door 3. Denoting in red the door the host chooses to reveal, we therefore have the four outcomes CGG, CGG, GCG, and GGC, the first two occurring with probability 1/6 (for a total probability of 1/3), and the last two occurring with probability 1/3 each. When accounting for the host's choice, it's clear that we don't have a uniform distribution, and although the information that there is a goat behind door 3 does reduce the possible outcomes to CGG and GCG, it's clear that these possibilities are not equally likely. To correct the second line of reasoning, we note that the event B to condition on is not that there is a goat behind door 3, but that *the host has chosen to reveal that there is a goat behind door* 3 (noting that the latter event is a subset of the former, since the host may choose to reveal that there is a goat behind door 2 instead of door 3 in the case CGG). We therefore have

$$\begin{aligned}\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} &= \frac{\mathbb{P}(\mathrm{CG{\color{red}G}})}{\mathbb{P}(\mathrm{CG{\color{red}G}}) + \mathbb{P}(\mathrm{GC{\color{red}G}})} \\ &= \frac{1/6}{1/6 + 1/3} = \frac{1}{3},\end{aligned}$$

from which we conclude that **switching doors from 1 to 2 leads to a win with probability 2/3 rather than 1/3, and is therefore advantageous**.

Intuitively, one may reason that it is more likely for the host to have revealed a goat behind door 3 in the case GCG (in which case the host only has one choice of door to reveal) than it is for the host to have revealed a goat behind door 3 in the case CGG (in which case the host may choose to reveal door 2 instead), and it's therefore more likely that switching doors will lead to a win.

The following similar problem serves as a good test of one's understanding of the above problem, and of one's awareness of the pitfalls to be careful of in such situations:

Problem

Three prisoners, A, B, and C, have applied for parole. The parole officer has decided to release two of the three prisoners at random, and the prisoners do not know which two. Prisoner A, however, is friends with one of the guards who knows who will be released. Prisoner A asks the guard to tell them the name of a prisoner (other than A) who will be released, and the guard answers "B will be released". Based on this, Prisoner A reasons that now only two outcomes are possible – either A and B are released or B and C are released – and therefore his chance of being released has dropped from $\frac{2}{3}$ to $\frac{1}{2}$. Is this reasoning correct? If not, what is the correct probability?

Solution

Let U denote the event that prisoner A will be released, and let V denote the event that the guard tells prisoner A that prisoner B will be released (noting that this is not the same as the event that prisoner B will be released, for if prisoners B and C are both being released, the guard may choose to tell prisoner A that prisoner C will be released instead). Denoting in red the person who will not be released and in blue the person who the guard reveals will be released, we have the four outcomes ABC, ABC, ABC, and ABC, with the first two occurring with probability 1/6, and the last two occurring with probability 1/3 each. We therefore have

$$\begin{aligned}\mathbb{P}(U|V) &= \frac{\mathbb{P}(U \cap V)}{\mathbb{P}(V)} = \frac{\mathbb{P}(\mathrm{ABC})}{\mathbb{P}(\mathrm{ABC}) + \mathbb{P}(\mathrm{ABC})} \\ &= \frac{1/3}{1/3 + 1/6} = \frac{2}{3},\end{aligned}$$

so **the probability of prisoner A being released is still $\mathbf{\frac{2}{3}}$**.

Intuitively, one may reason that it is more likely for the host to have revealed that prisoner B is being released because both prisoners A and B are being released (in which case the guard only has one choice of which prisoner to release information about) than it is because prisoners B and C are the ones being released (in which case the host may choose to reveal information about either prisoner B or prisoner C).

Drunken passenger

Problem

100 people are in line waiting to board a plane. They each hold a ticket to one of the 100 seats on the flight – say the nth passenger has a ticket for seat n. Being drunk, the first passenger in line picks a seat uniformly at random. The other 99 passengers are sober and will sit in their proper seat unless it's occupied, in which case they also choose an available seat uniformly at random. What is the probability that the last person in line ends up in their assigned seat?

Solution

The last person ends up in their assigned seat if and only if seat 1 gets taken before seat 100. Each time a passenger has to make a random choice of where to sit and these two seats are still available, it is equally likely for them to choose either seat 1 or 100, and it is therefore equally likely for seat 1 or seat 100 to be taken first. The answer is therefore $\mathbf{\frac{1}{2}}$.

