CMStatistics 2023

Quantile Importance Sampling

Jyotishka Datta [1]

December 18, 2023

Virginia Tech

Outline of My Talk

Quantile Importance Sampling[2]

- Evidence estimation.
- Riemann sum estimators: Yakowitz et al. [1978]: $O(n^{-4})$ rate.
- Nested sampling [Skilling, 2006]: high-dimensional integrals!
- Use Riemann sum estimators for nested sampling.

[2] Joint work with Nick Polson.

Evidence Estimation

- Interest: approximation of integrals of the form:

$$\psi = \mathbb{E}_F L(\boldsymbol{\theta}) = \int_{\chi} L(\boldsymbol{\theta}) dF(\boldsymbol{\theta})$$

  where $L(\boldsymbol{\theta})$ is a measurable function of interest and $F(\cdot)$ is a finite measure, such as a probability.

- Evidence estimation, i.e.

$$m(\mathbf{x}) = \int_{\Theta} L(\mathbf{x} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

  where $L(\mathbf{x} \mid \boldsymbol{\theta})$ is the likelihood with observed data $\mathbf{X} = \mathbf{x}$, $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter of interest and $\pi(\boldsymbol{\theta})$ is the prior over $\boldsymbol{\theta}$.

- Bayes factors for model comparison, numerical analysis, statistical mechanics etc [Llorente et al., 2020].

Evidence Estimation

- Llorente et al. [2020] classifies evidence estimation into four categories:
    1. deterministic approximation,
    2. density estimation,
    3. importance sampling, and
    4. vertical representation, which encompasses **nested sampling** [Skilling, 2006].
- I'll show that the Riemann sum estimator due to Yakowitz et al. [1978] can be used in the context of nested sampling [Skilling, 2006] to achieve a $O(n^{-4})$ rate of convergence, faster than the usual Ergodic Central Limit Theorem.

Outline

Merge two ideas:

Riemann sum estimators [Yakowitz et al., 1978]

Outline

Merge two ideas:

Riemann sum estimators [Yakowitz et al., 1978]

$+$

Nested sampling [Skilling, 2006]

$\Downarrow$

Outline

Merge two ideas:

Riemann sum estimators [Yakowitz et al., 1978]

$+$

Nested sampling [Skilling, 2006]

$\Downarrow$

Quantile Importance Sampling: QIS

Riemann sum estimators

Naive Monte Carlo

- Naïve Monte Carlo or the *usual* importance sampling method:
- The empirical average of a sample $(x_1, \ldots, x_n)$ drawn from either the target density f or a proposal density g:

$$\hat{\psi}_{IS} = \frac{1}{n} \sum_{i=1}^{n} \frac{L(x_i) f(x_i)}{g(x_i)} \rightarrow \int L(x) f(x) dx. \tag{1}$$

- $\hat{\psi}_{IS}$ converges to the true value ψ at a $O(n^{-1})$ rate.
- The density function $g(\cdot)$ is selected such that the weights $w(x_i)$ are approximately constant.
- Many, many alternatives and strategies to increase efficiency.

Riemann sums integrals over $[0,1]$

- Yakowitz et al. [1978] proposed a weighted Monte-Carlo scheme.
- Consider the Riemann sum estimator:

$$\hat{\psi}_Y = \sum_{i=0}^{n-1} \frac{L(u_{(i)}) + L(u_{(i+1)})}{2} (u_{(i+1)} - u_{(i)}), \text{ for } \psi = \int_0^1 L(x)dx, \tag{2}$$

where $\{u_{(i)}\}_{i=1}^{n}$ are ordered uniform draws from $f = \mathbb{I}_{[0,1]}$.

Riemann sums integrals over $[0, 1]$

- Yakowitz et al. [1978] proposed a weighted Monte-Carlo scheme.
- Consider the Riemann sum estimator:

$$\hat{\psi}_Y = \sum_{i=0}^{n-1} \frac{L(u_{(i)}) + L(u_{(i+1)})}{2}(u_{(i+1)} - u_{(i)}), \text{ for } \psi = \int_0^1 L(x)dx, \tag{2}$$

where $\{u_{(i)}\}_{i=1}^{n}$ are ordered uniform draws from $f = \mathbb{I}_{[0,1]}$.

