Master 2 Mathématiques
Parcours recherche

TD2 – Large deviations and concentration inequalities*

Exercise 1 : Exponential variables

We consider an i.i.d. sequence $(X_k)_{k\in\mathbb{N}}$ of exponential variables with parameter λ.

1) Justify that the X_k's have finite log-MGF on $(-\infty, \lambda)$, and compute their log-MGF $\Lambda(t)$.

2) Compute its Legendre transform $\Lambda^\star(x)$.

3) Justify that the distribution of $S_n := \frac{1}{n}\sum_{k=1}^n X_k$ satisfies a LDP and give its rate function and speed.

Answer :

1) The log-MGF for $Exp(\lambda)$ variables can be computed and is equal to

$$\Lambda_{Z_1}(t) = \begin{cases} \log\left(\frac{\lambda}{\lambda-t}\right) & \text{if } t < \lambda \\ +\infty & \text{otherwise} \end{cases}.$$

2) Its legendre transform is given by

$$\Lambda^\star_{Z_1}(x) = \lambda x - 1 - \log \lambda x.$$

3) This is a consequence Cramér's Theorem. The speed is n and the good rate function $\Lambda^\star_{Z_1}$. □

Exercise 2 : Cramér's theorem and Poisson tail

We want to prove that the tail of the Poisson distribution decays faster than exponentially. Show that given $X \sim Poi(t)$, we have for any $C > 0$

$$\limsup_{k\to\infty} \mathbb{P}(X > k)e^{Ck} = 0.$$

*For any typo/question, please contact me at *clement.erignoux@inria.fr*. The exercise sheets will be put on the webpage, http://chercheurs.lille.inria.fr/cerignou/homepage.html in the "teaching" section.

Answer : We write that $\mathbb{P}(X > k) = \mathbb{P}(Z_1 + \cdots + Z_k < 1)$, where $Z_i \sim Exp(\lambda)$ are independant exponential variables. We want an upper bound on the probability that those are small, so that we cannot use chernoff, and need to resort to a large deviations estimate. The large deviations functional for exponential variables with parameters λ (see previous exercise) is given by

$$\Lambda^\star_{Z_1}(x) = \lambda x - 1 - \log \lambda x,$$

which diverges as $x \to 0$. Choose k_0 large enough so that $\Lambda^\star_{Z_1}(1/k_0) > 2C$, and write

$$\mathbb{P}(X > k) \leq \mathbb{P}\left(\frac{Z_1 + \cdots + Z_k}{k} < 1/k\right) \leq \mathbb{P}\left(\frac{Z_1 + \cdots + Z_k}{k} < 1/k_0\right).$$

In particular,

$$\limsup \frac{1}{k} \log \mathbb{P}(X > k) \leq -\Lambda^\star_{Z_1}(1/k_0) < -2C,$$

which proves the result, since then

$$\limsup \frac{1}{k} \log\left(\mathbb{P}(X > k)e^{Ck}\right) = C + \limsup \frac{1}{k} \log \mathbb{P}(X > k) \leq -C.$$

□

Exercise 3 : Random walks

1) We consider a symmetric *discrete time* random walk (S_k). Prove that for any $n \in \mathbb{N}$, and any vanishing sequence $\varepsilon_k \to 0$,

$$\mathbb{P}\left(S_{nk^2} \geq k/\varepsilon_k\right) \underset{k\to\infty}{\longrightarrow} 0.$$

2) (i) We now consider a symmetric *continuous time* rate 1 random walk $(X_t)_{t\geq 0}$. Prove that for any $t > 0$, and any vanishing sequence $\varepsilon_k \to 0$,

$$\mathbb{P}\left(X_{tk^2} \geq k/\varepsilon_k\right) \underset{k\to\infty}{\longrightarrow} 0.$$

