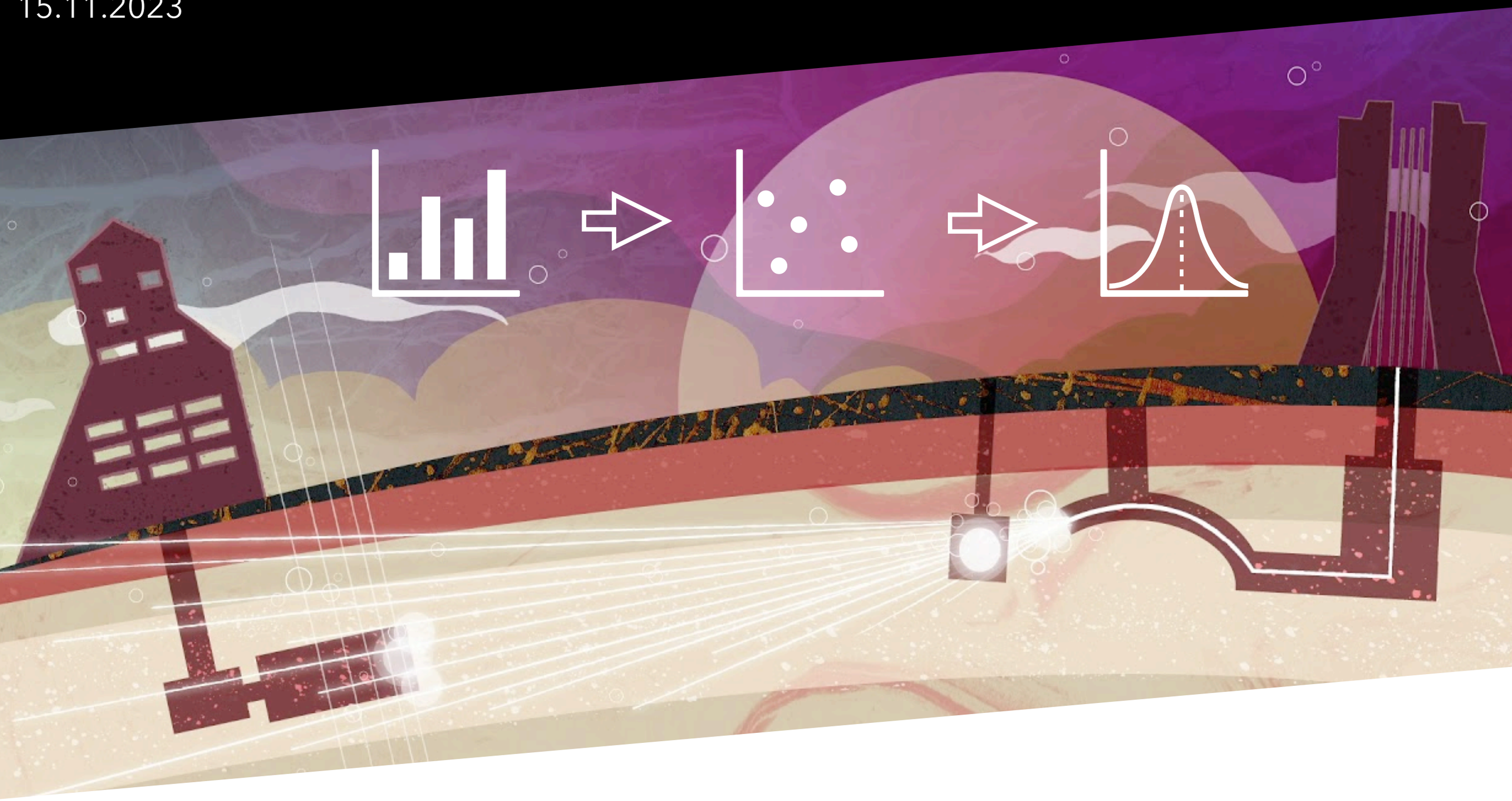
DUNE-France workshop #2

IJCLab, France

15.11.2023

Leïla Haegel

IP2I Lyon / CNRS



Bayesian inference for neutrino physics

Bayes theorem definition

$$\boxed{P(H|D)} = \frac{P(D|H) \quad P(H)}{P(D)}$$

Probability of the hypothesis H given the data D
= "posterior probability on H"

Bayes theorem definition

Probability of observing the data D according to hypothesis H
= "likelihood"

$$P(H|D) = \frac{P(D|H)\ P(H)}{P(D)}$$

Probability of the hypothesis H given the data D
= "posterior probability on H"

Bayes theorem definition

Probability of observing the data D according to hypothesis H
= "likelihood"

Probability of the hypothesis H
= "prior probability"

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Probability of the hypothesis H given the data D
= "posterior probability on H"

Bayes theorem definition

Probability of observing the data D according to hypothesis H = "likelihood"

Probability of the hypothesis H = "prior probability"

$$P(H|D) = \frac{P(D|H)\,P(H)}{P(D)}$$

Probability of the hypothesis H given the data D = "posterior probability on H"

Probability of the data D independently of the hypothesis H = " evidence"

Bayes theorem derivation

- **Derivation from conditional probabilities**
  - Probability to observe A and B:
    $P(A \cap B) \;=\; P(A)\; P(B\,|\,A)$

Bayes theorem derivation

- **Derivation from conditional probabilities**
  - Probability to observe A and B:

$$P(A \cap B) = P(A)\,P(B\,|\,A)$$

Probability to observe A

Probability to observe B if A is observed

Bayes theorem derivation

- **Derivation from conditional probabilities**
  - Probability to observe A and B:

$$P(A \cap B) \;=\; P(A)\; P(B\,|\,A) \;=\; \underbrace{P(B)}_{\text{Probability to observe } B}\;\underbrace{P(A\,|\,B)}_{\text{Probability to observe } A \text{ if } B \text{ is observed}}$$

Bayes theorem derivation

- **Derivation from conditional probabilities**
  - Probability to observe A and B:

$$P(A \cap B) \;=\; P(A)\; P(B\,|\,A) \;=\; P(B)\; P(A\,|\,B)$$

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

Bayes theorem derivation

- **Derivation from conditional probabilities**
  - Probability to observe A and B:

$$P(A \cap B) \;=\; P(A)\; P(B\,|\,A) \;=\; P(B)\; P(A\,|\,B)$$

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

- **Physical interpretation**
  - In physics, we often have A as the hypothesis H, and B as the data D:

$$P(H\,|\,D) = \frac{P(D\,|\,H)\;\; P(H)}{P(D)}$$

  - The posterior probability is the prior probability when it is weighted by the likelihood of observing the data, normalised by the probability of observing any data

Bayesian inference

- **Bayesian inference is the process of updating the probability on a statement**
  - Evaluation of the posterior probability on H according the data D
  - Bayes theorem actualises the prior probability according to the evidence
  - Also referred to as "updating belief on H "

Bayesian inference

- **Bayesian inference is the process of updating the probability on a statement**
  - Evaluation of the posterior probability on H according the data D
  - Bayes theorem actualises the prior probability according to the evidence
  - Also referred to as "updating belief on H "

- **Example from neutrino physics**
  - We do not know δ_{CP} → our prior probability on the parameter is flat
  - We compute the likelihood of observing the ν_e and $\bar{\nu}_e$ spectra according to several δ_{CP} hypothetical values
  - We measure the posterior probability for each hypothesis
  - We estimate which value of δ_{CP} corresponds to the highest posterior probability

Posterior probability sampling

- **Simple 1 parameter case:**
  - Chose a grid, and evaluate the posterior probability at each point e.g. $P(\delta_{CP,i}|D)$ for $i \in [-\pi, \pi]$ with steps of $\pi/100$
  - Easy to extend to 4 oscillation parameters

Posterior probability sampling

- **Simple 1 parameter case:**
  - Chose a grid, and evaluate the posterior probability at each point e.g. $P(\delta_{CP,i}|D)$ for $i \in [-\pi, \pi]$ with steps of $\pi/100$
  - Easy to extend to 4 oscillation parameters

- **But: systematics!**
  - At each step, one needs to evaluate the posterior probability varying the systematical parameters as well (throws)
  - Neutrino physics → $\mathcal{O}(100)$ systematical parameters (flux, interaction model, detector response…)
  - The real posterior probability is: $P(\delta_{CP,i}, \vec{\varsigma_j}|D)$ where $\vec{\varsigma_j}$ is the vector of systematical uncertainties
  - Grid searches can become computationally very expensive (many points to evaluate)

Markov Chain Monte-Carlo

- **Markov Chain Monte-Carlo (MCMC)**
  - Alternate procedure to grid to sample the space of oscillation parameters $\overrightarrow{\vartheta}$ and systematics parameters $\vec{\varsigma}$
  - Semi-random walk in the parameter space
  - Stochastic model: randomness of throws
  - Sequential process: the state of a throw only depends on the throw before

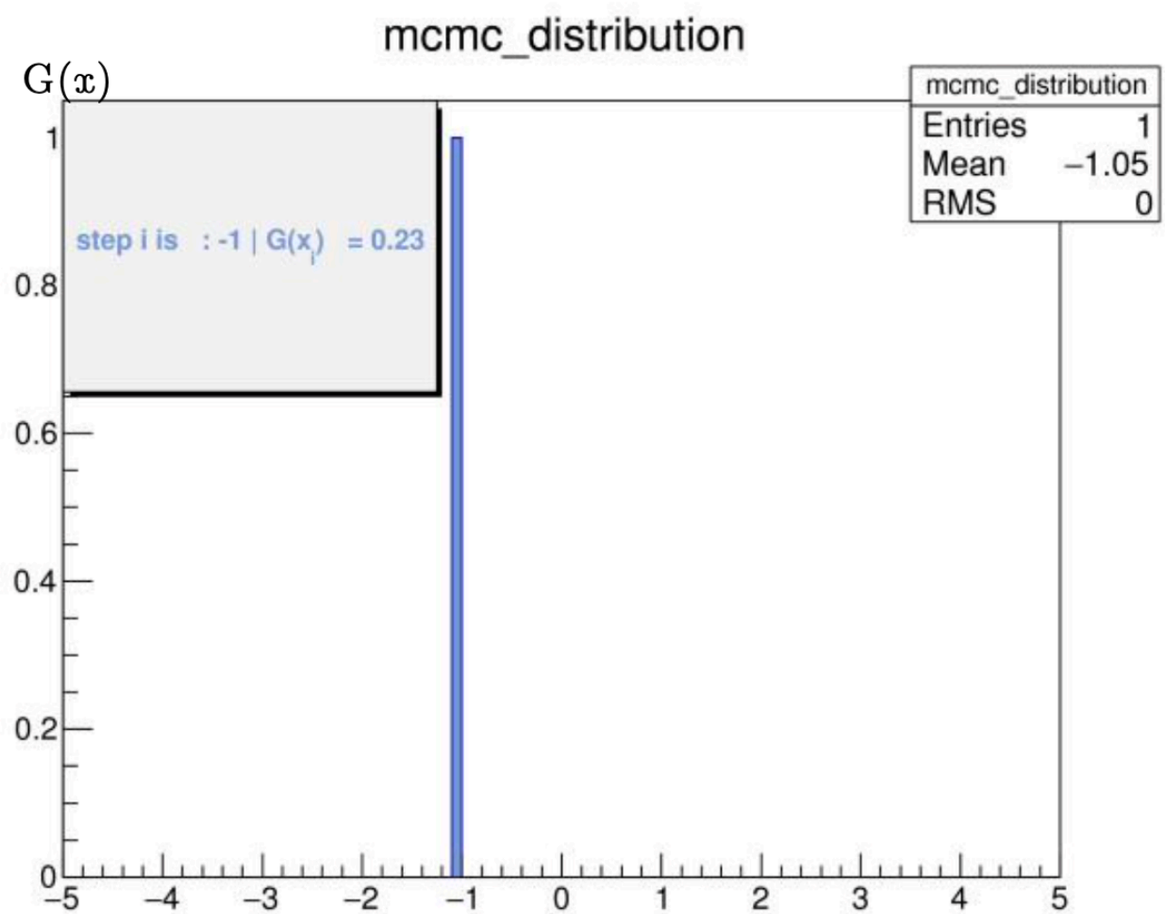
Markov Chain Monte-Carlo

- **Markov Chain Monte-Carlo (MCMC)**
  - Alternate procedure to grid to sample the space of oscillation parameters $\overrightarrow{\vartheta}$ and systematics parameters $\vec{\varsigma}$
  - Semi-random walk in the parameter space
  - Stochastic model: randomness of throws
  - Sequential process: the state of a throw only depends on the throw before