Yakowitz (1978)

For $f \equiv \mathcal{U}(0, 1)$, the Riemann sum estimator $\hat{\psi}_Y$ in (2) attains a $O(n^{-4})$ rate of convergence.

Riemann sums for general f

- Yakowitz's estimator (2) was extended to a general (prior) density $f(\cdot)$ by Philippe [1997], Philippe and Robert [2001].

Riemann sums for general f

- Yakowitz's estimator (2) was extended to a general (prior) density $f(\cdot)$ by Philippe [1997], Philippe and Robert [2001].
- Take ordered samples:

$$x_{(1)} \doteq F^{-}(u_{(1)}) \leq x_{(2)} \doteq F^{-}(u_{(2)}) \leq \cdots \leq x_{(n)} \doteq F^{-}(u_{(n)}),$$

- Construct the Riemann sum estimator

$$\hat{\psi}_R = \sum_{i=0}^{n-1} L(x_{(i)}) f(x_{(i)})(x_{(i+1)} - x_{(i)}). \tag{3}$$

- This attains a convergence rate of $O(n^{-2})$, while admitting a vanishing bias of $O(n^{-1})$ order.
- Philippe [1997], Philippe and Robert [2001] argue this convergence is 'far from formal', i.e. noticeable within a short range.

Demonstration

Two integrals:

- $\mathrm{Beta}(a, b)$ integral, for $a = 3$, $b = 3$,

- $\mathbb{E}(1/(1 + U))$ for $U \sim \mathrm{Exp}(1)$.

Demonstration

Two integrals:

- $\mathrm{Beta}(a, b)$ integral, for $a = 3$, $b = 3$,
- True ψ: Beta normalizing constant $\mathrm{Beta}(3, 3)$.

- $\mathbb{E}(1/(1 + U))$ for $U \sim \mathrm{Exp}(1)$.
- True ψ: $E_1(x) = \int_x^\infty e^{-t}/t dt$, and is given by:

$$\psi = \int_0^\infty \frac{e^{-x}}{1+x} dx = e \cdot \int_1^\infty \frac{e^{-t}}{t} dt = e \cdot E_1(1).$$

