Probability and Statistics

MAT-2020-1, Spring 2026

Supplementary Notes

Pritam Chandra

General advice for writing mathematics:

1. Do not use symbols unless they make the readers job easier. Give preference to words and sentences wherever possible.
2. Notation should be minimal and precise, as long as it doesn't hinder interpretation.
3. Solving a problem is not the end. Spend time refining, simplifying etc. Generally one should consider themselves to be at their reader's service.

Conventions. The notation $A \subset B$ will mean that the set A is either contained in or equal to B. Strict containment will be denoted by $A \subsetneq B$.

Table of contents

1 Sets

Some elementary binary set operations are unions, intersections, set differences, symmetric differences of sets etc. Among these, unions and intersections are commutative, associative, distributive. The union distributes over intersections (and the vice versa) as follows

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \text{ and } A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Question 1. Which of the mentioned operations are not commutative, associative?

Set complement is unary operator.

Proposition 1. *Show that for any two finite sets A_1, A_2 we have*

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|.$$

Proof. If X, Y are disjoint then $|X \cup Y| = |X| + |Y|$. Now note that for two general sets A_1, A_2, we note that A_1 and $A_2 \setminus A_1$ are disjoint. And

$$A_1 \cup (A_2 \setminus A_1) = A_1 \cup (A_2 \cap \overline{A_1}) = A_1 \cup A_2 .$$

So $|A_1 \cup A_2| = |A_1| + |A_2 \setminus A_1|$. Similarly, note that $A_2 \setminus A_1$ and $A_1 \cap A_2$ are disjoint. And

$$(A_2 \setminus A_1) \cup (A_1 \cap A_2) = A_2 .$$

Theorem 1. *(Principle of inclusion and exclusion) Let $A_1, \ldots, A_n$ be finite sets. Then*

$$|A_1 \cup \cdots \cup A_n| = \sum_{k=1}^{n} (-1)^{k+1} \left[\sum_{1 \leq j_1 \leq \cdots \leq j_k \leq k} |A_{j_1} \cap \cdots \cap A_{j_k}| \right].$$

2 Counting

What exactly are permutations and combinations … and what is the relationship between them? The number $C(n, r)$, also written as, $\binom{n}{r}$ is the number of subsets of size r of a set of size n. The number $P(n, r)$, also written as, ${}^nP_r$ is the number of "ordered" subsets of size r of a set of size n. We have

$${}^nP_r = \frac{n!}{(n-r)!} \text{ and } r! \binom{n}{r} = {}^n P_r .$$

Theorem 2. *(Pascal's Identity) When $n, r \in \mathbb{N}$ and $r < n$ we have*

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

Proof. There are different ways to see this. One can give an algebraic proof. Here we will use the method of "counting in two different ways". First, we will use the identification of $\binom{n}{r}$ being the number of size r subsets of $\{1, \ldots, n\}$. Now, such subsets either contain the element n, or they don't. If a subset doesn't contain the element n, then they are a size r subset of $\{1, \ldots, n-1\}$. The number of such subsets is $\binom{n-1}{r}$. If a subset contains the element n, then the rest of the elements form a size $r-1$ subset of $\{1, \ldots, n-1\}$. The number of such subsets is $\binom{n-1}{r-1}$. Thus we get the required identity. $\square$

Remark 1. In fact, it turns out that with the initial conditions $f(n,n)=1$ and $f(n,0)=1$, the only solution to $f(n,r)=f(n-1,r)+f(n-1,r-1)$ for all n, $r\in\mathbb{N}_0$ is $f(n,r)=\binom{n}{r}$.

2.1 Counting integer solutions of an equation

How many solutions are there to the equation

$$x_1+\cdots+x_r=n \tag{1}$$

such that $x_j\in\mathbb{N}$ for all j? Another version of the question is asked by requiring the solutions to be whole (nonnegative) integers instead of natural numbers. Further refinments are also made by replacing the equality with an inequality, or restricting the variables with some inequality constraint.

The standard way to count this is with some equivalent of the "stars and bars" method. We will describe this method later. Such arguments are clever, efficient, easy to understand once described, but perhaps not so easy to come up with. We will try to describe a more "regular", albeit longer, approach to count the solutions.

Let the number of natural number solutions of (1) be $S(n,r)$. Note that if x_r takes the value t, then $x_1+\cdots+x_{r-1}=n-t$. So we obtain the recurrence

$$S(n,r)=S(n-1,r-1)+S(n-2,r-1)+\cdots+S(1,r-1)$$

with the understanding that $S(m,s)=0$ if $s>m$. Computing $S(n-1,r)$ with the same recurrence we observe that the above equation reduces to

$$S(n,r)=S(n-1,r)+S(n-1,r-1).$$

This immediately reminds us of the Pascal identity. However, the initial conditions here are $S(n,1)=1$ and $S(n,n)=1$. As a work around, consider the function f defined as

$$f(n,r)=S(n+1,r+1).$$

The recurrence for S can be alternatively written as

$$S(n+1,r+1)=S(n,r+1)+S(n,r),$$

along with $S(n+1,n+1)=1$ and $S(n+1,1)=1$. Substituting f for S in this equation, we get the recurrence

$$f(n,r)=f(n-1,r)+f(n-1,r-1)$$

with the initial conditions $f(n,n)=1$ and $f(n,0)=1$. This immediately tells us that

$$S(n+1,r+1)=f(n,r)=\binom{n}{r} \quad \text{or} \quad S(n,r)=\binom{n-1}{r-1}.$$

The “coins” and bars technique.

To illustrate this technique we study a slightly modified problem of finding the number of *nonnegative* integer solutions of the equation (1). Earlier we required each variable to be at least 1, and now we allow them to be 0 as well.

To do this, let us instantiate the problem by assuming that we are distributing n objects (say coins) among r variables (say bins). Imagine arranging the coins next to each other in a straight line. For illustration, suppose $n = 9$, and $r = 4$.

$$\circledcirc\ \circledcirc\ \circledcirc\ \circledcirc\ \circledcirc\ \circledcirc\ \circledcirc\ \circledcirc\ \circledcirc$$

One can distribute the coins by introducing $r - 1$ dividers into the arrangement of coins. This will divide the coins into r sections, and each section can be collected into a bin.

$$\circledcirc\ \circledcirc\ \circledcirc\ |\ |\ \circledcirc\ \circledcirc\ \circledcirc\ \circledcirc\ \circledcirc\ |\ \circledcirc$$

For example, the above arrangement corresponds to the solution $3 + 0 + 5 + 1 = 9$. Thus, every distribution corresponds uniquely to a arrangement of $n + r - 1$ objects, of which n are coins and $r - 1$ dividers. Conversely, every such arrangement corresponds to a distribution. So it is sufficient to count the number of arrangements. To do this simply choose the position of the dividers, and place the coins in the remaining spots. Thus, the number of solutions is $\binom{n+r-1}{r-1}$.

Inequality constraints.

Let us consider the number of solutions of (1) where the constraints on the variables are inequalities. That is, each x_i satisfies a constraint of the form $p_i \leq x_i \leq q_i$. A constraint like $p_i = 0$ or $q_i = n$ is trivial.

