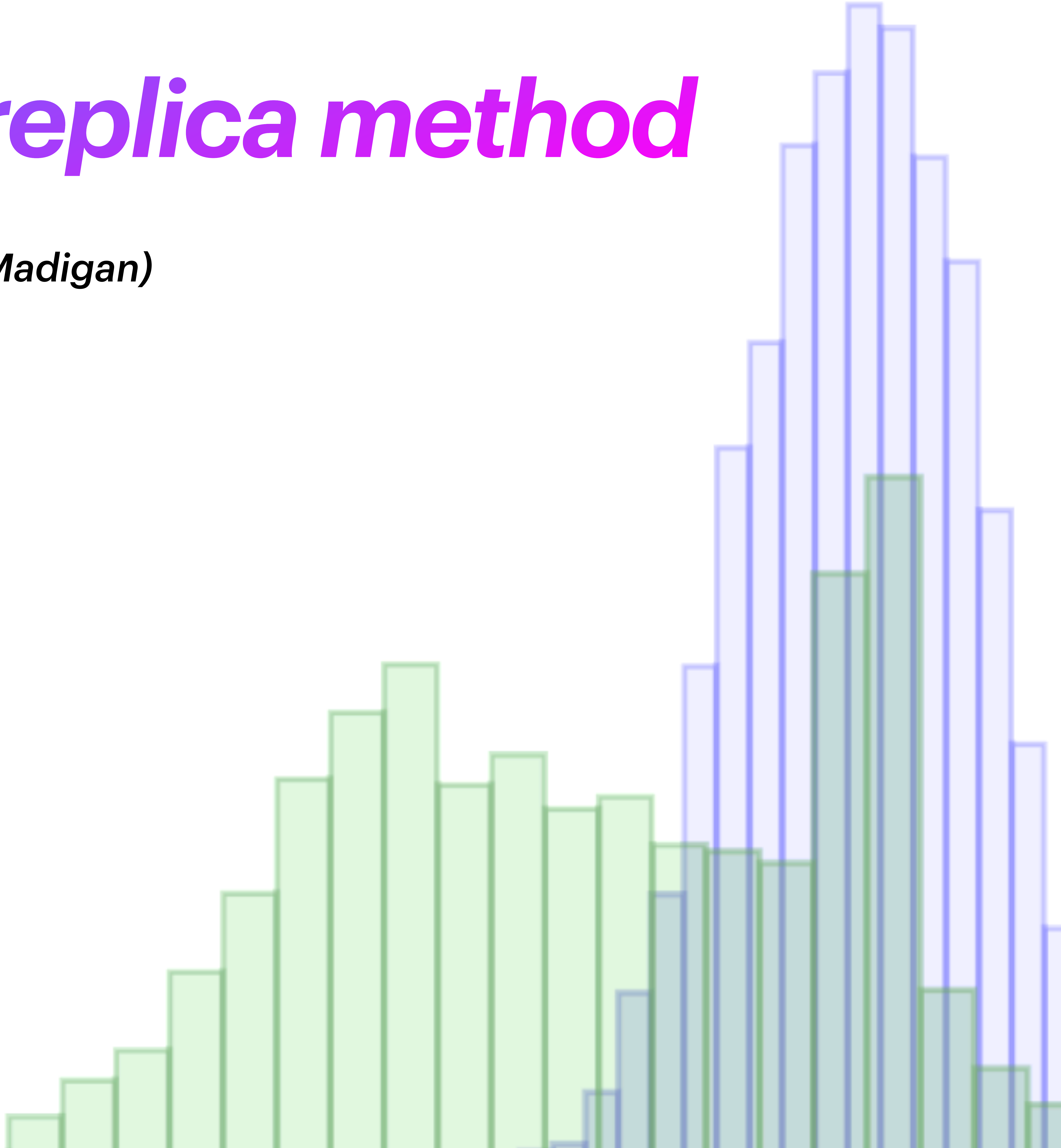
A study of the MC replica method

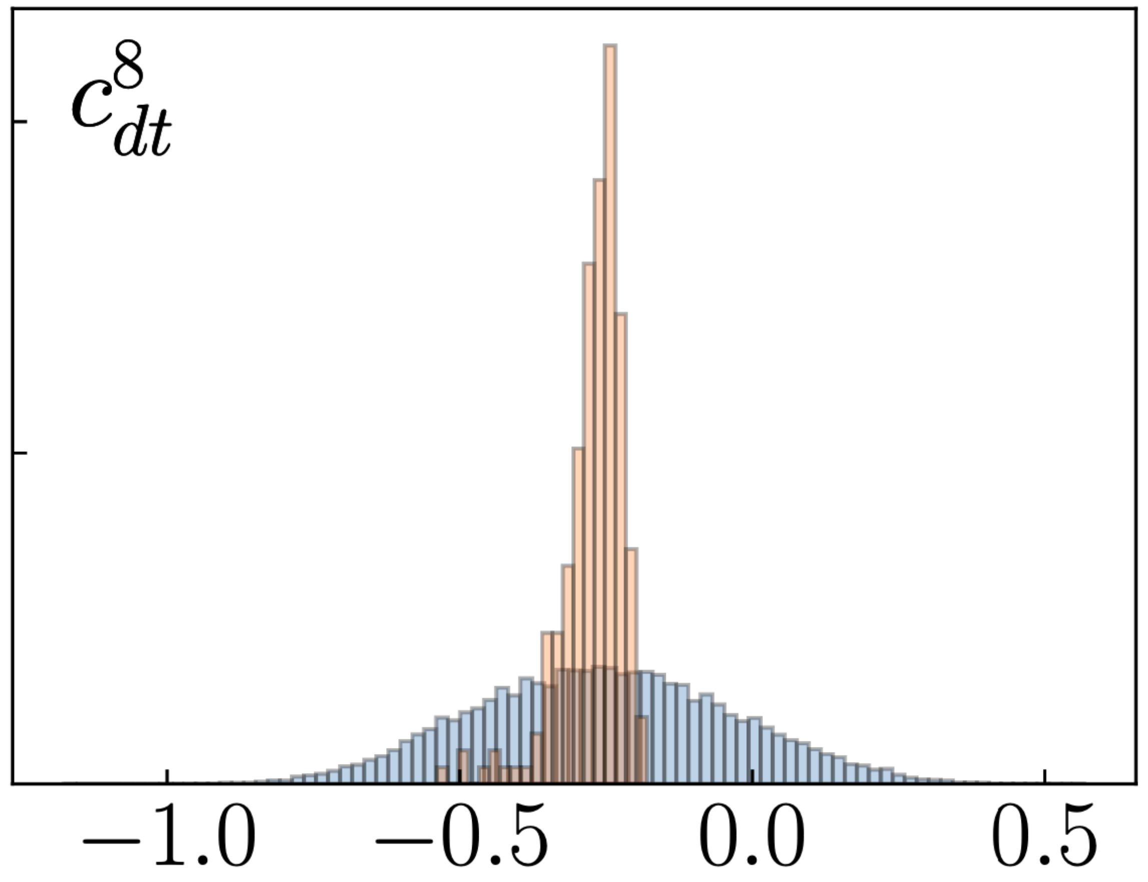
for PDF4LHC 2024, CERN, based on 2404.10056
(with Mark Costantini, Luca Mantani and Maeve Madigan)

James Moore,
Lucy Cavendish College,
University of Cambridge



The talk in a nutshell...

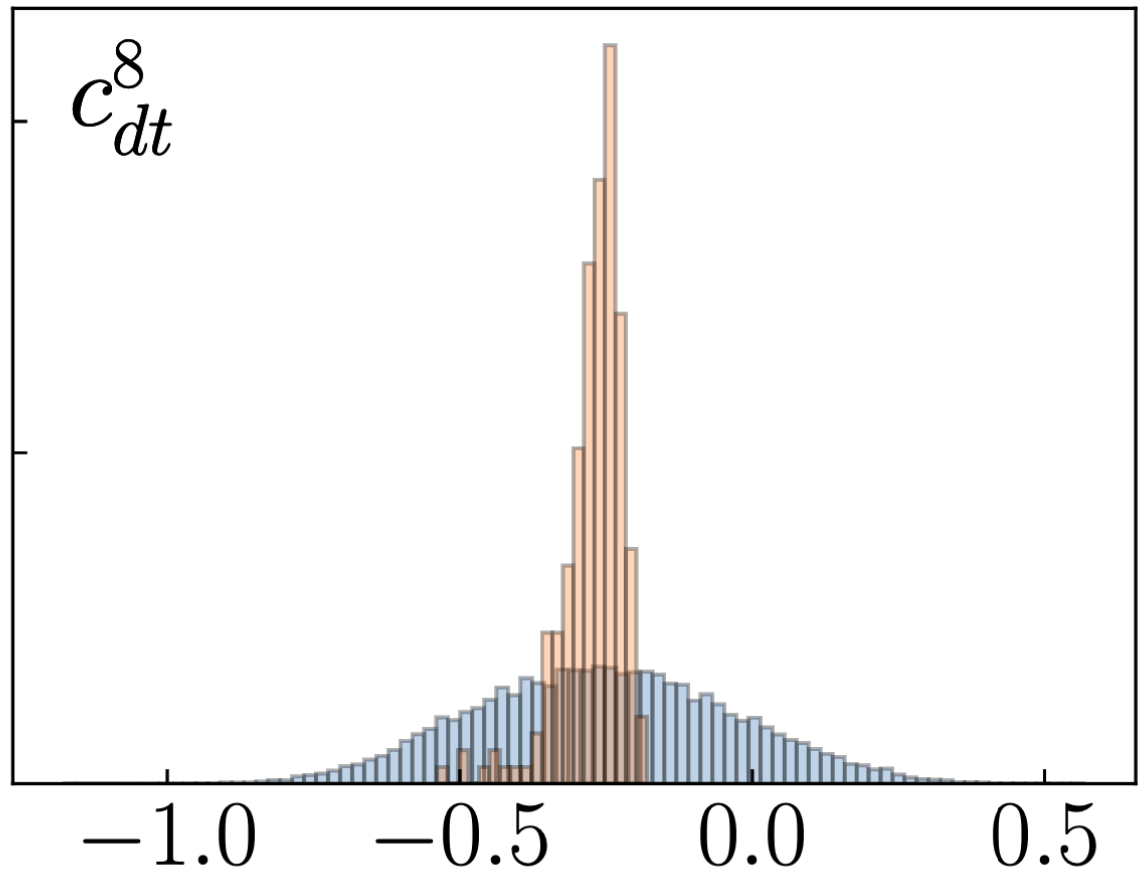
1. Why study the Monte Carlo replica method?

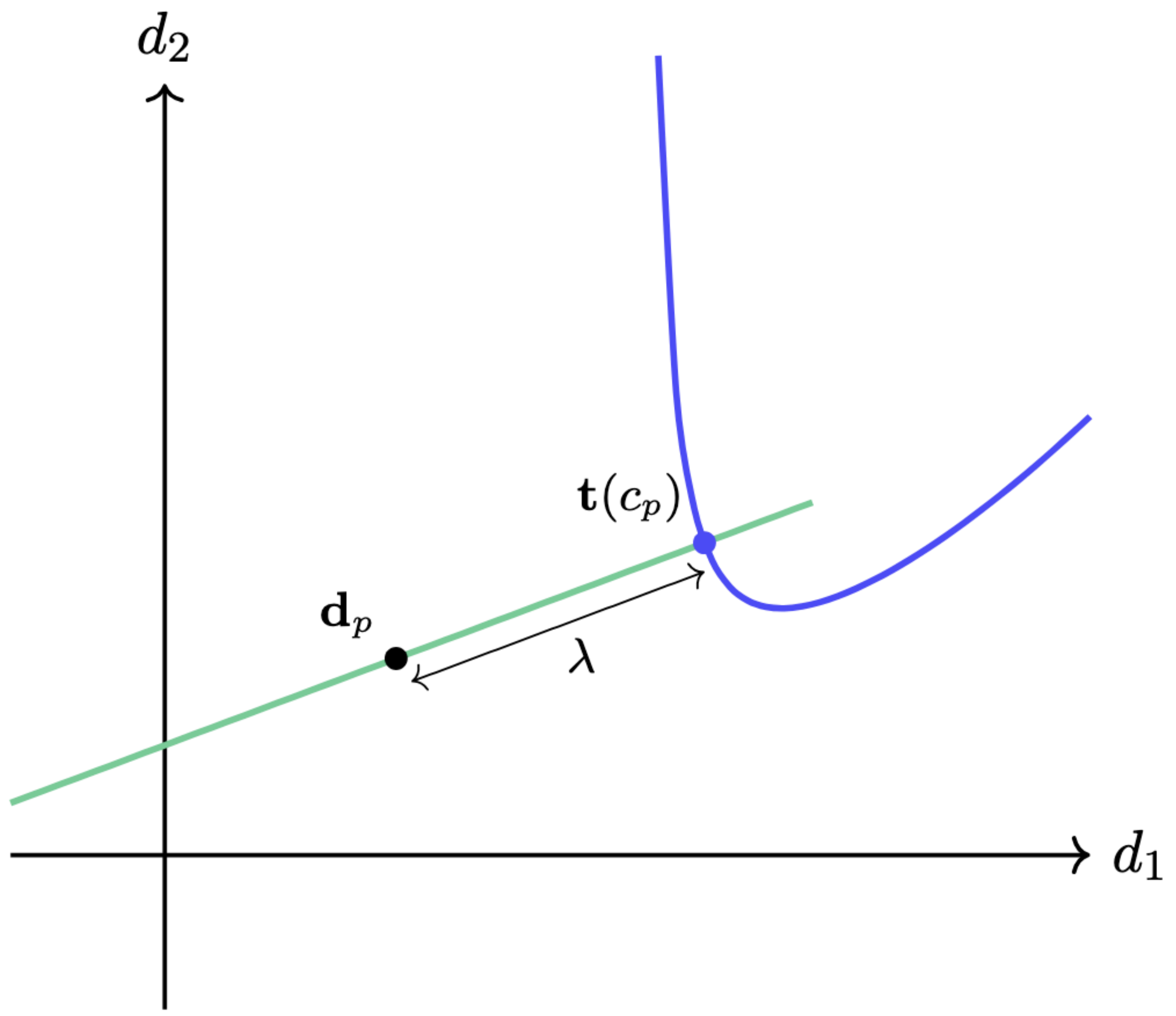


The talk in a nutshell...

1. Why study the Monte Carlo replica method?

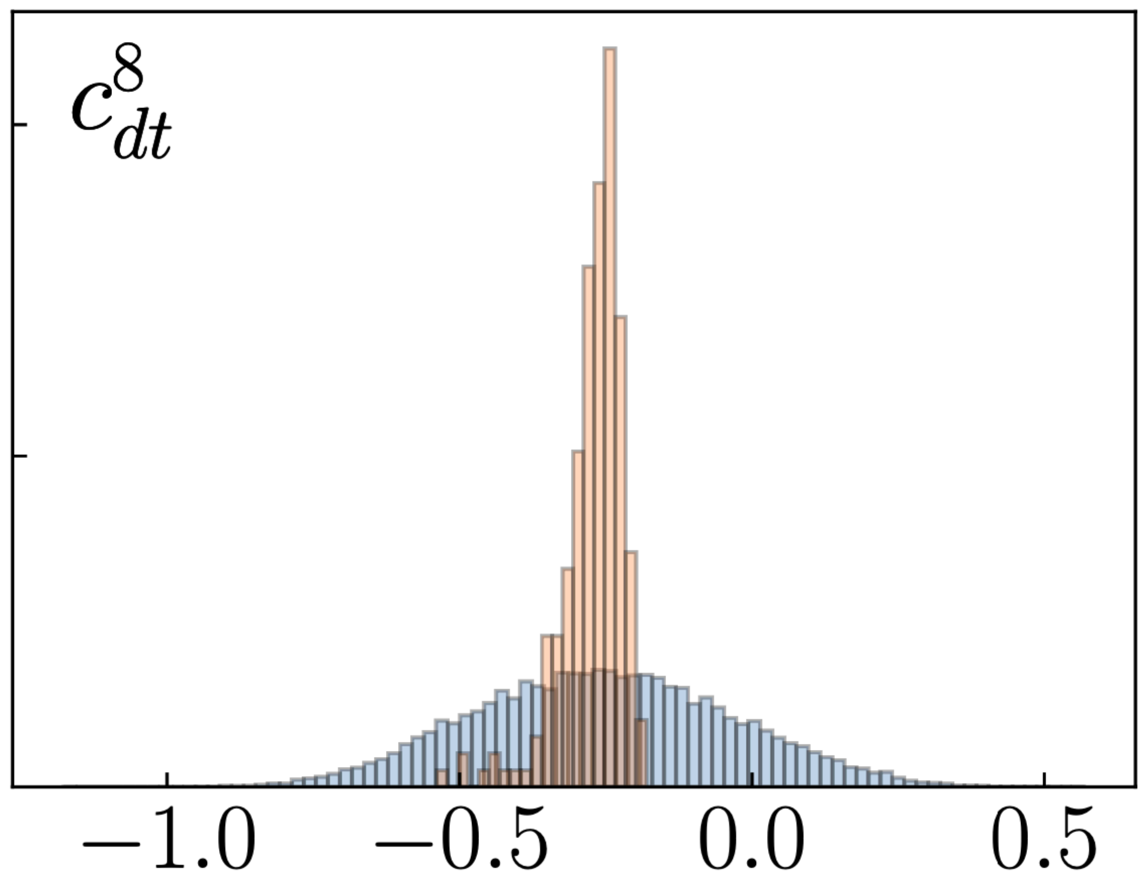
2. A geometric approach to the MC replica method

