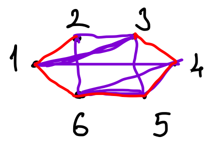
Dogacan Sertbaş

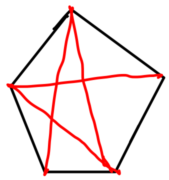
Probabilistic Number Theory

§0 Prologue (Ramsey Theory)

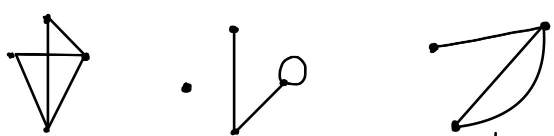
Example. Numbers are person, edges are relationship.



Example.



Definition. A <u>simple graph</u> $G=(V,E)$ is a tuple, where V denotes the set of vertices, and E is the set of edges which consists unordered pairs $\{x,y\}$ such that $x,y \in V$
$(E \subseteq \mathcal{P}_2(V) = \{\{x,y\} : x,y \in V\})$



simple graph   not   not

Definition. A complete graph K_n is a simple graph on n vertices $(|V|=n)$ such that $\forall x,y \in V$ $\{x,y\} \in E$.

Definition. A two-coloring of the edges K_n is a map $\chi : E(K_n) \longrightarrow \{\text{Black}, \text{White}\}$ which assigns either Black or White to each edge.

? Let $n \geq 1$ and $j,k \in \{1,\dots,n\}$. For any two-coloring χ of K_n, does K_n contain either a white monochromatic K_j or a black monochromatic K_k?

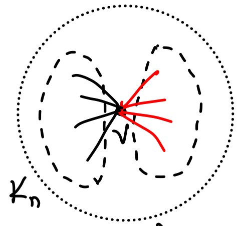
$\hookrightarrow$ $R(j,k) := \min \{ n \in \mathbb{N} \mid K_n$ contains either a white K_j or a black K_k for any two coloring of $K_n \} = \min S$

If S is nonempty, then Ramsey Theorem $\forall j,k \in \mathbb{N}_{\geq 1}$ $R(j,k) < \infty$ and
$$R(j,k) \leq \binom{j+k-2}{j-1}$$

Proof. We will prove it by induction on $j+k$. Note that
$$R(j,1) = R(1,k) = 1$$
$$1 = R(j,1) \leq \binom{j-1}{j-1} \quad \text{or} \quad 1 = R(1,k) = \binom{k-1}{0}$$

Now suppose that the statement is true for $j+k-1$. Let $n = R(j-1,k) + R(j,k-1)$ and χ be any two-coloring of K_n. Take any vertex v of K_n.

Define



$$\mathcal{W} = \{\{w, v\} : \chi(\{w, v\}) = \text{White}\}$$
$$\mathcal{B} = \{\{b, v\} : \chi(\{b, v\}) = \text{Black}\}$$

Observe that $|\mathcal{W}| + |\mathcal{B}| = n-1 = R(j-1, k) + R(j, k-1) - 1$, then $|\mathcal{W}| \geq R(j-1, k)$ $(*)$ or $|\mathcal{B}| \geq R(j, k-1)$ $(+)$

(Suppose not, $|\mathcal{W}| < R(j-1, k) \iff |\mathcal{W}| \leq R(j-1, k) - 1$
$|\mathcal{B}| < R(j, k-1) \iff |\mathcal{B}| \leq R(j, k-1) - 1$)

It gives $|\mathcal{W}| + |\mathcal{B}| \leq R(j-1, k) + R(j, k-1) - 2 = n-2$.

Case 1. $*$ holds, there there is a white monochromatic K_{j-1} lies in $\mathcal{W}$ or there is a black monochromatic K_k lies in $\mathcal{W}$.

If the second case holds, then a black K_k lies in K_n. Otherwise, since all the elements of $\mathcal{W}$ are connected with white edges to , this means that there is a monochromatic white K_j lies K_n (Similar situation also occurs when $(+)$ holds)

Thus
$$R(j,k) \leq R(j-1, k) + R(j, k-1) \leq \binom{j-1+k-2}{k-1} + \binom{j+k-3}{k-2} = \binom{j+k-2}{k-1} \qquad \square$$

$R(2,2) = 2$ $\qquad R(4,4) = 18$ $\qquad 102 \leq R(6,6) \leq 165$

$R(3,3) = 6$ $\qquad 43 \leq R(5,5) \leq 49$

Remark. For any $k \geq 1$, we have $R(k,k) \leq \binom{k+k-2}{k-1} = \binom{2k-2}{k-1} \leq \sum_{i=0}^{2k-2} \binom{2k-2}{i} 1^i \cdot 1^{2k-2-i} = 4^{k-1}$

? What can we say about lower case bound for $R(k,k)$ depending on k?

