1

X = random variable, $\mathbb{E}X$ = expected value of X

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

$M_X(t) = \mathbb{E}(e^{tX}),\ t \in \mathbb{R}$

moment generating function

$\mathbb{E}X^p$ = p'th moment

$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{\frac{1}{p}},\ p \in (0,\infty)$

$\|X\|_{L^\infty}$ = ess sup $|X|$.

probability space: $(\Omega, \Sigma, \mathbb{P})$

sample space, the set of all possible outcomes

set of events, an event being a set of outcomes in the sample space

A probability function which assigns a probability to each event.

$$\mathcal{L}^p(\Omega, \Sigma, \mathbb{P}) = \{ X : \|X\|_{\mathcal{L}^p} < \infty \}$$

If $p \in [1, \infty]$, $\|X\|_{\mathcal{L}^p}$ is a <u>norm</u>

and $\mathcal{L}^p$ is a <u>Banach space</u>

please look up what a "norm" and a "Banach space" is. When you do, you may encounter other unfamiliar notions, so you just keep scrambling!

The exponent $p=2$ is special because in this case, $\mathcal{L}^2$ is not only a Banach space, but also a <u>Hilbert space</u>

please look it up!

The <u>inner product</u> on $\mathcal{L}^2$ is given by

$$\langle X, Y \rangle_{\mathcal{L}^2} = \mathbb{E} XY, \quad \|X\|_{\mathcal{L}^2} = \left(\mathbb{E}|X|^2 \right)^{\frac{1}{2}}$$

③

$$\underset{\substack{\shortparallel \\ \text{standard} \\ \text{deviation}}}{\sigma(X)} = \|X - \mathbb{E}X\|_{L^2} = \sqrt{Var(X)}$$

$$\underset{\substack{\text{///} \\ \text{covariance of} \\ X \,\&\, Y}}{cov(X,Y)} = \mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big) = \langle X-\mathbb{E}X, Y-\mathbb{E}Y \rangle_{L^2}$$

gives a geometric interpretation of covariance:

the more $X - \mathbb{E}X$ and $Y - \mathbb{E}Y$ are alligned, the larger their inner product and covariance.

We now discuss some inequalities.

Definition: We say that a function φ is convex if

$$\varphi(\lambda x + (1-\lambda) y) \leq \lambda \varphi(x) + (1-\lambda)\varphi(y) \quad \forall \lambda \in [0,1]$$

and all $x, y \in \text{domain}(\varphi)$.

Jensen's inequality: X random variable and φ convex, then

$$\varphi(\mathbb{E} X) \leq \mathbb{E} \varphi(X)$$

to be proved later this week

Corollary: If $0 \leq p \leq q \leq \infty$,

$$\|X\|_{L^p} \leq \|X\|_{L^q}$$

Work it out!!

Minkowski: $\|X+Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p}$

if $p \in [1, \infty]$.

to be proved later this week

(5)

Cauchy-Schwarz:

$$|\mathbb{E}XY| \le \|X\|_{L^2} \|Y\|_{L^2}$$

Hölder: $p, q \in (1, \infty)$ $\frac{1}{p} + \frac{1}{q} = 1$, then

$$|\mathbb{E}XY| \le \|X\|_{L^p} \|Y\|_{L^q}$$

Cumulative distribution function (CDF)

$$F_X(t) = \mathbb{P}\{X \le t\}, \quad t \in \mathbb{R}$$

We shall often work w/ $\mathbb{P}\{X > t\} = 1 - F_X(t)$

Key identity: X non-negative random variable. Then

$$\mathbb{E}X = \int_0^\infty \mathbb{P}\{X > t\}\, dt$$

Proof: $\mathbb{E}X = \mathbb{E} \int_0^\infty 1_{\{t<X\}}(t)\, dt$, where

$$1_{\{t<X\}}(t) = \begin{cases} 1, & t < X \\ 0, & \text{otherwise} \end{cases}$$

(6)

Why is that? Because

$$x = \int_0^x 1\, dt = \int_0^\infty 1_{\{t<x\}}(t)\, dt$$

Going back,

$$\mathbb{E}X = \mathbb{E}\int_0^\infty 1_{\{t<X\}}(t)\, dt$$

$$= \int_0^\infty \mathbb{E}\, 1_{\{t<X\}}(t)\,dt = \int_0^\infty \mathbb{P}\{t<X\}\,dt.$$

Changing the order of integration and expectation requires proof. The key result is the Fubini-Tonelli theorem.

