EXTREME BANDITS

A.CARPENTIER@STATSLAB.CAM.AC.UK *and* MICHAL.VALKO@INRIA.FR

UNIVERSITY OF CAMBRIDGE

Inria informatics mathematics

ANOMALY DETECTION

Goal: Find **extreme** values

- outburst of the network activity
- peak water flow
- biosurveillance

Challenge: **Heavy tails** of real-world distributions

New algorithm: EXTREMEHUNTER

Analysis: **Finite-time** performance guarantees

Prior work: Either heuristics or only asymptotic guarantees for parametric distributions

EXTREME REGRET

Protocol for learner π - Every time step

- each of the K *arms* emits a sample $X_{k,t} \sim P_k$
- learner π chooses some arm I_t
- learner π receives only $X_{I_t,t}$ (bandit setting)

Reward of learner π

- overall reward is the highest value found in n steps

$$G_n^\pi = \max_{t \le n} X_{I_t,t}$$

Reward for pulling the optimal arm $*$

- overall reward is the highest value found in n steps

$$\mathbb{E}\left[G_n^*\right] = \max_{k \le K} \mathbb{E}\left[\max_{t \le n} X_{k,t}\right]$$

Extreme regret in the bandit setting

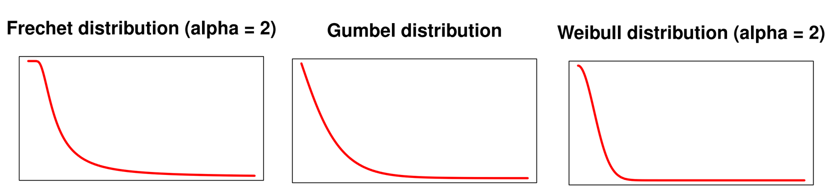
$$\mathbb{E}\left[R_n^\pi\right] = \mathbb{E}\left[G_n^*\right] - \mathbb{E}\left[G_n^\pi\right] = \max_{k \le K} \mathbb{E}\left[\max_{t \le n} X_{k,t}\right] - \mathbb{E}\left[\max_{t \le n} X_{I_t,t}\right]$$

EXTREMAL TYPES FOR MAXIMA

What is the distribution of the maximum?

- Fisher-Tippett-Gnedenko theorem
  - analogue of central limit theorem for averages
  - limiting distribution for maximum is one of the three

Fréchet	**Gumbel**	**Weibull**
$1 - e^{-x^{-\alpha}}$	$1 - e^{-e^{-x}}$	$e^{-x^{\alpha}}$



- α-Fréchet is defined as

$$P(x) = \exp\left\{-\left(\frac{x-m}{s}\right)^{\alpha}\right\}.$$

- FTP theorem: 'P converges to an α-Fréchet distribution' $\equiv$ '$1 - P$ is a $-\alpha$ regularly varying function in the tail'

***Approximately* α-Pareto distribution (slightly more restrictive)**

$$\lim_{x \to \infty} \frac{|1 - P(x) - Cx^{-\alpha}|}{x^{-\alpha}} = 0$$

- does ensure a limit
- does not ensure a convergence rate

Second order **Pareto** condition also known as the *Hall condition*

$$\left|1 - P(x) - Cx^{-\alpha}\right| \le C' x^{-\alpha(1+\beta)}$$

- also **ensures a convergence rate**
- equivalent to $P(x) = 1 - Cx^{-\alpha} + \mathcal{O}\left(x^{-\alpha(1+\beta)}\right)$
- similar to approximate Pareto for small β

How does it characterize a tail?

- β is the rate of the convergence (when x diverges to infinity) of the tail of P to the tail of $1 - Cx^{-\alpha} \equiv$ Pareto
- α is the heaviness of the tail
- the smaller the α, the heavier the tail

Learn αs and identify the smallest one among the sources.

EXPECTATION OF THE MAXIMUM

What is the expectation of the maximum of a 2$^{\text{nd}}$-order Pareto?

Theorem 1. *Let $X_1, \dots, X_n$ be n i.i.d. samples drawn according to (α, β, C, C')-second order Pareto distribution P. If $\alpha > 1$, then:*

$$\left|\mathbb{E}(\max_i X_i) - (nC)^{\frac{1}{\alpha}}\Gamma\left(1 - \frac{1}{\alpha}\right)\right| \le \frac{4D_2}{n}(nC)^{\frac{1}{\alpha}} + \frac{2C' D_{\beta+1}}{C^{\beta+1} n^{\beta}}(nC)^{\frac{1}{\alpha}} + B$$

