Probability Theory

Soham Pyne

IISER-BHOPAL

October 25, 2020

Putting a "Measure" on Randomness

- Define Randomness!
- Trying to put a **Measure** on Randomness
- Incorporating the randomness in terms of mathematical language.

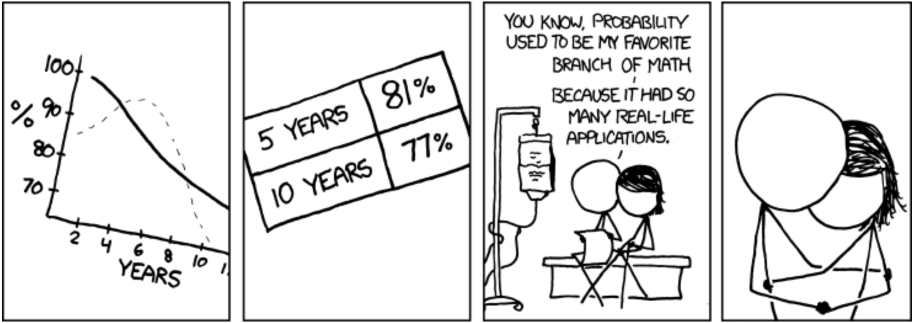
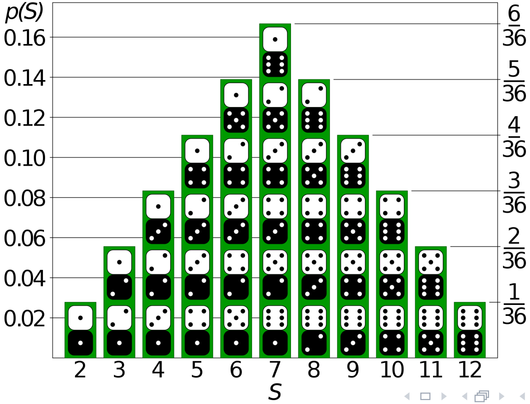


Figure: Credits:xkcd.

Discrete Space

example

Tossing two six sided fair die simultaneously



Discrete Space

Definition

A discrete probability space is a pair of (Ω, p) where Ω is a finite or countable set and $p : \Omega \to \mathrm{R}_+$ is a function such that $\sum_{\omega \in \Omega} p_\omega = 1$. For any subset $A \subset \Omega$ define $p(A) = \sum_{\omega \in A} p_\omega$

- We need the basic notion of countable sums
- Notion of convergence and divergence

Length of an Interval

Example

Suppose we consider the sample space. $\Omega = [0, 1]$ then can we justify statements like if $A = [0.1, 0.6]$ then $P(A) = 0.5$ because saying things like "$P([0.1, 0.6]) = \sum_{\omega \in [0.1,0.6]} p_\omega$" makes no sense.

- Cantor set
- set of rationals

Infinite coin toss

example

We consider a sequence of infinite fair coin tosses.

$$\Omega = \{0,1\} \times \{0,1\} \times \dots \tag{1}$$

- What is the probability that we get three consequetive heads?
- What is the probability that first two tosses are tails?

What to expect

Let us again take $\Omega = [0,1]$ and for any set $A \subseteq [0,1]$ define "length"

$$\mathbf{P}^*(A) = \inf\{\sum_{k=1}^{\infty} |I_k| : \text{each } I_k \text{ is an interval and } A \subset \bigcup_{k=1}^{\infty} I_k\} \qquad (2)$$

Example:Rationals

Let $A = \mathbf{Q} \cap [0,1]$ and fixing $\epsilon > 0$ such that $I_k = [r_k - \epsilon 2^{-k}, r_k + \epsilon 2^{-k}]$ then $A \subseteq \cup_k I_k$ and $\sum_k |I_k| = 2\epsilon$ which shows $\mathbf{P}^*(A) = 0$ as ϵ is arbitrary.

