Supplementary Materials:
Thicken, Sample, Project: Monte Carlo Computation for Implicit Statistical Models

David Kahle

Jonathan D. Hauenstein

S1 The Variety Normal Distribution as Shifting Surfaces

The variety normal (VN) distribution admits an interesting interpretation as a four step procedure of selecting a random surface, intersecting it with a fixed surface, projecting the intersection, and randomly selecting a point in the resulting set. Let $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ denote the projection onto the first n coordinates, namely $(x_1, \ldots, x_n, x_{n+1}) \mapsto (x_1, \ldots, x_n)$, and $\mathcal{G}_f = \left\{ \big(\boldsymbol{x}, f(\boldsymbol{x})\big) : \boldsymbol{x} \in \mathbb{R}^n \right\} \subset \mathbb{R}^{n+1}$ be the graph of the function $f : \mathbb{R}^n \to \mathbb{R}$, which is an n-dimensional hypersurface in $\mathbb{R}^{n+1}$. For $c \in \mathbb{R}$, let $\mathcal{G}_c$ denote the flat n-dimensional hyperplane in $\mathbb{R}^{n+1}$ at height c parallel to the $\boldsymbol{x}$-axis. Sampling the VN distribution then admits the following five perspectives, among others, each corresponding to different selections of hypersurfaces derived from manipulations of the equation $\frac{g}{||\nabla_{\boldsymbol{x}} g||} - X = 0$ with $X \sim \mathcal{N}\left(0, \sigma^2\right)$. The first is the one that motivated our previous discussion.

I. $\mathcal{G}_{\overline{g}-X} \cap \mathcal{G}_0$: randomly shift $\mathcal{G}_{\overline{g}}$ up/down in $\mathbb{R}^{n+1}$ by a $\mathcal{N}\left(0, \sigma^2\right)$ draw, intersect it with the $\boldsymbol{x}$-axis, project the intersection using π onto the first n components, and sample uniformly from the resulting set.

II. $\mathcal{G}_{\overline{g}} \cap \mathcal{G}_X$, randomly shift $\mathcal{G}_0$ up/down in $\mathbb{R}^{n+1}$ by a $\mathcal{N}\left(0, \sigma^2\right)$ draw, intersect it with $\mathcal{G}_{\overline{g}}$, project the intersection using π onto the first n components, and sample uniformly from the resulting set.

III. $\mathcal{G}_g \cap \mathcal{G}_{X||\nabla_{\boldsymbol{x}} g||}$, which has two interpretations:

   (1) differentially scale the random flat plane $\mathcal{G}_X$ by $||\nabla_{\boldsymbol{x}} g||$, intersect it with $\mathcal{G}_g$, project the intersection using π onto the first n components, and sample uniformly from the resulting set, or

   (2) randomly scale $\mathcal{G}_{||\nabla_{\boldsymbol{x}} g||}$ with a $\mathcal{N}\left(0, \sigma^2\right)$ draw, intersect it with $\mathcal{G}_g$, project the intersection using π onto the first n components, and sample uniformly from the resulting set.

IV. $\mathcal{G}_{g-X||\nabla_{\boldsymbol{x}} g||} \cap \mathcal{G}_0$: intersect a carefully selected random hypersurface in $\mathbb{R}^{n+1}$ with the n-dimensional $\boldsymbol{x}$-axis, project the intersection using π onto the first n components, and sample uniformly from the resulting set.

V. $\mathcal{G}_{g^2-X^2||\nabla_{\boldsymbol{x}} g||^2} \cap \mathcal{G}_0$, similar to the previous perspective but where the hypersurface is a variety, which is of interest as it is more algebraic. This approximates the offset variety [Horobet and Weinstein, 2018].

These perspectives are illustrated in Figure S1.

