Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P7789-cd

Polynomial Approximations to the Sliced-Normal Density

Luis G. Crespo[†*] and Steven M. Snyder[†].

***Abstract*— Sliced-Normal (SN) distributions can be used to characterize strongly dependent random variables having multiple modes. Furthermore, their polynomial structure renders common uncertainty quantification tasks, such as identifying the true Most Probable Point (MPP) of failure, computationally tractable. However, the need to accurately estimate the normalization constant has limited their applicability to problems with a moderately large number of uncertain parameters. This note develops lower polynomial bounds and polynomial approximations to the SN joint density that enable computing such a constant analytically. Furthermore, we propose a polynomial class of distributions that exhibit not only the same versatility of the SNs but also the integrability benefits of the proposed approximations. This paper presents the mathematical framework supporting such developments along with easily reproducible examples.**

I. Introduction

The characterization of data is of great significance to uncertainty quantification, model calibration, and robust system identification. Variability in measured data arises from random variation in physical parameters, varying operating conditions, model-form uncertainty, and measurement error. In contrast, variability in calculated data arises from numerical, approximation, and convergence errors.

Modeling the parameter dependencies observed in data enables the analyst to eliminate unnecessary conservatism in the analysis and robust design of systems depending on the underlying data-generating process [1], [2], [3]. Dependencies are often modeled using copulas [2], which are estimated by determining an individual marginal distribution for each variable separately, and then adding a dependence structure describing the joint behavior. In the multivariate case, this behavior is often modeled by the pair-copula decomposition approach called C-vine and R-vine copulas. Several parametric and non-parametric families of copula models can be used to estimate dependencies. The acceptability of a dependency model depends strongly on the selection of an appropriate family of copulas. Unfortunately, standard families of copulas might fail to accurately describe complex dependencies.

In contrast to copulas, sliced normal distributions model the marginals and their joint behavior simultaneously [4], [5], [6]. Sliced distributions have shown to be more versatile than most copula families since they can handle non-monotonic dependencies. Furthermore, whereas the selection of a copula structure is a cumbersome process requiring extensive expertise, the selection of a sliced distribution only requires prescribing the degree of a polynomial transformation to a higher dimensional space. More importantly, their structure makes them amenable to rigorous analysis and robust design methods based on Sum of Squares (SOS) polynomial optimization. This is in contrast to mixture models of other distribution classes, for which the level sets of the joint density are mathematically intractable, thereby preventing rigorous uncertainty quantification. The developments in [7], which are applicable to reliability analysis problems having polynomial limit state functions and SNs, enable the *accurate* identification of *all* the *Most Likely Points* (MLP) of failure. These points, which include the MPP used in FORM and SORM, are guaranteed to be correct thereby preventing the incorrect estimation of the failure probability resulting from gradient-based search algorithms not converging to the true MPP. Many engineering problems in structural reliability and controls render polynomial limit state functions.

⋆This work was supported by the Human Research Program for radiation protection from NASA.
[†]NASA Langley Research Center, Hampton, VA, 23681, USA.
[*]Corresponding author, luis.g.crespo@nasa.gov.

II. Sliced Distributions

Sliced distributions having a polynomial basis allow for the characterization of multivariate data as both a vector of continuous and possibly dependent random variables, or as a semi-algebraic, data enclosing set. A brief introduction to essential features of such distributions is presented next. The joint density of a sliced distribution is

$$f_\delta(\delta; d, \theta, \Delta) = \begin{cases} \exp\left(-\phi(z;\theta)\right)/c(\theta) & \text{if } \delta \in \Delta, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $\theta \in \mathbb{R}^{n_\theta}$ is the shaping parameter,

$$c(\theta) = \int_\Delta \exp\left(-\phi(z;\theta)\right)\, d\delta \tag{2}$$

is the normalization constant, Δ is a hyper-rectangular support set, and

$$z = t(\delta; d) \tag{3}$$

is a polynomial mapping from physical space $\delta \in \mathbb{R}^{n_\delta}$ to feature space $z \in \mathbb{R}^{n_z}$. In particular, $t(\delta; d) : \mathbb{R}^{n_\delta} \to \mathbb{R}^{n_z}$ with $n_z = \binom{n_\delta + d}{n_\delta} - 1$, is the vector of monomials in δ of degree less than or equal to d. The support set Δ, assumed to be hyper-rectangular, and d will be fixed hereafter.

Different functional forms of $\phi(z;\theta)$ give rise to different subclasses of sliced distributions. Next we introduce the sliced-normal subclass.

A. Sliced Normals

Denote as S^n_{++} the space of symmetric positive definite matrices in $\mathbb{R}^{n\times n}$. The density of the Normal vector z is

$$f_z(z; \mu, P) = \exp\left(-\frac{(z-\mu)^\top P (z-\mu)}{2}\right)/c_1, \tag{4}$$

where $c_1 = \sqrt{(2\pi)^{n_z} \det(P^{-1})}$ is the integration constant, $\mathbb{R}^{n_z}$ is the support set, $\mu \in \mathbb{R}^{n_z}$ is the mean, and $P \in S^{n_z}_{++}$ is the precision matrix.

A Sliced Normal (SN) is given by (1) and

$$\phi(z;\theta) = (z-\mu)^\top P(z-\mu), \tag{5}$$

where $\theta = \{\mu, P\}$ is the shaping parameter. Equation (5) is a SOS of polynomials in δ of degree $2d$.

