Likelihood and Bayesian inference

Michael Matschiner
Ludwig Maximilian University of Munich

Inference methods

1. Maximum likelihood
2. Bayesian inference
3. Simulation-based inference
4. Machine learning

Georgia Tsambos,
Martin Petr

Andrew Kern

Likelihood

Probability

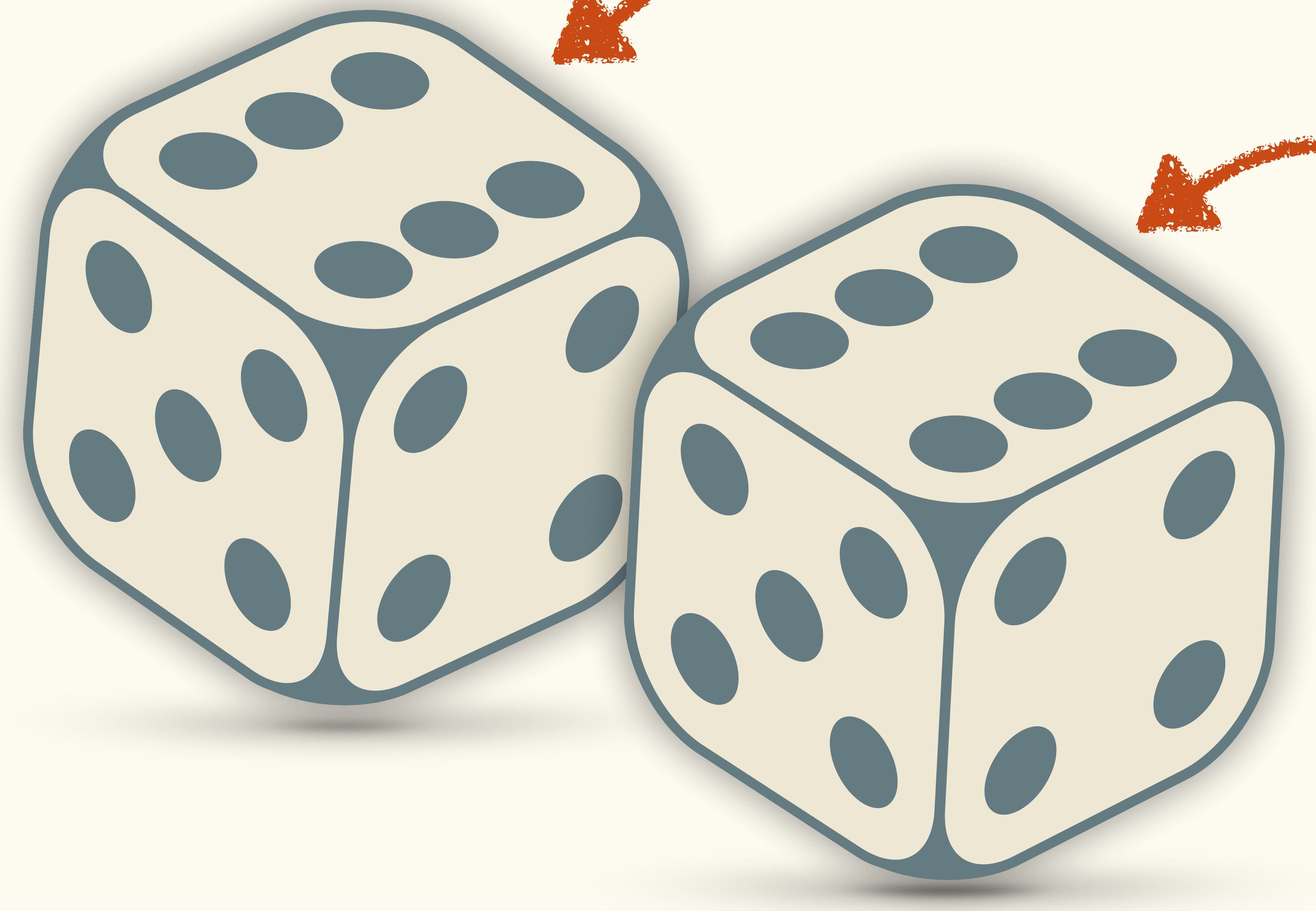
Probability

$P = {}^{1}/_{6}$

Probability

$P(\text{⚅}) = {}^{1}/_{6}$

Probability

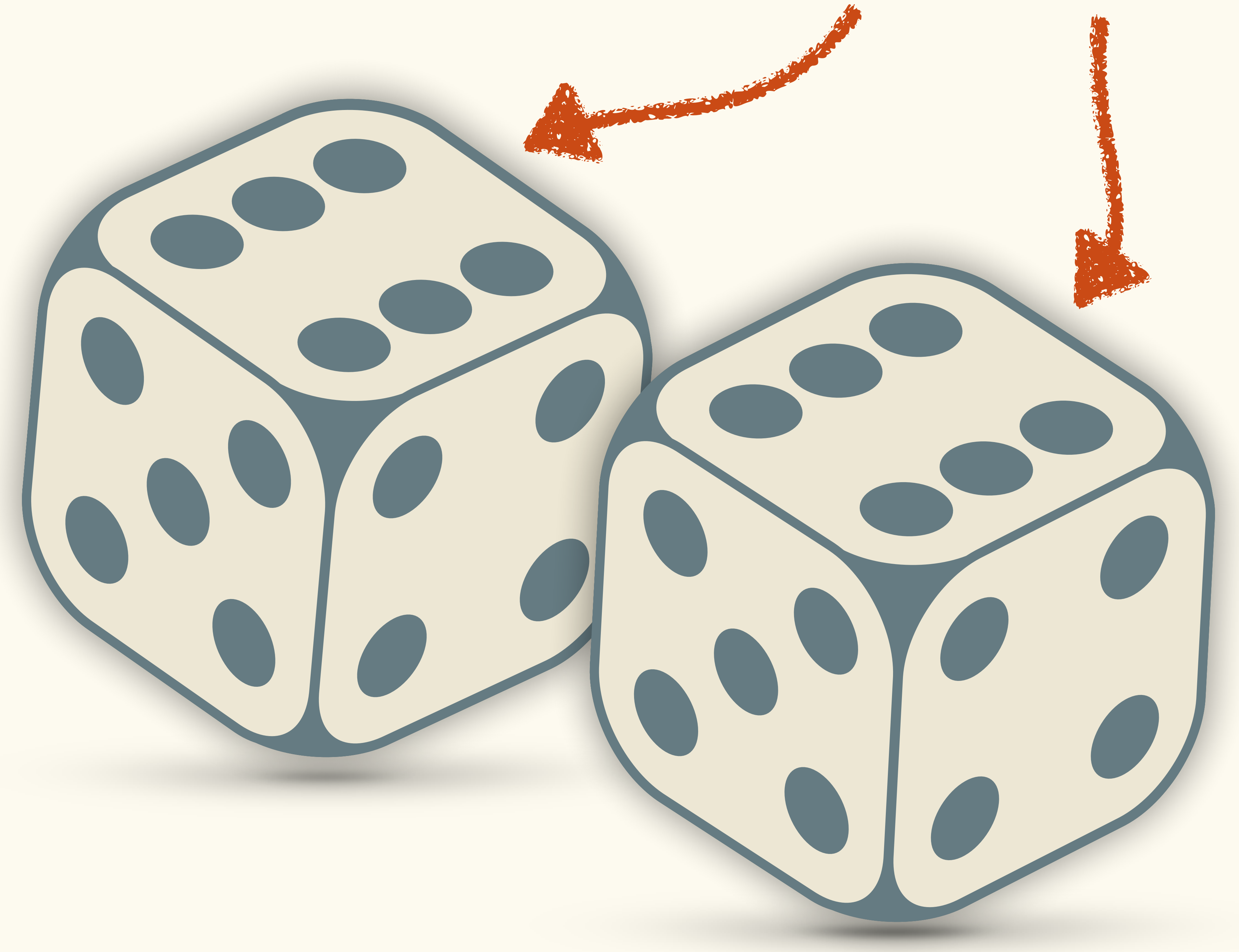


$P(\text{⚅}) = 1/6$

$P(\text{⚅}) = 1/6$

Probability

P(⚅ & ⚅) $= {}^{1}/_{6} \times {}^{1}/_{6} = {}^{1}/_{36}$



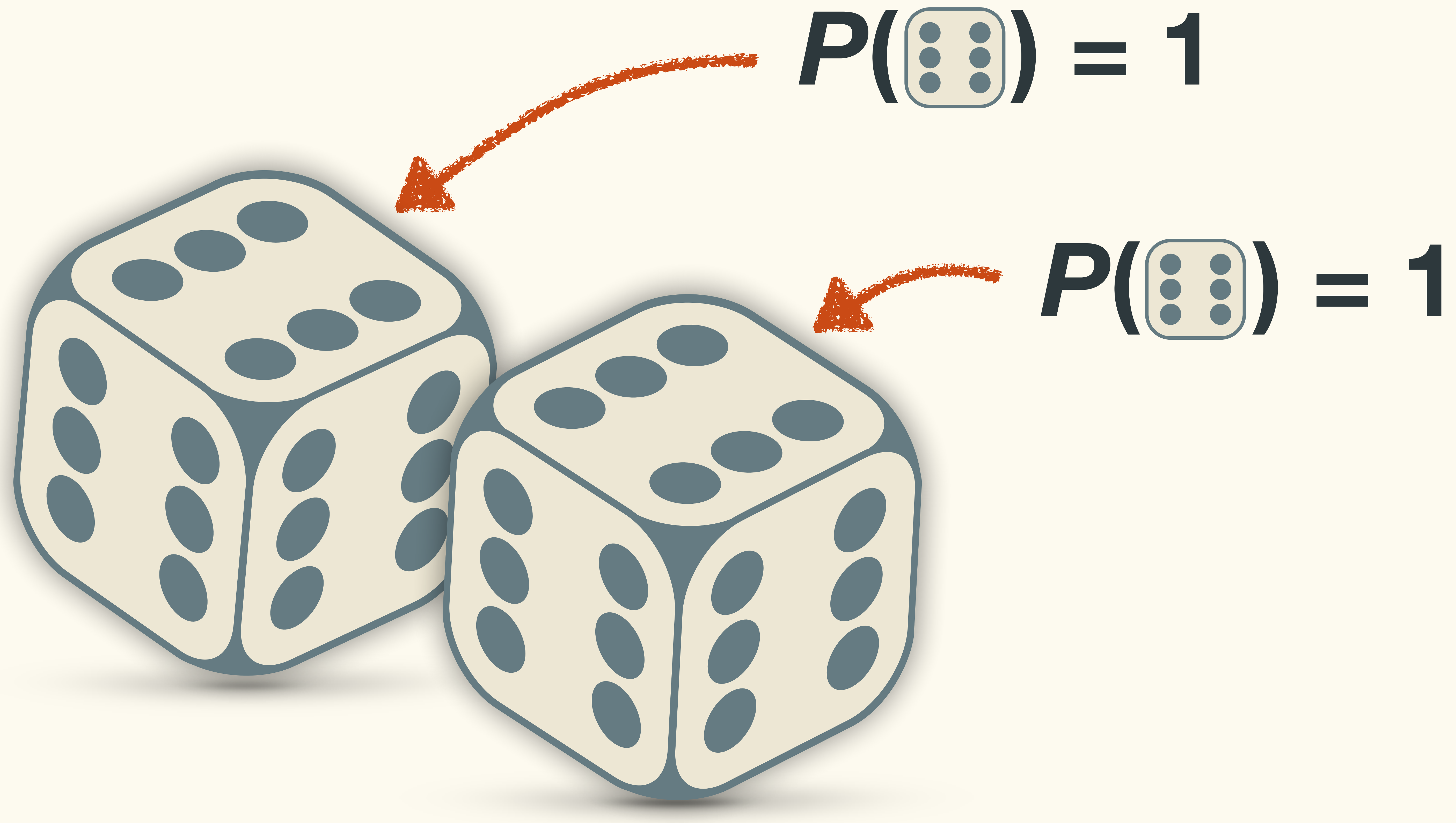
Probability

$P($⚅$) = ?$

Probability

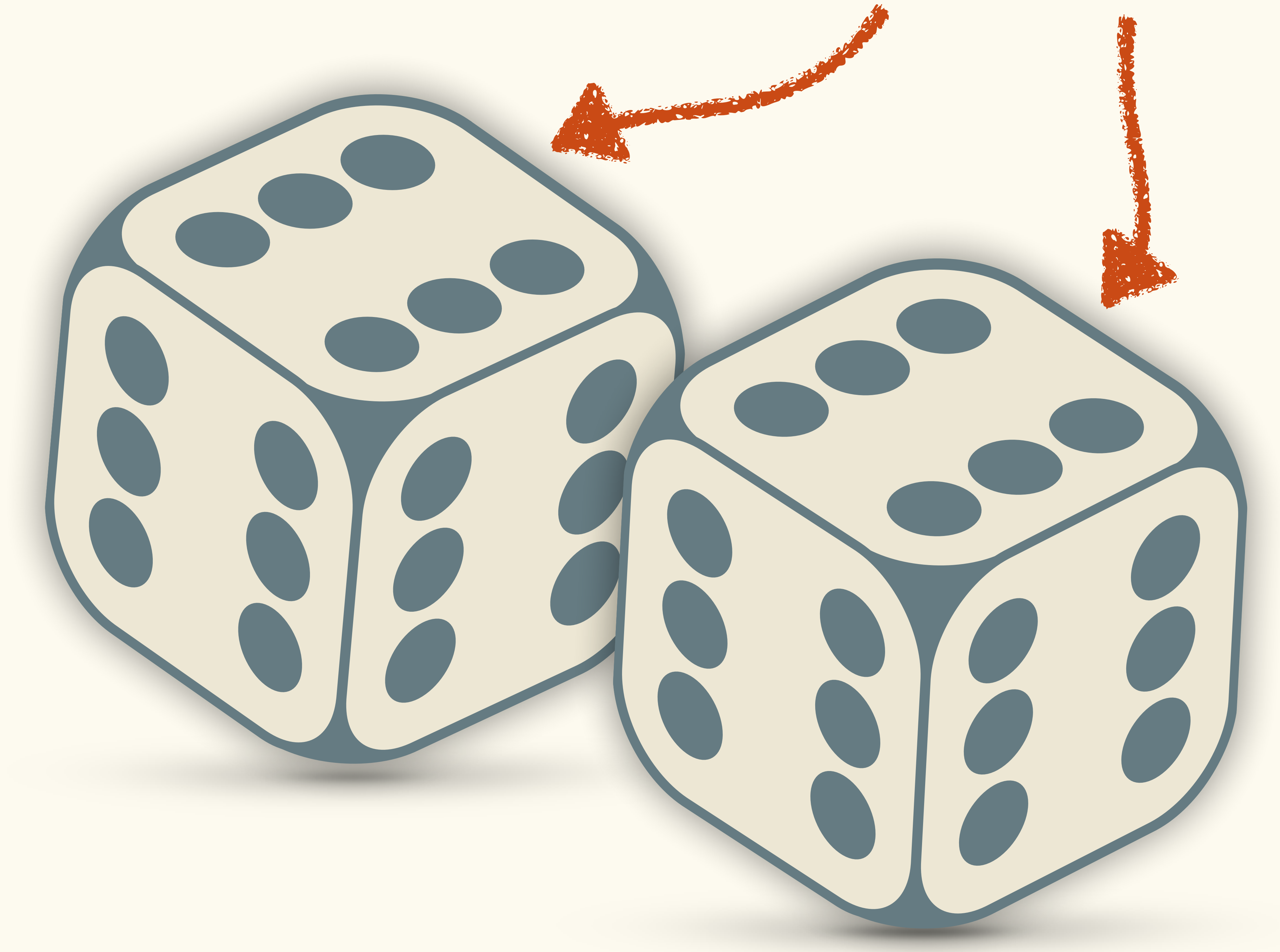
$P($⚅$) = 1$

Probability



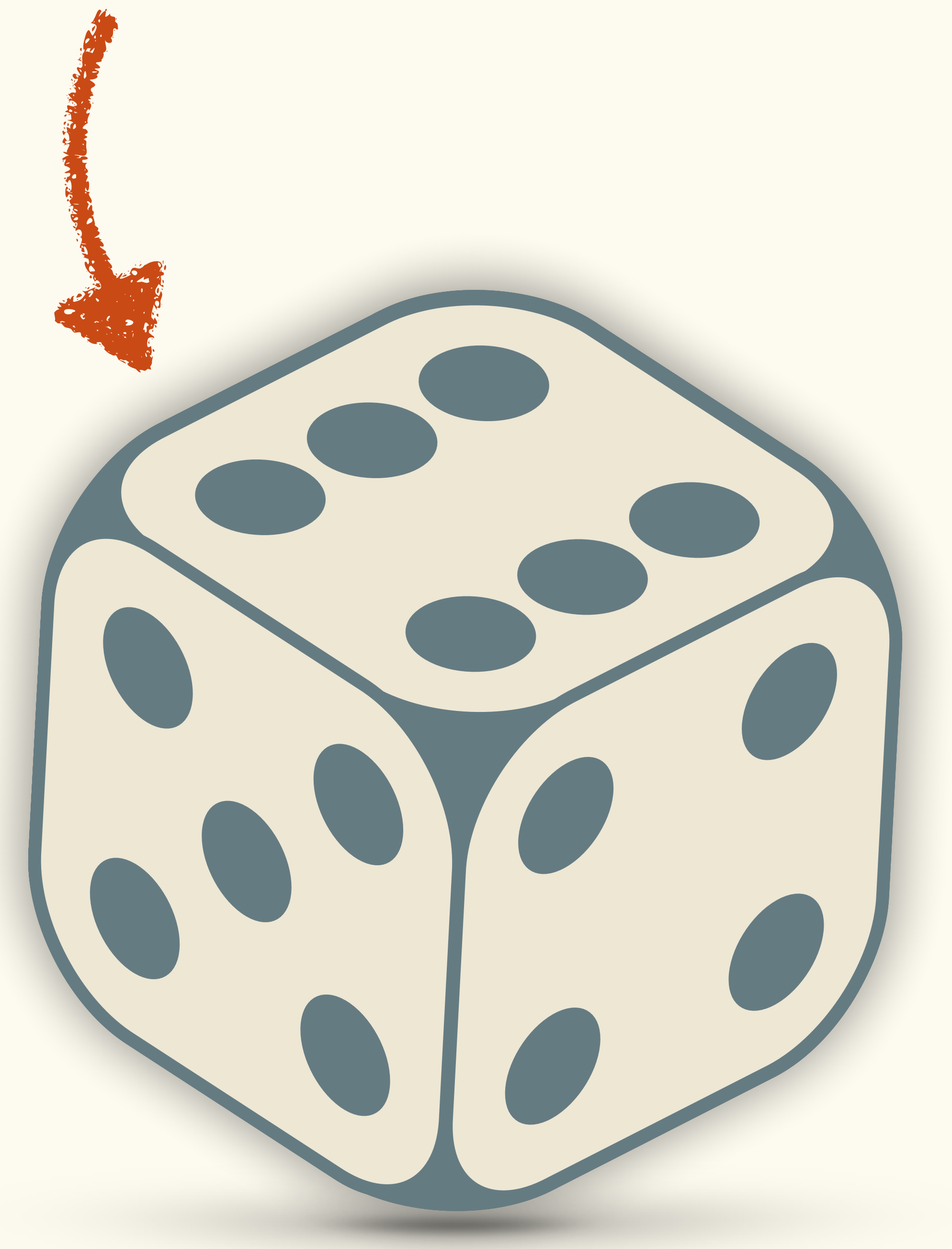
Probability

P(⚅ & ⚅) = 1 × 1 = 1

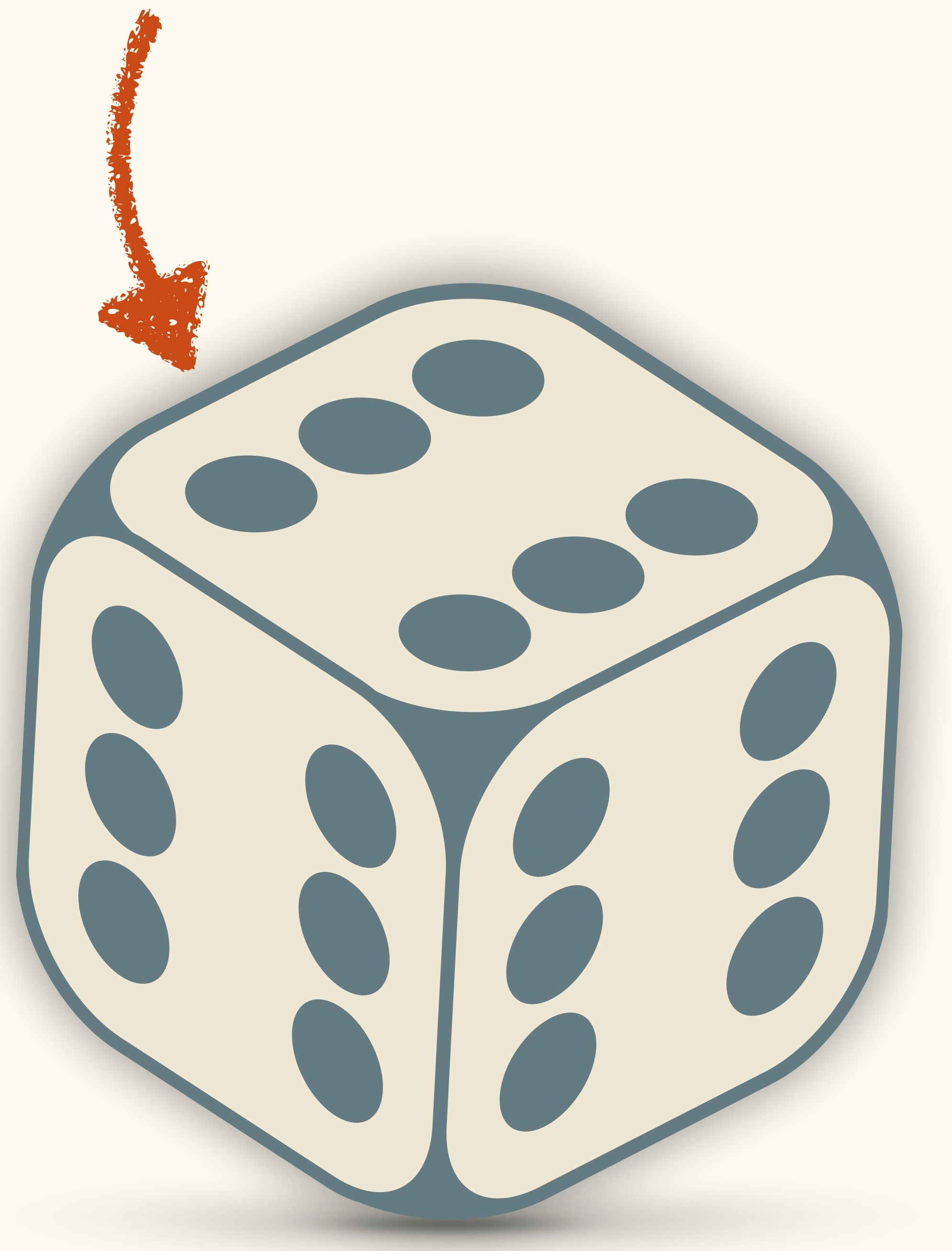


Probability

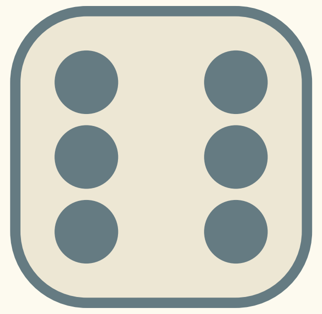
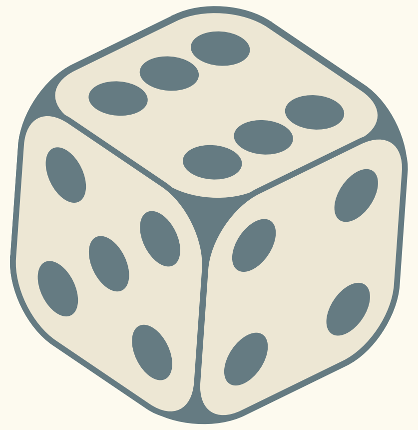
$P(⚅) = {}^1/_6$



$P(⚅) = 1$

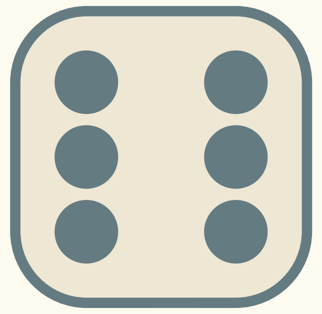
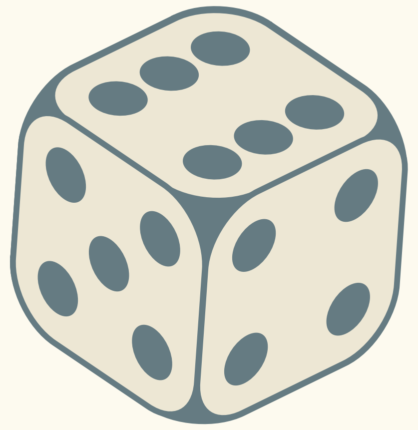


Probability

P( | ) = 1/6

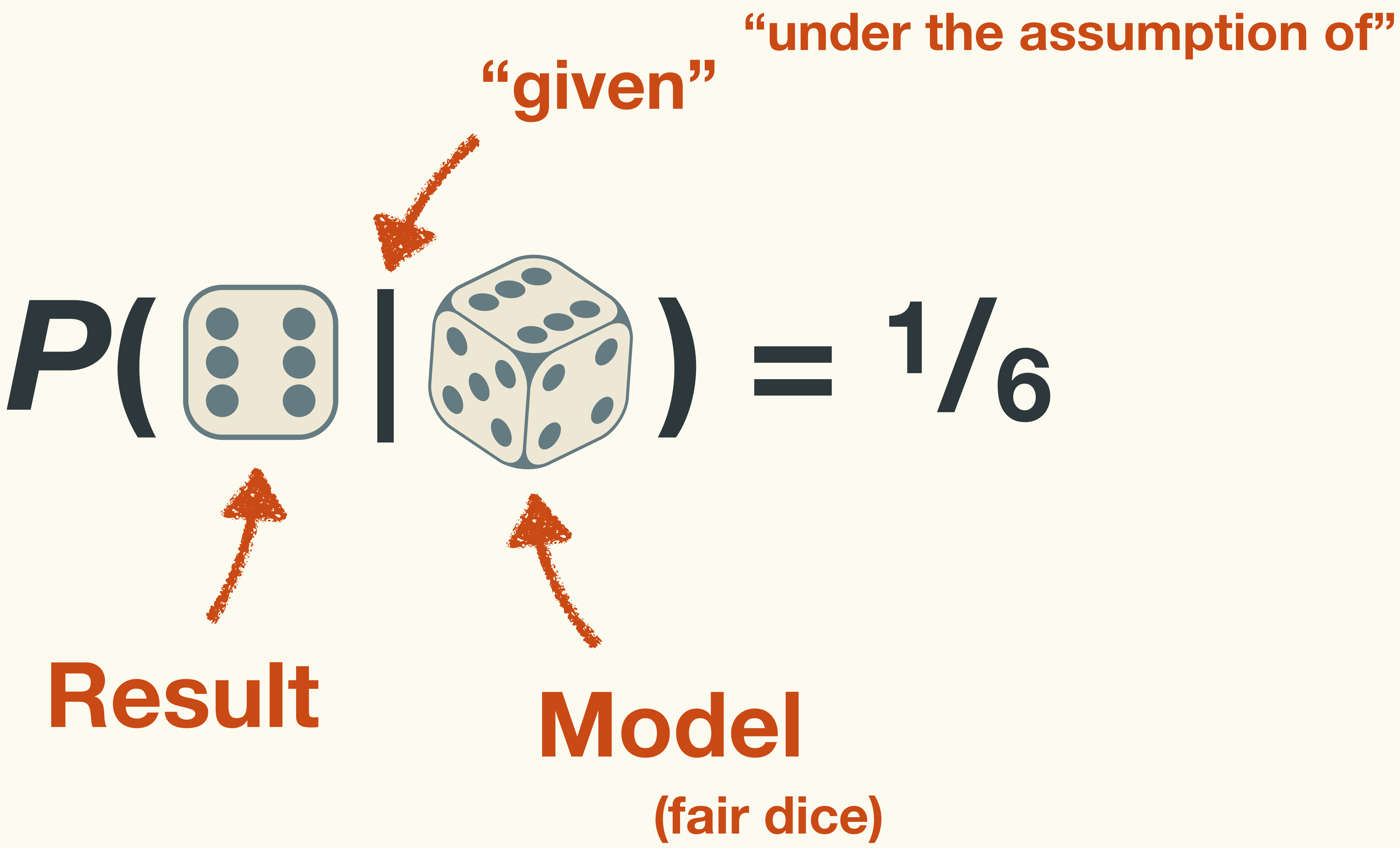
Observation

Probability

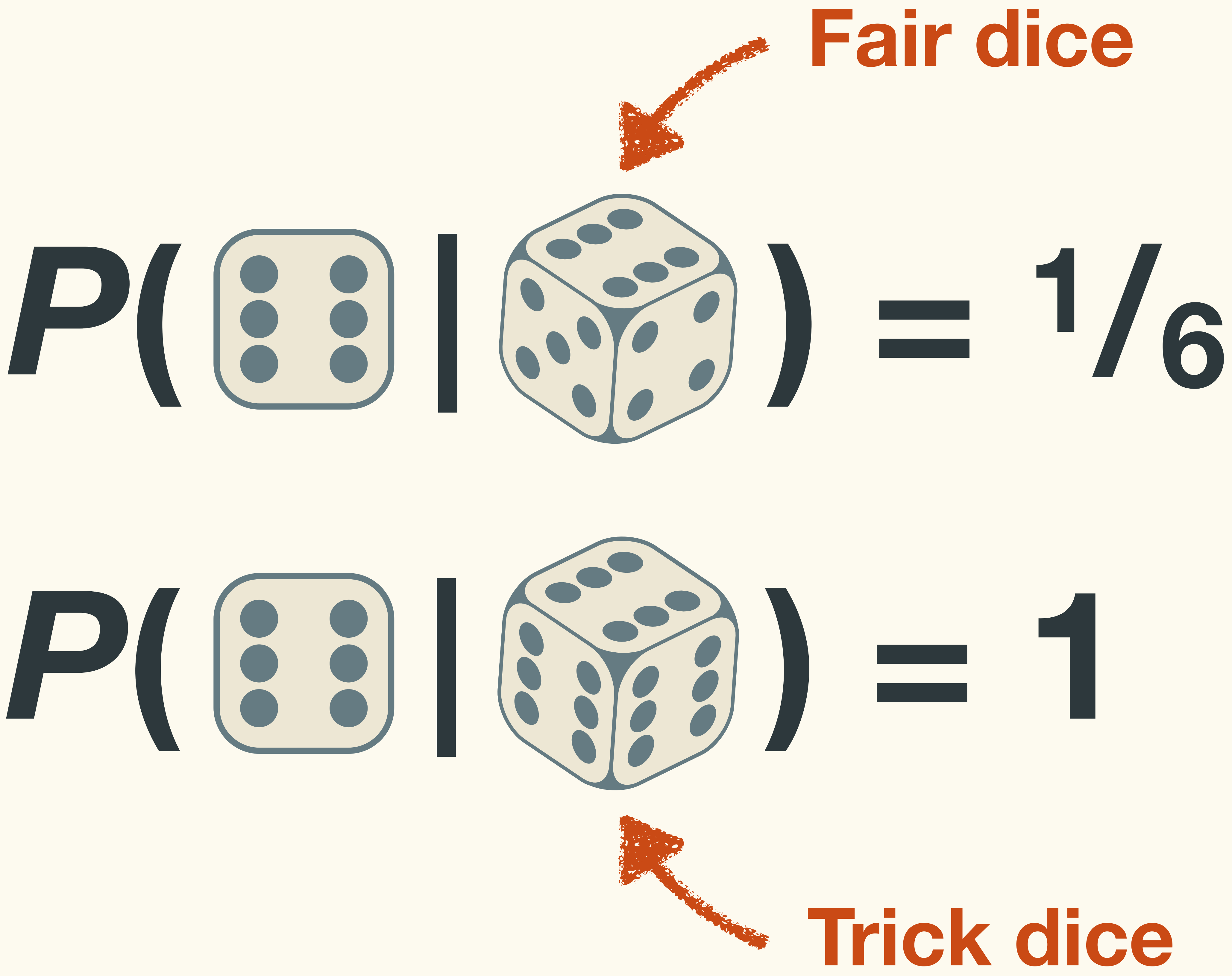
P( | ) = 1/6

Result

Probability

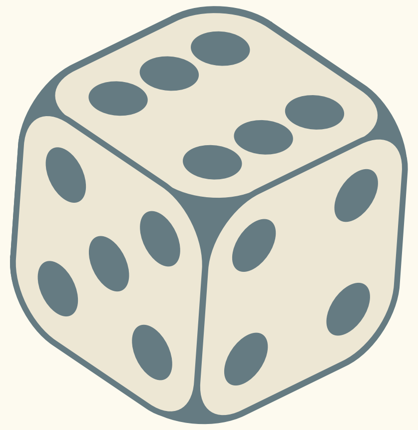
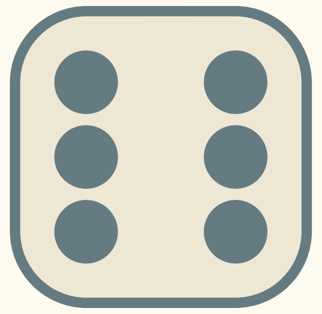


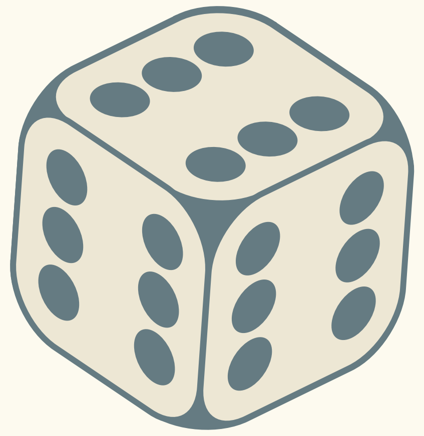
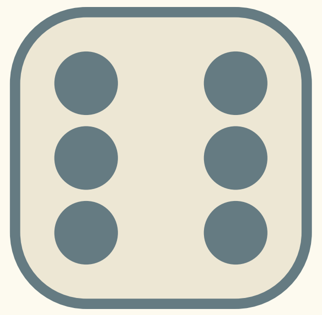
Probability



Likelihood

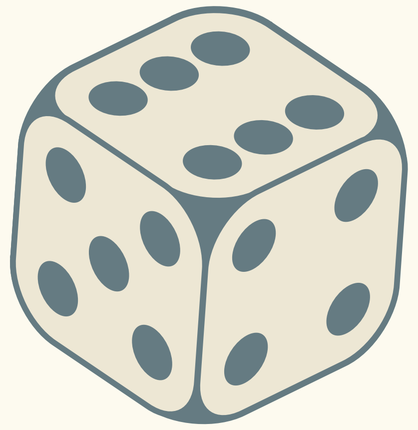
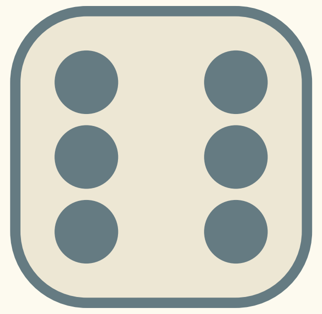
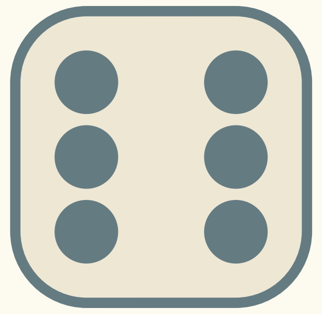
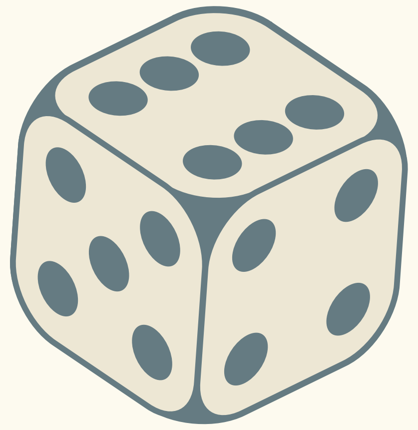
Fair dice

$L($  $|$  $) = 1/6$

$L($  $|$  $) = 1$

Trick dice

Likelihood

L( | ) = *P*( | )

