THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

COIN
Cosmostatistics Initiative

Rafael S. de Souza,

in collaboration with E. E. O. Ishida

Researcher at the University of North Carolina at Chapel Hill

Chapel Hill, USA

Astrophysics + Statistics + Machine Learning

Interdisciplinary science development

Chair of the Cosmostatistics Initiative (COIN)

Day 1: Overview

1. Interpretation

    Bayesian x Frequentists

2. Parameter inference

    Markov Chain Monte Carlo

3. Language/tools

    JAGS, ...

https://github.com/RafaelSdeSouza/Bayes_ESTEC
or
https://rafaelsdesouza.github.io/Bayes_ESTEC/

"Begin at the beginning,"
the King said, very gravely,
"and go on till you come to the end:
then stop".

Alice in Wonderland, Lewis Carroll (1865)

In the beginning ...

Statistics ⟷ *State*

New Latin: *Statisticum collegium*

Italian: *statista* ("statesman")

German: *Statistik (science of state)*

Taxes

Medicine

In the beginning ...

Statistics ⟷ *State*

Mean

Variance

Frequency

Taxes

Medicine

In the beginning ...

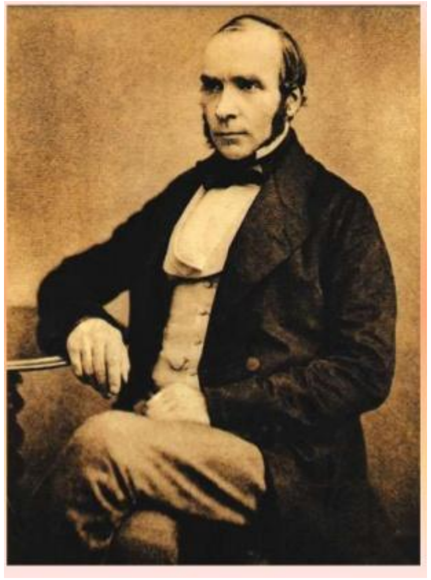
Statistics ⟷ *State*

Taxes

Medicine

In the beginning …

Statistics ⟷ *State*



John Snow (1813 – 1858)
The father of epidemiology

Used data to prove that cholera was transmitted via contaminated water

Taxes

Medicine

… then, gambling!

Blaise Pascal

1654

Pierre de Fermat

Atoine Gambaud asked for help in games of chance for Pascal.

Gambling led, indirectly, to the birth of probability theory. In the mid-17th century, an exchange of letters between Blaise Pascal and Pierre de Fermat–laid the foundation for probability.

APS NEWS

July 2009 (Volume 18, Number 7)

This Month in Physics History

July 1654: Pascal's Letters to Fermat on the "Problem of Points"

http://homepages.wmich.edu/~mackey/Teaching/145/probHist.html

... then, gambling!

Blaise Pascal

1654

Pierre de Fermat

Christian Huygens

On Reasoning in Games of Chance, 1657

http://homepages.wmich.edu/~mackey/Teaching/145/probHist.html

… then, gambling!

Blaise Pascal

1654

Pierre de Fermat

Christian Huygens

On Reasoning in Games of Chance, 1657

Chance

http://homepages.wmich.edu/~mackey/Teaching/145/probHist.html

… then, gambling!

Blaise Pascal

1654

Pierre de Fermat

Christian Huygens

On Reasoning in Games of Chance, 1657

Probability

http://homepages.wmich.edu/~mackey/Teaching/145/probHist.html

… then, gambling!

Blaise Pascal

1654

Pierre de Fermat

Christian Huygens

On Reasoning in Games of Chance, 1657

An Essay towards solving a Problem in the Doctrine of Chances, 1763

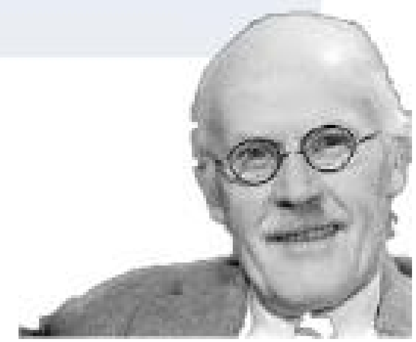
Thomas Bayes

http://homepages.wmich.edu/~mackey/Teaching/145/probHist.html

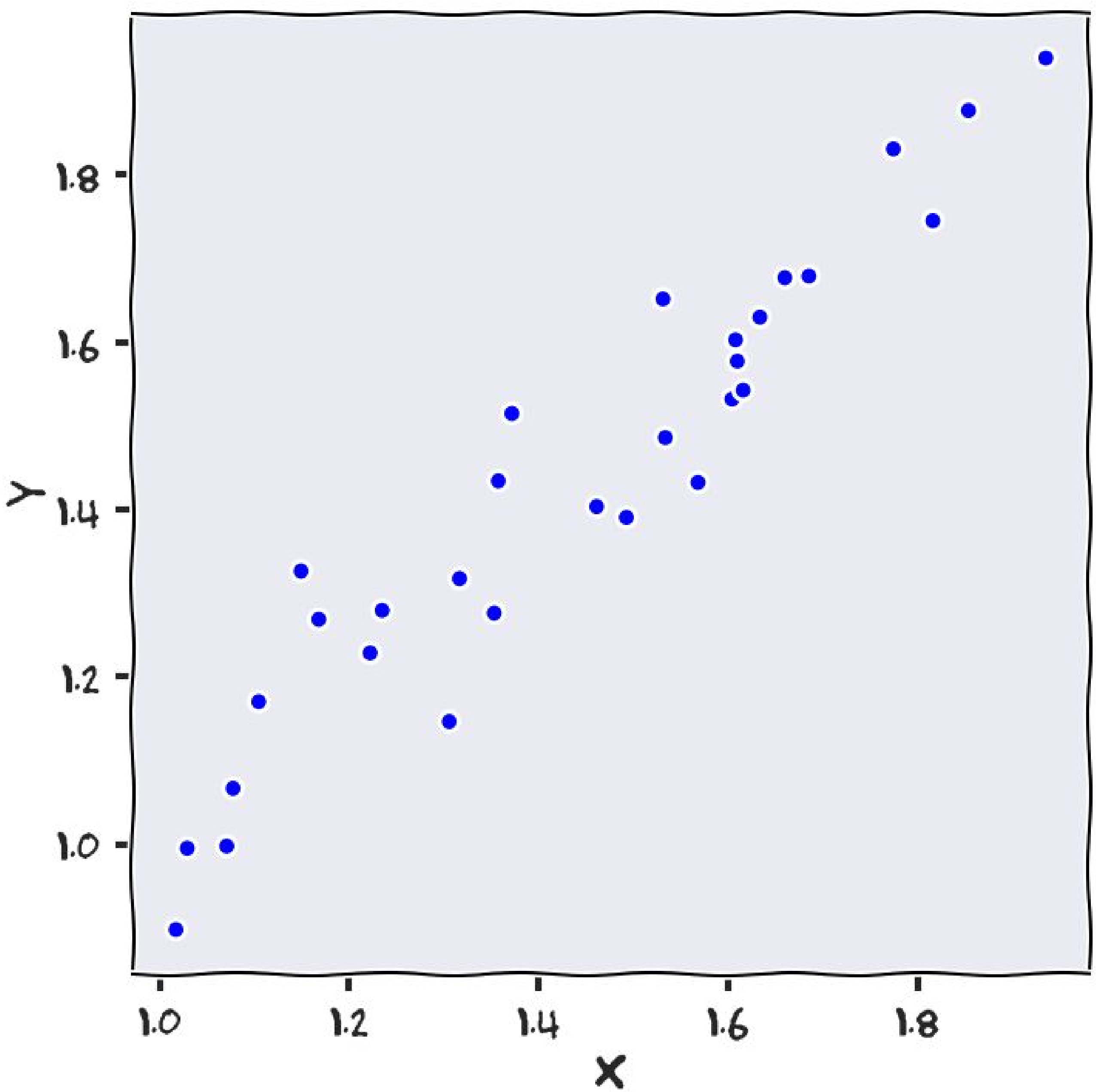
What is a probability?