To estimate θ from the data sequence $\mathcal{D} = \left\{\delta^{(i)}\right\}_{i=1}^{n}$, the log-likelihood of $\mathcal{D}$ given by

$$\mathcal{L}\left(\theta, \mathcal{D}_\delta\right) = \sum_{i=1}^{n} \log f_\delta(\delta^{(i)};\theta) \tag{6}$$

can be maximized. This leads to the non-convex program

$$\theta^\star = \underset{\theta\in\Theta}{\operatorname{argmin}} \left\{ n \log c(\theta) + \sum_{i=1}^{n} \phi\left(z^{(i)};\theta\right) \right\}, \tag{7}$$

where $z^{(i)} = t\left(\delta^{(i)}; d\right)$. The shaping parameter $\theta^\star$ is called the *Maximum Likelihood* (ML) estimate.

B. Primal Sliced Normals

The *Primal* SN subclass [5] is given by (1) and

$$\phi(z;\theta) = (z-\mu^\star)^\top C \operatorname{diag}\{\theta\} C^\top (z-\mu^\star), \tag{8}$$

where $\theta = \lambda \in \mathbb{R}^{n_z}$ with $\lambda \geq 0$ is the shaping parameter,

$$\mu^\star = \frac{1}{n}\sum_{i=1}^{n} z^{(i)}, \tag{9}$$

$$P^\star = \left(\frac{1}{n}\sum_{i=1}^{n} \left(z^{(i)} - \mu^\star\right)\left(z^{(i)} - \mu^\star\right)^\top \right)^{-1}, \tag{10}$$

and $CC^\top$ is the Cholesky decomposition of $P^\star$. The ML estimate of primal SNs is given by the solution to the convex optimization program,

$$\lambda^\star = \underset{\lambda\geq 0}{\operatorname{argmin}} \left\{ n \log c(\lambda) + \sum_{i=1}^{n} \phi\left(z^{(i)};\lambda\right) \right\}. \tag{11}$$

The key advantage of the primal SNs over the generic SNs is that their ML estimate results from solving a convex optimization program in a number of decision variables that grows linearly with the dimension of feature space. This is in sharp contrast to the generic SNs, which have a number of decision variables that increase polynomially with such a dimension. Numerical experiments have shown that primal SNs perform as well as generic SNs in most cases while being less susceptible to the approximation error caused by inaccurate estimates of $c_\delta(\theta)$.

III. Normalization

The normalization constant (2) is a multidimensional integral that can not be evaluated analytically in general. Therefore, sampling methods, Bayesian quadrature, or sparse grids approximations [8] are required to solve (7) and (11). Details on the former approximation are given next.

A. Monte Carlo Integration

A sampling-based approximation to $c(\theta)$ is

$$c_{\text{sam}}(m,\theta) = \frac{\text{Vol}(\Delta)}{m} \sum_{i=1}^{m} \exp\left(-\phi\left(t(\delta_u^{(i)};d);\theta\right)\right), \tag{12}$$

where $U = \left\{\delta_u^{(i)}\right\}_{i=1}^{m}$ are samples uniformly distributed over Δ, and $\text{Vol}(\Delta)$ is the volume of Δ. Low-discrepancy Quasi-Monte Carlo sampling, such as the Halton or Sobol sequences, yields more accurate c_{sam} estimates than standard Monte Carlo sampling.

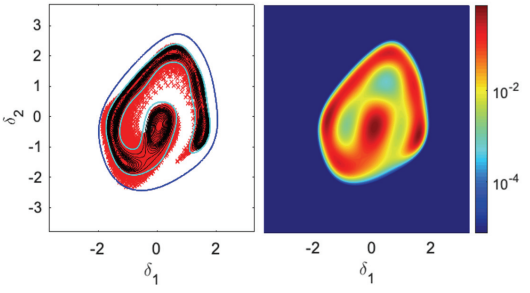

Fig. 1. ML SN computed with $m = 8e4$ samples. The dataset is shown as red crosses on the left subplot along with level sets of the joint density. An upper view of the joint density is shown on the right.

It can be shown that the approximation (12) preserves the convexity of (11) for any value of m. However, small values of m make the error $|c_{\text{sam}}(m,\theta) - c(\theta)|$ large, thereby leading to erroneous ML estimates. This is the case when $m/\text{Vol}(\Delta)$ is too small.

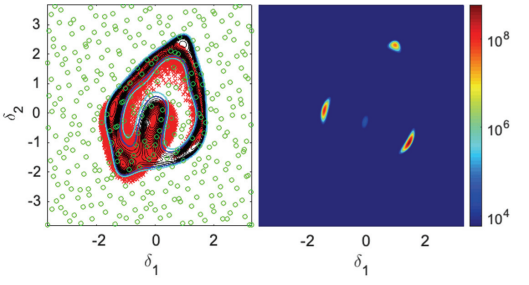

Fig. 2. ML SN computed with $m = 360$ samples. Such samples are shown as green circles on the left subplot.

Large approximation errors lead to erroneous ML estimates whose corresponding density exhibits spurious spikes at under-sampled regions of Δ. In this case the true likelihood of the data might be significantly lower than the computed likelihood at the ML estimate. Thus, even though the numerical search for the ML estimate converges to a sound numerical minimum, the resulting distribution might only attain a suboptimal performance. Figures 1 and 2 show the distributions corresponding to the ML estimate for two values

of m. Whereas the SN in Figure 1 characterizes the dataset well thanks to a sufficiently large m, the distribution in Figure 2 does not. Note the spurious spikes in the latter case.

