m

Lecture 3: Reading Sections 1.6–1.7. ①

Subject : Date:

Recall Markov's Inequality

Lem: Let Y be a non-negative r.v. with mean $\mathbb{E}Y$. Then for every $y>0$,

$$\Pr(Y \geq y) \leq \frac{\mathbb{E}Y}{y}.$$

Pf: $$\mathbb{1}\{Y \geq y\} \leq \frac{Y}{y} \quad \forall y>0.$$

Take Expectations on both sides $\Rightarrow \mathbb{E}\,\mathbb{1}\{Y\geq y\} \leq \frac{\mathbb{E}Y}{y} \Rightarrow \Pr(Y\geq y) \leq \frac{\mathbb{E}Y}{y}$.

Chebyshev's Inequality

Lem: Let Z be a rv with finite mean $\mathbb{E}Z$ & finite variance $Var(Z)=\sigma_Z^2$. Then

$$\Pr\left((Z-\mathbb{E}Z)^2 \geq y\right) \leq \frac{\sigma_Z^2}{y} \quad \forall y>0.$$

Rmk: What's the prob that Z deviates from its mean by $\geq c$ standard deviations for some $c>0$?

$$\Pr\left(|Z-\mathbb{E}Z| > c\sigma_Z\right) = \Pr\left((Z-\mathbb{E}Z)^2 > c^2\sigma_Z^2\right) \leq \frac{1}{c^2}$$

Rmk: Chebyshev's ineq. applies to rv with finite mean & finite variance.

Markov's inequality applies to only non-negative rvs with only finite mean (i.e., the variance could be ∞)

Pf: Apply Markov's ineq to the non-negative r.v. $(Z-\mathbb{E}Z)^2$, i.e.,

$$\Pr\left((Z-\mathbb{E}Z)^2 \geq y\right) \leq \frac{\mathbb{E}[(Z-\mathbb{E}Z)^2]}{y} = \frac{Var(Z)}{y} = \frac{\sigma_Z^2}{y}$$

as desired.

Subject: Date:

Chernoff bounds.

Let $X_1, X_2, \dots, X_n$ be i.i.d. rvs with mean 0 & variance σ^2. Fix $\varepsilon > 0$.

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \varepsilon\right) = \Pr\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 > \varepsilon^2\right) \le \frac{\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right]}{\varepsilon^2}$$

$$= \frac{\sigma^2}{n\varepsilon^2}$$

If X_i's each have variance $\sigma^2 = 1/4$ & $\varepsilon = 0.01$, it takes $10^6 \cdot \frac{1}{4}$ $= 250{,}000$ samples to get $\Pr(\dots) \le 0.01$.

Do we need this # of samples to get $\Pr(\dots) \le 0.01$?

MGF of a rv X: $g_X(r) = \mathbb{E}[e^{rX}]$, $r \in \mathbb{R}$.

Assume that $g_X(r)$ exists for all r in a nbd of 0.

Note that $g_X(0) = 1$.
$g_X'(r) = \mathbb{E}[X e^{rX}]$, $g''(r) = \mathbb{E}[X^2 e^{rX}], \dots, g^{(k)}(r) = \mathbb{E}[X^k e^{rX}]$.
& $g^{(k)}(0) = \mathbb{E}[X^k]$
Thus the function g generates the moments $\mathbb{E}[X^k]$ of X.

Markov's inequality immediately gives

$$\Pr(e^{rZ} \ge y) \le \frac{\mathbb{E}[e^{rZ}]}{y} = \frac{1}{y} g_Z(r).$$

Replace y by $e^{rb} \Rightarrow \underbrace{\Pr(e^{rZ} \ge e^{rb})}_{\Pr(Z \ge b)} \le \exp(-rb)\, g_Z(r) \quad \forall r > 0.$

In the other direction, for $r \le 0$,
$$\Pr(Z \le b) = \Pr(e^{rZ} \ge e^{rb}) \le \exp(-rb)\, g_Z(r), \quad r \le 0.$$

Subject:

Date:

Going back to the original problem, (One-sided version)

Fix $r>0$. $P\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \varepsilon\right) = P\left(\sum_{i=1}^{n} X_i > n\varepsilon\right)$

$= P\left(\exp\left(r\sum_{i=1}^{n} X_i\right) \geq \exp(nr\varepsilon)\right) \leq E\left[\exp\left(r\sum_{i=1}^{n} X_i\right)\right]\exp(-nr\varepsilon)$

$= \left(g_X(r)\right)^n \exp(-nr\varepsilon)$

Define the semi-invariant MGF (also called cumulant generating function)

$\gamma_X(r) := \ln g_X(r)$.