The **lower bound** constraints are easier to handle. For any variable x_i with a lower bound $x_i \geq p_i$ is dealt with by introducing a new variable $y_i = x_i - p_i$. This is visualized by imagining that p_i coins are removed from the collection at the beginning and placed in the ith bin. Once all such substitutions have been made, we obtain the following new equation with the y-variables

$$y_1 + \cdots + y_r = n - (p_1 + \cdots + p_r).$$

Note that the constraint $x_i \geq p_i$ now reduces to $y_i \geq 0$. This is familiar territory and can be solved using the previous methods.

Upper bounds on variables are a little more tricky to handle. Let us first consider a single upper bound constraint, say $x_1 \leq q_1$. There could be multiple lower bound constraints. Let the universal set (say of size N) be the set of solutions of (1) subject to all the lower bound constraints. Let A be the subset of this universal set with the added constraint $x_1 \leq q_1$. Then

$$|A| = N - |\bar{A}|$$

where $\bar{A}$ is the set of solutions with the previous lower bound constraints along with the new lower bound $x_1 \geq q_1$. This is again familiar territory.

Suppose, now, we had two upper bound constraints, say $x_1 \leq q_1$ and $x_2 \leq q_2$. We deal with them again by defining the corresponding subsets. It is important to be precise. Again, let the universal set be of size N which contains the solutions of (1) only with the lower bound constraints. Let A_1 be the subset of this universal set with the added constraint that $x_1 \leq q_1$. Similarly, consider A_2 for the constraint $x_2 \leq q_2$. We are looking for $|A_1 \cap A_2|$. We have

$$\begin{aligned} |A_1 \cap A_2| &= |A_1| + |A_2| - |A_1 \cup A_2| \\ &= |A_1| + |A_2| - |\overline{\bar{A}_1 \cap \bar{A}_2}| \\ &= N - |\bar{A}_1| + N - |\bar{A}_2| - N + |\bar{A}_1 \cap \bar{A}_2| \\ &= N + |\bar{A}_1 \cap \bar{A}_2| - |\bar{A}_1| - |\bar{A}_2| \,. \end{aligned}$$

The sets on the right hand side of the equation are now addressable using our previous techniques. For more than two upper bounds, we employ the general principle of inclusion and exclusion. We summarise our counts of the solutions of (1) in the following table.

Constraint	Count
$x_j \in \mathbb{N}$ for all j	$\binom{n-1}{r-1}$
$x_j \in \mathbb{N}_0$ for all j	$\binom{n+r-1}{n}$
$x_j \geq p_j$	replace x_j with the variable $y_j = x_j - p_j$
$x_j \leq q_j$	Let A_j be set of solutions with no upper bound constraints except $x_j \leq q_j$. Count $\lvert A_1 \cap \cdots \cap A_r\rvert$ using set completements.

2.2 Derangements

We denote by D_n the number of "derangements" of n, where

$$D_n = |\{\sigma \in \mathrm{Sym}(n) \colon \sigma(j) \neq j \text{ for all } 1 \leq j \leq n\}|.$$

In words, a deragement is a permutation (of n objects) such every object is displaced from its original position. We count D_n using the principle of inclusion and exclusion. For $1 \leq j \leq n$, let A_j denote the set

$$A_j = \{\sigma \in \mathrm{Sym}(n) \colon \sigma(j) = j\},$$

the collection of permuations that keep the jth object in its place. We have

$$\begin{aligned}
D_n &= |\overline{A_1 \cup \cdots \cup A_n}| \\
&= n! - |A_1 \cup \cdots \cup A_n| \\
&= n! - \sum_{j=1}^{n} (-1)^j \sum_{i_1 < \cdots < i_j} |A_{i_1} \cap \cdots \cap A_{i_j}| \\
&= n! - \sum_{j=1}^{n} (-1)^j \binom{n}{j} (n-j)! \\
&= n! \sum_{j=0}^{n} \frac{(-1)^j}{j!},
\end{aligned}$$

where the fourth equality is because there are $\binom{n}{j}$ choices for the indices $i_1, \ldots, i_j$, and for each such choice the number of permutations that fixes all these positions $(n-j)!$.

Proposition 2. *Let D_k be the number of derangements of k elements. We have*

$$n! = \sum_{j=0}^{n} \binom{n}{j} D_{n-j}.$$

Proof. We give a combinatorial argument for this identity. The left hand side counts the total number of arrangements of n objects. The set of all arrangements can be partitioned into $n+1$ collections, the jth collection comprising those arrangements which keep exactly j objects fixed (for $0 \leq j \leq n$). The jth term on the right hand side counts size of the jth collection. The j objects that are fixed can be chosen in $\binom{n}{j}$ ways. Each of the remaining $n-j$ objects must be displaced from their position. This is done in D_{n-j} ways. $\square$

3 Push-forward measures

Suppose we have a set X equipped with some "randomness" μ. That is, (X, μ) is a probability space. Then we can produce a notion of randomness on any set Y as long as we have a function $f: X \longrightarrow Y$. We say that we can "push the measure μ on X forward" to a measure, say $f_* \mu$, on Y.

Formally, let $\mathcal{P}(X)$ denote the set of all possible measures on X. Then every function $f: X \longrightarrow Y$ induces a function $f_*: \mathcal{P}(X) \longrightarrow \mathcal{P}(Y)$. For any $\mu \in \mathcal{P}(X)$, the measure $f_*\mu \in \mathcal{P}(Y)$ is defined as

$$f_*\mu(y) = \mu(f^{-1}(\{y\})) \quad \text{for all } y \in Y.$$

A caution here that the notation f^{-1} should not be confused with the inverse of an invertible function. Here, input to f^{-1} can be any subset of Y. By definition, $f^{-1}(B) = \{x \in A: f(x) \in B\}$. When B is a singleton, say $\{y\}$, for convenience $f^{-1}(\{y\})$ may be written as $f^{-1}(y)$.

The sets $f^{-1}(\{y\})$ for $y \in Y$ are called the **fibers** of f.

Question 2. Do we need any restriction on f to define its push forward f_*?

Proposition 3. *Let $f: X \longrightarrow Y$. Then $f_*: \mathcal{P}(X) \longrightarrow \mathcal{P}(Y)$ is injective if and only if f is injective. It is surjective if and only if f is surjective.*

Proof. Suppose f is injective. Pick any $x \in X$. Let $y = f(x)$. As f is injective we have $f^{-1}(\{y\}) = \{x\}$. Now, let μ_1 and μ_2 be any two probability measures on X. Observe that

$$\begin{array}{rccc} & f_* \mu_1 & = & f_* \mu_2 \\ \Longrightarrow & \mu_1(f^{-1}(\{y\})) & = & \mu_1(f^{-1}(\{y\})) \\ \Longrightarrow & \mu_1(x) & = & \mu_2(x). \end{array}$$

As this is true for all x, we conclude that $\mu_1 = \mu_2$. This shows that f_* is injective.

Conversely, suppose f is not injective. That is, there exist x_1, x_2 in X such that $x_1 \neq x_2$ and $f(x_1) = f(x_2) = y$ (say). Consider the Dirac measures[1] δ_{x_1} and δ_{x_2} on X. Clearly, $\delta_{x_1} \neq \delta_{x_2}$. But, note that $f_* \delta_{x_1} = f_* \delta_{x_2} = \delta_y$. So, f_* is not injective.

