The Cramér - Chernoff method ... and some exponential bounds

Jon A. Wellner

Department of Statistics, Box 354322
University of Washington
Seattle, WA 98195-4322

e-mail: jaw@stat.washington.edu

Abstract: We develop the Cramér - Chernoff method of deriving exponential bounds and use it to derive several classical exponential bounds

1. The Cramér-Chernoff method

Suppose that X is a random variable, and we want a bound for $P(X > x)$. Let $r > 0$. Then we can bound the tail probability $P(X > x)$ as follows. Since $r > 0$ and the exponential function is monotone increasing, application of Markov's inequality yields

$$\begin{aligned} P(X > x) &= P(rX > rx) = P(\exp(rX) > \exp(rx)) \\ &\le \exp(-rx) E \exp(rX). \end{aligned}$$

Since this is true for every $r >$ we can choose $r > 0$ to minimize the bound. Thus we conclude that

$$P(X > x) \le \inf_{r>0} \exp(-rx) E \exp(rX).$$

Example 1. Let $Z \sim N(0,1)$. Then since $E\exp(rZ) = e^{r^2/2}$, the Cramér-Chernoff method gives

$$P(Z > z) \le \inf_{r>0} e^{-rz} E \exp(rZ) = \inf_{r>0} \exp(-rz + r^2/2) = \exp(-z^2/2)$$

by choosing $r = z$. In this particular case we can do better. Note that for the bound we derived, $P(Z > 0) = 1/2$ while $\exp(-0^2/2) = 1$. It turns out that for $Z \sim N(0,1)$ an argument geared to this case yields

$$P(Z > z) \le \frac{1}{2} e^{-z^2/2} \quad \text{for all} \quad z > 0. \tag{1.1}$$

Furthermore, with ϕ and Φ denoting the standard normal density and distribution function,

$$\begin{aligned} P(Z > z) &= 1 - \Phi(z) = \int_z^\infty \phi(y) dy \le \int_z^\infty \frac{y}{z} \phi(y) dy \\ &= \frac{1}{z} \int_z^\infty -\phi'(y) dy \quad \text{since} \quad \phi'(y) = -y\phi(y) \\ &= \frac{1}{z} \phi(z). \end{aligned}$$

This inequality is known as *Mills' ratio*. Thus we have

$$P(Z > z) \leq \min\left\{\frac{1}{2}\exp(-z^2/2), \frac{1}{z\sqrt{2\pi}}\exp(-z^2/2)\right\}.$$

Note that the very general Cramér - Chernoff method captured the basic structure of these special bounds for the tail of the standard normal distribution.

Example 2. Now suppose that $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. Rademacher random variables (i.e. $P(\epsilon_1 = \pm 1) = 1/2$), and $a_1, \ldots, a_n \in \mathbb{R}$. Then with $S_n \equiv \sum_{j=1}^n a_j \epsilon_j$, we want to bound $P(S_n \geq t)$. Here the Cramér - Chernoff method yields, for $t > 0$,

$$P(S_n > t) \leq \inf_{r>0} \frac{Ee^{rS_n}}{e^{rt}}. \tag{1.2}$$

By the independence of the Rademacher random variables ϵ_j,

$$E\exp(rS_n) = \prod_{j=1}^n Ee^{ra_j\epsilon_j} = \prod_{j=1}^n \left\{\frac{1}{2}e^{ra_j} + \frac{1}{2}e^{-ra_j}\right\}. \tag{1.3}$$

