Analysis of Local Stability and Global Dynamics of Heavy-Tailed SGDs

$$* \begin{cases} W_0^\eta(x) = x \\ W_{j+1}^\eta(x) = W_j^\eta(x) + \eta \cdot a(W_j^\eta(x)) + \eta \cdot \sigma(W_j^\eta(x)) Z_j \quad \forall j \geq 1 \end{cases}$$

iid $RV_{-\alpha}$

$$\mathbb{W}^\eta(x) \triangleq \{ W^\eta_{\lfloor t/\eta \rfloor}(x) : t \in [0,T] \}$$

$$\varphi_b(x) = \frac{x}{|x|} \min\{b, |x|\}$$

$$* \begin{cases} W_0^{\eta|b}(x) = x \\ W_{j+1}^{\eta|b}(x) = W_j^{\eta|b}(x) + \varphi_b\left(\eta \cdot a(W_j^{\eta|b}(x)) + \eta \cdot \sigma(W_j^{\eta|b}(x)) Z_j\right) \quad \forall j \geq 1 \end{cases}$$

$$\mathbb{W}^{\eta|b}(x) \triangleq \{ W^{\eta|b}_{\lfloor t/\eta \rfloor}(x) : t \in [0,T] \}$$

* Assumptions

A1. $\mathbb{P}(|Z_1| > x) = x^{-\alpha} L(x)$

A2. $|\sigma(x) - \sigma(y)| \vee |a(x) - a(y)| \leq D \cdot |x - y|$

A3. $\sigma(x) > 0 \quad \forall x$

* For this lecture, we will specialize $a(\cdot) = -f'(\cdot)$

so that $\mathbb{W}^\eta(x)$ is SGD

$*$ Assume w.l.o.g. $f'(0) = 0$

and $f'(x) \cdot x > 0$ for $x \in I \setminus \{0\}$ so that

$I = [-s_\ell, s_r]$

$$\begin{cases} y_0(x) = x \\ \dfrac{dy_t(x)}{dt} = -f'(y_t(x)) \quad \forall t > 0 \end{cases} \Longrightarrow \lim_{t\to\infty} y_t(x) = 0$$

$*$ $\tau^\eta(x) \triangleq \min\{ j \geq 0 : W_j^\eta(x) \notin I \}$

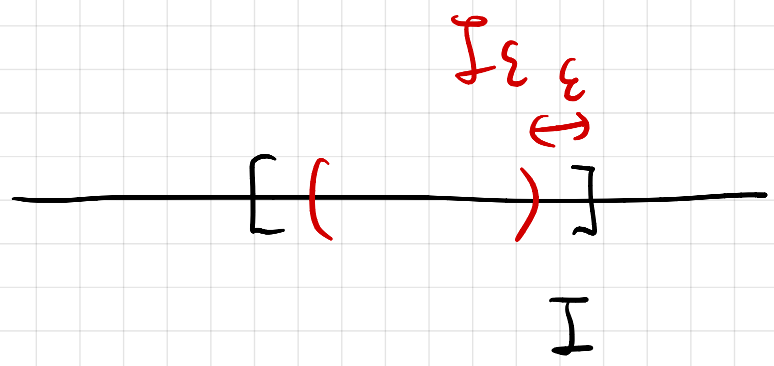
$\tau^{\eta|b}(x) \triangleq \min\{ j \geq 0 : W_j^{\eta|b}(x) \notin I \}$

$*$ $\left.\begin{array}{l} \check{C}(\cdot\,; x) \\ \check{C}_k^b(\cdot\,; x) \end{array}\right\}$: measures on $\mathbb{R}$