Now suppose f is surjective. We always have

$$X = \bigsqcup_{y \in Y} f^{-1}(\{y\}).$$

Now, let ν be any measure on Y. Consider the measure μ on X such that

$$\mu(x) = \frac{\nu(f(x))}{|f^{-1}(\{f(x)\})|}.$$

Then for any $y \in Y$ we have

$$f_* \mu(y) = \sum_{x \in f^{-1}(\{y\})} \mu(x) = \sum_{x \in f^{-1}(\{y\})} \frac{\nu(y)}{|f^{-1}(\{y\})|} = \nu(y).$$

1. For any $x \in X$, the Dirac measure $\delta_x \in \mathcal{P}(X)$ is defined as $\delta_x(x) = 1$ and $\delta_x(\tilde{x}) = 0$ for all $\tilde{x} \neq x$.

Note here surjectivity is critical for these set of equalities to hold. Otherwise, for $y \in Y$ without any pre-image under f, the probability $f_*\mu(y)$ will necessarily have to be 0.

Now, conversely, suppose f is not surjective. Let $y \in Y$ such that y has no pre-image in X under f. Consider the Dirac measure δ_y on Y. Suppose there is a measure μ on X such that $\delta_y = f_* \mu$. Then

$$1 = \delta_y(y) = \mu(f^{-1}(\{y\})) = \mu(\phi) = 0.$$

This is a contradiction. Or, no such μ can exist. Hence, f_* is not surjective. $\square$

Push forward measures behave well with respect to compositions and products.

Lemma 1. *Let $f: X \longrightarrow Y$ and $g: Y \longrightarrow Z$. Then $(g \circ f)_* = g_* \circ f_*$. In words, the composition of push forward measures, is the push forward of compositions.*

Proof. To show the above equality, we will show that for any measure $\mu \in \mathcal{P}(X)$ we have $(g \circ f)_* \mu = g_*(f_* \mu)$. This in turn will be showed by showing

$$(g \circ f)_* \mu(z) = g_*(f_* \mu)(z) \text{ for every } z \in Z.$$

Indeed note that

$$(g \circ f)_* \mu(z) \;=\; \mu((g \circ f)^{-1}(z)) \;=\; \mu(f^{-1}(g^{-1}(z))),$$

and we also have

$$g_*(f_* \mu)(z) \;=\; (f_* \mu)(g^{-1}(z)) = \sum_{y \in g^{-1}(z)} \mu(f^{-1}(g^{-1}(y))) \;=\; \mu(f^{-1}(g^{-1}(z))). \quad \square$$

Proposition 4. *Let (X_1, μ_1) and (X_2, μ_2) be two probability spaces. Let f and g be functions such that $f: X_1 \longrightarrow Y_1$ and $g: X_2 \longrightarrow Y_2$. Then*

$$(f_* \mu_1) \times (g_* \mu_2) = (f \times g)_*(\mu_1 \times \mu_2).$$

In words, the product of two push forward measures is the push forward of the product of the measures.

Proposition 5. *Let (X, μ) be a probability space where μ is the uniform measure. Let $f: X \longrightarrow Y$ such that $|f^{-1}(y)|$ is constant for all $y \in Y$. Then $f_*\mu$ is uniform on Y.*

3.1 Generating uniform permutations and subsets of a set

We present further illustrations of the push-forward measure by studying the problem of generating random permutations, or random subsets of a set. We start with the following definitions.

$[n]$	$\{1, 2, \ldots, n\}$
$\mathrm{Sym}(S)$	Set of all bijections from $[\lvert S\rvert]$ to S, or, set of all permutations of the elements of S
$\binom{S}{k}$ for $k \leq n$	$\{A \subset S \colon \lvert A\rvert = k\}$
$\binom{S}{k_1, \ldots, k_r}$ for $k_1 + \cdots + k_r = n$	$\left\{(A_1, \ldots, A_r) \colon \lvert A_i\rvert = k_i \text{ and } A = \bigsqcup_{i=1}^{r} A_i\right\}$, or, set of all **ordered** partitions of S into subsets of size $k_1, \ldots, k_r$ respectively

The number $\binom{n}{k_1, \ldots, k_r}$, called the **multinomial coefficient** corresponding to n, $k_1, \ldots, k_r$, denotes the cardinality of the set $\binom{S}{k_1, \ldots, k_r}$.

Given we have uniform distributions on $[n]$, we will describe recipes for generating uniform distributions on the other sets.

Let $S = \{s_1, \ldots, s_n\}$. Consider the probability space $[k]$, for $k \leq n$, equipped with the uniform distribution. Let $\mathcal{S}$ be the probability space $[1] \times [2] \times \cdots \times [n]$. We define the function $f \colon \mathcal{S} \longrightarrow \mathrm{Sym}(S)$ in the following steps. Let $i = (i_1, \ldots, i_n)$ be any element of $\mathcal{S}$. Let $f(i) = \sigma$. Then σ is the following arrangement. First place s_1. Place s_2 on the location $i_2 \in [2]$ denoting either before or after the object s_1. Place s_3 on the location $i_3 \in [3]$, which is either at the beginning, or between the two placed objects, or at the end, and so on.

Showing f is a bijection, using proposition 5, is sufficient to prove a unifrom measure on $\mathcal{S}$ pushes forward to the uniform measure on $\mathrm{Sym}(S)$.

Further, we can propagate the uniform measure on $\mathrm{Sym}(S)$ to the uniform measure on $\binom{S}{k}$. Let $g \colon \mathrm{Sym}(S) \longrightarrow \binom{S}{k}$ such that for any $s \in S$ we have

$$g((s_1, s_2, \ldots, s_n)) = \{s_1, \ldots, s_k\}.$$

Thus for any $A \in \binom{S}{k}$ we have

$$\lvert g^{-1}(A)\rvert = k!\,(n-k)!,$$

a number independent of A. Thus, again by proposition 5 we obtain a uniform measure on $\binom{S}{k}$. In a similar way, one can push forward the uniform measure from $\mathrm{Sym}(S)$ to $\binom{S}{k_1, k_2, \ldots, k_r}$ via the function g, such that for any $s \in S$ we have

$$g((s_1, \ldots, s_n)) = \Big(\{s_1, \ldots, s_{k_1}\}, \{s_{k_1+1}, \ldots, s_{k_1+k_2}\}, \ldots, \{s_{\sum_{j=1}^{r-1} k_j + 1}, \ldots, s_{\sum_{j=1}^{r} k_j}\}\Big).$$

Example 1. *Suppose a standard pack of 52 cards is shuffled and distributed uniformly into 4 hands of 13 each. What is the probability that each hand gets an Ace?*

Consider the act of the distributing the cards in the following manner. Randomly shuffle the cards. Make 4 contiguous partitions of 13 cards each and collect them as the hands. Let S be the set of all racks (permutations) of 52 cards equipped with the uniform measure μ. Let

$$T = \{(t_1, t_2, t_3, t_4)\colon 1 \leq t_1 < t_2 < t_3 < t_4 \leq 52\}.$$

Then $|T| = \binom{52}{4}$. Consider the map $f\colon S \longrightarrow T$ such that $f(s) = (t_1, t_2, t_3, t_4)$ where t_1 is the position of the first ace in the rack s, and t_2 the position of the second, and so on. Pick any $t = (t_1, t_2, t_3, t_4) \in T$. Then $f^{-1}(t)$ is set of those racks in which the Aces occupy positions t_1, t_2, t_4 and t_4. One immediately calculates $|f^{-1}(t)| = 4!48!$. Therefore,

$$f_*\mu(t) = \mu(f^{-1}(t)) = \frac{|f^{-1}(t)|}{|S|} = \frac{1}{\binom{52}{4}}.$$

This shows that $f_*\mu$ is the uniform measure on T. Each hand getting an Ace is described by the event $E \subset T$ where

$$E = \{(t_1, t_2, t_3, t_4)\colon 13(i-1) < t_i \leq 13i \quad \text{for } i = 1, \ldots, 4\}.$$

Clearly $|E| = 13^4$. Hence, the required probability is

$$f_*\,\mu(E) = \frac{|E|}{|T|} = \frac{13^4}{\binom{52}{4}}. \qquad \square$$

One might question the introduction of push-forward formalism in such problems where the quantity of interest can be counted directly. In this example, originally one had to count the number of hands such that each hand contained an Ace. The formalism reduced the problem to counting the number of permutations of 52 cards such that the Aces appear in 4 specific given locations. One would generally agree that the latter is a “simpler” (or rather, a “more precise”) quantity than the former, and thus the chance of a counting error is reduced. While solving such problems, one learns to imbibe sufficient formalism such that what remains can be reliably handled by intuition. Although, in many occassions, intuition is altogether misleading, or multiple intuitions are in conflict.