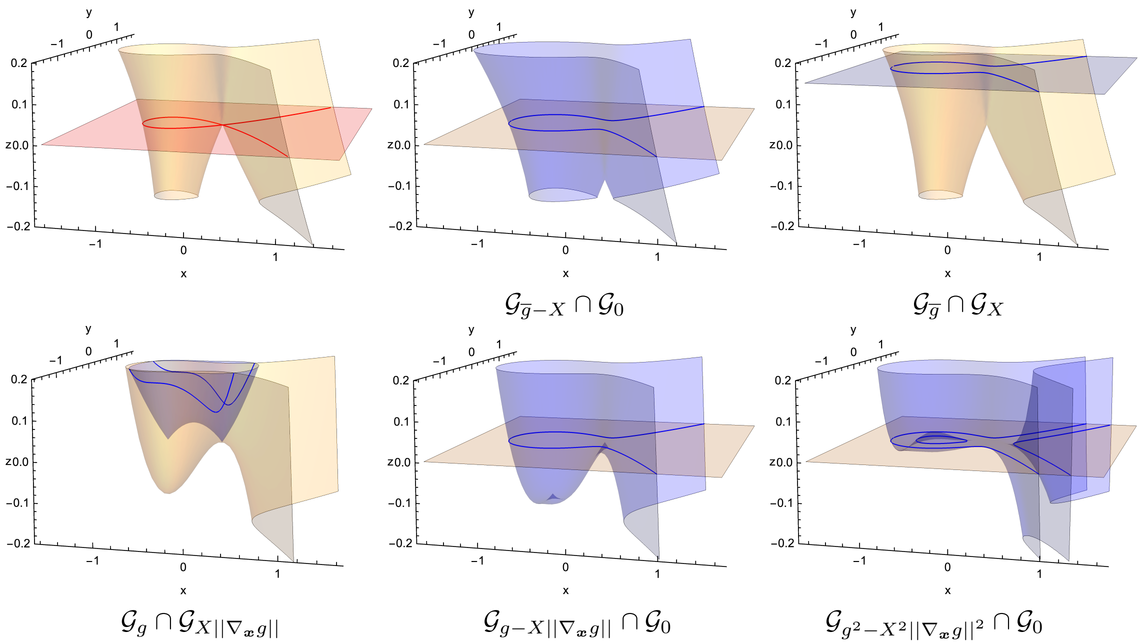

Figure S1: Geometrically, the variety normal distribution can be seen from various perspectives as selecting a point at the intersection of a fixed and a random surface (projected into $\mathbb{R}^n$). Perspectives I.–V. are illustrated, with the random surface in blue. In each case, the random surface illustrated was generated with $X = .15$, i.e. at 1.5 "standard deviations" from the variety with $\sigma = .1$.

S2 The Whitney Umbrella Example

The VN distribution is not only defined for distributions on $\mathbb{R}^2$, but any dimensional real space; the distributions just become more difficult to view. In $\mathbb{R}^3$, the Whitney umbrella is define by the polynomial $g(x, y, z) = x^2 - y^2 z$ and is illustrative of the general character of these distributions. Figure S2 provides 3D density visuals of this distribution with different amounts of variability. As the Whitney umbrella is not compact, a truncation is used.

A helpful way to think of these distributions is as probabilistic thickenings of the variety into the ambient space $\mathbb{R}^n$. As the Whitney umbrella demonstrates, the thickening expands all components of the variety,

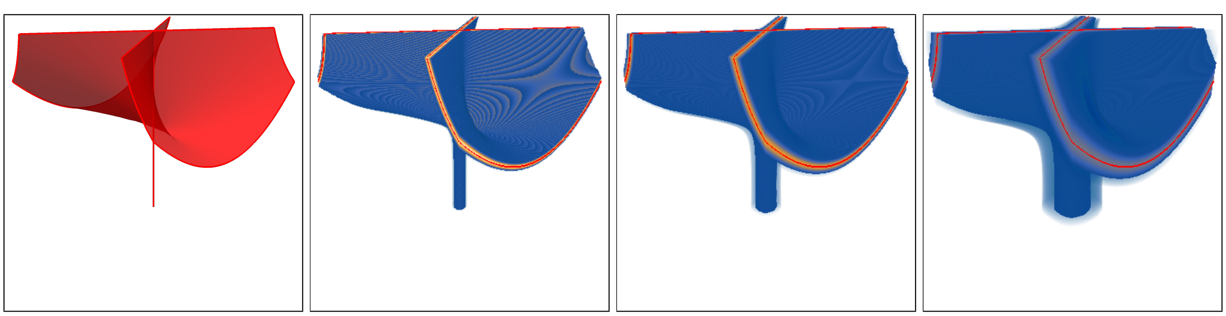

Figure S2: (Left) The Whitney umbrella defined by $g(x, y, z) = x^2 - y^2 z$ is the union of a one-dimensional handle and a two-dimensional canopy in $\mathbb{R}^3$. (Right) Three 3D density plots of the truncated VN distribution defined by the Whitney umbrella with $\sigma = .025$, $.05$, and $.10$.

regardless of dimension, into the full dimension of the ambient space. Moreover, the thickening is locally uniform across the variety in the sense that, away from singularities, two points equidistant from the variety have arbitrarily close probability density provided σ^2 is selected sufficiently small. By construction, probability density decreases exponentially as one moves orthogonally away from the variety–in the direction parallel to the gradient. It is also approximately symmetric, with asymmetries arising from the curvature in $g(\boldsymbol{x}|\boldsymbol{\beta})$ relative to the size of σ^2.

