Bregman Deviations of Generic Exponential Families

Patrick Saux[1]
(Joint work with Sayak Ray Chowdhury[2], Odalric-Ambrym Maillard[1] and Aditya Gopalan[3])

[1] Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 - CRIStAL, F-59000, Lille, France
[2] Boston University, Boston, Massachussetts, United States
[3] Indian Institute of Science, Bangalore, India

May 19, 2022

Table of Contents

Confidence sets

Let $X_1, \ldots, X_t$ i.i.d random variables distributed as $X \sim \nu \in \mathcal{P}(\mathbb{R}^d)$.

Where is  $= \mathbb{E}[X]$?

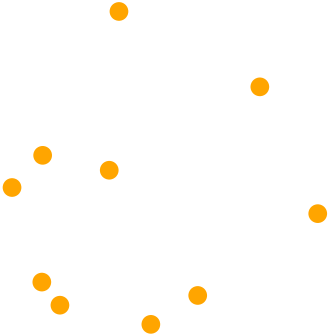
t=3

Confidence sets

Let $X_1, \ldots, X_t$ i.i.d random variables distributed as $X \sim \nu \in \mathcal{P}(\mathbb{R}^d)$.

Where is  $= \mathbb{E}[X]$?

t=10

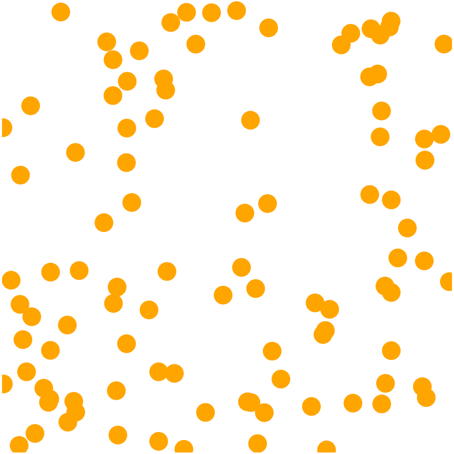


Confidence sets

Let $X_1, \ldots, X_t$ i.i.d random variables distributed as $X \sim \nu \in \mathcal{P}(\mathbb{R}^d)$.

Where is  $= \mathbb{E}[X]$?

t=100

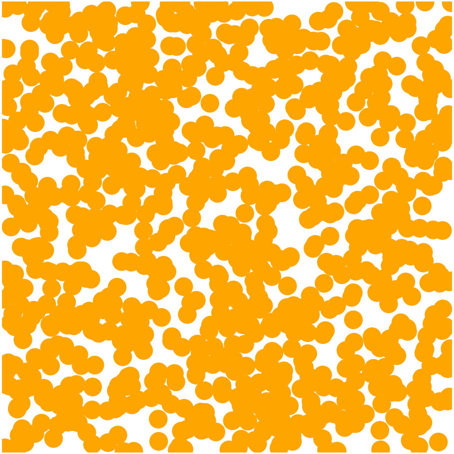


Confidence sets

Let $X_1, \ldots, X_t$ i.i.d random variables distributed as $X \sim \nu \in \mathcal{P}(\mathbb{R}^d)$.

Where is  $= \mathbb{E}[X]$?

t=1000

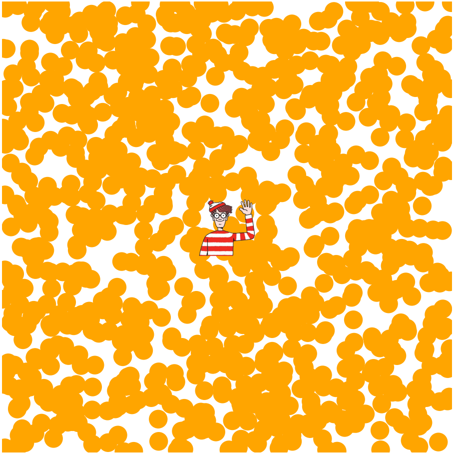


Confidence sets

Let $X_1, \ldots, X_t$ i.i.d random variables distributed as $X \sim \nu \in \mathcal{P}(\mathbb{R}^d)$.

Where is  $= \mathbb{E}[X]$?

t=1000



Confidence sets

We want **confidence sets**, not just estimation!

Confidence sets

We want **confidence sets**, not just estimation!
For $\delta \in (0,1)$ and $t \in \mathbb{N}$, we need a set $\mathcal{C}_t^\delta$ such that $\mathbb{P}\left(\mu \in \mathcal{C}_t^\delta\right) \geq 1 - \delta$.

