Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

Milad Bakhshizadeh

Stanford University

May 24, 2023

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三回= のへで

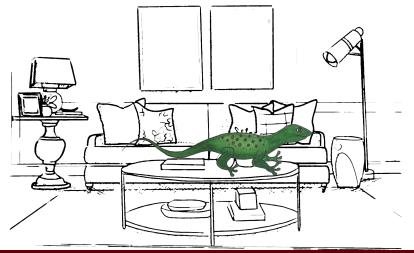
Intro ●0	Heavy-tail challenge 000	Setup 00	Truncation method	Optimality 000	End oo

うかの 正則 ふゆやえゆや (日本)

Milad Bakhshizadeh

Inaugural CAMDA Conference

Intro	Heavy-tail challenge	Setup	Truncation method	Optimality	End
●0	000	00		000	00

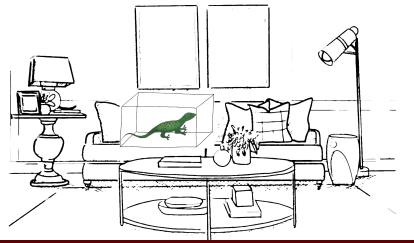


#### Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

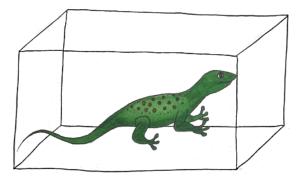
Intro	Heavy-tail challenge	Setup	Truncation method	Optimality	End
●0	000	00		000	00



#### Milad Bakhshizadeh

Inaugural CAMDA Conference

Intro	Heavy-tail challenge	Setup	Truncation method	Optimality	End
●0	000	00		000	00

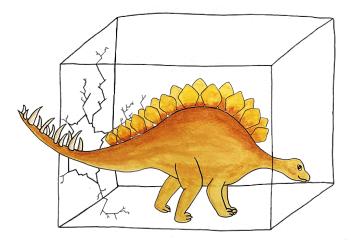


#### <ロ> < 四> < 回> < 回> < 回> < 回> < 回< のへの

Milad Bakhshizadeh

Inaugural CAMDA Conference

Intro	Heavy-tail challenge	Setup	Truncation method	Optimality	End
●0	000	00		000	oo



Milad Bakhshizadeh

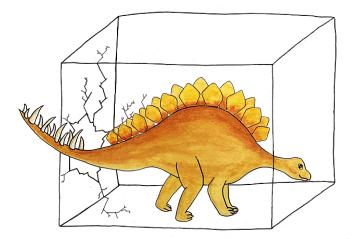
Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

= 990



### Uncertainty : How to make the Dinosaur a lizard?!



Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

-



### What is this talk about?

- 1. Heavy-tail makes it challenging to bound uncertainty
- 2. We can control heavy-tail by truncation
- 3. Truncation gives optimal bound

ELE DOG



### Part 1: problem setup

#### Heavy-tail distributions: the dinosaur



#### Inaugural CAMDA Conference

Milad Bakhshizadeh

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

ъ

Heavy-tail challenge End 000 Heavy-tails don't have finite MGF SubGaussian distributions  $\blacktriangleright \mathbb{P}(|X| > t) \le \exp(-ct^2)$  $|X||_p = O(\sqrt{p})$ •  $\mathbb{E}\left[\exp\left(\lambda X\right)\right] \leq \exp\left(c\lambda^{2}\right)$ 0.3 formal distribution

Heavy-tailed distributions

- $\blacktriangleright \mathbb{E}\left[\exp\left(\lambda X\right)\right] = \infty, \ \forall \lambda > 0$
- e.g. Weibull with k < 1,
  - $\mathcal{N}(0,1)^{lpha} \ lpha > 2$ , Log-Normal, ...

#### Milad Bakhshizadeh

Inaugural CAMDA Conference

4

0.2

0.1

2

-2

-4

E SQA

Heavy-tail challenge End 000 Heavy-tails don't have finite MGF SubGaussian distributions  $\blacktriangleright \mathbb{P}\left(|X| > t\right) \leq \exp\left(-ct^2\right)$  $|X||_p = O(\sqrt{p})$ •  $\mathbb{E}\left[\exp\left(\lambda X\right)\right] \leq \exp\left(c\lambda^{2}\right)$ formal distribution Heavy tailed Heavy-tailed distributions 0.1

• 
$$\mathbb{E}\left[\exp\left(\lambda X\right)\right] = \infty, \ \forall \lambda > 0$$

• e.g. Weibull with 
$$k < 1$$
,

$$\mathcal{N}(0,1)^{lpha} \ lpha >$$
 2, Log-Normal, ..