Probability

$P(A) = ?$

Probability

$P(\mathbf{A}) = 1/4$

$P(\mathbf{A}) = 1$

Probability

A, C, G, T

$P($ **A** $|$  $) = 1/4$

$P($ **A** $|$  $) = 1$

Only A

Likelihood

A, C, G, T

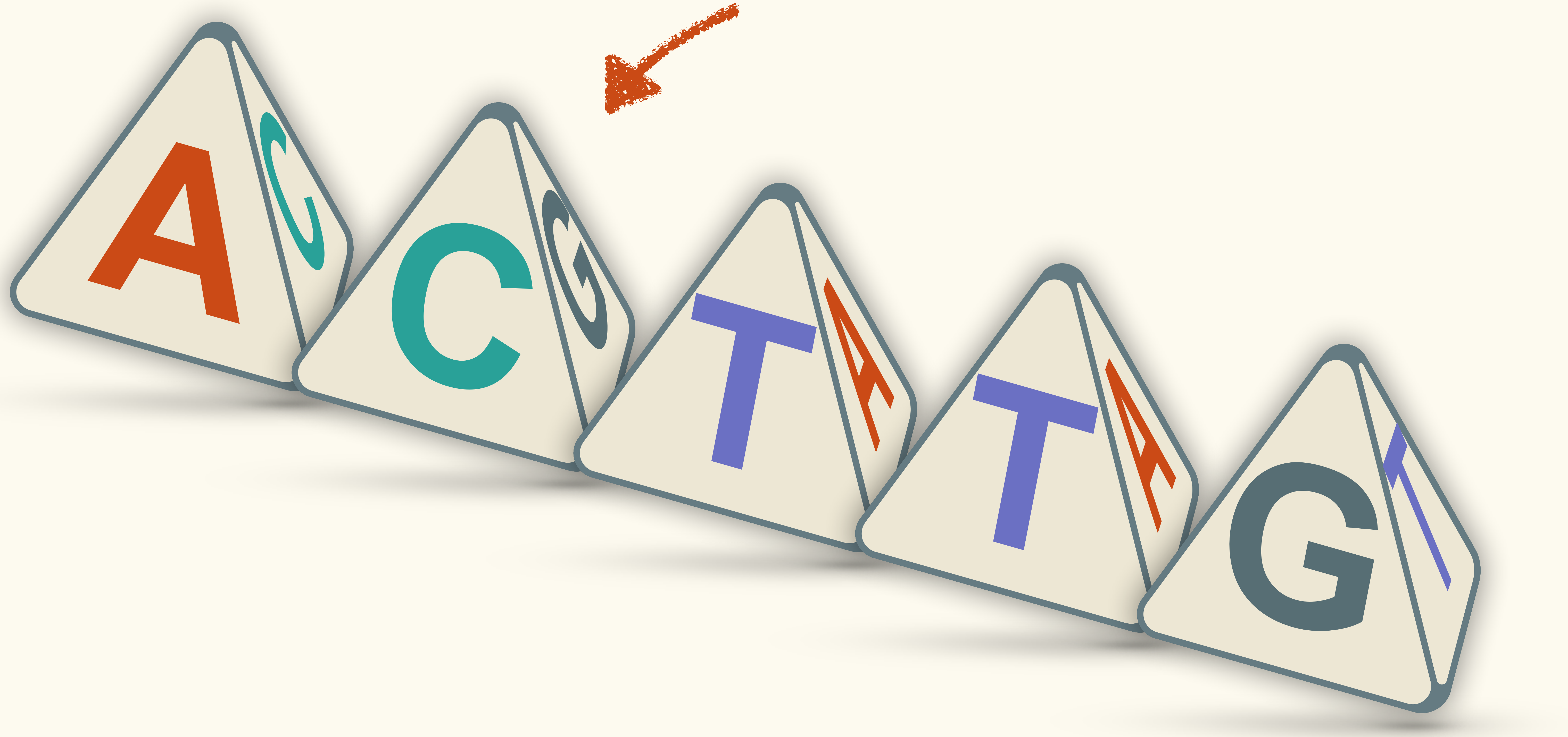
$L($  $|\, \mathrm{A}) = 1/4$

$L($  $|\, \mathrm{A}) = 1$

Only A

Probability

$P(\text{ACTTG}) = \ ?$



Probability

A, C, G, T

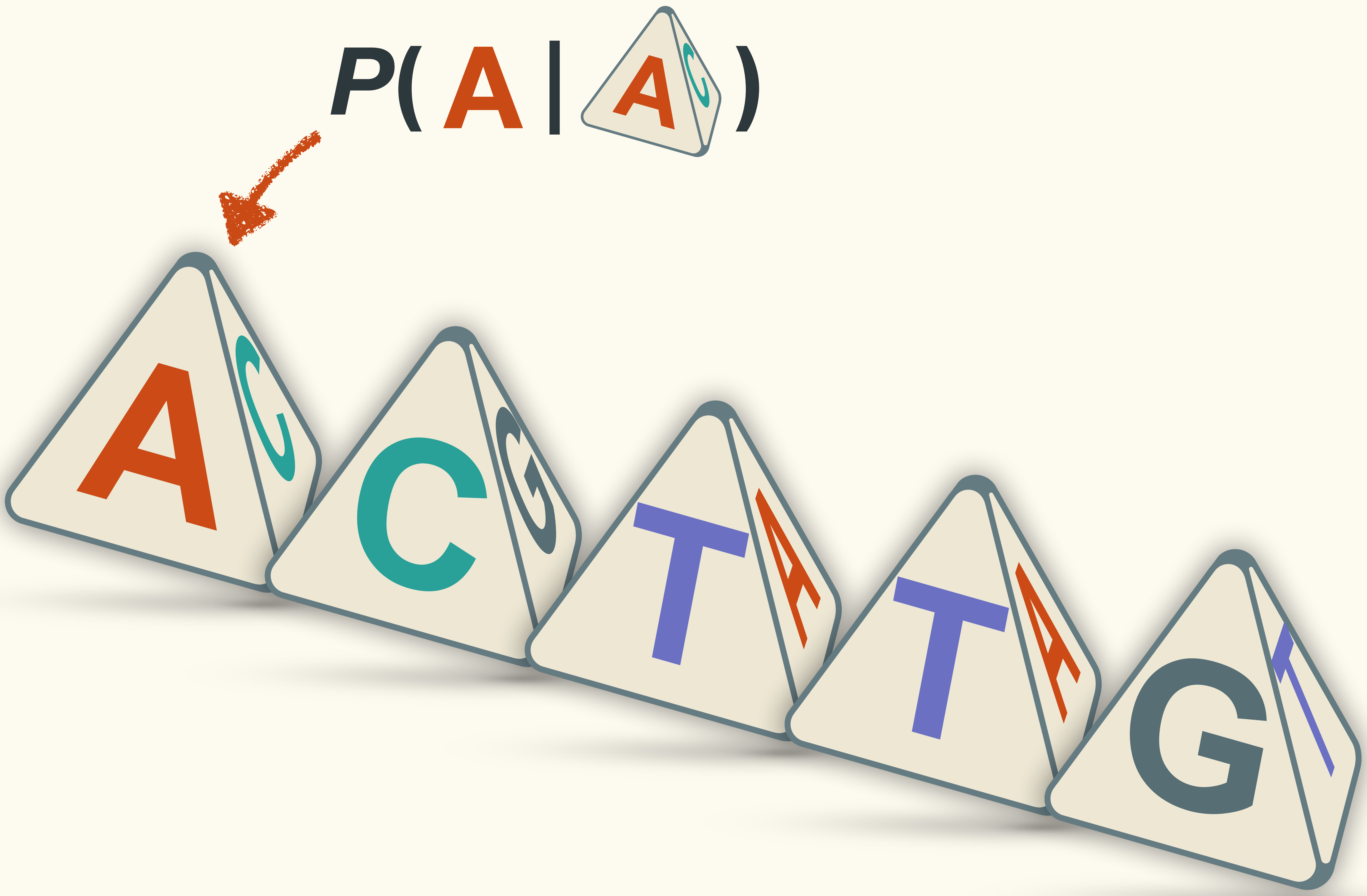
$P($ACTTG $\mid$  $) = \ ?$

Probability

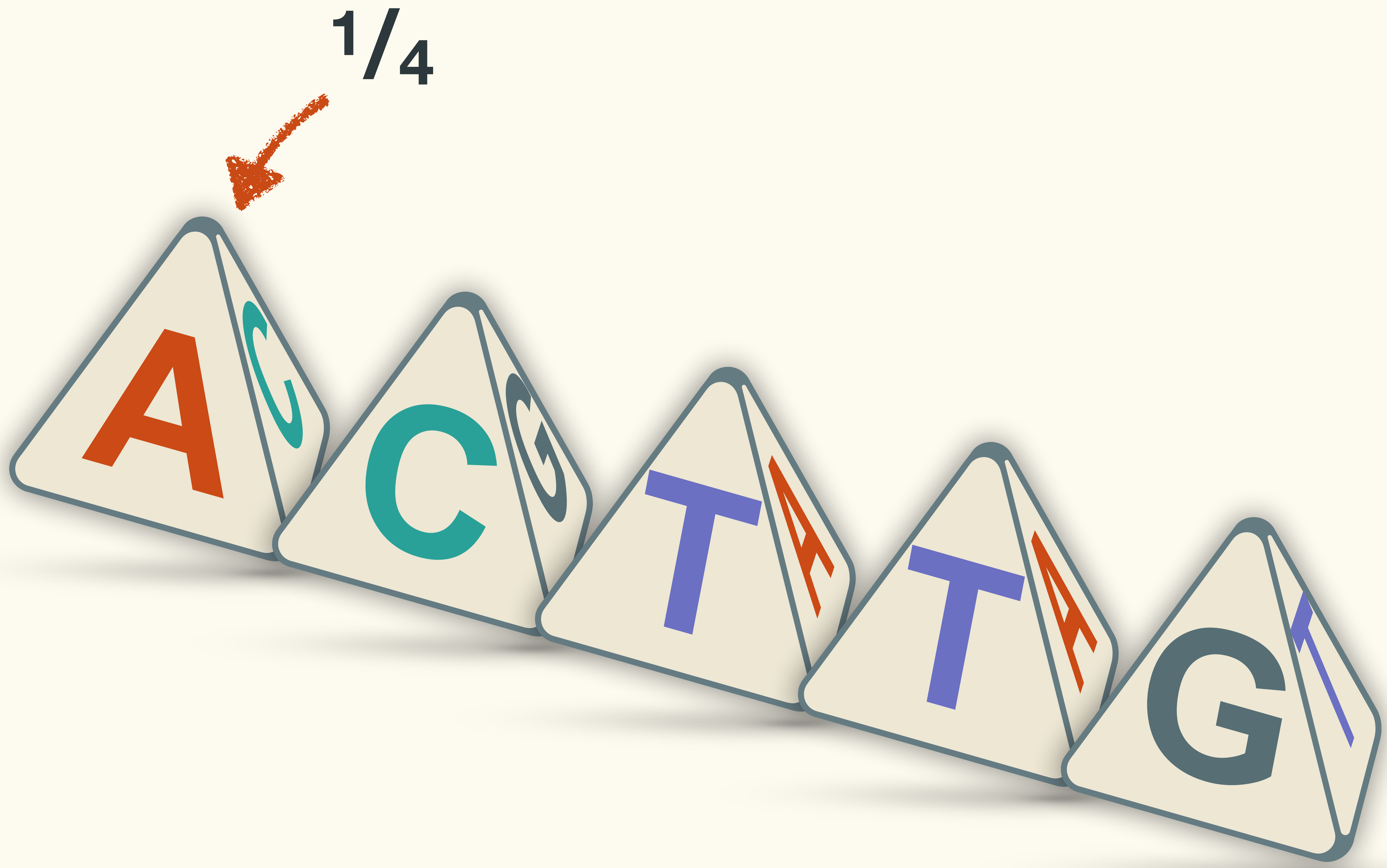
A, C, G, T

$P($ A & C & T & T & G |  $) = ?$

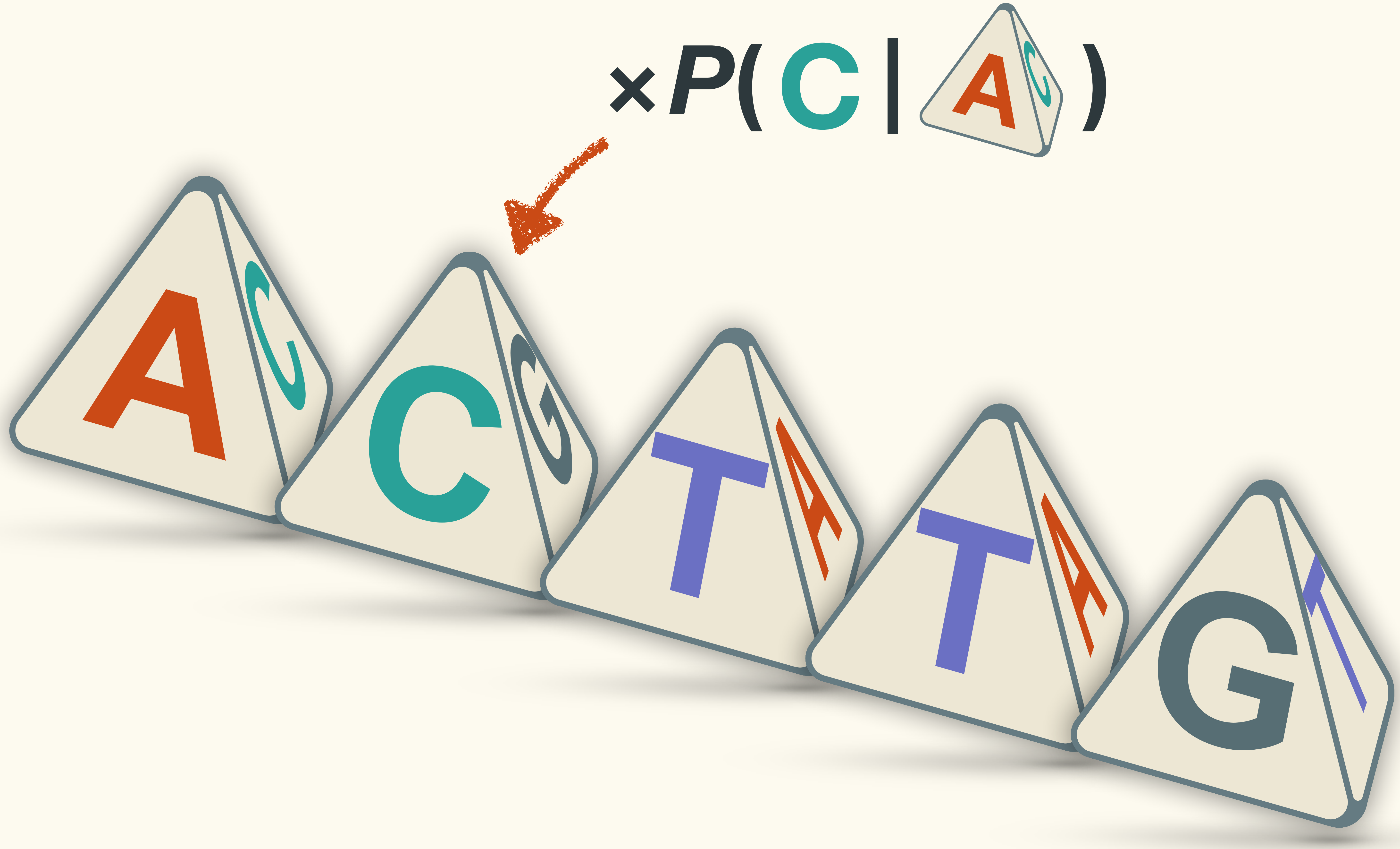
Probability



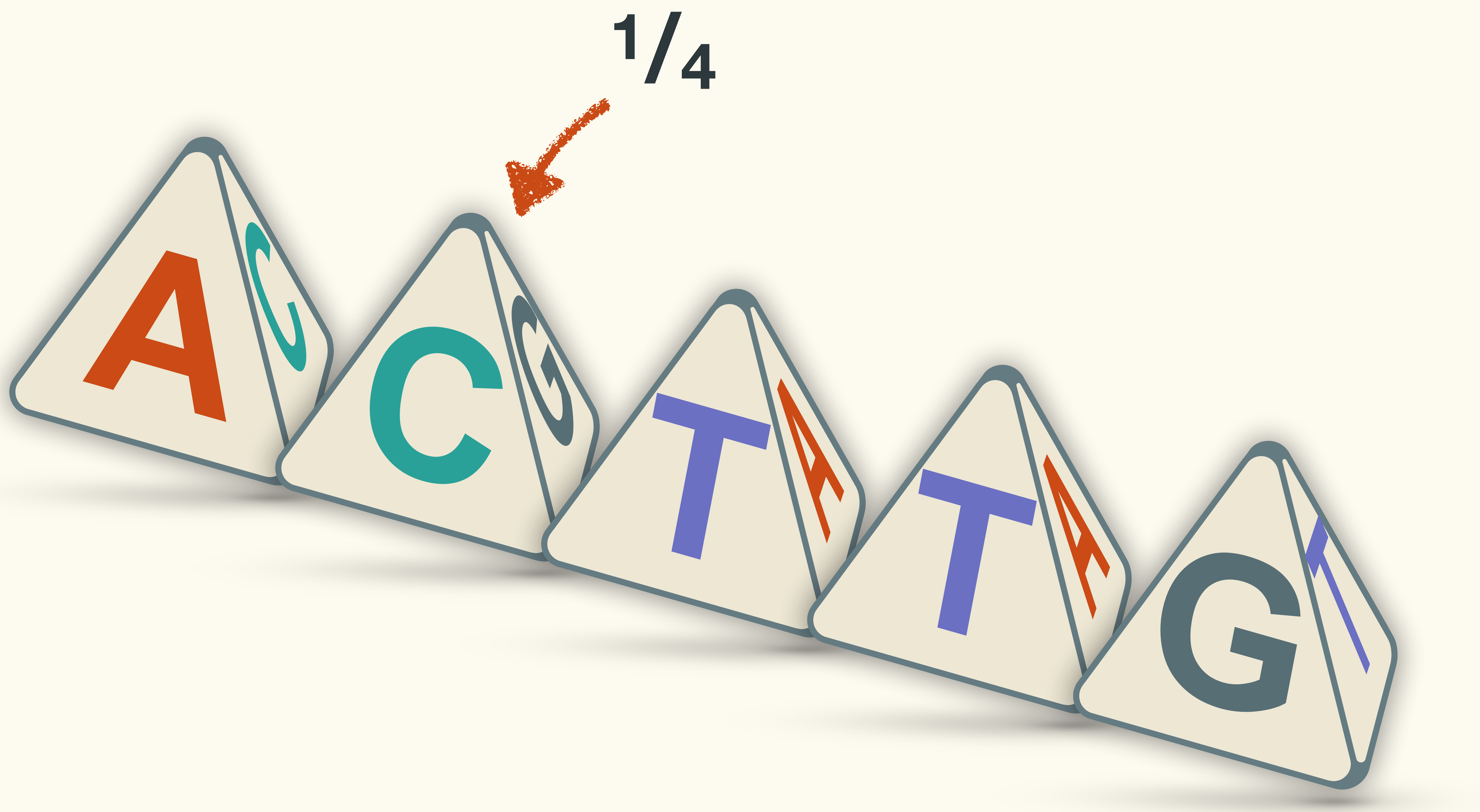
Probability



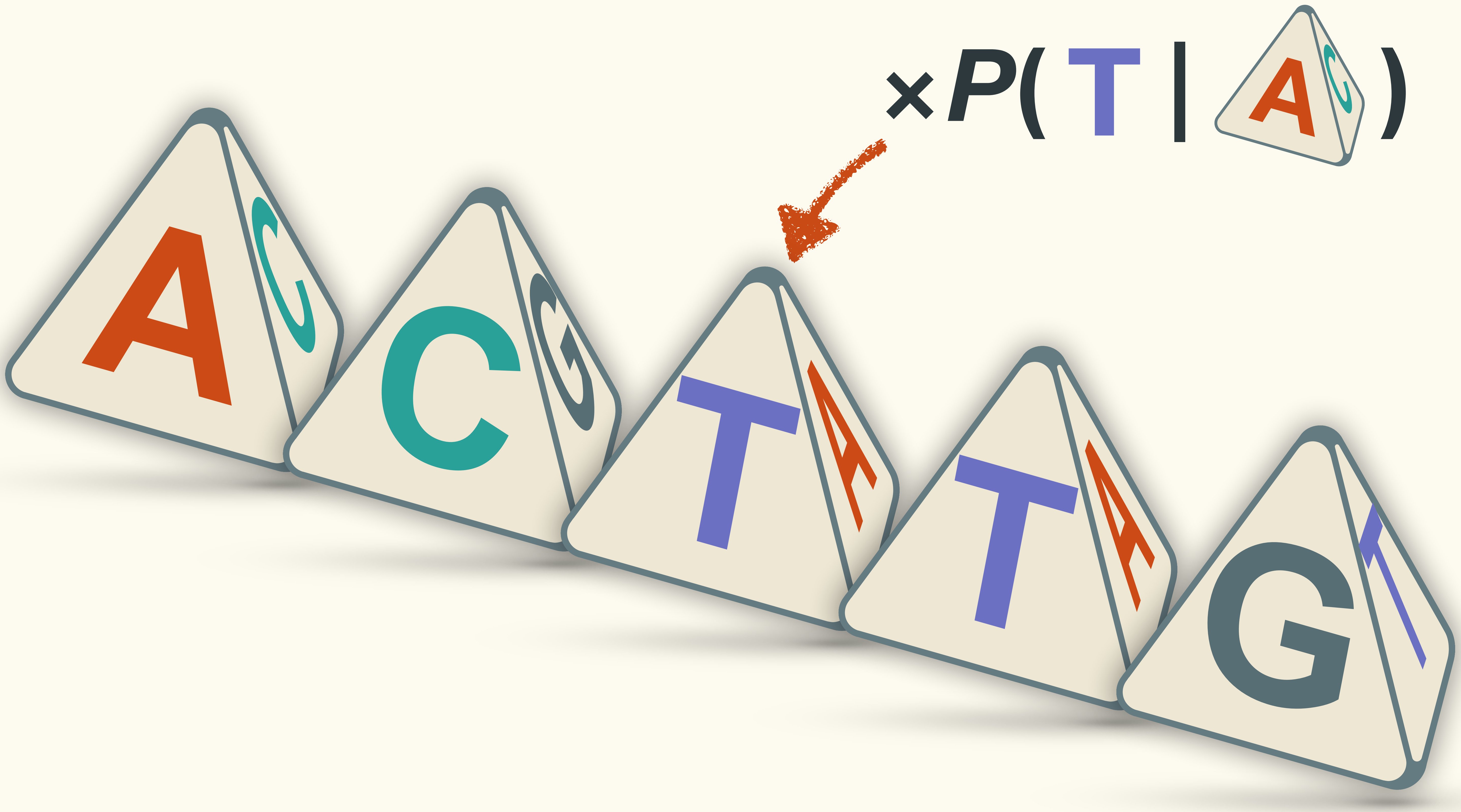
Probability



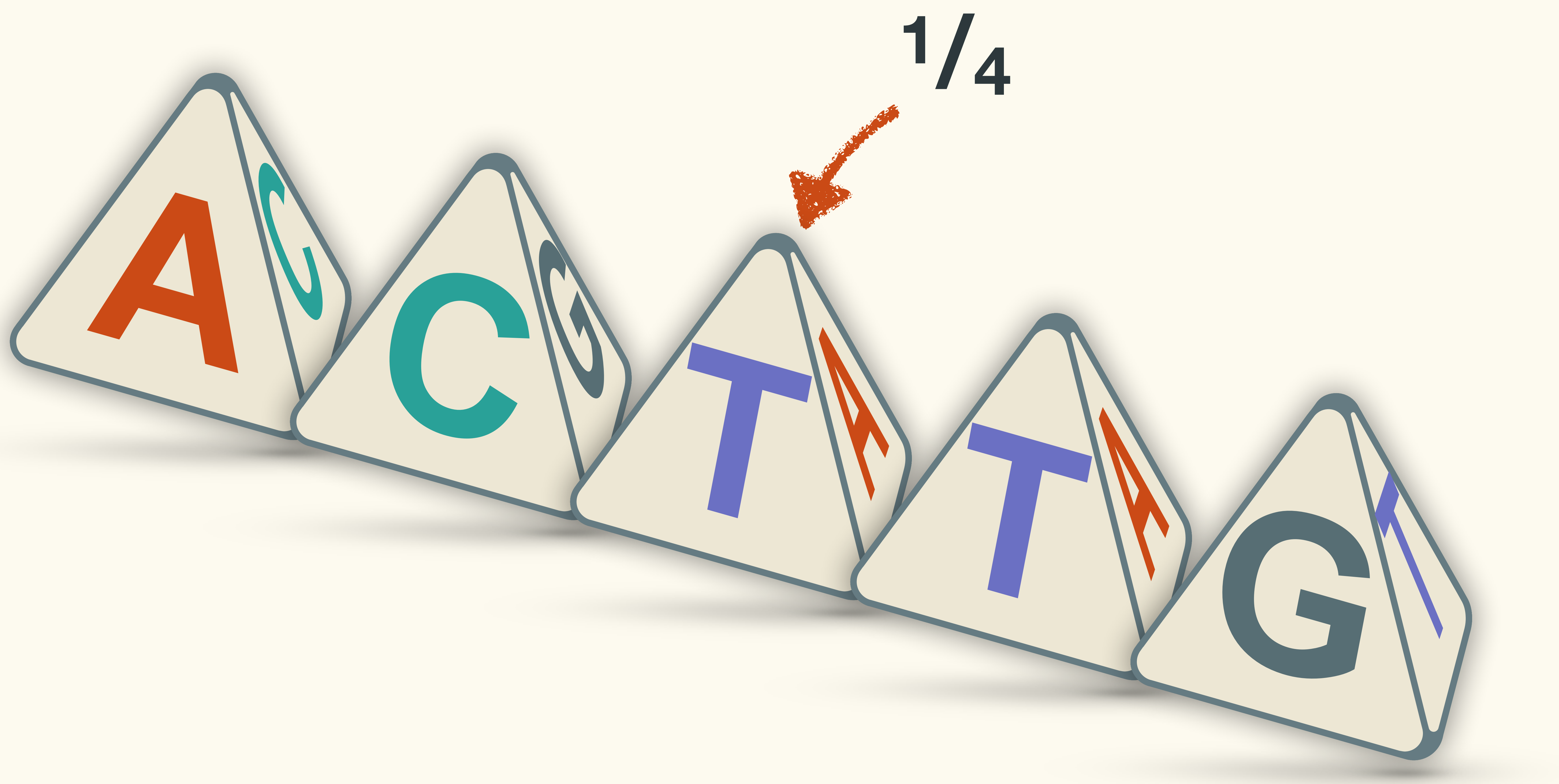
Probability



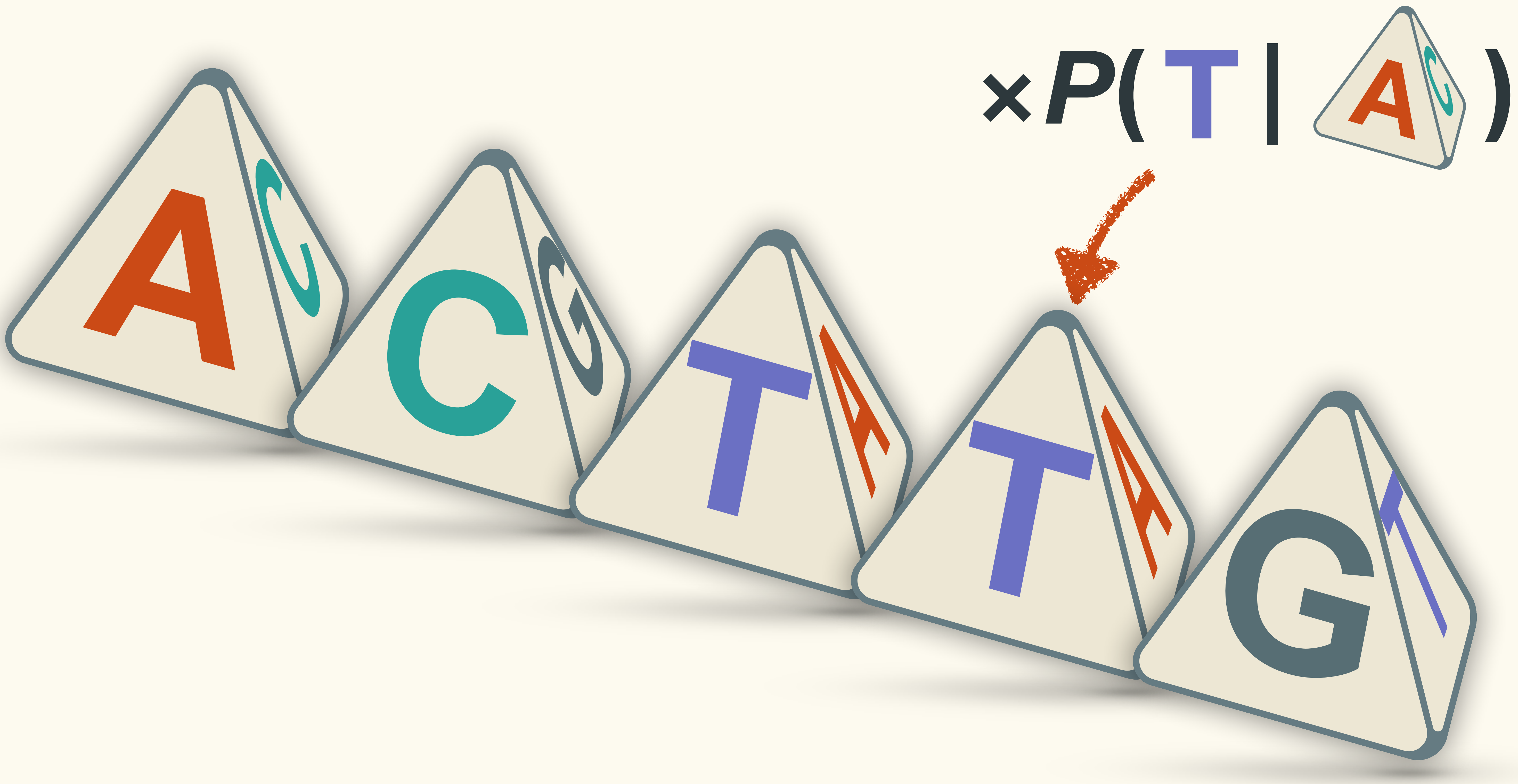
Probability



Probability

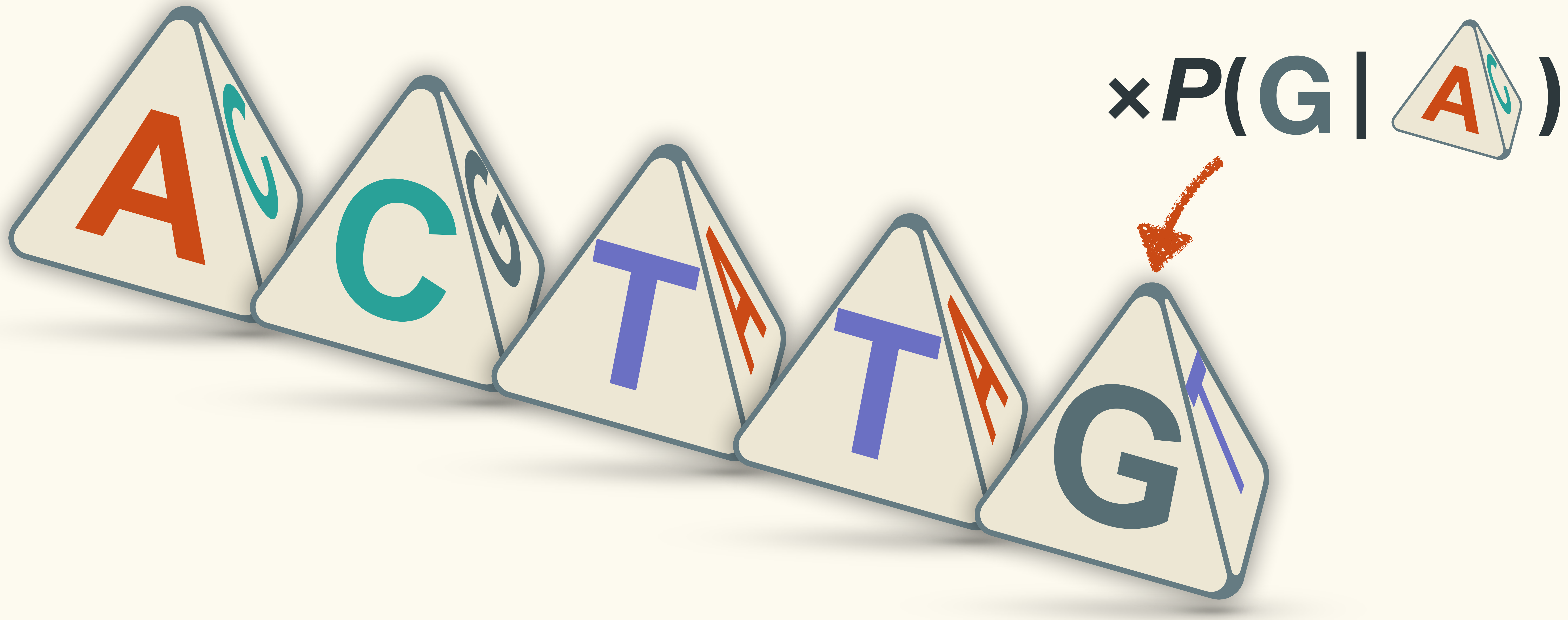


Probability

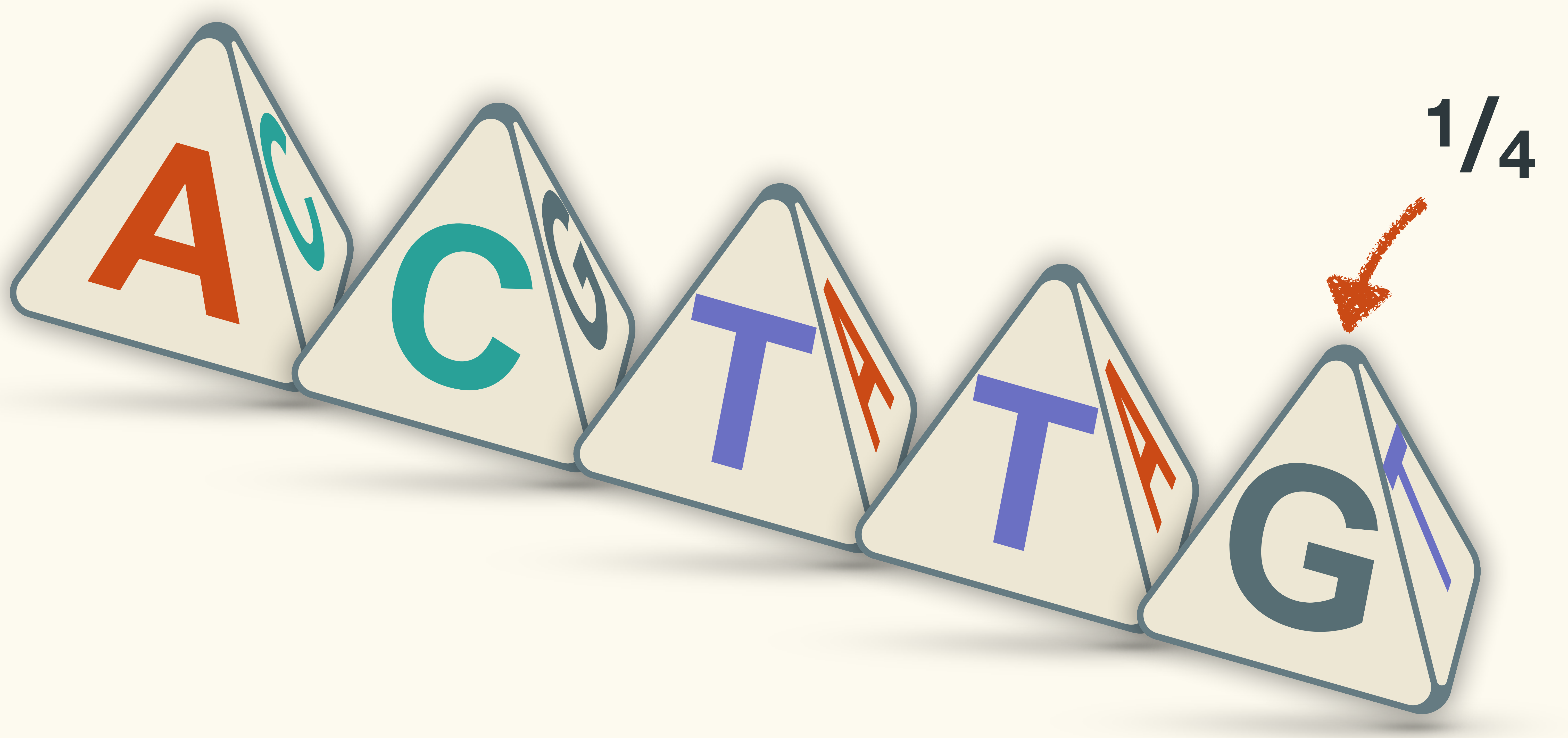


Probability

Probability

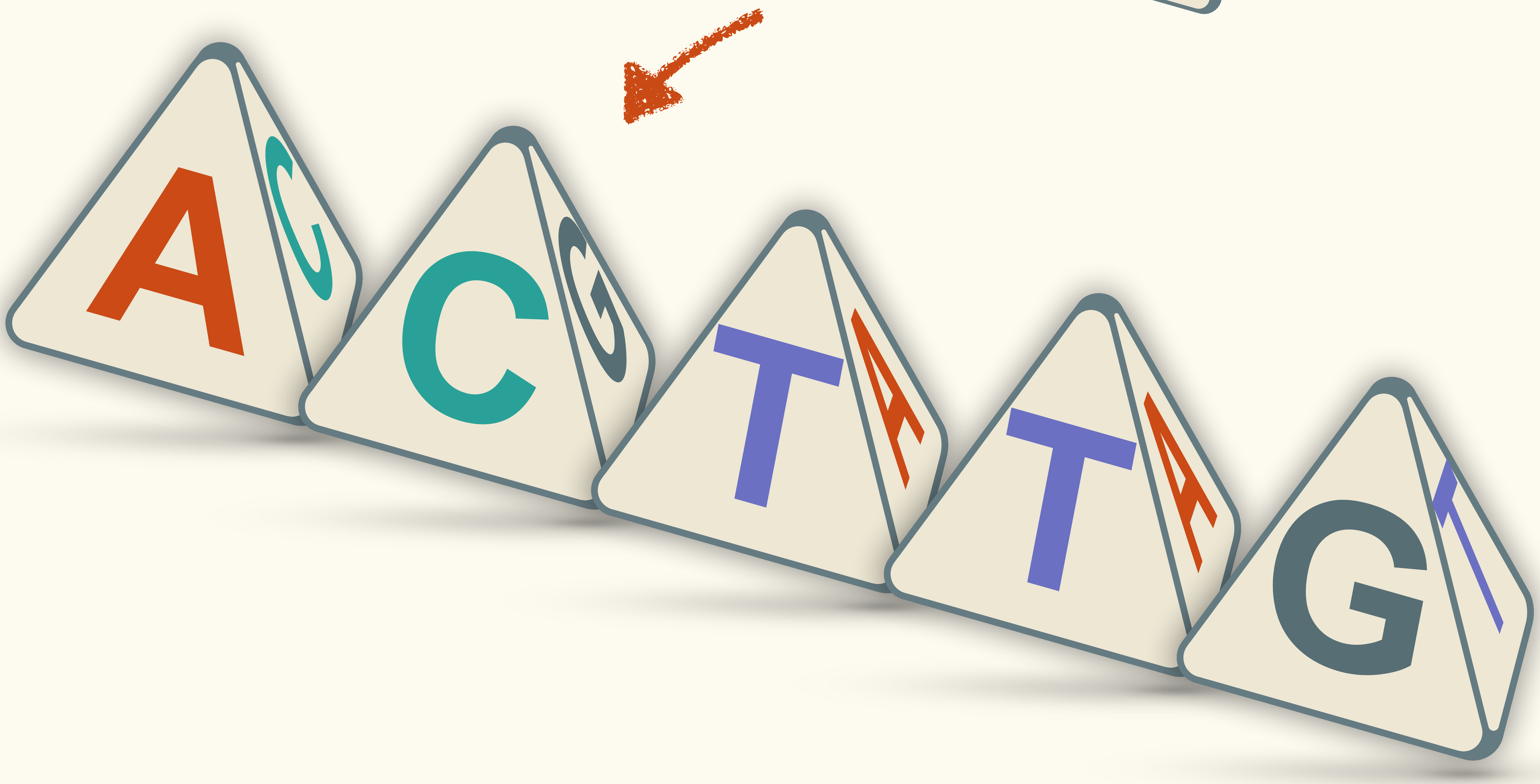


Probability

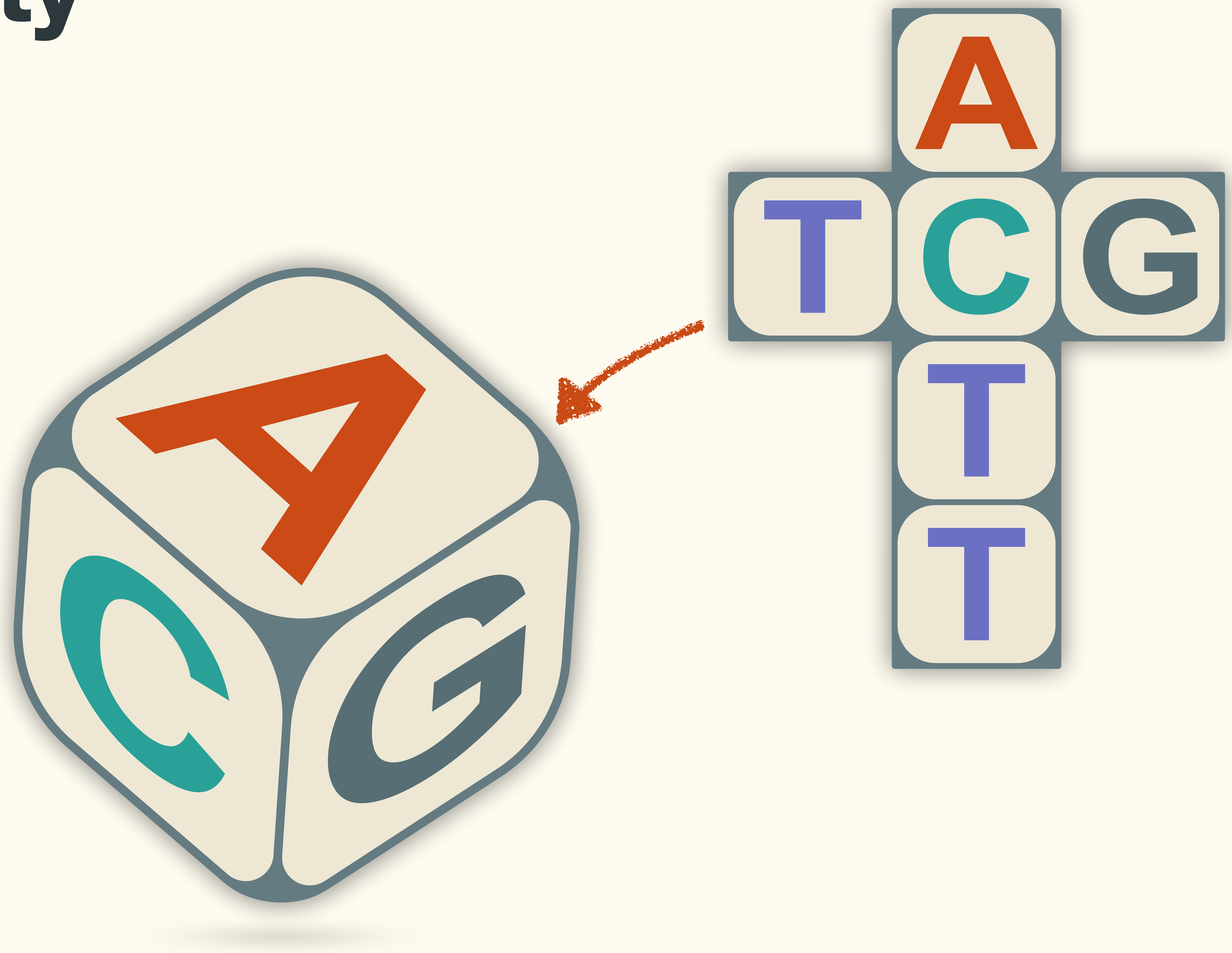


Probability

$P(A \,\&\, C \,\&\, T \,\&\, T \,\&\, G \mid$  $) = (^1/_4)^5$

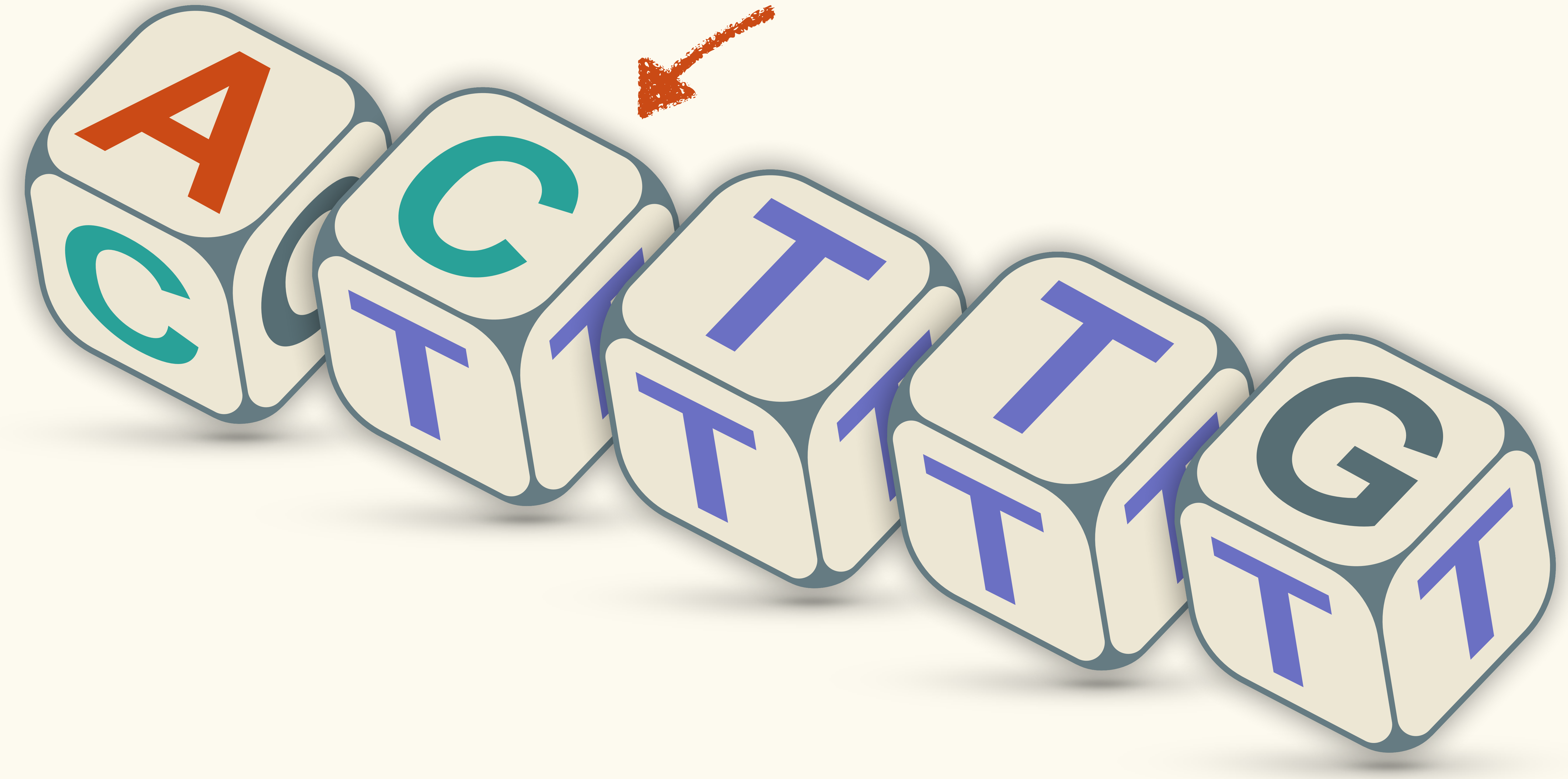


Probability



Probability

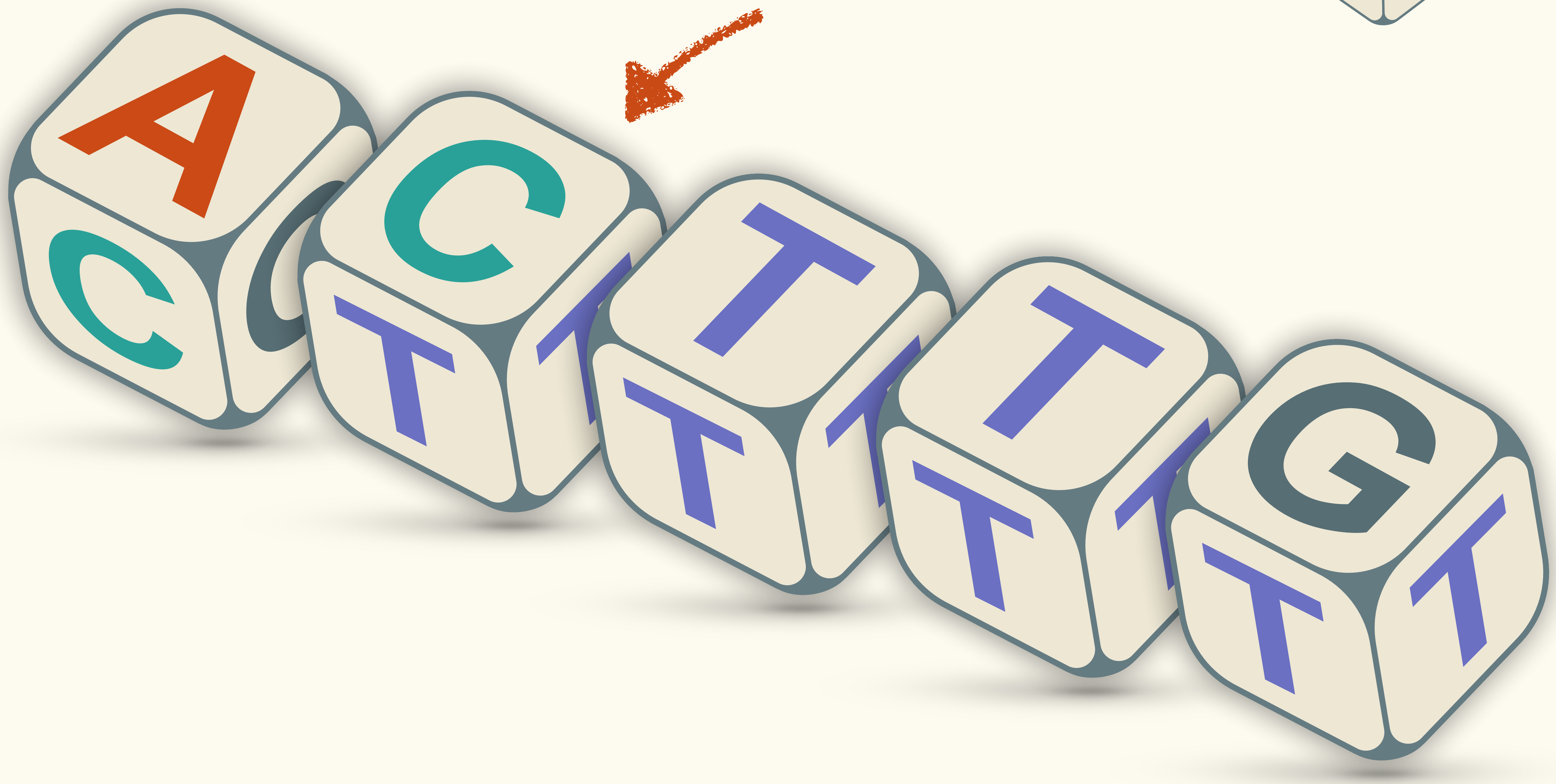
$P(\text{ACTTG}) = ?$



Probability

f(T) > f(A), f(C), f(G)