Riemann vs. Naïve: Beta

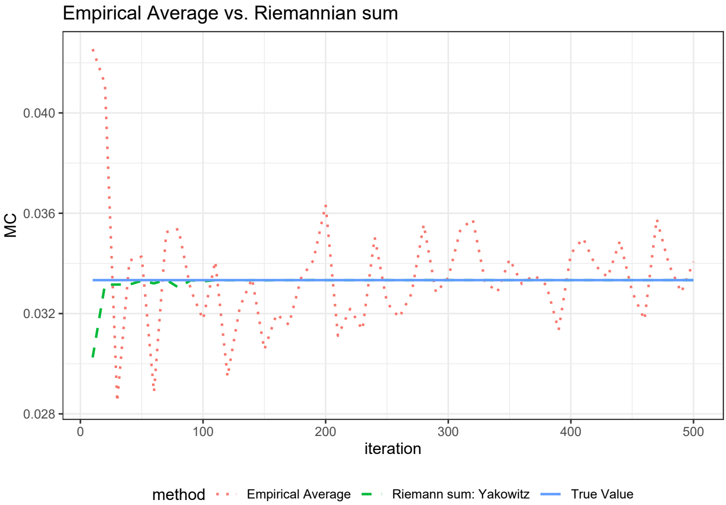


Figure 1: Estimating Beta$(3, 3)$

Riemann vs. Naïve: Exponential

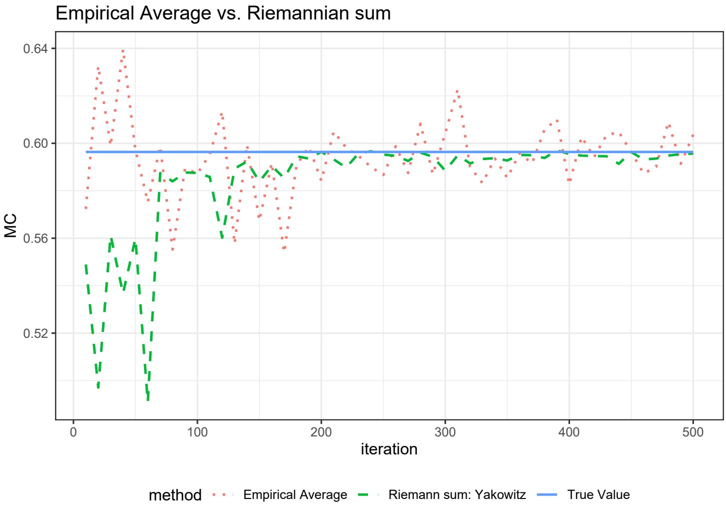


Figure 2: Estimating $\mathbb{E}(1/(1+U))$ for $U \sim \mathrm{Exp}(1)$

Riemann estimators

- Despite the much faster convergence rates, the Yakowitz estimator has limited use.
- Extending trapezoid rules to higher dimension would be thwarted by the ‘curse of dimensionality’.
- An overlooked fact was that high-dimensional integrals can be written as uni-dimensional integrals over the positive real line or over the unit interval (Nested Sampling).
- **Nested sampling** is a perfect situation to invoke the Yakowitz estimator.

Nested Sampling

Lorenz identity

We start with the Lorenz identity that is at the heart of nested sampling and the quantile importance sampling.

Nested Sampling

- Let $\mathbf{X} \sim P(\mathbf{x})$ and $Y \equiv L(\mathbf{X})$, called the *likelihood ordinate*.
- Let $F_Y(y) = \mathbb{P}\{L(\mathbf{X}) \leq y\}$ be the CDF of Y, and the upper cumulant or the survival function for Y is:

$$Z(y) = 1 - F_Y(y) = \int_{L(\mathbf{x})>y} \mathrm{d}P(\mathbf{x}). \tag{4}$$

- Let $\Lambda(s)$ denote the pseudo-inverse of $Z(y)$:

$$\Lambda(s) = \sup\{y : Z(y) > s\} \ , \tag{5}$$

Volume

We can write the evidence ψ (for any p with $\mathbf{x} \in \mathbb{R}^p$) as

$$\psi = \int_0^\infty Z(y)\mathrm{d}y = \int_0^1 \Lambda(s)\mathrm{d}s. \tag{6}$$

The function $Z(y) \in [0,1]$ is also called the *volume variable* in Ashton et al. [2022], as it is the volume enclosed by the likelihood contour.

Lorenz identity

Lorenz identity

Lemma

Since $\mathbf{X} \sim P$, *and* $U \sim \mathcal{U}(0,1)$, *it follows that* $\Lambda(U) \overset{\mathrm{D}}{=} L(\mathbf{X})$, *that is the likelihood ordinates are distributionally same as the* $\Lambda(u)$ *values at uniform grid points. It also follows from the definition of* $\Lambda \equiv Z^{-1}$, *that* $Z\{L(\mathbf{X})\} \sim \mathcal{U}(0,1)$.

Proof

- To quickly see why (6) holds true, consider the following result:

$$\psi = \int_{\chi} L(x)\mathrm{d}P(x) = \int_{\chi}\int_{0}^{\infty} \mathbb{I}\{y < L(x)\}\mathrm{d}y\mathrm{d}P(x) = \int_{0}^{\infty} Z(y)\mathrm{d}y \tag{7}$$

Proof

- To quickly see why (6) holds true, consider the following result:

$$\psi = \int_{\chi} L(x)\mathrm{d}P(x) = \int_{\chi}\int_{0}^{\infty} \mathbb{I}\{y < L(x)\}\mathrm{d}y\mathrm{d}P(x) = \int_{0}^{\infty} Z(y)\mathrm{d}y \tag{7}$$

- Intuitively, $\Lambda(s)$ is the pseudo-inverse of $Z(y)$, i.e. it gives the value y such that s is the fraction of prior draws with likelihood ordinates larger than y.

