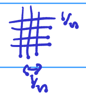
Lec 9 : I am going to skip two sections.

1) 2.6 Subadditive ergodic theorem. → Outlines Kingman's subadditive ergodic theorem. Analytic tools necessary for this theorem

2) 2.7 Gromov - Hausdorff convergence



$\left(\frac{\mathbb{Z}^d}{n}, \frac{T(nx,ny)}{n}\right)$ is a metric space on a lattice with spacing $\frac{1}{n}$.

In what sense does it converge to $(\mathbb{R}^d, \mu(x-y))$

This is the content of this section. It doesn't really give you anything new on the lattice, but it helps on non amenable things like Cayley graphs where subadditive ergodic theory does not directly apply.

On the lattice it doesn't give us anything new.
| Cayley graphs.

But it does use a so-called "concentration" inequality. → There are general analytic inequalities.

Lemma 2.37 : Fix a box of size M and $\epsilon > 0$

Then $\exists\, C, \alpha$ st

$$P\left(\exists\, x,y \in [-M,M]^d : \left| \overbrace{\frac{T(nx,ny)}{n}}^{\text{random qty}} - \overbrace{\mu(x-y)}^{\text{thing its converging to}} \right| > \epsilon \right)$$

$$\leq \underbrace{Ce^{-n^\alpha}}$$

Borel-Cantelli: $\frac{T(nx,ny)}{n} \longrightarrow \mu(x-y)$ as

$\forall\, (x,y) \in [-M,M]^d$

Lets take the standard strong law: $\{X_i\}_{i=1}^{n}$ $\int_0^\infty e^{\lambda u}\, dF(u) < \infty$

$S_n = X_1 + \cdots + X_n$ and $E[e^{\lambda X}] < \infty$ for $-\delta < \lambda < \delta$

We know

$$\mathbb{P}\left(\omega : \lim_{n\to\infty} \frac{S_n}{n} = E[X_1]\right) = 1 \qquad \left(P\left(\left| \frac{S_n}{n} - E[X] \right| > t \right) \longrightarrow 0 \right)$$

conv. in probability.

But whats the rate at which this convergence happens?

$\frac{S_n}{n} - E[X_1] < t$ $t<0$

$$\mathbb{P}\left(\frac{S_n}{n} - E[X_1] > t \right) \qquad t > 0$$

free parameter.

$$= \mathbb{P}\left(e^{\theta(S_n - nE[X_1])} > e^{n\theta t} \right) \qquad \theta > 0 \qquad P(A) = E[1_A]$$

$$= \mathbb{E}\left[1_{\overbrace{e^{\theta(S_n - nE[X_1])} > e^{n\theta t}}^{A}} \right]$$

ratio >1 on the set A, by definition.

Markov inequality

Tchebyshev, Chernoff trick.

$$\le \mathbb{E}\left[\frac{e^{\theta(S_n - nEX_1)}}{e^{n\theta t}} 1_A \right] \le e^{-n\theta t}\, \mathbb{E}\left[e^{\theta \tilde{S}_n} \right] \quad - \#1$$

$1_A \le 1$

where $\tilde{S}_n = \sum \tilde{X}_i$ $\tilde{X}_i = X_i - E[X_i]$

These variables are still iid, so

$$E\left[e^{\theta(\hat{X}_1+\hat{X}_2+\cdots+\hat{X}_n)}\right] = \prod_{i=1}^{n} E\left[e^{\theta\tilde{X}_1}\right]$$

$$/\!/$$

$$E\left[e^{\theta\tilde{S}_n}\right] = E\left[e^{\theta\tilde{X}_1}\right]^n = e^{n\log M(\theta)}$$

$$\downarrow$$

$$M(\theta) := E\left[e^{\theta\tilde{X}_1}\right]$$

from 1

$$\boxed{P(A) \leqslant e^{-\theta t n + n\log M(\theta)}}$$

$$\leqslant e^{-n\sup_\theta(\theta t - \log M(\theta))} = e^{-nI(t)}$$

$$\underbrace{I(t)} := \sup_\theta\left(\theta t - \log M(\theta)\right)$$

This is the legendre transform of $\log M(\theta)$ and is called the large deviation rate fn for X_1.

→ We saw it in the last class in Biggin's theorem.