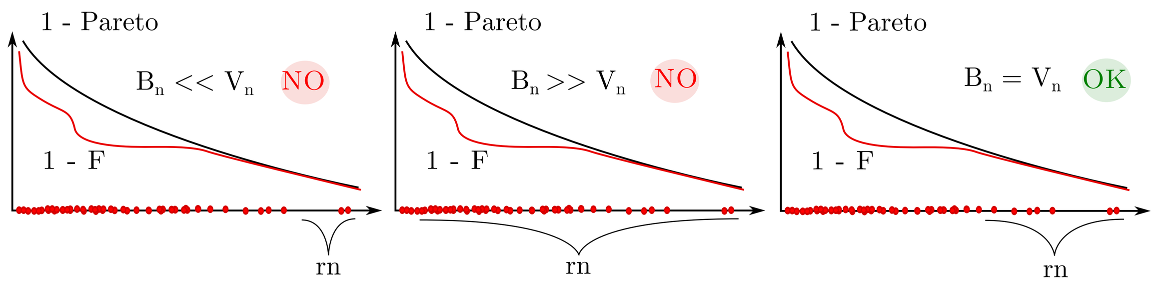
where $D_2, D_{1+\beta} > 0$ are some universal constants.

EFFICIENT ESTIMATORS

Estimator of α

Traditional Hill's estimator with r sample fraction

$$\widehat{\alpha}_t^H = \left(\frac{1}{rn}\sum_{u=1}^{r} \log \frac{X_{(n-i+1)}}{X_{(n-rn+1)}}\right)^{-1}$$



Adaptive (to β) tail estimator with concentration

$$\widehat{a}_t = \log\left(\frac{1}{n}\sum_{u=1}^{n} \mathbf{1}\left\{X_u > e^r\right\}\right) - \log\left(\frac{1}{n}\sum_{u=1}^{n} \mathbf{1}\left\{X_u > e^{r+1}\right\}\right) \quad (1)$$

$\widehat{a}_t$ comes with a high probability finite-time concentration guarantee

$$\left|\frac{1}{\alpha_k} - \widehat{h}_{k,t}\right| \le D\sqrt{\log(1/\delta)}\,T_{k,t}^{-b/(2b+1)} = B_1(T_{k,t}) \quad (2)$$

Associated estimator of constant C

$$\widehat{C}_{k,t} = T_{k,t}^{1/(2b+1)}\left(\frac{1}{T_{k,t}}\sum_{u=1}^{T_{k,t}} \mathbf{1}\left\{X_{k,u} \ge T_{k,t}^{\widehat{h}_{k,t}/(2b+1)}\right\}\right) \quad (3)$$

$\widehat{C}_{k,t}$ has also a high probability finite-time concentration guarantee

$$\left|C_k - \widehat{C}_{k,t}\right| \le E\sqrt{\log(T_{k,t}/\delta)}\log(T_{k,t})T_{k,T}^{-b/(2b+1)} = B_2(T_{k,t}) \quad (4)$$

EXTREMEHUNTER**'s UCB index**

$$\left(\left(\widehat{C}_{k,t} + B_2\left(T_{k,t}\right)\right)n\right)^{\widehat{h}_{k,t} + B_1(T_{k,t})} \bar{\Gamma}\left(\widehat{h}_{k,t}, B_1\left(T_{k,t}\right)\right) \quad (5)$$

EXTREMEHUNTER

Input and Initialization:
b: where $b \le \beta_k$ for all $k \le K$
N: minimum number of pulls of each arm
$T_k \leftarrow 0$ for all $k \le K$
$\delta \leftarrow \exp(-\log^2 n)/(2nK)$
Run:
for $t = 1$ **to** n **do**
  for $k = 1$ **to** K **do**
    if $T_k \le N$ **then**
      $B_{k,t} \leftarrow \infty$
    else
      estimate $\widehat{h}_{k,t}$ using (1) that verifies (2)
      estimate $\widehat{C}_{k,t}$ using (3) that verifies (4)
      update $B_{k,t}$ using (5) with (2) and (4)
    end if
  end for
  Play arm $k_t \leftarrow \arg\max_k B_{k,t}$
  $T_{k_t} \leftarrow T_{k_t} + 1$
end for

REGRET BOUND FOR EXTREMEHUNTER

Theorem 2. *Assume that the distributions of the arms are respectively $(\alpha_k, \beta_k, C_k, C')$ second order Pareto with $\min_k \alpha_k > 1$. If $n \ge Q$, the expected extreme regret of* EXTREMEHUNTER *is bounded from above as:*

$$\mathbb{E}\left[R_n\right] \le L(nC^*)^{\frac{1}{\alpha^*}}\left(\frac{K}{n}\log(n)^{\frac{2b+1}{b}} + n^{-\log(n)(1-\frac{1}{\alpha^*})} + n^{\frac{-b}{(b+1)\alpha^*}}\right)$$

where $L, Q > 0$ are some constants depending on $(\alpha_k, C_k)_k, C'$, and b.

