Classical quadrature rules via Gaussian processes

Toni Karvonen & Simo Särkkä
Department of Electrical Engineering and Automation
Aalto University, Finland

Quadrature rules

Classical quadrature

In classical quadrature, the aim is to integrate many polynomials exactly. Given any \( n \) distinct nodes \( x_1, \ldots, x_n \in \Omega \subset \mathbb{R} \), it is possible to construct a quadrature rule

\[
Q(f) = \sum_{i=1}^{n} w_i f(x_i) \approx \int_{\Omega} f(x) \, d\mu
\]

that is exact for all polynomials of degree at most \( n - 1 \) by solving the weights from \( Vw = c \), where \( \{V_i\}_{i=1}^{n} = (x_i^{-1}) \) is the invertible Vandermonde matrix and \( \{c_i\} = f_i x_i^{-1} \, d\mu(x) \).

A classical quadrature rule is of degree \( m \) if it is exact for all polynomials of degree at most \( m \). One can do better than degree \( n - 1 \) with \( n \) points. An \( n \)-point Gaussian quadrature rule is of degree \( 2n - 1 \). This rule is unique (given the measure \( \mu \)) and its weights \( w_i \) are positive.

Bayesian quadrature

In Bayesian quadrature [1,2], the integrand \( f \) is assigned a Gaussian process prior \( f \sim \mathcal{N}(0,k) \) with a positive-definite covariance kernel \( k \). The data \( \mathcal{D} = \{(x_i, f(x_i)), \ldots, (x_n, f(x_n))\} = (X, y) \) induce a Gaussian posterior \( f | \mathcal{D} \) and, consequently, a Gaussian posterior on the integral \( \int_{\Omega} f(x) \, d\mu | \mathcal{D} \) with the mean and variance

\[
Q_B(f) = \mathbb{E} \left( \int_{\Omega} f(x) \, d\mu | \mathcal{D} \right) = y^T K^{-1} k(X), \\
V_B = \text{Var} \left( \int_{\Omega} f(x) \, d\mu | \mathcal{D} \right) = \int_{\Omega} k(x) \, d\mu - \int_{\Omega} k(x) \, K^{-1} k(x) \, d\mu
\]

where \( k(x) = \int_{\Omega} k(x, x) \, d\mu(x), \) \( k_j(x) = k(x, x_j) \), and \( |K|_{ij} = k(x_i, x_j) \). Following the philosophy of probabilistic numerics [3], the integral posterior variance \( V_B \) can be used in quantifying the uncertainty inherent to the numerical integral approximation \( Q_B(f) \).

A numerical experiment

- We work with \( \Omega = [-1, 1] \) and \( d\mu = \frac{1}{2} dx \).
- The orthogonal polynomials are the Legendre polynomials so \( k^p \) is constructed using them.
- We set \( n = 4 \) and select the four nodes as those of the unique Gaussian quadrature rule.
- For \( p = 4, \ldots, 8 \), the kernel \( k^p \) yields a Bayesian quadrature rule that coincides with the Gaussian one by Theorem 1.
- The posterior processes for (for \( p = 4 \) the posterior variance vanishes) are depicted here.
- Note that the posterior variances are non-zero even though \( V_B = 0 \) by Theorem 1.

Conclusions and discussion

- All classical quadrature rules in one dimension can be interpreted as Bayesian quadrature rules if the kernel is selected suitably.
- In particular, Gaussian quadrature rules are unique optimal Bayesian rules when \( p = m = 2n \).
- In the paper we also provide some interesting multivariate generalisations.
- It is likely that the assumption on \( \varphi_i \) being orthogonal polynomials is not necessary.

Main results

Objectives

Särkkä et al. [4] have shown that many classical quadrature rules popular in non-linear Kalman filtering can be interpreted as Bayesian quadrature rules if the kernel is selected suitably. We clarify and extend their analysis.

Results

Let \( \varphi_0, \ldots, \varphi_{p-1} \) be polynomials that form a basis of the space of polynomials of degree at most \( p - 1 \). We define \( k^p \), the polynomial kernel of degree \( p \), as

\[
k^p(x, x') = \sum_{i=0}^{p-1} \varphi_i(x) \varphi_i(x').
\]

A Bayesian quadrature rule is said to coincide with a classical quadrature rule if the rules have the same nodes and weights.

**Theorem 1.** Let \( \varphi_0, \ldots, \varphi_{p-1} \) be the orthogonal polynomials. Consider the Bayesian quadrature rule with the kernel \( k^p \) and nodes \( X \) and the classical quadrature rule that is of degree \( m - 1 \) and uses the same nodes. Then these rules coincide if and only if \( n = m \). If the rules coincide, \( V_B = 0 \).

When \( p < m = 2n \), there are in general multiple optimal Bayesian quadrature rules (i.e. rules whose nodes globally minimise \( V_B \)). For \( p = m = 2n \), uniqueness of Gaussian quadrature rules results in the following corollary.

**Corollary 2.** When \( \varphi_i \) are the orthogonal polynomials there is a unique \( n \)-point optimal Bayesian quadrature rule for the kernel \( k^{2n} \). This is the Gaussian quadrature rule for the measure \( \mu \).

References