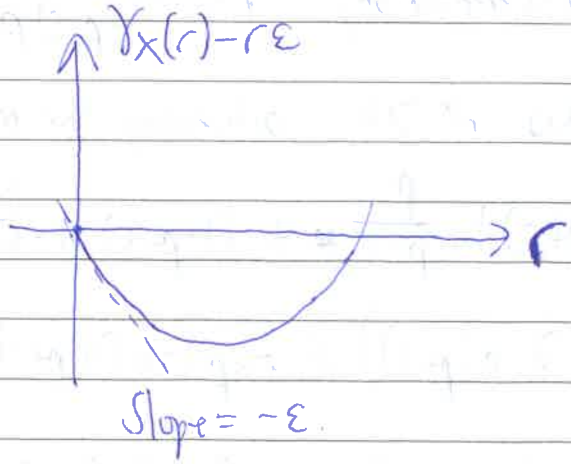
$\Rightarrow P\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \varepsilon\right) \leq \exp\left(n(\gamma_X(r) - r\varepsilon)\right)$

This bound is <u>exponential</u> in n for fixed ε & fixed r.

To obtain the smallest upper bd, we minimize the upper bd (or the exponent $\gamma_X(r) - r\varepsilon$ over all $r>0$)

Note that $\gamma_X(r) - r\varepsilon\Big|_{r=0} = 0$, $\frac{d}{dr}(\gamma_X(r) - r\varepsilon) = EX - \varepsilon = -\varepsilon < 0$.

This means that $r \mapsto \gamma_X(r) - r\varepsilon$ must be negative for sufficiently small r.



$\gamma_X(0) = 0$
$\gamma_X'(0) = EX$
$\gamma_X''(0) = Var(X)$

Subject : Date:

Thm: (Chernoff bound) Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. r.v.s & let $S_n = X_1 + \dots + X_n$. Assume MGF $g_X(r)$ exists $\forall r \in \mathbb{R}$. Then

$$\Pr(S_n \geq n\varepsilon) \leq \exp(n\mu_X(\varepsilon)) \quad \text{where}$$

$$\mu_X(\varepsilon) = \inf_{r \geq 0} \{\gamma_X(r) - r\varepsilon\}$$

Furthermore $\mu_X(\varepsilon) < 0$ for $\varepsilon > \mathbb{E}X$ & $\mu_X(\varepsilon) = 0$ $\forall \varepsilon \leq \mathbb{E}X$.

Ex: $X \in \{0,1\}$ $\Pr(X=1) = p$, $\Pr(X=0) = 1-p = q$.

$g_X(r) = q + pe^r$, $\forall r \in \mathbb{R}$, $\gamma_X(r) = \ln(q + pe^r)$

Consider $\Pr[S_n \geq n(p+\varepsilon)] = \Pr\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq p+\varepsilon\right)$;

According to the Chernoff bound, this probability

$$\leq \exp(n\mu_X(p+\varepsilon)) \quad \text{where} \quad \mu_X(p+\varepsilon) = \inf_{r \geq 0}\{\gamma_X(r) - r(p+\varepsilon)\}$$

$$= \inf_{r \geq 0}\{\ln(q + pe^r) - r(p+\varepsilon)\}.$$

After some differentiation, $r^* = \ln \frac{(p+\varepsilon)(1-p)}{(1-p-\varepsilon)p}$.

For $\varepsilon > 0$, this $r^* > 0$, achieving the min over $r \geq 0$.

$$\mu_X(p+\varepsilon) = (p+\varepsilon)\ln\frac{p}{p+\varepsilon} + (1-p-\varepsilon)\ln\frac{1-p}{1-p-\varepsilon} = -D(p+\varepsilon \| p)$$

$$\Rightarrow \Pr\left(\sum_{i=1}^{n} X_i \geq n(p+\varepsilon)\right) \leq \exp(-nD(p+\varepsilon\|p))$$

If we want this to be ≤ 0.005 & $\varepsilon = 0.01$, $p = q = \frac{1}{2}$ so $\operatorname{Var} X = \frac{1}{4}$, we need $n \geq (\ln 200)/D(p+\varepsilon\| p) \approx 263$ samples.

Subject: Date:

Laws of Large Numbers.

