$$\sum_{k=0}^{m} \binom{n-1+k}{k} = \binom{m+n}{n}$$

Count the number of $(m+n)$-paths from $(0,0)$ to $(0,0)$ using steps $(1,0)$ and $(0,1)$ on $\mathbb{Z}_m \times \mathbb{Z}_n$. Let k be the row of the first return to column 0.STOP

```
\usepackage{youngtab}
\young(1,5,3)
```

STOP

$X_1, \ldots, X_n$

$$E(\max_i X_i) =$$

$$\int_0^\infty E(\max_i X_i | \sum_i X_i = t) \frac{d}{dt} P(\sum_i X_i \leq t) dt =$$

$$\int_0^\infty tE(\max_i X_i | \sum_i X_i = 1) \frac{d}{dt} P(\sum_i X_i \leq t) dt =$$

$$E(\max_i X_i | \sum_i X_i = 1) E(\sum_i X_i)$$

If $P(X_i \leq t) = 1 - e^{-t}$ for all i, then $E(\max_i X_i) = \sum_{k=1}^n \binom{n}{k} \frac{(-1)^{k+1}}{k}$ and $E(\sum_i X_i) = n$

Furthermore the lengths X_i are the lengths of a partition of $[0, \sum_i X_i]$ into n parts chosen uniformly at random.

Thus $E(\max_i X_i) = \frac{1}{n} \sum_{k=1}^n \binom{n}{k} \frac{(-1)^{k+1}}{k}$ for the lengths X_i of a uniformly random chosen partition of $[0,1]$ into n parts.

Todo: simplify $\sum_{k=1}^n \binom{n}{k} \frac{(-1)^{k+1}}{k}$ using EKHAD

(Nothing above is properly justified.)STOP

Let $P(f)$ be the Newton polytope (or -gon) of the polynomial f.

$$P(fg) = P(f) + P(g)$$

(Minkowski sum). For the non-obvious inclusion, argue that the lex-largest monomial is the product of the two lex-largest monomials of the factors (and thus a single product and not a sum of products) and that lex-order can be defined in any direction, not just in an axis-parallell manner. STOP

The graph of the result of identifying x and $-x$ in the dodecahedron is the Petersen graph. Of course.

What is the communication complexity of finding the median of $A \cup B$ when Alice knows A and Bob knows B? It is $O(\log(n))$ when $A, B \subseteq \{1, \dots, n\}$.

For each n, find n lattice vectors $v_i = (x_i, y_i)$, x_i, y_i integers, such that the sums $\sum_{i \in S} v_i$ are all distinct but as compact(undefined) as possible.

Percolation Voronoi (not Voronoi percolation) [after thinking about this: it seems not much can be said.] STOP

Minimizing the sum of squares of the errors is more natural than minimizing the sum of the absolute values of the errors if the errors are Gaussian (as physical errors are). This is because minimizing the sum of squares is a maximum likelihood estimator. (todo: remove opinion) STOP

The relation $(s_i s_j)^{m_{ij}} = e$ makes sense for reflections in arbitrary hyperplanes, not just linear ones (bjÃ¶rner-brenti notation). STOP

A path in PR^2 can be thought of as a continuous movement of a pair of antipodal points on S^2. Restricting attention to one of these points (this is possible, even though the pair being unordered, since the two points will never equal each other) gives a path in S^2, which either ends in its starting point or the antipode of the starting point. (And a path in S^2 obviously gives a path in PR^2.) This is the use of the concept of 'covering space'.
STOP

Knowledge of $\lambda_1^k + \cdots + \lambda_n^k$ for $1 \le k \le n$ as an ordered sequence is equivalent to knowing $\lambda_1, \dots, \lambda_n$ as a multiset. If A is the adjacency matrix of a graph G, the k:th term in the sequence is the number of closed paths of length k in G (with marked 'starting' point). STOP

(non-un-)Knot around three points a,b,c such that removing any of a, b or c gives an un-knot: (aba'b')c(aba'b')'c' = aba'b'cbab'a'c', where ' is inverse of an operation and 'X' is the operation of going clockwise around point X. (Strictly these are computations in the fundamental group of the triply punctured plane, which happens to be a free three-group on generators a,b,c.)

Let $\sigma : [n] \to [n]$. The number of graphs G on $[n]$ such that $xy \in E(G)$ iff $\sigma(x)\sigma(y) \notin E(G)$ is $2^{\cdots}$. Replacing 'graph' by 'r-coloring of the edges of K_n' replaces '2' by r.

proposition: if $Z_1, \dots, Z_n$ are exponentially distributed with same means, then $(X_1, \dots, X_n)$, where $X_i = \frac{Z_i}{Z_1 + \cdots + Z_n}$, is uniformly distributed (wrt. lebesgue) on $\{(z_1, \dots, z_n) : z_i \ge 0$ for all i, $z_1 + \cdots + z_n = 1\}$. proof. If W is a random variable, then p_W is its density function.

