TD 5: From Diffusion to Large Deviations - Solutions

Baptiste Coquinot *Contact:* baptiste.coquinot@ens.fr *Webpage:* https://coquinot.fr

14 octobre 2021

1 Diffusion Equation. In this exercise we solve explicitly the diffusion equation

$$\partial_t u = D\partial_x^2 u \tag{1}$$

over $\mathbb{R}^+ \times \mathbb{R}$ with initial condition $u(t=0,x) = u_0(x)$.

1. Show that we have the following similarity principle: if $u(t,x)$ is solution for $t > 0$ then $v(t,x) = u(\lambda t, \sqrt{\lambda}x)$ is solution too.
2. Heuristically, why can we expect to be able to restrict the partial differential equation in an ordinary differential equation with variable $s = \frac{x}{\sqrt{Dt}}$?
3. We define $u(t,x) = v\left(\frac{x}{\sqrt{Dt}}\right)$. Find an equation for v.
4. Solve the equation in v by imposing $u_0(x=0)$.
5. The previous solution is only partial because we have imposed the invariance by the dilatation λ but u_0 may be not non constant. To get the complete solution, the idea is then to perform a convolution between the kernel
$$K(t,x) = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \tag{2}$$
and the initial condition u_0. Check that K is solution of the diffusion equation and give the general solution.
6. (Bonus) Solve the diffusion equation using a Fourier transform.

Correction

1. We calculate $\partial_t v(t,x) = \lambda \partial_t u(\lambda t, \sqrt{\lambda}x) = D\lambda \partial_x^2 u(\lambda t, \sqrt{\lambda}x)$ and $\partial_x^2 v(t,x) = \lambda \partial_x^2 u(\lambda t, \sqrt{\lambda}x)$. Thus, $\partial_t v = D\partial_x^2 v$ indeed.
2. If u is solution, by taking $\lambda = \frac{x}{\sqrt{Dt^3}}$, which is not constant, we find that $u(s, Ds^{3/2})$ which has only has one parameter s. This gives us the idea to look for a solution u of one variable.
3. We calculate
$$\partial_t u(t,x) = -\frac{x}{2\sqrt{Dt^3}} v'\left(\frac{x}{\sqrt{Dt}}\right) \tag{3}$$
and
$$\partial_x^2 u(t,x) = \frac{1}{Dt} v''\left(\frac{x}{\sqrt{Dt}}\right). \tag{4}$$
Thus,
$$v''(s) = -\frac{s}{2} v'(s). \tag{5}$$
4. By usual techniques we find: $v'(s) = v'(0)e^{-s^2/4}$ and then
$$v(s) = v'(0)\int_{-\infty}^{s} e^{-s^2/4} \mathrm{d}s = \sqrt{4\pi} v'(0) \int_{-\infty}^{x/\sqrt{Dt}} \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \mathrm{d}x. \tag{6}$$
When $t \to 0$, $\frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \to \delta_0$. Imposing the initial condition, we then get
$$u(t,x) = u_0(0) \int_{-\infty}^{x/\sqrt{Dt}} \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} \mathrm{d}x. \tag{7}$$
5. We easily check that $\partial_t K = D\partial_x^2 K$. Thus, the complete solution writes
$$u(t,x) = \int_{\mathbb{R}} K(t, x-y) u_0(y) \tag{8}$$
because K is solution and converges to the Dirac when $t \to 0$.

2 Central Limit Theorem. We consider a series (X_n) of random variables which are independent and of identical distribution. We assume X to have moments of order 2 (*i.e.* the variance is well defined). We denote μ the mean and σ is mean-squared value of the distribution. We finally denote $S_n = \sum_{i=1}^n X_i$.

1. Prove Markov identity: let A be a positive random variable and $\epsilon > 0$, then
$$\mathbb{P}(A \geq \epsilon) \leq \frac{\mathbb{E}(A)}{\epsilon}. \tag{9}$$

2. Deduce the law of Large Numbers:
$$\frac{1}{n} S_n \longrightarrow \mu \tag{10}$$
where the limit is defined by: $\forall \epsilon > 0$, for large enough n, $\mathbb{P}(|\frac{1}{n}S_n - \mu| \geq \epsilon) \leq \epsilon$. A stronger version (the almost sure convergence) of this theorem can also be proved.

