Nested Expectations with Kernel Quadrature

Zonghao Chen[1] Masha Naslidnyk[1] François-Xavier Briol[1]

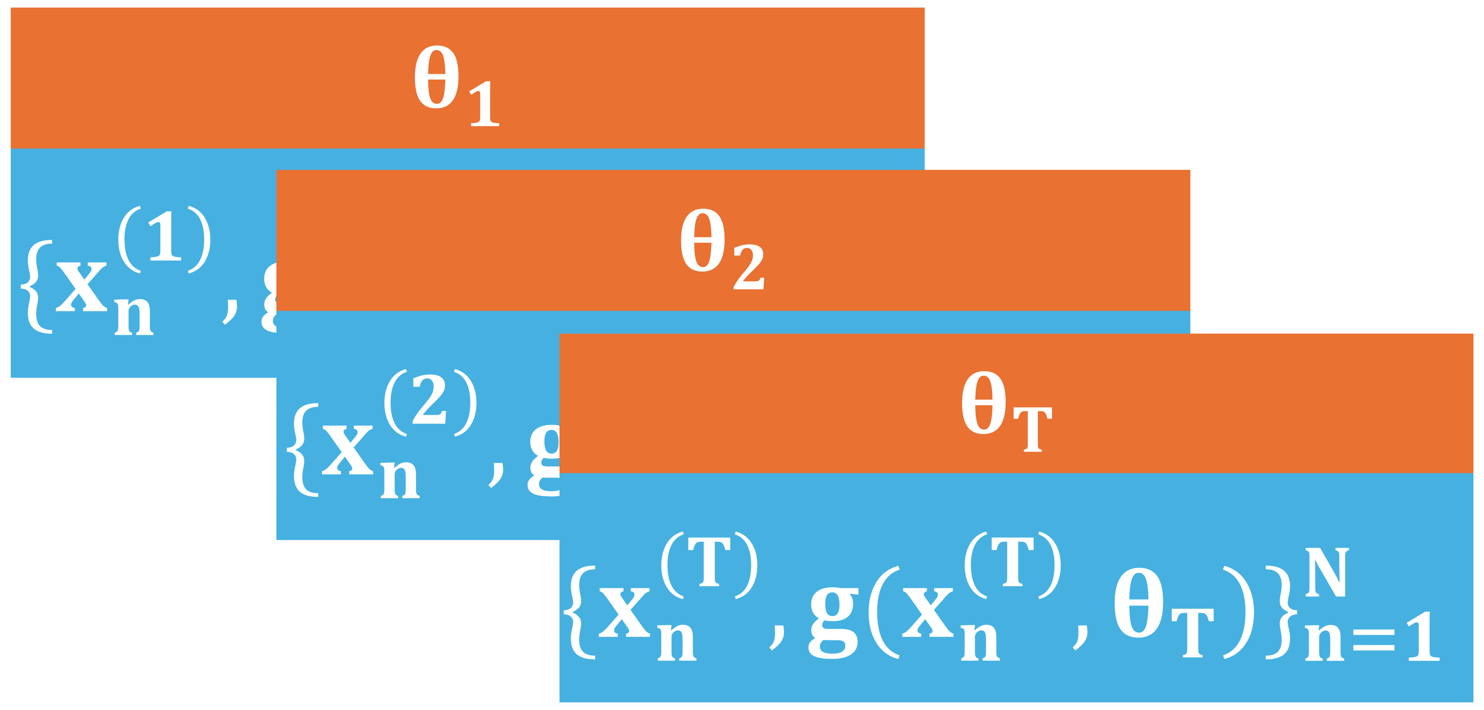
[1] University College London @Hudson19990518

TL;DR

We are interested in estimating nested expectation.

$$I := \mathbb{E}_{\theta \sim \mathbb{Q}}\left[f\left(\mathbb{E}_{X\sim \mathbb{P}_\theta}[g(X,\theta)]\right)\right]$$
$$= \underbrace{\int_\Theta f\Big(\underbrace{\int_{\mathcal{X}} g(x,\theta)p_\theta(x)dx}_{\text{inner conditional expectation}}\Big) q(\theta)d\theta}_{\text{outer expectation}}.$$

with T samples in Θ and N samples in $\mathcal{X}$.