$*$ $r = s_\ell \wedge s_r$ $\qquad J_b^* = \lceil r/b \rceil$

* **Thm**

$$\limsup_{\eta\downarrow 0}\ \sup_{x\in I_\varepsilon} \mathbb{P}\left(C_b^*\,\eta\cdot\left(\frac{1}{\eta}\mathbb{P}(|Z|>1/\eta)\right)^{J_b^*}\tau^{\eta|b}(x) > t\ ;\ W^{\eta|b}_{\tau^{\eta|b}}(x)\in B\right) \le \frac{\check{C}^b_{J_b^*}(B^-)}{C_b^*}\exp(-t)$$



$$\liminf_{\eta\downarrow 0}\ \inf_{x\in I_\varepsilon} \mathbb{P}\left(C_b^*\,\eta\cdot\left(\frac{1}{\eta}\mathbb{P}(|Z|>1/\eta)\right)^{J_b^*}\tau^{\eta|b}(x) > t\ ;\ W^{\eta|b}_{\tau^{\eta|b}}(x)\in B\right) \ge \frac{\check{C}^b_{J_b^*}(B^\circ)}{C_b^*}\exp(-t)$$

$$C_b^* \triangleq \check{C}^b_{J_b^*}(I^c) \in (0,\infty)$$

* Thm

$$\limsup_{\eta\downarrow 0}\ \sup_{x\in I_\varepsilon} \mathbb{P}\Big(C^*\eta\cdot\big(1/\eta\ \mathbb{P}(|Z|>1/\eta)\big)\,\tau^{\eta}(x) > t\ ;\ W^{\eta}_{\tau^{\eta}}(x)\in B\Big)$$

$$\le \frac{\check{C}(B^-)}{C}\exp(-t)$$

$$\liminf_{\eta\downarrow 0}\ \inf_{x\in I_\varepsilon} \mathbb{P}\Big(C^*\eta\cdot\big(1/\eta\ \mathbb{P}(|Z|>1/\eta)\big)\,\tau^{\eta}(x) > t\ ;\ W^{\eta}_{\tau^{\eta}}(x)\in B\Big)$$

$$\ge \frac{\check{C}(B^\circ)}{C}\exp(-t)$$

$$C \triangleq \check{C}(I^c) \in (0,\infty)$$

$\ast$ Def) Asymptotic Atom $\{A(\varepsilon) : \varepsilon > 0\}$ of M.C. $\{V_j^\eta(x) : j \geq 0\}$

assoc. with I, $\{I(\varepsilon) : \varepsilon > 0\}$

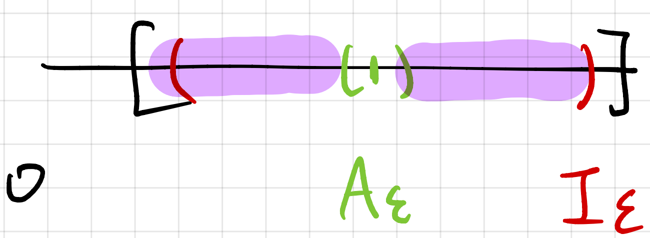
$C(\cdot)$, $\gamma(\cdot)$

Given B, $\exists\ \delta_B,\ \varepsilon_B,\ T_B$ s.t.

$$C(B^\circ) - \delta_B(\varepsilon, T) \leq \liminf_{\eta \downarrow 0} \frac{\inf_{x \in A(\varepsilon)} \mathbb{P}\left(\eta \cdot \tau^\eta_{I(\varepsilon)^c}(x) \leq T;\ V^\eta_{\tau^\eta_{I(\varepsilon)^c}}(x) \in B\right)}{\gamma(\eta)\, T/\eta}$$