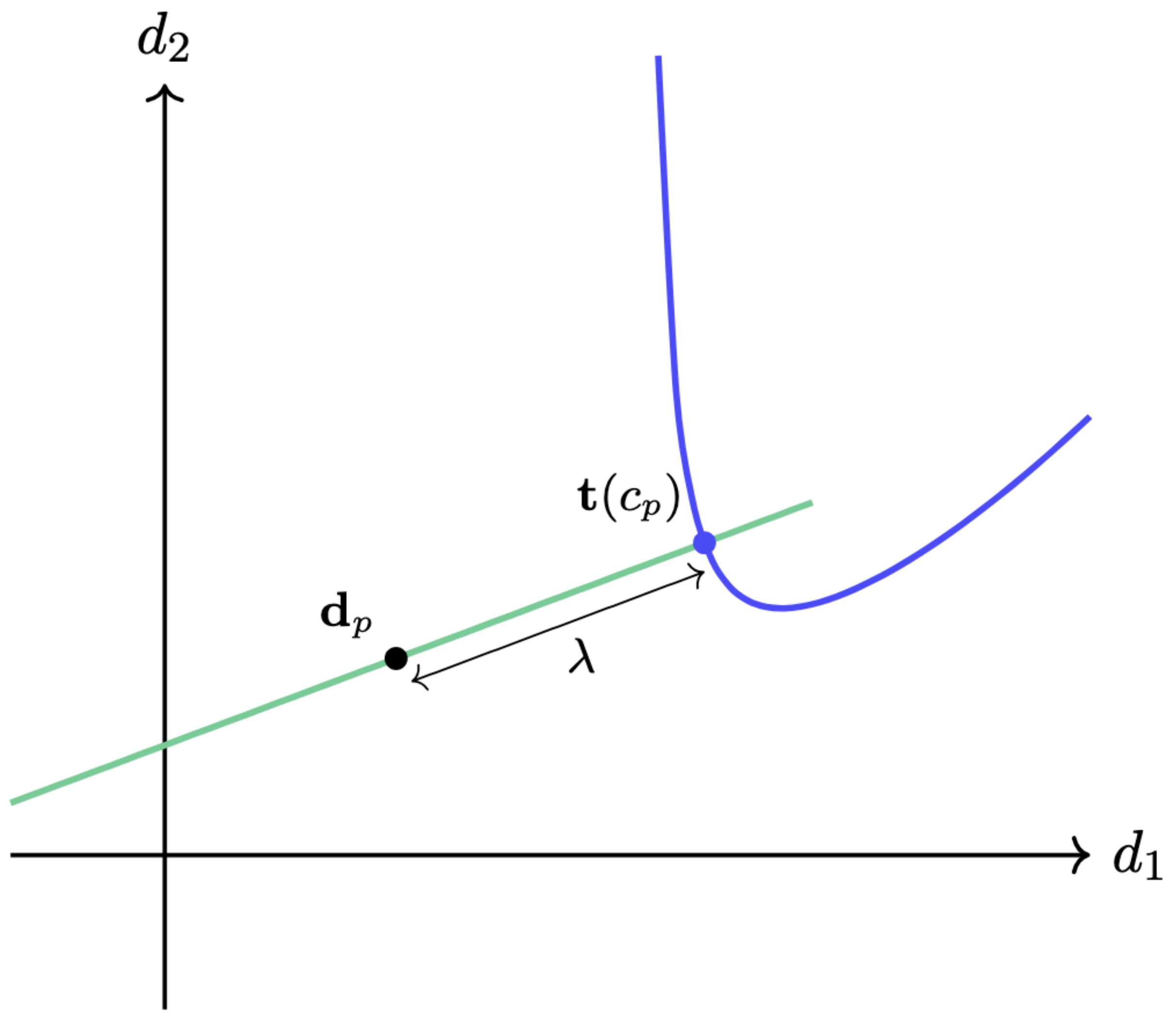


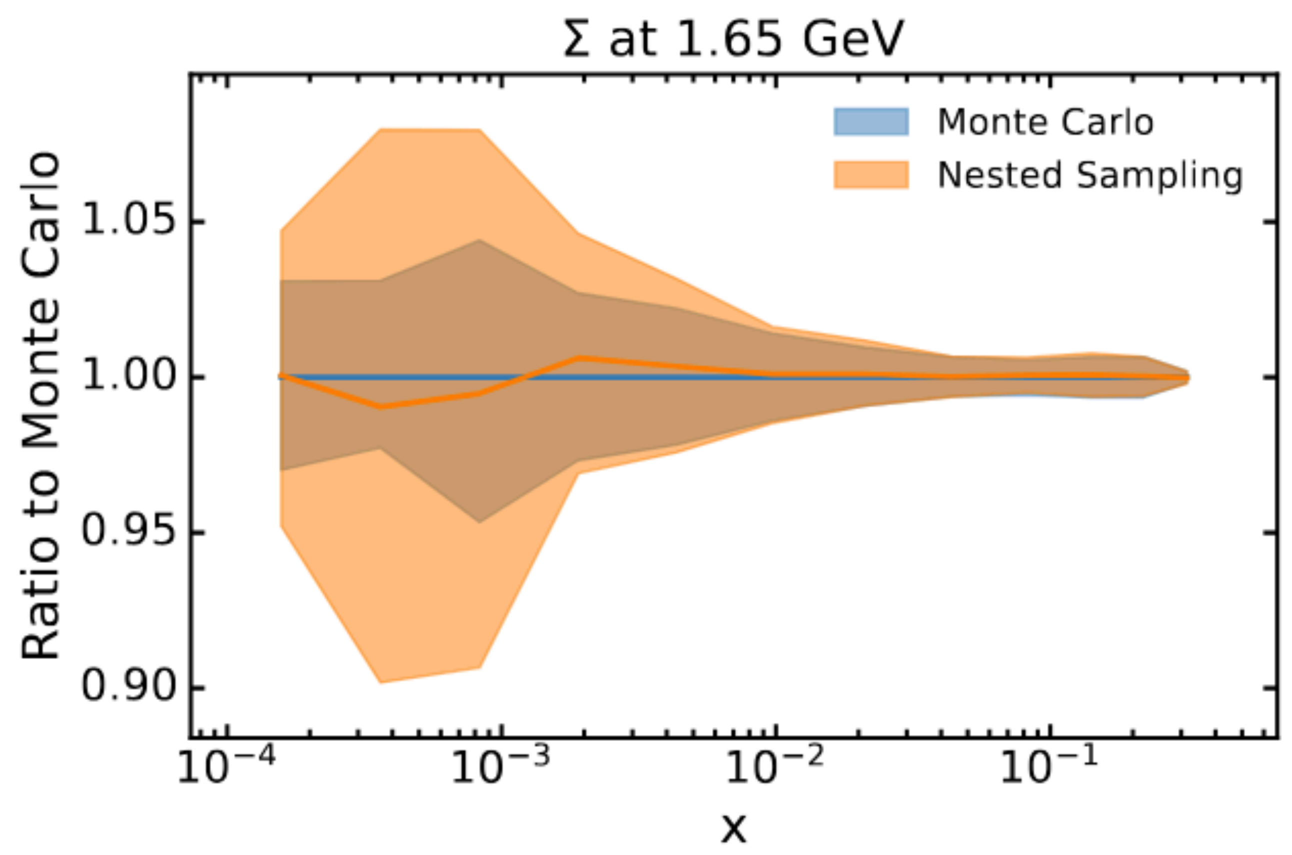


The talk in a nutshell...

1. Why study the Monte Carlo replica method? → 2. A geometric approach to the MC replica method → 3. Relevance in PDF fits







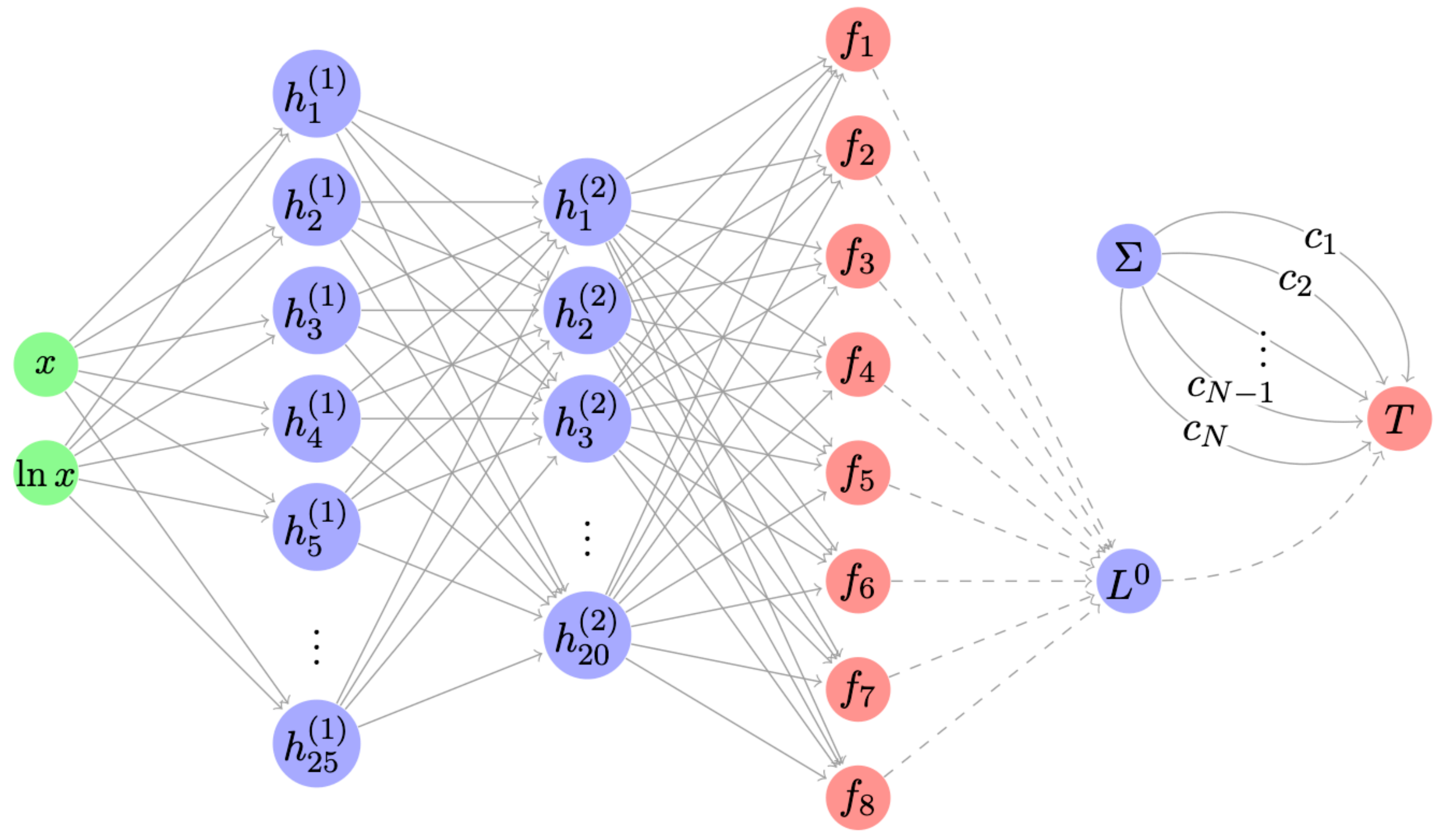
1. - Why study the Monte Carlo replica method?

New Physics in the top sector...?

- Back in the **winter of 2022**, Maria Ubiali's team in Cambridge, together with Juan Rojo, were working on a comprehensive analysis of the impact of **top data** on simultaneous PDF-EFT fits - see 2303.06159.

New Physics in the top sector...?

- Back in the **winter of 2022**, Maria Ubiali’s team in Cambridge, together with Juan Rojo, were working on a comprehensive analysis of the impact of **top data** on simultaneous PDF-EFT fits - see 2303.06159.



- This study was facilitated through our fantastic new **SIMUnet code**, which is publicly available for use at: https://hep-pbsp.github.io/SIMUnet. **Go check it out!**

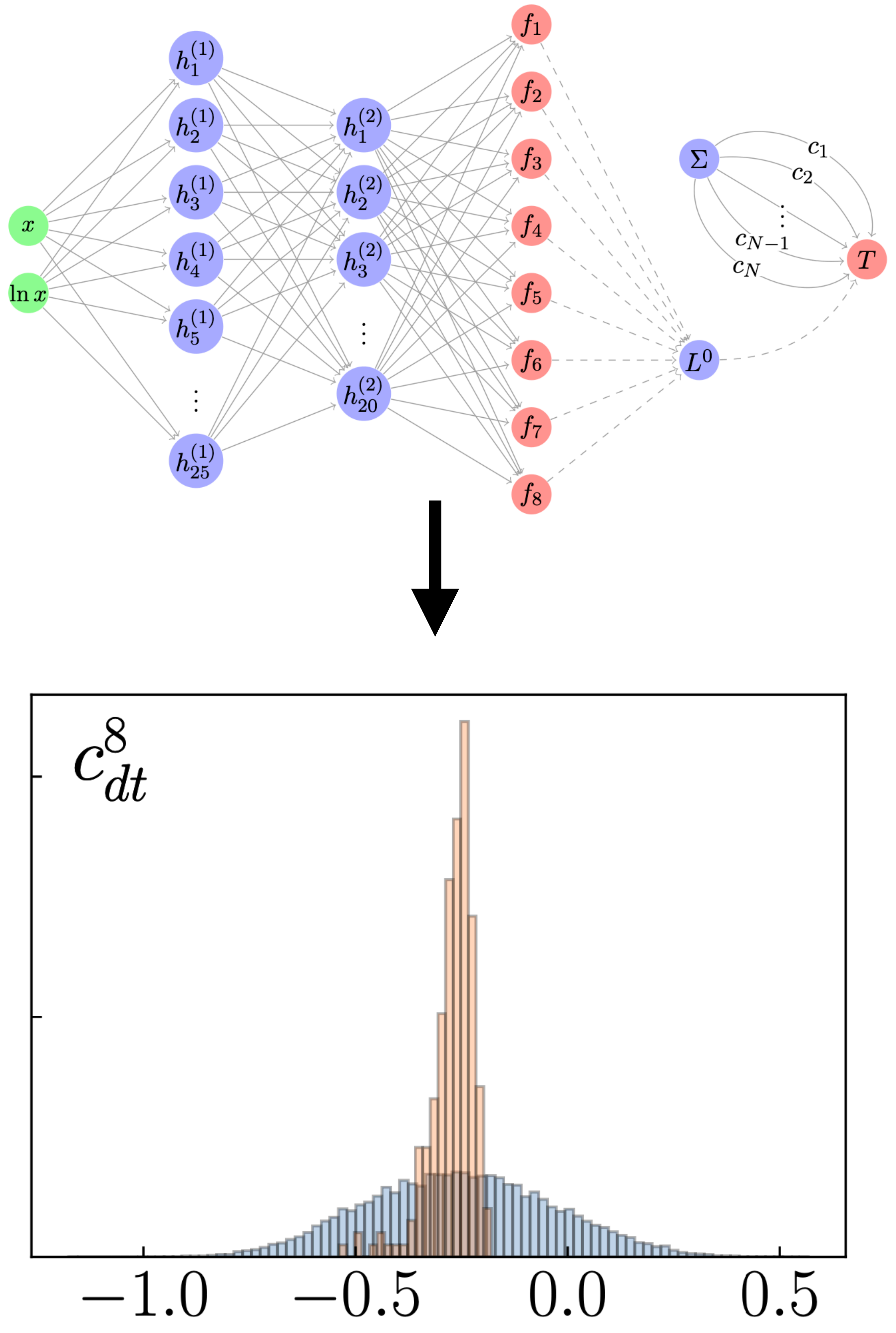
SIMUnet

New Physics in the top sector...?

- Back in the **winter of 2022**, Maria Ubiali's team in Cambridge, together with Juan Rojo, were working on a comprehensive analysis of the impact of **top data** on simultaneous PDF-EFT fits - see 2303.06159.

- This study was facilitated through our fantastic new **SIMUnet code**, which is publicly available for use at: https://hep-pbsp.github.io/SIMUnet. **Go check it out!**

- We had a bit of a shock, though, when we ran the **quadratic SMEFT results**, and discovered **new physics at the 7σ level**!



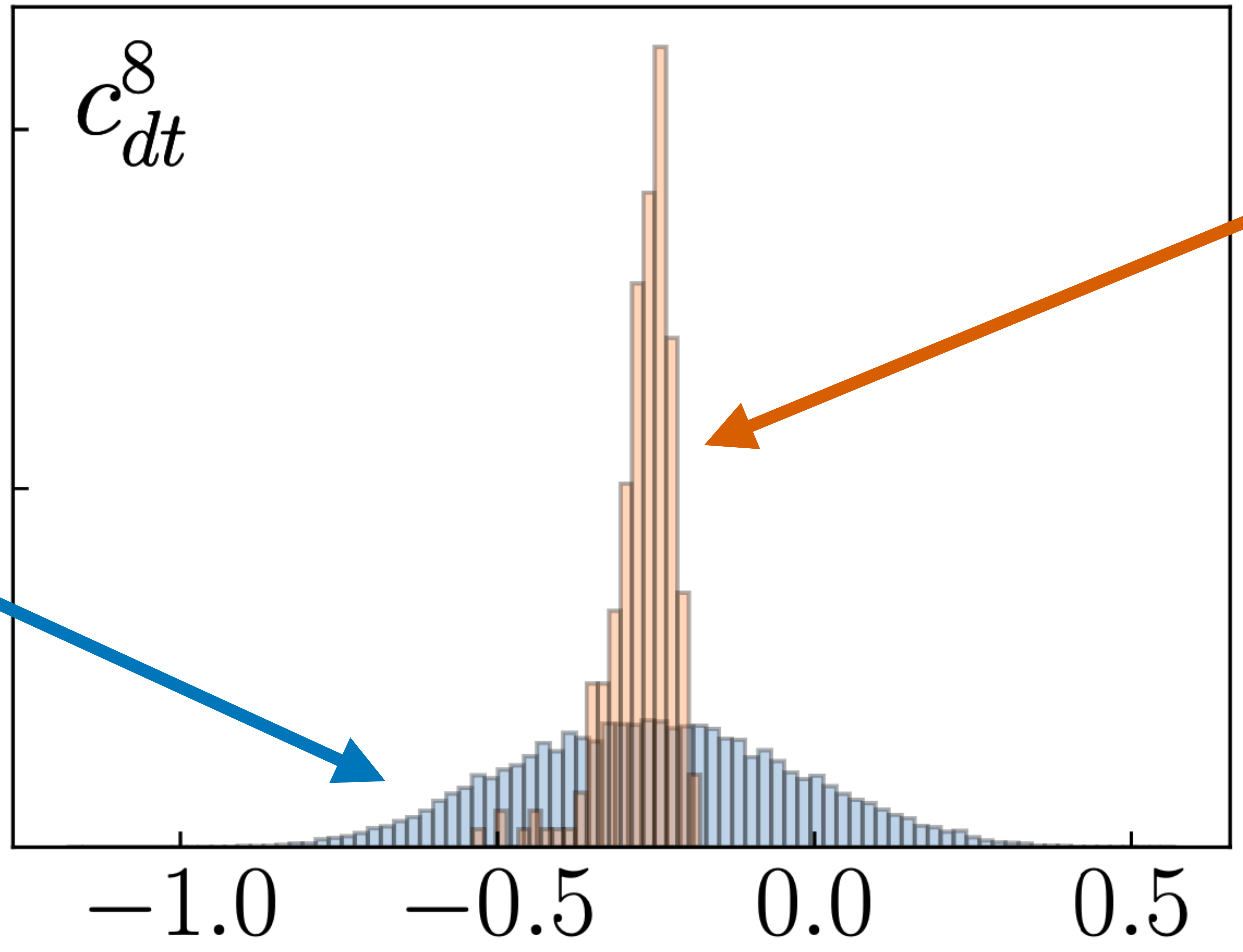
So, where's our Nobel prize?

So, where's our Nobel prize?

- After searching through optimiser learning rates, initialisations, and checking datasets twice, we traced the issue to be **methodological** in nature.

So, where's our Nobel prize?

- After searching through optimiser learning rates, initialisations, and checking datasets twice, we traced the issue to be **methodological** in nature.
- We realised this by **cross-checking with the SMEFiT code** (2302.06660) - we saw that **the Monte Carlo option disagreed with the Bayesian option**.



