Hoeffding's inequality for sub-Gaussian random variables, confidence intervals determination

Thibault Bourgeron

August 9, 2020

Abstract

Hoeffding's inequality is generally stated for compactly supported random variables. The proof relies on an upper bound of the Laplace transform of such variables. It can be extended immediately to the space of variables having this property, the sub-Gaussian variables, which includes Gaussian variables for instance. We discuss how confidence intervals obtained with Hoeffding's inequality compare to more classically obtained intervals.

Keywords: Hoeffding's inequality, large deviations, concentration inequalities, Chernoff method, Laplace transform, moment-generating function, Legendre transform, sub-Gaussian variables, confidence interval

1 Hoeffding's inequality for sub-Gaussian random variables

1.1 Light tails, sub-Gaussian variables

Sub-gaussian distributions are *ad hoc* light tailed distributions; this lightness is measured in term of their Laplace.

Definition (sub-Gaussian). An integrable real random variable X is said to be *sub-Gaussian with parameter* σ, where $\sigma > 0$, if we have:

$$\forall s > 0 \qquad \mathbb{E}[\exp(s(X - \mathbb{E}[X]))] \leqslant \exp\left(\frac{1}{2}\sigma^2 s^2\right).$$

For instance, an $\mathcal{N}(m, \sigma^2)$ Gaussian random variable is sub-Gaussian with parameter σ (and the inequality in the definition holds as an equality).

Proposition. The sum of independent sub-Gaussian variables with parameters σ_k is sub-Gaussian with parameter $\left(\sum_{k=1}^n \sigma_k^2\right)^{1/2}$.

Chernoff method. Let X be an integrable real random variable. We assume the existence of the Legendre transform of the log Laplace transform of X defined on $\{s > 0\}$. We denote: φ_X^* the Legendre transform of: $\varphi_X(s) = \ln \mathbb{E}[\exp(s(X - \mathbb{E}[X]))]$ defined for $s > 0$. We have the upper tail inequality:

$$\forall t \in \mathbb{R} \qquad \mathbb{P}(X - \mathbb{E}[X] > t) \leqslant \exp(-\varphi_X^*(t)).$$

The same is true for the lower tail, and for both tails at the same time up to a factor 2, *cf.* proof of Hoeffding's inequality.

Proof. Thanks to Markov's inequality, we have, for any $s > 0$:

$$\begin{aligned}\mathbb{P}(X - \mathbb{E}[X] > t) &= \mathbb{P}(\exp(s(X - \mathbb{E}[X]) > \exp(st)))\\ &\leqslant \mathbb{E}[\exp(s(X - \mathbb{E}[X]))]\exp(-st) = \exp(\varphi_X(s) - st),\end{aligned}$$

and we get the result by minimizing the rhs over $s > 0$.

For instance, for an $\mathcal{N}(m, \sigma^2)$ Gaussian random variable X, we have:

$$\varphi_X(t) = \frac{\sigma^2 t^2}{2}, \qquad \varphi_X^*(t) = \frac{t^2}{2\sigma^2}.$$

More generally, a sub-Gaussian random variable X with parameter σ has exponentially decaying tails:

$$\forall t \in \mathbb{R} \qquad \mathbb{P}(X - \mathbb{E}[X] > t) \leqslant \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

1.2 Hoeffding's lemma and Hoeffding's inequality

Compactly supported random variables are light tailed in the previous sense.

Hoeffding's lemma (1963). A random variable which is (almost surely) supported in $[a, b]$, where: $a \leqslant b$ are finite, has a standard deviation less or equal to $\frac{1}{2}(b-a)$ and is sub-Gaussian with parameter $\frac{1}{2}(b-a)$.

