John Boozer and Austin Mohr
Computational Complexity

Problem 1

Prove that if (E, D) is an encryption scheme with message size m and key size $n < m$, then there exist two messages $x, x' \in \{0,1\}^m$ such that $\mathbf{E}_{U_n}(x)$ is not the same distribution as $\mathbf{E}_{U_n}(x')$. (In other words, perfect secrecy is only possible if the key length is at least the message length.)

Proof. Suppose, to the contrary, that $\mathbf{E}_{U_n}(x) = \mathbf{E}_{U_n}(x')$ for all plaintexts x, x'. Let y be some ciphertext in the range of E. For every plaintext x, there is some key $k \in \{0,1\}^n$ such that $\mathsf{E}_k(x) = y$. Now, $\mathsf{E}_k(x) = \mathsf{E}_k(x')$ implies that $x = x'$. Hence, the available 2^n keys are insufficient to map all 2^m distinct plaintexts to the ciphertext y. Therefore, it must be that there exist two messages $x, x' \in \{0,1\}^m$ such that $\mathbf{E}_{U_n}(x)$ is not the same distribution as $\mathbf{E}_{U_n}(x')$. □

Problem 3

Prove: If $\{f_k\}_{k\in\{0,1\}^*}$ is a pseudorandom function family, then for every polynomial $\ell(n)$, the function G that maps $k \in \{0,1\}^n$ to $f_k(1), \ldots, f_k(\ell(n))$ is a secure pseudorandom generator.

Proof. Suppose not. Then there is an evil polynomial time algorithm A so that for all ϵ there exists n such that

$$\left|\Pr[A(G(U_n)) = 1] - \Pr[A(U_{nl(n)}) = 1]\right| \geq \epsilon(n)$$

We create the oracle equipped polynomial time algorithm B^{f_k} which computes $f_k(1), \ldots, f_k(l(n))$ then shoves it into A. $A(G(k)) = B^{f_k}(1^n)$, so

$$\Pr_{k\in_R\{0,1\}^n}[A(G(k)) = 1] = \Pr_{k\in_R\{0,1\}^n}[B^{f_k}(1^n) = 1]$$

Also if g is randomly selected then the distribution of $g(1), \ldots, g(l(n))$ is uniform. Thus,

$$\Pr_{g\in_R F_n}[B^g(1^n) = 1] = \Pr_{k\in_R\{0,1\}^{nl(n)}}[A(k) = 1]$$

Finally,

$$\left|\Pr_{k\in_R\{0,1\}^n}[B^{f_k}(1^n) = 1] - \Pr_{g\in_R F_n}[B^g(1^n) = 1]\right| = \left|\Pr[A(G(U_n)) = 1] - \Pr[A(U_{nl(n)}) = 1]\right| \geq \epsilon(n)$$

So $\{f_k\}$ is not a pseudorandom function family, causing a contradiction. □

Problem 4

Show that if f is a one-way permutation then so is

$$f^{(k)} = \underbrace{f \circ f \circ \cdots \circ f}_{k}$$

where $k = n^c$ for some fixed $c > 0$. Furthermore, show that the assumption that f is a permutation is necessary.

Proof. Assume $f^{(k)}$ is not a one-way function. Then let A be a polynomial time algorithm so that for all ϵ, there exists an n such that

$$\Pr_{\substack{x\in_R\{0,1\}^n \\ y=f^{(k)}(x)}}[f^{(k)}(A(y)) = y] \geq \epsilon(n)$$

Since f and $f^{(k)}$ are both permutations,

$$\begin{aligned}
\Pr_{\substack{x\in_R\{0,1\}^n \\ y=f^{(k)}(x)}}[f^{(k)}(A(y)) = y] &= \Pr_{y\in_R\{0,1\}^n}[f^{(k)}(A(y)) = y] \\
&= \Pr_{\substack{x\in_R\{0,1\}^n \\ y=f(x)}}[f^{(k)}(A(y)) = y] \\
&= \Pr_{\substack{x\in_R\{0,1\}^n \\ y=f(x)}}[f(f^{(k-1)}(A(y))) = y]
\end{aligned}$$

$f^{(k-1)}(A(y))$ can be computed in polynomial time. First compute $A(y)$, which is polynomial time, then apply f $k-1$ times, which takes n^c times the running time of f, which is polynomial time. This contradicts that f is a one-way function.

Furthermore, if a one-way function exists, then a one-way function exists whose outputs and inputs are the same length, specifically f_U. Let g be the function that computes f_U, then outputs a string of length $2n$, consisting of $f_U(x)$ listed twice. Then the length of he output of $f^{(k)}$ is $2^{n^c}n$, which is not printable in polynomial time, much less computable in polynomial time. □

Problem 5

Suppose $x \in \{0,1\}^m$ is an unknown vector. Let $r_1, \ldots, r_m \in \{0,1\}^m$ be randomly chosen, and $x \odot r_i$ revealed to us for all $i \in [m]$. Describe a deterministic algorithm to reconstruct x from this information, and show that the probability (over the choice of the r_i's) is at least $1/4$ that it works. (This shows that if $r_1, \ldots, r_m$ are fully independent then we cannot guess $x \odot r_1, \ldots, x \odot r_m$ with probability much better than 2^{-m}, and hence it was crucial to move to a merely pairwise independent collection of vectors in the proof of the Goldreich-Levin Theorem).

Proof. If $r_1, \ldots, r_m$ are linearly independent vectors then for each $i \in [m]$:

- Determine the linear combination $\sum_{j=1}^m c_j r_j$ which gives the vector e_i with a 1 in the ith position and 0 elsewhere.
- Compute the ith coordinate of x as follows:

$$x_i = x \odot e_i = x \odot \sum_{j=1}^m c_j r_j = \sum_{j=1}^m c_j (x \odot r_j)$$

Now the probability that $r_1, \ldots, r_m$ are linearly independent is:

$$\prod_{i=1}^m \left(\frac{2^i - 1}{2^i}\right) \geq \frac{1}{2} \prod_{i=2}^m \frac{2^{i-1}+1}{2^{i-1}+2} = \frac{2^{m-1}+1}{8} \prod_{i=2}^{m-1} \frac{2^{i-1}+1}{2^i+2} = \frac{2^{m-1}+1}{2^{m+1}} > \frac{1}{4}$$

