Practical MCMC

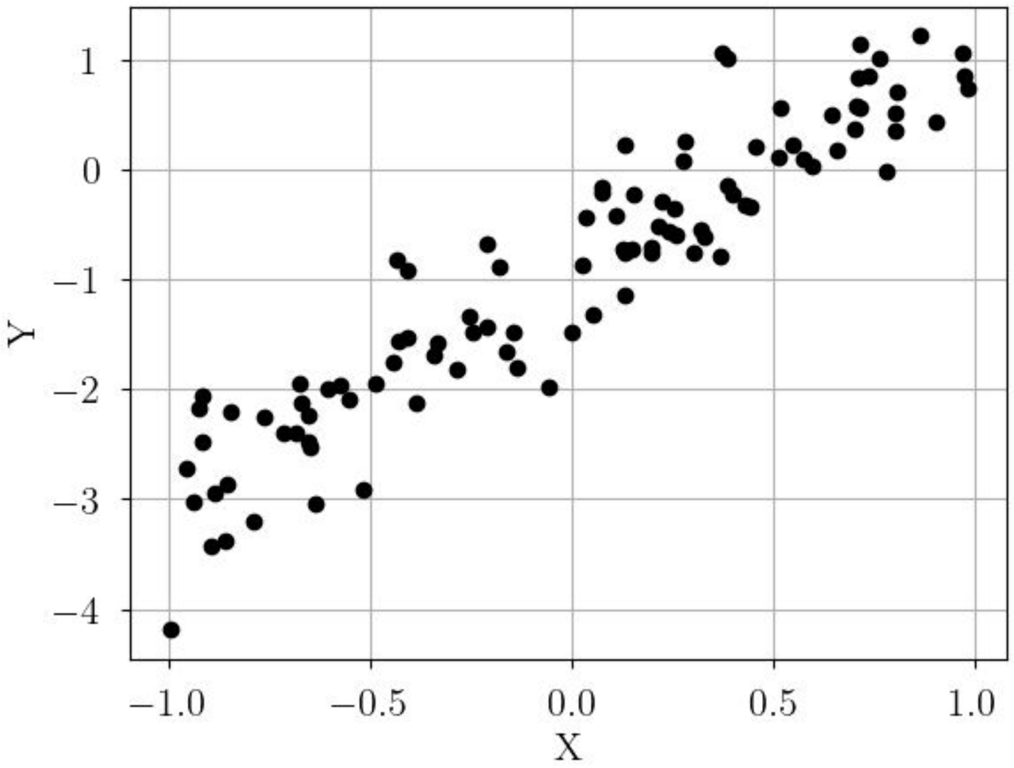
Trey V. Wenger

Math Methods & Computation Coffee

9 October 2023

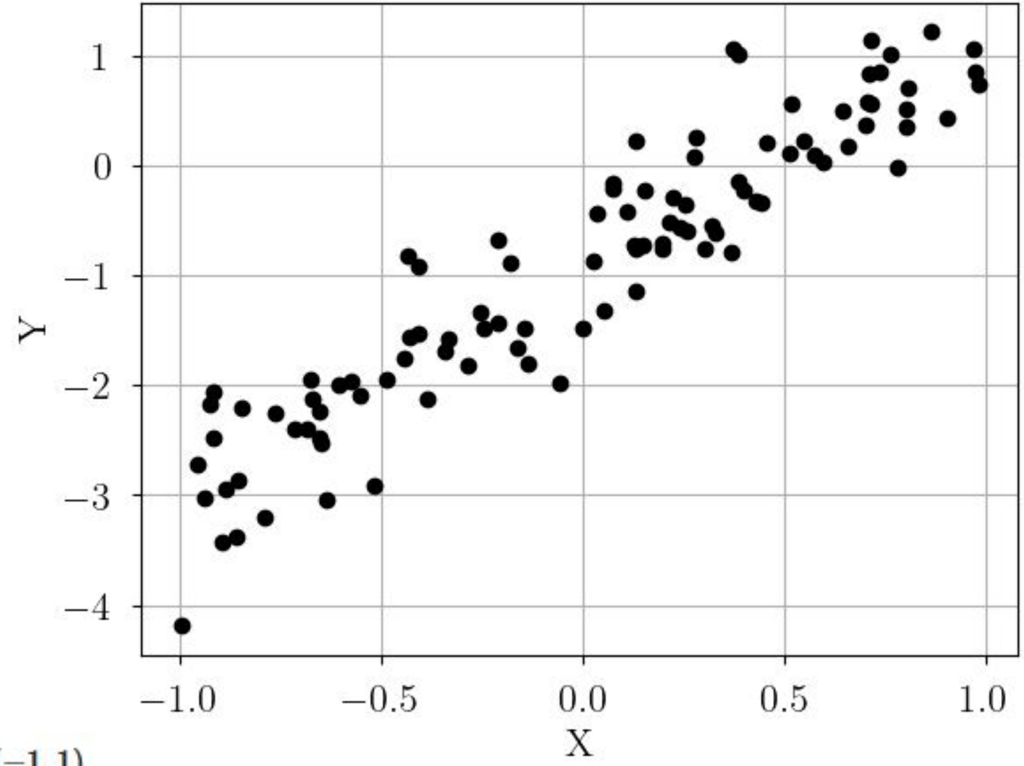
Example 1: Fitting a Line

- Observables:
  - X (independent variable)
  - Y (dependent variable)
  - *S (error)*
- Want to infer:
  - M (slope)
  - B (intercept)
  - *S (intrinsic scatter)*
- What do you do?
- What assumptions do you make?



Example 1: Fitting a Line

- Observables:
  - X (independent variable)
  - Y (dependent variable)
  - *S (error)*
- Want to infer:
  - M (slope)
  - B (intercept)
  - *S (intrinsic scatter)*
- What do you do?
- What assumptions do you make?
  - Linear process
  - Uncorrelated
  - Gaussian scatter



$$X \sim \mathrm{Uniform}(-1, 1)$$
$$Y \sim MX + B + \mathrm{Normal}(0, S)$$

Ordinary Least Squares is Maximum Likelihood

Likelihood = Likelihood that data (Y_obs) observed at (X_obs) come from linear model (Y_model) with parameters (M, B, S)

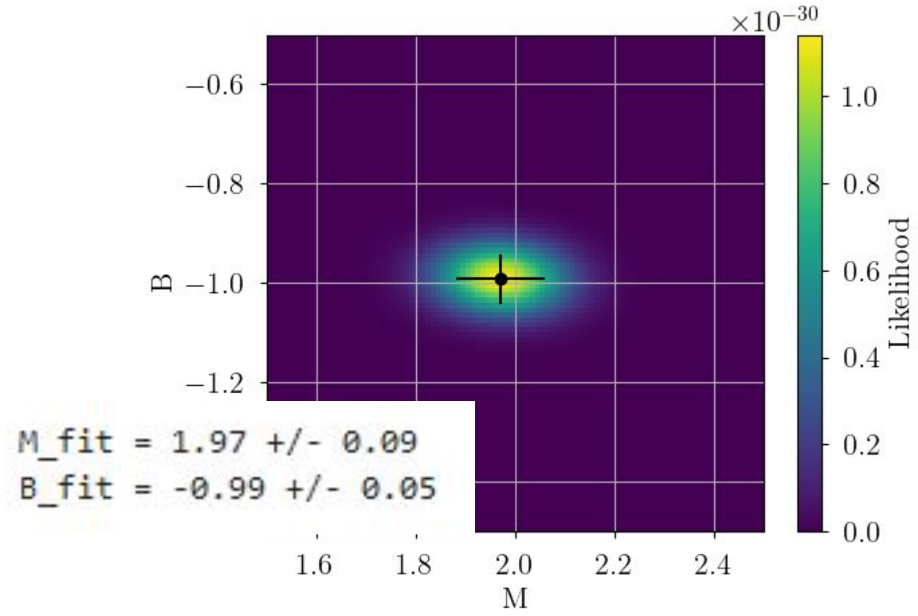
We're assuming a Gaussian scatter, so the likelihood that the data are drawn from the model:

$$P(Y_{\mathrm{obs}}|X_{\mathrm{obs}}, M, B, S) = \prod \frac{1}{\sqrt{2\pi S^2}} e^{-\frac{(Y_{\mathrm{model}}(X_{\mathrm{obs}}, M, B) - Y_{\mathrm{obs}})^2}{2S^2}}$$

$$\log P(Y_{\mathrm{obs}}|X_{\mathrm{obs}}, M, B, S) \propto \sum \frac{(Y_{\mathrm{model}}(X_{\mathrm{obs}}, M, B) - Y_{\mathrm{obs}})^2}{S^2} = \sum R^2$$

Ordinary Least Squares is Maximum Likelihood

Likelihood = Likelihood that data (Y_obs) observed at (X_obs) come from linear model (Y_model) with parameters (M, B, S)



Could build a grid of parameters to search for maximum likelihood.