- **Metropolis-Hastings algorithm**
  - Most generic implementation of Markov Chain Monte-Carlo (MCMC)
  - The semi-random walks is *proportional* to the target distribution
  - The collection of steps are *samples* from the posterior distribution
  - Ensure 2 specific properties of the MCMC:
    Aperiodicity: do not oscillate between same values
    Ergodicity: can converge to the stationary distribution

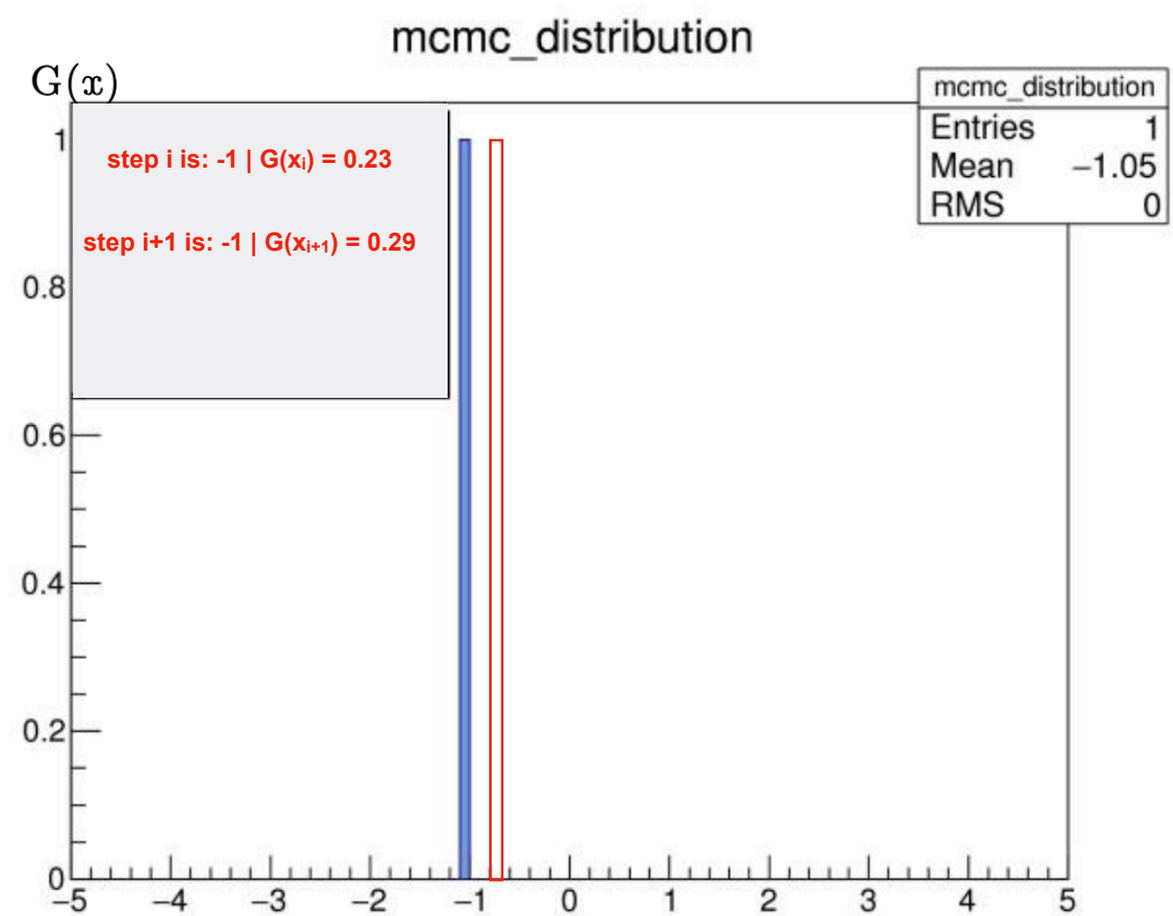
Metropolis-Hastings MCMC

- **Demonstration for target distribution = Gaussian distribution** $G(x)$
  - First step $i = 1$: start with a random choice of hypothetical value x_i



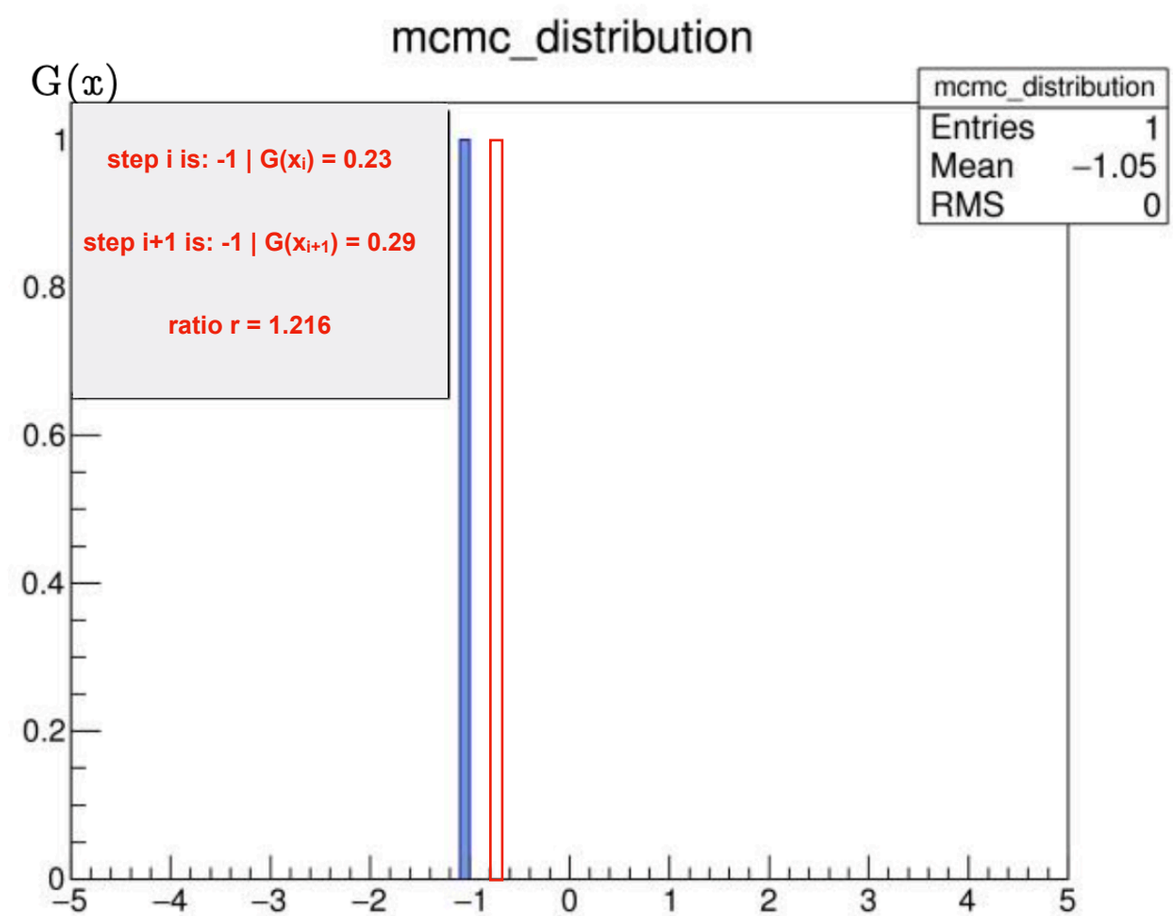
Metropolis-Hastings MCMC

- **Demonstration for target distribution = Gaussian distribution** $G(x)$
  - First step $i = 1$: start with a random choice of hypothetical value x_i
  - Propose a new step $i + 1$: using the *jump function* $J(x_i + 1 \,|\, x_i)$