§1 Basic Probability Theory.

1.1 Probability Spaces.

Definition. Let X be non-empty set and $\mathcal{A}$ be a collection of subsets of X ($\mathcal{A} \subseteq \mathcal{P}(X)$). $\mathcal{A}$ is called σ-algebra if

i $\emptyset \in \mathcal{A}$ $\qquad$ ex. $\mathcal{A} = \mathcal{P}(X)$, $\mathcal{A} = \{\emptyset, X\}$

ii $\forall A \in \mathcal{A}$ $\quad A^c \in \mathcal{A}$

iii. $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Definition. A probability space $(\Omega, \mathcal{F}, P_r)$ is a tuple where Ω is nonempty set which stands for the set of possible outcomes $\mathcal{F}$ is a σ-algebra on Ω where $\mathcal{F}$ contains the all possible events and $P_r : \mathcal{F} \to [0,1]$ is a function such that

i. $P_r(\Omega) = 1$

ii. For any sequence $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ of disjoint set, we have $P_r\left(\dot{\bigcup}_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P_r(E_i)$ (countable additivity)

Example. (tossing a coin) $\Omega = \{H, T\}$, $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset, \{H\}, \{T\}, \{H,T\}\}$. If we flip a fair coin, then $P_r(\{H\}) = P_r(\{T\}) = 1/2$. $P_r : \mathcal{F} \to [0,1]$ where $\emptyset \mapsto 0$, $\{H\} \mapsto 1/2$, $\{T\} \mapsto 1/2$, $\Omega \mapsto 1$.

Example. (throw 2 dice) $\Omega = \{(i,j) : 1 \le i,j \le 6\}$. If $E \in \mathcal{F}$ represents the event that of the faces of two dice is 7, then

$$E = \{(1,6),(2,5),(3,4),\dots,(6,1)\}$$

If we deal fair dice, then $P_r(E) = \frac{|E|}{36} = \frac{6}{36} = 1/6$

Properties

P1. $\forall E \in \mathcal{F}$, $P_r(E^c) = 1 - P_r(E)$.

P2. $P(\emptyset) = 0$.

P3. $\forall E, F \in \mathcal{F}$, $E \subseteq F$, then $P_r(E) \le P_r(F)$ $\left(F = (F \setminus E) \dot{\cup} E\right)$

P4. (Inclusion Exclusion Property).

$$P_r(E \cup F) = P_r(E) + P_r(F) - P_r(E \cap F) \quad \left(E \cup F = (E \setminus F) \dot{\cup} (F \cap E) \cup (F \setminus E)\right)$$

P5. (Boole's Inequality) If $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$, then $P_r\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} P_r(A_i)$.

Proof. Define $B_1 := A_1$ and $B_k := A_k \setminus \bigcup_{i=1}^{k-1} A_i$. Observe that $\forall k \in \mathbb{N}$, $B_k \subseteq A_k$. Also $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$. Note that all B_i's are disjoint. Hence,

$$P_r\left(\bigcup_{i=1}^{\infty} A_i\right) = P_r\left(\dot{\bigcup}_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) \le \sum_{i=1}^{\infty} P_r(A_i)$$

$\square$

Corollary 1. Let $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ be a sequence of events. If $\sum_{i=1}^{\infty} P_r(A_i) < 1$, then $P_r\left(\bigcap_{i=1}^{\infty} A_i^c\right) > 0$

In other words, there is a an event where none of A_i's occur.

Proof (sketch).

$$P_r\left(\bigcap_{i=1}^{\infty} A_i^c\right) = P_r\left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c\right) = 1 - P_r\left(\bigcup_{i=1}^{\infty} A_i\right) \ge 1 - \sum_{i=1}^{\infty} P_r(A_i) > 0.$$

Probabistic Method.

Main Idea. To show the occurance of an event $\mathcal{D}$, prove that $\Pr(\mathcal{D})>0$

↳ (Diagonal) Ramsey Numbers.

Theorem 1. For any $k\in\mathbb{N}_{\geq 2}$, let $r(k)=\max\left\{n\in\mathbb{N} \mid \binom{n}{k}2^{1-\binom{k}{2}}<1\right\}$. Then $R(k,k)>r(k)$. In particular, $R(k,k)>\frac{k\cdot 2^{k/2}}{2e}$

Proof. (Erdős, '47) Remember that Boole's Sieve:

$$\sum_{i\geq 1}\Pr(A_i)<1 \quad \text{implies that} \quad \Pr\left(\bigcap_{i\geq 1}A_i^c\right)>0.$$

