Concentration Inequalities

Mahsa Asadi
Inria Lille - Nord Europe, MAGNET team

Nov 21, 2020

Abstract

This report includes a summary of basics of probability and statistics along with rudimentary concentration inequalities. This is a report in progress and is going to be completed along the way of my PhD. I thank Dr. Odalric Maillard for his feedbacks and helping me learn this concepts. This is an in progress document. So, I would be grateful to know about any possible errors you may find.

Chapter 1

Concentration Inequalities

In this chapter, we are going to probe the world of concentration inequalities and explore some cases for which we can have confidence guarantees using high level understanding of concentration inequalities. More detailed and advanced investigation of the topic can be found at [1]. We are going to provide bound for:

- A Random Variable
- Sum of Independent Random Variables:

$$\sum_{i=1}^{n} X_i \tag{1.1}$$

- A Function of different independent random variables

$$f(X_1, X_2, \dots, X_n) \tag{1.2}$$

and also guarantee find out how probable it would be if we bound.

1.1 Probability Prerequisites

There exists a number of concepts in probability that are required before going into concentration inequalities:

σ-algebra: Let us denote the outcome space by Ω and define $\sigma-$algebra to be a set $\mathcal{F} \subset 2^{\Omega}$ which satisfies three properties of: including Ω, being closed under complement and countable union.

1.1.1 Boole's inequality(Union Bounds)

For a finite countable set of events $A_1, A_2, A_3, \ldots, A_n$, we have:

$$\mathbb{P}(\cup_i^n A_i) \leq \sum_i^n \mathbb{P}(A_i)$$

Proof. Using induction: For $n = 1$, $\mathbb{P}(A_1) = \mathbb{P}(A_1)$. Let us assume that the inequality holds for $n = k$, and prove for $n = k + 1$. Since $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, for $n = k + 1$:

$$\mathbb{P}(\cup_i^{k+1} A_i) = \mathbb{P}(\cup_i^k A_i) + \mathbb{P}(A_{k+1}) - \mathbb{P}(\cup_i^k A_i - A_{k+1})$$

The first part is bounded using induction base case, and the last part can be eliminated since we are searching for an upper bound. therefore,

$$\leq \sum_i^k \mathbb{P}(A_i) + \mathbb{P}(A_{k+1}) = \sum_i^{k+1} \mathbb{P}(A_i)$$

1.2 One Random Variable

We want to find out the probability δ that a random variable X is bounded by ϵ and based on the different assumption we make on X's corresponding distribution, we can find different guarantees [2]. More concretely,

$$\mathbf{P}(X \geq \epsilon) \leq \delta \tag{1.3}$$

1.2.1 Markov inequality

For any random variable $X \geq 0$,

$$\mathbf{P}(X \geq \epsilon) \leq \frac{\mathbf{E}[X]}{\epsilon} \tag{1.4}$$

Proof. Since $X \geq 0$

$$\mathbf{E}[X] = \int_0^\infty xp(x)d_x = \int_0^\epsilon xp(x)d_x + \int_\epsilon^\infty xp(x)d_x$$

$$\geq \int_\epsilon^\infty xp(x)d_x \geq \epsilon\mathbf{P}(X \geq \epsilon)$$

Therefore,

$$\mathbf{P}(X \geq \epsilon) \leq \frac{\mathbf{E}[X]}{\epsilon}$$

Extending Markov: Markov inequality can be employed for *non-negative non-decreasing* functions like ϕ defined on $\mathbb{R}$, with the random variable X as:

$$\mathbf{P}(X \geq \epsilon) \leq \mathbf{P}(\phi(X) \geq \phi(\epsilon)) \leq \frac{\mathbf{E}[\phi(X)]}{\phi(\epsilon)}$$

Proof. Function ϕ is said to be non-decreasing if for every $a < b$ on the domain, $\phi(a) < \phi(b)$. Therefore, if $X \geq \epsilon$ and ϕ is non-decreasing, then $\phi(X) \geq \phi(\epsilon)$. Now, if $(X \geq \epsilon) \subseteq (\phi(X) \geq \phi(\epsilon))$, then $\mathbf{P}(X \geq \epsilon) \leq \mathbf{P}(\phi(X) \geq \phi(\epsilon))$.

Adding the non-negativity assumption, we can apply Markov inequality and the result is obtained.

Let us see few applications of this extension: the Chernoff method and Chebyshev inequality.

1.2.2 Chernoff bound Technique

For any $t > 0$,

$$\mathbf{P}(X \geq \epsilon) = \mathbf{P}(e^{\lambda X} \geq e^{\lambda \epsilon}) \leq \frac{\mathbf{E}[e^{\lambda X}]}{e^{\lambda \epsilon}} \tag{1.5}$$

Proof. The idea of chernoff bound technique is to write down an equivalent inequality replacing the random variable with the moment generating function of the random variable as demonstrated above which is obtained by choosing $\phi(Y) = e^{tY}$.

1.2.3 Chernoff Method

Using chernoff bound and therefore, assumptions of markov inequality, we know:

$$\mathbb{P}\big(X \geq \epsilon\big) \leq \exp^{-\lambda\epsilon} \mathbf{E}[e^{\lambda X}] \tag{1.6}$$

Which holds true $\forall \lambda \geq 0$. We want to choose a λ that gives the tightest bound for the above equation. Taking the logrithm of rhs, we need to find:

$$\Psi_X^{\star}(t) = \inf_{\lambda \geq 0} -\lambda\epsilon + \log \mathbf{E}[e^{\lambda\epsilon}] = \sup_{\lambda \geq 0} \lambda\epsilon - \log \mathbf{E}[e^{\lambda\epsilon}]$$

Let us define $\Psi_X(\lambda) = \log \mathbb{E} \exp^{\lambda\epsilon}$ and using jensen inequality the convexity of exp function, we get $\Psi_X(\lambda) \geq \mathbb{E} \log \exp^{\lambda X} = \lambda \mathbb{E} X$. Therefore, we can get $\forall \lambda < 0 : \lambda\epsilon - \log \mathbb{E} \exp^{\lambda\epsilon} \leq 0$ for $t \geq \lambda \mathbb{E} X$ and $\Psi_X^{\star}(t)$ can be extended to:

$$\Psi_X^{\star}(t) = \sup_{\lambda \geq \mathbb{R}} \lambda\epsilon - \Psi_X(\lambda)$$

which is called the Cramer transform with the dual function of $\Psi_X(\lambda)$. This gives us the bound:

$$\mathbb{P}\big(X \geq \epsilon\big) \leq e^{-\Psi_X^{\star}(t)}$$

1.2.4 Chebyshev's Inequality

For any random variable X, we can get:

$$\mathbf{P}(|X - \mathbf{E}X| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2} \tag{1.7}$$

Proof. This is another extension of Markov's inequality obtained by choosing $\phi(t) = t^2$ and defining a random variable $Y = |X - \mathbf{E}X|$. Therefore,

$$\mathbf{P}(|X - \mathbf{E}X| \geq \epsilon) = \mathbf{P}(Y \geq \epsilon) \leq \mathbf{P}(Y^2 \geq \epsilon^2) \leq \frac{\mathbf{E}(|X - \mathbf{E}X|^2)}{\epsilon^2} = \frac{Var(X)}{\epsilon^2} \tag{1.8}$$

Bibliography

[1] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[2] Dimitri P Bertsekas and John N Tsitsiklis. Introduction to probability vol. 1. 2002.

[3] David Williams. *Probability with martingales*. Cambridge university press, 1991.