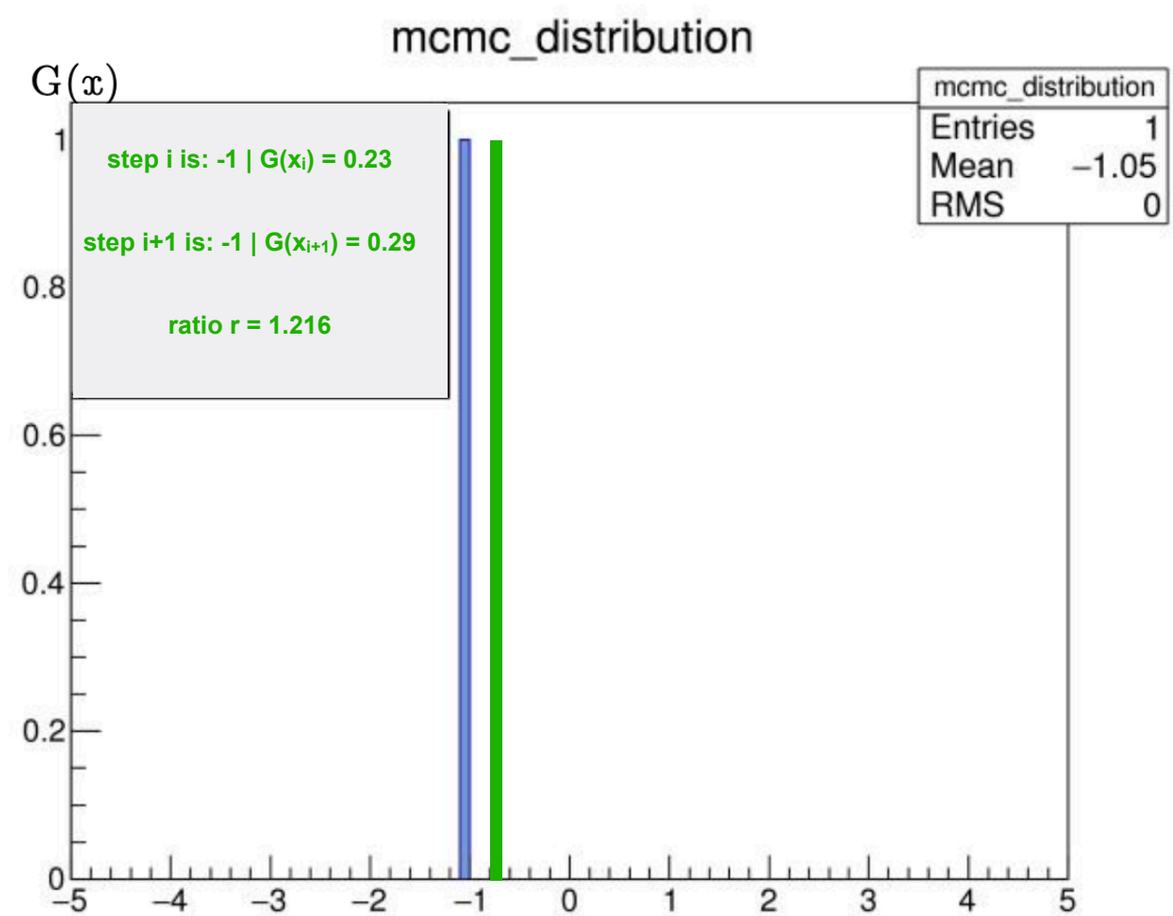
Metropolis-Hastings MCMC

- **Demonstration for target distribution = Gaussian distribution** $G(x)$
  - First step $i = 1$: start with a random choice of hypothetical value x_i
  - Propose a new step $i + 1$: using the *jump function* $J(x_i + 1 \,|\, x_i)$
  - Compute the Metropolis-Hastings ratio r: $r = \dfrac{G(x_{i+1})\ J(x_i \,|\, x_{i+1})}{G(x_i)\ J(x_{i+1}) \,|\, x_i}$



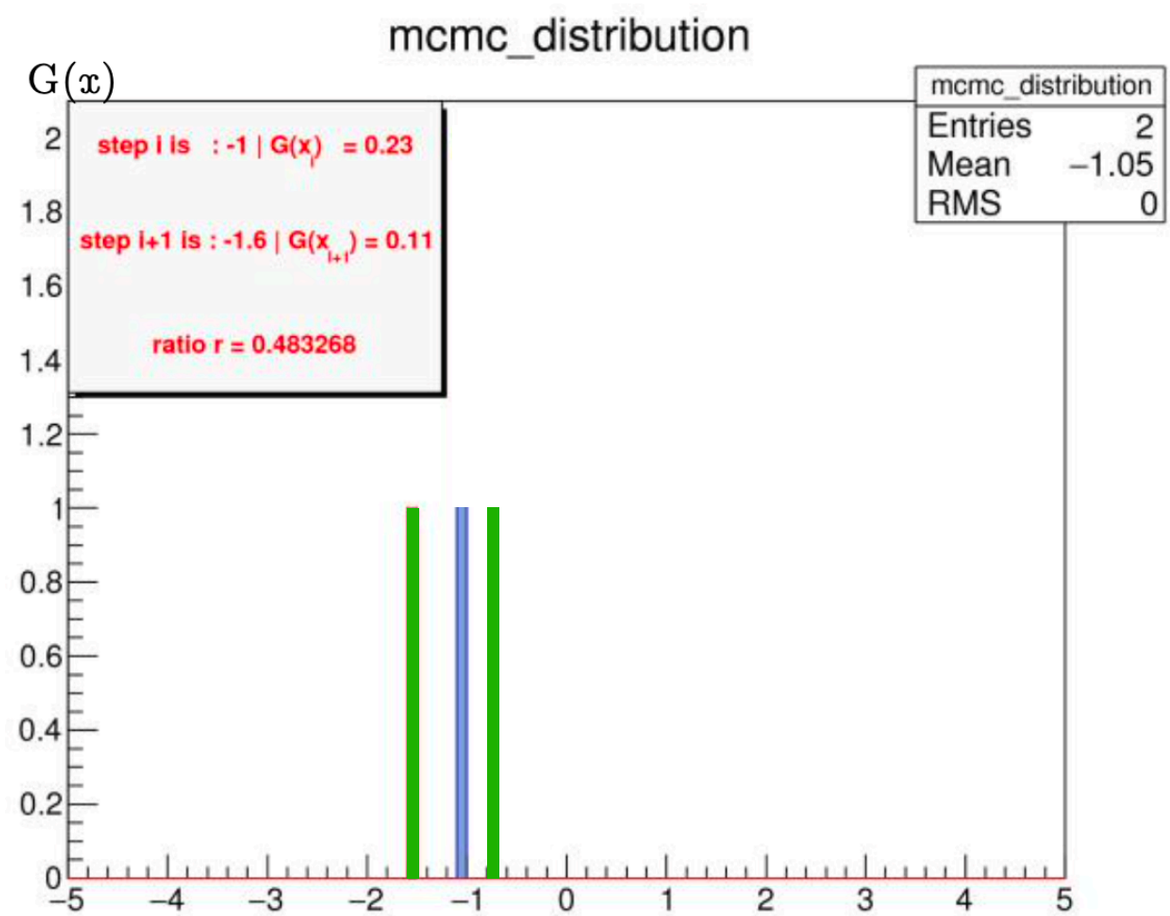
Metropolis-Hastings MCMC

- **Demonstration for target distribution = Gaussian distribution** $G(x)$
  - First step $i = 1$: start with a random choice of hypothetical value x_i
  - Propose a new step $i + 1$: using the *jump function* $J(x_i + 1 \,|\, x_i)$
  - Compute the Metropolis-Hastings ratio r: $r = \dfrac{G(x_{i+1})\ J(x_i \,|\, x_{i+1})}{G(x_i)\ J(x_{i+1}) \,|\, x_i}$
  - Apply the *acceptance function* $A(x_{i+1}, x_i)$:
    - $r \geq 1 \rightarrow$ accept step $i + 1$



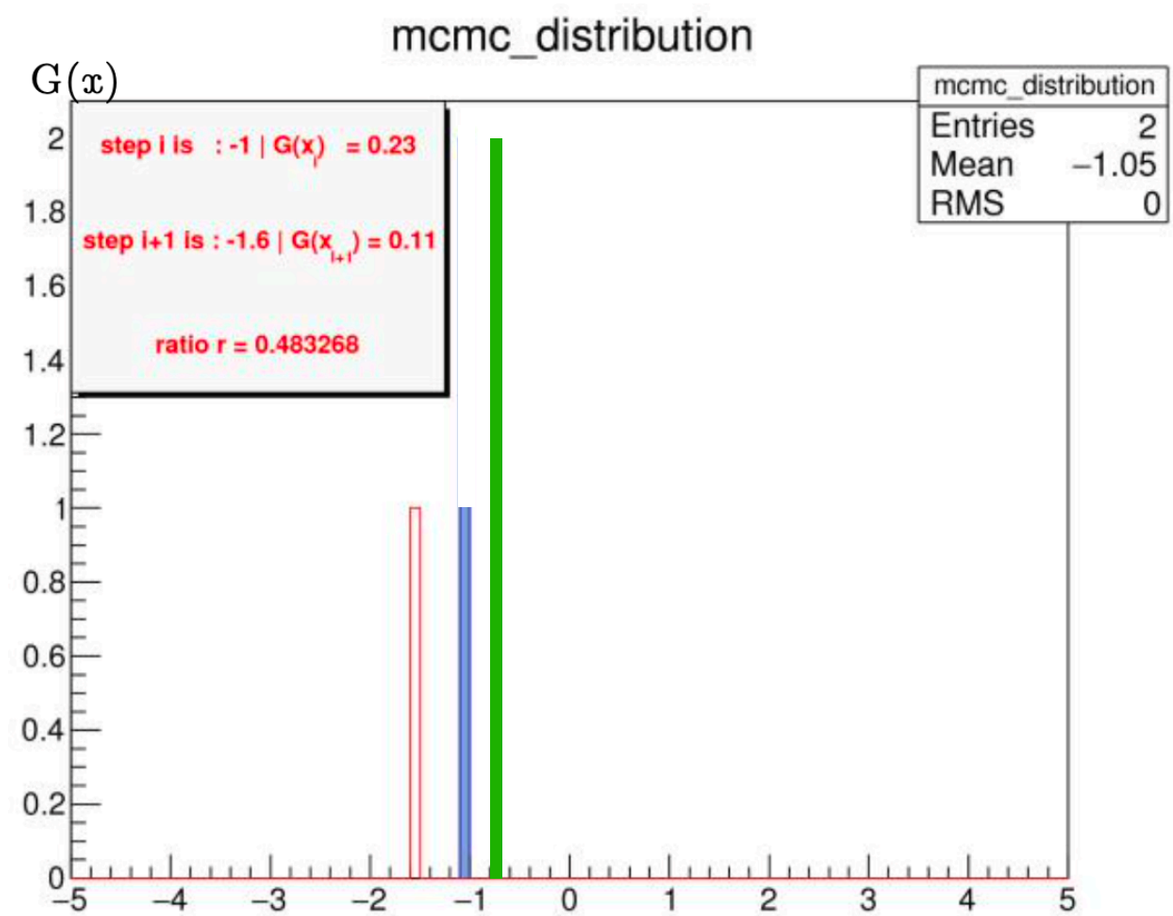
Metropolis-Hastings MCMC

- **Demonstration for target distribution = Gaussian distribution** $G(x)$
  - First step $i = 1$: start with a random choice of hypothetical value x_i
  - Propose a new step $i + 1$: using the *jump function* $J(x_i + 1 \,|\, x_i)$
  - Compute the Metropolis-Hastings ratio r: $r = \dfrac{G(x_{i+1})\ J(x_i \,|\, x_{i+1})}{G(x_i)\ J(x_{i+1}) \,|\, x_i}$
  - Apply the *acceptance function* $A(x_{i+1}, x_i)$:
    - $r \geq 1 \rightarrow$ accept step $i + 1$
    - $r < 1 \rightarrow$ throw a number $u \in U(0,1)$
      $r \geq u \rightarrow$ accept step $i + 1$