Our Contributions:

1. We propose a novel estimator called *nested kernel quadrature* (NKQ)

$$\hat{I}_{\text{NKQ}} = \sum_{t=1}^{T} w_t^{\Theta} f\Big(\sum_{n=1}^{N} w_{n,t}^{\mathcal{X}} g(x_n^{(t)},\theta_t)\Big)$$

2. We prove NKQ has fast rate of convergence.

$$\left|\hat{I}_{\text{NKQ}} - I\right| = \tilde{\mathcal{O}}_P\left(N^{-\frac{s_\mathcal{X}}{d_\mathcal{X}}}\right) + \tilde{\mathcal{O}}_P\left(T^{-\frac{s_\Theta}{d_\Theta}}\right)$$

Full paper & code:
`https://www.arxiv.org/pdf/2502.18284`

Nested Kernel Quadrature

Kernel Quadrature

Consider $J = \mathbb{E}_{X\sim\pi}[h(X)]$. Suppose $h \in \mathcal{H}_\mathcal{X}$. KQ is a quadrature algorithm based on kernel interpolation (regression). Denote the kernel mean embedding as $\mu_\pi(\cdot) = \mathbb{E}_{X\sim\pi}[k_\mathcal{X}(X,\cdot)]$ and the Gram matrix as $\boldsymbol{K}_\mathcal{X}$.

$$\hat{J}_{\text{KQ}} = \mu_\pi(x_{1:N})(\boldsymbol{K}_\mathcal{X} + N\lambda \mathbf{I}_N)^{-1} h(x_{1:N})$$

Nested Kernel Quadrature

- **Stage I**: Estimate T inner expectations $J(\theta_t) = \mathbb{E}_{X\sim\mathbb{P}_{\theta_t}}[g(X,\theta_t)]$ with KQ.

$$\hat{J}_{\text{KQ}}(\theta_t) = \mu_{\mathbb{P}_{\theta_t}}(x_{1:N})(\boldsymbol{K}_\mathcal{X}^{(t)} + N\lambda_\mathcal{X}\mathbf{I}_N)^{-1} g(x_{1:N}^{(t)},\theta_t)$$

- **Stage II**: Denote $\hat{F}_{\text{KQ}}(\theta_t) = f(\hat{J}_{\text{KQ}}(\theta_t))$ for $t = \{1,\ldots,T\}$. Estimate the outer expectation again with KQ as if $\hat{F}_{\text{KQ}}(\theta_t)$ are the true observations!

$$\hat{I}_{\text{NKQ}} := \mu_\mathbb{Q}(\theta_{1:T})(\boldsymbol{K}_\Theta + T\lambda_\Theta \mathbf{I}_T)^{-1}\hat{F}_{\text{KQ}}(\theta_{1:T})$$

Sample Efficiency

The number of function evaluations or samples required to ensure $|I - \hat{I}| \leq \Delta$. Smaller exponents r in Δ^{-r} indicate a more efficient method.

NMC	$\mathcal{O}(\Delta^{-3})$ or $\mathcal{O}(\Delta^{-4})$
NQMC	$\mathcal{O}(\Delta^{-2.5})$
MLMC	$\mathcal{O}(\Delta^{-2})$
NKQ (Theorem)	$\tilde{\mathcal{O}}(\Delta^{-\frac{d_\mathcal{X}}{s_\mathcal{X}}-\frac{d_\Theta}{s_\Theta}})$

Main Theorem

Suppose the following assumptions hold:

1. The kernels $k_\mathcal{X}$ and k_Θ are Sobolev kernels of smoothness $s_\mathcal{X} > d_\mathcal{X}/2$ and $s_\Theta > d_\Theta/2$.
2. $D_\theta^\beta g(\cdot,\theta) \in W_2^{s_\mathcal{X}}(\mathcal{X})$ for all $\beta \in \mathbb{N}^{d_\Theta}$ with $|\beta| \leq s_\Theta$.
3. $g(x,\cdot) \in W_2^{s_\Theta}(\Theta)$ and $\theta \mapsto p_\theta(x) \in W_2^{s_\Theta}(\Theta)$.
4. f has bounded derivatives up to and including order $s_\Theta + 1$.

For large N, T, with probability $\geq 1 - 8e^{-\tau}$,

$$|\hat{I}_{\text{NKQ}} - I| \leq \tau\left(N^{-\frac{s_\mathcal{X}}{d_\mathcal{X}}}(\log N)^{\frac{s_\mathcal{X}+1}{d_\mathcal{X}}} + T^{-\frac{s_\Theta}{d_\Theta}}(\log T)^{\frac{s_\Theta+1}{d_\Theta}}\right)$$

Empirical Evaluations

- Synthetic Experiment.
- Option pricing in mathematical finance.
- Uncertainty decision making in health economics.
- Bayesian Optimization.