The inability of $c_{\text{sam}}(m,\theta)$ to accurately approximate $c(\theta)$ for large values of n_δ restricts the SN's applicability to moderately large n_δ values, e.g., $n_\delta \leq 10$. The dependency on d will be removed from the notation hereafter.

B. A Polynomial Lower Bound to the SN Density

In this section we propose a polynomial approximation to the integrand of (2) with ϕ given by (5). This approximation enables the analytical calculation of the normalization constant, thereby eliminating the anomalies caused by using (12) with an insufficiently large number of samples m. This approximation leverages the functional form of the following bound.

Proposition 1: A lower bound to the un-normalized SN density is

$$e^{-\phi(t(\delta);\,\theta)} \geq e^{-\underline{\phi}}\ \alpha(\delta;\theta,\underline{\phi},\overline{\phi}), \tag{13}$$

where

$$\underline{\phi} = \min_{\delta\in\Delta} \phi\left(t(\delta);\theta\right), \tag{14}$$

$$\overline{\phi} = \max_{\delta\in\Delta} \phi\left(t(\delta);\theta\right), \tag{15}$$

and

$$\alpha(\delta;\theta,\underline{\phi},\overline{\phi}) = 1 - z(\delta;\theta,\underline{\phi}) + \hat{\gamma}(\overline{\phi})\ z(\delta;\theta,\underline{\phi})^2. \tag{16}$$

This latter expression is fully prescribed by the invertible positive-definite function

$$\hat{\gamma}(\overline{\phi}) = \left(e^{-\overline{\phi}} - 1 + \overline{\phi}\right)\ /\ \overline{\phi}^2, \tag{17}$$

and $z(\delta;\theta,\underline{\phi}) = \phi(t(\delta);\theta) - \underline{\phi}$.

Proof: It was shown in [9] that for any $x \in [0,1]$ and $a < 0$,

$$e^{ax} \geq ax(1-x) + x^2(e^a - 1) + 1.$$

Equation (13) results from choosing $x = z/\overline{\phi}$, $a = -\overline{\phi}$ and performing algebraic manipulations. ■

Note that α is a linear combination of polynomial squares. Furthermore, notice that the global optimum of the polynomial programs (14) and (15) can be found efficiently by using SOS optimization. It can be shown that the critical points of $\alpha(\delta;\theta,\underline{\phi},\overline{\phi})$ coincide with those of $\phi(\delta;\theta)$. Furthermore, the global minima of α is

$$\underline{\alpha} = \min_{\delta} \alpha(\delta;\theta,\underline{\phi},\overline{\phi}) = 1 - \frac{1}{4\hat{\gamma}(\overline{\phi})}. \tag{18}$$

Note that the bound in (13) might be less than zero, thereby violating the positive definiteness requirement of the joint density. Further notice that this bound as well as the developments that follow are applicable to problems for which $n_\delta > 1$, even though the examples are only one-dimensional.

IV. Approximations to the SN density

Next we propose two polynomial approximations to the SN density based on (13), one based on bounding $\exp(-\phi(t(\delta);\theta))$ directly and another one based on combining bounds to each polynomial square comprising ϕ.

A. An SOS-based Approximation

An approximation to the SN density is

$$\exp\left(-\phi(t(\delta);\theta)\right) \approx \beta(\delta;\theta,\underline{\phi},\gamma,q), \tag{19}$$

where

$$\begin{aligned}\beta(\delta;\theta,\underline{\phi},\gamma,q) &= e^{-\underline{\phi}}\left(\frac{\alpha(\delta;\theta,\underline{\phi},\hat{\gamma}^{-1}(\gamma)) - \underline{\alpha}}{1-\underline{\alpha}}\right)^q,\\ &= e^{-\underline{\phi}}\left(1 - 2\gamma z(\delta;\theta)\right)^{2q},\end{aligned} \tag{20}$$

$\gamma \in \mathbb{R}$, and $q \in \mathbb{N}$. Note that the base is equal to one when $\alpha(\delta) = 1$, i.e., values of δ where the likelihood of the SN is large, and that the base is equal to zero when $\alpha(\delta) = \underline{\alpha}$, i.e., values of δ where the likelihood of the SN is overly small.

The SOS in $\phi(\delta;\theta)$ approaches infinity as $\|\delta\|$ increases. That is also the case for $z(\delta;\theta)$ and $\beta(\delta;\theta)$. The error in the calculation of the integration constant caused by this divergence can be eliminated by reducing the integration domain of (2). This reduction can be done as follows.

Denote as $S(\theta) \subseteq \Delta$ the connected set containing a parameter point satisfying (14) such that $\beta(\delta;\theta,\underline{\phi},\gamma,q) > 0$. Because the approximation error is moderate for all δ in $S(\theta)$, we will use a subset of $S(\theta)$ as the integration domain in (2). To this end, consider the hyper-rectangular set

$$R(\rho) = \left\{\delta\ :\ c - m\rho \leq \delta \leq c + m\rho\right\}, \tag{21}$$

where $\rho \in \mathbb{R}^+$, $c \in \Delta$ and $m > 0$ are vectors in $\mathbb{R}^{n_\delta}$, and m is a unit vector. Consider the optimization program

$$\begin{aligned}\langle \rho^\star, \delta^\star\rangle = \underset{\rho>0,\ \delta\in\Delta}{\operatorname{argmin}}\ & \rho \\ \text{subject to:}\quad & 1 = 2\gamma z(\delta;\theta,\underline{\phi}),\\ & |\delta_i - c_i| \leq \rho m_i \text{ for } i = 1,\dots n_\delta.\end{aligned} \tag{22}$$

Hence, $R(\rho^\star)$ is the greatest hyper-rectangle centered at c with diagonal m fully contained in $S(\theta)$. Additional details of this technique are available in [10]. The chosen values for c and m should lead to the rectangle having the greatest volume. Making c equal to the geometric center of Δ and m equal to the half diagonal of Δ are generally good choices.