→ Prove that this is convex

→ Cramér's theorem.

You can also prove the lower bound in this case

So $P(A) \approx e^{-nI(t)}$ or more precisely

$$-\frac{\log P(A)}{n} \longrightarrow I(t)$$

→ deviation of the random qty in the limit.

This is Cramer's theorem, and you can read about it online. The lower bound is interesting.

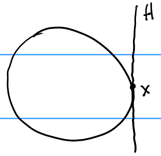
★ Good reading opportunity.

So the theorem we have above is a large deviation bound

2.8 Some definitions about curvature.

$$B = \{x : \mu(x) \le 1\}$$

H is a supporting hyperplane of B at x if H contains x and B intersects at most one component of H^c

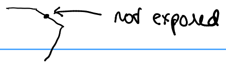


$\mu(x)$ is a convex fn. and hence it is differentiable a.e. In fact the set of non differentiability pts is at most countable. (Rademacher's theorem)

Thus B has supporting hyperplanes at all points except for a countable bad set.

Extreme point of ∂B : x Cannot be written as a linear combination $x = \lambda z_1 + (1-\lambda) z_2$ for $0 < \lambda < 1$

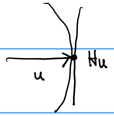
Exposed point : $\exists$ a hyperplane H at x that intersects B only at x_0 .



Curvature exponent : Assume μ is differentiable, and let H_u be the supporting hyperplane at u.

We say μ has curvature exponent $\kappa(u) > 0$ in direction u if $\exists$ c_1, c_2, ϵ st $\forall$ $z+u \in H$, $|z| < \epsilon$

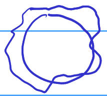
$$c_1 |z|^{\kappa(u)} \leq |\mu(z+u) - \mu(u)| \leq C_2 |z|^{\kappa(u)} \qquad - (\#1)$$

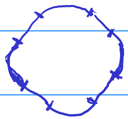


Uniformly curved : If #1 holds for fixed c_1 and ϵ for all u, then its called uniformly curved.

If the limit shape is a circle, then $\kappa(u) = 2$.

Newman generalized this idea to non differentiable shapes but its not that important now. Many theorems are proved using such uniform curvature assumptions.





Fluctuations

We have seen that $T(0,x) = \mu(x) + \underbrace{o(|x|)}_{\text{2nd order fluctu}} \longleftarrow$ $o(|x|)$ follows $\frac{T(0,x)}{|x|} \to \mu\left(\frac{x}{|x|}\right)$

— #2

shape theorem | subadditive ergodic

Recall that we expect (universality)

$$\frac{T(0,x) - \mu(x)}{C\,|x|^{1/3}} \Rightarrow F_{GUE}$$

(Tracy-Widom distribution) → from random matrix theory.

$\frac{S_n}{n} \to E[X_1]$

$\frac{S_n - nE[X_1]}{n^{1/2}} \xRightarrow{CLT}$ Gaussian

$1/2$ → scale of the fluctuations.

So the first thing one would want to show is that the error in #2 is

$$o(|x|) \sim |x|^{\chi} \qquad \chi = \frac{1}{3} \text{ in } d=2$$

$$\mathbb{E}\left[\,|S_n - nE[X_1]|^2\right]^{1/2} = Var(S_n)^{1/2}$$

$$= \sigma\sqrt{n}$$

We can divide this error into two parts:

1) $T(0,x) - \mathbb{E}[T(0,x)]$ random fluctuation
2) $\mathbb{E}[T(0,x)] - \mu(x)$ non random "

$Var(S_n) = Var(X_1) + Var(X_2) + \cdots$

$= n\,Var(X_1) = n\sigma^2$

$\mathbb{E}\left[(S_n - \mathbb{E}[S_n])^2\right]$

(In the case of the strong law $\mathbb{E}[T(0,x)]$ and $\mu(x)$ are the same "$n\,\mathbb{E}[X_1]$"

We expect that:

$$Var(T(0,x)) \sim |x|^{2\chi} \qquad \chi = \frac{1}{3}$$

and

$$\mathbb{E}(T(0,x)) - \mu(x) \sim |x|^{\chi}$$

This is known to be true in the solvable model → Exponential wts last percolation.

of course.