Frequentist	Bayesian
Probability is a long-run average	Probability is a degree of belief
There is a true Model, the Data is a random realization	The Data is true/fixed, Models have probabilities
Probability of the data given a hypothesis (Likelihood)	Probability of a hypothesis given the data
Each repeated experiment/observation starts from ignorance	Can incorporate prior knowledge: probabilities can be updated


Jerzy Neyman


Harold Jeffereys

A first example:



$$y = a\,x + b$$

Ordinary Least Squares: $$\{a, b\} \leftarrow \min\left[\sum_{i=1}^{N} \frac{(y_i - (ax_i + b))^2}{\sigma^2}\right]$$

$y = a x + b$

Ordinary Least Squares:

$$\{a, b\} \leftarrow \min \left[\sum_{i=1}^{N} \frac{(y_i - (ax_i + b))^2}{\sigma^2} \right]$$

Main assumptions:

- independent observations (iid)
  no correlations
- all the points are equally valid
  no outliers
- errors are Gaussian distributed
  homoscedasticity
- no exterior information
  no priors

https://www.albert.io/blog/key-assumptions-of-ols-econometrics-review/

$$y = a\,x + b$$

"Physical" model

Ordinary Least Squares:

$$\{a, b\} \leftarrow \min \left[\sum_{i=1}^{N} \frac{(y_i - (ax_i + b))^2}{\sigma^2} \right]$$

Main assumptions:

- independent observations (iid)
  no correlations
- all the points are equally valid
  no outliers
- errors are Gaussian distributed
  homoscedasticity
- no exterior information
  no priors

$$\left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\} \text{Statistical model}$$

https://www.albert.io/blog/key-assumptions-of-ols-econometrics-review/

$$y = a\,x + b$$

Model of the mean

Ordinary Least Squares: $$\{a, b\} \leftarrow \min \left[\sum_{i=1}^{N} \frac{(y_i - (ax_i + b))^2}{\sigma^2} \right]$$

Main assumptions:

- independent observations (iid)
  no correlations
- all the points are equally valid
  no outliers
- errors are Gaussian distributed
  homoscedasticity
- no exterior information
  no priors

Model
for the
uncertainties

https://www.albert.io/blog/key-assumptions-of-ols-econometrics-review/

Frequentist approach

If $\{X, Y\}$ are random variables:

$X \sim \mathit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \mathit{Normal}(0, \sigma^2)$

or

$Y \sim \mathit{Normal}(aX + b, \sigma^2)$

Frequentist approach

If $\{X, Y\}$ are random variables:

$X \sim \textit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \textit{Normal}(0, \sigma^2)$

or

$Y \sim \textit{Normal}(aX + b, \sigma^2)$

Likelihood for 1 point:

$$\mathcal{L}(x, y|a, b, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - (ax + b))^2}{2\sigma^2}\right]$$

Frequentist approach

If {X, Y} are random variables:

X ~ *Uniform*(1.0, 2.0)

Y = aX + b + ε

ε ~ *Normal*(0, σ^2)

or

Y ~ *Normal*(aX + b, σ^2)

Likelihood for 1 point:

$$\mathcal{L}(x, y|a, b, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - (ax + b))^2}{2\sigma^2}\right]$$

Likelihood for the complete data set:

$$\begin{aligned}\mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - (ax_i + b))^2}{2\sigma^2}\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - (ax_i + b))^2}{\sigma^2}\right]\end{aligned}$$

Frequentist approach

If $\{X, Y\}$ are random variables:

$X \sim \textit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \textit{Normal}(0, \sigma)$

or

$Y \sim \textit{Normal}(aX + b, \sigma)$

Likelihood for 1 point:

$$\mathcal{L}(\boxed{x, y|a, b, \sigma}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - (ax + b))^2}{2\sigma^2}\right]$$

Likelihood for the complete data set:

$$\mathcal{L}(\boxed{\vec{x}, \vec{y}|a, b, \sigma}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - (ax_i + b))^2}{2\sigma^2}\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - (ax_i + b))^2}{\sigma^2}\right]$$

Maximum Likelihood

$$\theta_{\mathrm{ML}} \equiv \max_{\theta} \mathscr{L}(\theta)$$

If Gaussian

$$\mathscr{L}(\mu,\sigma) = p(\hat{x}|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\frac{(\hat{x}_i-\mu)^2}{\sigma^2}\right)$$

$$\mathscr{L} = L_0 \exp\left(-\chi^2/2\right)$$

$$\chi^2 = \sum_{i=1}^{N} \frac{(\hat{x}_i-\mu)^2}{\sigma^2}$$

Confidence intervals

Taylor series expansion around maximum of a general likelihood function:

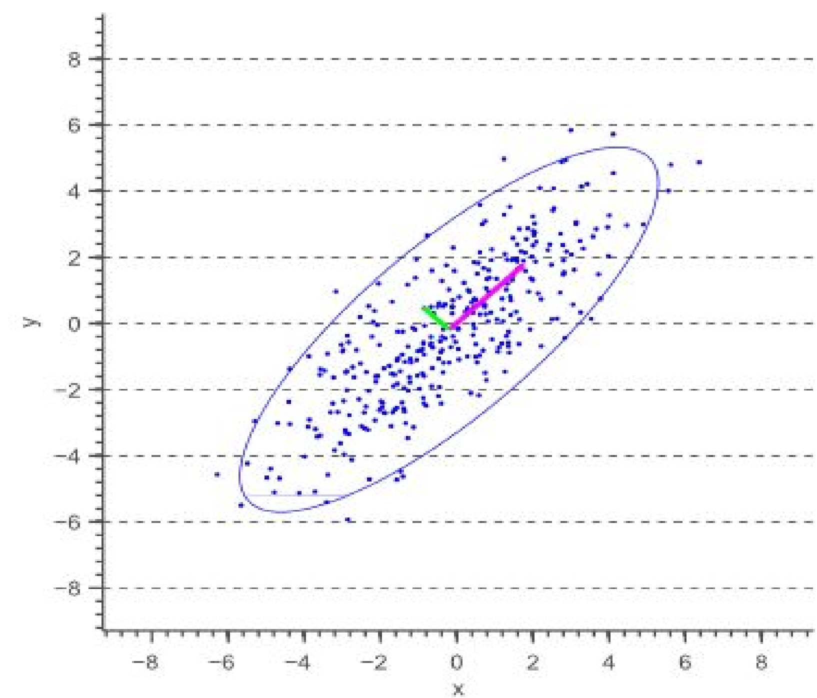
$$\ln \mathcal{L}(\theta) = \ln \mathcal{L}|_{\hat{\theta}} + \cancel{(\theta - \hat{\theta}) \frac{\partial \ln \mathcal{L}}{\partial \theta}\Big|_{\hat{\theta}}} + \frac{1}{2}(\theta - \hat{\theta})^2 \frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\Big|_{\hat{\theta}} + \ldots \qquad \frac{\partial \ln \mathcal{L}}{\partial a}\Big|_{\hat{\theta}} = \frac{\partial \ln \mathcal{L}}{\partial b}\Big|_{\hat{\theta}} = 0$$

