Lecture 1 : Introduction

We will start with a simple combinatorial problem. Consider $\{-1,1\}^{1000}$. How many elements

$$x \in \{-1,1\}^{1000}$$

satisfy

$$\left|\sum_{i=1}^{1000} x_i\right| \geq 50?$$

More generally, for any $n \in N$ and $\lambda > 0$ how many elements

$$x \in \{-1,1\}^n$$

satisfy

$$\left|\sum_{i=1}^{n} x_i\right| \geq \lambda n?$$

The answer is given by the binomial distribution. We are only seeking approximations. This is a question that we will spend a fair deal of time on this quarter. Today we will be satisfied with a crude upper bound.

Fact: For any $r \in \mathbb{R}$

$$(r+1)^2 + (r-1)^2 = (r^2+2r+1) + (r^2-2r+1) = 2(r^2+1).$$

For $x \in \{-1,1\}^n$ we write $S_n(x) \sum_{i=1}^n x_i$ and for $m < n$ we write $x|_m$ for the restriction of x to the first m terms.

Lemma 1.0.1

$$\sum_{x \in \{-1,1\}^n} (S_n(x))^2 = n2^n.$$

Proof: By induction. It is easy to check that it is true for $n = 1$. Assume it is true for n.

For each $y \in \{-1,1\}^{n+1}$ we will write it as a pair (x,a) where $x \in \{-1,1\}^n$ and $a \in \{-1,1\}$.

$$\begin{aligned}
\sum_{y\in\{-1,1\}^{n+1}} \left(\sum_{i=1}^{n+1} y_i\right)^2 &= \sum_{y\in\{-1,1\}^{n+1}} (S_n(y|_n) + y_{n+1})^2 \\
&= \sum_{X\in\{-1,1\}^n} (S_n(X)+1)^2 + (S_n(x)-1)^2 \\
&= 2\left(\sum_{x\in\{-1,1\}^n} (S_n(x))^2 + 1\right) \\
&= 2(n2^n + 2^n) \\
&= (n+1)2^{n+1}.
\end{aligned}$$

■

Fix $\lambda > 0$ and $n \in \mathbb{N}$. Let

$$A_{\lambda,n} = \{x \in \{-1,1\}^n : \ |S_n(x)| > \lambda n\}.$$

Then

$$n2^n = \sum_{x\in\{\}^n} S_n(x)^2 \geq \sum_{x\in A_{\lambda,n}} s(x)^2 \geq |A_{\lambda,n}|(\lambda n)^2.$$

Rearranging we get that

$$A_{\lambda,n} \leq \frac{2^n}{\lambda^2 n}.$$

We have proven the following theorem.

Theorem 1.0.2 *For any* $\lambda > 0$

$$\lim_{n\to\infty} \frac{A_{\lambda,n}}{2^n} \to 0.$$

This is a version of the weak law of large numbers.

Key Concepts

1. variance
2. large deviations
3. Chebychev's inequality

There is one real drawback of the approach that we took. It works fine for finite statements but it doesn't allow us to make statements about infinite objects like $\{-1, 1\}^\infty$.

What questions do we want to ask? For example what fraction of $x \in \{-1, 1\}^\infty$ satisfy

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i \to 0?$$

Although it is possible to answer this question using discrete statements, we will introduce measure theory which will make such questions possible to state formally. One question which we will be interested in is for what fraction of $x \in \{-1, 1\}^{\mathbb{N}}$ does

$$\lim_{n\to\infty} \frac{1}{n} S_n(x) = 0?$$

Since the set of x which satisfy this condition is infinite, as is the set of x which don't satisfy the condition. We cannot answer this question like we did with counting arguments above. With measure theory this question will be easy to state.

But now we show how to give an answer to a similar question which can be analyzed by combinatorial means.

Define

$$B_{\epsilon,N_0,N} = \left\{ x \in \{-1, 1\}^N : \text{there exists m such that } n_0 < m \le N \text{ with } \frac{1}{m} S_m(x)| > \epsilon \right\}.$$

We wish to show that for every $\epsilon > 0$ there exists N_0 such that $\frac{1}{2^N}|B_{\epsilon,N_0,N}$ for all $N > N_0$. This is a finite version that indicates that for "most" $x \in \{-1, 1\}^{\mathbb{N}}$ we have that $\frac{1}{n} S_n(x) \to 0$. We won't quite be able to do this today but there is a slight modification of our argument can be made to work.