'spiked' Wilson coefficient distribution from using SIMUnet or SMEFiT Monte Carlo option

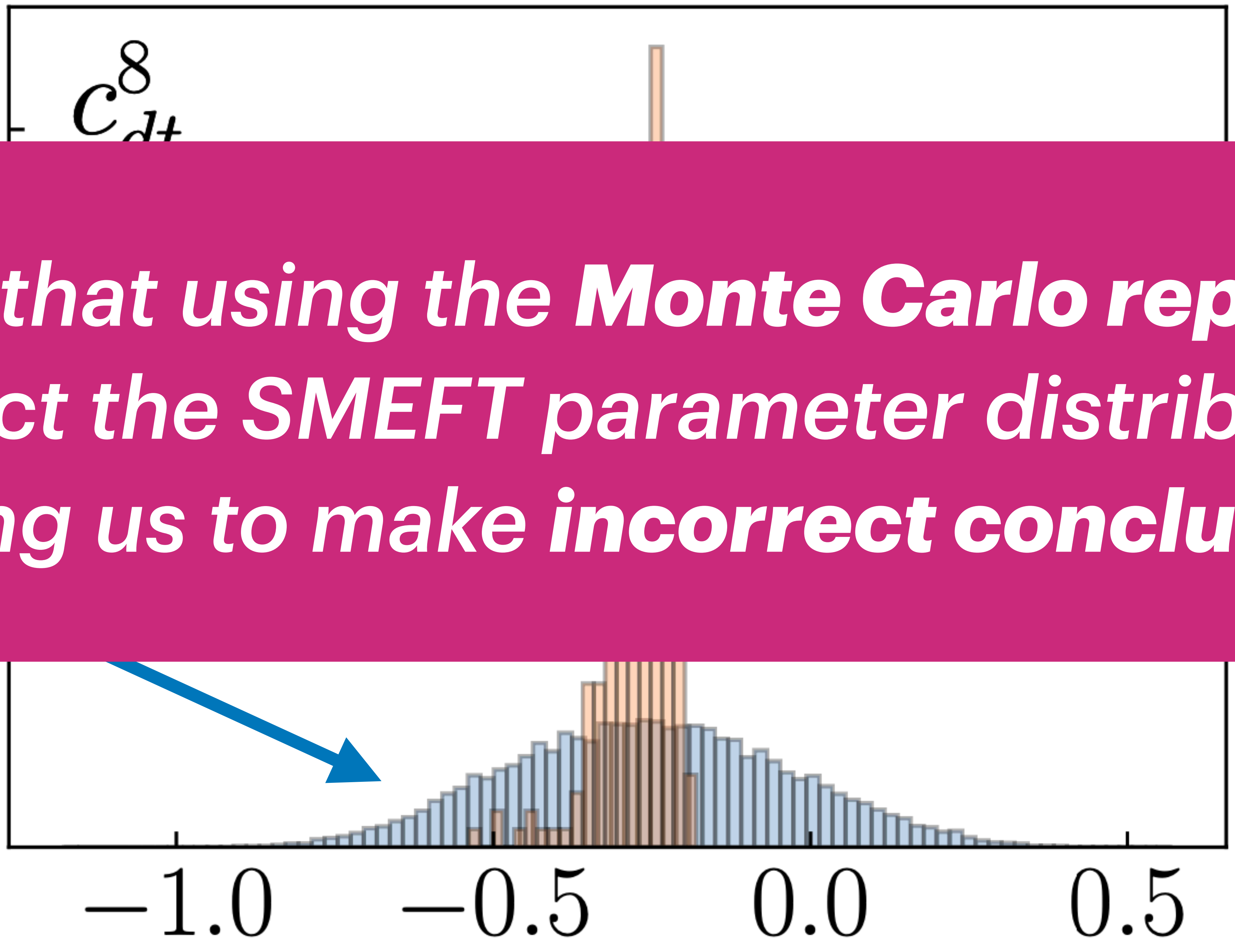
distribution from SMEFiT Bayesian option (using the Nested Sampling algorithm), with a wide uniform prior

So, where's our Nobel prize?

- After searching through optimiser learning rates, initialisations, and checking datasets twice, we traced the issue to be **methodological** in nature.
- We realised this by **cross-checking with the SMEFiT code** (2302.06660) - we saw that **the Monte Carlo option disagreed with the Bayesian option**.

'spiked' Wilson coefficient distribution from
... SMEFiT Monte Carlo option

distribution from S...
(using the Nestec...
with a wid...



It appeared that using the Monte Carlo replica method to construct the SMEFT parameter distributions was leading us to make incorrect conclusions.

Key questions for the talk...

1. **Why was there a disagreement between the distributions - can this be understood mathematically?**

Key questions for the talk...

1. **Why was there a disagreement between the distributions - can this be understood mathematically?**

2. **Could using the Monte Carlo replica method in PDF fits give different distributions, as compared to using a Bayesian methodology?**

2. - A geometric approach to the MC replica method

Bayesian vs MC analyses

- The disagreement shown in our top fit distributions motivated us to compare the Monte Carlo replica method with a fully Bayesian method on a **deeper mathematical level**. To present the results, we first set up the framework.

Bayesian vs MC analyses

- The disagreement shown in our top fit distributions motivated us to compare the Monte Carlo replica method with a fully Bayesian method on a **deeper mathematical level**. To present the results, we first set up the framework.

- Imagine we are given a vector of experimental central data values $\mathbf{d} \in \mathbb{R}^{N_{\mathrm{dat}}}$ by the experimentalists, together with an experimental covariance matrix Σ. We would like to compare this to our **theory predictions**:

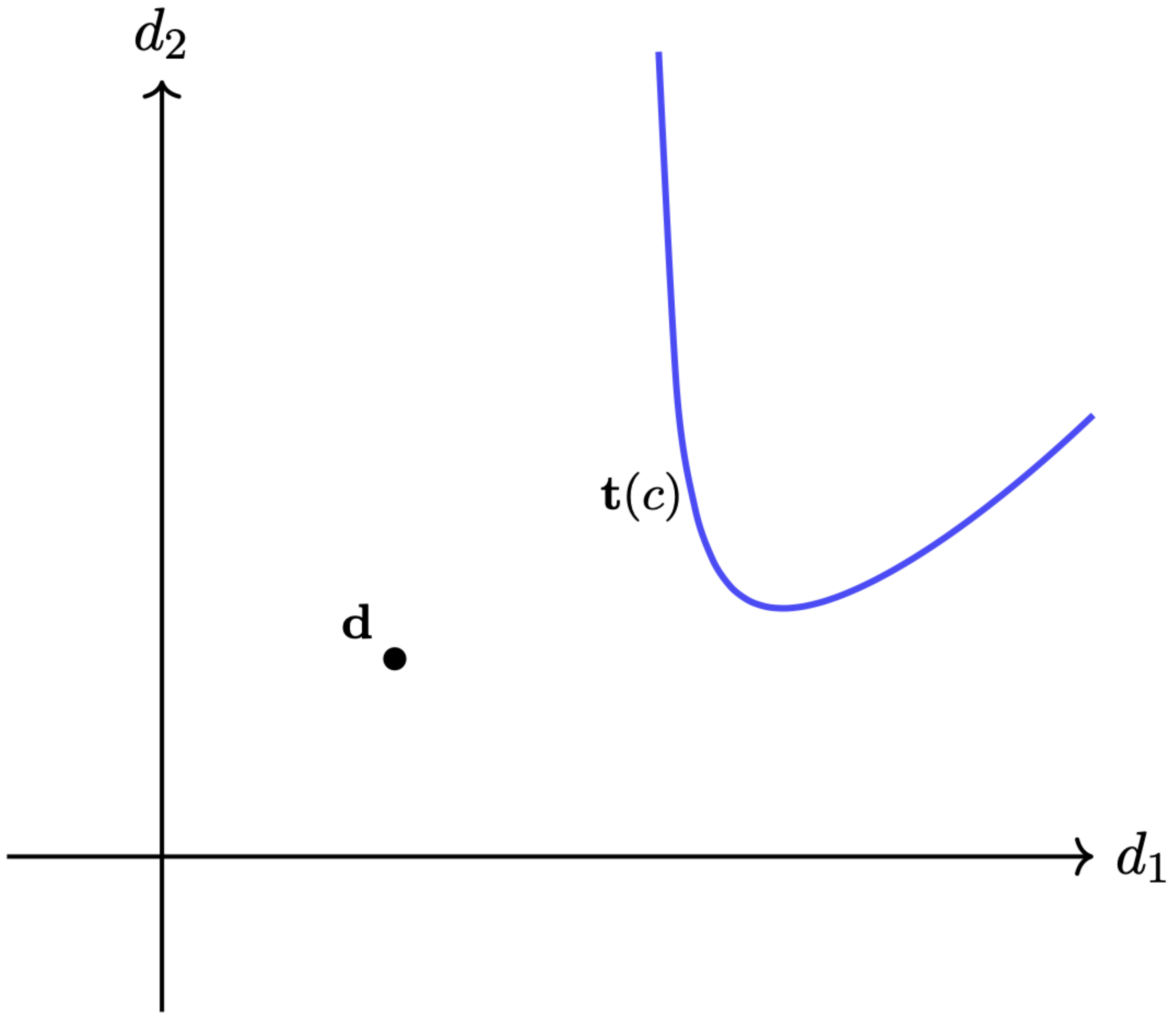
$$\mathbf{t} : \mathbb{R}^{N_{\mathrm{param}}} \to \mathbb{R}^{N_{\mathrm{dat}}}$$

which can be viewed as a vector function $\mathbf{t}(\mathbf{c})$ of some parameters $\mathbf{c} \in \mathbb{R}^{N_{\mathrm{param}}}$. The parameters in SMEFT fits are **Wilson coefficients**, and in PDF fits they are the **parameters of the PDF model under consideration**.

Bayesian vs MC analyses

- If the number of parameters is smaller than the number of data, we can view the theory function $\mathbf{t}(\mathbf{c})$ as **carving out a surface in data-space** $\mathbb{R}^{N_{\text{dat}}}$.

- *Right*: an example, with **two data points** and a theory prediction (in blue) depending on a **single parameter** c. The **observed data** is at the point $\mathbf{d}$.



Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method

- The parameter distributions are given by **Bayes' theorem**:

$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c})p(\mathbf{d}\,|\,\mathbf{c})$$
$$= \pi(\mathbf{c})\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T\Sigma^{-1}(\mathbf{d}-\mathbf{t}(\mathbf{c}))\right),$$

where $\pi(\mathbf{c})$ is some **prior distribution**, and we use the fact that the experimental data is **normally distributed**.

Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method

- The parameter distributions are given by **Bayes' theorem**:

$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c})p(\mathbf{d}\,|\,\mathbf{c})$$
$$= \pi(\mathbf{c})\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T\Sigma^{-1}(\mathbf{d}-\mathbf{t}(\mathbf{c}))\right),$$

where $\pi(\mathbf{c})$ is some **prior distribution**, and we use the fact that the experimental data is **normally distributed**.

- The **exponent** (in pink) can be understood as the distance (with respect to the metric Σ) from the data, $\mathbf{d}$, to a point on the theory surface, $\mathbf{t}(\mathbf{c})$.

Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method

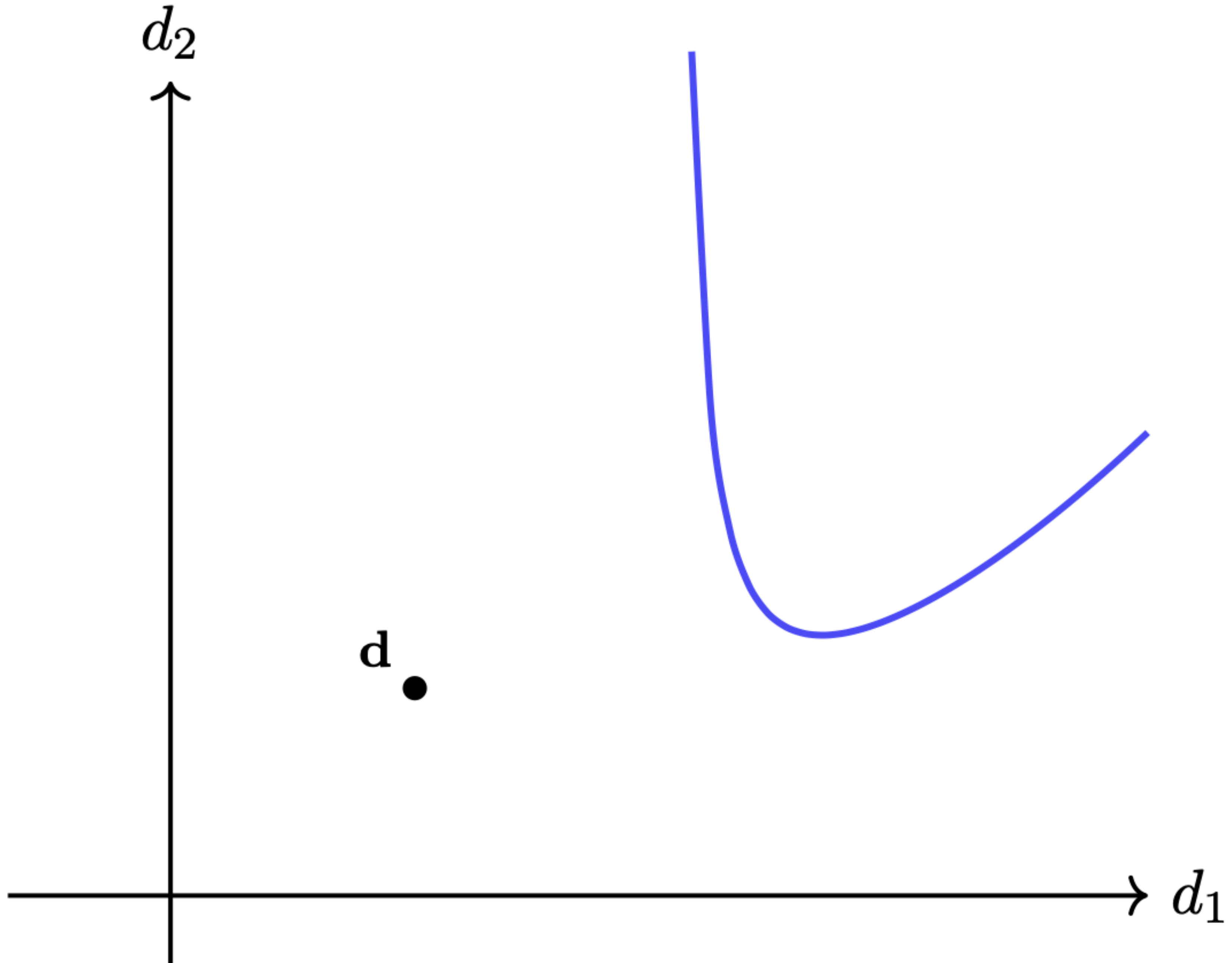
- The parameter distributions are given by **Bayes' theorem**:

$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c})p(\mathbf{d}\,|\,\mathbf{c})$$
$$= \pi(\mathbf{c})\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T\Sigma^{-1}(\mathbf{d}-\mathbf{t}(\mathbf{c}))\right),$$

where $\pi(\mathbf{c})$ is some **prior distribution**, and we use the fact that the experimental data is **normally distributed**.

- The **exponent** (in pink) can be understood as the distance (with respect to the metric Σ) from the data, $\mathbf{d}$, to a point on the theory surface, $\mathbf{t}(\mathbf{c})$.

- Hence, we can understand the Bayesian method as follows: **for each parameter value**, we compute the Σ**-distance** to the **data**. The posterior parameter distribution is the **exponential** of $-1/2$ multiplied by these distances, **scaled by the prior**.



Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method

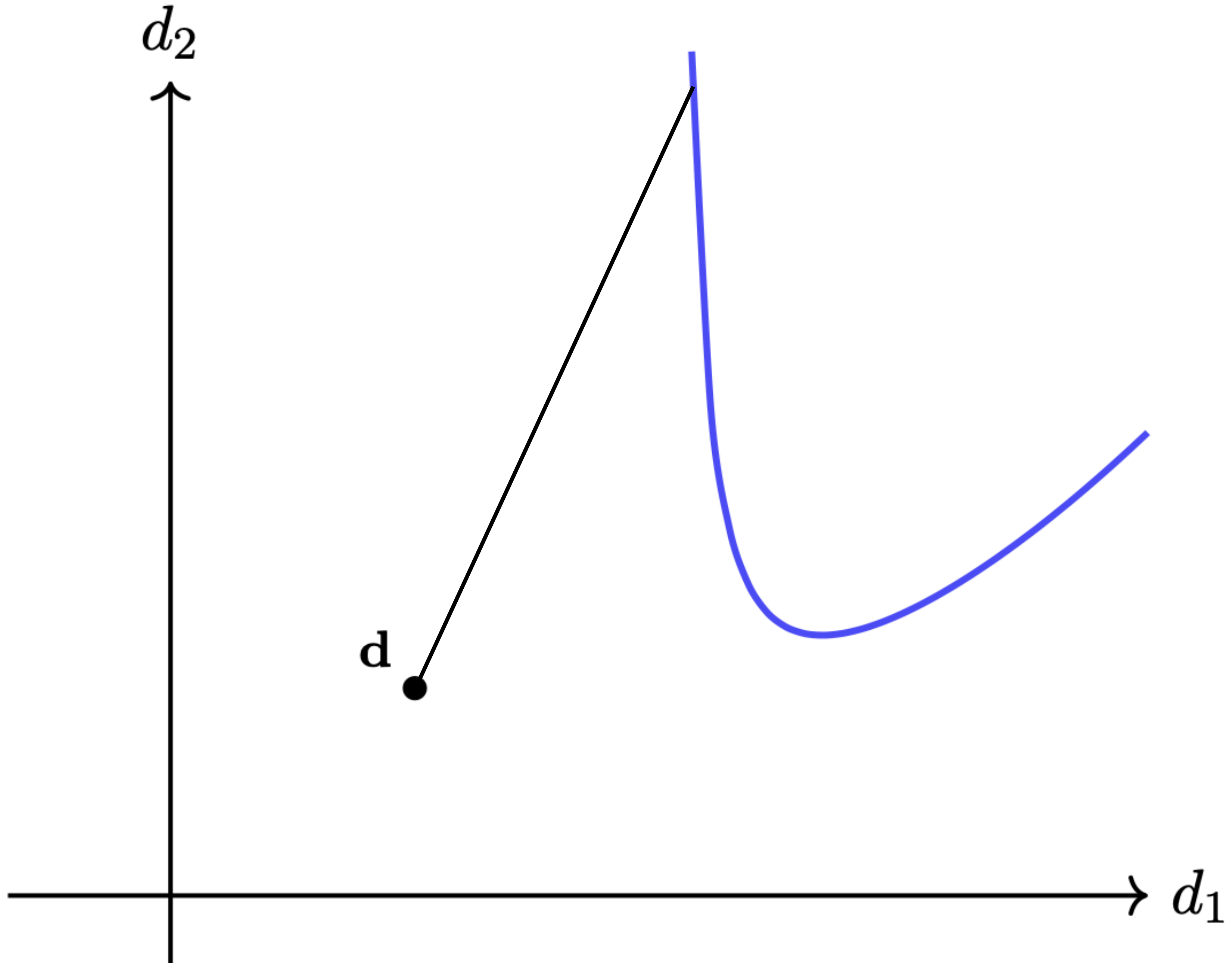
- The parameter distributions are given by **Bayes' theorem**:

$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c})p(\mathbf{d}\,|\,\mathbf{c})$$
$$= \pi(\mathbf{c})\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T\Sigma^{-1}(\mathbf{d}-\mathbf{t}(\mathbf{c}))\right),$$

where $\pi(\mathbf{c})$ is some **prior distribution**, and we use the fact that the experimental data is **normally distributed**.

- The **exponent** (in pink) can be understood as the distance (with respect to the metric Σ) from the data, $\mathbf{d}$, to a point on the theory surface, $\mathbf{t}(\mathbf{c})$.

- Hence, we can understand the Bayesian method as follows: **for each parameter value**, we compute the **Σ-distance** to the **data**. The posterior parameter distribution is the **exponential** of $-1/2$ multiplied by these distances, **scaled by the prior**.



Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method

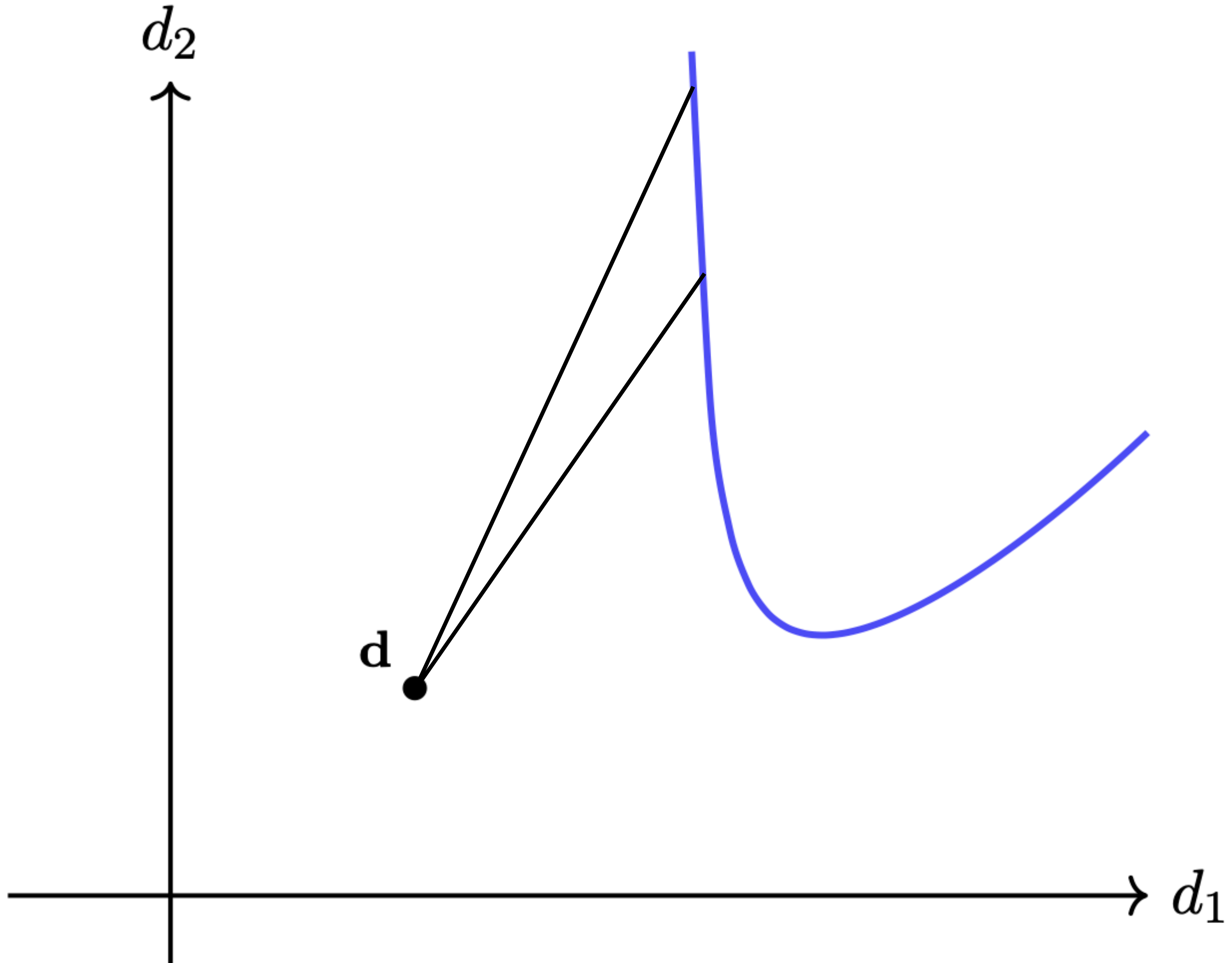
- The parameter distributions are given by **Bayes' theorem**:

$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c})p(\mathbf{d}\,|\,\mathbf{c})$$
$$= \pi(\mathbf{c})\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T\Sigma^{-1}(\mathbf{d}-\mathbf{t}(\mathbf{c}))\right),$$

where $\pi(\mathbf{c})$ is some **prior distribution**, and we use the fact that the experimental data is **normally distributed**.

- The **exponent** (in pink) can be understood as the distance (with respect to the metric Σ) from the data, $\mathbf{d}$, to a point on the theory surface, $\mathbf{t}(\mathbf{c})$.

- Hence, we can understand the Bayesian method as follows: **for each parameter value**, we compute the Σ**-distance** to the **data**. The posterior parameter distribution is the **exponential** of $-1/2$ multiplied by these distances, **scaled by the prior**.



Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method

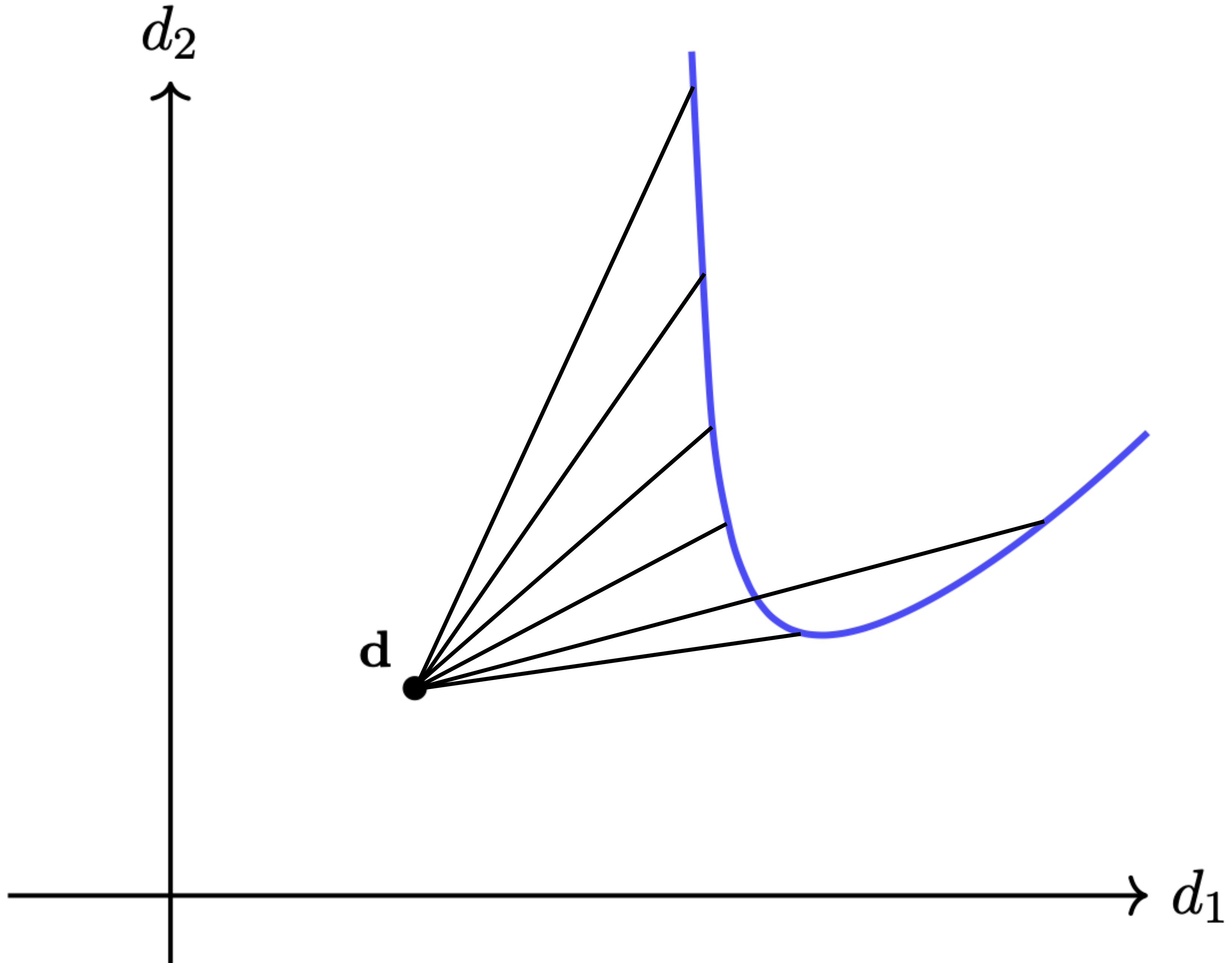
- The parameter distributions are given by **Bayes' theorem**:

$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c})p(\mathbf{d}\,|\,\mathbf{c})$$
$$= \pi(\mathbf{c})\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T\Sigma^{-1}(\mathbf{d}-\mathbf{t}(\mathbf{c}))\right),$$

where $\pi(\mathbf{c})$ is some **prior distribution**, and we use the fact that the experimental data is **normally distributed**.

- The **exponent** (in pink) can be understood as the distance (with respect to the metric Σ) from the data, $\mathbf{d}$, to a point on the theory surface, $\mathbf{t}(\mathbf{c})$.

- Hence, we can understand the Bayesian method as follows: **for each parameter value**, we compute the Σ**-distance** to the **data**. The posterior parameter distribution is the **exponential** of $-1/2$ multiplied by these distances, **scaled by the prior**.



Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

MC replica method

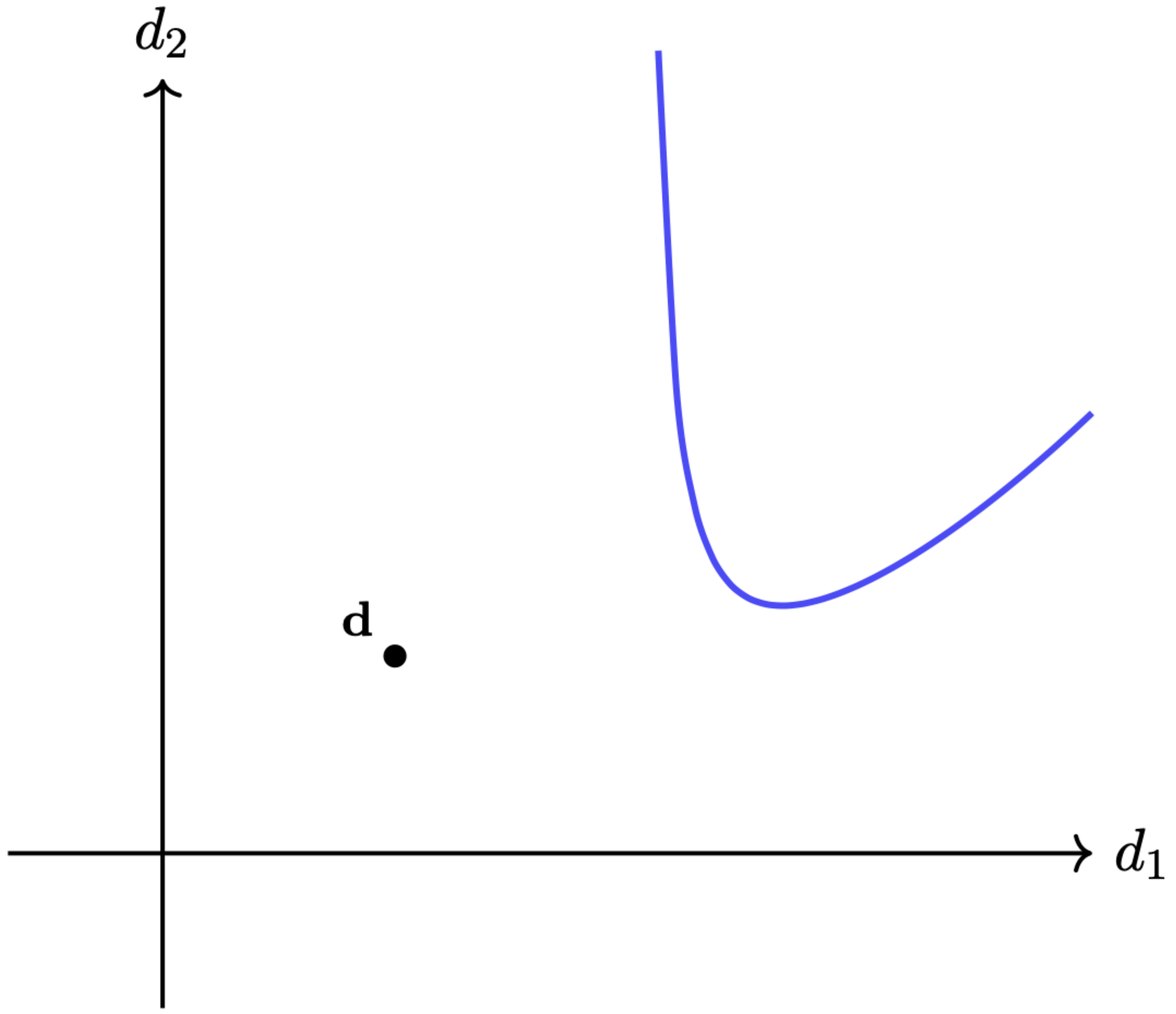
- The MC replica method is instead based on the principle of **fitting pseudodata.**

Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

MC replica method

- The MC replica method is instead based on the principle of **fitting pseudodata.**
- We **throw random pseudodata points** about the experimental data $\mathbf{d}$, according to a multivariable normal distribution centred on the experimental data $\mathbf{d}$, and with the experimental covariance Σ.

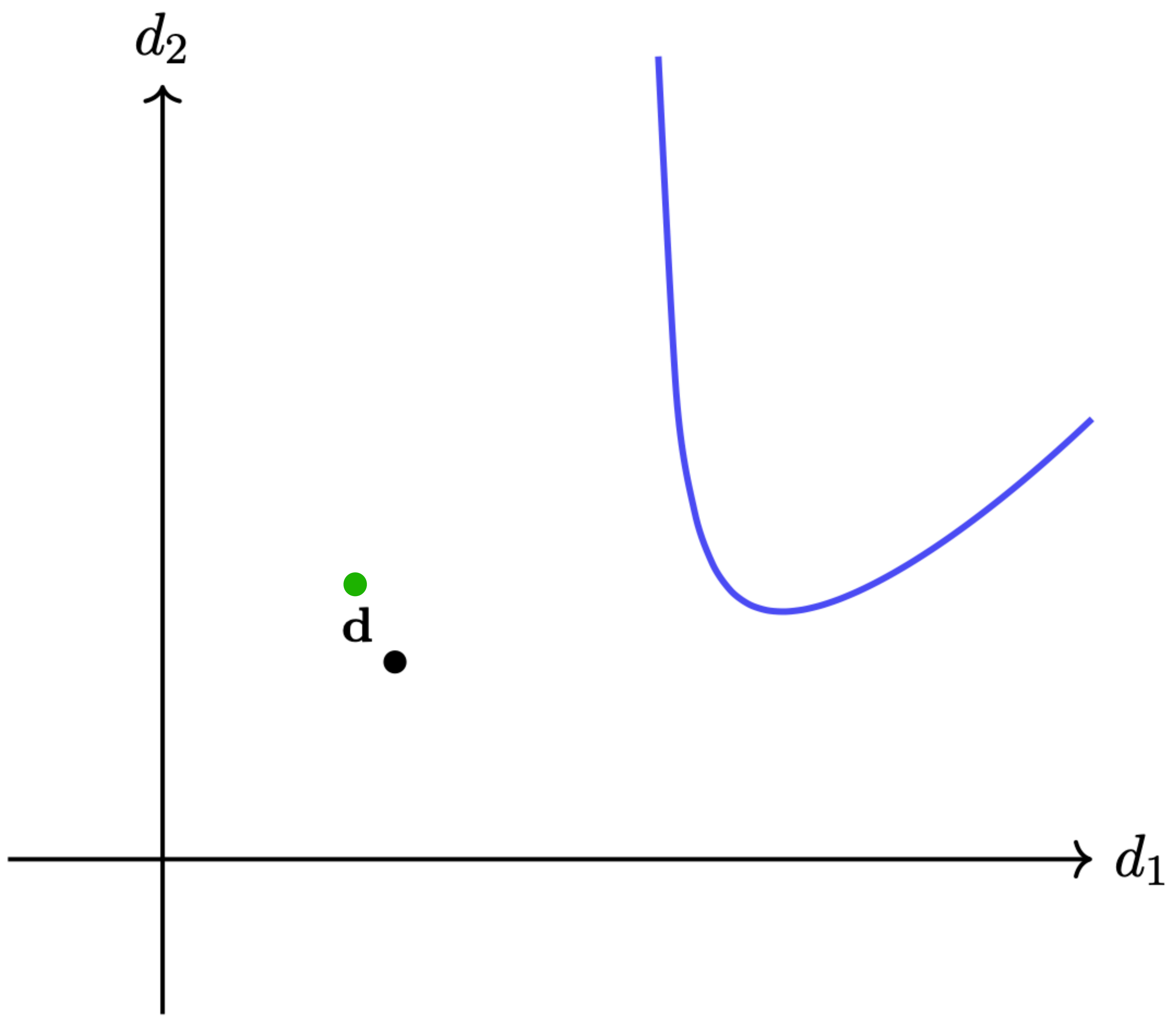


Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

MC replica method

- The MC replica method is instead based on the principle of **fitting pseudodata.**
- We **throw random pseudodata points** about the experimental data $\mathbf{d}$, according to a multivariable normal distribution centred on the experimental data $\mathbf{d}$, and with the experimental covariance Σ.

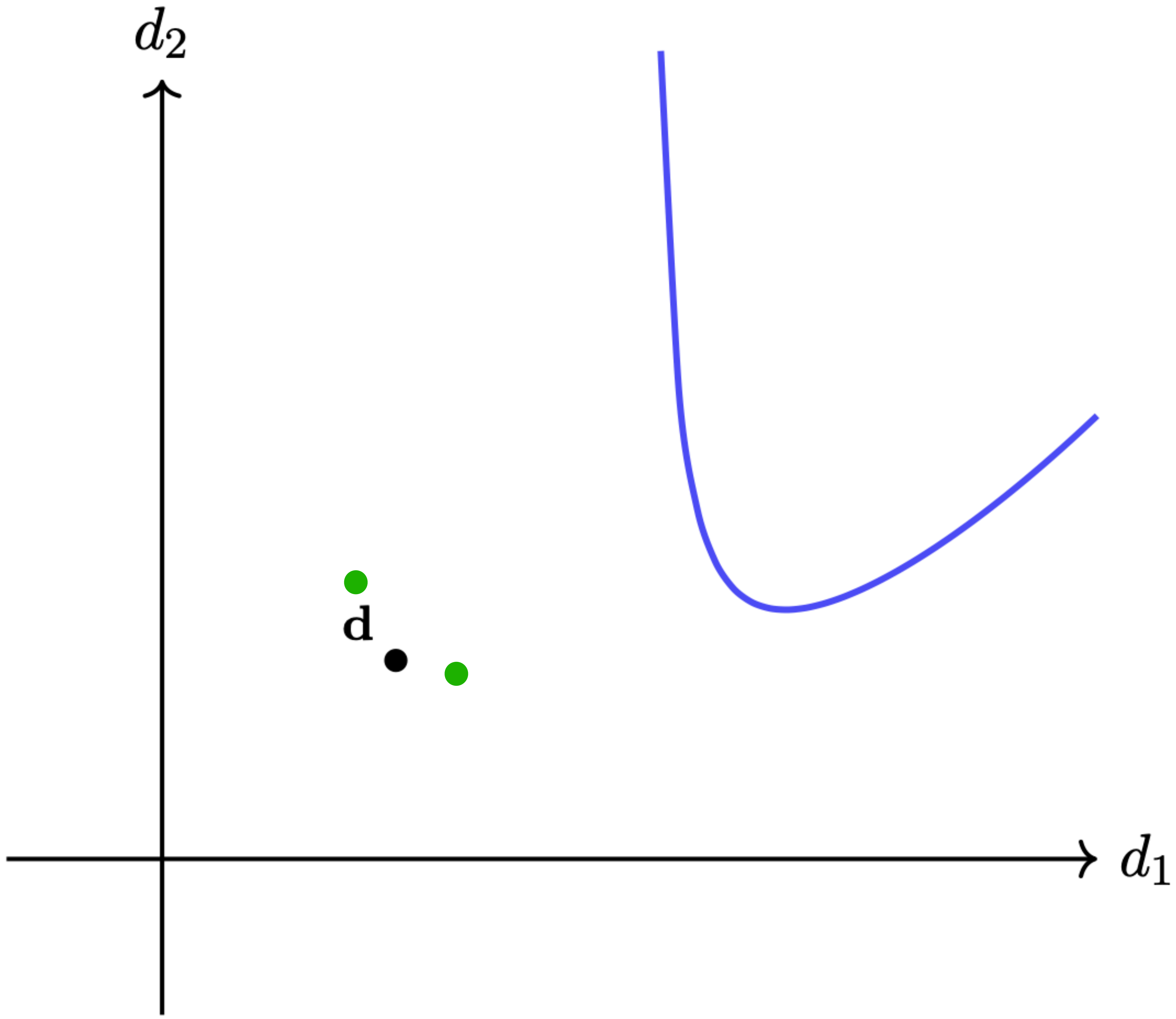


Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

MC replica method

- The MC replica method is instead based on the principle of **fitting pseudodata.**
- We **throw random pseudodata points** about the experimental data $\mathbf{d}$, according to a multivariable normal distribution centred on the experimental data $\mathbf{d}$, and with the experimental covariance Σ.

