INDEPENDENT RANDOM VARIABLES, COVARIANCE, VARIANCE

CHEBYSHEV’S AND MARKOV’S INEQUALITIES

Let \((\Omega, P)\) be a probability space. Random variables \(X_1, \ldots, X_k : \Omega \to \mathbb{R}\) are independent if

\[
(\forall \alpha_1, \ldots, \alpha_k \in \mathbb{R}) \left( P \left( \bigwedge_{i=1}^k X_i = \alpha_i \right) = \prod_{i=1}^n P(X_i = \alpha_i) \right).
\]

Random variables satisfying Eq. (1) are sometimes called fully independent or mutually independent to emphasize the distinction of independence from pairwise independence.

The most important aggregate of a random variable \(X\) is the expected value, defined

\[
E(X) = \sum_{a \in \Omega} X(a) \cdot P(a).
\]

**DO 17.1.** If \(X, Y\) are independent then \(E(X \cdot Y) = E(X) \cdot E(Y)\).

**Definition 17.2.** The covariance of random variables \(X, Y\) is defined

\[
\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y).
\]

**DO 17.3.** If \(X, Y\) are independent then \(\text{Cov}(X, Y) = 0\).

**Definition 17.4.** Random variables \(X, Y\) are uncorrelated if \(\text{Cov}(X, Y) = 0\).

So if \(X, Y\) are independent then they are uncorrelated. The converse is false.
HW 17.5. (6 points) Find random variables $X,Y$ which are uncorrelated but not independent. Make $|\Omega|$ as small as possible. — First you have to state your sample space and the probability distribution. Then you need to define your random variables $X,Y$, calculate $E(X), E(Y),$ and $E(XY)$, show that $X,Y$ are uncorrelated. Finally you need to prove they are not independent.

DO 17.6. If $X_1, \ldots, X_k$ are independent then $E(\prod_i X_i) = \prod_i E(X_i)$.

How much does the random variable $X$ tend to deviate from its expected value? One measure of this is the mean deviation: $E(|X - E(X)|)$. For reasons of mathematical simplicity, we find it much easier to work with a related quantity, the variance.

**Definition 17.7.** The variance of a random variable $X$ is the quantity

$$\text{Var}(X) = E\left( (X - E(X))^2 \right).$$

DO 17.8. $\text{Var}(X) \geq 0$. Prove: The variance is zero if and only $X$ is almost constant, i.e., there is a number $r$ such that $P(X = r) = 1$.

DO 17.9. $\text{Var}(X) = E(X^2) - (E(X))^2$.

DO 17.10. Notice that $\text{Var}(X) = \text{Cov}(X, X)$.

**Corollary 17.11 (Cauchy–Schwarz inequality).** $E(X^2) \geq (E(X))^2$.

DO 17.12. Show that the inequality in the preceding problem is equivalent to the following form of the Cauchy–Schwarz inequality. For vectors $x, y \in \mathbb{R}^k$,

$$|x^T y| \leq \|x\| \cdot \|y\|.$$  \hspace{1cm} (4)

HW 17.13. (4 points) $\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_i \sum_j \text{Cov}(X_i, X_j)$.

**Corollary 17.14.** If $X_1, \ldots, X_k$ are pairwise independent, then

$$\text{Var} \left( \sum_i X_i \right) = \sum_i \text{Var}(X_i).$$

HW 17.15. (3+6 points) Let $G = ([n], E)$ be a 3-regular graph. Let $X_1, \ldots, X_n$ be independent unbiased Bernoulli trials (probability of success = $1/2$). For every edge $e = \{i, j\} \in E$ let $Y_e = X_i X_j$. Let $Z = \sum_{e \in E} Y_e$. Determine

(a) $E(Z)$

(b) $\text{Var}(Z)$.

You answers should be simple expressions in terms of $n$.