Proof. Let Z be such a random variable. To obtain the first part, we write:

$$\operatorname{Var}(Z) \leqslant \mathbb{E}\left[\left(Z - \frac{a+b}{2}\right)^2\right] \leqslant \left(\frac{b-a}{2}\right)^2.$$

To establish the second part, we can assume that $\mathbb{E}[Z] = 0$. Indeed, changing Z in $Z - \mathbb{E}[Z]$ does not change the sub-Gaussian constant nor the translation invariant parameter $\frac{1}{2}(b-a)$. Let us denote: $\varphi(s) = \ln \mathbb{E}[\exp(sZ)]$, the log Laplace transform of Z. As Z is compactly supported it is an analytic function on $\{s > 0\}$. Simple computations lead to:

$$\forall s \geqslant 0 \qquad \varphi'(s) = \frac{\mathbb{E}[Z\exp(sZ)]}{\mathbb{E}[\exp(sZ)]}, \quad \varphi''(s) = \frac{\mathbb{E}[Z^2\exp(sZ)]}{\mathbb{E}[\exp(sZ)]} - \left(\frac{\mathbb{E}[Z\exp(sZ)]}{\mathbb{E}[\exp(sZ)]}\right)^2.$$

Denoting: Q the law of Z, $f_s(z) = \frac{\exp(sz)\,\mathbb{1}_{z\in[a,b]}}{\int_a^b \exp(su)\,du}$, $dQ_s = f_s(Z)\,dQ$, the previous equalities can be written as:

$$\forall s \geqslant 0 \qquad \varphi'(s) = \mathbb{E}_{Q_s}[Z], \quad \varphi''(s) = \operatorname{Var}_{Q_s}[Z].$$

The first part implies: $\forall s \geqslant 0 \quad \varphi''(s) \leqslant \left(\frac{b-a}{2}\right)^2$, which implies the result using: $\varphi(0) = \varphi'(0) = 0$.

Hoeffding's inequality. Let X_k's be sub-Gaussian variables with parameters σ_k, for $k = 1, \ldots, n$. We assume that the X_k's are independent. We denote: $\overline{X} = \frac{1}{n}\sum_{k=1}^n X_k, \overline{m} = \frac{1}{n}\sum_{k=1}^n \mathbb{E}[X_k], \overline{\sigma^2} = \frac{1}{n}\sum_{k=1}^n \sigma_k^2$. Then, we have:

$$\forall t \in \mathbb{R} \qquad \mathbb{P}\left(\left|\overline{X} - \overline{m}\right| > t\right) \leqslant 2\exp\left(-\frac{nt^2}{2\overline{\sigma^2}}\right).$$

Proof. It is enough to prove the one-sided inequality:

$$\forall t \in \mathbb{R} \qquad \mathbb{P}\left(\overline{X} - \overline{m} > t\right) \leqslant \exp\left(-\frac{nt^2}{2\overline{\sigma^2}}\right).$$

Indeed, once this proved, changing all the X_i's to $-X_i$, gives: $\mathbb{P}\left(-\overline{X} + \overline{m} > t\right) \leqslant \exp\left(-\frac{nt^2}{2\overline{\sigma^2}}\right)$, and the final result is a consequence of: $(|U| > t) = (U > t) \cup (U < -t)$ where $U = \overline{X} - \overline{m}$.

Thanks to the proposition, $\sum_{k=1}^n X_k$ is a sub-Gaussian variable with parameter $\left(\sum_{k=1}^n \sigma_k^2\right)^{1/2}$. Thanks to the Chernoff method, we have:

$$\mathbb{P}\left(\overline{X} - \overline{m} > t\right) = \mathbb{P}\left(\sum_{k=1}^n (X_k - \mathbb{E}[X_k]) > nt\right) \leqslant \exp\left(-\frac{(nt)^2}{2\sum_{k=1}^n \sigma_k^2}\right).$$

Examples.

- If the X_k's are independent variables supported in $[a, b]$, we get:

$$\mathbb{P}\left(\left|\overline{X} - \overline{m}\right| > t\right) \leqslant 2\exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$

  In particular, this applies to Bernoulli variables with $a = 0, b = 1, \overline{\sigma^2} = 1$.