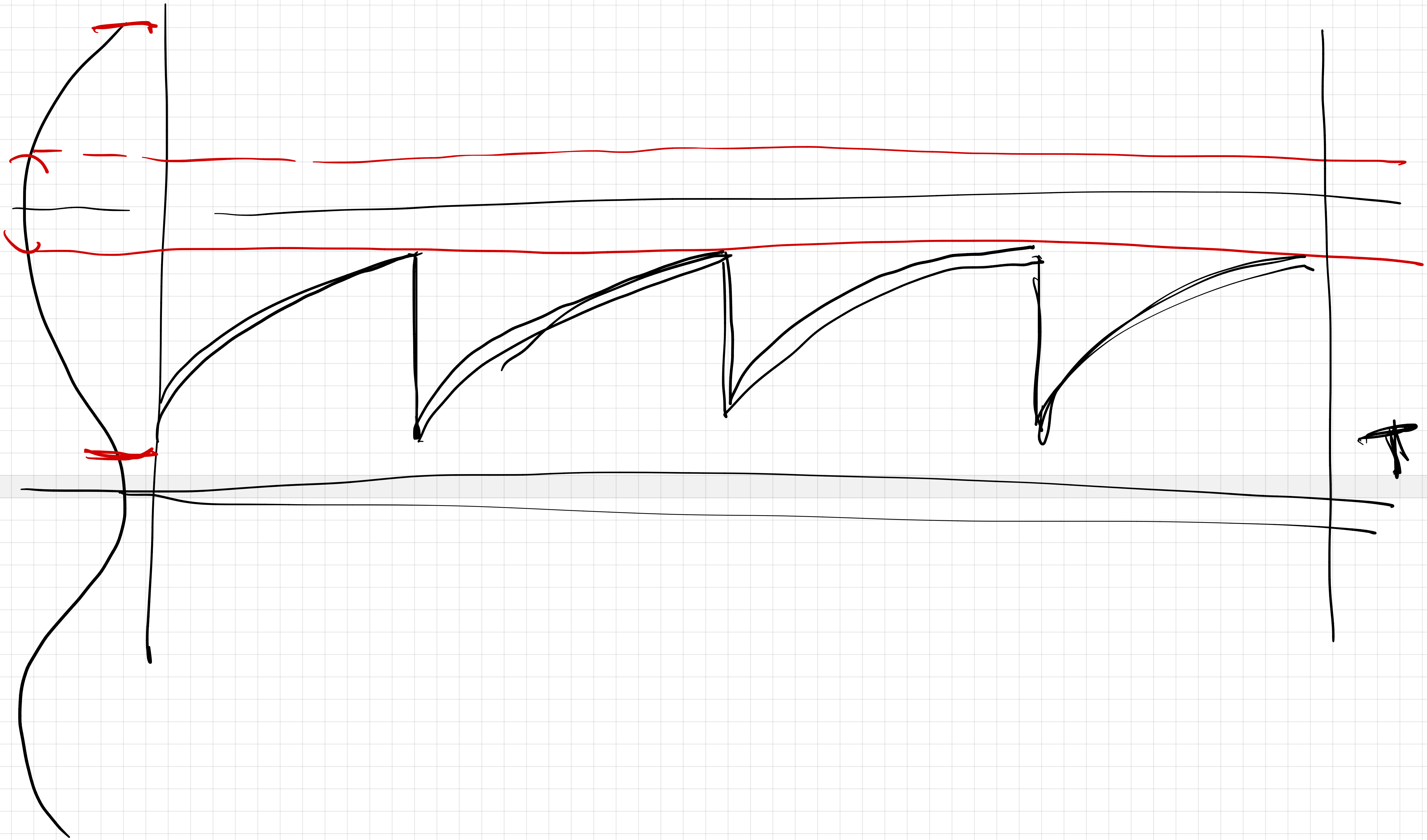
$$\leq \limsup_{\eta \downarrow 0} \frac{\sup_{x \in A(\varepsilon)} \mathbb{P}\left(\tau^\eta_{I(\varepsilon)^c}(x) \leq T/\eta;\ V^\eta_{\tau^\eta_{I(\varepsilon)^c}}(x) \in B\right)}{\gamma(\eta)\, T/\eta} \leq C(B^-) + \delta_B(\varepsilon, T)$$

$$\limsup_{\eta \downarrow 0} \frac{\sup_{x \in A(\varepsilon)} \mathbb{P}\left(\tau^\eta_{(I_\varepsilon \setminus A_\varepsilon)^c}(x) > T/\eta\right)}{\gamma(\eta)\, T/\eta} = 0$$



$$\liminf_{\eta \downarrow 0}\ \inf_{x \in I(\varepsilon)} \mathbb{P}\left(\tau^\eta_{A(\varepsilon)}(x) > T/\eta\right) = 1$$

for any $\varepsilon \leq \varepsilon_B$, $T \geq T_B$ where $\lim_{\varepsilon \to 0} \lim_{T \to \infty} \delta_B(\varepsilon, T) = 0$