Now the quantity in brackets on the right side is just $\cosh(ra_j)$ where

$$\begin{aligned}
\cosh(v) &= \frac{1}{2}(e^v + e^{-v}) \\
&= \frac{1}{2}\sum_{k=0}^\infty 2\frac{v^{2k}}{(2k)!} \leq \sum_{k=0}^\infty \frac{(v^2/2)^k}{k!} \\
&= \exp(v^2/2)
\end{aligned}$$

where the inequality follows from $(2k)! \geq 2^k k!$ for $k \geq 0$. Using this bound in (1.3) yields

$$E\exp(rS_n) \leq \exp(r^2\|a\|^2/2).$$

Plugging this into (1.2) yields

$$P(S_n > t) \leq \inf_{r>0} \exp(-rt + r^2\|a\|^2/2) = \exp\left(-\frac{t^2}{2\|a\|^2}\right)$$

by choosing $r = t/\|a\|^2$. This is sometimes known as *Hoeffding's inequality* since it follows from similar exponential bounds obtained by Hoeffding [1963]. It is a typical example of a *sub-Gaussian* tail bound.

Example 3. (A Poisson tail probability bound) Before proceeding to more general exponential bounds it is helpful to see what the Cramér-Chernoff method yields for a centered Poisson(ν) random variable. If $Y \sim$ Poisson(ν), then we have

$$P(Y - \nu \geq t) \leq \inf_{r>0} e^{-rt} E(\exp(r(Y-\nu)))$$

so we need to calculate the moment generating function of $Y - \nu$:

$$\begin{aligned}
E\exp(r(Y-\nu)) &= e^{-r\nu}\sum_{k=0}^\infty e^{rk}e^{-\nu}\frac{\nu^k}{k!} = e^{-r\nu-\nu}\sum_{k=0}^\infty \frac{(\nu e^r)^k}{k!} \\
&= e^{-r\nu-\nu}\exp(\nu e^r),
\end{aligned}$$

and it follows that

$$\psi_{Y-\nu}(r) \equiv \log E e^{r(Y-\nu)} = \nu(e^r - 1 - r) = \nu\varphi(r),$$

so that

$$\psi_{Y-\nu}(r) - rt = \nu(e^r - 1 - r) - rt$$

and the latter is minimized by $r^* = r^*(t) = \log(1 + t/\nu)$. Thus the resulting bound becomes

$$P(Y - \nu \ge t) \le \exp(-\nu h(1 + t/\nu)) \tag{1.4}$$

where $h(v) \equiv v(\log v - 1) + 1$.

The bound in the last display become more understandable if we define $\psi(x) \equiv (2/x^2)h(1+x)$ so that $(x^2/2)\psi(x) = h(1+x)$, and the bound in (1.4) becomes

$$P(Y - \nu \ge t) \le \exp\left(-\nu \frac{t^2/\nu^2}{2} \psi\left(\frac{t}{\nu}\right)\right) \quad = \quad \exp\left(-\frac{t^2}{2\nu}\psi\left(\frac{t}{\nu}\right)\right).$$

Here the function ψ satisfies $\psi(0) = 1$ and $\psi(x) \sim 2x^{-1}\log x$ as $x \to \infty$. Furthermore,

$$\psi(x) \ge \frac{1}{1 + x/3} \quad \text{for all} \quad x \ge 0, \tag{1.5}$$

so a less refined form of the bound is

$$P(Y - \nu \ge t) \le \exp\left(-\frac{t^2}{2(\nu + t/3)}\right).$$

This last inequality has the form of a Bernstein type inequality.

2. The exponential bounds of Bennett and Bernstein

In this section we first derive an exponential bound due to Bennett [1962]. We then derive a further (simpler) exponential bound which is due to Bernstein [1946].

Theorem. (Bennett's inequality) Suppose that $X_1, \ldots, X_n$ are independent random variables with finite variances and $X_j \le b$ for all $1 \le j \le n$. Let $S_n \equiv \sum_{i=1}^n (X_i - EX_i)$. and $\nu \equiv \sum_{i=1}^n E(X_i^2)$. Then, for all $r > 0$,

$$\log E e^{rS_n} \le n \log\left(1 + \frac{\nu}{nb^2}\varphi(br)\right) \le \frac{\nu}{b^2}\varphi(br)$$

where $\varphi(u) = e^u - 1 - u$. Furthermore, for any $t > 0$,

$$P(S_n > t) \le \exp\left(-\frac{\nu}{b^2} h\left(1 + \frac{bt}{\nu}\right)\right) \tag{2.1}$$

where $h(v) \equiv v(\log v - 1) + 1$.