Marbles in a jar

Problem

You are taking out marbles one by one from a jar that has 10 red marbles, 20 blue marbles, and 30 green marbles in it. What is the probability that there are at least 1 blue marble and 1 green marble left in the jar when you have taken out all the red marbles?

Solution

Let B denote the event that the last marble drawn from the jar is blue, and let G denote the event that the last marble drawn from the jar is green, and note that for the given event E to occur, it must be the case that either B or G has occurred, and it follows by the law of total probability that $\mathbb{P}(E) = \mathbb{P}(B)\mathbb{P}(E|B) + \mathbb{P}(G)\mathbb{P}(E|G)$. Now, each marble is equally likely to be the last one drawn, so $\mathbb{P}(B) = 20/60 = 1/3$ and $\mathbb{P}(G) = 30/60 = 1/2$. Given that the last marble drawn was blue, the condition that there is at least 1 blue marble and 1 green marble left in the jar when all red marbles have been removed becomes equivalent to the condition that the last marble removed from among the red and green marbles is green. But among these marbles, we again argue that each marble is equally likely to be the last one drawn, so we have $\mathbb{P}(E|B) = 30/40 = 3/4$, and similar reasoning leads to $\mathbb{P}(E|G) = 20/30 = 2/3$. The answer is therefore $\mathbb{P}(E) = \frac{1}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{2}{3} = \mathbf{\frac{7}{12}}$.

Consecutive heads

A very natural and classical question is that of how many times one should expect to flip a fair coin until one sees n heads in a row. This is often asked for some specific n, but we will address the problem in full generality. One nice option is to use the martingale betting trick outlined in the previous section (to be added later), but we provide two alternative approaches here for the sake of seeing a variety of arguments.

Problem

You flip a fair coin repeatedly until you get n heads in a row. What is the expected number of coin flips?

Solution

Let N_n denote the number of tosses required to see n heads in a row. We have by the law of total expectation that $\mathbb{E}[N_n] = \mathbb{E}[\mathbb{E}[N_n|N_{n-1}]]$, where

$$\begin{aligned}\mathbb{E}[N_n|N_{n-1}] &= \frac{1}{2}(N_{n-1} + 1) + \frac{1}{2}(N_{n-1} + 1 + \mathbb{E}[N_n]) \\ &= N_{n-1} + 1 + \frac{1}{2}\mathbb{E}[N_n]\end{aligned}$$

(also by the law of total expectation, since the next toss is either a heads or tails, each with probability 1/2). It follows that $\mathbb{E}[N_n] = \mathbb{E}[N_{n-1}] + 1 + \frac{1}{2}\mathbb{E}[N_n]$, from which we derive the recurrence relation $\mathbb{E}[N_n] = 2(\mathbb{E}[N_{n-1}] + 1)$, which is more conveniently written as $\mathbb{E}[N_n] + 2 = 2(\mathbb{E}[N_{n-1}] + 2)$. Observing that $N_1 \sim \text{Geom}(1/2)$ has expectation $\mathbb{E}[N_1] = 2$, iterating the recurrence relation leads to

$$\mathbb{E}[N_n] + 2 = 2^{n-1}(\mathbb{E}[N_1] + 2) = 2^{n+1},$$

hence $\mathbb{E}[N_n] = \mathbf{2^{n+1} - 2}$.

The following alternative approach is best suited for when the problem is asked for some specific small value of n, since otherwise one must rely on the identity $\sum_{j=1}^{n} j\left(\frac{1}{2}\right)^j = 2 - \frac{n+1}{2^n}$ which is difficult to remember (but can be derived by differentiating the standard identity for $\sum_{j=0}^{n} x^j$).

