1 Solutions to Inclusion-exclusion problems

Let $A_1, ..., A_n$ be events in a probability space. Let σ_j denote $\sum_{1\le i_1<\cdots<i_j\le n} P(\cap_{\ell=1}^{j} A_{i_\ell})$.
I1) Show that the probability of *exactly* q of these $A_1, ..., A_n$ events occuring is $\sum_{j=q}^{n}\binom{j}{q}(-1)^{q+j}\sigma_j$. Interpret this result for $q = 0$.

The $q = 0$ case is just the ordinary Inclusion-Exclusion formula, where none of the events happen.
The probability of exactly q events happening is

$$P\Big(\bigcup_{\substack{M:|M|=q\\ M\subseteq[n]}}\Big((\cap_{i\in M}A_i)\cap(\cap_{i\notin M}\overline{A_i})\Big)\Big),$$

where the union is disjoint union; therefore it is equal to

$$\sum_{\substack{M:|M|=q\\ M\subseteq[n]}} P\Big((\cap_{i\in M}A_i)\cap(\cap_{i\notin M}\overline{A_i})\Big).$$

We have to show that this term is identical to $\sum_{j=q}^{n}\binom{j}{q}(-1)^{q+j}\sigma_j$. We check it by substituting events of probability 0 or 1 into the A_i's.
Case 1. Every A_i has probability zero. Both expressions have value 0.
Case 2. $m > 0$ of the A_i's have probability 1, the others have probability zero. What we want to evaluate is 0, if $m \neq q$, and 1 if $m = q$. The summation suggested as solution is $\sum_{j=q}^{n}\binom{j}{q}(-1)^{q+j}\binom{m}{j}$. Indeed, if $m = q$, then $\sum_{j=q}^{n}\binom{j}{q}(-1)^{q+j}\binom{q}{j}$ has only one nonzero term, namely $j = q$, which gives value 1. For $m < q$ the summation has only zero terms. For $m > q$, we have: $\sum_{j=q}^{n}\binom{j}{q}(-1)^{q+j}\binom{m}{j} = \sum_{j=q}^{n}\binom{m}{q}\binom{m-q}{j-q}(-1)^{q+j} = \binom{m}{q}\sum_{j=q}^{n}\binom{m-q}{j-q}(-1)^{j-q} = 0$ by the binomial theorem.

I2) Show that for *any* $\ell \ge 1$ natural number,

$$\binom{\ell+1}{2}P\Big(\bigcup_{i=1}^{n}A_i\Big)\ge \ell\sigma_1-\sigma_2.$$

Spell it out as

$$\binom{\ell+1}{2}P\Big(\bigcup_{i=1}^{n}A_i\Big)\ge \ell\sum_i P(A_i)-\sum_{i<j}P(A_i\cap A_j).$$

Use the method of substituting events of probability 1 or 0. If all events have probability zero, both sides are 0. If $m > 0$ events have probability 1, the LHS is $\binom{\ell+1}{2}$. The RHS is $\ell m - \binom{m}{2}$. It is easy to see that $\binom{\ell+1}{2} \ge \ell m - \binom{m}{2}$ is equivalent to $(\ell - m)^2 + (\ell - m) \ge 0$. Though it fails if $0 < \ell - m < 1$, it hold for integer ℓ, m.

I3) Assume that no more than d of the $A_1, ..., A_n$ events can occur simultaneously. Show that

$$P\Big(\bigcup_{i=1}^{n}A_i\Big)\le \sigma_1-\frac{2}{d}\sigma_2.$$

Now we have to be more carelul as the method of substituting 0-1 probability events *does not* allow conditions "no more than d of the events can occur simultaneously". Instead, we show that a corresponding inequality holds for the indicator functions pointwise:

$$\chi_{\cup_{i=1}^{n}A_i}\le\sum_i\chi_{A_i}-\frac{2}{d}\sum_{i<j}\chi_{A_i\cap A_j}.$$

Assume that a point belongs to none of the events, then both functions take value 0 in this point. Assume that a point belongs $d \ge m \ge 1$ of the events, then the LHS takes value 1, while the RHS takes value $m - \frac{2}{d}\binom{m}{2}$, The inequality $1 \ge m - \frac{2}{d}\binom{m}{2}$ is equivalent to $m(m-1) \le d(m-1)$, which is obviously true.

From the inequality for the indicator functions the inequality for probability follows by taking expectation of both sides. This keeps the inequality.

I4) (a) Compute by inclusion-exclusion the number of surjections from an n-element set to an m-element set. (b) Interpret your result for $m = n$.

I5) Let us be given a fixed simple graph G on n labelled vertices. A *good colouration* of the vertices with k colours assigns to every vertex a colour from the set $\{1, 2, ..., k\}$, such that endvertices of any edge get different colours. Show that the number of good colourations of G with k colours is a polynomial of k.

Apply inclusion-exclusion for the events A_e $(e \in E(G))$, where A_e means that the two endpoints of edge e receive the same color in a k-colouration. It is easy to see that $|A_{e_1} \cap \cdots \cap A_{e_m}| = k^c$, where c is the number of components of the graph that has vertex set $V(G)$ and edge set $\{e_1, ..., e_m\}$. Though we obtain lots of monomials in variable k, their number is fully determined by the given graph G, so the result will be a polynomial of k.