Therefore, (2) will be approximated by

$$c(\theta) \approx \int_{R(\rho^\star)\cap\Delta} \beta(\delta;\theta,\underline{\phi},\gamma,q)\ d\delta. \tag{23}$$

Hence, approximating the normalization constant (2) entails solving the polynomial optimization programs (14) and (15), solving (22), and integrating the resulting polynomials over a hyper-rectangular domain.

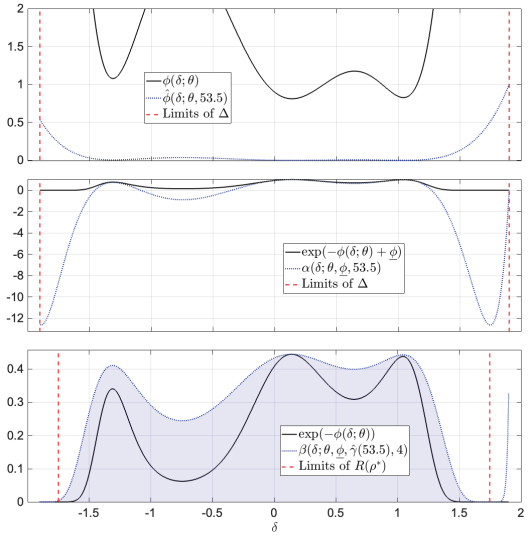

Fig. 3. Normalized and un-normalized SOS (top), lower bound (middle) and approximation (bottom) to the density of a SN in $n_\delta = 1$.

Example 1: Consider an SN of degree 3 in $n_\delta = 1$ with $\mu = [0, 1, -0.2471]^\top$ and

$$P = \frac{1}{2}\begin{bmatrix} 12.8696 & -0.5301 & -7.3345 \\ -0.5301 & 1.8166 & 0.5984 \\ -7.3345 & 0.5984 & 4.5731 \end{bmatrix}.$$

Figure 3 shows variables needed to compute the bound and the approximation above. Define the function

$$\hat{\phi}(\delta;\theta,u) = \frac{\phi(t(\delta);\theta) - \underline{\phi}}{u - \underline{\phi}}. \tag{24}$$

Note that the range of $\hat{\phi}(\delta;\theta,\overline{\phi})$ shown at the top subplot is $[0,1]$. It can be shown that the closer $\hat{\phi}(\delta;\theta,\overline{\phi})$ is to zero, the more accurate is the approximation[1]. Further notice in the middle subplot that even though the bound $\alpha(\delta)$ is loose for most δ in Δ, it preserves the tri-modal structure of the SN. The bottom subplot shows that the approximation $\beta(\delta)$ preserves the tri-modal strucure of the bound. The approximation diverges where the SN density takes on very small values, i.e., when $\hat{\phi}(\delta;\theta,\overline{\phi})$ approaches one. The normalization constant will be computed by integrating (19) over the reduced integration domain $R(\rho^\star)\cap\Delta$, thereby preventing this divergence from impacting the integration. The limits of this domain are shown as dashed-lines at the bottom subplot. The same concepts are illustrated in the example below, various values of γ are considered (see Figure 5).

[1]Non-integer q values might yield better approximations. However, the inability to analytically integrate the resulting non-polynomial approximation renders this choice unsuitable.

B. An Optimal SOS-based Approximation

Numerical studies showed that the accuracy of the polynomial approximation (20) depends strongly on γ, with the choice $\hat{\gamma}(\underline{\phi})$ often leading to poor approximations, e.g., see Figure 3. Figure 4 shows approximations corresponding to three different γ values (or equivalently to different u values satisfying $\hat{\gamma}(u) = \gamma$). Note that the range of $\hat{\phi}(\delta;\theta,u)$ in (24) exceeds one in two out of the three cases, with one of them, $u = 20$, leading to the best fit. This fit is good at the δ range where the SN's density is sufficiently far from zero. Remarkably, the changes in a single parameter reduce the approximation error uniformly over a full range of δ values (as opposed to the changes obtained by increasing the degree of a Taylor expansion).

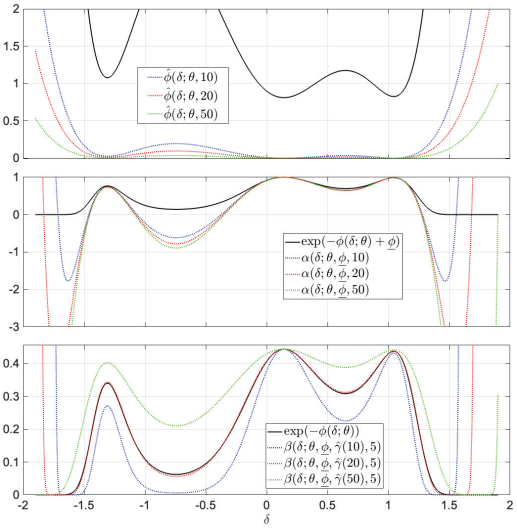

Fig. 4. Normalized and un-normalized SOS (top), lower bounds (middle) and approximations (bottom) to the density of a SN in $n_\delta = 1$ for several values of u.