□

Problem 6

Let V_n denote the binomial distribution on n points with probability $1/3$, i.e.,

$$\mathbf{P}_{V_n}(w) = \left(\frac{1}{3}\right)^{\mathrm{wt}(w)} \left(\frac{2}{3}\right)^{n-\mathrm{wt}(w)}$$

for each $w \in \{0,1\}^n$, where $\mathrm{wt}(w)$ denotes the number of 1's in w. Call a polynomial-time computable function $G : \{0,1\}^* \to \{0,1\}^*$ of stretch $\ell(n)$ a "secure biased pseudorandom generator" if, for every probabilistic polynomial time algorithm A, there exists a negligible function $\epsilon : \mathbb{N} \to [0,1]$ such that, for all n,

$$\left|\mathbf{P}[A(G(U_n)) = 1] - \mathbf{P}[A(V_{\ell(n)}) = 1]\right| < \epsilon(n).$$

Show that, if a secure biased pseudorandom generator exists, then a secure pseudorandom generator of stretch $\alpha\ell(n)$ exists for some fixed $\alpha > 0$. (Hint: You may want to use Chernoff-type bounds: see Corollary A.15 in Arora-Barak.) In particular, provide a *deterministic* polynomial-time algorithm which transforms the output of a secure biased pseudorandom generator into that of a secure pseudorandom generator.

Proof. Let $\alpha = 7/18$ (We'll use it later). To transform the biased pseudorandom number x into a pseudorandom number y, we proceed along x taking two digits at a time. If we find a 00, then the next digit of y is 0, if we find a 01 or 10, then the next digit of y is 1, otherwise do nothing. We continue until either we have $\alpha l(n)$ bits, in which case we output the bits we've collected, or we get to the end of x, in which case we output all zeros. Call this pseudorandom generator g. We need to show that this a secure pseudorandom generator.

Suppose not. Then there is an algorithm A that ruins its day. Let's take an arbitrary negligible ϵ. $\epsilon(n) + 2e^{n/72}$ is also a negligible function so there exists an n such that

$$\Pr[A(g(U_n)) = 1] - \Pr[A(U_{\alpha l(n)}) = 1] \geq \epsilon(n) + 2e^{n/72}$$

Build algorithm B to be the algorithm that takes a biased pseudorandom number, uses our method to compute a pseudorandom number y, then return $A(y)$. Then $B(G(x)) = A(g(x))$, so $\Pr[A(g(U_n)) = 1] = \Pr[B(G(U_n)) = 1]$.

Now $\Pr[B(V_{l(n)}) = 1]$ depends on $A(0^{\alpha l(n)})$. Let E be the event that the input for B has too many 11 pairs, and so our conversion process returns $0^{\alpha l(n)}$. Then if $A(0^{\alpha l(n)}) = 0$

$$\Pr[B(V_{l(n)}) = 1] = (1 - \Pr(E))Pr[A(U_{\alpha l(n)}) = 1]$$

otherwise $A(0^{\alpha l(n)}) = 1$ and

$$\Pr[B(V_{l(n)}) = 1] = (1 - \Pr(E))Pr[A(U_{\alpha l(n)}) = 1] + \Pr(E)$$

Either way

$$\Pr[A(U_{\alpha l(n)}) = 1] - \Pr(E) \leq \Pr[B(V_{l(n)}) = 1] \leq \Pr[A(U_{\alpha l(n)}) = 1] + \Pr(E)$$

Now the expectation for the number of 11 pairs is $l(n)/18$. And I rigged α (this is where we use $7/18$) so that event E is having more than $l(n)/9$ pairs. So we use Chernoff bounds to estimate $\Pr(E)$ with $c = 1$ and $\mu = l(n)/18$.

$$\Pr(E) \leq 2e^{n/72}$$

Hence

$$\Pr[A(U_{\alpha l(n)}) = 1] - 2e^{n/72} \leq \Pr[B(V_{l(n)}) = 1] \leq \Pr[A(U_{\alpha l(n)}) = 1] + 2e^{n/72}$$

Now to drive it home,

$$\begin{aligned}\left|\Pr[B(G(U_n)) = 1] - \Pr[B(V_{l(n)}) = 1]\right| &\geq \left|\Pr[A(g(U_n)) = 1] - \Pr[A(U_{\alpha l(n)}) = 1]\right| - 2e^{n/72} \\ &\geq \epsilon(n)\end{aligned}$$

Thus G is not a secure biased pseudorandom generator, which is a contradiction. □

1. Let $S \subset \mathbb{Z}_p^\times$ be such that $S = -S$, where p is prime. Define the "circulant graph" G_S by $V(G_S) = \mathbb{Z}_p$ and $\{x, y\} \in E(G_S)$ iff there exists $s \in S$ so that $x + s = y$. Define the k^{th} discrete Fourier coefficient of a function $f : \mathbb{Z}_p \to \mathbb{C}$, for $k \in \mathbb{Z}_p$, by

$$\hat{f}(k) = \frac{1}{\sqrt{p}} \sum_{j=0}^{p-1} e^{2\pi ijk/p} f(j).$$

Show that

$$\lambda(G) = \frac{\sqrt{p}}{|S|} \max_{k \neq 0} |\hat{\chi}_S(k)|,$$

where χ_S is the characteristic function of S. Hint: First prove that $\{f_k\}_{k \in \mathbb{Z}_p}$ is an orthonormal basis of $\mathbb{C}^{\mathbb{Z}_p}$, where $f_k(j) = \exp(2\pi ijk/p)/\sqrt{p}$.

Proof.

$$< f_k, f_k >= \sum_{j=0}^{p-1} \frac{e^{2\pi ijk/p}}{\sqrt{p}} \frac{e^{-2\pi ijk/p}}{\sqrt{p}} = \sum_{j=0}^{p-1} \frac{1}{p} = 1$$

Let $k \neq l$

$$< f_k, f_l >= \sum_{j=0}^{p-1} \frac{e^{2\pi ijk/p}}{\sqrt{p}} \frac{e^{-2\pi ijl/p}}{\sqrt{p}} = \sum_{j=0}^{p-1} \frac{e^{2\pi ij(k-l)/p}}{p} = 0$$

So $\{f_k\}_{k\in\mathbb{Z}_p}$ is an orthonormal basis.