Proof. By homogeneity we may assume that $b = 1$; if not, divide each X_j by b. Note that $u^{-2}\varphi(u)$ is a non-decreasing function of $u \in \mathbb{R}$ where we extend the function by continuity at 0. Hence for all $r > 0$ and $1 \le i \le n$

$$e^{rX_i} - 1 - rX_i \le X_i^2(e^r - 1 - r).$$

Taking expectations across the inequality in the last display yields

$$Ee^{rX_i} - 1 - rEX_i \le E(X_i^2)\varphi(r).$$

Since the cumulant generating function of a sum of independent random variables is the sum of the their cumulant generating functions, it follows that

$$\begin{aligned}
\psi_{S_n}(r) &\equiv \log Ee^{rS_n} \\
&\le \sum_{i=1}^n \left\{\log\left(1 + rEX_i + E(X_i^2)\varphi(r)\right) - rEX_i\right\} \\
&\le n\left\{\log\left(1 + r\frac{\sum_1^n EX_i}{n} + \frac{\nu}{n}\varphi(r)\right) - r\frac{\sum_1^n EX_i}{n}\right\} \\
&\qquad \text{by concavity of the log} \\
&\le n\left(\frac{\sum_1^n EX_i}{n} + \frac{\nu}{n}\varphi(r)\right) - r\frac{\sum_1^n EX_i}{n} \\
&\qquad \text{since } \ \log(1+u) \le u \\
&= \nu\varphi(r)
\end{aligned}$$

which has the same form as in the Poisson example in the previous section. After optimizing the choice of r the upshot is that (2.1) holds □

By using the function $\psi(x) = 2x^{-2}h(1+x)$ as in the last paragraph of section 1, we get an inequality which is called Bernstein's inequality:

Corollary: Suppose that the hypotheses of Bennett's inequality hold. Then

$$\begin{aligned}
P(S_n \ge t) &\le \exp\left(-\frac{\nu}{b^2}\frac{(bt/\nu)^2}{2}\psi\left(\frac{bt}{\nu}\right)\right) \\
&= \exp\left(-\frac{t^2}{2\nu}\psi\left(\frac{bt}{\nu}\right)\right) \\
&\le \exp\left(-\frac{t^2}{2\nu}\frac{1}{1+(bt)/(3\nu)}\right) \\
&= \exp\left(-\frac{t^2}{2(\nu + bt/3)}\right)
\end{aligned}$$

For a nice treatment of these and other exponential bounds via the Chernoff - Cramér method see Boucheron, Lugosi and Massart [2013]. For more on the functions h and ψ, see Shorack and Wellner [1986], chapters 10 and 11. For applications of the Hoeffding and Bernstein inequalities see van der Vaart and Wellner [1996] sections 2.2 - 2.5. For an introduction to exponential bounds for martingales and further applications, see Bercu, Delyon and Rio [2015].

3. Some exercises

Exercise 1. Plot the functions $h(v)$ and $\psi(v)$ for $v \ge 0$. Does the function ψ extend to $[-1, \infty)$?
Exercise 2. Find an upper bound for $P(-(Y - \nu) \ge t)$.
Exercise 3. Prove that (1.1) holds.
Exercise 4. Show that (1.5) holds.

References

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *J. Am. Statist. Assoc.* **57** 33-45.

Bercu, B., Delyon, B. and Rio, E. (2015). *Concentration inequalities for sums and martingales. SpringerBriefs in Mathematics*. Springer, Cham. MR3363542

Bernstein, S. N. (1946). *The Theory of Probabilities*. Gastehizdat Publishing House, Moscow.

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities*. Oxford University Press, Oxford. MR3185193

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York. MR838963

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes. Springer Series in Statistics*. Springer-Verlag, New York With applications to statistics. MR1385671