Draft version January 30, 2020
Typeset using LaTeX default style in AASTeX62

FITTING MODELS TO DATA WITH BOTH INTRINSIC AND EXTRINSIC UNCERTAINTIES IN TWO DIMENSIONS: EXTINCTION AND ABSORPTION MODELS AND THE LV RELATION

Adam S. Trotter,[1] Daniel E. Reichart,[1] and Nicholas C. Konz[1]

[1] *Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3255, USA*

ABSTRACT

Lorem ipsum dolor sit amet, omnesque eloquentiam at vel. Ad nostrud erroribus eum, purto integre an sea. Esse adipisci sensibus est id. Per homero semper civibus eu, mei modo petentium ei, elitr debitis sensibus ne nam. No eam accusam dissentiet.

Keywords: methods: statistical — dust, extinction — intergalactic medium — gamma-ray burst: general

1. INTRODUCTION

Before we can construct a generalized model for extinction by dust and absorption by gas for a GRB, or any point source of radiation, it is necessary to establish a solid statistical foundation. In Chapter **??**, we develop a new, very generally applicable statistic for fitting model distributions to data in two dimensions, where the data have both intrinsic uncertainties (i.e., error bars) and extrinsic scatter (i.e., sample variance) that is greater than can be attributed to the intrinsic uncertainties alone. A model distribution is described by the convolution of a (possibly asymmetric) 2D probability distribution that characterizes the extrinsic scatter, or "slop", in the data in both the x- and y-dimensions with a model curve $y_c(x;\vartheta_m)$, where $\{\vartheta_m\}$ is a set of M model parameters that describe the shape of the curve. The task in fitting a model distribution to a data set is to find the values of the parameters $\{\vartheta_m\}$, and of the parameters that characterize the slop, that maximize the joint probability, or likelihood function, of the model distribution and the data set. First, we derive a very general probabilistic formulation of the joint probability for the case where both the intrinsic uncertainties and the slop are normally distributed (§2). We then demonstrate that there is a choice of rotated coordinate systems over which the joint probability integrals may be evaluated, and that different choices of this rotated coordinate system yield fundamentally different statistics with fundamentally different properties, and yield fundamentally different predictions for any given model parameterization and data set (§2.3).

We show that one choice of rotated coordinates yields the traditional statistic advocated by **?**, hereafter D05, and another the statistic advocated by **?**, hereafter R01. The D05 statistic is *non-invertible*, i.e., it yields different model fits depending on whether one chooses to fit a model distribution to y vs. x or to x vs. y (§2.4.1). The R01 statistic *is* invertible, but as we will demonstrate, suffers from a fatal flaw in that it does not reduce to χ^2 in the 1D limiting case of zero extrinsic scatter, and zero intrinsic uncertainty in either the x- or y-dimensions (§2.4.4). We introduce a new statistic, TRK, that is both invertible and reduces to χ^2 in these 1D limiting cases. We demonstrate that the best-fit model distribution predicted by the TRK statistic, in the case of normally distributed intrinsic uncertainties and extrinsic scatter, is geometrically equivalent to the distribution that minimizes the sum of the squares of the radial distances of each data point centroid from the model curve, measured in units of the 1σ radius of the 2D convolved intrinsic and extrinsic Gaussian error ellipse (§2.3).

However, unlike the D05 statistic, the TRK statistic is not *scalable*, i.e., it yields different best-fit model distributions depending on the choice of basis for each coordinate axis (§2.4.6). We demonstrate that, in the limit of data that are entirely dominated by extrinsic scatter (i.e., zero intrinsic error bars), the predictions of TRK at extremes of scaling the x- and y-axes is equivalent to the predictions of D05 under an inversion of the x- and y-axes (§2.5). However, D05 is limited to this binary choice of inversion or non-inversion, while TRK is free to explore a continuous range of scales between the two limiting cases. Moreover, when the data are dominated by intrinsic rather than extrinsic uncertainty,

Corresponding author: Adam Trotter
astrotter@gmail.com

the range of scales that result in physically meaningful model distributions is significantly smaller, while the predictions of D05 are the same under inversion whether the data are dominated by intrinsic or extrinsic uncertainty. In the limit of data that are entirely dominated by intrinsic uncertainty (i.e., zero slop), this range of scales reduces to a single physically meaningful scale. We discuss these behaviors in the context of linear model distributions fit to various randomized data sets. We define a new correlation coefficient, $R^2_{\rm TRK}$ (§2.5.1), that is analogous to the well-known Pearson correlation coefficient R^2_{xy}, and describe an algorithm for arriving at an "optimum scale" for a fit with TRK to a given data set, which yields linear fits that fall approximately midway between the inverted and non-inverted D05 fits (§2.5.2). Lastly, we generalize this algorithm to non-linear fits (§2.5.3), and to fits to data with asymmetric intrinsic or extrinsic uncertainties (§2.7.2 and Appendices A & B).

We next apply the TRK statistic to fit empirical model distributions to the distribution of and correlations between observed interstellar extinction parameters in the Milky Way and Magellanic Clouds (§**??**), and to observed IGM Lyα forest flux deficits as a function of redshift in a sample of QSOs (§**??**). The fitted probability distributions of the secondary model parameters describing these distributions define a set of priors that, in turn, constrain the values of the parameters comprising a comprehensive model for sources of line-of-sight extinction due to dust and absorption due to gas of a GRB afterglow, or of any extragalactic point source of radiation. These model fits, and the GRB model fits themselves, are all obtained using the highly flexible genetic algorithm software package, *Galapagos* (§**??**; Foster & Reichart 2011, in preparation). Only when all significant sources of extinction and absorption are taken into account is it possible to accurately model the intrinsic emission of a GRB afterglow. Furthermore, the fitted values of extinction and absorption parameters are physically interesting in their own right, providing information about the amount and properties of source-frame dust and gas. By carefully modeling changes in extinction as a function of time, it may be possible to detect signatures of dust modification by the burst (at early times), and illumination of different populations of dust as the GRB jet expands laterally (at later times).

We then demonstrate a second application of the T19 statistic: characterizing the observed GRB luminosity-variability (LV) relation...

2. FITTING A MODEL TO DATA IN TWO DIMENSIONS

First, we consider the very general problem of fitting a model to a two-dimensional set of data points. Each data point may have *intrinsic scatter*, or *statistical uncertainty*, in both dimensions, which can vary from point to point. A data point's intrinsic scatter is commonly referred to as its *error bars*. The data set as a whole may also have *extrinsic scatter*, or *sample variance*, in both dimensions, i.e., scatter that is greater than can be accounted for by the error bars alone. Unlike statistical uncertainty, which is specified for each data point, extrinsic scatter is something that typically must be parameterized and fit to. Hence, it is part of the model. In this paper, we colloquially refer to this extrinsic scatter as *slop*.

Slop, or sample variance, is a catch-all quantification of our ignorance of all "higher-order" physical processes that superimpose variations on the overall trend of the data and its intrinsic scatter, that we either can not or choose not to model explicitly. For example, in fitting models to GRB afterglows, we typically model the intrinsic flux density F_ν as a series of smoothly broken power laws in time. But there may be lower-level, shorter-timescale variability superimposed on these smooth light curves, e.g., due to clumpiness in the circumburst medium or variability of the central engine. We typically model such variability to be normally distributed in $\log F_\nu$, and describe it with a single slop parameter, $\sigma_{\log F_\nu}$, which we fit to as a free model parameter. In the following formulation of the very general case of fitting models to data in two dimensions, we assume that slop is normally distributed in both dimensions, and can be described by two parameters (σ_x, σ_y) that are constant over the entire data set.[1]

In the following, we discuss three approaches for defining the likelihood function for this most general two-dimensional case: that of **?**, hereafter D05, that of **?**, hereafter R01, and a new approach, TRK (Trotter, Reichart & Kunz 2019, in preparation), which is the main topic of this section of the paper. All three statistics are based on the same fundamental formulation of the likelihood function in terms of the joint probability of the intrinsic and extrinsic probability distributions of the data set with the model, which we present below. We discuss the differences between the three statistics in §2.3.

[1] Of course, more complicated parameterizations of slop are possible, including non-normal distributions and distributions that vary across the data set. For example, there may be a component of unmodeled variability in a GRB afterglow that is normally distributed in linear flux space, due to radio scintillation in the interstellar medium along the line of sight. In that case, our model would include an additional slop parameter, σ_{F_ν}, and the unmodeled lower-level, shorter-timescale variability would be described by a convolution of linear- and log-normal probability distributions. Such refinements are introduced as needed (see, e.g., §§2.7 & **??**).

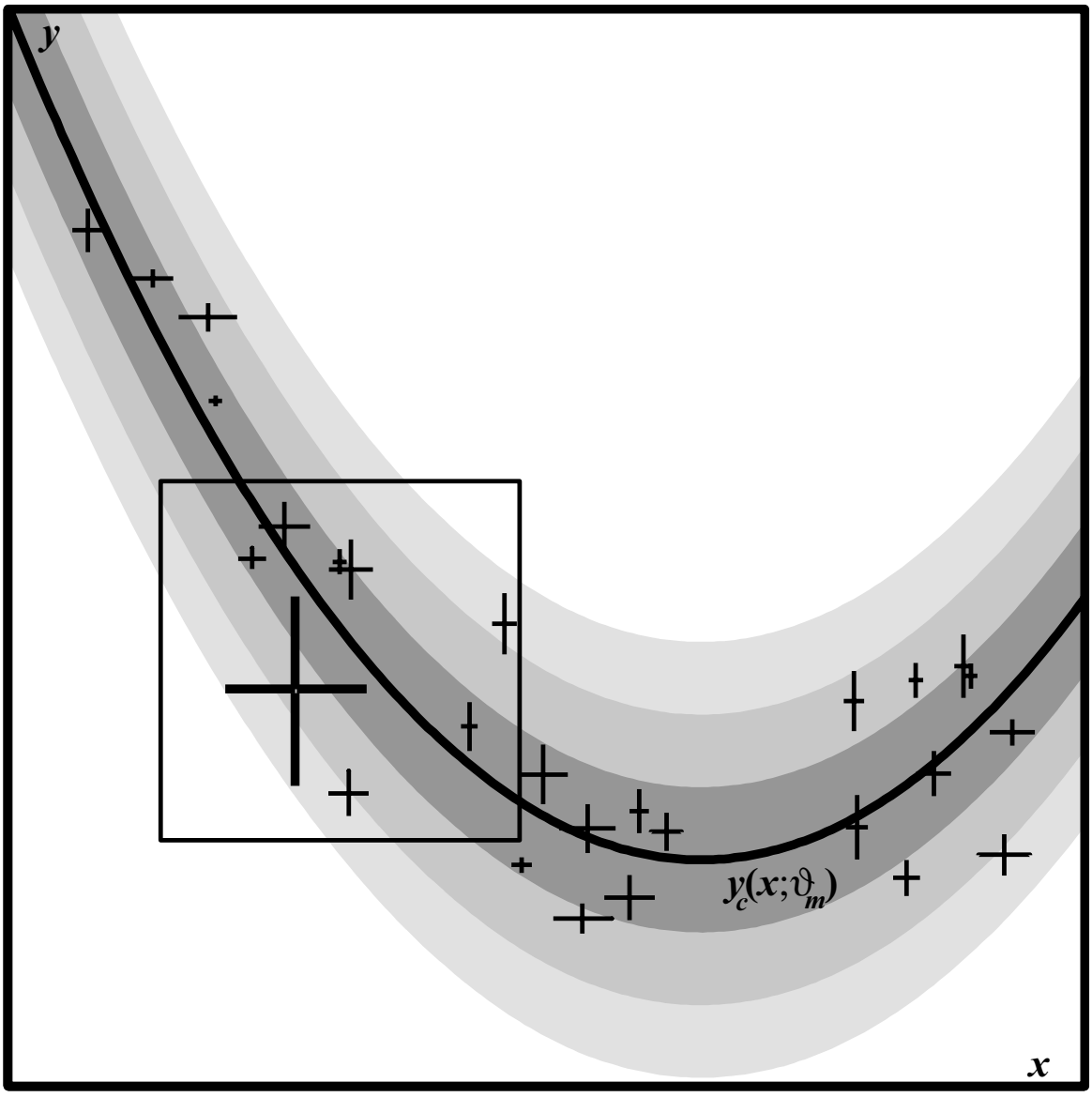


Figure 1. Illustration of a two-dimensional data set and model distribution. Inset box is expanded in Figure 2.

2.1. *Analytical Approximation of the Likelihood Function*

The following derivation follows closely that of R01, which formed the basis of much of the new investigation in this paper. We wish to quantify the fitness of a model to a two-dimensional set of N measurements. The data points have centroids at $\{x_n, y_n\}$, and normally distributed intrinsic scatter, or error bars, $\{\sigma_{x,n}, \sigma_{y,n}\}$. We describe the model itself not as a curve, but rather as a *model distribution*, which we treat mathematically as a continuous distribution of points, or a relative probability distribution, $g(x, y)$ along a curve $y_c(x; \vartheta_m)$, convolved with a two-dimensional Gaussian probability distribution function. The functional form of the curve is defined by a set of M parameters $\{\vartheta_m\}$. The width of the convolving Gaussian in each dimension is (σ_x, σ_y), and reflects the extrinsic scatter, or slop, in the data. In all that follows, we assume that slop in each dimension is constant across the entire data set. We also assume that both the intrinsic error bars and extrinsic slops in the x- and y-dimensions are uncorrelated. Figure 1 illustrates a generic two-dimensional data set and model distribution. The inset box is expanded in Figure 2, which illustrates the overlapping intrinsic probability distribution of a single data point with the slop-convolved model probability distribution.

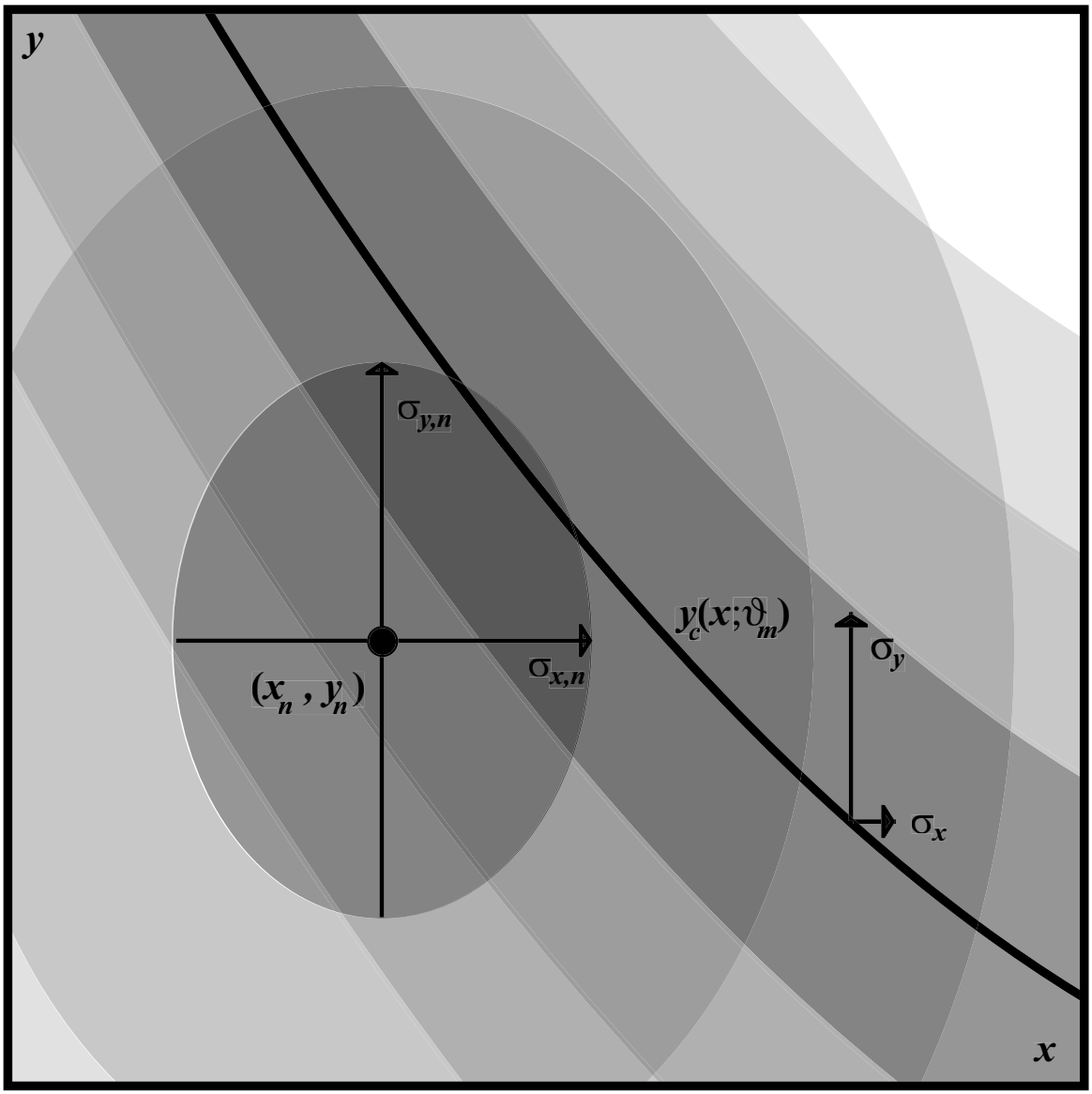


Figure 2. Expansion of the inset box of Figure 1, illustrating the intrinsic probability distribution of a single data point and the model distribution. The data point (x_n, y_n) has intrinsic error bars $(\sigma_{x,n}, \sigma_{y,n})$. The model distribution is the convolution of a probability density $g(x, y)$ along a curve $y_c(x; \vartheta_m)$ with a two-dimension Gaussian of width (σ_x, σ_y) that represents the extrinsic scatter, or slop, in the data set. Shading represents the 1, 2 and 3σ widths of the intrinsic probability distribution of the data point, and of the extrinsic slop-convolved model distribution.

In order to proceed, the effectively one-dimensional model curve must be expressed in terms of a one-dimensional delta function. But at this point, we are completely free to choose the direction in which that delta function is defined. Most generally, the probability density along the curve can be expressed as $g(x, y)\delta(v_n - v_{c,n}(u_n; \vartheta_m))$. Here, the coordinates (u_n, v_n) are related to (x, y) by a simple rotation of the axes, and $v_{c,n}(u_n; \vartheta_m)$ describes the curve $y_c(x; \vartheta_m)$ in this rotated coordinate system. Note that, in general, the choice of (u_n, v_n) may be *different for each data point*. As we will see, the choice of rotated coordinates (u_n, v_n) is critical in determining the properties of the statistic we derive; the three statistics D05, R01 and TRK differ from each other only in the choice of (u_n, v_n). Convolving the probability density along the model curve with the two-dimensional Gaussian representing extrinsic scatter in the

data, we obtain the *model distribution function*:

$$p_n^{\rm mod}(x',y'|\vartheta_m,\sigma_x,\sigma_y) = \int_{u_n}\int_{v_n} g(x,y)\delta(v_n - v_{c,n}(u_n;\vartheta_m))G(x',x,\sigma_x)G(y',y,\sigma_y)dv_n du_n\,, \tag{1}$$

where G denotes the normalized Gaussian function:

$$G(x',x,\sigma_x) \equiv \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left[-\frac{1}{2}\frac{(x'-x)^2}{\sigma_x^2}\right], \tag{2}$$

and where the integrals are all definite integrals from $-\infty$ to $+\infty$. For now, we continue to express the Gaussians, and $g(x,y)$ in terms of unrotated coordinates (x,y).

However, in reality, the model distribution is not what is observed. Let $f(x',y')$ represent the *data selection function*, which describes the efficiency at which the actual data set in question samples the model distribution. Then, the *observed* model distribution is:

$$\begin{aligned} p_n^{\rm obs}(x',y'|\vartheta_m,\sigma_x,\sigma_y) = \int_{u_n}\int_{v_n} & f(x',y')g(x,y)\delta(v_n - v_{c,n}(u_n;\vartheta_m))\times \\ & G(x',x,\sigma_x)G(y',y,\sigma_y)dv_n du_n\,. \end{aligned} \tag{3}$$

We now consider the *joint probability* of a single data point with the observed model distribution of Equation 3. The intrinsic probability distribution of data point n is given by the two-dimensional Gaussian:

$$p_n^{\rm int}(x',y'|x_n,y_n,\sigma_{x,n},\sigma_{y,n}) = G_n(x',x_n,\sigma_{x,n})G_n(y',y_n,\sigma_{y,n})\,. \tag{4}$$

The joint probability of data point n with the model distribution is then obtained by integrating the product of $p_n^{\rm obs}$ and $p_n^{\rm int}$ over the (x',y') plane:

$$\begin{aligned} & p_n(\vartheta_m,\sigma_x,\sigma_y|x_n,y_n,\sigma_{x,n},\sigma_{y,n}) = \\ & \int_{x'}\int_{y'}\int_{u_n}\int_{v_n} f(x',y')g(x,y)\delta(v_n - v_{c,n}(u_n;\vartheta_m))\times \\ & \quad G(x',x,\sigma_x)G(y',y,\sigma_y)G_n(x',x_n,\sigma_{x,n})G_n(y',y_n,\sigma_{y,n})dv_n du_n dy' dx'\,. \end{aligned} \tag{5}$$

To achieve a computationally practical statistic, an analytic result is desirable. To this end, we now make a number of important simplifying assumptions. First, we assume that the selection function $f(x',y')$ varies slowly with respect to the scale of the intrinsic uncertainties $(\sigma_{x,n},\sigma_{y,n})$ and extrinsic scatter (σ_x,σ_y), and so can be approximated as a constant in the (x',y') integral. In this case, the integral can be evaluated analytically, and the joint probability becomes:

$$\begin{aligned} & p_n(\vartheta_m,\sigma_x,\sigma_y|x_n,y_n,\sigma_{x,n},\sigma_{y,n}) \approx \\ & f(x_n,y_n)\int_{u_n}\int_{v_n} g(x,y)\delta(v_n - v_{c,n}(u_n;\vartheta_m))G_n(x,x_n,\Sigma_{x,n})G_n(y,y_n,\Sigma_{y,n})dv_n du_n\,, \end{aligned} \tag{6}$$

where $(\Sigma_{x,n},\Sigma_{y,n})$ are the quadrature sums of the intrinsic and extrinsic scatters:

$$\begin{aligned} \Sigma_{x,n} &\equiv \left(\sigma_{x,n}^2 + \sigma_x^2\right)^{1/2} \\ \Sigma_{y,n} &\equiv \left(\sigma_{y,n}^2 + \sigma_y^2\right)^{1/2}\,. \end{aligned} \tag{7}$$

In other words, the joint probability of data point n with the model distribution is proportional to the integral of the effectively one-dimensional probability density along the model curve through a two-dimensional *convolved* Gaussian whose widths are the quadrature sums of the intrinsic uncertainties and the slops in each direction (i.e., Figure 2 is equivalent to Figure 3).

We now make two further important simplifying assumptions: (1) the intrinsic probability density along the model curve $g(x,y)$ varies slowly with respect to the scale of the convolved error ellipse $(\Sigma_{x,n},\Sigma_{y,n})$; and (2) the model curve $y_c(x;\vartheta_m)$ (or $v_{c,n}(u_n;\vartheta_m)$) deviates slowly from linearity over this same scale. The first assumption allows us to pull

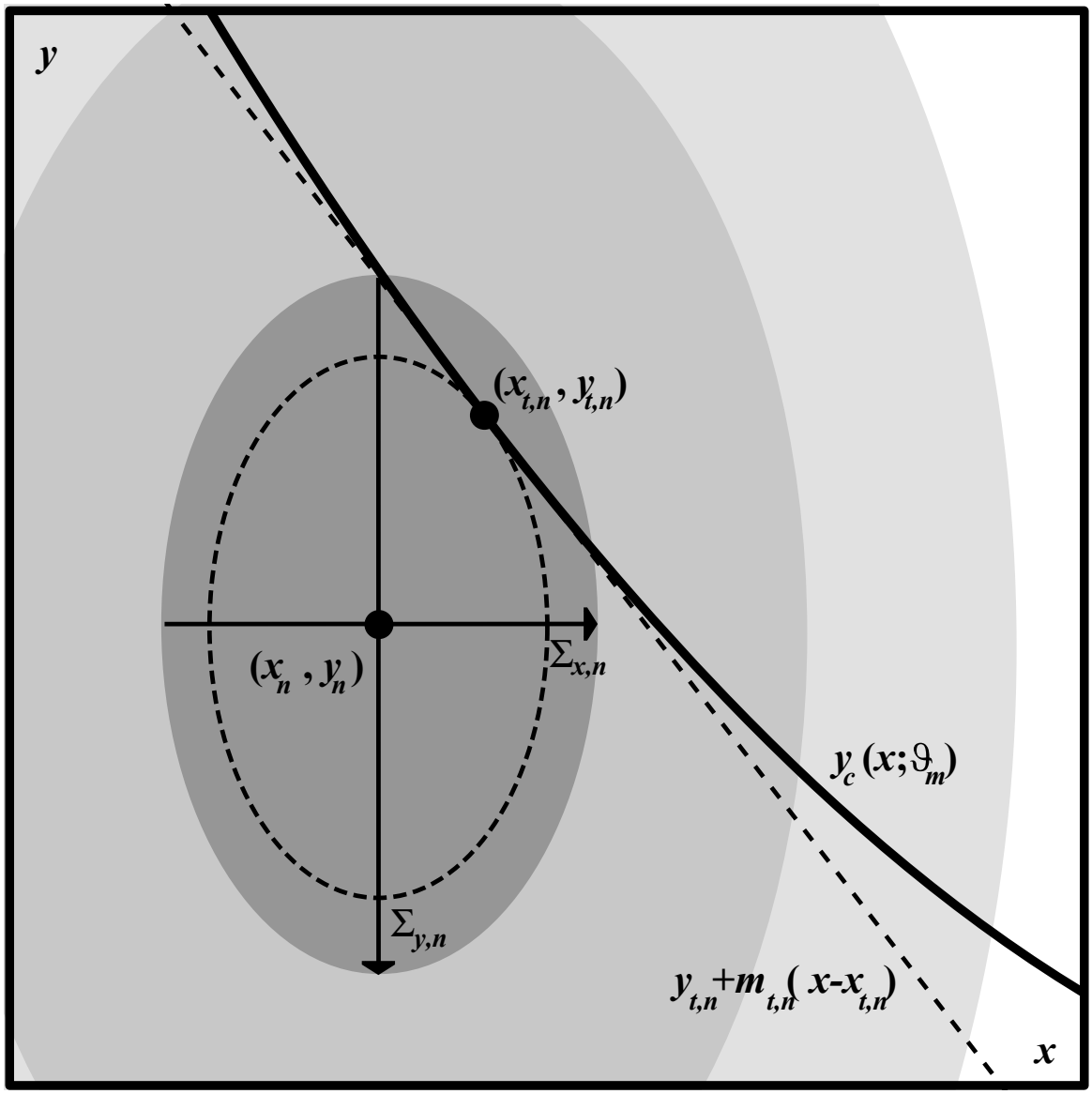


Figure 3. Illustration of the convolved error ellipse and tangent line approximation. The data point (x_n, y_n) is assigned convolved error bars $(\Sigma_{x,n}, \Sigma_{y,n})$. The model distribution is replaced by the curve $y_c(x)$. For non-linear models, this curve is approximated as a line $y_{t,n}(x)$ tangent to the convolved error ellipse at point $(x_{t,n}, y_{t,n})$.