Not so easy

Outer measure is not finitely additive

There exist a subset $A \subseteq [0,1]$ such that $\mathbf{P}^*(A) = 1$ and $\mathbf{P}^*(A^c) = 1$

Further we also demand the notion of countable additivity.

$$\mathbf{P}(\cup_{n=1}^{\infty} A_n) = \sum_n \mathbf{P}(A_n) \tag{3}$$

Curious case

There does not exist any function $f : 2^{[0,1]} \to [0,1]$ such that f is countably additive and $f([a,b]) = b - a$ for all $[a,b] \subseteq [0,1]$

That is why we do not say "Draw a number at random from $[0,1]$"

Setting up the space

Definition : Probability space

A probability space is a triple$(\Omega, \mathcal{F}, \mathbf{P})$ such that

- The sample space Ω is a arbitrary nonempty set.
- The σ-algebra $\mathcal{F}$ is a set of subsets of Ω such that $(i)\phi, \Omega \in \mathcal{F}, (ii)$ if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}, (iii)$ if $\{A_n\}_{n\in\mathbf{N}} \in \mathcal{F}$ then $\cup A_n \in \mathcal{F}$. That is $\mathcal{F}$ or set of measurable sets is closed under complementation,countable unions and contain the empty set.
- A probability measure is any function $\mathbf{P} : \mathcal{F} \to [0, 1]$ is such that if $A_n \in \mathcal{F}$ and pairwise disjoint then $\mathbf{P}(\cup_n A_n) = \sum_n \mathbf{P}(A_n)$ such that $\mathbf{P}(\Omega) = 1$. $\mathbf{P}(A)$ is called the probability of A.

Probability and Measure

Probability is called a measure :

- If we allow $\mathbf{P}$ to take values in $[0, \infty]$.
- We drop the requirement $\mathbf{P}(\Omega) = 1$

They have same basic property but probability has the properties

- Independence
- Conditional Expectation

Which makes it richer. This framework includes the old one of the discrete probability spaces and here we have examples where we can includes all subsets of Ω is the $\sigma-$algebra.

Important properties:

- **F** is closed under finite and countable unions, intersections, differences, symmetric differences.
- **Countable subadditivity:** If $A_n \in \mathcal{F}$ then $\mathbf{P}(\cup_n A_n) \leq \sum_n \mathbf{P}(A_n)$
- **Continuity from below and above** If $A_n \in \mathcal{F}$ and $A_n \uparrow (\downarrow) A$ then $\mathbf{P}(A_n) \uparrow (\downarrow) \mathbf{P}(A)$.
- If $A_n \in \mathcal{F}$ then

$$\limsup A_n = \{\omega : \omega \text{ belongs to infinitely many } A_n\}$$
$$\liminf A_n = \{\omega : \omega \text{ belongs to all but finitely many } A_n\}$$

are also in $\mathcal{F}$.

Generated σ-algebra

- If S is a collection of subsets of Ω then there exists a smallest σ-algebra $\mathcal{F}$ containing S. $\mathcal{F} = \sigma(S) = \bigcap_{i \in I} \mathcal{F}_i$
- We take $S =$ set of all left open right closed intervals in $[0, 1]$. Then $\sigma(S)$ is called the borel σ-algebra on $[0, 1]$ and there is a probability measure $\mathbf{P}$ satisfying $\mathbf{P}(a, b] = b - a$

Terminologies

Let S be a collection of subsets of Ω

- $\pi -$ **system** : if $A, B \in S \implies A \cap B \in S$
- $\lambda -$ **system** : $\Omega \in S$, If $A, B \in S$ and $A \subseteq B \implies B \backslash A \in S$ also if $A_n \uparrow A$ and $A_n \in S \implies A \in S$
- **Algebra** :$\phi, \Omega \in S$. If $A \in S \implies A^c \in S$ and $A, B \in S \implies A \cup B \in S$
- **Monotone class**: If $A_n \uparrow A$ or $A_n \downarrow A$ then $A \in S$
- $\sigma -$ **algebra** : If $\phi, \Omega \in S$ and $A \in S \implies A^c \in S$ also $A_n \in S \implies \cup_n A_n \in S$.

Sierpinski-Dynkin $\pi - \lambda$ theorem

Theorem:

Let Ω be a set and let $\mathcal{F}$ be a set of subsets of Ω.

1. $\mathcal{F}$ is a $\sigma-$algebra if and only if it is a π system as well as a λ system.
2. If S is a $\pi-$ system, then $\lambda(S) = \sigma(S)$

Proof

Suppose $\mathcal{F}$ is a π-system as well as a λ-system. Then $\Omega \in \mathcal{F}$ and if $A \in \mathcal{F}$, then $A^c = \Omega \backslash A \in \mathcal{F}$. If $A_n \in \mathcal{F}$, then the finite unions $B_n := \cup_{k=1}^{n} A_k = (\cap_{k=1}^{n} A_k^c)^c \in \mathcal{F}$. The countable union $\cup A_n$ is an increasing sequence of B_n hence belongs to $\mathcal{F}$ by the λ-property.