Asymptotically:

$$C_t^\delta = \left[\widehat{\mu}_t - \frac{\sigma^2}{\sqrt{t}}\Phi^{-1}(1-\delta/2),\ \widehat{\mu}_t + \frac{\sigma^2}{\sqrt{t}}\Phi^{-1}(1-\delta/2)\right]$$

is such that

$$\lim_{t \to +\infty} \mathbb{P}\left(\mu \in \mathcal{C}_t^\delta\right) = 1 - \delta\,,$$

where $\widehat{\mu}_t = \frac{1}{t}\sum_{s=1}^t X_s$.

✖ Does not tell us anything about the small sample size regime...

Nonasymptotic confidence sets

We need some assumptions...

Sub-ψ distributions:

$$\forall \lambda \in \mathcal{I} \subseteq \mathbb{R}_+, \ \log \mathbb{E}_{X\sim\nu}\left[e^{\lambda(X-\mu)}\right] \leq \psi(\lambda)\,.$$

Proposition (Chernoff bound)

$$\mathbb{P}\left(\widehat{\mu}_t - \mu \geq \psi_*^{-1}\left(\frac{1}{t}\log\frac{1}{\delta}\right)\right) \leq \delta\,,$$

where $\psi_*(u) = \sup_{\lambda\in\mathcal{I}} \lambda u - \psi(\lambda)$ *is the Fenchel-Legendre conjugate of* ψ*.*

Examples:

- $\nu = \mathcal{N}(\mu, \sigma^2)$, $\psi(\lambda) = \sigma^2\lambda^2/2$ ·
- ν has bounded support, $\psi(\lambda) = diam(Supp(\nu))^2\lambda^2/8$ ·
- $\nu = \chi^2(k)$, $\psi(\lambda) = (1-2\lambda)^{-k/2}$, $\mathcal{I} = \left(0, \frac{1}{2}\right)$.

Nonasymptotic confidence sets

Proof:

$$\begin{aligned}
\mathbb{P}\left(\sum_{s=1}^{t} X_s - \mu \geq tu\right) &= \mathbb{P}\left(\prod_{s=1}^{t} e^{\lambda(X_s-\mu)} \geq e^{t\lambda u}\right) \\
&\leq e^{-t\lambda u}\mathbb{E}\left[\prod_{s=1}^{t} e^{\lambda(X_s-\mu)},\right] \quad \text{(Markov)} \\
&= e^{-t\lambda u}\prod_{s=1}^{t}\mathbb{E}\left[e^{\lambda(X_s-\mu)}\right] \quad \text{(independence)} \\
&\leq e^{-t\lambda u+t\psi(\lambda)} \quad \text{(sub-}\psi\text{)},
\end{aligned}$$

then optimise in $\lambda \in \mathcal{I}$. $\square$

Nonasymptotic confidence sets

Other possible assumptions:

- **Fully parametric**
  - If you know the quantiles, use them!
- **Bounded**
  - With control of moments:
    Bennett, Berstein Boucheron et al. [2013], Bentkus [2004],
  - With empirical estimators of moments:
    Bernstein Maurer and Pontil [2009],
    Bentkus Kuchibhotla and Zheng [2021].
  - With only boundedness: Phan et al. [2021].
- **Self-normalised sums**
  - Bercu et al. [2015], Bercu and Touati [2019]

Is this all for mean estimation?

Detour: stochastic bandit

(a) 15% chance to win 10€

(b) 5% chance to win 10€

The casino does not tell you which one is the winner!

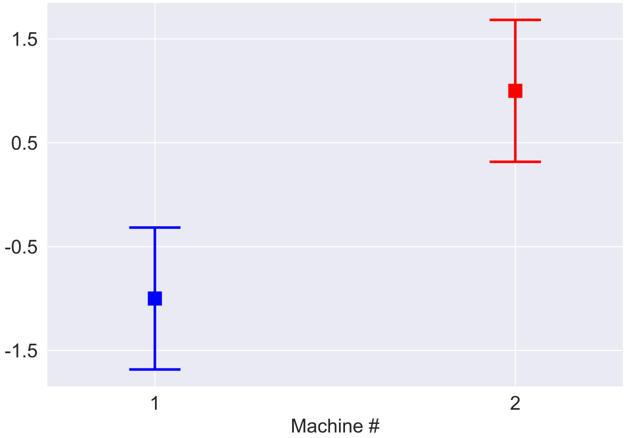
Optimism in face of uncertainty

- **Optimism**: play the machine with highest *plausible* reward.
- More formally:
  - play $\operatorname{argmax}_{k=1,2} U^k_{\boldsymbol{\tau_k}}(\delta)$,
  - τ_k: number of pulls to machine k (random stopping time),
  - $U^k_{\boldsymbol{\tau_k}}(\delta)$ is a measurable function of $\boldsymbol{\tau_k}$ samples from machine k such that

$$\mathbb{P}\left(\mu \geq U^k_{\boldsymbol{\tau_k}}(\delta)\right) \leq \delta.$$

