

Confidence Intervals for Randomized Quasi-Monte Carlo Estimators

Pierre L'Ecuyer, Marvin K. Nakayama, Art B. Owen, Bruno Tuffin

EURO 2024, Copenhagen

Université 
de Montréal

 NJIT
New Jersey Institute
of Technology


STANFORD

 Inria

Monte Carlo (MC) integration

A simulation model produces a random output X . We want to estimate

$$\mu = \mathbb{E}[X] = \mathbb{E}[f(\mathbf{U})] = \int_{(0,1)^s} f(\mathbf{u}) d\mathbf{u} = \int_0^1 \cdots \int_0^1 f(u_1, \dots, u_s) du_1 \cdots du_s$$

where $f : (0, 1)^s \rightarrow \mathbb{R}$ and $\mathbf{U} = (U_1, \dots, U_s)$ is a uniform r.v. over $(0, 1)^s$.

(Crude) Monte Carlo method:

1. Generate n independent copies of $\mathbf{U} \sim \mathcal{U}(0, 1)^s$, say $\mathbf{U}_1, \dots, \mathbf{U}_n$;
2. Estimate μ by $\hat{\mu}_n = (1/n) \sum_{i=1}^n f(\mathbf{U}_i)$.

Monte Carlo (MC) integration

A simulation model produces a random output X . We want to estimate

$$\mu = \mathbb{E}[X] = \mathbb{E}[f(\mathbf{U})] = \int_{(0,1)^s} f(\mathbf{u}) d\mathbf{u} = \int_0^1 \cdots \int_0^1 f(u_1, \dots, u_s) du_1 \cdots du_s$$

where $f : (0, 1)^s \rightarrow \mathbb{R}$ and $\mathbf{U} = (U_1, \dots, U_s)$ is a uniform r.v. over $(0, 1)^s$.

(Crude) Monte Carlo method:

1. Generate n independent copies of $\mathbf{U} \sim \mathcal{U}(0, 1)^s$, say $\mathbf{U}_1, \dots, \mathbf{U}_n$;
2. Estimate μ by $\hat{\mu}_n = (1/n) \sum_{i=1}^n f(\mathbf{U}_i)$.

Almost sure convergence as $n \rightarrow \infty$ (strong law of large numbers).

Can use central limit theorem to compute a confidence interval on μ :

$$\mathbb{P} \left[\mu \in \left(\hat{\mu}_n - \frac{c_\alpha S_n}{\sqrt{n}}, \hat{\mu}_n + \frac{c_\alpha S_n}{\sqrt{n}} \right) \right] \approx 1 - \alpha,$$

where S_n^2 is any consistent estimator of $\sigma^2 = \text{Var}[f(\mathbf{U})]$. Converges at slow rate: $\mathcal{O}(n^{-1/2})$.

Quasi-Monte Carlo (QMC)

Replace the independent random points \mathbf{U}_i by a set of **deterministic** points

$P_n = \{\mathbf{u}_0, \dots, \mathbf{u}_{n-1}\}$ that cover $(0, 1)^s$ **more evenly**. Approximate

$$\mu = \int_{(0,1)^s} f(\mathbf{u})d\mathbf{u} \quad \text{by} \quad \bar{\mu}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{u}_i).$$

Quasi-Monte Carlo (QMC)

Replace the independent random points \mathbf{U}_i by a set of **deterministic** points $P_n = \{\mathbf{u}_0, \dots, \mathbf{u}_{n-1}\}$ that cover $(0, 1)^s$ **more evenly**. Approximate

$$\mu = \int_{(0,1)^s} f(\mathbf{u}) d\mathbf{u} \quad \text{by} \quad \bar{\mu}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{u}_i).$$

For various spaces \mathcal{H} of functions f , one has

$$|E_n| := |\bar{\mu}_n - \mu| \leq \mathcal{D}_n(P_n) \cdot \mathcal{V}(f) \quad (\text{worst-case error bound})$$

for all $f \in \mathcal{H}$, where $\mathcal{V}(f)$ is the **variation** of f in \mathcal{H} and $\mathcal{D}(P_n)$ measures the **discrepancy** between the empirical distribution of P_n and the uniform distribution over $(0, 1)^s$.