S3 Near Roots

Before considering how to guarantee normalizability of a HVN or VN for a given g, it helps to understand the opposite end of the spectrum: when g does not have an exact root nearby, but evaluate close to zero. In order to to understand this case, a univariate $g \in \mathbb{R}[x]$ is instructive.

Consider $g = x^2 - c$, where $c > 0$. Clearly, $\mathcal{V}(g) = \{\pm\sqrt{c}\}$. Thus, $\widetilde{p}(x|g, \sigma^2)$ is maximized at $\pm\sqrt{c}$ for both the HVN and VN distributions. These distributions have somewhat different character, however: the HVN case resembles a mixture of normals, and the VN case has a probabilistic chasm between the two roots.

One potential issue arising with the use of the HVN or VN distributions in practice is that of near roots, values of $\boldsymbol{x}$ such that $g(\boldsymbol{x})$ is close to zero. The basic intuition is that if where $g \approx 0$, $\exp\{-g^2\} \approx 1$, and thus the variety normal distributions will place at least some probability mass near these locations. To illustrate, consider the cubic polynomial $g = x\big((s-2\epsilon)x^2 + (3\epsilon - 2s)x + s\big) \in \mathbb{R}[x]$ for positive numbers $s \geq 3\epsilon > 0$, constructed so that $g(0) = 0$, $g'(0) = s$, $g(1) = \epsilon$, $g'(1) = 0$, and $g''(1) \geq 0$. If ϵ is small, g has a near root at $x = 1$, but the condition $s \geq 3\epsilon$ guarantees $\mathcal{V}(g) = \{0\}$, i.e., it has a single root. Hence, s governs the slope of the curve at $x = 0$. Clearly, if ϵ is small, $x = 1$ will look very much like a root of g and the distributions may have a hard time distinguishing between the real root at $x = 0$ and the false root at $x = 1$. They may place high probability mass on both with the amount of probability mass placed on $x = 1$ depends on σ as shown in Figure S3.