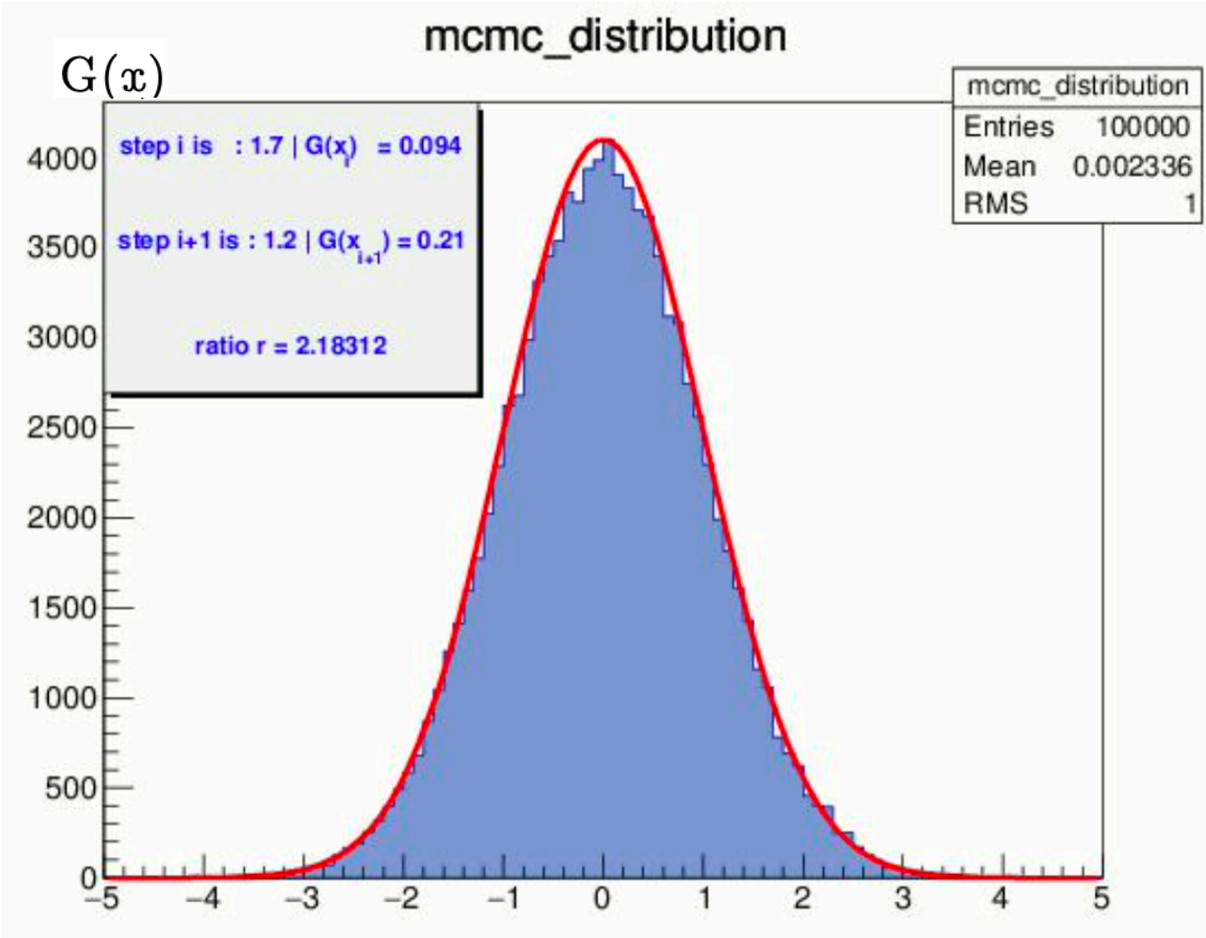
Metropolis-Hastings MCMC

- **Demonstration for target distribution = Gaussian distribution** $G(x)$
  - First step $i = 1$: start with a random choice of hypothetical value x_i
  - Propose a new step $i + 1$: using the *jump function* $J(x_i + 1 \,|\, x_i)$
  - Compute the Metropolis-Hastings ratio r: $r = \dfrac{G(x_{i+1})\ J(x_i \,|\, x_{i+1})}{G(x_i)\ J(x_{i+1}) \,|\, x_i}$
  - Apply the *acceptance function* $A(x_{i+1}, x_i)$:
    - $r \geq 1 \rightarrow$ accept step $i + 1$
    - $r < 1 \rightarrow$ throw a number $u \in U(0,1)$
      - $r \geq u \rightarrow$ accept step $i + 1$
      - $r < u \rightarrow$ reject step $i + 1$
        count again step i



Metropolis-Hastings MCMC

- **Demonstration for target distribution = Gaussian distribution** $G(x)$
  - First step $i = 1$: start with a random choice of hypothetical value x_i
  - Propose a new step $i + 1$: using the *jump function* $J(x_i + 1 \,|\, x_i)$
  - Compute the Metropolis-Hastings ratio r: $r = \dfrac{G(x_{i+1})\ J(x_i \,|\, x_{i+1})}{G(x_i)\ J(x_{i+1}) \,|\, x_i}$
  - Apply the *acceptance function* $A(x_{i+1}, x_i)$:
    - $r < 1 \rightarrow$ throw a number $u \in U(0,1)$
      $r \geq u \rightarrow$ accept step $i + 1$
      $r < u \rightarrow$ reject step $i + 1$
      count again step i
    - $r \geq 1 \rightarrow$ accept step $i + 1$
  - Iterate process until obtaining enough step to analyse the distribution



Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\, r)$

$$r \geq 1 \rightarrow A(x_{i+1}, x_i) \;=\; 1$$

$$r < 1 \rightarrow A(x_{i+1}, x_i) \;=\; r$$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\, r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\, r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$
  - We can derive the detailed balance equation:

    $G(x_i)\;\; T(x_{i+1}\,|\,x_i) \;= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; A(x_{i+1}, x_i)$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\; r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$
  - We can derive the detailed balance equation:

    $$G(x_i)\;\; T(x_{i+1}\,|\,x_i) \;= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; A(x_{i+1}, x_i)$$
    $$= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min(1,\; r)$$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\; r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$
  - We can derive the detailed balance equation:

$$
\begin{aligned}
G(x_i)\;\; T(x_{i+1}\,|\,x_i) \;&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; A(x_{i+1}, x_i) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min(1,\; r) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min\left(1,\; \frac{G(x_{i+1})\;\; J(x_i\,|\,x_{i+1})}{G(x_i)\;\;\; J(x_{i+1}\,|\,x_i)}\right)
\end{aligned}
$$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\; r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$
  - We can derive the detailed balance equation:

$$
\begin{aligned}
G(x_i)\;\; T(x_{i+1}\,|\,x_i) \;&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; A(x_{i+1}, x_i) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min(1,\; r) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min\left(1,\; \frac{G(x_{i+1})\;\; J(x_i\,|\,x_{i+1})}{G(x_i)\;\;\; J(x_{i+1}\,|\,x_i)}\right) \\
&= min\left(G(x_i)\; J(x_{i+1}\,|\,x_i),\; G(x_{i+1})\; J(x_i\,|\,x_{i+1})\right)
\end{aligned}
$$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\; r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$
  - We can derive the detailed balance equation:

$$
\begin{aligned}
G(x_i)\;\; T(x_{i+1}\,|\,x_i) \;&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; A(x_{i+1}, x_i) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min(1,\; r) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min\left(1,\; \frac{G(x_{i+1})\;\; J(x_i\,|\,x_{i+1})}{G(x_i)\;\;\; J(x_{i+1}\,|\,x_i)}\right) \\
&= min(\,G(x_i)\; J(x_{i+1}\,|\,x_i),\; G(x_{i+1})\; J(x_i\,|\,x_{i+1})\,) \\
&= G(x_{i+1})\; J(x_i\,|\,x_{i+1})\, A(x_i, x_{i+1})
\end{aligned}
$$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\; r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$
  - We can derive the detailed balance equation:

$$
\begin{aligned}
G(x_i)\;\; T(x_{i+1}\,|\,x_i) \;&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; A(x_{i+1}, x_i) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min(1,\; r) \\
&= G(x_i)\;\; J(x_{i+1}\,|\,x_i)\;\; min\left(1,\; \frac{G(x_{i+1})\;\; J(x_i\,|\,x_{i+1})}{G(x_i)\;\;\; J(x_{i+1}\,|\,x_i)}\right) \\
&= min(\,G(x_i)\; J(x_{i+1}\,|\,x_i),\; G(x_{i+1})\; J(x_i\,|\,x_{i+1})\,) \\
&= G(x_{i+1})\; J(x_i\,|\,x_{i+1})\, A(x_i, x_{i+1}) \\
&= G(x_{i+1})\; T(x_i\,|\,x_{i+1})
\end{aligned}
$$

Detailed balance equation

- **The *detailed balance equation* ensure that the samples follow the target distribution**
  - The *acceptance function* is: $A(x_{i+1}, x_i) \;=\; min(1,\, r)$
  - Defining the probability to transition to the step x_{i+1}, i.e. the *transition probability*:

    $T(x_{i+1}\,|\,x_i) = J(x_{i+1}\,|\,x_i)\; A(x_{i+1}, x_i)$
  - We can derive the detailed balance equation:

    $G(x_i)\;\; T(x_{i+1}\,|\,x_i) \;=\; G(x_{i+1})\; T(x_i\,|\,x_{i+1})$
  - Interpretation: if we propose a step with $G(x_{i+1}) \;> G(x_i)$

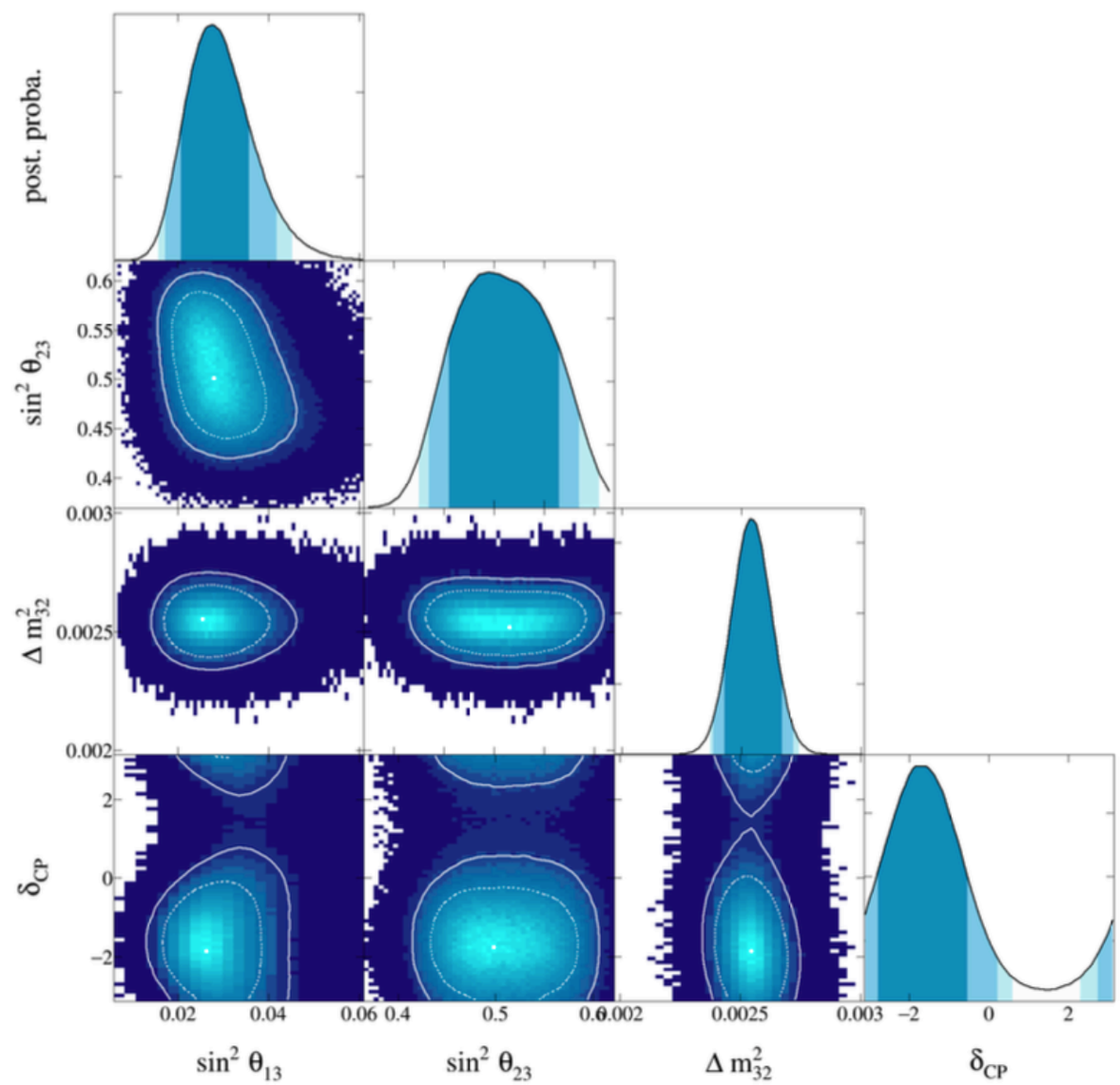
    The acceptance function is: $A(x_{i+1}, x_i) \;=\; 1$

