Generalizing the Unscented Ensemble Transform to Higher Moments

Deanna Easley (Advisor Tyrus Berry)
Department of Mathematics
George Mason University

Our goal is to estimate the expectation of the expected values of some functionals on multivariate non-Gaussian distributions,

$$\mathbb{E}[f(X)] = \int f(x)p(x)\,dx.$$

Our idea is to do this by generating a small number of quadrature points for the weighted integral, i.e. $\mathbb{E}[f(X)] \approx \sum_{i=1}^{N} w_i f(x_i)$ where x_i are nodes and w_i are weights for $i = 1, \ldots, N$. Note that the largest value of m so that all polynomials of degree m and below are integrated exactly is called degree-of-exactness.

Background

Scaled Unscented Transform

Julier's Idea: Suppose we choose the right nodes so that our quadrature has degree of exactness 2, i.e. matches the first two moments exactly.

The σ–points of the Scaled Unscented Transform (SUT)

Suppose we are given the first two moments, the mean $\mu \in \mathbb{R}^d$ and the covariance $C \in \mathbb{R}^{d\times d}$. Then for some $\beta \in \mathbb{R}$ the **σ–points** are defined by

$$\sigma_0 = \mu \quad \text{and} \quad \sigma_i = \begin{cases} \mu + \beta\sqrt{C}_i & \text{if } i = 1, \ldots, d \\ \mu - \beta\sqrt{C}_{i-d} & \text{if } i = d+1, \ldots, 2d \end{cases}$$

Note: The choice of β can have significant impact on the effectiveness of the transform and $\sum_{i=1}^{d} \sqrt{C}_i \sqrt{C}_i^\top = C$

Empirical mean and Empirical covariance

For $\beta = \sqrt{d}$, we have this derivation

$$\mathbb{E}[X] = \frac{1}{2d}\sum_{i=1}^{2d}\sigma_i = \frac{1}{2d}\sum_{i=1}^{d} 2\mu = \frac{1}{2d}(2d\mu) = \mu$$

$$\mathbb{E}[(X-\mu)(X-\mu)^\top] = \frac{1}{2d}\sum_{i=1}^{2d}(\sigma_i - \mu)(\sigma_i-\mu)^\top = \frac{1}{2d}(2dC\mu) = C$$

Our goal is to generalize the unscented transform to higher moments.

Tensors

Tensors are basically multidimensional matrices. For example the skewness is

$$S = \int (x-\mu)^{\otimes 3} p(x)\,dx, \qquad S_{ijk} = \int (x-\mu)_i (x-\mu)_j (x-\mu)_k p(x)\,dx$$

Similarly, the kurtosis is defined as $K = \int (x-\mu)^{\otimes 4} p(x)\,dx$ where

$$(x-\mu)^{\otimes n} = \underbrace{(x-\mu)\otimes(x-\mu)\otimes\cdots\otimes(x-\mu)}_{n \text{ times}}$$

Note that moments are symmetric, i.e. $M_{i_1\cdots i_n} = M_{p(i_1\cdots i_n)}$ for any permutation p. Motivated by the unscented transform we first need to find the generalized rank 1 decomposition,

$$S = \sum_i x_i^{\otimes 3} \qquad K = \sum_i x_i^{\otimes 4}$$

It's been shown that finding the best rank 1 decomposition is NP-complete.

Approximate Rank 1 Decomposition

A generalization of Normalized Power Iteration (NPI) known as the Symmetric Higher-Order Power Method (S-HOPM) [2] is used to find the eigenvector u associated to the largest eigenvalue λ in absolute value of the tensor T. A theorem of [2] states that $\lambda u^{\otimes k}$ is the best rank 1 approximation of T, namely

$$\|T - \lambda u^{\otimes k}\|$$

is minimized over all possible λ, $\|u\| = 1$. It was suggested in [2] that repeatedly subtracting the rank 1 approximations may result in an approximate rank 1 decomposition. The following theorem shows that this process converges.

Theorem

Let T be a k-order symmetric tensor with size n, i.e. $T \in \mathbb{R}^{n^k}$. Consider the process of finding an approximate rank-1 decomposition of T by starting from $T_0 = T$ and setting $T_{\ell+1} = T_\ell - \lambda_\ell v_\ell^{\otimes k}$ where λ_ℓ is the largest eigenvalue of T_ℓ and v_ℓ is the associated eigenvector. Then $\|T_\ell\|_F \to 0$ and there exists a constant $c \in (0,1]$ such that $\lambda_0 \geq c|T_{i_1\ldots i_k}|$ and for $r = \sqrt{1 - \frac{c^2}{n^k}} \in (0,1)$

$$\frac{\|T_{\ell+1}\|_F}{\|T_\ell\|_F} \leq r \quad \text{and} \quad T = \sum_{\ell=1}^{L}\lambda_\ell v_\ell^{\otimes k} + \mathcal{O}(r^L)$$

for all $L \in \mathbb{N}$.