Let A be the random walk matrix of G_S. $A_{l,j} = \frac{\chi_S(l-j)}{|S|}$ Then the jth entry of Af_k is

$$\begin{aligned}\frac{1}{|S|}\sum_{l=0}^{p-1}\chi_S(l-j)\frac{e^{2\pi ilk/p}}{\sqrt{p}} &= \frac{e^{2\pi ijk/p}}{\sqrt{p}}\frac{1}{|S|}\sum_{l=0}^{p-1}\chi_S(l-j)e^{2\pi i(l-j)k/p} \\ &= \frac{\sqrt{p}}{|S|}\hat{\chi}_S(k)\frac{e^{2\pi ijk/p}}{\sqrt{p}}\end{aligned}$$

So f_k is an eigenvector with the eigenvalue $\frac{\sqrt{p}}{|S|}\hat{\chi}_S(k)$. Since $\lambda(G)$ is the second largest eigenvalue, with the largest occurring with f_0, then taking the maximum of the eigenvalues of $f_1, \ldots, f_{p-1}$ gives us $\lambda(G)$. Note that since f_k is a basis of eigenvectors, we are assured of getting all of the eigenvalues. □

2. Let A and B be symmetric stochastic matrices. Prove that $\lambda(A+B) \leq \lambda(A) + \lambda(B)$.

Proof. Let $\mathbf{w}$ be the vector that maximizes $||(A+B)\mathbf{v}||_2$ over all vectors $\mathbf{v} \in \mathbf{1}^\perp$.

$$\begin{aligned}\lambda(A+B) &= ||(A+B)\mathbf{w}||_2 \\ &= ||A\mathbf{w} + B\mathbf{w}||_2 \\ &\leq ||A\mathbf{w}||_2 + ||B\mathbf{w}||_2 \text{ (by the triangle inequality)}\end{aligned}$$

It is clear that $||A\mathbf{w}||_2 \leq \max_{\mathbf{v}\in\mathbf{1}^\perp} ||A\mathbf{v}||_2$ and $||B\mathbf{w}||_2 \leq \max_{\mathbf{v}\in\mathbf{1}^\perp} ||B\mathbf{v}||_2$, hence

$$\begin{aligned}\lambda(A+B) &\leq ||A\mathbf{w}||_2 + ||B\mathbf{w}||_2 \\ &\leq \max_{\mathbf{v}\in\mathbf{1}^\perp} ||A\mathbf{v}||_2 + \max_{\mathbf{v}\in\mathbf{1}^\perp} ||B\mathbf{v}||_2 \\ &= \lambda(A) + \lambda(B)\end{aligned}$$

□

3. Prove that, for every n-vertex d-regular graph, there is some subset S of $n/2$ vertices so that $|E(S,\overline{S})| \leq dn/4 + O(1)$. Conclude that no (n,d,ρ)-edge expander family exists if $\rho > 1/2$.

Proof. Let S be a random subset of $\frac{n}{2}$ vertices. The probability of an edge belonging to $E(S,\overline{S})$ is the same as the probability that its endpoints belong to different subsets (with respect to S and $\overline{S}$). Now, the presence of loops in G can only lower this probability, making the desired upper bound easier to achieve. Assume, then, that G does not contain loops. Fix a vertex v in G. Each of its d neighbors has probability $\frac{\frac{n}{2}}{n-1} = \frac{1}{2} + \frac{1}{2(n-1)}$ of belonging to the opposite subset (since there are no loops, we know that v is not a neighbor of itself, hence the denominator of $n-1$). We see that v contributes an average of $d(\frac{1}{2} + \frac{1}{2(n-1)})$ edges to $E(S,\overline{S})$. Taken over all vertices of G (with a factor of $\frac{1}{2}$ to account for the double-counting of edges), the expected value of $|E(S,\overline{S})|$ is

$$\begin{aligned}\frac{n}{2}\cdot d(\frac{1}{2} + \frac{1}{2(n-1)}) &= \frac{nd}{4} + \frac{nd}{4(n-1)} \\ &= \frac{nd}{4} + O(1)\end{aligned}$$

Therefore, there must be some subset S such that $|E(S,\overline{S})| \leq \frac{nd}{4} + O(1)$.

Now, suppose $\rho > \frac{1}{2}$. A family of graphs $\{G_n\}$ is an (n,d,ρ)-edge expander family if, for each G_n, any subset S of at most $\frac{n}{2}$ vertices has the property that $|E(S,\overline{S})| \geq \rho d|S|$. If we let $|S| = \frac{n}{2}$ and use the given ρ, we require that $|E(S,\overline{S})| > \frac{nd}{4}$. By the first part of the problem, we know that, for large n, we can find some S that breaks this bound. Hence, there is no edge expander family for $\rho > \frac{1}{2}$. □

4. Let G be an (n, D, ρ)-edge expander and G' be a (D, d, ρ')-edge expander, for $\rho, \rho' > 0$. Prove that $G ® G'$ is a $(nD, 2d, \rho^2\rho'/80)$-edge expander.

Proof. Let $H = G ® G'$ and let S be a subset of $V(H)$ of at most $\frac{nD}{2}$ vertices. We can view H as being made up of n D-vertex clusters, and so we can partition the vertices in S based on the cluster to which they belong. To establish the claim, we must show that there are $\frac{1}{80}\rho^2\rho' \cdot 2d|S|$ edges leaving S.

Define S_i to be the set of vertices of S belonging to cluster i. We partition the indices $1, \ldots, n$ of the clusters into two sets I' and I'' as follows: If $|S_i| \leq (1 - \frac{1}{4}\rho)D$, put i in I'. Otherwise, put i in I''. Finally, define $S' = \bigcup_{i \in I'} S_i$ and $S'' = \bigcup_{i \in I''} S_i$. Observe that $S_{\prime} \cup S_{\prime\prime} = S$.

Case $|S'| \geq \frac{1}{10}\rho|S|$

Since G' is a (D, d, ρ')-edge expander, there are at least $\frac{1}{4}\rho\rho' d|S_i|$ edges leaving S_i within each cluster. Taken over all clusters, we see there are $\frac{1}{4}\rho\rho' d|S'|$ edges leaving $S' \subset S$. Since $|S'| \geq \frac{1}{10}\rho|S|$, we conclude that there are at least $\frac{1}{80}\rho^2\rho' \cdot 2d|X|$ edges leaving S in H.