(ii) We want a stronger estimate. Prove that for any $t > 0$, and any $n > 0$,

$$k^n \mathbb{P}\left(X_{tk^2} \geq k \log(k)^2\right) \underset{k\to\infty}{\longrightarrow} 0.$$

hint : consider the variables $Y_q = min(A_k, X_{t(q+1)} - X_{tq})$, and show using the previous exercise

$$\mathbb{P}\left(X_{tk^2} \neq \sum_{q=0}^{k^2-1} Y_q\right) \to 0.$$

Answer :

1) This is a direct consequence of the CLT: for any n, any $\varepsilon > 0$, and any k large enough such that $\varepsilon_k < \varepsilon$

$$\mathbb{P}\left(S_{nk^2} \geq k/\varepsilon_k\right) \leq \mathbb{P}\left(S_{nk^2} \geq k/\varepsilon\right),$$

which converges as $k \to \infty$ to $\mathbb{P}(\mathcal{N}(0, \sigma_n^2) > 1/\varepsilon)$, where $\sigma_n^2 = n$ is the variance of S_n. This is true for any ε, we have the result by letting $\varepsilon \to 0$.

2) (i) The same estimate is true, with σ_n^2 replaced by σ_t^2 the variance of S_t.

(ii) By Markov property, X_{tk^2} is the sum of k^2 i.i.d. variables distributed as X_t. We want to apply Hoeffding's inequality, but in continuous time, X_t is not bounded. However, Assume that X_t was bounded by some constant A_k, we would have by Hoeffding's inequality

$$\mathbb{P}\left(X_{tk^2} \geq k \log(k)^2\right) \leq \exp\left(\frac{-2k^2 \log(k)^4}{k^2 A_k^2}\right).$$

We then define Y_q as proposed, and apply Hoeffding'inequality to the Y_q's, so that

$$\mathbb{P}\left(X_{tk^2} \geq k \log(k)^2\right) \leq \mathbb{P}(\exists q \leq k^2 - 1, Y_q \neq X_{t(q+1)} - X_{tq}) + \exp\left(\frac{-2 \log(k)^4}{A_k^2}\right).$$

By union bound, the first term is less than $k^2 \mathbb{P}(Poi(t) > A_k) \ll k^2 e^{-CA_k}$ for any positive constant C. In particular, for any C,

$$\mathbb{P}\left(X_{tk^2} \geq k \log(k)^2\right) = o\left(k^2 e^{-CA_k} + e^{-2\log(k^4)/A_k^2}\right).$$

Multiplying by k^n, one can choose $A_k = \log k$ and C large enough for both terms to vanish, which proves the result. □

Exercise 4

Fix $p \in (0, 1)$, and consider an i.i.d. sequence $(X_k)_{k\in\mathbb{N}}$ of *Bernoulli*(p) variables. Using Hoeffding's inequality, build for any $\alpha \in (0, 1)$ give an α-confidence interval, i.e. an interval $C_{p,n,\alpha}$ such that

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k \in C_{p,n,\alpha}\right) \geq 1 - \alpha.$$

What does the interval become using the first Bernstein inequality ?

Answer : We write

$$\mathbb{P}\left(-\varepsilon \leq \sum_{k=1}^{n} X_k - np \leq \varepsilon\right) = 1 - \mathbb{P}\left(\sum_{k=1}^{n}[X_k - p] \geq \varepsilon\right) - \mathbb{P}\left(\sum_{k=1}^{n}[p - X_k] \geq \varepsilon\right).$$

By Hoeffding's inequality,

$$\mathbb{P}\left(\sum_{k=1}^{n}[X_k - p] \geq n\varepsilon\right) \leq \exp\left(-\frac{2n^2\varepsilon^2}{n}\right) = e^{-2n\varepsilon^2},$$

and the same is true for the second bound, so that if $e^{-2n\varepsilon^2} = \alpha/2$, we must have

$$\varepsilon = \sqrt{-\frac{\log(\alpha/2)}{2n}},$$

which does not depend on p.