**Proposition 17.16 (Markov’s Inequality).** Let $X$ be a nonnegative random variable. Then for all $a > 0$ we have

$$P(X \geq a) \leq \frac{E(X)}{a}.$$  \hspace{1cm} (5)
Proof.

\[
E(X) = \sum_{t \in \mathbb{R}^+} t \cdot P(X = t) \geq \sum_{t \geq a} t \cdot P(X = t) \\
\geq a \cdot \sum_{t \geq a} P(X = t) = a \cdot P(X \geq a).
\]

The result now follows by substituting \( a \cdot E(X) \) in place of \( a \). □

**Definition 17.17.** The **standard deviation** of \( X \) is \( \sigma(X) = \sqrt{\text{Var}(X)} \).

**Proposition 17.18** (Chebychev’s Inequality). For any random variable \( X \) and any \( b > 0 \) we have

\[
P\left( |X - E(X)| \geq b \right) \leq \frac{\text{Var}(X)}{b^2} = \left( \frac{\sigma(X)}{b} \right)^2.
\]  

(6)

**Proof.** Let \( Y = (X - E(X))^2 \), so \( Y \geq 0 \). By Markov’s inequality,

\[
P(Y \geq b^2) \leq \frac{E(Y)}{b^2} = \frac{\text{Var}(X)}{b^2}.
\]

This is the simplest example of a **concentration inequality**: it says that that the value of \( X \) tends to be close to its expected value as long as the standard deviation is small.

---

**PROBABILLITY GENERATING FUNCTIONS AND REAL-ROOTED POLYNOMIALS**

Today we shall see further far-reaching consequences of real-rootedness.

**Definition 17.19.** The **generating function** of a sequence \( a = (a_0, a_1, \ldots) \) is the function \( f_a(t) = \sum_k a_k t^k \). If the sequence is finite then \( f_a \) is a polynomial.

**DO 17.20.** Prove that the generating function of the Fibonacci numbers has the following closed-form expression:

\[
\sum_{k=0}^{\infty} F_k t^k = \frac{t}{1 - t - t^2}.
\]  

(7)

Our main interest will be in the **matching generating function**

\[
m_G(t) = \sum_{k=1}^{\nu(G)} m_k(G) \cdot t^k,
\]  

(8)

where \( m_k(G) \) is the number of \( k \)-matchings of the graph \( G \). By the Heilmann–Lieb theorem, this polynomial is real-rooted. We shall indicate far-reaching implications of this fact on the distribution of the numbers \( m_k(G) \); under general conditions on the graph \( G \), these will be shown to be asymptotically normal.

**Definition 17.21.** Let \( X \) be a random variable. If \( \text{range}(X) \subseteq \mathbb{N} = \{0, 1, 2, \ldots\} \), then we shall say that \( X \) is a **counting variable**.
Remark 17.22. “Counting variable” is not a standard term. Even though the concept it describes is the most frequent type of discrete random variable, I could not find a commonly used term for it. The caveat is that if you use this term outside this class, you need to define it.

Definition 17.23. The probability generating function of a counting variable $X$ is defined by

$$f_X(t) = \sum_{k=0}^{\infty} P(X = k) \cdot t^k.$$ 

DO 17.24. If $f_X$ is a probability generating function then $f_X(1) = 1$.

HW 17.25. (5 points) If $X, Y$ are independent counting variables, then $f_{X+Y} = f_X \cdot f_Y$.

DO 17.26. More generally, if $X_1, \ldots, X_n$ are independent counting variables then

$$f_{\sum X_i} = \prod f_{X_i}.$$ 

The simplest examples of counting variables are Bernoulli trials; they take value 0 and 1 only.

DO 17.27. If $X$ is a Bernoulli trial with probability $p$ of success then

$$f_X = (1 - p) + pt.$$ (9)

Corollary 17.28. Let $X_i$ be independent Bernoulli random variables with probability of success $p_i$, and let $Y = \sum_{i=1}^{n} X_i$. Then the probability generating function of $Y$ is

$$f_Y = \prod_{i=1}^{n} ((1 - p_i) + p_i t)$$ (10)

DO 17.29. This polynomial is real-rooted; the roots are the negative numbers $-\alpha_i$ where

$$\alpha_i = \frac{1 - p_i}{p_i} = \frac{1}{p_i} - 1.$$ (11)

Let $g(t) = \prod_{i=1}^{n} (t + \alpha_i)$. Then

$$f_Y(t) = \frac{g(t)}{g(1)}.$$ (12)

We can read this observation backwards: instead of defining $Y$ as a sum of independent Bernoulli trials, we can decompose a given counting variable $Y$ into the sum of independent Bernoulli trials as long as $f_Y$ is real-rooted.

Corollary 17.30. If $Y$ is a counting variable and $f_Y$ is real-rooted, of degree $n$, then $Y$ is a sum of $n$ independent Bernoulli trials.
More precisely, the distribution of $Y$ is identical with the distribution of the sum of $n$ independent Bernoulli trials. Note that these Bernoulli trials will usually not be identically distributed.