Both depend on \mathcal{H} . Typically, we know how to construct P_n such that

$\mathcal{D}_n(P_n) = \mathcal{O}(n^{-\alpha}(\log n)^{s-1})$ for some $\alpha > 1/2$ (**low discrepancy** points).

When $\mathcal{V}(f) < \infty$, $|E_n|$ also converges at this rate, which beats the MC rate.

Difficulty: **very hard to estimate the error E_n** in practice.

Randomized quasi-Monte Carlo (RQMC)

We randomize P_n into $\tilde{P}_n = \{\mathbf{U}_1, \dots, \mathbf{U}_n\} \subset (0, 1)^s$ such that

- (1) the uniformity of P_n as a whole is preserved;
- (2) each point \mathbf{U}_i has the uniform distribution over $(0, 1)^s$.

Unbiased RQMC estimator of μ : $\hat{\mu}_{n,\text{rqmc}} = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{U}_i)$.

Variance estimation: Make R independent replicates of the RQMC estimator $\hat{\mu}_{n,\text{rqmc}}$, then estimate μ and $\text{Var}[\hat{\mu}_{n,\text{rqmc}}]$ by their sample mean and sample variance.

Variance of sample average is (approximately) $\mathcal{O}(R^{-1}n^{-2\alpha})$.

For a fixed computing budget Rn , One prefers a **small R** (e.g., $R = 10$) and large n .

How do we compute a confidence interval on μ ?

The CLT applies for $R \rightarrow \infty$, but we want a small R .

For R fixed and $n \rightarrow \infty$, the sample average is often far from normally distributed.

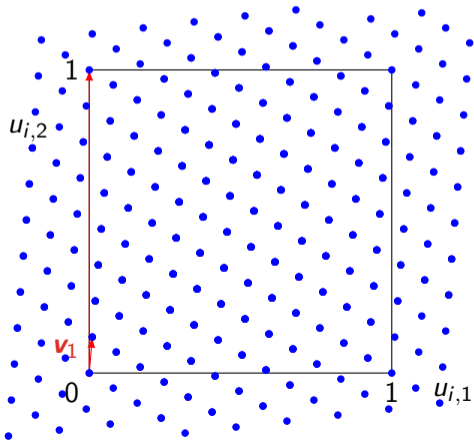
Then using the **bootstrap** appears as an appropriate approach. **We want to test that!**

Example: Lattice rule

Select $\mathbf{a} = (a_1, \dots, a_s) \in \{1, \dots, n-1\}^s$, put $\mathbf{v}_1 = \mathbf{a}/n$ and define

$$P_n = \{\mathbf{u}_i = i\mathbf{v}_1 \bmod 1, i = 0, \dots, n-1\}$$

Example: $s = 2$, $n = 101$, $\mathbf{v}_1 = (1, 12)/n$.

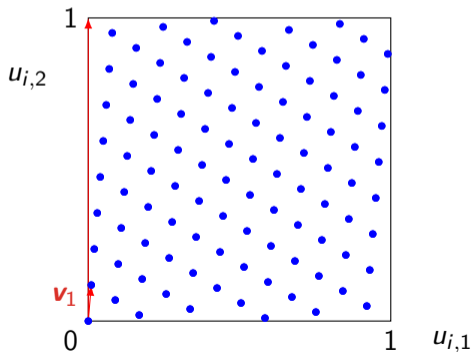


Example: Lattice rule

Select $\mathbf{a} = (a_1, \dots, a_s) \in \{1, \dots, n-1\}^s$, put $\mathbf{v}_1 = \mathbf{a}/n$ and define