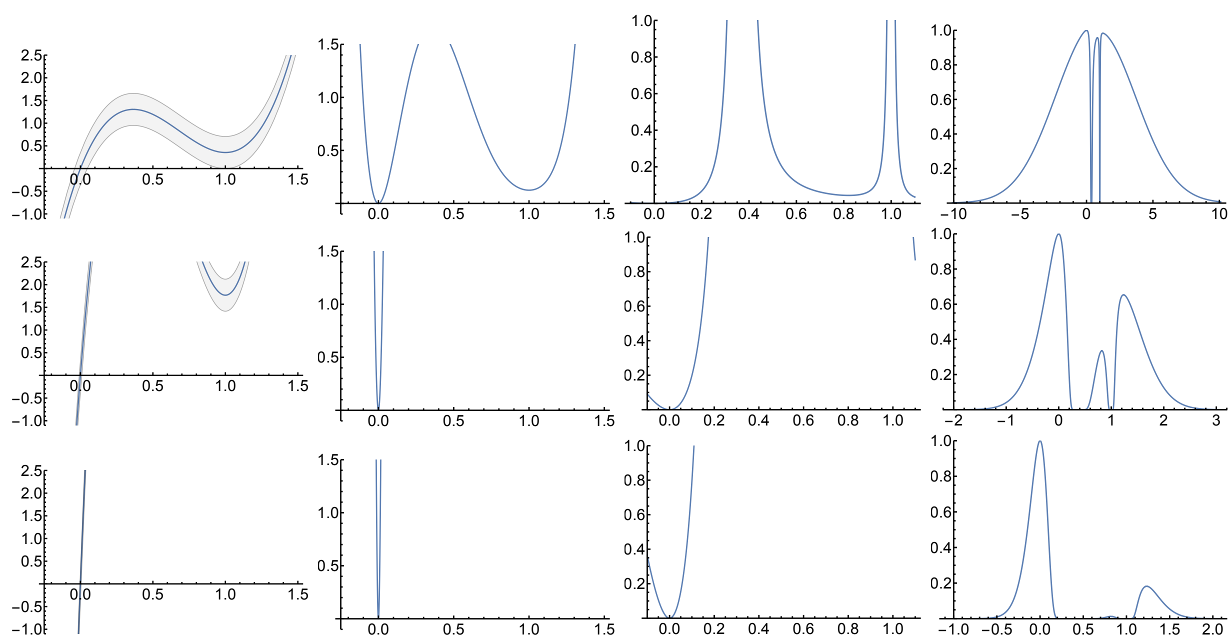

Figure S3: In univariate models, near roots are eliminable with a sufficiently small σ. For the cubic polynomial $g = x\big((10+\epsilon)x^2 - (20+3\epsilon)x + (10+3\epsilon)\big)$ with $\epsilon = \frac{1}{2}$, left to right: $\frac{g}{\sqrt{2}\sigma}$, $\frac{g^2}{2\sigma^2}$, $\frac{\bar{g}^2}{2\sigma^2}$, $\exp\{-\frac{\bar{g}^2}{2\sigma^2}\}$. Top to bottom: $\sigma = 1, \frac{1}{5}, \frac{1}{10}$. Note the x-axis scale change in the right plots.

S4 Detailed Normalizability Analysis

The following provides analysis of normalizability. A first obstruction to normalizability is already visible in a simple unbounded example such as $g(x, y) = y - x$ with $\overline{g}(x, y) = (y - x)/\sqrt{2}$. The corresponding VN pseudodensity is

$$\widetilde{p}(x, y|g, \sigma^2) = \exp\left\{-\frac{(y - x)^2}{4\sigma^2}\right\}.$$

This function is constant along lines parallel to $(1, 1)$, so its Lebesgue integral over $\mathbb{R}^2$ diverges. Thus, unbounded varieties can require truncation or tapering even in very simple cases.

S4.1 Tapering and truncation

Proposition S1. *Let $\widetilde{p}(\boldsymbol{x}|g, \sigma^2)$ be the HVN or VN pseudodensity of a nonzero or nonconstant (respectively) polynomial $g \in \mathbb{R}[\boldsymbol{x}]$, and let $f(\boldsymbol{x})$ the Lebesgue PDF of any distribution on $\mathbb{R}^n$. Then $\widetilde{h}(\boldsymbol{x}) := \widetilde{p}(\boldsymbol{x}|g, \sigma^2)f(\boldsymbol{x})$ is a normalizable pseudodensity on $\mathbb{R}^n$.*

Proof. Consider the HVN case. Since the product of two measurable functions is measurable, $\widetilde{h}(\boldsymbol{x})$ is measurable, $\widetilde{p}(\boldsymbol{x}|g, \sigma^2)$ being (Lebesgue) measurable since it is continuous. Moreover, $\widetilde{h}(\boldsymbol{x})$ is non-negative, since $\widetilde{p}(\boldsymbol{x}|g, \sigma^2)$ is strictly positive and $f(\boldsymbol{x})$ is non-negative. Thus, $\int_{\mathbb{R}^n} \widetilde{h}(\boldsymbol{x})\, d\boldsymbol{x} \in (0, \infty]$. Since $g(\boldsymbol{x}|\boldsymbol{\beta})^2 \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$, one has $\widetilde{p}(\boldsymbol{x}|g, \sigma^2) = \exp\{-\frac{g(\boldsymbol{x}|\boldsymbol{\beta})^2}{2\sigma^2}\} \leq 1$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Thus, $\widetilde{h}(\boldsymbol{x}) = \widetilde{p}(\boldsymbol{x}|g, \sigma^2)f(\boldsymbol{x}) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x}$ and, since the latter bounds finite volume over $\mathbb{R}^n$, so too must the former.

The same proof works for the VN case, noting that measureability derives from the fact that $\overline{g} = g/\,||\nabla g||$ is only undefined on a set of measure 0 where it can be arbitrarily redefined, $||\nabla g||$ being nonzero due to g being nonconstant. □

S4.2 Multivariate near roots and the failure of Gaussian covers

Although unboundedness guarantees nonexistence of the variety normal, boundedness itself may not be sufficient to guarantee normalizability of the VN due to the persistence of near zeros encountered in the multivariate setting. To understand why, we need to understand σ^2 and the influence of near zeros of g better.

