EE6106 (Lecture 5) (Date: 30 Jan 2024)

Convergence of random sequences

$\{X_n\}_{n\geq 1}$ ~ seq of random variables

- $X_n \rightarrow X$ almost surely (w.p. 1)

  ($X_n \xrightarrow{a.s.} X$) if

$$P(\lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$$

- $X_n \rightarrow X$ in probability

  ($X_n \xrightarrow{P} X$) if

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \quad \forall\, \varepsilon > 0$$

this is an "Event" since it is the difference of two random variables

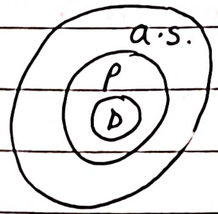
- $X_n \rightarrow X$ in distribution (weak convergence)

  ($X_n \xrightarrow{D} X$) if

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x)$$

at all points of continuousness x of $P(X\leq x)$

Note: $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$



$A_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\}$

$X_n \xrightarrow{P} X \iff \lim_{n\to\infty} P(A_n(\varepsilon)) = 0 \quad \forall\, \varepsilon > 0$

$X_n \xrightarrow{a.s.} X \iff \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} A_m(\varepsilon)\right) = 0 \quad \forall\, \varepsilon > 0$

The probability of ever having an ε-deviation in the future goes to zero.

Limit theorems

$$S_0 = 0$$
$$S_n = \sum_{1}^{n} X_i \quad (n \geq 1)$$

S_n is thus the random sum

Strong Law of Large Numbers (SLLN)

If $\{X_i\}$ are iid with finite mean μ,

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu$$

i.e. $P\left(\lim_{n \to \infty} \frac{S_n(\omega)}{n} = \mu(\omega)\right) = 1$

Weak Law of Large Numbers (WLLN)

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

i.e. $\lim_{n \to \infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0$

SLLN : running average will become very close to the expected average
S_n grows linearly with n (generally), with a slope equal to μ

$$S_n \approx n\mu + O(n)$$

$\downarrow$

we do not know the "nature" of sublinearity?

$\downarrow$

we do, under certain assumptions $\Rightarrow \sqrt{n}$

Central Limit Theorem (CLT)

If X_i are i.i.d. with finite mean μ, finite variance σ^2,

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{P} Z \rightarrow \mathcal{N}(0,1)$$

$\downarrow$

standard Gaussian

Date:

Moral : $S_n \approx n\mu + \sqrt{n}Z + o(\sqrt{n})$

$Z \to \mathcal{N}(0,\sigma^2) \to$ here the variance is σ^2 instead of 1

Thus, the typical deviations between S_n (running sum) & $n\mu$ are of the order $O(\sqrt{n})$.

Suppose we ask,

what is $\lim_{n\to\infty} \mathbb{P}(0 \leq S_n - n\mu \leq \sqrt{n})$ or otherwise

$$\lim_{n\to\infty} \mathbb{P}(a\sqrt{n} \leq S_n - n\mu \leq b\sqrt{n}) ?$$

A: This is simply $\mathbb{P}(a \leq Z \leq b)$, where Z is $\mathcal{N}(0,\sigma^2)$

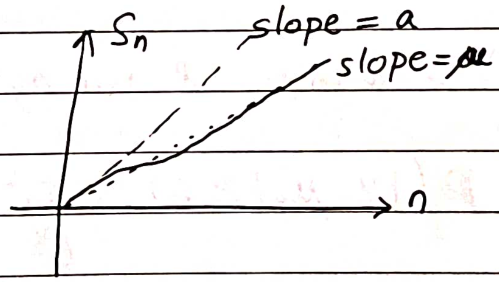
* The CLT thus says that "typical" deviations of S_n from $n\mu$ are $O(\sqrt{n})$.

Q: $P(S_n > na)$ ↳ $a > \mu$

↓

Order of "n" deviation

↳ what is the probability (large deviations)



Claim 1 : Goes to 0 as $n \to \infty$

Trivial consequence of the WLLN

$$P(S_n > na) = \mathbb{P}\left(\frac{S_n}{n} > a\right) = \mathbb{P}\left(\frac{S_n}{n} - \mu > \underbrace{a-\mu}_{>0}\right) = 0 \quad (\text{by WLLN})$$