Milad Bakhshizadeh

4

-2

-4

= 990

etup O Optimality

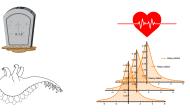
End 00

# Rise of the heavy-tails: A new era dawns

- Data is heavy-tailed
- Multiplication makes tail heavier
  - $\blacktriangleright$  XY, X<sup>n</sup>
  - ▶  $\mathcal{N}(0,1), \ \mathcal{N}(0,1)^2, \ \mathcal{N}(0,1)^3$

### Real applications

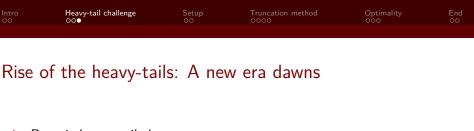
- Neural nets
- Phase retrieval



#### - ▲ □ ▶ ▲ @ ▶ ▲ 필 ▶ ▲ 필 ⊨ ~ 의식 @

Inaugural CAMDA Conference

Milad Bakhshizadeh



- Data is heavy-tailed
- Multiplication makes tail heavier
  - $\blacktriangleright XY, X^n$
  - $\mathcal{N}(0,1), \ \mathcal{N}(0,1)^2, \ \mathcal{N}(0,1)^3$

### Real applications

- Neural nets
- Phase retrieval

#### Preprints

LAST UPDATE	AUTHORS	TITLE
Apr. 2023	M. Bakhshizadeh	Algebra of Sub-Weibull Random Variables (download)

#### 

Inaugural CAMDA Conference

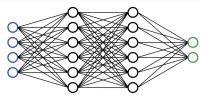
Milad Bakhshizadeh

# Rise of the heavy-tails: A new era dawns

- Data is heavy-tailed
- Multiplication makes tail heavier
  - $\blacktriangleright XY, X^n$
  - ▶  $\mathcal{N}(0,1), \ \mathcal{N}(0,1)^2, \ \mathcal{N}(0,1)^3$
- Real applications
  - Neural nets
  - Phase retrieval

#### Preprints





#### \*ロ> \*(型)> \*(目)> \*(目)> \*(目)> \*(の)

Inaugural CAMDA Conference

Milad Bakhshizadeh

Preprints

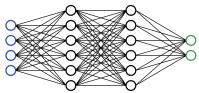
End 00

# Rise of the heavy-tails: A new era dawns

- Data is heavy-tailed
- Multiplication makes tail heavier
  - $\blacktriangleright$  XY, X<sup>n</sup>
  - ▶  $\mathcal{N}(0,1), \ \mathcal{N}(0,1)^2, \ \mathcal{N}(0,1)^3$
- Real applications
  - Neural nets
  - Phase retrieval

# LAST AUTHORS TITLE UWANT Agr.2023 M. Bababasehn Algebra of Sub-Weibull Random Variables (download)

< □ > < 同 >



#### Inaugural CAMDA Conference

#### Milad Bakhshizadeh

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics



### U-statistics, a low risk unbiased estimator

► 
$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < i_2 < ... < i_m \le n} h(X_{i_1}, ..., X_{i_m}), \quad h(\cdot) \text{ symmetric}$$
  
►  $X_i \text{ iid}$   
►  $U_n \to \mathbb{E}[h]$   
►  $h = X_1 \qquad \rightarrow \hat{\mu} = \bar{X}_n$   
►  $h = \frac{(X_1 - X_2)^2}{2} \rightarrow \hat{\sigma}^2$   
► How fast  $\mathbb{P}\left(|U_n - \mathbb{E}[h]| > \epsilon\right) \rightarrow 0$ ?  
► Sample size  $n$   
► High-dimensional statistics

Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics



### U-statistics, a low risk unbiased estimator

► 
$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < i_2 < ... < i_m \le n} h(X_{i_1}, ..., X_{i_m}), \quad h(\cdot) \text{ symmetric}$$
  
►  $X_i \text{ iid}$   
►  $U_n \to \mathbb{E}[h]$   
►  $h = X_1 \quad \rightarrow \hat{\mu} = \bar{X}_n$   
►  $h = \frac{(X_1 - X_2)^2}{2} \rightarrow \hat{\sigma}^2$   
► How fast  $\mathbb{P}\left(|U_n - \mathbb{E}[h]| > \epsilon\right) \rightarrow 0$ ?  
► Sample size  $n$   
► High-dimensional statistics

Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics



### U-statistics, a low risk unbiased estimator

Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

## Concentration inequality, a tool for uncertainty control

$$\blacktriangleright \mathbb{P}\left(|U_n| > \epsilon\right) \le \exp\left(-L(n,\epsilon)\right)$$

•  $\mathbb{E}[h] = 0$  (no generality loss)

Simple

Tight (asymptotically)

Recall

$$U_n = \frac{1}{\binom{n}{m}} \sum h(X_{i_1}, ..., X_{i_m})$$
  
 
$$\mathbb{E} \left[ \exp \left( \lambda h(X_1, ..., X_m) \right) \right] = \infty, \quad \forall \lambda > 0$$

Milad Bakhshizadeh

Inaugural CAMDA Conference

E SQA

## Concentration inequality, a tool for uncertainty control

$$\blacktriangleright \mathbb{P}(|U_n| > \epsilon) \le \exp(-L(n, \epsilon))$$

•  $\mathbb{E}[h] = 0$  (no generality loss)

Simple

Tight (asymptotically)

Recall

$$U_n = \frac{1}{\binom{n}{m}} \sum h(X_{i_1}, ..., X_{i_m})$$
$$\mathbb{E} \left[ \exp \left( \lambda h(X_1, ..., X_m) \right) \right] = \infty, \quad \forall \lambda > 0$$

Milad Bakhshizadeh

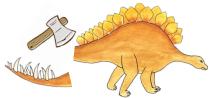
Inaugural CAMDA Conference

E SQA

Intro	Heavy-tail challenge	Setup	Truncation method	Optimality	End
00	000	00	●000	000	00
_					

#### Part 2

#### ► The Solution: Tail Truncation



#### もよい 正明 エルマ・エリ・エピ・トロ

Milad Bakhshizadeh

Inaugural CAMDA Conference



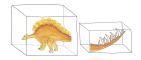
## Bound tail and body, separately

#### Define:

• 
$$\mathbf{k} = \lfloor \frac{n}{m} \rfloor$$

$$\blacktriangleright h_L = h \ \mathbb{1}(h \le L)$$

$$\blacktriangleright \mathbb{P}(h > t) \simeq \exp(-I(t)), \ I(t) \ll t$$



 $\blacktriangleright \mathbb{P}(U_n > t) \leq \mathbb{P}(U_n(h_L) > t) + \mathbb{P}(\exists i_j \mid h(X_{i_1}, ..., X_{i_m}) > L)$ 

#### Theorem (1)

$$\mathbb{P}\left(U_n > t\right) \lesssim \exp\left(-rac{kt^2}{2Var(h)}
ight) + \left(1 + \binom{n}{m}
ight)\exp\left(-I(kt)
ight)$$

Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

E SQA



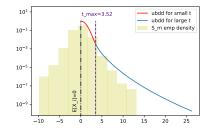
## There are two different regions of deviation

$$\mathbb{P}\left(U_n > t\right) \lesssim \exp\left(-\frac{kt^2}{2\operatorname{Var}(h)}\right) + \left(1 + \binom{n}{m}\right)\exp\left(-I(kt)\right)$$



- Small *t*, Gaussian decay
- Large t, like exp (-I(kt))

• Change point: 
$$kt^2 \simeq I(kt)$$



Milad Bakhshizadeh

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

Inaugural CAMDA Conference

-

**Derivation**  

$$\mathbb{P}\left(U_n > t\right) \leq \mathbb{P}\left(U_n(h_L) > t\right) + \mathbb{P}\left(\exists i_j \mid h(X_{i_1}, ..., X_{i_m}) > L\right)$$

$$\leq e^{-\lambda t} \mathbb{E}\left[e^{\lambda U_n(h_L)}\right] + \binom{n}{m}e^{-l(L)}$$

$$\models \text{ issues}$$

$$\Rightarrow \text{ dependent terms in } U_n(h_L)$$

$$\Rightarrow \mathbb{E}\left[e^{\lambda U_n(h_L)}\right] \xrightarrow{L \to \infty} \infty, \text{ fixed } \lambda$$

$$\Rightarrow \text{ Optimize } \lambda, L \text{ together}$$

$$\Rightarrow \text{ Sharpness}$$
First term 
$$\leq \exp\left(-\frac{kt^2}{2\nu(kt,\beta\frac{l(kt)}{kt})}\right) + \exp\left(-\beta l(kt)\max(\frac{1}{2}, c(t, \beta, k))\right)$$

$$\Rightarrow \nu(L, \eta) \triangleq \mathbb{E}\left[h_L^2 1(h \leq 0) + h_L^2\exp(\eta h_L)1(h > 0)\right] \rightarrow Var(h)$$