Change wrt dimension

χ = "fluctuations exponent"

$d=1$ | $\chi = 1/2$ (CLT) (Theorem)

$d=2$ | $\chi = 1/3$

$d \geq 3$ | unknown, no guesses.

Does $\chi(d) \longrightarrow 0$ as $d \to \infty$? Lots of debate here.

Analogous model for a binary tree (BRW) $\chi(d) \longrightarrow 0$

↳ lots of comments about this in

Newan - Piza (1995) Annals. of Prob.

Auffinger - Tang

First rigorous results:

$$\mathrm{Var}(T(0, ne_1)) \leq C\left(\frac{n}{\log^{1/p} n}\right)^2$$

Kesten, 1986 (Aspects, 5.16)

p is some dimension dependent #.

$\mathrm{Var}(S_n) \leq n^2$

$$\mathbb{E}\left(\sum \tilde{X}_i\right)^2 = \mathbb{E}\sum \tilde{X}_i \sum \tilde{X}_j$$

$$= \sum_{i,j} \mathbb{E}[\tilde{X}_i \tilde{X}_j]$$

$$\leq n^2\, \mathbb{E}[\tilde{X}_i]^2$$

$$= n^2 \sigma^2$$

$$\mathbb{E}[\underset{\to 0}{X_i}]\,\mathbb{E}[\underset{\to 0}{X_j}]$$

Independence & mean 0

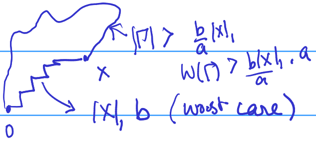
This is marginally better than trivial: If

$0 < a \leq \tau_e \leq b < \infty$ (bounded wts)

then Γ_{0x} the geodesic between 0 and x satisfies

length of the geodesic

$$|\Gamma_{0x}| \leq \frac{b}{a}|x|_1$$



So we can restrict to paths that lie in a box of size

$\frac{b}{a}|x|_1$

Then $T(0,x) = \sum_{i=1}^{|\Gamma|} X_i$ where X_i are the weights

that appear on Γ

$$Var(T(0,x)) = Cov\left(\sum_{i=1}^{|\Gamma|} X_i , \sum_{j=1}^{|\Gamma|} X_j\right)$$

$$= \sum_{i,j=1}^{|\Gamma|} Cov(X_i, X_j) \leq |\Gamma|^2 b^2 = C|x|_1^2$$

Covariance is bilinear.

$Cov(X,Y) = E[(X-EX)(Y-EY)]$

$Cov(X,X) = Var(X)$

$x = ne_1$ $Var(T(0,ne_1)) \leqslant Cn^2$

$\leqslant \left(\frac{n}{\log^{1/p} n}\right)^2 C$

Such a bound can be proved in general, without assuming that $\{\tau_e\}$ are bounded.

So Kesten's theorem is a very minor improvement.

(1993 Kesten)

Thm: If $\overbrace{F(0) < p_c}^{\text{no percolation}}$ and $\overbrace{E[\tau_e^2]}^{\text{2nd moment.}} < \infty$

$$Var(T(0,x)) \leq C|x|_1 \quad \forall d \geqslant 1$$

power was reduced from $2-\varepsilon$ to 1.

Then $\boxed{C_1} \leqslant Var(T(0,x)) \leqslant C_2 |x|_1$

where the lower bound holds if $\{\tau_e\}$ is not deterministic.

$\log|x|_1$ (Newman-Piza)

$\tau_e = 1$

$Var(T(0,x)) \sim |x|^{2/3}$

If you could prove

$Var(T(0,x)) \leq \underline{n}^{1-\varepsilon}$

S.Chatterjee

by Jean Bourgain.

Efron-Stein inequality (Kesten)

Before I get into this, I ought to review conditional expectation, and conditional Jensen.