$g(x, y)$ out of the (u, v) integral, because it is effectively a constant across the integral. The second assumption requires some finesse: We approximate the model curve $y_c(x; \vartheta_m)$ as a line $y_{t,n}(x)$ passing through the point $(x_{t,n}, y_{t,n})$ where the curve is tangent to the convolved error ellipse, with a slope $m_{t,n}$ equal to the slope of the curve at that tangent point (Figure 3):

$$y_c(x) \approx y_{t,n} + m_{t,n}(x - x_{t,n})\,. \tag{8}$$

To find the tangent point, consider an arbitrary point x along the model curve $y_c(x; \vartheta_m)$. If the curve at this point is a tangent point $(x_{t,n}, y_{t,n})$ to the error ellipse, then:

$$\frac{(y_c(x;\vartheta_m) - y_n)^2}{\Sigma_{y,n}^2} + \frac{(x - x_n)^2}{\Sigma_{x,n}^2} = \frac{(y_{t,n} - y_n)^2}{\Sigma_{y,n}^2} + \frac{(x_{t,n} - x_n)^2}{\Sigma_{x,n}^2}\,. \tag{9}$$

Note that the right-hand side of this equation is a constant. If we take the derivative of Equation 9, we find that the condition for $(x, y_c(x;\vartheta_m))$ to be a tangent point is equivalent to requiring:

$$\frac{d}{dx}\left(\frac{(y_c(x;\vartheta_m)-y_n)^2}{\Sigma_{y,n}^2}+\frac{(x-x_n)^2}{\Sigma_{x,n}^2}\right)=0\,. \tag{10}$$

In other words, the tangent point is equivalent to the point on the curve $y_c(x;\vartheta_m)$ that minimizes the radial distance to the error ellipse centroid, as measured in units of $(\Sigma_{x,n}, \Sigma_{y,n})$. Carrying out the derivative, we see that the tangent point $x_{t,n}$ is the value of x for which:

$$(y_c(x)-y_n)\frac{dy_c(x;\vartheta_m)}{dx}\Sigma_{x,n}^2+(x-x_n)\Sigma_{y,n}^2=0\,. \tag{11}$$

For a linear model, i.e., $y_c(x;\vartheta_m) = mx + b$, the tangent point coordinates can be obtained from Equation 11 analytically:

$$\begin{aligned} x_{t,n} &= \frac{m\Sigma_{x,n}^2\,(y_n-b)+\Sigma_{y,n}^2 x_n}{m^2\Sigma_{x,n}^2+\Sigma_{y,n}^2} \\ y_{t,n} &= mx_{t,n}+b\,. \end{aligned} \tag{12}$$

For non-linear model curves, the tangent point must be found numerically. Depending on the functional form of $y_c(x;\vartheta_m)$, this point can be bracketed and found by any number of numerical root-finding algorithms.[2] We use the Two-Point Newton Method of Tiruneh et al. (2013), which has proven to be both efficient and useful for a number of non-linear models. We note that in the case of non-monotonic curves, there may be two or more such tangent points for each data point; in those cases, we choose the one that maximizes the joint probability.

Finally, under all these simplifying assumptions, the joint probability of data point n with the model distribution of Equation 6 can be evaluated analytically by integrating over v_n:

$$\begin{aligned} &p_n(\vartheta_m,\sigma_x,\sigma_y|x_n,y_n,\sigma_{x,n},\sigma_{y,n}) \\ &\approx f(x_n,y_n)g(x_n,y_n)\int_{-\infty}^{\infty}G_n(x,x_n,\Sigma_{x,n})G_n(y_c(x;\vartheta_m),y_n,\Sigma_{y,n})du_n\,. \end{aligned} \tag{13}$$

Re-expressing $du_n = \frac{du_n}{dx}dx$, substituting $y_c(x;\vartheta_m) \approx y_{t,n} + m_{t,n}(x - x_{t,n})$, and integrating over x, we obtain:

$$\begin{aligned} &p_n(\vartheta_m,\sigma_x,\sigma_y|x_n,y_n,\sigma_{x,n},\sigma_{y,n}) \\ &\approx f(x_n,y_n)g(x_n,y_n)\frac{du_n}{dx}G_n\left[y_n, y_{t,n}+m_{t,n}(x_n-x_{t,n}), \sqrt{m_{t,n}^2\Sigma_{x,n}^2+\Sigma_{y,n}^2}\right]\,. \end{aligned} \tag{14}$$

As a reminder, (u_n, v_n) is the rotated coordinate system in which we define the fundamentally one-dimensional curve $\delta(v_n - v_{c,n}(u_n;\vartheta_m))$ that underlies the observed model distribution (Equation 1). In this derivation of the joint probability of a data point with the model distribution, there is no a priori reason that one choice of u_n should be preferable to another, or even that u_n should be the same for all data points. But, as we shall see in §2.4, *different choices of* $\mathrm{u_n}$ *yield different statistics*, entirely due to the factor $\frac{du_n}{dx}$ in the joint probability. And each resulting statistic exhibits different behaviors and yields different best-fit model distributions for a given data set.

The likelihood function $\mathcal{L}$ is defined as the joint probability of the model distribution with all the data points, and is simply the product of the individual joint probabilities:

$$\mathcal{L}\equiv\prod_{n=1}^{N}p_n(\vartheta_m,\sigma_x,\sigma_y|x_n,y_n,\sigma_{x,n},\sigma_{y,n})\,. \tag{15}$$

In practice, we find the best-fit model parameters by minimizing $-2\ln\mathcal{L}$:

$$-2\ln\mathcal{L}=-2\sum_{n=1}^{N}\ln p_n(\vartheta_m,\sigma_x,\sigma_y|x_n,y_n,\sigma_{x,n},\sigma_{y,n})$$

[2] R01 presents an alternative, iterative algorithm for determining the tangent point. We have found that this algorithm, though efficient when it succeeds, often fails to converge, especially for non-monotonic model curves.

$$= \sum_{n=1}^{N} \frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} - 2 \sum_{n=1}^{N} \ln \left(\frac{du_n}{dx} \frac{1}{\sqrt{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \right) + C \,, \tag{16}$$

where C is an (undetermined) constant for a given data set:

$$C = N \ln(2\pi) - 2 \sum_{n=1}^{N} \ln f(x_n, y_n) g(x_n, y_n) \,. \tag{17}$$

We choose to define the fitness in terms of $-2 \ln \mathcal{L}$ because it is analogous to χ^2 in certain simplifying cases. For instance, if there are no error bars in the x-direction, no slop in either direction, and $\frac{du_n}{dx} = 1$ for all points[3], then:

$$-2 \ln \mathcal{L} = \sum_{n=1}^{N} \frac{[y_c(x_n) - y_n]^2}{\sigma_{y,n}^2} + C \equiv \chi^2 + C \,. \tag{18}$$

In *any* case, the additive constant C in this formulation does not affect the relative fitness $-2 \ln \mathcal{L} - (-2 \ln \mathcal{L})_{\min}$ of a model with respect to the best fit; in this simple one-dimensional case, the relative fitness reduces to $\Delta\chi^2$.

2.2. *Summary of Simplifying Assumptions*

Before proceeding to discuss the different statistics that result from different choices of u_n, we summarize the simplifying assumptions that are common to all three in the analysis so far:

- Intrinsic scatter (error bars) in both dimensions is uncorrelated and normally distributed.
- Extrinsic scatter (slop) in both dimensions is uncorrelated and normally distributed.
- The selection function $f(x, y)$ and model probability density $g(x, y)$ vary slowly with respect to both the intrinsic and extrinsic scatter, and so can be approximated as constants over these scales.
- Non-linear model curves vary slowly with respect to the intrinsic and extrinsic scatter, and so can be approximated by lines tangent to convolved error ellipses over these scales.

Finally, consider the example of Figure 4. It is worth noting that although the large error bars of the data point in bold violate the latter two simplifying assumptions, its ability to help constrain the model parameters is clearly negligible, compared to the other data points. Consequently, in such cases the statistics we present below can still be taken as valid.

2.3. *Three Statistics: D05, R01 and TRK*

As hinted at above, the choice of path integral element du_n is, in fact, the only choice that differentiates the three statistics we explore in this paper. As we will see, each statistic has its own strengths and its own weaknesses. There may well be no two-dimensional statistic of this probabilistic form that is ideal for all model fitting cases. In this section, we describe the basic properties, including the functional form of the likelihood function, of three statistics. In §2.4 we discuss, compare and contrast their behavior in some detail for various special cases and coordinate transformations.

D05 presents a statistic for which du_n is simply dx, so that $\frac{du_n}{dx} = 1$ for all data points. In this case, the likelihood has the form:

$$\mathcal{L}^{\mathrm{D05}} \propto \prod_{n=1}^{N} \frac{1}{\sqrt{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \exp \left\{ -\frac{1}{2} \frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right\}$$

[3] As we will show in §§2.3 & 2.4.4, $\frac{du_n}{dx} = 1$ for all points for the D05 statistic, and $\frac{du_n}{dx} \to 1$ for the TRK statistic in this one-dimensional limiting case of no error bars or slop in the x-direction.

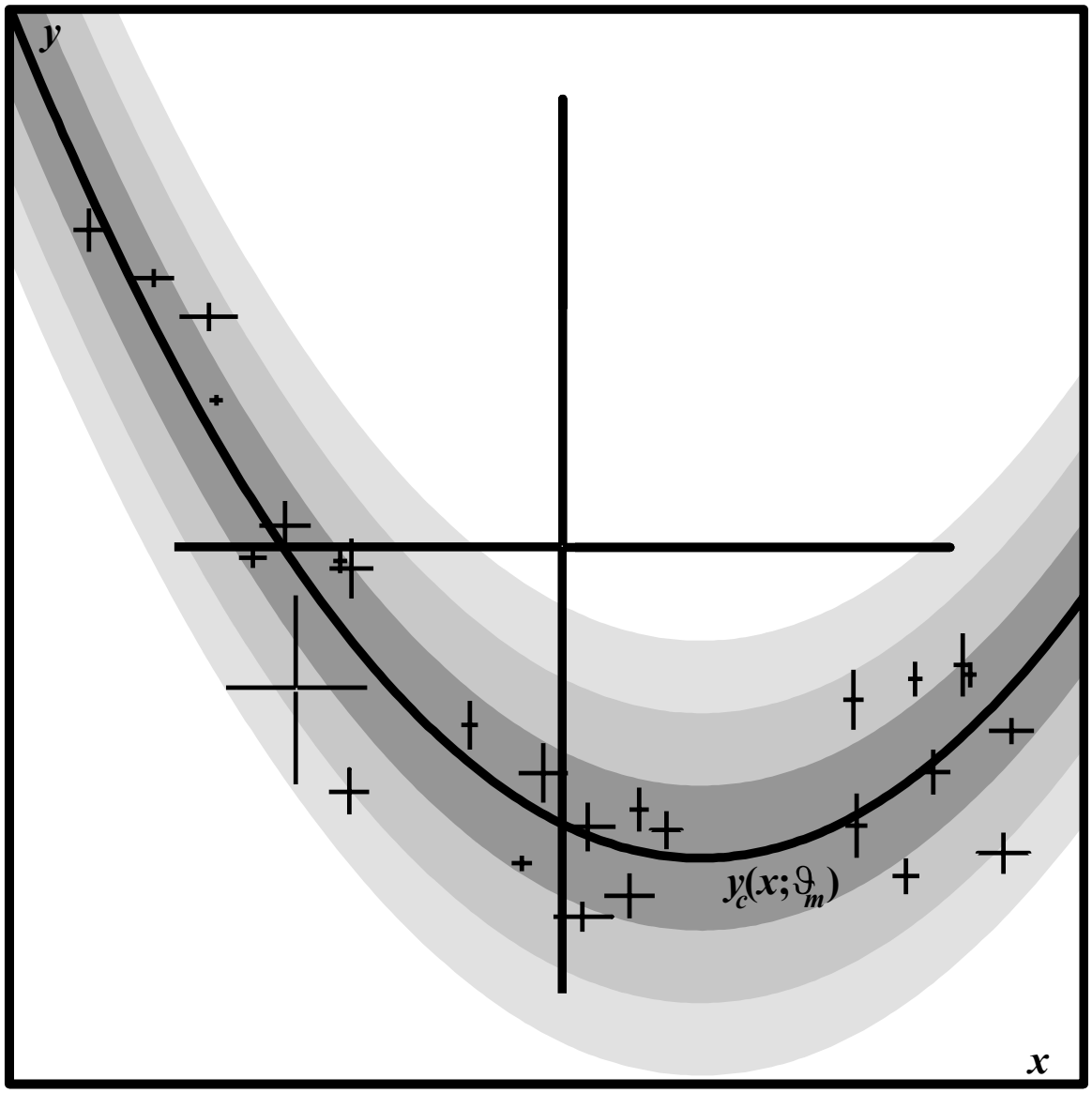


Figure 4. Illustration of a two-dimensional data set and model distribution, with a single anomalous data point (in bold).

$$-2\ln\mathcal{L}^{\mathrm{D05}} = \sum_{n=1}^{N} \frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2} + \sum_{n=1}^{N} \ln(m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2) + C\,. \tag{19}$$

The D05 statistic can be seen to be analogous to a one-dimensional χ^2 statistic in y, where the difference between the model and the data point is the difference between the tangent line at $x = x_n$ and y_n, and where the 1σ uncertainty in the convolved data point is replaced by the quadrature sum of $\Sigma_{y,n}$ and $\Sigma_{x,n}$ projected into the y-direction using the slope $m_{t,n}$. D05 differs from a traditional χ^2 statistic in that the denominator of the argument of the exponential, and the prefactor of the exponential are themselves functions of the slops (σ_x, σ_y), which are treated as free model parameters.

It should be noted that a common approach to fitting models to data with error bars in both dimensions, and with unmodeled extrinsic scatter, is to use a one-dimensional χ^2 statistic in y, with the intrinsic $\sigma_{x,n}$ error bars

projected into the y-direction, as here, but to ignore the prefactor and instead approximate the slop either by adding an additional uncertainty σ_y to the intrinsic error bars in quadrature, or even by multiplying the intrinsic error bars by some constant, and iterating until the reduced $\chi^2 = 1$. In general, however, this is not correct, even in the one-dimensional limiting case where error bars and slop are entirely in the y-direction. It amounts to holding σ_y fixed at its best-fit value, and is unnecessarily restrictive. Rather, since the exponential term in the likelihood of Equation 19 decreases as the slops approach zero, while the prefactor term decreases as the slops approach infinity, maximizing the likelihood with the slops as free parameters naturally constrains (σ_x, σ_y), without unnecessarily restricting them. In the one-dimensional limit, we refer to this more proper approach as a "χ^2-like" statistic.

As we will discuss in §2.4.1, D05 is *non-invertible*: a fit to x vs. y will, in general, yield a different model curve than a fit to y vs. x. R01 presented a first attempt at an *invertible statistic*, that is, a statistic that is invariant under an exchange of the x- and y-axes. The R01 statistic is formulated using $du_n = ds_n$, where $ds_n = (dx^2 + dy^2)^{1/2}$ is parallel to the tangent line, with slope $m_{t,n}$. The net effect is to add a multiplicative prefactor $\frac{du_n}{dx} = \sqrt{1 + m_{t,n}^2}$ to the joint probability:

$$\begin{aligned}
\mathcal{L}^{\mathrm{R01}} &\propto \prod_{n=1}^{N} \sqrt{\frac{1 + m_{t,n}^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \exp\left\{ -\frac{1}{2} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n}) \right]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right\} \\
-2 \ln \mathcal{L}^{\mathrm{R01}} &= \sum_{n=1}^{N} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n}) \right]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \\
&\quad - \sum_{n=1}^{N} \ln \left(\frac{1 + m_{t,n}^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right) + C \,.
\end{aligned} \tag{20}$$

It can be easily shown that $\mathcal{L}^{\mathrm{R01}}$ is analytically invariant under inversion of the x- and y-axes (see §2.4.1). However, R01 does not reduce to a one-dimensional χ^2-like statistic in the limit $\Sigma_{x,n} \to 0$ (see §2.4.4). One can also show that R01 is not invariant under multiplicative scaling transformations of either axis (see §2.4.6). The discovery of these limitations in R01 prompted us to explore other possible forms of the likelihood that preserve invertibility.

We present a new statistic, TRK, that is both invertible and reduces to a χ^2-like statistic in the one-dimensional limit. We propose the following functional form of the likelihood:

$$\begin{aligned}
\mathcal{L}^{\mathrm{TRK}} &\propto \prod_{n=1}^{N} \sqrt{\frac{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}{m_{t,n}^2 \Sigma_{x,n}^4 + \Sigma_{y,n}^4}} \exp\left\{ -\frac{1}{2} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n}) \right]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right\} \\
-2 \ln \mathcal{L}^{\mathrm{TRK}} &= \sum_{n=1}^{N} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n}) \right]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \\
&\quad - \sum_{n=1}^{N} \ln \left(\frac{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}{m_{t,n}^2 \Sigma_{x,n}^4 + \Sigma_{y,n}^4} \right) + C \,.
\end{aligned} \tag{21}$$

Geometrically, TRK is equivalent to choosing du_n to be perpendicular to the line segment connecting the data point centroid (x_n, y_n) and the tangent point of the curve $(x_{t,n}, y_{t,n})$ (that is, the delta function $\delta(v_n - v_{c,n}(u_n))$ is defined along the axis that connects the centroid and the tangent point). As with R01, it can easily be shown that this form of $\mathcal{L}^{\mathrm{TRK}}$ is analytically invertible (see §2.4.1). TRK reduces to a χ^2-like one-dimensional statistic in both limits $\Sigma_{x,n} \to 0$ (in which case it is equivalent to D05) and $\Sigma_{y,n} \to 0$ (see §2.4.4). Furthermore, in the case $\Sigma_{x,n} = \Sigma_{y,n}$, TRK is equivalent to R01. Most fortuitously, we find that *TRK is analogous to a one-dimensional χ^2-like statistic measured in the direction of the tangent point*, which we will now derive.

Figure 5 illustrates the geometry of the TRK statistic. The convolved error ellipse of a given point has a centroid at point $O \equiv (x_n, y_n)$, with 1σ widths $(\Sigma_{x,n}, \Sigma_{y,n})$. The tangent point of the model curve $y_c(x; \vartheta_m)$ is at point $T \equiv (x_{t,n}, y_{t,n})$. As described above, we approximate the curve $y_c(x; \vartheta_m)$ as a line through that point, with slope $m_{t,n} = \tan \theta_{t,n}$.

For TRK, we define the rotated coordinate system (u_n, v_n) such that u_n is perpendicular to the segment connecting the ellipse centroid and the tangent point. The u_n-axis is rotated by an angle $\psi_{t,n}$ with respect to the x-axis, and so differs from D05. The u_n-axis describes an angle $\psi_{t,n} - \theta_{t,n}$ with respect to the tangent line, and so differs from R01.

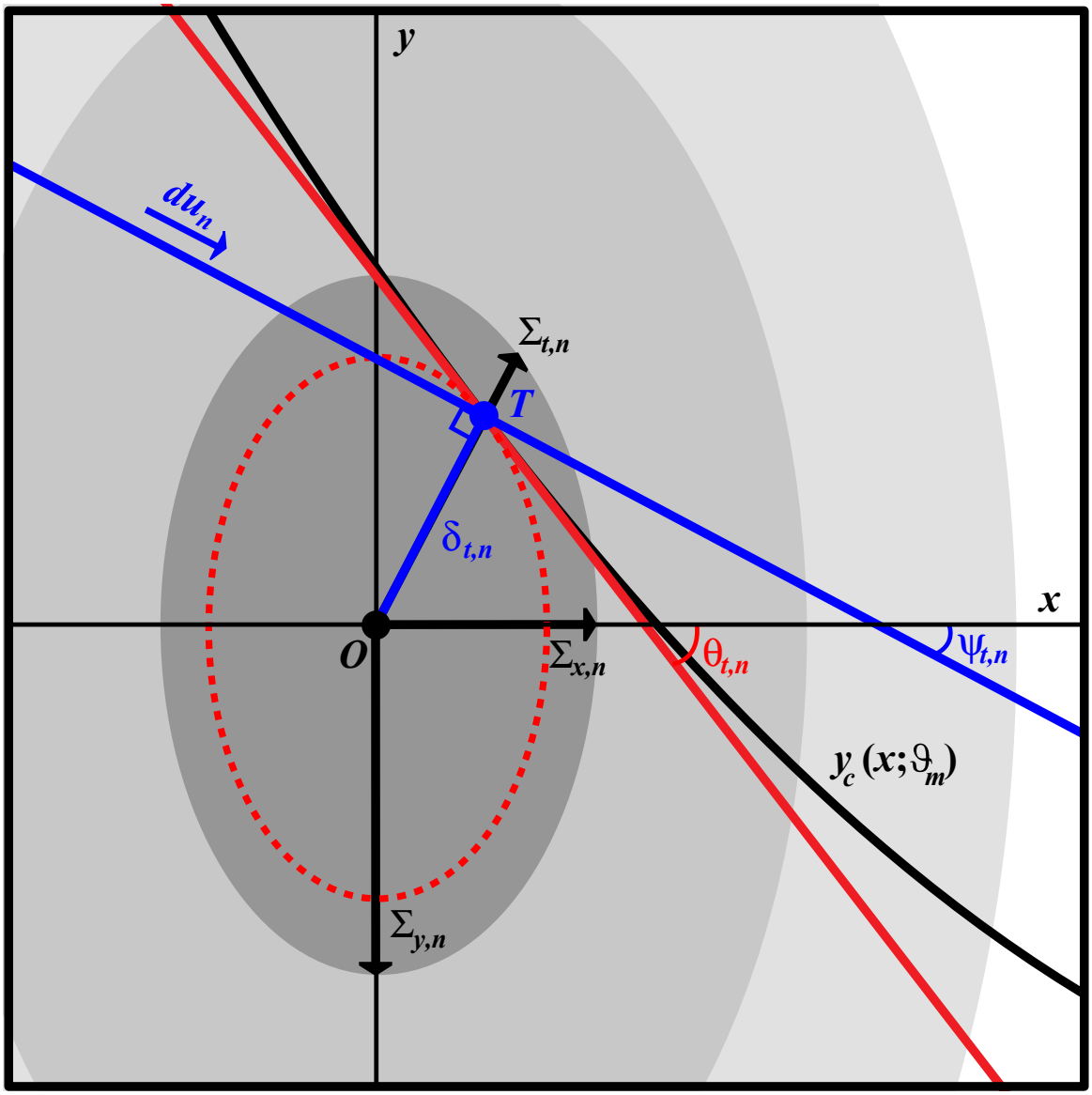


Figure 5. Illustration of the geometry of the TRK statistic. Data point centroid is at (x_n, y_n) (point O), with convolved widths $(\Sigma_{x,n}, \Sigma_{y,n})$. Model curve $y_c(x;\vartheta_m)$ is tangent to the convolved error ellipse at tangent point $(x_{t,n}, y_{t,n})$ (point T). Red line is the linear approximation of the model curve, with slope $m_{t,n} = \tan\theta_{t,n}$. Blue line indicates the direction du_n of path integration for the TRK statistic, perpendicular to the segment $\overline{OT}$. TRK is geometrically equivalent to a 1D χ^2-like statistic measured in the direction of the tangent point (see Equation 27).

It is not difficult to show that the radial distance of the tangent point from the centroid of the ellipse is:

$$\delta_{t,n} \equiv \overline{OT} = [y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})] \frac{\sqrt{m_{t,n}^2 \Sigma_{x,n}^4 + \Sigma_{y,n}^4}}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}\,, \tag{22}$$

and that the 1σ radius of the convolved error ellipse in the direction of the tangent point is:

$$\Sigma_{t,n} = \frac{\Sigma_{x,n}\Sigma_{y,n}}{\sqrt{\Sigma_{x,n}^2 \cos^2\psi_{t,n} + \Sigma_{y,n}^2 \sin^2\psi_{t,n}}}\,, \tag{23}$$

where

$$\cos^2\psi_{t,n} = \left(\frac{y_{t,n} - y_n}{\delta_{t,n}}\right)^2$$

$$\sin^2\psi_{t,n} = \left(\frac{x_{t,n}-x_n}{\delta_{t,n}}\right)^2 . \tag{24}$$

Substituting for $(x_{t,n}, y_{t,n})$ from Equation 12, we find that:

$$\Sigma_{t,n} = \sqrt{\frac{m_{t,n}^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}} , \tag{25}$$

and

$$\frac{\delta_{t,n}}{\Sigma_{t,n}} = \frac{y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})}{\sqrt{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}} . \tag{26}$$

Thus, we see that the expression for $\mathcal{L}^{\mathrm{TRK}}$ in Equation 21 can equivalently be expressed as:

$$\begin{aligned}\mathcal{L}^{\mathrm{TRK}} &\propto \prod_{n=1}^{N}\frac{1}{\Sigma_{t,n}}\exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\Sigma_{t,n}}\right)^2\right] \\ -2\ln\mathcal{L}^{\mathrm{TRK}} &= \sum_{n=1}^{N}\frac{\delta_{t,n}^2}{\Sigma_{t,n}^2} + 2\sum_{n=1}^{N}\ln\Sigma_{t,n} + C .\end{aligned} \tag{27}$$

In other words, *TRK is analogous to a one-dimensional χ^2-like statistic, measured in the direction of the tangent point to the convolved error ellipse of each data point.* For future reference, we can also express the D05 likelihood of Equation 19 in terms of $\Sigma_{t,n}$; note that the argument of the exponential is equivalent; however, the prefactor differs considerably from the intuitive form of Equation 27:

$$\mathcal{L}^{\mathrm{D05}} \propto \prod_{n=1}^{N}\left(\frac{\sqrt{m_{t,n}^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}\right)\frac{1}{\Sigma_{t,n}}\exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\Sigma_{t,n}}\right)^2\right] . \tag{28}$$

2.4. *Comparison of the D05, R01 and TRK Statistics*

As a reminder, the different statistics that we are considering here differ only in the choice of the rotated coordinate system (u_n, v_n) in which we define the delta function that traces the model curve $y_c(x;\vartheta_m)$, through the factor $\frac{du_n}{dx}$ in the likelihood function (Equation 16). In this section, we consider the choices of $\frac{du_n}{dx}$ that yield the D05, R01 and TRK statistics (Equations 19, 20 & 21), and evaluate their likelihood functions for: invertibility (§2.4.1); behavior in one-dimensional limiting cases (§2.4.4); the "two-point test" (§2.4.5); scalability (§2.4.6); and degeneracy of slop in linear fits (§2.4.7). We summarize the strengths and weaknesses of the three statistics in Table 1 (§2.4.8).

2.4.1. *Invertibility: The Motivation for R01 and TRK*

One desirable property of a 2D statistic is that it be *invertible*; that is, if the best-fit model curve to y vs. x is $y_c(x)$, the best-fit model curve to x vs. y should be $x_c(y) = y_c^{-1}(x)$. In the case of a linear fit, if the slope of the best-fit line to y vs. x is m_{xy}, a fit to x vs. y with an invertible statistic should give a best-fit slope $m_{yx} = 1/m_{xy}$. For non-linear model curves, $x_c(y)$ may or may not be an analytic function, and may have to be defined piece-wise or numerically; still, we expect, for an invertible statistic, that the slopes of the curves at each tangent point $m_{t,n} \to 1/m_{t,n}$ in the inverted space.