Alternative Solution

Let N denote the number of tosses required, and let T denote the toss on which the first tails is observed. By the law of total expectation, we have

$$\begin{aligned}
\mathbb{E}[N] &= \sum_{j=1}^{n} \mathbb{P}(T=j)\mathbb{E}[N|T=j] + \mathbb{P}(T>n)\mathbb{E}[N|T>n] \\
&= \sum_{j=1}^{n} \left(\frac{1}{2}\right)^j (j + \mathbb{E}[N]) + n\left(\frac{1}{2}\right)^n \\
&= \left(1 - \frac{1}{2^n}\right)\mathbb{E}[N] + \left(2 - \frac{n+1}{2^n}\right) + n\left(\frac{1}{2}\right)^n.
\end{aligned}$$

Solving for $\mathbb{E}[N]$ gives $\mathbf{2^{n+1} - 2}$.

Random rendezvous

Problem

Two friends are visiting a restaurant together. Each arrives at a random time within a specific hour-long window and stays for exactly half an hour. What is the probability that the friends meet?

Solution

Letting T_1 and T_2 denote the arrival times of the two friends, we have $T_i \sim \text{Unif}(0,1)$ for $i = 1, 2$. The event that the friends meet is equivalent to $|T_1 - T_2| \leq 1/2$. If we draw the square $[0,1]^2$, this complement of this region forms two right-angled triangles with legs of length $1/2$ in the top-left and bottom-right corners. These each have an area of $1/8$, and the complement therefore has probability $2 \cdot \frac{1}{8} = 1/4$. The friends therefore meet with probability $1 - \frac{1}{4} = \mathbf{\frac{3}{4}}$.

TH vs HT

Problem

We toss a fair coin 10 times. What is the probability that the pattern TH occurs strictly more often than the pattern HT?

Solution

The occurrence of a TH pattern signals the termination of a contiguous block of tails, and the occurrence of a HT pattern signals the termination of a contiguous block of heads. It follows that the patterns have the same number of occurrences if and only if there is an odd number of contiguous blocks, which happens if and only if the sequence of 10 flips begins and ends with the same outcome. This occurs with probability $1/2$, meaning the patterns TH and HT also occur a different number of times with probability $1/2$. By symmetry, it follows that TH occurs strictly more often than HT with probability $\mathbf{\frac{1}{4}}$.

Expected number/size of cycles in a permutation

Problem

What is the expected number of disjoint cycles in a random permutation of $\{1, 2, \cdots, n\}$?

Solution

Consider the process of constructing a random permutation by choosing where to send the first element (in one of n ways), and then choosing where to send the second element (in one of $n-1$ ways), and so on. When we decide where the ith element will be sent, a partially formed cycle may or may not be closed; we denote by X_i the random variable which takes the value 1 if the element i closes a cycle in this way and takes the value 0 otherwise. Then, it's clear that the number of disjoint cycles in the permutation is given by $N = \sum_{i=1}^n X_i$, hence $\mathbb{E}[N] = \sum_{i=1}^n \mathbb{E}[X_i]$. Now, in deciding where to send the ith element, there are $n-(i-1)$ remaining possibilities, exactly one of which will close a cycle. It follows that $\mathbb{E}[X_i] = \frac{1}{n-(n-i)}$, hence

$$\mathbb{E}[N] = \sum_{i=1}^{n} \frac{1}{n-(i-1)} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \frac{\mathbf{1}}{\mathbf{i}}.$$

Problem

What is the average number of elements in a cycle in a random permutation of $\{1, 2, \cdots, n\}$?

Solution

Intuitively, since the previous problem shows that there are $H_n := \sum_{i=1}^n \frac{1}{i}$ cycles on average, and there are n elements in total, the average number of elements in a cycle ought to be n/H_n. To make this more rigorous we seek to make the question more precise, and we elaborate that we wish to compute the average size across all disjoint cycles across all permutations of $\{1, 2, \cdots, n\}$. This is simply described given by the formula

$$\frac{\sum_{P\in S_n}\sum_{C\in P}|C|}{\sum_{P\in S_n}\sum_{C\in P}1},$$

where S_n denotes the symmetric group of all permutations of n elements, and the notation $\sum_{C\in P}$ refers to a summation over all disjoint cycles C in a given permutation P. Now, it's clear that

$$\sum_{P\in S_n}\sum_{C\in P}|C| = n!\cdot n,$$

and moreover, we have from the previous problem that

$$\frac{\sum_{P\in S_n}\sum_{C\in P}1}{n!} = H_n,$$

hence

$$\sum_{P\in S_n}\sum_{C\in P}1 = n!\cdot H_n.$$

It follows by taking quotients that the answer is indeed $\mathbf{n/H_n} = \mathbf{n}/\left(\sum_{\mathbf{i=1}}^{\mathbf{n}}\frac{\mathbf{1}}{\mathbf{i}}\right)$, as originally claimed.