The fundamental shifting surfaces perspective is that σ^2 shifts the graphs of g and $\overline{g}$ up and down to determine the HVN and VN distributions. Suppose $g = x - \mu$ as in the univariate normal case. It is natural to think of σ^2 as varying the root left and right on the x axis, but in fact it is more helpful to think of it as varying the line up and down. In this case, these are in fact the same since $dg = 1$ and $g = \overline{g}$, so σ^2 can be thought of either way and the effect is the same.

However, with other polynomials, the situation is different. This shifting up and down also has implications for near roots and roots with multiplicity. In particular, roots with multiplicity may not have them upon shifting, and those that don't have roots may be ascribed some probability mass due to their proximity to a root. The former case does not have dire implications as that region will still receive probability mass. However, the latter seems a bit more delicate.

Consider the polynomial $g = (x^2 - y^2)^2 + x^2 + y^2$, a multivariate sum of squares (SOS) polynomial. Notice that the x^2 and y^2 terms guarantee that $(x, y) = (0, 0)$ is the only possibility of a zero, and in fact $\mathcal{V}(g) = \{(0, 0)\}$. However, consider the relevant rational function

$$\frac{g^2}{||\nabla_{\boldsymbol{x}} g||^2} = \frac{((x^2 - y^2)^2 + x^2 + y^2)^2}{(2x + 4x\,(x^2 - y^2))^2 + (2y - 4y\,(x^2 - y^2))^2}.$$

The peculiar thing is that as $||\boldsymbol{x}||^2 = x^2 + y^2$ grows large, the function above need not. In other words, it feels natural to think that if $\deg g = d_g$, since the degree of the numerator is $2d_g$, even, and the degree of the denominator is $2d_g - 2$, and the highest degree terms have positive coefficients, that the further away you move from the origin at some point the function must go to ∞. Unfortunately, this is demonstrably not the case. In polar coordinates, $x = r\cos\theta$, $y = r\sin\theta$, and $\overline{g}$ is

$$\overline{g}_{r,\theta} = \frac{r^2\left(r^2\cos(4\theta) + r^2 + 2\right)^2}{4\left(8r^4 + 8r^2 + 8\left(r^4 + r^2\right)\cos(4\theta) + 4\right)}.$$

Consider the curve $(r, \theta(r))$ with $\theta(r) = \frac{1}{4}\cos^{-1}\left(\frac{1+r^2-r^4}{r^2+r^4}\right)$. Since $\lim_{r\to\infty}\theta(r) = \frac{\pi}{4}$, $(r, \theta(r))$ flows into the line $y = x$ in the first quadrant; it grows unbounded. (Note that $y = x$ is a factor of the first term's square.) In particular,

$$\overline{g}_{r,\theta(r)} \quad = \quad \frac{4r^4 + 3r^2}{16(r^2+1)^2} \quad = \quad \frac{1}{4} + O(r^{-2}),$$

so that along this curve, the function does not grow unbounded. This implies that there is no bivariate normal distribution, regardless of how large the variance is and how high the multiple, whose PDF covers that of the VN distribution with this g. The near-root of $\overline{g}$ along the lines $y = x$ and $y = -x$, while diminished by smaller σ, is not eliminable by it in the sense that a Gaussian can cover it.

S4.3 Coercivity and Gaussian domination

The Gaussian cover strategy requires a more stringent condition: that the normalized square $\overline{g}^2$ is 2-coercive. Coercivity is a condition from optimization that formalizes the asymptotic growth required for Gaussian domination [Bajbar and Stein, 2015, Bajbar and Behrends, 2017]. We use the following definition.

Definition S1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is coercive if $||\boldsymbol{x}|| \to \infty$ implies $f(\boldsymbol{x}) \to \infty$. A f is q-coercive if for all $c > 0$ there exists a $R \geq 0$ such that $||\boldsymbol{x}|| \geq R$ implies $f(\boldsymbol{x}) \geq c\,||\boldsymbol{x}||^q$.*

In particular, 2-coercivity yields an integrable Gaussian envelope for the VN pseudodensity.

