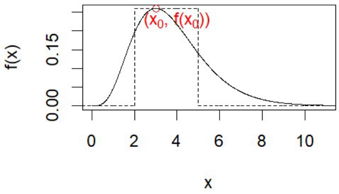
Algorithms and Data Structures

Quadrature
Laplace's method



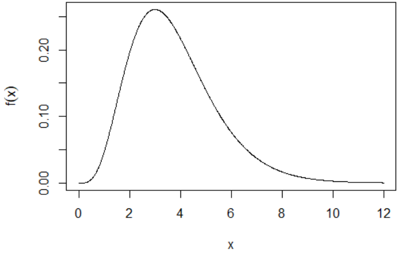
Learning goals

- Laplace's method

LAPLACE'S METHOD

Target: Approximate integral of function f with the following properties:

- The mass concentrates on a small area around a center and the function has very rapidly decreasing tails ("similarity" to the density of a normal distribution)
- The function we want to integrate is the density of a random variable that is approximately normally distributed



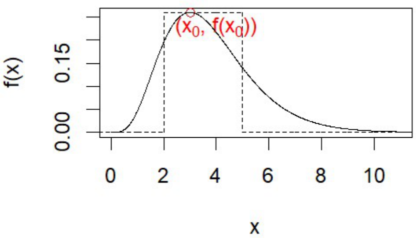
LAPLACE'S METHOD / 2

In particular, we assume that f

- Can only be positive
- Is two times continuously differentiable
- Has a **global maximum** at x_0

We could approximate the area underneath the graph of the function with a staircase function and represent the integral with a very simple formula that depends on $f(x_0)$:

$$\int f(x)\,dx \approx f(x_0) \cdot c$$



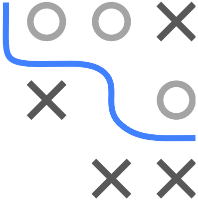
LAPLACE'S METHOD / 3

But instead of the step function we would like to choose a function that approximates f **better** and which has well-known properties.

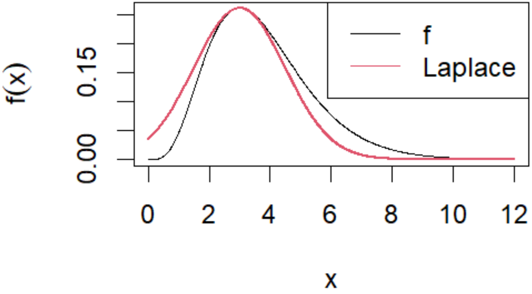
Idea: Approximate the integral using the density function of the normal distribution!

How? We center and scale the density function of the normal distribution such that it approximates f "best possible".

LAPLACE'S METHOD / 4



In other words: We determine **expectation** and **standard deviation** of a normal distribution such that the corresponding density function fits best possible to the function *f* we are interested in.



LAPLACE'S METHOD / 5

Mathematical derivation:

Let there be a function f with a global maximum at x_0.

We define $h(x) := \log f(x)$ as the logarithmized function and rewrite the integral

$$\int_a^b f(x)\, dx = \int_a^b \exp(\underbrace{\log f(x)}_{:=h(x)})\, dx$$

Using Taylor's theorem around x_0 we obtain

$$\int_a^b \exp\left(h(x)\right) \approx \int_a^b \exp\left(h(x_0) + h'(x_0)(x - x_0) + \frac{1}{2}h''(x_0)(x - x_0)^2\right) dx$$

LAPLACE'S METHOD / 6

x_0 is also the maximum of $h(x) = \log(f(x))$. Hence, $h'(x_0) = 0$ and the second summand disappears:

$$\int_a^b \exp\left(h(x)\right)\, dx \approx \int_a^b \exp\left(h(x_0) + \frac{1}{2}h''(x_0)(x - x_0)^2\right) dx$$

We take advantage of the fact that $\exp(x + y) = \exp(x)\exp(y)$

$$\int_a^b \exp\left(h(x_0)\right) \cdot \exp\left(\frac{1}{2}h''(x_0)(x - x_0)^2\right) dx$$

and pull the constant $\exp\left(h(x_0)\right)$ out of the integral

$$\exp\left(h(x_0)\right) \cdot \int_a^b \exp\left(\frac{1}{2}h''(x_0)(x - x_0)^2\right) dx$$