The choice of u is equivalent to the choice of γ via the function $\hat{\gamma}$. As such, we will consider the approximation $\beta(\delta)$ in (20) where γ and q are tuning parameters, i.e.,

$$\exp\left(-\phi(t(\delta);\theta)\right) \approx \beta(\delta;\theta,\underline{\phi},\gamma^\star,q^\star). \tag{25}$$

The value of γ is chosen to be

$$\gamma^\star = \underset{u>\underline{\phi}}{\operatorname{argmin}}\ e(\hat{\delta};\theta,\underline{\phi},\gamma,q), \tag{26}$$

where $\hat{\delta} \in \Delta$ and q are fixed, and

$$e(\delta;\theta,\underline{\phi},\gamma,q) = \Big(\exp(-\phi(t(\delta);\theta)) - \beta(\delta;\theta,\underline{\phi},\gamma,q)\Big)^2 \tag{27}$$

is the approximation error. Hence, we seek a value of γ that minimizes the error at $\hat{\delta}$. Even though such an error is local, the nature of the bound underlying the approximation renders

a global error reduction. $\hat{\delta}$ should be any point where the likelihood of the SN is sizable but not maximal, e.g., $0.01 < \hat{\phi}(\hat{\delta}, \theta, \overline{\phi}) < 0.1$. The solution to (26) is

$$\gamma^\star = \frac{1 - \exp\left(-\frac{\hat{\phi}-\underline{\phi}}{2q}\right)}{2(\hat{\phi} - \underline{\phi})}. \tag{28}$$

Hence, the value of γ that minimizes the prediction error is available in closed-form.

The prescription of $q^\star$ is studied next. Low-order polynomial approximations are preferable so $q = 1$ should be a starting choice. When the approximation error $e(\hat{\delta}; \theta, \underline{\phi}, \gamma^\star, q)$ is large at the δ points where the likelihood is sizable, q should be increased. Even though $e(\hat{\delta}; \theta, \underline{\phi}, \gamma^\star, q)$ might be zero for several q values, the error elsewhere will differ slightly. Numerical experiments indicate that as q increases, the approximation error decreases globally whereas the volume of $R(\rho^\star)$ increases. The power resulting from this empirical procedure will be denoted as $q^\star$.

The integration constant can be readily computed from Equation (23) after replacing γ with $\gamma^\star$. Even though the integration can be performed analytically for any q, the rapid increase in the number of monomials for the multi-variate case makes a small q preferable.

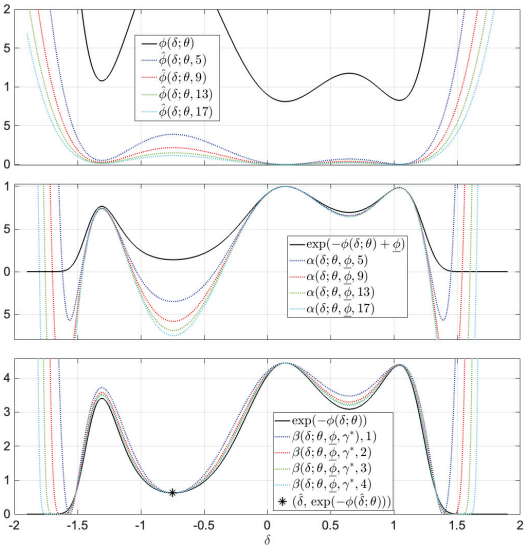

Fig. 5. Normalized and un-normalized SOS (top), lower bounds (middle) and optimal approximations (bottom) corresponding to several values of q. The parameter point at which the error is minimized is marked with a "$*$".

Example 2: Next we compute the optimal approximations in (20) using (26) for several values of q and the same SN used in the previous example. Figure 5 shows the corresponding bounds and approximations. Note that the approximation error decreases over a range of δ values (rather than at the vicinity of $\hat{\delta}$ only) as q is increased but the local error at $\hat{\delta} = -0.75$ is zero in all cases. Further notice that the greater the power q, the larger the subset of Δ where the optimal approximation is accurate.

Proposition 2: The proposed approximations converges to the SN as q increases, i.e.,

$$\lim_{q\to\infty} \beta(\delta; \theta, \underline{\phi}, \gamma^\star, q) = \exp(-\phi(t(\delta);\, \theta)), \tag{29}$$

for all $\delta \in \mathbb{R}^{n_\delta}$. Furthermore, the approximation error e approaches zero monotonically with q for all δ in a $\underline{\phi} \le \phi(\delta;\theta) \le 1/2\gamma^\star$.