Could use latin hypercube sampling to sample the likelihood and estimate the shape of the likelihood distribution.

Probabilities

In frequentist framework, we assume we know nothing about the model parameters before observing the data (uniform priors).

In Bayesian framework, we assume we know *something* (priors, which might be uniform) about the model parameters, and we update our knowledge (posterior) after observing the data. We characterize our knowledge as *probabilities.*

$$
\begin{aligned}
P_A(a) &\rightarrow \text{Probability of event a over range of possible events A} \\
P_{AB}(a, b) &\rightarrow \text{Joint probability of a and b} \\
P_{AB}(a|b) &\rightarrow \text{Conditional probability of a given b} \\
P_{AB}(a) &\rightarrow \text{Marginal probability of a over all B}
\end{aligned}
$$

Joint Probability

- Consider rolling two 6-sided dice. What’s the probability that the first roll is 2 and the second roll is 3?

Joint Probability

- Consider rolling two 6-sided dice. What's the probability that the first roll is 2 and the second roll is 3?

$$a \in A = \{1, 2, 3, 4, 5, 6\}$$
$$b \in B = \{1, 2, 3, 4, 5, 6\}$$
$$P_{AB}(a, b) = P_A(a) P_B(b)$$

a independent of b

$$P_{AB}(2, 3) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Marginal Probability

- Consider rolling two 6-sided dice. What’s the probability that the second roll is 6 no matter what the first roll was?

Marginal Probability

- Consider rolling two 6-sided dice. What's the probability that the second roll is 6 no matter what the first roll was?

$$P_{AB}(b) = \sum_{A} P_{AB}(a, b) = \int_{A} P_{AB}(a, b)\, da$$

$$P_{AB}(6) = \sum_{a=1}^{6} \frac{1}{36} = \frac{1}{6}$$

Conditional Probability

- Consider rolling two 6-sided dice. What’s the probability that the second roll is 6 given that the first roll is 2?

Conditional Probability

- Consider rolling two 6-sided dice. What's the probability that the second roll is 6 given that the first roll is 2?

$$P_{AB}(b|a) = \frac{P_{AB}(a,b)}{P_{AB}(a)}$$

Joint distribution

Marginal distribution

$$P_{AB}(6|2) = \frac{P_{AB}(6,2)}{P_{AB}(2)}$$

$$P_{AB}(6|2) = \frac{1/36}{1/6} = \frac{1}{6}$$

Bayes' Theorem

It's just conditional probabilities…

$$P_{AB}(b|a) = \frac{P_{AB}(a,b)}{P_{AB}(a)}$$

$$P_{AB}(a|b) = \frac{P_{AB}(a,b)}{P_{AB}(b)}$$

$$\boxed{P_{AB}(b|a) = \frac{P_{AB}(a|b) \times P_{AB}(b)}{P_{AB}(a)}}$$

Assuming evidence != 0

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Bayesian Inference is Likelihood x Prior

Bayes' Theorem

$$P(M,B,S|X_{\text{obs}},Y_{\text{obs}}) = \frac{P(Y_{\text{obs}}|X_{\text{obs}},M,B,S) \times P(M,B,S)}{P(Y_{\text{obs}})}$$

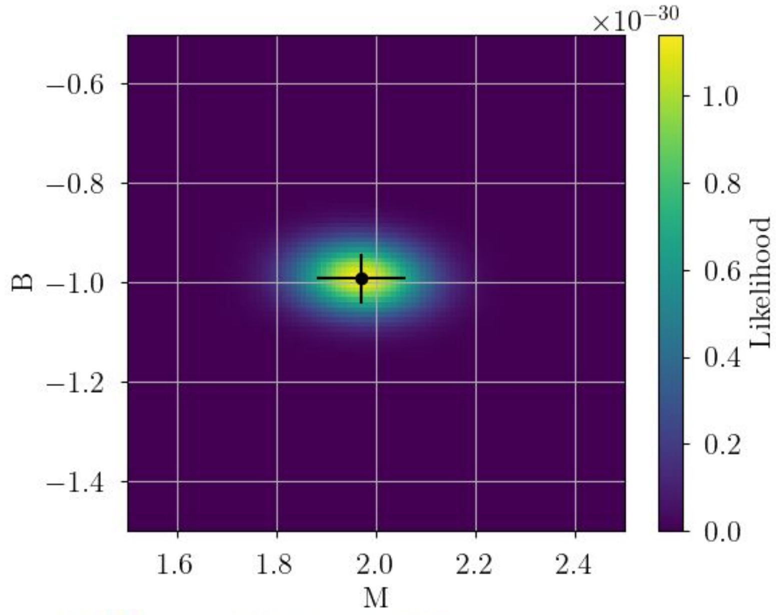
$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Maximum Likelihood == Max a posteriori for uniform prior

MCMC uses “algorithms” to sample the posterior distribution. With enough samples, you can infer the posterior distribution directly.

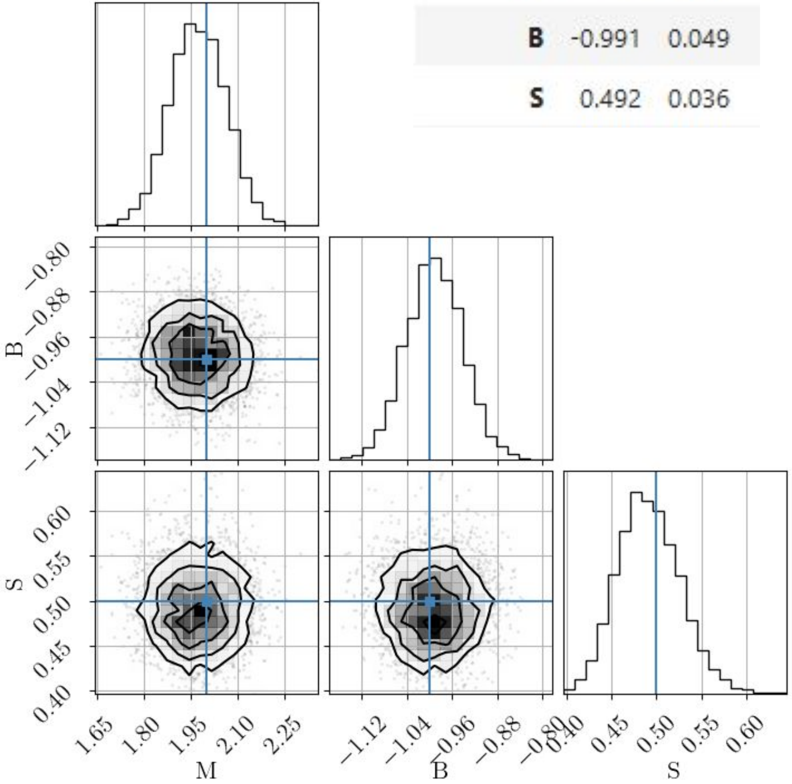
Maximum likelihood vs. Bayesian Inference

	mean	sd
M	1.970	0.089
B	-0.991	0.049
S	0.492	0.036




```
M_fit = 1.97 +/- 0.09
B_fit = -0.99 +/- 0.05
```