Monotone class theorem

Monotone class theorem:

Let Ω is a set and let S be a collection of subsets of Ω. If S is a an algebra then the monotone class generated by S is a $\sigma-$algebra. That is $\mathcal{M}(S) = \sigma(S)$.

Applications

If two probability measures on $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$ agree on all intervals, then are they same?

Lemma

Let S be a $\pi-$system of subsets of Ω and let $\mathcal{F} = \sigma(S)$. If $\mathbf{P}$ and $\mathbf{Q}$ are two probability measures on $\mathcal{F}$ such that $\mathbf{P}(A) = \mathbf{Q}(A)$ for all $A \in S$. Then $\mathbf{P}(A) = \mathbf{Q}(A)$ for all $A \in \mathcal{F}$.

Measurable function

Random variables

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $(X, \mathcal{G})$ be a set with a σ-algebra. A function $T : \Omega \to X$ is called a random variable or a measurable function if $T^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{G}$.

- Generally X is taken to be a metric space and $\mathcal{G} = \mathcal{B}_X$
- If $X = \mathbf{R}$ we have $\mathbf{R}$-valued random variable.
- If $X = \mathbf{R}^d$ we have a Random vector.

Neither Random nor a variable!

Let Ω be a set of all possible random outcomes of some experiment. Then a **Random Variable is a function assigning a numerical value to each outcome**.The randomness part is the **underlying uncertainty in picking an element $\omega \in \Omega$ as an outcome of the experiment**. Once that is done the RV deterministically picks a real number $T(\omega) \in \mathbf{R}$. We note the fact that a probability measure is associated with select subsets of Ω where as a RV is associated to each element of Ω.

Push forward of a measure:

Definition

If $T : \Omega_1 \to \Omega_2$ is a random variable and $\mathbf{P}$ is a probability measure on $(\Omega_1, \mathcal{F}_1)$, then defining $\mathbf{Q}(A) = \mathbf{P} \circ T^{-1}(A)$ we get a probability measure on $(\Omega_2, \mathcal{F}_2)$. $\mathbf{P} \circ T^{-1}(A)$ is called push forward of $\mathbf{P}$ under T.

- $T : \Omega \to \mathbf{R}$ is $(\mathcal{F}, \mathcal{B}(\mathbf{R}))$ measurable and $\mathbf{P} : \Omega \to [0, 1]$ is $(\mathcal{F}, \mathcal{B}([0, 1]))$ measurable.
- $\mathbf{P} \circ T^{-1}(A) = \mathbf{P}(\{\omega \in \Omega | T(\omega) \in A\})$.
- $\mathbf{Q}$ is indeed a measure since if we take disjoint sets $\{A_n\} \in \mathcal{B}(\mathbf{R})$ we have $T^{-1}(A_n)$ measurable hence countable additivity follows.

Measures on $\mathbf{R}^d$

Theorem

Let μ be a probability measure on $\mathbf{R}^d$. Then there exists a Borel function $T : [0,1] \to \mathbf{R}^d$ such that $\mu = \lambda \circ T^{-1}$. Where λ is a lebesgue measure on $[0,1]$.

The T defined as above may not be unique. So we associate a unique function to each probability measure.

Distribution function:

Let μ be a Borel probability measure on $\mathbf{R}^d$. The CDF of μ is defined as $F_\mu : \mathbf{R}^d \to [0,1]$ by $F_\mu(x) = \mu(R_x)$; $R_x := (-\infty, x_1] \times \ldots \times (-\infty, x_d]$ for $x = (x_1, \ldots, x_d)$.

Properties:

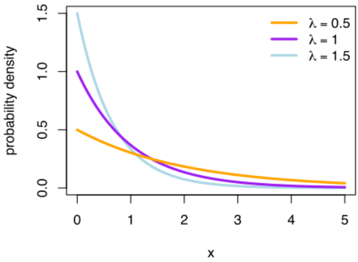
- F_μ is increasing in each coordinate.
- F_μ is right continuous in each coordinate.
- If $\min_i x_i \to -\infty$ then $F_\mu \to 0$ and $\min_i x_i \to +\infty$ then $F_\mu \to 1$

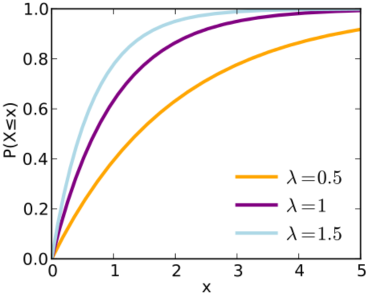
Theorem

A function $F : \mathbf{R}^d \to [0, 1]$ satisfying the three properties above. then there exists a unique borel probability measure μ on $\mathbf{R}^d$ such that $F_\mu = F$.