Markov inequality: X non-negative random variable and $t>0$,

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}X}{t}$$

we shall explore this thoroughly using python

Proof of Markov:

Fix $t > 0$. Write $X = X\,\mathbb{1}_{\{X \geq t\}} + X\,\mathbb{1}_{\{X < t\}}$

It follows that

$$\mathbb{E}X = \mathbb{E}X\,\mathbb{1}_{\{X \geq t\}} + \mathbb{E}X\,\mathbb{1}_{\{X<t\}}$$

$$\geq \mathbb{E}\,t\,\mathbb{1}_{\{X \geq t\}} = t\,\mathbb{P}\{X \geq t\}$$

$+ 0$

corresponding to $\mathbb{E}X\,\mathbb{1}_{\{X<t\}}$

We conclude that

$$\mathbb{P}\{X \geq t\} \leq \frac{1}{t}\mathbb{E}X, \text{ as claimed.}$$

Corollary: (Chebyshev) Let X be a random variable w/ mean μ and variance σ^2. Then for any $t > 0$,

$$\mathbb{P}\{|X-\mu| \geq t\} \leq \frac{\sigma^2}{t^2}$$

Please don't take my word for it and deduce Chebyshev from Markov.

Limit theorems:

We know that $Var(X_1 + \dots + X_N) = Var(X_1) + \dots + Var(X_N)$

if X_js are independent (check!)

still independent

If $\tilde{X_i}$s have the same distribution w/ mean μ and variance σ^2, then

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{\sigma^2}{N} \xrightarrow[N \to \infty]{} 0$$

In other words, the variance of the sample mean $\frac{1}{N}\sum_{i=1}^{N} X_i$ of the sample $\{X_1, \dots, X_N\}$ shrinks to 0 as $N \to \infty$.

This suggests that, for large N, the sample mean should concentrate around the expected value μ. The quantitative version of this statement is known as the

Strong Law of Large Numbers:

Let $X_1, X_2, \dots, X_N$ be independent identically distributed random variables w/ mean μ. Consider the sum

$$S_N = X_1 + X_2 + \dots + X_N$$

Then as $N \to \infty$, $\frac{S_N}{N} \to \mu$ almost surely.

Borel Law of Large Numbers:

Consider random variables $X_1, X_2, \ldots, X_n, \ldots$ independent, identically distributed and assume the values 0 & 1 w/ probability $\frac{1}{2}$.

Let $S_n = \sum_{k=1}^{n} X_k$. Then $\frac{S_n}{n} \xrightarrow{n \to \infty} \frac{1}{2}$

w/ probability 1

Hardy and Littlewood improved this result to

$$\limsup_{n \to \infty} \frac{\left| \frac{S_n}{n} - \frac{1}{2} \right|}{\sqrt{n \log(n)}} < \frac{1}{\sqrt{2}} \qquad (1914)$$

and Khinchin improved this to

$$\text{Prob}\left[\limsup_{n \to \infty} \frac{\left| \frac{S_n}{n} - \frac{1}{2} \right|}{\sqrt{n \log \log n}} \right] = 1$$

It might be fun to explore these numerically!

(10)

Another absolutely fundamental limit result is the Lindeberg-Lévy Central Limit Theorem:

Consider the standard normal distribution, denoted by $N(0,1)$, w/ density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}$$

Let $X_1, X_2, \dots, X_N, \dots$ be a sequence of independent identically distributed random variables. Let

$S_N = X_1 + \dots + X_N$ and normalize it, to obtain a random variable w/ zero mean and unit variance:

$$Z_N = \frac{S_N - \mathbb{E}S_N}{\sqrt{Var(S_N)}} = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} (X_i - \mu)$$

Then, as $N \to \infty$,

$$Z_N \to N(0,1) \text{ in distribution.}$$

i.e convergence of CDFs.

Interesting special case: $X_1, X_2, \ldots, X_N, \ldots$ independent, identically distributed random variables taking on values 0 & 1 w/ probability $(1-p)$ & p respectively (Bernoulli random variables)

Observe that $\mathbb{E} X_i = p$ and $\mathrm{Var}(X_i) = p(1-p)$

please make these calculations!

The Central Limit Theorem implies that

if $S_N = X_1 + \ldots + X_N$, as before, then

$$\frac{S_N - Np}{\sqrt{Np(1-p)}} \longrightarrow N(0,1) \text{ in distribution}$$

De Moivre-Laplace theorem

(18th century!!)

(12)

Life is not always this "easy". Suppose that X_i takes on 1 w/ probability p_i w/

$$p_i \xrightarrow[N \to \infty]{} 0 \quad \text{so fast that}$$

Ber(p)

S_N has mean $O(1)$ instead of $\sim N$.

Then the Central Limit Theorem fails! But what is true?

Definition: Z has a poisson distribution w/ parameter λ

random variable

$Z \sim \mathrm{Pois}(\lambda)$, if it takes values in $\{0, 1, 2, \dots\}$ w/ probability

$$\mathbb{P}\{Z = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Theorem: $X_{N,i}$, $1 \le i \le N$ independent random variables $X_{N,i} \sim \mathrm{Ber}(p_{N,i})$ and let $S_N = \sum_{i=1}^{N} X_{N,i}$. Assume that as $N \to \infty$,

$$\max_{i \le N} p_{N,i} \to 0 \text{ and } \mathbb{E} S_N = \sum_{i=1}^{N} p_{N,i} \to \lambda < \infty.$$

13

Then as $N \to \infty$,

$$S_N \to \mathrm{Pois}(\lambda) \text{ in distribution.}$$