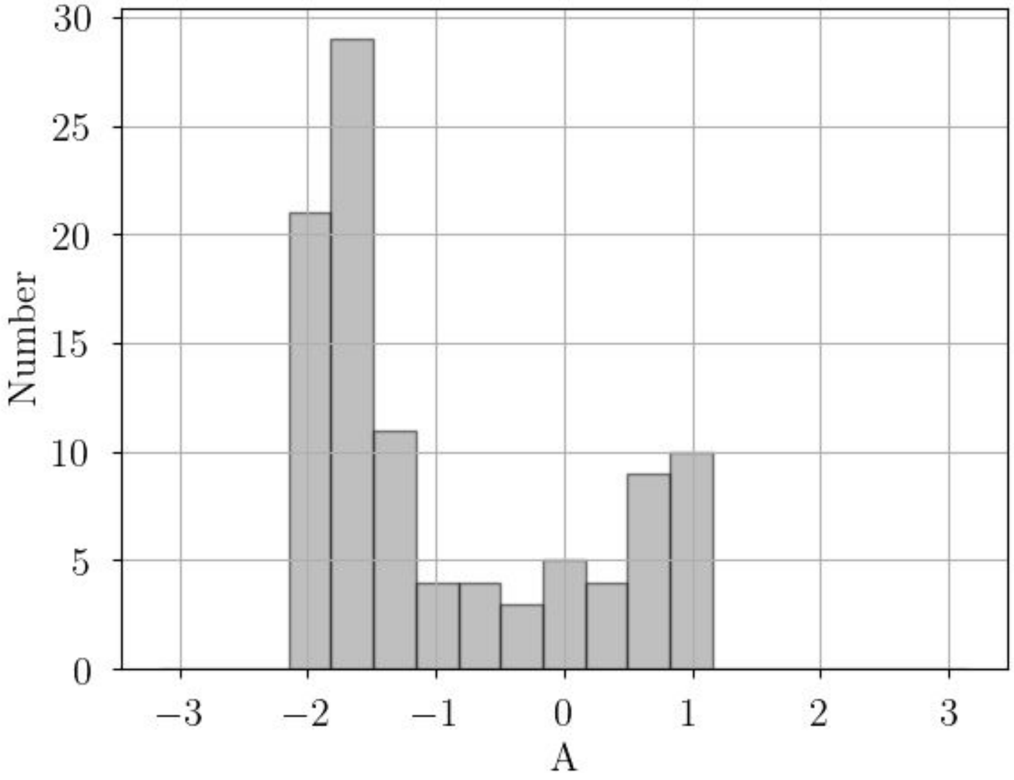
Is the likelihood distribution Gaussian?



The posterior distribution is!

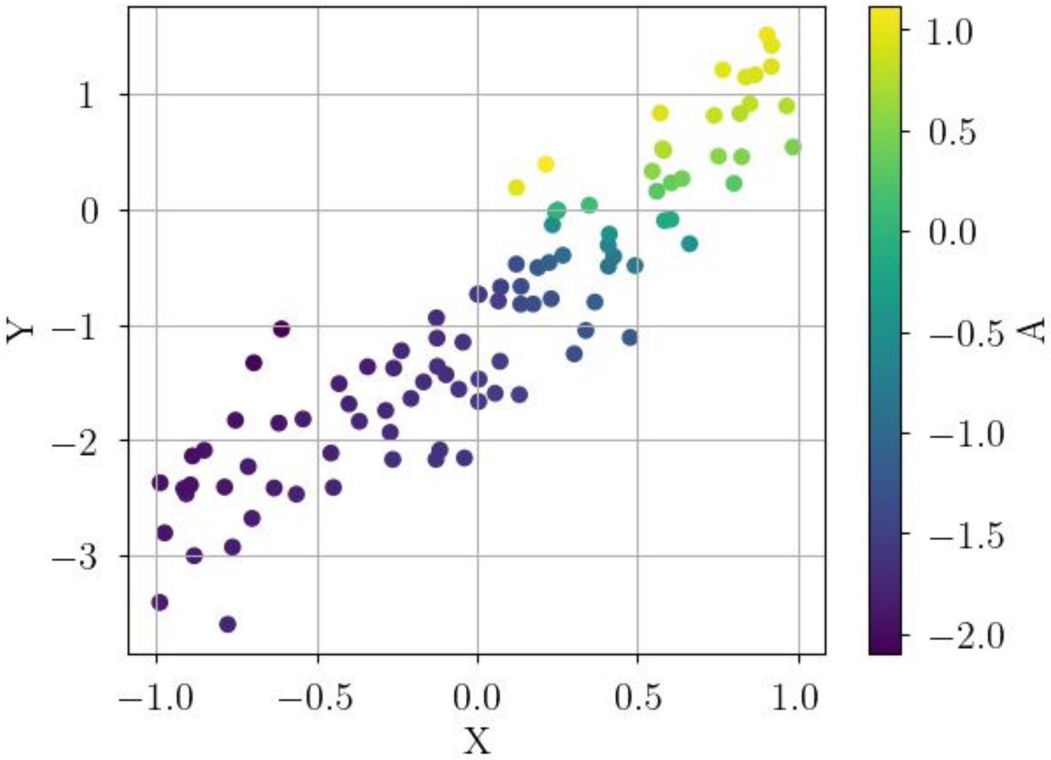
Example 2: Fitting a Line?

- Observables:
  - A; Y / X = tan A
  - E (measurement error of A)
- Know:
  - X between -1 and 1
  - Linear model
- Educated guess:
  - Slope is ???
  - Intercept is ???
- Want to infer:
  - M (slope)
  - B (intercept)
  - S (intrinsic scatter)



Example 2: Fitting a Line?

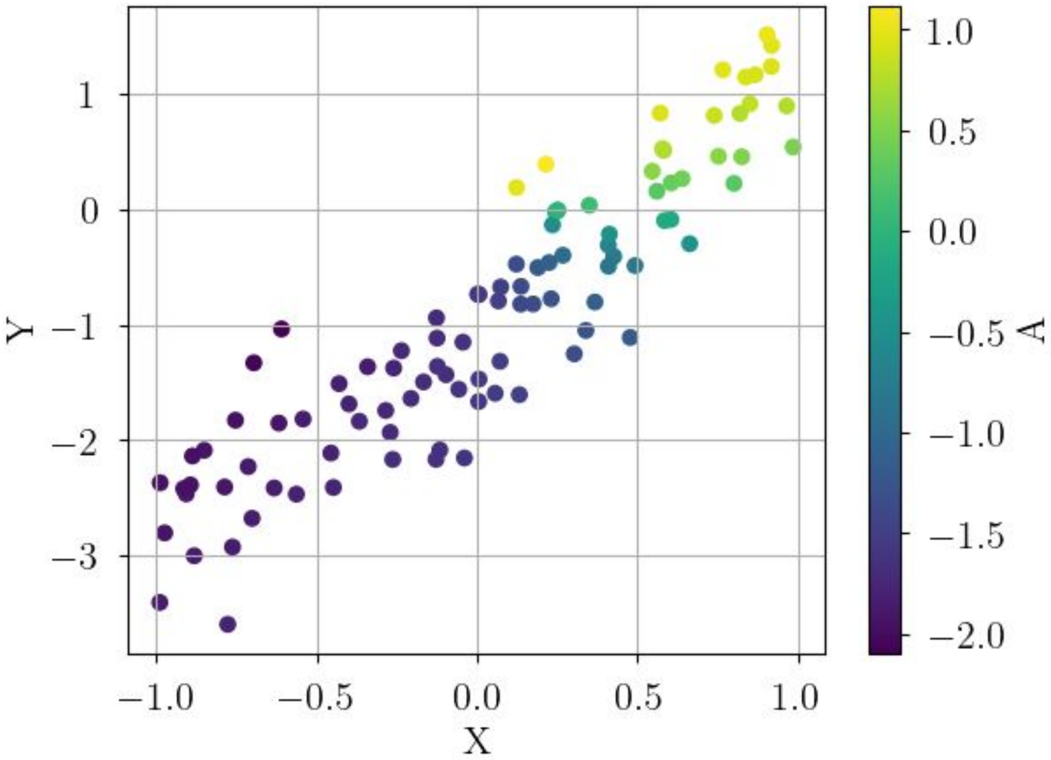
- Observables:
  - A; Y / X = tan A
  - E (measurement error of A)
- Know:
  - X between -1 and 1
  - Linear model
- Educated guess:
  - Slope is ~small positive
  - Intercept is ~small negative
- Want to infer:
  - M (slope)
  - B (intercept)
  - S (intrinsic scatter)



Trick to Bayesian Inference: How would you simulate?