A general likelihood function can be approximated to second order as a Gaussian around the ML value

$$\mathcal{L}(\theta) \approx \mathcal{L}(\hat{\theta}) \exp\left(-\frac{1}{2}\frac{(\theta - \hat{\theta})^2}{C_{\hat{\theta}}}\right)$$

$$C_{\hat{\theta}} = F^{-1} = \left(-\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\Big|_{\hat{\theta}}\right)^{-1}$$

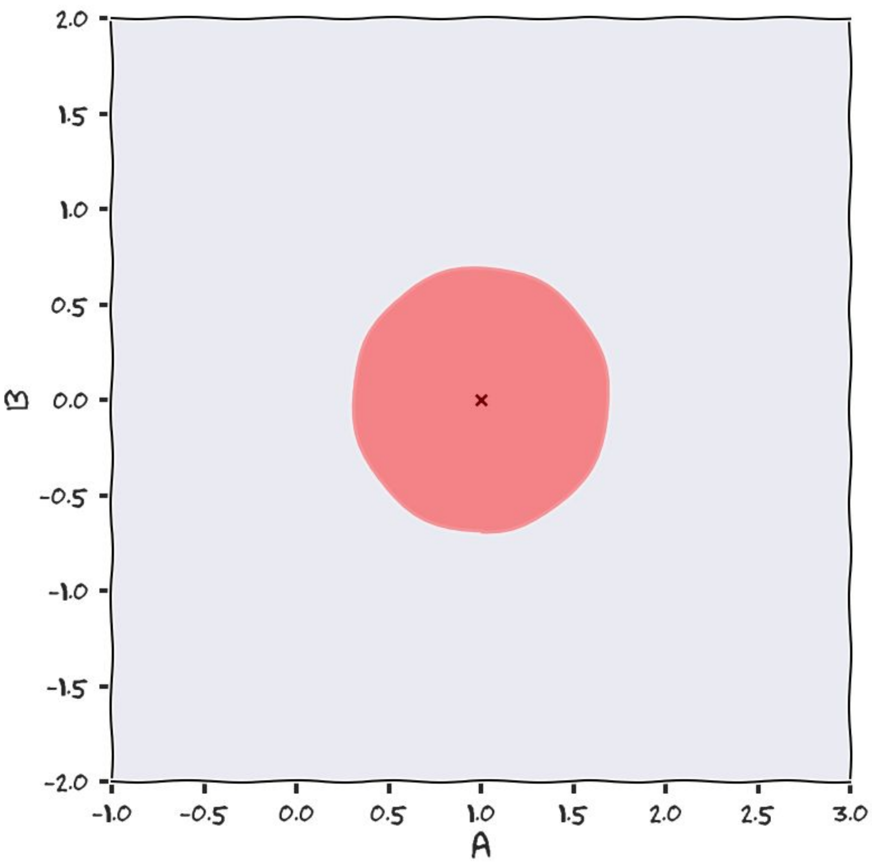
Fisher Matrix



y = a x + b

Frequentist:

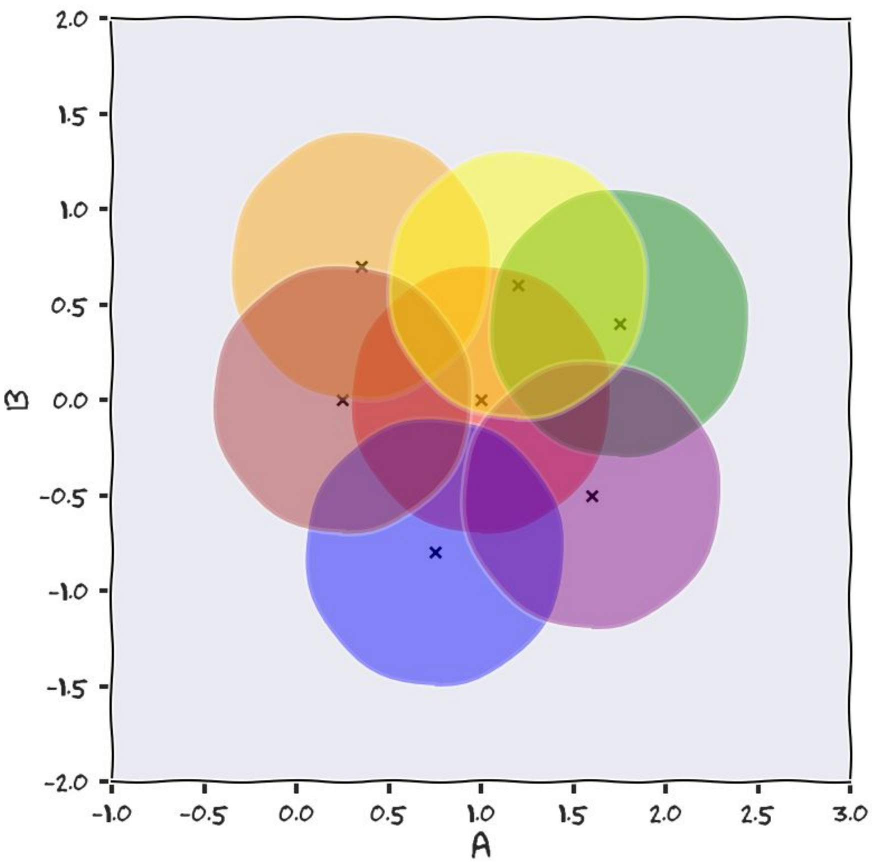
$$\mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - (ax_i + b))^2}{\sigma^2}\right]$$



$y = a\,x + b$

Frequentist:

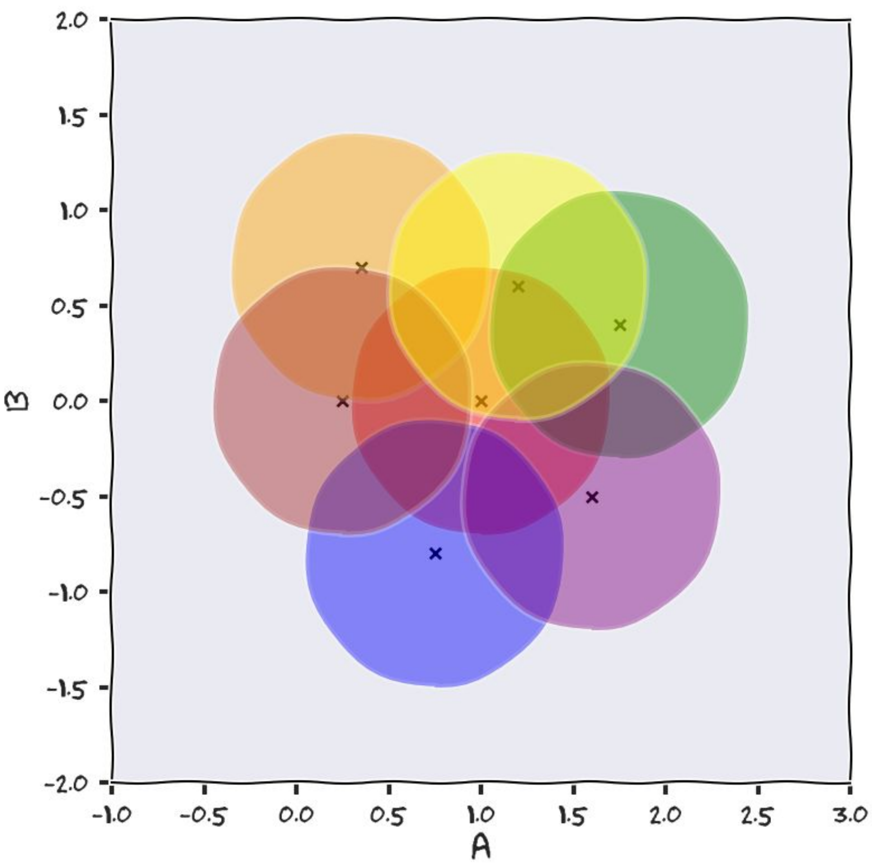
$$\mathcal{L}(\vec{x}, \vec{y} | a, b, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - (ax_i + b))^2}{\sigma^2}\right]$$



y = a x + b

Frequentist:

$$\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i-(ax_i+b))^2}{\sigma^2}\right]$$



Frequentist:

95% confidence interval →
95% of the intervals derived from possible data will contain the true value

Bayesian approach

If {X, Y} are random variables:

$X \sim \mathit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \mathit{Normal}(0, \sigma^2)$

or

$Y \sim \mathit{Normal}(aX + b, \sigma^2)$

+

$a \sim \mathit{Normal}(0, 5.0)$

$b \sim \mathit{Normal}(0, 5.0)$

Bayesian approach

If $\{X, Y\}$ are random variables:

$X \sim \mathit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \mathit{Normal}(0, \sigma^2)$

or

$Y \sim \mathit{Normal}(aX + b, \sigma^2)$

+

$a \sim \mathit{Normal}(0, 5.0)$

$b \sim \mathit{Normal}(0, 5.0)$

Bayes theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayesian approach

If $\{X, Y\}$ are random variables:

Bayes theorem

$$P(\vec{\theta}|\mathcal{D}) = \frac{P(\mathcal{D}|\vec{\theta})P(\vec{\theta})}{P(\mathcal{D})}$$

$X \sim \mathit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \mathit{Normal}(0, \sigma^2)$

or

$Y \sim \mathit{Normal}(aX + b, \sigma^2)$

$+$

$a \sim \mathit{Normal}(0, 5.0)$

$b \sim \mathit{Normal}(0, 5.0)$

Bayesian approach

If $\{X, Y\}$ are random variables:

Bayes theorem

$$P(a, b|\vec{x}, \vec{y}, \sigma) = \frac{\mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma)p(a)p(b)}{E(\vec{x}, \vec{y})}$$

$X \sim \mathit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \mathit{Normal}(0, \sigma^2)$

or

$Y \sim \mathit{Normal}(aX + b, \sigma^2)$

+

$a \sim \mathit{Normal}(0, 5.0)$

$b \sim \mathit{Normal}(0, 5.0)$

Bayesian approach

If {X, Y} are random variables:

$X \sim \textit{Uniform}(1.0, 2.0)$

$Y = aX + b + \varepsilon$

$\varepsilon \sim \textit{Normal}(0, \sigma^2)$

or

$Y \sim \textit{Normal}(aX + b, \sigma^2)$

+

$a \sim \textit{Normal}(0, 5.0)$

$b \sim \textit{Normal}(0, 5.0)$

Bayes theorem

$$P(a, b|\vec{x}, \vec{y}, \sigma) = \frac{\mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma)p(a)p(b)dadb}$$

Bayesian approach

If {X, Y} are random variables:

X ~ *Uniform*(1.0, 2.0)

Y = aX + b + ε

ε ~ *Normal*(0, σ²)

or

Y ~ *Normal*(aX + b, σ²)

+

a ~ *Normal*(0, 5.0)

b ~ *Normal*(0, 5.0)

Bayes theorem

$$\boxed{P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a\int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}}$$

$$\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i-(ax_i+b))^2}{\sigma^2}\right]$$

Bayesian approach

If {X, Y} are random variables:

X ~ *Uniform*(1.0, 2.0)

Y = aX + b + ε

ε ~ *Normal*(0, σ²)

or

Y ~ *Normal*(aX + b, σ²)

+

a ~ *Normal*(0, 5.0)

b ~ *Normal*(0, 5.0)

Bayes theorem

$$\boxed{P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}}$$

$$\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i-(ax_i+b))^2}{\sigma^2}\right]$$

$$p(a) = \frac{1}{\sqrt{2\pi(5)^2}} \exp\left[-\frac{(a-0)^2}{2\times(5^2)}\right]$$

$$p(b) = \frac{1}{\sqrt{2\pi(5)^2}} \exp\left[-\frac{(b-0)^2}{2\times(5^2)}\right]$$