For a given $n\leq r(k)$, let $(\Omega,\mathcal{F},\Pr)$ be a probability space Ω contains all possible two colorings of K_n. Assume that for any coloring $X\in\Omega$, we have

$$\Pr(\{X\})=1/2^{\binom{n}{2}}.$$

Remember that $\#E(K_n)=\binom{n}{2}$. Let S be any fixed set of vertices in K_n, where $|S|=k$. If A_S denotes the event that S constitutes a monochoromatic K_k, then

$$\Pr(A_S)=\underbrace{\prod_{i\in E(S)}\frac{1}{2}}_{\text{being white}}+\underbrace{\prod_{i\in E(S)}\frac{1}{2}}_{\text{being black}}=2^{1-\binom{k}{2}}$$

Note that

$$\Pr\Big(\bigcup_{\substack{S\\|S|=k}}A_S\Big)\leq\sum_{\substack{S\subseteq V(K_n)\\|S|=k}}\Pr(A_S)=\sum_{\substack{S\subseteq V(K_n)\\|S|=k}}2^{1-\binom{k}{2}}=2^{1-\binom{k}{2}}\sum_{\substack{S\subseteq V(K_n)\\|S|=k}}1=\binom{n}{k}2^{1-\binom{k}{2}}$$

Since $n\leq r(k)$, we have $\binom{n}{k}2^{1-\binom{k}{2}}<1$. By previous Corollary, there exists a two coloring that none of K's are monochromatic. Hence $R(k,k)>r(k)$ by definition.

To find a precise lower bound observe that

$$\binom{n}{k}=\frac{n\cdot(n-1)\cdots(n-k+1)}{k!}<\frac{n^k}{k!}$$

Also, $\log(k!)=\log\left(\prod_{j=1}^{k}j\right)=\sum_{j=1}^{k}\log j>\int_1^k\log t\,dt=t\log t-t\Big|_1^k=k\log k-(k-1)$

$>k\log k-k$

This implies $k!=e^{\log k!}=\left(\frac{k}{e}\right)^k$. Hence $\binom{n}{k}2^{1-\binom{k}{2}}<\frac{n^ke^k}{k^k}\cdot\frac{2}{2^{\frac{k(k-1)}{2}}}$. So if $\frac{n^ke^k}{k^k}\cdot\frac{2}{2^{\frac{k(k-1)}{2}}}\leq 1$, or equivalently $n^k\leq\frac{2^{\frac{k(k-1)}{2}}\cdot k^k}{2e^k}$. We see that there is a two-coloring of K, which does not contain any monochromatic K_k. As $k\geq 2$

$$\frac{k\cdot 2^k}{\frac{2e}{k\cdot e^{k/2}\,2e}}\leq\frac{k\cdot 2^{k/2}}{2^{1/k}e\sqrt{2}}.$$

Thus, if $n\leq\frac{k\cdot e^{k/2}}{2e}$, then $n\leq\frac{k\cdot 2^{\frac{k-1}{2}}}{2^{1/k}e}$. Hence, $n^k\leq\frac{k^k\cdot 2^{\frac{k(k-1)}{2}}}{2e^k}$ $\left(\binom{n}{k}2^{1-\binom{k}{2}}<1\right)$

Then,

ITP: Leon, Coq
Isabelle

$$R(k,k) > \frac{k \cdot 2^{k/2}}{2e}$$

$\square$

Best known Bounds for $R(k,k)$

(Stirling's formula : $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$)

$$(1+o(1))\frac{k\sqrt{2}}{e}2^{k/2} \le R(k,k)$$

where $o(1)$ denotes a function $f(k)$ such that $\lim_{k\to\infty} f(k) = 0$. Recently, $\exists \varepsilon > 0$ such that

$$R(k,k) < (4-\varepsilon)^k$$

(2022 2023 by I.T.P.

Remark. Note that there are $2^{\binom{n}{2}}$ many possible two-colorings of K_n. To obtain a monochromatic K_k in K_n, we need to count the number of possible colorings which contains a monochromatic K_k,

$$\underbrace{\binom{n}{k}}_{\text{choosing k-vertices}} \cdot \underbrace{2}_{\text{choosing color}} \cdot \underbrace{2^{\binom{n}{2}-\binom{k}{2}}}_{\text{color the rest}}$$

If $\binom{n}{k} 2\, 2^{\binom{n}{2}-\binom{k}{2}} < 2^{\binom{n}{2}}$, then there exists a two-coloring of K_n where it doesn't contain any monochromatic K_k. Hence, if

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

then there is a coloring where K_n does not contain any monochromatic K_k.

1.2 Conditional Probability and Independence

Definition. Let $(\Omega, \mathcal{F}, P_r)$ be a probability space, and $A, B \in \mathcal{F}$ two events such that $P_r(B) > 0$. We define

$$P_r(A|B) = \frac{P_r(A\cap B)}{P_r(B)}$$

as the probability of A given that B has occured

[H.W]. Remark.