$$\Rightarrow c(t, \beta, k) \triangleq 1 - \frac{\beta}{2t}\frac{l(kt)}{kt}\nu(kt, \beta\frac{l(kt)}{kt}) \rightarrow 1$$

$$= 1$$

$$= 1$$

$$= 1$$

Intro 00	Heavy-tail challenge 000	Setup 00	Truncation method 000●	Optimality 000	End 00
Deriv	ration				
	$\mathbb{P}\left(U_n>t ight)\leq\mathbb{P}$	$(U_n(h_L) > t$	$(\exists i_j \mid h(X_{i_1},,a_{i_k}))$	$(X_{i_m}) > L)$	
	$\leq e^{-1}$ issues	$\lambda t \mathbb{E}\left[\mathrm{e}^{\lambda U_n(h_n)}\right]$	$\binom{n}{m} = \binom{n}{m} e^{-I(L)}$		
	depender	nt terms in ${\cal L}$	$V_n(h_L)$		
	$\blacktriangleright \mathbb{E}\left[\mathrm{e}^{\lambda U_n(t)}\right]$	$\left  \xrightarrow{L \to \infty} \infty \right $	, fixed $\lambda$		
	Optimize	$\dot{\lambda}, L$ togethe	er		
	Sharpnes	S			
Fir	$rst term \leq \exp\left(-\frac{1}{2\nu}\right)$	$\left(\frac{kt^2}{(kt,\beta\frac{l(kt)}{kt})}\right) +$	$-\exp\left(-\beta l(kt)\max\left(\frac{1}{2}\right)\right)$	$\left[\frac{1}{2}, \boldsymbol{c}(t, \beta, k)\right)$	
	$\blacktriangleright$ $v(L,\eta) \triangleq \mathbb{E}\left[h_L^2 1(h_L^2)\right]$	$h \leq 0) + h_L^2 q$	$\exp(\eta h_L) 1(h > 0) ] =$	Var(h)	
	• $c(t,\beta,k) \triangleq 1 - \frac{\beta}{2t}$	$\frac{I(kt)}{kt}v(kt,\beta$	$rac{l(kt)}{kt})  ightarrow 1$ . The second secon	☞ · · · · · · · · · · · · · · · · · · ·	

Milad Bakhshizadeh

Intro 00	Heavy-tail challenge 000	Setup 00	Iruncation method 000●	Optimality 000	End 00
Deriv	ation				
	$\mathbb{P}\left(U_n>t ight)\leq\mathbb{P}$	$(U_n(h_L)>t$	$\mathbb{P}$ ) + $\mathbb{P}$ ( $\exists i_j \mid h(X_{i_1},, i_{i_j})$	$(X_{i_m}) > L)$	
	≤ e <sup>−</sup> ► issues	$-\lambda t \mathbb{E}\left[\mathrm{e}^{\lambda U_n(h_l)}\right]$	$\binom{n}{m} e^{-I(L)}$		
	depende	nt terms in $U$	$I_n(h_L)$		
	$\blacktriangleright \mathbb{E}\left[\mathrm{e}^{\lambda U_n(\lambda)}\right]$	$h_L$ ) $\xrightarrow{L \to \infty} \infty$	, fixed $\lambda$		
	Optimize	e $ec{\lambda}, L$ togethe	er		
	Sharpnes	55			
Fir	st term $\leq \exp\left(-\frac{1}{2\nu}\right)$	$\frac{kt^2}{(kt,\beta\frac{l(kt)}{kt})}\right) +$	$-\beta I(kt) \max(\frac{1}{2})$	$\frac{1}{2}, \boldsymbol{c(t,\beta,k)})$	
1	• $v(L,\eta) \triangleq \mathbb{E}\left[h_L^2 1(h)\right]$	$h \leq 0) + h_L^2 q$	$\exp(\eta h_L) 1(h > 0) \big] -$	→ Var(h)	
I	• $c(t,\beta,k) \triangleq 1 - \frac{\beta}{2t}$	$\frac{l(kt)}{kt}v(kt,\beta)$	(kt) / kt)  ightarrow 1	∄ → ∢ ≣ → ∢ ≣ → 星	= ∽ <b></b> < ભ

Truncation method

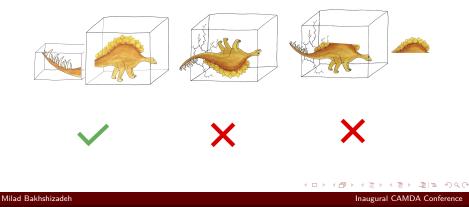
Milad Bakhshizadeh

Inaugural CAMDA Conference

Intro	Heavy-tail challenge	Setup	Truncation method	Optimality	End
00	000	00	0000	●00	oo