    The transition probability is: $T(x_i\,|\,x_{i+1}) = \dfrac{G(x_{i+1})}{G(x_i)}$

    → The probability to jump back on the previous step is proportional to the ratio of $G(x)$ value

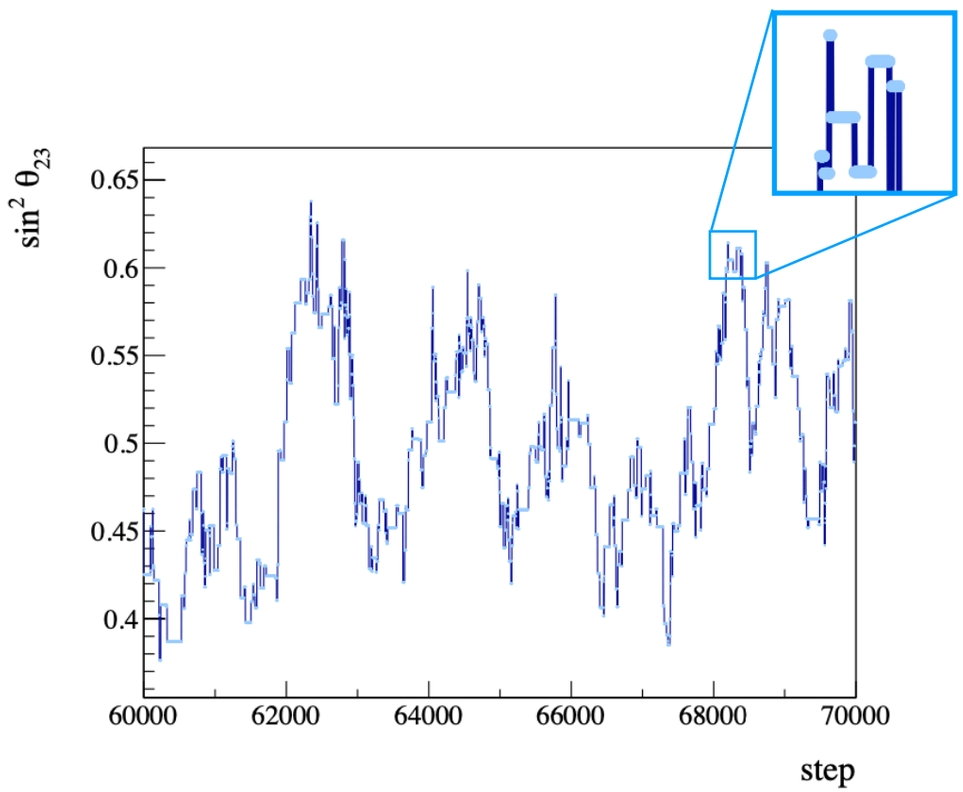
Application for neutrino physics

- **In the case of neutrino physics**
  - The target distribution is the posterior probability on the oscillation parameters $\vec{\vartheta}$ and systematics parameters $\vec{\zeta}$
  - All parameters are treated the same, they just have different prior probabilities
  - All parameters are inferred at the same time
    → joint analysis of near and far detector values

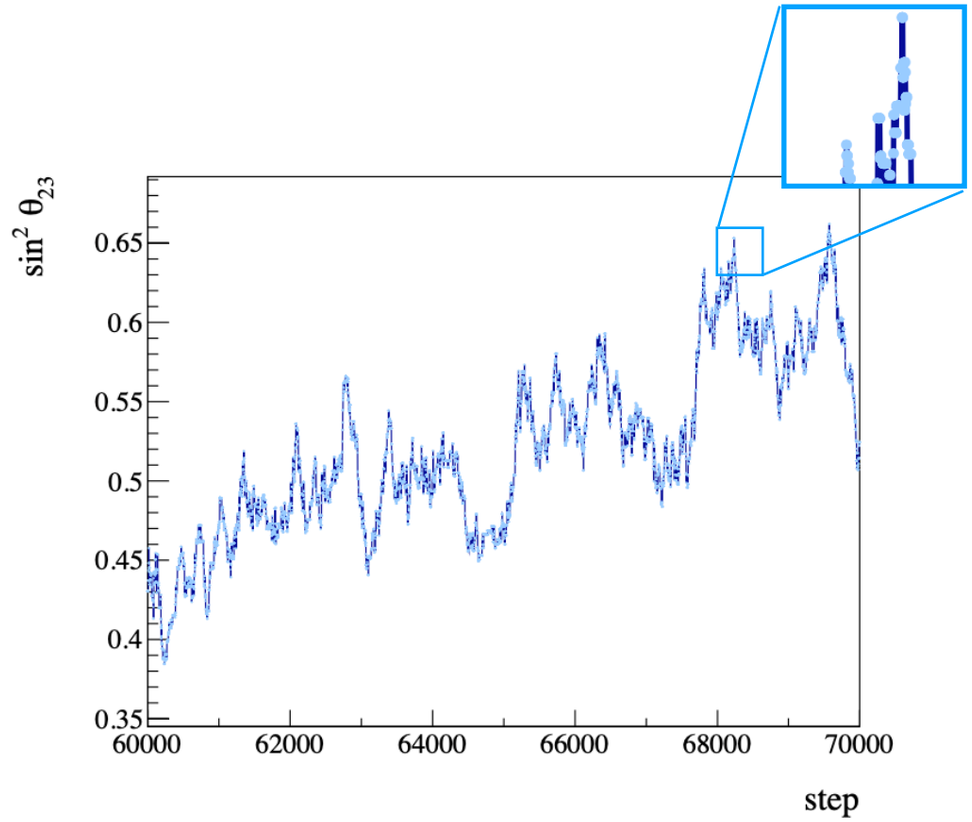


Step size

- **Jump function parameter**
  - The jump function can be symmetrical → Metropolis algorithm
    or asymmetrical → Hastings addition
  - The jump function has a width parameter:
    → this is referred to as the *step size*
    → its value is heuristic, although literature exist about its optimisation
    → strongly impacts the convergence rate of the chain



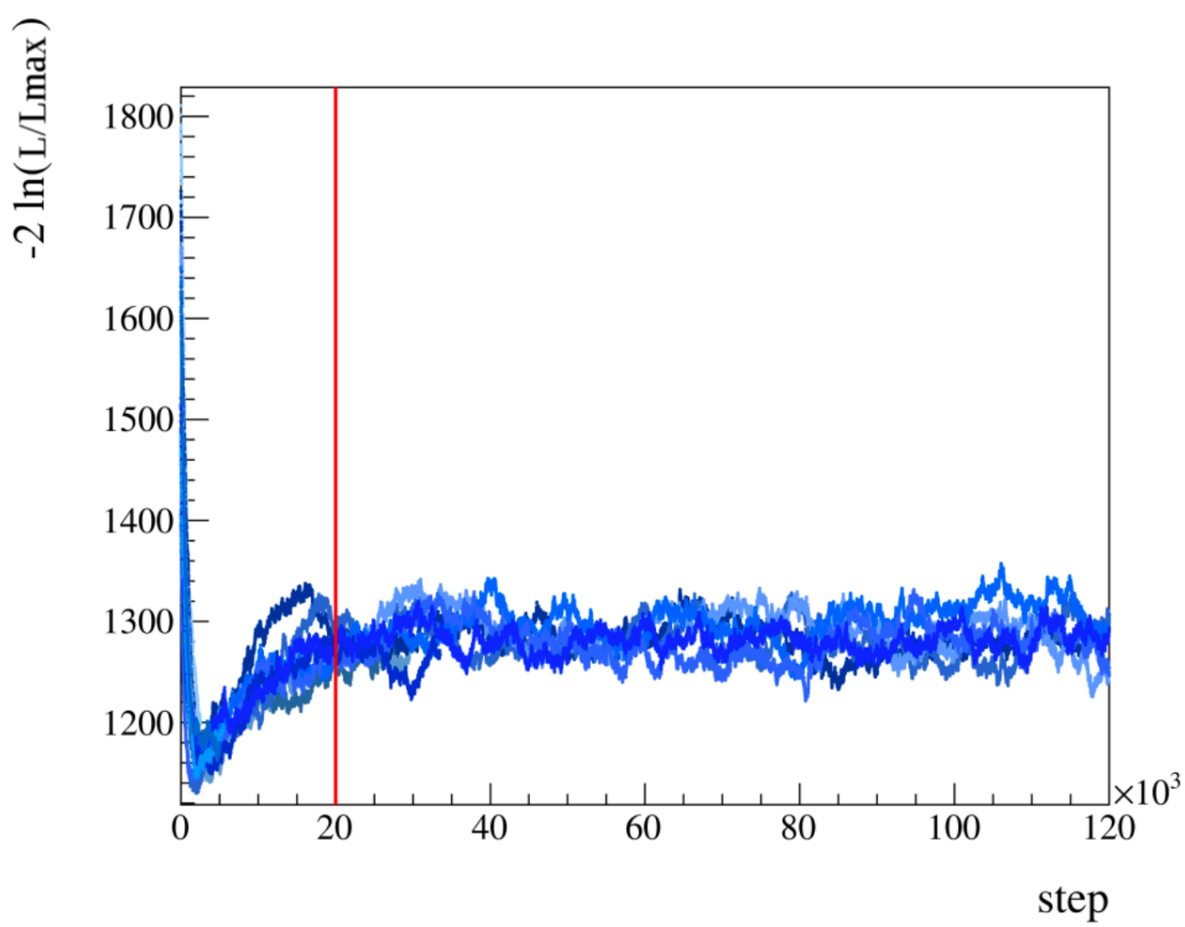
(c) $\sin^2\theta_{23}$, large step scale