1. Notice that for any $H \in \mathcal{F}$ w/ $P_r(H) > 0$, the function

$$P_{r_H} : \mathcal{F} \to [0,1] \text{ defined by } A \longmapsto P_{r_H}(A|H)$$

is also a probability measure on $(\Omega, \mathcal{F}, P_{r_H})$

2. $\forall \in \mathcal{F}$, $P_r(A^c|H) = 1 - P_r(A|H)$

The Chain Rule for the Conditional Probabilities

For any $n \geq 2$, let $A_1, \ldots, A_n$ be the events whose intersection has strictly positive probability. Then

$$P_r(A_1 \cap \cdots \cap A_n) = P_r(A_n \mid A_{n-1} \cap \cdots \cap A_1)$$
$$\cdot P_r(A_{n-1} \mid A_{n-2} \cap \cdots \cap A_1)$$
$$\cdot P_r(A_2 \mid A_1) \cdot P_r(A_1)$$

$$= \prod_{i=1}^{n} P_r\left(A_i \mid \bigcap_{j=1}^{i} A_j\right)$$

Proof. Note that $A_1 \cap \cdots \cap A_n \subseteq A_1 \cap \cdots \cap A_{n-1} \subseteq \cdots \subseteq A_1$. Hence

$$0 < P_r\left(\bigcap_{i=1}^{n} A_i\right) \leq P_r\left(\bigcap_{i=1}^{n-1} A_i\right) \leq \cdots \leq P_r(A_1)$$

By induction, the rule follows. [H.W]

The Total Probability Law.

Let $(\Omega, \mathcal{F}, P_r)$ be a probability space and $\{A_i\}_{i=1}^{k} \subseteq \mathcal{F}$ a partition of Ω where $P(A_i) > 0$ for each $i \in \mathbb{N}$ (and possibly $k = \infty$) (i.e. $\Omega = \dot{\bigcup} A_i$). Then $\forall H \in \mathcal{F}$, $P_r(H) = \sum_{i=1}^{k} P_r(A \cap H) = \sum_{i=1}^{k} P_r(H \mid A_i) \cdot P_r(A_i)$

Definition. Let $A, B \in \mathcal{F}$ be two events in the probability space $(\Omega, \mathcal{F}, P_r)$. These events are called indipendent if

$$P_r(A \cap B) = P_r(A)\, P_r(B)$$

Remark.

1) If A and B are independent w/ $P_r(B) > 0$, then $P_r(A \mid B) = P_r(A)$
2) If A and B are independent, A and B^c are independent

Definition. Let $\{A_i\}_{i=1}^{n} \subseteq \mathcal{F}$ be events. We say that $A_1, \ldots, A_n$ are pairwise independent if $\forall i, j \in \{1, \ldots, n\}$ $i \neq j$ implies that A_i and A_j are independent. Also $A_1, \ldots, A_n$ are mutually independent. If for any subset $K \subseteq \{1, \ldots, n\}$

$$P_r\left(\bigcap_{k \in K} A_k\right) = \prod_{k \in K} P_r(A_k)$$

Moreover, we say that A is independent of the events $\{A_j \mid j \in I\}$ for some $I \subseteq \{1, \ldots, n\}$ if $\forall K \subseteq J$

$$P_r\left(A_0 \cap \bigcap_{k \in K} A_k\right) = P_r(A_0) \cdot P_r\left(\bigcap_{k \in K} A_k\right)$$

Remark. If A, B, C are independent events, then C^c are pairwise independent with A, B and $A\cap B$, so A^c, B^c, C^c are independent. In general, $A_1, \dots, A_n$ are independent implies that $A_1^c, \dots, A_n^c$ are independent.

Proposition If $A_1, \dots, A_n$ are independent and $Pr(A_i) < 1$ for each $i \in \{1, \dots, n\}$ then $Pr\left(\bigcap_{i=1}^n A_i^c\right) > 0$.

Proof. If $A_1, \dots, A_n$ are independent, then so $A_1^c, \dots, A_n^c$. Thus

$$Pr\left(\bigcap_{i=1}^n A_i^c\right) = \prod_{i=1}^n Pr(A_i^c) = \prod_{i=1}^n \left(1 - Pr(A_i)\right) > 0 \quad \text{as } Pr(A_i) \in [0,1].$$

Lovasz's Local Lemma. (The Theory of Probobity – S.S. Venkatesh)

$A_1, \dots, A_n$ events where each A_i's mutually independent except at most d of $\{A_1, \dots, A_n\} \setminus \{A_i\}$

$$\forall i \in \{1, \dots, n\} \quad Pr(A_i) \leq p. \quad 4pd \leq 1 \implies Pr\left(\bigcap_{i=1}^n A_i^c\right) > 0.$$