- If the X_k's are independent $\mathcal{N}(m_k, \sigma_k^2)$ Gaussian variables, we get:

$$\mathbb{P}\left(\left|\overline{X} - \overline{m}\right| > t\right) \leqslant 2\exp\left(-\frac{nt^2}{2\overline{\sigma^2}}\right).$$

  One could apply directly the Chernoff method to the variable $\overline{X}$ which is distributed as $\mathcal{N}\left(\overline{m}, \frac{\overline{\sigma^2}}{n}\right)$ and obtain the same bound: in this case (in which all inequalities are equalities) nothing has been lost by applying Hoeffding's inequality.

2 Application to confidence interval determination

Let us assume X_k, $k = 1, \dots, n$ are integrable independent identically distributed variables. Using Monte-Carlo simulations, we are looking at determining the mean value $m = E[X_1]$, with an error less or equal to $\varepsilon > 0$ and a confidence greater or equal to $\alpha \in (0,1)$. In other words, denoting: $\overline{X_n} = \frac{1}{n}\sum_{k=1}^n X_k$, we are interested in determining n such that:

$$\mathbb{P}\left(\left|\overline{X_n} - m\right| \leqslant \varepsilon\right) \geqslant \alpha.$$

2.1 Exact confidence intervals: Chebyshev and Hoeffding

- If the random variables are in L^2, and $\sigma^2 = \mathrm{Var}(Z_1)$, using Chebyshev's inequality, we have:

  $$\mathbb{P}\left(\left|\overline{X_n} - m\right| > \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2}.$$

  To make the rhs less or equal to $1-\alpha$, it is enough to choose n greater or equal to:

  $$\frac{1}{1-\alpha}\sigma^2\varepsilon^{-2}.$$

- If the random variables are sub-Gaussian (Gaussian, compactly supported) with parameter σ, using Hoeffding's inequality, we have:

  $$\mathbb{P}\left(\left|\overline{X_n} - m\right| > \varepsilon\right) \leqslant 2\exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right).$$

  To make the rhs less or equal to $1-\alpha$, it is enough to choose n greater or equal to:

  $$2\ln\left(\frac{2}{1-\alpha}\right)\sigma^2\varepsilon^{-2}.$$

2.2 Asymptotic confidence intervals: central limit theorem

If the random variables are in L^2, and n is *a priori* big enough (...), the central limit theorem yields:

$$\mathbb{P}\left(\left|\overline{X_n} - m\right| \leqslant \varepsilon\right) = \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\left|\overline{X_n} - m\right| \leqslant \frac{\sqrt{n}\varepsilon}{\sigma}\right) \approx \mathbb{P}\left(|N| \leqslant \frac{\sqrt{n}\varepsilon}{\sigma}\right) = 2F\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right) - 1,$$

where N is an $\mathcal{N}(0,1)$ random variable, F its cumulative distribution function. To make the rhs greater or equal to α, it is enough to choose n greater or equal to:

$$F^{-1}\left(\frac{\alpha+1}{2}\right)\sigma^2\varepsilon^{-2}.$$

Remark. This asymptotic bound is rigorous if the correct speed $\sqrt{n}$ is used:

$$\mathbb{P}\left(\left|\overline{X_n} - m\right| \leqslant \frac{\varepsilon}{\sqrt{n}}\right) \to 2F\left(\frac{\varepsilon}{\sigma}\right) - 1.$$

With this scaling the previous exact bounds write:

$$\mathbb{P}\left(\left|\overline{X_n} - m\right| \leqslant \frac{\varepsilon}{\sqrt{n}}\right) \leqslant \min\left(\frac{\sigma^2}{\varepsilon^2},\, 2\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)\right).$$