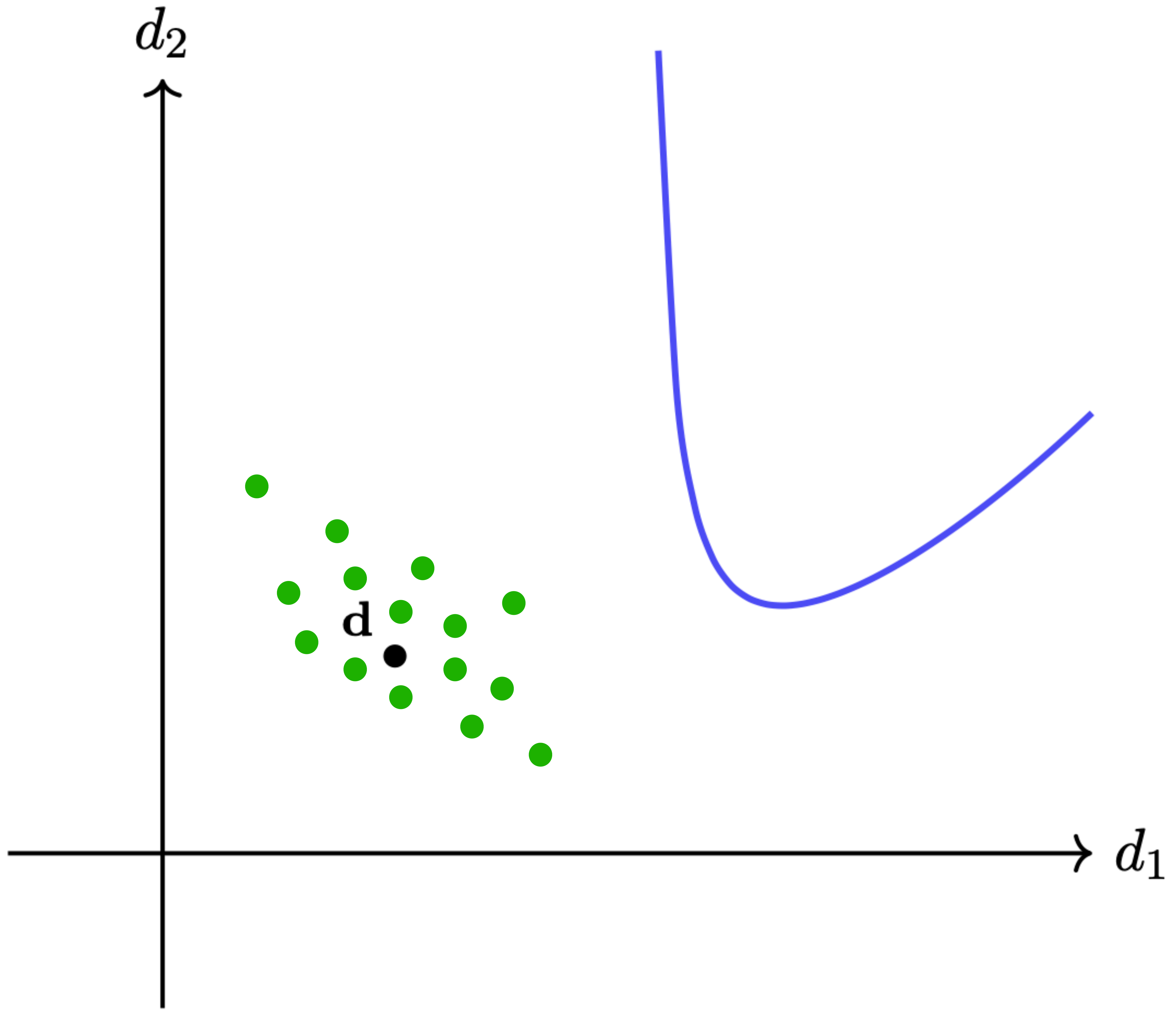


Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

MC replica method

- The MC replica method is instead based on the principle of **fitting pseudodata.**
- We **throw random pseudodata points** about the experimental data $\mathbf{d}$, according to a multivariable normal distribution centred on the experimental data $\mathbf{d}$, and with the experimental covariance Σ.

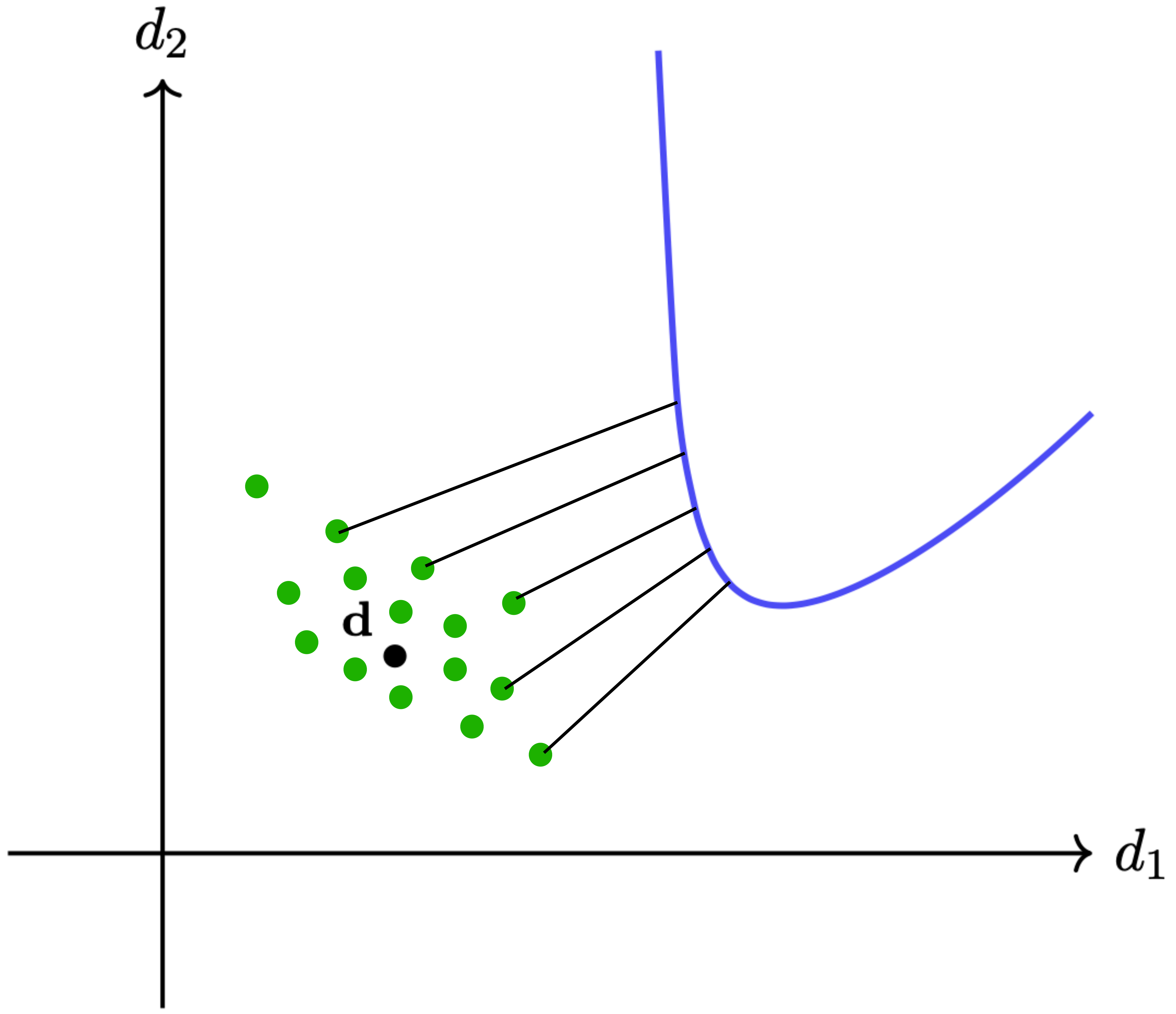


Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

MC replica method

- The MC replica method is instead based on the principle of **fitting pseudodata.**
- We **throw random pseudodata points** about the experimental data $\mathbf{d}$, according to a multivariable normal distribution centred on the experimental data $\mathbf{d}$, and with the experimental covariance Σ.
- For each pseudodata point, we **minimise** - we compute the **closest point** on the theory surface (in the Σ-distance), and thus obtain **associated parameter values**.

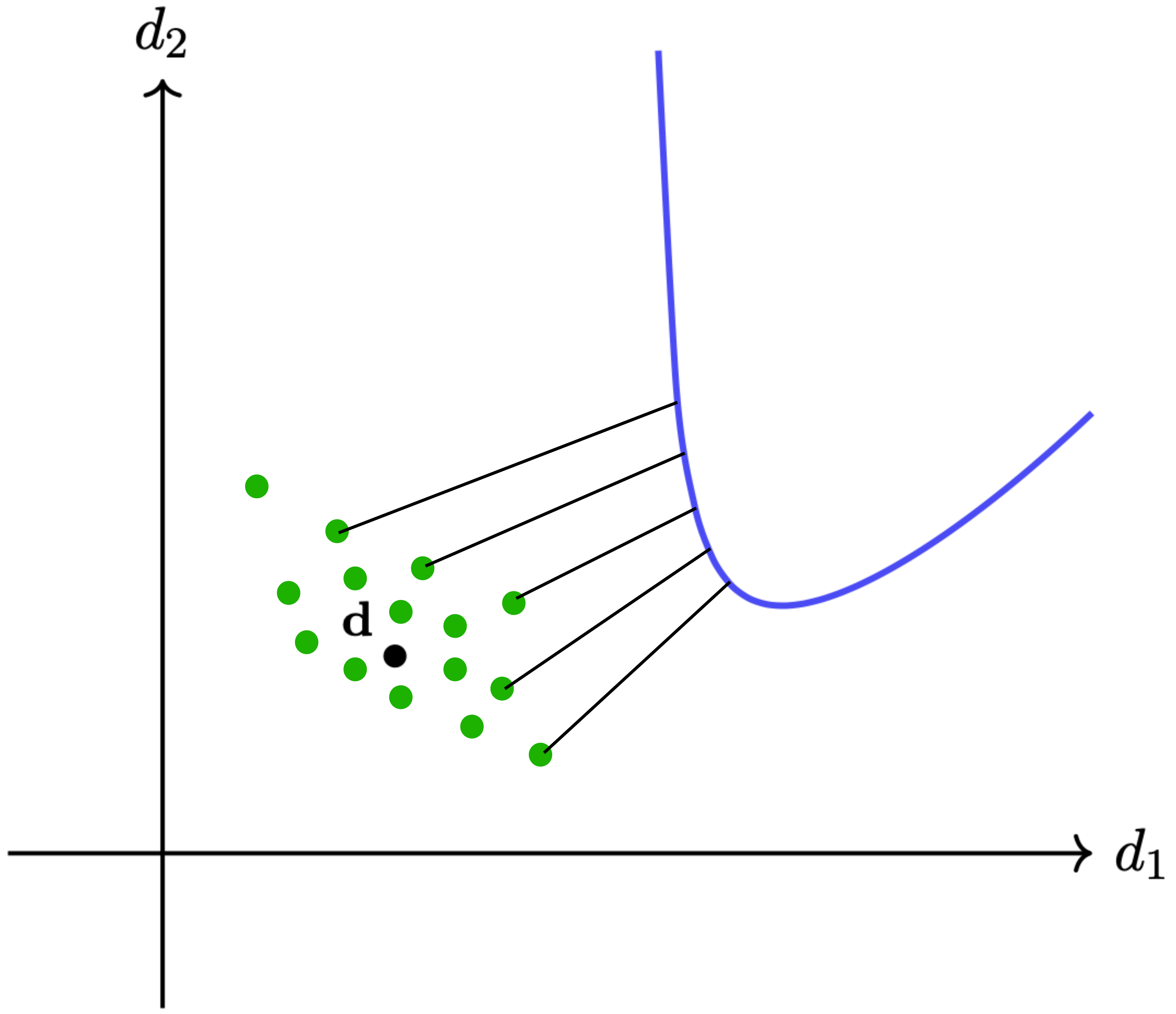


Bayesian vs MC analyses

- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

MC replica method

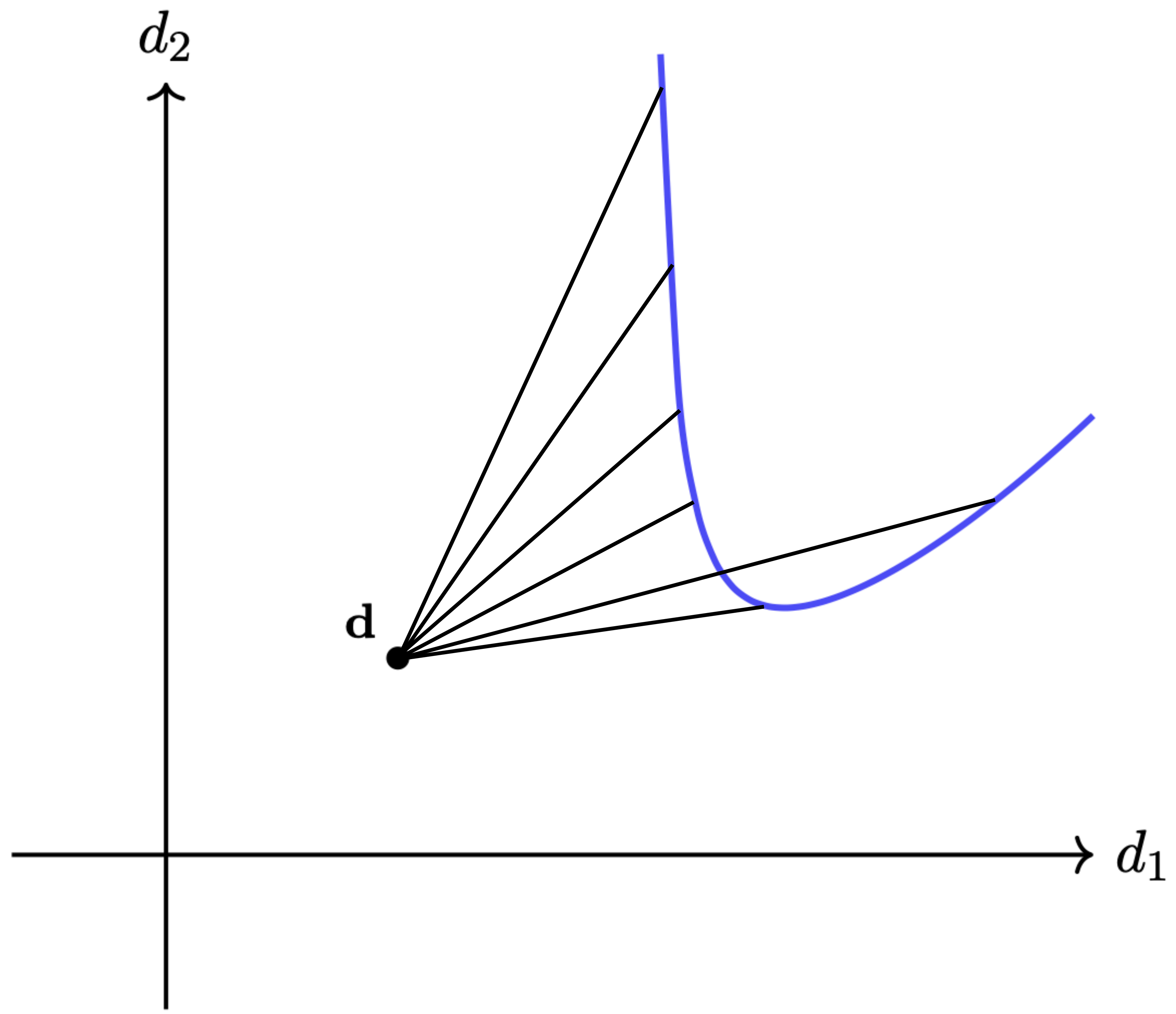
- The MC replica method is instead based on the principle of **fitting pseudodata.**
- We **throw random pseudodata points** about the experimental data $\mathbf{d}$, according to a multivariable normal distribution centred on the experimental data $\mathbf{d}$, and with the experimental covariance Σ.
- For each pseudodata point, we **minimise** - we compute the **closest point** on the theory surface (in the Σ-distance), and thus obtain **associated parameter values**.
- Repeating for **large amounts of pseudodata** gives an **approximation** to the **parameter distributions**.



Bayesian vs MC analyses

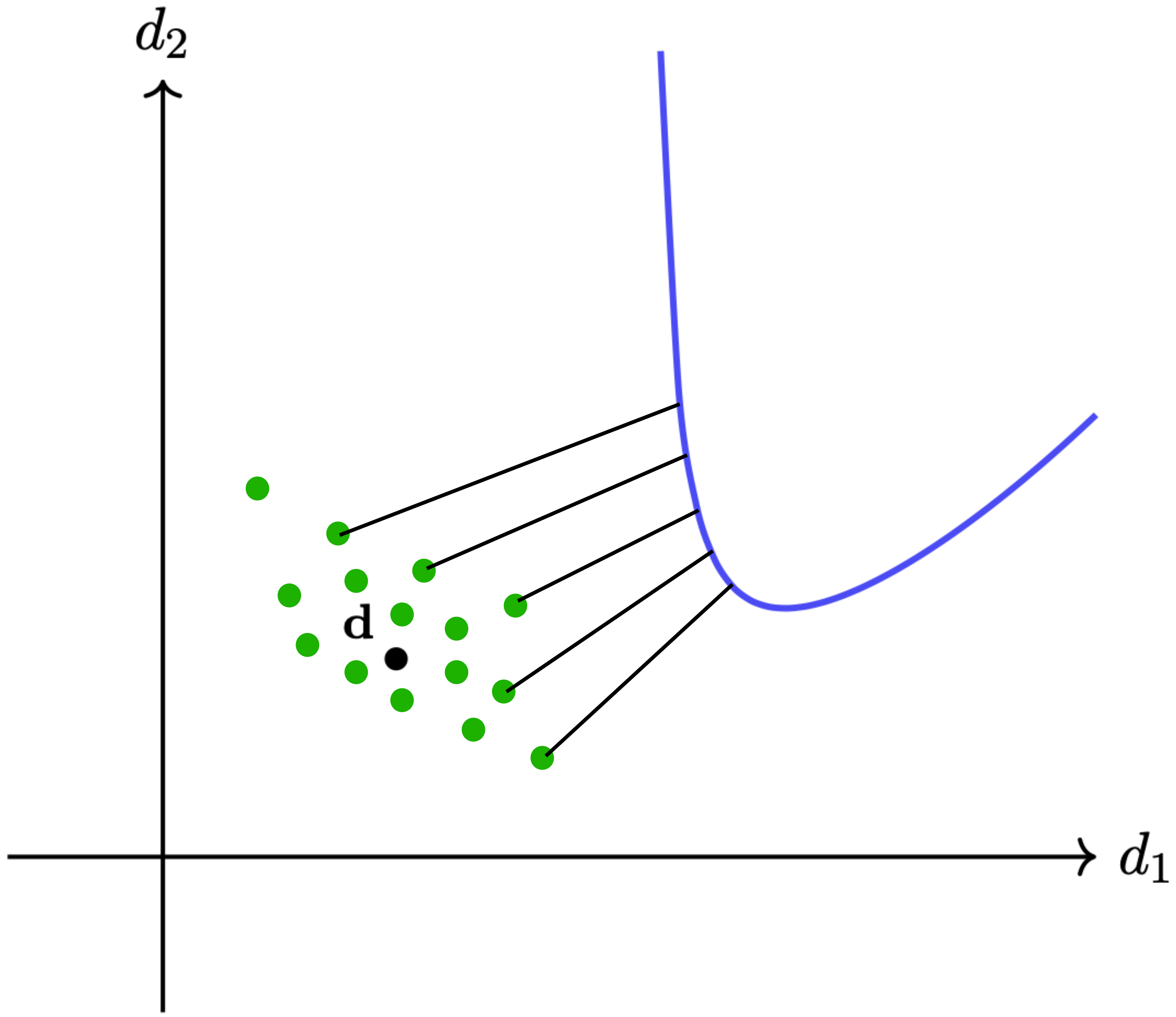
- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method