Proposition S2. *If $\overline{g}^2$ is 2-coercive, then the un-normalized VN density in Definition 2 of the main text is integrable on $\mathbb{R}^n$ and hence normalizable.*

Proof. Let

$$\widetilde{p}(\boldsymbol{x}|g, \sigma^2) = \exp\left\{-\frac{\overline{g}(\boldsymbol{x})^2}{2\sigma^2}\right\}.$$

Since $\overline{g}^2$ is 2-coercive, taking the constant $c = 1$ in the definition yields an $R \geq 0$ such that $||\boldsymbol{x}|| \geq R$ implies

$$\overline{g}(\boldsymbol{x})^2 \geq ||\boldsymbol{x}||^2.$$

Hence, outside the closed ball $\overline{B}_R^n$,

$$\widetilde{p}(\boldsymbol{x}|g, \sigma^2) \leq \exp\left\{-\frac{||\boldsymbol{x}||^2}{2\sigma^2}\right\}.$$

The right-hand side is the un-normalized density of an isotropic multivariate normal distribution and is integrable on $\mathbb{R}^n$.

Inside $\overline{B}_R^n$ we simply use the bound $0 \leq \widetilde{p}(\boldsymbol{x}|g, \sigma^2) \leq 1$, so

$$\int_{\overline{B}_R^n} \widetilde{p}(\boldsymbol{x}|g, \sigma^2)\, d\boldsymbol{x} \leq \lambda(\overline{B}_R^n) < \infty.$$

Combining the integral over $\overline{B}_R^n$ with the Gaussian-dominated tail integral over its complement proves that $\widetilde{p}(\boldsymbol{x}|g, \sigma^2)$ is integrable on $\mathbb{R}^n$. $\square$

S5 Reduction of the MVN to the VN

The following verifies that when $m = 1$ and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$, the multivariety normal distribution reduces to the variety normal distribution. Define $\boldsymbol{d} = \nabla_{\boldsymbol{x}} g(\boldsymbol{x}|\boldsymbol{\beta})$ and note that in the $m = 1$ case, $\boldsymbol{g}(\boldsymbol{x}|\boldsymbol{B}) = g(\boldsymbol{x}|\boldsymbol{\beta}) = g$, $\boldsymbol{J}_{\boldsymbol{x}} = \boldsymbol{d}'$, and $\boldsymbol{J}_{\boldsymbol{x}}^{+} = \boldsymbol{d}(\boldsymbol{d}'\boldsymbol{d})^{-1} = \frac{1}{||\boldsymbol{d}||^2}\boldsymbol{d}$. Therefore, when $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$,

$$\begin{aligned}
\exp\left\{-\frac{1}{2}(\boldsymbol{J}_{\boldsymbol{x}}^{+}\boldsymbol{g})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{J}_{\boldsymbol{x}}^{+}\boldsymbol{g})\right\} &= \exp\left\{-\frac{1}{2}\left(\left(\frac{1}{||\boldsymbol{d}||^2}\boldsymbol{d}\right)g\right)'(\sigma^2\boldsymbol{I}_n)^{-1}\left(\left(\frac{1}{||\boldsymbol{d}||^2}\boldsymbol{d}\right)g\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{g}{||\boldsymbol{d}||}\left(\frac{\boldsymbol{d}}{||\boldsymbol{d}||}\right)'\left(\frac{\boldsymbol{d}}{||\boldsymbol{d}||}\right)\frac{g}{||\boldsymbol{d}||}\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{g}{||\nabla_{\boldsymbol{x}} g||}\right)^2\right\} \quad = \quad \exp\left\{-\frac{\bar{g}^2}{2\sigma^2}\right\},
\end{aligned}$$

which is the PDF of the variety normal distribution. More generally, if the system is not overdetermined, $\boldsymbol{J}_{\boldsymbol{x}}$ typically has m independent rows so that, when $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$,

$$p(\boldsymbol{x}|\boldsymbol{g}, \boldsymbol{\Sigma}) \quad \propto \quad \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{g}'(\boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{J}_{\boldsymbol{x}}')^{-1}\boldsymbol{g}\right\}. \tag{S1}$$

In particular, when $\boldsymbol{g} = \boldsymbol{x} - \boldsymbol{\mu}$, $m = n$ and $\boldsymbol{J}_{\boldsymbol{x}} = \boldsymbol{I}_n$, the distribution reduces to the multivariate normal. The individual varieties of the g_i's are the n hyperplanes of dimension $n - 1$ that run parallel to the coordinate axes and intersect at the point $\boldsymbol{\mu}$.