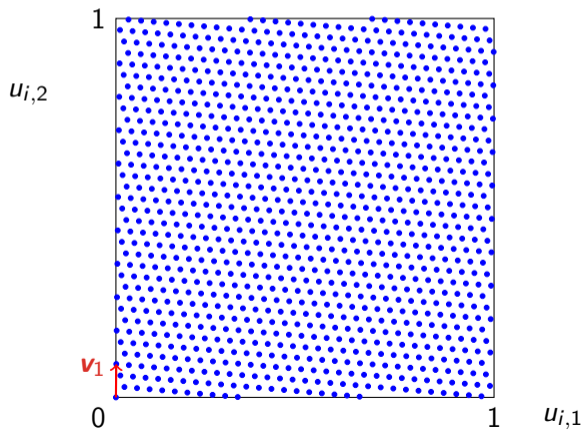
$$P_n = \{\mathbf{u}_i = i\mathbf{v}_1 \bmod 1, i = 0, \dots, n-1\}$$

Example: $s = 2$, $n = 101$, $\mathbf{v}_1 = (1, 12)/n$.



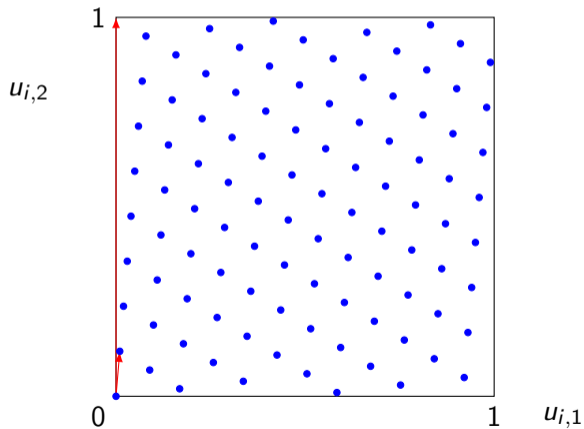
Another example: $s = 2$, $n = 1021$, $\mathbf{v}_1 = (1, 90)/n$.

$$P_n = \{\mathbf{u}_i = i\mathbf{v}_1 \bmod 1 : i = 0, \dots, n-1\}.$$



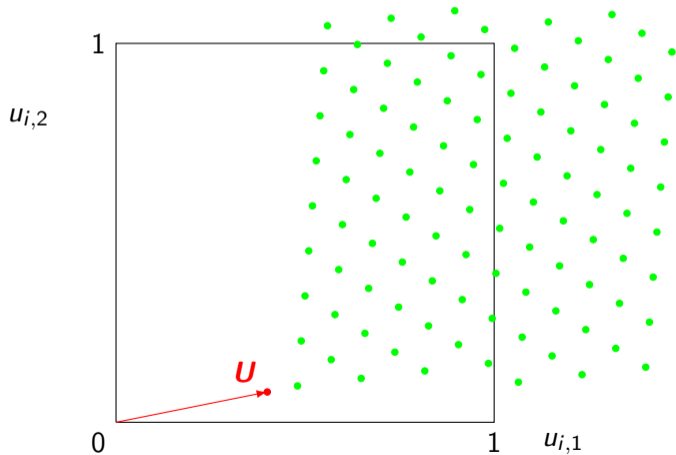
Randomly-Shifted Lattice

Example: lattice with $s = 2$, $n = 101$, $\mathbf{v}_1 = (1, 12)/101$.



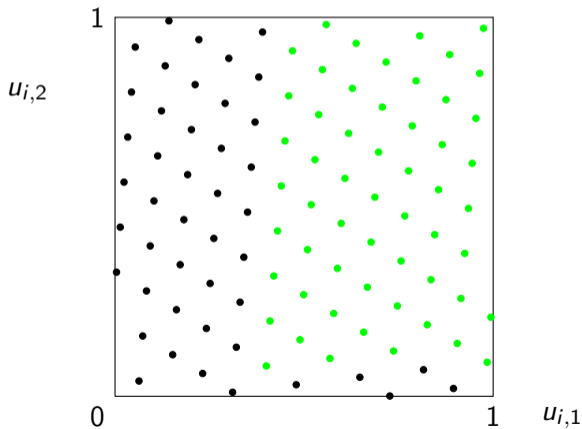
Randomly-Shifted Lattice

Example: lattice with $s = 2$, $n = 101$, $\mathbf{v}_1 = (1, 12)/101$.