Proof

- To quickly see why (6) holds true, consider the following result:

$$\psi = \int_{\chi} L(x) \mathrm{d}P(x) = \int_{\chi} \int_0^{\infty} \mathbb{I}\{y < L(x)\} \mathrm{d}y \mathrm{d}P(x) = \int_0^{\infty} Z(y) \mathrm{d}y \tag{7}$$

- Intuitively, $\Lambda(s)$ is the pseudo-inverse of $Z(y)$, i.e. it gives the value y such that s is the fraction of prior draws with likelihood ordinates larger than y.
- That is, $\{s < Z(y)\}$ if and only if $\{y < \Lambda(s)\}$.

$$\psi = \int_0^{\infty} \int_0^1 \mathbb{I}\{s < Z(y)\} \mathrm{d}s \mathrm{d}y = \int_0^{\infty} \int_0^1 \mathbb{I}\{y < \Lambda(s)\} \mathrm{d}s \mathrm{d}y = \int_0^1 \Lambda(s) \mathrm{d}s \, . \tag{8}$$

Couple of remarks

1. Critically, we do not have to assume that either $F^{-1}(s)$ or $\Lambda(s)$ are available in closed form, as we can find an unbiased estimate of this by simulating the Lorenz curve.
2. It also follows from the Lorenz identity that the implied distribution of the likelihood ordinates are the same as the Lorenz curve $\Lambda(\cdot)$ evaluated at uniform grid points.

$$\boxed{Z(L(\mathbf{X})) \overset{\mathrm{D}}{=} \mathcal{U}(0,1), \text{ and } \Lambda(U) \overset{\mathrm{D}}{=} L(\mathbf{X}).}$$

Algorithm [Ashton et al., 2022]

Basic idea

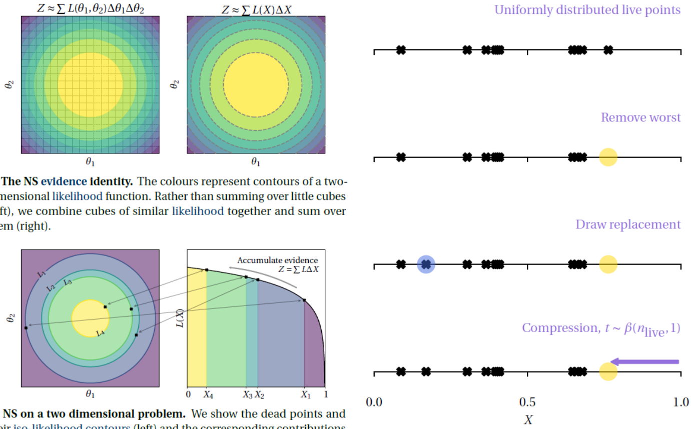
"NS starts with an ensemble of n_{live} random locations Θ drawn from the prior, $\pi(\Theta)$, each of which has its likelihood $L(\Theta)$ which we can place in ascending order. Crudely, if we discarded the lowest half of the values, the survivors would be random samples taken within the restricted volume $L >$ median(L), which would statistically be roughly half the original volume. This allows us to make a statistical estimate of the volume variable $Z(y)$. Repeating that n_{iter} times would yield compression by a factor of about $2^{n_{iter}}$. This is the exponential behaviour required to overcome the curse of dimensionality"'.