I6) (a) How many m-element families of k-subsets of an N-element set have the property, that their union is the underlying N-element set?
(b) Show the identity

$$\sum_{i\geq 0} \binom{2n}{i} (-1)^{i+1} \binom{\binom{2n-i}{2}}{n} = (2n-1)!!.$$

I7) Let A_1, A_2,..., A_n be subsets of an n-element set, such that for all $i < j$ we have $|A_i \cap A_j| \leq 1$. Show that $\sum_{i=1}^{n} |A_i| = O(n^{3/2})$.

$$\sum_{i=1}^{n} |A_i| = \sum_{i:|A_i|<100\sqrt{n}} |A_i| + \sum_{i:|A_i|\geq 100\sqrt{n}} |A_i|.$$

The first sum is easily estimated by $100n^{3/2}$, as there at most n sets. We will show that the second sum has at most $\sqrt{n}$ terms, and therefore this sum is estimated by $n^{3/2}$. Assume that we have some m sets pairwise iintersecting in at most 1 element, each with at least $100\sqrt{n}$ elements. From the Bonferroni inequality $n = |\cup_{i=1}^{n} A_i| \geq \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| \geq 100m\sqrt{n} - \binom{m}{2}$. This line makes a contradiction when $m = \lfloor \sqrt{n} \rfloor$.

I8) Assume that the prime factorization of n is $p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ with different prime factors. Compute the number of integers in $[1, n]$ that are relative primes to n.

I9) There are n points on a straight line. Assume that any distance among these points occurs only once or twice. Show that at least $(n-1)/2$ distances occur only once.

Denote the points by $P_1, P_2, ..., P_n$ as they come in this order. Let A_i denote the set of distances measured from P_i to $P_{i+1}, ..., P_n$. Clearly $|A_i| = n - i$ and $|A_i \cap A_j| \leq 1$ for $i \neq j$ by the condition: for $|A_i \cap A_j| \geq 2$, the distance of P_i and P_j would occur at least 3 times.

Let d_1 denote the number of distances occuring once, and d_2 denote the number of distances occuring twice. We have

$$d_1 + 2d_2 = \binom{n}{2} \tag{1.1}$$

azonos

and

$$d_1 + d_2 = |\cup_{i=1}^{n} A_i| \geq |A_1| + (|A_2| - 1) + (|A_3| - 2) + ... + (|A_n| - (n-1)), \tag{1.2}$$

differenc

as A_i shares at most $i - 1$ elements with $\cup_{j=1}^{i-1} A_j$. In fact, stopping after the last non-negative term in the RHS of (1.2) still gives a lower bound. This means the greatest I such that $n - i \geq i - 1$. We obtain

$$d_1 + d_2 \geq \sum_{1\leq i\leq \frac{n+1}{2}} (n + 1 - 2i) \geq \frac{n^2 - 1}{4} \tag{1.3}$$

differenc

by computing the sum separately for odd and even n. (For even n, the sum is $n^2/4$, for odd n the sum is $(n^2-1)/4$.) Subtracting from (1.3) the half of (1.1), we obtain

$$\frac{d_1}{2} \geq \frac{n^2 - 1}{4} - \frac{1}{2}\binom{n}{2} = \frac{n - 1}{4}$$

we finally obtain the required result.

Note that the result is tight for $2m+1$ points. Take $P_0, ..., P_m$ on a line in this order such that every distance occurs exactly once, and even more, for any $0 < i < j$ and $0 \leq k < \ell \leq n$

$$\overline{P_i P_j} + 2\overline{P_0 P_i} \neq \overline{P_k P_\ell}.$$

Add the mirror images of the points with respect to P_0, to obtain $2m+1$ points. The distances between a point and its mirror image occur exactly once, while all other distances occur exactly twice.

I10) (Area of a parallelogram problem - was shown in class with drawing.) Let us be given an $ABCD$ paralleogram of unit area. Divide the sides into three equal pieces by adding new points in the cyclic order $AEE'BFF'CGG'DHH'$. Define four new points as intersection of segments: $I = \overline{CE} \cap \overline{BH}$, $J = \overline{CE} \cap \overline{DF}$, $K = \overline{GA} \cap \overline{DF}$, $L = \overline{GA} \cap \overline{HB}$. What is the area of the $IJKL$ parallelogram?

Using a sloppy notation not distinguishing between a polygon and its area, we have

$$ABCD - FCD - ABH - EBC - AGD = IJKL - AHL - EIB - FJC - GKD$$

and

$$ABCD - IJKL = ABC + AKD + DCJ + CIB =$$

$$\frac{9}{4}(AHL + EIB + FJC + GKD) = \frac{9}{4}(IJKL + FCD + ABH + EBC + AGD - ABCD).$$

The information $ABCD = 1$, $FCD = ABH = EBC = AGD = \frac{1}{3}$ combined with the equation

$$ABCD - IJKL = \frac{9}{4}(IJKL + FCD + ABH + EBC + AGD - ABCD)$$

gives $IJKL = \frac{1}{13}$.

I11) n persons deposited their hats in the cloakroom, and by bookkeeping error, they get back one hat each, such that any distribution of the n hats happens with the same probability.

(a) What is the probability of the event that nobody gets back his own hat?

(b) What is the limit of this probability as n goes to infinity?