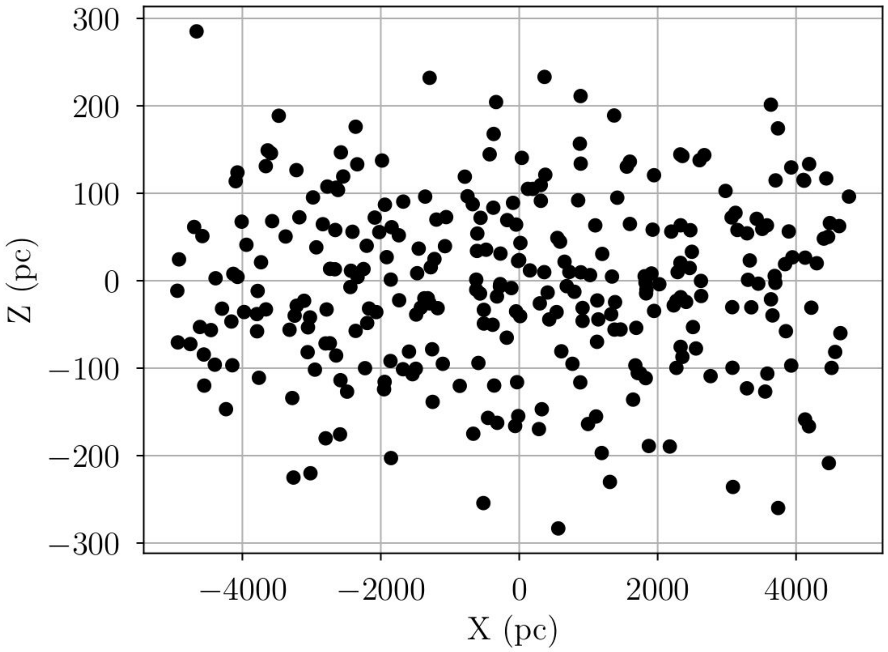
- Observables:
  - A; Y / X = tan A
  - E (measurement error of A)
- Know:
  - X between -1 and 1
  - Linear model

$$X \sim \mathrm{Uniform}(-1, 1)$$
$$Y \sim MX + B + \mathrm{Normal}(0, S)$$
$$A = \mathrm{arctan2}(Y, X) + \mathrm{Normal}(0, E)$$

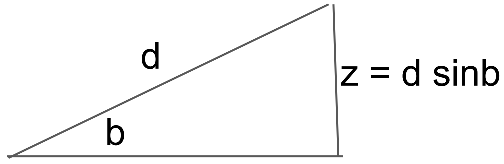


Crovisier (1978): The Problem

- Can we infer the shape of the vertical distribution of some objects in the Galactic plane?
- Given: observations of position on the sky.
- Assumption: plane-parallel Galaxy



The Crovisier (1978) method



Kinematics of Neutral Hydrogen Clouds in the Solar Vicinity from the Nançay 21-cm Absorption Survey

J. Crovisier

Département de Radioastronomie, Observatoire de Meudon, F-92190 Meudon, France

Received February 8, 1978

r is random and cannot be known individually for each cloud, but if we assume that the mean cloud density only depends on the distance $|z|$ from the Galactic plane (the plane-parallel model) and that the Sun lies in this plane, then

$$\langle r \rangle = \langle |z| \rangle / \sin |b| \qquad (4)$$

where $\langle |z| \rangle$ is the first central moment of the z-distribution of the cloud medium (its average distance from the plane).

By central limit theorem, mean observed distance = distance expectation value

$$\text{let } z = |z| \in \mathbb{R}^+$$

$$\boxed{P(z)|dz| = P(d)|dd|}$$

Transformation of probabilities in 1-D

$$P(z) = P(d)\left|\frac{dd}{dz}\right|$$

$$P(z) = \frac{P(d)}{\sin|b|}$$

$$\langle z \rangle = \int_0^\infty z P(z)\, dz$$

First moment = Expectation value (distribution mean)

$$\langle z \rangle = \int_0^\infty d \sin|b| \frac{P(d)}{\sin|b|} \sin|b|\, dd$$

$$\langle z \rangle = \sin|b| \int_0^\infty d P(d)\, dd$$

$$\langle z \rangle = \sin|b| \langle d \rangle$$

The Crovisier (1978) method

Random variable

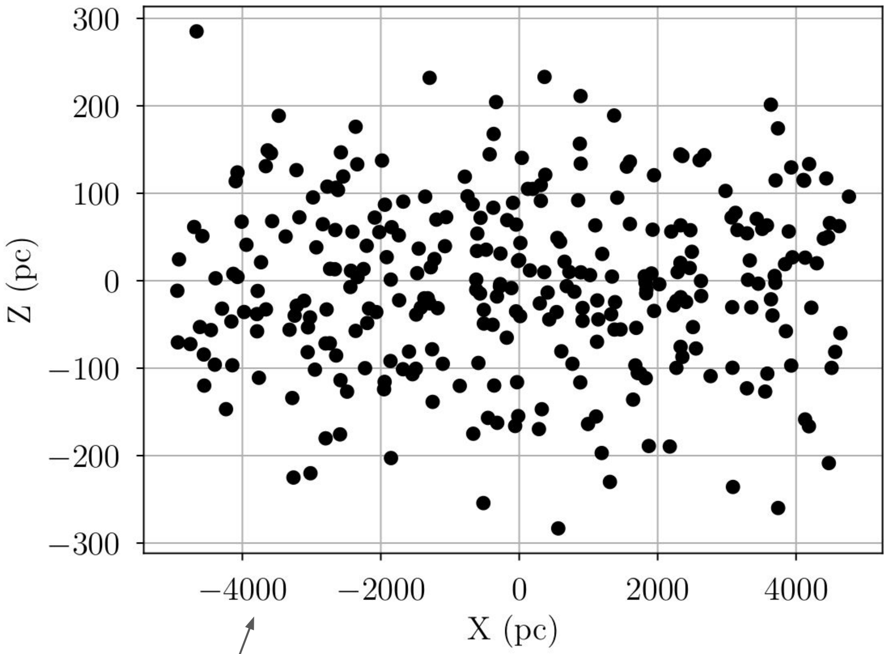
$$P_Z(|z|) = \frac{2}{\sqrt{2\pi\sigma_z^2}} e^{-z^2/(2\sigma_z^2)}$$

Probability space: values that random variable can take

$$\langle |z| \rangle = \int_0^\infty |z| P_Z(|z|)\, dz = \sqrt{\frac{2}{\pi}}\sigma_z$$

First moment = expectation value (distribution mean)