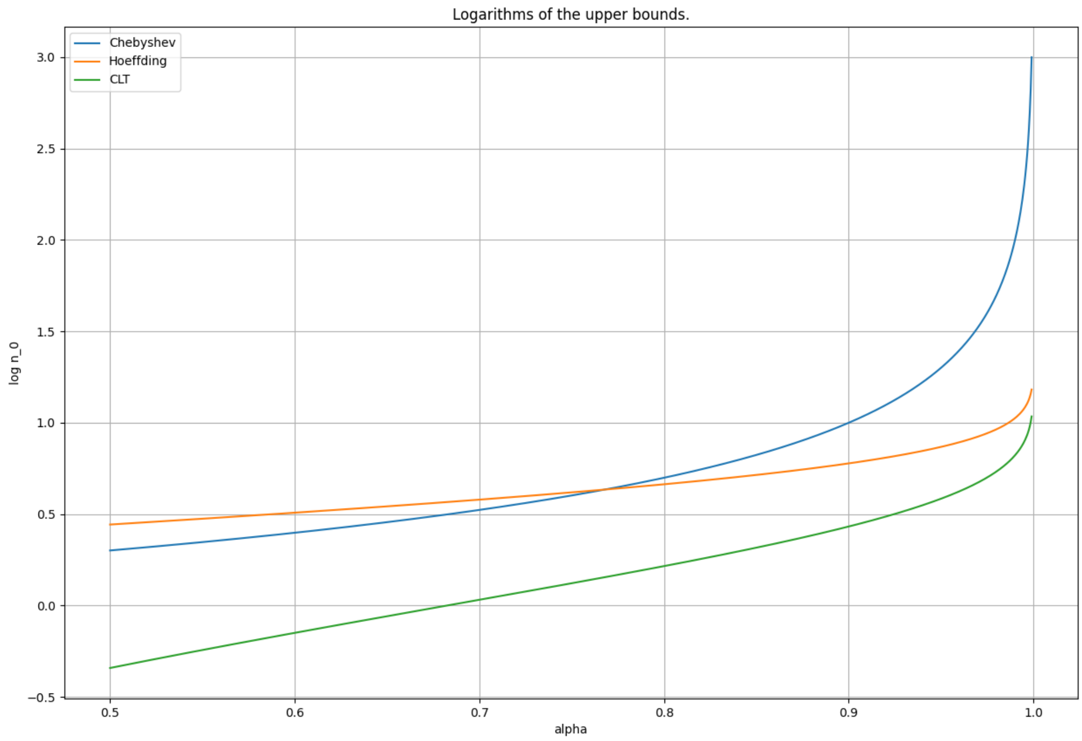

Figure 1: Decimal logarithms of the bounds. For a confidence level $\alpha \geqslant 0.7678$, the lowest bound is obtained thanks to the TCL, then comes Hoeffding's, then Chebyshev's.

2.3 Comparison of the bounds

The previous bounds can be described as being proportional to:

- the quantity σ^2, which is the variance for Chebyshev's and CLT derived bounds, and, a sub-Gaussian constant for the Hoeffding's derived bound (which can also be interpreted as a variance, *cf.* proof of Hoeffding's lemma),
- the inverse of the square of the error (ε^{-2}): to add one decimal, *i. e.* divide the error by a factor 10, 100 times more variables are needed,
- a universal function of α which is increasing and diverges as $\alpha \to 1^-$.

inequality	confidence term	σ term	error term
Chebyshev	$\frac{1}{1-\alpha}$	σ^2	ε^{-2}
Hoeffding	$2\ln\left(\frac{2}{1-\alpha}\right)$	σ^2	ε^{-2}
CLT	$F^{-1}\left(\frac{\alpha+1}{2}\right)$	σ^2	ε^{-2}

To rank these bounds it is enough to fix $\varepsilon = 1$ as they all have the same dependancy in ε. The parameter σ^2 may change from a distribution to another and from a bound to another (if it is the variance or a sub-Gaussian constant). For the sake of simplicity, let us fix it to 1. Figure 1 allows comparing the bounds as functions of α.

In practice the TCL bound is used; it is the best possible one as it entails the asymptotic behaviour of $\overline{X_n}$. The best exact bound is Hoeffding's (but it requires the sub-Gaussian condition to hold). To improve these bounds, nothing can be done on the precision term as it is prescribed. Other confidence terms can be obtained using other concentration inequalities (and requiring other hypotheses on the distributions). Improving the variance term is the goal of variance reduction techniques (finding a distribution which has the same mean, a smaller variance and the same simulation cost).