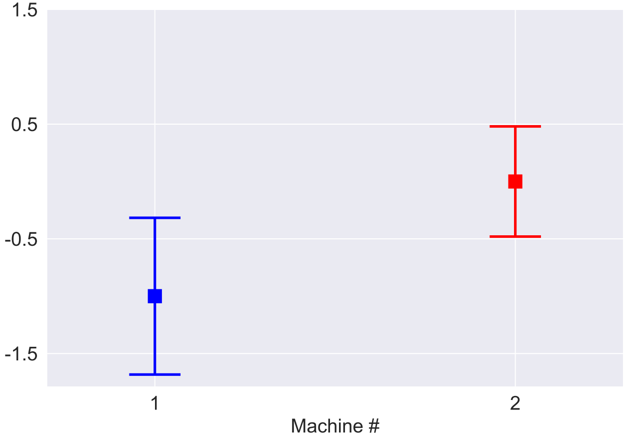
Concentration bound for sample of **random** size!

Optimism in face of uncertainty



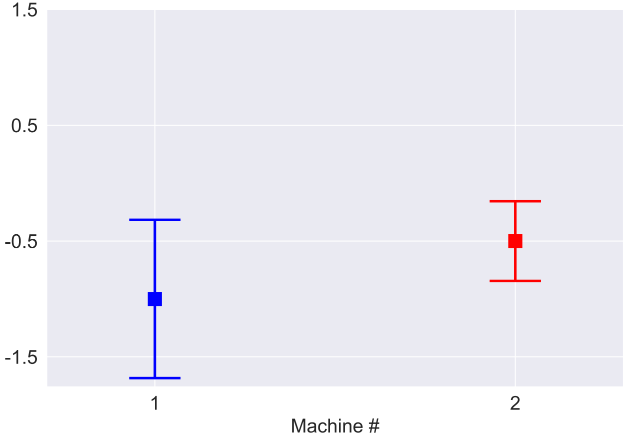
Machine 1: 5 pulls. Machine 2: 5 pulls.

Optimism in face of uncertainty



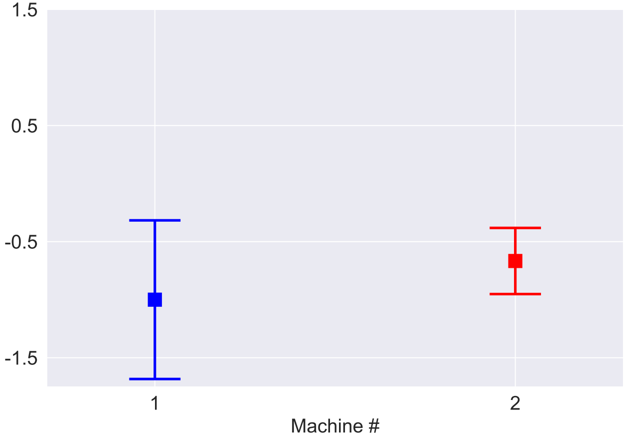
Machine 1: 5 pulls. Machine 2: 10 pulls.

Optimism in face of uncertainty



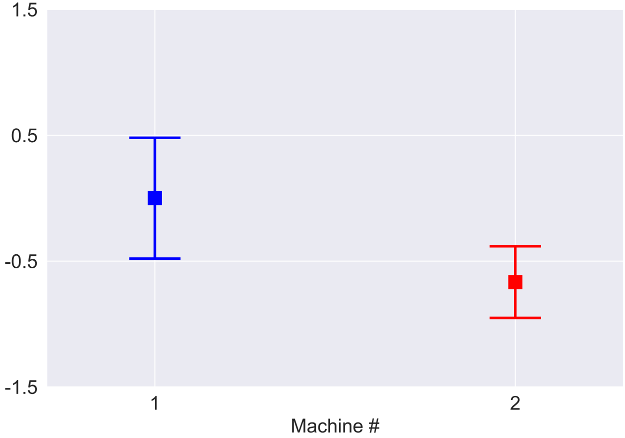
Machine 1: 5 pulls. Machine 2: 20 pulls.