* **Thm** $\exists$ Asymptotic Atom then

$$C(B^\circ)\cdot e^{-t} \leq \liminf_{\eta\downarrow 0} \inf_{x\in I(\varepsilon)} \mathbb{P}\left(\gamma(\eta)\tau_{I^c}^{\eta} > t,\ V^{\eta}_{\tau_{I^c}^{\eta}}(x)\in B\right)$$

$$\leq \limsup_{\eta\downarrow 0} \sup_{x\in I(\varepsilon)} \mathbb{P}\left(\gamma(\eta)\tau_{I^c}^{\eta} > t,\ V^{\eta}_{\tau_{I^c}^{\eta}}(x)\in B\right) \leq C(B^-)\cdot e^{-t}$$

* $h_k : (x, \vec{z}, \vec{t}) \longmapsto \xi.$

$x \in \mathbb{R}$

$\vec{t} = (t_1, t_2, \cdots, t_k) \in (0,T]^{k\uparrow}$

$= \{(t_1, \cdots, t_k) \in \mathbb{R}^k : t_1 < t_2 < \cdots < t_k\}$

$\vec{z} = (z_1, z_2, \cdots, z_k)$

$\xi \in \mathbb{D}[0,T]$

"gradient flow w/ k jumps"

$$\begin{cases} \xi_0 = x \\ \dfrac{d\xi_s}{ds} = a(\xi_s) & \forall s \in [0,T],\ s \neq t_1, \cdots, t_k \\ \xi_s = \xi_{s-} + \sigma(\xi_{s-}) \cdot z_j & \text{if } s = t_j \text{ for } j = 1, 2, \cdots, k \end{cases}$$

* $\mathbb{D}_k(A) = h_k(A, \mathbb{R}^k, [0,T]^{k\uparrow})$ $\qquad \mathbb{D}_{-1}(A) = \emptyset$

* $\mathbb{D}_{<k}(A) = \bigcup_{j=1}^{k-1} \mathbb{D}_j(A)$

* $\lambda(\eta) = \eta^{-1} \, \mathbb{P}(|z| > \eta^{-1})$

* $C_k(\,\cdot\,;x) \triangleq \int \mathbb{1}\{h_k(x,\vec{z},\vec{t}) \in \cdot\,\}\, \nu_\alpha^k(d\vec{z}) \times \mathcal{L}_T^{k\uparrow}(d\vec{t})$

$$= \mathbb{E}\left[\nu_\alpha^{\vec{J}}\{\vec{z}\in\mathbb{R}^k : h_k(x,\vec{z},\vec{t}) \in \cdot\,\}\right]$$

* $\nu_\alpha(x,\infty) = x^{-\gamma}$
* ν_α^k : k-fold product measure of ν_α

* <u>Thm</u> Supp A: cpt. For any $B \subseteq \mathbb{D}$ bdd away from $\mathbb{D}_{<k}(A)$,

$$\inf_{x\in A} C_k(B^\circ;x) \leq \liminf_{\eta\downarrow 0} \frac{\inf_{x\in A}\mathbb{P}(W^\eta(x)\in B)}{\left(1/\eta\, \mathbb{P}(|Z|>1/\eta)\right)^k}$$

$$\leq \limsup_{\eta\downarrow 0} \frac{\sup_{x\in A}\mathbb{P}(W^\eta(x)\in B)}{\left(1/\eta\, \mathbb{P}(|Z|>1/\eta)\right)^k} \leq \sup_{x\in A} C_k(B^-;x)$$

* $h_k^b : (x, \vec{z}, \vec{t}) \longmapsto \xi.$

$$\begin{cases} \xi_0^b = x \\ \dfrac{d\xi_s^b}{ds} = a(\xi_s^b) & \forall s \in [0,T],\ s \neq t_1, \cdots, t_k \\ \xi_s^b = \xi_{s-}^b + \varphi_b\big(\sigma(\xi_{s-}) \cdot z_j\big) & \text{if } s = t_j \text{ for } j = 1, 2, \cdots, k \end{cases}$$

"gradient flow w/ k jumps"

bdd by b

* $\mathbb{D}_k^b(A) = h_k^b(A, \mathbb{R}^k, [0,T]^{k\uparrow})$ $\qquad \mathbb{D}_{-1}^b(A) = \emptyset$

