Generalizing classical probability bounds via convex optimization: Chernoff bound[1]

Presenter: Xiaopeng Li
UNI: xl3040

December 6, 2020

Introduction

Let X be random variable on $\mathbf{R}$. Classical Chernoff bound:

$$\mathbf{Prob}(X \geq u) \leq \inf_{\lambda \geq 0} \{\mathbf{E}\, e^{\lambda(X-u)}\}$$

or equivalently by taking logarithm,

$$\log \mathbf{Prob}(X \geq u) \leq \inf_{\lambda \geq 0} \{-\lambda u + \log \mathbf{E}\, e^{\lambda X}\} \tag{1}$$

Some observations:

- $\log \mathbf{E}\, e^{\lambda X}$ is always convex;
- If $X \sim \mathcal{N}(0,1)$, $\mathbf{Prob}(X \geq u) \leq e^{-u^2/2}$ for $u \geq 0$;

Our goal is to generalize the classical chernoff bound for X on arbitrary subset $C \subset \mathbf{R}^n$, i.e., using convex optimization to find an upper bound for $\mathbf{Prob}(X \in C)$.

Formulation

Let $\lambda \in \mathbf{R}^m$, $\mu \in \mathbf{R}$, and $f\colon \mathbf{R}^m \mapsto \mathbf{R}$ given by $f(z) = e^{\lambda^{\mathrm{T}} z + \mu}$.

- If $\lambda^{\mathrm{T}} z + \mu \geq 0$, $f(z) \geq \mathbb{1}_C(z)$ for all $z \in \mathbf{R}^m$;
- If $f(z) \geq \mathbb{1}_C(z)$, $\mathbf{Prob}(X \in C) = \mathbf{E}\, \mathbb{1}_C(X) \leq \mathbf{E}\, f(X)$;
- Taking logarithm, $\log \mathbf{Prob}(X \in C) \leq \mu + \log \mathbf{E}\, e^{\lambda^{\mathrm{T}} X}$.

The general Chernoff bound can be obtained by

$$
\begin{aligned}
\log \mathbf{Prob}(X \in C) &\leq \inf\{\mu + \log \mathbf{E}\, e^{\lambda^{\mathrm{T}} X} \mid \mu \geq -\lambda^{\mathrm{T}} z,\ \forall\, z \in C\} \\
&= \inf_{\lambda,\, \mu}\{\mu + \log \mathbf{E}\, e^{\lambda^{\mathrm{T}} X} \mid \mu \geq \sup_{z \in C}(-\lambda^{\mathrm{T}} z)\} \\
&= \inf_{\lambda}\{\sup_{z \in C}(-\lambda^{\mathrm{T}} z) + \log \mathbf{E}\, e^{\lambda^{\mathrm{T}} X}\} \\
&= \inf_{\lambda}\{S_C(-\lambda) + \log \mathbf{E}\, e^{\lambda^{\mathrm{T}} X}\} \qquad (2)
\end{aligned}
$$

where S_C is the support function of C. Recall that support function of any set is convex, so (2) is a **convex** optimization problem.

Example

Let $X \in \mathbf{R}^m$ and $X \sim \mathcal{N}(\mathbf{0}, I_m)$, where I_m is $m \times m$ identity matrix. Take $C = \{x \in \mathbf{R}^m \mid Ax \le b\}$ to be a nonempty polyhedron.

- $\log \mathbf{E}\, e^{\lambda^{\mathrm{T}} x} = \lambda^{\mathrm{T}} \lambda / 2$;
- $S_C(y) = \sup\{y^{\mathrm{T}} x \mid Ax \le b\}$;
- Apply strong duality, $S_C = \inf\{b^{\mathrm{T}} u \mid A^{\mathrm{T}} u = y,\ u \ge 0\}$.

Therefore, the generalized chernoff bound described by (2) is

$$\log \mathbf{Prob}(X \in C) \le \inf_{\lambda,\, u} \{b^{\mathrm{T}} u + \lambda^{\mathrm{T}} \lambda / 2 \mid A^{\mathrm{T}} u + \lambda = 0,\ u \ge 0\}$$

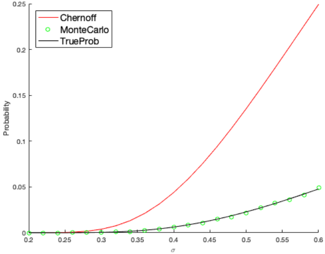
which requires to solve a quadratic programming problem:

$$\begin{array}{ll} \underset{u,\,\lambda}{\text{minimize}} & \lambda^{\mathrm{T}} \lambda / 2 + b^{\mathrm{T}} u \\ \text{s.t.} & A^{\mathrm{T}} u + \lambda = 0 \\ & u \ge 0 \end{array} \iff \begin{array}{ll} \underset{u}{\text{minimize}} & \|A^{\mathrm{T}} u\|_2^2 / 2 + b^{\mathrm{T}} u \\ \text{s.t.} & u \ge 0 \end{array}$$

Numerical Experiment

To illustrate the above example, take

- $X \in \mathbf{R}^2$ and $X \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_2)$;
- $C = \{x \in \mathbf{R}^2 \mid x_1 \in [-5, 5],\ x_2 \in [1, 5]\}$;



- TrueProb is calculated by probability density function of X; MonteCarlo is obtained by Monte Carlo simulation;
- The Chernoff bound becomes much higher than true probability when σ is large.

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.