Optimism in face of uncertainty



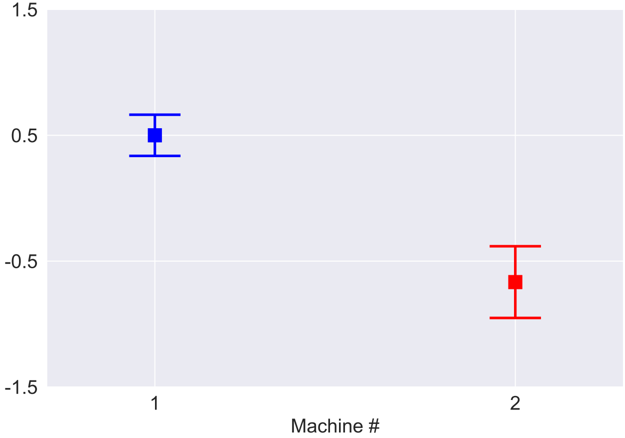
Machine 1: 5 pulls. Machine 2: 30 pulls.

Optimism in face of uncertainty



Machine 1: 10 pulls. Machine 2: 30 pulls.

Optimism in face of uncertainty



Machine 1: 100 pulls. Machine 2: 30 pulls.

Table of Contents

Martingale and stopping time

- Stopping times are hard to deal with...
- ... but go well with martingales!

For $t \in \mathbb{N}$, assume we know an invertible, nondecreasing $F_t : \mathbb{R} \to \mathbb{R}_+$ s.t

(i) $M_t = F_t(S_t)$ defines a supermartingale adapted to the filtration $\mathcal{F}_t = \sigma(X_s, s \leq t)$, with $S_t = \sum_{s=1}^{t} X_s - \mu$;

(ii) $\mathbb{E}[M_0] \leq 1$.

Then for any $\mathcal{F}$-stopping time τ,

$$\begin{aligned}
\mathbb{P}\left(S_\tau \geq F_\tau^{-1}\left(\frac{1}{\delta}\right)\right) &= \mathbb{P}\left(M_\tau \geq \frac{1}{\delta}\right) \\
&\leq \delta \mathbb{E}[M_\tau] \quad \text{(Markov)} \\
&\leq \delta \mathbb{E}[M_0] \quad \text{(Doob)} \\
&\leq \delta \,.
\end{aligned}$$

To find a martingale

The sub-ψ assumption is really a martingale condition:

$$\forall \lambda \in \mathcal{I},\ M^\lambda = \left(e^{\lambda S_t - t\psi(\lambda)}\right)_{t\in\mathbb{N}} \text{ is a nonnegative supermartingale.}$$

Proposition

For any $\mathcal{F}$-stopping time τ,

$$\forall \lambda \in \mathcal{I},\ \mathbb{P}\left(\widehat{\mu}_\tau - \mu \geq \frac{1}{\lambda}\left(\frac{1}{\tau}\log\frac{1}{\delta} + \psi(\lambda)\right)\right) \leq \delta\,.$$

Cannot optimise in λ for all values of τ!
(optimising for a fixed time t_0 recovers the Fenchel-Legendre formula).

To find a good martingale

For any mixture density $q(\lambda)$ over $\mathcal{I}$,

$$M = \left(\int_{\mathcal{I}} e^{\lambda S_t - t\psi(\lambda)} q(\lambda) d\lambda\right)_{t\in\mathbb{N}}$$ is also a nonnegative supermartingale.

Proposition (Chernoff-Laplace mixture bound (sub-Gaussian))

If $\psi(\lambda) = \frac{\sigma^2\lambda^2}{2}$, then for any $\mathcal{F}$-stopping time τ and any $c > 0$,

$$\mathbb{P}\left(\widehat{\mu}_\tau - \mu \geq \sigma\sqrt{\frac{2\left(1+\frac{c}{\tau}\right)\log\left(\frac{\sqrt{\frac{\tau}{c}+1}}{\delta}\right)}{\tau}}\right) \leq \delta\,.$$

Remark: this corresponds to the mixture distribution $\mathcal{N}\left(0, \frac{1}{c}\right)$.