Case $|S'| \leq \frac{1}{10}\rho|S|$

We see immediately that $|S''| > (1 - \frac{1}{10}\rho)|S|$. For $i \in I''$, $|S_i| \geq (1 - \frac{1}{4}\rho)D$, and so $\frac{|S''|}{D} \leq |I''| \leq \frac{|S''|}{(1-\frac{1}{4}\rho)D}$. Furthermore, since $|S''| \leq \frac{1}{2}nD$, we see that $|I''| \leq \frac{2}{3}n$. Since G is an (n, D, ρ)-edge expander, we can find a set of edges M with $|M| \geq \frac{1}{2}\rho D|I''|$ between the vertices of I' and I'' (that is, the sets X_i with i belonging to I' and I'', respectively). Now, there are $d|M| \geq \frac{1}{\rho}dD|I''|$ edges in H connecting vertices in the cycles whose index belongs to I' with vertices in the clusters whose index belongs to I''. For $i \in I''$, $|S_i| \geq (1 - \frac{1}{4}\rho)D$. Hence, at most $\frac{1}{4}\rho dD|I''|$ of the $d|M|$ edges connect a vertex not in S_i with a vertex in one of the clusters whose index belongs to I'. So, for $i \in I''$, at least $\frac{1}{4}\rho dD|I''|$ edges connect vertices of the S_i with vertices of the aforementioned clusters. The number of edges connecting the remaining clusters (having $i \in I''$) with S' is at most $d|S'|$. Since $|S'| \leq \frac{1}{10}\rho|S| \leq \frac{1}{6}\rho D|I''|$, there are at most $\frac{1}{6}\rho dD|I''|$ corresponding edges. Hence, at least $\frac{1}{12}\rho dD|I''|$ edges connect vertices of the S_i ($i \in I''$) with the complement of S_j ($j \in I'$) inside the cycles with index j. Since $|I''| \geq \frac{|S''|}{D}$ and $|S''| \geq \frac{1}{2}|S|$, we conclude that there at least $\frac{1}{48}\rho \cdot 2d|S|$ edges leaving S, proving the claim. □

5. A (countably) infinite locally finite graph G is said to be an f-expander, for $f : \mathbb{N} \to \mathbb{R}^+$, if

$$\min_{S : n \leq |S| < \infty} |E(S, \overline{S})| = \Theta(f(n))$$

Give an example of a 1-expander, an $n^{2/3}$-expander and an n-expander. (Of course, you need to provide proofs that they work.)

Claim. Define P to be the path beginning at a root r and extending infinitely in one direction. P is a 1-expander.

Proof. Since the graph is infinite and connected, $|E(S, \overline{S})| \geq 1$ for any S. Let $|S| = n$. If we construct S by selecting the n vertices closest to r (including r itself), we see that $|E(S, \overline{S})| = 1 = \Theta(1)$. Hence, P is a 1-expander. □

Claim. Define T to be an infinite binary tree with root r. T is an n-expander.

Proof. Suppose first that the graph induced by S (call it T') is disconnected. Define the e-degree of a vertex (subgraph) to be the number of edges that vertex (subgraph) contributes to $E(S, \overline{S})$. By translating this lowest component of T' upward so that it becomes connected with another component, we can only reduce the e-degree of both components. Hence, the S which gives the minimum value of $|E(S, \overline{S})|$ must induce a connected subgraph.

Observe next that if we translate T' so that it is rooted at r (that is, we rechoose S so that we have a new tree that is isomorphic to T' but rooted at r), then $|E(S, \overline{S})|$ can only be reduced. This is because the number of edges in $E(S, \overline{S})$ that extend *below* the tree remains unchanged, while some edges in $E(S, \overline{S})$ that extended *above* the tree have been removed during the tranlsation. Hence, the overall e-degree of T' is reduced. We will assume, then, that T' is rooted at r.

Now, let $|S| = n$. We observe next that $|E(S, \overline{S})| = n + 1$. To see this, build T' by first choosing r and successively appending edges as prescribed by S. When we have only r, $|E(S, \overline{S})| = 2$. At the next stage, we append an leaf to r. This decreases the e-degree of r by 1, but the leaf itself has e-degree of 2, resulting in a net gain of one edge in $E(S, \overline{S})$. Continuing in this way, we see that $|E(S, \overline{S}| = n + 1 = \Theta(n)$. □

Claim. Define T to be the infinite three-dimensional lattice. T is an $n^{\frac{2}{3}}$-expander.

Proof. The minimization of $|E(S, \overline{S})|$ is equivalent to the minimization of the "surface area" (i.e. the number of vertices of S exposed to vertices of $\overline{S}$), which occurs when the graph induced by S forms a cube. Let $|S| = n$ with $n = k^3$. The number of vertices on the surface is around $6k^2 = 6n^{\frac{2}{3}} = \Theta(n^{\frac{2}{3}})$. □

1. Show directly that $P_1(s) \Rightarrow P_1(t)$ for all $t \leq s$.

Proof. Given G, let H_s be a subgraph of G on s vertices such that $\#(H_s \sqsubset G) = n^s 2^{-\binom{s}{2}}(1 + o(1))$. Let H_{s-1} be a graph on $s - 1$ vertices. Each embedding of H_{s-1} can be extended to n embeddings of s-vertex graphs by adding one of the n vertices as the s^{th} vertex. We count these extended embeddings and divide by n.

Each embedding is a copy of our original graph plus an additional vertex. There are 2^{s-1} such graphs, each embedded $n^s 2^{-\binom{s}{2}}(1 + o(1))$ times. So

$$\begin{aligned}
\#(H_{s-1} \sqsubset G) &= \frac{2^{s-1}}{n} \cdot \#(H_s \sqsubset G) \\
&= \frac{2^{s-1}}{n}(n^s 2^{-\binom{s}{2}}(1 + o(1))) \\
&= n^{s-1} 2^{-\binom{s-1}{2}}(1 + o(1)).
\end{aligned}$$

□

2. Show that, for a sequence of G's with $n \to \infty$,

$$P_1(3) \not\Rightarrow P_1(4).$$

Proof. Let G_n be comprised of two K_n and one $K_{n,n}$ connected in the following way:

- there are no edges between the K_n
- between the $K_{n,n}$ and each K_n, every possible edge occurs with probability $\frac{1}{2}$

To see that G_n satisfies $P_1(3)$, observe first that it suffices to count only copies of K_3 and P^1 (here, we use P^j to denote a path of length j to distinguish from the property names), since G_n is self-complementary. Hence, if the number of K_3 is correct, then the number of $\overline{K_3}$ is correct. Similarly, if the number of P^1 is correct, then the number of $\overline{P^1} = P^2$ is correct.

We have the following cases for the K_3.

If all three vertices belong to a K_n, then there are $2 \cdot n \cdot n \cdot n = 2n^3$ copies of K_3 (2 choices for which K_n the vertices belong to and n choices for each vertex).