Higher Order Unscented Ensemble

Suppose we are given the first 4 moments: $\mu \in \mathbb{R}^d$, $C \in \mathbb{R}^{d^2}$, $S \in \mathbb{R}^{d^3}$, and $K \in \mathbb{R}^{d^4}$ such that $S = \sum_{i=1}^{J}\tilde{v}_i^{\otimes 3}$ and $K = \sum_{i=1}^{L} s_i \tilde{u}_i^{\otimes 4}$. Now denote $\tilde{\mu} = \sum_{i=1}^{J}\tilde{v}_i$, $\tilde{C} = \sum_{i=1}^{L} s_i \tilde{u}_i^{\otimes 2}$, $\hat{\mu} = \left(1 - d\beta^{-2} - \hat{L}\delta^{-4}\right)\mu - \gamma^{-2}\tilde{\mu}$, where $\hat{L} = \sum_{i=1}^{L} s_i$ and $\hat{C} = C - \frac{1}{\rho^2}\tilde{C}$ for any $\rho > \sqrt{\frac{\lambda^{\tilde{C}}_{\max}}{\lambda^{C}_{\min}}}$.

The 4 moment σ–points of the Higher Order Unscented Transform (HOUT)

We define the 4 **moment σ-points** by

$$\sigma_i = \begin{cases} \mu & \text{if } i = -2 \\ \mu + \alpha\hat{\mu} & \text{if } i = -1 \\ \mu - \alpha\hat{\mu} & \text{if } i = 0 \\ \mu + \beta\sqrt{\hat{C}}_i & \text{if } i = 1, \ldots, d \\ \mu - \beta\sqrt{\hat{C}}_{i-d} & \text{if } i = d+1, \ldots, 2d \\ \mu + \gamma\tilde{v}_{i-2d} & \text{if } i = 2d+1, \ldots, 2d+J \\ \mu - \gamma\tilde{v}_{i-2d-J} & \text{if } i = 2d+J+1, \ldots, 2d+2J \\ \mu + \delta\tilde{u}_{i-2d-2J} & \text{if } i = 2d+2J+1, \ldots, 2d+2J+L \\ \mu - \delta\tilde{u}_{i-2d-2J-L} & \text{if } i = 2d+2J+L+1, \ldots, N \end{cases}$$

and the corresponding weights by $w_{-2} = 1 - d\beta^{-2} - \hat{L}\delta^{-4}$, $w_{-1} = \frac{1}{2\alpha}$, $w_0 = -\frac{1}{2\alpha}$, $w_1, \ldots, w_{2d} = \frac{1}{2\beta^2}$, $w_{2d+1}, \ldots, w_{2d+J} = \frac{1}{2\gamma^3}$, $w_{2d+J+1}, \ldots, w_{2d+2J} = -\frac{1}{2\gamma^3}$, and

$$w_i = \begin{cases} \frac{1}{2\delta^4}s_{i-2d-2J} & \text{if } i = 2d+2J+1, \ldots, 2d+2J+L \\ \frac{1}{2\delta^4}s_{i-2d-2J-L} & \text{if } i = 2d+2J+L+1, \ldots, N \end{cases}$$

For convenience, we denote $N = 2(d+J+L)$.

Theorem

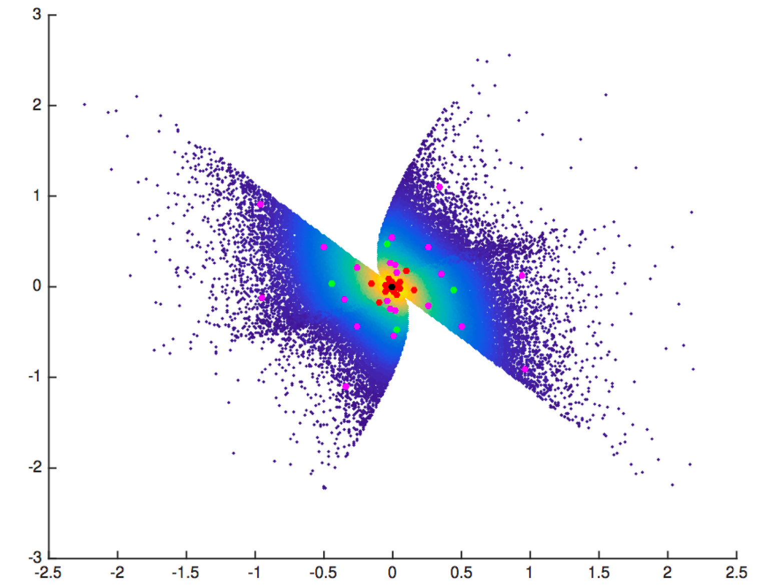
Given the four moment σ-points associated with μ, C, S, and K, then for any $\delta > \sqrt{\frac{\lambda^{\tilde{C}}_{\max}}{\lambda^{C}_{\min}}}$ as defined above and α, β, γ we have

$$\sum_{i=-1}^{N} w_i\sigma_i = \mu \qquad \sum_{i=-1}^{N} w_i(\sigma_i - \mu)^{\otimes 3} = S + \alpha^2\hat{\mu}^{\otimes 3}$$

