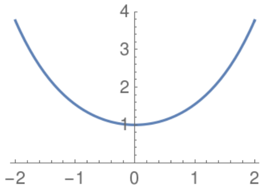
Sample Solutions Exercise 6

Exercise 6.1: Transforming probability densities (6P)



Consider the curve $y = \cosh x$. The aim of this exercise is to decorate this curve in the interval $x \in [x_1, x_2]$ with random points in such a way that the density of the points is uniform along the curve. Since the slope of the curve varies, this means that the x-coordinates of these points are *not* uniformly distributed.

(a) Compute the probability density $p(x)$ of the x-coordinates. Hint: the density of the points per arc length of the curve has to be constant. (2P)

(b) Normalize the probability density on the interval $x \in [x_1, x_2]$. (1P)

(c) Find a function $f : z \mapsto x = f(z)$ such that it maps a uniform probability density $p(z) = const$ to the non-uniform probability density $p(x)$ calculated in (a)-(b). (2P)

(d) Adjust the integration constant in (c) in such a way that f maps $[z_1, z_2] \mapsto [x_1, x_2]$ with $f(z_1) = x_1$ and $f(z_2) = x_2$. Specialize the result for a standard random number $(z_1 = 0, z_2 = 1)$. (1P)

Sample Solution

(a) The infinitesimal line element $\mathrm{d}\ell$ of the curve $y = f(x) = \cosh x$ is given by (1P)

$$\mathrm{d}\ell = \sqrt{1 + (f'(x))^2}\,\mathrm{d}x = \sqrt{1 + \sinh^2(x)}\,\mathrm{d}x = \cosh(x)\,\mathrm{d}x$$

We start with the ansatz

$$|p(\ell)\,\mathrm{d}\ell| = |p(x)\,\mathrm{d}x|$$

where $p(\ell) = C$ is constant along the curve. Using this ansatz we get (1P)

$$p(x) = C\left|\frac{\mathrm{d}\ell}{\mathrm{d}x}\right| = C\cosh x.$$

(b) The normalization can be computed straight-forwardly:

$$\int_{x_1}^{x_2} p(x)\,\mathrm{d}x = C(\sinh x_2 - \sinh x_1) \quad\Rightarrow\quad C = \frac{1}{\sinh x_2 - \sinh x_1}$$

giving

$$p(x) = \frac{\cosh x}{\sinh x_2 - \sinh x_1}.$$

(c) Again we start with

$$|p(z)\,\mathrm{d}z| = |p(x)\,\mathrm{d}x|,$$

where $z \in [z_1, z_2]$ is uniformly distributed, meaning that $p(z) = \frac{1}{z_2 - z_1}$ is constant. Again we get a first-order differential equation

$$\frac{\mathrm{d}z}{\mathrm{d}x} = z'(x) = \frac{p(x)}{p(z)} = \frac{z_2 - z_1}{\sinh x_2 - \sinh x_1} \cosh x \,,$$

hence

$$z(x) = \frac{z_2 - z_1}{\sinh x_2 - \sinh x_1} \sinh x + \tilde{C}, .$$

where $\tilde{C}$ is the integration constant. However, what we need is not $z(x)$ but the inverse function $x(z)$:

$$x(z) = \operatorname{arcsinh}\left[\frac{\sinh x_2 - \sinh x_1}{z_2 - z_1}(z - \tilde{C})\right]$$

(d) Assume that f maps $[z_1, z_2]$ onto $[x_1, x_2]$ with

$$x(z_1) = z_1 \,, \quad x(z_2) = z_2 \,.$$

It is easy to see that the difference $\sinh(x(z_2)) - \sinh(x(z_1) = \sinh(x_2 - \sinh(x_1)$ is always true, hence both equations are not independent. Taking one of them we can compute the integration constant

$$\tilde{C} = \frac{z_2 \sinh x_1 - z_1 \sinh x_2}{\sinh x_2 - \sinh x_1} \,.$$

Plugging that in we arrive at the final result

$$x(z) = \operatorname{arcsinh}\left[\frac{(z - z_1)\sinh x_2 - (z - z_2)\sinh x_1}{z_2 - z_1}\right].$$

For a standard random number generator $z \in [0, 1]$ with $z_1 = 0$ and $z_1 = 1$ this specializes to

$$x(z) = \operatorname{arcsinh}\Big[\sinh x_1 + z\big(\sinh x_2 - \sinh x_1\big)\Big] \,.$$

Exercise 6.2: Relative entropy (2P)

The relative entropy $H(p||q)$ of two probability distributions $p_1, \ldots, p_N$ and $q_1, \ldots, q_N$, which is also known as Kullback-Leibler-Divergence $D(p||q)$, measures how different the probability distributions are. It is defined by

$$H(p||q) = D(p||q) = \sum_i p_i \ln \frac{p_i}{q_i}.$$

(a) Use Jensen's inequality to show that $H(p||q) \geq 0$. (1P)