Bus wait

Problem

You are waiting for a bus which runs on a fixed schedule of appearing every 10 minutes. However, the driver, independently between appearances, may want to refill on gas. The driver refills on gas with 10% probability per trial, independently between trials. If the driver fills up on gas, 1 hour is added to his travel time. If you arrive at a uniformly random time throughout the day, what is the expected time until the next bus appears (in minutes)?

Solution

We use the law of total expectation, but note that the driver refilling on gas with probability 10% doesn't imply that there is a 10% probability that you will arrive during a trip in which the driver has refilled on gas (as is apparent by considering the behaviour in the limit as the extra time taken to refill gas approaches infinity). Instead, we argue that on average, 1 in every 10 trips takes 70 minutes and 9 in every 10 trips takes 10 minutes, hence in the average a 160-minute interval, 70 of those minutes will belong to a trip which had a refill and the other 90 will belong to a normal trip. The probability of arriving during a refill trip is therefore $70/160 = 7/16$, and the probability of arriving during a normal trip is $90/160 = 9/16$. Given that you arrive during a trip of a certain length, your arrival time will be uniformly distributed within that length of that trip, so your expected wait time will be half the length of the trip itself. It follows by the law of total expectation that the answer is $\frac{7}{16} \cdot 35 + \frac{9}{16} \cdot 5 = \mathbf{\frac{145}{8}}$.

To be added later:

- Ants on a rod and variants (ants on a circle; infectious ants)
- Breaking a stick to form a triangle (include 'Uniform triangle')
- Population that has babies until a girl is born
- Chess tournaments/other bracket problems
- Coupon collector problem
- Rabbit hop/frog jump problems
- Gridworld problems, especially staying on or below the diagonal
- Lines dividing a plane
- Burning ropes brainteaser

Addressing Common (Personal) Mistakes

This section is dedicated to pinpointing mistakes I personally tend to make when solving interview problems (or have tended to make in the past). Each mistake has an associated explanation, and is accompanied by some example problems in which that mistake could be made if one isn't careful. In some cases, these are problems on which I made the mistake myself, and in other cases, I had already learned to avoid the mistake when solving the problem for the first time, but wished to include the example nonetheless to further reinforce the correct methods. Each solution demonstrates the flawed method presented as if it were a genuine solution attempt, followed by a corrected method which should be used in its place.

Using complicated methods to compute simple probabilities

It is often the case that complicated probabilities are computed by enumerating all favourable outcomes as well as all outcomes in general, and taking the ratio of the two quantities. Despite this, we can often draw on more elementary insights (such as symmetry) to compute simple probabilities, in which case the combinatorial approach wastes time and demonstrates poor understanding.

Problem

Suppose 20 people whose heights follow some unknown continuous distribution are arranged in a single-file line. We then stand at the front of the line and observe that we can see someone's head if they are taller than everyone that comes before them. Let X denote the number of visible heads. Compute $\mathbb{E}[X]$.

Solution

The approach is clear enough: for each position i in the sequence, we denote by X_i the indicator random variable which is 1 if person i's head is visible. Then, $X = \sum_{n=1}^{20} X_n$, hence $\mathbb{E}[X] = \sum_{n=1}^{20} \mathbb{E}[X_n]$. Now, $X_n = 1$ if and only if person n is taller than persons $1, \cdots, n-1$, and we must therefore compute the probability of this event.

The Mistake

Each of the $n!$ orderings of the heights of persons $1, \cdots, n$ are equally likely. Fixing person n to be the tallest, there are then $(n-1)!$ ways to order the heights of the remaining people. Thus, person n is the tallest with probability $(n-1)!/n! = 1/n$.

The Correction

Of persons $1, \cdots, n$, each is equally likely to be the tallest; in particular, person n is the tallest with probability $1/n$.