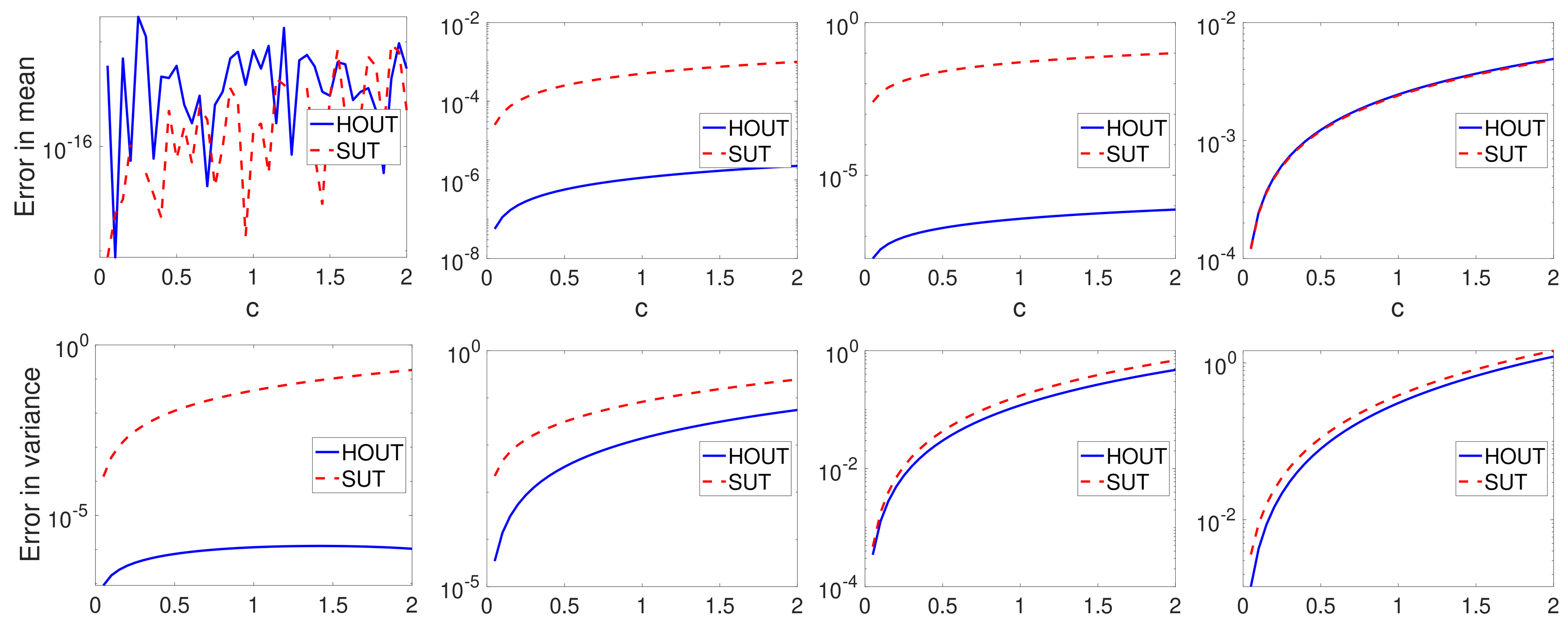
$$\sum_{i=-1}^{N} w_i(\sigma_i - \mu)^{\otimes 2} = C \qquad \sum_{i=-1}^{N} w_i(\sigma_i-\mu)^{\otimes 4} = K + \beta^2\sum_{i=1}^{d}\sqrt{\hat{C}}_i^{\otimes 4}$$

Results

Given a distribution as shown on the right, our ensemble's σ-points are shown in black, red, green and magenta. The black σ-points show our estimate of μ, the red our estimate of C, the green our estimate of S and the magenta our estimate of K.



We conducted a numerical experiment where we generated a two-dimensional distribution and passed our ensemble through several polynomial functions of the form $f(x) = ax + bcx^n$ for $n = 2, 3, 4, 5$ where a and b are made random 1×2 vectors. To show the influence of the strength of the nonlinearity, we sweep through different values of c. We compare the HOUT and SUT for estimating the mean and variance of the output of each of these polynomials.



$f(x) = ax + bcx^2$ $\quad$ $f(x) = ax + bcx^3$ $\quad$ $f(x) = ax + bcx^4$ $\quad$ $f(x) = ax + bcx^5$

We can see from the figure above that, as expected, the HOUT is exact for the means up to $n = 4$ and for the variances up to $n = 2$ due to having degree of exactness four. For higher degree polynomials, the HOUT has comparable or better performance.

Bibliography

- S. Julier and J. K. Uhmann, *A general method for approximating nonlinear transformations of probability distributions.* Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford, 1996.
- E. Kofidis and P. Regalia. *On the Best Rank-1 Approximation of Higher-Order Supersymmetric Tensors.* SIAM J. MATRIX ANAL. APPL., Vol. 23, No. 3, pp. 863-884. 2002.