Pensées 4: Matrix Exponentiation, Basic Exterior Calculus, Maximum Entropy PDFs

0x10af

Matrix Exponentiation Recovers the Analytic Continuation of the Trigonometric Functions

$$\cos\varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}$$

$$\sin\varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i} = \frac{-ie^{i\varphi} + ie^{-i\varphi}}{2}$$

$$\exp\left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\varphi\right) = \begin{bmatrix} \frac{e^{i\varphi}+e^{-i\varphi}}{2} & \frac{ie^{i\varphi}-ie^{-i\varphi}}{2} \\ \frac{-ie^{i\varphi}+ie^{-i\varphi}}{2} & \frac{e^{i\varphi}+e^{-i\varphi}}{2} \end{bmatrix} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

Double Exterior Derivative and Green’s Theorem

It is true for any k-form ω that:

$$dd\omega = 0$$

But we particularly focus on the 0-form f:

$$ddf = 0$$

$$df = \partial_i f dx^i$$

$$ddf = \partial_i \partial_j f dx^i \wedge dx^j$$

$$\partial_a \partial_b f = \partial_b \partial_a f \Rightarrow ddf = 0$$

The contraction of symmetric and skew-symmetric indices necessarily is 0, many of these identities are derived from this fact. Green’s theorem can be easily derived from the generalized Stokes’s theorem, and related to $dd\omega = 0$.

$$\int_{\partial\Omega} \omega = \int_{\Omega} d\omega$$

$$v = v_i dx^i$$

$$\int_{\partial\Omega} v = \int_{\Omega} dv = \int_{\Omega} \partial_i v_j dx^i \wedge dx^j$$

The integral over the boundary is closed, thus it being zero follows from v being conservative ($\exists f\ v = df$). The integral of $dv = ddf$ being zero also follows from $dd\omega = 0$.

$$v = df \Rightarrow \int_{\partial\Omega} v = \int_{\Omega} dv = \int_{\Omega} \partial_i \partial_j f dx^i \wedge dx^j = 0$$

Some Vector Calculus Identities

The curl does not need to use the covariant derivative because of the symmetry of the covariant indices of the connection coefficients and the skew-symmetry of the Levi-Civita tensor.

$$\varepsilon^{ijk}\nabla_i v_j = \varepsilon^{ijk}\left(\partial_i v_j - \Gamma^l_{ij} v_l\right) = \varepsilon^{ijk}\partial_i v_j$$

The divergence of the curl of a vector field is zero; the Levi-Civita tensor is like a constant with respect to the covariant derivative which is why it can be pulled out and again the symmetry makes it zero.

$$\nabla_l \varepsilon^{ijk} = 0$$
$$\nabla \cdot \nabla \times v = \nabla_k \varepsilon^{ijk}\nabla_i v_j = \varepsilon^{ijk}\nabla_k \nabla_i v_j = 0$$

The curl of the gradient of a function is zero, closely related to ddf and the part about Green's theorem.

$$\nabla \times \nabla f = \varepsilon^{ijk}\nabla_i \nabla_j f = 0$$

Some Maximum Entropy Distributions

Lagrange multipliers are used to find the solution, our constraints are the definition of a PDF and having mean μ.

$$L = \sum_{x=0}^{\infty} p(x)\log p(x) - \lambda_0\left(\sum_{x=0}^{\infty} p(x) - 1\right) - \lambda_1\left(\sum_{x=0}^{\infty} x p(x) - \mu\right)$$

Deriving each term with respect to $p(x)$. I'm not sure if the chain rule needs to be used for the third term, but $\frac{\partial x}{\partial p(x)}$ makes no sense and I got the right answer.

$$\log p(x) + 1 - \lambda_0 - \lambda_1 x$$
$$p(x) = \exp(\lambda_1 x + \lambda_0 - 1) = ab^x$$

We have the form of the expression with unknown constants a and b, so we must use the conditions to find them. We use solutions of power series and sticking it into Wolfram Alpha.

$$\sum_{x=0}^{\infty} ab^x = \frac{a}{1-b} = 1$$
$$\sum_{x=0}^{\infty} ab^x x = \frac{ab}{(b-1)^2} = \mu$$
$$a = \frac{1}{\mu+1}$$
$$b = \frac{\mu}{\mu+1}$$
$$\mu = \frac{1-p}{p}$$
$$p(x) = p(1-p)^x$$

It can be seen that the PDF of the distribution is equivalent to that of a geometric distribution (shifted by one from some conventions).

The normal distribution maximizes entropy with a given mean and variance, μ and σ^2. The third term is the integral version of the variance and also includes the mean, so the expectation is not necessary to include as a bound.

$$L = \int_{-\infty}^{\infty} p(x) \log p(x) \,\mathrm{d}x - \lambda_0 \left(\int_{-\infty}^{\infty} p(x) \,\mathrm{d}x - 1 \right) - \lambda_1 \left(\int_{-\infty}^{\infty} p(x)(x-\mu)^2 \,\mathrm{d}x - \sigma^2 \right)$$

$$\log(p(x)) + 1 + \lambda_0 + \lambda_1 (x-\mu)^2$$

$$p(x) = \alpha e^{-\beta(x-\mu)^2}$$

Differentiating we find a correct-looking form for our function, note that it is necessary that $-\beta < 0$, so I defined β as positive and used the negative of it.

$$\int_{-\infty}^{\infty} \alpha e^{-\beta(x-\mu)^2} \,\mathrm{d}x = \alpha \int_{-\infty}^{\infty} e^{-\beta(x-\mu)^2} \,\mathrm{d}x = \alpha \sqrt{\frac{\pi}{\beta}} = 1$$

$$\int_{-\infty}^{\infty} \alpha e^{-\beta(x-\mu)^2} (x-\mu)^2 \,\mathrm{d}x = \alpha \int_{-\infty}^{\infty} e^{-\beta z^2} z^2 \,\mathrm{d}z = \frac{\alpha\sqrt{\pi}}{2\beta^{\frac{3}{2}}} = \sigma^2$$

We have all our conditions (through annoying integrals) and can now solve.

$$\alpha = \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$\beta = \frac{1}{2\sigma^2}$$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

It makes sense that so many things tend to be approximately normal, not adding any information beyond μ and σ^2 makes the most “reasonable” or least informative distribution normal, the differences from that normal distribution are extra information added on which cannot be assumed and will not be expected to be true generally.