Most generally, for an invertible statistic, the likelihood should be analytically invariant under the substitutions $x_n \leftrightarrow y_n$, $x_{t,n} \leftrightarrow y_{t,n}$, $\Sigma_{x,n} \leftrightarrow \Sigma_{y,n}$, and $m_{t,n} \leftrightarrow 1/m_{t,n}$. If we perform these substitutions for the D05 joint probability (Equation 19), we find the inverted D05* likelihood:

$$\begin{aligned}\mathcal{L}^{\mathrm{D05*}} &\propto \prod_{n=1}^{N}\frac{1}{\sqrt{\Sigma_{y,n}^2/m_{t,n}^2 + \Sigma_{x,n}^2}}\exp\left\{-\frac{1}{2}\frac{[x_n - x_{t,n} - (y_n - y_{t,n})/m_{t,n}]^2}{\Sigma_{y,n}^2/m_{t,n}^2 + \Sigma_{x,n}^2}\right\} \\ &= \prod_{n=1}^{N}\frac{m_{t,n}}{\sqrt{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}}\exp\left\{-\frac{1}{2}\frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}\right\} \\ &= \left(\prod_{n=1}^{N} m_{t,n}\right)\mathcal{L}^{\mathrm{D05}} ,\end{aligned} \tag{29}$$

that is, $\mathcal{L}^{\mathrm{D05}*} \neq \mathcal{L}^{\mathrm{D05}}$. Note that in this form of the inverted D05* likelihood, all quantities are still defined in the original, uninverted y vs. x space; the best-fit curve obtained by maximizing $\mathcal{L}^{\mathrm{D05}*}$ should be plotted in the original space. Note also that the argument of the exponential, which is the same for all three statistics (D05, R01 and TRK), is analytically invertible under these substitutions; the non-invertibility of D05 resides in the prefactor. However, it can be easily shown that the prefactors of R01 and TRK (Equations 20 & 21) are analytically invertible, so that inverted $\mathcal{L}^{\mathrm{R01}*} = \mathcal{L}^{\mathrm{R01}}$ and inverted $\mathcal{L}^{\mathrm{TRK}*} = \mathcal{L}^{\mathrm{TRK}}$.

2.4.2. *Demonstration of Invertibility and D05 Bias: Gaussian Random Clouds*

Consider the following simple, but illuminating, example: linear fits to circularly symmetric, slop-dominated, Gaussian random clouds of points. We generated ensembles of zero-mean, symmetric $\sigma_x = \sigma_y \equiv \sigma = 1$ clouds of $N = 100$ points each. The independent Gaussian random coordinates (x_n, y_n) of each point in each cloud were generated via the Box-Muller transformation on independent, uniformly distributed random numbers in the range $[0, 1]$. The data points were assigned zero intrinsic error bars $\sigma_{x,n} = \sigma_{y,n} = 0$.

In this simple example, since there is no intrinsic uncertainty in the data, $\Sigma_{x,n} = \sigma_x = \sigma$ and $\Sigma_{y,n} = \sigma_y = \sigma$ for all points, where (σ_x, σ_y) are the values of the fitted slop parameters. Since we are fitting a linear model $y_c(x) = b + mx$, the tangent point slope for all points is simply equal to the slope of the fitted line, $m_{t,n} = m \equiv \tan\theta$. Hence, the TRK likelihood of Equation 27 reduces to:

$$\mathcal{L}^{\mathrm{TRK}} \propto \prod_{n=1}^{N} \frac{1}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\sigma}\right)^2\right], \tag{30}$$

while the D05 likelihood of Equation 28 reduces to:

$$\begin{aligned}\mathcal{L}^{\mathrm{D05}} &\propto \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{1+m^2}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\sigma}\right)^2\right] \\ &= \left(\frac{1}{\sqrt{1+m^2}}\right)^N \mathcal{L}^{\mathrm{TRK}} \\ &\equiv \left(\cos^N\theta\right) \mathcal{L}^{\mathrm{TRK}}\,. \end{aligned} \tag{31}$$

Note, again, that the argument of the exponential is always the same for all of the statistics we are considering; any difference in their predictions lies entirely in the prefactor. What Equation 31 tells us is that, for a given guess of a best-fit linear slope $m \equiv \tan\theta$, the likelihood of D05 differs from that of TRK by a factor $\cos^N\theta$. Since any model-fitting algorithm we might choose is solely concerned with maximizing the likelihood function (or, equivalently, with minimizing $-2\ln\mathcal{L}$), it should be clear that, in this simple example, D05 will attempt to maximize $\cos\theta$, and hence minimize the slope, so that $\theta \sim 0$, to the degree that the particular data set allows. Under TRK, however, there is no a priori bias towards $\theta = 0$.

TRK is analytically invertible, but D05 is not. If we perform the above analysis using the inverted D05* likelihood of Equation 29, we see that $\mathcal{L}^{\mathrm{D05}*} = m^N \mathcal{L}^{\mathrm{D05}}$, so that $\mathcal{L}^{\mathrm{D05}*} = \left(\sin^N\theta\right)\mathcal{L}^{\mathrm{TRK}}$. Fits using the inverted D05* statistic will thus be biased to $\theta = 90°$, as plotted in the original, uninverted space. TRK *will* choose a particular best-fit θ, depending on the particular accidental correlations in a given random cloud (no given cloud with a finite number of data points is truly perfectly symmetric); this slope will be the same under inversion (see Figure 6). However, if we conduct fits to an ensemble of many such Gaussian random clouds, we should expect to find a uniform probability distribution of fitted θ for TRK, while for D05, the probability distribution should be proportional to $\cos^N\theta$; we find that this is indeed the case (see Figure 7).

2.4.3. *The Correlation Coefficient R^2_{xy} as a Measure of Invertibility*

The subject of invertibility can also be approached via the well-known Pearson correlation coefficient R^2_{xy} (**?**). R^2_{xy} can be interpreted as a measure of invertibility or non-invertibility of D05, or any other statistic under an inversion of the x- and y-axes (see, e.g., **?**). First, consider the case of fitting a line to a set of N data points $\{x_n, y_n\}$, with no intrinsic uncertainties (i.e., zero error bars) and no consideration of extrinsic scatter. In this case, linear regression gives the slope m_{xy} of the best-fit line to y vs. x:

$$m_{xy} = \frac{N\sum(x_n y_n) - \sum x_n \sum y_n}{N\sum(x_n^2) - (\sum x_n^2)}\,, \tag{32}$$

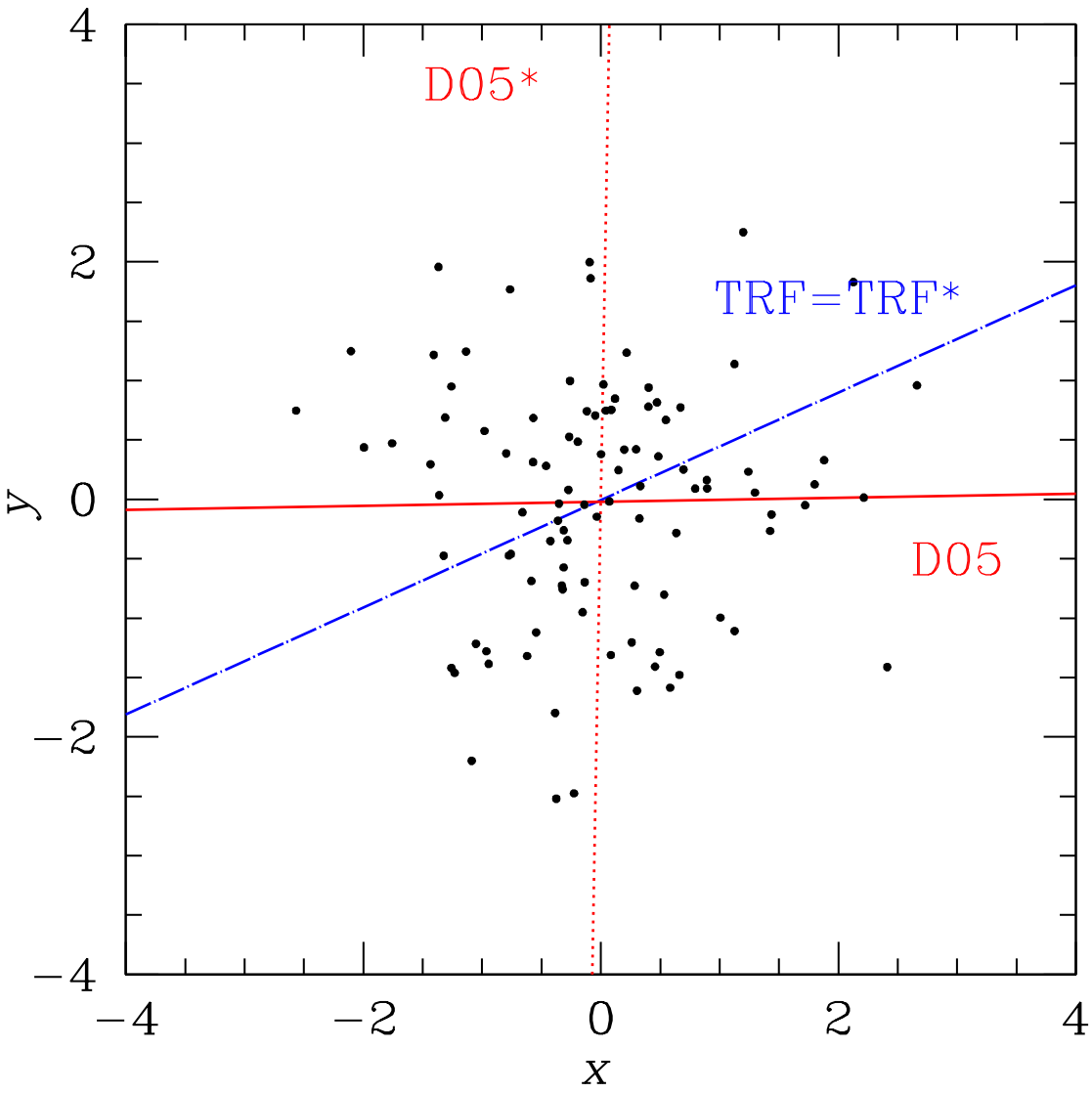


Figure 6. Linear fits to a slop-dominated Gaussian random cloud of $N = 100$ points with zero intrinsic error bars. Point coordinates generated with zero mean, $\sigma = 1$ Gaussian random variables in both dimensions. Solid red line shows fit with D05 statistic; dotted red line is fit with inverted D05* statistic. Blue line shows fit and inverted fit with TRK statistic; for an invertible statistic, the two are the same.

while the slope m_{yx} of the best-fit line to x vs. y (as measured in the inverted space) is:

$$m_{yx} = \frac{N\sum(x_n y_n) - \sum x_n \sum y_n}{N\sum(y_n^2) - (\sum y_n^2)}\,, \tag{33}$$

where the summations are all over the range $n = \{1 \ldots N\}$. In this simple linear regression case, Pearson's correlation coefficient R_{xy}^2 can be expressed as:

$$R_{xy}^2 \equiv \frac{[N\sum(x_n y_n) - \sum x_n \sum y_n]^2}{\left[N\sum(x_n^2) - (\sum x_n)^2\right]\left[N\sum(y_n^2) - (\sum y_n)^2\right]}\,. \tag{34}$$

By Equations 32, 33 & 34, Pearson's correlation coefficient is thus:

$$R_{xy}^2 \equiv m_{xy} m_{yx}\,. \tag{35}$$

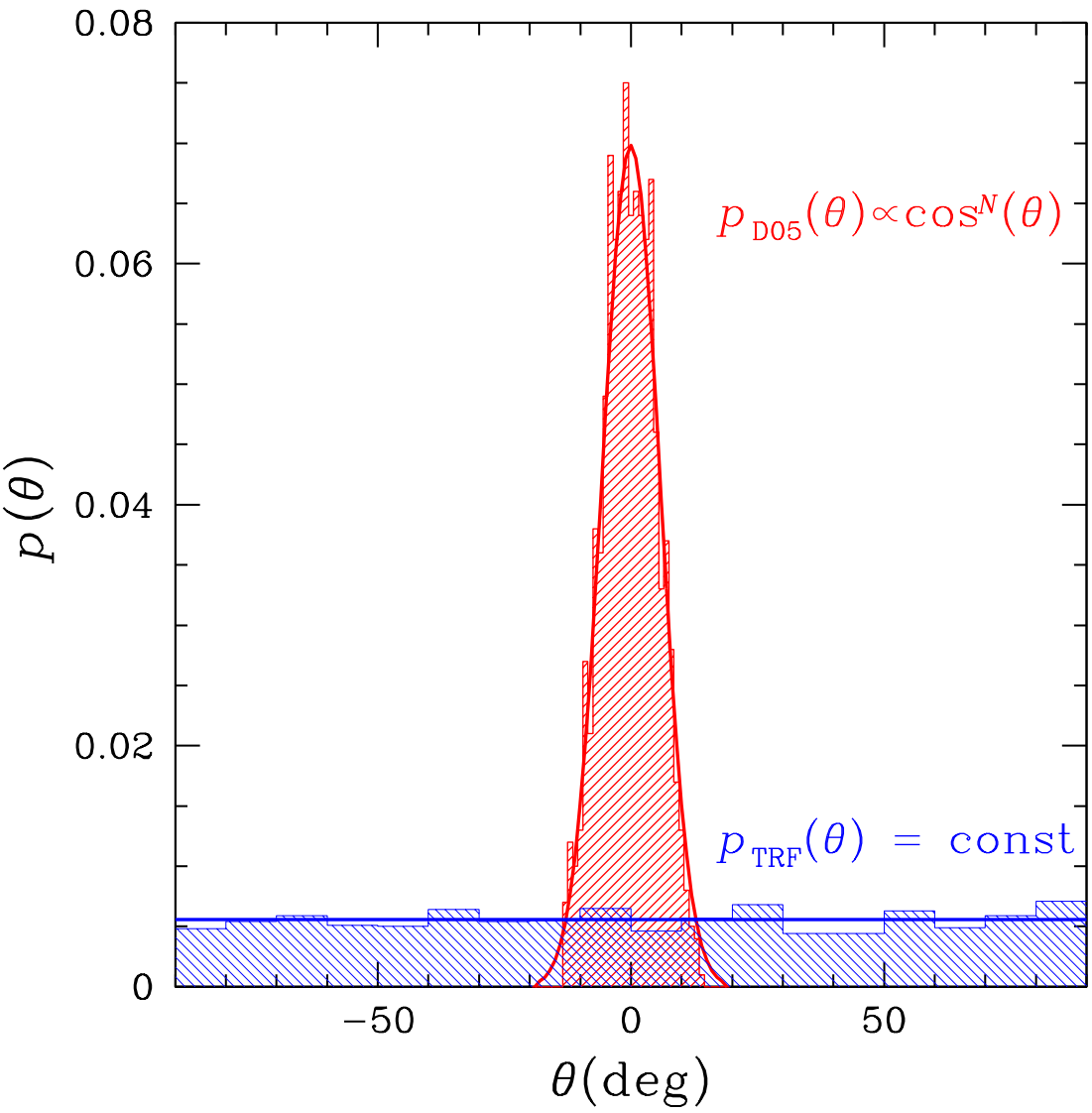


Figure 7. Histogram of fitted slope θ for an ensemble of 1000 zero-mean, $\sigma = 1$ Gaussian random clouds of $N = 100$ points each, for both D05 and TRK statistics. Solid lines indicate expected theoretical distribution. For TRK, $p(\theta)$ is a constant; for D05, $p(\theta)$ is biased towards $\theta = 0$ by a factor $\cos^N \theta$.

The above derivation does not take into account any intrinsic error bars or extrinsic slop in the data. However, we can use Equation 35 to generalize the R^2_{xy} measure, where m_{xy} and m_{yx} may come from linear fits with different statistics to data with both intrinsic and extrinsic scatter. Consider again the simple example of linear fits to slop-dominated Gaussian random clouds of §2.4.2. Given the bias of D05 towards the x-axis, we expect $m_{xy} \to 0$ and $m_{yx} \to 0$. Hence, D05 will yield a generalized value of $R^2_{xy} \to 0$; this is what we might expect, given the clearly uncorrelated nature of these data sets.

More generally, for linear fits to any data set with the D05 statistic, $R^2_{xy} \leq 1$ always; this, in itself, is a consequence of the bias of D05 to shallower slopes discussed in the context of the Gaussian random clouds. If $R^2_{xy} \approx 1$, the results of the D05 fit can be trusted, even though the statistic is technically non-invertible. But the lower the value of R^2_{xy}, the less the fits of D05 (or any other non-invertible statistic) can be trusted.

For invertible statistics like R01 and TRK, since $m_{xy} = 1/m_{yx}$, $R^2_{xy} \equiv 1$ by definition. This means that the fitted results can always be trusted under inversion. [4] The fact that $R^2_{xy} \equiv 1$ does not mean that using TRK or R01 somehow magically makes a data set fully correlated. Rather, any lack of correlation will be reflected in the fitted uncertainty in m_{xy} (see §2.6.1), instead of in an R^2_{xy} value. It is important to note that this uncertainty in m_{xy} can be usefully propagated into other analyses, perhaps as part of a prior. A low value of R^2_{xy}, on the other hand, is useful only as a warning that the fitted result should not be trusted.

2.4.4. *Behavior in 1D Limit*

Another desirable property of any two-dimensional statistic is that it reduce to a χ^2-like statistic in the one-dimensional limits $\Sigma_{x,n} \to 0$ (all uncertainty in the y-direction) or $\Sigma_{y,n} \to 0$ (all uncertainty in the x-direction). In the former case, we require that:

$$\mathcal{L} \to \prod_{n=1}^{N} \frac{1}{\Sigma_{y,n}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{y,n}}{\Sigma_{y,n}}\right)^2\right], \tag{36}$$

and, in the latter, that:

$$\mathcal{L} \to \prod_{n=1}^{N} \frac{1}{\Sigma_{x,n}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{x,n}}{\Sigma_{x,n}}\right)^2\right], \tag{37}$$

where $\delta_{y,n}$ and $\delta_{x,n}$ are the distances between the data point centroid (x_n, y_n) and the model curve in the y- and x-directions, respectively. Again, we refer to these statistics as "χ^2-like", because they are analogous to the traditional one-dimensional χ^2 statistic, with the caveat that Σ_n is not fixed, but a function of a free model parameter, i.e., the slop in the relevant dimension (§2.3). It should be obvious by inspection that the argument of the exponentials in the D05, R01 and TRK likelihood functions (Equations 19, 20 & 21) satisfy both Equation 36 and 37 in both limiting cases. If $\Sigma_{x,n} = 0$, the tangent point will lie along the curve $y_c(x; \vartheta_m)$ at the x-value of the data point, i.e., at $(x_n, y_c(x_n; \vartheta_m))$. And if $\Sigma_{y,n} = 0$, the tangent point will lie along the curve at the y-value of the data point, i.e., at $(y_c^{-1}(y_n; \vartheta_m), y_n)$. Again, the behavior of the statistics is entirely determined by the form of the prefactor in the likelihood function.

It should also be obvious by inspection that D05 (Equation 19) reduces to Equation 36 when $\Sigma_{x,n} \to 0$, but that it does not reduce to Equation 37 when $\Sigma_{x,n} \to 0$; in the latter case, there is an extra factor of $1/m_{t,n}$ in the prefactor. To use D05 to fit to data in this one-dimensional limit, one must first invert the x- and y-axes; in this case, the inverted D05* statistic of Equation 29 reduces to Equation 37. TRK has the advantage that one does not need to keep track of whether one-dimensional error bars are in the x- or y-direction and perform the inversion; it works, automatically (Equation 21 reduces to Equations 36 & 37 in both limiting cases).

Note, however, that R01 (Equation 20) does not reduce to the one dimensional statistics in either limit; there are extra factors of $\sqrt{1 + m^2_{t,n}}$ when $\Sigma_{x,n} \to 0$, and $\sqrt{1 + 1/m^2_{t,n}}$ when $\Sigma_{y,n} \to 0$. We consider this a fatal flaw of R01; it was one of the motivating factors in developing a new invertible statistic in the first place.

2.4.5. *Do Two Points Really Define a Line?*

As a very basic reality check, we investigate the seemingly trivial problem of fitting a line to two data points. We proceeded under the assumption that, for a two-point data set, the concept of extrinsic scatter is meaningless, i.e., the slop parameters (σ_x, σ_y) are held fixed at zero. In this very simple case, our intuition leads us to expect the best-fit line to pass through the centroids of both data points' error ellipses; "two points define a line," after all. Or do they? We find that *none* of the three statistics, neither D05, R01, nor TRK, is capable of producing this intuitive result in all cases.

Figure 8 shows a series of linear fits to two data points with centroids (x_n, y_n) at $(-1, 1)$ and $(1, 1)$. The intrinsic error bars in each panel are a function of the parameter ρ, such that $\sigma_{x,n} = \sin\rho$, $\sigma_{y,n} = \cos\rho$; $\rho = 0°$ corresponds to the one-dimensional case $\sigma_{x,n} = 0$, $\sigma_{y,n} = 1$, while $\rho = 90°$ corresponds to the one-dimensional case $\sigma_{x,n} = 1$, $\sigma_{y,n} = 0$. The D05 statistic recovers the expected $\theta = 45°$ line only in the $\rho = 0°$ case, while the inverted D05* statistic succeeds only for $\rho = 90°$. The R01 statistic succeeds only in the $\rho = 45°$ case (i.e., when $\sigma_{x,n} = \sigma_{y,n}$). TRK, on the other

[4] Note, however, that these statistics may have other problems the user should be aware of. For instance, unlike D05, TRK fits are not invariant under a multiplicative scaling of the x- or y-axis; we discuss this behavior in §2.4.6, and introduce an alternative, scale-based correlation coefficient in §2.5.1.

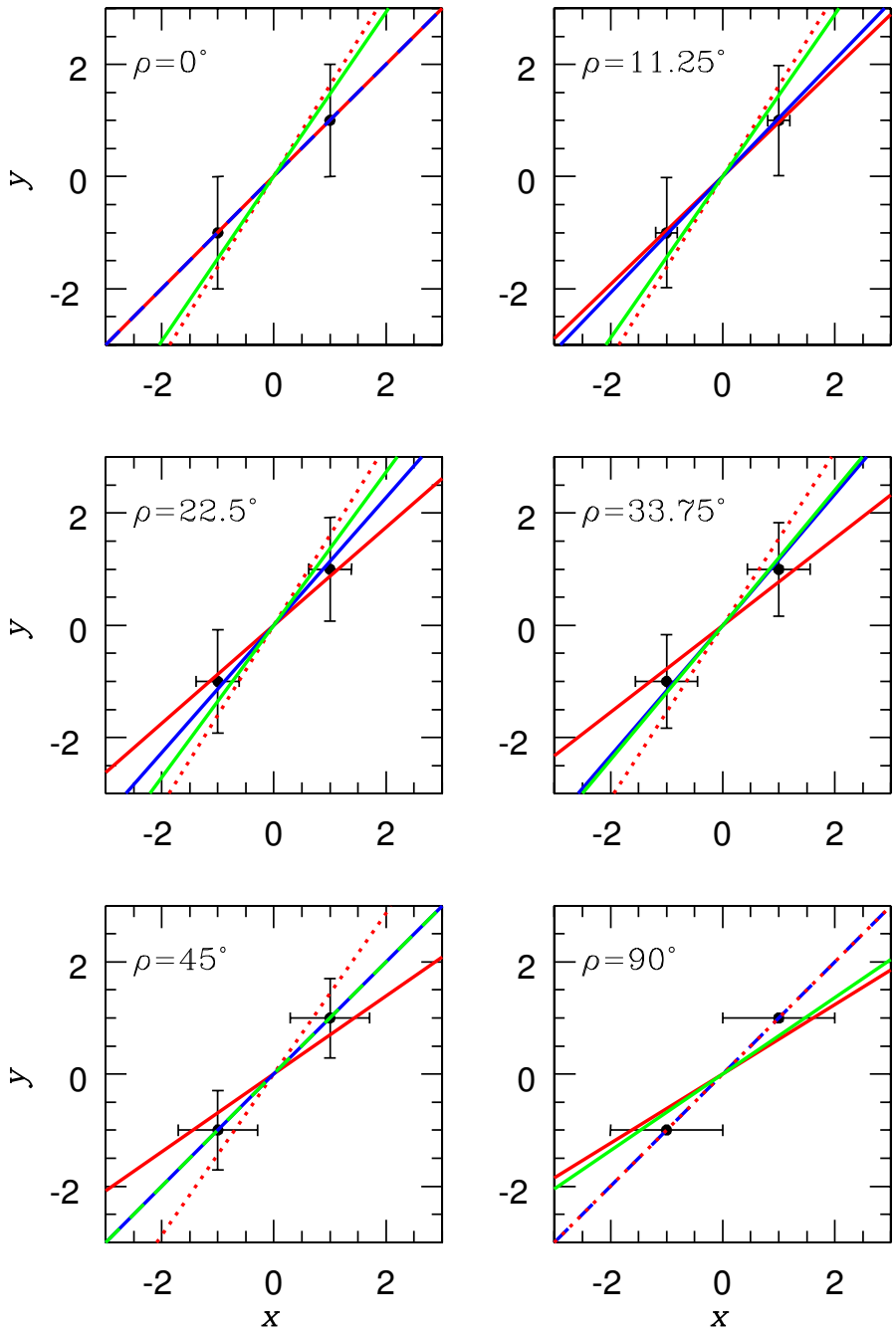


Figure 8. Linear fits to two data points, with centroids at $(-1,-1)$ and $(1,1)$, with a range of intrinsic uncertainties, $\sigma_{x,n} = \sin\rho$, $\sigma_{y,n} = \cos\rho$. Solid red line: D05. Dotted red line: inverted D05*. Green line: R01. Blue line: TRK.

hand, recovers $\theta = 45°$ for $\rho = 0°, 45°$ and $90°$. This result is not surprising, in light of the fact that TRK reduces to D05 (or inverted D05*) in both one-dimensional cases ($\Sigma_{x,n} \to 0$, $\Sigma_{y,n} \to 0$; § 2.4.4), and to R01 in the case where $\Sigma_{x,n} = \Sigma_{y,n}$. Figure 9 shows fitted θ vs. ρ for D05, inverted D05*, R01 and TRK. So, while none of these statistics is "perfect" in light of this two-point test, TRK at least succeeds, by itself, where all the others succeed separately.

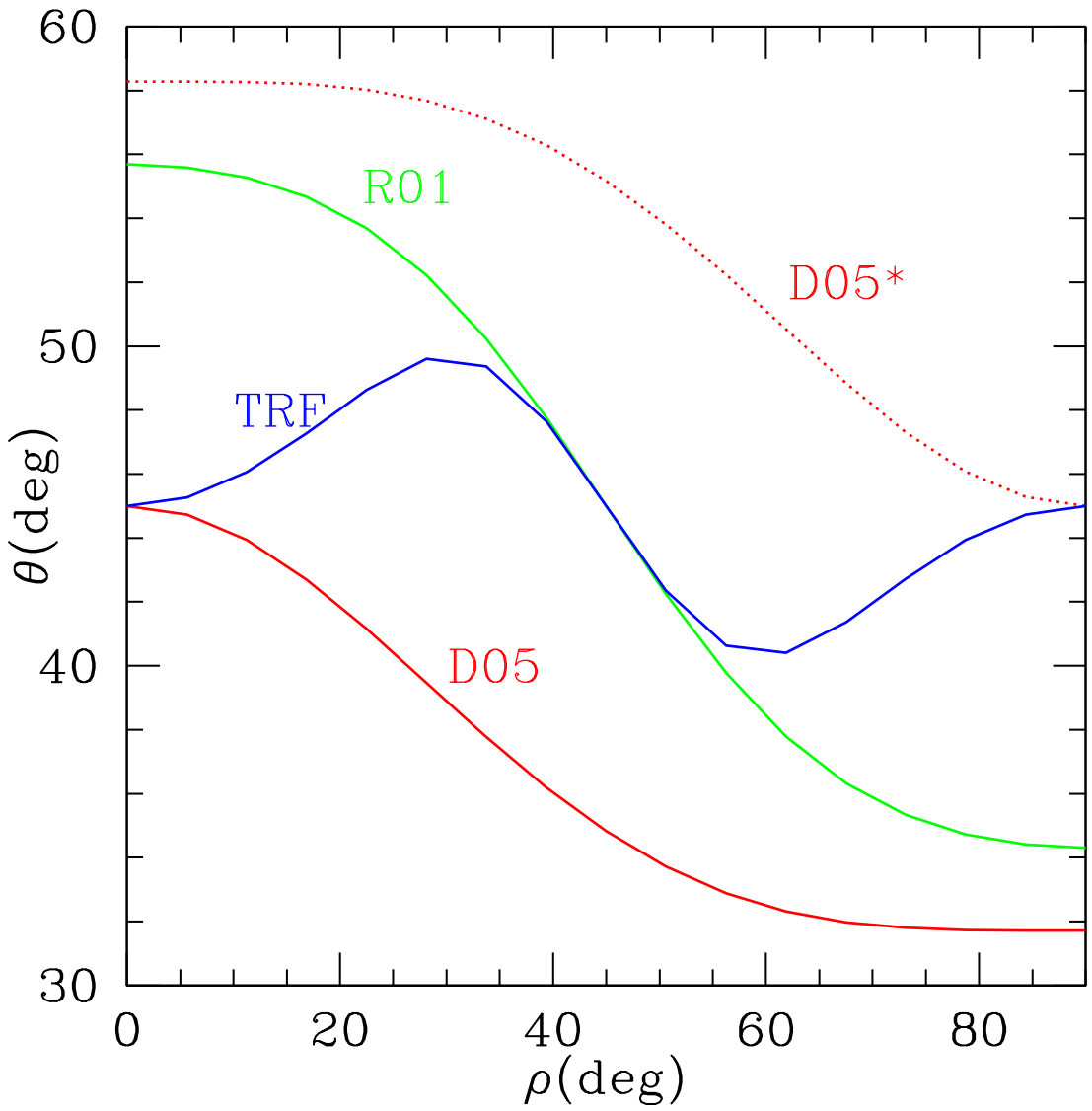


Figure 9. Best-fit linear position angle θ vs. axial ratio parameter ρ for two-point data sets with centroids at $(-1,-1)$ and $(1,1)$, for D05, inverted D05*, R01 and TRK statistics. Parameter ρ describes intrinsic uncertainties of each data point: $\sigma_{x,n} = \sin\rho$, $\sigma_{y,n} = \cos\rho$.