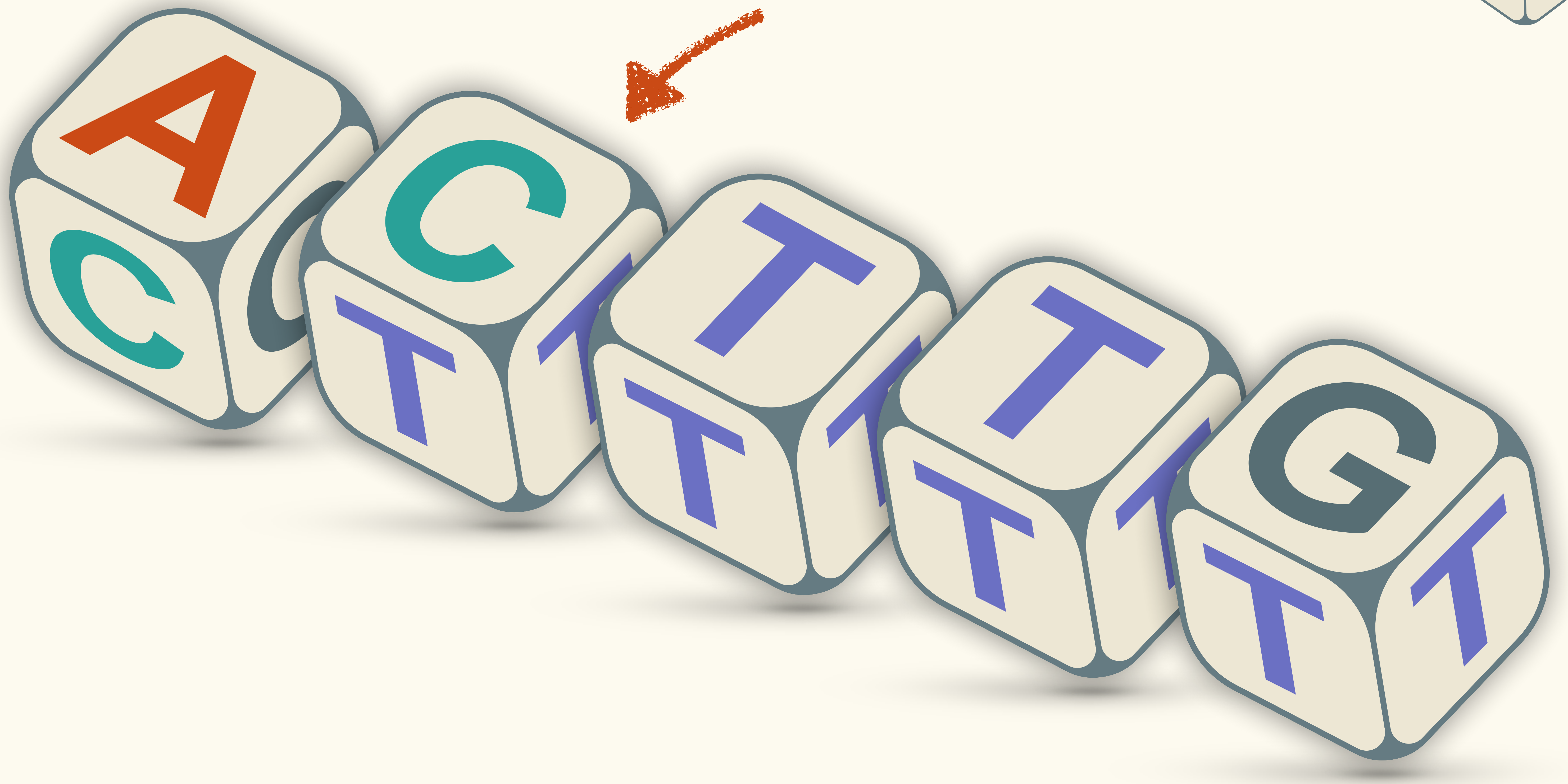
$P($ACTTG $|$  $) = ?$



Probability

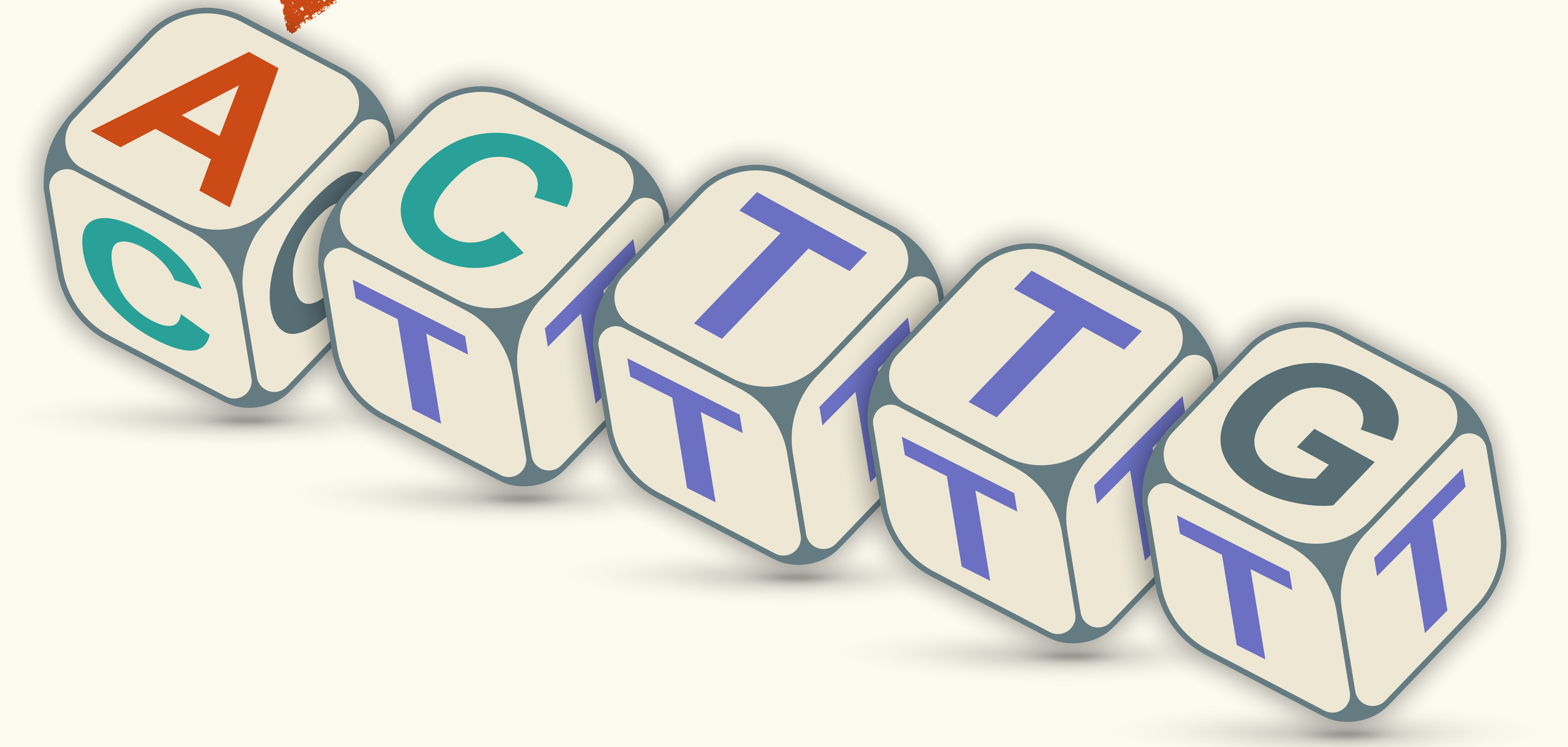
f(T) > f(A), f(C), f(G)