(a) $\sin^2\theta_{23}$, correct step scale

Burn-in

- **The Markov chain takes time to reach equilibrium**
  - The chain can start far from the target distribution
    → creates a bias towards initial values
  - The first values must be discarded: “burn-in”
  - The burn-in size can be determined from the trace of the chain



Autocorrelation

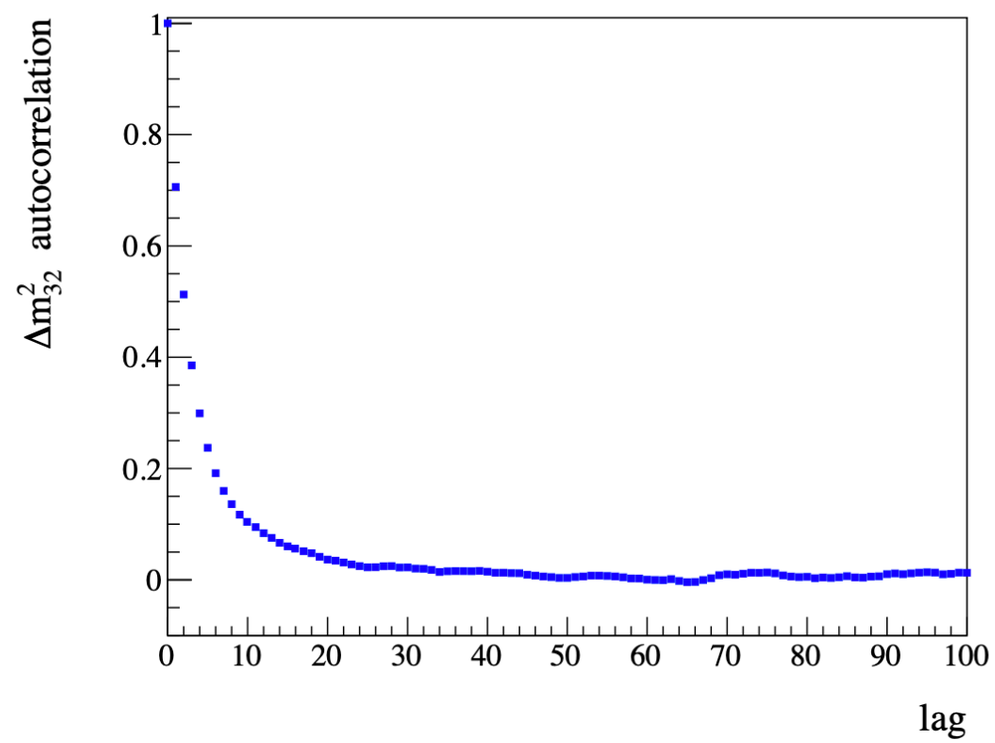
- **The steps are correlated between them**
  - Independent samples can be selected by subsampling the chain
  - Value of subsampling order can be determined from the autocorrelation function

$$\mathscr{A}(k) = \frac{\varrho(k)}{\varrho(0)}$$

where:

$$\varrho(k) = \mathbb{E}(x_i - \bar{x})\ \mathbb{E}(x_{i+k} - \bar{x})$$
$$= \frac{1}{N-k} \sum_{i}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})$$

$\mathbb{E}$ = expectation value



(b) Δm^2_{32}

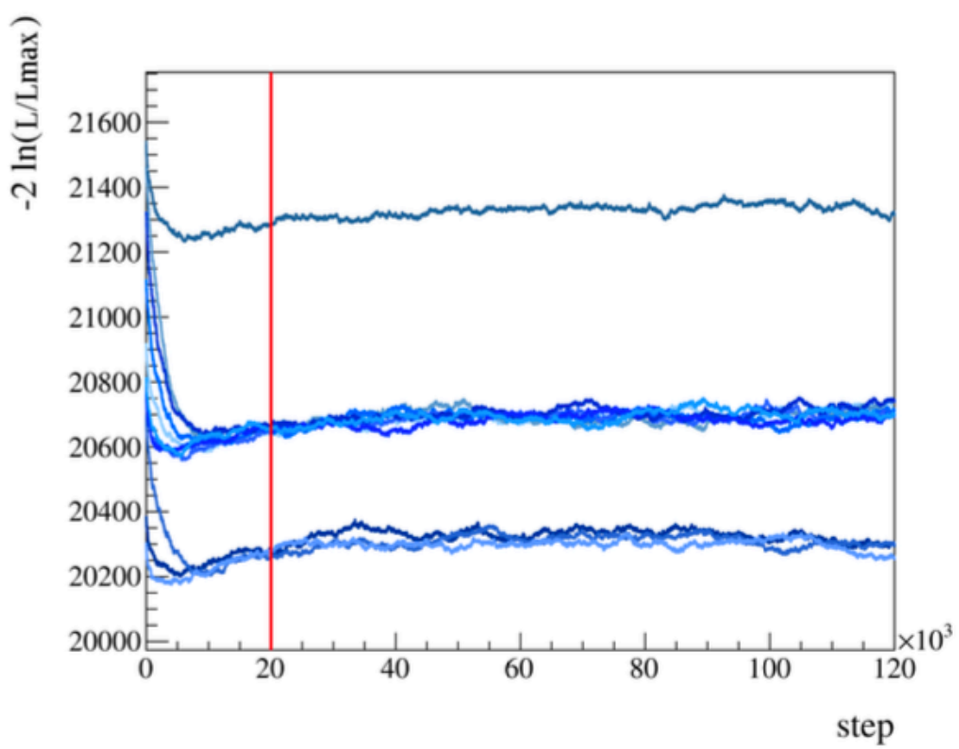
Convergence tests

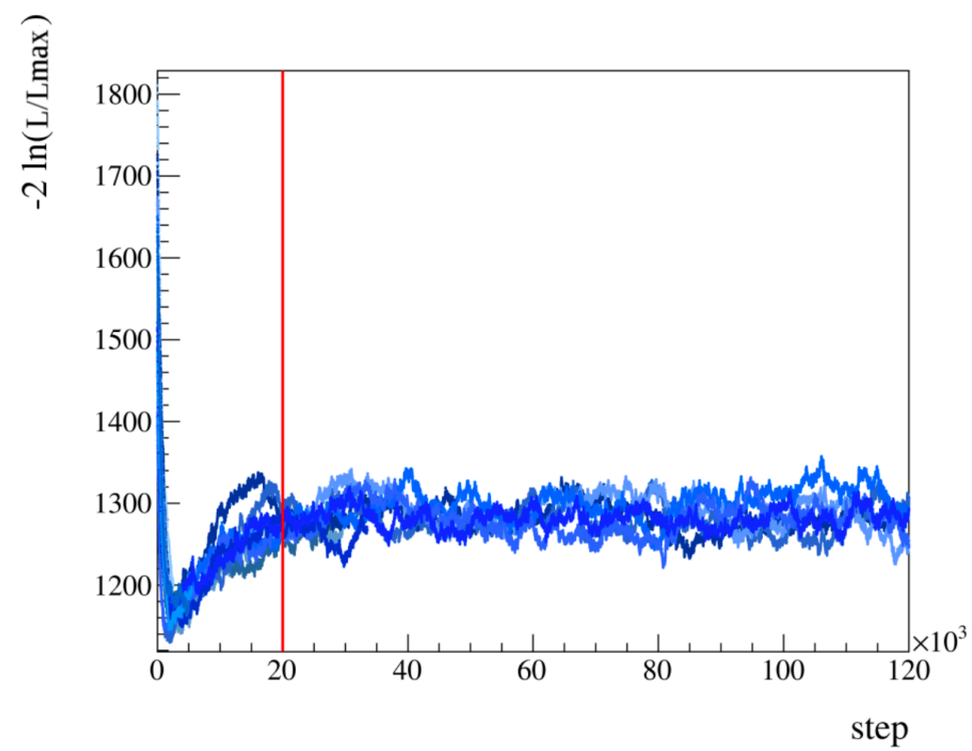
- **Ergodicity**
  - Are the chains spanning the entire value of parameter space?
  - Test: comparison of independent chains

Chains not properly tuned
Not ergodic

increase step size →

Chains properly tuned
Ergodic





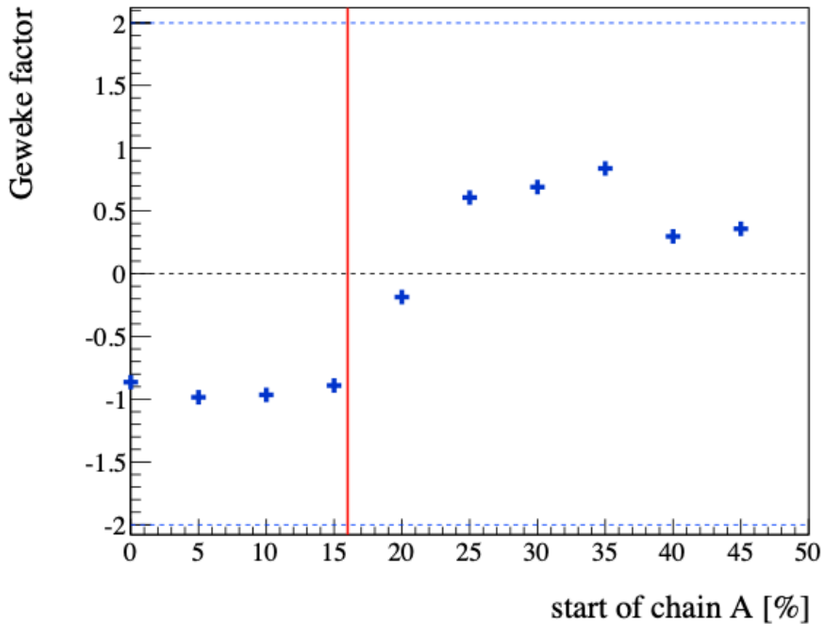
Convergence tests

- **Ergodicity**
  - Are the chains spanning the entire value of parameter space?
  - Test: comparison of independent chains

- **Geweke diagnostic**
  - Compare the beginning and the end of a Markov chain
  - Select 5% of the chain from its beginning and increment of 5%
    e.g. [0-5%], [5-10%], …, [45-50%]
    and compare with remaining 50% of the chain: [50-100%]
  - Useful to determine burn-in value and spot issues

