DATA 1010
In-class exercises
Samuel S. Watson
05 November 2018

Problem 1

Suppose that $f : [a, b] \to \mathbb{R}$ is a smooth function whose values on a grid of points `xs = a:ϵ:b` are stored in a vector `v`. Write a Julia expression which approximates $\int_a^b f$.

Solution

We sum the function's values and multiply by the grid increment: `sum(v)*step(xs)`. Each of the terms in this sum represent the area of a rectangle whose base is `step(xs)` and whose height is the value of the function at the corresponding grid point, so this is a Riemann sum approximation of the integral.

Problem 2

Suppose that $f : [a, b] \to \mathbb{R}$ is a probability density function whose values on `xs = a:ϵ:b` are stored in a vector `v`. Write a Julia expression which approximates $\mathbb{E}[X]$, where X is a random variable whose density is f.

Solution

The integral we seek to approximate is $\int_a^b x f(x)\, \mathrm{d}x$, so we calculate either `sum(x .* v)*step(xs)` (following the previous example) or `sum(x .* v)/sum(v)` (thinking of the values of `v` as weights in a weighted average).

These expressions yield approximately the same result since `sum(v)*step(xs)` is approximately $\int_a^b f$, which is 1 since f is a PDF.

Problem 3

Given a flower randomly selected from a field, let X_1 be its petal width in centimeters, X_2 its petal length in centimeters, and $Y \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$ its color. Let

$$\begin{array}{lll} \boldsymbol{\mu}_{\mathrm{R}} = \begin{bmatrix} 9 \\ 5 \end{bmatrix} & \boldsymbol{\mu}_{\mathrm{G}} = \begin{bmatrix} 4 \\ 10 \end{bmatrix} & \boldsymbol{\mu}_{\mathrm{B}} = \begin{bmatrix} 7 \\ 9 \end{bmatrix} \\ A_{\mathrm{R}} = \begin{bmatrix} 1.5 & -1 \\ 0 & 1 \end{bmatrix} & A_{\mathrm{G}} = \begin{bmatrix} 0.5 & 0.25 \\ 0 & 0.5 \end{bmatrix} & A_{\mathrm{B}} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}. \end{array}$$

Suppose that the joint distribution of X_1, X_2, and Y has the property that for any $A \subset \mathbb{R}^2$ and color $c \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$, we have

$$\mathbb{P}(A \times \{c\}) = p_c \int_{\mathbb{R}^2} f_c(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2,$$

where $(p_R, p_G, p_B) = (1/3, 1/6, 1/2)$ and f_c is the multivariate normal density with mean μ_c and covariance matrix $A_c A_c'$.

Find the best predictor of Y given $(X_1, X_2) = (x_1, x_2)$ (using the 0-1 loss function), and find a way to estimate that predictor using the given samples.

Solution

See the course text (second volume) for a solution.