**Proof.** Since $f_Y$ has non-negative coefficients, its roots must be negative; let us denote them $-\alpha_i$ where $\alpha_i > 0$. Let $g(t) = \prod_{i=1}^{n}(t + \alpha_i)$. Then $f_Y$ has the same roots (with the same multiplicities) as $g$ and therefore $g$ and $f_Y$ differ only in a scalar factor: $f_Y(T) = c \cdot g(t)$. Given that $f_Y(1) = 1$, it follows that $c = 1/g(1)$. Therefore

$$f_Y(t) = \frac{g(t)}{g(1)} = \prod_{i=1}^{n} \frac{t + \alpha_i}{1 + \alpha_i} = \prod_{i=1}^{n}(1 - p_i + p_it)$$

(13)

where

$$p_i = \frac{1}{1 + \alpha_i}.$$  

(14)

(Do: Verify the last equation!) Note that $0 < p_i < 1$ (since $\alpha_i > 0$), so we can view $p_i$ as the probability of success of a Bernoulli trial. But according to Cor. [17.28], the right-hand side of Eq. (13) is precisely the probability generating function of the sum of independent Bernoulli trials $X_i$ with probability $p_i$ of success. 

**Definition 17.31.** For a graph $G$, let $M_G$ be the set of matchings. This will be our sample space, with uniform distribution. Let $X_G : M_G \to \mathbb{R}$ be the random variable that counts the edges of the matching $x \in M_G$. We call $X_G$ the **matching counting variable** for $G$.

We now derive a powerful corollary of the Heilmann–Lieb theorem, the reality of the roots of the matching generating function.

**Corollary 17.32.** The matching counting variable is a sum of independent Bernoulli trials.

**Proof.** Let $m_G(t)$ be the matching generating function defined by Eq.[8], and let $X_G$ be the matching counting variable. Now $|M_G| = \sum_k m_k(G) = m_G(1)$, so the probability generating function of $X_G$ is

$$f_{X_G}(t) = \frac{m_G(t)}{m_G(1)}.$$  

(15)

(Do: Verify!) So this polynomial is real-rooted, and the result follows from Cor. [17.30].

---

**Central Limit Theorems**

**Definition 17.33.** Let $Y$ be a random variable with $\sigma(Y) > 0$ (so $Y$ is not almost constant). Then the centered version of $Y$ is the variable $U = Y - E(Y)$ and the normalized version of $Y$ is the variable

$$Z = \frac{Y - E(Y)}{\sigma(Y)}.$$  

(16)

**Do 17.34.** (a) $E(U) = 0$ and $\sigma(U) = \sigma(Y)$

(b) $E(Z) = 0$ and $\sigma(Z) = 1$
(c) Show that among all random variables of the form $aY + b$ ($a, b \in \mathbb{R}$), the normalized version $Z$ is the only one that satisfies (b).

HW+ 17.35. (6 points) Let $X$ be a Bernoulli trial with success probability $p$ ($0 < p < 1$) and let $U$ be the centered version of $X$. Prove: $E(|U|^3) < \text{Var}(X)$.

DO 17.36. Let $X$ be a Bernoulli trial with success probability $p$ ($0 < p < 1$) and let $Z$ be the normalized version of $X$. Determine $Z$.

Solution. We have $E(X) = p$ and $\text{Var}(X) = p(1-p)$, so

$$Z = \frac{X - p}{\sqrt{p(1-p)}} = \begin{cases} \sqrt{\frac{1-p}{p}} & \text{with probability } p \\ \sqrt{\frac{p}{1-p}} & \text{with probability } 1 - p \end{cases} \quad (17)$$

Definition 17.37. The cumulative distribution function (CDF) of a random variable $X$ is the function $F_X(t) = P(X \leq t)$.

Definition 17.38. The standard normal distribution is the distribution defined by the CDF

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \quad (18)$$

Remark 17.39. The “density” of this distribution is the function $\Phi'(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$, the standard bell curve.

Definition 17.40. Let $\{X_n\}$ be an infinite sequence of random variables, each defined on its own separate probability space. Assume the $X_n$ are not almost constant; let $Z_n$ denote the normalized version of $X_n$. Let $F_n$ denote the CDF of $Z_n$. We say that the sequence $\{X_n\}$ is asymptotically normal if $F_n$ approaches $\Phi$ uniformly, i.e., the distance $\sup_x |F(x) - \Phi(x)|$ approaches zero as $n \to \infty$.