§ Van der Warden's Theorem

For any given positive integers $r, k \geq 1$ there is a natural $N = N(r,k)$ s.t. if we color the set of integers $\{1, \dots, N\}$ with r colors, there is an arithmetic progression of length k whose elements are of the same color.

$$\left(\chi : \{1, \dots, N\} \longrightarrow \{1, \dots, r\} \quad a \longmapsto \chi(a) \right)$$

We define the van der Warden number $W(r,k)$ as the smallest N that satisfies vdW theorem

Theorem (first Lower Bond) For any $n \geq 2$ and $k \geq 2$ we have $W(r,k) \geq \sqrt{2}\, r^{\frac{k-1}{2}}$

Proof. Let $n < \sqrt{2} \cdot r^{\frac{k-1}{2}}$ and $(\Omega, \mathcal{F}, Pr)$ be a probability space where Ω consists of all possible colorings of $\{1, \dots, n\}$. Assume that any colorings occurs equally likely, i.e.,

$$\forall x \in \Omega \quad Pr(\{x\}) = \left(\frac{1}{r}\right)^n$$

as coloring each number is independent of coloring the others.

Let S be any k-term AP in $\{1, \dots, n\}$. Denote A_S the event that S is monochromatic. Then

$$Pr(A_S) = r^{1-k}$$

Also observe that the structure of an AP is determined by its first two terms

(Since if we know a and q, then we obtain a, q) Therefore, there are at most $\binom{n}{2}$ many possible choices for a and a+q. This means that the number of k-AP in $\{1, \dots, n\}$ is bounded by $\binom{n}{2}$.

$$\sum_{\substack{S \text{ is} \\ \text{a k-AP}}} Pr(A_S) = \sum_S r^{1-k} = \binom{n}{2} r^{1-k} < 1.$$

This implies that there is a coloring that $\{1, \dots, n\}$ does not contain any number k-AP

$$\frac{n \cdot (n-1)}{2} r^{k-1} < \frac{n^2}{2} r^{k-1} < 1 \iff n^2 \leq 2 r^{k-1}$$

Since $n < \sqrt{2}\, r^{\frac{k-1}{2}}$, $W(r,k) \not\leq n$. Hence $W(r,k) \geq \sqrt{2}\, r^{\frac{k-1}{2}}$. □

r \ k	1	2	3	...	7
2	1	3	9	Known	
3	1	4	27		
4	1	5	76		
.					
.					

→ Not Known

<u>Best Known Bounds</u> $r, k \geq 2$ $\frac{r^k}{ek}(1 + o_k(1)) < W(r,k)$.
using an improved version of Lovasz Local Lemma

$$W(r,k) < 2^{2^{r^{2^{2^{k+9}}}}} \quad (\text{Gower '08}).$$

Theorem (Improved Lower Bounds for $W(r,k)$) For all $r, k \geq 2$ $W(r,k) > \frac{r^k}{4rk}$

Proof. We use Lovasz' Local lemma and from the previous proof,

$$Pr(A_S) = r^{1-k}$$

for each k-AP. So, take $p = r^{1-k}$.

Also let S and T be two different k-AP in $\{1, \dots, n\}$. Notice that A_S and A_T are independent, if they do not share any common integer.

Now, we are going to count how many k-AP's can intersect with S fix any integer $x \in S$ and a natural number $q > 0$. Assume that T intersects with S at x and the difference between the consecutive terms of T is q. Observe that there are at most k such T's that intersects with at x (with the common difference q)

Also there are at most $\lfloor n/k \rfloor$ many choice for q.

$$1 \leq a \leq a+q \leq \dots \leq a+(k-1)q \leq n \overset{\text{H.W.}}{\Longrightarrow} q \leq \lfloor n/k \rfloor$$

Finally, there are k many different x's. So we get that there are at most nk many different

T's in $\{1,\dots,n\}$ for which A_S and A_T may be dependent, since

$$\underbrace{k \cdot \lfloor \tfrac{n}{k} \rfloor}_{\substack{\#\text{ of different}\\ T\text{'s where } x,q \text{ fixed}}} \cdot \overbrace{k}^{\#\text{ of different } x\text{'s}} \leq \underbrace{\tfrac{n}{k}}_{\#\text{ of different } q\text{'s}} \cdot k^2 = nk$$

If we take $d = nk$, by Lovasz's Local Lemma we obtain that

$$4pd = 4r^{1-k}nk \leq 1$$

implies there is a coloring where none of k-AP's are monochoramatic.

$$(*) \iff n \leq \frac{r^{k-1}}{4k}$$

Thus $W(r,k) > \dfrac{r^k}{4k}$

$\square$

Some Known Bounds

1. $W(r,k) \geq 2^k / k^\varepsilon$ for sufficiently large k (Szabo '90)
2. $W(2, p+1) \geq p \cdot 2^p$ for $p \in \mathbb{P}$ (Berlekamp '68)

§ Expectation and First Moment Method

Definition. A random variable (r,v) X is a map $X : \Omega \to \mathbb{R}$ on a probability space $X : \Omega \to \mathbb{R}$ where for each Borel set of reals E, we have $X^{-1}(E) \in \mathcal{F}$
(Here Borel set corresponds to an element of the Borel σ-algebra of reals which is the smallest σ algebra containing the standart topology of reals.)