3. We denote $\varphi_X(t) = \mathbb{E}\left(e^{itX}\right)$ the characteristic function of X. This is the Fourier transform of its distribution. Express $\varphi_{\sqrt{n}(\frac{1}{n}S_n - \mu)}$ as a function of φ_Y where $Y = X - \mu$.

4. Give an estimate at second order of φ_Y when $t \to 0$.

5. Deduce the limit of $\varphi_{\sqrt{n}(\frac{1}{n}S_n - \mu)}$ when $n \to +\infty$ for infinitesimal t.

6. We admit Levy's theorem which states if φ_{X_n} converges to the characteristic function of some distribution X_∞, then $X_n \to X_\infty$. Deduce an expansion of $\frac{1}{n}S_n$ at order $o\left(\frac{1}{\sqrt{n}}\right)$. This is the Central Limit Theorem (CLT).

Correction

1. We always have $A \geq \epsilon \mathbb{1}_{A \geq \epsilon}$. Taking the average: $\mathbb{E}(A) \geq \epsilon \mathbb{P}(A \geq \epsilon)$.

2. Let $\epsilon > 0$,
$$\mathbb{P}\left(|\frac{1}{n}S_n - \mu| \geq \epsilon\right) = \mathbb{P}\left(\left(\frac{1}{n}S_n - \mathbb{E}\left(\frac{1}{n}S_n\right)\right)^2 \geq \epsilon^2\right) \leq \frac{V\left(\frac{1}{n}S_n\right)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} < \epsilon \tag{11}$$
for n large enough.

3. Using the independence of the (X_n), we have:
$$\varphi_{\sqrt{n}((\frac{1}{n}S_n - \mu)}(t) = \mathbb{E}\left(e^{it(S_n - n\mu)/\sqrt{n}}\right) = \mathbb{E}\left(e^{it(X-\mu)/\sqrt{n}}\right)^n = \varphi_Y(t/\sqrt{n})^n. \tag{12}$$

4. When $t = 0$, $\varphi_Y(0) = \mathbb{E}(1) = 1$. Then, $\partial_t \varphi_Y(0) = \mathbb{E}(iY) = 0$ and $\partial_t^2 \varphi_Y(0) = \mathbb{E}\left(-Y^2\right) = -\sigma^2$. Thus,
$$\varphi_Y(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2). \tag{13}$$

5. Taking the limit,
$$\varphi_{\sqrt{n}((\frac{1}{n}S_n - \mu)}(t) = \varphi_Y(t/\sqrt{n})^n = (1 - \frac{1}{2n}\sigma^2 t^2 + o(t^2))^n \to \exp\left(-\frac{1}{2}\sigma^2 t^2 + o(t^2)\right). \tag{14}$$

6. We recognize the Fourier transform of a Gaussian: $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x$. We then deduce that $\sqrt{n}((\frac{1}{n}S_n - \mu) \to \mathcal{N}(0, \sigma^2)$. In other words, we have the expansion
$$\frac{1}{n}S_n = \mu + \frac{1}{\sqrt{n}}\mathcal{N}(0, \sigma^2) + o\left(\frac{1}{\sqrt{n}}\right) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) + o\left(\frac{1}{\sqrt{n}}\right). \tag{15}$$

2 Beyond the Central Limit Theorem: the Large Deviations. In practice some events do happen and can be physically crucial. The goal of Large Deviation theory is to estimate the probability of such rare events. We use the same notations than is the previous exercise and denote p the distribution of X. We also assume X not to be almost sure (*i.e.* its distribution is not a Dirac).