Remark 17.41. The classical Central Limit Theorem (De Moivre–Laplace) says that for the binomial distributions are asymptotically normal in the following sense. Fix $0 < p < 1$ and let $X_1, X_2, \ldots, X_n$ be independent Bernoulli trials with success probability $p$. Let $Y_n = \sum_{i=1}^{n} X_i$. Then the infinite sequence $\{Y_n\}$ is asymptotically normal.

The following result is one of many generalizations of the classical Central Limit Theorem.

Theorem 17.42 (Andrew Berry, Carl-Gustav Esséen, 1941/42). Let $(X_i)_{i=1}^{n}$ be independent random variables with respective standard deviation $\sigma(X_i) = \sigma_i$, and let $\rho_i$ denote the third moment of $U_i = X_i - E(X_i)$, defined by $\rho_i = E(|U_i|^3)$. Let $Y = \sum_{i=1}^{n} X_i$ and let $Z$ be the normalized version of $Y$. Let $\sigma = \sigma(Y) = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$. Then

$$(\forall x \in \mathbb{R}) \left( |F_Z(x) - \Phi(x)| < \sigma^{-3} \cdot \sum_{i} \rho_i \right). \quad (19)$$
The Berry–Esséen Theorem strengthens the classical result in several directions.

(a) It does not require the $X_i$ to be Bernoulli trials
(b) It does not require the $X_i$ to be identically distributed
(c) It gives a specific rate of convergence.

For our purposes, (a) will be irrelevant (our variables will be Bernoulli trials), but (b) and (c) are crucial.

Corollary 17.43. Let $(X_i)_{i=1}^n$ be independent Bernoulli trials with respective success probability $p_i$. Let $Y = \sum_{i=1}^n X_i$ and let $Z$ be the normalized version of $Y$. Let $\sigma := \sigma(Y) = \sqrt{\sum_{i=1}^n p_i(1 - p_i)}$. Then

$$
(\forall x \in \mathbb{R}) \left( |F_Z(x) - \Phi(x)| < 1/\sigma \right). 
$$

Proof. By exercise HW[17.35] we have $\rho_i < \sigma_i^2$. So the right-hand side of Eq. (19) is

$$
\frac{\sum_i \rho_i}{\sigma^3} < \frac{\sum_i \sigma_i^2}{\sigma^3} = \frac{1}{\sigma}.
$$

The following is an immediate corollary. (DO: Why?)

Corollary 17.44. For $n = 1, 2, \ldots$ let $Y_n$ be a sum of independent Bernoulli trials. If $\lim_{n \to \infty} \sigma_n = \infty$ then the sequence $Y_1, Y_2, \ldots$ is asymptotically normal.

Remark 17.45. Note that the Bernoulli trials of which $Y_n$ is the sum must be independent but they do not need to be identically distributed.

Remark 17.46. A conceptual clarification. The probability space for $Y_n$ is the same as the probability space for the Bernoulli trials of which $Y_n$ is the sum. However, the probability spaces associated with $Y_n$ for distinct values of $n$ are unrelated.

ASYMPTOTIC NORMALITY OF RANDOM MATCHINGS

Recall the matching generator function

$$
m_G(t) = \sum_{k=0}^{\nu(G)} m_k t^k,
$$

where $m_k$ is the number of $k$-matchings.

The result that $\mu_G$ is real-rooted implies that $m_G$ is also real-rooted. By Newton’s inequalities, it follows that the sequence $m_k, k = 1, \ldots, \nu(G)$ is log-concave, and hence they are unimodal. We shall see that much more can be said about the behavior of this sequence. Let us first consider a simple example.

Example 17.47. Let $G$ be the graph consisting of $n/2$ disjoint edges: $G = \frac{n}{2} \cdot K_2$. Then $m_k = \binom{n/2}{k}$.
So the sequence of these sequences is asymptotically normal by De Moivre–Laplace. A far-reaching generalization of this fact was observed by Chris Godsil in 1981. For a graph $G$ let $X_G$ denote the matching counting variable defined in Def. 17.31 so $X_G$ is the size of a random matching of $G$. Let $\sigma(G) := \sigma(X_G)$. Let us refer to the sequence $(m_0(G), m_1(G), \ldots, m_{\nu(G)}(G))$ as the matching sequence of $G$.

**Theorem 17.48** (Godsil, 1981). Let $G_n$ be an infinite sequence of graphs. If $\sigma(G_n) \to \infty$ then the matching sequences of the $G_n$ are asymptotically normal.