If exactly two vertices belong to a K_n, then there are $2 \cdot 3 \cdot n \cdot n \cdot 2n \cdot \frac{1}{4} = 3n^3$ copies of K_3 (2 choices for which K_n the two vertices belong to, 3 choices for which vertex of the K_3 belongs to the $K_{n,n}$, n possible choices for each of the two vertices within the K_n, and $2n$ choices for the vertex in the $K_{n,n}$ having a $\frac{1}{4}$ probability of successfully linking up with the vertices in the K_n).

If exactly one vertex belongs to a K_n, then there are $2 \cdot 3 \cdot n \cdot n \cdot n \cdot 2 \cdot \frac{1}{4} = 3n^3$ (2 choices for which K_n the two vertices belong to, 3 choices for which vertex of the K_3 belongs to the K_n, n possible choices for the vertex within the K_n, n choices for the vertex in the left partite set of the $K_{n,n}$, n choices for the vertex in the right partite set of the $K_{n,n}$, 2 ways to flip the partite sets, and a $\frac{1}{4}$ probability of successfully linking up with the vertex in the K_n).

In total, we determine there are $8n^3$ copies of K_3 in G_n (which has $4n$ vertices), as desired.

Next, we consider the following cases for the P^1. For convenience, we view P^1 as a single edge and an isolated vertex.

If the edge lies in a K_n, there are $2 \cdot n \cdot n \cdot (n + \frac{2n}{4}) = 3n^3$ copies of P^1 (2 choices for which K_n the edge belongs to, n choices for each of the two vertices comprising the edge, and $n + \frac{2n}{4}$ choices for the isolated vertex representing placement in the other K_n or the $\frac{1}{4}$ chance of successful placement in the $K_{n,n}$, respectively).

If the edge lies in the $K_{n,n}$, there are $2 \cdot n \cdot n \cdot n \cdot 2 \cdot \frac{1}{4} = n^3$ copies of P^1 (2 choices for which K_n the isolated vertex belongs to, n choices for the vertex in the left partite set of the $K_{n,n}$, n choices for the vertex in the right partite set of the $K_{n,n}$, 2 ways to flip the partite sets, and a $\frac{1}{4}$ probability of keeping the isolated vertex isolated).

If the edge lies between a K_n and the $K_{n,n}$, there are $2 \cdot n \cdot 2n \cdot 2 \cdot \frac{1}{2} \cdot (\frac{n}{2} + \frac{n}{2}) = 4n^3$ copies of P^1 (2 choices for which K_n the isolated vertex belongs to, n choices for the vertex in the K_n, $2n$ choices for the vertex in the $K_{n,n}$, 2 permutations of the labels of these vertices, a $\frac{1}{2}$ chance of successfully linking these vertices, and finally an $\frac{n}{2}$ chance of placing the isolated vertex in the same partite set of the $K_{n,n}$ or a $\frac{n}{2}$ chance of placing the isolated vertex in the other K_n).

In total, we determine there are $8n^3$ copies of P^1 in G_n (which has $4n$ vertices), as desired.

Next, we demonstrate that there are too many (more than $4n^4$) copies of K_4 in G_n, and so conclude that G_n does not satisfy $P_1(4)$. We again consider cases.

If all four vertices lie in a K_n, there are $2 \cdot n \cdot n \cdot n \cdot n = 2n^4$ copies of K_4 (2 choices for which K_n the vertices belong to, and n choices for each of the vertices within the K_n).

If exactly three of the vertices are in a K_n, there are $2 \cdot n \cdot n \cdot n \cdot 2n \cdot \frac{1}{8} \cdot 4 = 2n^4$ copies of K_4 (2 choices for which K_n the vertices belong to, n choices for each of these three vertices, $2n$ choices for the vertex in the $K_{n,n}$, and a $\frac{1}{8}$ probability of successfully linking all the vertices, and 4 ways to permute the labels).

If exactly two of the vertices are in a K_n, there are $2 \cdot n \cdot n \cdot n \cdot n \cdot \frac{1}{16} \cdot \binom{4}{2} = \frac{3}{4}n^4$ copies of K_4 (2 choices for which K_n the vertices belong to, n choices for each vertex within the K_n, n choices for the vertex in the left partite set of $K_{n,n}$, n choices for the vertex in the right partite set of $K_{n,n}$, a $\frac{1}{16}$ chance of successfully linking the vertices, and $\binom{4}{2}$ ways to permute the labels).

We have already accounted for more than $4n^4$ copies of K_4 in G_n. Hence, G_n does not satisfy $P_1(4)$.

Therefore, $P_1(3) \not\Rightarrow P_1(4)$. □

3. Show that there is a universal constant C so that for any collection of n line segments in the plane, there is some subset S of them with

$$\left| e(S) - \frac{|S|^2}{4} \right| \geq Cn^2.$$

where $e(S)$ means the number of pairs of segments in S which cross one another.

Proof. Define the incidence matrix G where each vertex represents a line segment in the plane and two vertices are adjacent if and only if their corresponding line segments intersect. Observe that the negation of P_4 implies the desired claim. Hence, it suffices to show that G is not quasirandom. We accomplish this by providing a graph that cannot occur as an induced subgraph of G, thereby violating $P_1(s)$.

Let H be a copy of K_5 in which every edge has been subdivided. Consider the "original" vertices of H (that is, those vertices that were present before the subdivisions). None of these vertices are adjacent in H, and so represent line segments in the plane that are pairwise non-intersecting. Call this set of line segments A. Now, the vertices arising as a result of the subdivisions specify another set of line segments (call this set of line segments B). The graph H specifies that each pair of line segments in A is intersected by a single line segment in B in a bijective manner and that all line segments in B are mutually non-intersecting. This is equivalent to drawing K_5 in the plane without intersection, which is impossible. Therefore, H cannot be a subgraph of G, which violates $P_1(15)$, thus asserting the desired claim. □

4. Show that a random graph is quasirandom.

Proof. We proceed by demonstrating that the random graph satisfies P_5. We seek to employ Chebyshev's Inequality,

$$\Pr\left(|X - E[X]| \geq k\sigma\right) \leq \frac{1}{k^2}$$

where our random variable X is the number of common neighbors. On an n-vertex graph, we take $\mu = \frac{n-2}{4}$ as the expected number of common neighbors of some fixed vertices v and w. To determine the total variance, consider the variance at a single vertex u. Now, u can either be a common neighbor of v, w or it cannot, so X can take on only values 0 or 1. Hence, we obtain

$$\begin{aligned} Var(X) &= E[X^2] - E[X]^2 \\ &= \frac{1}{4} - (\frac{1}{4})^2 \\ &= \frac{3}{16} \end{aligned}$$