2.4.6. *Scalability*

While the TRK statistic is invariant under an inversion of the x- and y-axes, reduces to a χ^2-like statistic in the 1D limit, and passes the two-point test in more cases than either D05 or R01, it is not invariant under an arbitrary rescaling of the x- or y-axes; in other words, the best-fit that TRK provides depends on the choice of units of measurement (or, equivalently, on the choice of basis when fitting to data in logarithmic space; see §2.7.1). Since there is no a priori reason that one choice of units (or logarithmic basis) is preferable to another, it is necessary to explore the behavior of TRK under all possible scalings in order to fully describe its predictions.

Consider the behavior of any of our statistics under an arbitrary rescaling of the x-axis by a factor s_x, and of the y-axis by a factor s_y. This is equivalent to replacing all coordinates (x, y) with $(s_x x, s_y y)$, all uncertainties (Σ_x, Σ_y) with $(s_x\Sigma_x, s_y\Sigma_y)$, and all slopes m with $(s_y/s_x)m$ in the expression for the likelihood (Equation 20, 19 or 21). By inspection, it is obvious that the R01, D05 and TRK likelihoods are invariant when both axes are scaled by the same factor, $s_x = s_y$. This means we can explore, without loss of generality, the full range of scale-dependent behavior of these statistics by considering only rescalings of the y-axis by a factor s (equivalently, we can define the scaling

factor $s \equiv s_y/s_x$; the results are the same). The D05 statistic is, in fact, fully scalable. If we substitute $y_n \to sy_n$, $y_{t,n} \to sy_{t,n}$, $\Sigma_{y,n} \to s\Sigma_{y,n}$, and $m_{t,n} \to sm_{t,n}$ in Equation 19, we obtain the scaled likelihood function:

$$\begin{aligned}\mathcal{L}_s^{\mathrm{D05}} &\propto \prod_{n=1}^{N} \frac{1}{s\sqrt{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \exp\left\{-\frac{1}{2}\frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}\right\} \\ &= \left(\frac{1}{s^N}\right)\mathcal{L}^{\mathrm{D05}} \\ &\propto \mathcal{L}^{\mathrm{D05}} \quad \text{for a fixed } s\,. \end{aligned} \tag{38}$$

The argument of the exponential is invariant under the scaling; note, again, that this argument is the same for all three statistics we are considering here. The only effect of scaling on the D05 statistic is the introduction of a constant factor $1/s$ in the prefactor. Therefore, D05 will yield exactly the same best-fit curve (when plotted in the original, unscaled coordinate space); the total fitness $-2\ln\mathcal{L}_s^{\mathrm{D05}}$ will simply differ from $-2\ln\mathcal{L}^{\mathrm{D05}}$ by a constant offset of $2N\ln s$. In other words, $-2\ln\mathcal{L}_s^{\mathrm{D05}} - (-2\ln\mathcal{L}_s^{\mathrm{D05}})_{\min}$ will differ from $-2\ln\mathcal{L}^{\mathrm{D05}} - (-2\ln\mathcal{L}^{\mathrm{D05}})_{\min}$ by zero.

However, the TRK likelihood of Equation 21 scales as:

$$\begin{aligned}\mathcal{L}_s^{\mathrm{TRK}} &\propto \prod_{n=1}^{N} \sqrt{\frac{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}{m_{t,n}^2\Sigma_{x,n}^4 + s^2\Sigma_{y,n}^4}} \exp\left\{-\frac{1}{2}\frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}\right\} \\ &\not\propto \mathcal{L}^{\mathrm{TRK}} \quad \text{for a fixed } s\,. \end{aligned} \tag{39}$$

That is, the scaling factor s cannot be factored out of the likelihood function. Consequently, different choices of s will result in different fitted values of the slop parameters (σ_x, σ_y), and, through $m_{t,n}$, the other model parameters ϑ_m. (It can be shown, similarly, that $\mathcal{L}^{\mathrm{R01}}$ also does not scale, but we do not investigate the behavior of R01 any further in this discussion, given its other limitations.) Furthermore, it is impossible to use the numerical value of the likelihood function to make any assertions as to the relative fitness of the best fits that TRK yields at different scales. In other words, the scaling factor s is not a model parameter that we can fit to. Its value cannot be determined by the data, but is rather a choice that must be made before fitting. The primary practical effect is that, for the TRK statistic, the best-fit model distribution, plotted in the original, unscaled space, is different for different choices of s.

Consider the scale dependence of slop direction, i.e., how the total slop is distributed between σ_x and σ_y. For the TRK statistic, this dependence is most vividly illustrated in the case of linear fits to slop-dominated data, i.e., when the intrinsic error bars on the data points $(\sigma_{x,n}, \sigma_{y,n}) = 0$. In the limit $s \to 0$, the likelihood of Equation 39 is dominated by a term of order $1/\sigma_x$ in the prefactor. To maximize the likelihood, the TRK fit will force $\sigma_x \to 0$, and Equation 39 reduces to:

$$\mathcal{L}_{s\to 0}^{\mathrm{TRK}} \to \prod_{n=1}^{N} \frac{1}{s\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\delta_{y,n}}{\sigma_y}\right)^2\right]. \tag{40}$$

In other words, in the limit $s \to 0$, TRK reduces to the one-dimensional, χ^2-like statistic of Equation 36, modulo a constant factor of $(1/s)^N$. In the limit $s \to \infty$, the likelihood of Equation 39 is dominated by a term of order $1/s\sigma_y$ in the prefactor. To maximize the likelihood, the TRK fit will force $\sigma_y \to 0$, and Equation 39 reduces to:

$$\mathcal{L}_{s\to \infty}^{\mathrm{TRK}} \to \prod_{n=1}^{N} \frac{1}{\sigma_x} \exp\left[-\frac{1}{2}\left(\frac{\delta_{x,n}}{\sigma_x}\right)^2\right]. \tag{41}$$

In other words, for linear fits to slop-dominated data, in the limit $s \to \infty$, TRK reduces to the one-dimensional, χ^2-like statistic of Equation 37.

Another way of looking at this scaling behavior is that, for linear fits to slop-dominated data, *the predictions of TRK at these two extremes of scale are equivalent to those of D05 under inversion.* As $s \to 0$, TRK yields the same results as D05, and as $s \to \infty$, it yields the same results as inverted D05*. However, there are two significant advantages to using TRK: (1) TRK can explore a *continuum* of scales in between those limits, while D05 is restricted to the binary choice of inversion or non-inversion. In §2.5, we will introduce a method for selecting an optimum midrange scale for TRK fits. (2) As we will see below, if the intrinsic error bars $(\sigma_{x,n}, \sigma_{y,n})$ are non-zero, the *physically meaningful range*

of the scale continuum is a subset of $[0, \infty]$. As the intrinsic scatter grows relative to the extrinsic scatter, this range of physically meaningful scales shrinks. In the limit where the data are entirely dominated by intrinsic scatter (i.e., both slops $(\sigma_x, \sigma_y) = 0$), this range reduces to a single physically meaningful scale, with a single best-fit curve.

2.4.7. *Degeneracy of Slop for Linear Models*

In fitting linear models to data with extrinsic scatter, or slop, intuition tells us that there should be a degeneracy between slop in the x- and y-directions. That is, if the best-fit line has slope m, the model distribution can be described equally well by a single slop parameter σ projected entirely into the y-direction ($\sigma_x = 0$, $\sigma_y = \sigma$), or entirely into the x-direction ($\sigma_x = \sigma/m$, $\sigma_y = 0$), or into any combination of the two such that $m^2\sigma_x^2 + \sigma_y^2 \equiv \sigma^2$. Now, all three statistics (Equations 19, 20, & 21) have the same argument in the exponential function, with a term in the denominator:

$$\begin{aligned}\Sigma_n^2 &\equiv m^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2 \\ &\equiv m^2\sigma_{x,n}^2 + \sigma_{y,n}^2 + m^2\sigma_x^2 + \sigma_y^2 \\ &\equiv m^2\sigma_{x,n}^2 + \sigma_{y,n}^2 + \sigma^2 \,. \end{aligned} \tag{42}$$

Therefore, the quantity Σ_n, and therefore the argument of the exponential for all three statistics, is degenerate in slop for linear fits.

The D05 likelihood (Equation 19) has a prefactor that is simply $1/\Sigma_n$, and so linear fits with D05 are fully degenerate in slop; the statistic has no preference for projecting it into one direction or the other. Similarly, the R01 (Equation 20) prefactor is $\sqrt{1+m^2}/\Sigma_n$, which also preserves slop degeneracy. The prefactor of TRK (Equation 21), on the other hand, is equal to $\Sigma_n/\sqrt{m^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}$. In this case, the degeneracy is broken, meaning that TRK will pick specific values for σ_x and σ_y, even in the slop-dominated linear case. We saw this in §2.4.6: In the limit $s \to \infty$, all slop is placed in the x-direction ($\sigma_y \to 0$), and in the limit $s \to 0$, all slop is placed in the y-direction ($\sigma_x \to 0$). This does *not* mean that TRK can magically disentangle σ_x from σ_y. Rather, how much slop is put into one direction or the other depends on the choice of scale. The non-scalability of TRK and its non-degeneracy of slop in linear fits are, in a very real sense, equivalent.

Since we can conduct a particular fit only at a fixed scale, it is important to keep in mind that, when using TRK, we should not interpret the best-fit values of (σ_x, σ_y) *physically*. Other values of slop in either direction are possible, given other choices of scale. We must instead view the fitted slops as an *empirical* parameterization of extrinsic scatter in the data at that particular scale. This is not a problem, considering that this is how we introduced slop in the first place (§2): as an empirical parameterization of extrinsic scatter about the curve $y_c(x; \vartheta_m)$ that has a physical explanation, albeit one that we cannot or choose not to model physically. For linear fits, the model distribution obtained by convolving the curve $y_c(x; \vartheta_m)$ with the two-dimensional Gaussian that represents the slop is the same, whether the slop is projected entirely in the x-direction, entirely in the y-direction, or in some combination of the two. There is no consequence of the TRK statistic's lack of slop degeneracy, except perhaps that it differs from our intuitive experience with χ^2-like statistics. It simply reflects the fact that, for linear fits with TRK, the directionality of slop, like its magnitude, is empirical.[5]

2.4.8. *Summary of Strengths and Weaknesses*

Table 1 summarizes the properties of the D05, R01 and TRK statistics discussed above. Clearly, each statistic has its own strengths and weaknesses. We consider TRK to be, on the whole, an improvement on R01; R01 is, in fact, just a special case of TRK when $\Sigma_{x,n} = \Sigma_{y,n}$, and will not be discussed further here. TRK has the advantage of invertibility over D05 (and reduces to D05 in both one-dimensional limits). On the other hand, D05 is invariant under scaling, while TRK is not. Also, in the case of TRK, the fitted values of slop (σ_x, σ_y), at any particular choice of scale, cannot be interpreted physically (§2.4.7), but only as empirical parameterizations of extrinsic scatter in the data at that scale. However, the advantage of TRK is that we can explore a continuous range of scales, while D05 is restricted to a binary choice of inversion or non-inversion. In the next section, we explore the scaling behavior of TRK in greater detail, and describe a practical method for performing and interpreting model fits using this statistic.

[5] For non-linear fits, in cases where the data require that there be some slop in both directions, the best-fit TRK curves at extremes of scale may enter otherwise unphysical regimes in order to accommodate all the data with a single non-zero slop parameter, and so the range of allowable scales may need to be artificially constrained (see, e.g., Figure 16 in §2.5).

Table 1. Summary of Statistical Properties

	D05	R01	TRK
Invertible?	No	Yes	Yes
Reduces to 1D χ^2?	Yes	No	Yes
Passes 2-point test?	If $\Sigma_x = 0$	If $\Sigma_x = \Sigma_y$	If $\Sigma_x = 0$, $\Sigma_y = 0$, or $\Sigma_x = \Sigma_y$
Scalable?	Yes	No	No
Slop degenerate if linear?	Yes	Yes	No

2.5. *Implementation of the TRK Statistic*

2.5.1. *A New Correlation Coefficient for the TRK Statistic*

As we discuss in §2.4.6, for slop-dominated linear data (i.e., intrinsic error bars equal zero), the behavior of TRK under extremes of scale is equivalent to the behavior of D05 under an inversion of the x- and y-axes. Specifically, as the scaling factor $s \to 0$, TRK becomes equivalent to the one-dimensional D05 statistic of Equation 36, and as $s \to \infty$, it is equivalent to the one-dimensional inverted D05* statistic of Equation 37. Figure 10 illustrates this limiting behavior. Data points were generated by adding independent zero-mean Gaussian random variables with standard deviation $\sigma_x = \sigma_y = 0.5$ to points randomly distributed along the line $y_c(x) = x$. The dashed-dotted lines show the non-inverted D05 fit (shallower line), and inverted D05* fit (steeper line). These fits are identical to the TRK best-fits in the scale limits $s \to 0$ and $s \to \infty$, respectively. With the D05 statistic, these two lines are the only possible fits to this data set. However, with TRK, we have the freedom to choose any intermediate scale s along a continuum. The solid blue line shows the TRK fit at an "optimum" intermediate scale, which we will discuss further below.

Figure 11 shows fits to an identical data set as in Figure 10, but with intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0.25$ assigned to each point. The dotted red lines show the D05 and inverted D05* fits, which are *identical* to those for the slop-dominated data set Figure 10. The dashed blue lines show TRK fits at the scales $s_{\min}$ and $s_{\max}$ at which the fitted slops $\sigma_x \to 0$ and $\sigma_y \to 0$, respectively. Clearly, in this case, the TRK statistic produces a range of fitted slopes that lies inside the boundaries defined by the fits using the D05 and inverted D05* statistics.

For fits to data with non-zero intrinsic uncertainties with the TRK statistic, fits that fall outside of the range obtained at these two limiting scales would require either that $\sigma_x^2 < 0$ (at smaller scales) or $\sigma_y^2 < 0$ (at larger scales). Consider the extreme limiting cases of $s \to 0$ and $s \to \infty$ in the linear fits shown in Figure 11. In the former case, in order to reproduce the fit obtained with D05, TRK would have to reduce to the one-dimensional statistic of Equation 36. The only way this could happen would be if the quantity $\Sigma_n^2 \equiv m^2(\sigma_{x,n}^2 + \sigma_x^2) + (\sigma_{y,n}^2 + \sigma_y^2)$ (Equation 42) were to reduce to $\sigma_{y,n}^2 + \sigma_y^2$, which would require the slop parameter $\sigma_x^2 < 0$ (for correlated data with non-zero slope). Similarly, in the latter case, to reproduce the inverted D05* fit, TRK would have to reduce to the one-dimensional statistic of Equation 37, which would require the slop parameter $\sigma_y^2 < 0$. We defined the model distribution as the convolution of a two-dimensional Gaussian with a model curve, where the 1σ widths of the Gaussian in the x- and y-directions are equal to (σ_x, σ_y). Simply put, there is no real, physically meaningful model distribution in the x-y plane that can be constructed with imaginary values of slop. Therefore, any scenario that requires $\sigma_x^2 < 0$ or $\sigma_y^2 < 0$ is unphysical, and the range $[s_{\min}, s_{\max}]$ represents the full, physically meaningful range of scales for fits with the TRK statistic.[6]

In general, for fits to data with intrinsic error bars, TRK will predict a range of best-fit curves under physically meaningful changes of scale that falls within the region bounded by D05 and inverted D05*. How do we quantify this behavior? As we discussed in §2.4.3, the linear correlation coefficient R_{xy}^2 can be interpreted as a measure of the reliability of a statistic under an inversion of the x- and y-axes. Specifically, for a linear fit, if the slope of the best-fit line to y vs. x is m_{xy}, and the slope of the line fit to x vs. y is m_{yx} (as measured in the inverted space), then we can define a generalized correlation coefficient $R_{xy}^2 \equiv m_{xy} m_{yx}$. From this definition of R_{xy}^2 as the product of the slopes of best-fit lines under an inversion of the x- and y-axes, it should be clear that for an invertible statistic like TRK,

[6] There remains the possibility that if, at a given scale, a fit requires $\sigma_x^2 < 0$ or $\sigma_y^2 < 0$, then the intrinsic error bars $\sigma_{x,n}$ or $\sigma_{y,n}$ are overestimated. If there is genuine cause for concern that the quoted statistical error bars for a given data set are overestimated, we technically have the option to introduce, as free model parameters, new multiplicative scaling factors $(f_x, f_y) \leq 1$ to the intrinsic error bars whenever $(\sigma_x^2, \sigma_y^2) < 0$, i.e., to replace $(\sigma_{x,n}, \sigma_{y,n})$ with $(f_x \sigma_{x,n}, f_y \sigma_{y,n})$. In this case, TRK will asymptote to the limiting behavior of one-dimensional D05, or inverted D05*, as $f_x \to 0$, or $f_y \to 0$, respectively. We do not explore this option further here, i.e., we assume that the given intrinsic error bars in the data sets we analyze are not overestimated, at least not significantly.

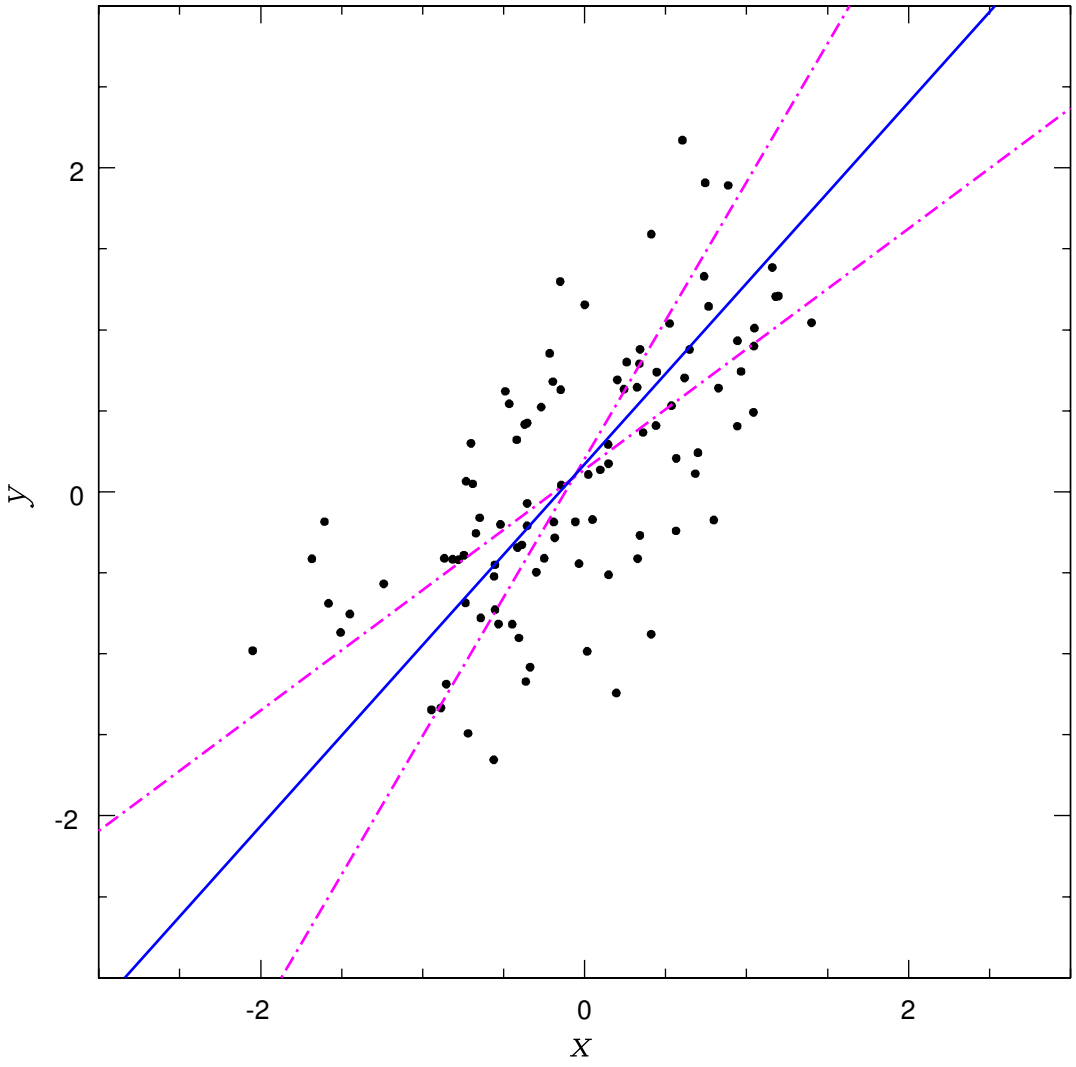


Figure 10. Linear fits to slop-dominated data. Data points generated by adding independent zero-mean Gaussian random variables with standard deviation $\sigma_x = \sigma_y = 0.5$ to $N = 100$ points randomly distributed between $(-1, -1)$ and $(1, 1)$ along the line $y_c(x) = x$. Data points are assigned zero intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0$. Dash-dotted lines: uninverted D05 fit/TRK fit as scale $s \to 0$ (shallow line), and inverted D05* fit/TRK fit as scale $s \to \infty$ (steeper line). Solid blue line: TRK fit at optimum scale.

$R^2_{xy} \equiv 1$ for all linear fits, since $m_{xy} = 1/m_{yx}$, by definition of invertibility. Thus, for an invertible statistic, R^2_{xy} can provide no statistical information.

We seek an analogous "correlation coefficient" for TRK, one that is a measure of the variance of the statistic's predictions under a change of scale. We require that this coefficient share the properties of the traditional R^2_{xy}, i.e., that $0 \leq R^2_{\rm TRK} \leq 1$, where $R^2_{\rm TRK} = 1$ corresponds to best-fit lines of identical slope (as plotted in the same scale space) for the two scales we are comparing, while $R^2_{\rm TRK} = 0$ corresponds to best-fit lines that are orthogonal. One option, if we were only concerned with linear fits, would be to define an analogous correlation coefficient in terms of the ratio of the slopes of the best-fit lines at the two extremes of scale. But even in the linear case, there are scenarios where this definition would fail. For instance, if the best-fit TRK line at one scale extreme has a slope $m = 0$, and at the other extreme has a slope just slightly greater or less than zero, any correlation coefficient defined in terms of some ratio of the two slopes would either equal zero or $\pm\infty$ – despite the fact that the two limiting lines are quite similar.

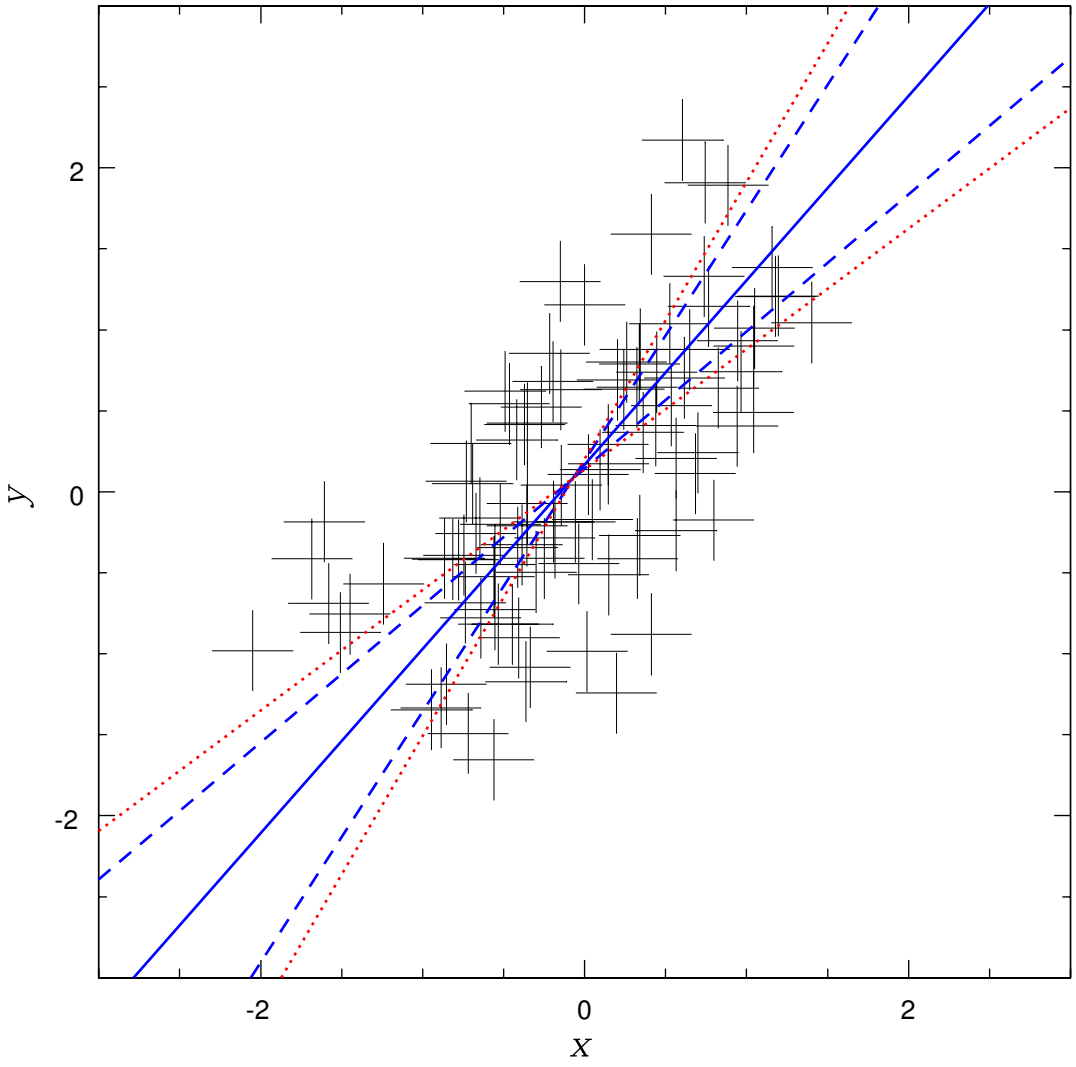


Figure 11. Linear fits to data with both slop and intrinsic uncertainties. The coordinates of each point are the same as in Figure 10. Each point was assigned symmetric intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0.25$. Dotted red lines: uninverted D05 fit (shallow line), and inverted D05* fit (steeper line); these fits are identical to those in Figure 10. Dashed blue lines: TRK fits at scale $s_{\min}$ at which fitted slop $\sigma_x \to 0$ (shallower line), and at scale $s_{\max}$ at which slop $\sigma_y \to 0$ (steeper line). Solid blue line: TRK fit at optimum scale.