DO 17.49. Show that this result follows by combining Cor. 17.32 (a consequence of the Heilmann–Lieb theorem) and Cor. 17.44 (a consequence of the Berry–Esséen theorem).

DO 17.50. Express $\sigma(G)$ in terms of the $m_k(G)$.

Godsil also gave rather general sufficient conditions that guarantee $\sigma(G_n) \to \infty$. One of these is the following.

**Theorem 17.51** (Godsil). Let $G_n$ be an infinite sequence of graphs without isolated vertices (i.e., $\deg_{\text{min}} \geq 1$). If $\deg_{\text{max}}(G_n)/|V(G_n)| \to 0$ then $\sigma(G_n) \to \infty$ and therefore the matching sequences of the $G_n$ are asymptotically normal.

Even though the conditions of this theorem do not hold for the complete graphs, Godsil showed that the complete graphs also satisfy the conclusion; in particular, the coefficients of the Hermite polynomials are asymptotically normal.

Recall that $X_G$ is the size of a random matching (picked uniformly from $M_G$). Godsil comments that a necessary condition for $\sigma(G_n) \to \infty$ is $E(X_{G_n}) \to \infty$. Another condition that is obviously necessary is that $\nu(G_n) \to \infty$. He then comments that “interestingly enough, the second of these conditions implies the first” and cites an observation by this instructor that proves this.

**Lemma 17.52** (Babai). The expected size of a random matching of $G$ is at least $\nu(G)/3$, i.e., $E(X_G) \geq \nu(G)/3$.

**CH 17.53.** Find a simple proof of this statement.

**BONUS+ 17.54.** (6 points) Find an infinite sequence $\{G_n\}$ of graphs such that $\nu(G_n) \to \infty$ but $\sigma(G_n) \to 0$.

Jeff Kahn (1998) significantly expanded Godsil’s study of the asymptotic behavior of the matching sequence, including cases when $\sigma(G)$ is bounded and the matching sequence is asymptotically Poisson.

The presentation so far in this class was based on the following paper.


---

**RAMSEY THEORY**

Recall the Erdős–Szekeres theorem:

\[
\binom{k + \ell}{k} \to (k + 1, \ell + 1). \tag{22}
\]
The diagonal case \((k = \ell)\) gives
\[
4^k > \binom{2k}{k} \rightarrow (k + 1)_2.
\]

Substituting \(n = 4^k\), i.e., \(k = \frac{1}{2} \log_2 n\), we obtain
\[
n \rightarrow \left(1 + \frac{1}{2} \log_2 n\right)_2. \tag{23}
\]

The question is, how far the \((1/2) \log_2 n\) from the best possible bound. Paul Erdős showed in 1949 that the order of magnitude is correct.

**Theorem 17.55 (Erdős).** \(n \rightarrow (1 + 2 \log_2 n)_2\).

**DO 17.56.** Show that Theorem 17.55 is equivalent to the following statement.

- For all \(n\) there exists a graph \(G_n\) with \(n\) vertices such that
  \[
  \max(\alpha(G_n), \alpha(\overline{G}_n)) < 1 + 2 \log_2 n. \tag{24}
  \]

Erdős’s proof was an early triumph of his “probabilistic method.” Rather than constructing such graphs, he showed that almost all graphs have the required property. What this means is that the probability that a random graph on \(n\) vertices satisfies Eq. (24) approaches 1 as \(n \rightarrow \infty\).

**Theorem 17.57 (Erdős).** Let \(G_n\) be a graph on \(n\) vertices chosen uniformly at random. Then
\[
P\left( \max(\alpha(G_n), \alpha(\overline{G}_n)) \geq 1 + 2 \log_2 n \right) \rightarrow 0 \text{ as } n \rightarrow \infty.
\]

**Proof.** Fix a set \(V\) of \(n\) vertices. Our sample space \(\Omega\) will be the set of all graphs with \(V\) as their set of vertices; the number of such graphs is \(|\Omega| = 2^{\binom{n}{2}}\). Our probability distribution will be uniform over \(\Omega\). In other words, we pick our graph \(G\) “uniformly at random” from \(\Omega\).