Fix $\epsilon > 0$, N_0 and N. For every $x \in B = B_{\epsilon,N_0,N}$ there exists an m such that $x|_m$ is in $A_{\epsilon,m}$. And for every element of $A_{\epsilon,N}$ there are exactly 2^{N-m} elements of B. Thus by our previous estimate

$$|B| \le \sum_{m=N_0+1}^{N} |A_{\epsilon,m}| 2^{N-m} \le \sum_{m=N_0+1}^{N} \frac{2^m}{\epsilon n} 2^{N-m} \le \sum_{m=N_0+1}^{N} \frac{2^N}{\epsilon m}.$$

If the series $\sum \frac{C}{m}$ were summable then we could just choose N_0 large enough so that the sum were less than ϵ and be done. Unfortunately that isn't true. In about two weeks we will return to this argument to show how it modify this argument to make it work.

We will define

$$\tilde{B}_{\epsilon,N_0,N} = \Bigg\{x \in \{-1,1\}^N : \text{there exists m such that } n_0 < m \leq N \text{ with } \frac{1}{m}S_m(x)| > \epsilon/2$$
$$\text{and } m \text{ is a perfect square}\Bigg\}.$$

We will then prove the following two statements:

1. for every $\epsilon > 0$ there exists N_0 such that $\frac{1}{2^N}|\tilde{B}_{\epsilon,N_0,N}| < \epsilon$ for all $N > N_0$ and
2. $B_{\epsilon,N_0,N} \subset \tilde{B}_{\epsilon,N_0,N}$.

Combining these two proves the result.

Lecture 2 : Ideas from measure theory

2.1 Probability spaces

This lecture introduces some ideas from measure theory which are the foundation of the modern theory of probability. The notion of a probability space is defined, and Dynkin's form of the monotone class theorem is presented.

Definition 2.1.1 *Let Ω be a set of points ω. In probability theory, Ω represents all possible outcomes of an experiment or observation.*

Example 2.1.2 *Tossing a coin has a set of outcomes $\Omega = \{Head, Tail\}$.*

Example 2.1.3 *The position of a body in a 3-D Euclidean space belongs to the set $\Omega = R^3$.*

A subset of Ω is called an event. It is natural to ask questions such as whether or not an outcome of a random experiment belongs to to a event. To do this, we need to define the events under consideration – we need to define a class of subsets of the space Ω. Since we'll want to talk about combinations of events, a systematic treatment will require this class of subsets to have some nice set-theoretic properties. The next definition spells this out precisely.

Definition 2.1.4 *A class $\mathcal{F}$ of subsets of a space Ω is called a* field *if it contains Ω itself and is closed under complements and finite unions. That is*

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$
3. $A, B \in \mathcal{F}$ *implies* $A \cup B \in \mathcal{F}$

Recall that $A \cap B = (A \cup B)^c$. Thus, a set of subsets $\mathcal{F}$ that is closed under complements is is closed under unions if and only if it is also is closed under intersections. Therefore, closure under union in the definition above could be replaced by closure under intersection.

Definition 2.1.5 *A class $\mathcal{F}$ of subsets of Ω is a* σ-field *if it is a field and if it is closed under the formation of countable unions. That is,*

1. *$\mathcal{F}$ is a field.*
2. *$A_1, A_2, ... \in \mathcal{F}$ implies $A_1 \cup A_2 \cup .... \in \mathcal{F}$.*

A field is closed under finite set theoretic operations whereas a σ-field is closed under countable set theoretic operations. In a problem dealing with probabilities, one usually deals with a small class of subsets $\mathcal{A}$, for example the class of subintervals of $(0, 1]$. It is possible that if we perform countable operations on such a class $\mathcal{A}$ of sets, we might end up operating on sets outside the class $\mathcal{A}$. Hence, we would like to define a class denoted by $\sigma(\mathcal{A})$ in which we can safely perform countable set-theoretic operations. This class $\sigma(A)$ is called the σ-field generated by $\mathcal{A}$, and it is defined as the intersection of all the σ-fields containing $\mathcal{A}$ (exercise: show that this is a σ-field). $\sigma(A)$ is the smallest σ-field containing A.