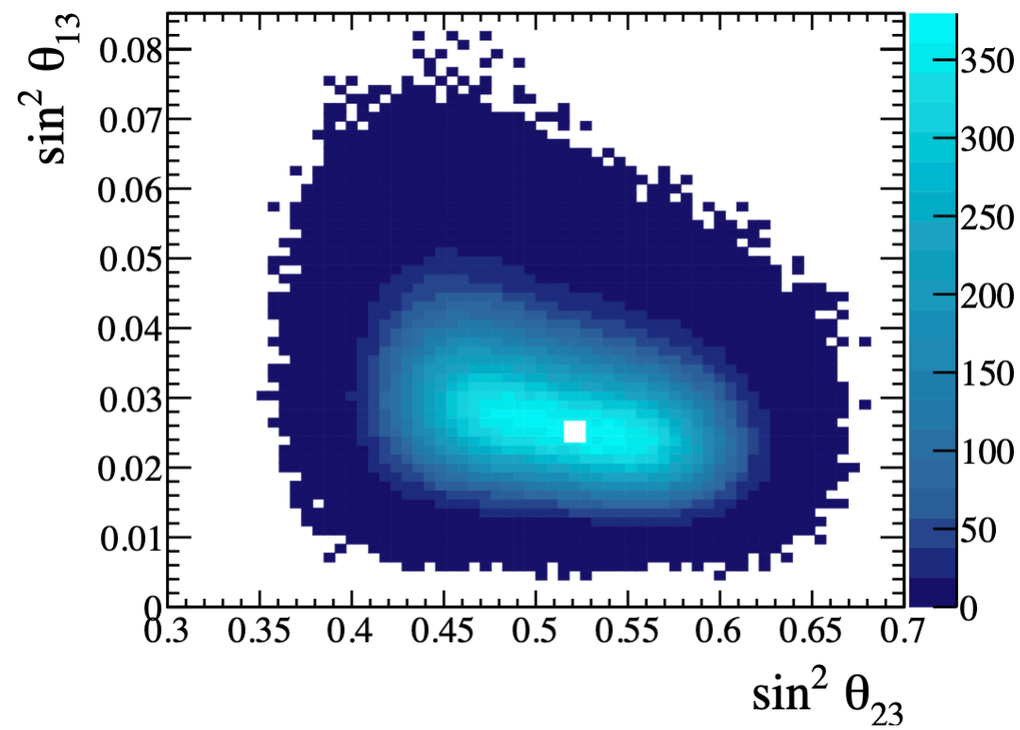
$$G = \frac{\bar{x}_{ini} - \bar{x}_{fin}}{\sqrt{\sigma(x)^2_{ini} + \sigma(x)^2_{fin}}}$$

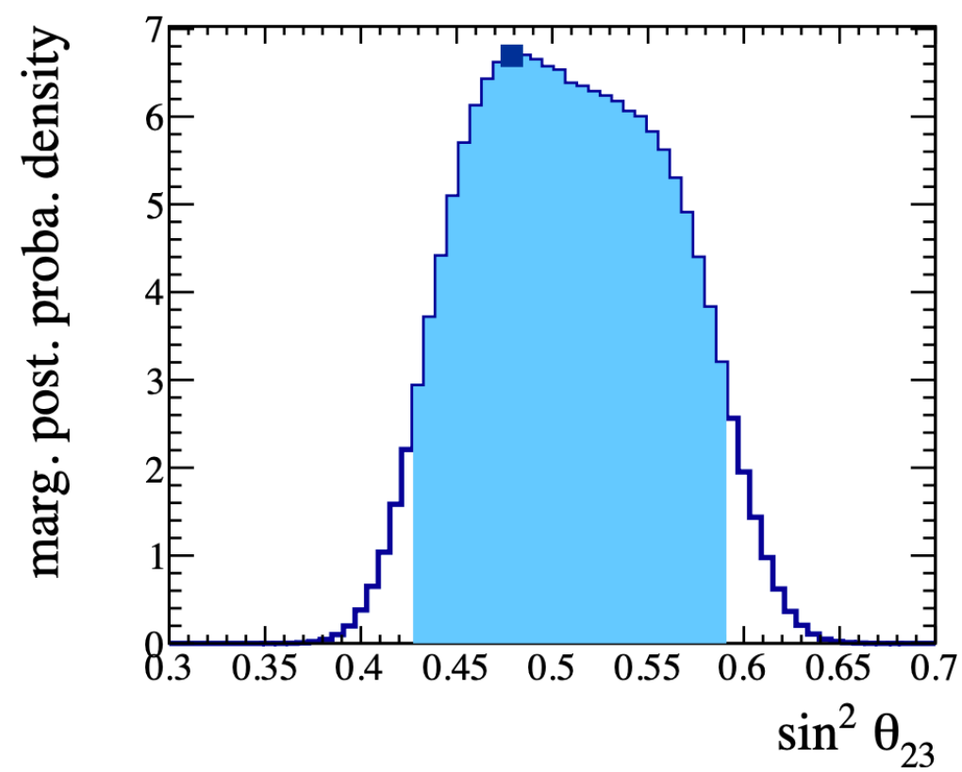
Note: 5% is not a hard rule, other binning can be chosen



Credible intervals

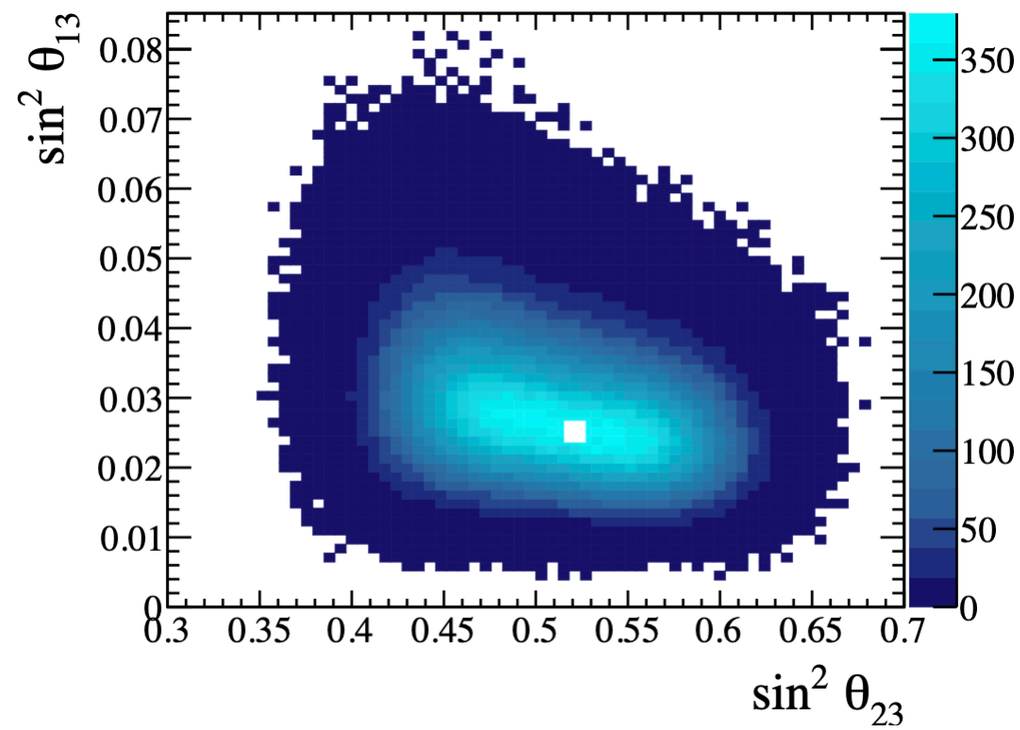
- **In Bayesian probabilities, results are given as *credible intervals***
  - Area where there is the highest probability that the true value lies in
  - E.g. there is a 90% probability that the true value of $\sin^2\theta_{23}$ is in [0.42, 0.59]



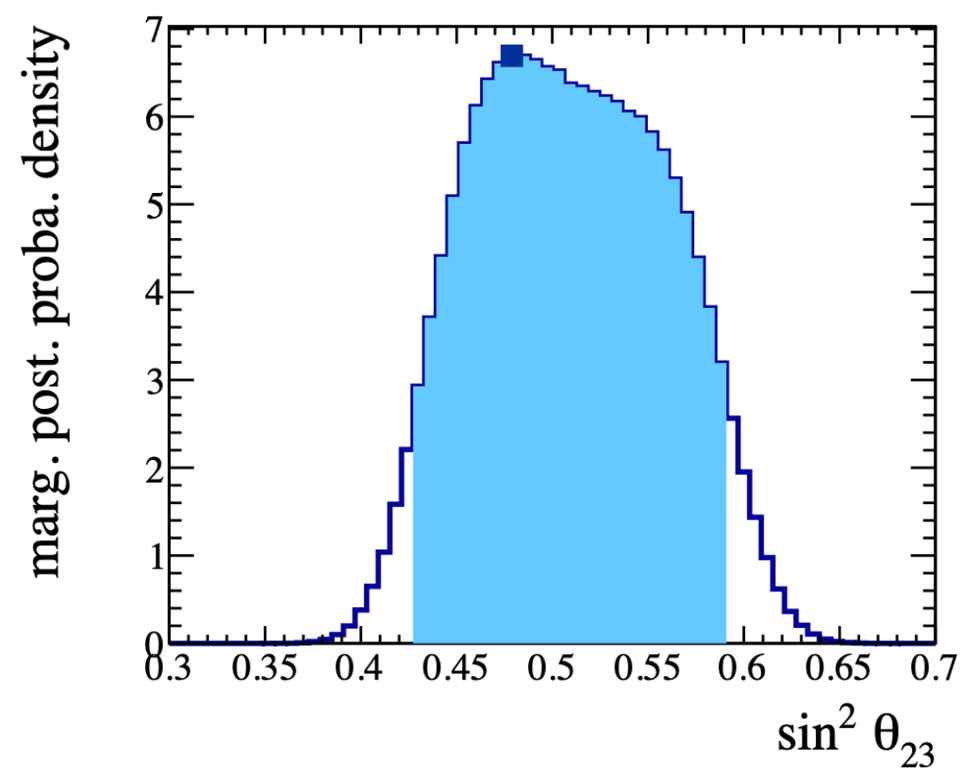


Credible intervals

- **In Bayesian probabilities, results are given as *credible intervals***
  - Area where there is the highest probability that the true value lies in
  - E.g. there is a 90% probability that the true value of $\sin^2\theta_{23}$ is in [0.42, 0.59]