Algorithm

Algorithm 1 Nested Sampling

```
**Require:** $n_{live}$, the number of live points; $\mathcal{L}$, the likelihood function; $p(\boldsymbol{\theta})$ prior.
**Ensure:** The evidence $\psi$ and other quantities of interest.
Initialize $\psi = 0$, and prior area/volume $X_0 = 1$.
Generate $n_{live}$ live points $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n \sim p(\boldsymbol{\theta})$.
Calculate the likelihood $\mathcal{L}_i = \mathcal{L}(\boldsymbol{\theta}_i)$, $i = 1, 2, \dots, n_{live}$ for each live points.
**repeat**
    Identify the live point $\boldsymbol{\theta}_{i^*}$ with the lowest likelihood, call it $\mathcal{L}_{i^*}$.
    Remove $\boldsymbol{\theta}_{i^*}$ from the set of live points.
    Sample a new point $\boldsymbol{\theta}_{\mathrm{new}}$ from $p(\boldsymbol{\theta})$ subject to the constraint $\mathcal{L} > \mathcal{L}_{i^*}$.
    Calculate the contraction factor $\Delta X_i$ enclosed by the likelihood contour, i.e.,
    $\Delta X_i = (1 - t) X_i$, where $t = \mathrm{e}^{-1/n_{live}}$ and $X_i = t X_{i-1}$.
    Update the evidence: $Z = Z + \mathcal{L}_{i^*} \Delta X_i$.
    $i \to i + 1$.
**until** stopping criterion is met, e.g. $\max \mathcal{L}(\theta) X_i > \mathrm{e}^{tol} Z$.
Adjust for the remainder of evidence $Z = Z + \frac{1}{n_{live}} \sum_{n=1}^{n_{live}} \mathcal{L}(\boldsymbol{\theta}_n) X_l$.
**return** Evidence $\psi$ and importance weights $p_i = \mathcal{L}_{i^*} \omega_i / Z, \ i = 1, \dots, I$.
```

Illustration [Ashton et al., 2022]


$Z \approx \sum L(\theta_1, \theta_2)\Delta\theta_1\Delta\theta_2$
$Z \approx \sum L(X)\Delta X$
θ_2
θ_1
L_1
L_2
L_3
L_4
Accumulate evidence
$Z = \sum L\Delta X$
$L(X)$
0
X_4
X_3
X_2
X_1
1
Uniformly distributed live points
Remove worst
Draw replacement
Compression, $t \sim \beta(n_{\text{live}}, 1)$
0.0
0.5
1.0
X

a | **The NS evidence identity.** The colours represent contours of a two-dimensional likelihood function. Rather than summing over little cubes (left), we combine cubes of similar likelihood together and sum over them (right).

b | **NS on a two dimensional problem.** We show the dead points and their iso-likelihood contours (left) and the corresponding contributions to the evidence integral (right). The volumes X_i are estimated statistically in NS.

c | **Compression in one iterate of NS.**

Figure 1 | **Illustrations of NS algorithm.**

Vertical Likelihood

Alternative presentation of nested sampling [Polson and Scott, 2014]:

The evidence $\psi = \int_0^1 \Lambda(s)\mathrm{d}s$ is approximated at $n_{\text{iter}} \doteq n$ grid points $1 \equiv s_0 > s_1 > s_2 > \cdots > s_n > 0)$ by a numerical integration rule:

$$\hat{\psi}_{NS} = \frac{1}{2}\sum_{i=1}^{n} w_i \Lambda(s_i) \text{ for } w_i = (s_{i-1} - s_i) \text{ or } w_i = \frac{1}{2}(s_{i-1} - s_{i+1}). \quad (9)$$

In the case of nested sampling, one can use the deterministic grid points $s_i = \exp(-i/n)$ for $i = 1, \ldots, m$ for a moderate n and a large m.

The rationale for the deterministic choice $s_i = \exp(-i/n)$ follows from two facts:

1. the 'volume' variable $Z(y)$ is uniformly distributed (see Lemma 1), and
2. the i^{th} order statistics from $n\,\mathcal{U}(0,1)$ random sample has a $\mathcal{B}eta(i, n-i+1)$ distribution.

Quantile Importance Sampling

Yakowitz + Nested

- Now, recall the Yakowitz estimator i.e. the Riemann sum.
- Let $\{U_{(i)}\}_{i=1}^{n}$ denote n ordered draws from $\mathcal{U}(0,1)$, with augmented fixed boundary points: $U_{(0)} \equiv 0$ and $U_{(n+1)} \equiv 1$.
- We can think of these as stochastic grid points on $[0,1]$, with $s_i = U_{(n+1-i)}$, with $s_{n+1} \equiv U_{(0)}$, and $s_0 = U_{(n+1)}$.
- The Quantile Importance Sampling estimator is:

$$\boxed{\hat{\psi}_{QIS} = \sum_{i=1}^{n} \frac{1}{2}(u_{(i+1)} - u_{(i-1)})\Lambda(u_{(i)})}$$

QIS

To apply Yakowitz's method for nested sampling, the main trick is to **reorder the quantiles of the likelihood ordinates**.