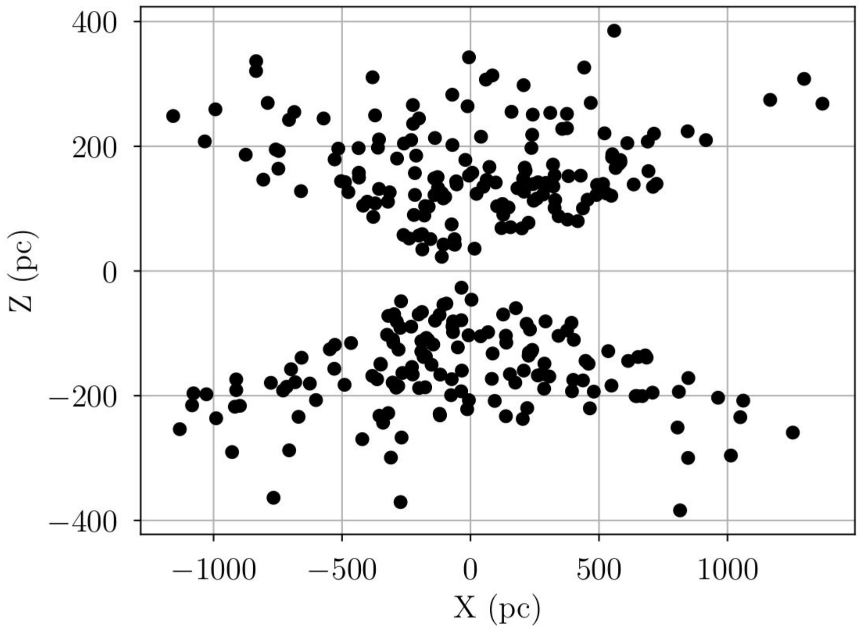
$$\sigma_z = \sqrt{\frac{\pi}{2}}\langle |z| \rangle = \sqrt{\frac{\pi}{2}}\langle d \sin |b| \rangle = 100\,\mathrm{pc}$$



Simulation result

The Crovisier (1978) method

included. In order to eliminate most non-local H I, the sample is restricted to sources at $|b| > 10$. This limitation in latitude will allow us to use simple laws for the local Galactic rotation and for the H I distribution; moreover it removes most of the complex spectra for which the Gaussian decomposition is speculative. The resulting sample consists of 299 velocity features.



$$\sigma_z = \sqrt{\frac{\pi}{2}}\langle|z|\rangle = \sqrt{\frac{\pi}{2}}\langle d\sin|b|\rangle = 200\,\mathrm{pc}$$

Simulation result IS WRONG!

The Crovisier (1978) method

included. In order to eliminate most non-local H I, the sample is restricted to sources at $|b| > 10$. This limitation in latitude will allow us to use simple laws for the local Galactic rotation and for the H I distribution; moreover it removes most of the complex spectra for which the Gaussian decomposition is speculative. The resulting sample consists of 299 velocity features.

$$\sigma_z = \sqrt{\frac{\pi}{2}}\langle|z|\rangle = \sqrt{\frac{\pi}{2}}\langle d\sin|b|\rangle = 200\,\mathrm{pc}$$

$$\text{let } z = |z| \in \mathbb{R}^+$$

$$P(z)|dz| = P(d)|dd|$$

$$P(z) = P(d)\left|\frac{dd}{dz}\right|$$

$$P(z) = \frac{P(d)}{\sin|b|}$$

$$\boxed{\langle z\rangle = \int_0^\infty zP(z)\,dz} \quad \text{Error!}$$

$$\langle z\rangle = \int_0^\infty d\sin|b|\frac{P(d)}{\sin|b|}\,\sin|b|\,dd$$

$$\langle z\rangle = \sin|b|\int_0^\infty dP(d)\,dd$$

$$\langle z\rangle = \sin|b|\langle d\rangle$$

We must consider the JOINT probability distribution between z and d

Transforming probabilities in multiple dimensions

Consider a uniform disk

$$P_X(x), P_Y(y) \sim \text{Uniform for } r < R_{\max}$$

$$P_{XY}(x,y) = P_X(x)P_Y(y) = A \text{ for } r < R_{\max}$$

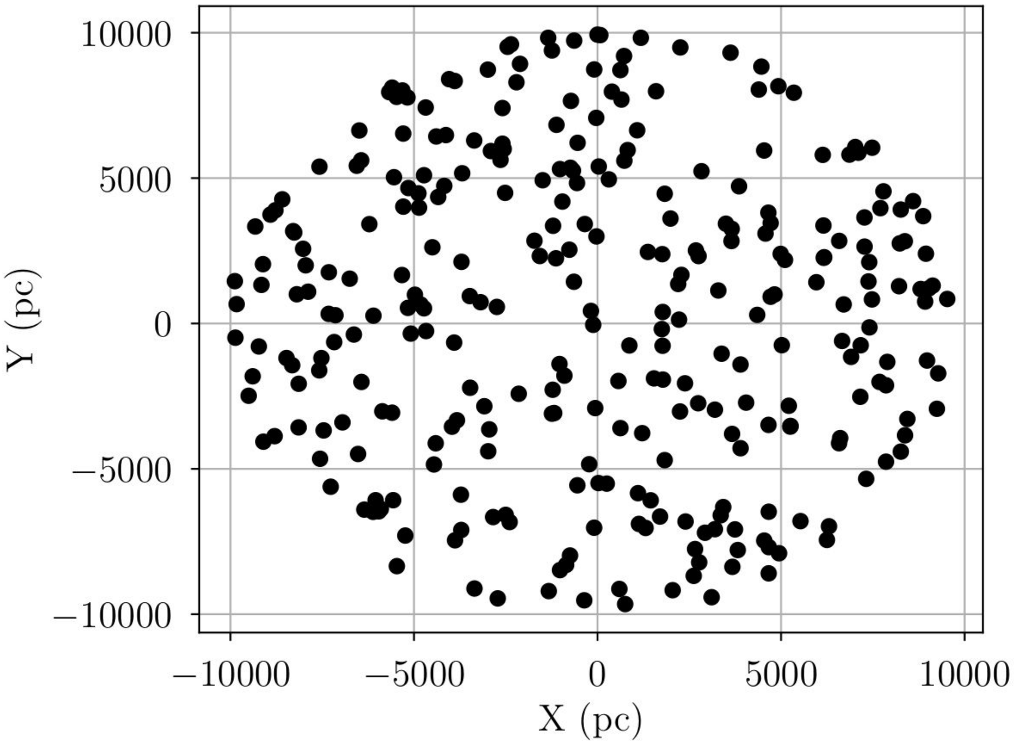
Joint distribution

$$r = \sqrt{x^2 + y^2}$$

$$\tan \ell = \frac{y}{x}$$

$$x = r\cos\ell$$

$$y = r\sin\ell$$



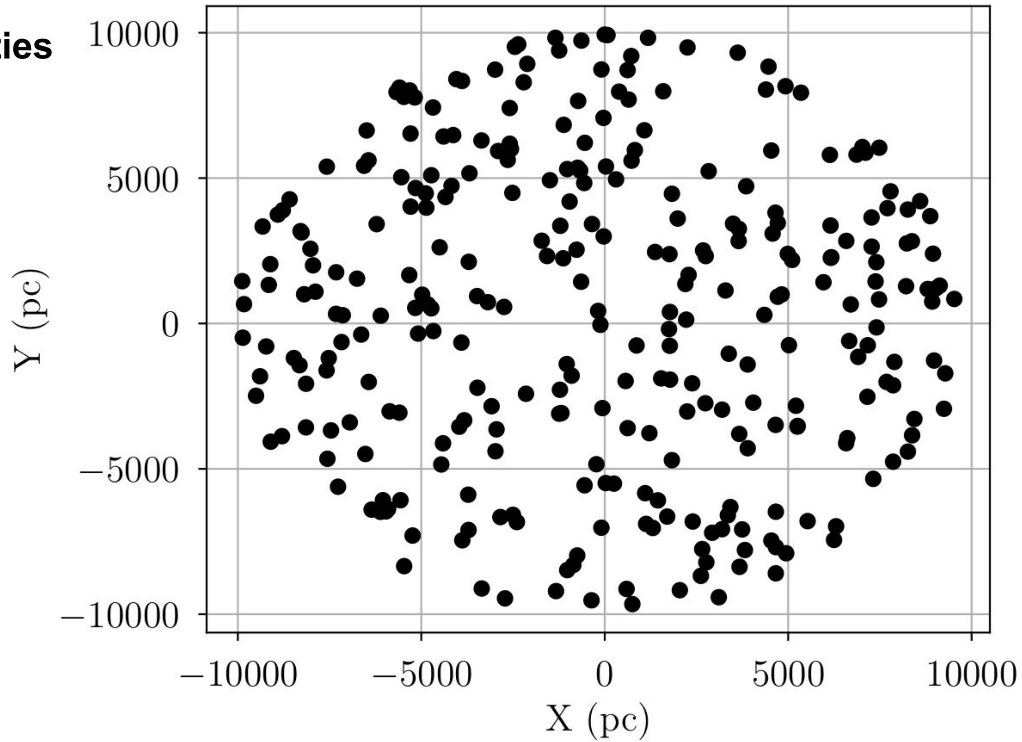
Transforming probabilities in multiple dimensions