CONCENTRATION INEQUALITIES

Markov inequality

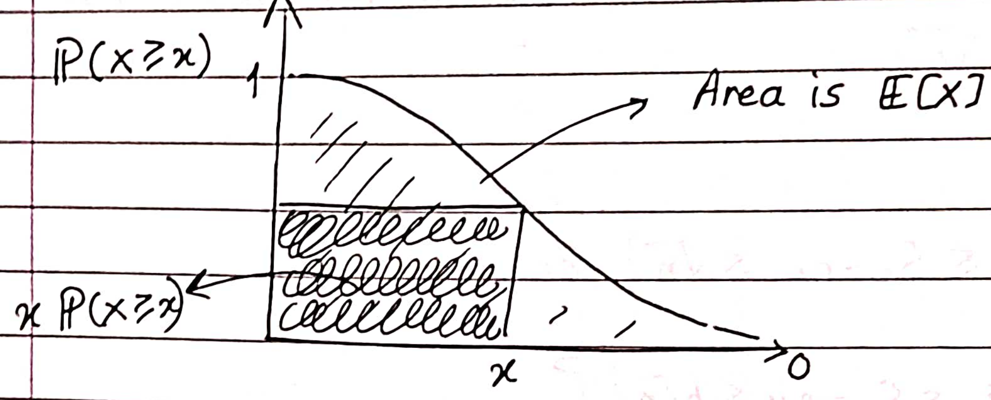
If X is a non-negative RV, $x > 0$

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[X]}{x}$$

Inequality becomes useful only when $x \gg \mathbb{E}[X]$ by a large margin.

For a non-negative RV,

$$\mathbb{E}[X] = \int_0^\infty P(X \geq t)\,dt$$



Then $x\,\mathbb{P}(X \geq x) \leq \mathbb{E}[X]$

Typically, we apply a monotone transformation to X & then apply Markov's inequality to both sides.

Corollary $Y \sim$ mean μ_Y & var σ_Y^2, $t > 0$

$$\mathbb{P}(|Y-\mu_Y| \geq t) \leq \frac{\sigma_Y^2}{t^2}$$ ... CHEBYSHEV INEQUALITY

Proof

$$\mathbb{P}(|Y-\mu_Y| \geq t) = \mathbb{P}(|Y-\mu_Y|^2 \geq t^2) \underset{M.I.}{\leq} \frac{\mathbb{E}[|Y-\mu_Y|^2]}{t^2}$$

Markov: true when mean is finite but get $1/t$ bound on tail
Chebyshev: variance must also be finite but get $1/t^2$ bound on tail

$$\mathbb{P}(Y > y) = \mathbb{P}(e^{sY} > e^{sy}) \leq \mathbb{E}[e^{sY}]e^{-sy}$$ ... CHERNOFF INEQUALITY

for $s > 0$ to ensure monotonic increase of e^x

Moment Generating Function

For a r.v. Y,

$$MGF_Y(s) = M_Y(s) = \mathbb{E}[e^{sY}]$$

Chernoff Bound :

$$\mathbb{P}(Y > y) \leq \inf_{s>0} \left[M_Y(s) e^{-sy}\right]$$

it might also be possible to find the explicit value of s as well.
The Chernoff Bound can also give an exponential bound
For Chernoff bound to hold, we need $\underbrace{M_Y(s) < \infty \text{ for some } s > 0}$

Y is 'LIGHT-TAILED'

also called SUB-EXPONENTIALITY

otherwise

Y is 'HEAVY-TAILED'

Q : Can we use the Chernoff bound to talk about $\mathbb{P}(S_n > na)$
Note: Using Markov inequality yields $\mathbb{P}(S_n > na) \leq \frac{\mu}{a}$

Assume $\{X_n\}_{n \geq 1} \sim$ iid, mean μ

$$\mathbb{P}(S_n > na) \underset{(s>0)}{\leq} M_{S_n}(s) e^{-sna} \qquad \text{(Chernoff bound)}$$

$a > \mu$

$$M_{S_n}(s) = \mathbb{E}[e^{s(X_1 + \cdots + X_n)}] = (M_X(s))^n \qquad \ldots \text{(iid)}$$

$$\therefore \mathbb{P}(S_n > na) \underset{s>0}{\leq} [M_X(s)]^n e^{-sna}$$

Define $\Lambda_X(s) = \log M_X(s)$

log MGF

$$\mathbb{P}(S_n > na) \leq e^{n\Lambda_X(s) - sna} = e^{-n[sa - \Lambda_X(s)]}$$

$$\Rightarrow \mathbb{P}(S_n > na) \leq e^{-n\left[\sup_{s>0}(sa - \Lambda_X(s))\right]}$$

as long as $\sup_{s>0}(sa - \Lambda_X(s))$ is positive
we can get an exponentially decaying term

M T W T F S S

Page No.:

Date:

YOUVA

★ For light tailed x, with $a > \mu$ it turns out that

$$\sup_{s>0} \left[sa - \Lambda_x(s) \right] > 0$$

It also turns out that this is the tightest bound possible. Thus the upper bound is asymptotically tight (Cramer's Theorem). Depending on the distribution of x, we can get the "customized" bounds. However, this is a downside when it comes to learning algorithms because now the distribution, and so, bounds are not available.