Bayesian approach

Likelihood

$$P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}$$

Bayesian approach

Likelihood

Priors

$$P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}$$

Bayesian approach

Likelihood

Priors

$$P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}$$

Evidence

Bayesian approach

Likelihood

Posterior

Priors

$$P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}$$

Evidence

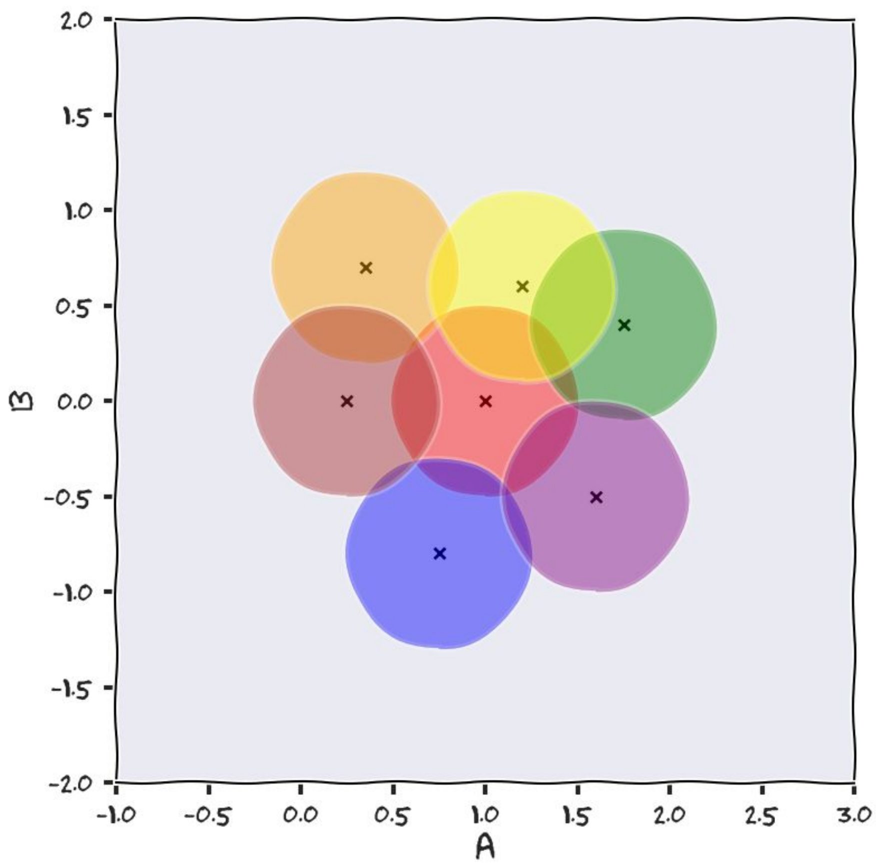
Bayesian approach

$$P(\boxed{a, b|\vec{x}, \vec{y}, \sigma}) = \frac{\mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x}, \vec{y}|a, b, \sigma)p(a)p(b)dadb}$$

y = a x + b

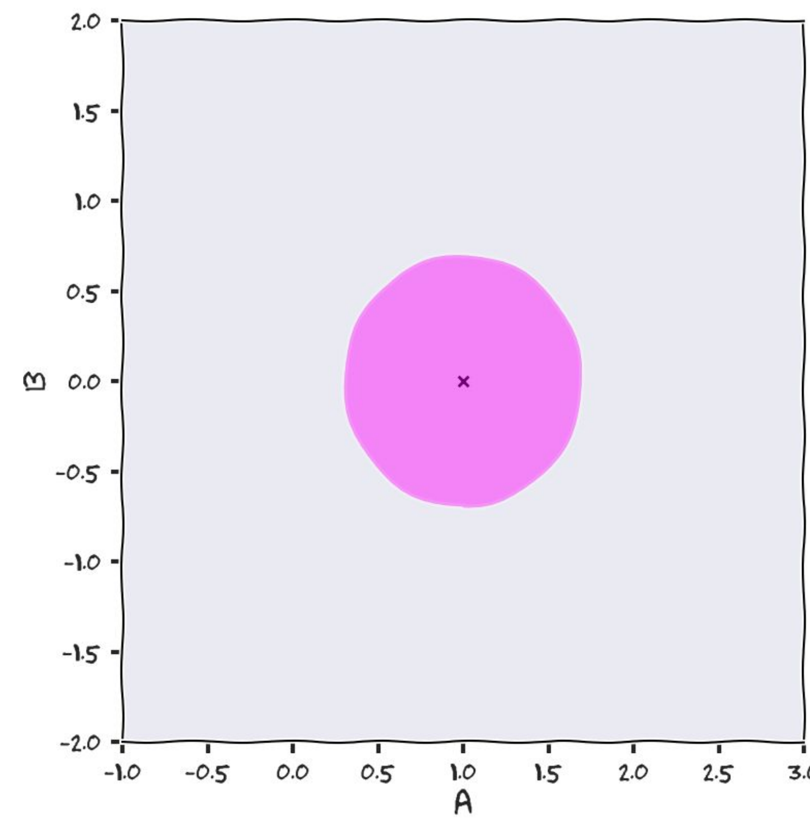
Frequentist:

$$\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i-(ax_i+b))^2}{\sigma^2}\right]$$



Bayesian:

$$P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a\int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}$$



Frequentist:

95% confidence interval →
95% of the intervals derived from possible data will contain the true value

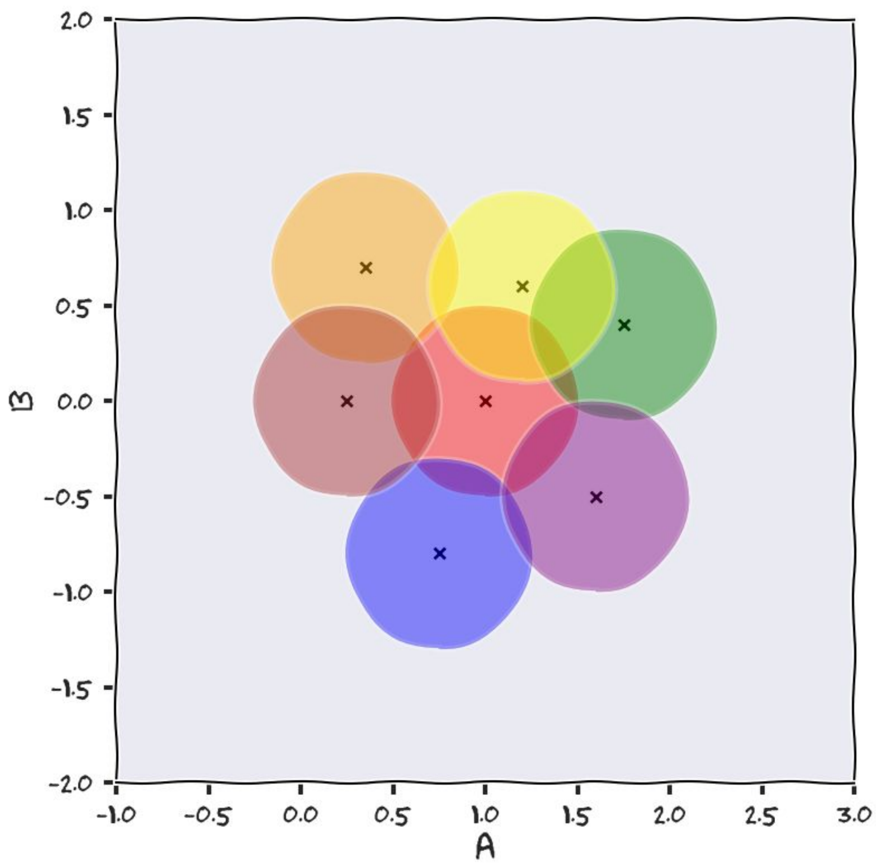
Bayesian:

95% credible interval →
There is a 95% chance that the true value is within this interval

y = a x + b

Frequentist:

$$\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i-(ax_i+b))^2}{\sigma^2}\right]$$