Examples

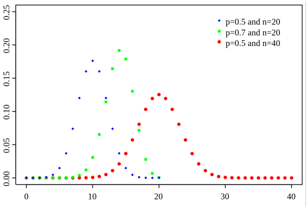
Exponential

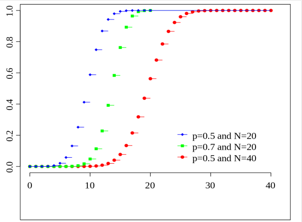




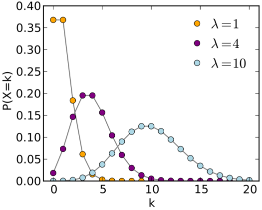
Examples

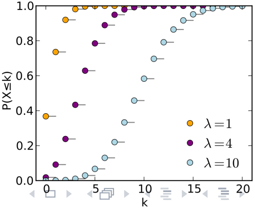
Binomial





Poisson





Lebesgue Integral

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. We define Expectation or lebesgue measure as follows:

- **Simple RV:** If $X = \sum_{i=1}^{n} c_i 1_{A_i}$ for some $A_i \in \mathcal{F}$. We define Expectation of X to be $E[X] := \sum_{i=1}^{n} c_i \mathbf{P}(A_i)$
- **Nonnegative RV:** If $X \geq 0$ then $E[X] := \sup\{E[S] : 0 \leq S \leq X, S \text{is simple RV}\}$. It can be nonnegative or $+\infty$.
- **Real valued:** If X is a real valued RV we define $E[X] := E[X^+] - E[X^-]$.

Properties

- if X, Y are two simple random variables then $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$ for $\alpha, \beta \in \mathbf{R}$.
- If $X \geq 0$ then $E[X] \geq 0$

Linearity of Non-negative random variables are found using:

Monotone convergence theorem

If $X_n(\omega)$ are nonnegative simple random variables increasing to $X(\omega)$ for all ω, then $E[X_n] \uparrow E[X]$

$$X_n(\omega) = \sum_{k=0}^{2^{2n}} \frac{k}{2^n} \mathbf{1}_{X(\omega) \in [k2^{-n}, (k+1)2^{-n}]} \tag{4}$$

Properties

Let $L^1(\Omega, \mathcal{F}, \mathbf{P})$ be the collection of all integrable random variables on Ω. then the expectation operator $E : L^1(\Omega, \mathcal{F}, \mathbf{P}) \to \mathbf{R}$. Then

- Linearity: If X, Y are integrable, then for any $\alpha, \beta \in \mathbf{R}$ the rv $\alpha X + \beta Y$ is also integrable and $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$.
- Positivity: If $X \geq 0$ then $E[X] \geq 0$ and $X \leq Y$implies $E[X] \leq E[Y]$.
- $|E[X]| \leq E[|X|]$
- $E[\mathbf{1}_A] = \mathbf{P}(A)$ for $A \in \mathcal{F}$.

PF

Lemma

Let $T : (\Omega_1, \mathcal{F}_1, \mathbf{P}) \to: (\Omega_2, \mathcal{F}_2, \mathbf{Q})$ be measurable and $\mathbf{Q} = \mathbf{P} \circ T^{-1}$. If X is an integrable rv on Ω_2, then $X \circ T$ is an integrable r.v. on Ω_1 and $E_{\mathbf{P}}[X \circ T] = E_{\mathbf{Q}}[X]$

Corollary

Let X_i be random variables on a common probability space. Then for any Borel measurable $f : \mathbf{R}^n \to \mathbf{R}$,the value of $E[f(X_1, \ldots, X_n)]$(if exists) depend only on the joint distribution of $X_1, \ldots, X_n$

Let $T = (X_1, \ldots, X) : \Omega \to \mathbf{R}^n$. Then $\mu := \mathbf{P} \circ T^{-1}$ is by definition the joint distribution of $X_1, \ldots, X_n$. The lemma gives $E_{\mathbf{P}}[f(X_1, \ldots, X_n)] = \int_{\mathbf{R}^n} f(t) d\mu(t)$ which depends only on measure μ.

