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.

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 $C > 0$

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

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 $\varepsilon_k \to 0$,

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

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 $\varepsilon_k \to 0$,

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

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

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

*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.

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

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

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 ?

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.