Using now the first Bernstein inequality, we obtain instead

$$\mathbb{P}\left(\sum_{k=1}^{n}[X_k - p] \geq n\varepsilon'\right) \leq \exp\left(-\frac{2n^2\varepsilon'^2}{np(1-p)+n/3}\right) = e^{-2n\varepsilon'^2},$$

which is a little bit better, and yields

$$\varepsilon' = \sqrt{-\frac{\log(\alpha/2)(p(1-p)+1/3)}{2n}} < \varepsilon',$$

□

Exercise 5

Fix $p \in (0,1)$, and consider an i.i.d. sequence $(X_k)_{k\in\mathbb{N}}$ of $Geom(p)$ variables. Define $Y_n = \sum_{k=1}^n kX_k$.

1) What is $\mathbb{E}(Y_n)$?

2) Show that for any $m > 3/2$,

$$\limsup_{n\to\infty} \mathbb{P}\left(Y_n \geq \frac{n(n+1)}{2p} + n^m\right) = 0.$$

Hint: as in the last question of Ex. 3, crop the X_k's and estimate the probability that cropping made a difference in the sum.

Exercise 6

Let μ be a distribution with support contained in $[a,b]$.

1) Justify that for an i.i.d. sequence $(X_n)_{n\in\mathbb{N}}$ distributed as μ, its empirical average $S_n = n^{-1}\sum_{k=1}^n X_k$ satisfies a large deviations principle.

2) Show that its rate function I satisfies $I = +\infty$ outside of $[a,b]$.

3) Show that if $\forall\varepsilon > 0$

$$\mu([a, a+\varepsilon)) > 0 \qquad \text{and} \qquad \mu((b-\varepsilon, b]) > 0,$$

then $I < +\infty$ on (a,b).

4) Show that $I(a) < +\infty \Leftrightarrow \mu(\{a\}) > 0$, and similarly with b.

Answer :

1) Since μ has bounded support, in particular $0 \leq e^{tX} \leq e^{t\max(a,b)}$ so that e^{tX} is integrable for any $t \in \mathbb{R}$ and its log-MGF is finite everywhere. By Cramér's theorem, the law of S_n/n satisfies a $LDP(n, \Lambda^\star)$ where $\Lambda^\star$ is the Legendre transform of the log-MGF of μ.

2) Fix x outside of $[a,b]$, $x > b$ for example, so that in particular $x > \mathbb{E}(X)$. By Cramér's theorem, we must have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n/n \geq x) = -\Lambda^\star(x).$$

But since x is outside the support of μ, $\mathbb{P}(S_n/n \geq x) = 0$, so that we must have $\Lambda^\star(x) = +\infty$. A similar arguments holds for $x < a$.

3) Assume that for any $\varepsilon > 0$, we have $\mu([a, a+\varepsilon)) > 0$. Fix $\varepsilon > 0$, we will show $\Lambda^\star(a+\varepsilon) < \infty$. Assume first that $\mathbb{E}(X) = a$, then $X = a$ a.s. because this is the only way for $\mathbb{E}(X)$ to be equal to a. In this case, $\Lambda^\star(a) = 0$, and in particular either $b = a$ or the condition is not satisfied. If $b = a$, the statement is trivial, so that we can now assume $\mathbb{E}(X) > a$. We then choose $\varepsilon < \mathbb{E}(X) - a$, by assumption

$$\mu([a, a+\varepsilon]) \geq \mu([a, a+\varepsilon)) \geq \delta > 0.$$

Note that $\mathbb{P}(S_n/n < a+\varepsilon) \geq \mathbb{P}(X_k \leq a+\varepsilon \ \forall k \leq n) = \delta^n$, therefore by Cramér's Theorem

$$\log \delta \geq \lim_{n\to\infty} \frac{1}{n} \mathbb{P}(S_n/n \leq a+\varepsilon) = -\Lambda^\star(a+\varepsilon),$$

thus $\Lambda^\star(a+\varepsilon) < \infty$ for any small ε.

4) Once again, this follows from Cramér's Theorem : Since $\mathbb{P}(S_n/n = a) = \mu(\{a\})^n$, we obtain by Cramér's Theorem

$$\Lambda^\star(a) = -\log(\mu(\{a\})),$$

which proves the equivalence.

□