GAUSSIAN PROCESSES
EXERCISE SHEET 5: ENTROPY AND CLT

Exercise 1.

Let X be a random variable with density $p_X(x)$, and let the Fisher information be defined as

$$J(X) = \mathbb{E}\left[\left(\frac{\partial}{\partial x}\log p_X(X)\right)^2\right] = \int \frac{[p_X'(x)]^2}{p_X(x)}\,dx.$$

1. Translation invariance. Let $Y = X + a$ for some constant $a \in \mathbb{R}$. Then the density of Y is

$$p_Y(y) = p_X(y-a),$$

and hence

$$p_Y'(y) = p_X'(y-a).$$

Therefore,

$$J(Y) = \int \frac{[p_Y'(y)]^2}{p_Y(y)}\,dy = \int \frac{[p_X'(y-a)]^2}{p_X(y-a)}\,dy.$$

With the change of variable $x = y - a$, we obtain

$$J(Y) = \int \frac{[p_X'(x)]^2}{p_X(x)}\,dx = J(X).$$

Thus, $J(X+a) = J(X)$.

2. Scaling property. Let $Y = aX$, where $a \neq 0$ is constant. Then

$$p_Y(y) = \frac{1}{|a|}p_X\left(\frac{y}{a}\right), \quad p_Y'(y) = \frac{1}{|a|}\cdot\frac{1}{a}p_X'\left(\frac{y}{a}\right) = \frac{\operatorname{sign}(a)}{a^2}p_X'\left(\frac{y}{a}\right).$$

Hence,

$$J(Y) = \int \frac{[p_Y'(y)]^2}{p_Y(y)}\,dy = \int \frac{(1/a^4)[p_X'(y/a)]^2}{(1/|a|)p_X(y/a)}\,dy = \frac{1}{a^2}\int \frac{[p_X'(x)]^2}{p_X(x)}\,dx,$$

where we used the substitution $x = y/a$. Therefore,

$$J(aX) = a^{-2}J(X).$$

□

Exercise 2.

(1) Let $Z_t = X + \sqrt{t}\,Y$. Its density is

$$p_t(x) = \int_{-\infty}^{\infty} p_X(u)\,p_{\sqrt{t}Y}(x-u)\,du = \int_{-\infty}^{\infty} p_X(u)\,\frac{e^{-\frac{(x-u)^2}{2t}}}{\sqrt{2\pi t}}\,du.$$

Differentiating under the integral sign gives

$$\begin{aligned} p_t'(x) &= \int p_X(u)\, \frac{-(x-u)}{t\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du, \\ p_t''(x) &= \int p_X(u)\, \frac{(x-u)^2 - t}{t^2\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du, \\ \partial_t p_t(x) &= \int p_X(u)\, \frac{(x-u)^2 - t}{2t^2\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du. \end{aligned}$$

Hence $\partial_t p_t = \frac{1}{2} p_t''$. Now

$$\frac{d}{dt} H(Z_t) = -\frac{d}{dt}\int p_t \log p_t = -\int \partial_t p_t\, \log p_t - \int \partial_t p_t.$$

The second integral vanishes since $\int p_t = 1$. Using $\partial_t p_t = \frac{1}{2} p_t''$,

$$\frac{d}{dt} H(Z_t) = -\frac{1}{2}\int p_t'' \log p_t = -\frac{1}{2}\Big[p_t' \log p_t\Big]_{-\infty}^{\infty} + \frac{1}{2}\int \frac{p_t'^2}{p_t}\, dx = -\frac{1}{2}\Big[p_t' \log p_t\Big]_{-\infty}^{\infty} + J(Z_t)$$

where integration by parts is used in the second equality. By the Cauchy–Schwarz inequality, we have

$$\begin{aligned} p_t'(x)^2 &= \left(\int p_X(u)\, \frac{-(x-u)}{t\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du\right)^2 \\ &\le \left(\int p_X(u)\, \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du\right)\left(\int p_X(u)\, \frac{(x-u)^2}{t^2\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du\right). \end{aligned} \tag{0.1}$$