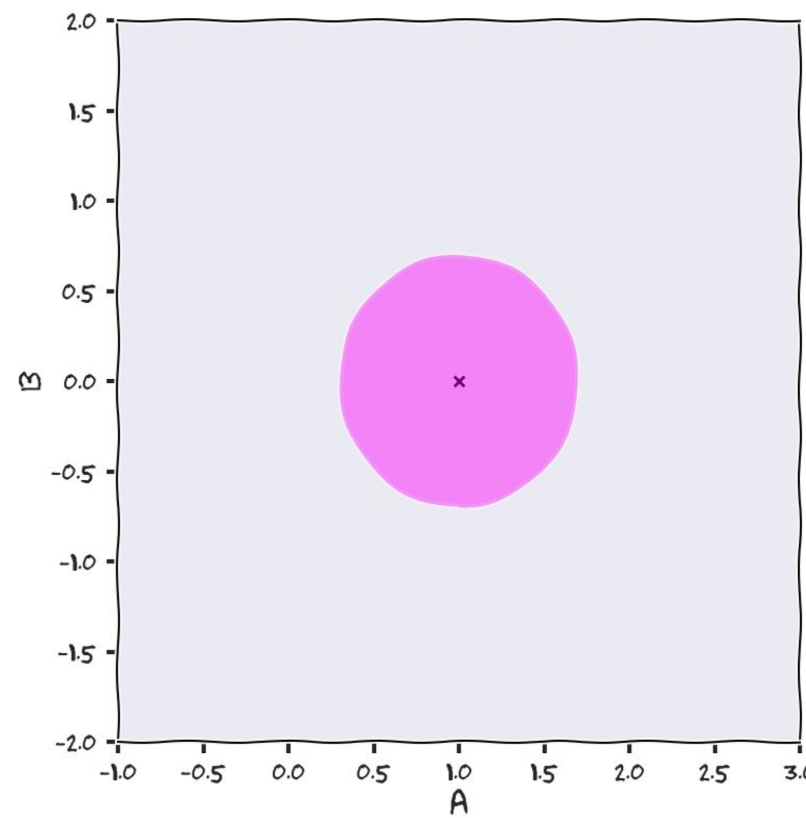
<u>Frequentist:</u>

95% confidence interval →

95% of the intervals derived from possible data will contain the true value

Bayesian:

$$P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a\int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)dadb}$$



belief

<u>Bayesian:</u>

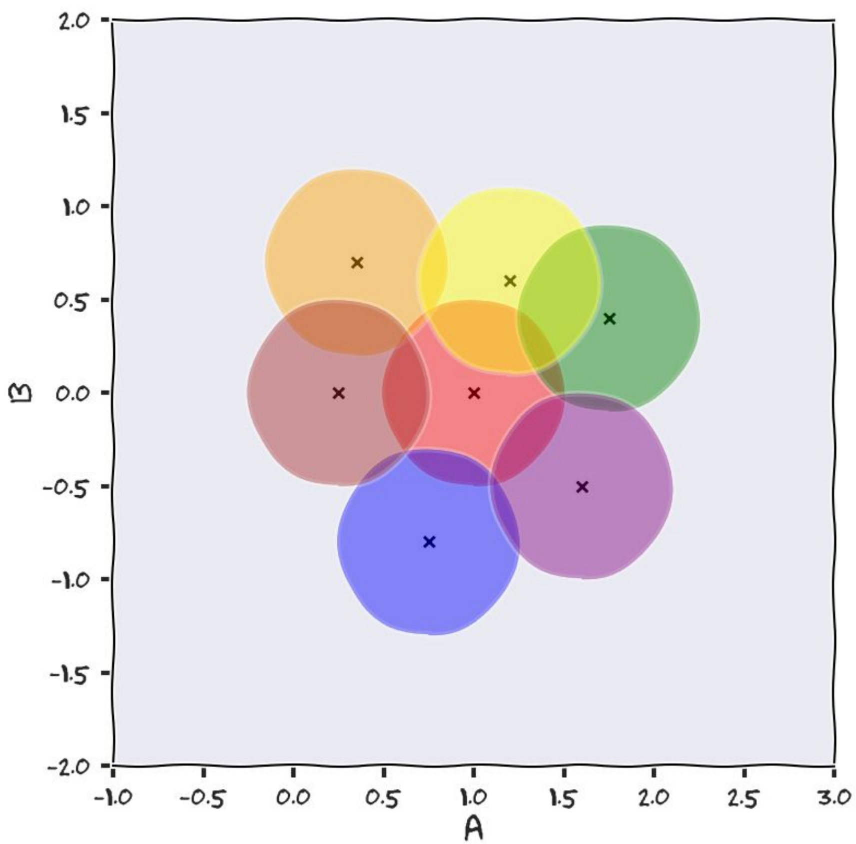
95% credible interval →

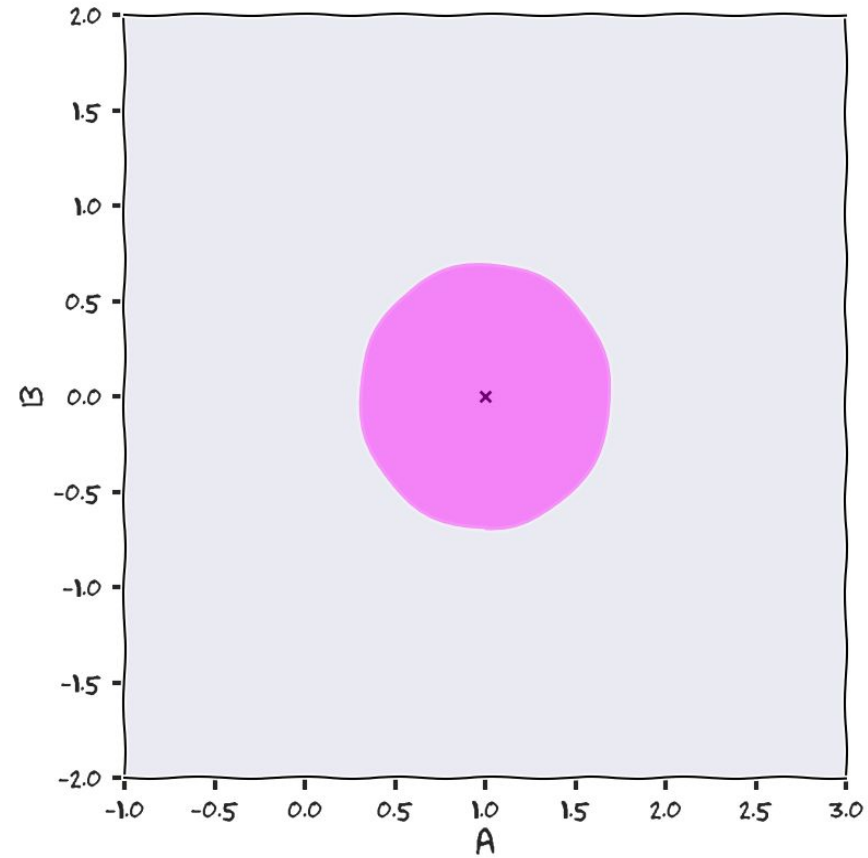
There is a 95% chance that the true value is within this interval

The Bayesian approach *is a process...*

$$P(a,b|\vec{x},\vec{y},\sigma) = \frac{\mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)}{\int_a \int_b \mathcal{L}(\vec{x},\vec{y}|a,b,\sigma)p(a)p(b)\,da\,db}$$