Here we denote

$$P_r(X \in S) := P_r(\{\omega \in \Omega : X(\omega) \in S\})$$

Example. (Flip Coin)

$$P_r(X=0) = P_r(X=1) = \frac{1}{2}$$

Example (uniform r.v.)

$$\forall i \in \{1,\dots,n\} \quad P_r(U=i) = 1/n$$

Example. (Indicator function as an r.v.)

Let $(\Omega, \mathcal{F}, P_r)$ be a probability space and $A \in \mathcal{F}$. Then the indicator function $\mathbb{1}_A : \Omega \to \mathbb{R}$ defined as $\omega \mapsto \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{else} \end{cases}$ is an r.v. (H.W show it)

If $|\Omega| = n < \infty$, the elements of $A \subseteq \Omega$ are chosen uniformly at random then we have

$$P_r(A) = P_r\left(\bigcup_{a \in A} \{a\}\right) = \sum_{a \in A} P_r(\{a\}) = \sum_{a \in A} \frac{1}{n} = \frac{1}{|\Omega|} \sum_{a \in A} 1 = \frac{|A|}{|\Omega|}$$

Independen Random Variables Set

Let $(\Omega, \mathcal{F}, P_r)$ be a probability space $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ r.v.s. For some given Borel set $A, B \subseteq \mathbb{R}$, we may want to consider the prob. of $X \in A$ given that $Y \in B$ where $P_r(Y \in B) > 0$.

In that case

$$P_r(X \in A \mid Y \in B) = \frac{P_r(X \in A, Y \in B)}{P_r(Y \in B)}$$

where $$P_r(X \in A, Y \in B) = P_r\left(\left\{ \omega \in \Omega : \begin{matrix} X(\omega) \in A \\ Y(\omega) \in B \end{matrix} \right\}\right)$$

Definition. Let $(\Omega, \mathcal{F}, P_r)$ be a probability space. $X, Y : \Omega \to \mathbb{R}$ r.v.s. We say that X and Y are (pairwise) independent if for any Borel sets $A, B \subseteq \mathbb{R}$

$$P_r(X \in A, Y \in B) = P_r(X \in A) P_r(Y \in B).$$

Definition. Let $(\Omega, \mathcal{F}, P_r)$ be a probability space and X be a discrete r.v. where $|\mathrm{Ran}(X)| < \infty$. The expected value or expectation of X is defined by

$$\mathbb{E}(X) = \sum_{x \in \mathrm{Ran}(X)} x \cdot P_r(X = x).$$

Measure Theory'de $\int_\Omega d\mu$.

~~X~~ Missed one class.

---

§4. Varianne & Second Moment Method.

Variance and Covariance.

Definition. Let $(\Omega, \mathcal{F}, \Pr)$ be a probability space and X a r.v. on Ω. Then the variance of X is defined as

$$\mathrm{Var}(X) := \mathbb{E}((X-\mathbb{E}(X))^2)$$

Remark. $\mathrm{Var}(X) \geq 0$ and standart deviation $SD(X) = \sqrt{\mathrm{Var}(X)}$.

Properties.

- $\mathrm{Var}(X) = 0 \iff \Pr(X = \mathbb{E}(X)) = 1$

$$(\Leftarrow)\ \mathrm{Var}(X) = \sum_{\substack{x \in \mathrm{Ran}(X) \\ x = \mathbb{E}(X)}} \cancelto{0}{((x-\mathbb{E}(X))^2)} \Pr(X=x) + \sum_{\substack{x \in \mathrm{Ran}(X) \\ x \neq \mathbb{E}(X)}} (x-\mathbb{E}(X))^2 \cancelto{0}{\Pr(X=x)} = 0.$$

H.W: Consider the situation $\Pr(X \neq \mathbb{E}(X)) > 0$.