Let $\{b_i\}_{i=1}^{\infty} \subset \mathbb{R}$ be a sequence of real numbers. We say that $b_i \to b$ as $i \to \infty$ if, $\forall \varepsilon > 0$, $\exists\, n_0 = n_0(\varepsilon) \in \mathbb{N}$ s.t.

$$|b_i - b| < \varepsilon \quad \forall\, i \geq n_0.$$

Random variables X_n are functions from the sample space Ω to the reals $\mathbb{R}$ (i.e., $X_n(\omega) \in \mathbb{R}$). Hence there are many ways a sequence of rvs (X_n) can converge to a limiting rv (function)

Weak law of large numbers.

Thm: $\forall n \geq 1$, let $S_n = X_1 + \cdots + X_n$ be the sum of n i.i.d. rvs each with finite variance $\sigma^2 < \infty$. Then $\forall\, \varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(\left| \frac{1}{n} S_n - \mathbb{E}X \right| \geq \varepsilon \right) = 0$$

We say that the empirical mean $\frac{1}{n} S_n$ converges in probability to $\mathbb{E}X$.

We write $\frac{1}{n} S_n \xrightarrow{P} \mathbb{E}X$ as $n \to \infty$.

Pf. By Chebyshev's inequality,

$$\Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathbb{E}X) \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

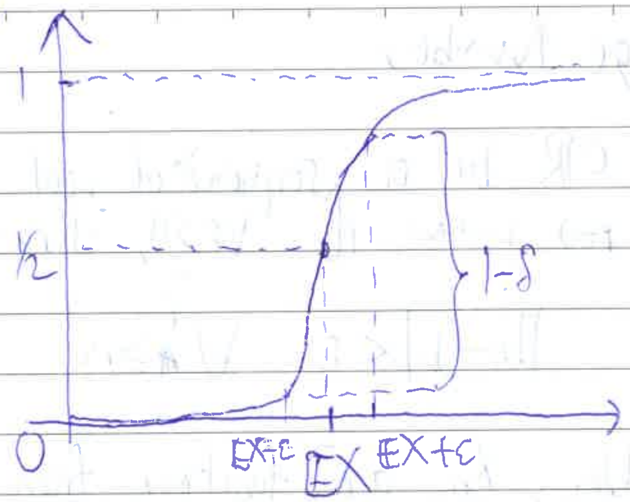
Note that

$$\Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathbb{E}X) \right| > \varepsilon \right) = F_{S_n/n}(\mathbb{E}X + \varepsilon) - F_{S_n/n}(\mathbb{E}X - \varepsilon)$$

$\simeq 1 - \delta$ when n is large enough

Subject : $F_{S_n/n}(\cdot)$

Date:



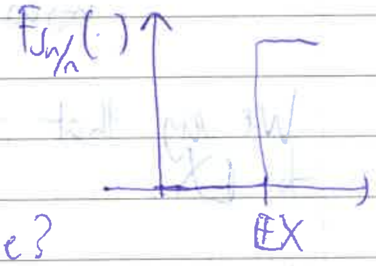
The bound of $O(1/n)$ is extremely loose in practice.

In fact, if the MGF of X exists,

$$\Pr\left(\left|\frac{1}{n}S_n - \mathbb{E}X\right| > \varepsilon\right) \le \exp\big(n\underbrace{\mu_X(\mathbb{E}X+\varepsilon)}_{-ve}\big) + \exp\big(n\underbrace{\mu_X(\mathbb{E}X-\varepsilon)}_{-ve}\big).$$

Hence, under the condition of existence of MGF, $P\left(\left|\frac{1}{n}S_n - \mathbb{E}X\right| > \varepsilon\right)$ converges to 0 exponentially fast.

The central limit theorem.



What does $F_{S_n/n}$ look like when n is large?
A step function (jump) at $\mathbb{E}X$ from 0 to 1. Not interesting.

Consider a different scaling. $Z_n := \frac{1}{\sqrt{n}\,\sigma}\sum_{i=1}^{n}(X_i - \mathbb{E}X)$.

Note that we divide the sum by $\sqrt{n}$ (instead of n). Normalize by σ after subtracting the mean.