$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c}) \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T \Sigma^{-1} (\mathbf{d}-\mathbf{t}(\mathbf{c}))\right)$$

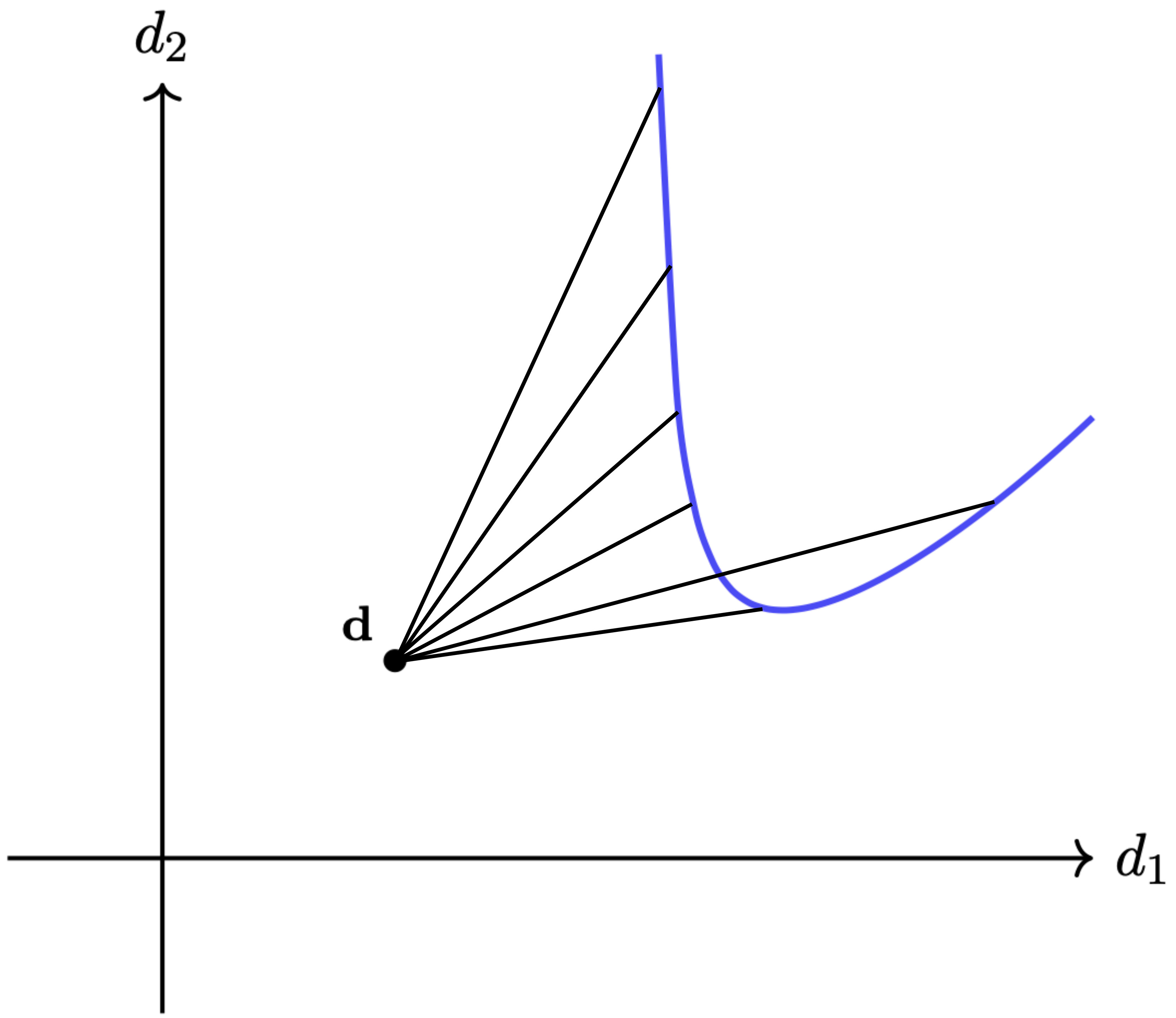
MC replica method



Bayesian vs MC analyses

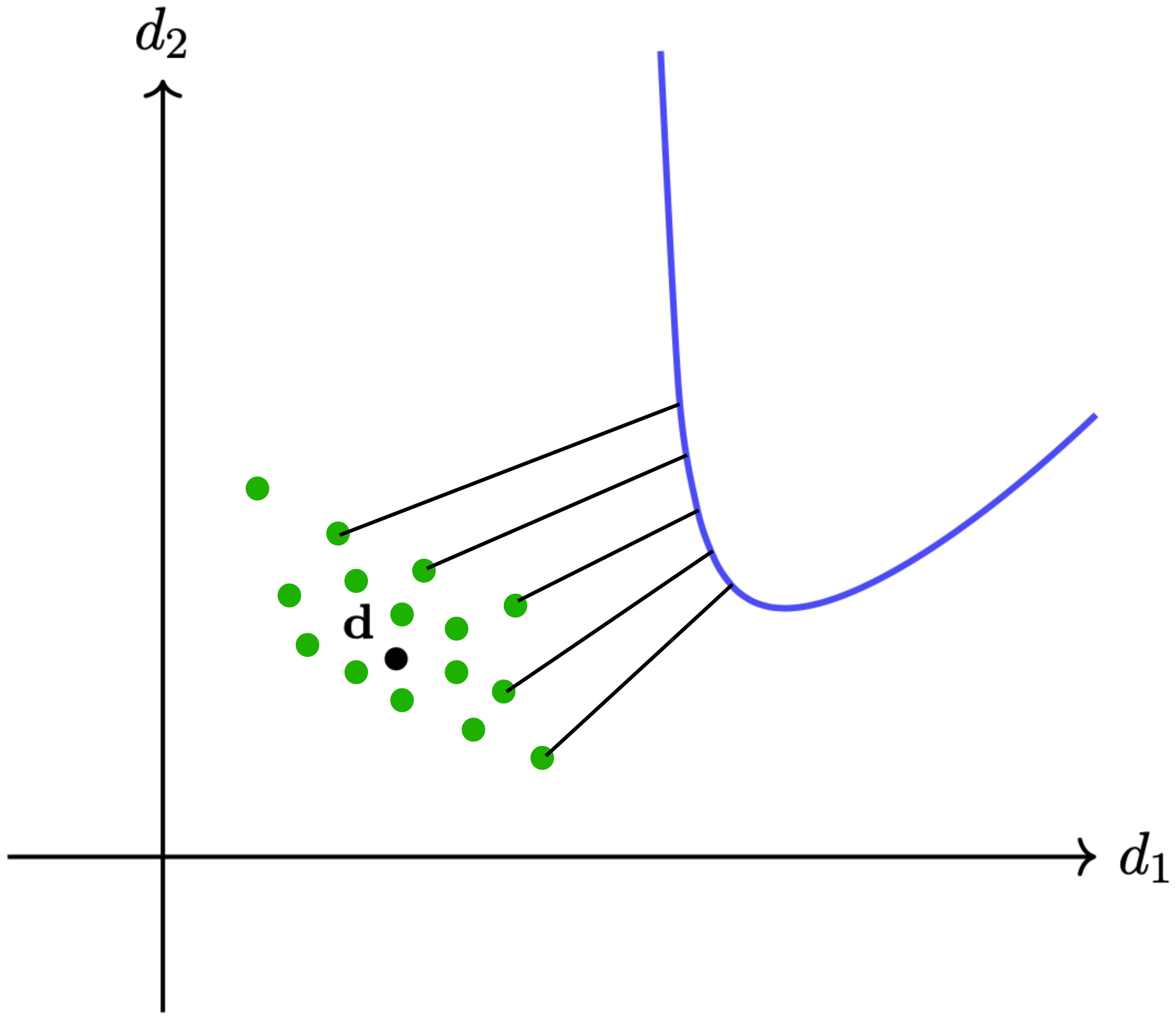
- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method



$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c}) \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T \Sigma^{-1} (\mathbf{d}-\mathbf{t}(\mathbf{c}))\right)$$

MC replica method

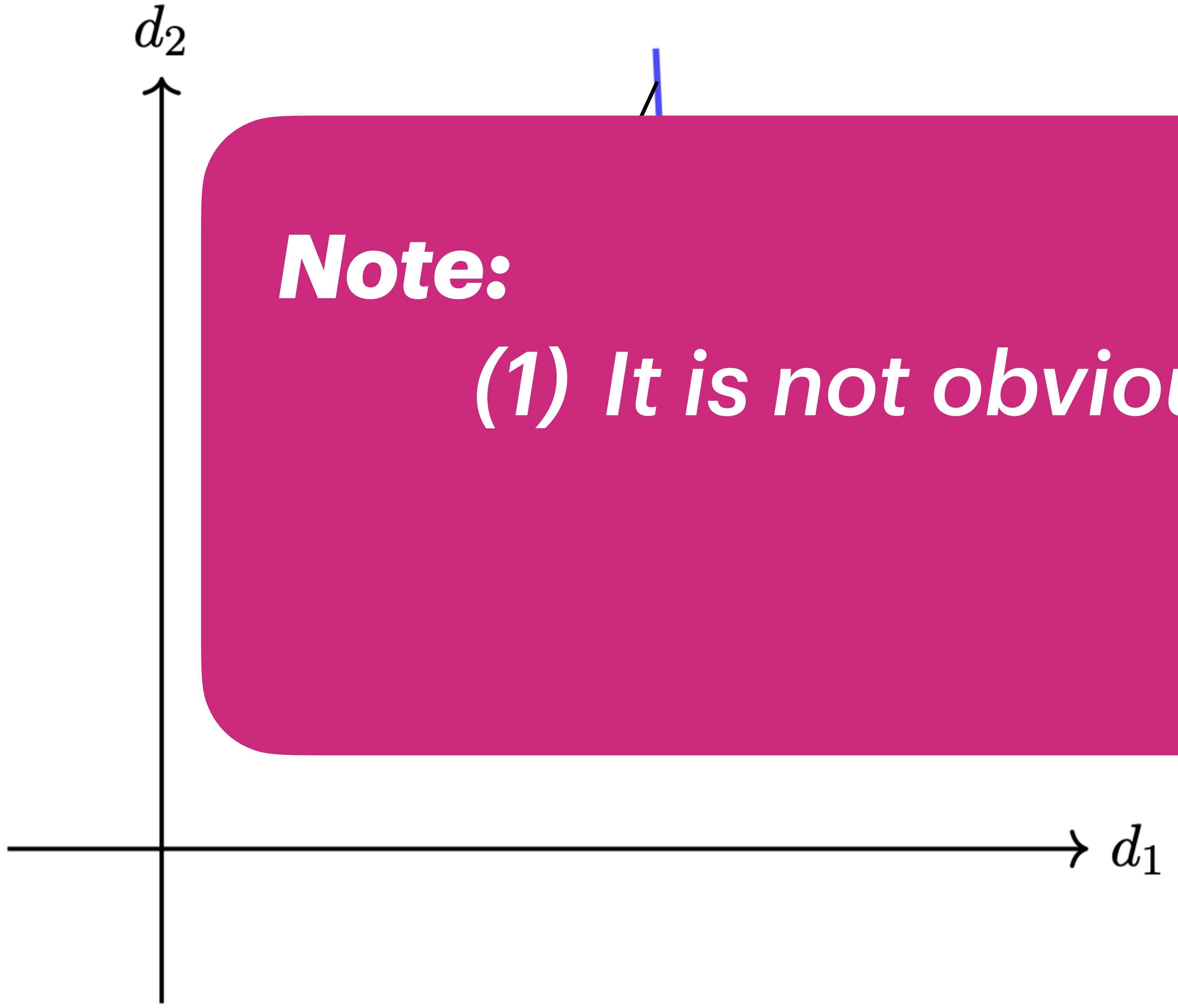


$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \ldots ?$$

Bayesian vs MC analyses

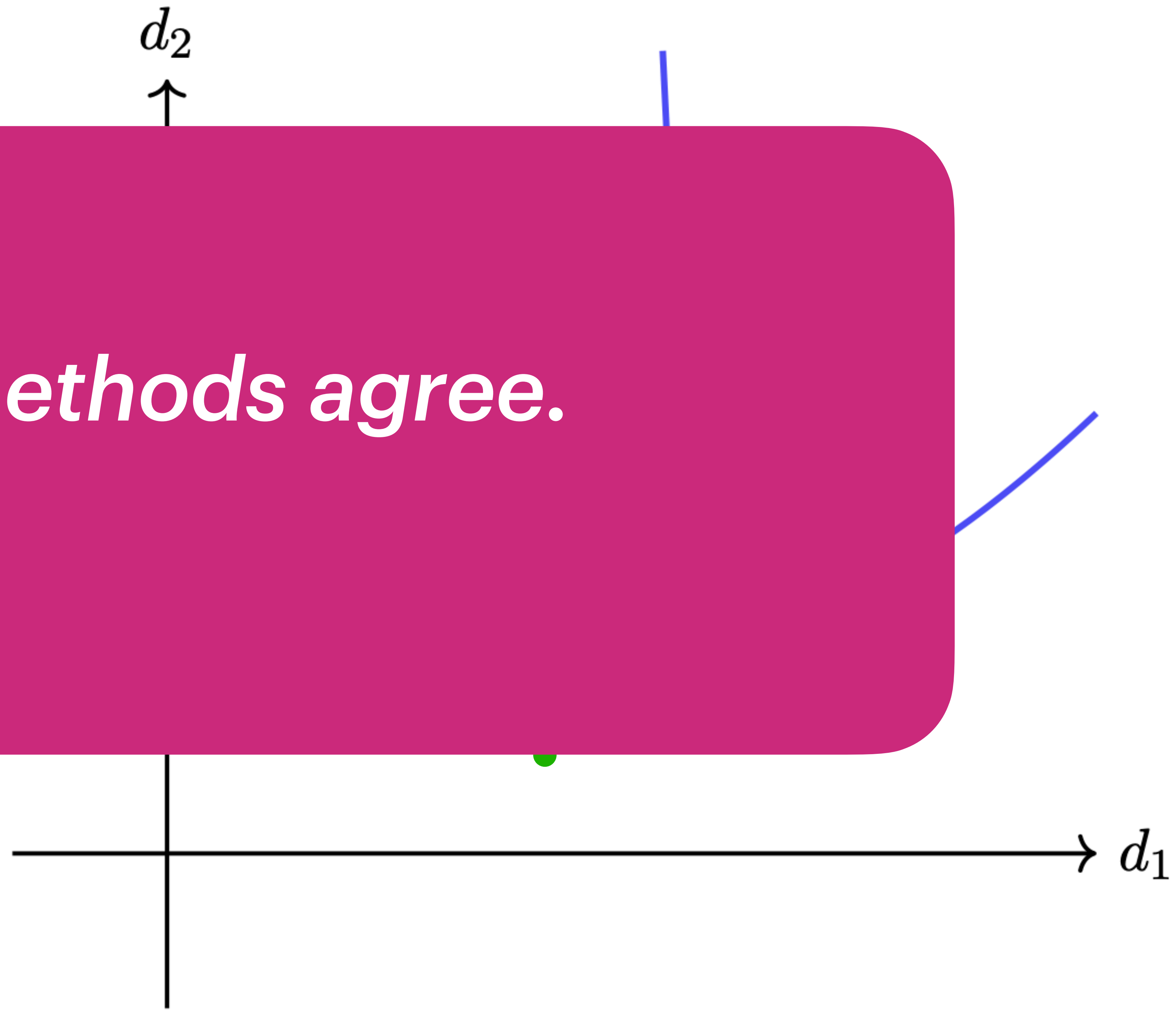
- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method



$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c})\exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T\Sigma^{-1}(\mathbf{d}-\mathbf{t}(\mathbf{c}))\right)$$

MC replica method



$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \ldots ?$$

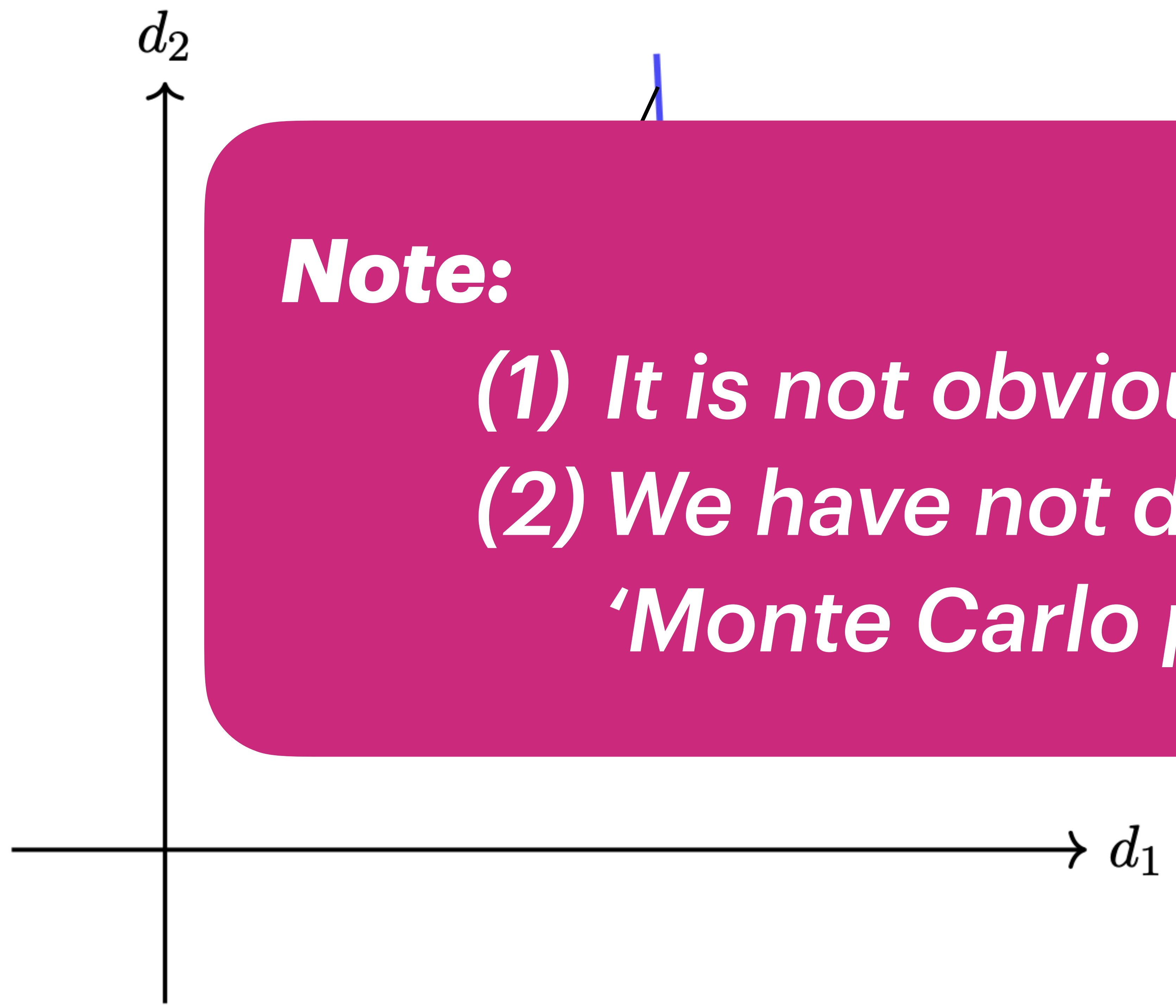
Note:

(1) It is not obvious that these methods agree.