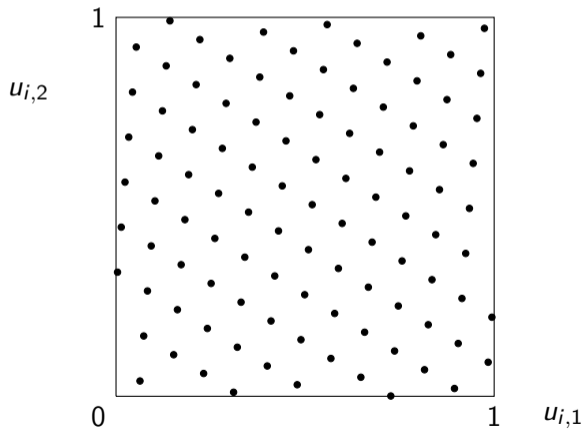
Randomly-Shifted Lattice

Example: lattice with $s = 2$, $n = 101$, $\mathbf{v}_1 = (1, 12)/101$.



Randomly-Shifted Lattice

Example: lattice with $s = 2$, $n = 101$, $\mathbf{v}_1 = (1, 12)/101$.



RQMC Point Sets Considered Here

- ▶ **Lat-RS**: Lattice rule of rank 1 with random shift (RS).
 - ▶ The i th point is $\mathbf{u}_i = (i\mathbf{a} \bmod n)/n$ with $\mathbf{a} = (a_1, \dots, a_s)$ for selected $a_j \in \{1, \dots, n-1\}$;
 - ▶ RS: Uniform random point $\mathbf{U} \in (0, 1)^s$ and add it to each point \mathbf{u}_i .
- ▶ **Lat-RSB**: Lattice rule + RS + baker
 - ▶ Add a baker's transformation after RS: replace each $u_{i,j}$ by $1 - 2|u_{i,j} - 1/2|$;
 - ▶ Brings continuity over the s -dimensional unit torus: may improve the convergence rate.
- ▶ **Sob-DS**: Digital net in base 2 with digital shift (DS)
 - ▶ Coord. j of point $i = \sum_{\ell=0}^{k-1} a_{i,\ell} 2^\ell$ is $u_{i,j} = \sum_{\ell=1}^w u_{i,j,\ell} 2^{-\ell}$ where $(u_{i,j,1}, \dots, u_{i,j,w})^t = \mathbf{C}_j \cdot (a_{i,0}, \dots, a_{i,k-1})^t \bmod 2$ and the \mathbf{C}_j are binary **generating matrices**.
 - ▶ DS: Make bitwise xor of random vector $\mathbf{U} : U(0, 1)^s$ with each point \mathbf{u}_i .
- ▶ **Sob-LMS**: Digital net in base 2 with left matrix scramble (LMS) + DS
 - ▶ Pre-multiply each \mathbf{C}_j by a random lower triangular $w \times w$ matrix \mathbf{L}_j with 1's on the diagonal and i.i.d. binary entries below the diagonal. Also add DS.
- ▶ **Sob-NUS**: Digital net in base 2 with nested uniform scramble (NUS)
 - ▶ Performs more involved nested uniform permutations of the digits (bits).
 - ▶ Obeys a CLT when $n \rightarrow \infty$, under some conditions.

Three types of confidence intervals (CI)

Standard Student-t interval. Assume that the RQMC estimator $\hat{\mu}_{n,\text{rqmc}}$ has a normal distribution, use the empirical mean \bar{y}_R and variance S_R^2 of its R replicates y_1, \dots, y_R to compute a CI by assuming that $\sqrt{R}(\bar{y}_R - \mu)/S_R \sim \text{Student}$ with $R - 1$ degrees of freedom.

Three types of confidence intervals (CI)

Standard Student-t interval. Assume that the RQMC estimator $\hat{\mu}_{n,\text{rqmc}}$ has a normal distribution, use the empirical mean \bar{y}_R and variance S_R^2 of its R replicates y_1, \dots, y_R to compute a CI by assuming that $\sqrt{R}(\bar{y}_R - \mu)/S_R \sim \text{Student}$ with $R - 1$ degrees of freedom.

Percentile (basic) bootstrap. If we knew the exact distribution of $\bar{y}_R - \mu$, we could compute an exact 95% CI just by using the 2.5% and 97.5% quantiles of that distribution. We don't, but we can approximate it by resampling B new approximate realizations of \bar{y}_R , each one by resampling R observations with replacement from the data y_1, \dots, y_R . If $\bar{y}^{*(1)} \leq \dots \leq \bar{y}^{*(B)}$ are the sorted values and $B = 1000$, the 95% CI has the form $(\bar{y}^{*(25)}, \bar{y}^{*(975)})$.

Three types of confidence intervals (CI)

Standard Student-t interval. Assume that the RQMC estimator $\hat{\mu}_{n,\text{rqmc}}$ has a normal distribution, use the empirical mean \bar{y}_R and variance S_R^2 of its R replicates y_1, \dots, y_R to compute a CI by assuming that $\sqrt{R}(\bar{y}_R - \mu)/S_R \sim \text{Student}$ with $R - 1$ degrees of freedom.

Percentile (basic) bootstrap. If we knew the exact distribution of $\bar{y}_R - \mu$, we could compute an exact 95% CI just by using the 2.5% and 97.5% quantiles of that distribution. We don't, but we can approximate it by resampling B new approximate realizations of \bar{y}_R , each one by resampling R observations with replacement from the data y_1, \dots, y_R . If $\bar{y}^{*(1)} \leq \dots \leq \bar{y}^{*(B)}$ are the sorted values and $B = 1000$, the 95% CI has the form $(\bar{y}^{*(25)}, \bar{y}^{*(975)})$.

Bootstrap t . Similar, but we approximate the distribution of $\sqrt{R}(\bar{y}_R - \mu)/S_R$ instead, by resampling B times R observations, computing the B realizations of $\sqrt{R}(\bar{y}_R^* - \bar{y}_R)/S_R^*$, and the appropriate quantiles t_L^* and t_U^* of their distribution. The CI is

$$\left(\bar{y}_R - S_R t_U^* / \sqrt{R}, \bar{y}_R - S_R t_L^* / \sqrt{R} \right).$$

Bootstrap properties (Hall 1988)

With skewness γ and kurtosis coefficient κ , the coverage error is:

$$\text{Normal theory: } (1/R)\varphi(z^{1-\alpha/2}) [0.14\kappa - 2.12\gamma^2 - 3.35] + \mathcal{O}(1/R^2),$$

$$\text{Percentile: } (1/R)\varphi(z^{1-\alpha/2}) [-0.72\kappa - 0.37\gamma^2 - 3.35] + \mathcal{O}(1/R^2),$$

$$\text{Bootstrap } t: (1/R)\varphi(z^{1-\alpha/2}) [-2.84\kappa + 4.25\gamma^2] + \mathcal{O}(1/R^2).$$

The bootstrap t has the advantage of missing the -3.35 constant. It has a large positive coefficient for γ^2 (extra coverage for skewed data) where the others have negative coefficients. The asymptotics predict that it will undercover when κ is large and $\gamma = 0$.

For R different values y_r , $\mathbb{P}(S_R = 0) = R^{1-R}$.

For $R = 5$, this gives $\mathbb{P}(S_R = 0) = 1/5^4 = 1/625$, which is not negligible.

Selected functions and set of experiments

- ▶ Five types of RQMC point sets: Lat-RS, Lat-RSB, Sob-DS, Sob-LMS, Sob-NUS.
- ▶ Each with $n = 2^k$ points for $k = 6, 8, 10, 12, 14$, and in $s = 4, 8, 16, 32$ dimensions.
- ▶ Bootstrap with $B = 1000$.
- ▶ Selected functions:
 1. **SumUeU** (smooth, additive): $f(\mathbf{u}) = -s + \sum_{j=1}^s u_j \exp(u_j)$.
 2. **MC2** (smooth): $f(\mathbf{u}) = -1 + (s - 1/2)^s \prod_{j=1}^s (x_j - 1/2)$.
 3. **PieceLinGauss** (piecewise linear, continuous, Gaussian inputs):

$$f(\mathbf{u}) = \max \left(s^{-1/2} \sum_{j=1}^s \Phi^{-1}(u_j) - \tau, 0 \right) - \varphi(\tau) + \tau \Phi(-\tau).$$
 4. **IndSumNormal** (discontinuous, infinite variation):

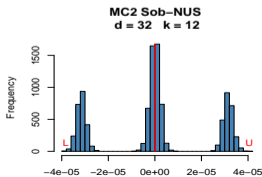
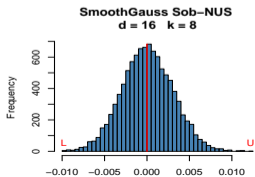
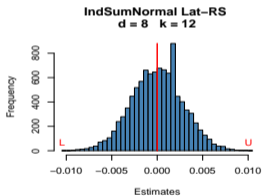
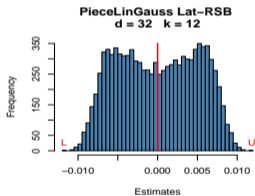
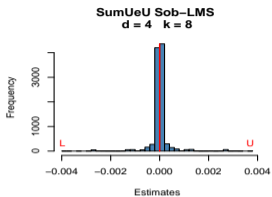
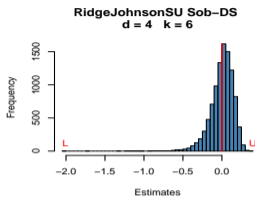
$$f(\mathbf{u}) = -\Phi(1) + \mathbb{I}\{s^{-1/2} \sum_{j=1}^s \Phi^{-1}(u_j) \geq 1\},$$
 5. **SmoothGaus** (smooth, bounded, monotone):

$$f(\mathbf{u}) = -\Phi(1/\sqrt{2}) + \Phi(1 + s^{-1/2} \sum_{j=1}^s \Phi^{-1}(u_j)).$$
 6. **RidgeJohnsonSU** (heavy-tailed): $f(\mathbf{u}) = -\eta + F^{-1}(s^{-1/2} \sum_{j=1}^s u_j)$ where F is the CDF of the Johnson's SU distribution with skewness -5.66 and kurtosis 96.8 for any s .

Results

- ▶ Experiments
 - ▶ **2400 tasks**: 6 integrands, 5 RQMC methods, 4 dimensions, 5 RQMC sample sizes and 4 values of R (5, 10, 20, 30).
 - ▶ For each case, we computed 1000 independent CIs at 95%, and declared a “failure” if there was less than 92.7% coverage (less than 927/1000 of the intervals covered the true value). If the true coverage is at least 94%, this happens with probability less than 4%.
- ▶ Results
 - ▶ **The percentile method failed for 1698 (70.75%) of those 2400 tasks**
 - ▶ Not well suited to very small sample sizes
 - ▶ Not well regarded for setting confidence intervals for the mean.
 - ▶ **The bootstrap t method failed 81 times**
 - ▶ 74 for Sob-LMS on SumUeU (44 times) or MC2 (30 times); spiky histograms, see next slide
 - ▶ Interval of infinite length if $S_R = 0$: 21 times for IndSumNormal with $R = 5$. Discrete distribution, fewer than 2^k different values.
 - ▶ **The plain Student t confidence interval method failed only 3 times.**
 - ▶ Fails only when $R = 5$ (bootstrap t has coverage higher than 95% then)
 - ▶ Coverage higher than 97% 81 times (SumUeU and MC2).