4 Examples of caution while counting

Example 2. *In how many ways can one distribute n identical candies among Alice and Bob such that Alice gets no more than p candies, and Bob no more than q?*

One identifies this as counting the number of nonnegative integer solutions of the equation

$$x_1+x_2=n \quad \text{such that} \quad x_1\leq p \text{ and } x_2\leq q.$$

We use the techniques discussed in section 2.1. Let A_1 be the set of solutions where $x_1\leq p$, and A_2 where $x_2\leq q$. Let N be the total number of solutions without any restrictions. Then

$$\begin{aligned}|A_1\cap A_2| &= |A_1|+|A_2|-|A_1\cup A_2|\\ &= |A_1|+|A_2|-|\overline{\bar{A}_1\cap\bar{A}_2}|\\ &= N-|\bar{A}_1|+N-|\bar{A}_2|-N+|\bar{A}_1\cap\bar{A}_2|\\ &= N+|\bar{A}_1\cap\bar{A}_2|-|\bar{A}_1|-|\bar{A}_2|\,.\end{aligned}$$

Solving these we get

$$|A_1\cap A_2|=\binom{n+1}{1}+\binom{n+1-(p+q+2)}{1}-\binom{n-p}{1}-\binom{n-q}{1}.$$

Now naively calculating this using $\binom{m}{1}=m$ **equates the expression to 0**. This does not seem right, and there must be a mistake. Note that we have not assumed any information about p,q. The identity $\binom{m}{1}=m$ holds only when $m\geq 1$, and otherwise it $\binom{m}{1}$ is 0. So, for example, the second term in the expression is nonzero only if $n\geq p+q+2>p+q$. But the question doesn't make sense only when $n\leq p+q$. Hence this term is always 0. Similarly, one assumes that $p,q\leq n$. Then the expression evaluates to

$$n+1-(n-p)-(n-q)=p+q+1-n.$$

It turns out that in this particular question that one may benefit to not jump to the standard approach. As there are only two variables, fixing x_1 automatically fixes $x_2=n-x_1$. Applying the conditions $x_1\leq p$ and $x_2=n-x_1\leq q$ reduced to the condition

$$n-q\leq x_1\leq p.$$

The number of such choices is $p-(n-q)+1$ which is exactly the required solution.

Example 3. *Suppose one wishes to generate an approximate uniform distribution on a set of size m using only fair coin. How many coin tosses are needed such that the generated distribution is at most "ε distance away" from the true uniform distribution?*

To begin with, we must have a notion of "distance" between two distributions. Let μ and ν be elements of $B(X)$. A standard notion of distance between μ and ν is as follows:

$$\operatorname{dist}(\mu, \nu) = \max_{x \in X} |\mu(x) - \nu(x)|\,.$$

This captures the worst difference in the masses assigned to a point of X by μ and ν.

We know k tosses of the fair coin can be used to generate a uniform distribution μ_n on $[n]$ where $n = 2^k$. Suppose we push μ forward to $\{0, 1, \ldots, m-1\}$ via the function f where $f(i) = i \bmod m$. Let $n = mq + r$. For $0 \leq j \leq m-1$ we have

$$|f^{-1}(j)| = \begin{cases} \frac{q}{n}, & j \leq r \\ \frac{q+1}{n}, & j > r \end{cases}.$$

Therefore, we have

$$\operatorname{dist}(f_* \mu_n, \mu_m) = \max\left(\left|\frac{q}{n} - \frac{1}{m}\right|, \left|\frac{q+1}{n} - \frac{1}{m}\right|\right) = \frac{1}{nm} \max(r, m-r) \geq \frac{1}{2n} = \frac{1}{2^{k+1}},$$

where the second last inequality is simply the fact that the maximum of 2 numbers is larger than their average. So, to get $\operatorname{dist}(f_* \mu_n, \mu_m) < \varepsilon$ it is sufficient to have

$$\frac{1}{2^{k+1}} < \varepsilon \Longleftrightarrow k \geq -(1 + \log_2 \varepsilon)\,.$$

Why is there no dependence on m? It cannot be that we can get arbitrarily close to μ_m for any m by pushing forward μ_n for a fixed n. This is because for all the calculations to go through, we need $n \geq m$, without which the entire model fails. The dependence is truly captured in the b

$$k \geq \max\{-1 - \log_2 \varepsilon, \log_2 n\}.$$

5 Kernels and joint measures in the language of linear algebra

Let Ω, Θ be two finite sets. Let $\mathcal{P}(\Theta)$ be the set of all probability measures on Θ. Then a map $\kappa\colon \Omega \longrightarrow \mathcal{P}(\Theta)$ is called a **kernel**. That is, for each $\omega \in \Omega$ its image $\kappa(\omega)$ is a probability measure on Θ. Further, for any measure μ on Ω, we define the **barycentric push-forward** of μ under κ as the measure

$$\kappa_* \mu = \sum_{\omega \in \Omega} \mu(\omega)\, \kappa(\omega).$$

So, $\kappa_* \mu$ is an average of the measures $\kappa(\omega)$ weighted by the probability masses $\mu(\omega)$. The measure $\kappa_*\mu$ can be visualized to be realizing the two step experiment: sample an ω from Ω with probability $\mu(\omega)$, then sample θ from Θ with probability $\kappa(\omega)$.

Suppose $|\Omega| = n$ and $|\Theta| = m$. Then any measure μ on Ω can be identified with a vector in $\mathbb{R}^n$ indexed by the elements of Ω. The entry μ_ω is the probability mass $\mu(\omega)$. Thus the entries of the vector μ lie in the interval $[0, 1]$ and all the entries add up to 1. Such vectors are called **probability vectors**. In the same spirit, the kernel κ can be represented by the $m \times n$ matrix K whose rows are indexed by elements of Θ and columns by elements of Ω. The entry

$$K_{\theta,\omega} = [\kappa(\omega)](\theta).$$

Again, the entries of K lie in the interval $[0, 1]$, and each column of K is a probability vector (in this case, the ωth column is the probability vector $\kappa(\omega)$). Such matrices are called **stochastic matrices**.

Let e_k be the k-dimensional vector whose entries are all 1. Thus, a vector $v \in \mathbb{R}_+^n$ is a probability vector if and only if $\langle v, e_n \rangle = 1$. Here $\mathbb{R}_+$ is the set of nonnegative real numbers. Similarly, a $m \times n$ matrix with entries in $\mathbb{R}_+$ is stochastic if $M^T e_m = e_n$.