Explanation

- The change of variable result shows the irrelevance of the underlying probability measure.
- All our questions can be about finite or infinite collection of random variables.
- The answers depend on the joint distributions of random variables and not on any other details of the underlying probability space.

eg.
Let $X_1, X_2, \ldots$ random variables have gamma distribution with mean and variance ....

Change of variables for densities

Density function

A borel probability measure μ on $\mathbf{R}^n$ has a desity $f : \mathbf{R}^n \to \mathbf{R}_+$ if f is a borel measurable function and $\mu(A) = \int_A f(x)dm(x) = \int_{\mathbf{R}^n} f(x)\mathbf{1}_A dm(x)$

Proposition

Let U, V be open sets of $\mathbf{R}^n$ and let $T : U \to V$ be a bijective smooth function such that $T^{-1} : V \to U$ is also smooth. Let X be a random vector on some probability space, taking values on U and assume that its distribution has density f wrt Lebesgue measure on U. Let $Y = T(X)$, so that Y takes values in V. then Y has density g wrt Lebesgue measure on V where $g(x) = f(T^{-1}x)|detJ[T^{-1}](x)|$.

Example

Suppose (X_1, X_2) has density $f(x, y) = e^{-x-y}$ on $\mathbf{R}^2_+$. Then to find the distribution of $X_1 + X_2$

- we consider the bijection $T(x_1, x_2) = (x_1 + x_2, x_2)$. onto $V = \{(u, v) : u > v > 0\}$ and $T^{-1}(u, v) = (u - v, v)$.
- The determinant of Jacobian is 1.
- The determinant of Jacobian is 1. Hence the density of $(Y_1, Y_2) = T(X_1, X_2)$ is given by $g(u, v) = f(u - v, v)\mathbf{1}_{u>v>0} = e^{-u}\mathbf{1}_{u>v>0}$.
- This gives joint density of (Y_1, Y_2). Integrating out v we get the density of $Y_1 = X_1 + X_2$ which is $\int_0^u e^{-u} dv = ue^{-u}$

Something

- Given two measure spaces μ, ν on $(\Omega, \mathcal{F})$ if $\nu(A) = \int_A f d\mu$ then we say ν has a density f with respect to μ

Now the question is about **existence and uniqueness** of such densities given μ, ν on $(\Omega, \mathcal{F})$

- **Absolutely Continuous:** μ is absolutely continuous to ν or $\mu \ll \nu$ if $\mu(A) = 0$ whenever $\nu(A) = 0$
- **Mutually singular:** $\mu \perp \nu$ if there is a set $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$

We see absolute continuity is a necessary condition for the existence of density function.

Radon Nikodym theorem

Theorem

Suppose μ and ν are two finite measure spaces on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, then ν has a density with respect to μ.

Teaser

Independence

Conditional probability

Independence

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.

- Let $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k$ be sub-sigma algebras of $\mathcal{F}$. We say $\mathcal{G}_i$ are independent if for every $A_1 \in \mathcal{G}_1, \ldots, A_k \in \mathcal{G}_k$ we have $\mathbf{P}(A_1 \cap A_2 \cap \ldots \cap A_k) = \mathbf{P}(A_1) \ldots \mathbf{P}(A_k)$.
- Random variables $X_1, X_2, \ldots$ on $\mathcal{F}$ are said to be independent if $\sigma(X_1), \ldots, \sigma(X_n)$ are independent. $\sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}_{\mathbf{R}}\}$
- An arbitrary collection of σ-algebras $\mathcal{G}_{i \in I} \in \mathcal{F}$are said to be independent if every finite sub collection of them is independent. Same for random variables.

Conditional probability

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.

- Let X be a random variable which takes finitely many values $a_1, a_2, \ldots, a_k$ with $\mathbf{P}\{X = a_k\} > 0$. Then law of total probability says for $A \in \mathcal{F}$

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A|X = a_k)\mathbf{P}(X = a_k) \tag{5}$$

Where $\mathbf{P}(A|X = a_k) = \frac{\mathbf{P}(A \cap \{X = a_k\})}{\mathbf{P}(\{X = a_k\})}$

Now suppose X takes uncountably many values then we want to generalize

$$\mathbf{P}(A) = \int \mathbf{P}(A|X = t)d\mu_X(t) \tag{6}$$

Conditional probability

Fix $A \in \mathcal{F}$ and set $\nu_A(I) = \mathbf{P}\{A \cap \{X \in I\}\}$ for $I \in \mathcal{B}_{\mathbf{R}}$. Then ν is a probabilit measure on $\mathbf{R}$. Also $\nu(A) \ll \mu_X$ which allows to use Radon Nikodym theorem to get

$$\mathbf{P}(A \cap \{X \in I\}) = \int_I f_A(t) d\mu_X(t) \tag{7}$$

In particular $\mathbf{P}(A) = \int_{\mathbf{R}} f_A(t) d\mu_X(t)$. Then we may define $f_A(t)$ as the conditional probability $\mathbf{P}(A|X = t)$