1. Let $a \geq \mu$. Prove that for any $\lambda \geq 0$, we have
$$\mathbb{P}\left(\frac{1}{n}S_n \geq a\right) \leq e^{n(m(\lambda) - \lambda a)} \tag{16}$$
where $m(\lambda) = \ln\left(\mathbb{E}\left(e^{\lambda X}\right)\right) = \ln\left(\int e^{\lambda x} \mathrm{d}p(x)\right)$.

2. We now want to optimize our estimate. We denote $f(\lambda) = m(\lambda) - \lambda a$. Study the form of this function and deduce that it admits a unique minimum for some $\lambda \geq 0$.

3. We denote $\bar{\lambda}$ this minimum and $\Omega(a) = f(\bar{\lambda})$ the grand deviations function. Similarly, we define Ω for $a \leq \mu$. When is Ω maximum? What is its maximum value?

4. Calculate $m'(\bar{\lambda})$ and $\Omega'(a)$. Do these relations remind you something?

5. Study the function Ω.

Correction

1. We calculate

$$\mathbb{P}\left(\frac{1}{n}S_n \geq a\right) = \int_{\frac{1}{n}S_n \geq a} \prod_{i=1}^{n} \mathrm{d}p(x_i) \leq \int_{\frac{1}{n}S_n \geq a} \prod_{i=1}^{n} \mathrm{d}p(x_i) e^{n\lambda(\frac{1}{n}S_n - a)} \tag{17}$$

$$\leq \int \prod_{i=1}^{n} \mathrm{d}p(x_i) e^{n\lambda(\frac{1}{n}S_n - a)} = e^{n(m(\lambda) - \lambda a)}. \tag{18}$$

2. We calculate $f'(\lambda) = \frac{\mathbb{E}(Xe^{\lambda X})}{\mathbb{E}(e^{\lambda X})} - a$. In particular, $f'(0) = \mu - a \leq 0$. Moreover,

$$f''(\lambda) = \frac{\mathbb{E}\left(X^2 e^{\lambda X}\right)}{\mathbb{E}\left(e^{\lambda X}\right)} - \frac{\mathbb{E}\left(Xe^{\lambda X}\right)^2}{\mathbb{E}\left(e^{\lambda X}\right)^2} = \mathbb{E}_\lambda(X^2) - \mathbb{E}_\lambda(X)^2 = V_\lambda(X) \geq 0 \tag{19}$$

where we have identified the expectancy given a modified distribution. We deduce f is convex and even strictly convex because X is not almost sure. Then, f admits a unique minimum which is positive because $f'(0) \leq 0$.

3. By definition, for $a \geq \mu$, $\mathbb{P}\left(\frac{1}{n}S_n \geq a\right) \leq e^{n\Omega(a)}$ and when $a \leq \mu$, $\mathbb{P}\left(\frac{1}{n}S_n \leq a\right) \leq e^{n\Omega(a)}$. These inequalities are almost equalities thanks to our optimization, we then expect Ω to be maximum for $a = \mu$. For sure, $\Omega \leq 0$. Moreover, going back to question 2 we see that for $a = \mu$, $\bar{\lambda} = 0$ then $\Omega(\mu) = 0$. Thus, $a = \mu$ is indeed the maximum.

4. Using that $\bar{\lambda}$ is the minimum of f we deduce $f'(\bar{\lambda}) = m'(\bar{\lambda}) - a$, *i.e.* $m'(\bar{\lambda}) = a$. Thus, $\Omega'(a) = m'(\bar{\lambda})\frac{\mathrm{d}\bar{\lambda}}{\mathrm{d}a} - a\frac{\mathrm{d}\bar{\lambda}}{\mathrm{d}a} - \bar{\lambda} = -\bar{\lambda}$. These are the same relations than between the Lagrangian and the Hamiltonian. $f = L$, $\Omega = H$, $a = v$ and $\bar{\lambda} = p$. This is natural because both cases are Legendre transforms.

5. We have $\Omega''(a) = -\frac{\mathrm{d}\bar{\lambda}}{\mathrm{d}a} \leq 0$, which is clear on a graph. We deduce that Ω is concave with its maximum at μ.