### Part 3

#### ▶ Tail truncation is optimal (in several cases)





For large same size, the bound is tight

$$\mathbb{P}\left(U_n > t\right) \lesssim \exp\left(-\frac{kt^2}{2Var(h)}\right) + \left(1 + \binom{n}{m}\right) \exp\left(-I(kt)\right)$$



Optimality

obo

#### Theorem (2)

$$\lim_{k\to\infty}\frac{-\log\mathbb{P}(U_n>t)}{I(kt)}=1$$

Assumptions

*kt<sup>2</sup>* ≫ *l(kt) l(t)* ≥ *c* 
$$\sqrt[\infty]{t}$$
 → sub-Weibull
(*h* > *t*) ≃ (∃*i* |*X<sub>i</sub>*| > *f(t)*)

r

•  $h = |X_1 - X_2|, (X_1 - X_2)^2, \max(|X_1|, |X_2|, ..., |X_m|)$ 

#### Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

End



## For large same size, the bound is tight

$$\mathbb{P}\left(U_n > t\right) \lesssim \exp\left(-\frac{kt^2}{2Var(h)}\right) + \left(1 + \binom{n}{m}\right)\exp\left(-I(kt)\right)$$

r



### Theorem (2)

$$\lim_{k\to\infty}\frac{-\log\mathbb{P}(U_n>t)}{I(kt)}=1$$



#### Milad Bakhshizadeh

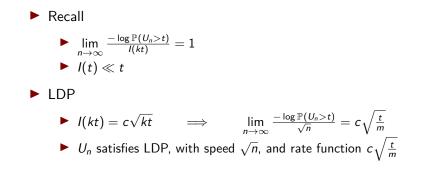
Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

ELE DOG

End

# It yields Large Deviation Principle (LDP)

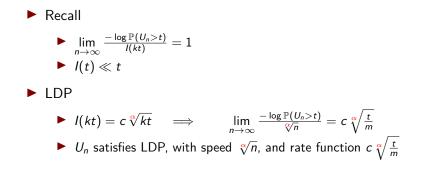


Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

# It yields Large Deviation Principle (LDP)



Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

## Future Work: This was a piece of the puzzle

Extend Heavy-tailed Analysis toolbox

Why only U-statistics?

$$\blacktriangleright F_n = F(X_1, ..., X_n) \xrightarrow{n \to \infty} \mathbb{E}[F]$$

- Applications
  - Finance

. . .

- Differential Privacy
- Asymptotic Hypothesis Testing
  - Bahadur Efficiency

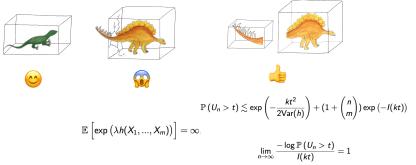
#### Milad Bakhshizadeh

Inaugural CAMDA Conference

E SQA

### Thanks

Truncation can turn dinos to lizards, so heavy-tails better beware - they're next in line for a makeover!



Cartoon characters credit: Mina Latifi

Milad Bakhshizadeh

Inaugural CAMDA Conference

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

EL OQO

## Another application

Phase retrieval

$$\mathbf{y} = |X\beta| + \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$
  
$$\hat{\beta} = \arg\min_{\mathbf{b}} \left\| \mathbf{y}^2 - (X\mathbf{b})^2 \right\|^2$$

Milad Bakhshizadeh

Inaugural CAMDA Conference

A D > 
 A D >
 A

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

三日 のへで

The extreme event:

One 
$$X_i$$
 large  $\implies \binom{n-1}{m-1}$  kernel terms large

▶ Recall:
■ U<sub>n</sub> = 1/(n) ∑ h(X<sub>i1</sub>,...,X<sub>im</sub>)
■ (∃i |X<sub>i</sub>| > f(t)) ≃ (h > t)
■ 1
■ 2 (∃i s.t. |X<sub>i</sub>| > f(kt))
■ (𝔅) ≃ exp (-I(kt))
■ 𝔅 ⇒ (U<sub>n</sub> > t), U<sub>n</sub> > 1/(n) (n-1)/(kt = m/n)/(kt = m/n)/(kt

Milad Bakhshizadeh

Exponential tail bounds and Large Deviation Principle for Heavy-Tailed U-Statistics

Inaugural CAMDA Conference