Now, indeed the probability vector $\kappa_*\mu \in \mathbb{R}^m$ indexed by elements of Θ is given by the matrix vector product $K\mu$. Further, we know that a measure μ on Ω together with a kernel $\kappa: \Omega \longrightarrow \mathcal{P}(\Theta)$ gives rise to a joint measure, say $\mathbb{P}$, on $\Omega \times \Theta$. Here

$$\mathbb{P}(\omega, \theta) = \mu(\omega)[\kappa(\omega)](\theta).$$

Again, $\mathbb{P}$ is naturally represented by a $m \times n$ matrix $J = J(\mu, \kappa)$. We have

$$J_{\theta,\omega} = \mathbb{P}(\omega, \theta) = K_{\theta,\omega}\, \mu_\omega.$$

The matrix J then relates to the matrix K and the vector μ as

$$J = K \operatorname{diag} \mu,$$

where $\operatorname{diag} \mu$ is $n \times n$ the diagonal matrix whose diagonal entries are the entries of the vector μ. One can verify that the vector of column sums of J is the measure μ, and the vector of row sums the measure $\kappa_*\mu$. This can be equivalently stated as

$$J e_n = \kappa_*\mu \text{ and } J^T e_m = \mu.$$

Example 4. Consider the measure $\mu \in \mathcal{P}(\Omega)$ and $\nu \in \mathcal{P}(\Theta)$ related through the kernel $\kappa: \Omega \longrightarrow \mathcal{P}(\Theta)$. That is, $\nu = \kappa_*\mu$. A natural question is the following: is there a kernel $\lambda: \Theta \longrightarrow \mathcal{P}(\Omega)$ such that $\mu = \lambda_* \nu$? Note that λ here would depend on both κ and μ. It is not immediately obvious if such a "reverse kernel" λ exists, and if it does, how to compute it. The language of linear algebra turns out to be useful in this situation.

Treat μ and ν like probability vectors and let K be the kernel corresponding to κ. For simplicity we further assume that ν has no nonzero coordinate, that is, $\nu(\theta) > 0$ for all $\theta \in \Theta$.[2] Then $\nu = K\mu$. Let $J = K \operatorname{diag} \mu$ be the joint distribution matrix arising out of μ and κ. Does there exists a kernel $\lambda\colon \Theta \longrightarrow \mathcal{P}(\Omega)$ whose joint distribution with ν is described by the matrix J^T? In other words, is there an $n \times m$ stochastic matrix L such that $L \operatorname{diag} \nu = J^T$? A candidate L is solved from this relation as follows:

$$L \operatorname{diag} \nu = J^T \Longrightarrow L = J^T (\operatorname{diag} \nu)^{-1} = (\operatorname{diag} \mu)\, K^T (\operatorname{diag} \nu)^{-1}.$$

We first verify that

$$\lambda_* \nu = L\nu = J^T (\operatorname{diag} \nu)^{-1} \nu = J^T e_m = \mu.$$

What remains to be verified is that L is a stochastic matrix. Indeed

$$L^T e_n = (\operatorname{diag} \nu)^{-1} J e_n = (\operatorname{diag} \nu)^{-1} \nu = e_m.$$

We briefly describe the situation when $\operatorname{diag} \nu$ is not invertible. First of all, if $\nu_\theta = 0$, then the θth column of $(\operatorname{diag} \nu)\, K^T (\operatorname{diag} \nu)^\dagger$ is also 0. So, this time we define the reverse kernel L such that the θth column of L and $(\operatorname{diag} \nu)\, K^T (\operatorname{diag} \nu)^\dagger$ agree whenever $\nu_\theta \neq 0$. When $\nu_\theta = 0$, we fill the θth column of L with any arbitrary probability vector.

Now, recall that ν_θ is the sum of the θth row of J. So if $\nu_\theta = 0$, as J only has nonnegative entries, the entire θth row of J must be 0. The vector $(\operatorname{diag} \nu)^\dagger \nu \neq e_m$ anymore, as some of its entries are 0. But $J^T (\operatorname{diag} \nu)^{-1} \nu = J^T e_m$ still holds. This is because, for indices of e_m which do not agree with $(\operatorname{diag} \nu)^{-1} \nu$, the corresponding rows of J (columns of J^T) are 0. These indices have no contribution. Similarly, by construction we already know that whenever $\nu_\theta = 0$, the θth column of L sums up to 1. The conclusion holds for the other columns as they are identical to the invertible case.

6 Coupling — an example

Two independent experiments are to be compared. Coupling is a method of curating an artificial dependence between the experiments so that they can be compared through the induced joint distribution. Kernels are the right tool to model this. Let us see an example.

Example 5. *An urn contains m red and n green balls, where $m, n \geq 1$. A sequence of $k \leq m + n$ balls is drawn with replacement, and separately a sequence of k balls is drawn without replacement. Show that*

$$\Pr\!\left(\begin{array}{c}\textit{all } k \textit{ balls are red,}\\ \textit{with replacement}\end{array}\right) \geq \Pr\!\left(\begin{array}{c}\textit{all } k \textit{ balls are red,}\\ \textit{without replacement}\end{array}\right).$$

2. This assumption enables us to invert the matrix $(\operatorname{diag} \nu)$. Without the assumption, the arguments can be made using the pseudo-inverse $(\operatorname{diag} \nu)^\dagger$ instead of the inverse $(\operatorname{diag} \nu)^{-1}$ with minor modifications. The pseudo-inverse of the diagonal matrix $\operatorname{diag}(\alpha_1, \ldots, \alpha_k, 0, \ldots, 0)$, where $\alpha_j \neq 0$ for all j, is the diagonal matrix $\operatorname{diag}(\alpha_1^{-1}, \ldots, \alpha_k^{-1}, 0, \ldots, 0)$.

We solve this using **coupling**. Let R be the set of the red balls and G the green ones. Let $B=R\cup G$ be the set of all the balls of size $N=m+n$. Then B^k represents the collection of all sequences of k balls drawn with replacement and, say, this is drawn by Bob. Let Ω_k represent the collection of sequences where the k balls are drawn without replacement. That is $\Omega_k=\mathrm{Sym}_k(B)$. Say Alice is drawing this sequence. Define a kernel $\kappa\colon\Omega_k\longrightarrow\mathcal{P}(B^k)$ such that

$$\kappa(\omega)=\nu_{\omega,1}\times\cdots\times\nu_{\omega,k} \quad \text{where} \quad \nu_{\omega,j}=[1-\varepsilon_j]\,\delta_{\omega_j}+\frac{\varepsilon_j}{j-1}\,[\delta_{\omega_1}+\cdots+\delta_{\omega_{j-1}}]$$

for $j\geq 2$, and $\nu_{\omega,1}=\delta_{\omega_1}$. Here, $\varepsilon_j=\frac{j-1}{N}$. In words this coupling is described as follows. Bob looks at Alice's sequence ω, ball by ball. For the jth ball, with probability $1-\varepsilon_j$, Bob copies Alice's ball ω_j, and with probability ε_j, Bob uniformly chooses one of the previous balls $\omega_1,\ldots,\omega_{j-1}$ drawn by Alice.