Note that

$$p_t(x) = \int p_X(u)\, \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du.$$

Since $(x-u)^2 e^{-(x-u)^2/(2t)} \le \frac{2t}{e}$, it follows that

$$p_t'(x)^2 \lesssim p_t(x),$$

so that

$$|p_t'(x)| \lesssim \sqrt{p_t(x)}.$$

This implies that the boundary term in integration by parts vanishes:

$$\big[p_t' \log p_t\big]_{-\infty}^{\infty} = 0.$$

Moreover, dividing the inequality (0.1) by $p_t(x)$, we get

$$\frac{p_t'(x)^2}{p_t(x)} \le \int p_X(u)\, \frac{(x-u)^2}{t^2\sqrt{2\pi t}}\, e^{-\frac{(x-u)^2}{2t}}\, du,$$

which shows that the Fisher information

$$J(Z_t) = \int \frac{p_t'(x)^2}{p_t(x)}\, dx$$

is finite.

(2) Let $Z_t = e^{-t}X + \sqrt{1-e^{-2t}}\,Y$. Then

$$p_t(x) = \int_{-\infty}^{\infty} p_X(u)\, \frac{1}{\sqrt{2\pi(1-e^{-2t})}} \exp\Big(-\frac{(x-e^{-t}u)^2}{2(1-e^{-2t})}\Big)\, du.$$

We check that p_t satisfies the PDE $\partial_t p_t = p_t'' + (x p_t)'$.

Let $s(t) = 1 - e^{-2t}$ and write

$$\phi(x,t;u) = \frac{1}{\sqrt{2\pi s(t)}} \exp\Big(-\frac{(x-e^{-t}u)^2}{2s(t)}\Big).$$

Then $p_t(x) = \int p_X(u)\phi(x,t;u)\,du$. For fixed u, set $y = x - e^{-t}u$ so that $\partial_t y = e^{-t}u$ and $s'(t) = 2e^{-2t}$. We compute:

$$\partial_x \phi = -\frac{y}{s}\phi, \qquad \partial_x^2 \phi = \Big(\frac{y^2}{s^2} - \frac{1}{s}\Big)\phi,$$

$$\partial_t \phi = \phi\Big(-\frac{s'}{2s} + \frac{y^2 s'}{2s^2} - \frac{y\,\partial_t y}{s}\Big) = \phi\Big(-\frac{e^{-2t}}{s} + \frac{e^{-2t}y^2}{s^2} - \frac{e^{-t}u\,y}{s}\Big).$$

Using $e^{-t}u = x - y$, we have $-e^{-t}u\,y/s = -(xy - y^2)/s$, so

$$\partial_t \phi = \phi\Big(-\frac{e^{-2t}}{s} + \frac{e^{-2t}y^2}{s^2} - \frac{xy}{s} + \frac{y^2}{s}\Big) = \phi\Big(\frac{y^2}{s^2}(e^{-2t}+s) - \frac{e^{-2t}}{s} - \frac{xy}{s}\Big).$$

Since $e^{-2t} + s = 1$, this simplifies to

$$\partial_t \phi = \phi\Big(\frac{y^2}{s^2} - \frac{1-s}{s} - \frac{xy}{s}\Big).$$

On the other hand,

$$\partial_x^2 \phi + (x\phi)' = \phi\Big(\frac{y^2}{s^2} - \frac{1}{s} + 1 - \frac{xy}{s}\Big),$$

so indeed $\partial_t \phi = \partial_x^2 \phi + (x\phi)'$. Integrating over u gives

$$\boxed{\partial_t p_t(x) = p_t''(x) + (xp_t(x))'.}$$

We now differentiate the entropy:

$$\frac{d}{dt}H(Z_t) = -\int \partial_t p_t \,\log p_t = -\int \big(p_t'' + (xp_t)'\big)\log p_t.$$

Integrating by parts and the first part of this exercise gives

$$-\int p_t'' \log p_t = \int \frac{(p_t')^2}{p_t} = J(Z_t),$$

$$-\int (xp_t)'\log p_t = -[xp_t \log p_t]_{-\infty}^{\infty} + \int x p_t' = -[xp_t\log p_t]_{-\infty}^{\infty} + [xp_t]_{-\infty}^{\infty} - \int p_t = -1.$$

Here for the boundary term in integration by parts, we note that Z_t also has a finite variance, hence $[xp_t]_{-\infty}^{\infty} = [xp_t \log p_t]_{-\infty}^{\infty} = 0$ (one can see that $p_t = o(1/x^2)$). Thus

$$\frac{d}{dt}H(Z_t) = J(Z_t) - 1.$$

□

Exercise 3.