Summing over all possible u, we obtain a total variance of $\frac{3}{16}(n-2)$. Hence, $\sigma = c\sqrt{n}$, some constant c. Returning to Chebyshev, we have

$$\begin{aligned} \Pr\left(|X - E[X]| \geq k\sigma\right) &\leq \frac{1}{k^2} \\ \Pr\left(\left|X - \frac{n-2}{4}\right| \geq kc\sqrt{n}\right) &\leq \frac{1}{k^2} \\ \Pr\left(\left|X - \frac{n-2}{4}\right| = o(n)\right) &\geq \frac{k^2-1}{k^2} \end{aligned}$$

for all $k = o(\sqrt{n})$, as desired. □

5. Define the Paley Graph G_p, for p a prime congruent to 1 mod 4, by setting $V(G_p) = \mathbb{Z}_p$ and $\{x, y\} \in G_p$ iff $x - y$ is a quadratic residue (i.e., $x - y = s^2$ for some $s \in \mathbb{Z}_p$). Show that G_p is quasirandom.

Proof. We count the number of common neighbors for any two vertices $v, w \in G_p$. In other words, we count the number of solutions to the equation

$$v - w = s + t,$$

where s and t are quadratic residues. Since $p \equiv 1 \pmod 4$, $-t$ is a quadratic residue whenever t is a quadratic reside. Hence, we can consider the equivalent equation

$$v - w = s - t.$$

Consider,

$$\begin{aligned} v - w &= (x+h)^2 - x^2 \\ v - w &= 2hx + h^2 \\ v - w - h^2 &= 2hx \\ (2h)^{-1}(v - w - h^2) &= x \end{aligned}$$

So long as $h \neq 0$, we get one solution for each h (not necessarily distinct).

If a is a solution to

$$(x+h)^2 - x^2 = v - w$$

then $-a$ is a solution to

$$(x-h)^2 - x^2 = v - w,$$

so if we have a, we don't need $-a$. So, we throw out exactly half of our solutions.

If a is a solution to

$$(x+h)^2 - x^2 = v - w$$

then it is also a solution to

$$(x - 2a - h) - x^2 = v - w.$$

So, a occurs at least twice in our list of solutions so long as $h \neq -2a - h$. Now, $h = -2a - h$ implies $a^2 = v - w$, which happens only once. Furthermore, if $a = 0$, then $x - 2a - h = x - h$, and we have already thrown out either the solution to $(x-h)^2 - x^2 = v - w$ or $(x+h)^2 - x^2 = v - w$. So, we throw out nearly half of our solutions. With $p-1$ possibilities for h, if $v - w$ is not a quadratic residue, we have at most $\frac{p-1}{4}$ distinct solutions. If $v - w$ is a quadratic residue, we have at most $\frac{p+3}{4}$ distinct solutions. This gives P_5, and so G_p is quasirandom. □

1. Two n-sided dice with sides labeled 1 through n are rolled, resulting in the i.i.d. random variables X and Y. Let $Z = X + Y$. Compute $H(X)$, $H(X,Y)$, $H(Z)$, $H(X|Z)$, $I(X;Y)$ and $I(X;Y|Z)$.

We compute the entropy of X, we compute directly.

$$\begin{aligned} H(X) &= -\sum_{i=1}^{n} \mathbf{P}(X = i) \log\left(\mathbf{P}(X = i)\right) \\ &= -\sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{n}\right) \\ &= \log(n) \end{aligned}$$

We compute the the joint entropy of X and Y directly.

$$\begin{aligned} H(X,Y) &= H(X) + H(Y \mid X) \\ &= H(X) + H(Y) && \text{(since } X \text{ and } Y \text{ are independent)} \\ &= 2\log(n) \end{aligned}$$

To determine the entropy of $Z = X + Y$, observe that there are a total of n^2 possible outcomes. There is no way to get $Z = 1$. For $2 \leq k \leq n+1$, there are $k-1$ ways to achieve $Z = k$ (for a given k, we have the results $(1, k-1)$, $(2, k-2)$, $\cdots$, $(k-1, 1)$). For $n+2 \leq k \leq 2n-1$, we see by a similar argument that there are $2n - (k-1)$ ways. Observe that for sums other than $n+1$, there is a kind of parity (both 2 and $2n$ have 1 possibility, 3 and $2n-1$ have 2 possibilities, etc.). Combining this information, we have

$$\begin{aligned}
H(Z) &= -\sum_{i=2}^{2n} \mathbf{P}\left(Z = i\right) \log\left(\mathbf{P}\left(Z = i\right)\right) \\
&= -2\sum_{i=2}^{n+1} \mathbf{P}\left(Z = i\right) \log\left(\mathbf{P}\left(Z = i\right)\right) + \mathbf{P}\left(Z = n+1\right) \log\left(\mathbf{P}\left(Z = n+1\right)\right) \\
&= -2\sum_{i=2}^{n+1} \frac{i-1}{n^2} \log\left(\frac{i-1}{n^2}\right) + \frac{n}{n^2} \log\left(\frac{n}{n^2}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i \log\left(\frac{i}{n^2}\right) + \frac{1}{n} \log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i \left(\log(i) - \log(n^2)\right) + \frac{1}{n} \log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\left(\sum_{i=1}^{n} i \log(i) - \sum_{i=1}^{n} i \log(n^2)\right) + \frac{1}{n} \log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i \log(i) + \frac{2}{n^2} \cdot \frac{n(n+1)\log(n^2)}{2} + \frac{1}{n} \log\left(\frac{1}{n}\right) \\
&= -\frac{2}{n^2}\sum_{i=1}^{n} i \log(i) + \frac{(2n+1)\log(n)}{n}
\end{aligned}$$

To determine the entropy of X given the value of Z, we employ a similar parity argument to obtain

$$\begin{aligned}
H(X \mid Z) &= -\sum_{i=2}^{2n} \left(\sum_{j=1}^{n} \mathbf{P}(X = j \mid Z = i) \log(\mathbf{P}(X = j \mid Z = i))\right) \\
&= -2\sum_{i=2}^{n+1} \left(\frac{i-1}{n^2} \sum_{j=1}^{i-1} \frac{1}{i-1} \log\left(\frac{1}{i-1}\right)\right) + \frac{1}{n}\sum_{j=1}^{n} \frac{1}{n} \log\left(\frac{1}{n}\right) \\
&= -2\sum_{i=2}^{n+1} \left(\frac{i-1}{n^2} \log\left(\frac{1}{i-1}\right)\right) + \frac{1}{n} \log\left(\frac{1}{n}\right) \\
&= \frac{2}{n^2}\sum_{i=1}^{n} i \log(i) - \frac{1}{n} \log(n)
\end{aligned}$$