$P($A & C & T & T & G $\mid$  $) = ?$

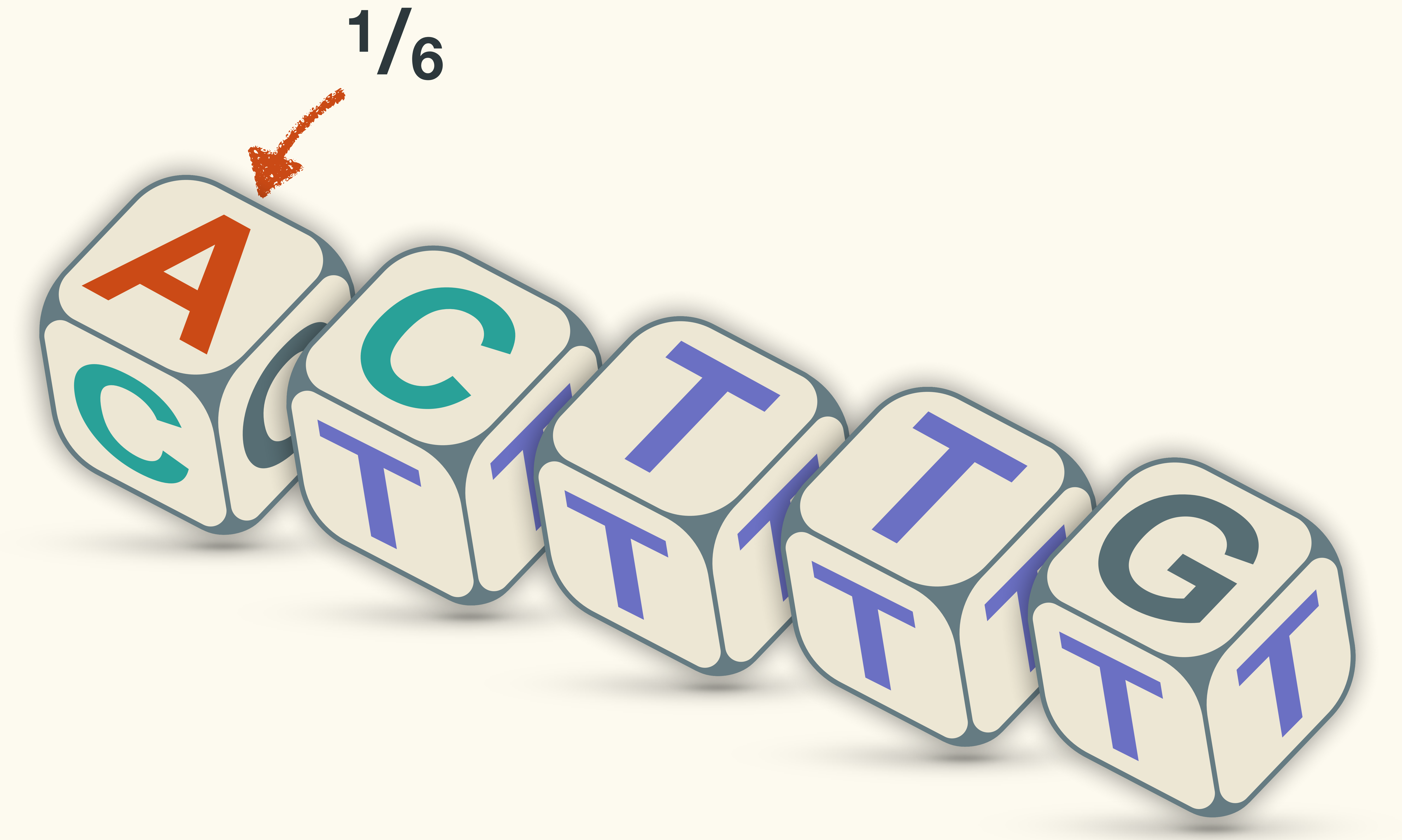


Probability

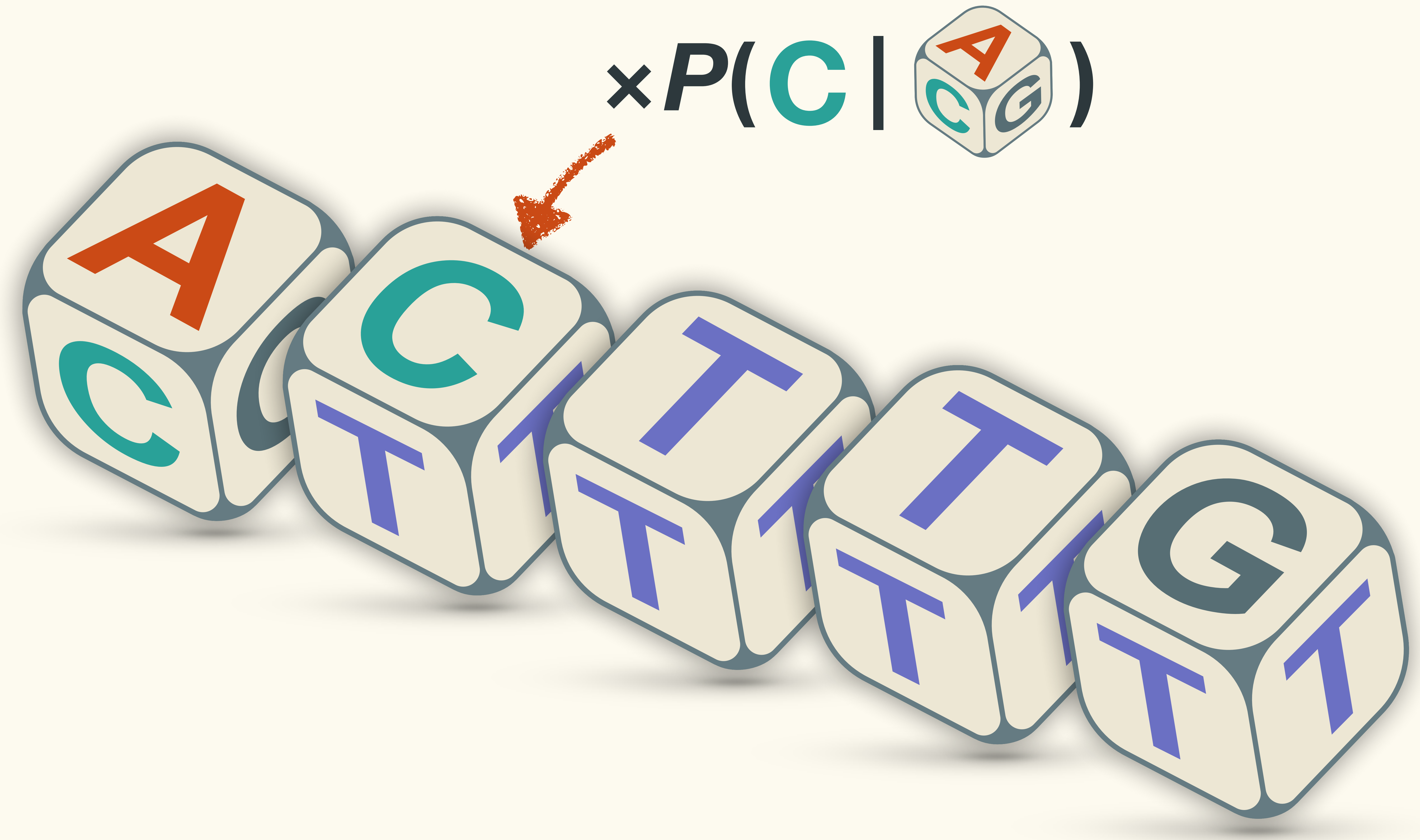
$P($A$\mid$  $)$



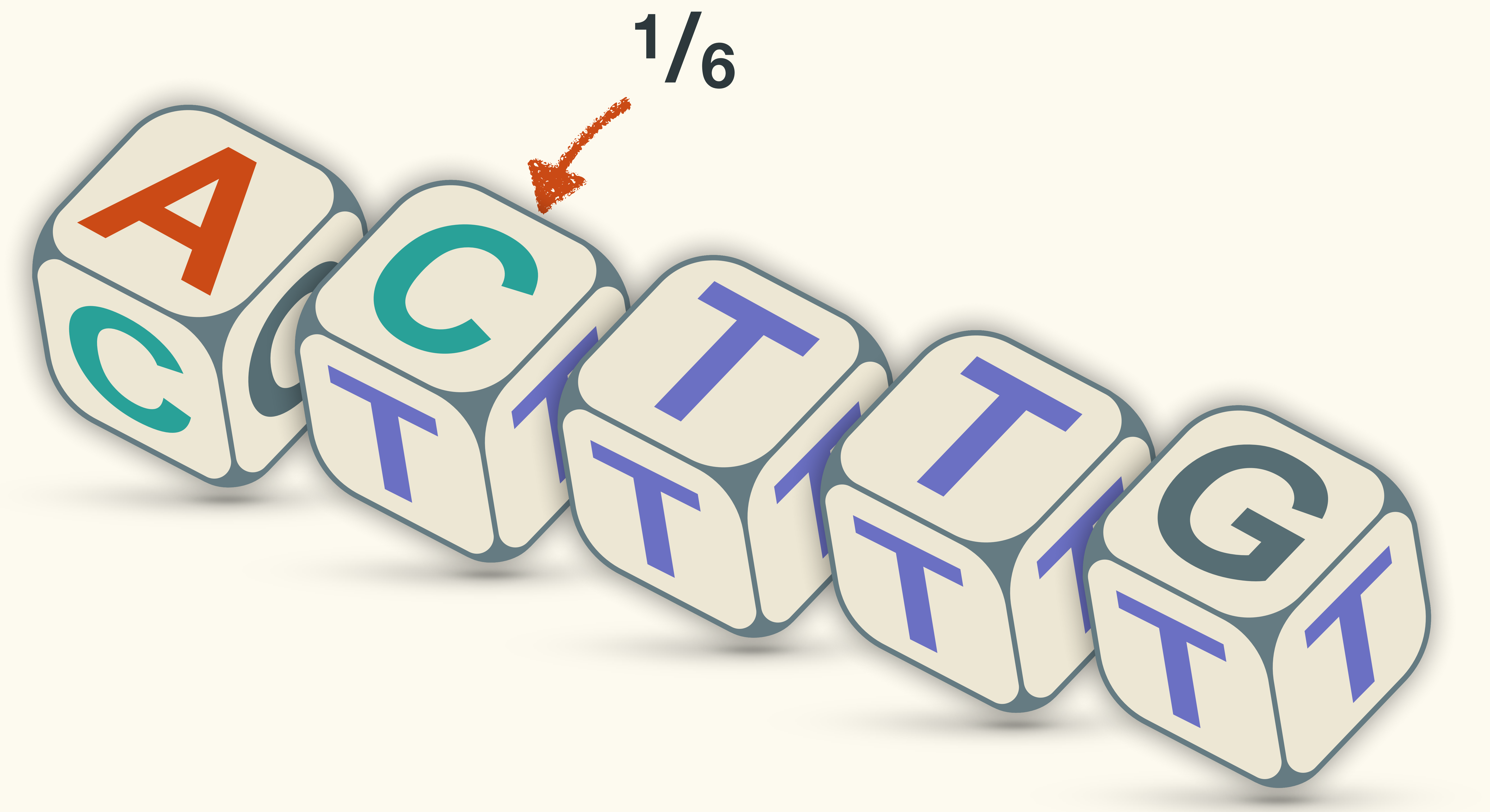
Probability



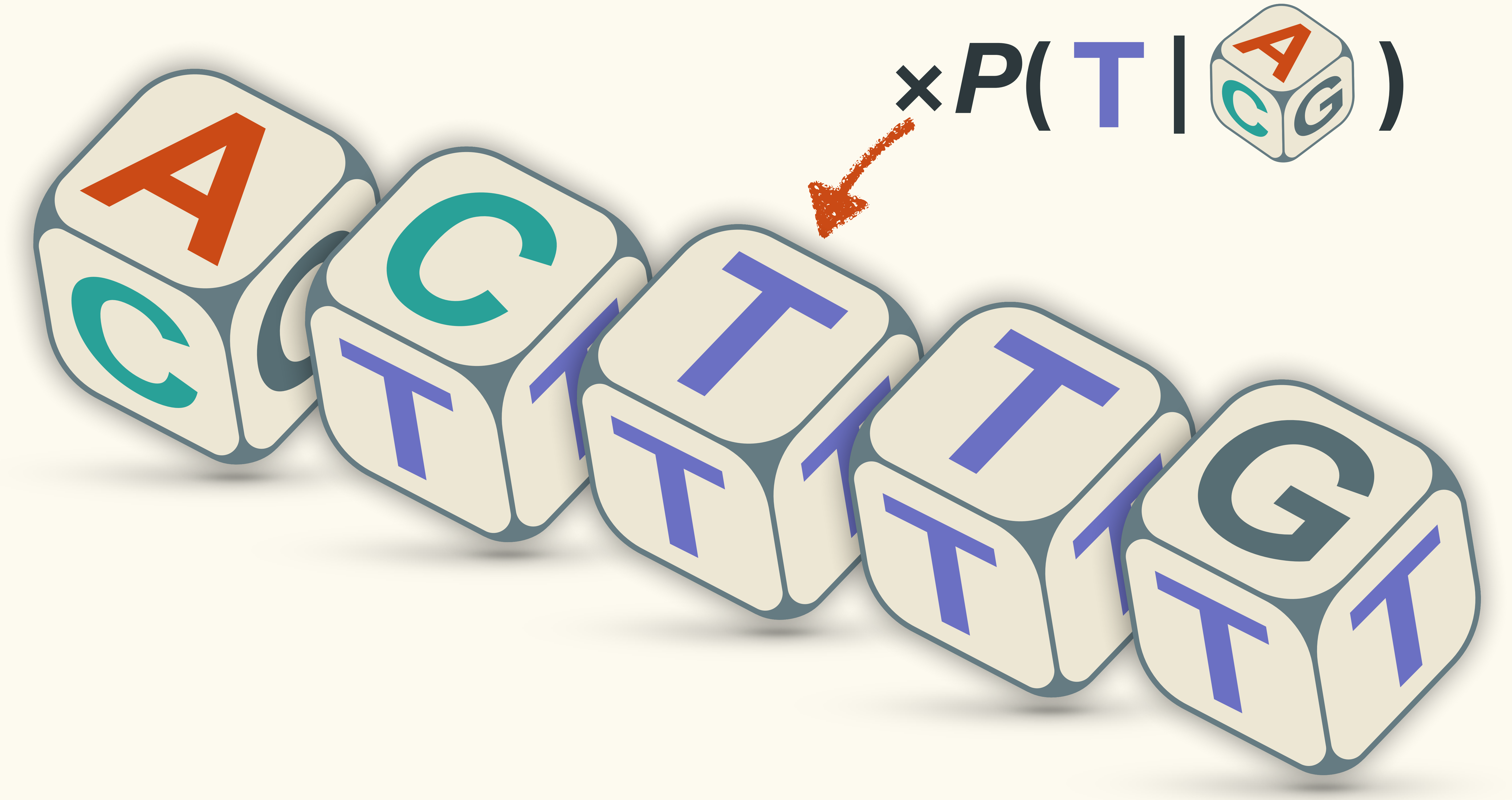
Probability



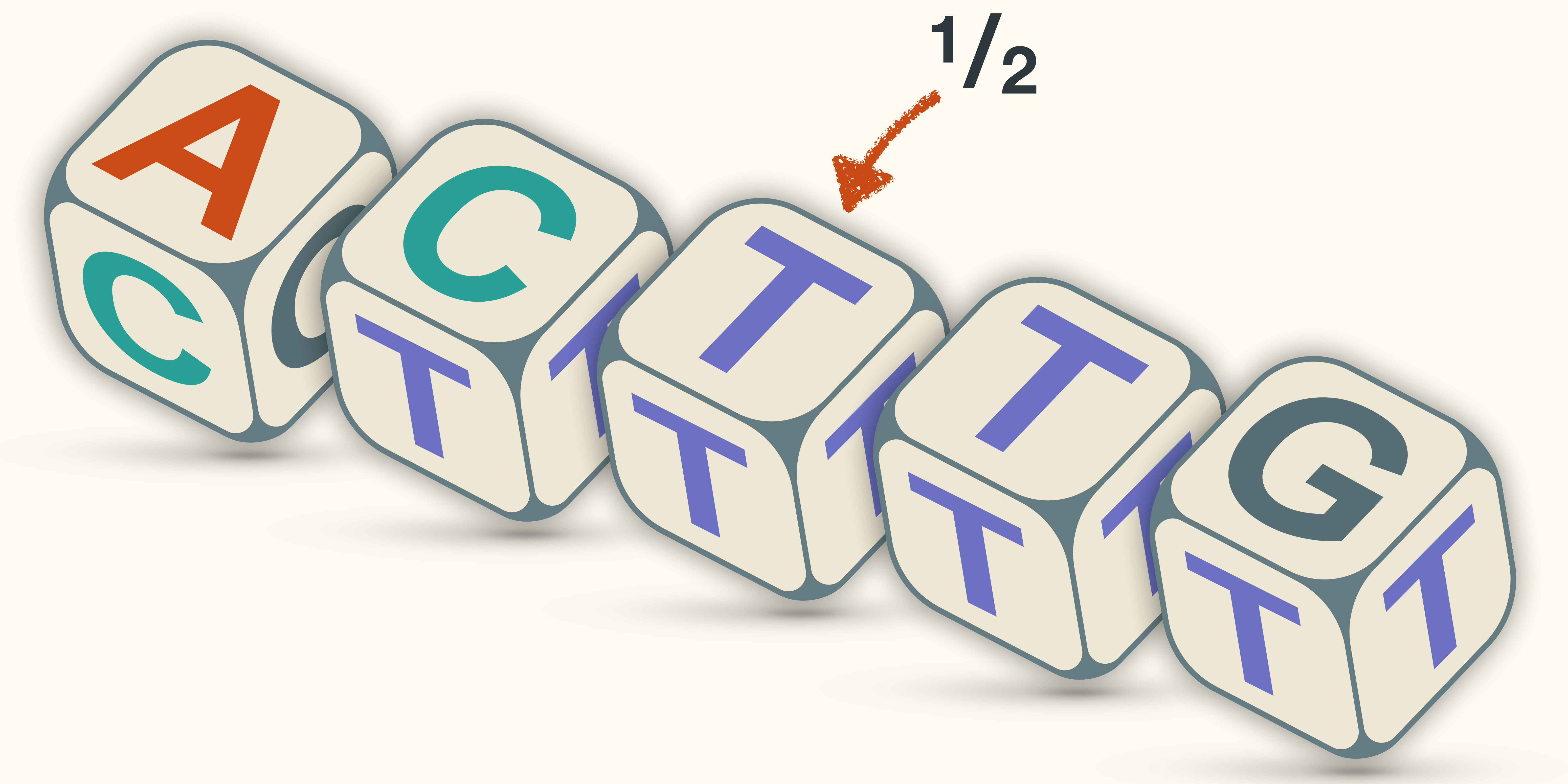
Probability



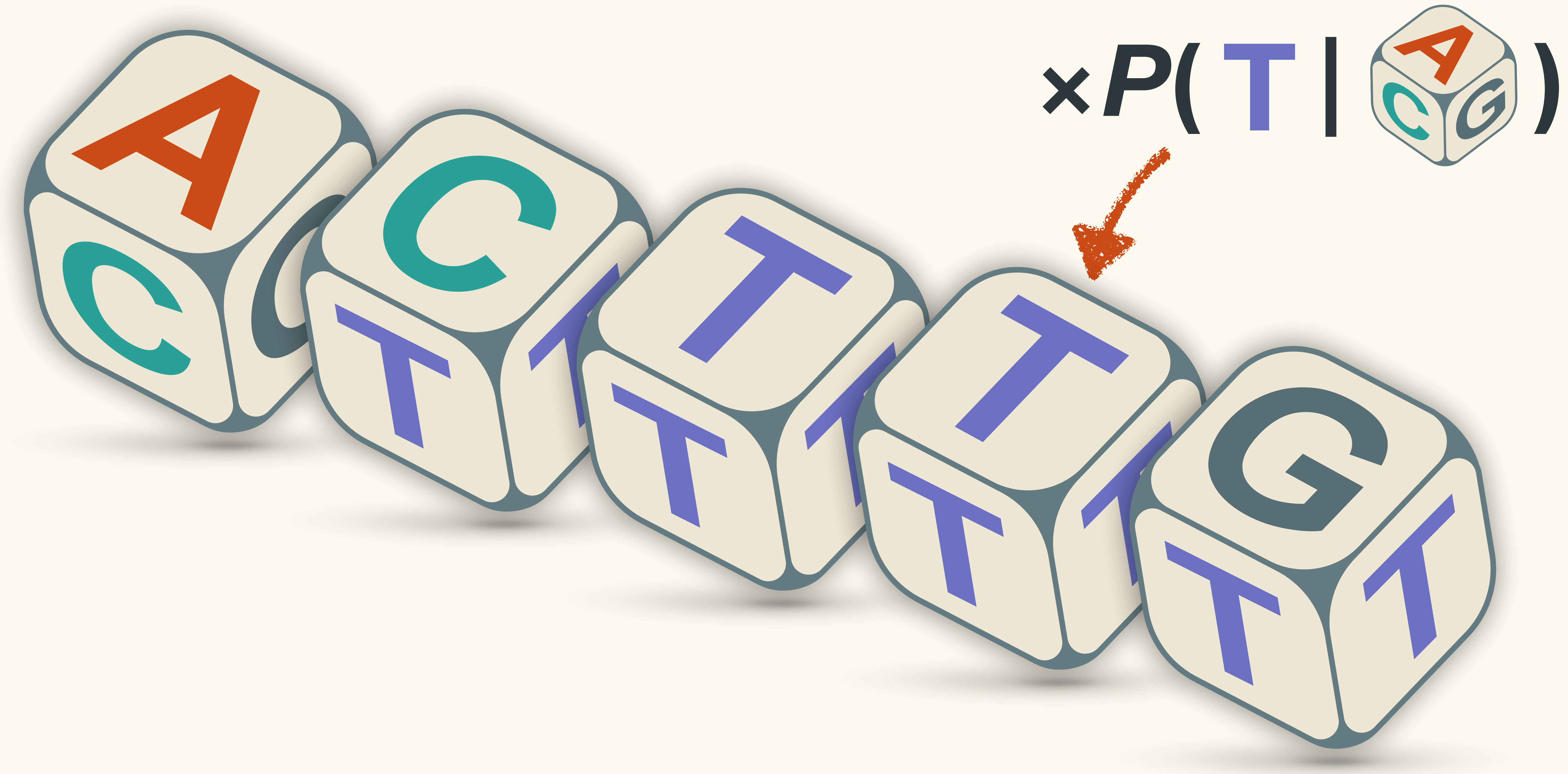
Probability



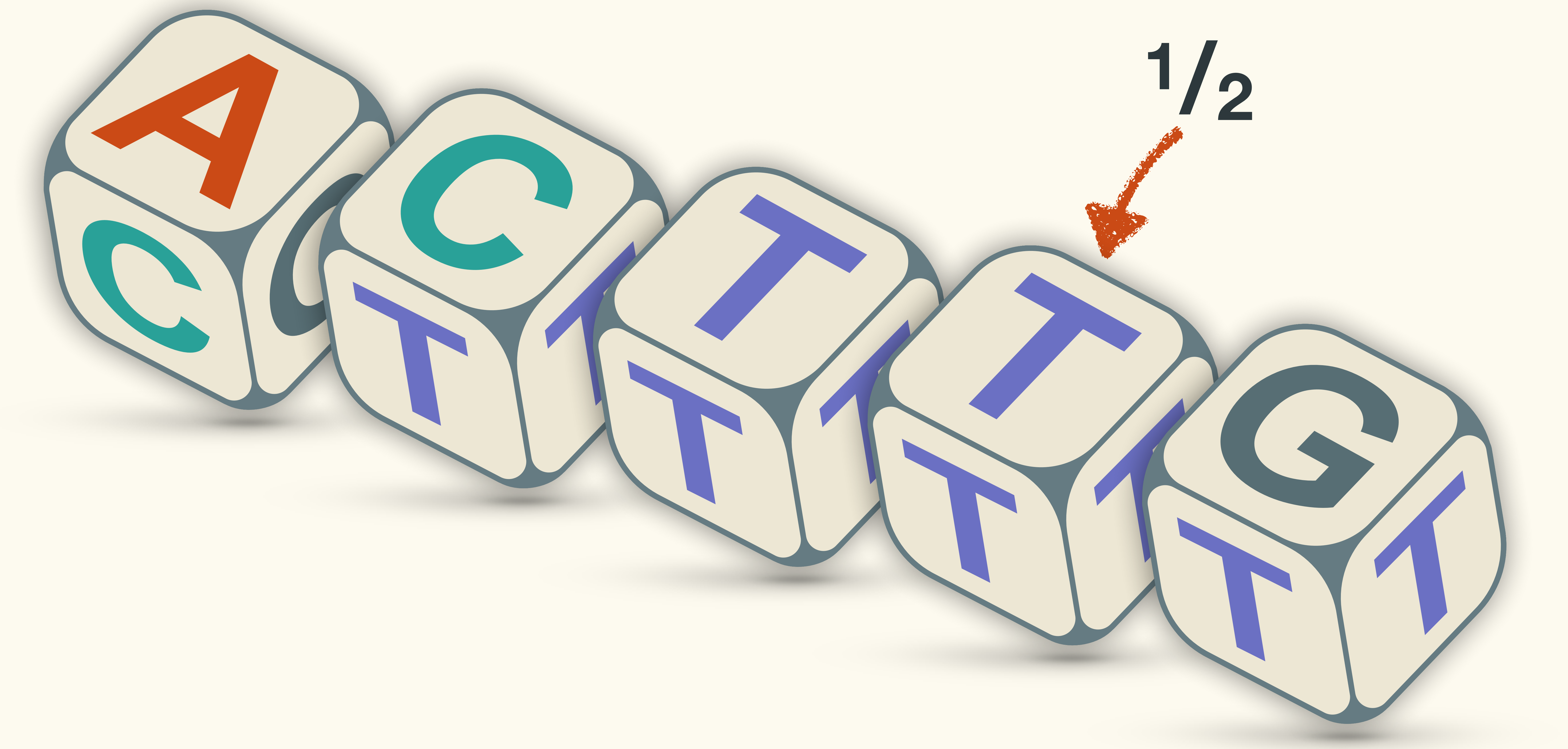
Probability



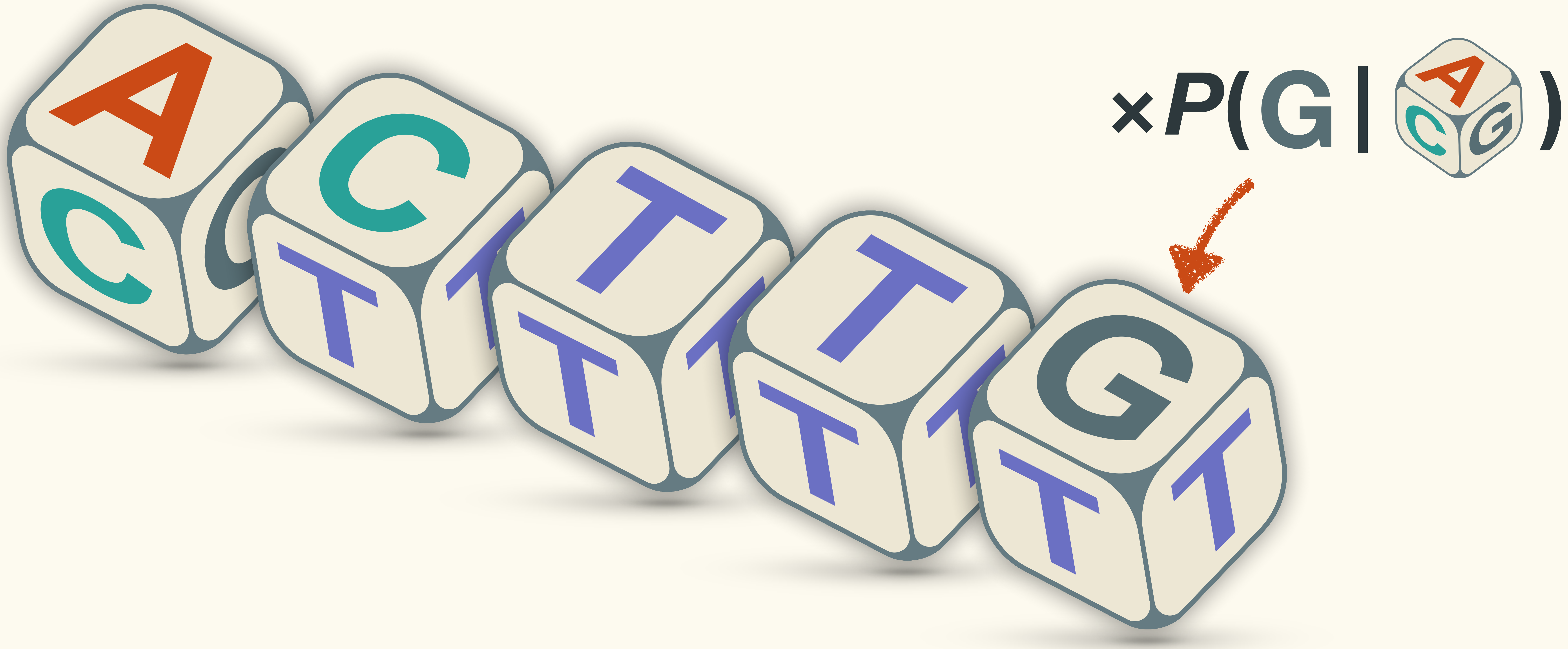
Probability



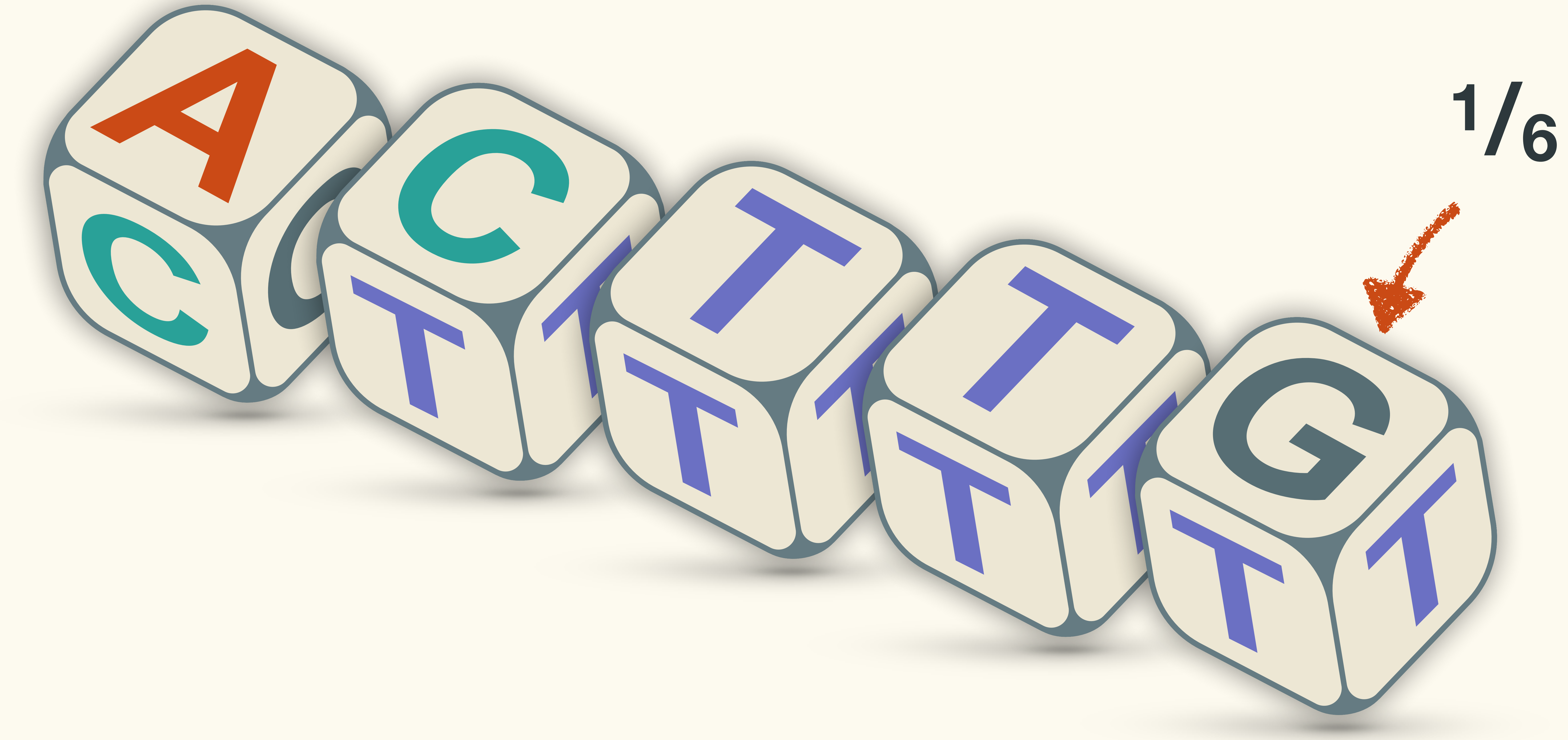
Probability



Probability

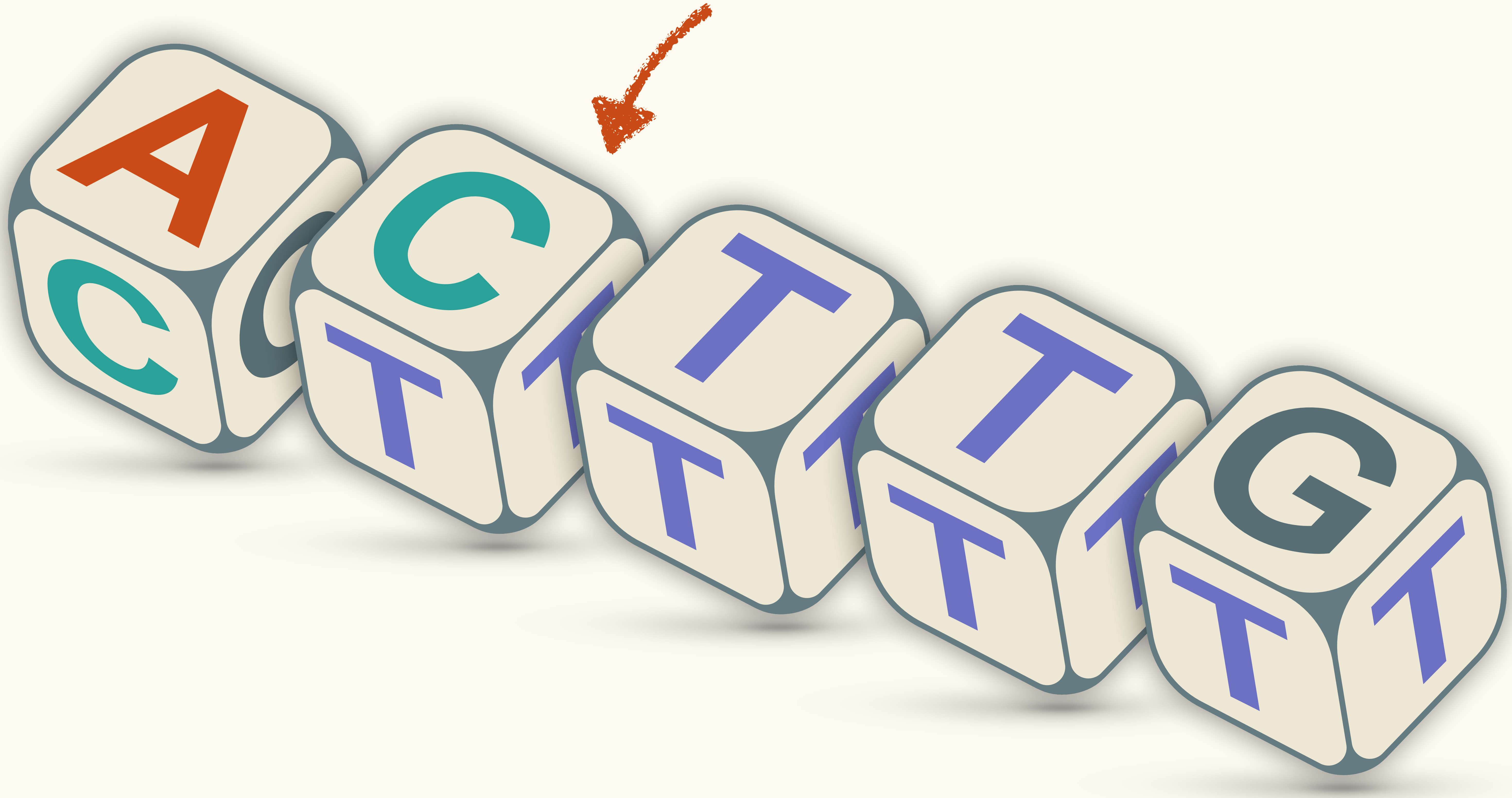


Probability



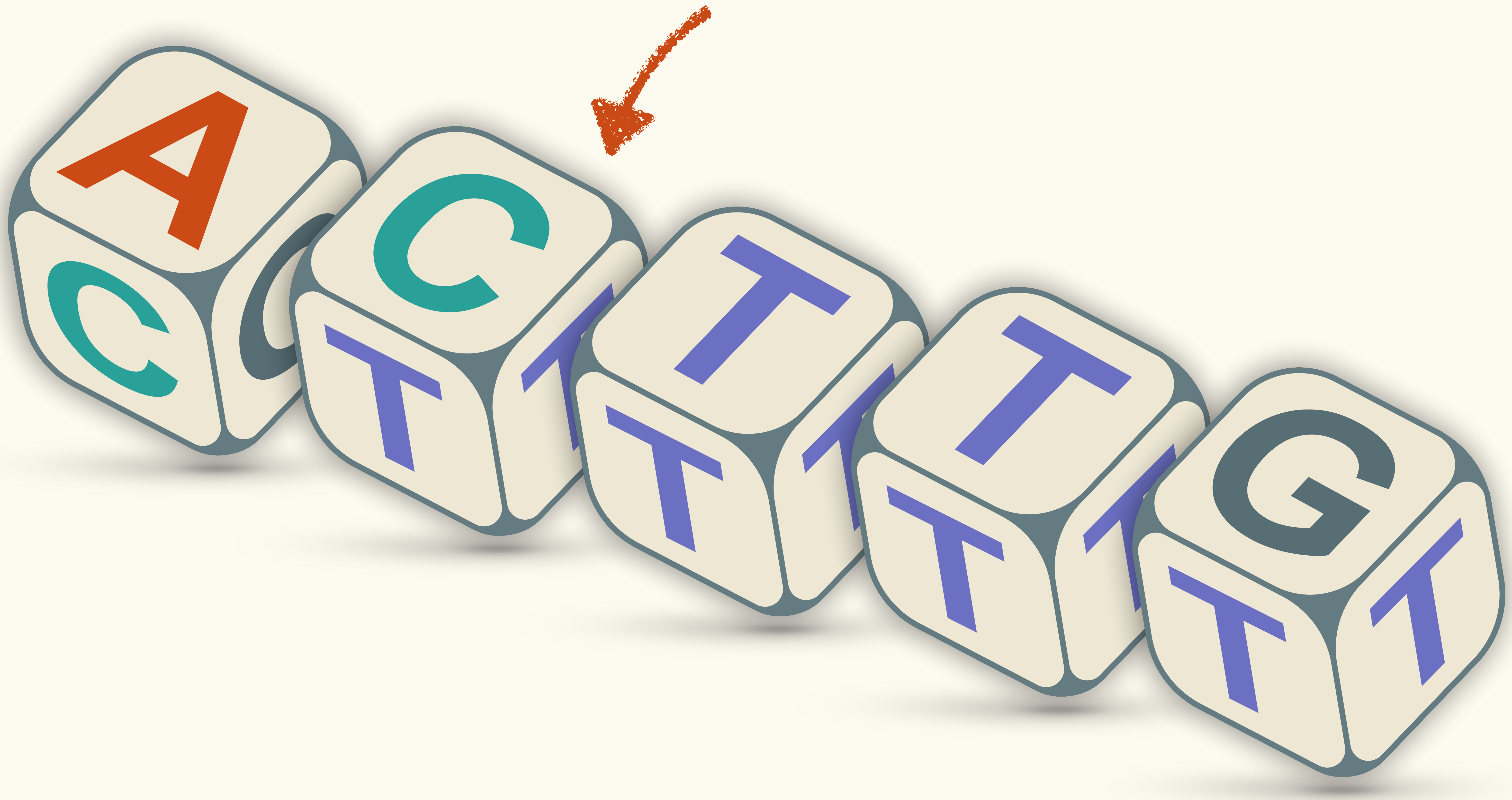
Probability

$P(A \,\&\, C \,\&\, T \,\&\, T \,\&\, G \mid$  $) = ?$



Probability

$P(A \,\&\, C \,\&\, T \,\&\, T \,\&\, G \mid$  $) = (^1/_6)^3 \, (^1/_2)^2$



Probability

$P($A & C & T & T & G $\mid$ 

$) = (^1/_4)^5$

$P($A & C & T & T & G $\mid$ 

$) = (^1/_6)^3 (^1/_2)^2$

Probability

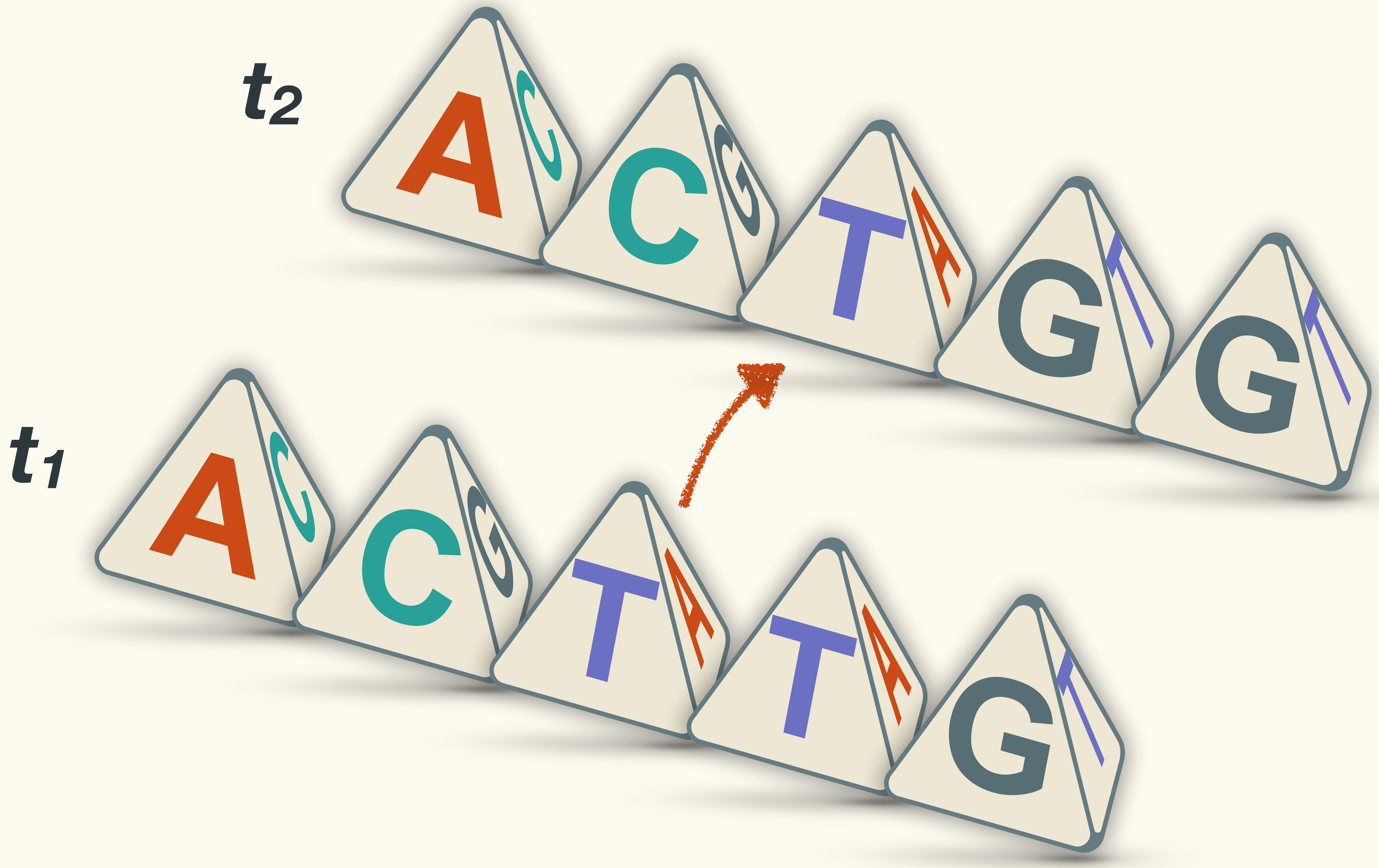
$P($ A & C & T & T & G | 

$) = 0.0010$

$P($ A & C & T & T & G | 

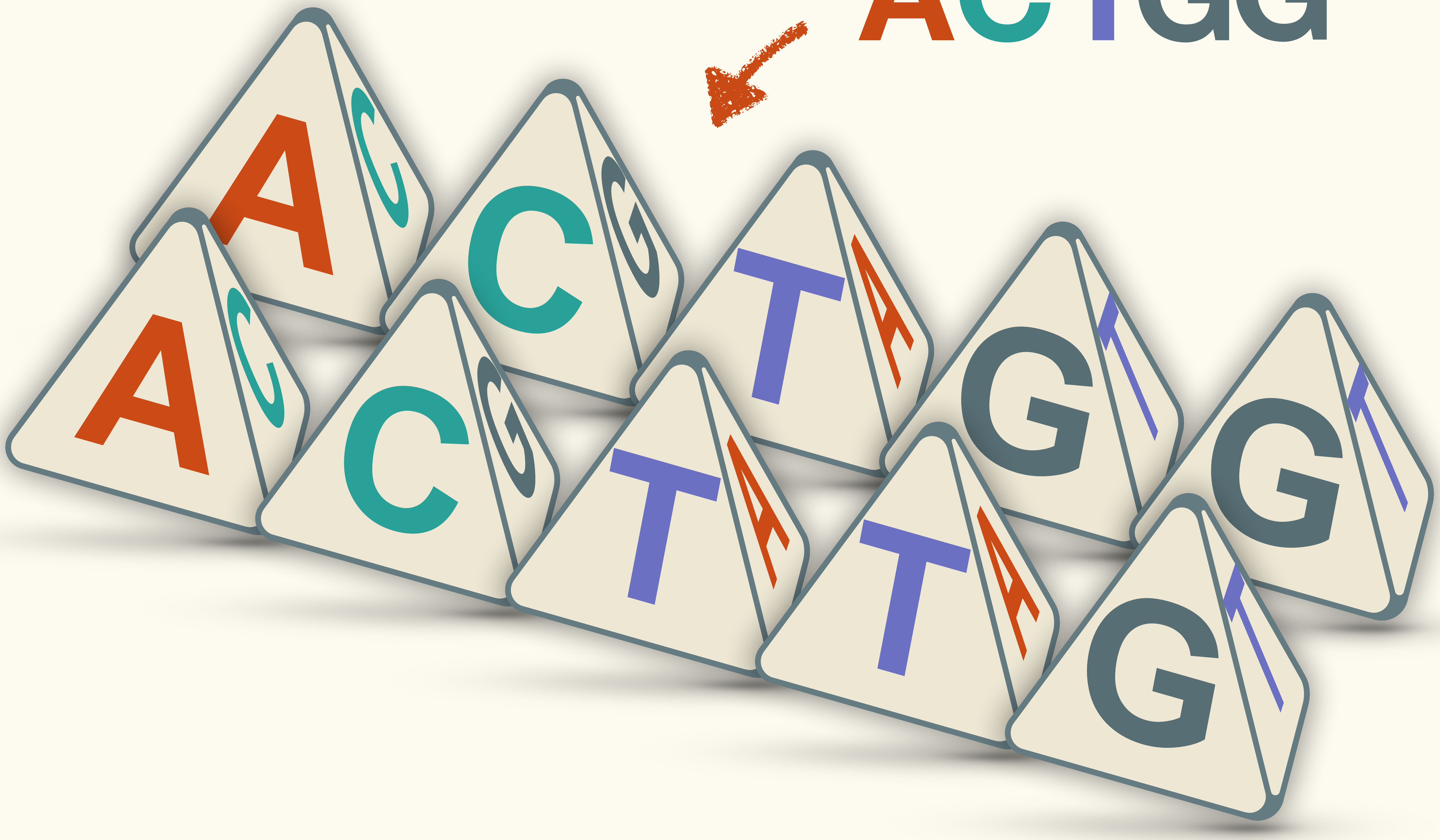
$) = 0.0012$

Probability

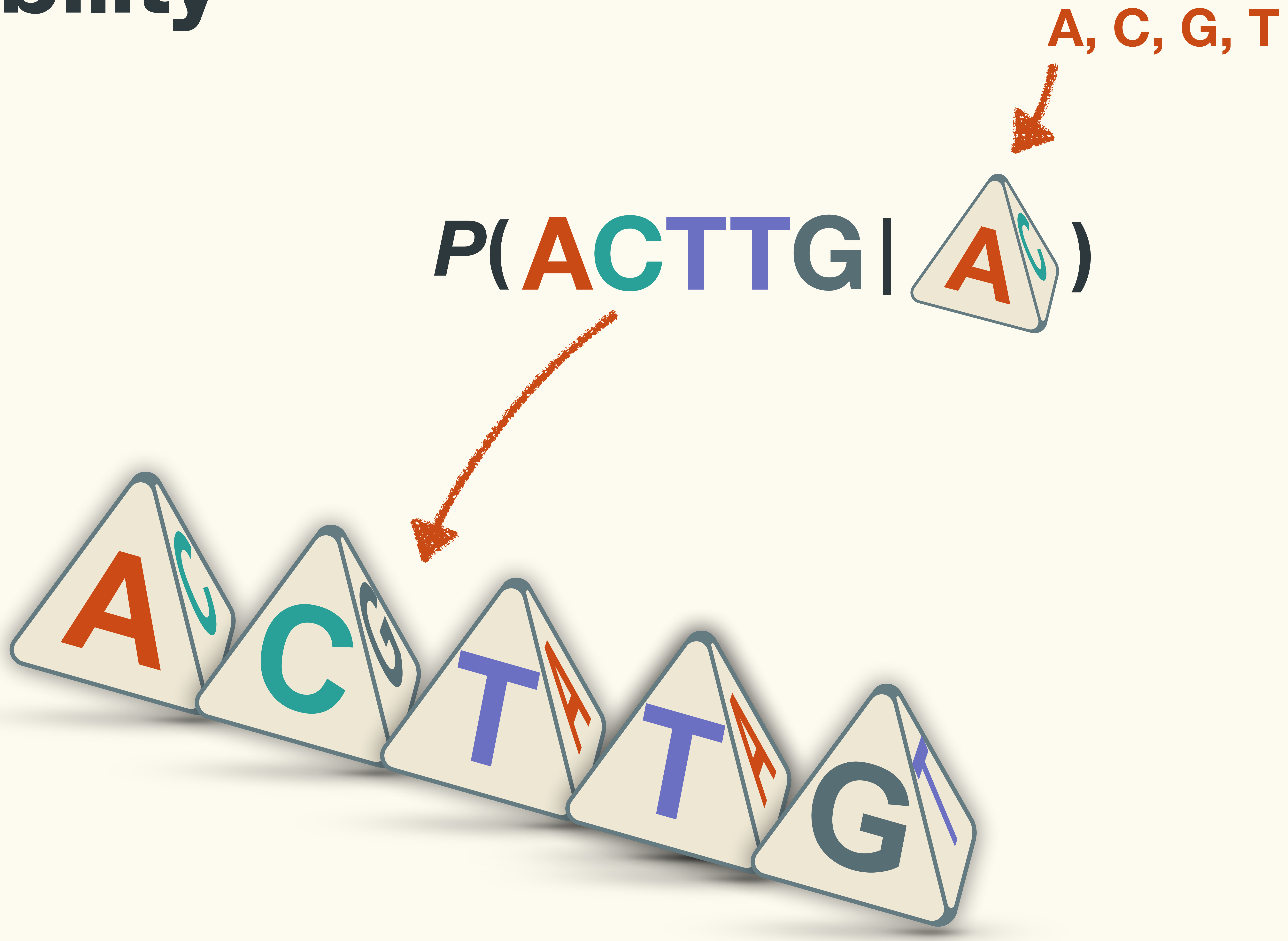


Probability

$$P\left(\begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right) = ?$$



Probability

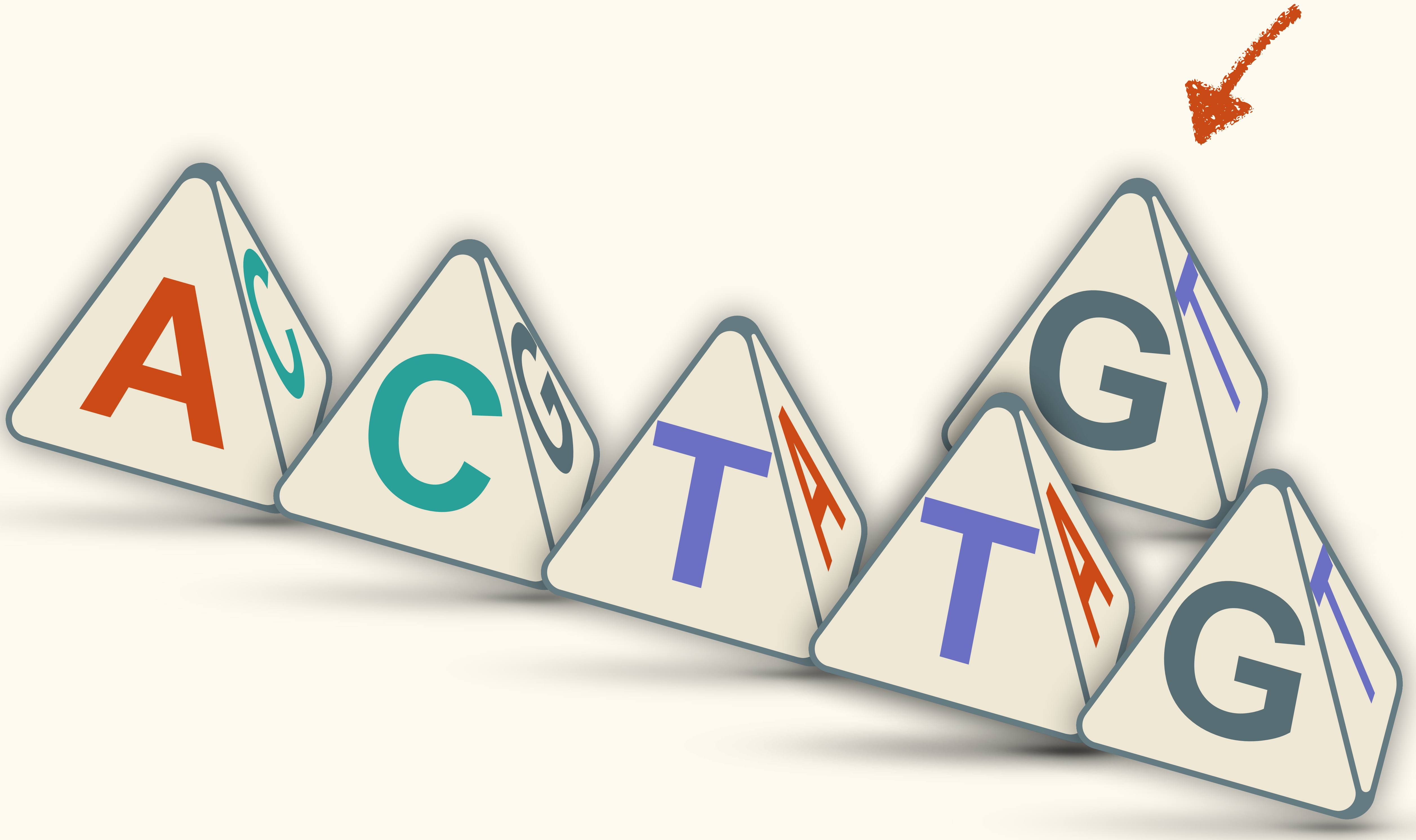


Probability

0.0010

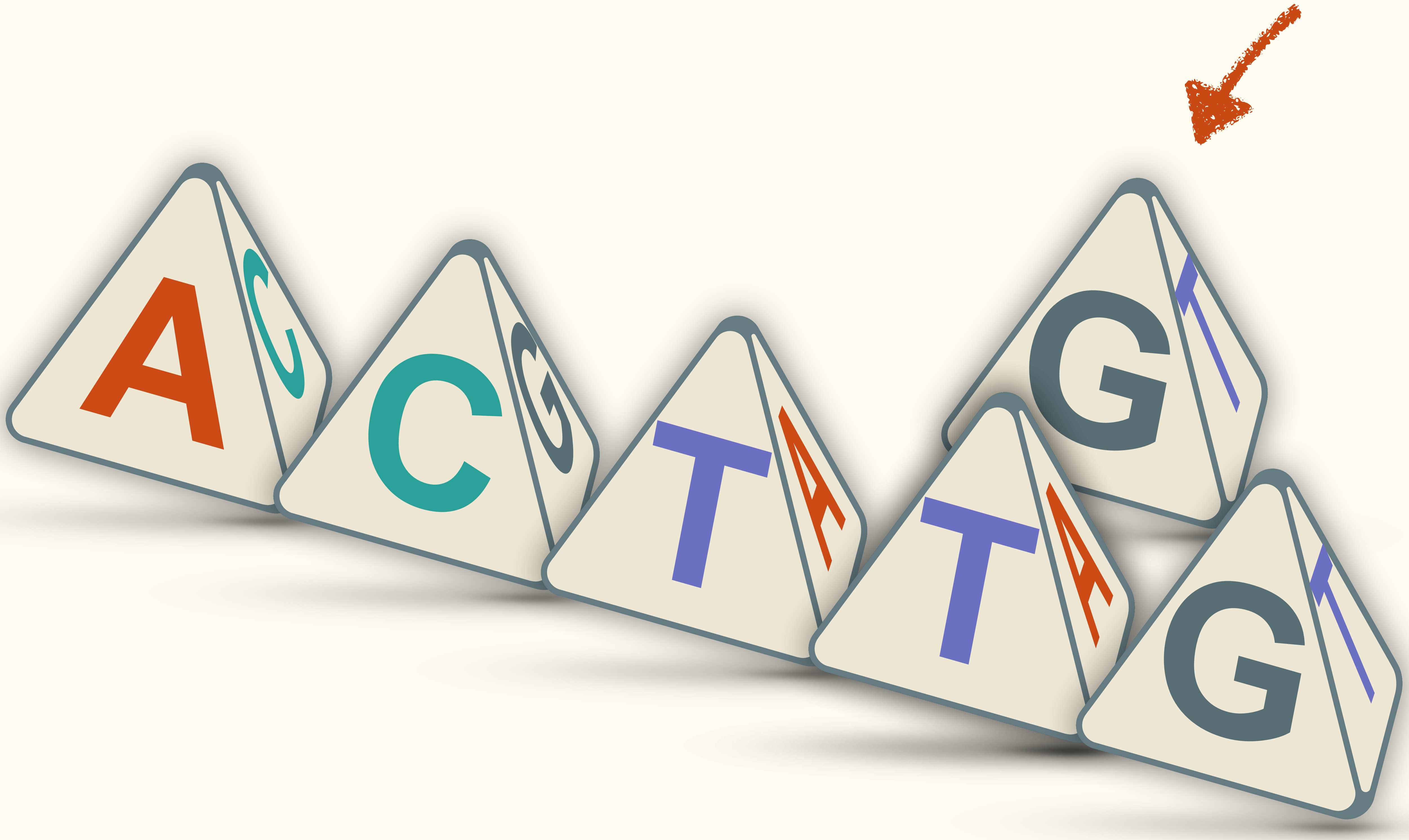
Probability

$\times P(\mathrm{T} \rightarrow \mathrm{G})$



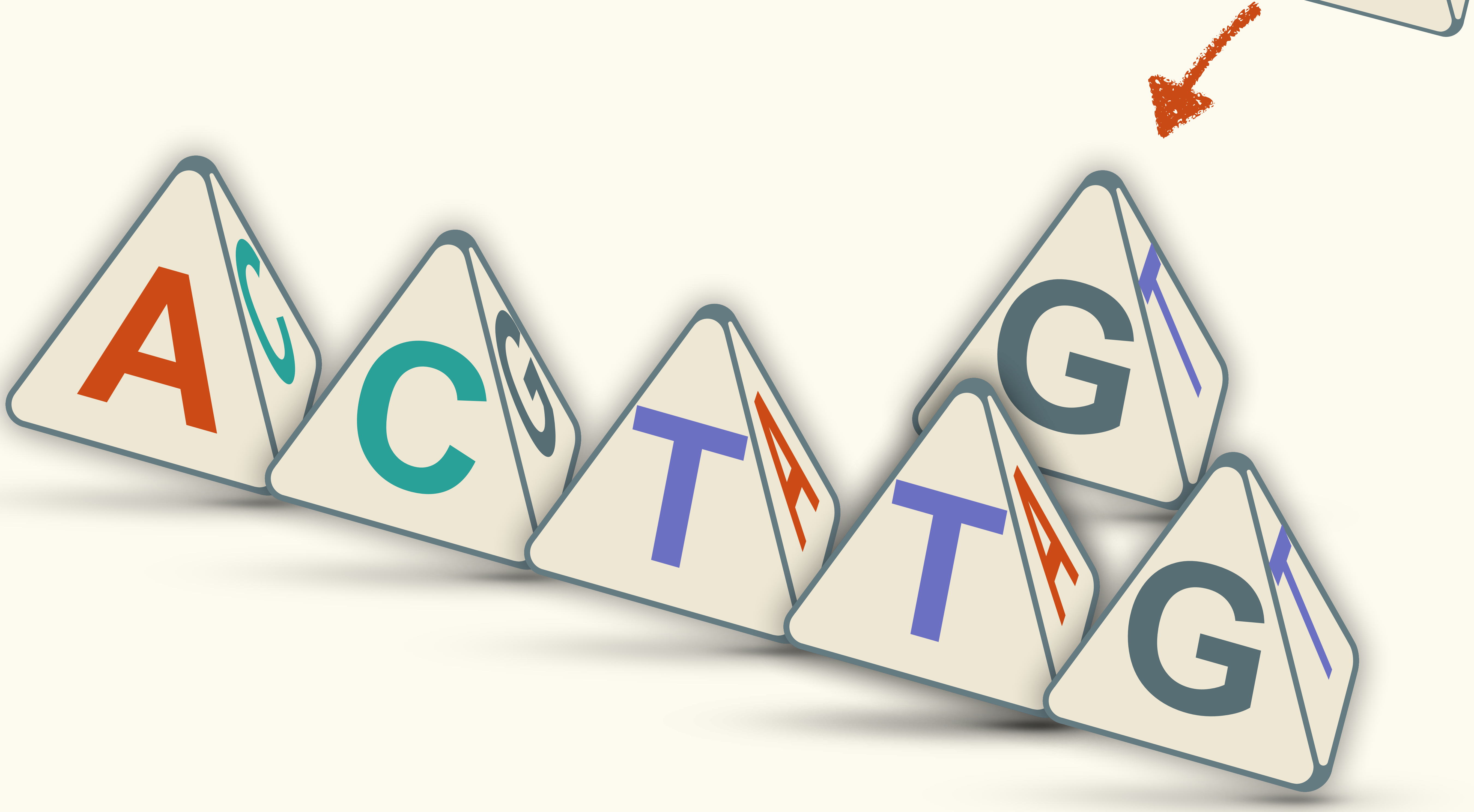
Probability

$\times P(T \to G \mid$  $)$

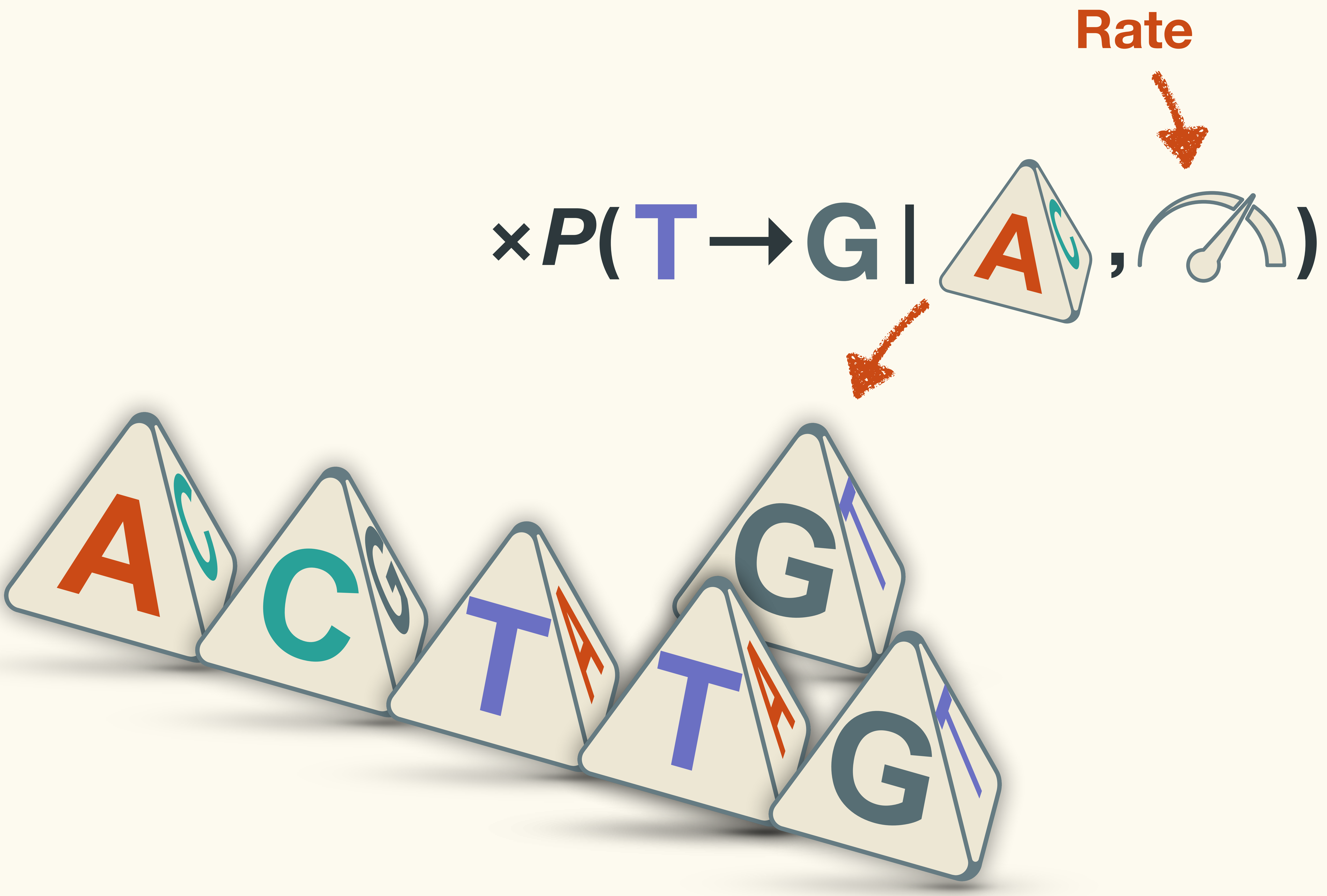


Probability

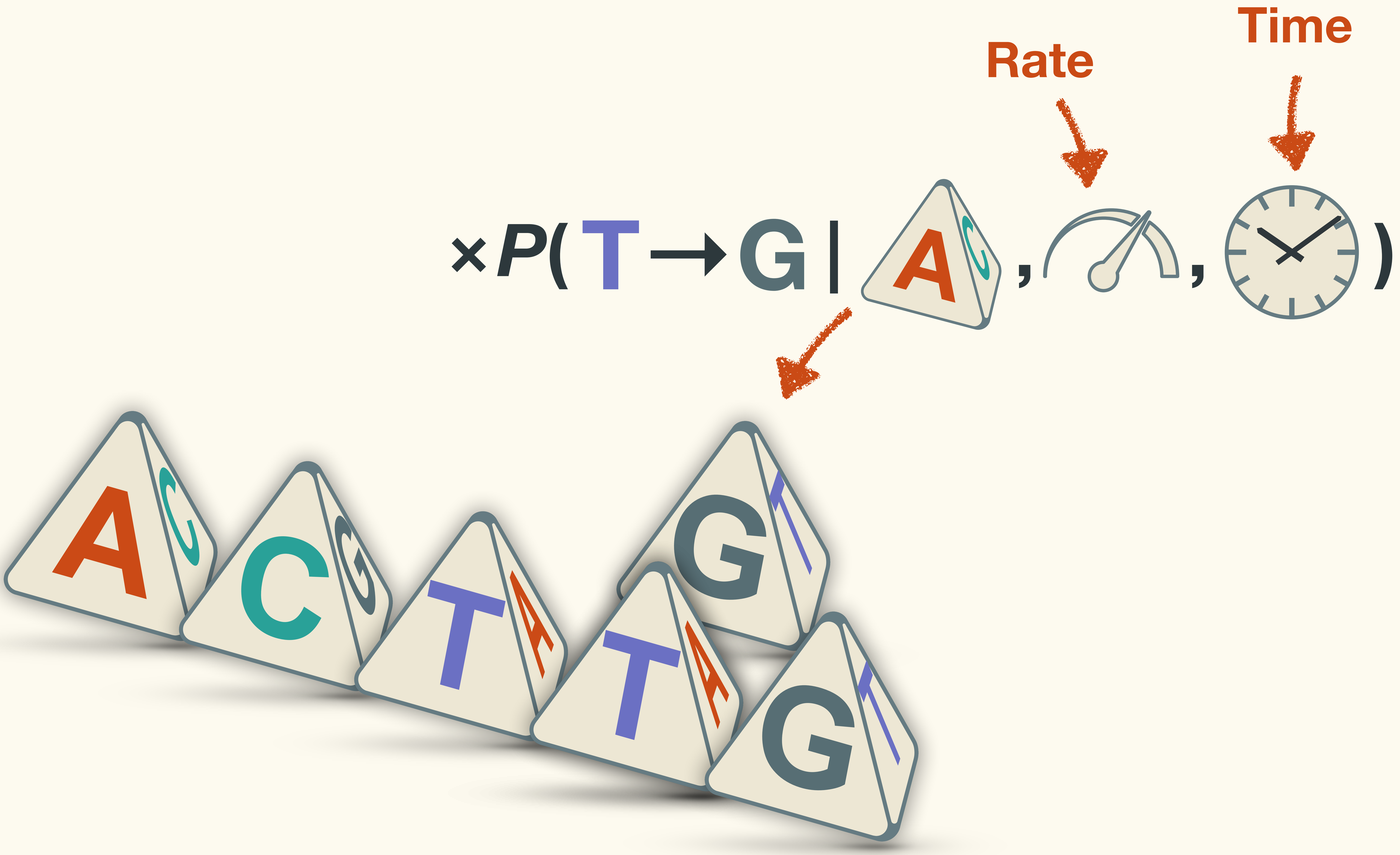
$\times P(\mathrm{T} \to \mathrm{G} \mid$  $, \textbf{?}\,)$



Probability



Probability



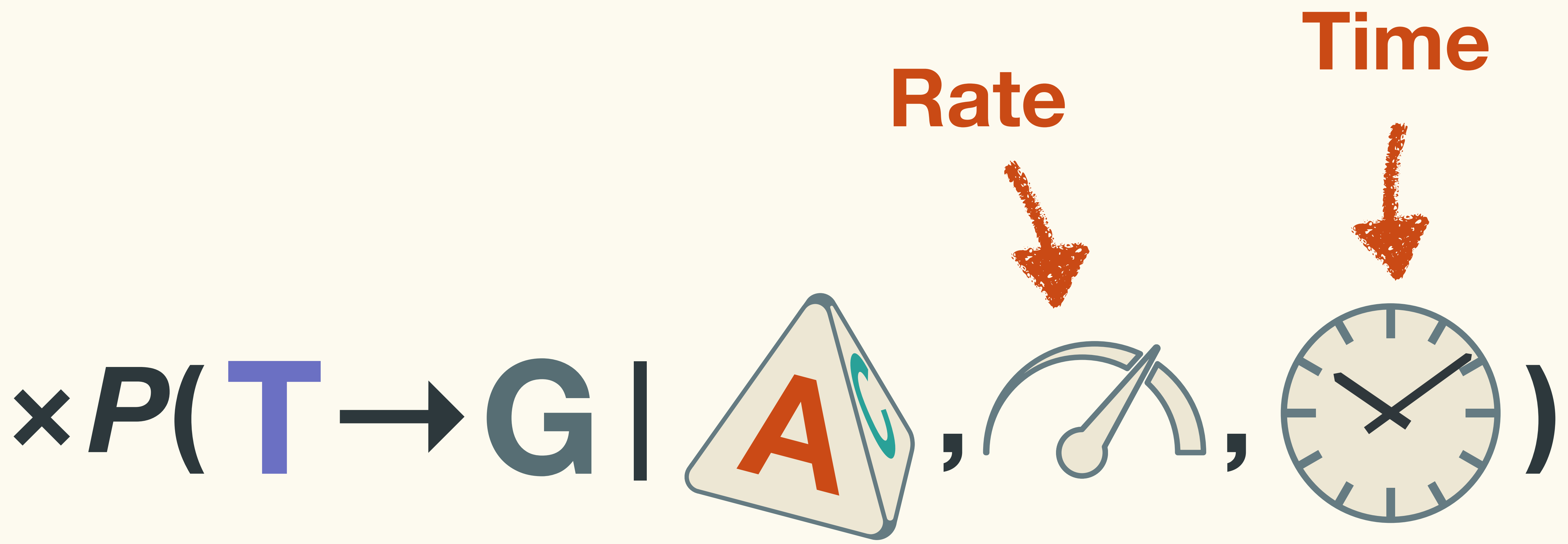
Probability

Time

Rate

$$(^{1}/_{4}) \times (1-e^{-4\alpha t})$$

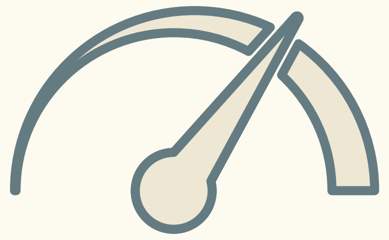
Probability



Probability

Rate

Time

$\times P($ T $\rightarrow$ G |  ,  = 1,  = 1 $)$

Probability

$$\times P(\mathrm{T} \to \mathrm{G} \mid \boxed{\mathrm{M}_1})$$

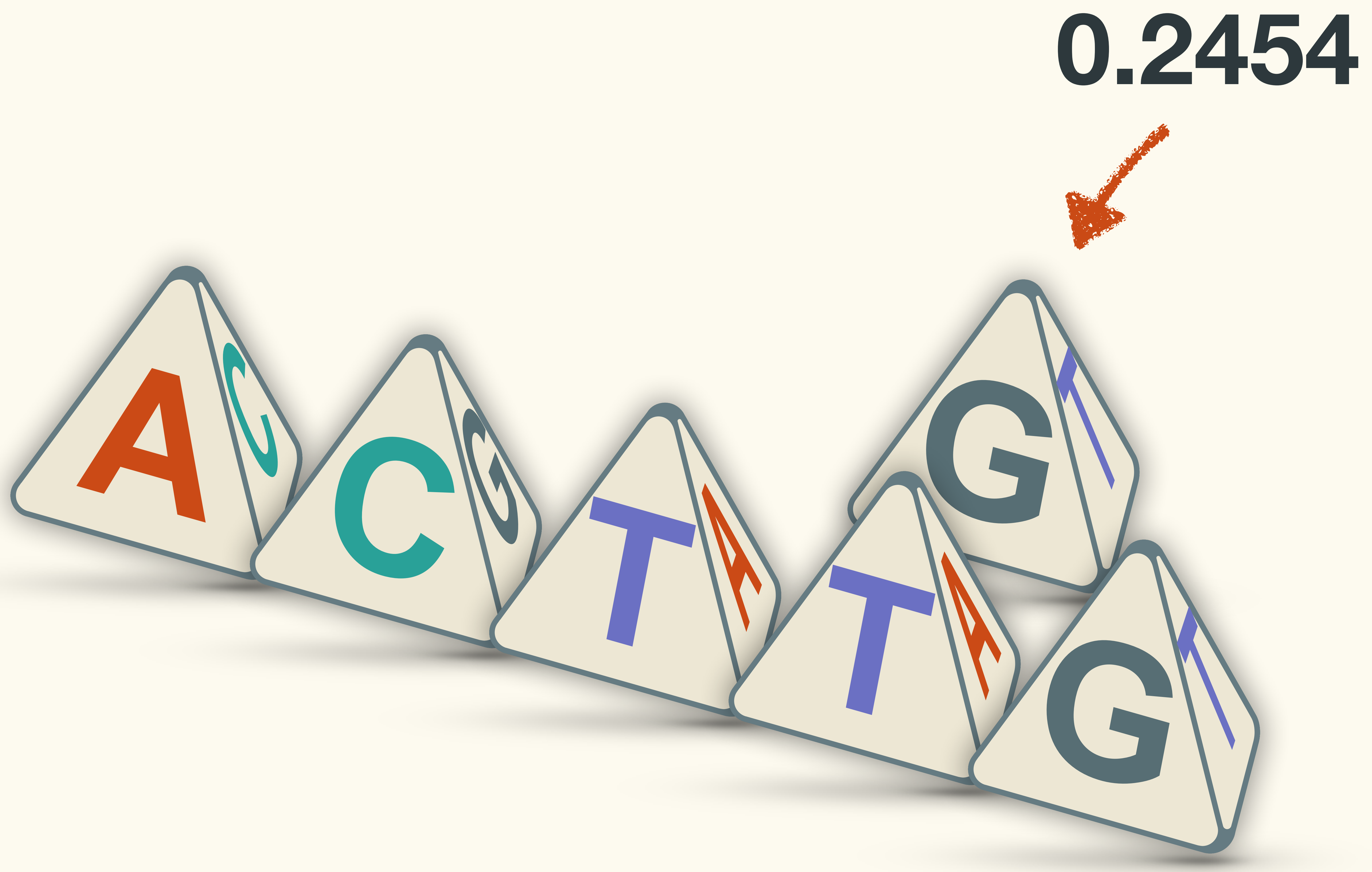
Probability

Time

Rate

$$(^1/_4) \times (1-e^{-4\alpha t})$$

Probability

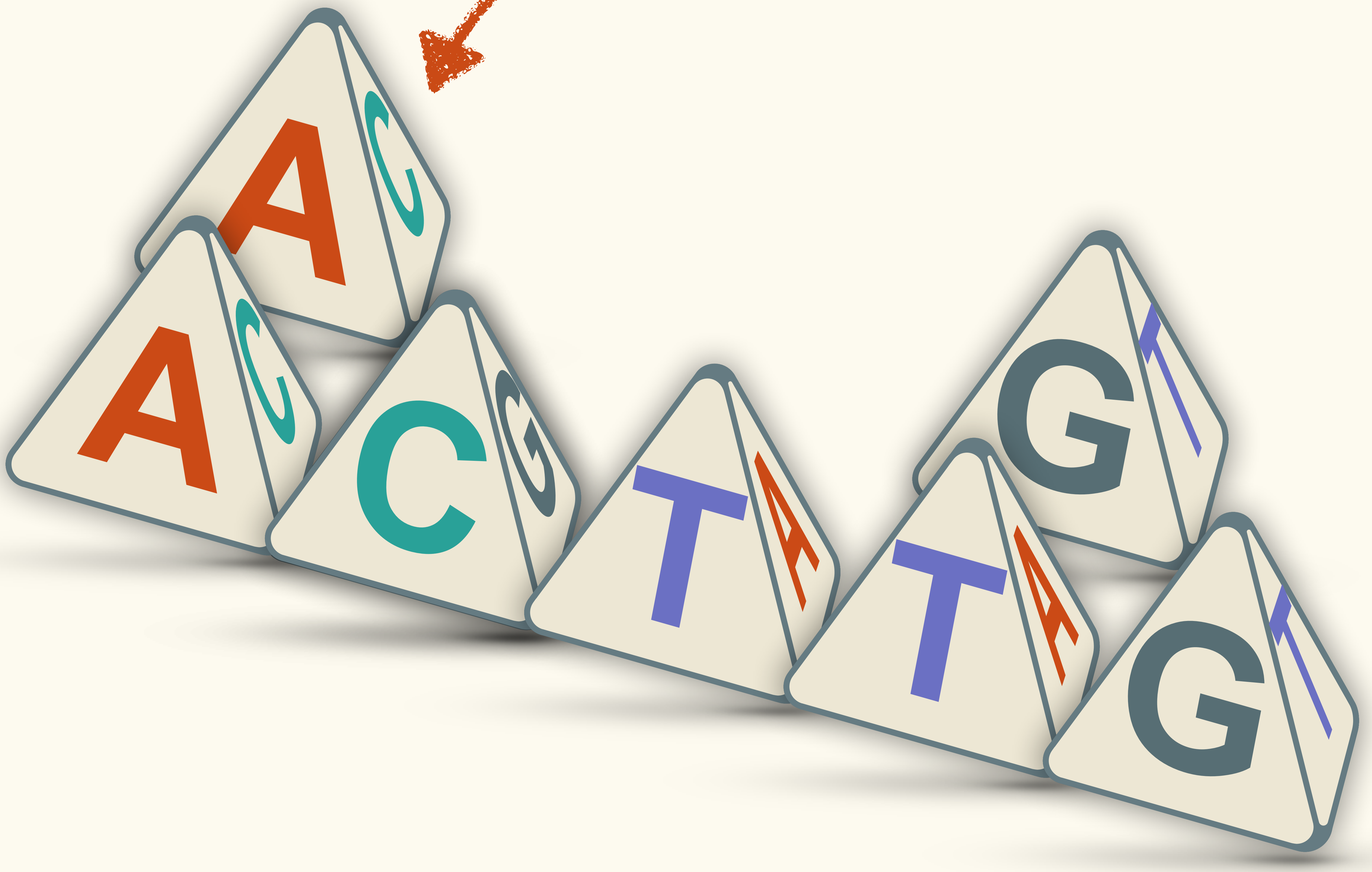


Probability

$$P\left(\begin{matrix}\mathrm{ACTTG}\\ \mathrm{ACTGG}\end{matrix} \,\middle|\, \boxed{M_1}\right) = ?$$

Probability

$$\times P(\mathrm{A} \rightarrow \mathrm{A} \mid \boxed{M_1})$$



Probability

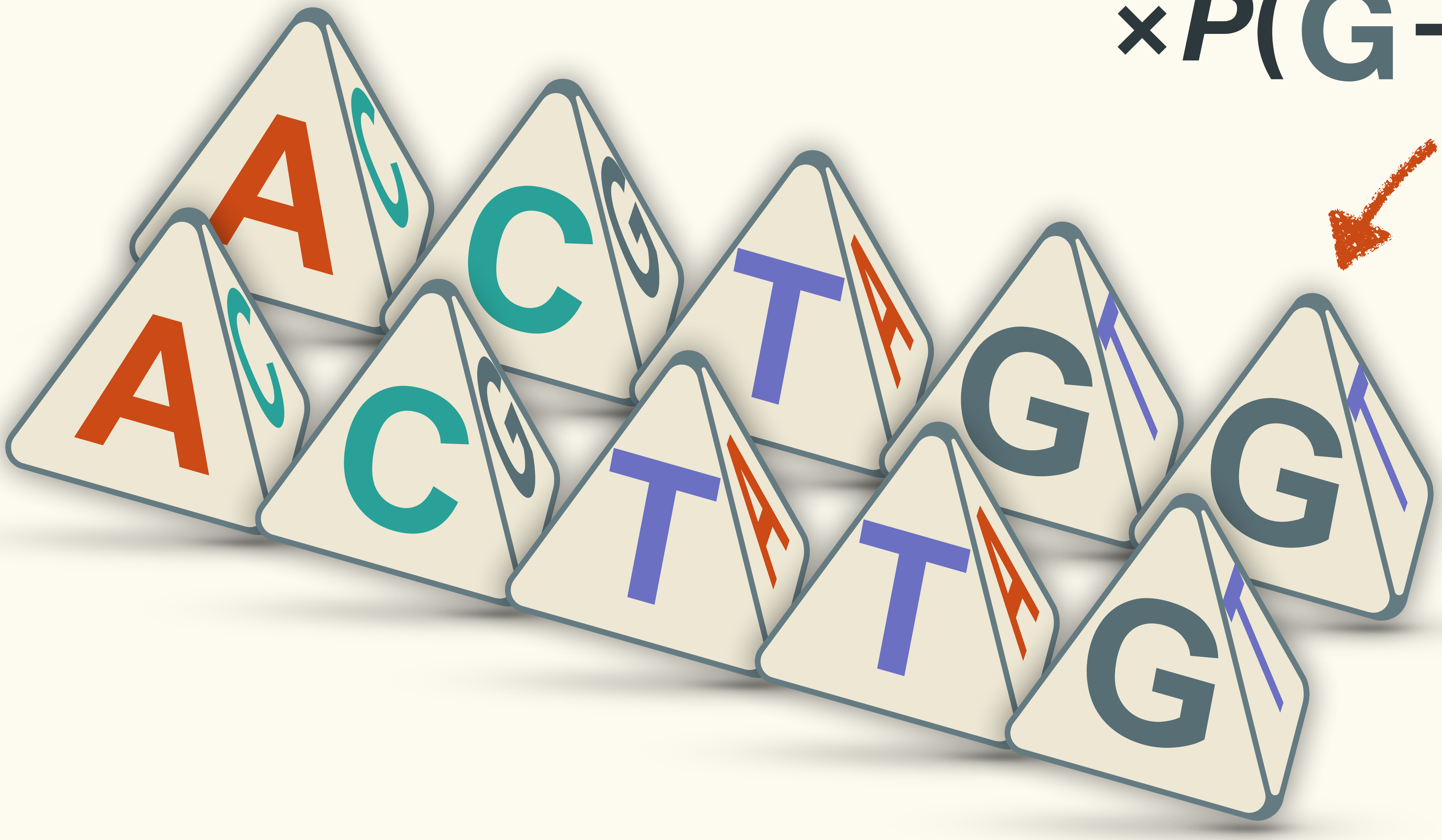
$$\times P(C \rightarrow C \mid \boxed{M_1})$$