Proof: Without loss of generality assume $\underline{\phi} = 0$. Substituting (26) into (20) yields

$$\beta(\delta; \theta, 0, \gamma^\star, q) = \left(1 - \frac{\phi}{\hat{\phi}}\left(1 - \exp\left(\frac{-\hat{\phi}}{2q}\right)\right)\right)^{2q}. \tag{30}$$

Taking the limit after exponentiating the logarithm and applying L'Hôpital's rule,

$$\lim_{q\to\infty} \beta = \exp\left(\lim_{q\to\infty} \frac{\log\left(1 - \frac{\phi}{\hat{\phi}}\left(1 - \exp\left(-\frac{\hat{\phi}}{2q}\right)\right)\right)}{\frac{1}{2q}}\right), \tag{31}$$

$$= \exp\left(\lim_{q\to\infty} \frac{\frac{\frac{\phi}{\hat{\phi}}\exp\left(-\frac{\hat{\phi}}{2q}\right)\frac{\hat{\phi}}{2q^2}}{1-\frac{\phi}{\hat{\phi}}\left(1-\exp\left(-\frac{\hat{\phi}}{2q}\right)\right)}}{\frac{-1}{2q^2}}\right), \tag{32}$$

$$= \exp\left(\lim_{q\to\infty} \frac{-\phi\exp\left(-\frac{\hat{\phi}}{2q}\right)}{1 - \frac{\phi}{\hat{\phi}}\left(1 - \exp\left(-\frac{\hat{\phi}}{2q}\right)\right)}\right), \tag{33}$$

$$= \exp(-\phi), \tag{34}$$

which proves the first claim.

The proof of the second claim is as follows. Differentiating the approximation error (27) with respect to q,

$$\frac{de}{dq} = \frac{d}{dq}\left(\exp(-\phi) - \beta\right)^2 \tag{35}$$

$$= -2\left(\exp(-\phi) - \beta\right)\frac{d\beta}{dq}. \tag{36}$$

Note that $\exp(-\phi) \ge \beta$ for $\hat{\phi} \le \phi \le 1/2\gamma^\star$ and $\exp(-\phi) \le \beta$ for $0 \le \phi \le \hat{\phi}$. Similarly differentiating β:

$$\frac{d\beta}{dq} = \frac{d}{dq}\left(1 - \frac{\phi\left(1 - \exp\left(-\frac{\hat{\phi}}{2q}\right)\right)}{\hat{\phi}}\right)^{2q}, \tag{37}$$

$$= \frac{d}{dq}\exp\left(2q\log\left(1 - \frac{\phi\left(1 - \exp\left(-\frac{\hat{\phi}}{2q}\right)\right)}{\hat{\phi}}\right)\right), \tag{38}$$

$$= \beta A, \tag{39}$$

where

$$A = \log\left(1 - \frac{\phi}{\hat{\phi}}\left(1 - \exp\left(-\frac{\hat{\phi}}{2q}\right)\right)\right) + \frac{\frac{\phi}{q}\exp\left(-\frac{\hat{\phi}}{2q}\right)}{1 - \frac{\phi}{\hat{\phi}}\left(1 - \exp\left(-\frac{\hat{\phi}}{2q}\right)\right)}. \quad (40)$$

Since β is positive on the domain of interest, the sign of $\frac{d\beta}{dq}$ is determined by the sign of A. There are two zeros of A in the domain, at $\phi = 0$ and $\phi = \hat{\phi}$, and A is continuous. For small $\phi < \hat{\phi}$, the log dominates and $A < 0$. For large $\phi > \hat{\phi}$, the second term dominates and $A > 0$. From (36) and the sign information for $\exp(-\phi) - \beta$ and A, we conclude that e is monotonically decreasing with q in the domain of interest. ■

Note that approximations to the SN based on Taylor expansions do not satisfy these properties, e.g., a Taylor expansion of degree three might be better than one of degree four for some $\delta \in \Delta$.

Hence, approximating the integration constant requires solving the optimization programs (14) and (22). The first of these two programs can be eliminated by further constraining the optimization program used to find ML estimate, e.g., by replacing (7) with

$$\min_{\theta\in\Theta,\,\delta\in\Delta} \left\{ n \log c(\theta) + \sum_{i=1}^{n} \phi\left(z^{(i)};\theta\right) + \phi(t(\delta);\theta) \right\}, \quad (41)$$

and using $\underline{\phi} = \phi(t(\delta^\star);\theta^\star)$.

C. Optimal Square-based Approximation

In contrast to the approximation in (19), which applies to the SOS directly, this section presents a framework that accounts for each individual square comprising $\phi(\delta;\theta)$ separately. A Choleski decomposition of the SOS $\phi(\delta;\theta)$ renders

$$\phi(\delta;\theta) = \sum_{i=1}^{n_z} s_i\left(t(\delta);\theta\right), \quad (42)$$

where s_i is a polynomial square in the variable δ. As before, SOS optimization can be used to solve

$$\underline{s}_i = \min_{\delta\in\Delta} s_i\left(t(\delta);\theta\right), \quad (43)$$

$$\overline{s}_i = \max_{\delta\in\Delta} s_i\left(t(\delta);\theta\right), \quad (44)$$

for $i = 1, \ldots n_z$. Applying the bound in (13) to each square and combining them yields

$$\exp(-\phi(t(\delta);\theta)) \approx \prod_{i=1}^{n_z} e^{-\underline{s}_i}\,\alpha_i(\delta;\theta,\underline{s}_i,\overline{s}_i) = h(\delta;\theta,\gamma), \quad (45)$$

where $\alpha_i(\delta;\theta,\underline{s}_i,\overline{s}_i) = 1 - (s_i(\delta;\theta) - \underline{s}_i) + \gamma_i(s_i(\delta;\theta) - \underline{s}_i)^2$ and $\gamma_i = (e^{-\overline{s}_i} - 1 + \overline{s}_i)/\overline{s}_i^2$. Note that the function $h(\delta)$ in (45) is not a lower bound of the SN density when the α_i's take on negative values. As before, we propose the approximation