We compute the mutual information of X and Y directly.

$$\begin{aligned}
I(X;Y) &= H(X) - H(X \mid Y) \\
&= H(X) - H(X) && \text{(since } X \text{ and } Y \text{ independent)} \\
&= 0
\end{aligned}$$

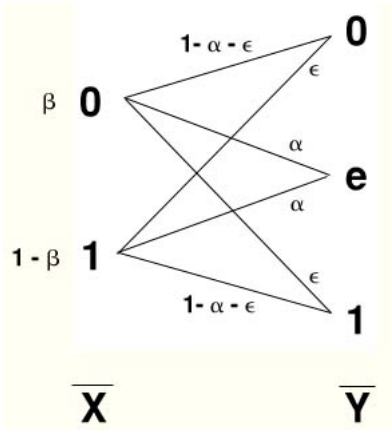
We compute the mutual information of X and Y given the value of Z directly.

$$\begin{aligned}
I(X;Y \mid Z) &= H(X \mid Z) - H(X \mid (Y,Z)) \\
&= H(X \mid Z)
\end{aligned}$$

since knowing both Y and Z completely determines X.

2. Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be ϵ and the probability of erasure be α. (Hence, the probability of correct transmission is $1-\alpha-\epsilon$.) What is the capacity of this channel?

We can visualize the channel in the following way



where e denotes erasure and we write the probability distribution of the input in terms of β. The capacity of the channel is given by

$$\max_{\mathbf{P}(X)} I(X;Y) = \max_{\mathbf{P}(X)} [H(Y) - H(Y \mid X)]$$

Now

$$\begin{aligned}
H(Y) &= -\sum_{y \in Y} \mathbf{P}\left(Y=y\right) \log\left(\mathbf{P}\left(Y=y\right)\right) \\
&= -[\mathbf{P}(Y=0)\log(\mathbf{P}(Y=0)) \\
&\quad + \mathbf{P}(Y=e)\log(\mathbf{P}(Y=e)) \\
&\quad + \mathbf{P}(Y=1)\log(\mathbf{P}(Y=1))] \\
&= -[(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon)\log(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon) \\
&\quad + (\beta\alpha+(1-\beta)\alpha)\log(\beta\alpha+(1-\beta)\alpha) \\
&\quad + (\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))\log(\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))] \\
&= -[(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon)\log(\beta(1-\alpha-\epsilon)+(1-\beta)\epsilon) \\
&\quad + \alpha\log(\alpha) \\
&\quad + (\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))\log(\beta\epsilon+(1-\beta)(1-\alpha-\epsilon))]
\end{aligned}$$

and

$$\begin{aligned}
H(Y \mid X) &= -\sum_{x \in X} \mathbf{P}(X = x) \left(\sum_{y \in Y} \mathbf{P}(Y = y \mid X = x) \log(\mathbf{P}(Y = y \mid X = x)) \right) \\
&= -\mathbf{P}(X = 0)[\mathbf{P}(Y = 0 \mid X = 0) \log(\mathbf{P}(Y = 0 \mid X = 0)) \\
&\quad + \mathbf{P}(Y = e \mid X = 0) \log(\mathbf{P}(Y = e \mid X = 0)) \\
&\quad + \mathbf{P}(Y = 1 \mid X = 0) \log(\mathbf{P}(Y = 1 \mid X = 0))] \\
&\quad - \mathbf{P}(X = 1)[\mathbf{P}(Y = 0 \mid X = 1) \log(\mathbf{P}(Y = 0 \mid X = 1)) \\
&\quad + \mathbf{P}(Y = e \mid X = 1) \log(\mathbf{P}(Y = e \mid X = 1)) \\
&\quad + \mathbf{P}(Y = 1 \mid X = 1) \log(\mathbf{P}(Y = 1 \mid X = 1))] \\
&= -\beta[(1 - \alpha - \epsilon) \log(1 - \alpha - \epsilon) \\
&\quad + \alpha \log(\alpha) \\
&\quad + \epsilon \log(\epsilon)] \\
&\quad - (1 - \beta)[\epsilon \log(\epsilon) \\
&\quad + \alpha \log(\alpha) \\
&\quad + (1 - \alpha - \epsilon) \log(1 - \alpha - \epsilon)] \\
&= -(\epsilon \log(\epsilon) + \alpha \log(\alpha) + (1 - \alpha - \epsilon) \log(1 - \alpha - \epsilon))
\end{aligned}$$

Since $H(Y \mid X)$ does not depend on β, we need only to maximize $H(Y)$, which occurs at $\beta = \frac{1}{2}$. So

$$\begin{aligned}
\max_{\mathbf{P}(X)} I(X;Y) &= \max_{\mathbf{P}(X)} [H(Y) - H(Y \mid X)] \\
&= -(1 - \alpha) \log(\frac{1}{2}(1 - \alpha)) + \epsilon \log(\epsilon) + \alpha \log(\alpha) + (1 - \alpha - \epsilon) \log(1 - \alpha - \epsilon)
\end{aligned}$$

3. Let $s_t(n)$ denote the number of graphs on n vertices not containing a K_t, $t \geq 3$. Show that

$$1 - \frac{1}{t-1} + o(1) \leq \frac{\log s_t(n)}{\log N} \leq 1 - \epsilon_t,$$

where $N = 2^{\binom{n}{2}}$ is the number of total graphs on n vertices and

$$\epsilon_t = \frac{1}{\binom{t}{2} 2^{\binom{t}{2}}}.$$

Hint: Shearer's Inequality.