It follows that $\mathbb{E}[\mathbf{X}] = \sum_{\mathbf{n=1}}^{\mathbf{20}} \frac{\mathbf{1}}{\mathbf{n}}$.

Another example along similar lines:

Problem

9 married couples (1 male and 1 female) are standing in a line. They all scatter around and form 9 pairs uniformly at random (noting that pairs can consist of members of the same gender). Find the expected number of the original couples that are paired up together.

Solution

For each $i = 1, 2, \cdots, 9$, let I_i denote the indicator random variable which is 1 if the ith married couple is paired up together. Then, by linearity of expectation and symmetry, the answer is $\mathbb{E}[\sum_{i=1}^{9} I_i] = \sum_{i=1}^{9} \mathbb{E}[I_i] = 9\mathbb{E}[I_1] = 9\mathbb{P}(I_1 = 1)$.

The Mistake

We assign each person a number at random, where there are two copies of each number from 1 to 9. Such an assignment determines a pairing of all 18 people, but we only care about who is in each pair – not what number they have received – so the total number of ways to form pairs out of all 18 people is $\frac{18!}{9!\cdot(2!)^9}$. By similar reasoning, if we specify that the first married couple stays together, there are $\frac{16!}{8!\cdot(2!)^8}$ ways to form pairs out of the remaining 16 people. Taking the ratio of these quantities gives $\mathbb{P}(I_1 = 1) = \frac{1}{17}$.

The Correction

Fix one of the people from couple 1. This person is equally likely to be paired with any one of the other 17 other people in the room, but only one of the other people is their original partner. Hence, $\mathbb{P}(I_1 = 1) = \frac{1}{17}$.

It follows that the answer is $\boldsymbol{\frac{9}{17}}$.

Not finding the right quantity to 'count on'

Sometimes, a difficult counting problem can be simplified by finding the right quantity to 'count on', in the sense that we partition according to the distinct possibilities which can be observed for this quantity. In easier problems, I tend not to have an issue with this, but when it comes to more difficult problems where the partition is less evident, sometimes I don't think to look for one.

Problem

A particle starts at $(4, 4)$. Each turn, it either moves 1 unit in the $-x$ direction, 1 unit in the $-y$ direction, or 1 unit in each of the $-x$ and $-y$ directions, each with probability $\frac{1}{3}$. The particle repeatedly takes turns until it hits the x or y axes for the first time. Find the probability that the particle hits the origin.

Solution

It's clear that for the particle to hit the origin, it must first reach the point $(1, 1)$ and then travel diagonally to $(0, 0)$.

The Mistake

How do we compute the probability of the particle first reaching $(1, 1)$? If it could only travel in either the $-x$ or $-y$ directions, this would be easy: we count the total number of distinct arrangements of three $-x$ steps and three $-y$ steps. But the given situation is more complicated: a diagonal step constitutes an inseparable $(-x, -y)$ pair, and it's not clear how to incorporate these into the count. Best off biting the bullet and recursively computing the probability of reaching $(1, 1)$ from any given starting point using the law of total probability...

The Correction

We 'count on' the number of diagonal steps taken. That is, we enumerate the paths from $(4,4)$ to $(1,1)$ by partitioning according to how many diagonal steps there were. If there are no diagonal steps, the problem reduces to the easy case mentioned above, and there are $\binom{6}{3,3} = 20$ possibilities; if there is one diagonal step, there must be an additional two steps in each of the $-x$ and $-y$ directions, and these can be arranged in $\binom{5}{1,2,2} = 30$ ways; if there are two diagonal steps, there must be one additional step in each of the $-x$ and $-y$ directions, and these can be arranged in $\binom{4}{1,1,2} = 12$ ways; and there is one way to get from $(3,3)$ to $(1,1)$ by taking three diagonal steps.

The particle reaches $(1,1)$ with probability $\frac{20}{3^6} + \frac{30}{3^5} + \frac{12}{3^4} + \frac{1}{3^3} = \frac{245}{3^6}$, and from there it travels to $(0,0)$ with probability $\frac{1}{3}$. The answer is therefore $\boldsymbol{\frac{245}{3^7}}$.