1. Draw m samples from the prior $\mathbf{x}_1, \ldots, \mathbf{x}_m \sim p(\mathbf{x})$,

QIS

To apply Yakowitz's method for nested sampling, the main trick is to **reorder the quantiles of the likelihood ordinates**.

1. Draw m samples from the prior $\mathbf{x}_1, \ldots, \mathbf{x}_m \sim p(\mathbf{x})$,
2. Calculate the likelihood ordinates at the prior values: $Y_1 = L(x_1), \ldots, Y_m = L(x_m)$, and sort the sequence of Y_j's obtained thus, *i.e.* $Y_{(1)}, \ldots, Y_{(m)}$,

QIS

To apply Yakowitz's method for nested sampling, the main trick is to **reorder the quantiles of the likelihood ordinates**.

1. Draw m samples from the prior $\mathbf{x}_1, \ldots, \mathbf{x}_m \sim p(\mathbf{x})$,
2. Calculate the likelihood ordinates at the prior values: $Y_1 = L(x_1), \ldots, Y_m = L(x_m)$, and sort the sequence of Y_j's obtained thus, *i.e.* $Y_{(1)}, \ldots, Y_{(m)}$,
3. Generate $n(< m)$ 'live' points from $\mathcal{U}(0, 1)$, and sort them: $u_{(1)} < u_{(2)} < \ldots < u_{(n)}$. Augment $u_{(i)}$'s with $u_{(0)} = 0$, and $u_{(n+1)} = 1$.

QIS

To apply Yakowitz's method for nested sampling, the main trick is to **reorder the quantiles of the likelihood ordinates**.

1. Draw m samples from the prior $\mathbf{x}_1, \ldots, \mathbf{x}_m \sim p(\mathbf{x})$,
2. Calculate the likelihood ordinates at the prior values: $Y_1 = L(x_1), \ldots, Y_m = L(x_m)$, and sort the sequence of Y_j's obtained thus, *i.e.* $Y_{(1)}, \ldots, Y_{(m)}$,
3. Generate $n(< m)$ 'live' points from $\mathcal{U}(0,1)$, and sort them: $u_{(1)} < u_{(2)} < \ldots < u_{(n)}$. Augment $u_{(i)}$'s with $u_{(0)} = 0$, and $u_{(n+1)} = 1$.
4. Evaluate the sample quantiles for Y at the points $u_{(i)}$:

   $\Lambda(u_{(i)}) \overset{\mathrm{D}}{=} Y_{\lceil m u_{(i)} \rceil}$, $i = 1, \ldots, n$, with $\Lambda(0) = \Lambda_{\max}$ and $\Lambda(1) = \Lambda_{\min}$, and use the formula:

$$\hat{Z}_{QIS} = \left[\sum_{i=1}^{n+1} (u_{(i)} - u_{(i-1)}) \frac{\Lambda(u_{(i-1)}) + \Lambda(u_{(i)})}{2} \right]. \qquad (10)$$

Rate of convergence

QIS can achieve a $O(n^{-4})$ rate of convergence for evidence using the results of [Yakowitz et al., 1978].