50-50 review.

$$f(X_1 \dots X_n) \rightarrow \mathbb{R}$$

$$\mathbb{E}[f(X_1, \dots X_n) \mid \overbrace{\underset{\substack{\shortparallel \\ u_1}}{X_1}, \underset{\substack{\shortparallel \\ u_2}}{X_2}}^{\sigma(X_1, X_2)}]$$

If X_i are iid

$$\int f(\overbrace{\bar{u}_1, \bar{u}_2}^{\text{freeze}}, u_3 \dots u_n) \, dF_3(u_3) \cdots df_n(u_n)$$

Given $\{X_i\}$ rvs

Let $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ called filtration.

$$Z = f(X_1, \dots, X_{n-1})$$

$$\mathbb{E}[Z \mid \mathcal{F}_n] = \mathbb{E}[Z \mid X_1, \dots X_{n-1}]$$

$$= \mathbb{E} \, \mathbb{E}[Z \mid X_1, \dots X_n] \mid X_1, \dots X_n]$$

Tower property

$$\mathbb{E}[f(X_1, \dots X_n)]$$

$$= \mathbb{E}[\mathbb{E}[f(X_1, \dots X_n) \mid X_1, X_2]]$$

What is this property called?

POLL: Radon-Nikodym theorem, Tower property, its obvious.

Jensens inequality: If ϕ is convex and $\phi(X)$ is integrable, where X is some random variable

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X])$$

$$\phi(u) = u^2$$

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$$

$$\int u^2 df(u) \geq \left(\int u \, df(u)\right)^2$$

$$\mathbb{E}[\underbrace{f(X_1,\cdots X_n)}\,|\,\underset{\text{random}}{\boxed{X_1,X_2}}]$$

Conditional Jensen: Let Σ be a σ-algebra ϕ be convex then

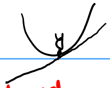
$$\mathbb{E}[\phi(X)\,|\,\Sigma] \geqslant \phi(\overset{\text{random\#}}{\mathbb{E}[X|\Sigma]})$$

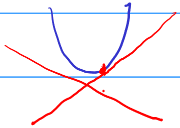
conditional version of Jensens inequality

If its concave reverse is true.

How to prove: There are many, but one is to use the fact that ϕ is supported below by linear fns and in fact

$$\boxed{\phi(y) = \sup_{ax+b\le\phi(x)\forall x} ay+b}$$





$$\mathbb{E}[\,aX+b\,|\,\Sigma] \overset{\text{linearity of integral}}{=} aE[X|\Sigma]+b$$

$$\mathbb{E}[aX+b|\Sigma] \overset{\text{below }\phi}{\leq} \mathbb{E}[\phi(X)|\Sigma]$$

$$\Rightarrow \quad \mathbb{E}[\phi(X)|\Sigma] \geqslant a\,\mathbb{E}[X|\Sigma]+b$$

Take a sup over a,b st $ax+b \leq \phi(x)$ $\forall x$

Slick proof.

and thus

$$\mathbb{E}[\phi(x)|\Sigma] \geqslant \phi(\mathbb{E}[X|\Sigma])$$

Back to Efron-Stein inequality. This is a discrete version of the Poincare inequality.

POLL: How many of you have seen this?

The Poincare inequality on some nice bounded set D on $\mathbb{R}^d$ with Lebesgue measure is:

Poincare constant. → Spectral-gap inequality / Poincare inequality.

$$\| f - \frac{1}{vol(D)} \int f\,dx \|_2 \leq \overset{\text{Poincare constant.}}{C_D} \| \, |\nabla f|_2 \|_2$$

Euclidean distance.

Need f to be differentiable.

There are L^p versions of the Poincare inequality. To generalize it to unbounded sets, you integrate over some finite measure $\rho(x)dx$ instead of Lebesgue mean dx.

Gaussian measure
$C_D = 1$.

However the Poincare inequality only holds for some $\rho(x)dx$, and there is a lot of theory to determine when a Poincare ineq holds (I dont remember this at the moment). It's also called a spectral-gap inequality. I have a certain outlook on this, but this is a whole other course.

Efron-Stein ("discrete Poincare")

Let $X_1, \dots X_n$ be iid, X_i' be an indep. copy of X_i for each i

$f: \mathbb{R}^n \to \mathbb{R}$ be in $L^2(\Omega)$

$$f \equiv f(X_1, \dots, X_n) \qquad f_i := f(X_1, \dots \underbrace{X_i'}, \dots X_n)$$

swap out ith coordinate with an independent copy.