Conditional probability

Let $\mathbf{P}, \mathbf{Q}$ be probability measures on $(\Omega, \mathcal{F})$ such that $\mathbf{Q} \ll \mathbf{P}$. Then there exists a $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ such that for $A \in \mathcal{F}$

$$\mathbf{Q}(A) = \int_A X d\mathbf{P} \tag{8}$$

Consider the restrictions $\mathbf{P}', \mathbf{Q}'$ to $\mathcal{G} \subseteq \mathcal{F}$. Naturally $\mathbf{Q}' \ll \mathbf{P}'$ and using Radon Nikodym theorem

$$\mathbf{Q}'(A) = \int_A X' d\mathbf{P}' = \mathbf{Q}(A) = \int_A X' d\mathbf{P} \tag{9}$$

Conditional probability

Now for any integrable random variable Y on $(\Omega, \mathcal{F}, \mathbf{P})$ we can similarly define using $Y = Y^+ - Y^-$ for any $\mathcal{G} \subseteq \mathcal{F}$

$$\int_A Y d\mathbf{P} = \int_A Y' d\mathbf{P} \tag{10}$$

This Y' is called the conditional expectation of Y with respect to $\mathcal{G}$ and denoted by $E[Y|\mathcal{G}]$. We note the fact that conditional expectation is a **Random variable** whereas Expectation is a **Number**.

First moment method or Markov's inequality

Markov's inequality:

Let $X \geq 0$ be a random variable. For any $t > 0$, we have $\mathbf{P}(X \geq t) \leq \frac{E[X]}{t}$.

Paley-Zygmund inequality:

For any non negative random variable X, and any $0 \leq \alpha \leq 1$, we have

$$\mathbf{P}(X > \alpha E[X]) \geq (1-\alpha)^2 \frac{E[X]^2}{E[X^2]} \tag{11}$$

Explanation

- If X has finite variance then $\mathbf{P}(|X - E[X]| \geq t) = \mathbf{P}((|X - E[X]|^2 \geq t^2) \leq \frac{var(X)}{t^2}$ which is called Chebyshev's inequality.
- Higher the moments exist we get better asymptotic bounds. If $E[e^{\lambda X}] < infty$ for $\lambda > 0$. We get $\mathbf{P}(X > t) = \mathbf{P}(e^{\lambda X} > e^{\lambda t}) \leq e^{-\lambda t} E[e^{\lambda X}]$
- An application of cauchy-schwartz gives $E[X1_{X > \alpha\mu}]^2 \leq E[X^2]E[1_{X>\alpha\mu}] = E[X^2]\mathbf{P}(X > \alpha\mu)$ and the result follows from $\mu = E[X1_{X<\alpha\mu}] + E[1_{X>\alpha\mu}] \leq \alpha\mu + E[1_{X>\alpha\mu}]$

$$\mathbf{P}(X > \alpha\mu) \leq \frac{E[X1_{X>\alpha\mu}]^2}{E[X^2]} \geq (1-\alpha)^2 \frac{E[X]^2}{E[X^2]} \tag{12}$$

Borel Cantelli Lemmas

If we have a sequence of events A_n on a common probability space,
$\limsup A_n = \{\omega \in \Omega |$set of all ω belonging to infinitely many events A_n $\}$

BC Lemmas

Let A_n be the events on a common probability space.

- If $\sum_n \mathbf{P}(A_n) < \infty$, then $\mathbf{P}(A_n -$ infinitely often$) = 0$.
- If A_n are independent and $\sum_n \mathbf{P}(A_n) = \infty$, then $\mathbf{P}(A_n -$ infinitely often$) = 1$