Lemma

Let the evidence $\psi \equiv \int_{\chi} L(\mathbf{x}) dF(\mathbf{x}) = \int_0^1 \Lambda(s) ds$*. Assume that the cumulant* $\Lambda(s)$ *has a continuous derivative and that the second derivative* $\Lambda''(s)$ *exists and is bounded in absolute value over the unit interval. Let* $U_{(0)} \equiv 0, U_{(n+1)} \equiv 1$*, and* $\{U_{(i)}\}_{i=1}^{n}$ *be the ordered statistics from* $\mathcal{U}(0,1)$ *distribution (with* $U_{(i)} \geq U_{(i-1)}, i = 1, \ldots, n+1$*). Consider the QIS:*

$$\hat{\psi}_{QIS} = \frac{1}{2} \left(\sum_{i=1}^{n+1} (U_{(i)} - U_{(i-1)}) \{ \Lambda(U_{(i-1)}) + \Lambda(U_{(i)}) \} \right).$$

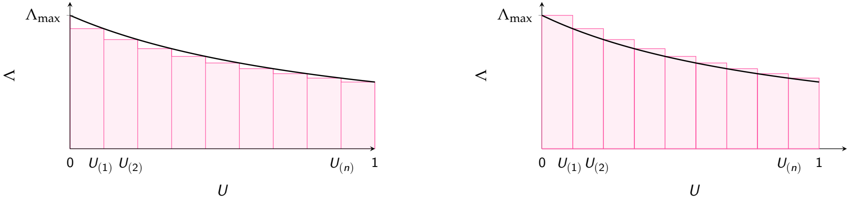
For some constant M*,*

$$\mathbb{E}[(\psi - \hat{\psi}_{QIS})^2] \leq M / n^4, \qquad \textit{for all } n \geq 1.$$

Remarks

$\hat{\psi}_{QIS}$ admits natural upper and lower bounds:

$$\sum_{i=1}^{n}(U_{(i)} - U_{(i-1)})\Lambda(U_{(i)}) \leq \hat{\psi}_{QIS} \leq \sum_{i=1}^{n}(U_{(i+1)} - U_{(i)})\Lambda(U_{(i)}) + U_{(1)}\Lambda_{\max}.$$



(a) Lower bound for $\int_0^1 \Lambda(u)du$.

(b) Upper bound for $\int_0^1 \Lambda(u)du$.

Skilling [2006]: $\Lambda_{\max}$ is independent of the nested sampling algorithm, and there could be intervals of negligible size containing very large likelihood values (see Fig. **??**) unless precluded by separate global analysis, and the error could be at most $O(n^{-1})$.

Example 1

As an illustrative example, we take Gaussian likelihood and Gaussian prior, *i.e.*,

$$L(x \mid \mu_1, \sigma_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},$$
$$p(x \mid \mu_2, \sigma_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \mathrm{e}^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

The integral $\psi = \int L(x)p(x)dx$ is available in closed form.

Compare QIS with the traditional Monte Carlo approach, as well as the original nested sampling approach with both Riemann summation and rectangular summation.

We take $n = 20$ live points for for $m = 1000$ samples from the prior, and 100 replicates to calculate the root mean squared errors (RMSE), $\{1/r\sum_{j=1}^{r}(\hat{\psi}_j - \psi)\}^{1/2}$ and mean absolute predictive error (MAPE), $1/r\sum_{j=1}^{r}|(\hat{\psi}_j - \psi)/\psi|$.

QIS vs. others

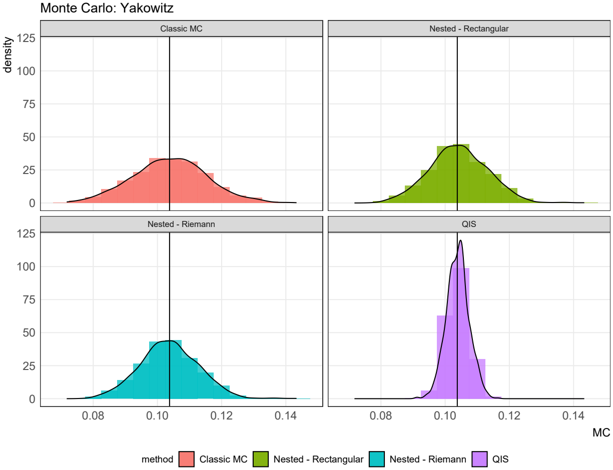


Figure 4: Comparison of three candidate schemes: Yakowitz or QIS (ordered uniform), Skilling's original scheme ($\exp(-i/N)$) and the classic Monte Carlo approach.