$$\mathrm{Var}(f) \le \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[(\overbrace{f - f_i}^{\text{difference}})^2\right] = \sum_{i=1}^{n} \left[\mathbb{E}\left[(f-f_i)^2 \mathbb{1}_{\{f > f_i\}}\right]\right]$$

$$\mathbb{E}\left[(f-f_i)^2 \mathbb{1}_{\{f > f_i\}}\right] = \mathbb{E}\left[(f-f_i)^2 \mathbb{1}_{\{f_i < f\}}\right]$$

$$\| |\nabla f|_2 \|_2^2 = \sum_{i=1}^{n} \int |\partial_i f|^2 \, dx$$

Pf: The proof is always the same. There are many ways to look at it, and they all involve "slowly replacing $X_1, \dots X_n$ by $X_1' \dots X_n'$ one by one." This also appears in the Lindeburg proof technique, and is my favorite proof of the CLT. This idea is not so obvious to see here, but its quite obvious when proving the Gaussian Poincare inequality.

Ex: Let $f(X_1, \dots X_n) = \sum_{i=1}^{n} X_i$ if X_i are iid

POLL: $\mathrm{Var}(f)$ is $O(\cdots)$

n^2	n	$n^{1/2}$	Constant
A	B	C	D

$\mathrm{Var}(f) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = n\sigma^2$

Answer is B

$\sum_{i=1}^{n} X_i$ $\sum_{i=1}^{n} x_i - x_i + x_i'$

$$\mathbb{E}\left[(f - f_i)^2\right] = \mathbb{E}\left[(X_i - X_i')^2\right] \leq 4 m_2$$

$f \approx E[X_i]n + \sqrt{\text{var} f}$

$= E[X_i]n + \sigma\sqrt{n}$

Cauchy-Schwarz

$$E[X_i^2 + X_i'^2 + 2X_i X_i'] \leq 4E[X_i^2]$$

where $m_2 = \mathbb{E}[X_1^2]$

should σ^2

Thus $\text{var}(f) \leq \boxed{2 m_2} n$ (not bad, right order)

Efron-Stein is pretty good in this case.

Ex:

$f = \max(X_1, .. X_i, X_n)$

$f_i = \max(X_1, \dots X_i', \dots X_n)$

→ Extreme laws.

⋆ Good student reading

$f \approx$ 1st order + fluctuations

We assume $\{X_i\}$ have ~~continuous~~ distribution.

and are iid

POLL: What is the order of $\text{Var}(f)$?

n^3	n^2	$n^{1/2}$	Constant
A	B	C	D

None of these. I will show you towards the end.

Should 'expect' smaller variance for $f = \max(X_1, \dots, X_n)$ compared to $f = \sum_{i=1}^{n} X_i$

Ex: Prove that there is a unique maximum in $(X_1, \dots X_n)$ when X_i have continuous distribution

Apply Efron-Stein to the max function $\longrightarrow$ RHS

$$\frac{1}{2}\mathbb{E}[(f-f_i)^2] = \mathbb{E}[(f-f_i)^2 \mathbb{1}_{\{X_i > X_i'\}}]$$

$$\mathbb{E}[(f-f_i)^2] = \mathbb{E}[(f-f_i)^2 \mathbb{1}_{\{X_i > X_i'\}}] + \mathbb{E}[(f-f_i)^2 \mathbb{1}_{\{X_i' > X_i\}}]$$

$$= \mathbb{E}[(f-f_i)^2 \mathbb{1}_{\{X_i > X_i'\}} \mathbb{1}_{\{X_i \text{ is max}\}}]$$

In the worst case $f_i = X_i'$

$f \geq f_i \geq X_i'$ $\quad f - f_i \leq X_i - X_i'$

$$\leq \mathbb{E}[(X_i - X_i')^{\overset{\swarrow p}{2}} \mathbb{1}_{\{X_i > X_i'\}} \mathbb{1}_{\{X_i \text{ is max}\}}]$$

— (#2)

Lets assume further $a < X_i < b$ a.s $\Rightarrow X_i - X_i' \leq (b-a) = C$

$$\leq C\,\mathbb{E}[\overset{\leq 1}{\mathbb{1}_{\{X_i > X_i'\}}} \mathbb{1}_{\{X_i \text{ is max}\}}] \leq C\,\mathbb{P}(X_i \text{ is max})$$

Thus $\mathrm{Var}(f) \leq C \sum_{i=1}^{n} \mathbb{P}(X_i \text{ is max}) = C \sum_{i=1}^{n} \frac{1}{n}$