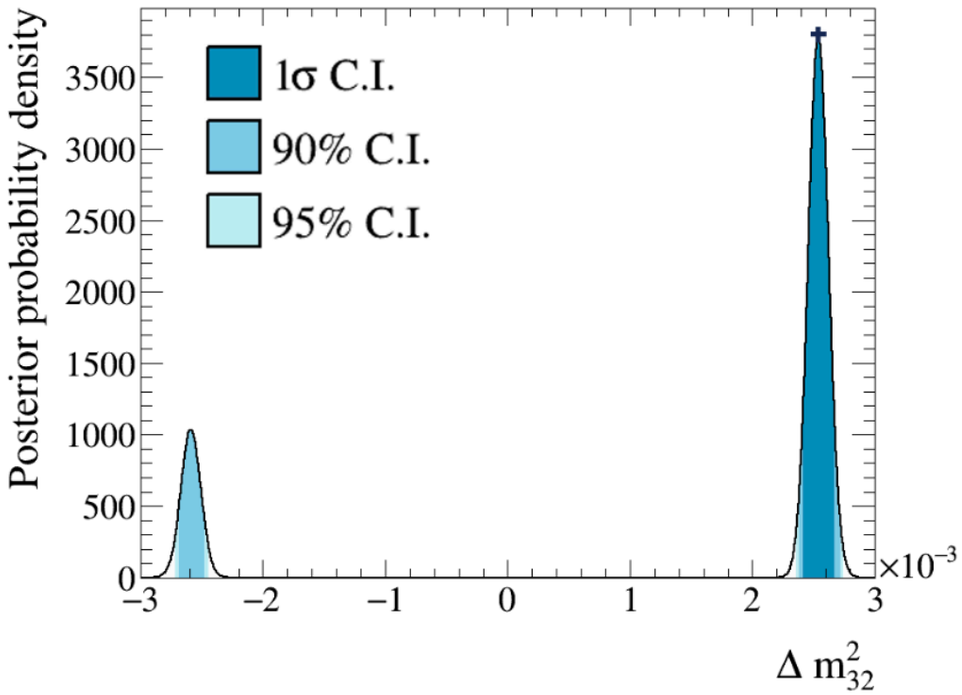


- **The posterior probabilities are automatically marginalised**
  - When projecting in lower dimensions than the Markov chain, the shape of the posterior probabilities of the other parameters is included in the integral



Bayes factor

- **Comparison of 2 hypotheses**
  - If we have 2 hypotheses H_1 and H_2, we can compare them with the *Bayes factor*, i.e. the ratio of marginalised likelihood
  - Bayes factor: $B_F = \dfrac{P(D|H_1)}{P(D|H_2)}$
  - If the prior probabilities are the same, this is equivalent to the ratio of posterior probabilities
  - Example: the Bayes factor for normal ordering is $B_F = 3.72$ on this plot



Changing the prior

- **The posterior probability can be evaluated for a different definition of the prior**
  - Equivalent to a variable change of the distribution:
    prior in $x \rightarrow$ prior in $y = f(x)$
  - Need to evaluate the Jacobian of the transformation:

$$P(H(x)) \;\rightarrow\; P(H(y)) \;=\; P(H(x))\;|J(y)|$$
$$= \; P(H(x))\;\left|\frac{\partial x}{\partial y}\right|$$

  - Can be extended to multi-variable cases

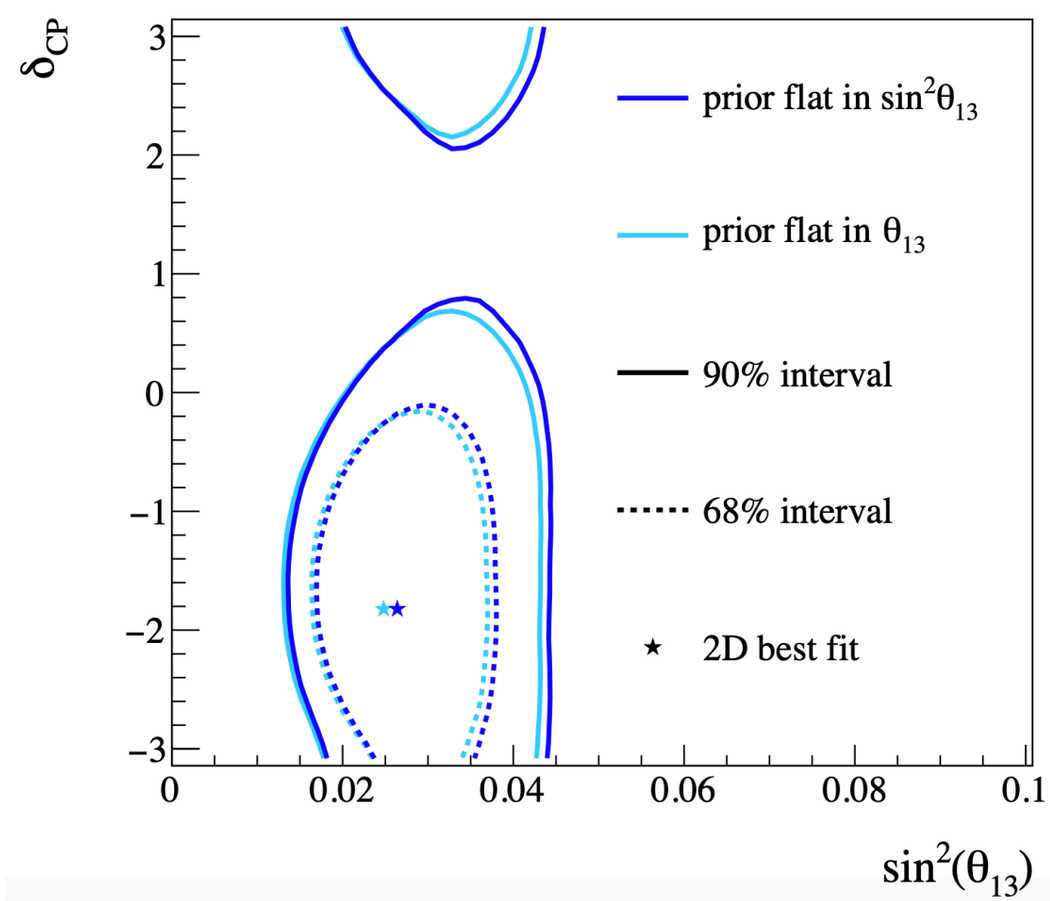
Changing the prior

- **The posterior probability can be evaluated for a different definition of the prior**
  - Equivalent to a variable change of the distribution:
    prior in $x \rightarrow$ prior in $y = f(x)$
  - Need to evaluate the Jacobian of the transformation:

$$P(H(x)) \rightarrow P(H(y)) = P(H(x))\,|J(y)|$$
$$= P(H(x))\,\left|\frac{\partial x}{\partial y}\right|$$

  - Can be extended to multi-variable cases

- **A useful way to:**
  - Check the robustness of the prior



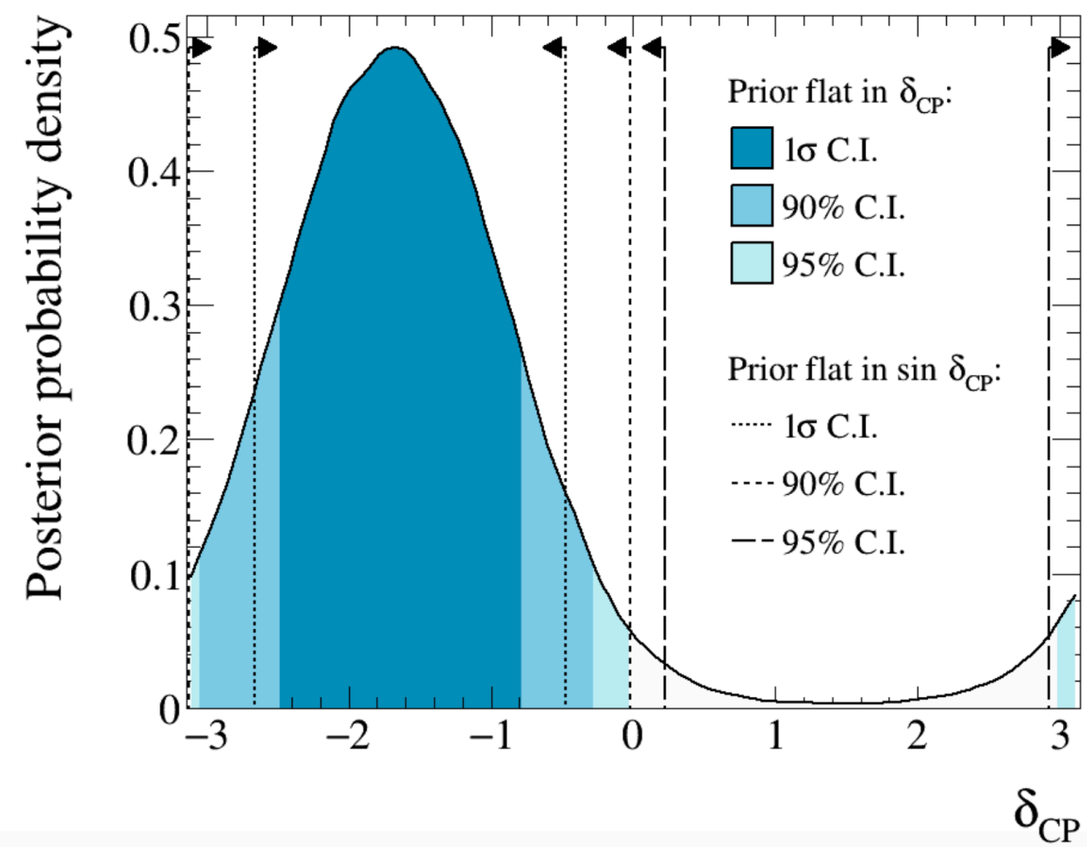
Changing the prior

- **The posterior probability can be evaluated for a different definition of the prior**
  - Equivalent to a variable change of the distribution:
    prior in $x \rightarrow$ prior in $y = f(x)$
  - Need to evaluate the Jacobian of the transformation:

$$P(H(x)) \rightarrow P(H(y)) = P(H(x))\,|J(y)|$$
$$= P(H(x))\,\left|\frac{\partial x}{\partial y}\right|$$

  - Can be extended to multi-variable cases

- **A useful way to:**
  - Check the robustness of the prior
  - Answer a different question
    e.g. what is the probability of CP-violation (instead of what is the δ_{CP} value)



Comparison with frequentist stat.

	Frequentist	**Bayesian**
probability	frequency of occurence	degree of belief
parameters	fixed (once chosen)	uncertain
observation	fluctuates	certain (once observed)

The two approaches in a nutshell:

- frequenstist $\rightarrow$ probability of observation, given a model
- bayesian $\rightarrow$ probability of a model, given an observation

Methodologies

- *frequenstist:* estimates frequencies, by emulating repetitions of the experiment (toys) for a given parameter, using the **likelihood** as PDF
- *bayesian:* exploits the Bayes theorem to compute the posterior $P(para|obs)$, using the prior $P(para)$ and $P(obs|para)$ - the **likelihood**

3. Both approaches get unifed when

- there is an infinite number of measurements
- the prior is uniform: $P(par|obs) = A \times \mathcal{L}(par; obs)$

  (same equation, but its meaning and the question it addresses *are* different)

Slide by Romain Madar, School of Statistics 2022

Conclusion

- **Bayesian inference consist in computing a posterior probability density**
  - Update the probability of a hypothesis according to the information on the data
  - Markov Chain Monte-Carlo is a useful tool to sample high dimensional cases
  - Can infer any shape of posterior probabilities

- **The process requires careful tuning**
  - Asymptotically, MCMC properties ensure that it will converge to the target distribution
  - We do not have infinite time, neither an infinite number of processors
  - Ensuring convergence is key to the process
    → convincing ourselves that the output is the needed one is not easy!
  - Extensive literature about it, but no « one-solution-fit-all »
  - Does not mean it should no be used! But not blindly