Probability

$$\times P(\mathrm{T} \rightarrow \mathrm{T} \mid \boxed{M_1})$$

Probability

$$\times P(G \rightarrow G \mid \boxed{M_1})$$



Probability

$$\times P(\mathrm{A} \rightarrow \mathrm{A} \mid \boxed{M_1})$$

Probability

Rate

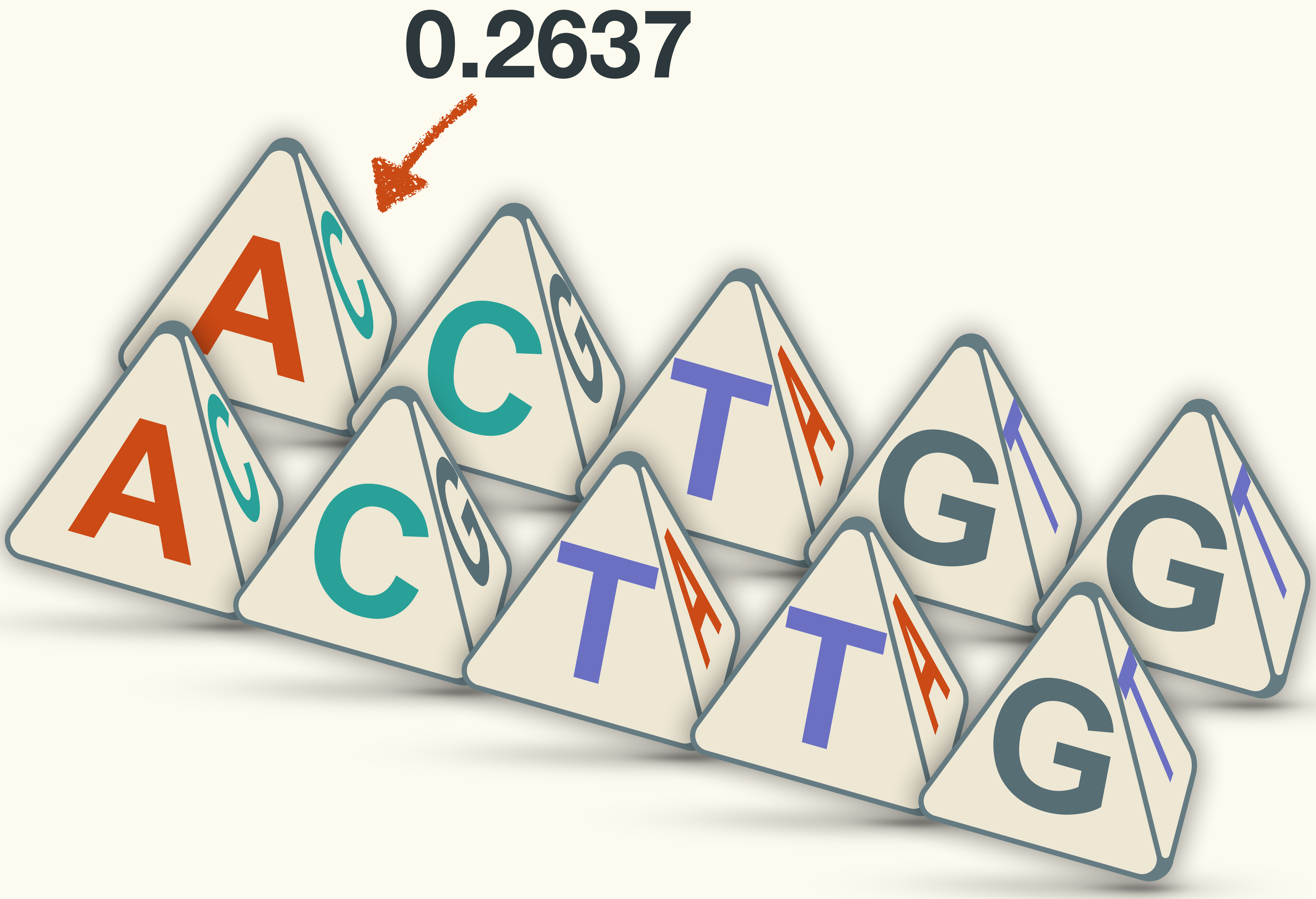
$$\left(\tfrac{1}{4}\right) \times \left(1-e^{-4\alpha t}\right) + e^{-4\alpha t}$$

Probability

Time

$$(^1/_4) \times (1-e^{-4\alpha t}) + e^{-4\alpha t}$$

Probability

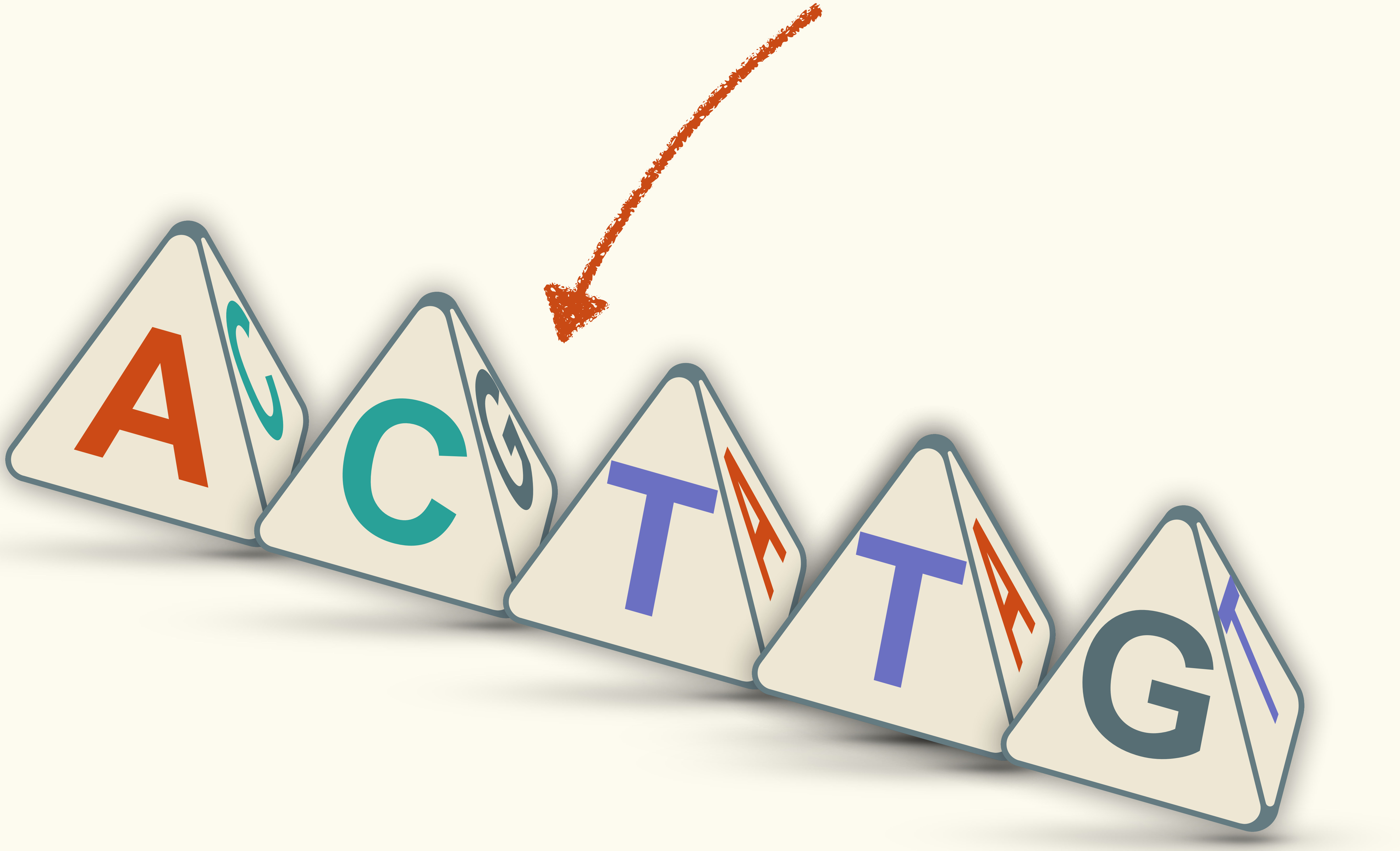


Probability

$$P\left(\begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix} \,\middle|\, M_1\right) = ?$$

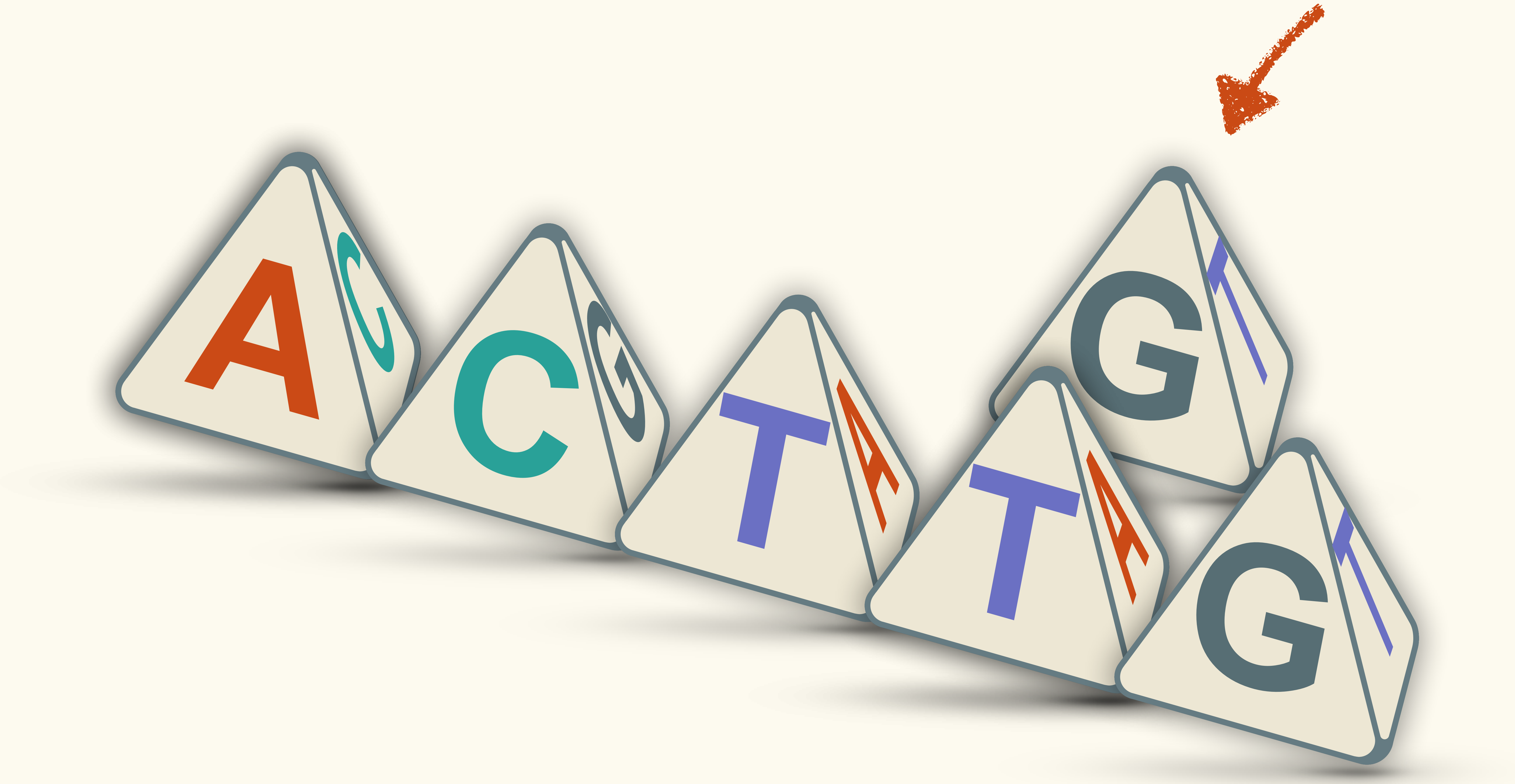
Probability

0.0010



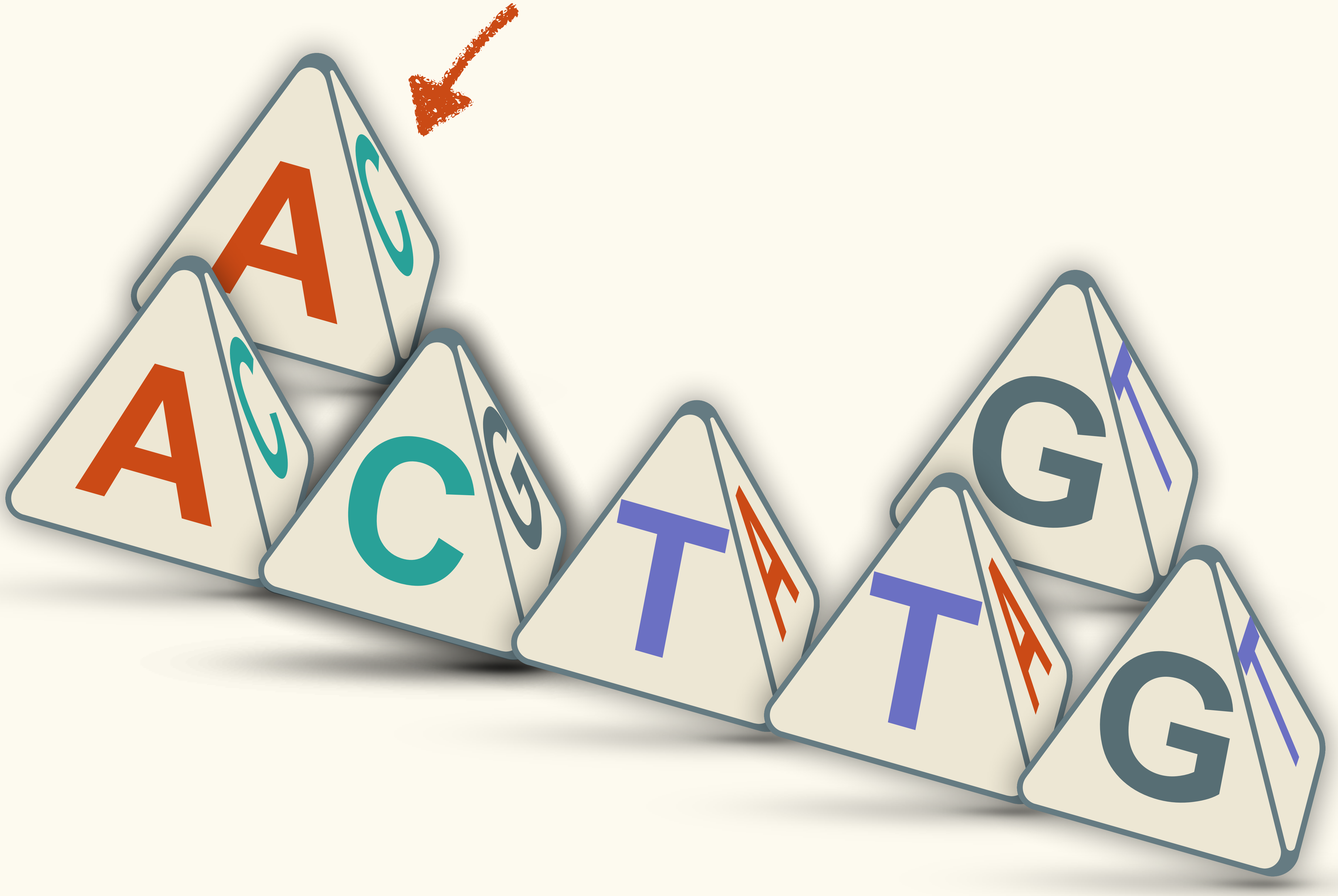
Probability

× 0.2454

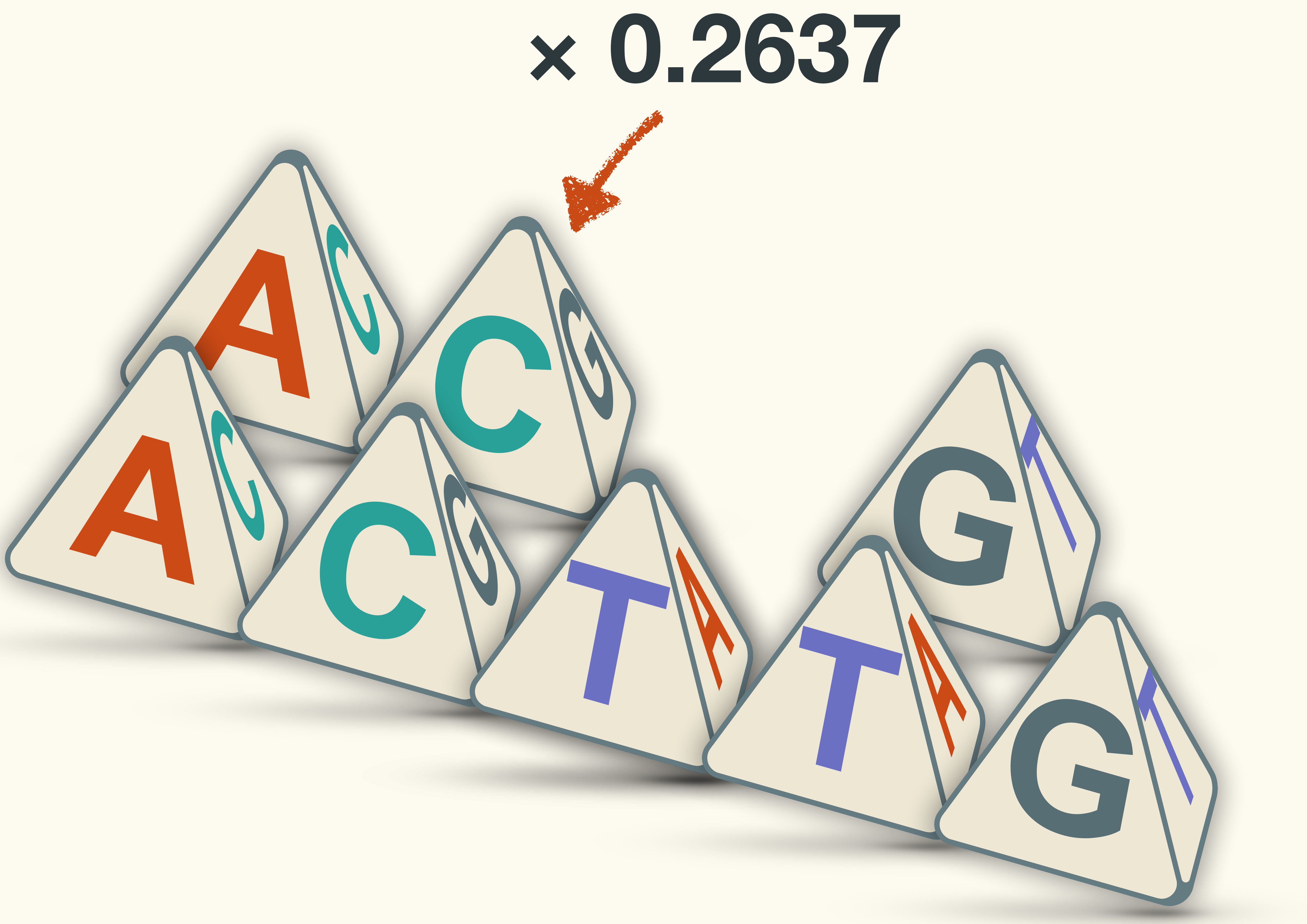


Probability

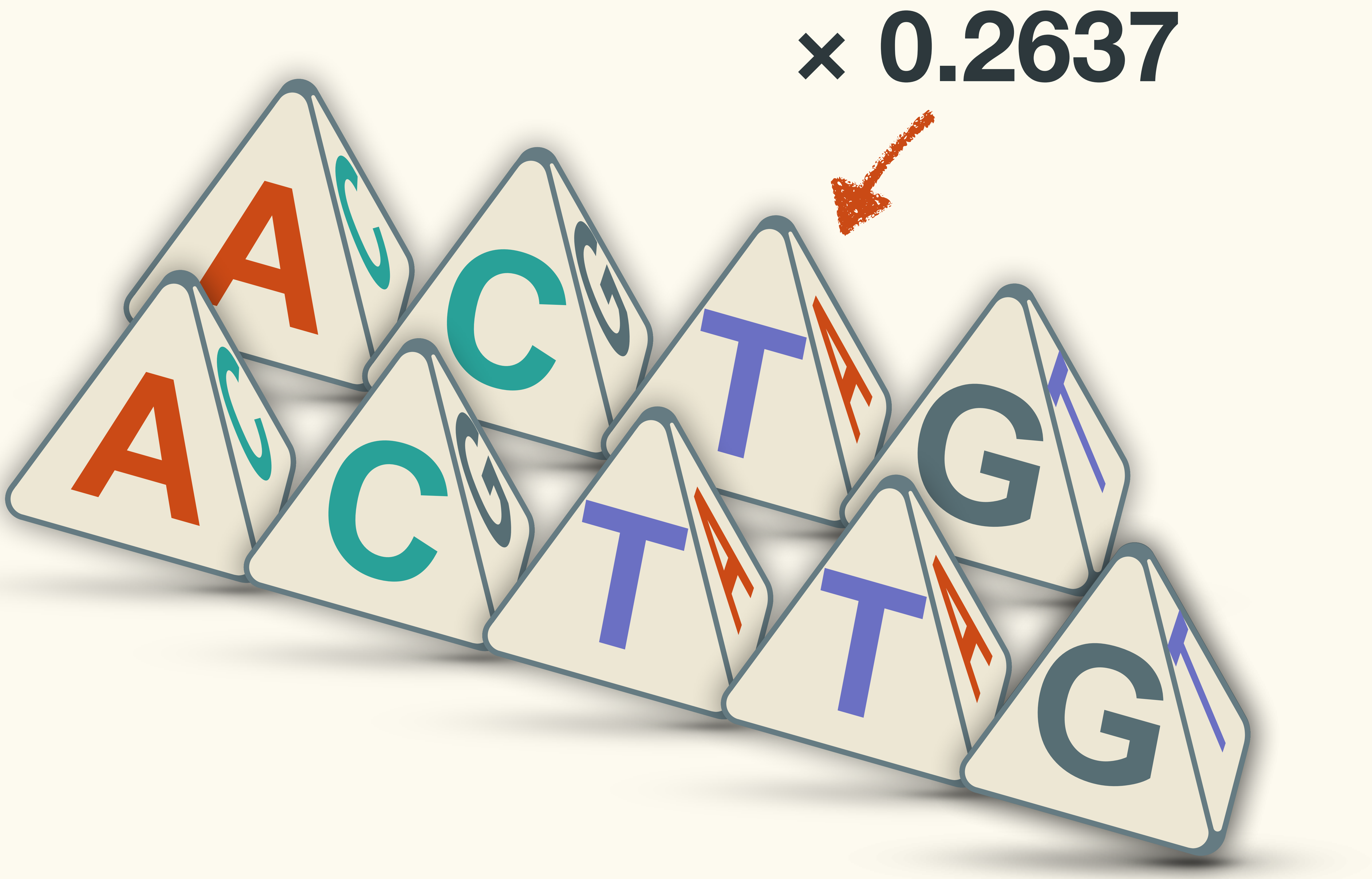
× 0.2637



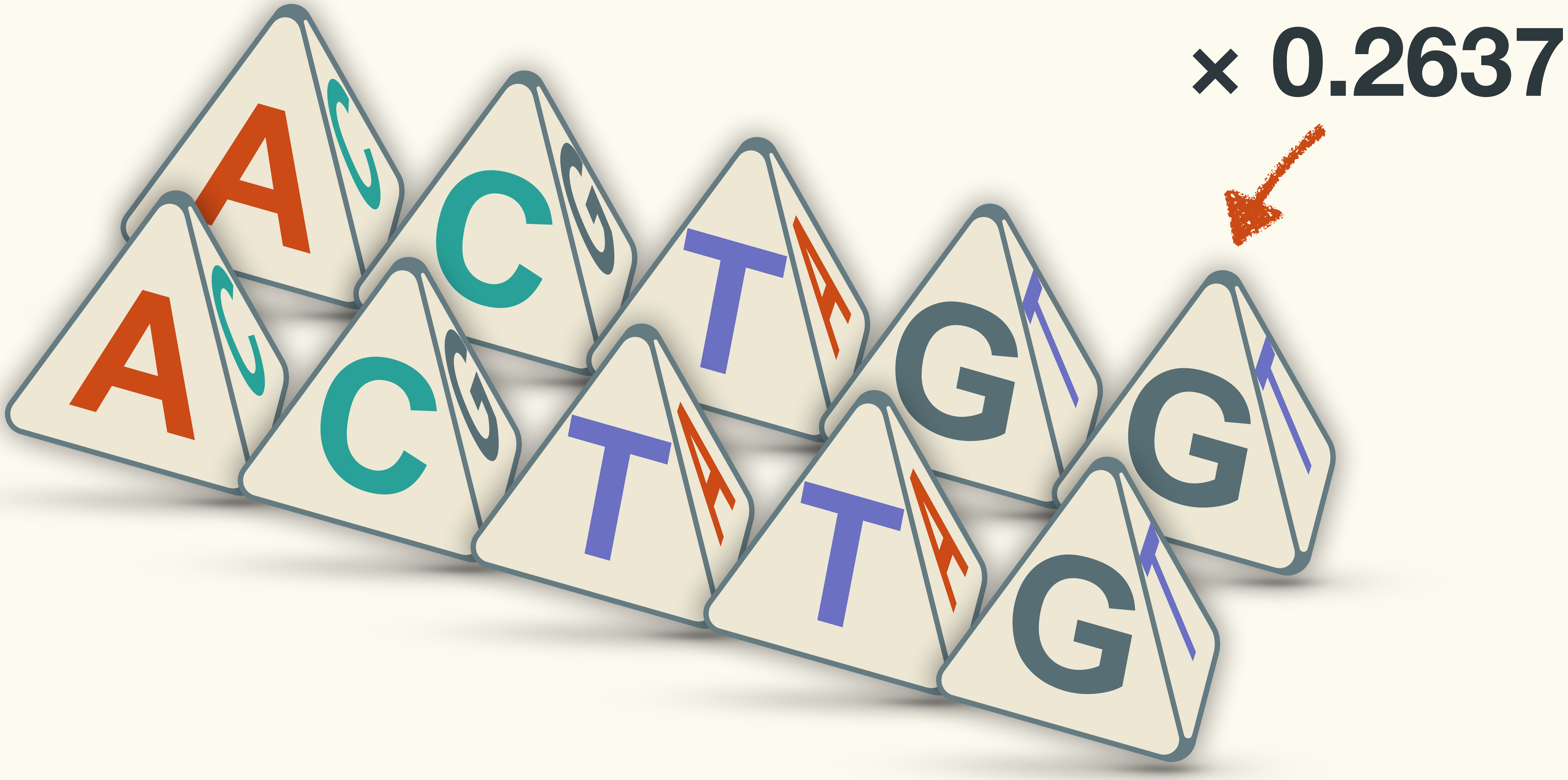
Probability



Probability

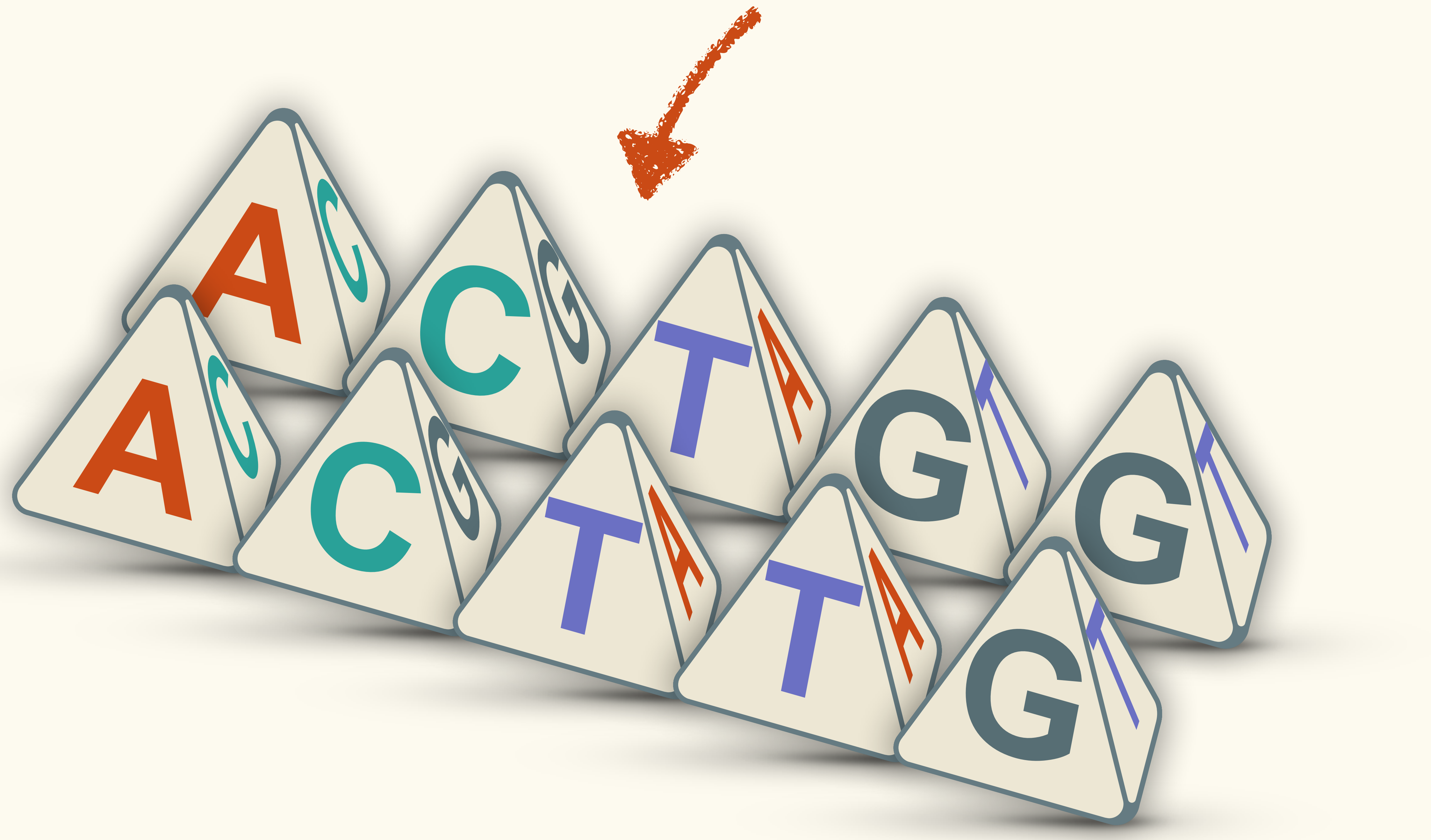


Probability



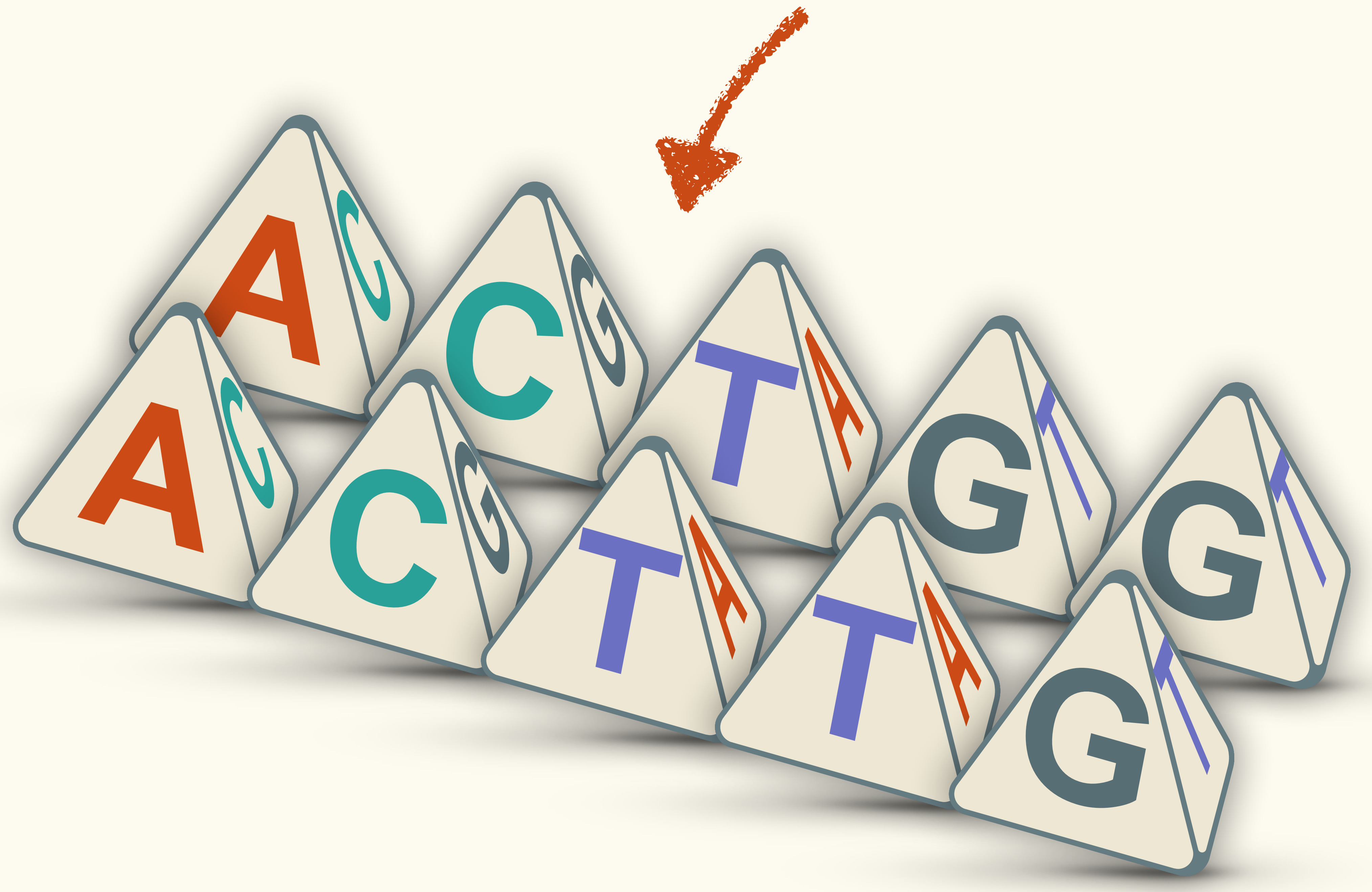
Probability

$$P\left(\begin{matrix}\text{ACTTG}\\\text{ACTGG}\end{matrix} \,\middle|\, \boxed{M_1}\right) = 0.0000015$$



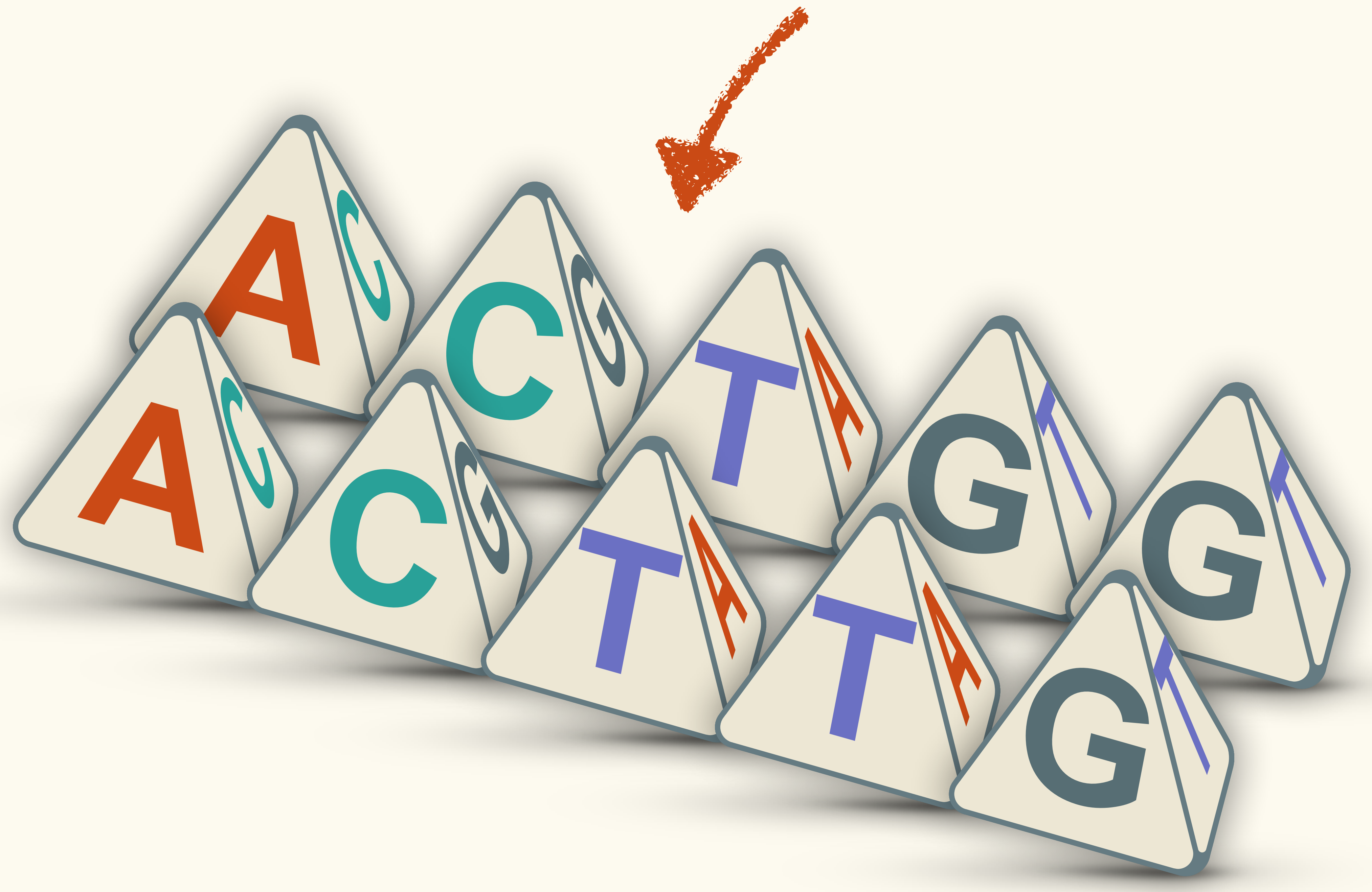
Likelihood

$$L\left(\boxed{M_1} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right) = 0.0000015$$



Likelihood

$$\log\left(L\left(\boxed{M_1} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -13.4$$

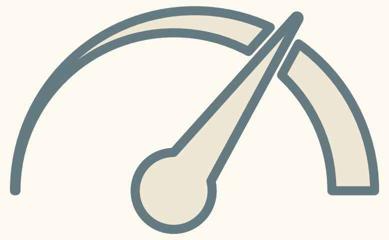


Likelihood

$$\log\left(L\left(\boxed{M}_1 \,\middle|\, \begin{matrix}\mathrm{ACTTG}\\ \mathrm{ACTGG}\end{matrix}\right)\right) = -13.4$$

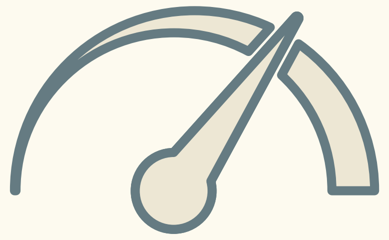
Likelihood

$$\log\left(L\left(\boxed{M_2} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -13.4$$

,  = 0.1,  = 1

Likelihood

$$\log\left(L\left(\boxed{M_2} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -10.3$$

,  = 0.1,  = 1

Likelihood

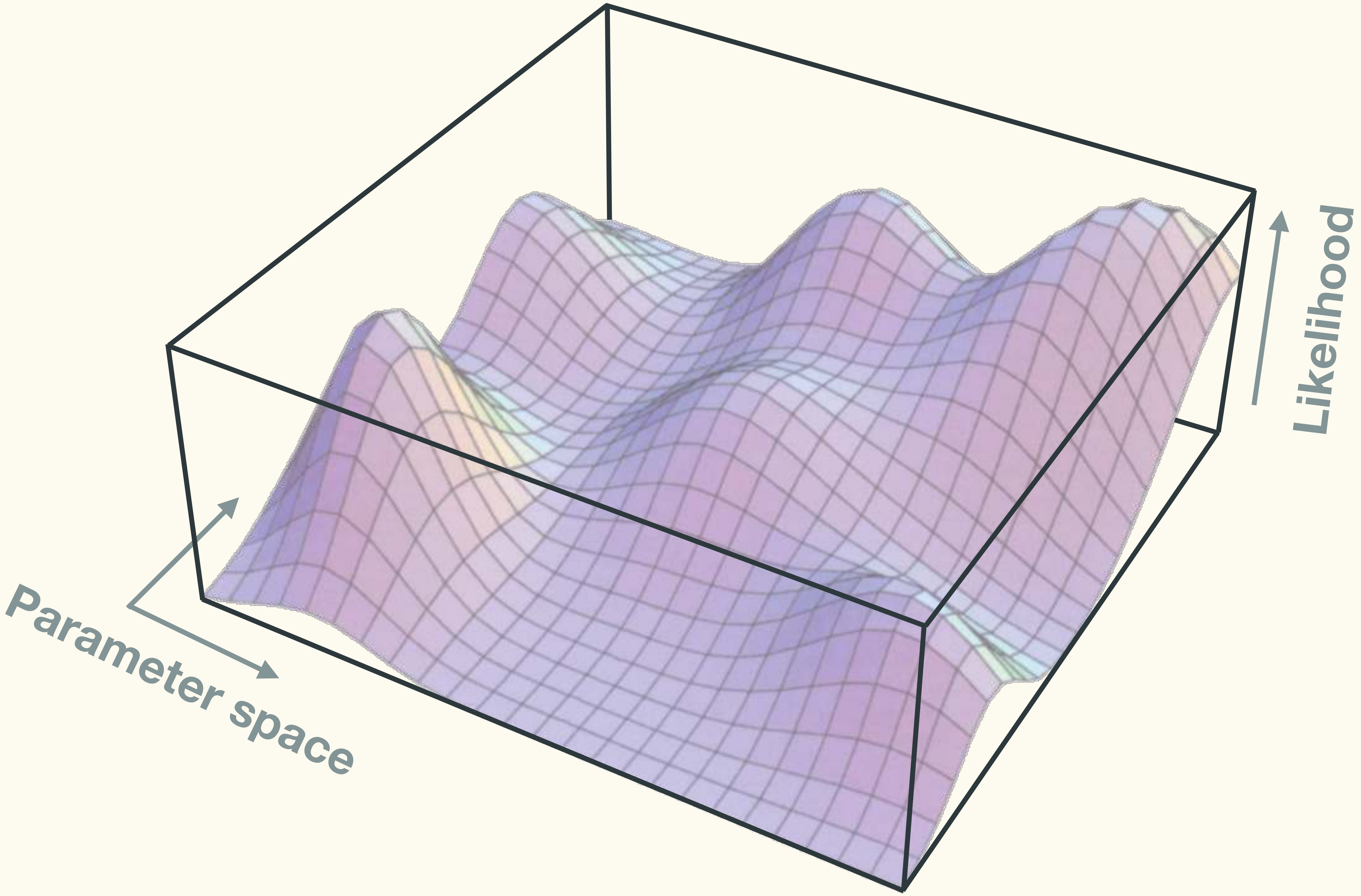
$$\log\left(L\left(\boxed{M_1} \,\middle|\, \begin{matrix}\mathrm{ACTTG}\\ \mathrm{ACTGG}\end{matrix}\right)\right) = -13.4$$

$$\log\left(L\left(\boxed{M_2} \,\middle|\, \begin{matrix}\mathrm{ACTTG}\\ \mathrm{ACTGG}\end{matrix}\right)\right) = -10.3$$