Explanation

- For any N, we have $\mathbf{P}(\cup_{n=N}^{\infty} A_n) \leq \sum_{n=N}^{\infty} \mathbf{P}(A_n)$, which goes to 0 as $N \to \infty$. Hence $\mathbf{P}(\limsup A_n) = 0$
- for $N < M$ we have $\mathbf{P}(\cup_{n=N}^{M} A_n) = 1 - \prod_{n=N}^{M} \mathbf{P}(A_n)$. Since $\sum_n \mathbf{P}(A_n) = \infty$ it follows that $\prod_{n=N}^{M} \mathbf{P}(A_n) \leq \prod_{n=N}^{M} e^{-\mathbf{P}(A_n)} \to 0$, for fixed N as $M \to \infty$.Therefore $\mathbf{P}(\cup_N^{\infty} A_n) = 1$ for all N. Then $\mathbf{P}(A_n i.o.) = 1$

Kolmogorov's Zero-one law

Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{F}_n$ be sub-σ-algebras of $\mathcal{F}$. Then the tail σ-algebra of the sequence $\mathcal{F}_n$ is defined to be $\mathcal{T} := \cap_n \sigma(\cup_{k \geq n} \mathcal{F}_k)$. For a sequence of random variables $X_1, X_2, \ldots$ the tail σ-algebra is also defined as the tail of the sequence $\sigma(X_n)$ denoted by $\mathcal{T}(X_1, X_2, \ldots)$

- If an event A is in the tail of $(X_k)_{k \geq 1}$ then $A \in \sigma(X_n, X_{n+1}, \ldots)$ for any n. That is the tail of the sequence is sufficient to tell you whether an event has occurred or not.Let A be the event that infinitely many heads.

...

Theorem

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space

- If $\mathcal{F}_n$ is a sequence of independent sub-*sigma*-algebras of $\mathcal{F}$ then the tail σ-algebra is trivial.
- If X_n are independent random variables, and A is a tail event, the $\mathbf{P}(A) = 0$ or $\mathbf{P}(A) = 1$

Bernstein/Hoeffding inequality

Since Chebyshev's inequality is rather weak in giving the bounds. Sergei Bernstein was first to exploit the full power of the Chebyshev's inequality and precise lemma was given by Hoeffding.

Hoeffding's inequality

Let $X_1, X_2, \ldots, X_n$ be independent sequence of random variables having zero mean. Assume that $|X_k| \leq a_k$ a.s. for some positive numbers a_k. Then $S = \sum_{k=0}^{n} X_k$ and $A = \sqrt{a_1^2 + \ldots + a_n^2}$, we have $\mathbf{P}(S \geq tA) \leq e^{-\frac{1}{2}t^2}$ for any $t > 0$.

Explanation

- Hoeffding inequality is a much better bound than $1/t^2$.
- we notice the bound $e^{-1/2t^2}$ which has a resemblance with the standard normal distribution.
- This is a non asymptotic statement showing S behaves in some ways like a Gaussian.
- If $|X_k| \leq a_k$ bout not necessarily zero mean, then we can apply Hoeffding's inequality to $Y_k = X_k - E[X_k]$. Since $|X_k| \leq a_k$ we have $|Y_k| \leq 2a_k$ we get slighly weaker result.

$$\mathbf{P}(S - E[S] \geq t\sqrt{a_1^2 + \ldots + a_n^2}) \leq e^{\frac{1}{8}t^2} \tag{13}$$

Lovasz's local lemma

We want to get lower bounds to the probability of events.one attempt to know about the happening of certain events(intersection)is the following. Since this can imply happening of certain other events.

Lemma

Let $A_1, A_2, \ldots, A_n$ be events in a common probability space. Assume each A_k is independent of all except atmost d of the other $A_i's$. Further that $\mathbf{P}(A_k) \geq 1 - p$ for all k. If $4dp < 1$, then $\mathbf{P}(A_1 \cap \ldots \cap A_n) \geq (1 - 2p)^n$.

proof:

Kolmogorov's Maximal inequality

Maximum of random variables can be much larger than any random variables. However the partial sums $S_1, S_2, \ldots$ are not independent. Kolmogorov found this amazing inequality

Lemma

Let X_n be independent andom variables with finite variance and $E[X_n] = 0$ for all n. Then $\mathbf{P}(\max_{k \leq n} |S_k| > t) \leq t^{-1/2} \sum_{k=1}^{n} var(X_k)$

We see this implies chebyshev's inequality and that gives an upperbound for one of the $|S_1|, \ldots, |S_n|$ is greater than or equal to t.

Applications

- **Coupon collector problem:** A shelf in IISERB library has n different books(n is very large). And you take out one book each day to read randomly and replace it in its position. How many days will it take to pick each book atleast once?