S6 Additional MVN Examples

An added nuance of the MVN distribution not present in the VN distribution is that of covariance, or rather what would be covariance in a multivariate normal distribution. The covariance for MVN distributions is far more complex than the parameters themselves and not of primary interest, like other moments for variety normal distributions. Figure S4 considers the variety corresponding to the polynomial $g = x^2 + y^2 - 1$ using $\boldsymbol{\Sigma} = \begin{bmatrix}\sigma_1 & 0 \\ 0 & \sigma_2\end{bmatrix}\begin{bmatrix}1 & \rho \\ \rho & 1\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 \\ 0 & \sigma_2\end{bmatrix}$ with $\sigma_1 = \sigma_2 = .2$ and $\rho = 0$, $.9$, and $-.9$. When $\rho = 0$, the variability about $\mathcal{V}(\boldsymbol{g})$ is isotropic; this is the "independence" case. When it is positive, variability is oriented towards regions of same-signed coordinates, and the opposite when it is negative. Figure S5, the one dimensional variety generated by the two polynomials $g_1 = x^2 + y^2 - 1$ and $g_2 = z$, illustrates the effect of "correlation" among the three variables, which can be very complex.

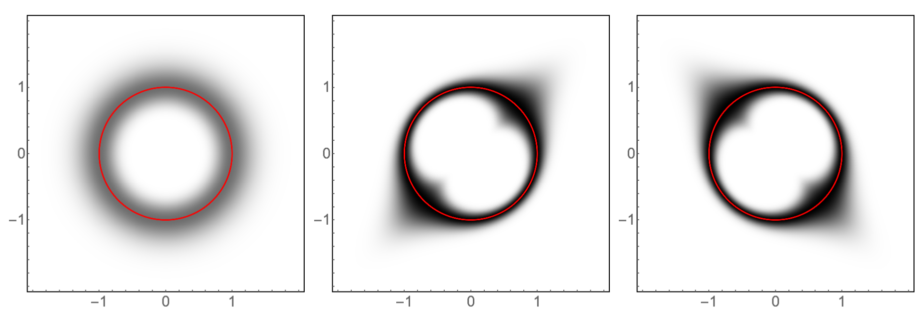

Figure S4: Multivariety normal densities with $g = x^2 + y^2 - 1$, $\sigma_1 = \sigma_2 = .2$, and $\rho = 0, .9, -.9$.

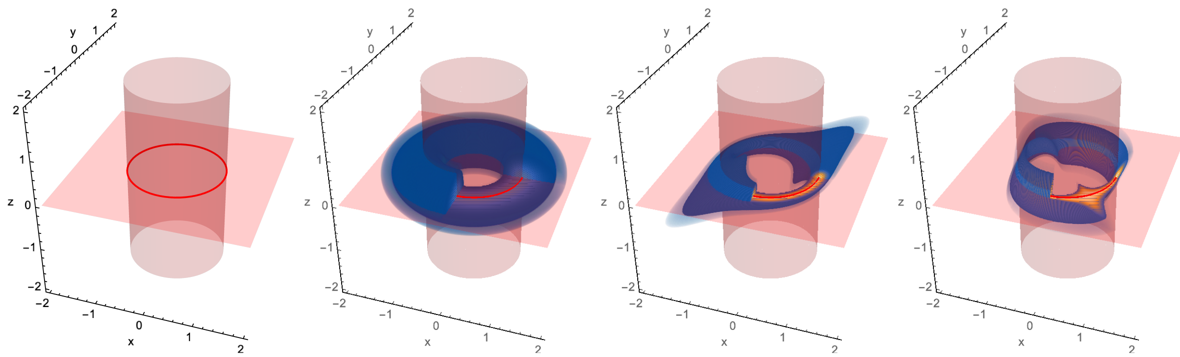

Figure S5: From left to right, the varieties of $g_1 = x^2 + y^2 - 1$ and $g_2 = z$ and their intersection $\mathcal{V}(g_1, g_2)$, the unit circle in the xy plane; and three 3D density plots of the MVN distribution with $\sigma_1 = \sigma_2 = \sigma_3 = .2$ and $\rho_{12} = \rho_{13} = \rho_{23} = 0$, $\rho_{12} = \rho_{13} = \rho_{23} = .9$, and $\rho_{12} = -.9$, $\rho_{13} = .5$, and $\rho_{23} = -.1$. Aesthetic scales are consistent across graphics, and the $x, z \geq 0$, $y \leq 0$ orthant has been removed to see inside the regions.

S7 HMC Derivation for the Variety Normal Distribution

This section provides the full derivation of Hamilton's equations for the variety normal distribution, summarized in Section 4 of the main text.