(b) Show that $H(p||p) = 0$ and $H(p||u) = \ln N - H(p)$, where u is the uniform (constant) probability distribution. (1P)

Sample Solution

(a) Since $-\ln x$ is a convex function, we can apply Jensen's inequality as discussed in the lecture:

$$H(p||q) = \sum_i p_i \ln \frac{p_i}{q_i} = -\sum_i p_i \ln \frac{q_i}{p_i} = \Big\langle -\ln \frac{q_i}{p_i} \Big\rangle$$

$$\geq -\ln \Big\langle \frac{q_i}{p_i} \Big\rangle = -\ln \Big(\sum_i p_i \frac{q_i}{p_i} \Big) = -\ln \Big(\underbrace{\sum_i q_i}_{=1} \Big) = 0.$$

(b)

$$H(p||p) = \sum_i p_i \underbrace{\ln \frac{p_i}{p_i}}_{ln 1 = 0} = 0.$$

$$u_i = 1/N \quad \Rightarrow \quad H(p||u) = \sum_i p_i \ln \frac{p_i}{1/N} = \underbrace{\sum_i p_i}_{=1} \ln N + \underbrace{\sum_i p_i \ln p_i}_{=-H(p)} = \ln N - H(p).$$

Exercise 6.3: Entropy of overlapping probability densities (4P)

Let $p(x)$ be a given continuously differentiable probability density. Let us shift it to the right and to the left by $x \to x \pm a$ and create a mixed probability density of the form

$$q(x) := \frac{1}{2}\Big(p(x-a) + p(x+a)\Big).$$

The purpose of this exercise is that the entropy is minimal for $a = 0$ where both of them match, that is, the entropy is minimized in the case of a perfect overlap.

(a) Show that $\int_{-\infty}^{+\infty} \Big(p'(x+a) - p'(x-a) \Big)\, \mathrm{d}x = 0 \quad \forall a.$ (1P)

(b) Prove that the entropy $H_a = -\int_{-\infty}^{+\infty} q(x) \ln q(x)\, \mathrm{d}x$ is extremal for $a = 0$. (1P)

(c) Show that the extremum at $a = 0$ is a minimum. You can assume that $p(x)$ and all its derivatives vanish at $x = \pm\infty$. (2P)

Note: Is this minimization of overlapping functions useful? Yes, it is, you can use it to tune your piano, see http://piano-tuner.org. Available for Windows, Mac, Android, and iOS.

Sample Solution

(a)

$$\int_{-\infty}^{+\infty} \Big(p'(x+a) - p'(x-a) \Big)\, \mathrm{d}x = \int_{-\infty}^{+\infty} \frac{\mathrm{d}}{\mathrm{d}a} \Big(p(x+a) + p(x-a) \Big)$$

$$= \frac{\mathrm{d}}{\mathrm{d}a} \Big[\int_{-\infty}^{+\infty} p(x+a)\, \mathrm{d}x + \int_{-\infty}^{+\infty} p(x-a)\, \mathrm{d}x \Big] = \frac{\mathrm{d}}{\mathrm{d}a}(1+1) = 0$$

(b) The entropy reads:

$$H_a = -\int_{-\infty}^{+\infty} \frac{1}{2}\Big(p(x-a)+p(x+a)\Big) \ln \frac{1}{2}\Big(p(x-a)+p(x+a)\Big).$$

As a necessary condition for an extremal point, $\frac{\mathrm{d}}{\mathrm{d}a}H_a = 0$ has to vanish. Carrying out the derivative we obtain

$$\frac{\mathrm{d}}{\mathrm{d}a}H_a = \frac{1}{2}\int_{-\infty}^{+\infty} \Big[1+\ln\Big(\frac{1}{2}\big(p(x-a)+p(x+a)\big)\Big)\Big]\cdot\Big(p'(x+a)-p'(x-a)\Big) = 0$$

Clearly, the round bracket vanishes for $a = 0$, hence we have an extremum here.

(c) A straight-forward calculation yields

$$\frac{\mathrm{d}^2}{\mathrm{d}a^2}H_a\Big|_{a=0} = -\int_{-\infty}^{+\infty} \big(1+\ln p(x)\big)p''(x)\ \mathrm{d}x$$

We integrate this expression by parts:

$$\frac{\mathrm{d}^2}{\mathrm{d}a^2}H_a\Big|_{a=0} = \underbrace{-\Big[\big(1+\ln p(x)\big)p'(x)\Big]_{-\infty}^{+\infty}}_{=0 \text{ since } p'(\pm\infty)=0} + \int_{-\infty}^{+\infty} \underbrace{\frac{1}{p(x)}}_{\geq 0}\underbrace{p'(x)p'(x)}_{\geq 0}\ \mathrm{d}x\ > 0$$

Hence we have a minimum.

(Σ = 12P)