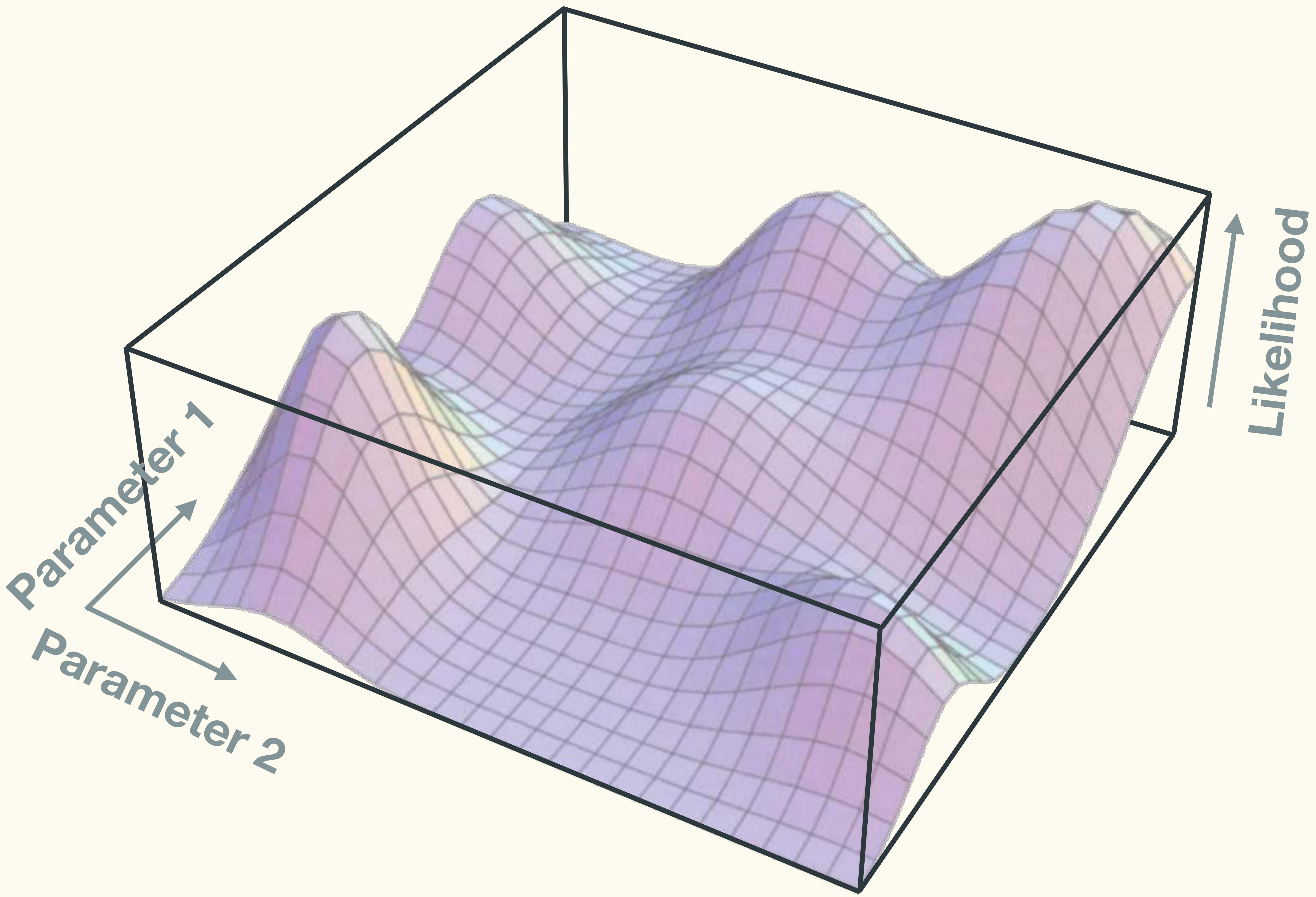
Estimating model parameters using likelihood

Maximum likelihood

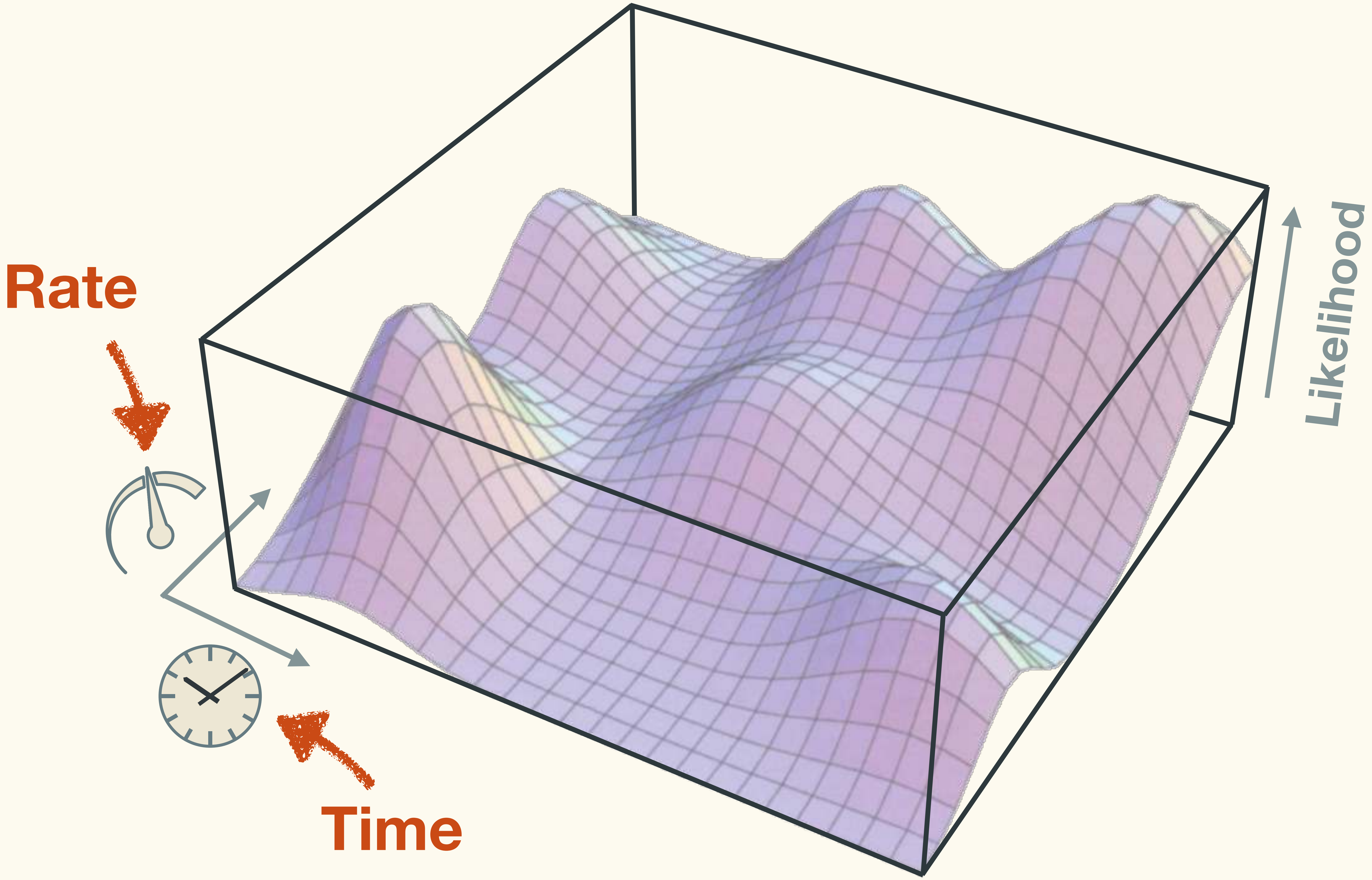
Likelihood surface



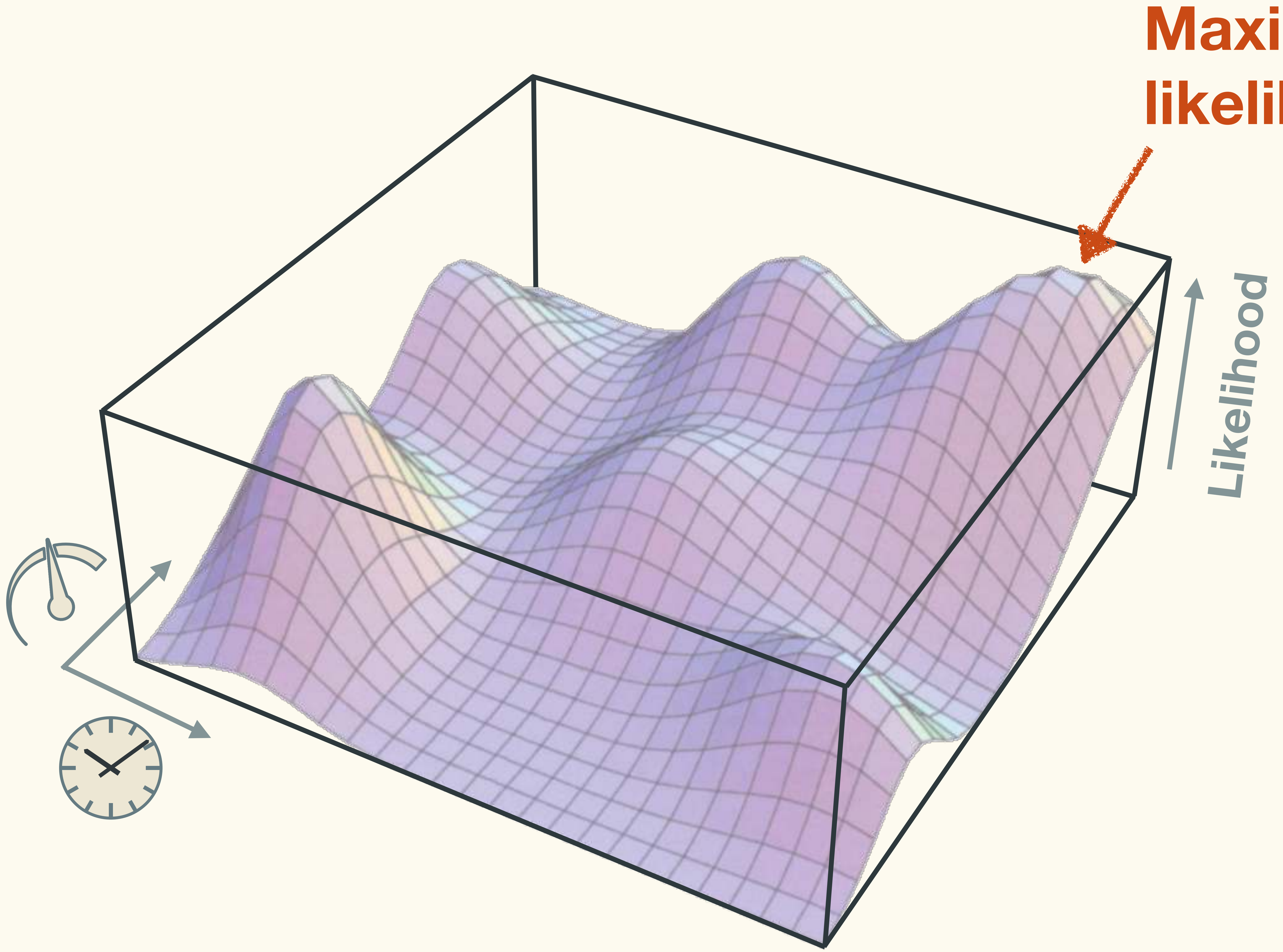
Likelihood surface



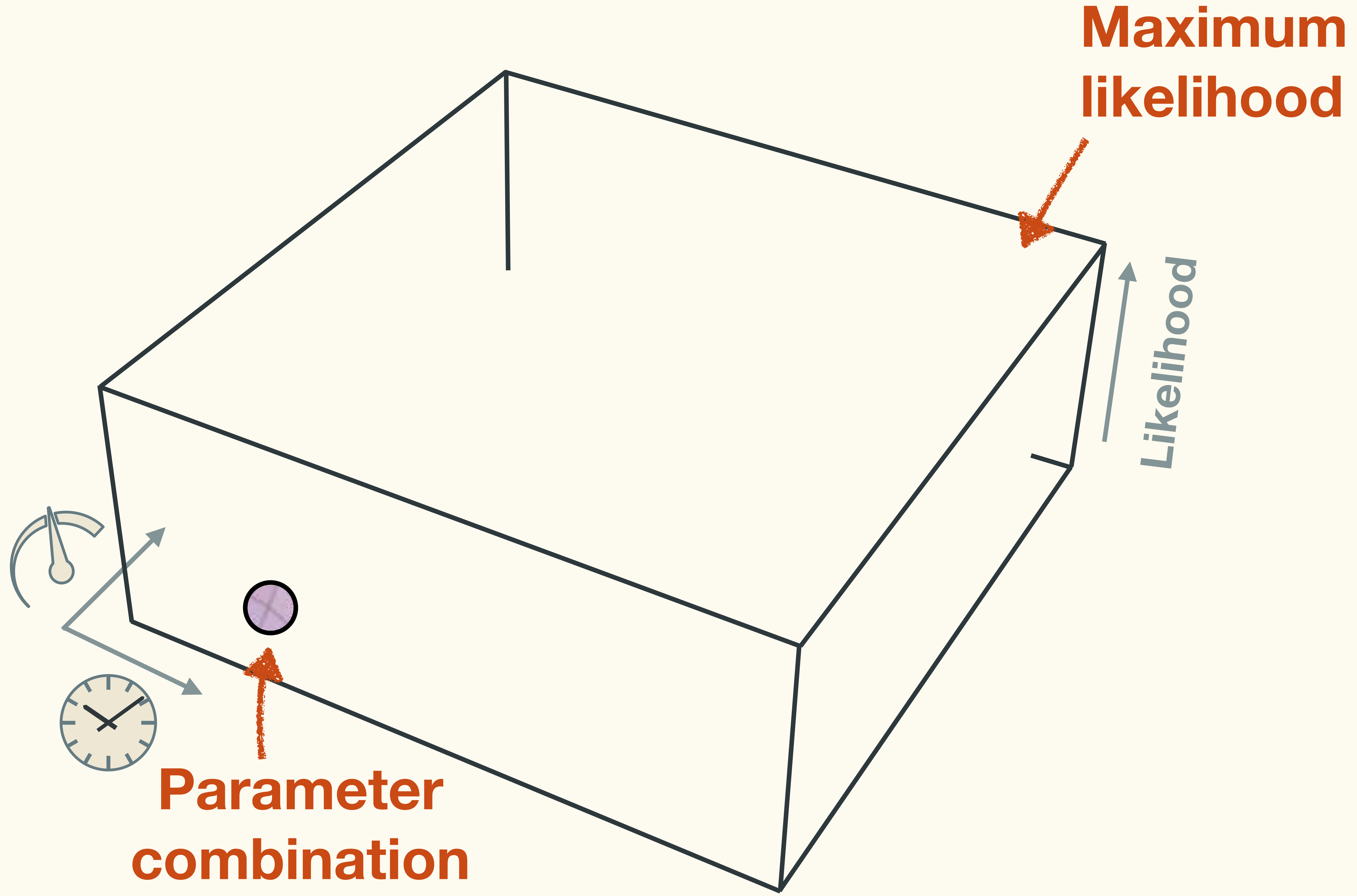
Likelihood surface



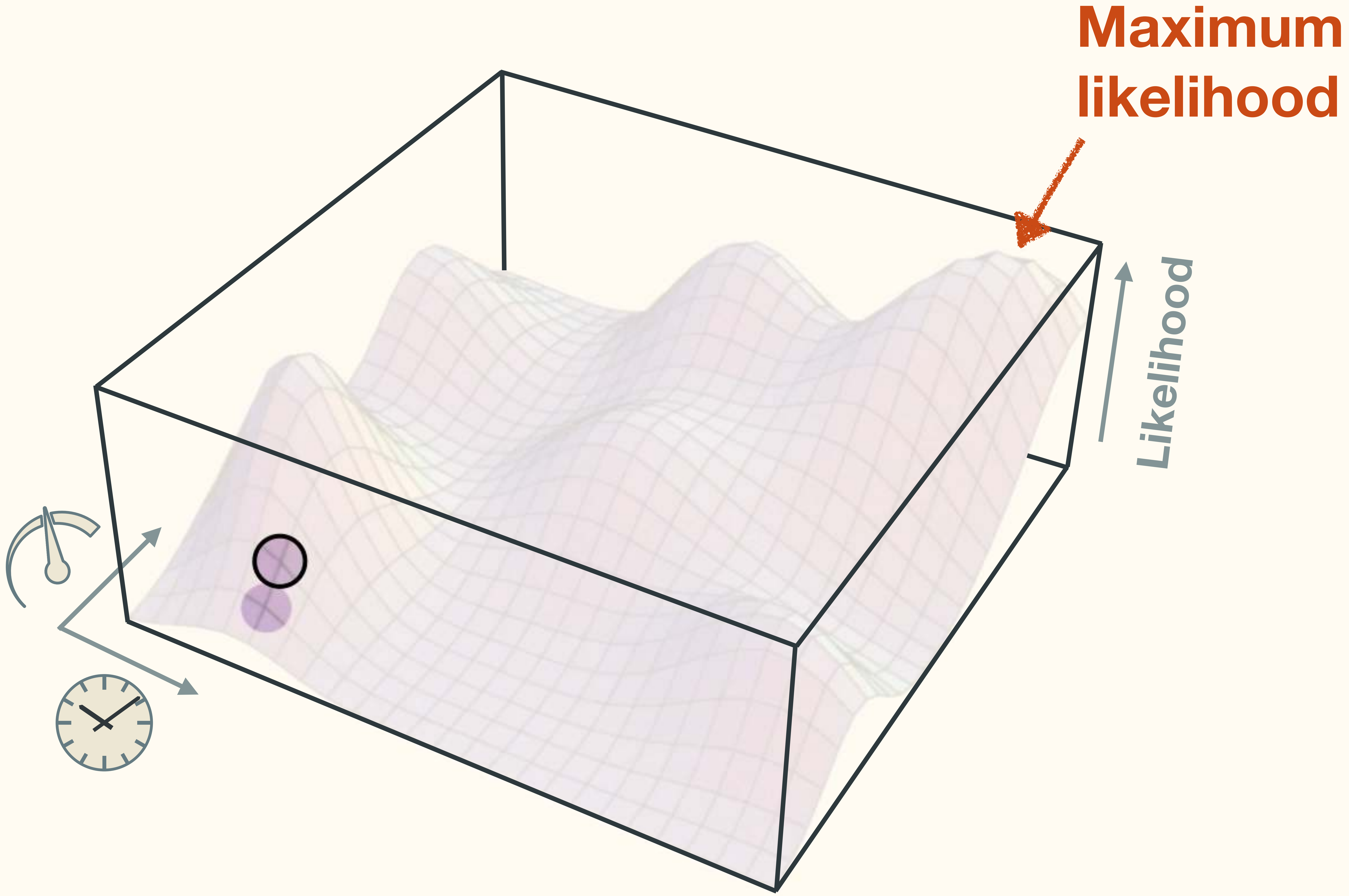
Likelihood surface



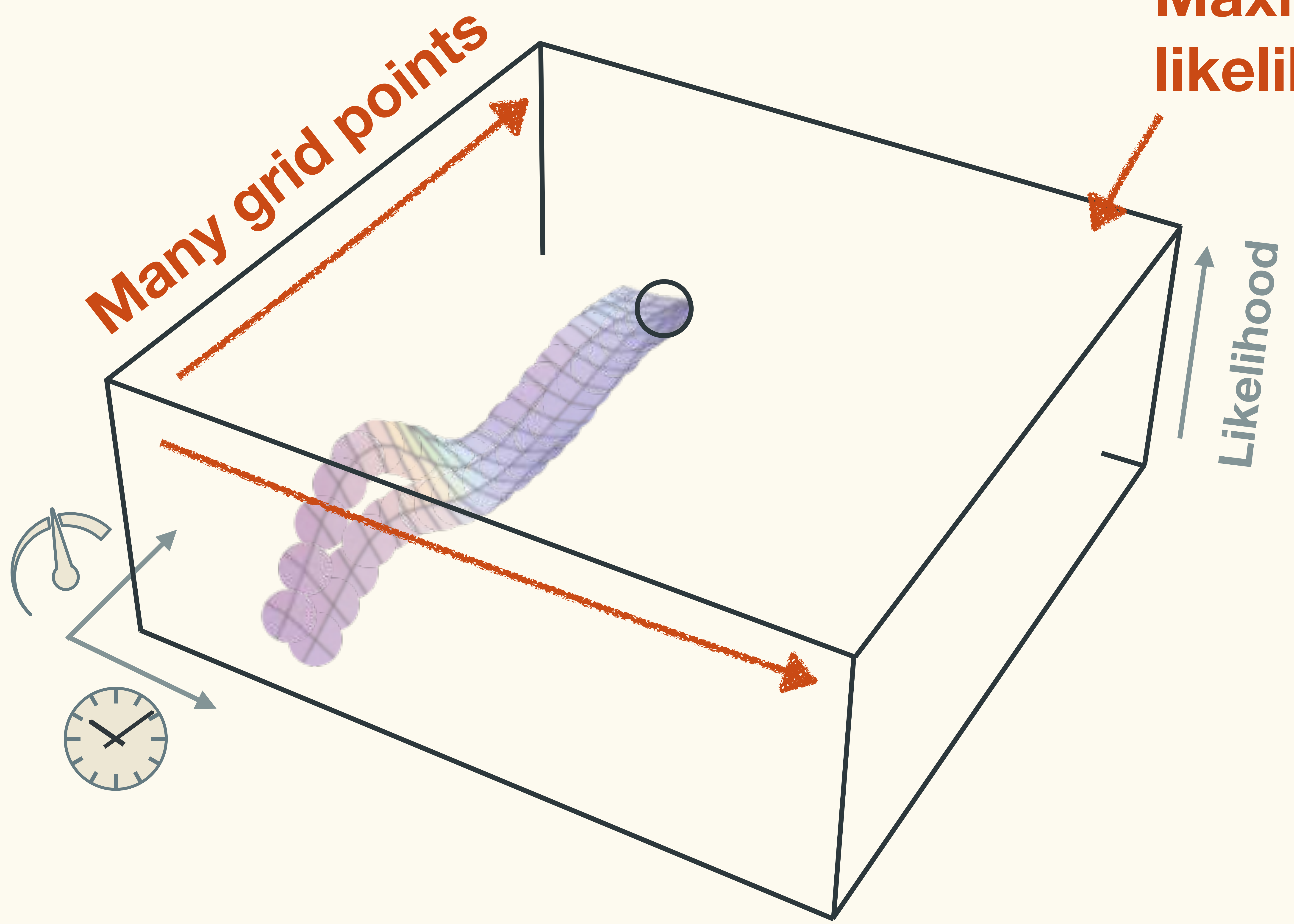
Likelihood surface



Grid search

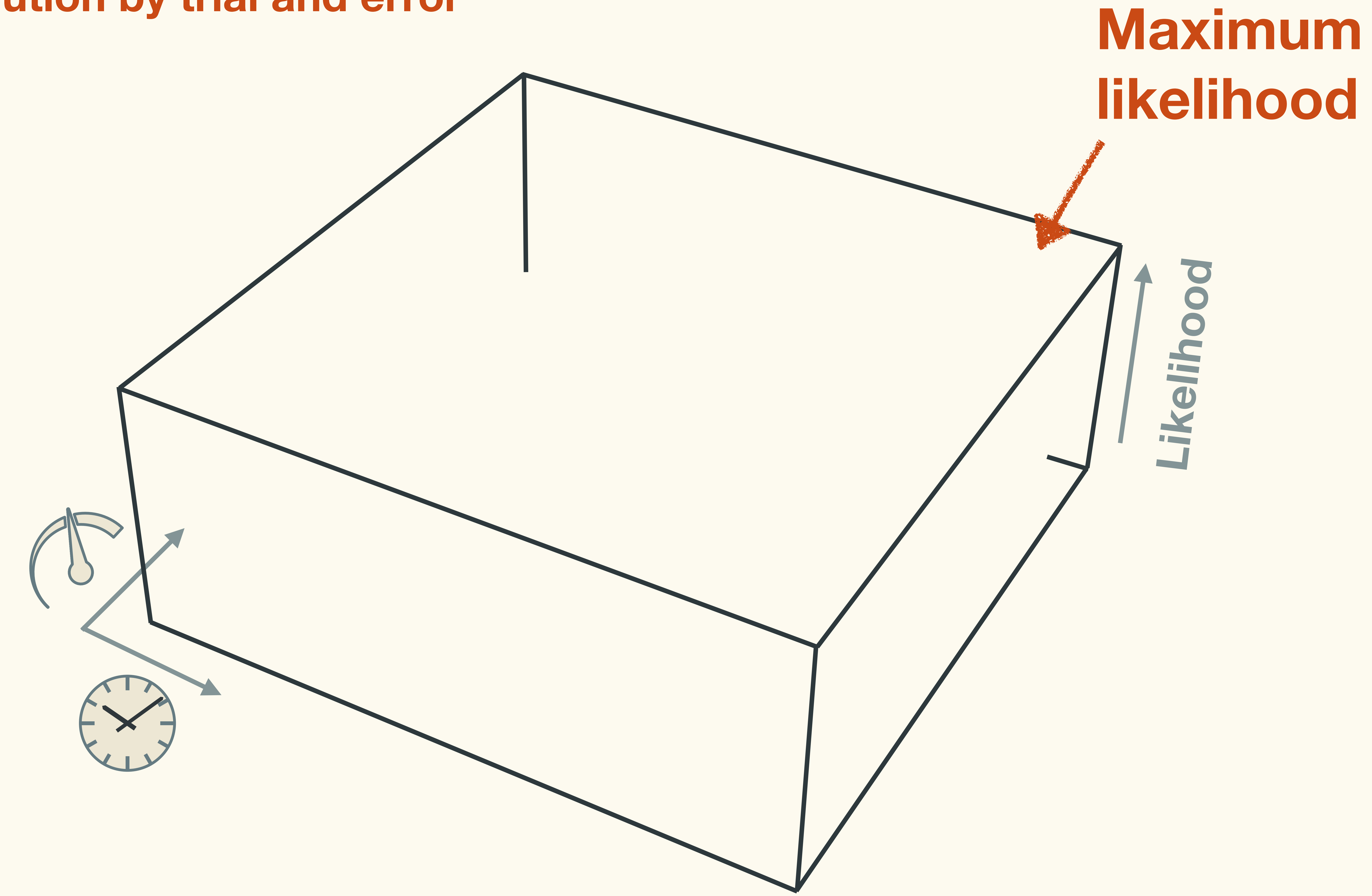


Grid search

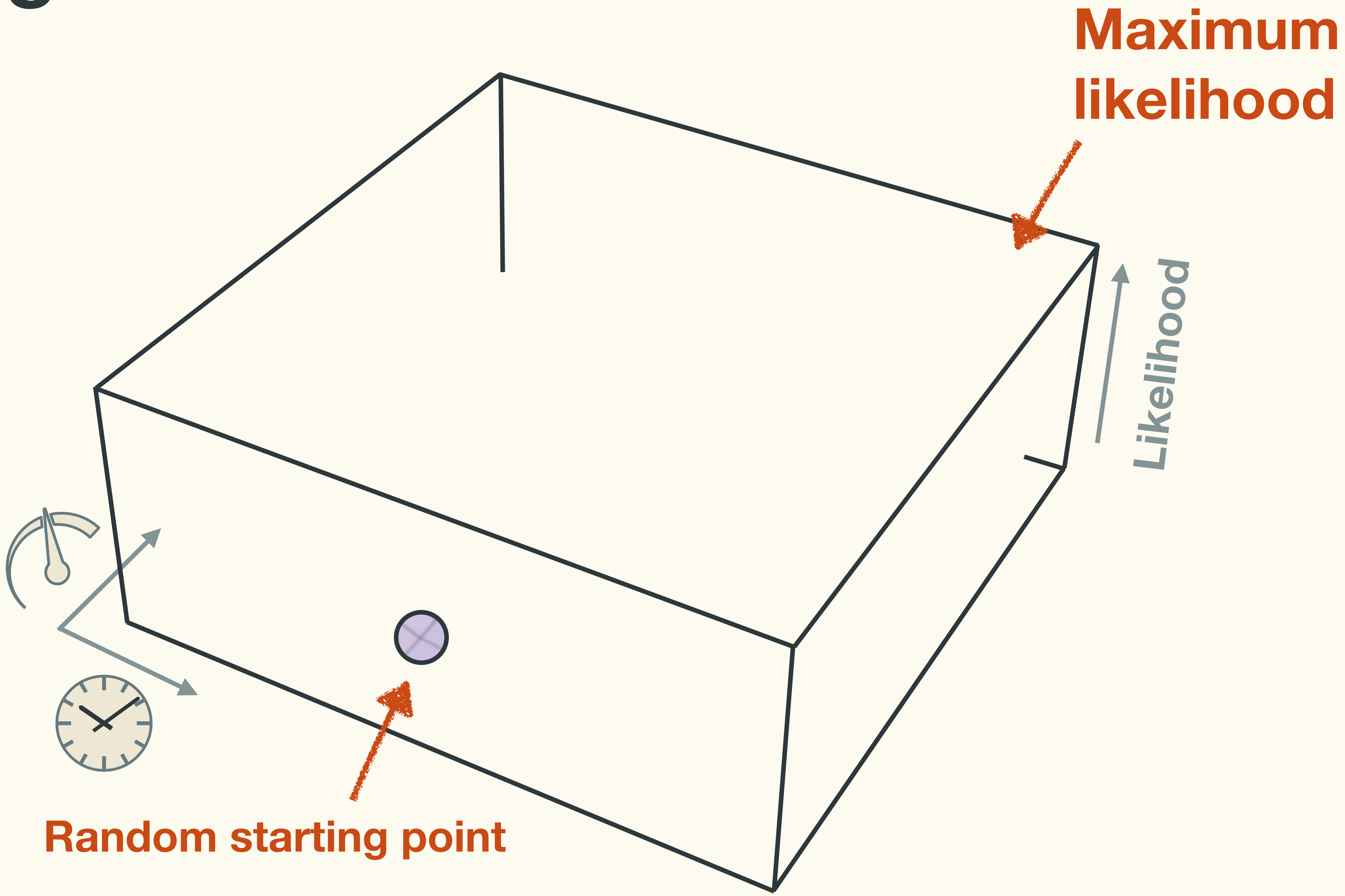


Heuristic search

“proceeding to a solution by trial and error”