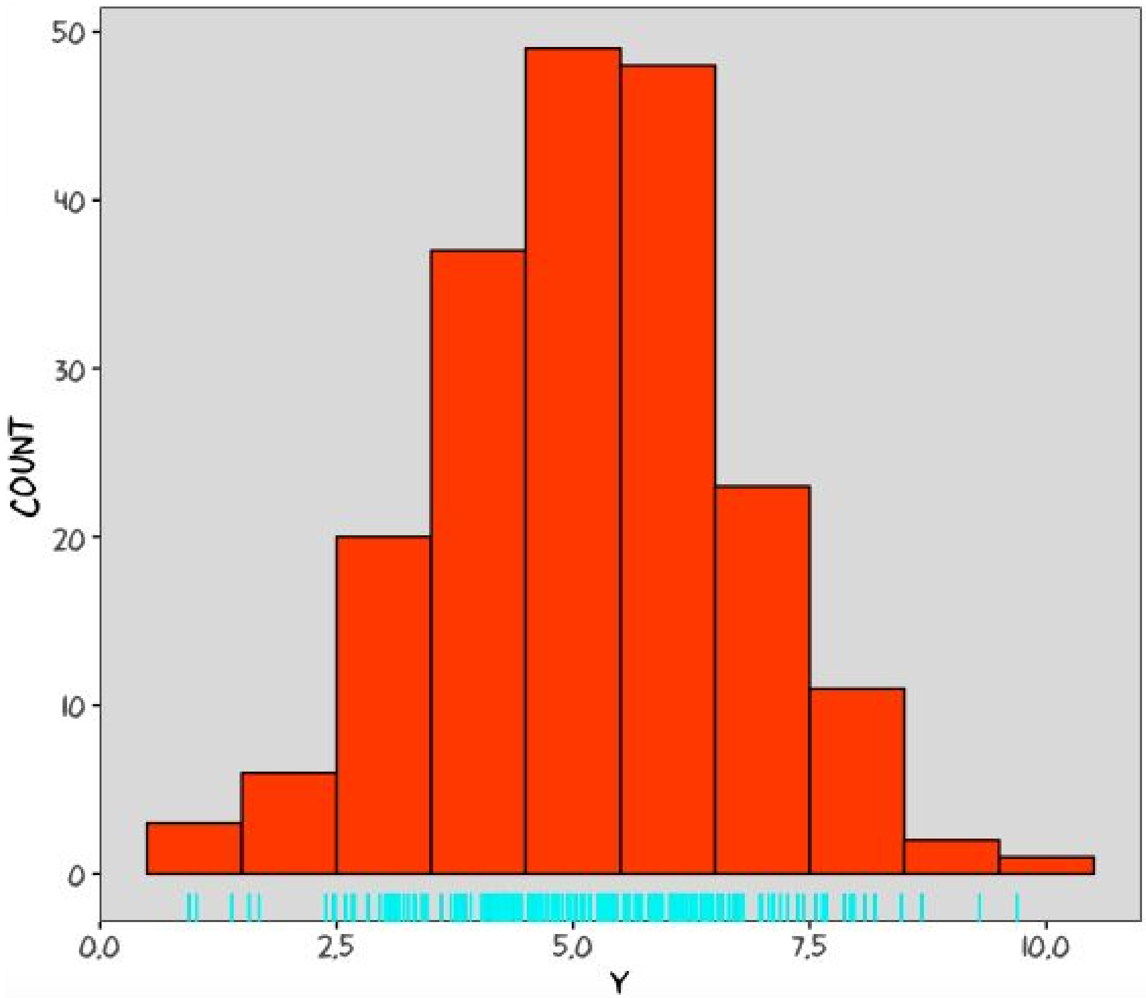
Confidence vs bayesian Intervals, practical examples





The mean of a Gaussian

Y ~ Normal(mu = 5, sd = 1.5)



Frequentist Approach

Unbiased estimator of the mean:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The **sampling distribution of** observed frequency of the x

$$f(\bar{x} \,||\, \mu) \propto \exp\left[\frac{-(\bar{x}-\mu)^2}{2\sigma_\mu^2}\right]$$

Standard error of the mean,

$$\sigma_\mu = \sigma_x/\sqrt{N}$$

The 95% confidence interval (two standard deviations), which roughly covers 95% of the area under the Gaussian curve.

$$CI_\mu = \left(\bar{x} - 2\sigma_\mu,\ \bar{x} + 2\sigma_\mu\right)$$

Bayesian Approach

Bayes' theorem:

$$P(\mu \mid D) = \frac{P(D \mid \mu)P(\mu)}{P(D)}$$

Prior:

$$P(\mu) \propto 1$$

Likelihood:

$$P(D \mid \mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[\frac{(\mu - x_i)^2}{2\sigma_x^2}\right]$$

Posterior:

$$P(\mu \mid D) \propto \exp\left[\frac{-(\mu - \bar{x})^2}{2\sigma_\mu^2}\right]$$

The 95% credible interval, the shortest interval that contains 95% of the probability.

$$CR_\mu = \left(\bar{x} - 2\sigma_\mu,\ \bar{x} + 2\sigma_\mu\right)$$

Example II: Truncated exponential

The Enterprise force field operates without failure until depletion of dilithium.

If under attack by Klingons, full depletion may occur after a time θ, after which failures may happen following the exponential failure law.

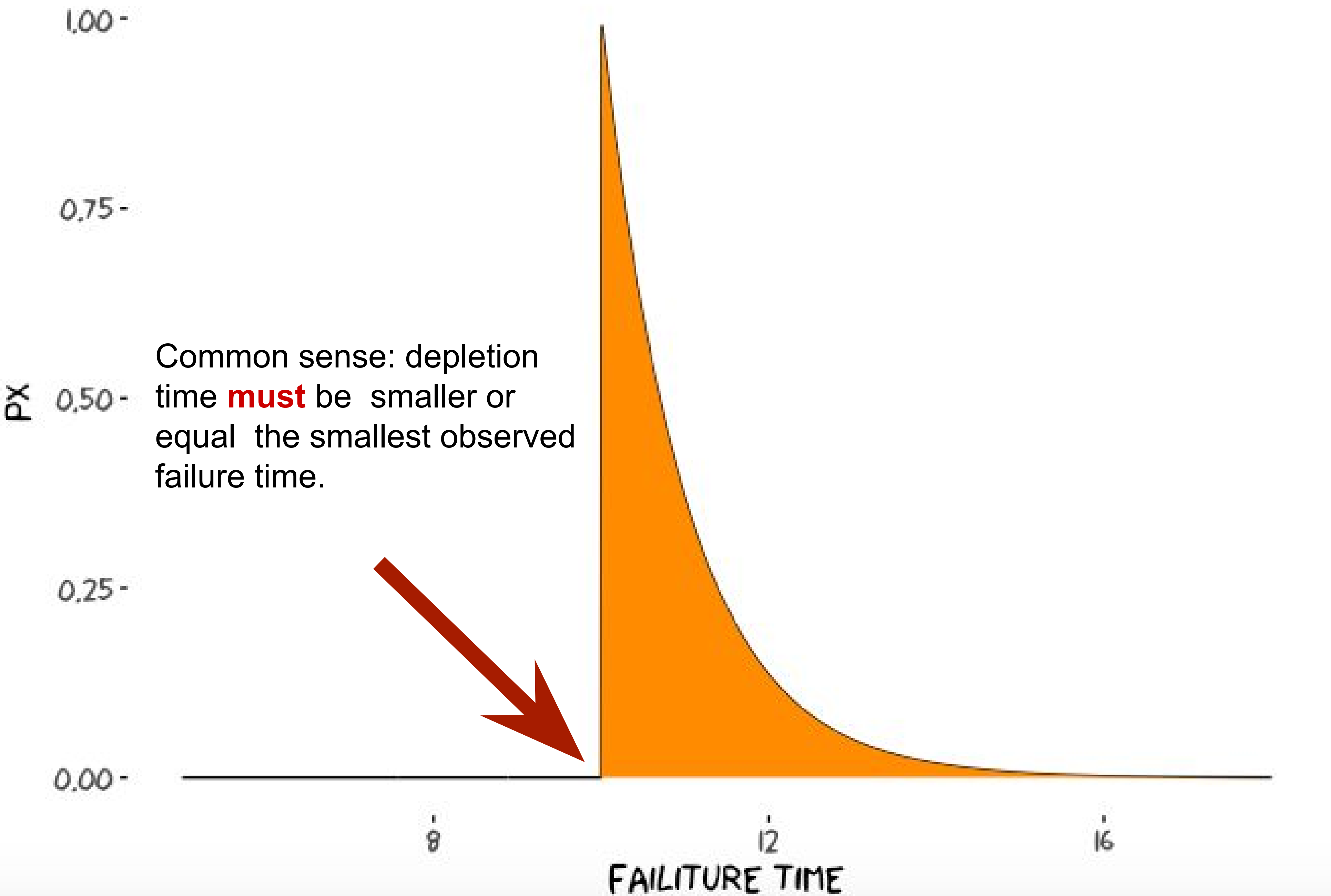
It is not feasible to observe the depletion rate directly, but one can observe the resulting failures after θ.