Table of Contents

Exponential families

Parametric family indexed by $\theta \in \Theta$ (open set) of distributions ν_θ over $\mathbb{R}^d$ given by

$$\frac{d\nu_\theta}{d\nu_{\theta_\circ}}(x) = h(x) e^{\langle \theta, F(x) \rangle - \mathcal{L}(\theta)} .$$

- F: feature function (of $x \in \mathbb{R}^d$),
- $\mathcal{L}$: log-partition function (of $\theta \in \Theta$), convex, twice differentiable.
  - Assume $\det \nabla^2 \mathcal{L}(\theta) > 0$ for all $\theta \in \Theta$.

Bregman divergence:

$$\begin{aligned} \mathcal{B}_{\mathcal{L}}(\theta', \theta) &= \mathcal{L}(\theta') - \mathcal{L}(\theta) - \langle \theta' - \theta, \nabla \mathcal{L}(\theta) \rangle \\ &= KL\left(\nu_\theta \| \nu_{\theta'}\right) . \end{aligned}$$

Examples

Gaussian $\mathcal{N}\left(\mu, \sigma^2\right)$ with known variance σ^2

$$\theta = \mu, \Theta = \mathbb{R}\,,$$
$$\mathcal{B}_{\mathcal{L}}(\theta', \theta) = \frac{\left(\theta' - \theta\right)^2}{2\sigma^2}$$

Gaussian $\mathcal{N}\left(\mu, \sigma^2\right)$

$$\theta = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^{\top}, \Theta = \mathbb{R} \times \mathbb{R}^*_-\,,$$
$$\mathcal{B}_{\mathcal{L}}(\theta', \theta) = \frac{1}{2}\log\frac{\theta_2}{\theta'_2} + \frac{\theta'_2}{2\theta_2} - \theta'_2\left(\frac{\theta'_1}{2\theta'_2} - \frac{\theta_1}{2\theta_2}\right)^2 - \frac{1}{2}.$$

Bernoulli $\mathcal{B}(p)$

$$\theta = p, \Theta = (0, 1)\,,$$
$$\mathcal{B}_{\mathcal{L}}(\theta', \theta) = \theta\log\frac{\theta}{\theta'} + (1 - \theta)\log\frac{1 - \theta}{1 - \theta'}$$

Bregman supermartingale

Lemma

For $\theta \in \Theta$ and λ s.t $\theta + \lambda \in \Theta$,

$$\log \mathbb{E}_\theta \left[e^{\langle \lambda, F(X) - \mathbb{E}_\theta[F(X)] \rangle} \right] = \mathcal{B}_{\mathcal{L}}(\theta + \lambda, \theta) \,.$$

Consequence: let $\widehat{\mu}_t = \frac{1}{t} \sum_{s=1}^{t} F(X_s)$ and $\mu = \mathbb{E}_\theta [F(X)]$, then

$\forall \lambda,\ M^\lambda = \left(e^{\langle \lambda, t(\widehat{\mu}_t - \mu) \rangle - t\mathcal{B}_{\mathcal{L}}(\theta+\lambda, \theta)} \right)_{t \in \mathbb{N}}$ is a nonnegative (super)martingale.

Mixture: for $c > 0$,

$$q_\theta(\lambda | c) \propto e^{\langle \theta + \lambda, c \nabla \mathcal{L}(\theta) \rangle - c\mathcal{L}(\theta)} \,,$$

$$M_t = \int M_t^\lambda q_\theta(\lambda | c) d\lambda \,.$$

Bregman-Laplace confidence set

Regularised parameter estimate:

$$\widehat{\theta}_{t,c}(\theta) = \nabla\mathcal{L}^{-1}\left(\frac{t}{t+c}\widehat{\mu}_t + \frac{c}{t+c}\mathcal{L}(\theta)\right) .$$

Bregman information gain:

$$\gamma_{t,c}(\theta) = \log \frac{\int_\Theta e^{-c\mathcal{B}_\mathcal{L}(\theta',\theta)}d\theta'}{\int_\Theta e^{-(t+c)\mathcal{B}_\mathcal{L}(\theta',\widehat{\theta}_{t,c}(\theta))}d\theta'} .$$

Theorem (Bregman-Laplace mixture bound for exponential families)

For any $\mathcal{F}$-stopping time τ and any $c > 0$,

$$\mathbb{P}\left((\tau+c)\mathcal{B}_\mathcal{L}\left(\theta, \widehat{\theta}_{\tau,c}(\theta)\right) \geq \log\frac{1}{\delta} + \gamma_{\tau,c}(\theta)\right) \leq \delta$$