LAPLACE'S METHOD / 7

Within the integral there is now an expression which "almost" corresponds to the density of a normal distribution with expectation $\mu := x_0$ and variance $\sigma^2 := -h''(x_0)^{-1}$:

$$
\begin{aligned}
\int_a^b f(x)dx &\approx \exp\left(h(x_0)\right) \cdot \int_a^b \exp\left(\frac{1}{2}h''(x_0)(x-x_0)^2\right) dx \\
&= \exp\left(h(x_0)\right) \cdot \int_a^b \exp\left(-\frac{1}{2}\frac{(x-x_0)^2}{-h''(x_0)^{-1}}\right) dx \\
&= \exp\left(h(x_0)\right) \cdot \int_a^b \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx
\end{aligned}
$$

$-h''(x_0)^{-1}$ must be truly positive to correspond to the variance of a normal distribution. Since $h(x)$ has a global maximum in x_0, the second derivative at this point is negative and therefore $-h''(x_0)^{-1} > 0$.

LAPLACE'S METHOD / 8

If we add (and cancel) the multiplicative constant $c = \frac{1}{\sqrt{2\pi\sigma^2}}$, we obtain

$$
\begin{aligned}
\int_a^b f(x)dx \quad &\approx \quad \frac{1}{c} \cdot \exp\left(h(x_0)\right) \cdot \int_a^b \underbrace{c \cdot \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)}_{\text{Density ND}} dx \\
&= \quad \frac{1}{c} \underbrace{\exp\left(h(x_0)\right)}_{f(x_0)} \cdot \int_a^b \phi_{\mu,\sigma^2}(x)\, dx \\
&= \quad \frac{1}{c} f(x_0) \cdot \left(\Phi_{\mu,\sigma^2}(b) - \Phi_{\mu,\sigma^2}(a)\right)
\end{aligned}
$$

where $\phi_{\mu,\sigma^2}(x)$ denotes the density and $\Phi_{\mu,\sigma^2}(x)$ the distribution function of a normal distribution with expectation μ and variance σ^2.

LAPLACE'S METHOD / 9

For integration limits $b = \infty$ and $a = -\infty$ Laplace's method of f is then

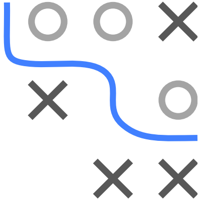
$$\int_{-\infty}^{\infty} f(x)dx \approx \frac{1}{c} \cdot f(x_0) \cdot \left(\Phi_{\mu,\sigma^2}(+\infty) - \Phi_{\mu,\sigma^2}(-\infty)\right)$$

$$= \sqrt{-\frac{2\pi}{h''(x_0)}} \cdot f(x_0)$$

with $h(x) = \log f(x)$.

Laplace's method thus corresponds to a value that only depends on the maximum of the function $f(x_0)$ and the curvature of the logarithmic function $h''(x_0)$.

LAPLACE'S METHOD / 10



Laplace's method also works well in higher dimensions. For $f : \mathbb{R}^m \to \mathbb{R}$ with global maximum in $\boldsymbol{x}_0$ the generalized form is given by

$$I(f) \approx (2\pi)^{m/2} \det(-H_f(x_0)^{-1})^{1/2} \exp(f(x_0))$$

where $H_f(x_0)$ denotes the Hessian matrix of f at x_0. Since x_0 is a global maximum, $H_f(x_0)$ is negative definite.

The problem of integration is reduced to

- Solving an optimization problem $\to$ find x_0
- Determining the second derivative $h''(x)$ (or generally the Hessian matrix $H_f(\boldsymbol{x})$) at the optimal position x_0.

Instead of integration, an optimization problem must now be solved, which is often much easier and faster.