Clearly $\mathbb{E}Z_n = 0$, $\mathrm{Var}(Z_n) = 1$. $\forall n \in \mathbb{N}$

Subject :

Date:

Thm: $\{X_i\}_{i=1}^{\infty}$ sequence of i.i.d. rvs with finite mean $\mathbb{E}X$ and finite variance σ^2. Then $\forall z \in \mathbb{R}$

$$\lim_{n\to\infty} P_r(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Rmk: We ~~say~~ say that $Z_n \xrightarrow{d} N(0,1)$ as $n\to\infty$

Proof sketch of CLT. Assume $\mathbb{E}X=0$, $\mathrm{Var}\,X=1$.

$$\mathbb{E}[e^{rZ_n}] = \mathbb{E}\left[e^{r\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i}\right] = \left(g_X\left(\frac{r}{\sqrt{n}}\right)\right)^n.$$

Consider $\frac{1}{n}\log \mathbb{E}[\exp(rZ_n)] = \log g_X(r/\sqrt{n})$

Taylor expansion: $g_X(r/n) = g_X(0) + g_X'(0)\, r/\sqrt{n} + \frac{g_X''(0)}{2} r^2/n + \cdots$

$$= 1 + r^2/n.$$

$$\Rightarrow \frac{1}{n}\log \mathbb{E}[\exp(rZ_n)] = \log(1 + r^2/n + \cdots) \approx r^2/n.$$

$$\Rightarrow \mathbb{E}[e^{rZ_n}] \approx \exp(r^2/2) \quad \text{as } n\to\infty.$$

But $\exp(r^2/2)$ is the MGF of a standard Gaussian so Z_n converges in distribution to a standard Gaussian.

Def: A sequence of rvs $\{Z_i\}_{i=1}^{\infty}$ converges in distribution to Z if

$$\lim_{n\to\infty} F_{Z_n}(z) = F_Z(z) \quad \forall z \text{ s.t. } F_Z(z) \text{ is continuous at } z.$$

By CLT, $Z_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}X) \xrightarrow{d} Z \sim N(0,1)$

Subject :

Date:

Def: A sequence of rvs $\{Z_i\}$ converges in prob. to Z if

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P\left(|Z_n - Z| > \varepsilon\right) = 0.$$

Eg: $\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to EX$ as $n\to\infty$.

Fact: Convergence in prob $\Rightarrow$ Convergence in distribution.

Def: A sequence of rvs $\{Z_i\}_{i=1}^{\infty}$ converges to Z in mean square if

$$\lim_{n\to\infty} E\left[(Z_n - Z)^2\right] = 0.$$

Eg: $\frac{S_n}{n} \to EX$ in MS.

Pf. $E\left[\left(\frac{S_n}{n} - EX\right)^2\right] = Var\left(\frac{S_n}{n}\right) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} Var\left(\sum X_i\right)$

$= \frac{1}{n^2}\sum_{i=1}^{n} Var\, X_i = \frac{\sigma^2}{n} \to 0$ as desired.

Def: $\{Z_i\}_{i=1}^{\infty}$ converges to Z with probability 1 (almost surely) if

$$P\left(\left\{\omega\in\Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\right\}\right) = 1.$$

We write this as $Z_n \xrightarrow{a.s} Z$ or $Z_n \xrightarrow{wp\,1} Z$.

Fact: $Z_n \to Z$ w.p. 1 if $\forall \varepsilon > 0$,

$$\lim_{m\to\infty} P\left(|Z_n - Z| < \varepsilon \text{ for all } n \geq m\right) = 1.$$

$$\Longleftrightarrow \quad P\left(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\{|Z_n - Z| < \varepsilon\}\right) = 1.$$

w.p. 1 convergence means that the set of sample paths $\omega\in\Omega$ that converge to $Z(\omega)$ (in the sense of a seq. converging to a limit) has probability 1.

Subject:

Date:

Ex: $X_n = \begin{cases} 0 & \text{w.p. } 1-1/n \\ 1 & \text{w.p. } 1/n \end{cases}$ X_n are independent

$\mathbb{E}\, X_n^2 = 1 \cdot \frac{1}{n} + 0 \cdot (1 - 1/n) = \frac{1}{n} \to 0.$

$\Rightarrow$ $X_n \to 0$ in mean square

$\Pr(|X_n - 0| > \epsilon) = \Pr(X_n > \epsilon) = \frac{1}{n} \to 0. \Rightarrow X_n \to 0$ in prob.

Does it converge to 0 w.p. 1?

$$P(|X_n - 0| < \varepsilon \ \forall n \geq m) = \lim_{n\to\infty} \prod_{i=m}^{n} \left(1 - \frac{1}{i}\right)$$

$$= \lim_{n\to\infty} \prod_{i=m}^{n} \left(\frac{i-1}{i}\right)$$

$$= \lim_{n\to\infty} \frac{m-1}{m} \cdot \frac{m}{m+1} \cdots \frac{n-1}{n} = \lim_{n\to\infty} \frac{m-1}{n} = 0.$$

Thus, X_n does not converge to 0 almost surely.