* $\mathbb{D}_{<k}^b(A) = \bigcup_{j=1}^{k-1} \mathbb{D}_j^b(A)$

* $\lambda(\eta) = \eta^{-1}\, \mathbb{P}(|z| > \eta^{-1})$

* $$C_k^b(\cdot\,;x) \triangleq \int \mathbb{1}\{h_k^b(x,\vec{z},\vec{t}) \in \cdot\}\, \nu_\alpha^k(d\vec{z}) \times \mathcal{L}_T^{k\uparrow}(d\vec{t})$$
$$= \mathbb{E}\left[\nu_\alpha^j\{\vec{z} \in \mathbb{R}^k : h_k^b(x,\vec{z},\vec{t}) \in \cdot\}\right]$$

* $\nu_\alpha(x,\infty) = x^{-\gamma}$

* ν_α^k : k-fold product measure of ν_α

* Thm Supp A: cpt. For any $B \subseteq \mathbb{D}$ bdd away from $\mathbb{D}_{<k}^b(A)$,

$$\inf_{x\in A} C_k^b(B^\circ;x) \leq \liminf_{\eta\downarrow 0} \frac{\inf_{x\in A} \mathbb{P}(W^{\eta|b}(x) \in B)}{\left(1/\eta\ \mathbb{P}(|Z| > 1/\eta)\right)^k}$$

$$\leq \limsup_{\eta\downarrow 0} \frac{\sup_{x\in A} \mathbb{P}(W^{\eta|b}(x) \in B)}{\left(1/\eta\ \mathbb{P}(|Z| > 1/\eta)\right)^k} \leq \sup_{x\in A} C_k^b(B^-;x)$$

* This is almost directly from

$$\left(\tfrac{1}{\eta}\mathbb{P}(|Z| > 1/\eta)\right)^{-k} \mathbb{P}\left(W^{\eta}(x) \in \cdot\right) \to C_k(\cdot \mid x)$$

in $\mathbb{M}\left(\mathbb{D} \setminus \mathbb{D}_{<k}(A)\right)$ uniformly in $x \in A$

what is this ?

$$\left(\tfrac{1}{\eta}\mathbb{P}(|Z| > 1/\eta)\right)^{-k} \mathbb{P}\left(W^{\eta|b}(x) \in \cdot\right) \to C^{b}_k(\cdot \mid x)$$

in $\mathbb{M}\left(\mathbb{D} \setminus \mathbb{D}^{b}_{<k}(A)\right)$ uniformly in $x \in A$

§ $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ – Convergence

Lindskog, Roy, Resnick. 2014. Probability Survey (LRR'14)

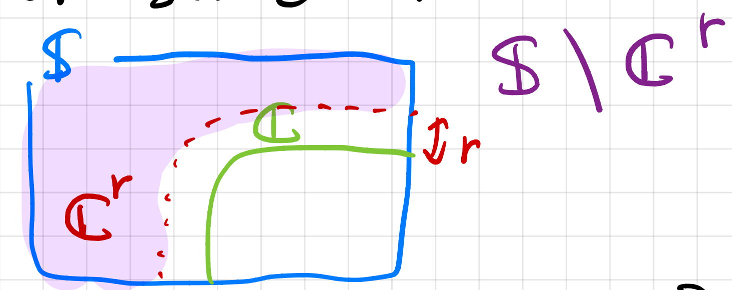
* $(\mathbb{S}, d)$: Complete Separable Metric Space

* $\mathbb{C} \subseteq \mathbb{S}$: closed

$$\{x \in \mathbb{S} : d(x,\mathbb{C}) < r\} \qquad d(x,\mathbb{C}) = \inf_{y\in\mathbb{C}} d(x,y)$$

* <u>Def</u> $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$: space of Borel measures μ s.t. $\mu(\mathbb{S}\setminus\mathbb{C}^r) < \infty$ $\forall r>0$, where $\mathbb{C}^r = \{x \in \mathbb{S} : d(x,\mathbb{C}) < r\}$

* <u>Def</u> $C(\mathbb{S}\setminus\mathbb{C})$: space of bdd cont' real valued non-negative ftns on $\mathbb{S}\setminus\mathbb{C}$ supported on $\mathbb{S}\setminus\mathbb{C}^r$ for some $r>0$



* We topologize $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ with subbasis

$$\Big\{\{\nu \in \mathbb{M}(\mathbb{S}\setminus\mathbb{C}) : \underbrace{\nu(f)}_{=\int f(x)\,\nu(dx)} \in G\} : f \in C(\mathbb{S}\setminus\mathbb{C}),\ G : \text{open in } \mathbb{R}_+\Big\}$$

That is, $\mu_n \to \mu$ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ if $\mu_n(f) \to \mu(f)$ $\forall f \in C(\mathbb{S}\setminus\mathbb{C})$.