General formula for transforming probabilities

$$\boxed{P_{RL}(r,\ell) = |J| P_{XY}(x,y)}$$

$$|J| = \det \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \ell} & \frac{\partial y}{\partial \ell} \end{vmatrix} = \det \begin{vmatrix} \cos\ell & \sin\ell \\ -r\sin\ell & r\cos\ell \end{vmatrix} = r$$

$$P_{RL}(r,\ell) = r P_{XY}(x,y) = Ar = \frac{r}{\pi R_{\max}^2}$$



Transforming Probabilities

General formula for transforming probabilities

$$\boxed{P_{RL}(r,\ell) = |J| P_{XY}(x,y)}$$

$$|J| = \det\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \ell} & \frac{\partial y}{\partial \ell} \end{vmatrix} = \det\begin{vmatrix} \cos\ell & \sin\ell \\ -r\sin\ell & r\cos\ell \end{vmatrix} = r$$

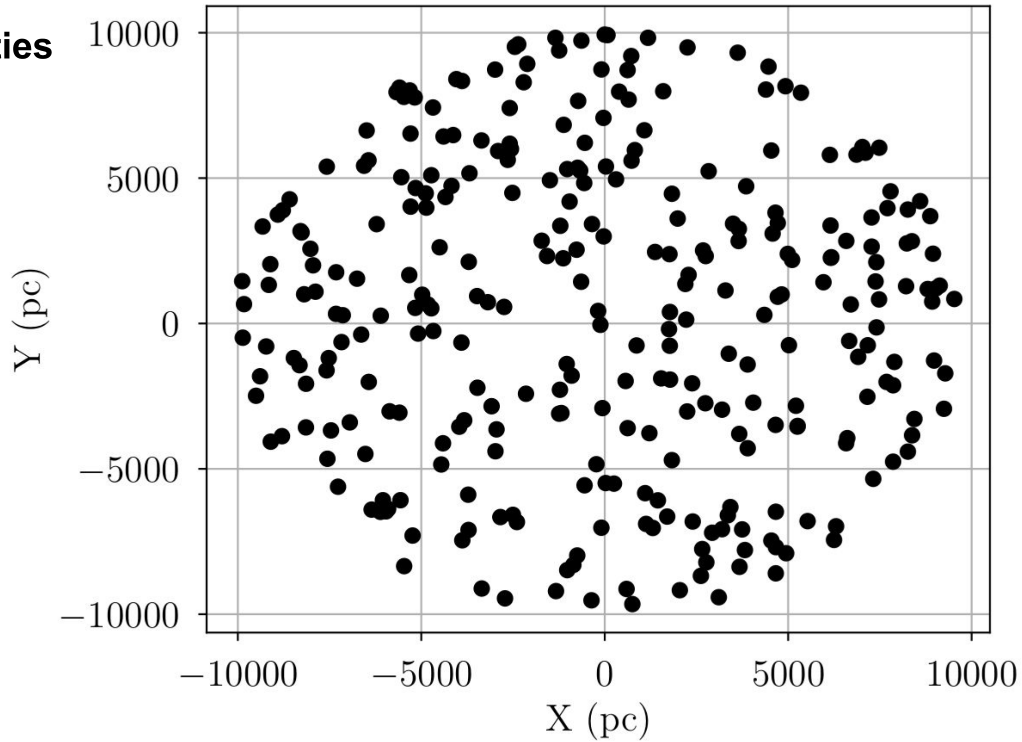
$$P_{RL}(r,\ell) = rP_{XY}(x,y) = Ar = \frac{r}{\pi R_{\max}^2}$$

Consider the direction l = 0 degrees

General formula for conditional probability

$$\boxed{P_{R|L}(r|\ell=0) = \frac{P_{RL}(r,\ell=0)}{P_L(\ell)}}$$

Conditional probability: probability of r given l.



Transforming Probabilities

General formula for transforming probabilities

$$\boxed{P_{RL}(r,\ell) = |J| P_{XY}(x,y)}$$

$$|J| = \det \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \ell} & \frac{\partial y}{\partial \ell} \end{vmatrix} = \det \begin{vmatrix} \cos\ell & \sin\ell \\ -r\sin\ell & r\cos\ell \end{vmatrix} = r$$