Remarks

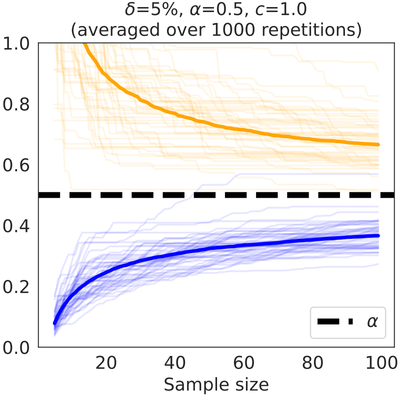
$$\mathbb{P}\left((\tau + c)\mathcal{B}_{\mathcal{L}}\left(\theta, \widehat{\theta}_{\tau,c}(\theta)\right) \geq \log\frac{1}{\delta} + \gamma_{\tau,c}(\theta)\right) \leq \delta$$

- Laplace's method for approximating integrals: when $t \to +\infty$,

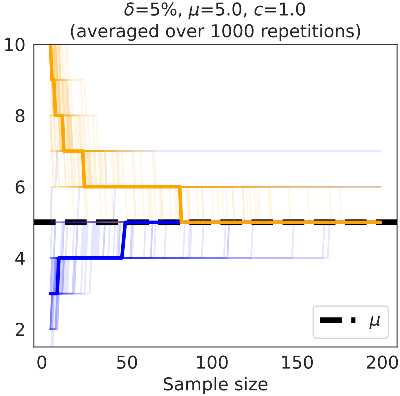
$$\gamma_{t,c}(\theta) = \frac{\dim\Theta}{2}\log\left(1 + \frac{t}{c}\right) + \mathcal{O}(1)\,.$$

- Implicit confidence set...
  - ...but convex: easy numerical solution.

Numerical experiments



(a) Pareto



(b) Chi-square

Numerical experiments

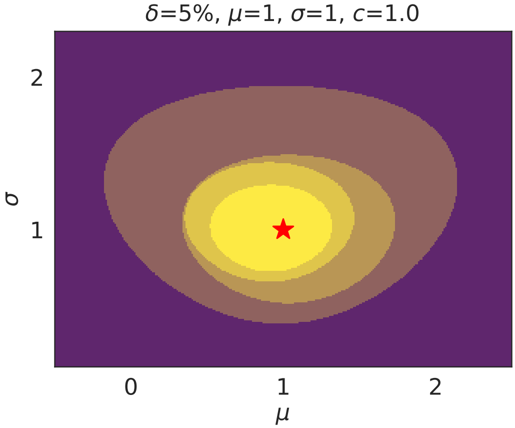
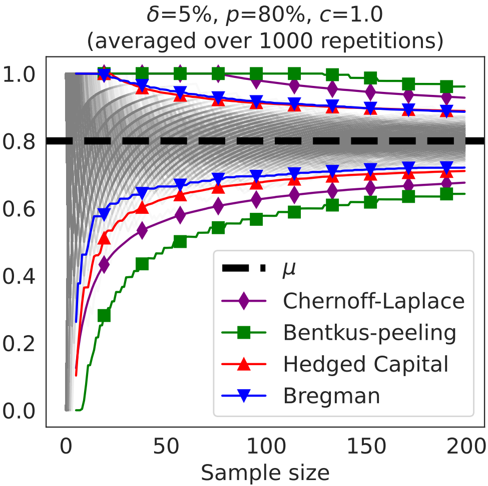


Figure: Gaussian (mean and variance) for $t \in \{10, 25, 50, 100\}$ observations

Numerical experiments



Comparison of median confidence envelopes around the mean for $\mathcal{B}(0.8)$. Grey lines are trajectories of empirical means $\widehat{\mu}_n$.

Examples

Table 2: Summary of Bregman confidence sets given by Theorem 3.2 for representative families
Throughout, the following notations are used:

$S_n = \sum_{t=1}^n X_t$, $\widehat{\mu}_n = \frac{S_n}{n}$, $Z_n(\mu,\sigma) = \frac{1}{\sigma^2}\sum_{t=1}^n (X_t-\mu)^2$,
$S_n^{(k)} = \sum_{t=1}^n X_t^k$, $L_n = \sum_{t=1}^n \log X_t$, $K_n = \sum_{t=1}^n \log X_t/2$,
$I(a,b) = \int_{-\infty}^{+\infty} e^{-ae^\theta+b\theta} d\theta$,
$J(a,b) = \int_0^\infty \exp\left(-a\log\Gamma\left(\frac{k}{2}\right)+b\frac{k}{2}\right) dk$ (if $k\in\mathbb{R}_+$),
$J(a,b) = \sum_{k'=1}^\infty \exp\left(-a\log\Gamma\left(\frac{k'}{2}\right)+b\frac{k'}{2}\right)$ (if $k\in\mathbb{N}$).