Bayesian vs MC analyses

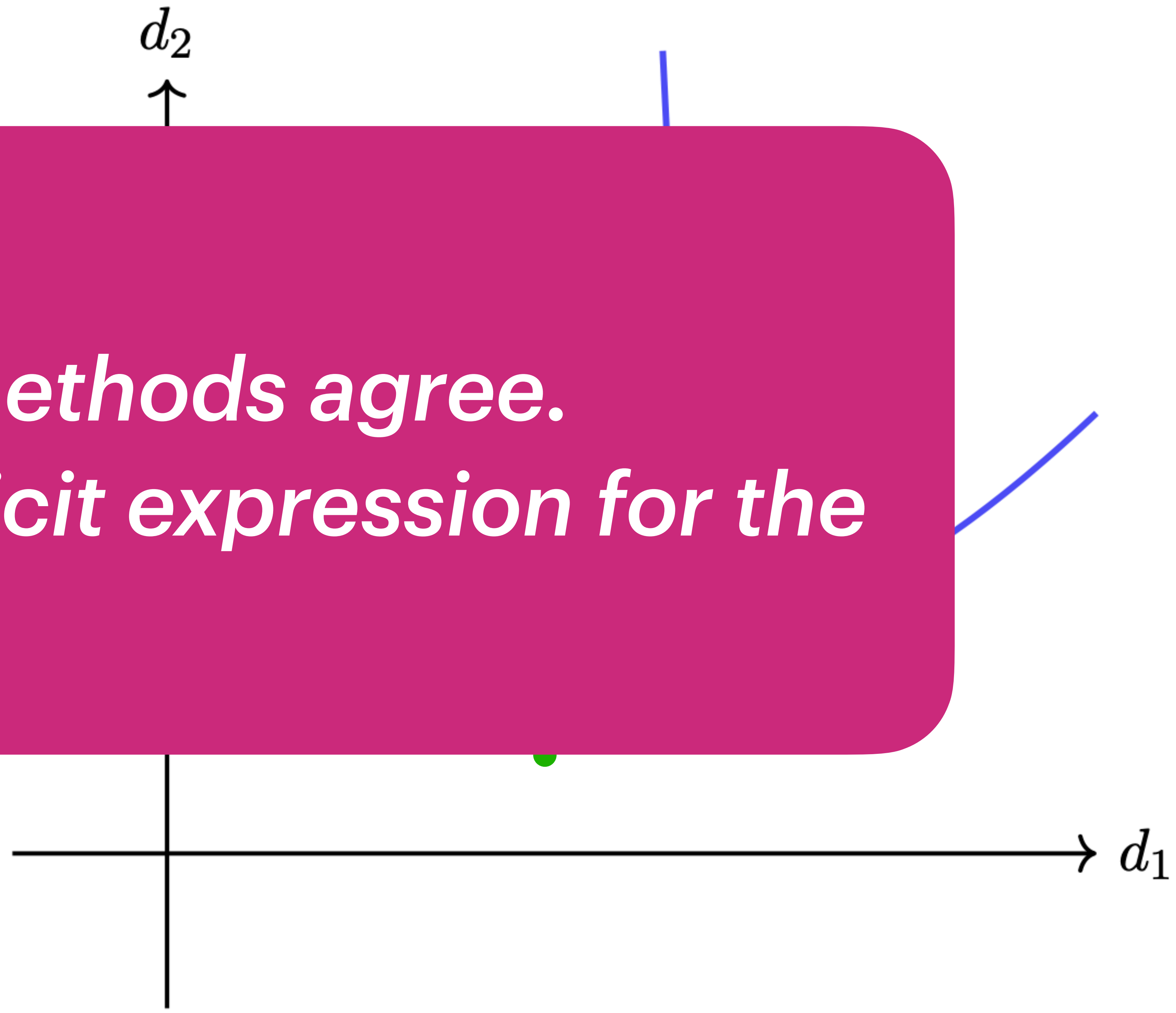
- Now, we can present the **Bayesian** and **Monte Carlo** methods **geometrically**.

Bayesian method



$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \pi(\mathbf{c}) \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{t}(\mathbf{c}))^T \Sigma^{-1} (\mathbf{d}-\mathbf{t}(\mathbf{c}))\right)$$

MC replica method



$$p(\mathbf{c}\,|\,\mathbf{d}) \propto \ldots ?$$

Note:

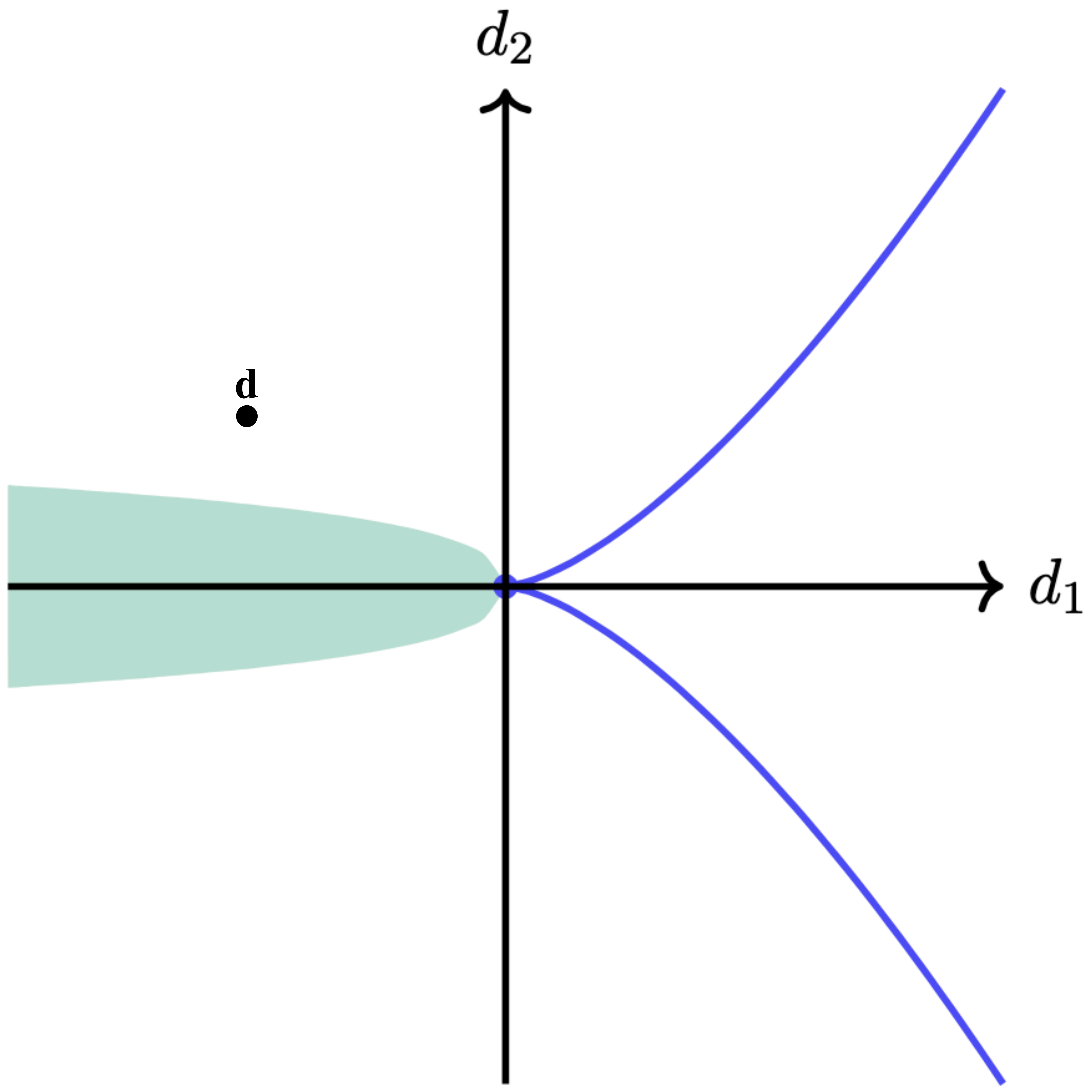
(1) It is not obvious that these methods agree.

(2) We have not derived an explicit expression for the 'Monte Carlo posterior'.

(1) Where might we expect disagreement?

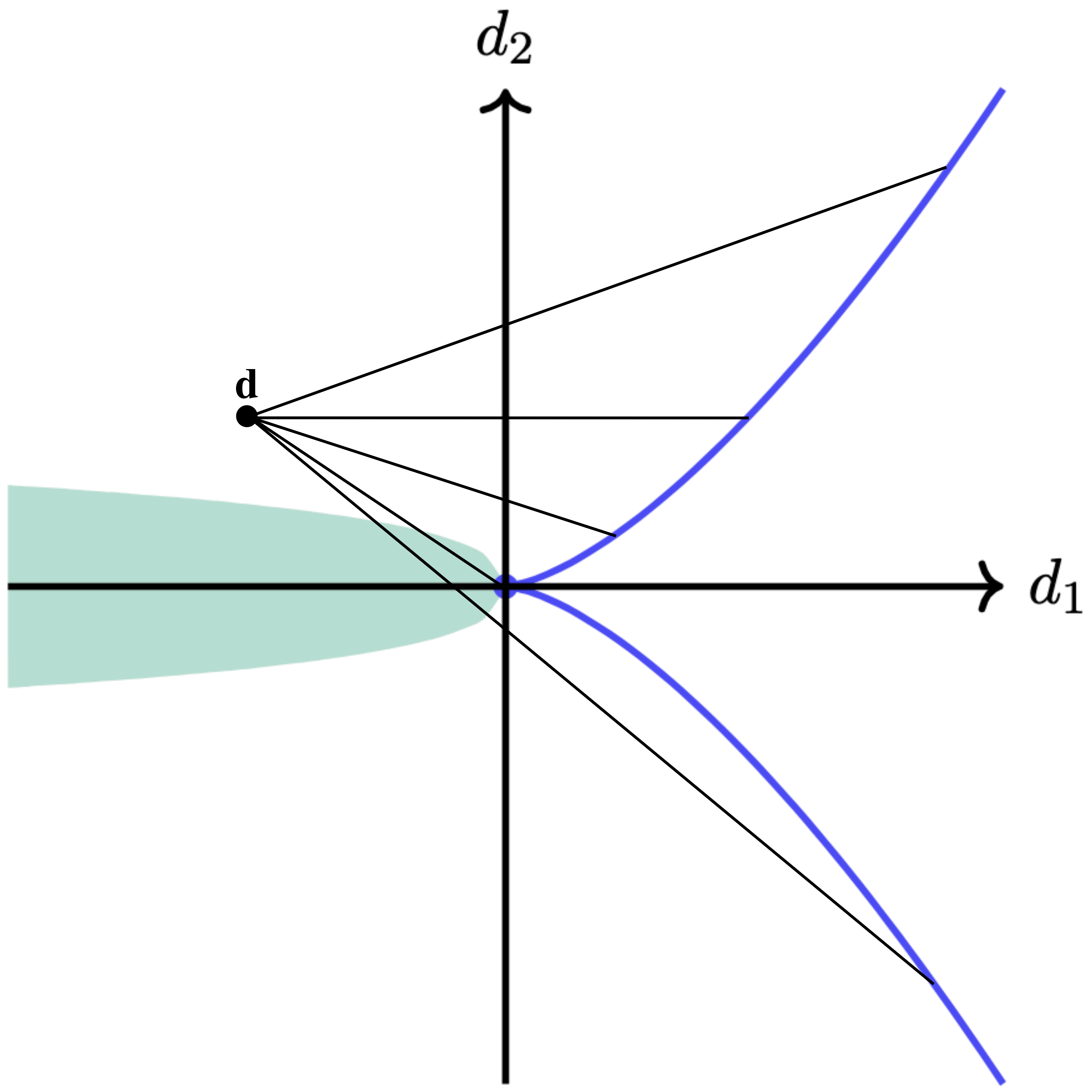
- This geometric understanding helped us to see that we might expect the methods to disagree near a point of **high curvature** on a theory surface.

- On the right, we show the **non-linear theory surface** $\mathbf{t}(c) = (c^2, c^3)^T$ in blue. The green region is the **set of all points** whose **closest point** on the theory surface is the **origin**.



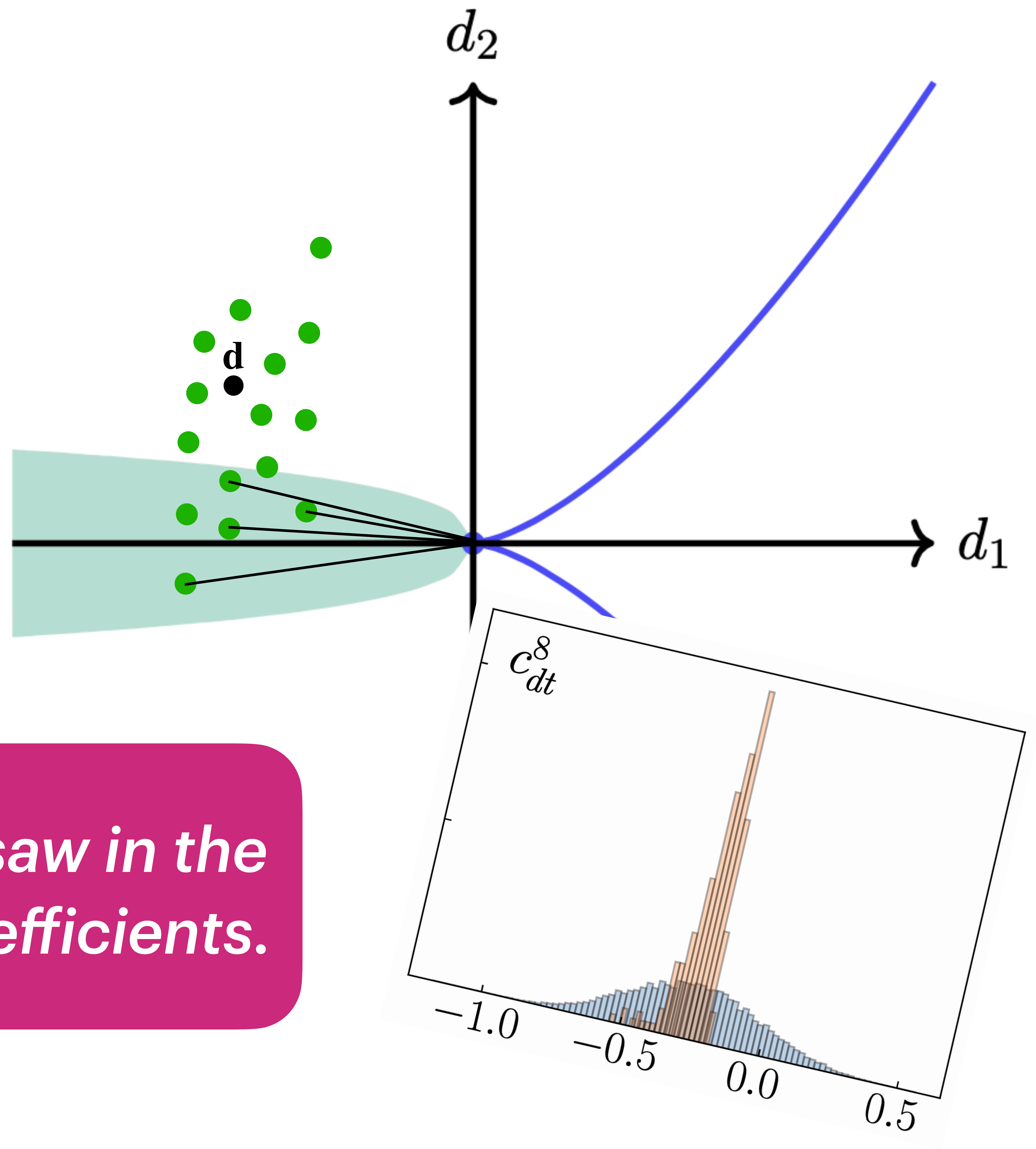
(1) Where might we expect disagreement?

- If we use the **Bayesian method** to analyse this problem, points near the cusp are treated **like any other points**.



(1) Where might we expect disagreement?

- If we use the **Bayesian method** to analyse this problem, points near the cusp are treated **like any other points**.
- On the other hand, if we throw pseudodata near $\mathbf{d}$, some proportion enters the green '**basin of attraction**', and is unfairly **drawn towards the cusp**.



*This is the origin of the '**spiked peaks**' we saw in the Monte Carlo distributions of the Wilson coefficients.*

(2) The Monte Carlo posterior

- Examples such as the one we have just seen are **characteristic of the general behaviour** of the **Monte Carlo posterior**, which was derived **explicitly** in our paper. The maths is hard, and the result is not easy to understand either:

$$
\begin{aligned}
&\exp\left(-\frac{1}{2}\chi^2_{\mathbf{d}_0}(\mathbf{c})\right) \\
&\quad \cdot \int d^{N_\parallel(\mathbf{c})}\mathbf{u}\ \delta\left(\mathbf{c}-\mathbf{f}(\mathbf{u})\right) \int_{\Lambda(\mathbf{c})} d^{N_\perp(\mathbf{c})}\boldsymbol{\lambda} \left|\det\left(\frac{\partial \mathbf{t}}{\partial \mathbf{c}}(\mathbf{f}(\mathbf{u}))\frac{\partial \mathbf{f}}{\partial \mathbf{u}} + \frac{\partial\left(\Sigma M \boldsymbol{\lambda}\right)}{\partial \mathbf{c}}(\mathbf{f}(\mathbf{u}))\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\middle|\Sigma M(\mathbf{f}(\mathbf{u}))\right)\right| \\
&\quad \cdot \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T M(\mathbf{c})^T \Sigma M(\mathbf{c})\boldsymbol{\lambda} + \boldsymbol{\lambda}^T M(\mathbf{c})^T(\mathbf{d}_0 - \mathbf{t}(\mathbf{c}))\right), \qquad (2.17)
\end{aligned}
$$

(2) The Monte Carlo posterior

- Examples such as the one we have just seen are **characteristic of the general behaviour** of the **Monte Carlo posterior**, which was derived **explicitly** in our paper. The maths is hard, and the result is not easy to understand either:

Bayesian result, for uniform flat prior

$$\exp\left(-\frac{1}{2}\chi^2_{\mathbf{d}_0}(\mathbf{c})\right)$$

$$\cdot \int d^{N_\parallel(\mathbf{c})}\mathbf{u}\, \delta\left(\mathbf{c}-\mathbf{f}(\mathbf{u})\right) \int_{\Lambda(\mathbf{c})} d^{N_\perp(\mathbf{c})}\boldsymbol{\lambda} \left|\det\left(\frac{\partial \mathbf{t}}{\partial \mathbf{c}}(\mathbf{f}(\mathbf{u}))\frac{\partial \mathbf{f}}{\partial \mathbf{u}} + \frac{\partial\left(\Sigma M \boldsymbol{\lambda}\right)}{\partial \mathbf{c}}(\mathbf{f}(\mathbf{u}))\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\middle|\Sigma M(\mathbf{f}(\mathbf{u}))\right)\right|$$

$$\cdot \exp\left(-\frac{1}{2}\boldsymbol{\lambda}^T M(\mathbf{c})^T \Sigma M(\mathbf{c})\boldsymbol{\lambda} + \boldsymbol{\lambda}^T M(\mathbf{c})^T(\mathbf{d}_0 - \mathbf{t}(\mathbf{c}))\right), \qquad (2.17)$$

Jacobian 'transformation factor' which is enhanced in regions of high curvature

delta functions, which cannot be integrated out in case of singular points, like cusp

(Read the paper for the full, careful derivation: https://arxiv.org/pdf/2404.10056.)