We need to verify that this coupling is valid. Let μ be the uniform distribution on Ω_k. We are required to show that

$$\kappa_*\,\mu(\theta)=\frac{1}{|\Omega_k|}\sum_{\omega\in\Omega_k}\prod_{j=1}^{k}\nu_{\omega,j}(\theta_j)=\frac{1}{N^k} \quad \text{for all } \theta\in\Theta.$$

Define $S_k(\theta)=\sum_{\omega\in\Omega_k}\prod_{j=1}^{k}\nu_{\omega,j}(\theta_j)$. Then, equivalently, using $|\Omega_k|={}^N P_k$ we are required to show that

$$S_k(\theta)=\frac{{}^N P_k}{N^k} \quad \text{for all } \theta\in\Theta.$$

We will show this by inducton on k. Choose any $\theta\in\Theta$. First observe that

$$S_1(\theta)=\sum_{\omega\in B}\delta_{\omega}(\theta_1)=1=\frac{{}^N P_1}{N},$$

as θ_1 equals exactly one member of B. Thus, the base case holds. For the inductive step, we decompose $S_k(\theta)$ as a product of the sums

$$S_k(\theta)=\left[\sum_{\omega\in\Omega_{k-1}}\prod_{j=1}^{k-1}\nu_{\omega,j}(\theta_j)\right]\left[\sum_{\omega_k\in B_{\omega,k}}\nu_{\omega,k}(\theta_k)\right],$$

where $B_{\omega,k}=B-\{\omega_1,\ldots,\omega_{k-1}\}$. Note that the first sum is $S_{k-1}(\theta)$ and the second sum

$$\begin{aligned}\sum_{\omega_k\in B_{\omega,k}}\nu_{\omega,k}(\theta_k) &= \sum_{\omega_k\in B_{\omega,k}}\left[(1-\varepsilon_k)\delta_{\omega_k}(\theta_k)+\frac{\varepsilon_k}{k-1}\sum_{j=1}^{k-1}\delta_{\omega_j}(\theta_k)\right]\\ &= (1-\varepsilon_k)\sum_{\omega_k\in B_{\omega,k}}\delta_{\omega_k}(\theta_k)+\frac{N-k+1}{k-1}\,\varepsilon_k\sum_{j=1}^{k-1}\delta_{\omega_j}(\theta_k)\end{aligned}$$

itself can be expressed as a sum of two terms as shown above. Again, as ω_js are distinct, the number θ_k either equals exactly one member of $\{\omega_1,..., w_{k-1}\}$ or exactly one member of $B_{\omega,k}$. If $\theta_k \in B_{\omega,k}$ then the second term is 0 and the first one equals

$$1-\varepsilon_k = \frac{N-k+1}{N}.$$

If $\theta_k \in \{\omega_1, \ldots, w_{k-1}\}$, then the first term is 0, and this time the second equals

$$\frac{N-k+1}{k-1}\,\varepsilon_k = \frac{N-k+1}{N}.$$

In either case the second sum evaluates to the same number. Substituting this to the expression for $S_k(\theta)$ we get

$$S_k(\theta) = S_{k-1}(\theta)\,\frac{N-k+1}{N} = \frac{{}^N P_{k-1}}{N^{k-1}}\,\frac{N-k+1}{N} = \frac{1}{N^k}\,\frac{N!\,(N-k+1)}{(N-k+1)!} = \frac{{}^N P_k}{N^k}.$$

Indeed, $\kappa_*\,\mu$ is the uniform distribution on B^k. Now, the distribution μ and the kernel κ gives rise to a joint distribution, say $\mathbb{P}$, on $\Omega_k \times B^k$. Suppose the projection maps from $\Omega_k \times B^k$ onto Ω_k and B^k be Π_1 and Π_2 respectively. Define the events E and F as

$$E = \{\omega \in \Omega_k | \{\omega_1, \ldots, \omega_k\} \subset R\} \text{ and } F = \{\theta \in B^k | \{\theta_1, \ldots, \theta_k\} \subset R\}.$$

Further, let $\tilde{E} = \Pi_1^{-1}(E)$ and $\tilde{F} = \Pi_2^{-1}(F)$. We are required to show that

$$\begin{array}{rccc} & \mu(E) & \leq & \kappa_*\,\mu(F) \\ \text{or,} & (\Pi_1)_*\,\mathbb{P}\,(E) & \leq & (\Pi_2)_*\,\mathbb{P}\,(F) \\ \text{or,} & \mathbb{P}(\tilde{E}) & \leq & \mathbb{P}(\tilde{F}). \end{array}$$

Consider a point $(\omega, \theta) \in \tilde{E} - \tilde{F}$. Since $\theta \notin F$, there must exist a j for which $\theta_j \notin R$. But since $\omega \in E$ we have $\{\omega_1, \ldots, \omega_k\} \subset R$. Therefore, $\theta_j \neq \omega_i$ for $1 \leq i \leq k$. Thefore

$$\nu_{\omega,j}(\theta_j) = [1-\varepsilon_j]\,\delta_{\omega_j}(\theta_j) + \frac{\varepsilon_j}{j-1}\,[\delta_{\omega_1}(\theta_j) + \cdots + \delta_{\omega_{j-1}}(\theta_j)] = 0.$$

This shows that $\mathbb{P}(\omega, \theta) = \mu(\omega)\,[\kappa(\omega)](\theta) = 0$. We conclude that $\mathbb{P}(\tilde{E} - \tilde{F}) = 0$. This completes the solution.

7 Some quick thoughts on "independence"

Let Ω be a probability space equipped with the probability measure $\mathbb{P}$ and let A, $B \subset \Omega$ be two events. We say A and B are "independent" if

$$\mathbb{P}(A\,|\,B) = P(A) \text{ or } P(B|A) = P(B).$$

This says that the proportion of A in B is the same as the proportion of A in the entire space. It is entirely a property of the events, and not of the outcomes contained within. For example, suppose A and B are independent. It is well possible that $X \subset A$ is not independent with $Y \subset B$. It is possible that A is independent to B and C, but not to $B \cap C$. Consider the example

$$\Omega = \{1,2,3,4\}, \quad A = \{1,2\}, \quad B = \{2,3\}, \quad C = \{2,4\}, \quad X = \{2\}$$

with the uniform measure on Ω. Then

$$P(A|B) = \frac{1}{2} = P(A).$$

But $P(A|X) = 1 \neq P(A)$. So A and X are not independent. Similarly, A is independent to B and C, but not to $B \cap C = X$.

8 Examples of problem solving using probabilistic methods

Probabilistic approaches may be applied to problems from seemingly disparate areas. Such problems may appear to not involve any randomness at the outset. Yet, one can cleverly introduce randomness to find elegant solutions. Before we begin, let us briefly recall random variables and expectations. Just the linearity of expectation, though fundamental, turns out to be the essential ingredient in many of these solutions.

