C

Mathematical Tools

In this appendix you find a number of mathematical identities and definitions useful for statistical calculations.

C–1 Nabla operator, gradient, and Jacobi matrix

The **nabla operator** is the row vector

$$\nabla = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d})$$

containing the first order partial derivative operators.

The **gradient** of a univariate function $f(\boldsymbol{x})$ with argument $\boldsymbol{x} = (x_1, \ldots, x_d)^T$ is also a row vector and can be expressed using the nabla operator

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_d} \right) .$$

The **Jacobi matrix**

$$\boldsymbol{J_h}(\boldsymbol{x}) = \begin{pmatrix} \nabla h_1(\boldsymbol{x}) \\ \vdots \\ \nabla h_m(\boldsymbol{x}) \end{pmatrix}$$

is the generalization of the gradient for a vector-valued function

$$\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \ldots, h_m(\boldsymbol{x}))^T .$$

The Jacobi matrix is useful to linearly approximate $\boldsymbol{h}(\boldsymbol{x})$ around some $\boldsymbol{x}_0$ via

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) \approx \boldsymbol{h}(\boldsymbol{x}_0) + \boldsymbol{J_h}(\boldsymbol{x}_0)\, \Delta \boldsymbol{x} .$$

For a univariate function this becomes

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) \approx f(\boldsymbol{x}_0) + (\nabla f(\boldsymbol{x}_0))\; \Delta \boldsymbol{x} .$$

In integral calculus the Jacobi matrix appears when transforming the infinitesimal volume element $\boldsymbol{dx}$ under coordinate changes $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$. Specifically,

$$\boldsymbol{dx} = \left|\det\left(\boldsymbol{J}_{\boldsymbol{x}}(\boldsymbol{y})\right)\right| \boldsymbol{dy}\,,$$

where $\boldsymbol{J}_{\boldsymbol{x}}(\boldsymbol{y})$ is the Jacobi matrix of the inverse transformation $\boldsymbol{x}(\boldsymbol{y})$. Note that in this case the Jacobian $\boldsymbol{J}_{\boldsymbol{x}}(\boldsymbol{y})$ is a square matrix because coordinate transformations are one-to-one mappings (also the determinant is only defined for a square matrix).

C–2 Hessian matrix

The **Hessian** matrix collects second order partial derivatives of a univariate function and can be written in nabla notation as

$$\nabla^T \nabla f(\boldsymbol{x}) = \boldsymbol{H}_f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_d^2} \end{pmatrix}.$$

By construction it is square and symmetric.

The Hessian matrix is needed for quadratic approximation of a univariate function via the Taylor series

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) \\ &\approx f(\boldsymbol{x}_0) + (\nabla f(\boldsymbol{x}_0))\ \Delta \boldsymbol{x} + \frac{1}{2} (\Delta \boldsymbol{x})^T \boldsymbol{H}_f(\boldsymbol{x}_0)\, \Delta \boldsymbol{x}\,. \end{aligned}$$

If the function $f(\boldsymbol{x})$ is expanded around a local optimum x_{opt} with $\nabla f(\boldsymbol{x}_{\text{opt}}) = 0$ then the above further simplifies to

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_{\text{opt}} + \Delta \boldsymbol{x}) \approx f(\boldsymbol{x}_{\text{opt}}) + \frac{1}{2} (\Delta \boldsymbol{x})^T \boldsymbol{H}_f(\boldsymbol{x}_{\text{opt}})\, \Delta \boldsymbol{x}\,.$$

If the optimum is a mode (i.e. a local maximum) then $-\boldsymbol{H}_f(\boldsymbol{x}_{\text{mode}})$ is positive definite.

C–3 Laplace integral approximation

The method of Laplace for computing the integral $\int_{-\infty}^{\infty} f(\boldsymbol{x}) \boldsymbol{dx}$ relies on the quadratic approximation of $\log f(\boldsymbol{x})$. The function $f(\boldsymbol{x})$ must be positive everwhere and also have a maximum at $\boldsymbol{x}_{\text{mode}}$. Expanding $\log f(\boldsymbol{x})$ we get

$$\begin{aligned}\log f(\boldsymbol{x}) &= \log f(\boldsymbol{x}_{\text{mode}} + \Delta \boldsymbol{x}) \\ &\approx \log f(\boldsymbol{x}_{\text{mode}}) - \frac{1}{2}(\Delta \boldsymbol{x})^T \boldsymbol{K}^{-1}\, \Delta \boldsymbol{x}\end{aligned}$$

with $\boldsymbol{K} = -\left(\boldsymbol{H}_{\log f}(\boldsymbol{x}_{\text{mode}})\right)^{-1}$. With this the integral can be written

$$\begin{aligned}\int_{-\infty}^{\infty} f(\boldsymbol{x}) d\boldsymbol{x} &= \int e^{\log f(\boldsymbol{x})} d\boldsymbol{x} \\ &\approx f(\boldsymbol{x}_{\text{mode}}) \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_{\text{mode}})^T \boldsymbol{K}^{-1}\,(\boldsymbol{x}-\boldsymbol{x}_{\text{mode}})} d\boldsymbol{x} \\ &= (2\pi)^{d/2} \det(\boldsymbol{K})^{1/2} f(\boldsymbol{x}_{\text{mode}})\,,\end{aligned}$$

where d is the dimension of $\boldsymbol{x}$. Note that the integral in the last step corresponds to the normalization constant of the multivariate normal distribution.

C–4 Sterling's formula

Sterling's formula is a highly useful approximation for the factorial $n! = \prod_{i=1}^{n} i$ which we can also define via the gamma function as

$$n! = \Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = \int_0^\infty e^{n\log(x)-x} dx.$$

If we evaluate the integral by the Laplace method with $\log f(x) = n\log(x) - x$, $x_{\text{mode}} = n$, $f(x_{\text{mode}}) = n^n e^{-n}$ and $K = n$ we directly recover Sterling's formula

$$n! \approx \sqrt{2\pi n}\, n^n e^{-n}\,.$$

On the log scale it becomes

$$\log(n!) \approx \frac{1}{2}\log(2\pi) + (n + \frac{1}{2})\log(n) - n$$

which for large n further simplifies to

$$\log(n!) \approx n\log(n) - n\,.$$