$$\tilde{\beta}(\delta;\theta,\underline{s},\gamma,q) = \prod_{i=1}^{n_z} e^{-\underline{s}_i}\left(1 - 2\gamma_i\left(\phi(t(\delta);\theta) - \underline{s}_i\right)\right)^{2q_i}, \quad (46)$$

where $q_i \in \mathbb{N}$. In contrast to (20), the approximation (47) can only approach the SN density locally thereby requiring a multi-point calibration formulation. The desired approximation is

$$\exp(-\phi(t(\delta);\theta)) \approx \tilde{\beta}\left(\delta;\theta,\underline{s},\gamma^\star,q^\star\right), \quad (47)$$

where $\gamma^\star \in \mathbb{R}^{n_z}$ is given by

$$\gamma^\star = \underset{u>\underline{s}}{\operatorname{argmin}} \sum_{j=1}^{n_p}\sum_{i=1}^{n_z} e(\hat{\delta}^{(j)};\theta,\underline{s}_i,\gamma_i,q_i), \quad (48)$$

e is the approximation error in (27), $\left\{\hat{\delta}^{(j)}\right\}_{j=1}^{n_p}$ is a sequence of high likelihood training points, and $q^\star$ is prescribed according to the procedure outlined above.

Example 3: Figure 6 shows variables needed to compute the optimal square-based approximation for the same SN used previously. Note that ϕ is comprised of three squares, which are color coded in the subplots. We first consider the approximations resulting from the suboptimal value $\gamma = \hat{\gamma}$. Note that each term in the product (45) bounds a square from below, but the SN density is not bounded. This can be seen at $\delta = -1.7$, where the value of h in (45) exceeds the likelihood therein. In contrast to the SOS-based bound studied earlier, the critical points of $\phi(\delta;\theta)$ do not coincide with those of $h(\delta)$. Consequently, the three modes of the resulting approximation do not coincide with those of the probability density.

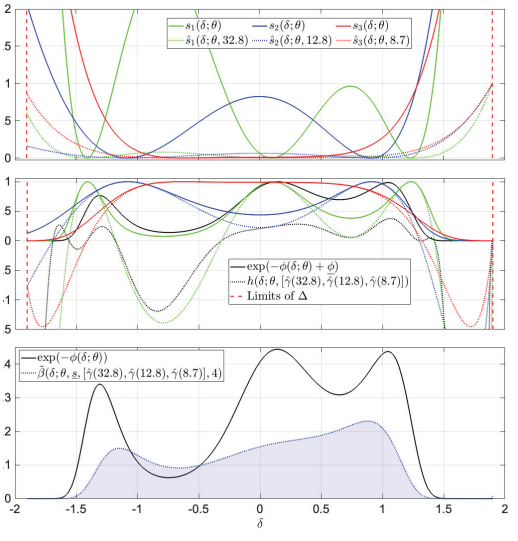

Fig. 6. Normalized and un-normalized squares (top), lower bounds to the contribution from individual squares to the SN (middle) and approximation to the density of a SN (bottom) in $n_\delta = 1$ for $u = \overline{s}$.

Figure 7 shows the same information in Figure 6 but for the optimal value $\gamma^\star$ corresponding to the $n_p = 2$ points

shown. Note that the approximation does not match the density at the training points, and that (29) no longer holds. Even though $\tilde{\beta}(\delta;\theta,\underline{s},\gamma^\star,q^\star)$ exhibits the desired tri-modal shape, the approximation error is fairly large, thereby rendering the SOS-based approximation $\beta(\delta;\theta,\underline{s},\gamma^\star,q^\star)$ preferable.

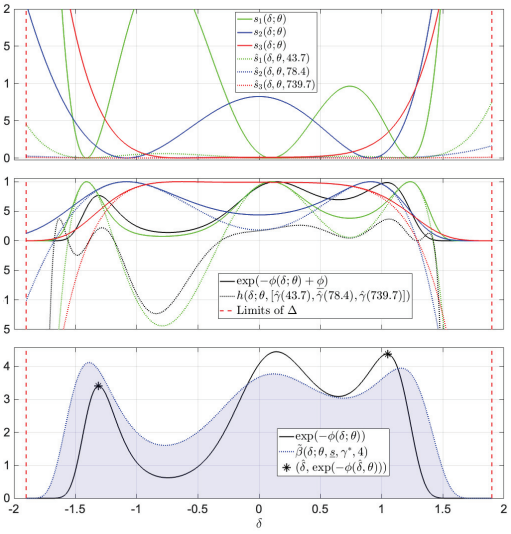

Fig. 7. Normalized and un-normalized squares (top), bounds of components of the SN corresponding to individual squares (middle) and optimal approximation (bottom) to the SN density.

D. Discussion

Having to solve (22) for each function call of (7) greatly increases the computational complexity of finding the ML estimate. As such, the strategy for computing the integration constant based on the above approximations should only be used when the number of samples m in (12) required to obtain an accurate estimate is too large to be computed efficiently, e.g., when $n_\delta > 10$.