Note that this divergence of R^2_{xy} never occurs with fits using the D05 statistic. For instance, consider a linear fit to *uncorrelated* data, i.e., a fit for which D05 gives a slope $m \approx 0$. For such a data set, whether the data are dominated by extrinsic scatter or intrinsic uncertainty, the inverted D05* fit will give a slope $m \to \infty$, and hence a correlation coefficient $R^2_{xy} \approx 0$. If the data are correlated, but dominated by extrinsic scatter, likewise, TRK will reproduce the D05 and inverted D05* fits at the extremes of scale $s \to 0$ and $s \to \infty$, respectively, as discussed above. However, if the data are dominated by intrinsic uncertainty, TRK will return similar slopes over the full range of physically meaningful scales; in the special case of $m \approx 0$, R^2_{xy} may then diverge. Pearson's correlation coefficient has stood the test of time in no small part thanks to the fact that linear fits with non-invertible statistics never encounter this failing case.

We introduce a new correlation coefficient, $R^2_{\rm TRK}$ that is a function of the *difference* of the slopes, as measured by position angles, of the two curves we are comparing, rather than the slope *ratio*. First consider the case of a linear fit.

Let the slope of the best-fit line at scale $s_{\min} = a$ be $m_a = \tan\theta_a$, and the slope of the best-fit line at scale $s_{\max} = b$ be $m_b = \tan\theta_b$, as plotted in the original scale space $s = 1$. Then we can define a scale-dependent correlation coefficient in terms of the difference $|\theta_a - \theta_b|$:

$$R^2_{\rm TRK}(a,b) \equiv \tan^2\left(\frac{\pi}{4} - \frac{|\theta_a - \theta_b|}{2}\right), \tag{43}$$

where the position angles θ describe the slopes in the original, unscaled space $s = 1$. By inspection, it is clear that $R^2_{\rm TRK}(a,b) = 1$ if $\theta_a = \theta_b$, while $R^2_{\rm TRK}(a,b) = 0$ if $\theta_a = \theta_b \pm \pi/2$. In order to find the "optimum scale" s_0, we consider the limiting behavior of the fits at the scales (a,b), defined either as $s \to 0$ and $s \to \infty$ (for slop-dominated data), or else as the scales at which the TRK fit becomes unphysical. Unphysical scales are typically those for which TRK would require fitted slops $(\sigma_x^2, \sigma_y^2) < 0$, though there are cases where best fits beyond a particular scale may be deemed unphysical for other reasons (see Figure 16, below). We next follow an iterative approach. The first approximation of s_0 is the scale at which:

$$R^2_{\rm TRK}(a,s_0) = R^2_{\rm TRK}(s_0,b) \equiv R^2_{\rm TRK}\,. \tag{44}$$

We then shift from the $s = 1$ space to this $s = s_0$ space, where the position angles of the lines are transformed as $\theta \to \tan^{-1}(s_0 \tan\theta)$. We repeat the analysis in this space, finding a new scale s_0' at which $R^2_{\rm TRK}(a,s_0') = R^2_{\rm TRK}(s_0',b) \equiv R^2_{\rm TRK}$, i.e.:

$$\begin{aligned} R^2_{\rm TRK} &\equiv \tan^2\left(\frac{\pi}{4} - \frac{\left|\tan^{-1}(s_0\tan\theta_a) - \tan^{-1}(s_0\tan\theta_{s_0'})\right|}{2}\right) \\ &= \tan^2\left(\frac{\pi}{4} - \frac{\left|\tan^{-1}(s_0\tan\theta_{s_0'}) - \tan^{-1}(s_0\tan\theta_b)\right|}{2}\right), \end{aligned} \tag{45}$$

where $\theta_{s_0'}$ is the position angle of the best-fit line at scale s_0', as measured in $s = 1$ space. We then set $s_0 = s_0'$, and repeat the procedure, scaling all fitted position angles into that s_0 space and finding a new value of s_0'. We find that s_0 and $R^2_{\rm TRK}$ converge to stable values to very high precision after only a few iterations. This final value of s_0 we take to be the *optimum scale* for TRK fits to the data set; it is at this optimum scale that we compute uncertainties in the fitted model parameters (see §§2.6.1 & 2.6.2).

2.5.2. *Properties and Behavior of $R^2_{\rm TRK}(a,b)$ and $R^2_{\rm TRK}$*

In the procedure described above, we defined both $R^2_{\rm TRK}(a,b)$ (Equation 43), which is a measure of the difference between the TRK fits at the two physically meaningful extremes of scale, and $R^2_{\rm TRK}$ (Equation 44), which is a measure of the difference between the fit at the optimum scale s_0 and the fit at either extreme. We find that when $R^2_{\rm TRK}(a,b)$ is defined in this way with this procedure, it has the additional, very fortuitous property that $R^2_{\rm TRK}(a,b) = R^2_{xy}$ for slop-dominated linear data, when computed at the optimum scale s_0. To demonstrate this fact empirically, we provide Figures 12 & 13. These two figures compare linear fits to slop-dominated data with a variety of slopes, for both loosely correlated ($\sigma_x = \sigma_y = 0.5$) and tightly correlated ($\sigma_x = \sigma_y = 0.2$) data sets. In both figures, the original data set (with different scalings of the same zero-mean Gaussian random variables) used in Figure 10 is rotated through a variety of position angles, with the rotations calibrated such that the best-fit D05/unscaled TRK line in the first panel of each figure gives $\theta = 0$. Note that, while $R^2_{\rm TRK}(a,b) = R^2_{xy}$ in all cases, $R^2_{\rm TRK} > R^2_{xy}$. This is simply a reflection of the fact that, with TRK, we are free to explore a continuum of scales between the two extremes, and to choose an optimum scale that gives a best-fit line that lies in between the limiting fits of D05/D05*.

Figure 14 shows the behavior of $R^2_{\rm TRK}(a,b)$ and $R^2_{\rm TRK}$ for the same $m \approx 1$ data set as in Figures 10 & 11, for a range of intrinsic error bars. In the slop-dominated case (Figure 14a), as we discussed above, the limiting behavior of TRK as $s \to \infty$ and $s \to 0$ is the same as that of D05 under inversion; in this case, $R^2_{\rm TRK}(a,b) = R^2_{xy} = 0.435$, where $R^2_{\rm TRK}(a,b)$ is computed at the optimum scale for this data set ($s_0 = 0.8861$). In the intermediate cases (Figure 14b,c) the range of physically meaningful scales for TRK narrows as the intrinsic error bars become larger relative to the slop, and $R^2_{\rm TRK}(a,b)$ and $R^2_{\rm TRK}$ increase accordingly. Figure 14d shows the limiting case of data entirely dominated by intrinsic error bars, i.e., TRK returns fitted slops $\sigma_x = \sigma_y = 0$ at scale $s_0 = 1$, and would require unphysical $(\sigma_x^2, \sigma_y^2) < 0$ for fits at all other scales, so that $R^2_{\rm TRK} \equiv 1$. In other words, for a data set dominated entirely by intrinsic uncertainty, the TRK statistic can be trusted completely; any uncertainty in the fitted slope will be fully reflected in the computed probability distributions of the slope parameter m (or position angle parameter $\theta = \tan^{-1} m$; see

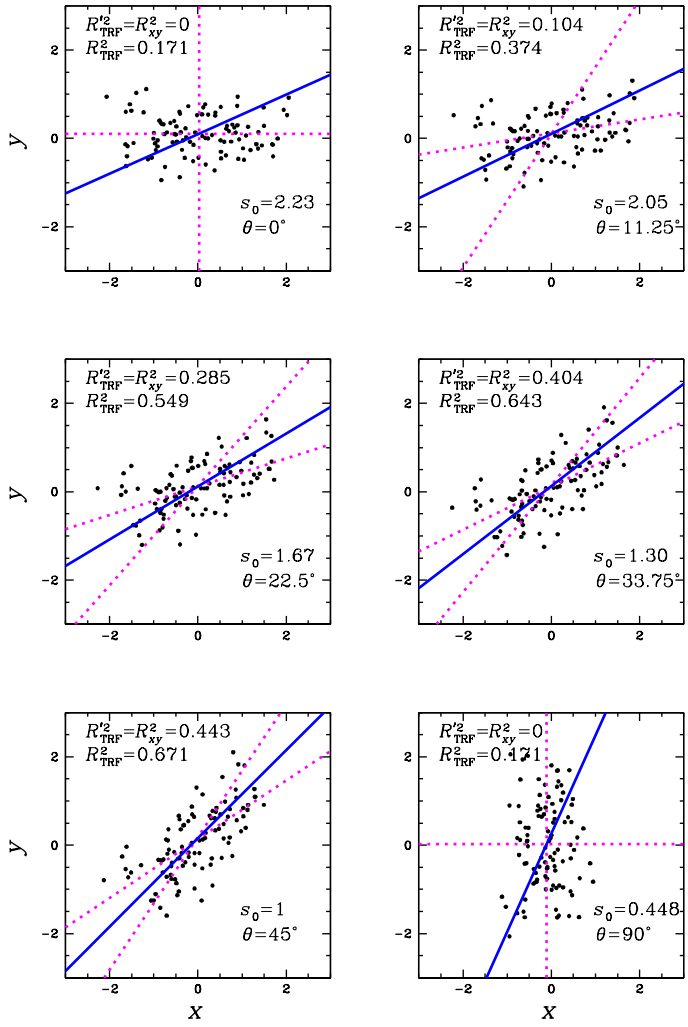


Figure 12. Equivalence of correlation coefficients $R^2_{\rm TRK}(a,b)$ and R^2_{xy} for linear fits to loosely correlated ($\sigma_x \approx \sigma_y \approx 0.5$) slop-dominated data with a range of slopes. The coordinates of each point are generated by rotation of the data points in Figure 10 over a range of angles. Rotation angles were chosen so that the best-fit D05/unscaled TRK fit gives $\theta = 0$ in the top left panel. Dashed-dotted lines: uninverted D05/low-scale TRK fit (shallow line), and inverted D05*/high-scale TRK (steeper line). Solid blue lines: TRK fits at the optimum scale s_0. For slop-dominated data, $R^2_{\rm TRK}(a,b) = R^2_{xy}$ at the optimum scale.

§2.6.1). Note that the behavior of D05 and inverted D05* is identical for all four cases, as is the value of $R^2_{xy} = 0.4348$: a value indicating that fits to these data sets with the D05 statistic should not be trusted. Finally, note that in all four cases, TRK fits at the optimum scale come closer to recovering the original $m = 1$ line from which these data sets were generated than either D05 or inverted D05*.

To summarize, for linear fits to entirely slop-dominated data, $R^2_{\rm TRK}(a,b) = R^2_{xy}$ when computed at the optimum scale s_0; the calibration of our intuition about R^2_{xy} carries over to $R^2_{\rm TRK}(a,b)$. The correlation coefficient $R^2_{\rm TRK}$ is always greater than $R^2_{\rm TRK}(a,b)$, because it takes advantage of the fact that scale is a continuum, while invertibility is a bimodal choice. Furthermore, fits to entirely slop-dominated data sets represent TRK's worst-case scenario. For data with non-zero intrinsic error bars, TRK requires unphysical $(\sigma_x^2, \sigma_y^2) < 0$ in order to reproduce the fits of D05, and the range of physically meaningful scales is reduced. Increasing the intrinsic uncertainty with respect to the extrinsic scatter only increases the reliability of TRK; fits to data that are entirely dominated by intrinsic uncertainty can be

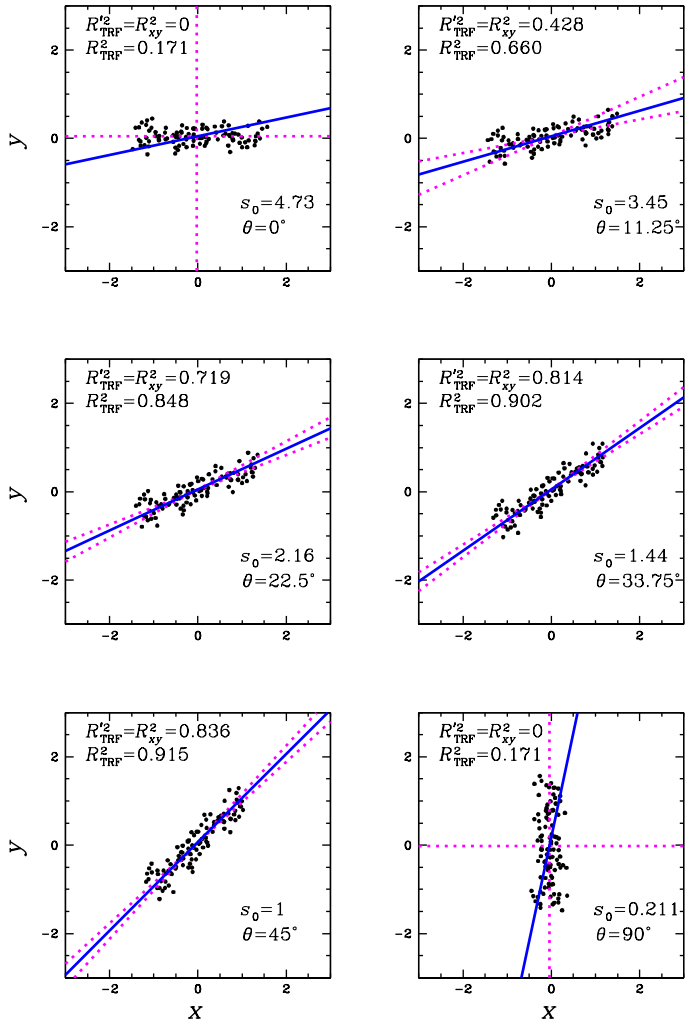


Figure 13. Equivalence of correlation coefficients $R^2_{\rm TRK}(a,b)$ and R^2_{xy} for linear fits to tightly correlated ($\sigma_x \approx \sigma_y \approx 0.2$) slop-dominated data with a range of slopes. The coordinates of each point are generated from the same zero-mean Gaussian random variables and random $y_c(x) = x$ data points as in Figure 10, but with standard deviation $\sigma_x = \sigma_y = 0.2$, rotated over a range of angles. Rotation angles were chosen so that the best-fit D05/unscaled TRK fit gives $\theta = 0$ in the top left panel. Dashed-dotted lines: uninverted D05/low-scale TRK fit (shallow line), and inverted D05*/high-scale TRK (steeper line). Solid blue lines: TRK fits at the optimum scale s_0. For slop-dominated data, $R^2_{\rm TRK}(a,b) = R^2_{xy}$ at the optimum scale.

trusted completely; any uncertainty in the fit will entirely be reflected in the computed probability distributions of the model parameters.

2.5.3. *TRK Fits to Non-Linear Data*

For non-linear fits to a set of N data points, we generalize the linear form of $R^2_{\rm TRK}(a,b)$ (Equation 43) as the average of N terms:

$$R^2_{\rm TRK}(a,b) \equiv \frac{1}{N}\sum_{n=1}^{N} \tan^2\left(\frac{\pi}{4} - \frac{|\theta_{t,n;a} - \theta_{t,n;b}|}{2}\right), \tag{46}$$

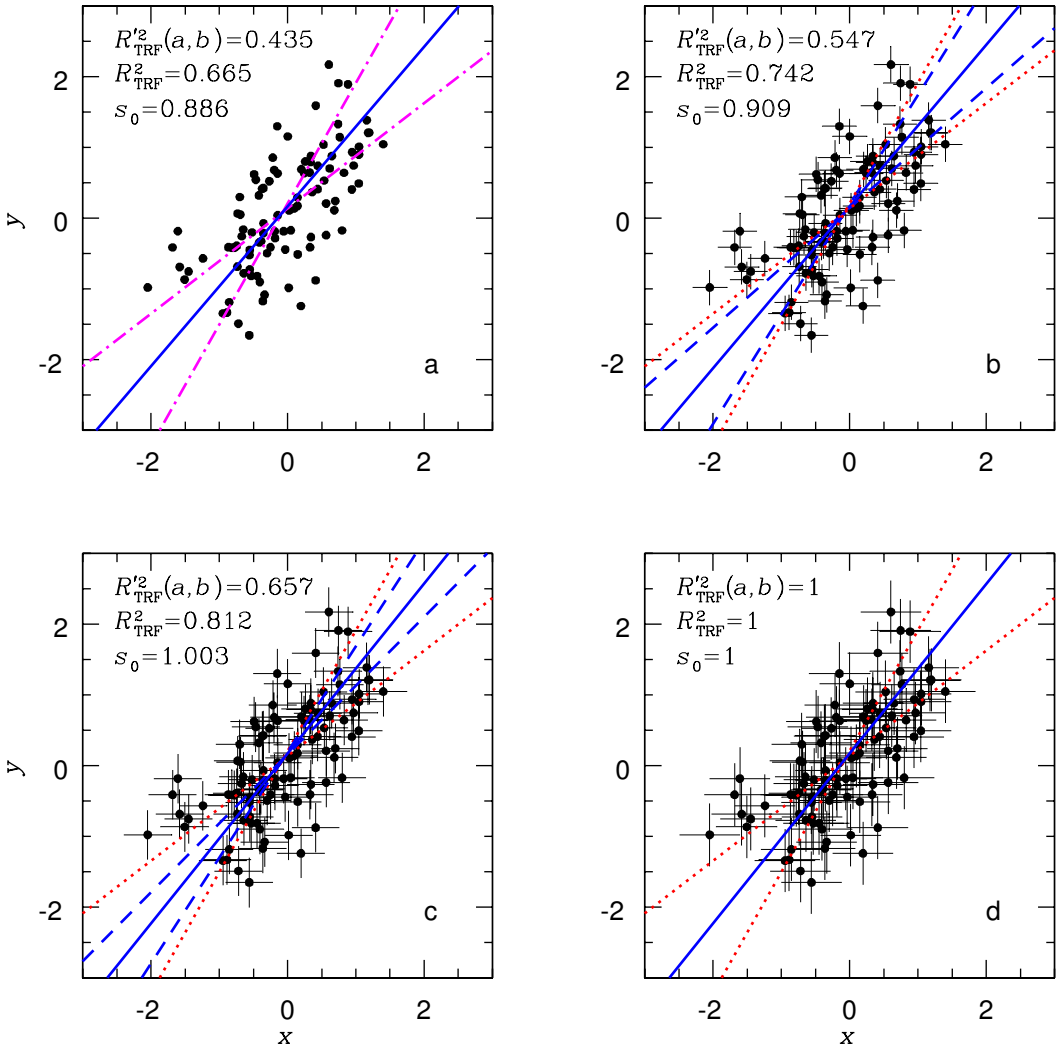


Figure 14. Comparison of correlation coefficients $R^2_{\rm TRK}(a,b)$ and $R^2_{\rm TRK}$ for linear fits to $m \approx 1$ data with a range of intrinsic uncertainties. The coordinates of each point are the same as in Figure 10. Each point was assigned symmetric intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0, 0.25, 0.35, 0.4435$. Dotted red lines: uninverted D05 fit (shallow line), and inverted D05* fit (steeper line); these fits are identical in all cases to those in Figure 10. Dotted blue lines: TRK fits at scale $s_{\rm min}$ at which fitted slop $\sigma_x \to 0$ (shallower line), and at scale $s_{\rm max}$ at which slop $\sigma_y \to 0$ (steeper line). Solid blue line: TRK fit at optimum scale. (a) is the same as the slop-dominated case of Figure 10; in this case $R^2_{\rm TRK}(a,b) = R^2_{xy} = 0.435$. (b) is the same as in Figure 11. (d) is the intrinsic uncertainty-dominated case, for which fitted slops in both directions, at all scales, equals zero, and $R^2_{\rm TRK} = 1$, meaning that the fit can be trusted completely; uncertainty in the slope will be fully reflected in its fitted probability distribution. Note that $R^2_{xy} = 0.435$ in all four cases.

Where $\theta_{t,n;a} = \tan^{-1} m_{t,n;a}$ and $\theta_{t,n;b} = \tan^{-1} m_{t,n;b}$ are the position angles of the best-fit curves at the tangent point to data point n at scales a and b, respectively. As with linear fits, we then iterate to find the optimum scale s_0 for which $R^2_{\rm TRK}(a, s_0) = R^2_{\rm TRK}(s_0, b) \equiv R^2_{\rm TRK}$.

Figure 15 illustrates the results of application of this iterative, pointwise, scale-dependent correlation coefficient to describe the non-linear correlation between observed interstellar extinction parameters R_V vs. c_2 for a sample of 441 stars in the Milky Way and Small and Large Magellanic Clouds, described in detail in Paper II. The solid line is the fit at the optimum scale s_0 for which $R^2_{\rm TRK}(s_{\rm min}, s_0) = R^2_{\rm TRK}(s_0, s_{\rm max})$, as measured in the s_0 scaled space. The dotted

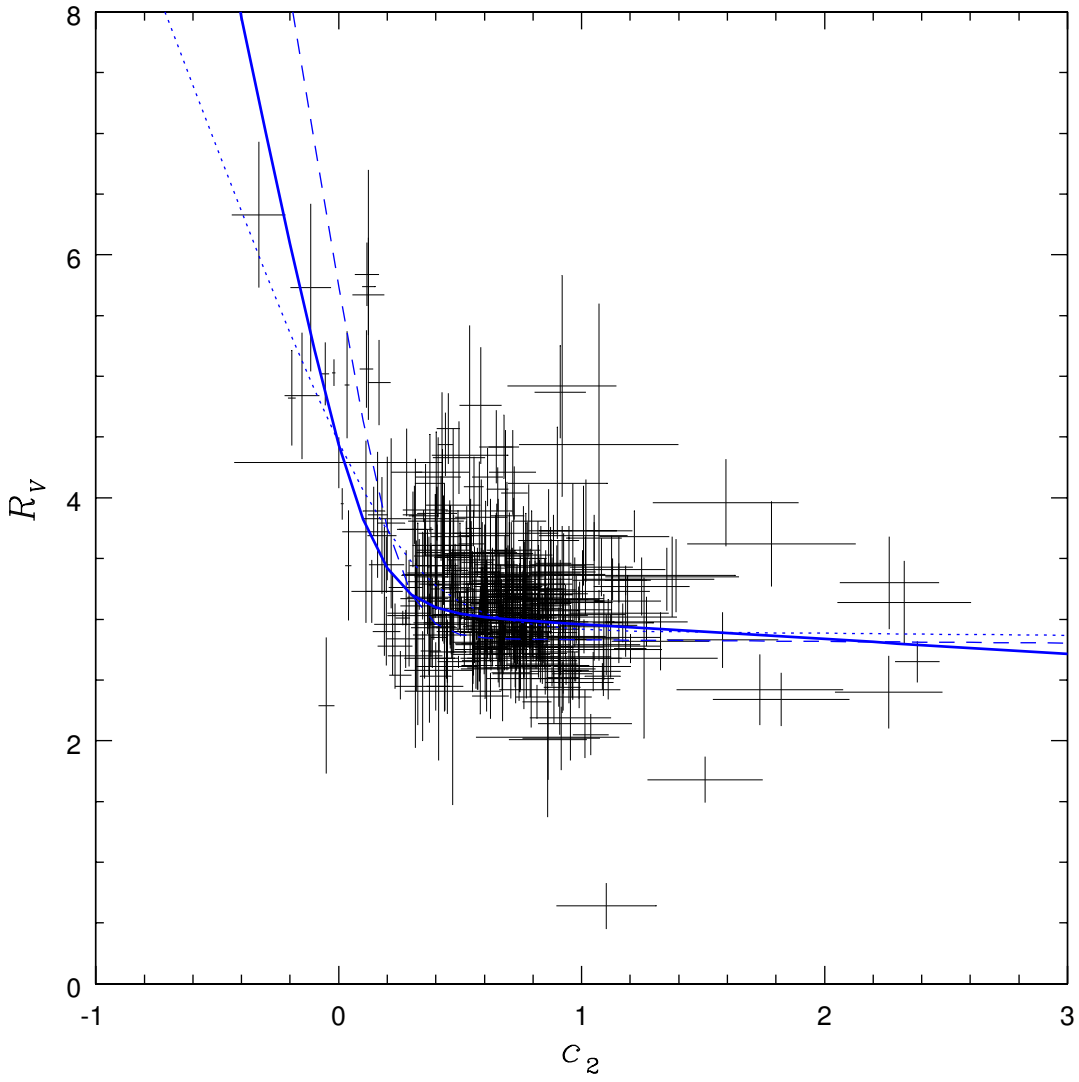


Figure 15. Example of TRK scale-dependence for a non-linear fit: observed interstellar extinction parameters R_V vs. c_2 (see §**??**). Dotted line: fit at $s_{\rm min}$, at which slop $\sigma_x \to 0$; Dashed line: fit at $s_{\rm max}$, the largest scale at which a physically meaningful best-fit curve can be obtained. Solid line is the fit at the scale s_0 for which $R^2_{\rm TRK}(s_{\rm min}, s_0) = R^2_{\rm TRK}(s_0, s_{\rm max})$. The best-fit slop envelopes for this model distribution are not shown here; see §**??** for a more thorough discussion of this fit.

line is the best TRK fit at the scale $s_{\rm min}$ at which slop $\sigma_x \to 0$. The dashed line is the best TRK fit at the largest scale $s_{\rm max}$ for which TRK recovers a physically meaningful curve.

This $s_{\rm max}$ is a smaller scale than the one at which the slop $\sigma_y \to 0$. Beyond $s_{\rm max}$, all best-fit curves, including local minima, discontinuously jump to a new family of curves where most of the data are fit by a line, but the right- or leftmost data are fit by a vertical segment (Figure 16). Such curves are the only way to accommodate the extrinsic scatter in the data with a model distribution that places slop primarily in the x-direction. Admittedly, since slop is purely an empirical parameterization of extrinsic scatter in the data at a given scale, these model *distributions* are empirically valid fits to the data at large scales. The problem is simply that the underlying *curves* are unphysical, i.e., there is simply no plausible physical distribution of dust that could produce such relationships between R_V and c_2.