Another example:

Problem

A 3×3 grid of light bulbs is formed, with each bulb turned on with probability $\frac{1}{2}$. Find the probability that no two adjacent light bulbs are turned on.

Solution

Clearly, there are 2^9 possible choices for combinations of the light bulbs being on and off. We seek to count those combinations for which no two adjacent light bulbs are turned on.

The Mistake

How does one effectively count these cases? Perhaps we ought to start with the first cell, i.e. the one in the top-left. That light bulb is either on or off. If it's off, there are four possibilities for the two adjacent squares; if it's on, there is only one possibility for the two adjacent squares. We then look at the squares adjacent to those, but this seems to be getting out of hand...

The Correction

We 'count on' the arrangement of light bulbs which are turned on in the middle row of the grid. With 0 denoting 'off' and 1 denoting 'on', the possibilities are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, and $(1,0,1)$. The first case leads to 5 possibilities for each of the top and bottom rows; the second and fourth cases lead to 3 possibilities for each of the top and bottom rows; the third case leads to 2^4 possibilities (since the four corner cells can be either on or off, and all other cells are determined); and the last case leads to 2^2 possibilities (since the top middle and bottom middle cells can be either on or off, and all other cells are determined).

The answer is therefore $\frac{25+18+16+4}{2^9} = \boldsymbol{\frac{63}{512}}$.

Forgetting simple ways to compute conditional probabilities/expectations

The definition of the conditional probability $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$ is so basic that I sometimes forget to even consider using it. Similarly, I rarely think to compute conditional expectations using $\mathbb{E}[X|A] = \int_{\mathbb{R}} x\mathbb{P}(X = x|A)\,dx$, though this issue is somewhat more insidious; it's more out of some lingering fear of conditional expectations than it is a simple oversight.

Problem

Four fair coins are laid out on a table and flipped. You receive an amount of dollars equal to the number of heads that appear. However, you have the option to re-flip all four coins at once one time for \$1. If you re-flip whenever you got 0 or 1 heads on the initial flip, what are your expected winnings?

Solution

Let X be the winnings from this process, and let Y denote the number of heads tossed in the initial round. We have $\mathbb{P}(Y \leq 1) = \frac{1}{2^4} + \frac{4}{2^4} = \frac{5}{16}$, hence $\mathbb{P}(Y > 1) = \frac{11}{16}$, and it follows by the law of total expectation that $\mathbb{E}[X] = \frac{5}{16}\mathbb{E}[X|Y \leq 1] + \frac{11}{16}\mathbb{E}[X|Y > 1]$.

The Mistake

Since we re-flip on $Y \leq 1$ at the expense of \$1, it's clear that $\mathbb{E}[X|Y \leq 1] = 1$. However, $\mathbb{E}[X|Y > 1]$ looks scary, so we use the law of total expectation on a more fine-grained level, writing $\mathbb{E}[X] = \frac{5}{16}\mathbb{E}[X|Y \leq 1] + \sum_{j=2}^{4} \mathbb{P}(Y = j)\mathbb{E}[X|Y = j]$, which can be computed without much difficulty.

The Correction

There's nothing inherently wrong with the above approach, and in fact it requires essentially the same computations as the following 'correction'. However, the point of bringing attention to this mistake is that avoiding conditional expectation computations betrays a lack of understanding of the material, and besides, what is one to do when such a computation genuinely is necessary? Regardless, there's no need to shy away from the conditional expectation. By definition, we have

$$\begin{aligned}\mathbb{E}[X|Y > 1] &= \sum_{j=2}^{4} j\mathbb{P}(X = j|Y > 1) = \sum_{j=2}^{4} j\frac{\mathbb{P}((X = j) \cap (Y > 1))}{\mathbb{P}(Y > 1)} \\ &= \sum_{j=2}^{4} j\frac{\mathbb{P}(Y = j)}{\mathbb{P}(Y > 1)} = 2\left(\frac{6}{11}\right) + 3\left(\frac{4}{11}\right) + 4\left(\frac{1}{11}\right) = \frac{28}{11}.\end{aligned}$$

The expected winnings are therefore $\frac{5}{16} + \frac{11}{16} \cdot \frac{28}{11} = \mathbf{\frac{33}{16}}$.