V. Sliced-Polynomial Distributions

Next we propose a polynomial class of distributions based on the SOS approximation to the SN in (19) called the *Sliced-Polynomials* (SP). These distributions not only inherit the versatility of the SNs but also allow for the analytical calculation of the integration constant. Consider the random vector $\delta \in \mathbb{R}^{n_\delta}$ with joint density

$$f_\delta(\delta;\psi,\underline{\phi},q,\Delta) = \begin{cases} \beta(\delta;\theta,0,\gamma,q)/c(\theta) & \text{if } \delta \in \Delta, \\ 0 & \text{otherwise,} \end{cases} \quad (49)$$

where β was defined in (20), $q \in \mathbb{N}$ is a fixed parameter, $\psi = \{\theta,\gamma\}$ is the shaping parameter, and c is the normalization constant. Notice that this distribution allows for the analytical calculation of the integration constant.

The ML estimate of ψ, $\psi^\star$, is the solution to

$$\max_{\psi\in\Psi,\ \delta\in\Delta} \ q\sum_{j=1}^{n}\log\left(1-2\gamma(\phi(t(\delta);\theta))\right) - n\log c(\theta), \quad (50)$$
$$\text{subject to:} \quad \phi\left(t(\delta);\theta\right) = 0,$$

where $\Psi = \Theta \times \mathbb{R}^+$. Note that the cost function of (50) is the log-likelihood of the data sequence D_δ, and that this formulation eliminates the need to solve (14) because $\phi = 0$. In contrast to the SNs, solving for the ML estimate in (50) does not require using (12). SP distributions exhibiting the diverging tails seen in Figure 5 will attain a lower likelihood of the data. The optimal SPs resulting from (50) will naturally ameliorate this anomaly. This outcome could also be attained by choosing a greater q and a support set Δ enclosing D_δ tightly.

Example 4: Next we compute SPs for the data sequence D_δ leading to the tri-modal SN used above. Figure 8 shows ML SPs of increasing degree q. Note that the SPs resemble the tri-modal SN. The greater q, the greater the log-likelihood of the data and the further the diverging tails. Note that the SPs for $q \geq 3$ attain a greater log-likelihood than the SN. Note that the maximization of the likelihood yields distributions for which the adverse effects of the diverging probability tails are naturally mitigated, e.g., the diverging tails for the ML SP estimate are outside the support set.

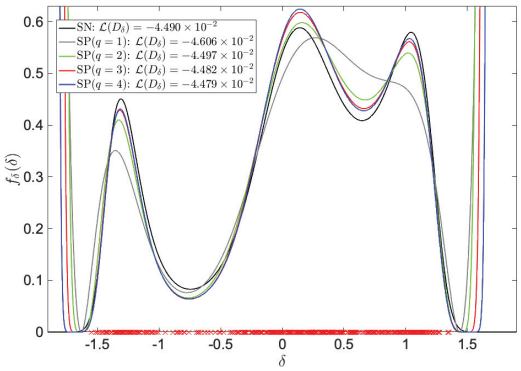

Fig. 8. Maximum Likelihood SN and SPs of varying degrees q. The support set of the SPs has been reduced in order to remove the diverging tails when computing the normalization constant. Elements of the data set D_δ are shown as red "×".

VI. Concluding Remarks

This paper proposes a framework to approximate the SN density with polynomials. Numerical studies show that the proposed approximation converges both globally and uniformly. This approximation enables calculating the integration constant analytically, thereby allowing for the estimation of SNs in higher dimensional spaces. Unfortunately, this process requires solving an auxiliary optimization program, thereby increasing the computational cost. Furthermore,

we propose a new class of polynomial distributions that leverage the versatility and convergence properties of the proposed approximations without requiring the solution of this auxiliary optimization program. The effectiveness and computational complexity of this class, as well as of the proposed approximation, will be evaluated in the future.

References

[1] Y. Wang, N. Zhang, C. Kang, and M. Miao, "An efficient approach to power system uncertainty analysis with high-dimensional dependencies," *IEEE Transactions on Power Systems*, vol. 33, no. 3, 2017.

[2] C. Czado, "Analyzing dependent data with vine copulas," *Springer International Publishing*, 2019.

[3] E. Torre, E. Marelli, P. Embrechts, and B. Sudret, "A general framework for data-driven uncertainty quantification under complex input dependencies using copulas," *Probabilistic Engineering Mechanics*, vol. 55, 2019.

[4] L. Crespo, B. Colbert, and S. Kenny, "On the quantification of aleatory and epistemic uncertainty using sliced-normal distributions," *Systems and Control Letters*, vol. 134, 2019.

[5] L. Crespo, B. Colbert, S. Slagel, and S. Kenny, "Robust estimation of sliced-exponential distributions," in *IEEE 60th CDC Conference*, 2021.

[6] L. Crespo, B. Colbert, S. Kenny, and D. Giesy, "Dimensionality reduction of sliced-normal distributions," *21st IFAC 2020 world congress, Berlin, Germany*, 2020.

[7] J. Hammond, L. G. Crespo, and F. Montomoli, "A distributionally robust data-driven framework to reliability analysis," *Structural Safety*, vol. 111, p. 102501, 2024.

[8] F. Briol, C. Oates, M. Girolami, and M. Osborne, "Frank-wolfe Bayesian quadrature: Probabilistic integration with theoretical guarantees," in *Advances in Neural Information Processing Systems 28*, 2015, pp. 1162–1170.

[9] C. Chesneau, Y. Bagul, and R. Manikrao, "On simple polynomial bounds for the exponential function," *Asia Pacific Journal of Mathematics*, vol. dod: 10.28924/APJM/9-6, 2022.

[10] L. Crespo, D. Giesy, and S. Kenny, "Robustness analysis and robust design of uncertain systems," *AIAA Journal*, vol. 46, no. 2, 2008.