$$P_{RL}(r,\ell) = r P_{XY}(x,y) = Ar = \frac{r}{\pi R_{\max}^2}$$

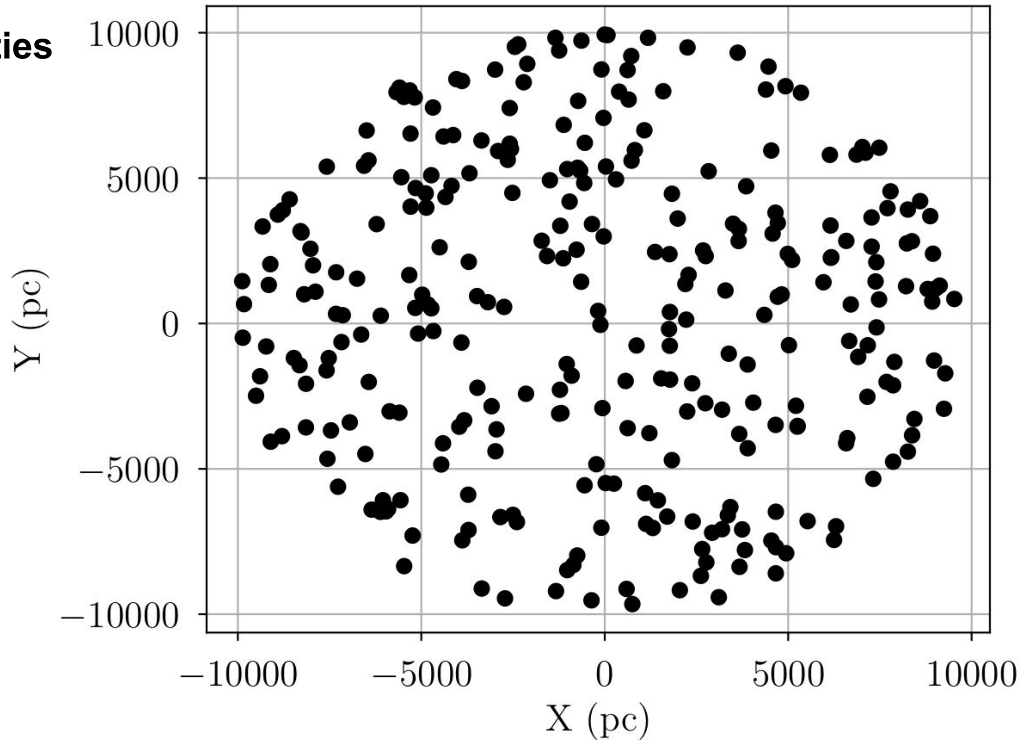
Consider the direction l = 0 degrees

$$P_{R|L}(r|\ell=0) = \frac{P_{RL}(r,\ell=0)}{P_L(\ell)}$$

General Formula for Marginal Probability

$$\boxed{P_L(\ell) = \int_0^{R_{\max}} P_{RL}(r,\ell)\,dr = \frac{1}{2\pi}}$$

Marginal probability: prob of l for all possible r



Transforming Probabilities

General formula for transforming probabilities

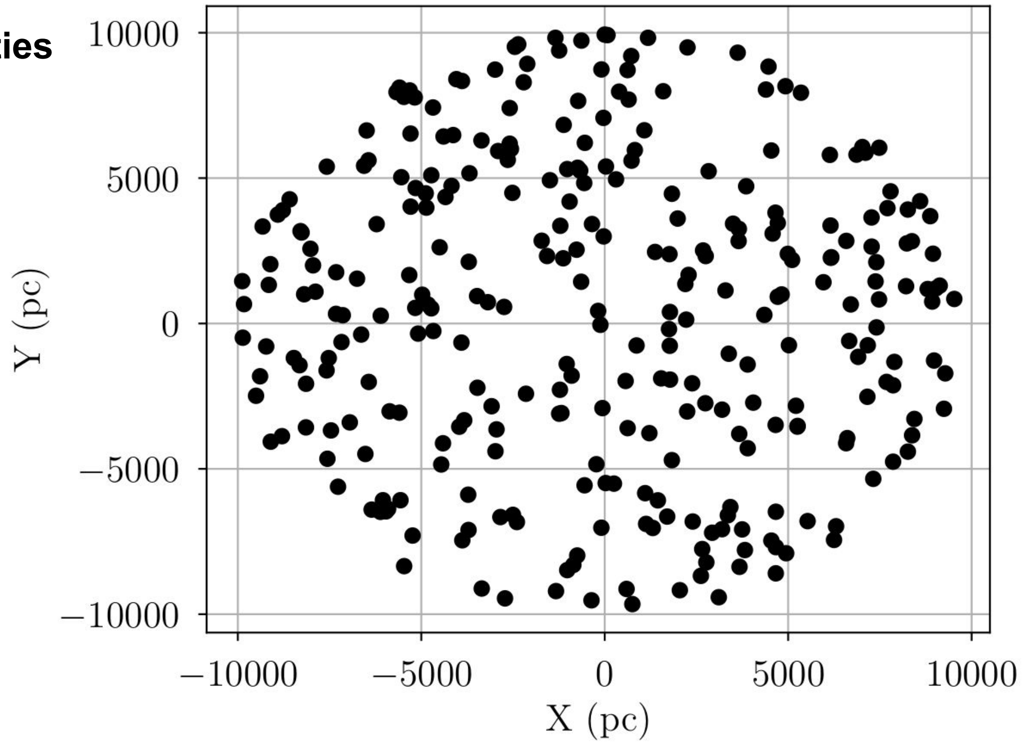
$$\boxed{P_{RL}(r,\ell) = |J| P_{XY}(x,y)}$$

$$|J| = \det \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \ell} & \frac{\partial y}{\partial \ell} \end{vmatrix} = \det \begin{vmatrix} \cos\ell & \sin\ell \\ -r\sin\ell & r\cos\ell \end{vmatrix} = r$$

$$P_{RL}(r,\ell) = r P_{XY}(x,y) = Ar = \frac{r}{\pi R_{\max}^2}$$

Consider the direction l = 0 degrees

$$P_{R|L}(r|\ell=0) = \frac{2r}{R_{\max}^2}$$



Transforming Probabilities

General formula for transforming probabilities

$$\boxed{P_{RL}(r,\ell) = |J| P_{XY}(x,y)}$$

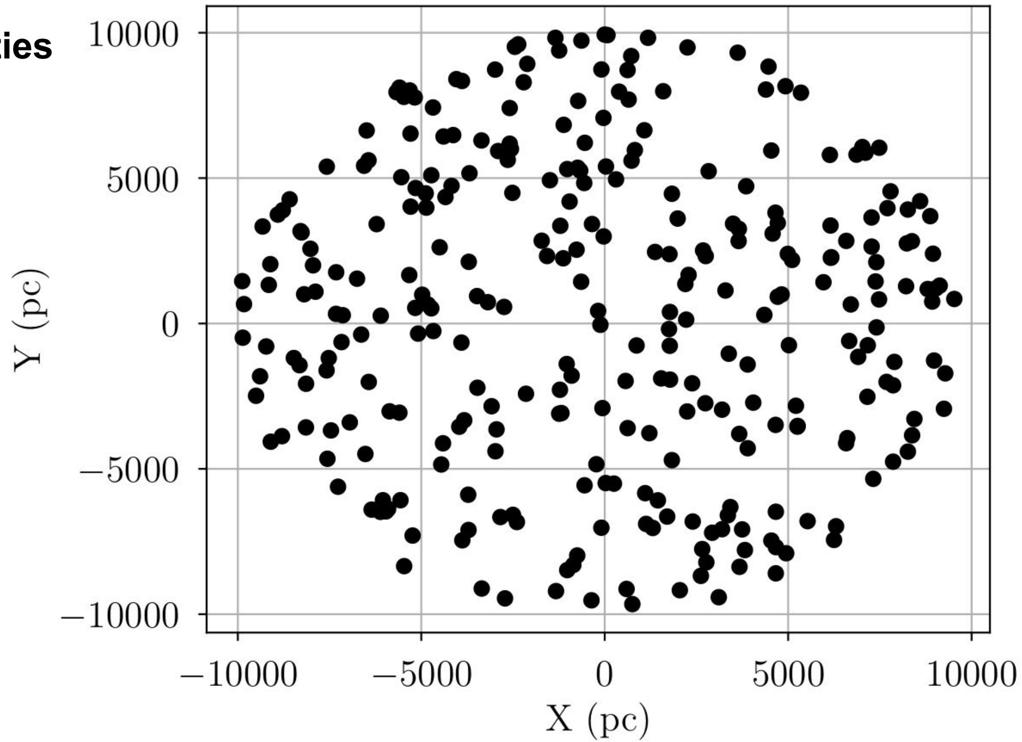
$$|J| = \det\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \ell} & \frac{\partial y}{\partial \ell} \end{vmatrix} = \det\begin{vmatrix} \cos\ell & \sin\ell \\ -r\sin\ell & r\cos\ell \end{vmatrix} = r$$

$$P_{RL}(r,\ell) = r P_{XY}(x,y) = Ar = \frac{r}{\pi R_{\max}^2}$$

Consider the direction l = 0 degrees

$$P_{R|L}(r|\ell=0) = \frac{2r}{R_{\max}^2}$$

$$P_{X|Y}(x|y=0) = \frac{1}{2R_{\max}}$$



Transforming Probabilities

General formula for transforming probabilities

$$\boxed{P_{RL}(r,\ell) = |J| P_{XY}(x,y)}$$

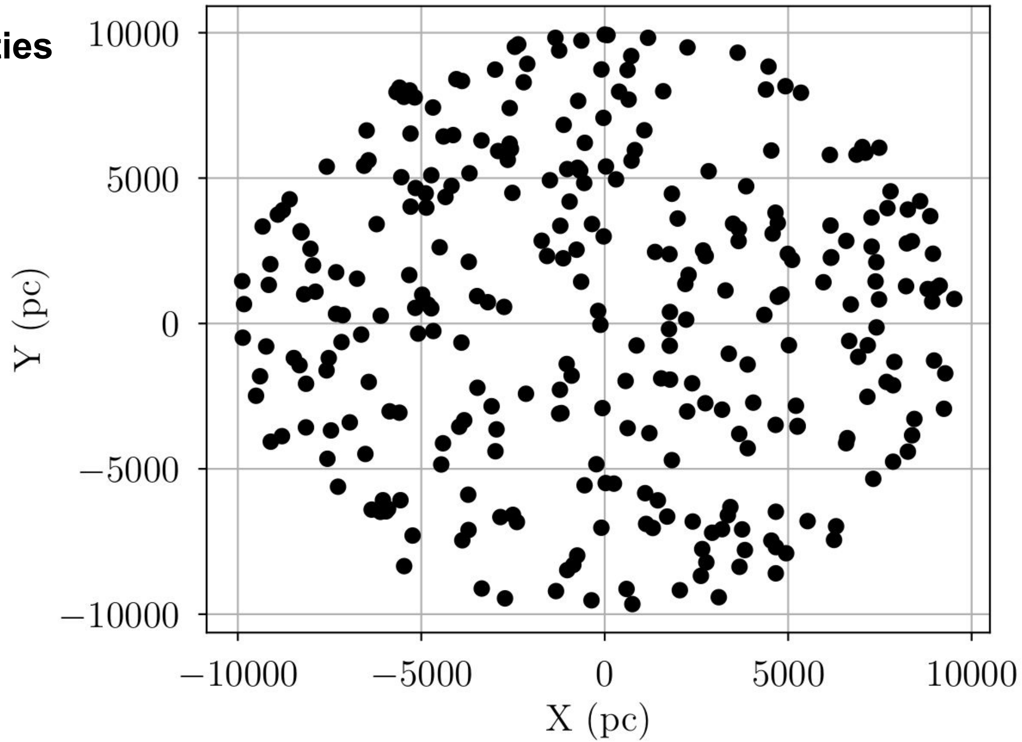
$$|J| = \det\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \ell} & \frac{\partial y}{\partial \ell} \end{vmatrix} = \det\begin{vmatrix} \cos\ell & \sin\ell \\ -r\sin\ell & r\cos\ell \end{vmatrix} = r$$