Subject : Date:

Intuitive difference btw convergence w.p. 1 and convergence in prob.

Recall $X_n \xrightarrow{p} X \iff \forall \epsilon > 0 \quad \lim_{n\to\infty} P(\{w \in \Omega : |X_n(w) - X(w)| \leq \epsilon) = 1.$

In contrast to a.s. convergence, i.e.,

$$P(\{w \in \Omega : \lim_{n\to\infty} X_n(w) = X(w)\}) = 1.$$

Big difference is that for a.s. convergence, limit is inside the probability.

Committee with 100 people. Each person is an $w \in \Omega = \{1, \dots, 100\}$. There is one meeting of the committee everyday throughout the year. We want to know the attendance of the committee.

Convergence a.s. $\iff$ Almost all members have perfect attendance.
~~$\iff$ Each meeting must be almost full~~

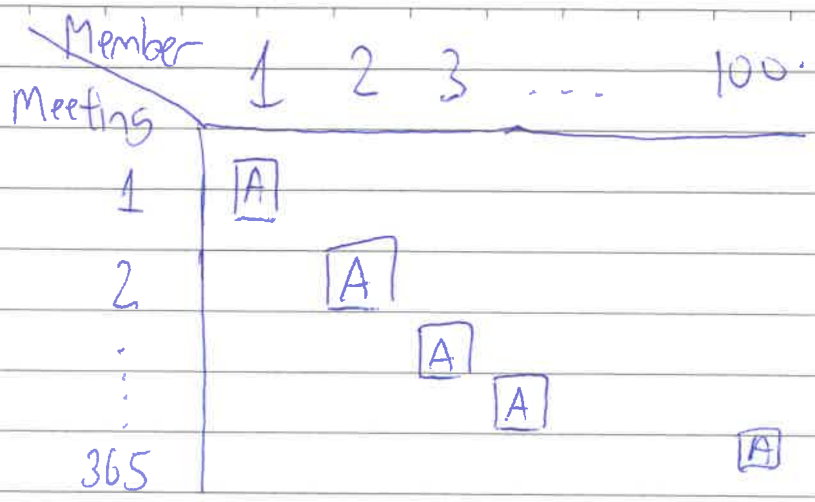
Convergence in probability $\iff$ All meetings were almost full.

If almost all members have perfect attendance, then each meeting must be almost full.

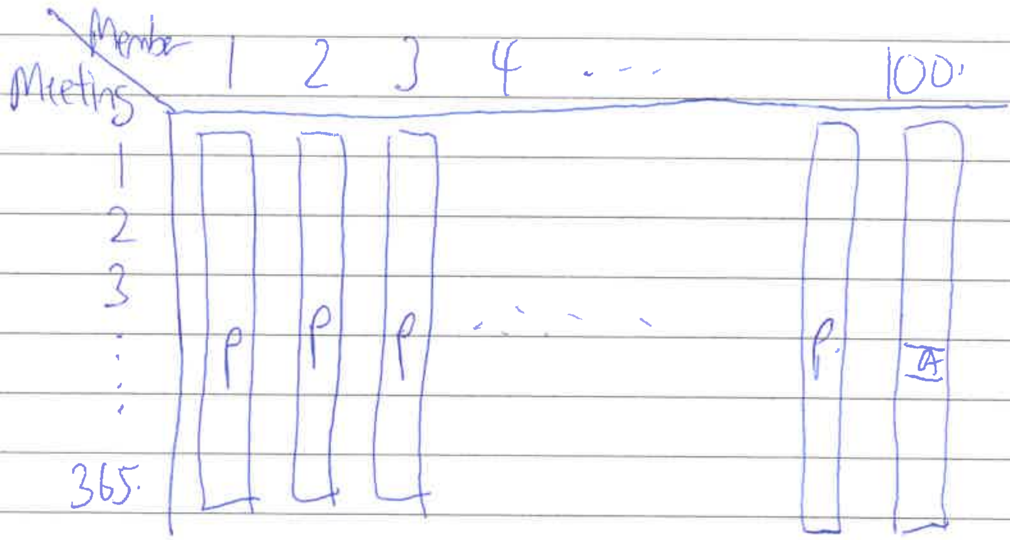
If all meetings were nearly full, it isn't that any member has perfect attendance

Subject :

Date:



All meetings were nearly full it isn't necessary that any member has perfect attendance.



Almost all members have perfect attendance $\Rightarrow$ Each mtg must be almost full.

Convergence a.s $\not\Rightarrow$ Conv. in prob.

Subject :

Date: