Error Estimate and Convergence Analysis of Moment-Preserving Discrete Approximations of Continuous Distributions

Ken'ichiro Tanaka[1] Alexis Akira Toda[2]

[1]School of Systems Information Science, Future University Hakodate

[2]Department of Economics, UC San Diego

Info-Metrics 2014

Question

- Finding discrete approximations of probability distributions and stochastic processes is practically important:
  - Solving dynamic economic models
    (e.g., optimal portfolio, consumption/saving, general equilibrium, etc.)
  - Estimating dynamic economic models
  - Option pricing by Monte Carlo simulations
- This paper: simple framework for discretizing distributions and stochastic processes.

Issues with existing methods

Existing methods use integration formula

$$\mathrm{E}[g(X)] = \int g(x)f(x)\mathrm{d}x \approx \sum_{i=1}^{M} w_{i,M}\, g(x_{i,M}) f(x_{i,M}).$$

- Newton-Cotes type or Gauss type quadrature (Miller & Rice 1983) works for only 1 dimension and the discrete set $D_M = \{x_{i,M}\}$ is constrained by the quadrature method.
- If D_M is arbitrarily given, approximation may not be accurate because moments of f not exact (Tauchen 1986, Adda & Cooper 2003).
- Multidimensional case with exact moments is computationally intensive (DeVuyst & Preckel 2007).

Problem

- Given
  1. continuous density $f : \mathbb{R}^K \to \mathbb{R}$,
  2. generalized moments $\bar{T} = \int T(x)f(x)\mathrm{d}x$,
     ($T : \mathbb{R}^K \to \mathbb{R}^L$ is moment defining function)
  3. arbitrary discrete points $D_M = \{x_{i,M}\}_{i=1}^M$,

  can we approximate f by a discrete distribution

$$P_M = \{p(x_{i,M})\}_{i=1}^M$$

  with exact moments $\bar{T}$?

- This paper: propose a numerical algorithm that is accurate and computationally very simple.

Solution

1. Start from some integration formula (e.g., trapezoidal)

$$\mathrm{E}[g(X)] = \int g(x)f(x)\mathrm{d}x \approx \sum_{i=1}^{M} w_{i,M} g(x_{i,M}) f(x_{i,M}).$$

2. Approximate f by solving the minimum Kullback-Leibler information problem

$$\min_{\{p(x_{i,M})\}} \sum_{i=1}^{M} p(x_{i,M}) \log \frac{p(x_{i,M})}{w(x_{i,M}) f(x_{i,M})}$$

$$\text{s.t. } \sum_{i=1}^{M} p(x_{i,M}) T(x_{i,M}) = \bar{T}, \quad \sum_{i=1} p(x_{i,M}) = 1, \quad p(x_{i,M}) \geq 0.$$

Dual problem

- Using Fenchel duality, the dual of the minimum K-L information problem is

$$\min_{\lambda \in \mathbb{R}^L} \left[-\left\langle \lambda, \bar{T} \right\rangle + \log \left(\sum_{i=1}^{M} w_{i,M} \, f(x_{i,M}) \, \mathrm{e}^{\left\langle \lambda, T(x_{i,M}) \right\rangle} \right) \right].$$

- Primal problem is constrained optimization with a large number of unknowns (M).
- Dual problem is unconstrained optimization with a small number of unknowns (L).

Tanaka & Toda (2013)

Theorem

Suppose that $\bar{T} \in \operatorname{conv} T(D_M)$*. Then the solution of the primal problem is given by*

$$p(x_{i,M}) = \frac{w_{i,M} f(x_{i,M}) \mathrm{e}^{\langle \bar{\lambda}_M, T(x_{i,M}) \rangle}}{\sum_{i=1}^{M} w_{i,M} f(x_{i,M}) \mathrm{e}^{\langle \bar{\lambda}_M, T(x_{i,M}) \rangle}},$$

where $\bar{\lambda}_M$ *is a solution of the dual problem.*

Theorem

Suppose that $\bar{T} \in \operatorname{int}(\operatorname{conv} T(D_M))$*. Then*

1. *the dual function is continuous and strictly convex, and*
2. *the solution* $\bar{\lambda}_M$ *uniquely exists.*

Assumptions

Assume that

1. $\bar{T} = 0 \in \operatorname{int}(\operatorname{conv} T(D_M))$,
   (Since $\int T(x)f(x)\mathrm{d}x = \bar{T} \iff \int(T(x) - \bar{T})f(x)\mathrm{d}x = 0$, we can set $\bar{T} = 0$ w.l.o.g.)
2. integration formula converges, so for any bounded g
$$\lim_{M\to\infty}\sum_{i=1}^{M} w_{i,M} f(x_{i,M}) g(x_{i,M}) = \int_{\mathbb{R}^K} f(x)g(x)\mathrm{d}x,$$
3. integration formula converges for $\|T\|$ as well:
$$\lim_{M\to\infty}\sum_{i=1}^{M} w_{i,M} f(x_{i,M}) \|T(x_{i,M})\| = \int_{\mathbb{R}^K} f(x)\|T(x)\| \mathrm{d}x =: I_{\|T\|} < \infty.$$

Notations for errors

- For any measurable g, let
$$E^{(\mathrm{a})}_{g,M} = \left|\sum_{i=1}^{M} w_{i,M} f(x_{i,M}) g(x_{i,M}) - \int_{\mathbb{R}^K} f(x) g(x) \mathrm{d}x\right|$$
be the error of the initial integration formula, and
$$E_{g,M} = \left|\sum_{i=1}^{M} p(x_{i,M}) g(x_{i,M}) - \int_{\mathbb{R}^K} f(x) g(x) \mathrm{d}x\right|$$
be the error of our approximation method.
- Previous assumptions imply $E^{(\mathrm{a})}_{g,M} \to 0$ for any bounded g and $E^{(\mathrm{a})}_{T,M}, E^{(\mathrm{a})}_{\|T\|,M} \to 0$ as $M \to \infty$.

Main result

Theorem

Let g be measurable with $|g(x)| \leq G$ and $\alpha > 0$ be large enough such that

$$C_\alpha := \inf_{\substack{\lambda \in \mathbb{R}^L \\ \|\lambda\|=1}} \frac{1}{2} \int_{\mathbb{R}^K} f(x) \left(\max\left\{0, \min\left\{\langle \lambda,\ T(x) \rangle, \alpha\right\}\right\}\right)^2 \mathrm{d}x > 0.$$

Then, for any ε with $0 < \varepsilon < C_\alpha$, there exists a positive integer M_ε such that for any M with $M \geq M_\varepsilon$, we have

$$E_{g,M} \leq E_{g,M}^{(\mathrm{a})} + G \left(E_{\mathbf{1},M}^{(\mathrm{a})} + 6 \frac{I_{\|T\|} + E_{\|T\|,M}^{(\mathrm{a})}}{C_\alpha - \varepsilon} E_{T,M}^{(\mathrm{a})} \right).$$

Hence $E_{g,M} = \mathrm{O}\left(\max\left\{E_{g,M}^{(\mathrm{a})}, E_{\mathbf{1},M}^{(\mathrm{a})}, E_{T,M}^{(\mathrm{a})}\right\}\right) \to 0$ as $M \to \infty$.

Numerical experiment 1

- Compute $\mathrm{E}[g(X)]$ for $g(x) = \mathrm{e}^x$ and $X \sim N(0,1)$ or $X \sim Be(2,4)$.
- For $X \sim N(0,1)$, start from trapezoidal formula with

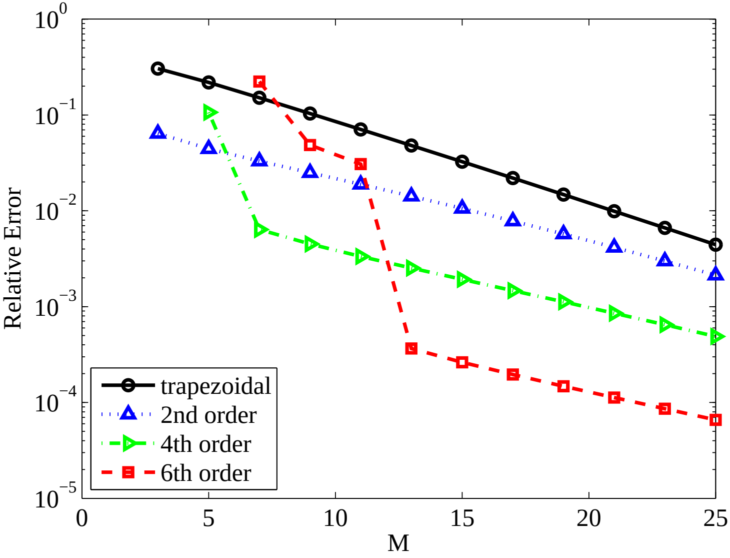
$$D_M = \{nh_M | n = 0, \pm 1, \pm 2, \dots, \pm N\},$$

  where $M = 2N+1$ and $h_M = 1/\sqrt{N}$.
- For $X \sim Be(2,4)$, start from trapezoidal formula with

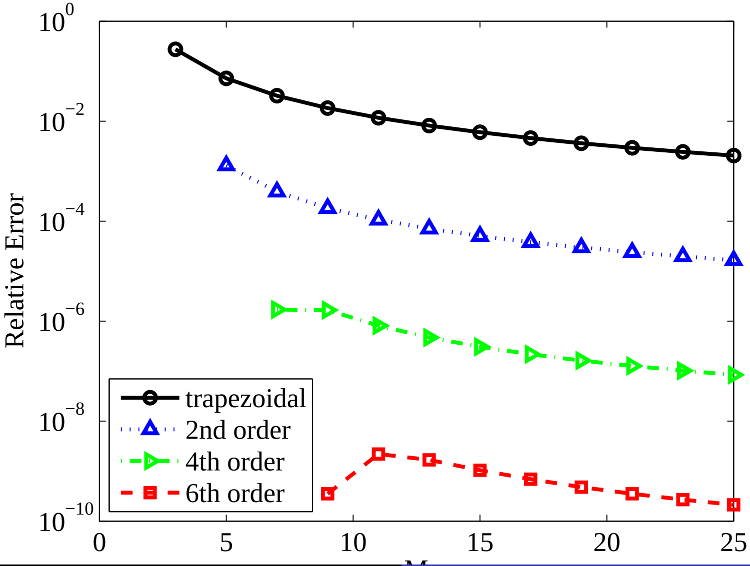
$$D_M = \{nh_M | n = 0, 1, 2, \dots, 2N\},$$

  where $M = 2N+1$ and $h_M = 1/2N$.
- In each case, $N = 1, \dots, 12$ and match up to L polynomial moments ($L = 2, 4, 6$).

Relative errors for the Gaussian case ($X \sim N(0,1)$)



Relative errors for the beta case ($X \sim Be(2,4)$)



Numerical experiment 2

- Consider VAR(1) process

$$y_t = Ay_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \Psi),$$

where $y_t = (z_t, g_t)$ and

$$A = \begin{bmatrix} 0.9809 & 0.0028 \\ 0.0410 & 0.9648 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 0.0087^2 & 0 \\ 0 & 0.0262^2 \end{bmatrix}.$$

(Example of Gospodinov & Lkhagvasuren (2014))

- Unconditional variance is

$$\Sigma = \begin{bmatrix} 0.0024 & 0.0024 \\ 0.0024 & 0.0127 \end{bmatrix}.$$

- Match 1st and 2nd conditional moments using algorithm. Generate 1000 samples of length 2,000,000, discard first 200,000 as burn in.

	$N=5$	$N=9$			
	TT	**Tau.**	**GL0**	**GL**	**TT**
Root mean squared error					
$\hat{\sigma}_z^2$	0.894	433	99	9	0.014
$\hat{\sigma}_g^2$	9.011	363	139	10	0.071
$\hat{\rho}_{zg}$	19.814	40	23	11	3.777
Bias					
$\hat{\sigma}_z^2$	0.893	433	99	-5	-0.001
$\hat{\sigma}_g^2$	9.010	362	138	-7	-0.004
$\hat{\rho}_{zg}$	-19.473	-38	-22	-6	-0.076
Standard deviation					
$\hat{\sigma}_z^2$	0.023	12	8	8	0.014
$\hat{\sigma}_g^2$	0.116	8	7	6	0.071
$\hat{\rho}_{zg}$	3.663	10	9	9	3.778

Table: All results scaled by 10^{-3}. Tau.: Tauchen (1986), GL0, GL: Gospodinov & Lkhagvasuren (2014) without & with moment targeting, TT: Tanaka & Toda.

Concluding remarks

Our method

- Starting from an initial integration formula, we "fine-tuned" probabilities by minimizing the K-L information subject to prescribed moment constraints.
- Although the primal problem is a constrained minimization problem with many unknowns, the dual is unconstrained with a small number of unknowns, hence computationally simple.

Results

- Provided an error bound and proved the weak convergence to the true density.
- Our method is much more accurate than existing methods.

Pathological case for $N(0,1)$

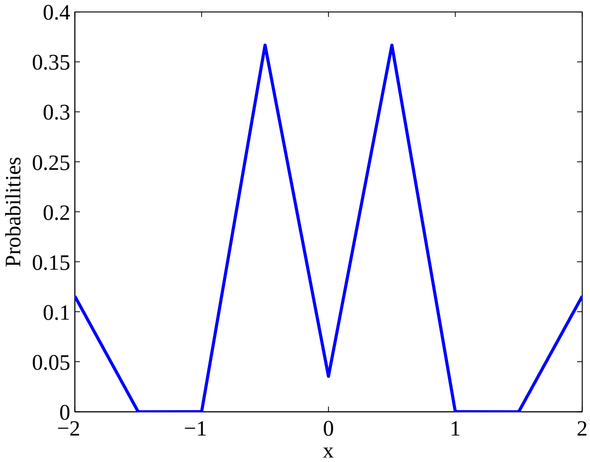


Figure: 6th order discrete approximation of $N(0,1)$ with $M = 9$ grid points.

Pathological case for $Be(2,4)$

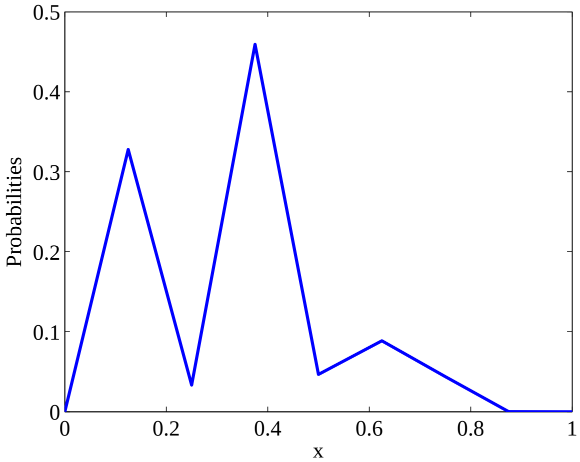


Figure: 6th order discrete approximation of $Be(2,4)$ with $M = 9$ grid points.