Definition 1. Let (Ω, μ) be a probability space and let $\Theta \subset \mathbb{R}$. Then any function $f: \Omega \longrightarrow \Theta$ is called a **random variable**. The push-forward measure $f_* \mu$ is called the distribution of the random variable f. Further, we denote the **expectation** of f by $\mathbb{E}_\mu f$ where

$$\mathbb{E}_\mu f = \sum_{\theta \in \Theta} \theta f_* \mu(\theta) = \sum_{\omega \in \Omega} f(\omega)\, \mu(\omega).$$

When μ is known from the context, we write $\mathbb{E}_\mu f$ as $\mathbb{E} f$. Expectation is linear on the space of random variables. That is, if $f_1, \ldots, f_n$ are random variables from taking values in Θ, then

$$\mathbb{E}\,(f_1 + \cdots + f_n) = \mathbb{E} f_1 + \cdots + \mathbb{E} f_n.$$

A quick observation here is as follows. There always exists an ω such that $\mu(\omega) \neq 0$ and $f(\omega) \geq \mathbb{E} f$. If not, then

$$\mathbb{E} f = \sum_{\omega \in \Omega} f(\omega)\, \mu(\omega) = \sum_{\omega : \mu(\omega) \neq 0} f(\omega)\, \mu(\omega) < \mathbb{E} f,$$

which is a contradiction. We will use this observation crucially in the coming examples. Examples 6 and 7 appeared in the special lecture conducted by Jaikumar Radhakrishnan.

Example 6. Let $G = (V, E)$ be a graph. The vertices are thought to be representing villages and the edges roads. Say $|V| = n$ and $|E| = m$. Every village $v \in V$ is assigned a service in $\{H, P\}$, where H represents a hospital and P a police station. Our assumption is that each village has exactly one of the two services, and none have both.

A road $e \in E$ is called "good" if it connects an H vertex with a P vertex. One can imagine why having two different services at the ends of a road is desirable. The proposition is that *for every graph there exists an assignment of services such that at least* $\frac{m}{2}$ *edges are good.* We will show this using probability theory.

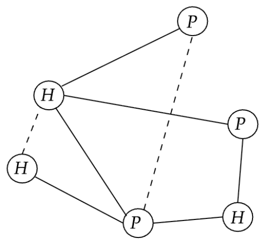

Figure 1. An example of a graph with services assigned to the vertices. The solid edges are "good" and the dotted ones are not.

Suppose to each vertex we assign a service H or P uniformly at random. Thus, every assignment is an outcome in the probability space $\Omega = \{H, P\}^n$ sampled according to the uniform distribution, say $\mathbb{U}$. For $e \in E$, define $f^e : \Omega \longrightarrow \{0, 1\}$ such that $f^e(\omega) = 1$ if e is a good edge under the assignment ω, and $f_e(\omega) = 0$ otherwise. Then

$$f^e_* \mathbb{U}(1) = \frac{1}{2}.$$

To see this, suppose $e = (v, w)$ and the services assigned to v and w is uniform among the choices $(H, H), (H, P), (P, H), (P, P)$. Out of these, two lead to good edges. Now, let

$$f = \sum_{e \in E} f^e.$$

Then $f(\omega)$ captures the number of good edges under the assignment ω. And

$$\mathbb{E} f = \sum_{e \in E} \mathbb{E} f^e = \frac{m}{2}.$$

Therefore, there must exist an assignment $\omega \in \Omega$ for which $f(\omega) \geq \frac{m}{2}$.

Example 7. *Consider another interesting graph problem. Let V be a set of n vertices. For convenience, assume n is even. What is the maximum possible number of edges in an edge set E such that the graph (V, E) has no triangles?*

Let us begin thinking of this question in relation to the previous example. Consider an assignment of services on a graph such that every edge turns out to be good. Observe that such a graph cannot have any triangles. Indeed, if the graph had a triangle, then at least two of vertices have to be assigned the same service, and the edge between them would not be good.

Now, let us partition V into two halves, that is, $V = V_H \sqcup V_P$ and $|V_H| = |V_P| = \frac{n}{2}$. Assign the service H to all vertices in V_H and P to all in V_P. Consider the edge set

$$E = \{(u, v) \colon u \in V_H \text{ and } v \in V_P\}.$$

Every edge in the graph (V, E) is good. And $|E| = \frac{n^2}{4}$. So, this is a graph on the vertices V without triangles and with $\frac{n^2}{4}$ edges. We will now argue using a probabilistic approach that one cannot do better.

Suppose $G = (V, E)$ is a graph on the vertices V without triangles. Let $|E| = m$. Consider the following experiment. Arbitrarily arrange the vertices in a line. Start from the left and proceed one "available" vertex after another. At each step, if you are on vertex v, delete all the vertices after v which are not neighbors of v. Equivalently, scan through the vertices following v and keep only its neighbors. "Availability" is determined by whether a vertex was deleted before it was reached.

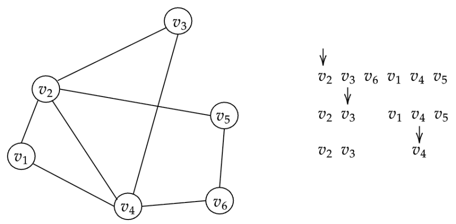

Figure 2. An illustration of running the experiment with an arbitrarily chosen initial ordering of the vertices.

Suppose at the end of this experiment more than three vertices are left, say

$$u \quad v \quad w \quad \cdots.$$

Then, v and w are neighbors of u, as they have not been removed. For the same reason w is a neighbor of v. Thus the verties u, v, w form a triangle in G. This cannot be the case. So, for every initial ordering, the experiment must end with at most 2 vertices.

Now consider the set $\Omega = \mathrm{Sym}(n)$ denoting all the possible arrangements of the vertices, equipped with the uniform distribution $\mathbb{U}$. For every vertex $v \in V$, consider the function $f^v \colon \Omega \longrightarrow \{0, 1\}$ such that $f^v(\omega) = 1$ if the vertex v survives at the end of the experiment starting with the arrangement ω. If v doesn't survive then $f^v(\omega) = 0$.

Consider another class of functions $g^v\colon \Omega \longrightarrow \{0,1\}$ such that $g^v(\omega)=1$ if all the non-neighbors of v lie to the right of v in the arrangement ω, and $g^v(\omega)=0$ otherwise. Observe that if no non-neighbor of v is to the left of v in ω, then v surely survives the experiment starting on the arrangment ω. It is the non-neighbors of v on v's left side which have a potential to delete v.[3] In other words, $g^v(\omega)=1$ implies $f^v(\omega)=1$. Or,

$$f^v(\omega) \geq g^v(\omega).$$

Define $f(\omega)=\sum_{v\in V} f^v(\omega)$. This captures the number of vertices surviving the experiment starting with the arrangement ω. Recall that $f(\omega)\leq 2$ for every ω. Hence, $\mathbb{E}\, f \leq 2$.

For each vertex $v\in V$, let d_v be the degree of v in the graph G. We will show later that $g^v_* \mathbb{U}(1)=\frac{1}{n-d_v}$. But proceeding with this expression for now we get

$$2 \geq \mathbb{E}\, f \geq \sum_{v\in V} \mathbb{E}\, g^v = \sum_{v\in V} (n-d_v)^{-1} \geq \frac{n}{n^2-\sum_{v\in V} d_v} = \frac{n^2}{n^2-2m}.$$