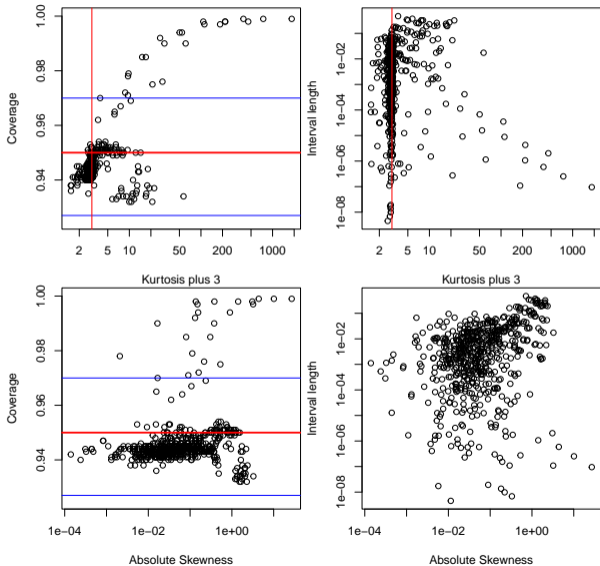
Histograms (mostly unusual ones)



- ▶ RidgeJohnsonSU: negatively skewed (other RQMC methods too)
- ▶ SumUeU (and MC2): “spike plus outliers”
- ▶ PieceLinGauss: bimodal (often for LAT+baker)
- ▶ IndSumNormal: Gaussian plus a spike near one value
- ▶ SmoothGauss: roughly Gaussian, as most of those in the data set
- ▶ MC2 Sob-NUS: untypical for NUS (more frequent for LMS).

CI Coverage and width, versus skewness and kurtosis, Student t intervals, $R = 10$

Coverage and length: standard t intervals and $R = 10$



- ▶ Some examples with high kurtosis, none with extreme skewness
- ▶ Standard CI known to have robust coverage in response to kurtosis but vulnerable to skewness.
- ▶ Kurtosis increases nominal coverage for the standard t intervals
- ▶ CI length decreasing with extreme kurtosis (Sob-LMS with SumUeU and MC2)
- ▶ Small R : rare outliers, CIs are extremely short and cover the true mean often enough.

Conclusions

- ▶ **Surprise:** Two-sided CIs using **Student-t distribution** work well for RQMC.
- ▶ We were expecting the bootstrap t to perform better, but it does not.
- ▶ Bootstrap t not working well for discrete random variables (SumNormalInd).
- ▶ The bootstrap t is known to have better coverage for one-sided intervals, not considered here. It is $\mathcal{O}(1/\sqrt{n})$ vs $\mathcal{O}(1/n)$ for other methods.

One last recommendation: Always look at the R individual averages that you obtained. They contain all the information that can be used by any method!

Thank you!

- ▶ P. L'Ecuyer, M. Nakayama, A. B. Owen, and B. Tuffin. Confidence Intervals for Randomized Quasi-Monte Carlo Estimators, *Proceedings of the 2023 Winter Simulation Conference*, IEEE Press, pages 445–456, 2023.
- ▶ *Monte Carlo and Quasi-Monte Carlo Methods 2022, 2020, 2018, ...* Springer-Verlag.
- ▶ J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, 2010.
- ▶ P. L'Ecuyer. Quasi-Monte Carlo methods with applications in finance. *Finance and Stochastics*, 13(3):307–349, 2009.
- ▶ P. L'Ecuyer. Randomized quasi-Monte Carlo: An introduction for practitioners. In P. W. Glynn and A. B. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2016*, 2017.
- ▶ P. L'Ecuyer, P. Marion, M. Godin, and F. Puchhammer. A Tool for Custom Construction of QMC and RQMC Point Sets. *Monte Carlo and Quasi-Monte Carlo Methods 2020*, A. Keller, Ed., Springer-Verlag, 51–70, 2022.
- ▶ A. B. Owen. *Practical Quasi-Monte Carlo*. Draft available at <https://artowen.su.domains/mc/practicalqmc.pdf>, 2023.
- ▶ H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1992.