- $\forall a, b \in \mathbb{R}$ $\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$. Consider the

$$\mathrm{Var}(ax+b) = \mathbb{E}((ax+b) - \mathbb{E}(ax+b)^2)$$
$$= a^2 (\mathbb{E}(X - \mathbb{E}(X))).$$

- Chebyshev's Inequality: Let X be an r.v. and $\lambda \in \mathbb{R}_{>0}$. Then

$$\Pr(|X - \mathbb{E}(X)|) \geq \lambda \sqrt{\mathrm{Var}(X)} \leq 1/\lambda. \qquad (*)$$

Proof. First, consider the case when $\mathrm{Var}(X) = 0$. This holds iff $X = \mathbb{E}(X)$ and prob 1 (by V(1)). So,

$$\Pr(|X - \mathbb{E}(X)| > 0) = 1 - \Pr(|X - \mathbb{E}(X)| \leq 0) = 1 - 1 = 0 < 1/\lambda.$$

Now assume that $\mathrm{Var}(X) > 0$. Define a new variable r.v. $Y = (X - \mathbb{E}(X))^2$. By Markov's Inequality, $\Pr(Y > \alpha) \leq \Pr(Y \geq \alpha) < \frac{\mathbb{E}(Y)}{\alpha}$

Choose $\alpha = \lambda^2 \mathrm{Var}(X) > 0$. Then $\Pr((X - \mathbb{E}(X))^2 > \lambda^2 \mathrm{Var}(X)) < \frac{\mathbb{E}((X-\mathbb{E}(X))^2)}{\lambda^2 \mathrm{Var}(X)}$

$$\iff \Pr(|X - \mathbb{E}(X)| > \lambda \sqrt{\mathrm{Var}(X)}) < 1/\lambda^2.$$

$\square$

Remark. $(*)$ asserts that $X = \mathbb{E}(X) + O((\mathrm{Var}(X))^{1/2})$ occurs with high probability.

Definition. Let X and Y be two r.v. on the same probability space $(\Omega, \mathcal{F}, \Pr)$. The convariance of X and Y is defined as $\mathrm{Cov}(X,Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))$

$$= \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).$$

Proposition. (C.1) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

(C-2) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.

(C.3) If X and Y are independent r.v.s., then $\mathrm{Cov}(X,Y)$.

(C.4) $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{\substack{1\le i,\, j\le n \\ i\neq j}} \mathrm{Cov}(X_i, X_j)$.

↳ (first compute $\mathrm{Var}(X+Y)$, then do ....)

(V.5) If $\{X\}_{i=1}^{n}$ is a set of pairwise independent r.v.s., then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

(V.6) Assume that B is some randomly generated subset of A. Then

$$\mathrm{Var}(|B|) = \sum_{a\in A} \Pr(a\in B)\,\Pr(a\in B)^2 + \sum_{\substack{a,a'\in A \\ a\neq a'}} \mathrm{Cov}(\mathbb{1}_{a\in B}\cdot \mathbb{1}_{a'\in B})$$

Book: Tao & Vu – Additive Combinatorics.

In particular, if the events $a\in B$ are pairwise independent, then

$$\mathrm{Var}(|B|) = \sum_{a\in A} \left(\Pr(a\in B) - P(a\in B)^2\right)$$

Book: Cojoru & Murty – An Introduction to Sieve Theory → Lecture Notes.

Hardy–Ramanujan Thr.

$$\omega(n) = \sum_{p\mid n} 1 \quad \text{— \# of prime divisors of } n.$$
$$= \sum_{p\le n} \mathbb{1}_{p\mid n}$$

Theorem. Let $f(n)$ tend to infinity arbitrarily slowly. Then

$$\#\left\{ k\le n : |\omega(k) - \log\log k > f(n)\sqrt{\log\log n} \right\} = o(n).$$

Erdős–Kac Theorem ('40) Let $\lambda\in\mathbb{R}$ be fixed. Then

$$\lim_{n\to\infty} \frac{1}{n}\left|\left\{k\in[1,n] : \omega(k) \geq \log\log n + \lambda\sqrt{\log\log n}\right\}\right| = \int_{\lambda}^{\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt$$

(Book: Alon & Spencer The Probabilistic Method.)

Its proof uses C3 and Central Limit Theorem.

Distinct Sums.

A set of positive integers $\{x_1,\dots,x_n\}$ has distinct sums, if $\forall S\subseteq\{1,\dots,n\}$ the sums $\sum_{i\in S} x_i$ are distinct. Let $f(n)$ denote the maximum k for which $\exists$ a set

$\{x_1,\dots,x_k\} \subseteq \{1,\dots,n\}$ with distinct sums.

Example. $\{1, 2, 2^2, 2^3, \dots, 2^{\log_2 n}\} \subseteq \{1,\dots n\}$. $f(n) \geq \lfloor \log_2 n \rfloor + 1 \geq \log n$.

