S26 AerE 5740: Optimal Control

Practice Set 01: Calculus of Variations in Probability

Instructor: Abhishek Halder

Problem 1

Probability density function (PDF) of an n-dimensional continuous random vector X is a deterministic function $u(x)$, where $x \in \Omega \subseteq \mathbb{R}^n$, such that $u(x) \geq 0$ and $\int_\Omega u(x)\mathrm{d}x = 1$. The expected value of any nonlinear function $\phi(x)$, denoted here as $\mathbb{E}\left[\phi(x)\right]$, is the integral

$$\mathbb{E}\left[\phi(x)\right] := \int_\Omega \phi(x)u(x)\mathrm{d}x.$$

An example of the expectation integral of above type is the Shannon entropy

$$\mathbb{E}\left[-\log u(x)\right] = -\int_\Omega u(x)\log u(x)\mathrm{d}x,$$

i.e., $\phi(\cdot) \equiv -\log u(\cdot)$. In this exercise, we will assume that Ω is a compact set, and denote its (Lebesgue) volume as $\mathrm{vol}(\Omega) := \int_\Omega \mathrm{d}x$.

(a) Maximizing entropy versus minimizing expectation.

Given a function $f : \Omega \mapsto \mathbb{R}_{\geq 0}$ and a scalar $\beta \geq 0$, we want to determine the PDF $u(x)$ that minimizes

$$I(u) = \mathbb{E}\left[\log u + \beta\, f\right] = \int_\Omega \big\{u(x)\log u(x) + \beta\, f(x)u(x)\big\}\mathrm{d}x.$$

By changing the value of β, we can trade-off the relative importance of maximizing entropy and minimizing the expected value of f. In optimization literature, β is called a regularizing parameter.

(a.1) Using Euler-Lagrange equation, determine the PDF u^{opt} that minimizes $I(u)$.

(a.2) By specializing your answer in (a.1) for $\beta = 0$, determine u_0^{opt}. Explain in one sentence why this result matches your intuition.

Solution for part (a):

(a.1) Let λ be the Lagrange multiplier associated with the normalization constraint $\int_\Omega u = 1$, which is an integral equality constraint. Then the augmented Lagrangian

$$L(x, u, \nabla u) + \lambda M(x, u, \nabla u) = u \log u + \beta f(x) u + \lambda\, u$$

results in the EL equation:

$$\frac{\partial}{\partial u}(L + \lambda M) - \nabla \cdot \frac{\partial}{\partial \nabla u}(L + \lambda M) = 1 + \log u + \beta f(x) + \lambda = 0,$$

which gives optimizer $u^{\text{opt}} = \exp\left(-\left(1+\lambda\right) - \beta f(x)\right)$. To determine λ, we employ the feasibility condition $\int_\Omega u = 1$, to get

$$\exp\left(-\left(1+\lambda\right)\right) = \frac{1}{\int_\Omega \exp\left(-\beta f(x)\right) \mathrm{d}x}.$$

Substituting the above back in the RHS of u^{opt}, we obtain

$$u^{\text{opt}} = \frac{\exp\left(-\beta f(x)\right)}{\int_\Omega \exp\left(-\beta f(x)\right) \mathrm{d}x},$$

which is the Gibbs PDF. Notice that we did not need to explicitly enforce the pointwise inequality constraint $u(x) \geq 0$ for all $x \in \Omega$, since the exponential function returns nonnegative values.

(a.2) By substituting $\beta = 0$ in the expression for u^{opt} obtained in part (a.1), we get that

$$u_0^{\text{opt}}(x) = \frac{1}{\text{vol}\left(\Omega\right)},$$

which is a uniform PDF supported over Ω. The interpretation is that the uniform PDF maximizes entropy over a compact support, which intuitively makes sense since in the absence of additional information, maximum entropy occurs when all values in Ω are equally likely.

(b) Expectation constrained entropy maximization.

Suppose now we would like to minimize $I(u) = \mathbb{E}\left[\log u\right]$ subject to the constraints $\mathbb{E}\left[g_i(x)\right] = a_i$, $i = 1, ..., m$, where the vectors $g(x) = (g_1(x), ..., g_m(x))^\top$ and $a = (a_1, ..., a_m)^\top$ are given.

(b.1) Prove that the minimizing PDF u^{opt} must be of exponential family, i.e.,

$$u^{\text{opt}}(x) = \frac{\exp\left(-\langle \lambda, g(x)\rangle\right)}{\int_\Omega \exp\left(-\langle \lambda, g(x)\rangle\right) \mathrm{d}x}$$

for some $m \times 1$ vector λ that is independent of x.

(b.2) Specializing your answer in part (b.1) for an 1D interval $\Omega = [a, b]$, show that the maximum entropy PDF that has a given mean $\mathbb{E}\left[x\right] = \overline{x} \in (a, b)$, must be of the form

$$u^{\text{opt}}(x) = \frac{e^{-\lambda x}}{Z}$$

where the constants Z and λ are to be found in terms of $a, b, \overline{x}$.

Solution for part (b):

(b.1) For $i = 1, ..., m$, let λ_i be the Lagrange multiplier associated with the i-th integral equality constraint $\mathbb{E}\left[g_i(x)\right] = a_i$. Furthermore, let μ be the Lagrange multiplier associated with the normalization constraint $\int_\Omega u = 1$. Letting $\lambda := (\lambda_1, ..., \lambda_m)^\top$, $\nu := (\mu, \lambda)^\top \in \mathbb{R}^{m+1}$, the augmented Lagrangian is

$$L + \langle \nu, M \rangle = u \log u + \mu u + \langle \lambda, g(x) u \rangle.$$

Therefore, the EL equation

$$\frac{\partial}{\partial u}\left(L + \nu M\right) - \nabla \cdot \frac{\partial}{\partial \nabla u}\left(L + \nu M\right) = 1 + \log u + \mu + \langle \lambda, g(x) \rangle = 0$$

yields optimizer $u^{\text{opt}} = \exp\left(-1 - \mu - \langle \lambda, g(x) \rangle\right)$. Imposing the feasibility constraint $\int_\Omega u = 1$ as in (a.1), we get that

$$\exp(-1 - \mu) = \frac{1}{\int_\Omega \exp\left(-\langle \lambda, g(x) \rangle\right) \mathrm{d}x},$$

which upon substituting back to u^{opt}, gives the desired result. As in (a.1), here again, we did not need to explicitly enforce $u \geq 0$ since the exponential function returns nonnegative values.

(b.2) Specializing the answer in part (b.1) for $\Omega = [a, b]$, $m = 1$, $g_1(x) = x$, $a_1 = \overline{x}$, we get $u^{\text{opt}}(x) = \dfrac{e^{-\lambda x}}{Z}$ where the constants λ, Z solve

$$Z = \int_a^b e^{-\lambda x} \mathrm{d}x = \frac{e^{-\lambda a} - e^{-\lambda b}}{\lambda}, \qquad Z\,\overline{x} = \int_a^b x e^{-\lambda x} \mathrm{d}x = \frac{e^{-\lambda a}\left(\lambda a + 1\right) - e^{-\lambda b}\left(\lambda b + 1\right)}{\lambda^2}.$$

Equivalently, λ, Z solve the system of coupled nonlinear equations:

$$\lambda Z = e^{-\lambda a} - e^{-\lambda b}, \qquad \lambda Z\,\overline{x} + Z = a e^{-\lambda a} - b e^{-\lambda b}.$$

From here, these two equations can be solved (numerically) to compute the unknowns λ, Z in terms of $a, b, \overline{x}$.