The characteristic function of $X+Y$ is $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$. If Y has a smooth density, then φ_Y decays faster than any polynomial and the same holds for φ_{X+Y}, which again implies that $X+Y$ has a smooth density by Fourier inversion.

Assume next that Y has a (not necessarily smooth) density and denote by μ_X the distribution of X. Then

$$f(x) := \int_{-\infty}^{\infty} p_Y(x-y)d\mu_X(y)$$

defines a non-negative density function. One easily checks by using Fubini's theorem that the Fourier transform of f coincides with the characteristic function of $X+Y$, and hence $X+Y$ has density f. Let us choose a compact interval $A \subset \mathbb{R}$ such that $\mu_X(A) > 0$. If $p_Y(x) > 0$ for all $x \in \mathbb{R}$, we then see that

$$f(x) \geq \int_A p_Y(x-y) d\mu_X(y) \geq \inf_{z \in x-A} p_Y(z) \mu_X(A) > 0.$$

In particular if Y is a Gaussian then the above hold ($\inf_{z \in x-A} p_Y(z)$ also exists when x is in a compact interval). Moreover by Fourier inversion the derivative of the density is uniformly bounded by $\frac{1}{2\pi} \int t \varphi_X(t) e^{-t^2/2}\, dt$. Let us write $X+Y = X+Y_1+Y_2$ with Y_1, Y_2 i.i.d. Gaussians with variance $\frac{1}{2}\mathbb{E}Y^2$. Then by the decrease of Fisher information, we have

$$\begin{aligned} \mathbb{E}\rho_{X+Y}(X+Y)^4 &= \mathbb{E}\rho_{X+Y_1+Y_2}(X+Y_1+Y_2)^4 \\ &= \mathbb{E}(\mathbb{E}[\rho_{Y_2}(Y_2)|X+Y_1+Y_2])^4 \\ &\leq \mathbb{E}\rho_{Y_2}(Y_2)^4 < \infty. \end{aligned}$$

This also implies that the Fisher information is finite. □

Exercise 4.

Differentiating $\mathbb{E}_\theta[T(X)] = \theta$ gives

$$1 = \frac{d}{d\theta}\mathbb{E}_\theta[T(X)] = \int T(x) \frac{\partial p(x,\theta)}{\partial \theta}\, dx = \mathbb{E}_\theta[T(X)\rho_\theta(X)].$$

Let $U(X) = T(X) - \theta$, then $\mathbb{E}_\theta[U(X)] = 0$. Note that $\mathbb{E}_\theta[\rho_\theta(X)] = \int \frac{\partial p(x,\theta)}{\partial \theta}\, dx = 0$, we have that $\mathbb{E}_\theta[T(X)\rho_\theta(X)] = \mathbb{E}_\theta[U(X)\rho_\theta(X)]$, thus by Cauchy-Schwarz we have

$$1 = \mathbb{E}_\theta[U(X)\rho_\theta(X)]^2 \leq \mathbb{E}_\theta[U(X)^2]\mathbb{E}_\theta[\rho_\theta(X)^2] = \mathrm{Var}_\theta(T(X))\, J(\theta).$$

And this gives the Cramér-Rao bound:

$$\mathrm{Var}_\theta(T(X)) = \mathbb{E}_\theta[(T(X)-\theta)^2] \geq \frac{1}{J(\theta)}.$$

For distributions with fixed variance, the Gaussian has minimal Fisher information, so the bound is largest and the Gaussian is hardest to estimate. □