Here the second last inequality is a consequence of the arithemtic-harmonic mean inequality. Rearranging the terms in the extreme sides of the equation gives

$$m \leq \frac{n^2}{4}.$$

It only remains to show that indeed $g^v_* \mathbb{U}(1)=\frac{1}{n-d_v}$. For each v, let N_v be the set of neighbors of V. Let $B_v=V\setminus N_v$. We have $v\in B_v$. Now, consider the map h^v: $\mathrm{Sym}(V)\longrightarrow \mathrm{Sym}(B_v)$ where $h_v(\omega)$ is the relative order the vertices in B_v in the ordering ω. For example, with $V=\{v_1,\ldots,v_6\}$, $B_{v_1}=\{v_3,v_4,v_6\}$ and $\omega=(v_4,v_2,v_1,v_6,v_3,v_5)$, we have

$$h^v(\omega)=(v_4,v_6,v_3).$$

For each $\sigma\in \mathrm{Sym}(B_v)$ the fiber $h_v^{-1}(\sigma)$ has size $\binom{n}{d_v} d_v!$. This is explained as follows. Among the n positions in an arrangement we are choosing $n-d_v$ positions for elements of B_v and placing them in the order mandated by σ, and in the rest of the positions we can arrange elements of N_v in $d_v!$ ways. The import of this discussion is that $h^v_* \mathbb{U}$ is uniform on $\mathrm{Sym}(B_v)$. In our final step we can realize f^v as

$$\underbrace{\Omega \xrightarrow{h^v} B_v \xrightarrow{q} \{0,1\}}_{f^v},$$

where $q(\sigma)=1$ if v appears at the very beginning in the arrangement σ, and 0 otherwise. So

$$f^v_* \mathbb{U}(1) = q^v_*(h^v_* \mathbb{U}(1)) = h^v_* \mathbb{U}(q^{-1}(1)) = \frac{|q^{-1}(1)|}{(n-d_v)!} = \frac{(n-d_v-1)!}{(n-d_v)!} = \frac{1}{n-d_v}.$$

3. We say "potential" as it is possible that v's non-neighbors on the left were deleted by their non-neighbors preceding them. This is why $f^v(\omega)$ need not be equal to $g^v(\omega)$, but nonetheless, always exceeds.

Example 8. Let A be any subset of $\mathbb{R}$. The set A is called **sum-free** if whenever $x, y, z \in A$, we have $x + y \neq z$. For exaple, the set of odd numbers is a sum-free set. The set of even numbers is not sum-free. In this context, we have the following proposition. *Let $S \subset \mathbb{N}$ be a finite subset. Then there is a sum-free subset T of S with $|T| \geq \frac{|S|}{3}$.*

We discuss a probalistic proof of this number theoretic proposition. Suppose $x, y, z \in \mathbb{R}$ such that $x + y = z$. Then

$$\{x\} + \{y\} - \{z\} = x - \lfloor x \rfloor + y - \lfloor y \rfloor - z + \lfloor z \rfloor \in \mathbb{Z}.$$

Let θ be any number in the interval $(0, 1]$. Define

$$S_\theta = \left\{ s \in S | \{\theta s\} \in \left(\tfrac{1}{3}, \tfrac{2}{3}\right) \right\}.$$

We claim that S_θ is a sum-free subset of S. Indeed, if $x, y, z \in S_\theta$, then

$$1 = \tfrac{2}{3} + \tfrac{2}{3} - \tfrac{1}{3} > \{\theta x\} + \{\theta y\} - \{\theta z\} > \tfrac{1}{3} + \tfrac{1}{3} - \tfrac{2}{3} = 0.$$

This shows that

$$\{\theta x\} + \{\theta y\} - \{\theta z\} \notin \mathbb{Z} \Longrightarrow \theta x + \theta y \neq \theta z \Longrightarrow x + y \neq z.$$

We will show that there exists a θ for which $|S_\theta| \geq \frac{|S|}{3}$. Let X be a random variable which has the uniform distribution over $(0, 1]$. For any $n \in \mathbb{N}$ and $y \in [0, 1]$,

$$\Pr\left[\{nX\} \leq y\right] = \sum_{j=0}^{n-1} \Pr\left[j \leq nX \leq j + y\right] = \sum_{j=0}^{n-1} \Pr\left[\tfrac{j}{n} \leq X \leq \tfrac{j+y}{n}\right] = \sum_{j=0}^{n-1} \tfrac{y}{n} = y.$$

So, the random variable $\{nX\}$ is also uniform over $(0, 1]$. Define the random variable F_s as

$$F_s \colon (0, 1] \longrightarrow \{0, 1\} \quad \text{where } F_s(\theta) = \begin{cases} 1 & \text{if} \quad \{s\theta\} \in \left(\tfrac{1}{3}, \tfrac{2}{3}\right) \\ 0 & \quad\text{otherwise} \end{cases}.$$

As $\{s\theta\}$ is distributed uniformly over $(0, 1]$

$$\mathbb{E} F_s = \Pr\left[\{\theta s\} \in \left(\tfrac{1}{3}, \tfrac{2}{3}\right)\right] = \tfrac{1}{3}.$$

Then, note that

$$|S_\theta| = \sum_{s \in S} F_s \text{ and } \mathbb{E}\,|S_\theta| = \sum_{s \in S} \mathbb{E} F_s = \frac{|S|}{3}.$$

Hence, there must be a θ, for which $|S_\theta| \geq \frac{|S|}{3}$. This completes the proof.

So, we immediately have the relation

$$n! = \sum_{j=0}^{n} \binom{n}{j} D_{n-j}.$$

Coming to the "psychic problem". Let X be the number of matches the psychic gets. Then

$$\mathbb{P}[X = k] = \binom{n}{k} \frac{D_{n-k}}{n!}.$$

And

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{n} k \binom{n}{k} \frac{D_{n-k}}{n!} \\
&= \frac{1}{(n-1)!} \sum_{k=1}^{n} \binom{n-1}{k-1} D_{n-k} \\
&= \frac{1}{(n-1)!} \sum_{j=0}^{n-1} \binom{n-1}{j} D_{(n-1)-j} \\
&= 1.
\end{aligned}$$

Question. Suppose n coins are being distributed among r bins. What is the probability that exactly k bins are empty.

Distance between distributions.

Question. How many coins tosses to generate a distribution on k elements that is ε distance away from uniform?

Experiment. Toss a coin n times to generate a number uniformly between 1 and $2^n = N$. Let $N = kq + r$. So we get every output either q or $q+1$ times. So

$$\varepsilon > \delta(\mu, \mathbb{U}) = \max\left(\left|\frac{1}{k} - \frac{q}{N}\right|, \left|\frac{1}{k} - \frac{q+1}{N}\right|\right) = \frac{1}{kN} \max(r, k-r) \geq \frac{1}{2N}$$

given that $N \geq k$. Hence $n \geq \log_2(2\varepsilon)$. Is this not surprising? As long as $N \geq k$, the number of tosses required only depends on ε.

** distance between distributions

** push forward $\mathrm{Sym}(S) \longrightarrow \binom{S}{k}$ is uniform

** Ace problem

** iterative placing generates uniform permutations using push forward measures
... because of bijection

** probability of getting a straight

** multinomial coefficients

* uniform distributions using a coin toss, measure of non-uniformness

* push forward measures

* derangements (psychic problem)

* rejection sampling

$$
\begin{aligned}
\mu_{n+1} &= \kappa_* \, \mu_n \\
&= \mu_n(A) \, \kappa(A) + \mu_n(B) \, \kappa(B) + \mu_n(C) \, \kappa(C) \\
&= a_n \left(\frac{\delta_B + \delta_C}{2} \right) + b_n \, \delta_C + c_n \, \delta_A \\
&= c_n \, \delta_A + \frac{a_n}{2} \, \delta_B + \left(\frac{a_n}{2} + b_n \right) \delta_C \\
&= a_{n+1} \, \delta_A + b_{n+1} \, \delta_B + c_{n+1} \, \delta_C.
\end{aligned}
$$