From previous data, one want to estimate the time θ of guaranteed safe operation… *

$$p(x \mid \theta) = \begin{cases} \exp(\theta - x) & , \quad x > \theta \\ 0 & , \quad x < \theta \end{cases}$$

Adapted from Jaynes, confidence intervals vs Bayesian intervals http://bayes.wustl.edu/etj/articles/confidence.pdf

Goal: estimate the depletion time from a series of observed failure times.



Frequentist Approach

Unbiased estimator of the mean:
$$\begin{aligned} E(x) &= \int_0^\infty x p(x) dx \\ &= \theta + 1 \end{aligned}$$

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} x_i - 1$$

The 95% confidence intervals
$$CI_{\text{large N}} = \left(\hat{\theta} - 2N^{-1/2},\ \hat{\theta} + 2N^{-1/2}\right)$$

For a sample of e.g. {14,18, 21} it will lead to **CI = {15.53,17.8}**

Bayesian approach: analytical

$$p(x \mid \theta) = \begin{cases} \exp(\theta - x) & , \quad x > \theta \\ 0 & , \quad x < \theta \end{cases}$$

Likelihood:

$$p(D \mid \theta) = \prod_{i=1}^{N} p(x \mid \theta)$$

Posterior:

$$p(\theta \mid D) \propto \begin{cases} N \exp[N(\theta - \min(D))] & , \quad \theta < \min(D) \\ 0 & , \quad \theta > \min(D) \end{cases}$$

Leads for the same sample BCI = {13,14}

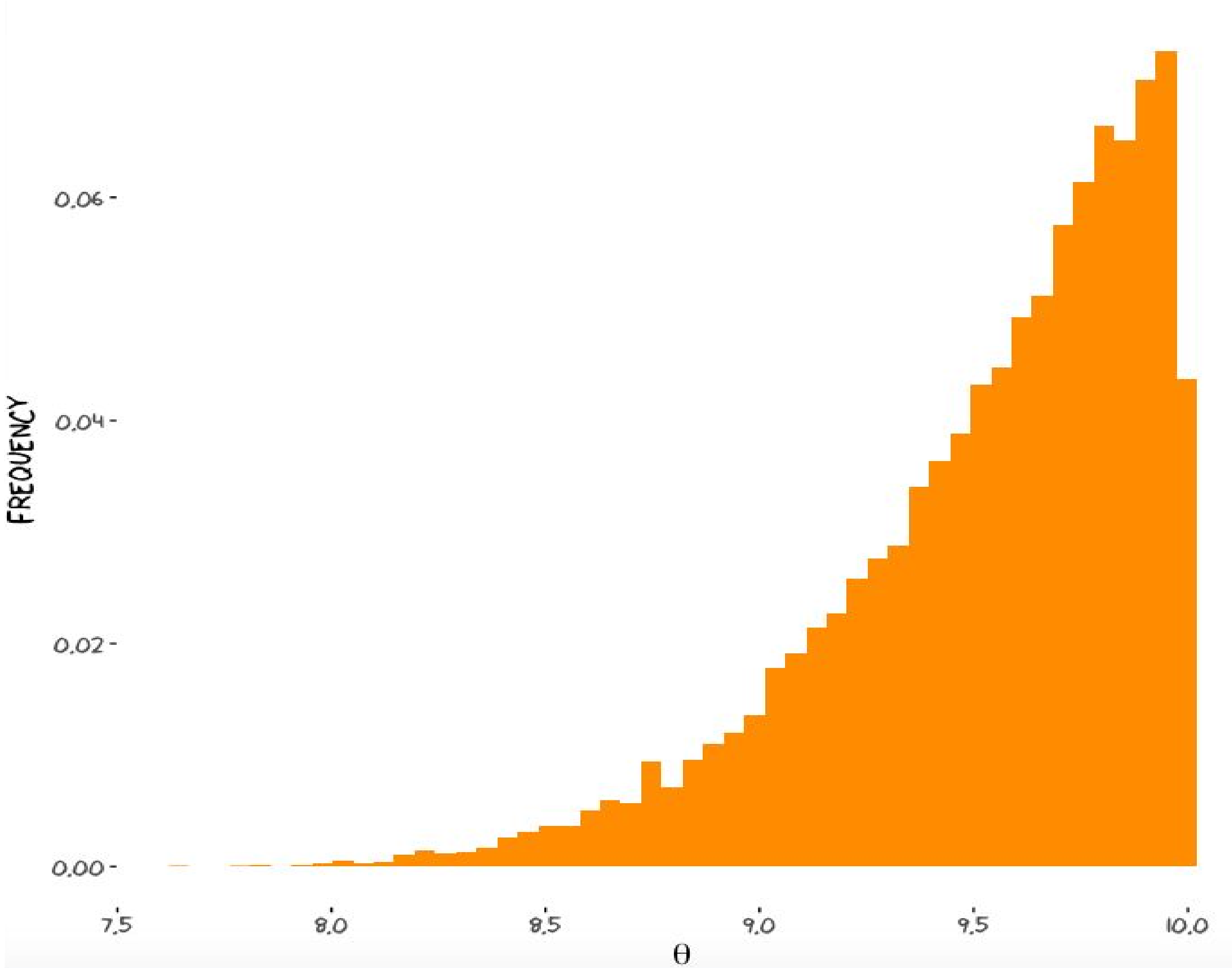
Interval:

$$\theta_2 = \min(D)$$

$$\theta_1 = \theta_2 + \frac{\log(1-f)}{N}$$

$$\int_{\theta_1}^{\theta_2} N \exp[N(\theta - \theta_2)] d\theta = f$$

Bayesian approach: MCMC



Next: Markov Chain Monte Carlo

MCMC is a numerical technique data allow us to sample from a target distribution