Definition 2.1.6 *A real-valued set function*[1] *$\mathbb{P}$ on a σ-field $\mathcal{F}$ is a* probability measure *if it satisfies the following conditions:*

1. *$0 \leq \mathbb{P}(A) \leq 1$ for $A \in \mathcal{F}$.*
2. *$\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$.*
3. *If $\{A_i\}_{i\in\mathbb{N}}$ are disjoint events in $\mathcal{F}$, then $\mathbb{P}(\bigcup_i A_i) = \sum_{i=0}^{\infty} \mathbb{P}A_i$.*

If $\mathcal{F}$ is a σ-field, then the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability measure space or simply a *probability space*. The countable additivity of the probability measure gives rise to the following properties that are stated in a theorem. Here, $A_i \uparrow A$ means that we are given a countable collection $\{A_i\}_{i\in\mathbb{N}}$, $A_i \subseteq A_j$ for $i < j$ and $\bigcup_{i=1}^{\infty} A_i = A$. Similarly, $A_i \downarrow A$ means $A_i \supseteq A_j$ for $i < j$ and $\bigcap_{i=1}^{\infty} A_i = A$.

Theorem 2.1.7 *Let $\mathbb{P}$ be a probability measure on a field $\mathcal{F}$.*

1. *Continuity from below: If A_n and A lie in $\mathcal{F}$ and $A_n \uparrow A$, then $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$.*
2. *Continuity from above: If A_n and A lie in $\mathcal{F}$ and $A_n \downarrow A$, then $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$.*
3. *Countable subadditivity: If $A_1, A_2...$ is a countable collection of events in $\mathcal{F}$,*
$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k). \tag{2.1}$$

[1] A set function is a function whose domain is a class of sets.

Example 2.1.8 *If $\mathcal{A}$ is the class of subintervals of $\Omega = (0, 1)$, then the sigma field generated by $\mathcal{A}$, denoted by $\mathcal{B}$, is called the collection of* Borel *sets of the unit interval. The probability space on a unit interval is then defined as $(\Omega, \mathcal{B}, \mathbb{P})$, where $\Omega = (0, 1)$, $\mathbb{P}(B) = \lambda(B)$ for $B \in \mathcal{B}$. Here λ denotes Lebesgue measure, for which $\lambda((a, b]) = b - a$.*

Caratheodory's Extension Theorem

Example $(0, 1]$

Let $\mathcal{A}$ be the collection of all sets of the form

$$A = (a_1, b_1 \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$$

for some $n \in \mathbb{N}$ and some sequence $0 \leq a_1 < b_1 \leq a_2 < \cdots < b_n$.

It is easy to check that $\mathcal{A}$ is an algebra and that $\mu(A) = \sum_{i=1}^n b_i - a_i$ is a measure on $\mathcal{A}$. Thus by Theorem we can extend μ to a measure on $\sigma(\mathcal{A})$.

Example $\{0, 1\}^{\mathbb{N}}$

Let $\mathcal{A}$ be the collection of all sets of the form

$$A = \{x : x_1 = a_1, \ldots, x_n = a_n\}$$

for some $n \in \mathbb{N}$ and some sequence $a \in \{0, 1\}^n$.

For $A \in \mathcal{A}$ we can define $\mu(A) = 2^{-n}$. It is easy to check that μ is a measure on the algebra generated by $\mathcal{A}$ (finite unions of elements in $\mathcal{A}$). Thus by Theorem we can extend μ to a measure on $\sigma(\mathcal{A})$.

We could also have chosen

$$\mu(A) = p^{\#\{i: a_i = 1\}}(1 - p)^{\#\{i: a_i = 0\}}$$

or

$$\mu(A) = p^{\#\{i: a_i = 1 \text{and } i \text{ is prime}\}}(1-p)^{\#\{i: a_i = 0\} \text{and } i \text{ is prime}} q^{\#\{i: a_i = 1 \text{and } i \text{ is not prime}\}}(1-q)^{\#\{i: a_i = 0\} \text{and } i \text{ is not prim}}$$

or any other of millions of choices.

Show that $B = \{x : \lim_{m \to \infty} \frac{1}{m} S_m(x) = .7\}$ is in $\sigma(\mathcal{A})$.