REFERENCES

[1] John Mark Agosta, Jaideep Chandrashekar, Mark Crovella, Nina Taft, and Daniel Ting. Mixture models of endhost network traffic. In *IEEE Proceedings of INFOCOM*, 2013.

[2] Alexandra Carpentier and Arlene K. H. Kim. Adaptive and minimax optimal estimation of the tail coefficient. *Statistica Sinica*, 2014.

[3] Alexandra Carpentier and Arlene K. H. Kim. Honest and adaptive confidence interval for the tail coefficient in the Pareto model. *Electronic Journal of Statistics*, 2014.

[4] Vincent A. Cicirello and Stephen F. Smith. The max k-armed bandit: A new model of exploration applied to search heuristic selection. *AAAI*, 2005.

ANALYSIS

Step 1 • Favorable high probability event ξ of interest.

Step 2 • Given ξ, we bound the estimates of α_k and C_k, and use them to bound the main upper confidence bound.

Step 3 • With high probability we do not pull suboptimal arms too often.
• Guarantees that the number of pulls of the optimal arms $*$ is on ξ equal to n up to a negligible term.

Step 4 • Lower bound the expectation of maximum of the collected samples.
• Straightforward in classical bandits by the linearity of the expectation. Challenging in extreme bandits.
• Show that the expectation of maximum on ξ is not far away from the one without conditioning on ξ.

Remarks:

- It is not possible to learn β.
- The larger β, the easier the problem (parametric for $\beta = \infty$).
- The smaller α^*, the easier the problem (easier to identify).

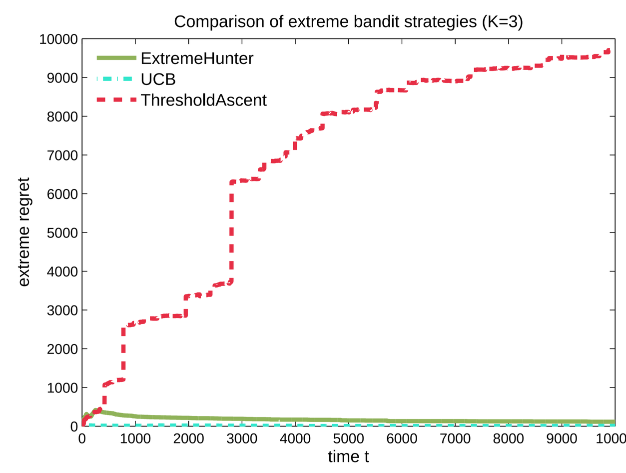
EXPERIMENTS

Comparison of extreme regret for:

- EXTREMEHUNTER
- UCB - mean-optimizing strategy
- THRESHOLDASCENT - state-of-the-art max-k strategy

Exact Pareto Distributions

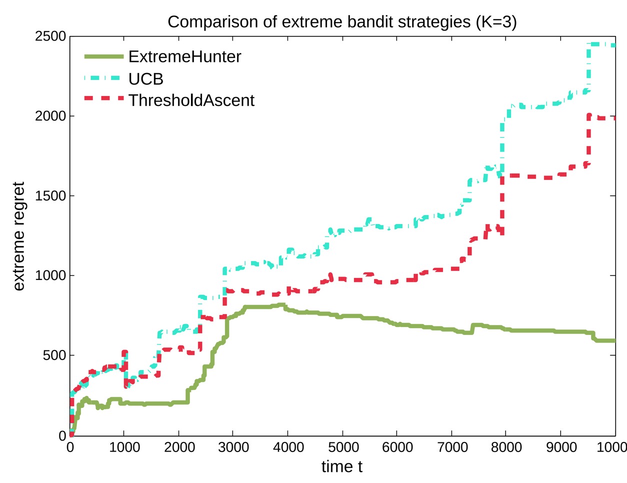
- 3 arms with $P_k(x) = 1 - x^{-\alpha_k}$, where $\alpha = [5, \mathbf{1.1}, 2]$
- the heaviest tail coincides with the largest mean



- easy parametric setup
- UCB performs well in terms of extreme regret

Approximate Pareto Distributions

- $P_1(x) = 1 - x^{-1.5}$ and $P_3(x) = 1 - x^{-3}$
- $P_2(x)$ is a mixture distribution
  - mixture weight of 0.8 of the Dirac(0)
  - mixture weight of 0.2 of $1 - x^{-1.1}$
- the second arm $\to$ the most heavy-tailed
- but the first arm $\to$ the largest mean



Computer Network Traffic Data

- heavy-tailed network traffic data
- collected from user laptops in the enterprise environment
- sample $\equiv$ number of network events in 4 seconds