$= C.$

Efron-Stein is telling us $\mathrm{Var}(f) \leq O(1)$

At #2 its worth nothing that you could use Hölder

Then

$$\mathrm{Var}(f) \leq C \sum_{i=1}^{n} \|X_i\|_{2p}^{1/2} \frac{1}{n^{1/q}} = n\,C \frac{1}{n^{1/q}} \|X_i\|_{2p}^{1/2}$$

$\leftarrow$ q norm

For Gaussian tails $P(X_1 > t) \approx Ce^{-bt^2}$ $\longrightarrow$ Can show this for a Gaussian

$\|X_1\|_p \approx p^{1/2}$ (Exercise)

$\frac{1}{p} + \frac{1}{q} = 1$

So $\mathrm{Var}(f) \leq C n^{1-\frac{1}{q}} \|X_i\|_{2p}^{1/2}$

$$= Cn^{1/p}\sqrt{2p}^{\,1/2} = e^{\frac{1}{p}\log n + \frac{1}{4}\log p}$$

If you chose $p = \log n \Rightarrow \mathrm{Var}(f) \leq e^{\log\log n} = \log n$ $\longrightarrow$ This is the wrong order

Is this the right order?

An enterprising student can perform this computation using extreme value theory quite easily.

POLL: Is $O(1)$ the correct order for $\mathrm{var}(f)$?

YES OR NO

→ Not the right order.

$$\mathbb{P}(\overbrace{\max(X_1,\dots,X_n)}^{G_n} \le t) \longrightarrow X_1 \le t, \; X_2 \le t, \dots \; X_n \le t$$

indep.
$$= \prod_{i=1}^{n} \mathbb{P}(X_i \le t) = F(t)^n \xrightarrow{n\to\infty ?} 0$$

scaling is not right

scale, center

$$\mathbb{P}(a_n G_n + b_n \le t) = F\left(\frac{t-b_n}{a_n}\right)^n$$

Assume for convenience $X_i \sim N(0,1)$ (Gaussian)

$$F(x) = \int_{-\infty}^{x} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,du$$

$$= \left(1 - \int_{\frac{t-b_n}{a_n}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\,du\right)^n$$

$$\boxed{\hat{=}} \left(1 - \frac{e^{-\left(\frac{t-b_n}{a_n}\right)^2 \frac{1}{2}}}{\sqrt{2\pi}}\right)^n \longrightarrow \left(1 - \frac{x}{n}\right)^n$$

For this to make any sense would need

$$\frac{n}{n}\,\frac{e^{-\left(\frac{t-b_n}{a_n}\right)^2 \frac{1}{2}}}{\sqrt{2\pi}} \;\hat{=}\; \frac{f(t)}{n}$$

The exponent is

$$-\underbrace{\frac{t^2}{a_n^2}} + \boxed{\frac{2b_n t}{a_n^2}} \; \boxed{- \frac{b_n^2}{a_n^2} + 2\log(n)} \longrightarrow kt$$

$$f(t) = e^{kt}$$

Other term

$\Rightarrow \quad \frac{b_n}{a_n} = \sqrt{2\log n} \quad \Big| \quad 2\frac{\sqrt{2\log n}}{a_n} t$

Must choose $a_n = k\sqrt{2\log n}$

$\Rightarrow \quad b_n = 4k \log n$

So this gives us the scale of the maximum: b_n is about $\log n$. The fluctuations are on scale ?

$b_n = \Theta(\log n)$

$\delta \approx k\sqrt{\log n} +$ 2nd term

$\delta = k \log n + \underbrace{\frac{1}{\sqrt{\log n}} \xi}$

$a_n \delta + b_n \to \xi$

$f \to \frac{\xi - b_n}{a_n}$

$\delta \to \frac{1}{a_n}\xi - \frac{b_n}{a_n}$

$\to \frac{1}{\sqrt{\log n}}\xi + k\sqrt{\log n}$

POLL:

$\sqrt{\log n}$	$\log n$	$\frac{1}{\sqrt{\log n}}$
A	B	C

Exercise: 1) Determine the scale for $X_i \sim Exp(1)$

2) Determine k and the form of the limiting distribution (Gumbel)

Comes under the umbrella of Extreme Value Theory.