* Portmanteau Thm for $\mathbb{M}(S\setminus C)$ Convergence (LRR '14)

The following conditions are equivalent:

(i) $\mu_n \to \mu$ in $\mathbb{M}(S\setminus C)$ as $n \to \infty$

(ii) $\mu_n(f) \to \mu(f)$ $\forall f \in C(\mathbb{S}\setminus\mathbb{C})$ uniformly continuous

(iii) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ $\forall F$: closed, bdd away from C

$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ $\forall G$: open, bdd away from C

(iv) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ $\forall A$: bdd away from C, $\mu(\partial A) = 0$

(v) $\mu_n^{(r)} \to \mu^{(r)}$ weakly in $\mathbb{S}\setminus\mathbb{C}^r$ for all but countable $r > 0$

(vi) $\exists r_i \searrow 0$ s.t. $\mu_n^{(r_i)} \to \mu^{(r_i)}$ weakly in $\mathbb{S}\setminus\mathbb{C}^{r_i}$ $\forall i$

(Wang & Rhee '24+)

Recall : $\mu_n \to \mu$ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ if $\mu_n(f) \to \mu(f)$ $\forall f \in \mathcal{C}(\mathbb{S}\setminus\mathbb{C})$

i.e., $|\mu_n(f) - \mu(f)| \to 0$

* Uniform $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence

$\mu_n^\theta, \mu^\theta \in \mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ $\forall \theta \in \Theta$

$\mu_n^\theta \to \mu^\theta$ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ uniformly in θ on Θ

if $\sup_{\theta\in\Theta} |\mu_n^\theta(f) - \mu^\theta(f)| \to 0$ $\forall f \in \mathcal{C}(\mathbb{S}\setminus\mathbb{C})$

* Portmanteau Thm for uniform $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ convergence

Supp. for any $\{\theta_n\}$ $\exists$ n_k and $\theta^* \in \Theta$ s.t.

$\mu^{\theta_{n_k}} \to \mu^{\theta^*}$ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$

i.e., $\{\mu^\theta : \theta \in \Theta\}$ is sequentially cpt.

Then the following conditions are equivalent :

(i) $\mu_n^\theta \to \mu^\theta$ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ uniformly in θ on Θ

(ii) $\sup_{\theta\in\Theta} \left|\mu_n^\theta(f) - \mu^\theta(f)\right| \to 0$ $\forall f \in \mathcal{C}(\mathbb{S}\setminus\mathbb{C})$ uniformly continuous

(iii) $\limsup_{n\to\infty} \sup_{\theta\in\Theta} \left(\mu_n^\theta(F) - \mu^\theta(F^\varepsilon)\right) \leq 0$ $\forall \varepsilon > 0$, $\forall F$: closed, bdd away from $\mathbb{C}$

$\liminf_{n\to\infty} \inf_{\theta\in\Theta} \left(\mu_n^\theta(G) - \mu^\theta(G^{-\varepsilon})\right) \geq 0$ $\forall \varepsilon > 0$, $\forall G$: open, bdd away from $\mathbb{C}$

Furthermore, any of the above implies

(iv) $\limsup_{n\to\infty} \sup_{\theta\in\Theta} \mu_n^\theta(F) \leq \sup_{\theta\in\Theta} \mu^\theta(F)$ $\forall F$: closed, bdd away from $\mathbb{C}$

$\liminf_{n\to\infty} \inf_{\theta\in\Theta} \mu_n^\theta(G) \geq \inf_{\theta\in\Theta} \mu^\theta(G^{-\varepsilon})$ $\forall G$: open, bdd away from $\mathbb{C}$