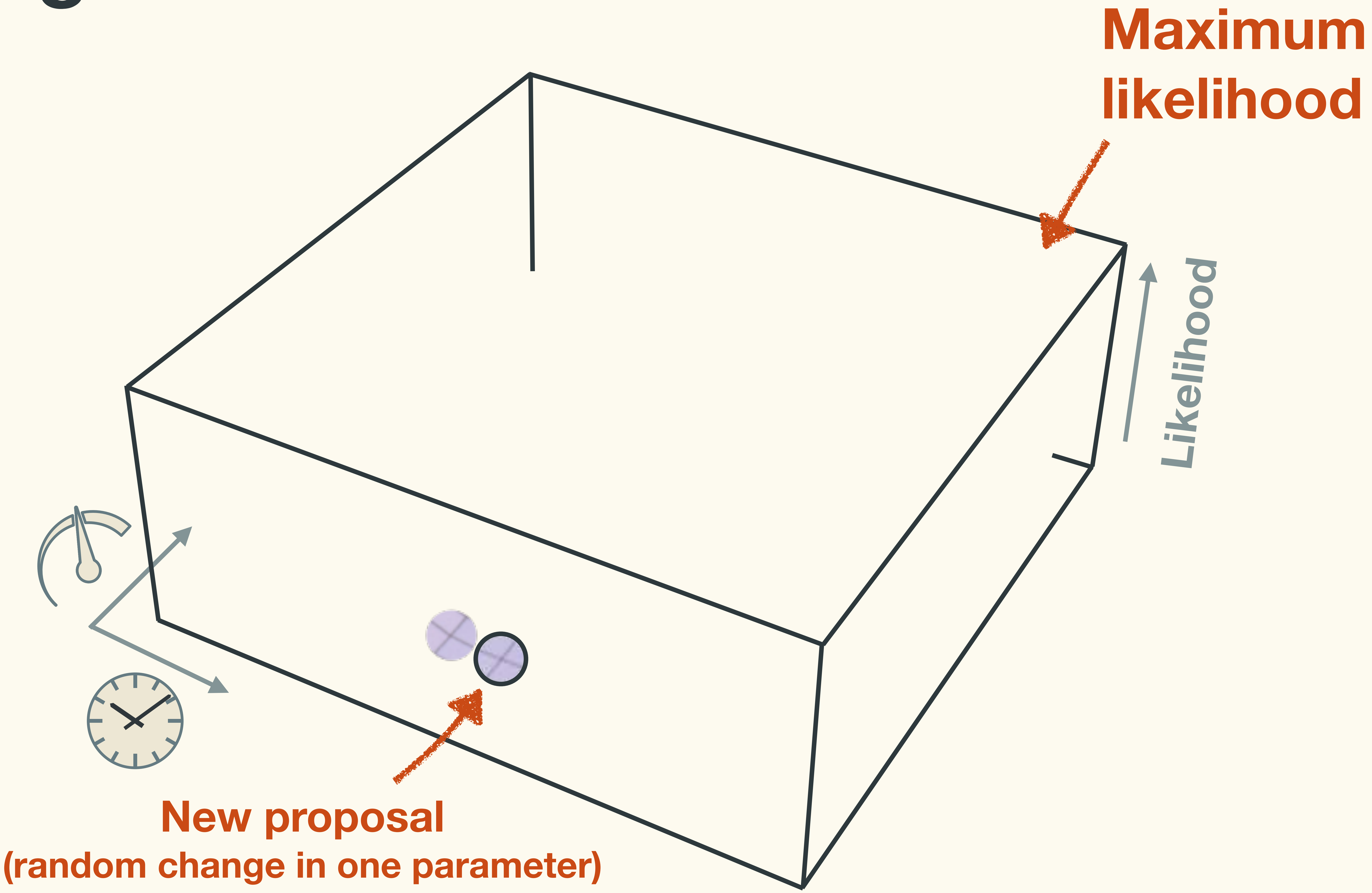


Heuristic search
Hill climbing



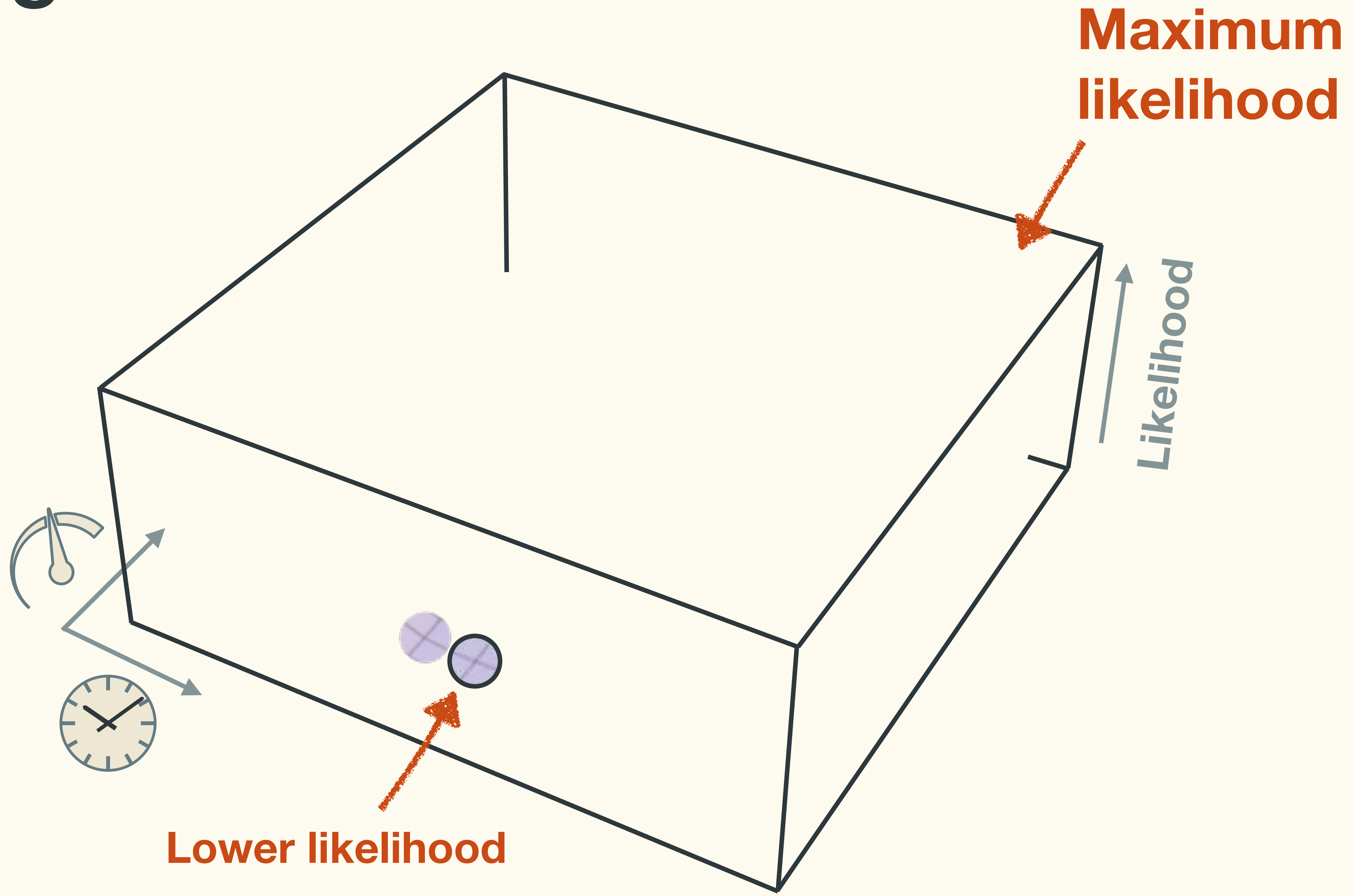
Heuristic search

Hill climbing



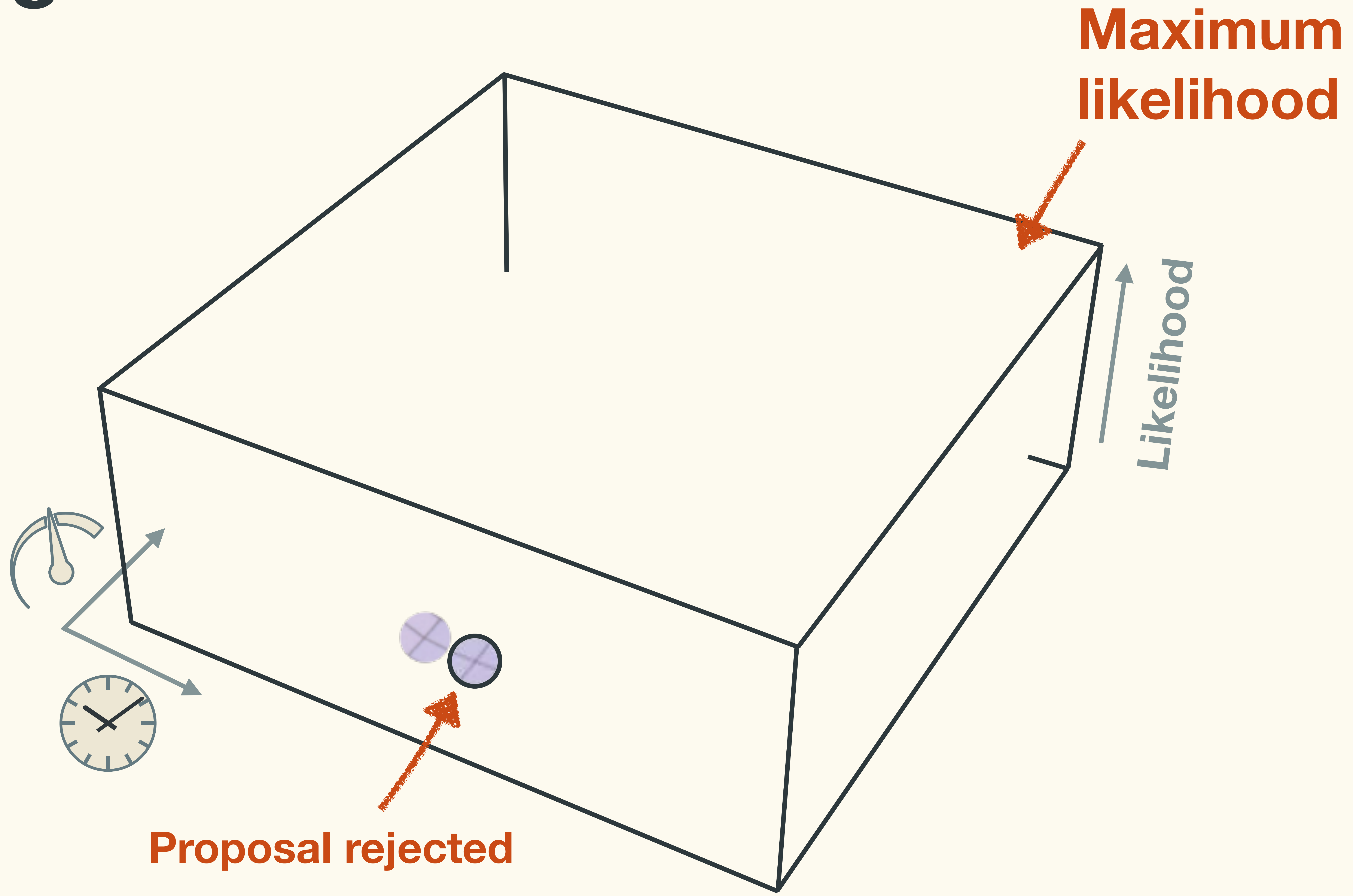
Heuristic search

Hill climbing



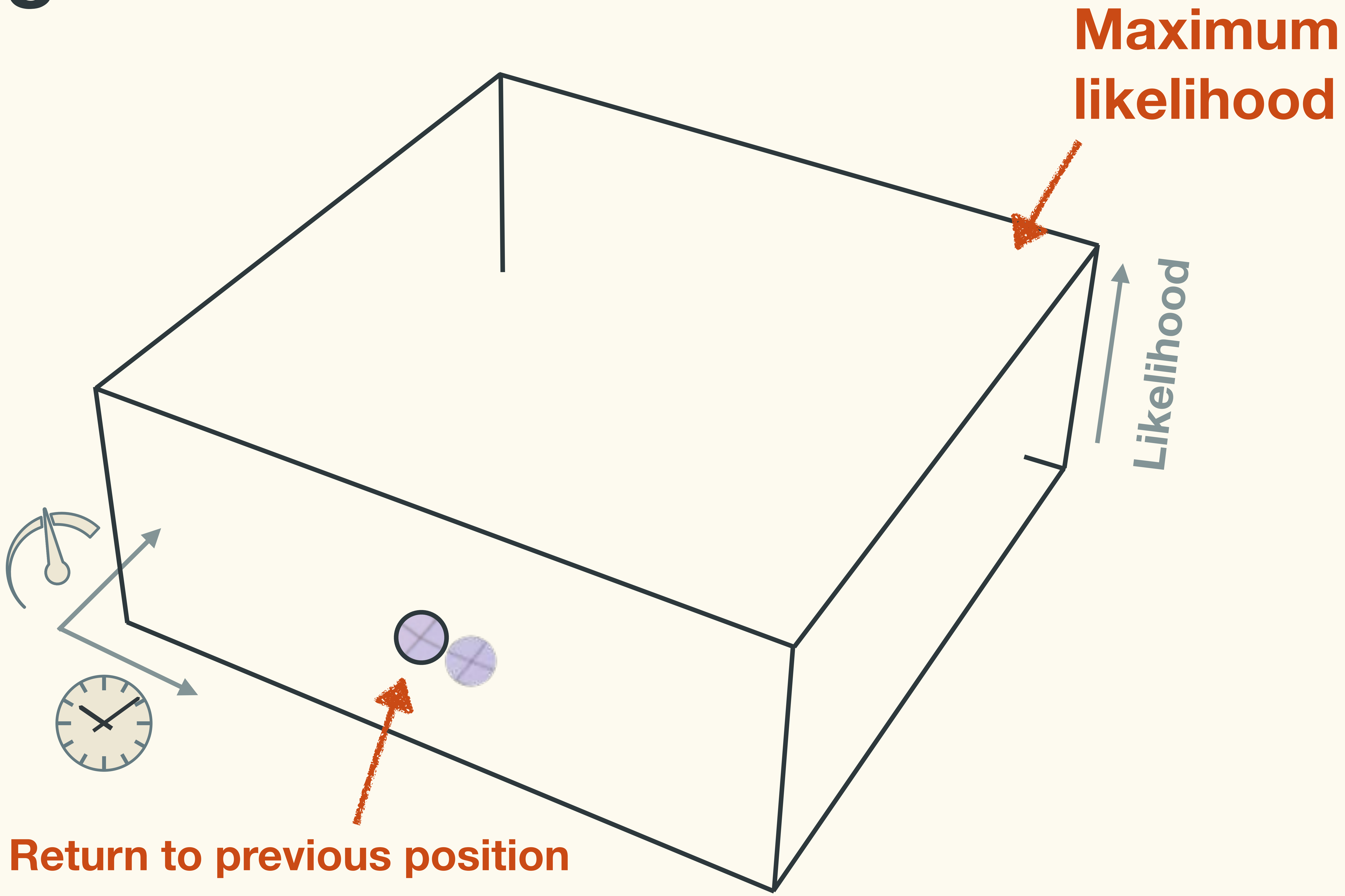
Heuristic search

Hill climbing



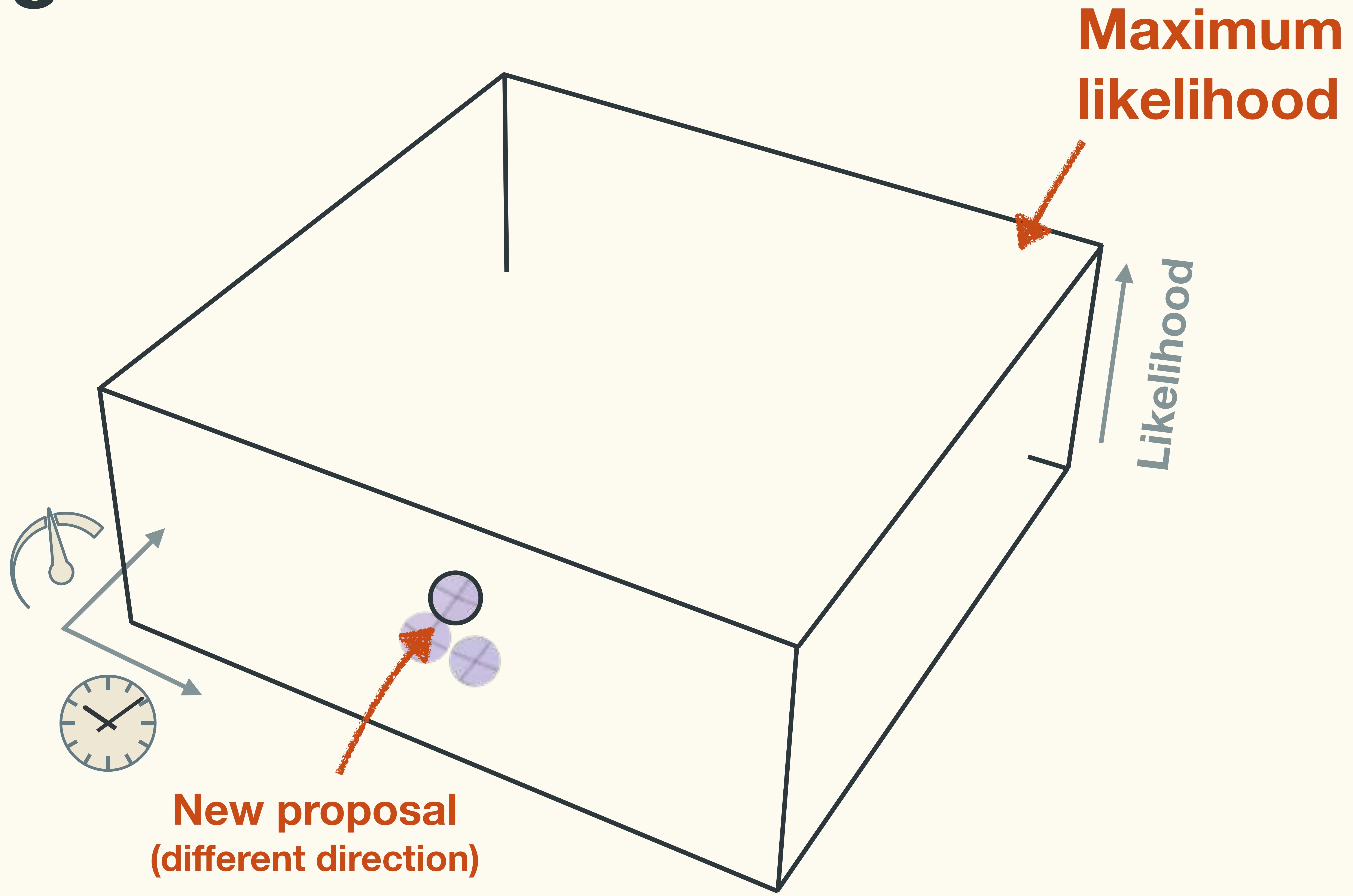
Heuristic search

Hill climbing



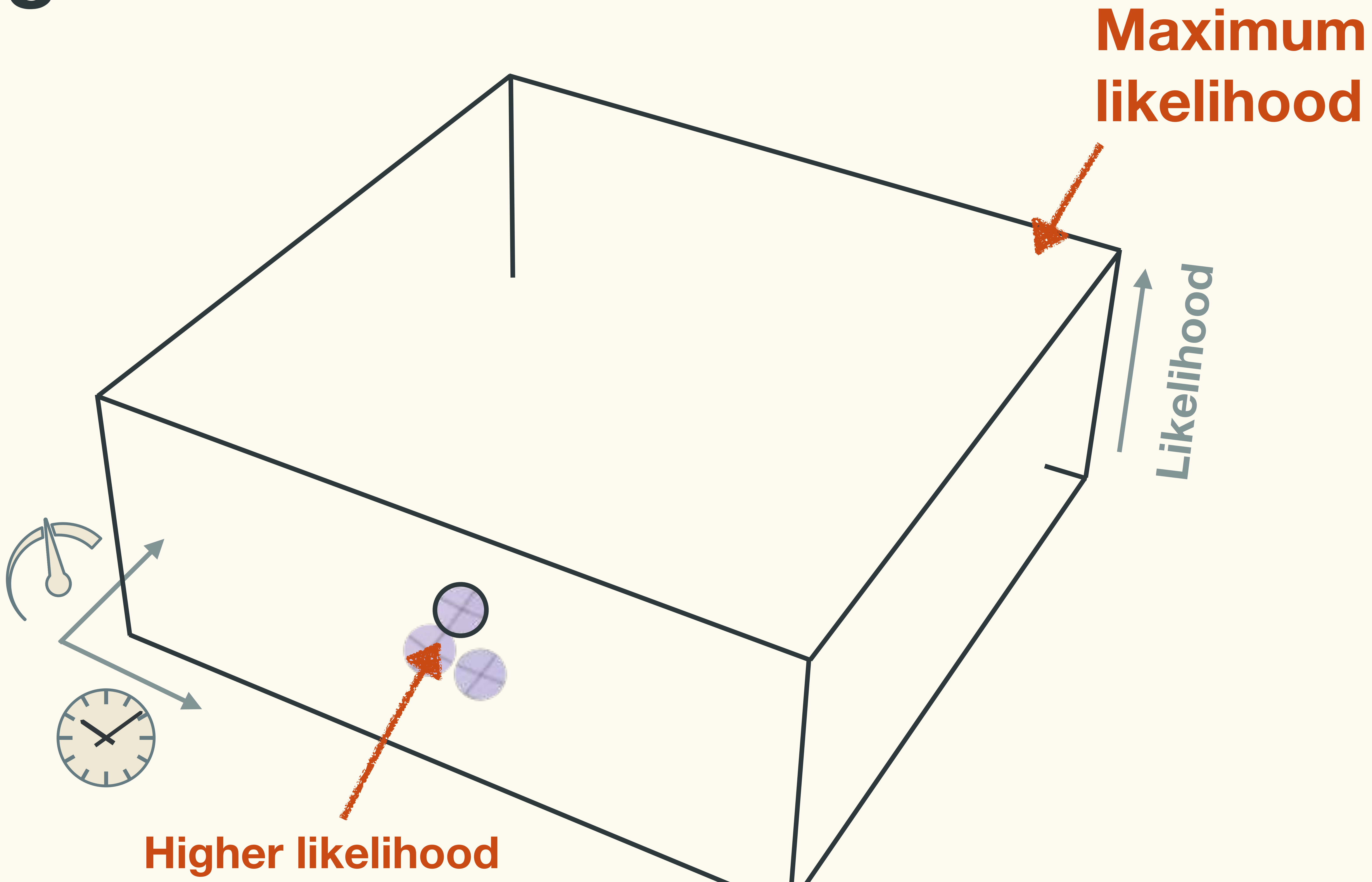
Heuristic search

Hill climbing



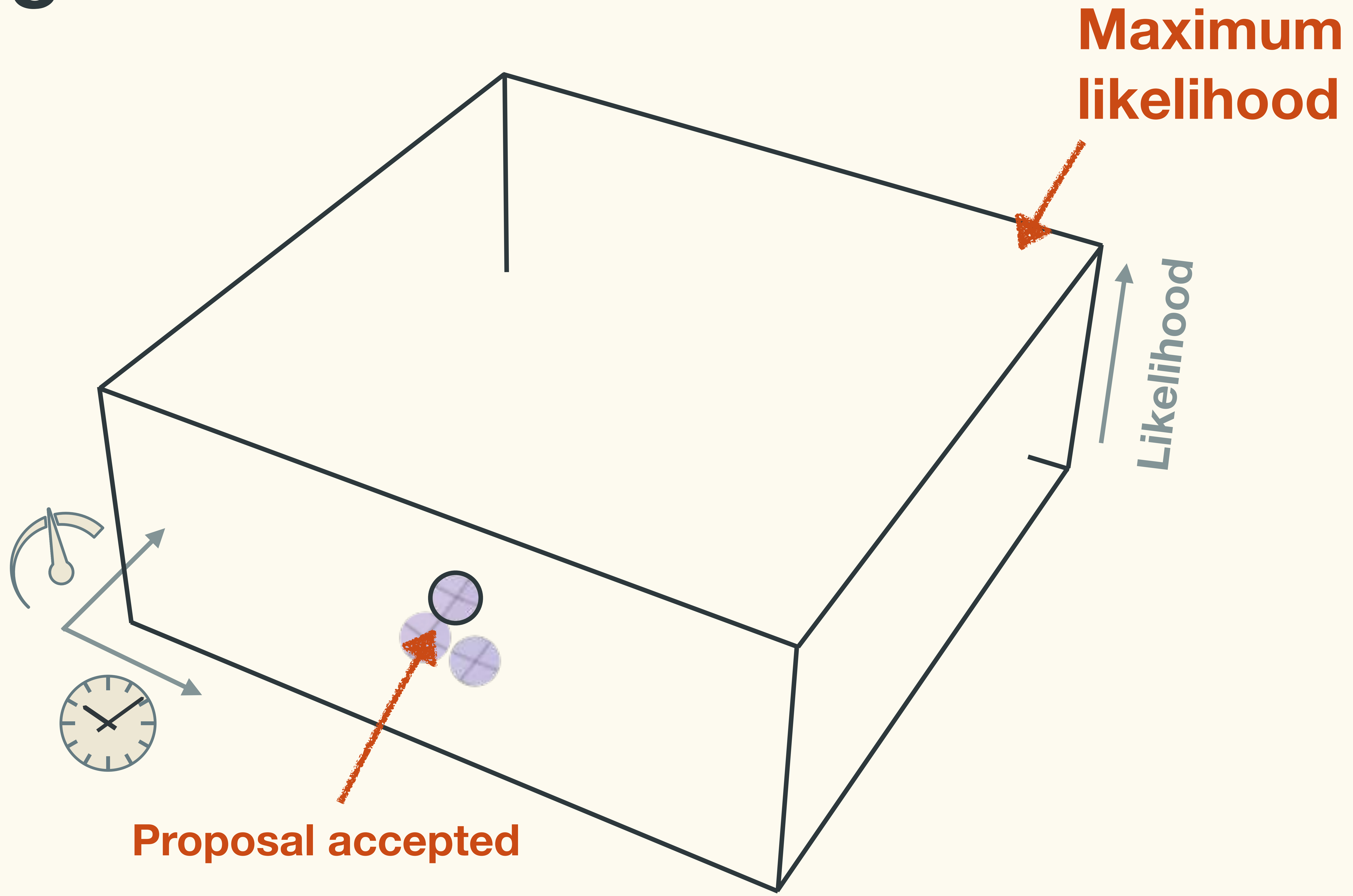
Heuristic search

Hill climbing



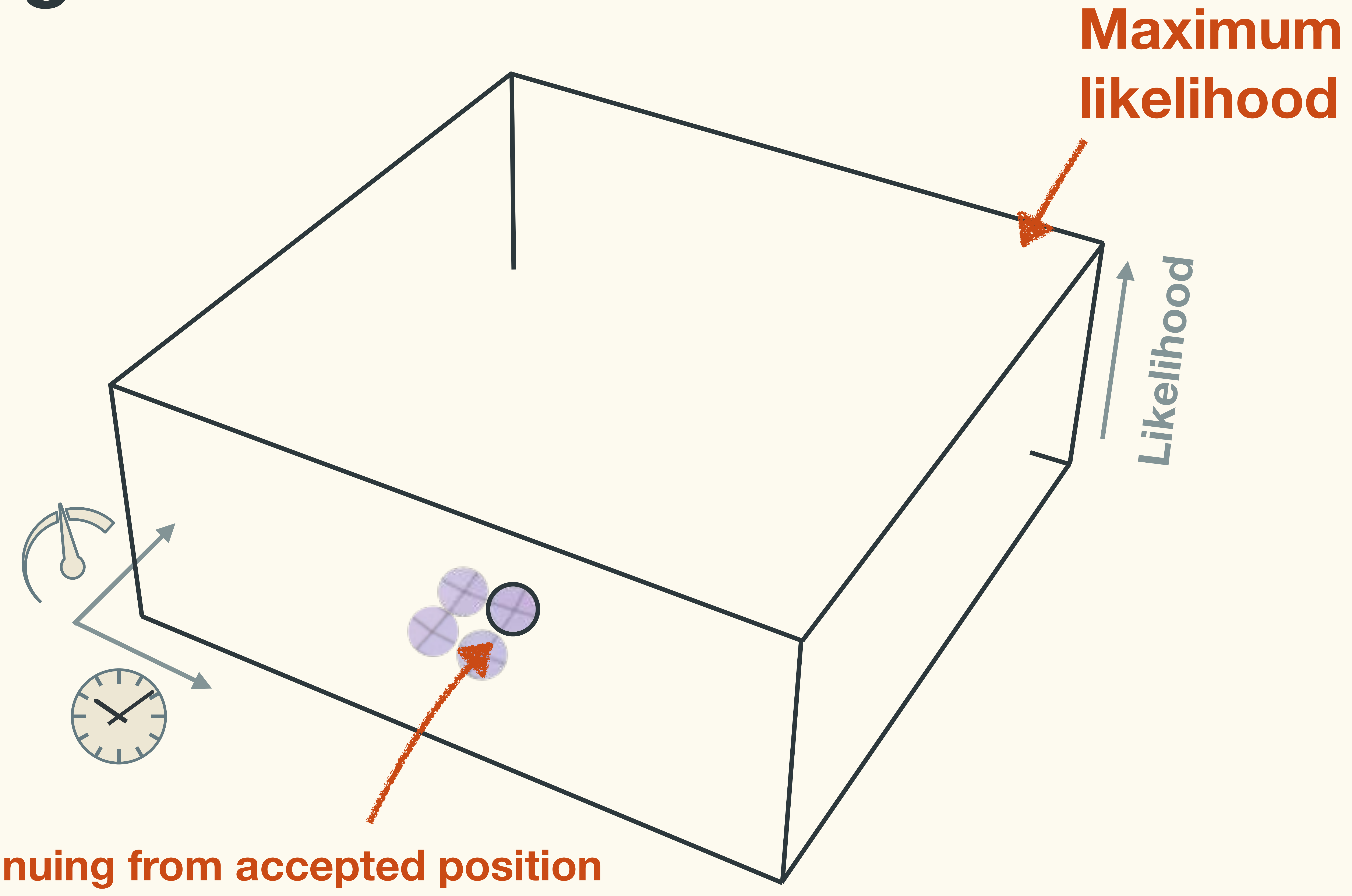
Heuristic search

Hill climbing



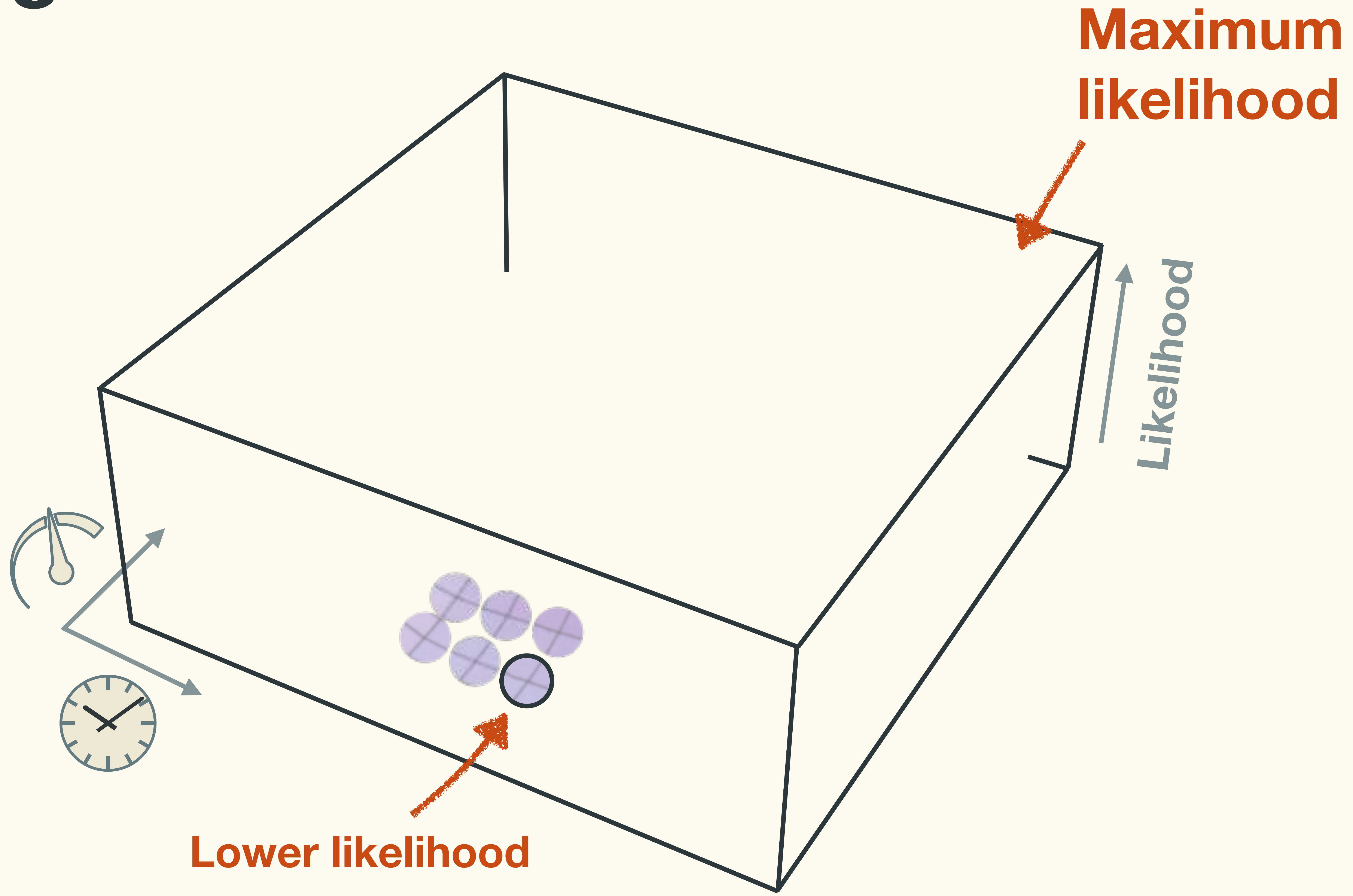
Heuristic search

Hill climbing



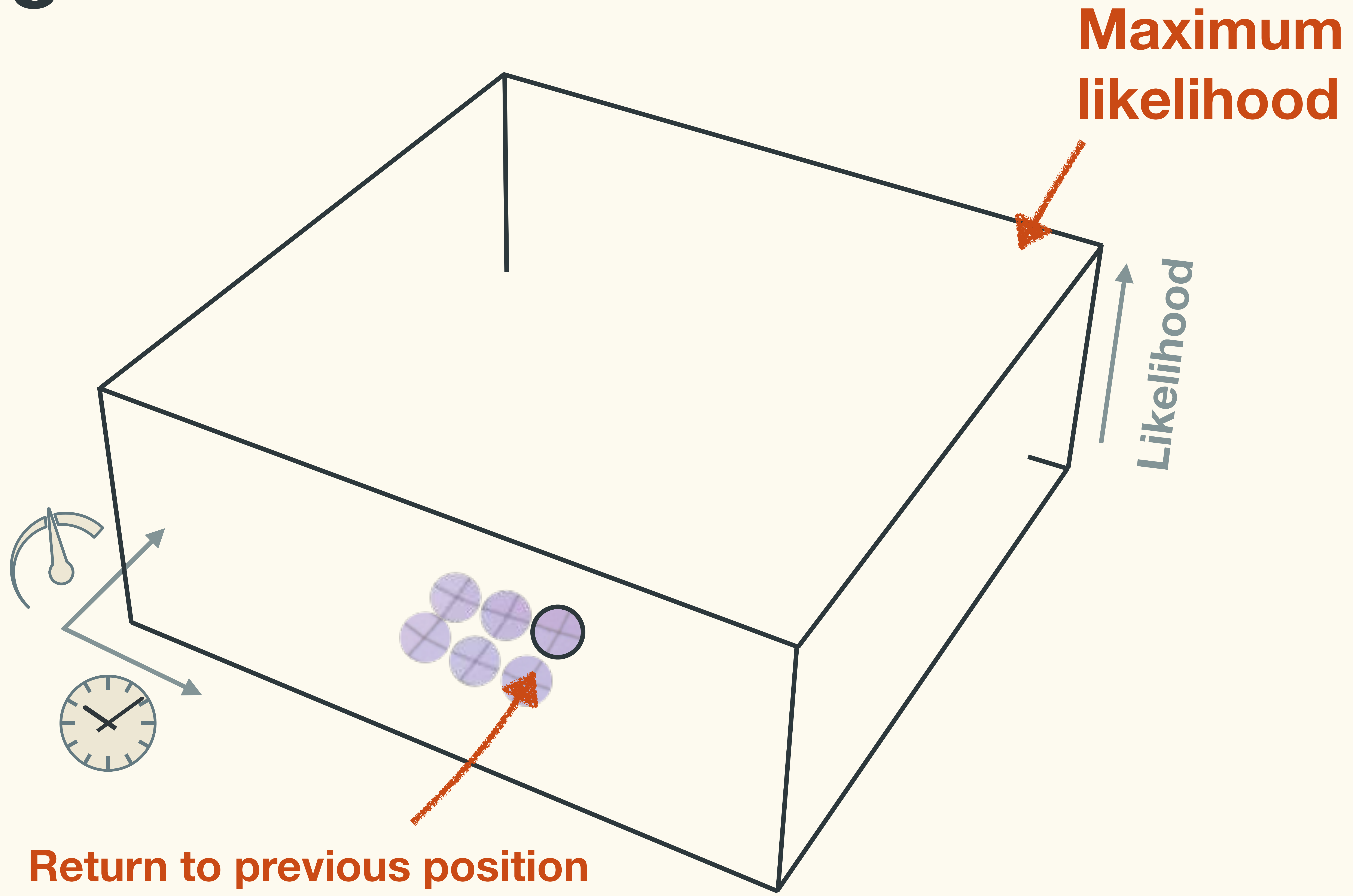
Heuristic search

Hill climbing



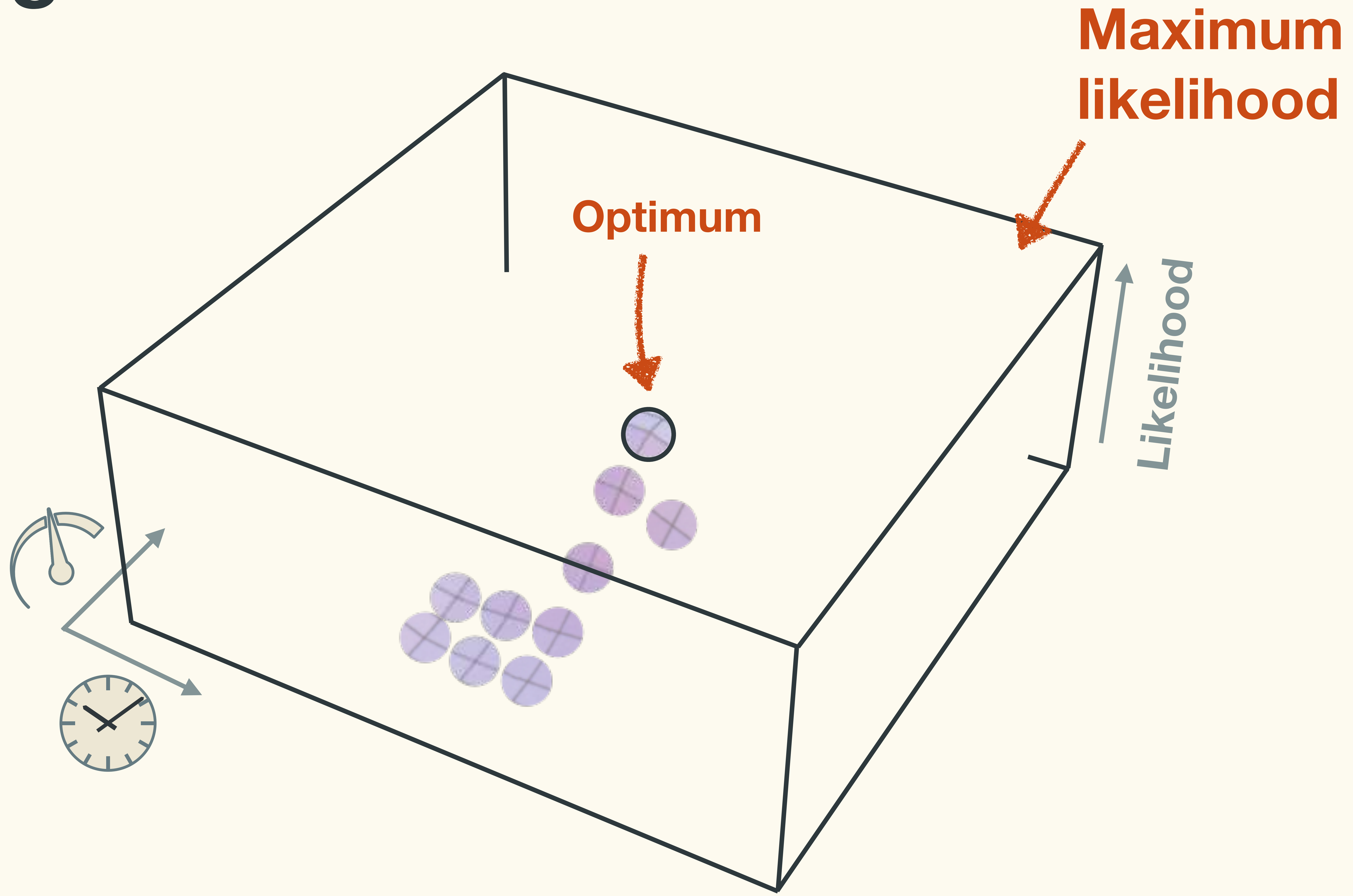
Heuristic search

Hill climbing



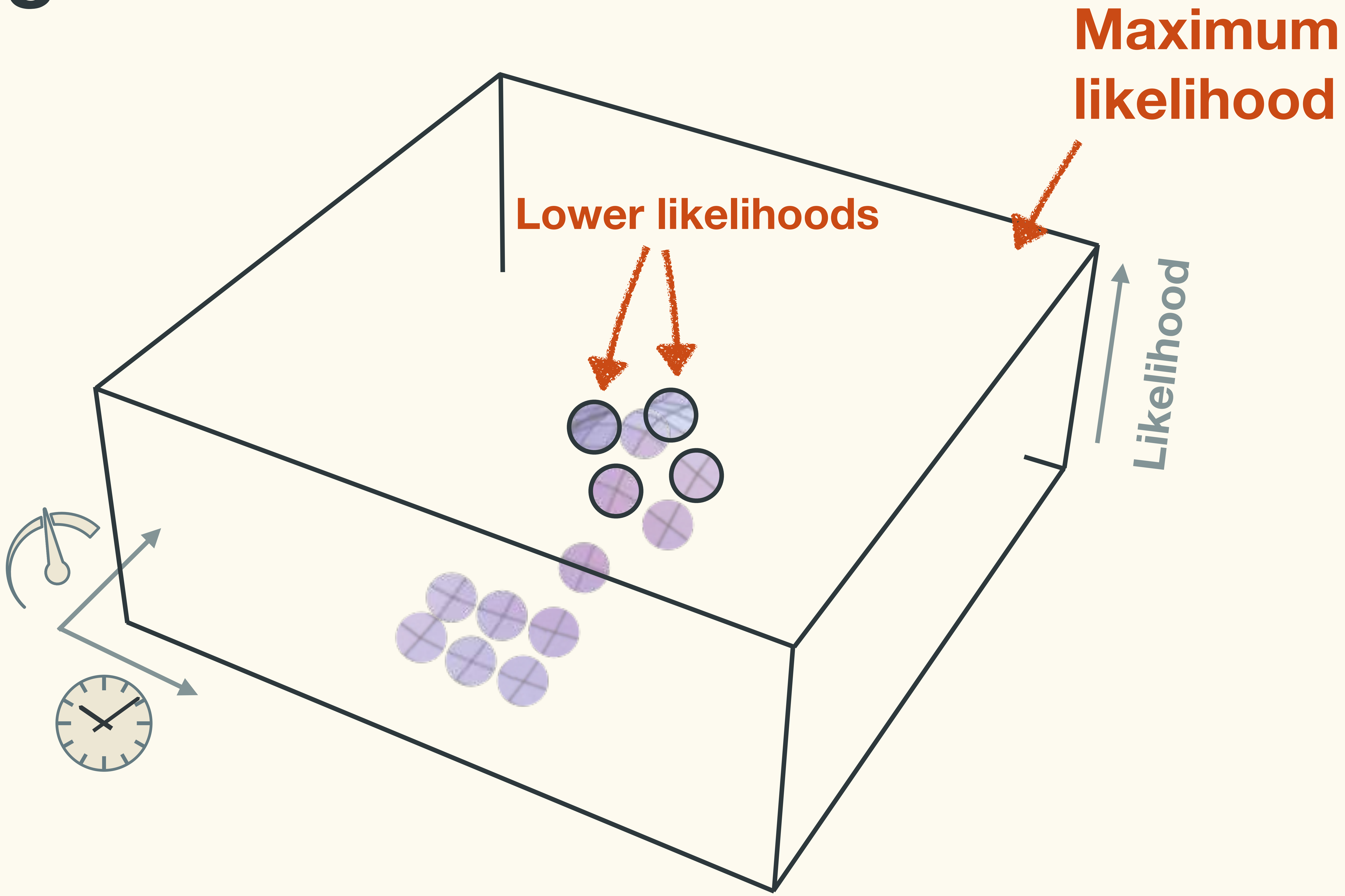
Heuristic search

Hill climbing



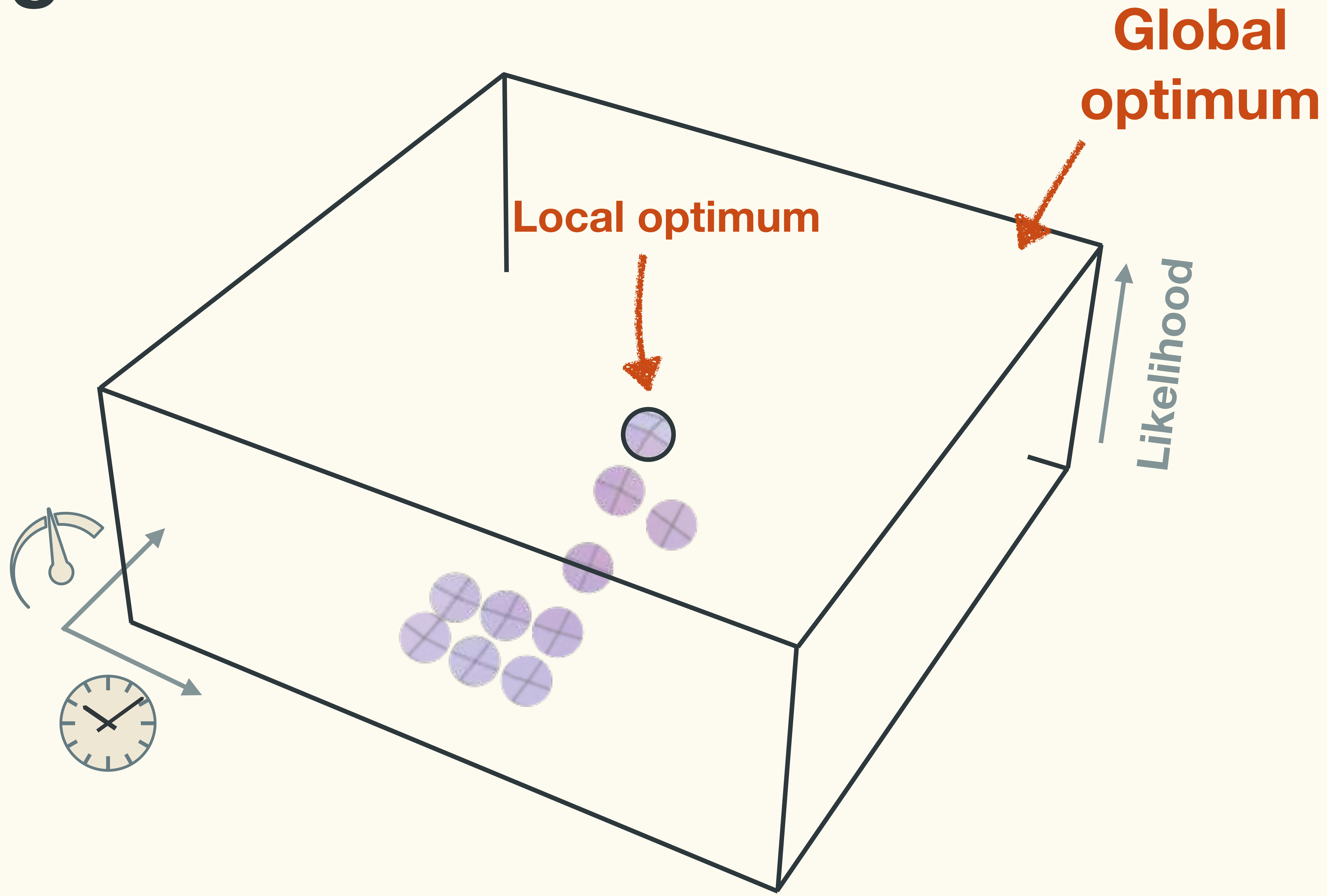
Heuristic search

Hill climbing



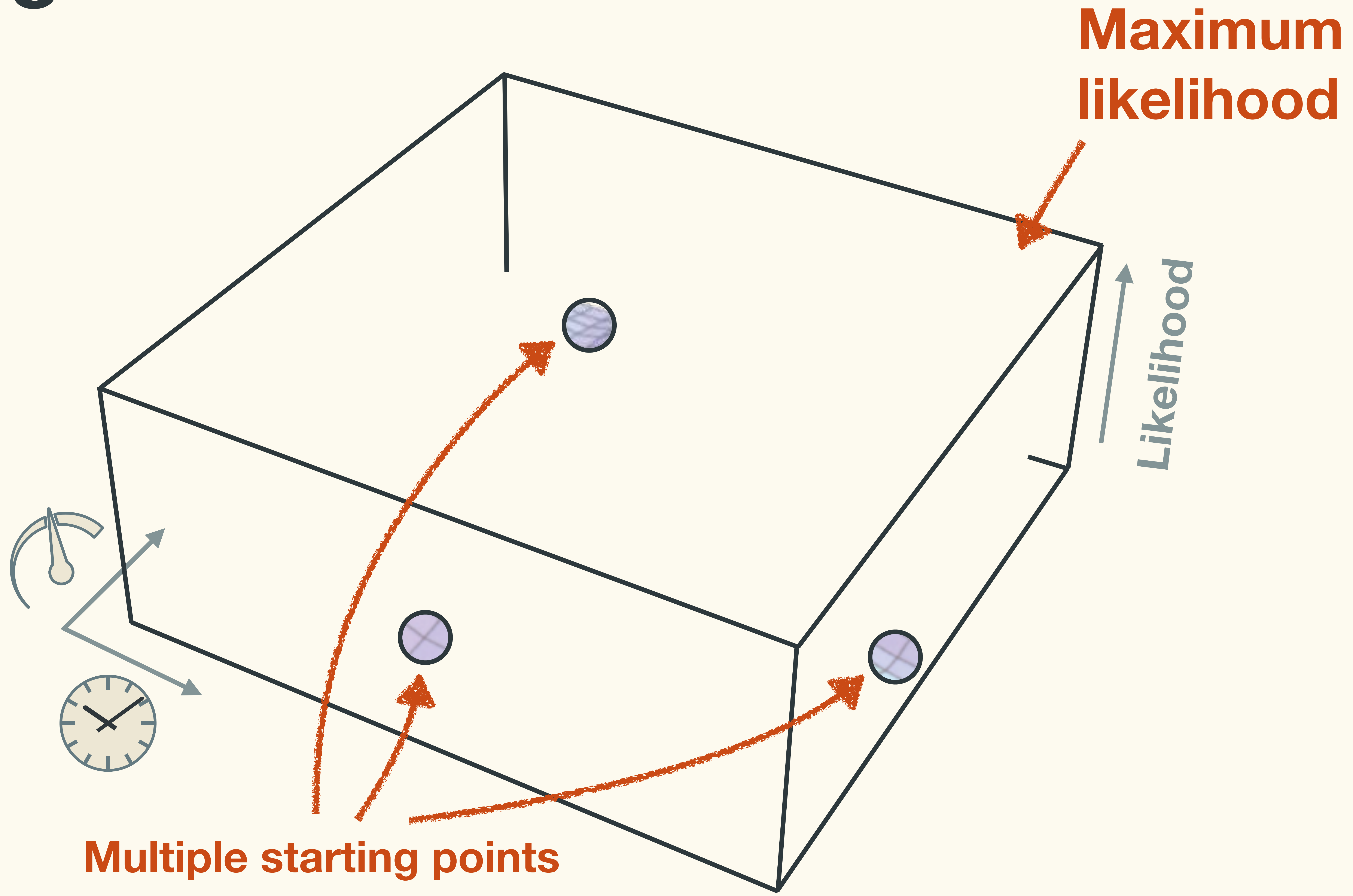
Heuristic search

Hill climbing

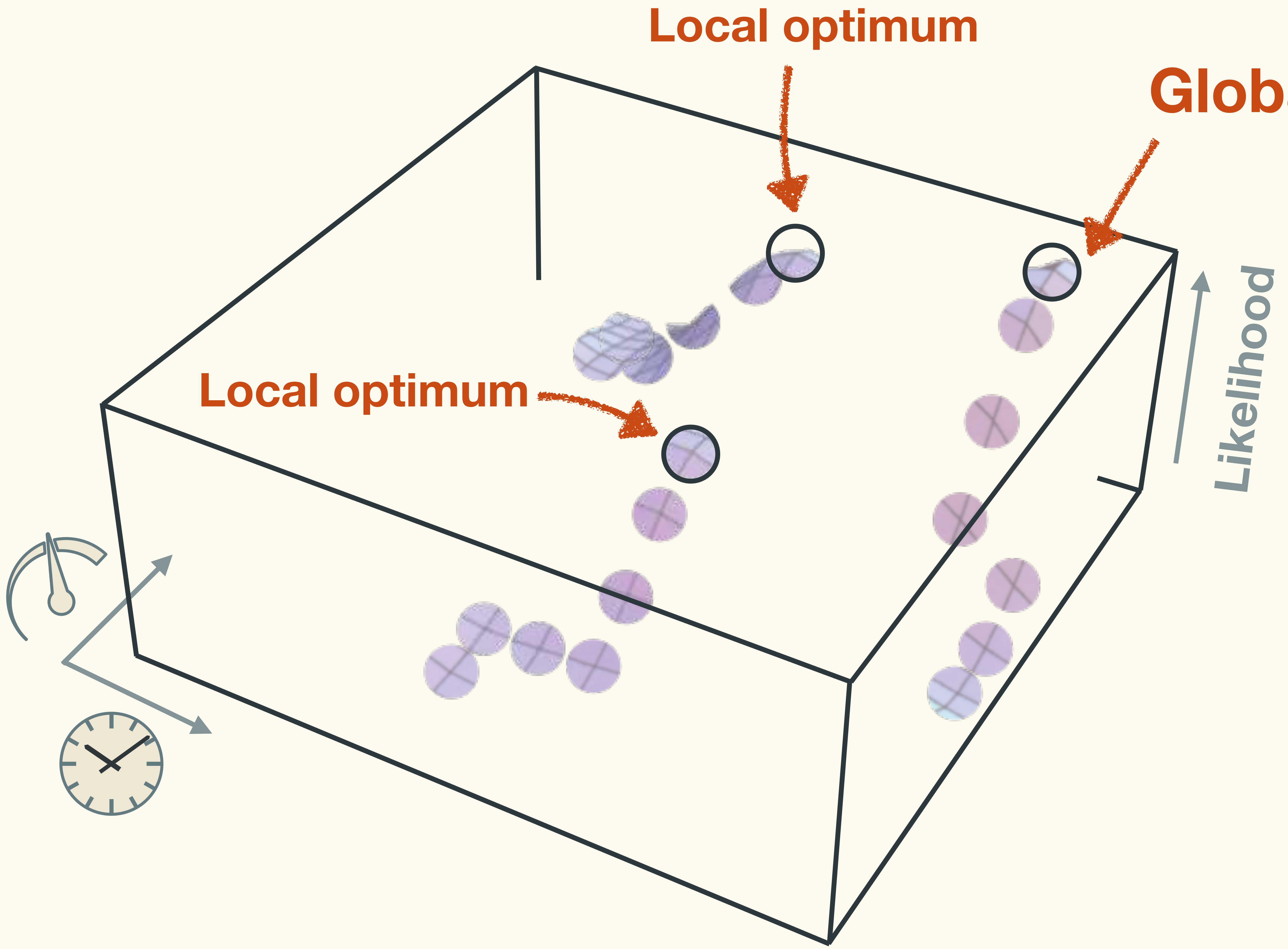


Heuristic search

Hill climbing

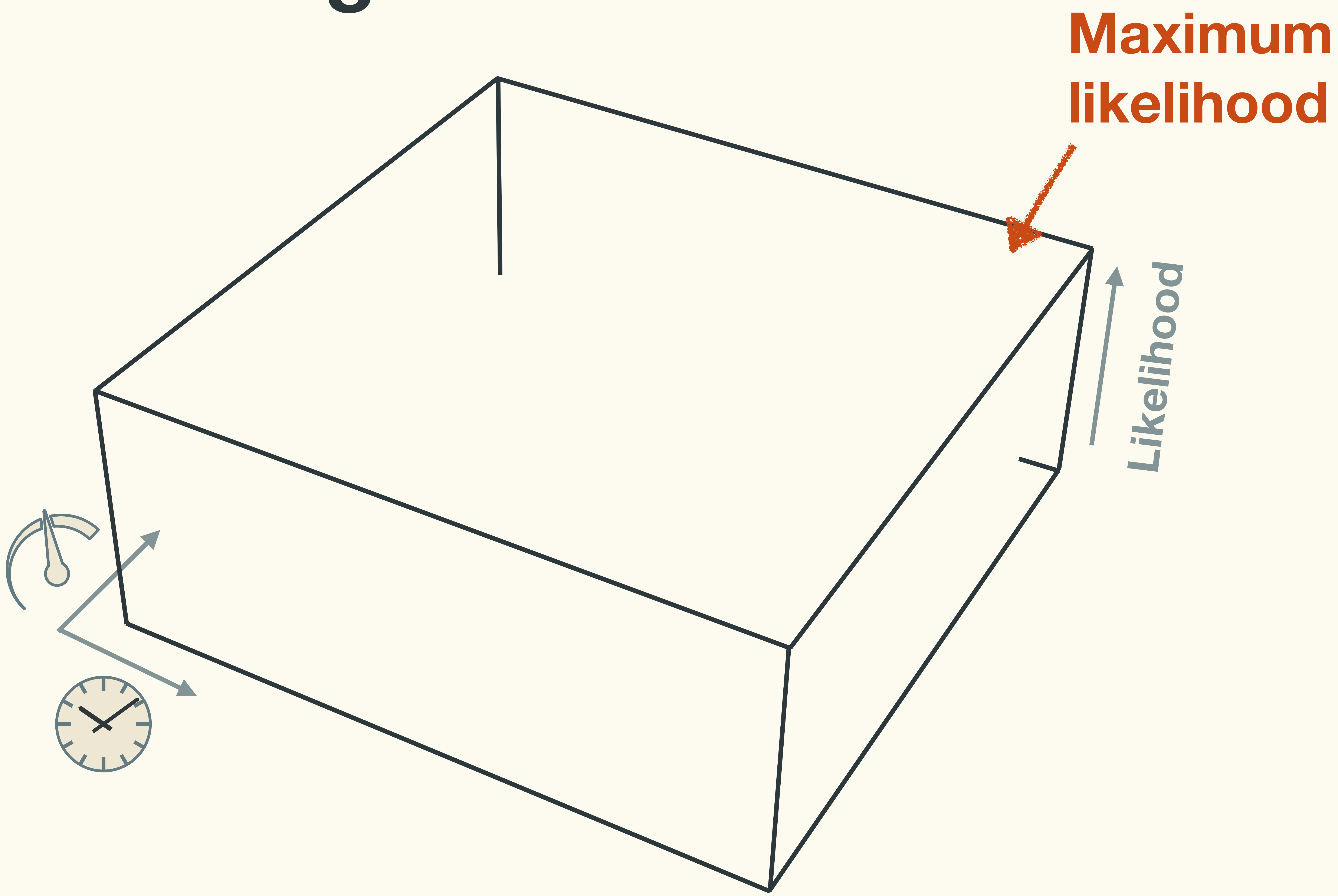


Heuristic search
Hill climbing



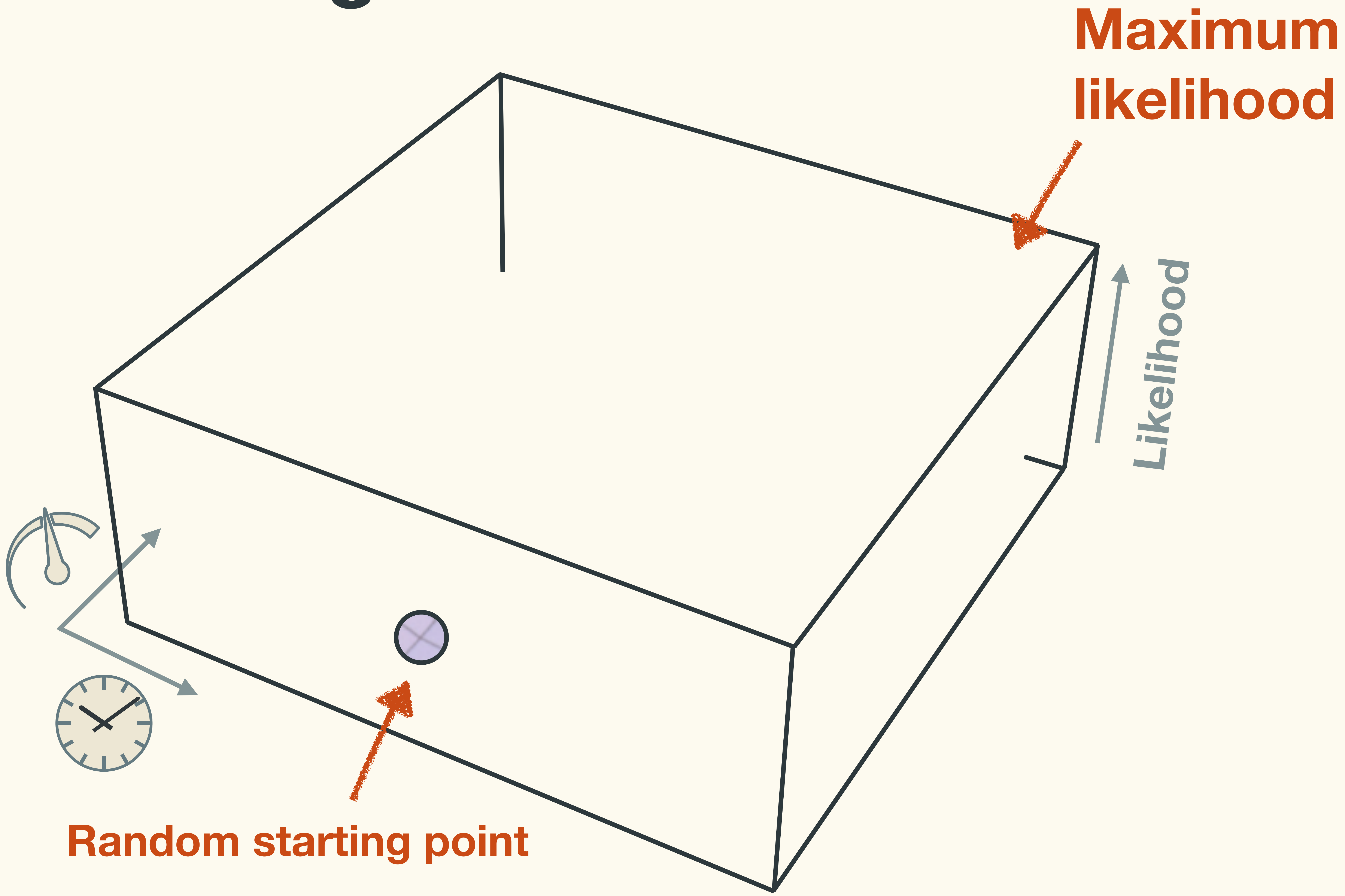
Heuristic search

Simulated annealing



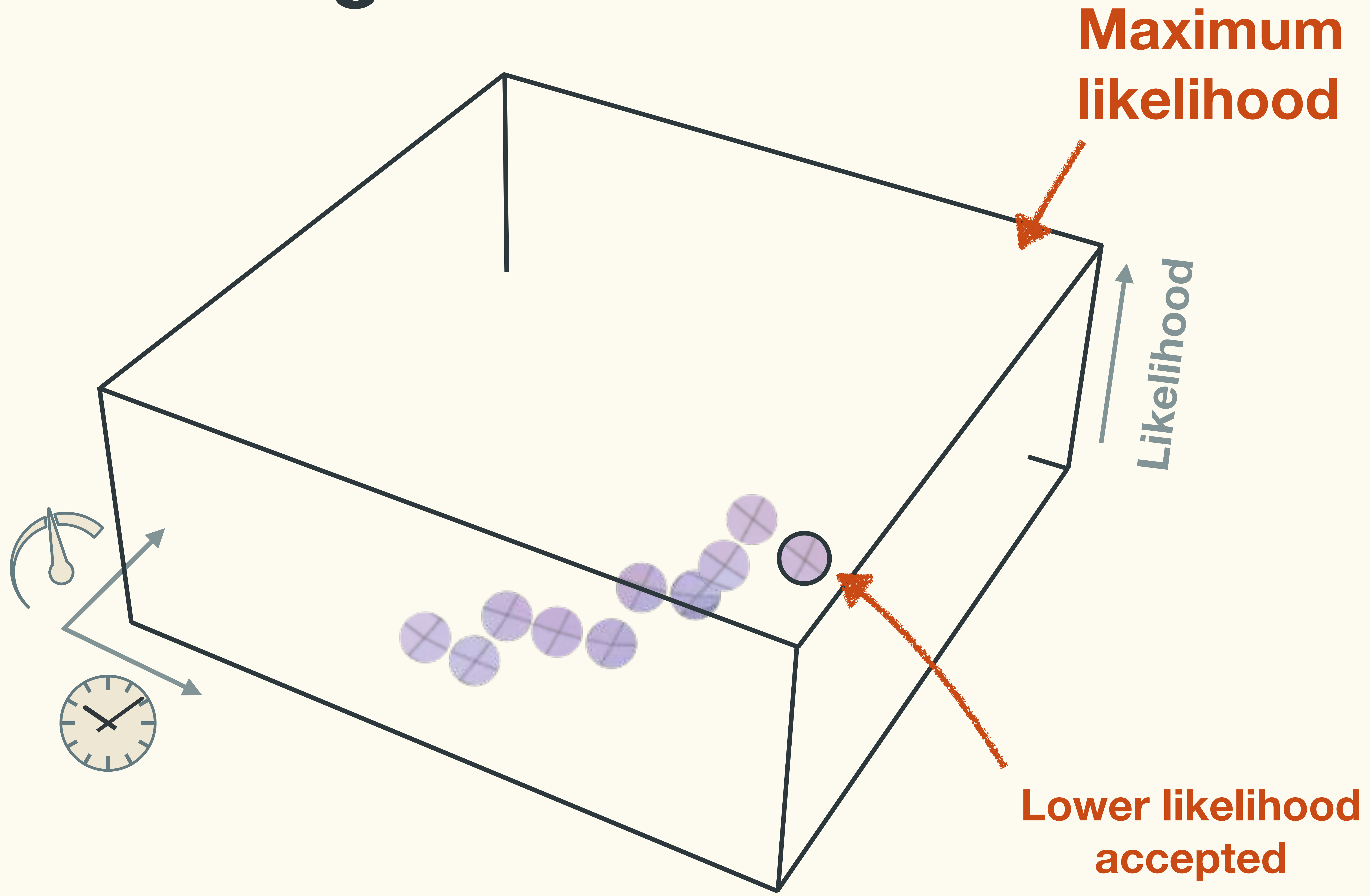
Heuristic search

Simulated annealing



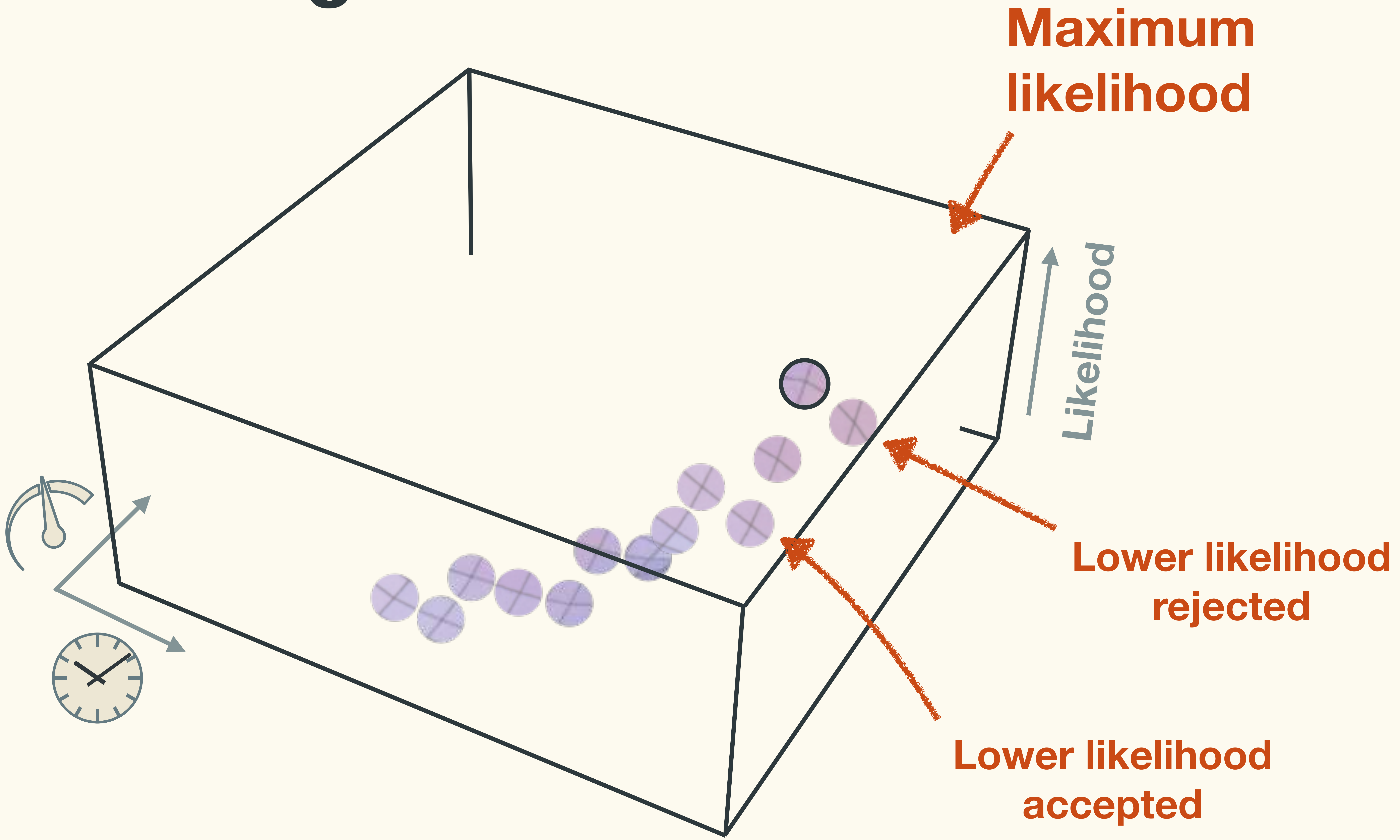
Heuristic search

Simulated annealing



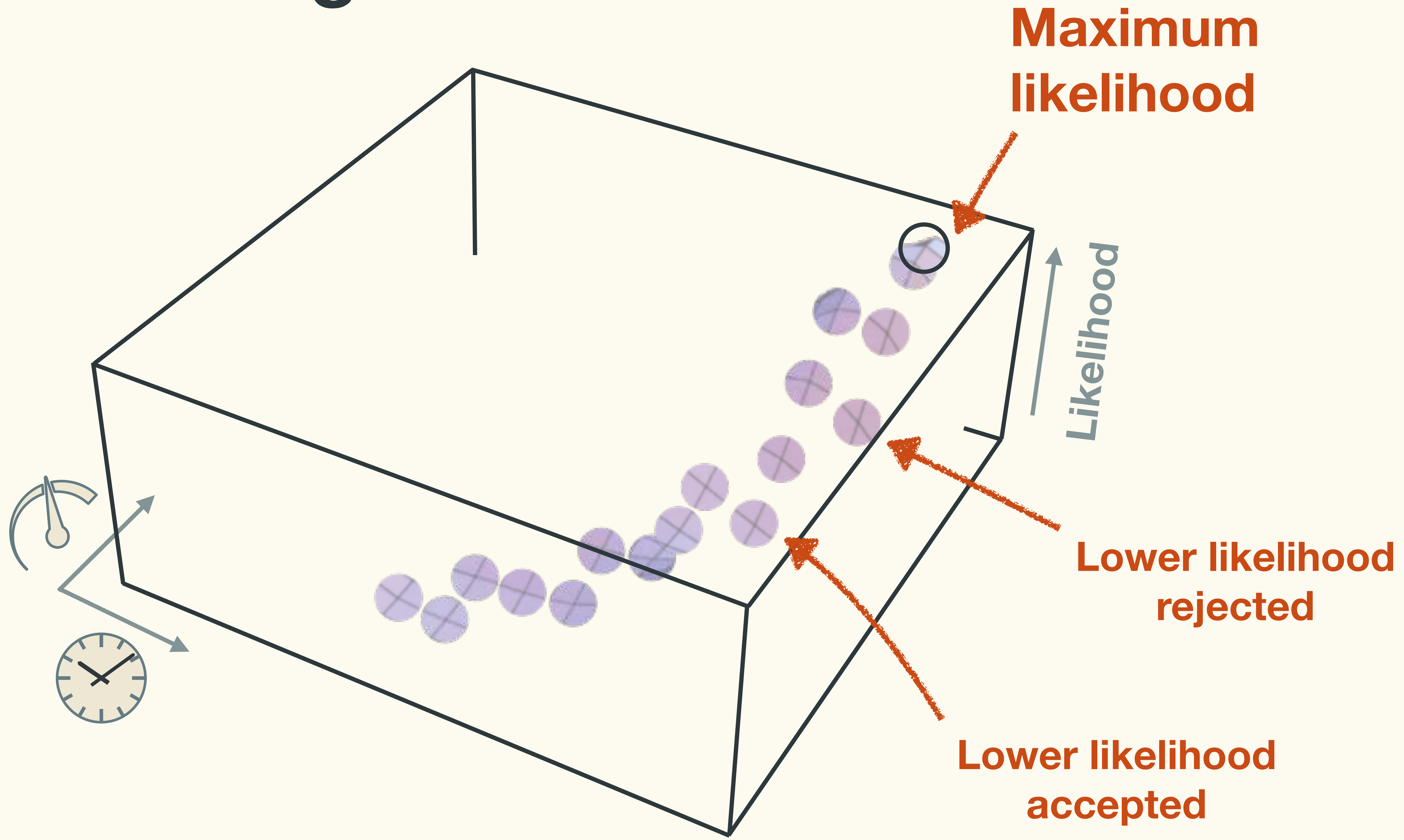
Heuristic search

Simulated annealing



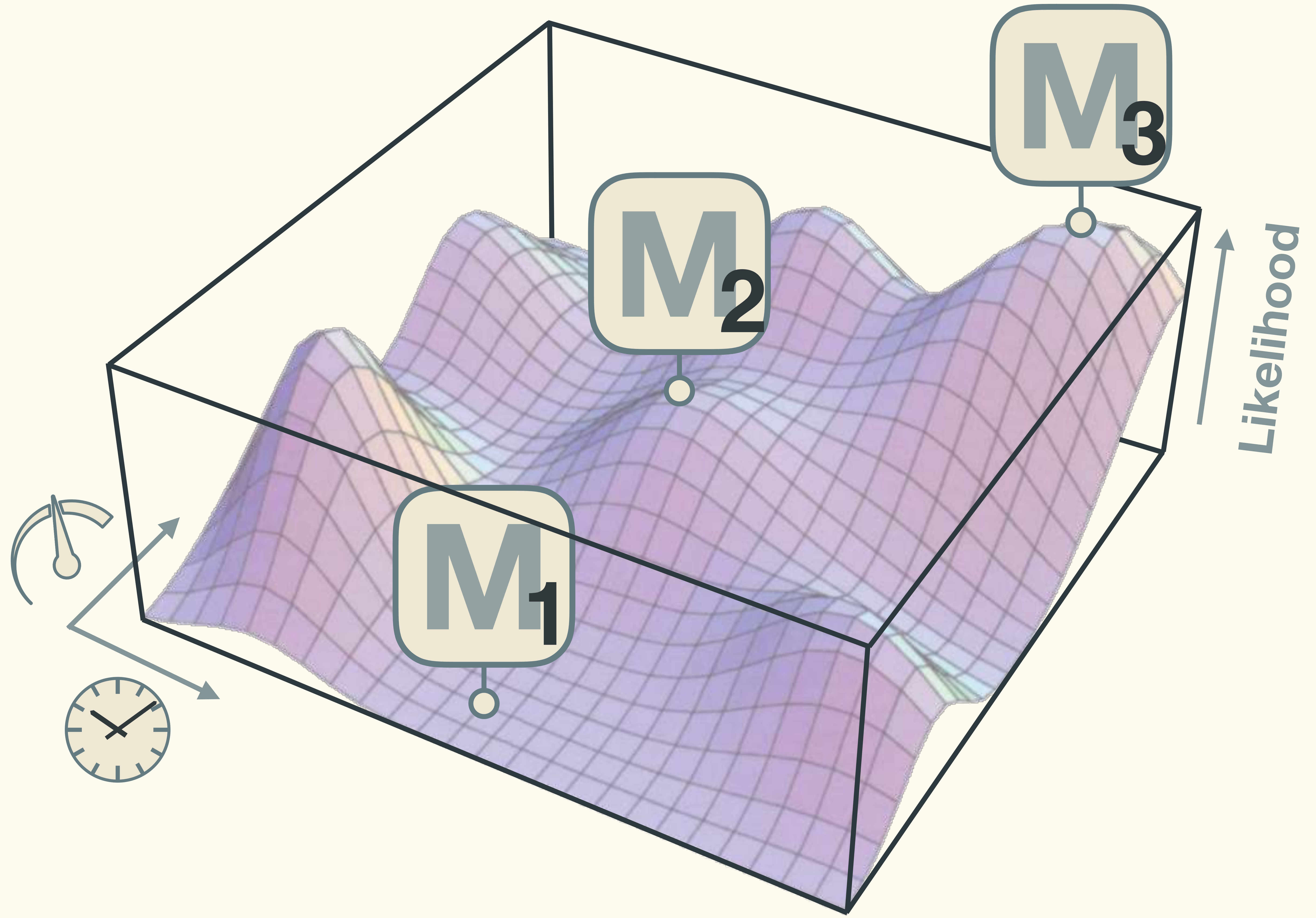
Heuristic search

Simulated annealing



Heuristic search

Simulated annealing



Likelihood

$$\log\left(L\left(\boxed{M}_1 \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -13.4$$

$$\log\left(L\left(\boxed{M}_2 \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -10.3$$