Proof. Let Y be the random variable representing choosing a random graph on n vertices without a K_t. Let X_i represent the ith edge of that graph.

$$log(S_t(n)) = H(Y) = H(X_1 \dots X_{\binom{n}{2}})$$

Let G_i be the set of possible edges in a t-vertex subgraph. This is a $\binom{n-2}{t-2}$ covering of the edges. So by Shearer's inequality:

$$
\begin{aligned}
H(X_1 \dots X_{\binom{n}{2}}) &\leq \frac{1}{\binom{n-2}{t-2}} \sum_{G_i} H(X_{G_i}) \\
&\leq \frac{1}{\binom{n-2}{t-2}} \sum_{G_i} log(2^{\binom{t}{2}} - 1) \\
&= \frac{\binom{n}{t}}{\binom{n-2}{t-2}} log(2^{\binom{t}{2}} - 1) \\
&= \frac{\binom{n}{2}}{\binom{t}{2}} log(2^{\binom{t}{2}} - 1)
\end{aligned}
$$

Now a little calculus tells us $log(2^{\binom{t}{2}} - 1) \leq \binom{t}{2} - \frac{1}{ln(2)2^{\binom{t}{2}}}$, since the derivative of $log(x)$ is $\frac{1}{ln(2)x}$ and $log(x)$ is concave down. Also $\binom{t}{2} - \frac{1}{ln(2)2^{\binom{t}{2}}} \leq \binom{t}{2} - \frac{1}{2^{\binom{t}{2}}}$. Thus

$$
log(S_t(n)) \leq \frac{\binom{n}{2}}{\binom{t}{2}} \binom{t}{2} - \frac{1}{2^{\binom{t}{2}}}
$$

$$
\frac{log(S_t(n))}{log(2^{\binom{n}{2}})} \leq 1 - \frac{1}{\binom{t}{2} 2^{\binom{t}{2}}}
$$

□

4. Find the capacity C of the union of two channels $(\mathcal{X}_1, p_1(y_1|x_1), \mathcal{Y}_1)$ and $(\mathcal{X}_2, p_2(y_2|x_2), \mathcal{Y}_2)$, where at each time, one can send a symbol over channel 1 or channel 2 but not both. Assume that the output alphabets are disjoint. Express your answer as a function of the two channel capacities C_1 and C_2.

Since the output alphabets are disjoint, we can assume without loss of generality that the input alphabets are also disjoint (for any letter $x \in \mathcal{X}_1 \cap \mathcal{X}_2$, remove it an add a letter x_1 to $\mathcal{X}_1$ only and x_2 to $\mathcal{X}_2$ only). Now,

$$
P(x) = \begin{cases} \lambda P_1(x) & : x \in \mathcal{X}_1 \\ (1-\lambda) P_2(x) & : x \in \mathcal{X}_2 \end{cases}
$$

where $\lambda \in [0, 1]$ is the probability of choosing to send through channel 1 (and so $(1-\lambda)$ is the probability of sending through channel 2). To determine C, we need

$$
\begin{aligned}
H(Y) &= -\sum_{y \in \mathcal{Y}} P(y) \log(P(y)) \\
&= -\sum_{y \in \mathcal{Y}_1} P(y) \log(P(y)) - \sum_{y \in \mathcal{Y}_2} P(y) \log(P(y)) \\
&= -\sum_{y \in \mathcal{Y}_1} \frac{\lambda P(y)}{\lambda} \log(\frac{\lambda P(y)}{\lambda}) - \sum_{y \in \mathcal{Y}_2} \frac{(1-\lambda) P(y)}{1-\lambda} \log(\frac{(1-\lambda) P(y)}{1-\lambda}) \\
&= -\lambda \sum_{y \in \mathcal{Y}_1} P_1(y) \log(P_1(y)) - \lambda \sum_{y \in \mathcal{Y}_1} P_1(y) \log(\lambda) \\
&\quad -(1-\lambda) \sum_{y \in \mathcal{Y}_2} P_2(y) \log(P_2(y)) - (1-\lambda) \sum_{y \in \mathcal{Y}_2} P_2(y) \log(1-\lambda) \\
&= \lambda H(Y_1) + (1-\lambda) H(Y_2) + \lambda \log\left(\frac{1}{\lambda}\right) + (1-\lambda) \log\left(\frac{1}{1-\lambda}\right)
\end{aligned}
$$

and

$$\begin{aligned}
H(Y \mid X) &= -\sum_{x\in\mathcal{X}} P(x)\left(\sum_{y\in\mathcal{Y}} P(y \mid x) log(P(y \mid x))\right) \\
&= -\left(\sum_{x\in\mathcal{X}_1} P(x) + \sum_{x\in\mathcal{X}_2} P(x)\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x) log(P(y \mid x))\right) \\
&= -\left(\lambda \sum_{x\in\mathcal{X}_1} \frac{P(x)}{\lambda} + (1-\lambda)\sum_{x\in\mathcal{X}_2} \frac{P(x)}{1-\lambda}\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x) log(P(y \mid x))\right) \\
&= -\left(\lambda \sum_{x\in\mathcal{X}_1} \frac{P(x)}{\lambda}\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x) log(P(y \mid x))\right) \\
&\quad -\left((1-\lambda)\sum_{x\in\mathcal{X}_2} \frac{P(x)}{1-\lambda}\right)\left(\sum_{y\in\mathcal{Y}} P(y \mid x) log(P(y \mid x))\right) \\
&= -\left(\lambda \sum_{x\in\mathcal{X}_1} \frac{P(x)}{\lambda}\right)\left(\sum_{y\in\mathcal{Y}_1} P(y \mid x) log(P(y \mid x))\right) \\
&\quad -\left((1-\lambda)\sum_{x\in\mathcal{X}_2} \frac{P(x)}{1-\lambda}\right)\left(\sum_{y\in\mathcal{Y}_2} P(y \mid x) log(P(y \mid x))\right) \\
&= \lambda H(Y_1 \mid X_1) + (1-\lambda) H(Y_2 \mid X_2)
\end{aligned}$$

Now,

$$\begin{aligned}
C &= \max_{P(X)} I(X;Y) \\
&= \max_{P(X)} H(X) - H(Y \mid X) \\
&= \max_{\lambda} \lambda H(Y_1) + (1-\lambda) H(Y_2) + \lambda \log\left(\frac{1}{\lambda}\right) + (1-\lambda)\log\left(\frac{1}{1-\lambda}\right) \\
&\quad - (\lambda H(Y_1 \mid X_1) - (1-\lambda) H(Y_2 \mid X_2)) \\
&= \max_{\lambda} \lambda C_1 + (1-\lambda) C_2 + \lambda \log\left(\frac{1}{\lambda}\right) + (1-\lambda)\log\left(\frac{1}{1-\lambda}\right)
\end{aligned}$$

Using calculus, we determine that this expression is maximized when $\lambda = \frac{1}{2^{b-a}+1}$.