(2) The Monte Carlo posterior

- Examples such as the one we have just seen are **characteristic of the general behaviour** of the **Monte Carlo posterior**, which was derived **explicitly** in our paper. The maths is hard, and the result is not easy to understand either:

Bayesian result, for
unifor

Key takeaway from (1) and (2):
As compared to a Bayesian method, the Monte Carlo replica method unfairly favours regions of ***high curvature****. In particular, its validity is* ***not guaranteed in non-linear models.***

delta
which
integr
case of singular
points, like cusp

3. - Relevance in PDF fits

PDF fitting is *non-linear*

- We originally saw the issue with the Monte Carlo replica method in the context of **quadratic SMEFT fits**, where we **wrongly concluded** the existence of New Physics because of our methodology for uncertainty propagation.

PDF fitting is *non-linear*

- We originally saw the issue with the Monte Carlo replica method in the context of **quadratic SMEFT fits**, where we **wrongly concluded** the existence of New Physics because of our methodology for uncertainty propagation.

- In PDF fitting, a large proportion of the data is **linear in the PDFs** (namely deep inelastic scattering data), but a growing proportion is **quadratic in the PDFs** (namely the proton-proton collision data). Further, it **contains no linear term**, so effectively has the '**highest curvature**' in the geometrical language we have been using.

PDF fitting is *non-linear*

- We originally saw the issue with the Monte Carlo replica method in the context of **quadratic SMEFT fits**, where we **wrongly concluded** the existence of New Physics because of our methodology for uncertainty propagation.

- In PDF fitting, a large proportion of the data is **linear in the PDFs** (namely deep inelastic scattering data), but a growing proportion is **quadratic in the PDFs** (namely the proton-proton collision data). Further, it **contains no linear term**, so effectively has the '**highest curvature**' in the geometrical language we have been using.

- It is natural to ask: ***could the use of the Monte Carlo method result in incorrect conclusions about PDF uncertainties, as we add more proton-proton data?***

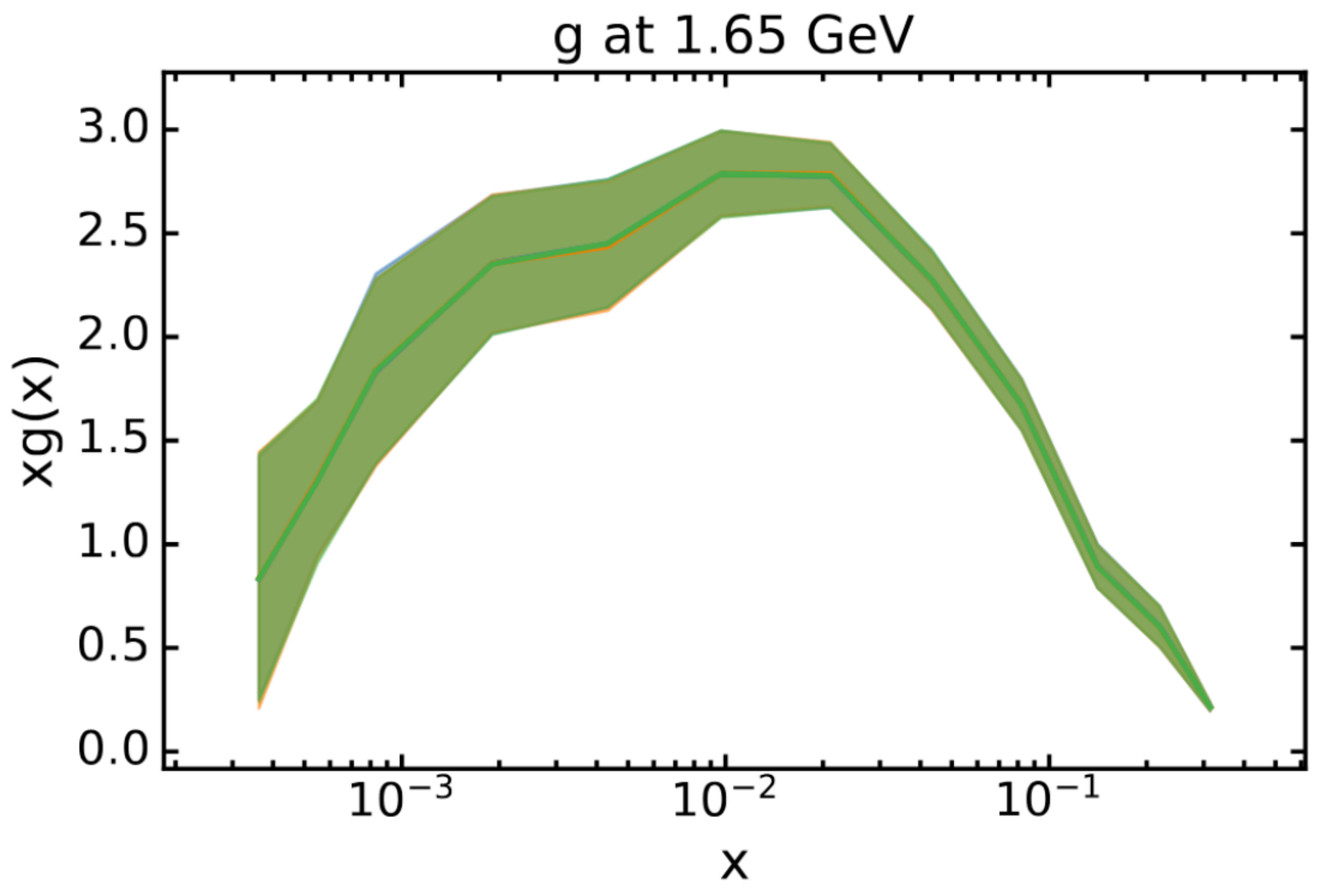
Fitting ‘kinky’ PDFs

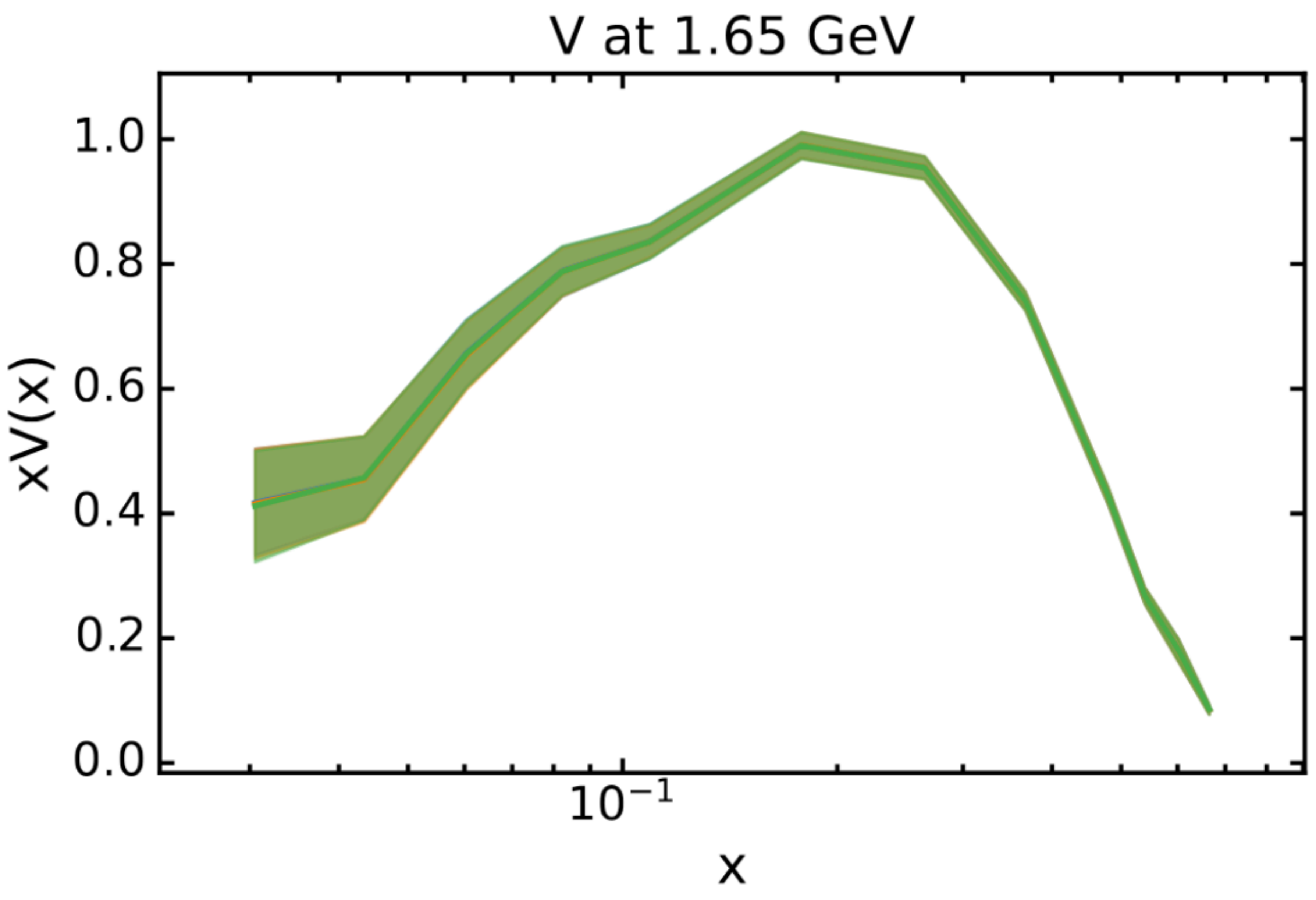
- Since **Bayesian fits** (through methods like **Nested Sampling**) suffer from the curse of dimensionality, we decided to investigate the effect on PDF fits in a **toy scenario**.

Fitting 'kinky' PDFs

- Since **Bayesian fits** (through methods like **Nested Sampling**) suffer from the curse of dimensionality, we decided to investigate the effect on PDF fits in a **toy scenario**.

- We consider the proton in terms of just the **singlet**, **gluon** and **valence PDFs**. Further, we parametrise each of the PDFs as **linear interpolants** on an x-grid comprising **12 grid points for each flavour**, hence 36 grid points in total. Examples of our 'kinky' PDFs are shown on the right.





Fitting 'kinky' PDFs

- We generated an **artificial copy of the complete NNPDF4.0 dataset***, with noise, based on a kinky PDF with values at the grid points taken from the NNPDF4.0 central PDF.

(* excluding jets for technical reasons)

Fitting ‘kinky’ PDFs

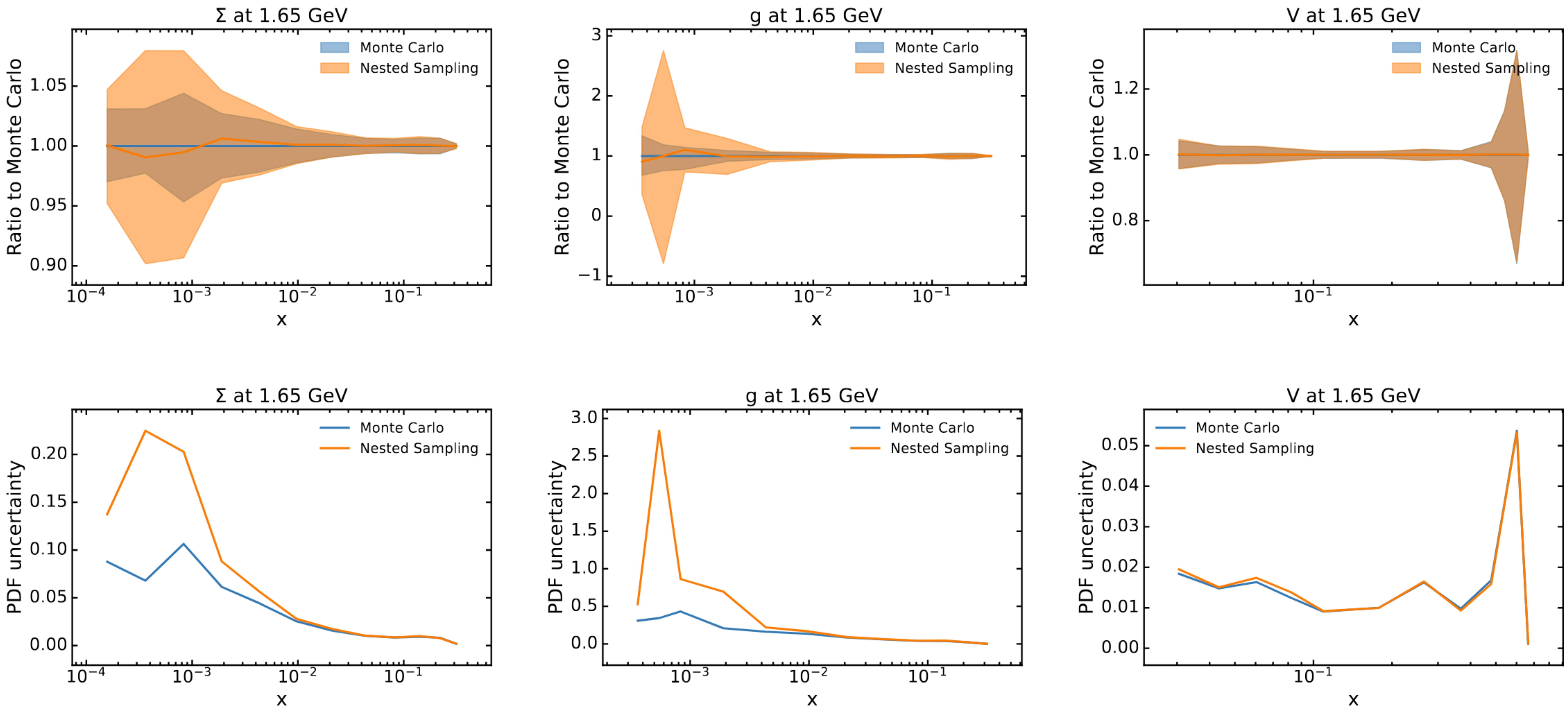
- We generated an **artificial copy of the complete NNPDF4.0 dataset***, with noise, based on a kinky PDF with values at the grid points taken from the NNPDF4.0 central PDF.

- We then fit the **dataset**, using a **fully Bayesian methodology** (with a large uniform prior), and using the **Monte Carlo replica methodology.**

(* excluding jets for technical reasons)

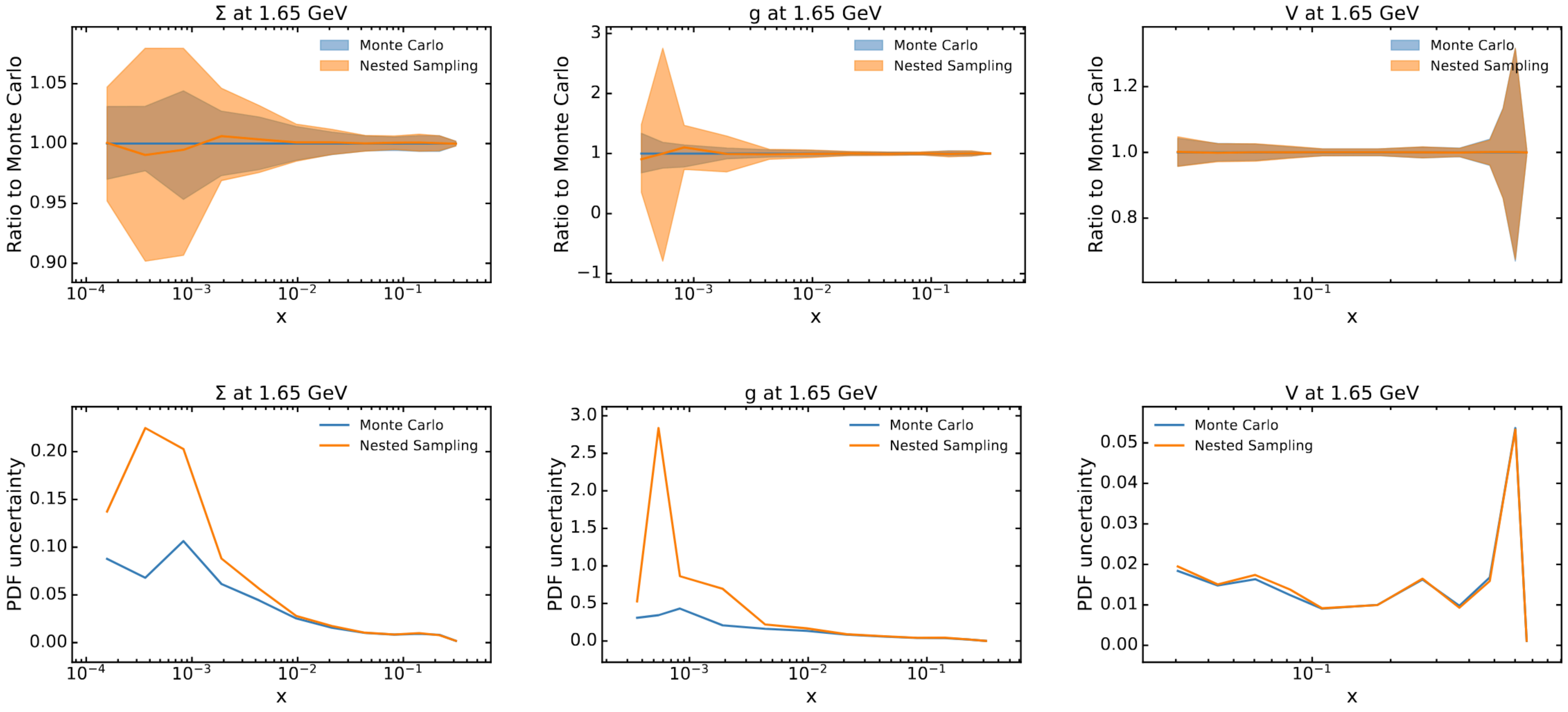
Results

- We found that, in our toy scenario, Monte Carlo does seem to **underestimate** errors relative to a fully Bayesian methodology.



Results

- Particularly at **lower x-values**, we see a reduction in uncertainties of **up to 60%** in the gluon PDF, when using Monte Carlo.



Takeaway

- However, it is important to note:

Takeaway

- However, it is important to note:
  - This is a **toy scenario** using **artificial pseudodata**. We have **not** performed a realistic PDF fit.

Takeaway

- However, it is important to note:
  - This is a **toy scenario** using **artificial pseudodata**. We have **not** performed a realistic PDF fit.
  - It is, a priori, **difficult to understand** the **weighting** assigned to the high curvature regions by the Monte Carlo posterior. It is therefore **difficult to know in advance** what the discrepancy between the Monte Carlo method and Bayesian method will be **before running the fit**.

Takeaway

- However, it is important to note:
  - This is a **toy scenario** using **artificial pseudodata**. We have **not** performed a realistic PDF fit.
  - It is, a priori, **difficult to understand** the **weighting** assigned to the high curvature regions by the Monte Carlo posterior. It is therefore **difficult to know in advance** what the discrepancy between the Monte Carlo method and Bayesian method will be **before running the fit**.

Hence: existing PDF fits are not invalidated by this study, it merely suggests a clear and present need for a future fully Bayesian PDF analysis.

Conclusions

Conclusions

- The **Monte Carlo replica method** can **unfairly bias** regions of **high curvature** in a given theoretical model.

Conclusions

- The **Monte Carlo replica method** can **unfairly bias** regions of **high curvature** in a given theoretical model.

- This can lead to **false conclusions** (originally identified in the SMEFT) in **inference problems.**

Conclusions

- The **Monte Carlo replica method** can **unfairly bias** regions of **high curvature** in a given theoretical model.

- This can lead to **false conclusions** (originally identified in the SMEFT) in **inference problems.**

- In a **toy PDF fit**, we showed that the Monte Carlo and Bayesian approaches **disagree**, with the Monte Carlo method **underestimating** uncertainties. Hence, there is a **clear and present need** to produce a **realistic, fully Bayesian PDF fit** in the near future.

Thanks for listening!
Questions?