Name	Parameters	Formula
Gaussian	$\mu\in\mathbb{R}$	$\frac{1}{n+c}\frac{(S_n-n\mu)^2}{2\sigma^2} \leqslant \log\frac{1}{\delta}+\frac{1}{2}\log\frac{n+c}{c}$
Gaussian	$\sigma\in\mathbb{R}_+$	$\frac{Q_n(\mu)}{2\sigma^2}-\left(\frac{n+c}{2}+1\right)\log\left(\frac{Q_n(\mu)}{2\sigma^2}+\frac{c}{2}\right)$ $\leqslant \log\frac{1}{\delta}+\log\frac{\Gamma(\frac{c}{2})}{\Gamma(\frac{n+c}{2})}-\log\left(\frac{n+c}{2}\right)-\frac{c}{2}\log\frac{c}{2}$
Gaussian	$\mu\in\mathbb{R}$ $\sigma\in\mathbb{R}_+$	$\frac{1}{2}Z_n(\mu,\sigma)-\frac{n+c+3}{2}\log\left(\frac{n}{n+c}Z_n(\widehat{\mu}_n,\sigma)+\frac{c}{n+c}Z_n(\mu,\sigma)+c\right)$ $\leqslant \log\frac{1}{\delta}-\frac{n}{2}\log 2-\left(\frac{c}{2}+2\right)\log c+\frac{1}{2}\log(n+c)$ $+\log\Gamma\left(\frac{c+3}{2}\right)-\log\Gamma\left(\frac{n+c+3}{2}\right)$
Bernoulli	$\mu\in[0,1]$	$S_n\log\frac{1}{\mu}+(n-S_n)\log\frac{1}{1-\mu}+\log\frac{\Gamma(S_n+c\mu)\Gamma(n-S_n+c(1-\mu))}{\Gamma(c\mu)\Gamma(c(1-\mu))}$ $\leqslant \log\frac{1}{\delta}+\log\frac{\Gamma(n+c)}{\Gamma(c)}$
Exponential	$\mu\in\mathbb{R}_+$	$\frac{S_n}{\mu}-(n+c+1)\log\left(\frac{S_n}{\mu}+c\right)$ $\leqslant \log\frac{1}{\delta}+\log\frac{\Gamma(c)}{\Gamma(n+c)}-\log(n+c)-c\log c$
Gamma	$\lambda\in\mathbb{R}_+$	$\frac{S_n}{\lambda}-((n+c)k+1)\log(\frac{S_n}{\lambda}+ck)$ $\leqslant \log\frac{1}{\delta}+\log\frac{\Gamma(ck)}{\Gamma((n+c)k)}-\log((n+c)k)-ck\log ck$
Weibull	$\lambda\in\mathbb{R}_+$	$\frac{S_n^{(k)}}{\lambda^k}-(n+c+1)\log\left(\frac{S_n^{(k)}}{\lambda^k}+c\right)$ $\leqslant \log\frac{1}{\delta}+\log\frac{\Gamma(c)}{\Gamma(n+c)}-\log(n+c)-c\log c$
Pareto	$\alpha\in\mathbb{R}$	$\alpha L_n-(n+c+1)\log(\alpha L_n+c)$ $\leqslant \log\frac{1}{\delta}+\log\frac{\Gamma(c)}{\Gamma(n+c)}-\log(n+c)-c\log c$
Poisson	$\lambda\in\mathbb{R}_+$	$n\lambda-S_n\log\lambda$ $\leqslant \log\frac{1}{\delta}+\log I(c,c\lambda)-\log I(n+c,S_n+c\lambda)$
Chi-square	$k\in\mathbb{N}$ or $k\in\mathbb{R}_+$	$n\log\Gamma\left(\frac{k}{2}\right)-\frac{k}{2}K_n-\log J\left(c,c\psi_0\left(\frac{k}{2}\right)\right)$ $+\log J\left(n+c,K_n+c\psi_0\left(\frac{k}{2}\right)\right)\leqslant\log\frac{1}{\delta}$

Questions?


THAT CONCLUDES MY TWO-HOUR PRESENTATION. ANY QUESTIONS?
scottadams@aol.com
www.dilbert.com
DID YOU INTEND THE PRESENTATION TO BE INCOMPREHENSIBLE, OR DO YOU HAVE SOME SORT OF RARE "POWER-POINT" DISABILITY?
© 2003 United Feature Syndicate, Inc.
ARE THERE ANY QUESTIONS ABOUT THE CONTENT?
THERE WAS CONTENT?