$$P_{RL}(r,\ell) = rP_{XY}(x,y) = Ar = \frac{r}{\pi R_{\max}^2}$$

Consider the direction l = 0 degrees

$$P_{R|L}(r|\ell=0) = \frac{2r}{R_{\max}^2}$$

$$P_{X|Y}(x|y=0) = \frac{1}{2R_{\max}}$$



Borel's Paradox: Conditional probabilities can do funny things under coordinate transformations.

Correcting the Crovisier (1978) Method

$$P_{RLZ}(r, \ell, |z|) = \frac{r}{\pi R_{\max}^2} P_Z(|z|)$$

$$d = \sqrt{r^2 + z^2}$$

$$\tan|b| = \frac{|z|}{r}$$

$$|z| = d \sin|b|$$

$$r = d \cos b$$

$$P_{DLB}(d, \ell, b) = \frac{r^2 \cos b}{\pi R_{\max}^2} P_Z(d \sin|b|)$$

Correcting the Crovisier (1978) Method

$$P_{RLZ}(r,\ell,|z|) = \frac{r}{\pi R_{\max}^2} P_Z(|z|)$$

$$P_{D|LB}(d|\ell,b) = \frac{P_{DLB}(d,\ell,b)}{P_{LB}(\ell,b)}$$

$$\begin{aligned} d &= \sqrt{r^2+z^2} \\ \tan|b| &= \frac{|z|}{r} \\ |z| &= d\sin|b| \\ r &= d\cos b \end{aligned}$$

$$\begin{aligned} P_{LB}(\ell,b) &= \int_0^{R_{\max}/\cos b} P_{DLB}(d,\ell,b)\,dd \\ &= \int_0^{R_{\max}\tan b} \frac{z^2\cos b}{\pi R_{\max}^2 \sin^3|b|} P_z(z)\,dz \\ &= \frac{\cos b}{\pi R_{\max}^2 \sin^3|b|}\mu_2 \quad \text{as} \quad R_{\max}\tan b \to \infty \end{aligned}$$

Second moment = variance

$$P_{DLB}(d,\ell,b) = \frac{r^2\cos b}{\pi R_{\max}^2} P_Z(d\sin|b|)$$

Correcting the Crovisier (1978) Method

$$P_{RLZ}(r, \ell, |z|) = \frac{r}{\pi R_{\max}^2} P_Z(|z|)$$

$$d = \sqrt{r^2 + z^2}$$

$$\tan|b| = \frac{|z|}{r}$$

$$|z| = d \sin|b|$$

$$r = d \cos b$$

$$P_{DLB}(d, \ell, b) = \frac{r^2 \cos b}{\pi R_{\max}^2} P_Z(d \sin|b|)$$

$$P_{D|LB}(d|\ell, b) = \frac{P_{DLB}(d, \ell, b)}{P_{LB}(\ell, b)}$$

$$P_{LB}(\ell, b) = \int_0^{R_{\max}/\cos b} P_{DLB}(d, \ell, b)\, dd = \frac{\cos b}{\pi R_{\max}^2 \sin^3|b|} \mu_2$$

$$P_{D|LB}(d|\ell, b) = \frac{d^2 \sin^3|b|}{\mu_2} P_Z(d \sin|b|)$$

$$\boxed{\langle d|\ell, b\rangle = \frac{\mu_3}{\mu_2 \sin|b|}}$$

Third moment = skewness

Correcting the Crovisier (1978) Method

$$P_{RLZ}(r,\ell,|z|) = \frac{r}{\pi R_{\rm max}^2} P_Z(|z|)$$

$$d = \sqrt{r^2+z^2}$$

$$\tan|b| = \frac{|z|}{r}$$

$$|z| = d\sin|b|$$

$$r = d\cos b$$

$$P_{DLB}(d,\ell,b) = \frac{r^2\cos b}{\pi R_{\rm max}^2} P_Z(d\sin|b|)$$

$$P_{D|LB}(d|\ell,b) = \frac{P_{DLB}(d,\ell,b)}{P_{LB}(\ell,b)}$$

$$P_{LB}(\ell,b) = \int_0^{R_{\rm max}/\cos b} P_{DLB}(d,\ell,b)\,dd = \frac{\cos b}{\pi R_{\rm max}^2 \sin^3|b|}\mu_2$$

$$P_{D|LB}(d|\ell,b) = \frac{d^2\sin^3|b|}{\mu_2} P_Z(d\sin|b|)$$

$$\langle d|\ell,b\rangle = \frac{\mu_3}{\mu_2\sin|b|} \neq \frac{\mu_1}{\sin|b|}$$

Crovisier (1978) result

Implications

For a half-normal distribution

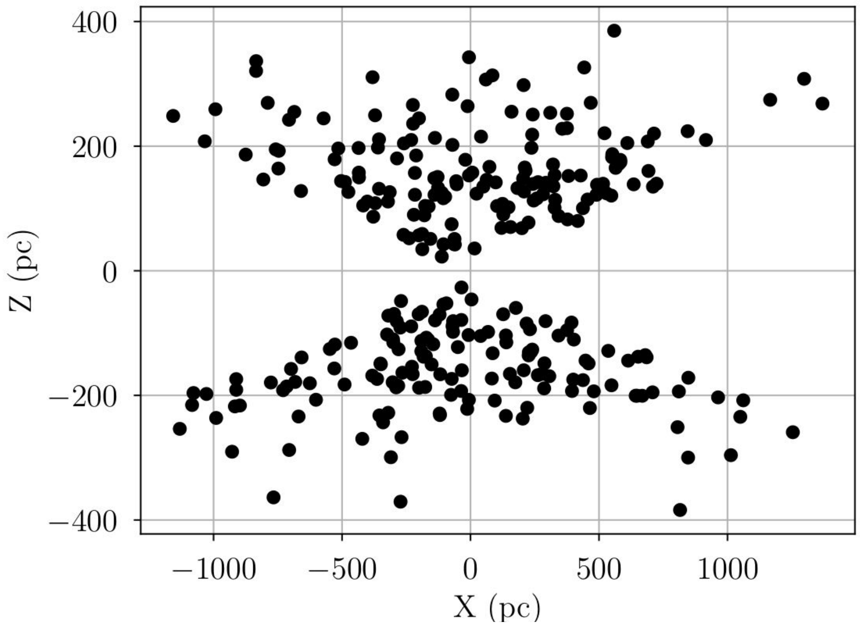
$$P_Z(|z|) = \frac{2}{\sqrt{2\pi\sigma_z^2}} e^{-z^2/(2\sigma_z^2)}$$

$$\mu_1 = \sqrt{\frac{2}{\pi}}\sigma_z$$ Mean

$$\mu_2 = \sigma_z^2$$ Variance

$$\mu_3 = 2\sqrt{\frac{2}{\pi}}\sigma_z^3$$ Skewness

$$\langle d|\ell, b\rangle = 2\frac{\mu_1}{\sin|b|}$$



Bingo!

$$\sigma_z = \frac{1}{2}\sqrt{\frac{\pi}{2}}\langle|z|\rangle = \frac{1}{2}\sqrt{\frac{\pi}{2}}\langle d \sin|b|\rangle = 100\,\mathrm{pc}$$