Computing the number of ways to insert k distinguishable objects between n objects in a roundabout way

Suppose we have n objects in a row, and we want to place k distinguishable objects between them. To compute the number of ways to do this, I gravitate towards the following argument: first, treat the k objects as indistinguishable, and count the number of ways to insert them between the n objects via a stars and bars-like argument. Then, choose an ordering of the k objects so they are again distinguishable. Overall, we find that there are

$$\binom{n+k}{n}k! = \frac{(n+k)!}{n!k!}k! = \frac{(n+k)!}{n!} = (n+1)(n+2)\cdots(n+k)$$

ways to do this. However, there is a more straightforward way to reach this conclusion: fix an ordering of the k objects, and imagine placing them one by one. There are $n+1$ gaps in which the first object can be placed; having placed it, there will then be $n+2$ gaps in which the second object can be placed, and so on, leading to the same answer of $(n+1)(n+2)\cdots(n+k) = \frac{(n+k)!}{n!}$. We note that this type of computation, if it arises, usually does so as a sub-step in the context of a more difficult problem. For example:

Problem

Consider the numbers $1, 2, \cdots, 13$ written down one by one in a random order. What is the probability that the sum of the numbers written so far during this process is at no point divisible by 3?

Solution

To ensure the sum of numbers written so far is never divisible by 3, we can obviously consider the numbers $1, 2, \cdots, 13$ reduced mod 3. Moreover, adding 0 mod 3 doesn't change the sum, so we can first arrange those numbers which are congruent to 1 or 2 mod 3 and then insert those which are congruent to 0 mod 3. It's clear that the only way to arrange the residue classes which are either 1 or 2 mod 3 without having a partial sum of 0 is $1, 1, 2, 1, 2, 1, 2, 1, 2$, and having done this, we can arrange the five numbers which are 1 mod 3 and the four numbers which are 2 mod 3 in $5! \cdot 4!$ ways. It remains to insert the four remaining numbers which are congruent to 0 mod 3.

The Mistake

Once the four 0's are inserted, there will be a total of 12 elements in the list excluding the first element (which cannot be a 0), and the locations of the four (indistinguishable) 0's among these 12 elements can be chosen in $\binom{12}{4}$ ways. Having done this, we can arrange the four numbers which are 0 mod 3 in $4!$ ways, and there is therefore a total of $5! \cdot 4! \cdot \binom{12}{4} \cdot 4!$ ways to arrange the numbers in such a way that the constraint is satisfied.

The Correction

The above works, but uses one more step than is necessary. Fix an ordering of the numbers which are 0 mod 3, say $3, 6, 9, 12$. Having placed all other numbers, there are 9 gaps in which the 3 can be inserted (since it can't be placed at the front), and having done this, there are then 10 gaps in which the 6 can be inserted, and so on. Overall, there are $9 \cdot 10 \cdot 11 \cdot 12$ ways to place $3, 6, 9, 12$, so there is a total of $5! \cdot 4! \cdot 9 \cdot 10 \cdot 11 \cdot 12$ ways to arrange the numbers in such a way that the constraint is satisfied.

Since there are $13!$ possibilities for the arrangement of the numbers in the list, the probability that they are arranged in such a way that the constraint is satisfied is $\frac{5! \cdot 4! \cdot 9 \cdot 10 \cdot 11 \cdot 12}{13!} = \mathbf{\frac{1}{182}}$.

Not identifying enough states in a symmetric situation

In Markov chain problems, it's often the case that one can identify certain cases which are equivalent by symmetry, thereby cutting down the number of equations one needs to solve. Occasionally, I'll fail to realise the full extent to which states can be identified in such a problem, and therefore not solve the problem with maximal efficiency.

Problem

How many times do we have to roll a fair 6-sided die until we roll two numbers in a row that differ by 2?

Solution

The Mistake

It's clear that there is some symmetry in the outcomes of the die roll, and the expected number of additional rolls required to meet the given condition will be the same when starting from an initial roll of either 1 or 6, or 2 or 5, or 3 or 4. We can therefore set up a system of three linear equations in three unknowns, and solve...