References I

V. Bentkus. On hoeffding's inequalities. The Annals of Probability, 32(2):1650–1673, 2004.

B. Bercu and T. Touati. New insights on concentration inequalities for self-normalized martingales. Electronic Communications in Probability, 24:1–12, 2019.

B. Bercu, B. Delyon, and E. Rio. Concentration inequalities for sums and martingales. Springer, 2015.

S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.

A. K. Kuchibhotla and Q. Zheng. Near-optimal confidence sequences for bounded random variables. In International Conference on Machine Learning, pages 5827–5837. PMLR, 2021.

A. Maurer and M. Pontil. Empirical bernstein bounds and sample variance penalization. 2009.

M. Phan, P. Thomas, and E. Learned-Miller. Towards practical mean bounds for small samples. In International Conference on Machine Learning, pages 8567–8576. PMLR, 2021.

Table of Contents

A bad idea: union bound

Say we have:

$$\forall t \in \{1, \ldots, T\},\ \mathbb{P}\left(\mu \in \mathcal{C}_t^\delta\right) \geq 1 - \delta\,.$$

Then:

$$\begin{aligned}
\mathbb{P}\left(\forall t \in \{1, \ldots, T\},\ \mu \in \mathcal{C}_t^\delta\right) &= 1 - \mathbb{P}\left(\bigcup_{t=1}^{T} \{\mu \notin \mathcal{C}_t^\delta\}\right) \\
&\geq 1 - \sum_{t=1}^{T} \mathbb{P}\left(\mu \notin \mathcal{C}_t^\delta\right) \\
&\geq 1 - T\delta\,.
\end{aligned}$$

A bad idea: union bound

Say we have:

$$\forall t \in \{1, \dots, T\},\ \mathbb{P}\left(\mu \in \mathcal{C}_t^\delta\right) \geq 1 - \delta\,.$$

Then:

$$\begin{aligned}\mathbb{P}\left(\forall t \in \{1, \dots, T\},\ \mu \in \mathcal{C}_t^\delta\right) &= 1 - \mathbb{P}\left(\bigcup_{t=1}^{T} \{\mu \notin \mathcal{C}_t^\delta\}\right) \\ &\geq 1 - \sum_{t=1}^{T} \mathbb{P}\left(\mu \notin \mathcal{C}_t^\delta\right) \\ &\geq 1 - T\delta\,.\end{aligned}$$

Proposition (Union bound)

$(\mathcal{C}_t^{\delta/T})_{t=1}^T$ *is a uniform confidence sequence at level* δ.

Turning a bad idea into a good one: time-peeling

Proposition (Doob's maximal inequality)

Let $(S)_{t\in\mathbb{N}}$ be a ***nonnegative supermartingale*** *w.r.t a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$. Then for all $p \geq 1$ and $\epsilon > 0$, it holds for all $T_0 < T$ that*

$$\mathbb{P}\left(\max_{T_0 \leq t \leq T} S_t \geq \epsilon\right) \leq \frac{\mathbb{E}\left[S_{T_0}^p\right]}{\epsilon^p}.$$

Idea: apply union bound over a geometric grid $(t_k)_{k\in\mathbb{N}}$ with $t_k = (1+\eta)^k$.

Turning a bad idea into a good one: time-peeling

Proposition (Doob's maximal inequality)

*Let $(S)_{t\in\mathbb{N}}$ be a **nonnegative supermartingale** w.r.t a filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$. Then for all $p \geq 1$ and $\epsilon > 0$, it holds for all $T_0 < T$ that*

$$\mathbb{P}\left(\max_{T_0\leq t\leq T} S_t \geq \epsilon\right) \leq \frac{\mathbb{E}\left[S_{T_0}^p\right]}{\epsilon^p}\,.$$

Idea: apply union bound over a geometric grid $(t_k)_{k\in\mathbb{N}}$ with $t_k = (1+\eta)^k$.

Proposition (Geometric time-peeling)

Let τ a $(\mathcal{F}_t)_{t\in\mathbb{N}}$-stopping time. Then for $\eta > 0$ and $\delta \in (0,1)$,

$$\mathbb{P}\left(\widehat{\mu}_\tau - \mu \geq \psi_*^{-1}\left(\frac{1+\eta}{\tau}\log\left(\frac{\log(\tau)\log\left((1+\eta)\tau\right)}{\delta\log^2(1+\eta)}\right)\right)\right) \leq \delta$$