LAPLACE'S METHOD: EXAMPLE

Application example: Bayesian computation

Given:

$$\begin{aligned} x|\lambda &\sim \text{Poisson}(\lambda) \quad \text{(Likelihood)} \\ \lambda &\sim \text{Gamma}(\alpha, \beta) \quad \text{(Prior)} \end{aligned}$$

Wanted: Posterior density of the parameter λ given n observations $\boldsymbol{x} = \left(x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right)$

$$\overset{\textit{Posterior}}{p(\lambda|\boldsymbol{x})} = \frac{\overset{\textit{Likelihood}}{p(\boldsymbol{x}|\lambda)} \cdot \overset{\textit{Prior}}{\pi(\lambda)}}{\int p(\boldsymbol{x}|\lambda) \cdot \pi(\lambda)\, d\lambda}$$

The density of the gamma distribution is given by $\pi_{\alpha,\beta}(\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda\beta)$

LAPLACE'S METHOD: EXAMPLE / 2

To keep the calculations simple, we calculate the posterior density for only **one** observation x.

The posterior density of λ given the observation x is (except for one constant)

$$p(\lambda|x) \propto \lambda^{x+\alpha-1} \exp\left(-\frac{\lambda}{1/\beta+1}\right) =: f(\lambda).$$

So to determine the posterior density $p(\lambda|x)$ **exactly**, we search for the normalization constant c, which ensures that $\int c \cdot f(\lambda)\, d\lambda = 1$, hence

$$\begin{aligned} c \cdot \int f(\lambda)\, d\lambda &= 1 \\ c &= \frac{1}{\int f(\lambda)\, d\lambda} \end{aligned}$$

LAPLACE'S METHOD: EXAMPLE / 3

Goal: Approximation of $\int f(\lambda)\, d\lambda$ with $f(\lambda) = \lambda^{x+\alpha-1} \exp\left(-\frac{\lambda}{1/\beta+1}\right)$

We calculate $h(\lambda) = \log f(\lambda)$

$$
\begin{aligned}
h(\lambda) &= \log\left(\lambda^{x+\alpha-1} \cdot \exp(-\frac{\lambda}{1/\beta+1})\right) \\
&= (x+\alpha-1)\log\lambda - \frac{\lambda}{1/\beta+1} \\
h'(\lambda) &= \frac{x+\alpha-1}{\lambda} - \frac{1}{1/\beta+1} \\
h''(\lambda) &= -\frac{x+\alpha-1}{\lambda^2}
\end{aligned}
$$

LAPLACE'S METHOD: EXAMPLE / 4

To approximate the integral using Laplace's method, we need $\lambda_0 := \arg\max f(\lambda)$ and $h''(\lambda_0)$, where $h(\lambda) := \log(f(\lambda))$.

The maximum of $f(\lambda)$ is the same as the maximum of $h(\lambda)$ (easier to calculate)

$$\begin{aligned} h'(\lambda) &= 0 \\ \frac{x+\alpha-1}{\lambda} - \frac{1}{1/\beta+1} &= 0 \\ \lambda_0 &= \frac{x+\alpha-1}{1/\beta+1} \end{aligned}$$

and thus

$$h''(\lambda_0) = -\frac{(1/\beta+1)^2}{x+\alpha-1}$$

LAPLACE'S METHOD: EXAMPLE / 5

We insert $\lambda_0 = \frac{x+\alpha-1}{1/\beta+1}$ and $h''(\lambda_0)$ into the formula for Laplace's method and obtain

$$\begin{aligned}
\int f(\lambda) d\lambda &\approx \sqrt{-\frac{2\pi}{h''(\lambda_0)}} \cdot f(\lambda_0) \\
&= \sqrt{2\pi} \cdot \frac{\sqrt{x+\alpha-1}}{1/\beta+1} \cdot f(\lambda_0)
\end{aligned}$$

Hence, the normalization constant c can be approximated by

$$c = \frac{1}{\int f(\lambda) d\lambda} \approx \frac{1}{\sqrt{2\pi}} \frac{1/\beta+1}{\sqrt{x+\alpha-1}} \cdot \frac{1}{f(\lambda_0)}$$

LAPLACE'S METHOD: EXAMPLE / 6

When calculating posterior distributions, Laplace's method provides a good approximation if

- The number n of observations is large
- The posterior distributions are roughly symmetric