To obtain an upper bound for $f(n)$, first observe that

$$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} x_k \leq nk \quad \text{for } f(n) = k$$

This means that each sum is less than nk. Since all the sums are distinct, we get $\forall S \subseteq \{1,\dots,k\}$, $\sum_{i\in S} x_i \in \{1, 2, \dots, nk-1\}$

Hence

$$\left\{\sum_{i\in S} x_i : S \subseteq \{1,\dots,k\}\right\} \subseteq \{1, 2, \dots, nk-1\}.$$

Thus, (since there are exactly 2^k subsets of $\{1,\dots,k\}$)

$$2^k = \left|\left\{\sum_{i\in S} x_i \mid S \subseteq \{1,\dots,k\}\right\}\right|$$
$$\leq |\{1, \dots, nk-1\}| < nk$$

Therefore $\dfrac{2^k}{k} = \dfrac{2^{f(n)}}{f(n)} < n \longrightarrow k < \log_2 n + \log_2 \log_2 n + 2$.

Theorem. Assume that the above settings hold. Then,

$$f(n) < \log_2 n + \frac{1}{2} \log_2 \log_2 n + 3$$

Merten's Theorem. $\underset{f(x)}{\sum_{p\leq x} \frac{1}{p}} \rightsquigarrow \underset{g(x)}{\log \log x}$ $\qquad \lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Example

$$\sum_{\substack{p\in\mathbb{P} \\ p \text{ known}}} 1/p = 4.$$

Proof (Erdős, Moser, '56). Fix any set $\{x_1,\dots,x_k\} \subseteq \{1,\dots,n\}$ with distinct sums. Let $\xi_1,\dots,\xi_k$ be independent and uniformly distributed r.v.s. on $\{0,1\}$. Define

$$X = \sum_{i=1}^{k} \xi_i x_i \quad - \text{ r.v.}$$

Since ξ_i's are uniformly dist., then $\forall i \in \{1,\dots,k\}$, $\Pr\{\xi_i = 0\} = \Pr\{\xi_i = 1\} = 1/2$.

Then $\mathbb{E}(X) = \dfrac{x_1 + \dots + x_k}{2}$

$$\mathrm{Var}(X) = \sum_{i=1}^{k} \mathrm{Var}(\xi_i x_i) = \sum_{i=1}^{k} x_i^2 \,\mathrm{Var}(\xi_i) = \sum_{i=1}^{k} x_i^2 \left(\mathbb{E}(\xi_i^2) - \mathbb{E}(\xi_i)^2\right)$$
$$= \sum_{i=1}^{k} x_i^2 \left(\sum_{x=0}^{1} x^2 \Pr(\xi_i = x) - \frac{1}{4}\right)$$

$$= \sum_{i=1}^{k} x_i^2/4 = \frac{x_1^2 + \cdots + x_k^2}{4} \leq \frac{n^2 k}{4}$$

This gives $\sqrt{\mathrm{Var}(x)} \leq \frac{n\sqrt{k}}{2}$

By CI, we put

$$\Pr(|X - \mathbb{E}(x)| \geq \lambda\sqrt{\mathrm{Var}(x)}) \leq 1/\lambda^2 \quad \text{for any } \lambda > 0. \text{ So,}$$

$$1 - \Pr(|X - \mathbb{E}(x)| < \lambda\sqrt{\mathrm{Var}(x)}) \leq 1/\lambda^2 \iff \Pr(|X - \mathbb{E}(x)| < \lambda\sqrt{\mathrm{Var}(x)}) \geq 1 - \frac{1}{\lambda^2}. \quad (*)$$

Observe that for any $s \in \{1, \ldots, n\}$

$$\Pr(X = s) \in \{0, \frac{1}{2^k}\}$$

Therefore,

$$\Pr(|X - \mathbb{E}(x)| < \lambda\sqrt{\mathrm{Var}(x)}) = \Pr\left(X \in (\mathbb{E}(x) - \lambda\sqrt{\mathrm{Var}(x)}, \mathbb{E}(x) + \lambda\sqrt{\mathrm{Var}(x)})\right)$$

Hence,

$$\Pr(|X - \mathbb{E}(x)| < \lambda\sqrt{\mathrm{Var}X}) \leq \frac{2\lambda\sqrt{\mathrm{Var}(X)} + 2}{2^k}$$

Combining this with $(*)$, we get

$$1 - 1/\lambda^2 \leq \frac{\lambda n\sqrt{k} + 1}{2^k} \quad \left(\text{as } \sqrt{\mathrm{Var}(X)} \leq \frac{n\sqrt{k}}{2}\right)$$

$$n \geq \left(\frac{\lambda^2 - 1}{\lambda^3}\right)\frac{2^k}{\sqrt{k}} - \frac{1}{\lambda\sqrt{k}} \qquad \text{Pick } \lambda = \sqrt{3}, \text{ the result follows.}$$

→ Chernoff's Bond

↳ Sidon Set

→ F.K.G Inequality ↳ additive N.T.