Finally, note that in the region $c_2 \gtrsim 1$, the slopes of the limiting curves in Figure 15 are close to zero. If, instead of the correlation coefficient $R^2_{\rm TRK}$ (Equation 46), we were to choose a correlation coefficient analogous to R^2_{xy}, defined in terms of an ensemble average of the ratios of tangent point slopes at different scales, this could result in terms

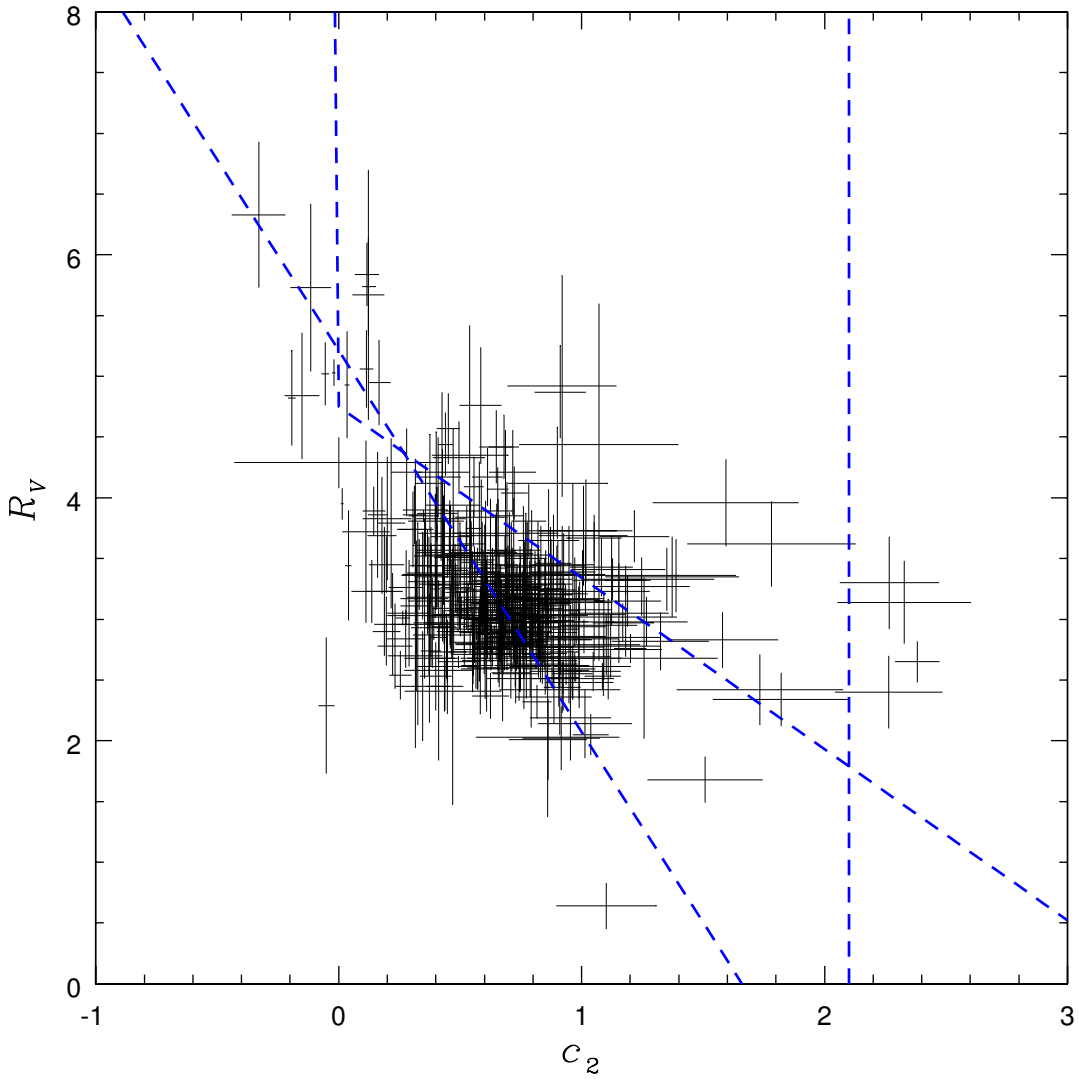


Figure 16. Examples of unphysical TRK fits to R_V vs. c_2 at scales $s > s_{\max}$, for which the slop $\sigma_y \to 0$. No physically plausible distributions of dust could result in such relations. The best-fit slop envelopes for this model distribution are not shown here; see Paper II for a more thorough discussion of this fit.

that approach either zero or infinity, effectively under- or over-weighting those data points relative to the others in the ensemble. As we discussed in §2.5.1, this is precisely the situation that $R^2_{\rm TRK}$ was designed to handle, by defining the correlation of two fitted curves in terms of the absolute difference of their tangent point position angles, rather than the ratio of their tangent point slopes.

2.6. *Model Parameter Estimation*

2.6.1. *Computing Probability Distributions and Error Bars of Fitted Parameters*

It is often assumed that the uncertainty in a fitted parameter is described by a $\Delta\chi^2$ distribution, as we had to do in the context of nested model comparisons in §**??**. However, in practice, this is not always the case, and may actually be a rarity.

The proper way to compute the probability distribution $p(\vartheta_i)$ of a fitted parameter ϑ_i is to compute the *marginalized probability distribution*. This requires marching through a dense grid over the entire M-dimensional parameter space, computing the posterior probability at each grid point, and then integrating over all dimensions except that of the

parameter in question:

$$p^{\mathrm{m}}(\vartheta_i) = \int_{\vartheta_1} \cdots \int_{\vartheta_{i-1}} \int_{\vartheta_{i+1}} \cdots \int_{\vartheta_M} p(\vartheta_m|DI) d\vartheta_m \,. \tag{47}$$

This is computationally impractical for $M \gtrsim 5$.

Instead, we use the well known Metroplis-Hastings Markov Chain Monte Carlo Method (MCMC, e.g. Hastings (1970)) to sample the actual joint posterior distribution of the model parameters (including slop), given $\mathcal{L}^{\mathrm{TRK}}$ and any priors of choice, following Bayes' Theorem. For the Markov Chain iteration, we use Gaussian proposal distributions for each of the M parameters with standard deviations of $\sigma_j = \sigma_1, \sigma_2, \ldots, \sigma_M$. In order to determine the best fit values of model parameters, including slop, we maximize the TRK likelihood (Equation 21) using a slightly modified version of the downhill simplex method of Nelder & Mead (1965).

One thing to note is that for this method, the success of the posterior sampling is very much dependent on the values of σ_j. As an example, for a simple single parameter model, a σ_1 that is too low will result in the Markov Chain not properly exploring the parameter space, resulting in a sample that is overly concentrated in one area. On the other hand, if σ_1 is too high, the sampler will "jump" around the parameter space, resulting in too discrete a sampling. A good way to measure the approximate success of this sampling is the Metropolis-Hastings acceptance ratio R, which is defined to be simply the number of trial points in the Markov Chain that were accepted to be added to the total sample, divided by the total number of trial points used. It follows that the low-σ_1 case corresponds to too high an R, while the high-σ_1 case corresponds to too low an R. As such, in order to properly sample the posterior, an optimal set of σ_j must be found.

In order for our algorithm to be easily generalizable, it is preferable for the selection of σ_j to be automated, so that the user does not need to know the best σ_j *a priori*, or find them through trial and error. A frequently used "rule of thumb" is that a good sample will have $R \simeq 0.44$. Following experimentation, we have converged on using $0.4 < R < 0.5$ as an optimal acceptance ratio. For the 1D case, R can be easily optimized with respect to σ_1, using any traditional 1D optimization method. However, for the case of any M-dimensional parameter space, finding a method to optimize R with respect to all M of the σ_j is more difficult. Given that R can not be represented explicitly as a function of σ_j, a number of multidimensional optimization methods are ruled out, such as the Gauss-Newton algorithm. We use the Adaptive Metropolis-Hastings method of Haario et al. (2001) to optimize σ_j (technically optimizing the covariance matrix and mean of the MCMC proposal distribution) due to its speed and generalizability, and it has proven to be a powerful and efficient tool to obtain good values of σ_j, even with higher-dimension parameter spaces.

Once the optimial σ_j are determined, we sample until we have a total of 100,000 points (excluding the initial 10,000 "burn-in" points that are discarded) in parameter space, which has proven to be plenty for the sample to converge to a strong approximation of the true posterior distribution. From here, we construct tabulated histograms of the samples for each parameter.

2.6.2. *Computing Error Bars of Fitted Parameters*

Assuming that the parameters follow symmetric or asymmetric Gaussian distributions, we are now in a position to compute projected confidence intervals, or error bars, on the best-fit values for the parameters. A 1-, 2- or 3σ confidence interval corresponds to a range of parameter values for which the integrated probability distribution is 68.2639%, 95.4500% or 99.7300% of the total integrated area under the distribution, respectively. Note that, strictly speaking, there are an infinite number of such ranges for each confidence level, depending on where one chooses to place the limits of integration. What we are interested in are the *most compact* confidence intervals for a given probability distribution.

The most compact confidence intervals are obtained by a technique we refer to as "lowering the bar". Figure 18 illustrates this concept, for a hypothetical bimodal probability distribution. Starting at the peak of the distribution, which corresponds to the best-fit parameter value, we lower an imaginary horizontal bar of constant $p(\vartheta_i)$ in small steps. At each step, we numerically interpolate the tabulated histogram values of $p(\vartheta_i)$ to find the values of ϑ_i for which $p(\vartheta_i)$ is equal to the height of the horizontal bar, and numerically integrate the area under the probability distribution between those values. We adjust the height of the horizontal bar until the this area equals 68.2639%, 95.4500% and 99.7300% of the total integrated area under the tabulated probability distribution, which gives us the most compact 1-, 2- and 3σ confidence intervals of parameter ϑ_i.

Note that, as in the example of Figure 18, there are times when one or more of these most compact confidence intervals may be disjoint. In this example, there are, in fact, two such 1σ confidence intervals, each roughly centered

on the two peaks in the bimodal distribution (however, the 2- and 3σ confidence intervals are contiguous). In such cases, one must be careful in reporting the best-fit parameter value and its uncertainty; strictly speaking, describing this distribution at the 1σ confidence level would require reporting two "best-fit" values of ϑ_i, each with its own error bars and relative weighted probability. In such cases, it is often better to present instead the actual probability distribution graphically, or in tabular form, with confidence intervals labeled as in Figure 18. If this probability distribution were intended to serve as a prior on ϑ_i in other model fits, we would most likely implement it by actually fitting a sum of two Gaussians model to the distribution. Fortunately, we find that in many cases, projected probability distributions fall monotonically from the best-fit peak, and can usually be well approximated by a Gaussian, or asymmetric Gaussian distribution (see, e.g., Figure **??** in §**??**). Utilizing such analytic approximations to prior probability distributions in our model-fitting algorithms can provide a significant advantage in numerical speed.

2.6.3. *Correlated Uncertainties in Fitted Parameters and Pivot Points*

We make efforts to minimize correlation among fitted parameters. For example, if we assume a linear model of the form $y(x) = mx + b$, the best-fit slope m will typically be correlated with the fitted intercept b. This is due entirely to the arbitrary choice of placing the intercept at $x = 0$; depending on the actual x-range of the data set, this choice can introduce a kind of lever arm effect whereby fitted m depends on b and vice versa. This correlation can be minimized by choosing a particular "pivot point", x_p, somewhere in the middle of the data range, so that $y(x) = m(x - x_p) + b_p$.

The optimum pivot point can be found by marching in m in a fine grid very close to the best-fit, refitting to b at each step. Consider two such linear fits, with best-fit parameters (b_1, m_1) and (b_2, m_2) (Figure 19). The pivot point x_p corresponds to the value of x at which these two lines intersect:

$$\begin{aligned} m_1 x_p + b_1 &= m_2 x_p + b_2 \\ x_p &= -\frac{b_1 - b_2}{m_1 - m_2} \equiv -\mu \,. \end{aligned} \tag{48}$$

By extension, when we plot the best-fit b_i vs. m_i for all the fits in the march, we find a clear linear correlation, with slope μ. The best choice of pivot point, to minimize the correlation between slope and intercept, is $x_p = -\mu$, so that $b_p = b - m\mu$. We then conduct the error bar marches with the model $y(x) = m(x - x_p) + b_p$ (holding x_p fixed) to determine projected probability distributions for m and b_p.

For more complicated, non-linear models, we can only attempt to parameterize them, and choose a coordinate basis, in such a way that certain segments are asymptotically linear, and analyze independently the fitted model parameters that describe these segments to find their individual pivot points. In practice, this minimizes the correlation between slope and intercept of each approximately linear segment, though it does not address the issue of cross-correlation among the multiple segments, or correlations among parameters that describe non-linear regions of the curve.

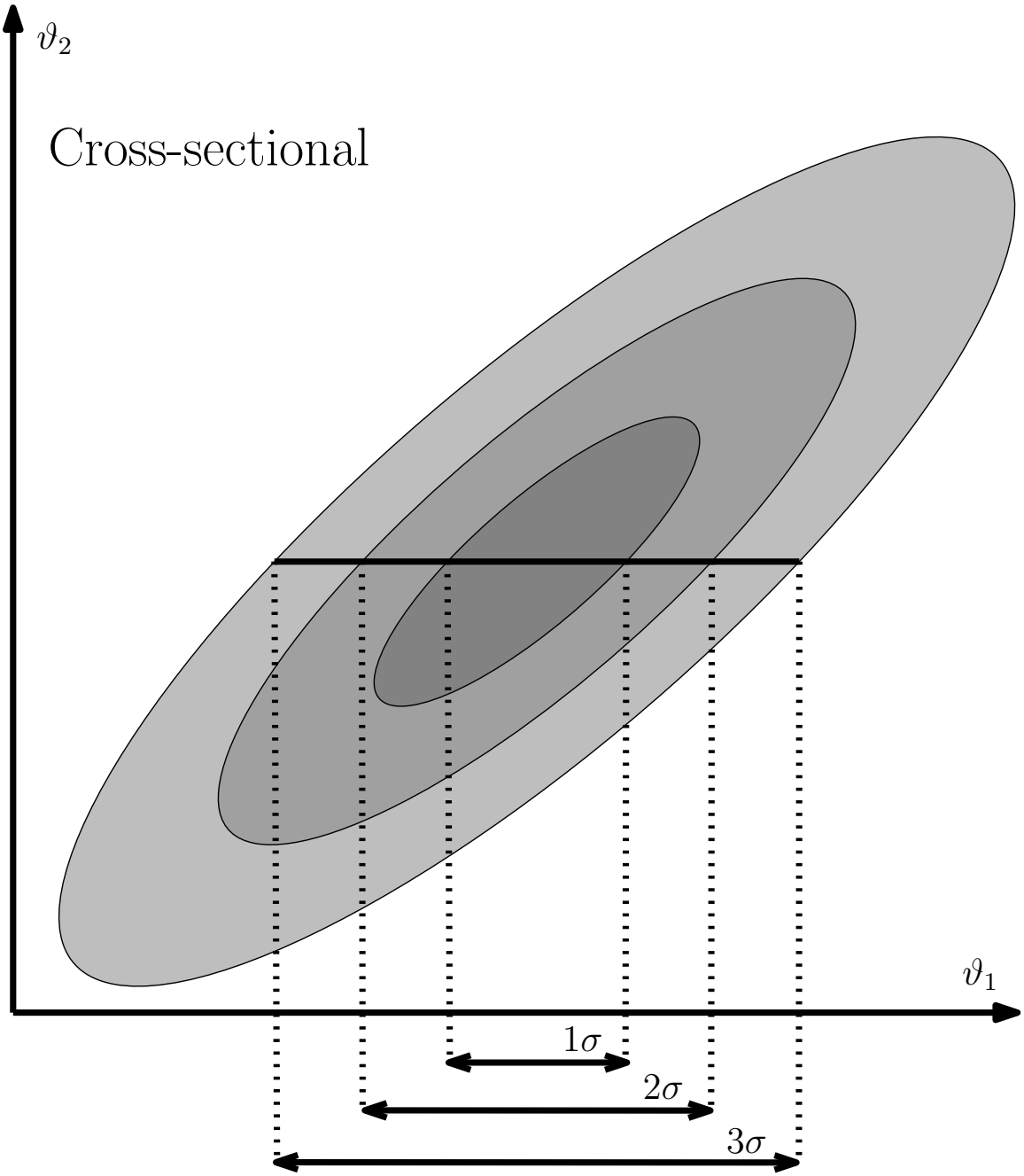
To determine the pivot point for some linearized model in practice, we take the following iterative approach. Given some initial guess for the pivot, we draw from the posterior distribution of the linearized model parameters with an MCMC sampler, and from this, draw a sufficiently large number of 2-combinations of sets of parameters. Then, for each of these combinations, we can compute a new pivot point using the parameters and the previous value of the pivot point, and weight them according to their uncertainties.[7] Finally, we take the weighted half-sample mode (e.g. Bickel & Frühwirth (2005)) of the generated pivot point distribution, use this as the next iteration of the pivot point, and repeat the cycle until convergence is achieved.

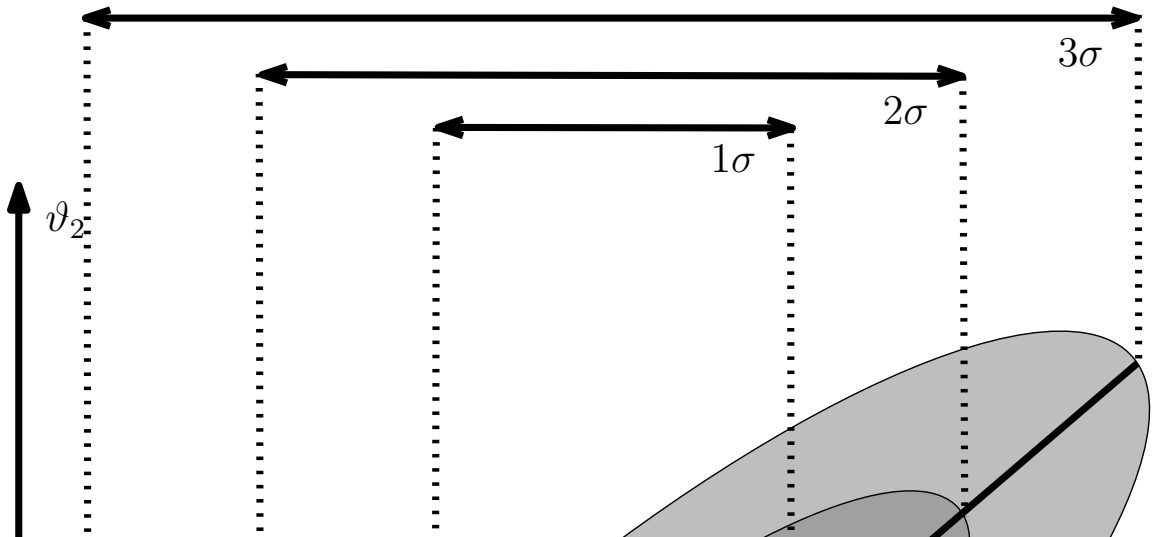
2.7. *Asymmetric Probability Distributions*

2.7.1. *Choice of Basis and Basis Transformations*

In all the preceding discussion of fitting model distributions to two-dimensional data sets, we have assumed that both the intrinsic uncertainties (error bars) and the extrinsic scatter (slop) are normally distributed in the space in which we conduct the fit. This assumption considerably simplifies the resulting statistics, allowing us to approximate the various probability integrals analytically. However, there are times when either the intrinsic uncertainty or the extrinsic scatter of the data in one or both dimensions are *not* both well described by a symmetric, Gaussian probability distribution

[7] For some pivot point, the weight w is defined to be proportional to the inverse square of the uncertainty of the pivot point, which is found via standard propagation of error. Assuming equal uncertainties for all sampled intercept parameters σ_b and the same for all sampled slope parameters σ_m, it follows that any dependence of on σ_b and σ_m within w cancels, such that only the parameter values themselves (as well as the value of the previous pivot point) are needed for the weight computation.


ϑ_2
Cross-sectional
ϑ_1
1σ
2σ
3σ


3σ
2σ
1σ
ϑ_2

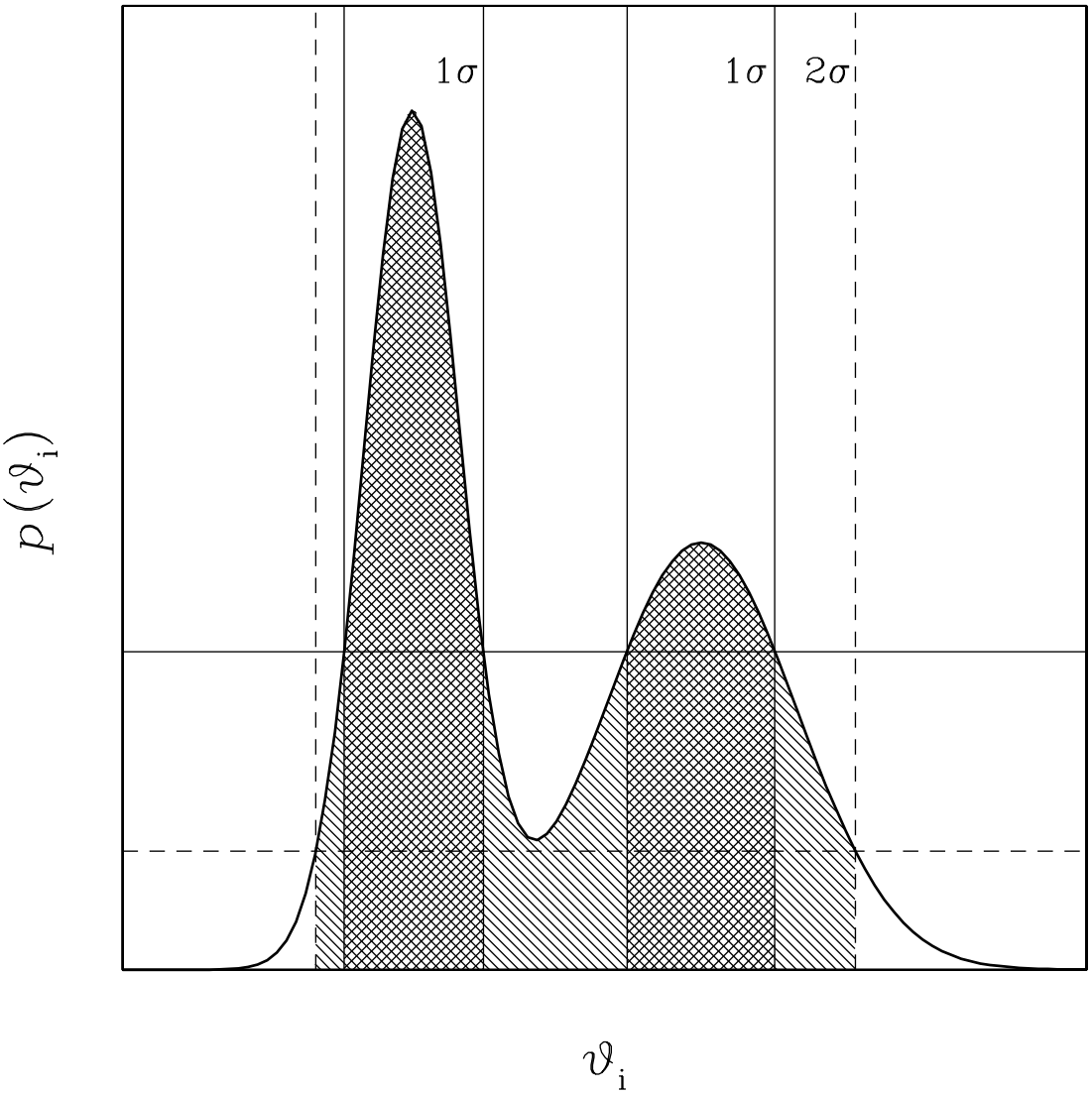


Figure 18. Illustration of confidence intervals for a bimodal fitted parameter probability distribution $p(\vartheta_i)$. Shaded regions show the most compact 1- and 2σ confidence intervals; the area of the shaded regions are 68.2689% and 95.4500% of the total area under the curve. This distribution has two, disjoint 1σ confidence intervals, but a single 2σ interval. The area of 3σ confidence region (not shown) would correspond to 99.7300% of the total area.

function in the same space, or *basis*. For example, in astronomy we typically measure the brightness of an object using a CCD camera, which is an intrinsically linear flux measuring instrument. The flux is measured in the form of photon counts in each CCD pixel element, and the uncertainty in the flux follows a Poisson distribution, which approaches a Gaussian distribution in the limit of count numbers $N \gtrsim 10$, so the intrinsic uncertainty in our measurements, to a good approximation, has a symmetric Gaussian probability distribution in linear flux space. However, the calibration of our flux measurements is typically done by comparison to standard stars in magnitude space, and uncertainties in that calibration are assumed to be normally distributed in magnitude, or log flux, space. Furthermore, the unmodeled physical processes that contribute to the extrinsic scatter (slop) in the observed flux of many astronomical sources are frequently assumed to have a normal distribution in log flux space as well. So, while the data have normal intrinsic

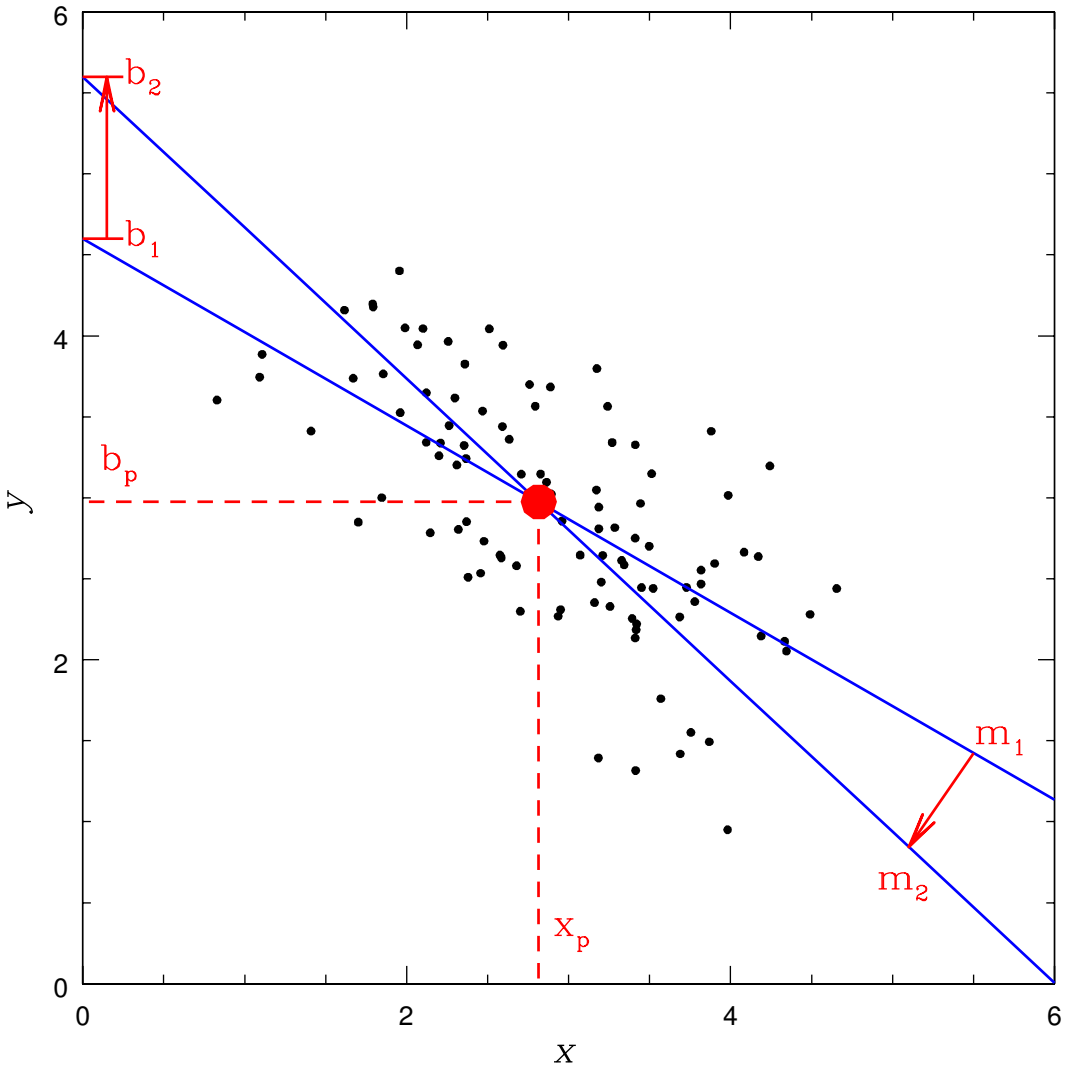


Figure 19. Illustration of linear pivot point determination. The two lines $y = m_1 x + b_1$ and $y = m_2 x + b_2$ are fits that might be obtained in an error bar march in the slope parameter m. Clearly, any uncertainty in m will be correlated with uncertainty in the intercept b, and vice versa. If we instead fit to $y = m(x - x_p) + b_p$, where the pivot point $x_p = -(b_1 - b_2)/(m_1 - m_2)$, correlations between uncertainties in m and b_p will be minimized.

uncertainties in linear flux, the model distribution is normally distributed in log flux; since we must choose one or the other basis in which to conduct the fit, we must transform one or the other of the distributions into that space. [8]

Most generally, if a probability distribution function in a particular basis x has the functional form $p(x)$, then in the transformed basis $f(x)$, conservation of probability requires that the transformed distribution $p(f(x))$ takes the form:

$$p(f(x))df(x) = p(x)dx$$
$$p(f(x)) = p(x)\left(\frac{df(x)}{dx}\right)^{-1}. \quad (49)$$

[8] For example, in Paper II we describe fits to observations of fading gamma-ray burst afterglows, where the model distribution has normally distributed slop in log flux space (corresponding to unmodeled stochastic variations in the GRB emission), but is fit to photometric CCD data with normally distributed intrinsic uncertainties in linear flux space. This requires that we transform the log-normal model distribution into that linear space. Besides being in line with an overall philosophy of "bringing the model to the data", this approach also allows us to easily include non-detections (typically reported as upper limits in log flux, or magnitude, space) in our data set; an upper limit in log flux may be approximated as a zero-mean Gaussian in linear flux space (See §2.7.3).