$p_X(x) = \int_0^\infty p_Z(tx)dt = \int_0^\infty e^{-(x_1 + \cdots + x_n)t} dt = \int_0^\infty e^{-t} dt$, which does not depend on x.

The following fact is implicit in many places, for example in the [theorem involving the word "abacus"]. However I needed some effort to state and prove it. fact: If λ is a partition, then if the hook number h occurs in λ, then all its positive integer divisors occur as hook numbers in λ. The only crucial observation is the following: the hook number h_{ij} counts the number of boxes in the unique 'snake' S_{ij} with one end in row i and the other in column j. This snake is removable, in the sense that $\lambda \backslash S_{ij}$ is a partiton diagram.

proposition: Suppose $d|h_{11}$. Then there are i, j such that $h_{ij} = d$. proof. Split the snake S_{11} into h_{11}/d consecutive subsnakes $S_1, \ldots, S_r$, each of size d, where $r = h_{11}/d$. It is sufficient to show that at least one of the S_i are removable, since then that snake will correspond to a hook, whose size will of course be d. Mark each end of each snake '+' or '-' in the following way. The left-bottom end of a snake is marked '+' iff removing the remainder(including the end itself) of the row containing the end leaves a partition. Similarly the top-right end is marked '+' if removing the remainder of the column containing that end leaves a partition. So a snake is removable if both ends (which may coincide (but only in a trivial case)) are marked '+'. We need to show that there is such a snake among the S_i. Let us think of the snakes as going from left to right (they are in fact going left - down to right-up) and call the marks on each snake the left or right mark.

The following two claims easily imply that some snake has a double +.

- If S_L, S_R (left, right) are two consecutive snakes, then there is a + to the right in S_L iff there is a '-' to the left in S_R.
- S_1 has a '+' to the left, and S_r has a '+' to the right.

Example:

9	7	5	4	$+$
8	6	4	$-$	S
6	$-$	S	$+$	
S	$+$			
$+$				

. Here S denotes the middle part of each size-three snake. The hooknumber 'behind' the leftmost S is 3.

Other implicit fact: Write $\lambda \to_t \lambda'$ if λ' is a partition and can be obtained from the partition λ by removing some snake of size t. For any partition λ and positive integer t, any sequence $\lambda \to_t \lambda_1 \to_t \cdots \to_t \lambda_0$ of maximal length ends in the same partition λ_0. By induction it is suffices to prove that for any partition λ together with λ_1, λ_2 such that $\lambda \to_t \lambda_1$ and $\lambda \to_t \lambda_2$, there is a partition λ_3 such that $\lambda_1 \to_t \lambda_3$ and $\lambda_2 \to_t \lambda_3$. This is not hard to prove but it is unexpected that it is possible to prove.

In fact, if $a_t(n)$ is the number of partitions of size n that can occur as λ_0's (*t-core partitions*), then

$$\sum_{n=0}^{\infty} a_t(n) x^n = \prod_{n=1}^{\infty} \frac{(1 - x^{tn})^t}{1 - x^n}.$$

entropy Suppose X is a random variable such that $P[X = i] = \frac{1}{n}$ for $i \in \{1, \ldots, n\}$. Then the entropy, or 'information content', $H(X)$, is defined to be $\sum_i p_i \log(\frac{1}{p_i})$.

Why?

Of course one can think about $H(\cdot)$ as simply being a nice way to associate a number to random variables, satisfying useful properties. I will think about this informally and try to motivate calling $H(X)$ the information content of X. We will choose the base 2 for the logarithm above and then the unit of $H(X)$ will be *bits*.

We need to informally motivate the following, after which the formula will follow from manipulations.

1. If X, Y are independent, then $H(X, Y) = H(X) + H(Y)$. Here (X, Y) is the random variable taking values (x, y) where $P[x, y] = P[X = x]P[Y = y]$.

If an experiment has information content 12.34 bits, then it is reasonable that 10 experiments have information content 123.4 bits.

2. If X is equal to the variable Y with probability p and to Z with probability $(1-p)$, then $H(X) = pH(Y) + (1-p)H(Z)$

Again thinking of X, Y, Z as 'experiments', I think the following description is convincing.

Suppose we perform many X-experiments, in total N. This can be done by first selecting randomly for each experiment whether it should be a Y or a Z. If N is large, the number of Y:s is close to pN. So counting the bits in two ways we have $NH(X) = pNH(Y) + (1-p)NH(Z)$.

3. The information content of a fair coin is 1 bit.

I'm fine with this statement.

The formula follows.