The Correction

The above works, but one can go one step further if one accounts for the condition we are actually trying to meet. They key property shared by 1 and 6 is not that they are symmetric about the centre of the distribution, but more specifically that they each have one other possibility which could be rolled in succession to terminate the game. But 2 and 5 also have this property, so in fact the numbers $1, 2, 5, 6$ functionally form a single state! Calling this state 1 and the state of having rolled a 3 or 4 state 2, we derive the system

$$\begin{aligned} E_1 &= 1 + \frac{4}{6}E_1 + \frac{1}{6}E_2, \\ E_2 &= 1 + \frac{2}{6}E_1 + \frac{2}{6}E_2, \end{aligned}$$

where E_i denotes the expected number of additional rolls when starting from state i.

Solving the above system gives $E_1 = 5$ and $E_2 = 4$. Hence, by the law of total expectation, the expected number of rolls when starting from nothing is $1 + \frac{4}{6}E_1 + \frac{2}{6}E_2 = \boldsymbol{\frac{17}{3}}$.

Overcomplicating the expectation of moving one step forward in a random walk

If one is interested in the expected number of steps it takes to reach position 1 when starting a random walk at position 0, there's a recursive trick which one could miss because it seems too good to be true.

Problem

Consider a particle that performs a random walk on the integers starting at position 0. At each step, the particle moves from position i to position $i + 1$ with probability $2/3$, and to position $i - 1$ with probability $1/3$. What is the expected number of steps until the particle reaches 1?

Solution

Let E_i denote the expected number of steps to reach 1 when starting from position i.

The Mistake

Clearly, by the law of total expectation, we have $E_0 = 1 + \frac{1}{3}E_{-1}$. Similarly, we have $E_{-1} = 1 + \frac{2}{3}E_0 + \frac{1}{3}E_{-2}$, but this is getting out of hand...

The Correction

Clearly, one may decompose a path from -1 to 1 as a path first from -1 to 0, and then from 0 to 1. Since traveling from -1 to 0 is symmetric to travelling from 0 to 1, it follows that $E_{-1} = 2E_0$, though this may seem too good to be true. To help see that this is the case, consider a random variable N which records the number of steps it takes to reach position 1 from -1 in a given iteration of the random walk, and decompose it as $N = N' + N''$, where N' denotes the number of steps it takes to reach 0 from -1 and N'' denotes the number of steps it takes to reach 1 from 0. Then, it becomes clear that $E_{-1} = \mathbb{E}[N] = \mathbb{E}[N'] + \mathbb{E}[N''] = 2E_0$.

The equation $E_0 = 1 + \frac{1}{3}E_{-1}$ now gives $E_0 = 1 + \frac{2}{3}E_0$, and we can solve to find $E_0 = \mathbf{3}$.

Forgetting to account for the cost of re-rolls

Some problems have the format of a gambler trying to maximise their profits when playing a game such as rolling a die, where the gambler has the option to forgo their winnings from the first round and replay the game for a fee. For some reason, I've repeatedly made the innocent and ridiculous mistake of forgetting to account for the fee after getting caught up in all the computations, so I'm including this example to help ensure I never do this again.

Problem

Alice rolls a fair 6-sided die with values 1 - 6. She sees the value showing up and is allowed to decide whether or not she wants to re-roll for a cost of \$1. If she decides to stop, Alice receives a payout equal to the upface of the die. If she rolls again, she received a payout equal to the upface of the second roll. Assuming optimal play by Alice, what is her expected payout?

Solution

The Mistake

The expected payout of a die roll is \$3.5, so surely Alice should re-roll if she initially sees either a 1, 2, or 3...

The Correction

While the expected payout of the first roll is \$3.5, when accounting for the fee, the expected payout of the second roll is \$2.5. It follows that Alice should re-roll only if she initially sees a value of 1 or 2, in which case her expected payout is \$2.5, and if Alice initially rolls a value of 3 or greater, she keeps it for an expected payout of $\$\frac{1}{4}(3+4+5+6) = \$4.5$.

By the law of total expectation, Alice's expected payout under the identified optimal strategy is $\$(\frac{2}{6}\cdot 2.5 + \frac{4}{6}\cdot 4.5) = \boldsymbol{\$\frac{23}{6}}$.