Let \(A \subseteq V\) be a set of size \(|A| = k\). Then
\[
P(A \text{ is independent}) = \frac{1}{2^{\binom{k}{2}}}. \tag{25}
\]

Using the union bound,
\[
P(\alpha(G) \geq k) \leq \sum_{A \subseteq V: |A| = k} P(A \text{ is independent}) = \binom{n}{k} \frac{1}{2^{\binom{k}{2}}}. \tag{26}
\]

Combining this result with **DO 17.59** we have
\[
P(\alpha(G) \geq k) \leq \frac{1}{k!} \cdot \frac{n^k}{2^{k(k-1)/2}} = \frac{1}{k!} \left(\frac{n}{2^{(k-1)/2}}\right)^k.
\]
If \( k \geq 1 + 2 \log_2 n \), then \( 2^{(k-1)/2} \geq n \), so we have
\[
P(\alpha(G) \geq k) \leq \frac{1}{k!}.
\]
The same statements apply to \( \overline{G} \). Making another appeal to the union bound, we obtain that if \( k \geq 1 + 2 \log_2 n \) then
\[
P(\alpha(G) \geq k \text{ or } \alpha(\overline{G}) \geq k) \leq \frac{2}{k!} \to 0 \text{ as } n \to \infty.
\]
Thus, for almost all graphs \( (P \to 1) \), \( \max(\alpha(G), \alpha(\overline{G})) < 1 + 2 \log_2 n \).

**DO 17.58.** Let us choose the graph \( G \) uniformly at random among the \( 2\binom{n}{2} \) graphs on a given set of \( n \) vertices. For each pair \( \{i, j\} \) of vertices \( i < j \) let \( X(i, j) \) denote the Bernoulli trial that is successful if \( \{i, j\} \in E(G) \). Prove that the probability of success is \( 1/2 \) and that these \( \binom{n}{2} \) Bernoulli trials are independent.

**DO 17.59.** \( \binom{n}{k} \leq \frac{n^k}{k!} \).

**BONUS+ 17.60.** (6 points) Prove: almost all graphs have diameter 2. Define what this statement means.

**BONUS+ 17.61.** (7 points) Prove: if \( G \) does not contain \( K_5 \) then \( \chi(G) = O(n^{3/4}) \). Hint. Erdős–Szekeres.

**BONUS+ 17.62** (Explicit Ramsey graph by Zsigmond Nagy, 1973). (7 points) Give a constructive proof of the relation \( \binom{k}{3} \to (k + 1)_2 \) using the following graph. Let \( V = \binom{[k]}{3} \) (the set of 3-subsets of \([k]\)) be the set of vertices. Make \( A, B \in V \) adjacent if \( |A \cap B| = 1 \).

**Remark 17.63.** Note that this shows \( n \to (cn^{1/3})_2 \) for some constant \( c \). Compare this with the construction by Abbott, previously assigned as a Challenge problem, that showed \( n \to cn^\alpha \) where \( \alpha = \log_5 2 = 0.43 \ldots \) — Nagy’s construction was later generalized by Peter Frankl and Richard M. Wilson (1980) to a constructive proof that \( n \to n^\varepsilon \) for any constant \( \varepsilon > 0 \).

**BONUS+ 17.64.** (8 points) A \((0, 1)\)-matrix is a matrix whose entries are \( a_{ij} \in \{0, 1\} \). A \( k \times \ell \) submatrix is obtained by selecting an arbitrary set of \( k \) rows and \( \ell \) columns and looking at the \( k\ell \) cells at their intersection. A submatrix is homogeneous if all of its entries are the same.

Prove: every \( n \times n \) \((0, 1)\)-matrix has a \( k \times k \) homogeneous submatrix where \( k \sim (1/2) \log_2 n \). (In an earlier version of this posting, I erroneously claimed \( k = \lceil(1/2) \log_2 n \rceil \). In fact, \( k \) will be slightly smaller but still asymptotically equal to \( (1/2) \log_2 n \).)

**HW+ 17.65.** (8 points) Prove that for all sufficiently large \( n \) there exists an \( n \times n \) \((0, 1)\)-matrix that has no homogeneous \( k \times k \) submatrix for \( k = \lceil 2 \log_2 n \rceil \). Hint. Use Erdős’s probabilistic method. State the size of the sample space you use. (The previous posting of this problem erroneously omitted the factor 2 before \( \log_2 n \). The due date for this problem is extended to Thursday, June 6, before class.)

**CH 17.66.** (15 points) For infinitely many values of \( n \), give a constructive proof that there exists an \( n \times n \) \((0, 1)\)-matrix that has no homogeneous \( k \times k \) submatrix for \( k > \sqrt{n} \).