In HMC, one constructs a Hamiltonian $H(\boldsymbol{x}, \boldsymbol{p}) = U(\boldsymbol{x}) + K(\boldsymbol{p})$ where $\boldsymbol{x}$ is the position (the candidate draw), $\boldsymbol{p}$ is an auxiliary momentum vector, the potential energy is $U(\boldsymbol{x}) = -\log \widetilde{p}(\boldsymbol{x})$, and the kinetic energy is $K(\boldsymbol{p}) = \boldsymbol{p}'\boldsymbol{M}^{-1}\boldsymbol{p}/2$ with mass matrix $\boldsymbol{M}$. Hamilton's equations are

$$\frac{dx_i}{dt} = \frac{\partial}{\partial p_i} H(\boldsymbol{x}, \boldsymbol{p}) \quad \text{and} \quad \frac{dp_i}{dt} = -\frac{\partial}{\partial x_i} H(\boldsymbol{x}, \boldsymbol{p}).$$

For the variety normal with $n \geq m$, $\boldsymbol{J}_{\boldsymbol{x}}^{+} = \boldsymbol{J}_{\boldsymbol{x}}'(\boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{J}_{\boldsymbol{x}}')^{-1}$ and $p(\boldsymbol{x}|\boldsymbol{g}, \boldsymbol{\Sigma}) \propto \exp\left\{-\frac{1}{2}\overline{\boldsymbol{g}}(\boldsymbol{x}|\boldsymbol{B})'\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{g}}(\boldsymbol{x}|\boldsymbol{B})\right\}$ so

$$U(\boldsymbol{x}) \;=\; \frac{1}{2}\overline{\boldsymbol{g}}(\boldsymbol{x}|\boldsymbol{B})'\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{g}}(\boldsymbol{x}|\boldsymbol{B}) \;=\; \frac{1}{2}\boldsymbol{g}(\boldsymbol{x}|\boldsymbol{B})'\Big((\boldsymbol{J}_{\boldsymbol{x}}^{+})'\boldsymbol{\Sigma}^{-1}\boldsymbol{J}_{\boldsymbol{x}}^{+}\Big)\boldsymbol{g}(\boldsymbol{x}|\boldsymbol{B}).$$

If $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$, this further reduces to

$$U(\boldsymbol{x}) \;=\; \frac{1}{2\sigma^2}\boldsymbol{g}(\boldsymbol{x}|\boldsymbol{B})'(\boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{J}_{\boldsymbol{x}}')^{-1}\boldsymbol{g}(\boldsymbol{x}|\boldsymbol{B}). \tag{S2}$$

If $\boldsymbol{M} = \sigma_m^2\boldsymbol{I}_n$, Hamilton's equations become

$$\frac{dx_i}{dt} \;=\; \frac{\partial}{\partial p_i}K(\boldsymbol{p}) \;=\; \frac{p_i}{\sigma_m^2}, \tag{S3}$$

$$\frac{dp_i}{dt} \;=\; -\frac{\partial}{\partial x_i}U(\boldsymbol{x}) \;=\; -\frac{1}{2\sigma^2}\frac{\partial}{\partial x_i}\left\{\boldsymbol{g}(\boldsymbol{x}|\boldsymbol{B})'(\boldsymbol{J}_{\boldsymbol{x}}\boldsymbol{J}_{\boldsymbol{x}}')^{-1}\boldsymbol{g}(\boldsymbol{x}|\boldsymbol{B})\right\}. \tag{S4}$$

Since $\boldsymbol{g} \in \mathbb{R}[\boldsymbol{x}]^m$ is polynomial, the Jacobian $\boldsymbol{J}_{\boldsymbol{x}}$ is available in closed form and the gradient of U needed by the leapfrog integrator can be computed exactly. In practice, Stan's automatic differentiation handles this computation so the user need only specify $U(\boldsymbol{x})$ and let the sampler compute the required derivatives. The mass matrix $\boldsymbol{M}$ is typically estimated during Stan's warmup phase as a diagonal or dense matrix.

References

Tomáš Bajbar and Sönke Behrends. How fast do coercive polynomials grow? Technical report, Technical report, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2017.

Tomas Bajbar and Oliver Stein. Coercive polynomials and their newton polytopes. *SIAM Journal on Optimization*, 25(3):1542–1570, 2015.

Emil Horobet and Madeleine Weinstein. Offset hypersurfaces and persistent homology of algebraic varieties. *arXiv preprint arXiv:1803.07281*, 2018.