Multivariate t

As an example of a higher-dimensional integral, we look at a multivariate t likelihood and a multivariate Gaussian prior with dimension $d = 50$, following Polson and Scott [2014].

The target integral is:

$$\psi = \int_{\mathbb{R}^d} (1 + \frac{\mathbf{x}^T\mathbf{x}}{\nu})^{-\frac{\nu+d}{2}} (\frac{\tau}{2\pi})^{d/2} \exp\{-\tau\mathbf{x}^T\mathbf{x}/2\} d\mathbf{x}. \qquad (11)$$

This integral can be written in terms of Kummer's confluent hypergeometric function of the second kind: $\psi = s^a U(a, b, s)$, where $a = (\nu + d)/2$, $b = \nu/2 + 1$, $s = \nu\tau/2$.

For the specific values used here: $d = 50, \tau = 1$ and $\nu = 2$, we can get: $\psi = U(26, 2, 1) = 1.95 \times 10^{-29}$.

QIS vs. others

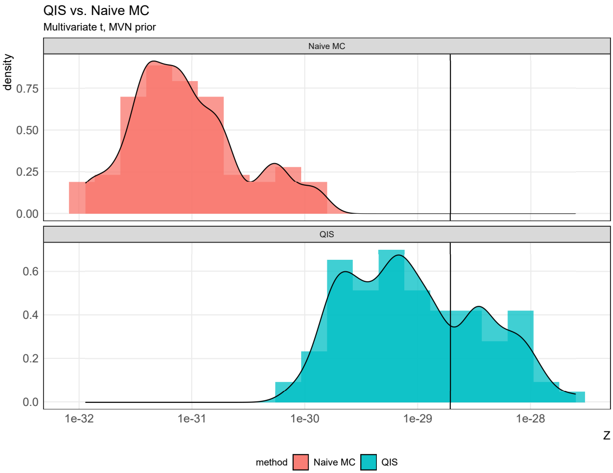


Figure 5: QIS vs. Naïve MC comparison for the multivariate *t*-multivariate Normal example.

Envoi

Take another look at the Riemann sum estimators for Monte Carlo integration in the context of nested sampling, a popular method for handling challenging high-dimensional integrals: aiding in model comparison and inference simultaneously.

We show that under certain conditions, the Riemann sum estimator achieves a much faster convergence rate $O(n^{-4})$ compared to the existing results based on CLT.

The other advantage of using the *quantile reordering trick* or the simulated Lorenz curve is that it obviates the need to know an analytical form for the $\Lambda(s)$ or $Z(L(\mathbf{x}))$ function.

However, there is one caveat to this result: if one does not know maximum of the likelihood ordinates, the last term can be $O(n^{-1})$, unfortunately.

Thank you!

Joint work with

Nick Polson

References

Greg Ashton, Noam Bernstein, Johannes Buchner, Xi Chen, Gábor Csányi, Andrew Fowlie, Farhan Feroz, Matthew Griffiths, Will Handley, Michael Habeck, et al. Nested sampling for physical scientists. *Nature Reviews Methods Primers*, 2(1):39, 2022.

Fernando Llorente, Luca Martino, David Delgado, and Javier Lopez-Santiago. Marginal likelihood computation for model selection and hypothesis testing: an extensive review. *arXiv preprint arXiv:2005.08334*, 2020.

Anne Philippe. Processing simulation output by riemann sums. *Journal of Statistical Computation and Simulation*, 59(4):295–314, 1997.

Anne Philippe and Christian P Robert. Riemann sums for mcmc estimation and convergence monitoring. *Statistics and Computing*, 11 (2):103–115, 2001.

Nicholas G Polson and James G Scott. Vertical-likelihood monte carlo. *arXiv preprint arXiv:1409.3601*, 2014.

John Skilling. Nested sampling for general bayesian computation. *Bayesian analysis*, 1(4):833–859, 2006.

S Yakowitz, JE Krimmel, and F Szidarovszky. Weighted monte carlo integration. *SIAM Journal on Numerical Analysis*, 15(6):1289–1300, 1978.