CCP

Let T_n be the number of days till each book is picked atleast once. Then T_n is concentrated around $n \log n$ in a window of size n by which we mean for any sequence of numbers $\theta_n \to \infty$, we have

$$\mathbf{P}(|T_n - nlogn| < n\theta_n) \to 1 \tag{14}$$

Branching process

Consider a Branching process which are i.i.d η. The process start with a single individual($Z_0 = 1$) and in subsequent generations each individual gives rise to number of offsprings with probability η. The process continues till there are individuals left.

- Random graphs
- Prime divisors

Weak law of large numbers:

Given a sequence of 100 coin tosses we expect that we will get 50 heads. Can we get some result that state precisely that it is unlikely that the number of heads is far from 50.

Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables. If $E[|X_1|] < \infty$ then for any $\delta > 0$ as $n \to \infty$ we have

$$\mathbf{P}(|\frac{1}{n}S_n - E[X_1]| > \delta) \to 0 \tag{15}$$

Application

Weirstrass approximation Theorem

The set of polynomials is dense in the space of continuous functions on an interval of real line with sup-norm metric.

Monte-carlo approximation

Accuracy of sample surveys

Explanation

- **Convergence in probability:** If Y_n be a sequence of random variables such that $\mathbf{P}(|Y_n - Y| \geq \delta) \to 0$ as $n \to \infty$ for every $\delta > 0$ then we say that Y_n converges to Y in probability($Y_n \xrightarrow{P} Y$). The WLLN precisely says this $\frac{1}{n}S_n \xrightarrow{P} E[X_1]$

Modes of convergence:

Let X_n, X be real valued random variables in a common probability sapce.

Definitons:

- **Almost sure convergence:** $X_n \xrightarrow{\text{a.s.}} X$ if $\mathbf{P}(\omega : \lim X_n(\omega) = X(\omega)) = 1$
- **In probability:**$X_n \xrightarrow{\text{P}} X$ if $\mathbf{P}(|X_n - X| > \delta) \to 0$ as $n \to \infty$ for any $\delta > 0$.
- L^p **convergence:**$X_n \xrightarrow{L^p} X$ if $E[|X_n - X|^p] \to 0$ for $0 < p \leq \infty$
- **Converges in distribution:**$X_n \xrightarrow{\text{d}} X$ if the distributions $F_{X_n} \to F_X$

Interrelations

$X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{P} X$

$$X_n \xrightarrow{P} X \Leftrightarrow \forall\delta > 0, \lim_{n\to\infty} \mathbf{P}(|X_n - X| > \delta) = 0 \tag{16}$$

$$X_n \xrightarrow{\text{a.s.}} X \Leftrightarrow \forall\delta > 0, \lim_{n\to\infty} \mathbf{P}(\limsup |X_n - X| > \delta) = 0 \tag{17}$$

We see that Fatou's lemma $\mathbf{P}(\limsup A_n) \geq \limsup \mathbf{P}(A_n)$ takes care of one direction. The reverse direction is not true in general but if we have fast enough convergence in probability.i.e. $\sum_n \mathbf{P}(|X_n - X| > \delta) < \infty$ then BC-lemma says $\mathbf{P}(|X_n - X| > \delta.i.o.) = 0$ which implies a.s. convergence.

Counter-example

If $X_n = 1$ with probability $1/n$ and 0 otherwise with probability $1 - 1/n$ then $\lim_{n\to\infty} \mathbf{P}(|X_n| > \epsilon) = \lim_{n\to\infty} \mathbf{P}(|X_n| = 1) = 0$

Distribution

Suppose X_n, X are defined on a common probability space then

- $X_n \xrightarrow{\text{P}}$ then $X_n \xrightarrow{\text{d}} X$
- $X_n \xrightarrow{\text{d}} X$ and X is constant a.s, then $X_n \xrightarrow{\text{P}} X$

L^p convergence

If $X_n \xrightarrow{L^p}$ which means $E[|X_n - X|^p] \to 0$ then convergence in probability is almost apparent from Markov's inequality.

$$\mathbf{P}(|X_n - X| > \delta) \leq \delta^{-p} E[|X_n - X|^p] \to 0 \tag{18}$$

Strong Law of Large numbers:

X_n are i.i.d with finite mean, then the weak law suggests that $\frac{S_n}{n} \xrightarrow{P} E[X_1]$. The strong law strengthens it to almost sure convergence.

Kolmogorov's SLLN

Let X_n be i.i.d with $E[|X_n|] < \infty$. Then $\frac{S_n}{n} \xrightarrow{a.s.} E[X_1]$. as $n \to \infty$.