Let Y be a number in $[n]$ chosen uniformly at random, and (S_i) be a partition of $[n]$ into r subsets. We define X to be the unique x such that $Y \in S_x$. Let $n_i = \#S_i$. Since Y is uniform, $H(Y) = \log n$ (at least for power-of-two n). The random variable Y is described by the independent(not really) pair of random variables (X, Z) where Z is Z_i with probability n_i/n, and Z_i is uniform in S_i. Hence $\log n = H(Y) = H(X) + H(Z) = H(X) + \sum_i \frac{n_i}{n} H(Z_i) = H(X) + \sum_i \frac{n_i}{n} \log n_i$, which gives $H(X) = \sum_i \frac{n_i}{n} \log(\frac{n}{n_i})$.

I'm not totally happy about these motivations. My objective was to understand how something can have a non-integer number of 'bits' as 'information content'. It is answered above only as 'it's the result of taking averages of integer numbers of bits'. Also, not all random variables can be described as X above.

mass below waist

How does one measure the weight of the part below the waist of a person without killing it?

attempt 1: Identify the mass distribution of the person with the function $f : \mathbb{R} \to \mathbb{R}^+$ with compact support. Here $f(h)dh$ is the mass of the crosscut of height dh at height h above the waist (h may be negative). We are looking for $\int_{-\infty}^{0} f(h)dh$ given some statistics $S(f)$ which can be measured without too much mess.

For $a \in \mathbb{R}$ let $S_a(f) = \int_{-\infty}^{\infty} hf(h+a)dh$. This is the angular momentum when balancing horizontally on a fixed point at $h = a$. Unfortunately, this just tells us the values $A = \int_{-\infty}^{\infty} hf(h)dh$ and $B = \int_{-\infty}^{\infty} f(h)dh$, and it is easy to construct two f:s with same values for A and B but different $\int_{-\infty}^{0} f(h)dh$.

attempt 2: Measure the force F needed to keep the waist at the water line in a pool. This just tells us what the difference of the total mass and the volume below the waist is (multiplying these by fixed physical constants so that the units are right). Clearly both these numbers are easily determined (independently of each other) in different ways. At least, we can find the mass under the waist if we somehow knew the mean density there.

attempt 3: Let $p_1, \ldots, p_n$ be points in space with masses $m_1, \ldots, m_n$. The gravitational field at a point q is $F(q) := \sum_i m_i/d(p_i, q)$. For n generic q_i, the vectors $v_i := (1/d(p_1, q_i), 1/d(p_2, q_i), \ldots, 1/d(p_n, q_i))$ are linearly independent(i think) so knowing $F(q_i)$ suffices to determine the m_i. This shows that the problem can be solved using only measurements of acceleration, but it is not very practical (or, more vaguely but more correct: using classical mechanics (no need for eg. thermodynamics)).

Putting inequalities into context

The Cauchy-Schwarz inequality is

$$\sum x_i y_i \leq \left(\sum_i x_i^2\right)^{1/2} \left(\sum_i y_i^2\right)^{1/2}.$$

By rescaling, it suffices to prove it for unit vectors. For unit vectors it follows from the AM-GM inequality: $x_i y_i \leq (x_i^2 + y_i^2)/2 \Rightarrow \sum x_i y_i \leq 1$ if $\sum_i x_i^2 + \sum_i y_i^2 = 1$.

Now, fix nonnegative integers m, n and consider an array x_{ij} $(1 \leq i \leq m, 1 \leq j \leq n)$ of positive reals together with a sequence $p_1, \ldots, p_m$ of positive reals with sum 1. By the $(m, \mathbf{p})$-Hölder inequality we mean the statement that for any values x_{ij}, we have

$$\sum_i \prod_j x_{ij} \leq \prod_j \left(\sum_i x_{ij}^{1/p_j}\right)^{p_j}.$$

Thus, the $(2, 1/2, 1/2)$-Hölder inequality is the Cauchy-Schwarz inequality, and the $(2, 1/p, 1/q)$-Hölder inequality is the classical Hölder inequality.

Note that by applying Cauchy-Schwarz over and over again to $\sum_i \prod_j x_{ij}$, we obtain $(m, \mathbf{p})$-Hölder inequalities for all $m, \mathbf{p}$ such that the p_i are powers of two (with negative integer exponents).

Now, choose k and let $p = \sum_{j \leq k} p_j$, $q = \sum_{j > k} p_j$, $x_{ij} = u_i^{p_j/p}$ for $j \leq k$ and $x_{ij} = v_i^{p_j/q}$ for $j > k$. Simplify. This shows that $(m, \mathbf{p})$-Hölder implies $(2, p, q)$-Hölder.

The previous two steps and the RHS being continuous in $\mathbf{p}$ imply $(2, p, q)$-Hölder for *all* $p, q \geq 0$ such that $p + q = 1$. Finally $(2, p, q)$-Hölder implies $(2, \mathbf{p})$-Hölder for all $\mathbf{p}$ by induction.