For example, if the model distribution is Gaussian in $\log x$ space, with a peak at $(\log x)_0$ and a width $\sigma_{\log x}$:

$$p(\log x) = G(\log x; (\log x)_0, \sigma_{\log x}) \,, \tag{50}$$

then, in linear x space, $p(x)$ will take the form:

$$\begin{aligned} p(x) &= G(\log x; (\log x)_0, \sigma_{\log x}) \left(\frac{d \log x}{dx} \right) \\ &= G(\log x; (\log x)_0, \sigma_{\log x}) \left(\frac{1}{x \ln 10} \right) \,. \end{aligned} \tag{51}$$

In the next section, we show that the pdf in Equation 51 can be approximated by an asymmetric Gaussian function for small to moderate values of $\sigma_{\log x}$.

2.7.2. *Asymmetric Gaussian Probability Distributions*

It can be shown analytically that, in linear x space, the transformed log-normal $p(\log x)$ of Equation 51 is a maximum at:

$$x^{\max} = 10^{((\log x)_0 - \sigma^2_{\log x} \ln 10)} \,. \tag{52}$$

In other words, while the log-normal distribution $p(\log x)$ is defined to peak at $(\log x)_0$, the transformed linear distribution $p(x)$ does *not* peak at $10^{(\log x)_0}$, but rather at a smaller value that is a function of the slop. If the slop $\sigma_{\log x} \lesssim 0.1$, $p(x)$ can be well approximated as a *skew-normal*, or *asymmetric Gaussian* function that peaks at $x^{\max}$ (Equation 52):

$$p(x) \approx G_A(x; x^{\max}, \sigma_{x\pm}) \,, \tag{53}$$

where:

$$G_A(x; x^{\max}, \sigma_{x\pm}) \equiv \frac{2}{\sqrt{2\pi}(\sigma_{x+} + \sigma_{x-})} \begin{cases} \exp\left[-\frac{1}{2} \left(\frac{x - x^{\max}}{\sigma_{x+}} \right)^2 \right] & \text{if } x \geq x^{\max} \\ \exp\left[-\frac{1}{2} \left(\frac{x - x^{\max}}{\sigma_{x-}} \right)^2 \right] & \text{if } x < x^{\max} \end{cases} \,, \tag{54}$$

and where the 1σ widths of the asymmetric Gaussian are given by the empirical formulas:

$$\begin{aligned} \sigma_{x+} &\approx x^{\max}(2.3029 \sigma_{\log x} + 2.6293 \sigma^2_{\log x} - 3.6945 \sigma^3_{\log x}) \\ \sigma_{x-} &\approx x^{\max}(2.3027 \sigma_{\log x} - 2.6544 \sigma^2_{\log x} - 4.0699 \sigma^3_{\log x}) \,. \end{aligned} \tag{55}$$

We emphasize that these are, indeed, *approximations*, and only valid for cases where the slop parameter $\sigma_{\log x} \lesssim 0.1$. In general, the transformed distribution's peak, and the most compact 1σ confidence interval around that peak must be found numerically. In some cases, the transformed probability distribution may then be expressed in terms of an asymmetric Gaussian distribution with a peak x_0 and widths $\sigma_\pm$ that reflect that confidence interval. Other functional transformations of basis may yield distributions that are not well approximated by asymmetric Gaussian distributions, and must be treated with care. Strictly speaking, the proper approach to dealing with generalized transformed, non-normal distributions is to evaluate the probability integrals numerically; unfortunately, this is prohibitively computationally expensive.

In cases where the model distribution, or a data point's intrinsic probability distribution, or both can be approximated by asymmetric Gaussian functions, we have formulated an empirical approximation for the joint probability. The expression for the joint probability of data point n with the model distribution (Equation 5) takes the form:

$$\begin{aligned} &p_n(\vartheta_m, \sigma_{x\pm}, \sigma_{y\pm} | x_n, y_n, \sigma_{x,n\pm}, \sigma_{y,n\pm}) = \\ &\int_{x'} \int_{y'} \int_{u_n} \int_{v_n} dv_n du_n dy' dx' f(x', y') g(x, y) \delta(v_n - v_{c,n}(u_n; \vartheta_m)) \times \\ &\quad G_A(x', x, \sigma_{x\pm}) G_A(y', y, \sigma_{y\pm}) G_A(x', x_n, \sigma_{x,n\pm}) G_A(y', y_n, \sigma_{y,n\pm}) \,. \end{aligned} \tag{56}$$

The x' and y' integrals can be shown to be equivalent to convolutions, i.e.:

$$\int_{x'} G_A(x', x, \sigma_{x\pm}) G_A(x', x_n, \sigma_{x,n\pm}) dx' \equiv G_A(x, 0, \sigma_{x\mp}) * G_A(x, x_n, \sigma_{x,n\pm}) \,. \tag{57}$$

In Appendix A, we describe an approximation of the convolution of two asymmetric Gaussian functions as a single asymmetric Gaussian:

$$G_A(x,0,\sigma_{x\mp}) * G_A(x,x_n,\sigma_{x,n\pm}) \approx G_A(x,x_n+\delta_{x,n},\Sigma_{x,n\pm})\,, \tag{58}$$

where

$$\Sigma^2_{x,n\pm} \equiv \sigma^2_{x\mp} + \sigma^2_{x,n\pm}\,, \tag{59}$$

and where $\delta_{x,n}$ is an empirical function of $\sigma_{x\mp}$ and $\sigma_{x,n\pm}$. In other words, the x' and y' integrals in the joint probability are approximately equivalent to shifting the data point centroid to $(x_n+\delta_{x,n}, y_n+\delta_{y,n})$, and adding the slops $(\sigma_{x\mp},\sigma_{y\mp})$ to the intrinsic error bars $(\sigma_{x,n\pm},\sigma_{y,n\pm})$ in quadrature.

For the one-dimensional, χ^2-like statistic we employ to fit light curves to GRB flux measurements, the joint probability then takes the simple form:

$$p_n \approx \frac{2}{\sqrt{2\pi}(\Sigma_{F_\nu,n+}+\Sigma_{F_\nu,n-})} \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{F_{\nu,n}+\delta_{F_{\nu,n}}-F^{\max}_\nu}{\Sigma_{F_{\nu,n}+}}\right)^2\right] & \text{if } F_{\nu,n}+\delta_{F_{\nu,n}} \geq F^{\max}_\nu \\ \exp\left[-\frac{1}{2}\left(\frac{F_{\nu,n}+\delta_{F_{\nu,n}}-F^{\max}_\nu}{\Sigma_{F_{\nu,n}-}}\right)^2\right] & \text{if } F_{\nu,n}+\delta_{F_{\nu,n}} < F^{\max}_\nu \end{cases}, \tag{60}$$

where $F^{\max}_\nu$ is the peak of the transformed log-normal model distribution (Equation 52).

For two-dimensional statistics, like TRK, evaluating the joint probability of an asymmetric model distribution with the asymmetric intrinsic uncertainty of a data point is somewhat more complicated. We proceed, as before, by approximating the model curve $y_c(x;\vartheta_m)$ by a line tangent to the asymmetric convolved (and shifted) error "ellipse", taking care to keep track of which quadrant the tangent point lies in. However, since the tangent line will, in general, cross through three quadrants of the asymmetric convolved ellipse, the joint probability integral over the range $[-\infty,\infty]$ of Equation 13 must be broken into three segments. We discuss this procedure in Appendix B.

2.7.3. *Treatment of Detections and Upper Limits in* $\log F_\nu$ *Space*

In our own GRB afterglow photometry, flux density measurements are reported in linear F_ν space, with symmetric Gaussian error bars. But we often need to plot these measurements in $\log F_\nu$ (or, perhaps, magnitude) space. If the intrinsic pdf of a flux density measurement has the symmetric Gaussian form $G(F_\nu;F_{\nu,n},\sigma_{F_{\nu,n}})$ in linear F_ν space, then, by Equation 49:

$$\begin{aligned} p(\log F_\nu) &= G(F_\nu;F_{\nu,n},\sigma_{F_{\nu,n}})\left(\frac{d\log F_\nu}{dF_\nu}\right)^{-1} \\ &= G(F_\nu;F_{\nu,n},\sigma_{F_{\nu,n}})\,(F_\nu \ln 10)\,. \end{aligned} \tag{61}$$

It can be shown analytically that the pdf of Equation 61 is a maximum at:

$$\log F^{\max}_{\nu,n} = \log\left[\frac{1}{2}\left(F_{\nu,n}+\sqrt{F^2_{\nu,n}+4\sigma^2_{F_{\nu,n}}}\right)\right]. \tag{62}$$

If a measurement $F_{\nu,n}$ has a sufficiently high signal-to-noise ratio, $p(F_\nu>0)\geq 0.9973$, or:

$$\int_0^\infty G(F_\nu;F_{\nu,n},\sigma_{F_{\nu,n}})dF_\nu \geq 0.9973\,, \tag{63}$$

then we transform the distribution using Equation 61. In other words, if the area under the normalized Gaussian for $F_\nu>0$ is equal to or greater than the area of the 3σ confidence interval, we treat the data point as detection. The transformed data point is plotted with a centroid value $\log F^{\max}_{\nu,n}$, given by Equation 62, and with asymmetric (non-Gaussian) error bars that must be determined numerically, using the "bar-lowering" technique of §2.6.2.

Measurements with low signal-to-noise ratios, $p(F_\nu>0)<0.9973$, or:

$$\int_0^\infty G(F_\nu;F_{\nu,n},\sigma_{F_{\nu,n}})dF_\nu < 0.9973\,, \tag{64}$$

are plotted as upper limits in $\log F_\nu$ space. The 3σ upper limit $\log F^{3\sigma}_{\nu,n}$ is the value for which:

$$\int_{-\infty}^{F^{3\sigma}_{\nu,n}} G(F_\nu;F_{\nu,n},\sigma_{F_{\nu,n}})dF_\nu = 0.9973\,. \tag{65}$$

Similarly, the 2σ and 1σ upper limits, $\log F_{\nu,n}^{2\sigma}$ and $\log F_{\nu,n}^{1\sigma}$, are those for which the integral in Equation 65 equals 0.9545 and 0.6827, respectively. Note, however, that by this definition, the 1-, 2- or 3σ limits may not exist in $\log F_\nu$ space. For instance, if $F_{\nu,n}^{1\sigma} < 0$, then $\log F_{\nu,n}^{1\sigma}$ is undefined. (In the rare case that even the 3σ upper limit is undefined, it may only be possible to plot that data point as a 4σ, or higher, limit in $\log F_\nu$ space.)

Again, we only transform our flux measurements into $\log F_\nu$ space for the purposes of data visualization. All GRB afterglow fits are conducted in linear F_ν space, using the approximation of the transformed log-normal model distribution discussed above, in §2.7.2. We chose this approach precisely because it allows us to include low signal-to-noise measurements, and even non-detections, in our data sets. In $\log F_\nu$ space, the intrinsic probability distributions of these measurements are highly asymmetric, and impossible to treat analytically – even leaving aside the fact that any fraction of the distribution lying below $F_\nu = 0$ is "thrown away" in the transformation from linear to log space.

We often are presented with the reverse scenario: measurements quoted as upper limits in $\log F_\nu$ (or magnitude) space in the literature. Most often, a 3σ upper limit $\log F_{\nu,n}^{3\sigma}$ is provided. In the absence of any other information about how they were computed, the best we can do is to assume that these upper limits correspond to zero-mean Gaussian distributions in linear F_ν space. The 1σ width of that distribution is the value of $\sigma_{F_{\nu,n}}$ for which:

$$\int_{-\infty}^{F_{\nu,n}^{3\sigma}} G(F_\nu; 0, \sigma_{F_{\nu,n}}) dF_\nu = 0.9973\,. \tag{66}$$

By this definition, $\sigma_{F_{\nu,n}} = F_{\nu,n}^{3\sigma}/2.7822$. Analytically, if 2- or 1σ upper limits are given, then $\sigma_{F_{\nu,n}} = F_{\nu,n}^{2\sigma}/1.6901$ or $F_{\nu,n}^{1\sigma}/0.4752$, respectively.[9]

ACKNOWLEDGEMENTS

thank u, next

APPENDIX

A. CONVOLUTION OF TWO ASYMMETRIC GAUSSIAN DISTRIBUTIONS

A.1. *Asymmetric 1D Probability Distributions*

In §2.1 we discuss the general problem of computing the joint probability of a model distribution, $p^{\rm mod}$, and a single measurement with an intrinsic probability distribution $p_n^{\rm int}$. Here, we consider the simpler case of a 1D statistic (we discuss the treatment of asymmetric distributions in a 2D statistic in Appendix B). The model distribution is described by some curve $y_c(x;\vartheta_m)$, which is a function of the independent variable x and some set of parameters $\{\vartheta_m\}$, and by some probability density function $p^{\rm mod}(y; y_c(x;\vartheta_m))$ about that curve. The measured data point is described by some value, y_n, and a probability density function $p_n^{\rm int}(y; y_n)$. The joint probability of the model distribution and the measurement is given by:

$$p_n = \int_{-\infty}^{\infty} p^{\rm mod}(y; y_c(x;\vartheta_m)) p_n^{\rm int}(y; y_n) dy\,. \tag{A1}$$

This integral may or may not yield an analytic expression for p_n, depending on the functional forms of the two probability density functions.

The most familiar, and simplest, scenario is when both probabilities are normally distributed, i.e., $p^{\rm mod}(y; y_c(x;\vartheta_m)) = G(y; y_c(x;\theta), \sigma_y)$ and $p_n^{\rm int}(y; y_n) = G(y; y_n, \sigma_{y,n})$, where G is the Gaussian function:

$$G(y; y_n, \sigma_{y,n}) \equiv \frac{1}{\sqrt{2\pi}\sigma_{y,n}} \exp\left[-\frac{1}{2}\left(\frac{y - y_n}{\sigma_{y,n}}\right)^2\right]. \tag{A2}$$

In this case, the joint probability integral of Equation A1 also has a Gaussian form:

$$p_n = \frac{1}{\sqrt{2\pi}\Sigma_{y,n}} \exp\left[-\frac{1}{2}\left(\frac{y_n - y_c(x;\vartheta_m)}{\Sigma_{y,n}}\right)^2\right], \tag{A3}$$

[9] A common, but technically incorrect, practice is simply to assume that $\sigma_{F_{\nu,n}} = F_{\nu,n}^{3\sigma}/3$ (or $F_{\nu,n}^{2\sigma}/2$ or $F_{\nu,n}^{1\sigma}/1$). However, we err on the side of caution and, given a 3σ upper limit, employ the more conservative estimation, given by Equation 66, for the width of the zero-mean Gaussian pdf in linear F_ν space.

where $\Sigma_{y,n}$ is the sum in quadrature of the sample variance, or "slop", and the intrinsic uncertainty:

$$\Sigma_{y,n} \equiv \sqrt{\sigma_y^2 + \sigma_{y,n}^2}\,. \tag{A4}$$

The joint probability integral can equivalently be expressed in terms of a convolution of a zero-mean Gaussian $G(y;0;\sigma_y)$ and the data point's intrinsic Gaussian pdf, evaluated at $y = y_c(x;\vartheta_m)$:

$$p_n = \int_{-\infty}^{\infty} \left[G(y;0,\sigma_y) * G(y;y_n,\sigma_{y,n})\right] \delta(y - y_c(x;\vartheta_m))dy\,, \tag{A5}$$

where $\delta(y - y_c(x;\vartheta_m))$ is the Kronecker delta function.

There are many instances where one or both probability distribution functions are not normally distributed. For instance, when fitting to GRB afterglows, we evaluate the statistic in linear flux space, but assume the model distribution is log-normally distributed. Similarly, while proper flux measurement uncertainties from a CCD image should be roughly normally (or Poisson) distributed in linear flux space (a CCD camera is an intrinsically linear flux-measuring device), we often must include in our data sets measurements that have less accurately been quoted with symmetric error bars in log flux (or magnitude) space. There are also situations where published data are quoted with asymmetric 1σ error bars, but with no obvious information as to the form of the underlying probability distribution.

Regardless, these data must be included in our analysis somehow. In the absence of other information, we must transform these log-normal distributions into linear space as best we can. One common practice is to compute a "symmetrized" distribution. If the quoted measurement and its uncertainty are $\log F_n(+\sigma_{\log F_n+}, -\sigma_{\log F_n-})$, then the transformed value may be approximated as:

$$F_n = 0.5\left(10^{\log F_n + \sigma_{\log F_n+}} + 10^{\log F_n - \sigma_{\log F_n-}}\right)\,, \tag{A6}$$

and its uncertainty as:

$$\sigma_{F_n} = 0.5\left(10^{\log F_n + \sigma_{\log F_n}} - 10^{\log F_n - \sigma_{\log F_n-}}\right)\,. \tag{A7}$$

However, this symmetric approximation is only reasonably accurate for high signal-to-noise data.

Ideally, we should always describe the probability distribution of a transformed normal distribution by a properly transformed pdf. Unfortunately, most such distributions result in joint probability integrals that cannot be evaluated analytically – and numerical integration may be prohibitively computationally expensive when repeatedly computing likelihoods in model fitting algorithms like Galapagos. Moreover, when confronted with published measurements with asymmetric error bars, but no information as to the true underlying probability distribution, by necessity we must assume some distribution in order to proceed. To this end, we have chosen to adopt and explore the properties of a particular "asymmetric Gaussian", or skew-normal, distribution, which has proven to be amenable to analysis. For instance, if a measurement is quoted with symmetric log-normal error bars, $\log y_n \pm \sigma_{\log y,n}$, and if $\sigma_{\log y,n} \lesssim 0.1$, we can approximate the distribution in linear y-space as a skew-normal, or asymmetric Gaussian, as described in §2.7. We define the normalized asymmetric Gaussian function as:

$$G_A(y;y_n,\sigma_{y,n\pm}) \equiv \frac{2}{\sqrt{2\pi}(\sigma_{y,n+} + \sigma_{y,n-})} \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{y-y_n}{\sigma_{y,n+}}\right)^2\right] & \text{if } y \geq y_n \\ \exp\left[-\frac{1}{2}\left(\frac{y-y_n}{\sigma_{y,n-}}\right)^2\right] & \text{if } y < y_n \end{cases}\,. \tag{A8}$$

This can also be expressed as:

$$G_A(y;y_n,\sigma_{y,n\pm}) \equiv \frac{2}{\sqrt{2\pi}(\sigma_{y,n+} + \sigma_{y,n-})} \exp\left[-\frac{1}{2}\left(\frac{y-y_n}{\sigma_{y,n\pm}}\right)^2\right] \times \Theta_\pm(y;y_n)\,, \tag{A9}$$

where $\Theta_\pm(y;y_n)$ is the Heaviside step function:

$$\Theta_+(y;y_n) \equiv \begin{cases} 1 & \text{if } y \geq y_n \\ 0 & \text{otherwise}\,, \end{cases}$$
$$\Theta_-(y;y_n) \equiv \begin{cases} 1 & \text{if } y < y_n \\ 0 & \text{otherwise}\,. \end{cases} \tag{A10}$$

If both the probability distributions in Equation A1 are approximated as asymmetric Gaussians (the symmetric Gaussian distribution being just a special case of G_A where $\sigma_{y+} = \sigma_{y-}$), the joint probability integral can still be expressed in terms of the convolution of two distributions, as in Equation A5:

$$p_n = \int_{-\infty}^{\infty} \left[G_A(y; 0; \sigma_{y\mp}) * G_A(y; y_n; \sigma_{y,n\pm}) \right] \delta(y - y_c(x; \vartheta_m)) dy \,, \tag{A11}$$

Note the reversal of signs in the first term of the convolution. Unfortunately, the convolution of two asymmetric

Gaussian functions does not, in general, yield an asymmetric Gaussian. Expanding the convolution integral:

$$\begin{aligned} &G_A(y;0;\sigma_{y\mp}) * G_A(y;y_n;\sigma_{y,n\pm}) = \\ &A\int_{-\infty}^{\infty}\left[e^{-\frac{1}{2}\left(\frac{y-y'}{\sigma_{y-}}\right)^2}\Theta_-(y';y) + e^{-\frac{1}{2}\left(\frac{y-y'}{\sigma_{y+}}\right)^2}\Theta_+(y';y)\right] \\ &\times\left[e^{-\frac{1}{2}\left(\frac{y'-y_n}{\sigma_{y,n-}}\right)^2}\Theta_-(y';y_n) + e^{-\frac{1}{2}\left(\frac{y'-y_n}{\sigma_{y,n+}}\right)^2}\Theta_+(y';y_n)\right]dy' , \end{aligned} \tag{A12}$$

where:

$$A \equiv \frac{2}{\pi(\sigma_{y+}+\sigma_{y-})(\sigma_{y,n+}+\sigma_{y,n-})} . \tag{A13}$$

This integral has four terms, which we denote by $I_{--}, I_{-+}, I_{+-}$and I_{++}. For example:

$$I_{--} = A\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\left(\frac{y-y'}{\sigma_{y-}}\right)^2+\left(\frac{y'-y_n}{\sigma_{y,n-}}\right)^2\right]}\Theta_-(y';y)\Theta_-(y';y_n)dy' . \tag{A14}$$

Expanding the argument of the exponential, and completing the square, we find this term equals:

$$I_{--} = A\frac{\sigma_{y-}\sigma_{y,n-}}{\Sigma_{--}}e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{--}}\right)^2} \times \begin{cases} \Phi\left[\frac{\sigma_{y-}}{\sigma_{y,n-}}\frac{(y-y_n)}{\Sigma_{--}}\right] & \text{if } y < y_n, \\ \Phi\left[\frac{\sigma_{y,n-}}{\sigma_{y-}}\frac{(y_n-y)}{\Sigma_{--}}\right] & \text{if } y \geq y_n, \end{cases} \tag{A15}$$

where:

$$\Sigma_{\pm\pm} \equiv \sqrt{\sigma_{y\pm}+\sigma_{y,n\pm}} , \tag{A16}$$

is the quadrature sum of the two relevant σ's, and:

$$\Phi(z) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}x^2}dx = \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right] , \tag{A17}$$

is the cumulative Gaussian distribution function. The remaining terms have a similar form:

$$I_{-+} = A\frac{\sigma_{y-}\sigma_{y,n+}}{\Sigma_{-+}}e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{-+}}\right)^2} \times \begin{cases} 0 & \text{if } y < y_n, \\ \Phi\left[\frac{\sigma_{y-}}{\sigma_{y,n+}}\frac{(y-y_n)}{\Sigma_{-+}}\right] - \Phi\left[\frac{\sigma_{y,n+}}{\sigma_{y-}}\frac{(y_n-y)}{\Sigma_{-+}}\right] & \text{if } y \geq y_n, \end{cases} \tag{A18}$$

$$I_{+-} = A\frac{\sigma_{y+}\sigma_{y,n-}}{\Sigma_{+-}}e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{+-}}\right)^2} \times \begin{cases} \Phi\left[\frac{\sigma_{y,n-}}{\sigma_{y+}}\frac{(y_n-y)}{\Sigma_{+-}}\right] - \Phi\left[\frac{\sigma_{y+}}{\sigma_{y,n-}}\frac{(y-y_n)}{\Sigma_{+-}}\right] & \text{if } y < y_n, \\ 0 & \text{if } y \geq y_n, \end{cases} \tag{A19}$$

$$I_{++} = A\frac{\sigma_{y+}\sigma_{y,n+}}{\Sigma_{++}}e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{++}}\right)^2} \times \begin{cases} 1-\Phi\left[\frac{\sigma_{y,n+}}{\sigma_{y+}}\frac{(y_n-y)}{\Sigma_{++}}\right] & \text{if } y < y_n, \\ 1-\Phi\left[\frac{\sigma_{y+}}{\sigma_{y,n+}}\frac{(y-y_n)}{\Sigma_{++}}\right] & \text{if } y \geq y_n. \end{cases} \tag{A20}$$

So, the convolution of two asymmetric Gaussian functions is the sum of these four terms:

$$G_A(y;0;\sigma_{y\mp}) * G_A(y;y_n;\sigma_{y,n\pm}) = I_{--} + I_{-+} + I_{+-} + I_{++} . \tag{A21}$$

A.2. *An Algorithm for Approximating the Convolution of Two Asymmetric Gaussians*

As noted before, the result of the convolution of two asymmetric Gaussian functions is not, in general, an asymmetric Gaussian function. We have extensively explored the properties of these convolution integrals for a wide range of values for $(\sigma_{y\pm}, \sigma_{y,n\pm})$, and found that an asymmetric Gaussian function of the form $G_A(y; y_n+\delta_{y,n}, \Sigma_{y,n\pm})$ is a reasonable approximation, where:

$$\Sigma_{y,n\pm} \equiv \sqrt{\sigma_{y\mp}^2+\sigma_{y,n\pm}} , \tag{A22}$$

and where the peak of the asymmetric Gaussian is shifted by some amount $\delta_{y,n}$ from the data point's most probable value, y_n. The joint probability of the model and the data point for a 1D statistic, in this approximation, is then:

$$p_n \approx G_A(y_c(x;\vartheta_m); y_n+\delta_{y,n}; \Sigma_{y,n\pm}) . \tag{A23}$$

In other words, the likelihood has the form of a χ^2-like statistic, but with the data point shifted from its nominal value by $\delta_{y,n}$, and its plus and minus 1σ error bars expanded in quadrature by the minus and plus components of the model slop, respectively. But what is $\delta_{y,n}$?

For a given set of values for $(\sigma_{y\pm}, \sigma_{y,n\pm})$, we estimated $\delta_{y,n}$ numerically, by finding the value that maximized the cross-correlation between the explicit form of the convolution integral (Equation A21) and the asymmetric Gaussian approximation (Equation A23). We tabulated $\delta_{y,n}$ for a wide range of values of $(\sigma_{y\pm}, \sigma_{y,n\pm})$, and searched for patterns in that four-dimensional space that might allow for an analytic approximation. We found that the magnitude of the shift scales as the maximum $\sigma_{\rm max} = \max\{\sigma_{y\pm}, \sigma_{y,n\pm}\}$, but also as a complicated function of various ratios of the four σ values.

We begin by identifying the maximum and minimum values of both the intrinsic and model σ's, and the maximum of all four values:

$$\begin{aligned}
\sigma_L &\equiv \max\{\sigma_{y+}, \sigma_{y-}\}\ , \\
\sigma_S &\equiv \min\{\sigma_{y+}, \sigma_{y-}\}\ , \\
\sigma_{n,L} &\equiv \max\{\sigma_{y,n+}, \sigma_{y,n-}\}\ , \\
\sigma_{n,S} &\equiv \min\{\sigma_{y,n+}, \sigma_{y,n-}\}\ , \\
\sigma_{\rm max} &\equiv \max\{\sigma_L, \sigma_{i,L}\}\ .
\end{aligned} \tag{A24}$$

We define the quantity ξ as the sum of the ratios of the smaller and larger σ's:

$$\xi \equiv \frac{\sigma_S}{\sigma_L} + \frac{\sigma_{n,S}}{\sigma_{n,L}}\,, \tag{A25}$$

and the quantity η as the difference of the ratios:

$$\eta \equiv \begin{cases} \frac{\sigma_{n,S}}{\sigma_{n,L}} - \frac{\sigma_S}{\sigma_L} & \text{if } \sigma_{n,L} < \sigma_L \\ \frac{\sigma_S}{\sigma_L} - \frac{\sigma_{n,S}}{\sigma_{n,L}} & \text{if } \sigma_{n,L} \geq \sigma_L\,. \end{cases} \tag{A26}$$

We also define the ratio r:

$$r \equiv \frac{\min\{\sigma_L, \sigma_{n,L}\}}{\max\{\sigma_L, \sigma_{n,L}\}}\,. \tag{A27}$$

Next, we transform the quantities ξ and η:

$$\xi' = \begin{cases} \xi & \text{if } \xi \leq 1 \\ 2 - \xi & \text{if } \xi > 1 \end{cases} \tag{A28}$$

and:

$$\eta' = \begin{cases} 0 & \text{if } \xi' = 0 \\ 2\xi' \left[\frac{1}{2}\frac{\eta}{\xi'} + 1\right]^{n(r)} - \xi' & \text{otherwise}\,, \end{cases} \tag{A29}$$

where the exponent $n(r)$ is the empirical function:

$$n(r) \equiv r^{-0.4087}\,. \tag{A30}$$

Next, we compute an intermediate shift δ_*:

$$\delta_* = \sigma_{\rm max} N(r)\left[f(\xi)g(\eta') + h(\xi)\right]\,, \tag{A31}$$

where N, f, g and h are empirical functions:

$$N(r) = -0.5326r^2 + 1.5307r + 0.0019 \tag{A32}$$

$$f(\xi) = \begin{cases} 0 & \text{if } \xi = 0\,, \\ 0.2454\xi^{-1.1452} & \text{if } \xi \leq 1\,, \\ 0.2454\xi^{-0.5203} & \text{if } \xi > 1\,, \end{cases} \tag{A33}$$

$$g(\eta') = \eta'^2 \tag{A34}$$

$$h(\xi) = -0.042\xi^2 - 0.1602\xi + 0.4884\,. \tag{A35}$$

We now must consider various cases. First, consider the case where one of the two distributions is symmetric, i.e., $\sigma_{y+} = \sigma_{y-}$ or $\sigma_{y,n+} = \sigma_{y,n-}$. Then, the magnitude of the shift $|\delta_{y,n}| = \delta_*$. The direction of the shift, which we denote by $i = \{-1, 1\}$, is determined by the direction of the largest σ of the asymmetric Gaussian:

$$i = \begin{cases} +1 & \text{if } \sigma_{n,L} = \sigma_{y,n+} \text{ or } \sigma_L = \sigma_{y-} \\ -1 & \text{if } \sigma_{\max} = \sigma_{y,n-} \text{ or } \sigma_{\max} = \sigma_{y+}\,, \end{cases} \tag{A36}$$

and $\delta_{y,n} = i \times \delta_*$.

If both distributions are asymmetric, there are two cases. The first case applies when the larger σ's of the slop and the intrinsic pdf are in opposite directions, i.e., when $\sigma_L = \sigma_{y-}$ and $\sigma_{n,L} = \sigma_{y,n+}$, or $\sigma_L = \sigma_{y+}$ and $\sigma_{n,L} = \sigma_{y,n-}$. In this case, as in the symmetric case above, the magnitude of the shift $|\delta_{y,n}| = \delta_*$, and the direction of the shift is:

$$i = \begin{cases} +1 & \text{if } \sigma_{\max} = \sigma_{y,n+} \text{ or } \sigma_{\max} = \sigma_{y-} \\ -1 & \text{if } \sigma_{\max} = \sigma_{y,n-} \text{ or } \sigma_{\max} = \sigma_{y+}\,, \end{cases} \tag{A37}$$

and $\delta_{y,n} = i \times \delta_*$.

The second case applies when the larger σ's of the slop and the intrinsic pdf are in the same direction, i.e., when $\sigma_L = \sigma_{y+}$ and $\sigma_{n,L} = \sigma_{y,n+}$, or $\sigma_L = \sigma_{y-}$ and $\sigma_{n,L} = \sigma_{y,n-}$. Here, the shift $\delta_{y,n}$ is obtained by multiplying $i \times \delta_*$ by an additional empirical function:

$$\delta_{y,n} = i \times \delta_* \times \sin\left(\frac{\pi}{2}\frac{\eta'}{\xi'}\right) \times \begin{cases} \xi^{0.7413} & \text{if } \xi \leq 1 \\ \xi^{-0.1268} & \text{if } \xi > 1\,, \end{cases} \tag{A38}$$

where i is given, as before, by Equation A37. Note that, since the sinusoidal term in Equation A38 may take negative values, the actual direction of the shift may be opposite in sign to i.[10]

B. EVALUATING THE JOINT PROBABILITY OF TWO ASYMMETRIC GAUSSIAN DISTRIBUTIONS IN TWO DIMENSIONS

We proceed, as in §2.1, by approximating the model curve $y_c(x; \vartheta_m)$ by a line tangent to the asymmetric convolved (and shifted) error "ellipse", taking care to keep track of which quadrant the tangent point lies in. However, since the tangent line will, in general, cross through three quadrants of the asymmetric convolved ellipse, the joint probability integral over the range $[-\infty, \infty]$ of Equation 13 must be broken into three segments.

Consider, for example, the case shown in Figure 21: a model distribution about the curve $y_c(x; \vartheta_m)$ with asymmetric Gaussian slops $(\sigma_{x\pm}, \sigma_{y\pm})$, and a data point at (x_n, y_n) with asymmetric Gaussian intrinsic uncertainties $(\sigma_{x,n\pm}, \sigma_{y,n\pm})$. As discussed in §2.7.2 and Appendix A, evaluation of the joint probability integral (Equation 5) requires a convolution of these two 2D asymmetric Gaussian distributions. The resulting convolved distribution can be approximated as a 2D asymmetric Gaussian, with 1σ widths $(\Sigma_{x,n\pm}, \Sigma_{y,n\pm})$, and a centroid at $(x_n + \delta_{x,n}, y_n + \delta_{y,n})$, where $\Sigma^2_{x,n\pm} = \sigma^2_{x\mp} + \sigma^2_{x,n\pm}$, $\Sigma^2_{y,n\pm} = \sigma^2_{y\mp} + \sigma^2_{y,n\pm}$, and where the centroid shifts $(\delta_{x,n}, \delta_{y,n})$ are given by the algorithm in Appendix A. In the case shown in Figure 21, the tangent point $(x_{t,n}, y_{t,n})$ lies in quadrant 1 of the asymmetric convolved error ellipse, where the curve has a slope $m_{t,n}$, and so, for the TRK statistic, the joint probability will be proportional to the factor (see §2.3):

$$\left(\frac{du_n}{dx}\right)_{++} = \frac{m^2_{t,n}\Sigma^2_{x,n+} + \Sigma^2_{y,n+}}{\sqrt{m^2_{t,n}\Sigma^4_{x,n+} + \Sigma^4_{y,n+}}}\,. \tag{B39}$$

In general, the factor $\left(\frac{du_n}{dx}\right)_{\pm\mp}$ is determined by which quadrant the tangent point lies in:

$$\left(\frac{du_n}{dx}\right)_{\pm\mp} = \frac{m^2_{t,n}\Sigma^2_{x,n\pm} + \Sigma^2_{y,n\mp}}{\sqrt{m^2_{t,n}\Sigma^4_{x,n\pm} + \Sigma^4_{y,n\mp}}}\,, \tag{B40}$$

where $\left(\frac{du_n}{dx}\right)_{++}$ corresponds to quadrant 1, $\left(\frac{du_n}{dx}\right)_{-+}$ to quadrant 2, $\left(\frac{du_n}{dx}\right)_{--}$ to quadrant 3, and $\left(\frac{du_n}{dx}\right)_{+-}$ to quadrant 4, and where the first subscript indicates the choice of $\Sigma_{x,n\pm}$, and the second indicates the choice of $\Sigma_{y,n\mp}$ in Equation B40.

[10] There is one specific case which requires special treatment. If both distributions are asymmetric, but if, by chance, $\sigma_{y,n-} = \sigma_{y-}$ exactly, and $\sigma_{y+} < \sigma_{y,n+} < \sigma_{y,n-}$, then the sign of i from Equation A37 must be reversed.

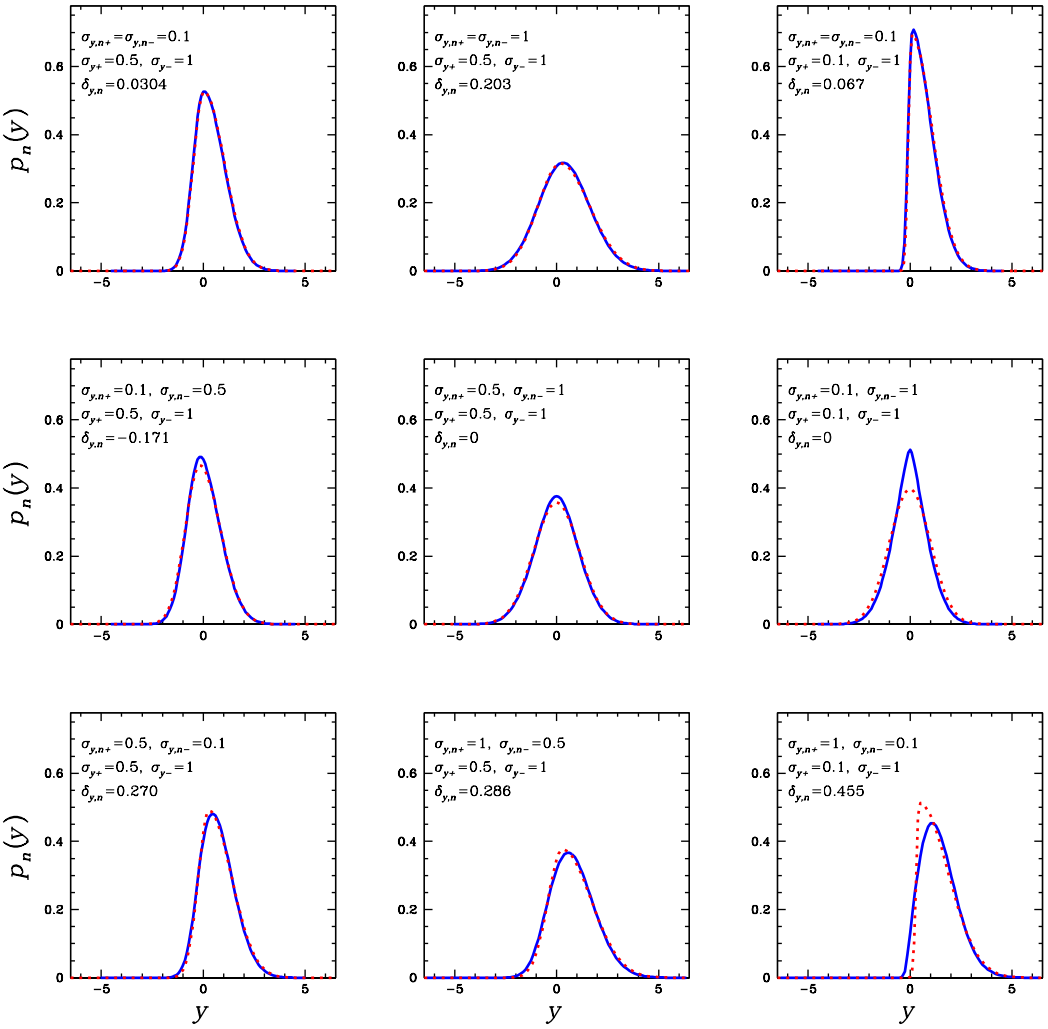


Figure 20. Asymmetric Gaussian approximations of the convolution of two asymmetric Gaussians, $G_A(y;0,\sigma_{y\mp}) * G_A(y;0,\sigma_{y,n\pm})$. Solid blue curves are the actual convolved distributions. Dotted red curves are the asymmetric Gaussian approximations $G_A(y;\delta_{y,n},\Sigma_{y,n\pm})$, where $\Sigma^2_{y,n\pm} \equiv \sigma^2_{y\mp} + \sigma^2_{y,n\pm}$.

The joint probability integral of Equation 13 must be broken into three segments, with limits $[-\infty, x_1]$, $[x_1, x_2]$ and $[x_2, \infty]$, where $x_1 = x_n + \delta_{x,n}$, and where x_2 is the point where the tangent line $y_{t,n} + m_{t,n}(x - x_{t,n}) = y_n + \delta_{y,n}$. The first segment of the integral, through quadrant 2, is:

$$I_{-+} = \left(\frac{du_n}{dx}\right)_{++} \frac{2\Sigma_{x,n-}\Sigma_{y,n+}}{\pi(\Sigma_{x,n+}+\Sigma_{x,n-})(\Sigma_{y,n+}+\Sigma_{y,n-})} \frac{\Phi(z_{-+}(x_1))}{\sqrt{m^2_{t,n}\Sigma^2_{x,n-}+\Sigma^2_{y,n+}}}$$
$$\times \exp\left\{-\frac{1}{2}\left[\frac{y_{t,n}-y_n-\delta_{y,n}-m_{t,n}(x_{t,n}-x_n-\delta_{x,n})}{\sqrt{m^2_{t,n}\Sigma^2_{x,n-}+\Sigma^2_{y,n+}}}\right]^2\right\}, \tag{B41}$$

where $\Phi(z)$ is the cumulative Gaussian function:

$$\Phi(z) \equiv \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx\,, \tag{B42}$$

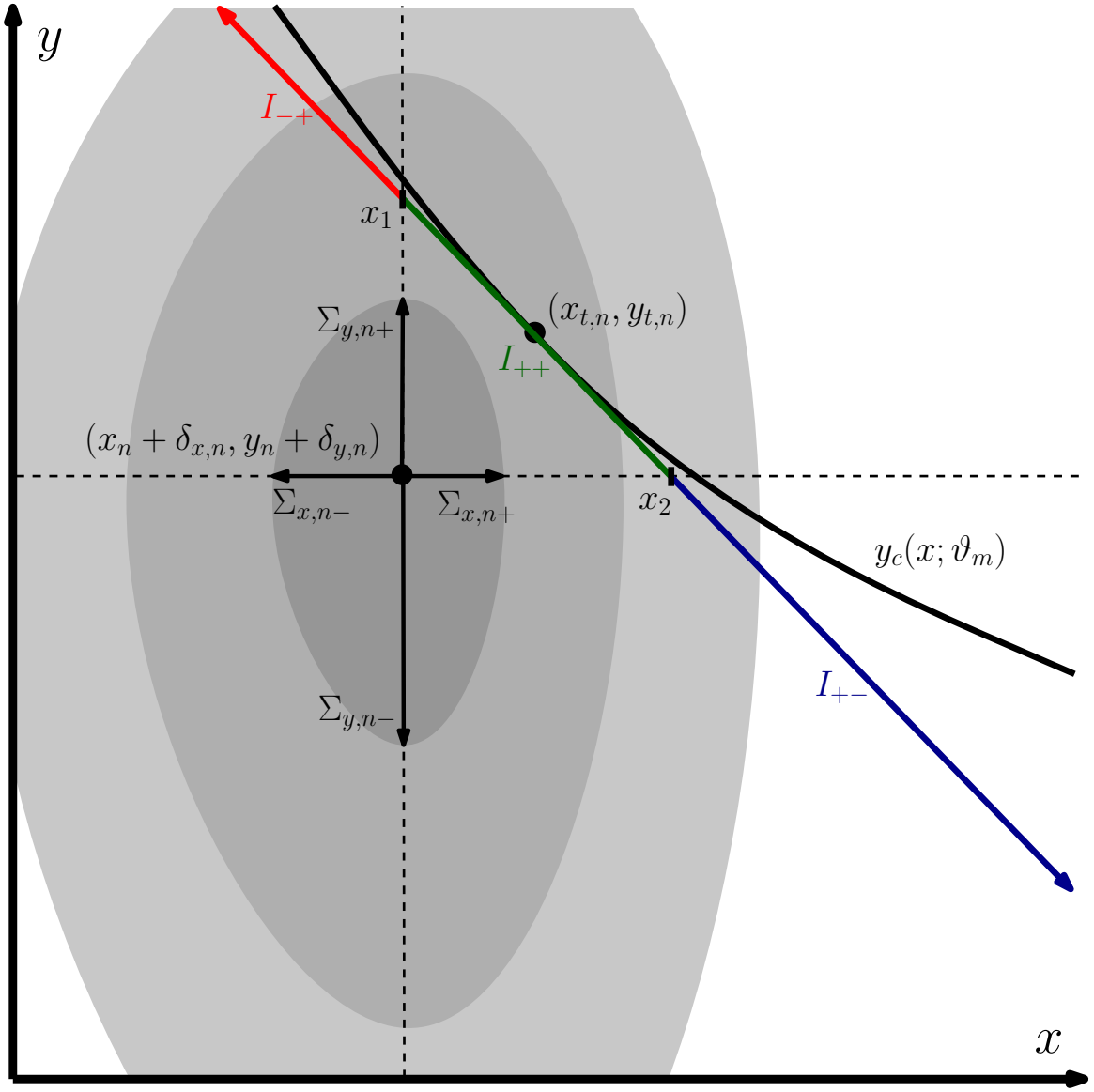


Figure 21. Geometry of the 2D joint probability integral for asymmetric probability distributions. The convolution of an asymmetric model distribution and the asymmetric intrinsic uncertainty of a data point is approximated by a 2D asymmetric Gaussian, with 1σ widths $(\Sigma_{x,n\pm}, \Sigma_{y,n\pm})$, and with a centroid at $(x_n+\delta_{x,n}, y_n+\delta_{y,n})$. The model curve $y_c(x;\vartheta_m)$ is approximated by a line tangent to the convolved ellipse at point $(x_{t,n}, y_{t,n})$. The joint probability integral must be broken into three segments. The integral I_{-+} (red) through quadrant 2 has limits $[-\infty, x_1]$. The integral I_{++} (green) through quadrant 1 has limits $[x_1, x_2]$. The integral I_{+-} (blue) through quadrant 4 has limits $[x_2, \infty]$.

and where the limit of integration x_1 is transformed to:

$$z_{-+}(x_1) = \frac{\Sigma_{y,n+}^2(x_1 - x_n - \delta_{x,n}) + m_{t,n}^2\Sigma_{x,n-}^2\left[x_1 - x_{t,n} - (y_n + \delta_{y,n} - y_{t,n})/m_{t,n}\right]}{\Sigma_{x,n-}\Sigma_{y,n+}\sqrt{m_{t,n}^2\Sigma_{x,n-}^2 + \Sigma_{y,n+}^2}}\,. \tag{B43}$$

This transformation of the limit from x to $z_{\pm\mp}(x)$ is generally applicable to integrations through all quadrants. In general:

$$z_{\pm\mp}(x) = \frac{\Sigma_{y,n\mp}^2(x - x_n - \delta_{x,n}) + m_{t,n}^2\Sigma_{x,n\pm}^2\left[x - x_{t,n} - (y_n + \delta_{y,n} - y_{t,n})/m_{t,n}\right]}{\Sigma_{x,n\pm}\Sigma_{y,n\mp}\sqrt{m_{t,n}^2\Sigma_{x,n\pm}^2 + \Sigma_{y,n\mp}^2}}\,, \tag{B44}$$

where $z_{++}(x)$ corresponds to integration through quadrant 1, z_{-+} to quadrant 2, z_{--} to quadrant 3, and z_{+-} to quadrant 4, and where the first subscript of $z_{\pm\mp}(x)$ indicates the choice of $\Sigma_{x,n\pm}$, and the second indicates the choice of $\Sigma_{y,n\mp}$ in Equation B44.

The second segment of the integral, through quadrant 1 between the limits $[x_1, x_2]$, is:

$$I_{++} = \left(\frac{du_n}{dx}\right)_{++} \frac{2\Sigma_{x,n+}\Sigma_{y,n+}}{\pi(\Sigma_{x,n+}+\Sigma_{x,n-})(\Sigma_{y,n+}+\Sigma_{y,n-})} \frac{[\Phi(z_{++}(x_2)) - \Phi(z_{++}(x_1))]}{\sqrt{m_{t,n}^2\Sigma_{x,n+}^2 + \Sigma_{y,n+}^2}} \times \exp\left\{-\frac{1}{2}\left[\frac{y_{t,n} - y_n - \delta_{y,n} - m_{t,n}(x_{t,n} - x_n - \delta_{x,n})}{\sqrt{m_{t,n}^2\Sigma_{x,n+}^2 + \Sigma_{y,n+}^2}}\right]^2\right\} . \tag{B45}$$

Finally, the third segment of the integral, through quadrant 4, is:

$$I_{+-} = \left(\frac{du_n}{dx}\right)_{++} \frac{2\Sigma_{x,n+}\Sigma_{y,n-}}{\pi(\Sigma_{x,n+}+\Sigma_{x,n-})(\Sigma_{y,n+}+\Sigma_{y,n-})} \frac{[1 - \Phi(z_{+-}(x_2))]}{\sqrt{m_{t,n}^2\Sigma_{x,n+}^2 + \Sigma_{y,n-}^2}} \times \exp\left\{-\frac{1}{2}\left[\frac{y_{t,n} - y_n - \delta_{y,n} - m_{t,n}(x_{t,n} - x_n - \delta_{x,n})}{\sqrt{m_{t,n}^2\Sigma_{x,n+}^2 + \Sigma_{y,n-}^2}}\right]^2\right\} . \tag{B46}$$

The total joint probability integral is then the sum of these three terms, $p_n = I_{-+} + I_{++} + I_{+-}$.

This procedure is easily generalized to cases where the tangent point falls in any of the other quadrants. In general, the integral will always consist of three segments through three different quadrants, one with limits $[-\infty, x_1]$, one with limits $[x_1, x_2]$, and one with limits $[x_2, \infty]$. The factor $\left(\frac{du_n}{dx}\right)_{\pm\mp}$ (Equation B40) is computed according to the quadrant the tangent point lies in, and is the same for all three integrals. Similarly, the normalizing factor $2/(\pi(\Sigma_{x,n+}+\Sigma_{x,n-})(\Sigma_{y,n+}+\Sigma_{y,n-}))$ will be the same for all three integrals. The first integral will take the form of Equation B41, but with a different choice of $(\Sigma_{x,n\pm}, \Sigma_{y,n\mp})$ in that expression and in the expression for the transformed limit $z_{\pm\mp}(x_1)$ (Equation B44). Likewise, the second integral will take the form of Equation B45, and the third the form of Equation B46, each expressed in terms of the values of $(\Sigma_{x,n\pm}, \Sigma_{y,n\mp})$ appropriate to the quadrant that that segment of the integral lies in. For example, if the tangent point lies in quadrant 4, all three integrals will be proportional to the factor $\left(\frac{du_n}{dx}\right)_{+-}$. The first integral, from $[-\infty, x_1]$, will be through quadrant 3, and have the form of Equation B41, but with $(\Sigma_{x,n-}, \Sigma_{y,n-})$. The second integral, from $[x_1, x_2]$, will be through quadrant 4, and have the form of Equation B45, but with $(\Sigma_{x,n+}, \Sigma_{y,n-})$. And the third integral, from $[x_2, \infty]$, will through quadrant 1, and have the form of Equation B46, but with $(\Sigma_{x,n+}, \Sigma_{y,n+})$.

C. COMPUTING 2D MODEL DISTRIBUTION ENVELOPES FOR PLOTTING

In general, for a best-fit curve $y_c(x;\vartheta_m)$, described by M parameters $\{\vartheta_m\}$, and with best-fit slop parameters (σ_x, σ_y), the $n\sigma$ slop envelope is defined as the locus of points that is tangent to the set of ellipses of widths $(n\sigma_x, n\sigma_y)$, centered on each point along the curve. Consider a point on the curve, $(x, y_c(x;\vartheta_m))$, at which the curve has a slope $m_c(x;\vartheta_m) = \frac{d}{dx}y_c(x;\vartheta_m)$. An $n\sigma$ ellipse centered on that point is described by the equation:

$$\frac{\Delta x_{n\sigma}^2}{\sigma_x^2} + \frac{\Delta y_{n\sigma}^2}{\sigma_y^2} = n^2 , \tag{C47}$$

where $(\Delta x_{n\sigma}, \Delta y_{n\sigma})$ are the distances from the point $(x, y_c(x;\vartheta_m))$. Differentiating Equation C47 with respect to x, and solving for $\frac{dy}{dx}$ gives the slope of the ellipse at $(\Delta x_{n\sigma}, \Delta y_{n\sigma})$:

$$m = -\frac{\sigma_y^2}{\sigma_x^2}\frac{\Delta x_{n\sigma}}{\Delta y_{n\sigma}} . \tag{C48}$$

We are interested in the point on the ellipse $(\Delta x_{n\sigma_c}, \Delta y_{n\sigma_c})$ at which the slope is equal to that of the curve $m_c(x;\vartheta_m)$. From Equation C48, this is equivalent to the point for which:

$$\Delta x_{n\sigma_c} = -m_c(x,\vartheta_m)\frac{\sigma_x^2}{\sigma_y^2}\Delta y_{n\sigma_c} . \tag{C49}$$

Substituting Equation C49 into Equation C47, and solving for $\Delta y_{n\sigma_c}$ yields:

$$\Delta y_{n\sigma_c} = \pm \frac{n\sigma_y}{\sqrt{1 + \frac{m_c^2(x;\vartheta_m)\sigma_x^2}{\sigma_y^2}}} \,. \tag{C50}$$

Together, Equations C50 & C49 describe the offset of the $n\sigma$ slop envelope from each point on the model curve. The model distribution envelopes shown in Figures **??**, **??** & **??** were computed using this technique.

If the fitted uncertainties in the model curve parameters themselves are sufficiently great, one may wish to plot combined slop and model parameter uncertainty envelopes. There is no ideal technique for plotting such envelopes; one is simply to approximate them by projecting the uncertainty of each parameter to first order into the y-direction, and adding those projected uncertainties in quadrature to the slop envelopes given by Equations C49 & C50:

$$\begin{aligned} \Delta y_{n\sigma_c;\, n\sigma_{\vartheta_m}} &= \left\{ \Delta y_{n\sigma_c}^2 + \sum_{k=1}^{M} \left[n\sigma_{\vartheta_k} \times \frac{\partial}{\partial \vartheta_k} y_c(x;\vartheta_m) \right]^2 \right\}^{1/2} , \\ \Delta x_{n\sigma_c;\, n\sigma_{\vartheta_m}} &= \Delta x_{n\sigma_c} \,. \end{aligned} \tag{C51}$$

Note that Equations C49, C50 & C51 assume that both the slop and the uncertainties in the model parameters are described by symmetric Gaussian functions. For asymmetric and/or non-Gaussian distributions, some care must be taken to ensure that each offset is of the proper magnitude and in the proper direction. Finally, we note that it is inappropriate to plot combined slop and model parameter uncertainty envelopes together with data. *Only best-fit slop envelopes should be plotted with data.* However, the combined envelopes can be plotted separately, and are useful for indicating the total uncertainty as a function of x. However, in the case of the fits presented in Chapter **??**, we do not do so, since the two envelopes are effectively indistinguishable.

REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Bickel, D. R. & Frühwirth, R. 2005, On a Fast, Robust Estimator of the Mode: Comparisons to Other Robust Estimators with Applications

Corrales, L. 2015, ApJ, 805, 23

Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137

Haario, H., Saksman, E. & Tamminen, J. 2001, An Adaptive Metropolis Algorithm.

Hanisch, R. J., & Biemesderfer, C. D. 1989, BAAS, 21, 780

Hastings, W. K. 1970, Monte Carlo sampling methods using Markov chains and their applications.

Lamport, L. 1994, LaTeX: A Document Preparation System, 2nd Edition (Boston, Addison-Wesley Professional)

Nelder, J. A. & Mead, R. 1965, A simplex method for function minimization.

Schwarz, G. J., Ness, J.-U., Osborne, J. P., et al. 2011, ApJS, 197, 31

Tiruneh, A. T., Ndlela, W. N. & Nkambule, S. J. 2013, A Two-Point Newton Method Suitable for Nonconvergent Cases and with Super-Quadratic Convergence

Vogt, F. P. A., Dopita, M. A., Kewley, L. J., et al. 2014, ApJ, 793, 127