Likelihood

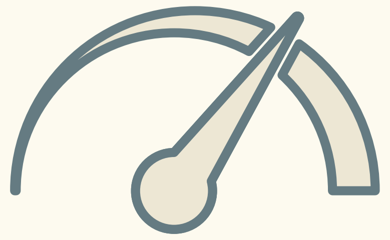
$$\log\left(L\left(\boxed{M_1} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -13.4$$

$$\log\left(L\left(\boxed{M_2} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -10.3$$

$$\log\left(L\left(\boxed{M_3} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -5.4$$

Likelihood

$$\log\left(L\left(\boxed{M_3} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -5.4$$

,  = 0.2,  = 0.01

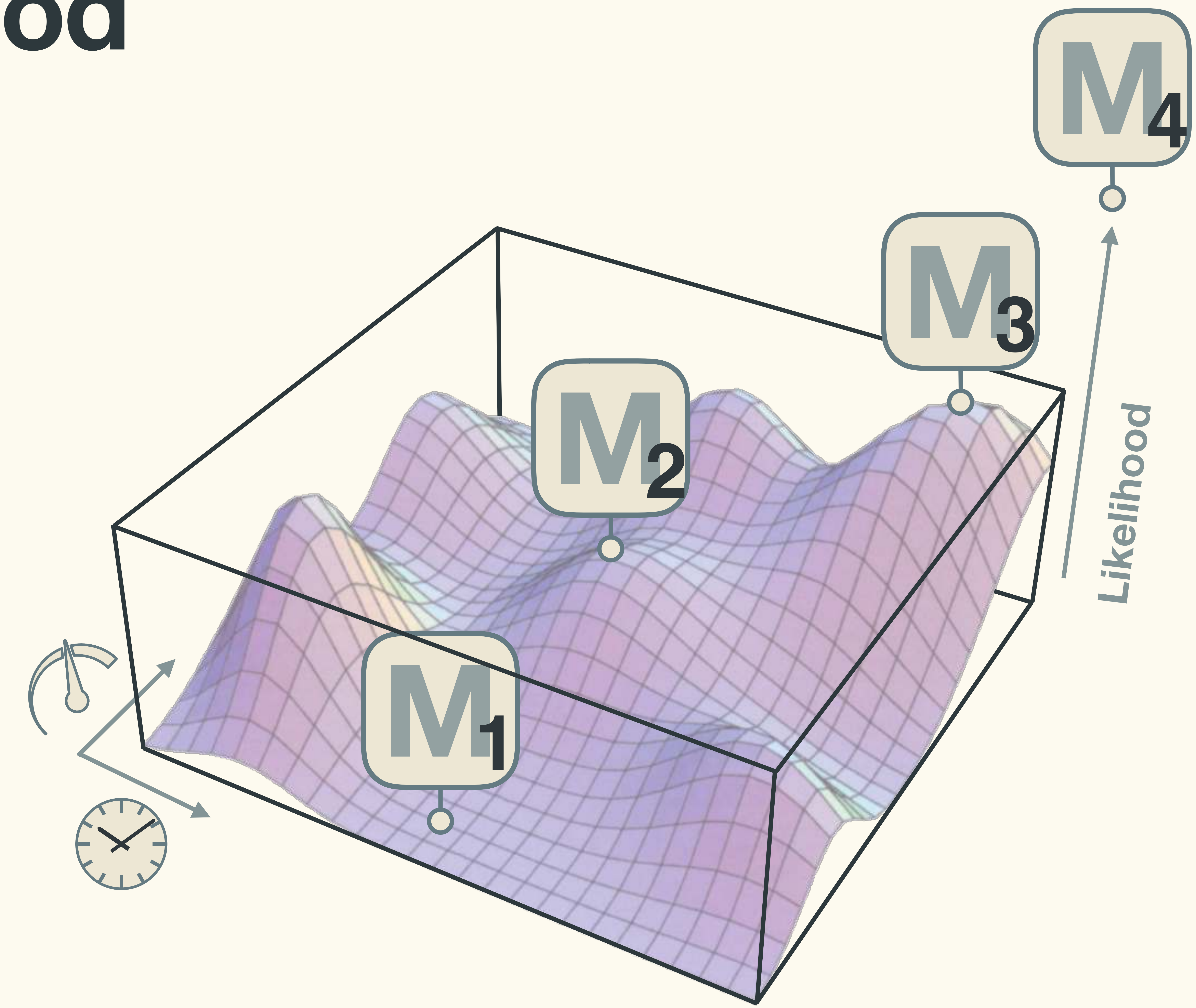
Likelihood

$$\log\left(L\left(\boxed{M_3} \,\middle|\, \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -5.4$$

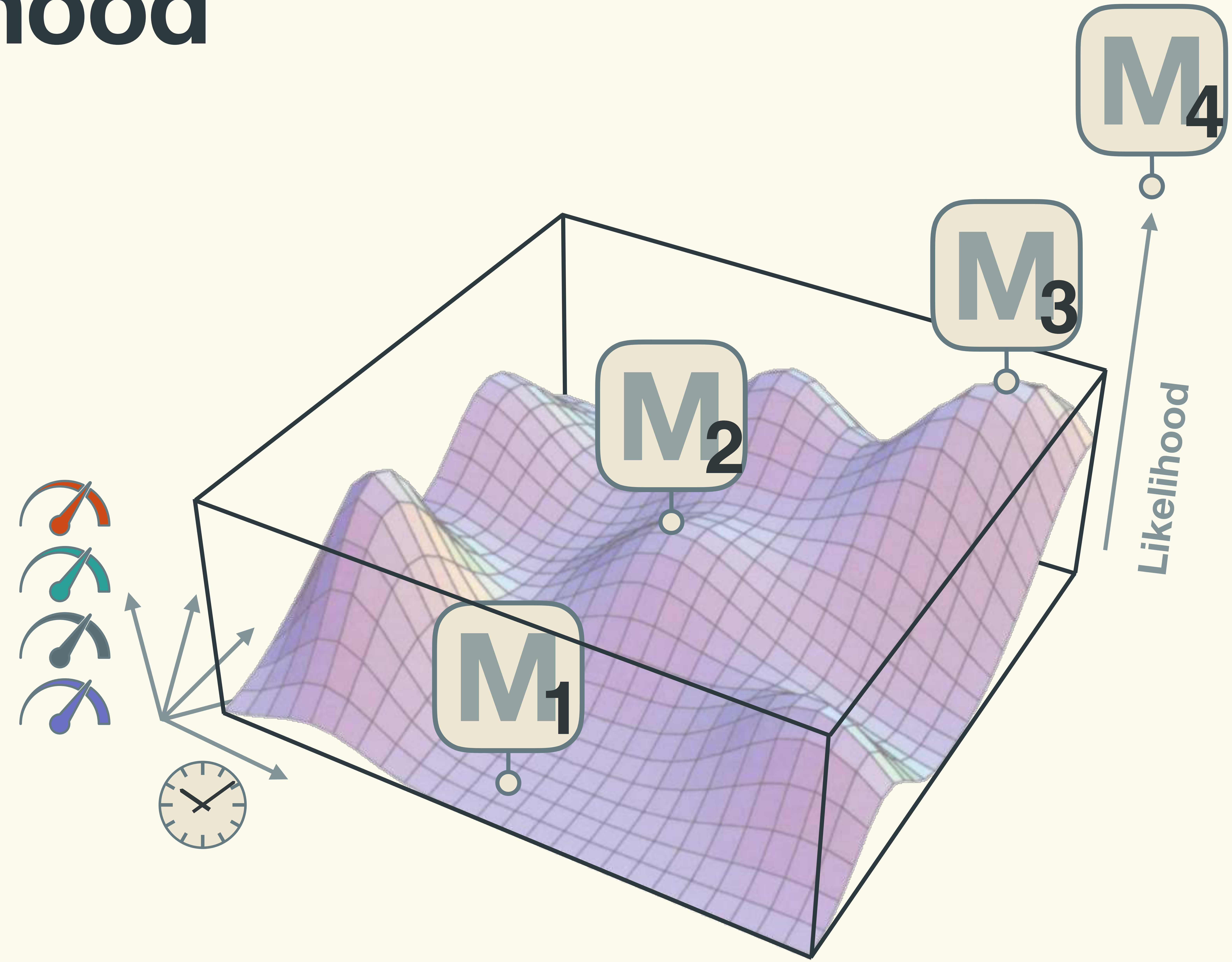
Likelihood

$$\log\left(L\left(\boxed{M_3} \mid \begin{matrix}\text{ACTTG}\\ \text{ACTGG}\end{matrix}\right)\right) = -5.4$$

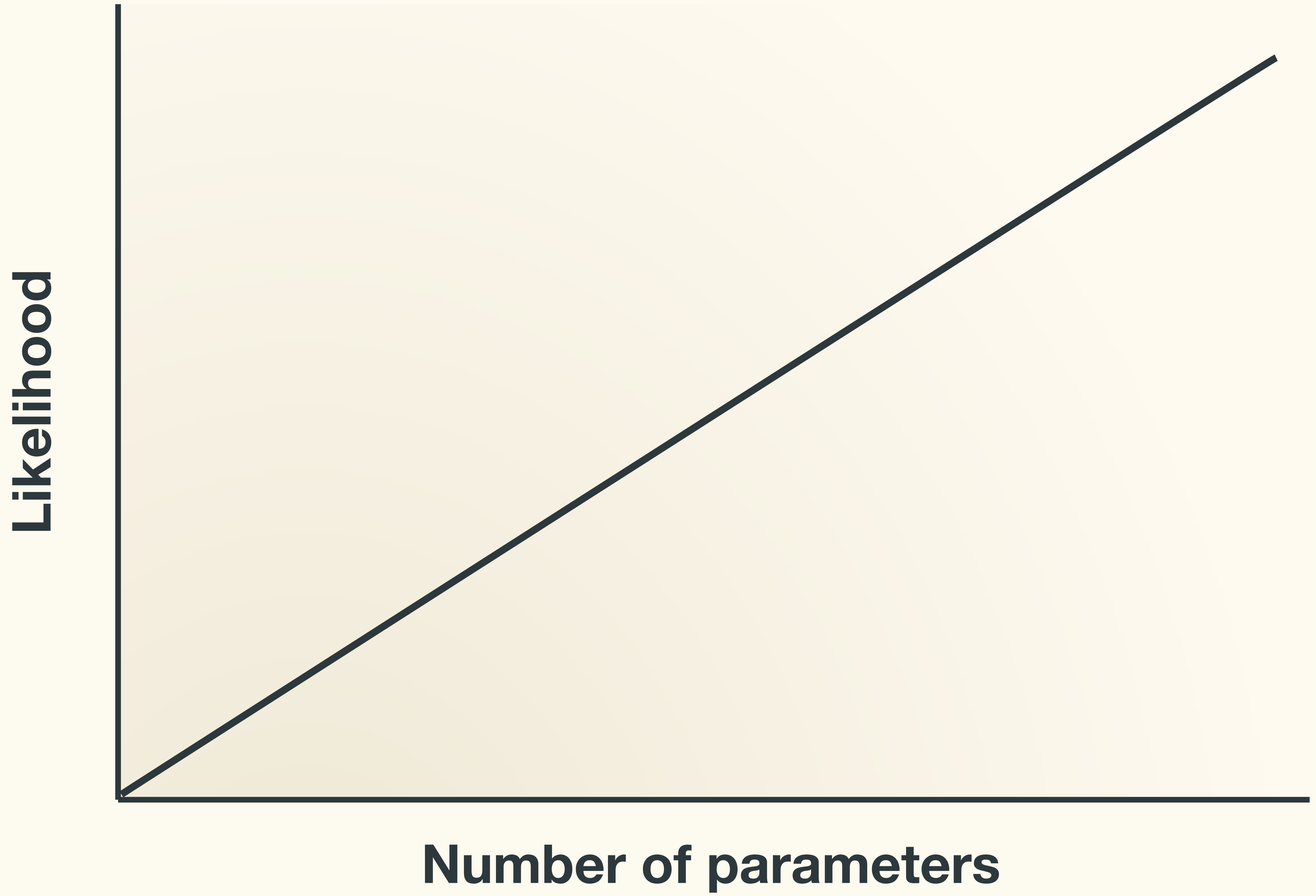
Likelihood



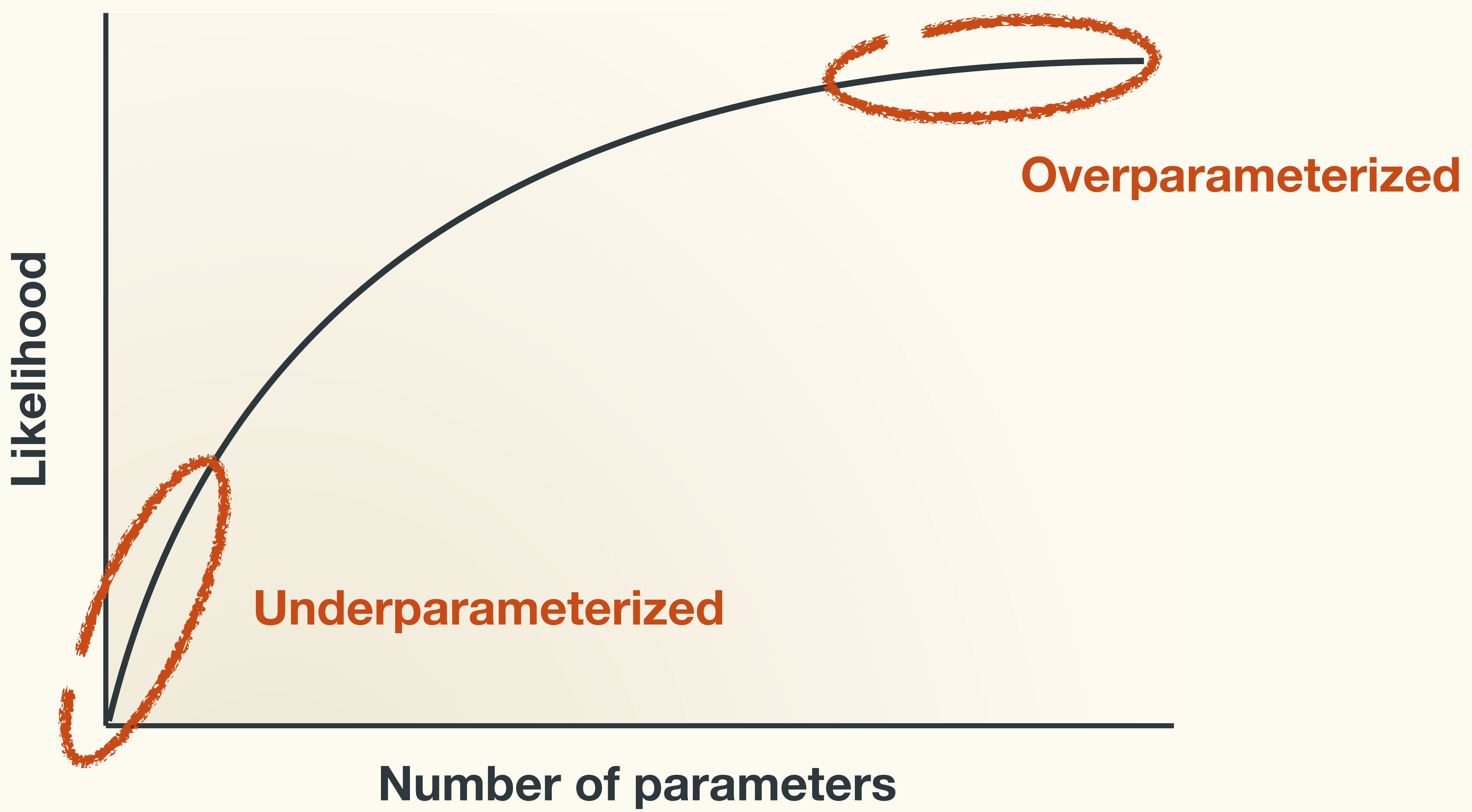
Likelihood



Likelihood

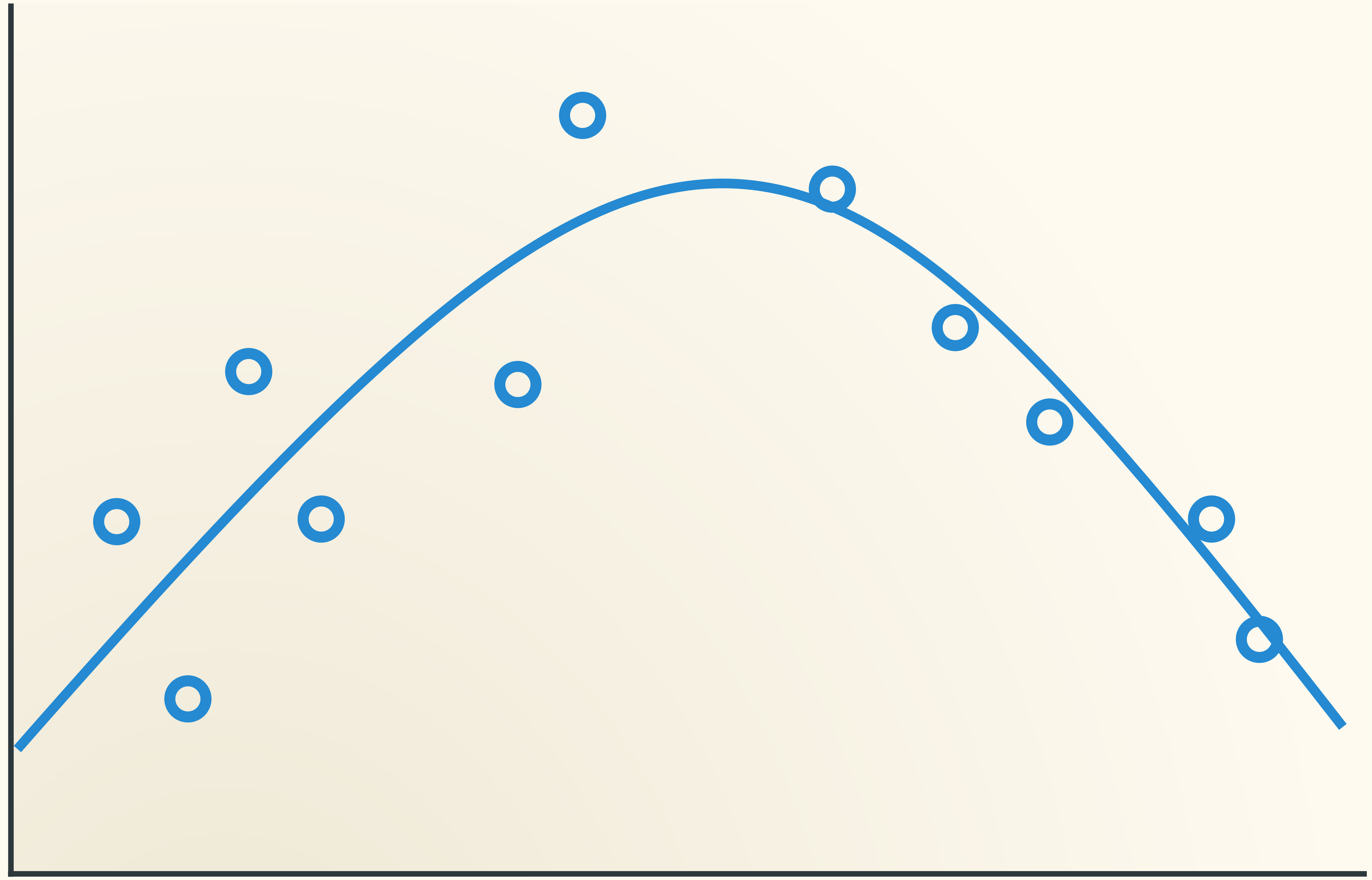


Likelihood

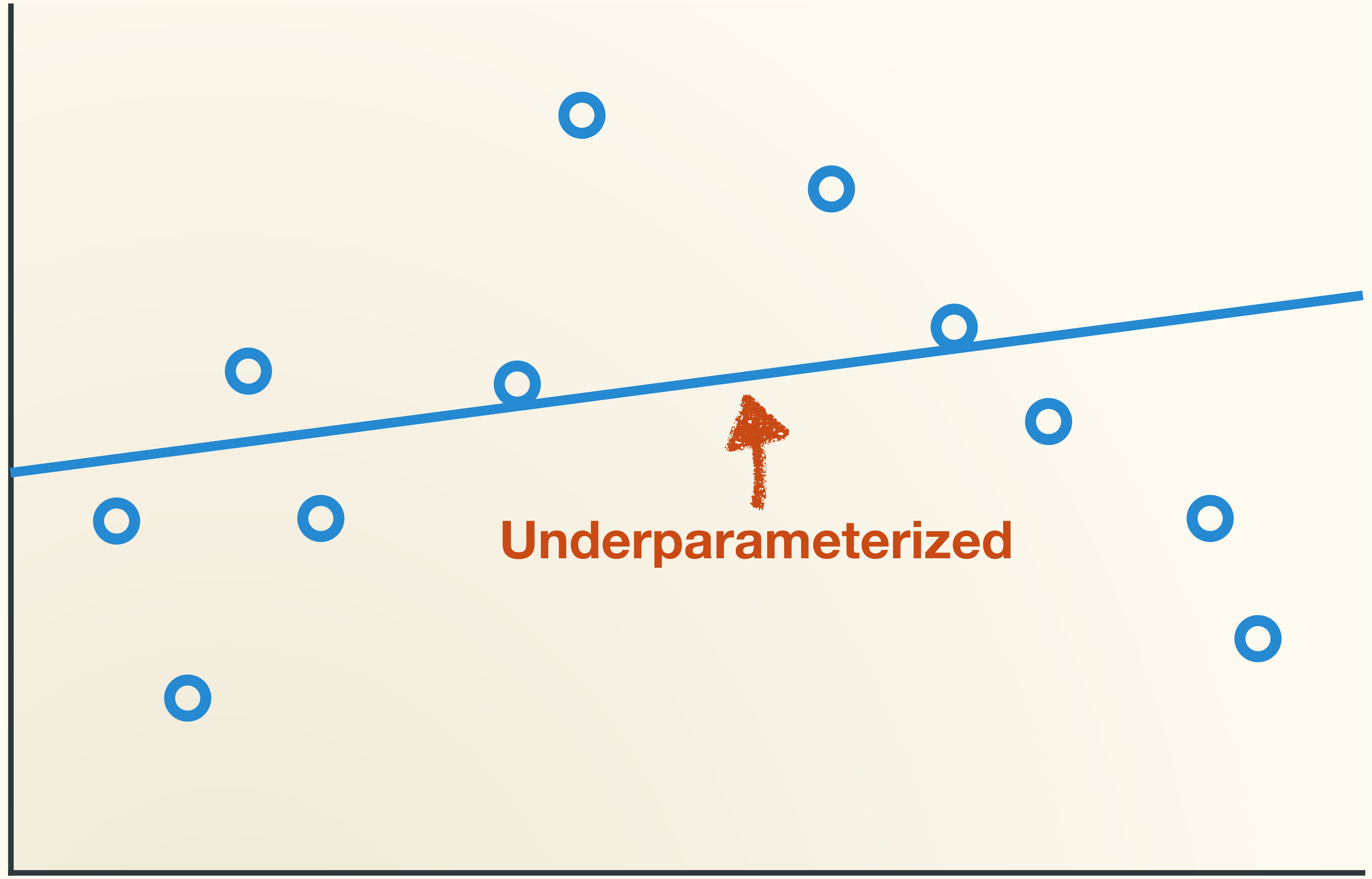


Overparameterized
Underparameterized
Likelihood
Number of parameters

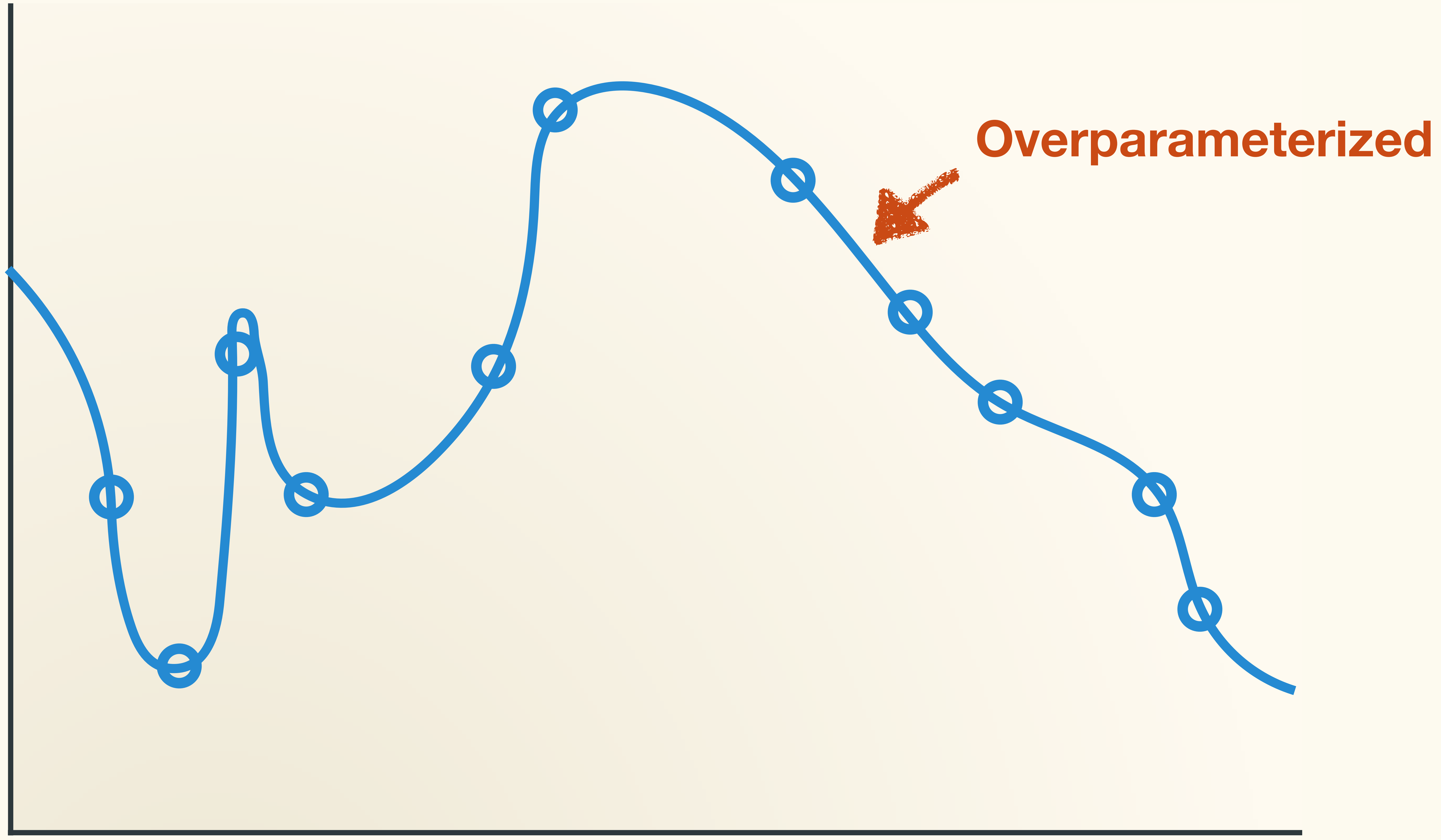
Parameterization



Parameterization



Parameterization



Likelihood ratio test

$$LRT = 2 \log \left(\frac{L \text{ (Complex model)}}{L \text{ (Simple model)}} \right)$$

Likelihood ratio test

$$LRT = 2\log\left(\frac{\boxed{M_4}}{\boxed{M_3}}\right)$$

Compared to
Chi-square score

Akaike information criterion

$$\text{AIC} = 2k - 2\left(\log(L)\right)$$

Number of parameters

Akaike information criterion

$$\mathrm{AIC}\left(\boxed{M_4}\right) = 2k - 2\left(\log\left(L \mid \boxed{M_4}\right)\right)$$

Akaike information criterion

$$\mathrm{AIC}\left(\boxed{M_4}\right) = 2k - 2\left(\log\left(L \mid \boxed{M_4}\right)\right)$$

$$\mathrm{AIC}\left(\boxed{M_3}\right) = 2k - 2\left(\log\left(L \mid \boxed{M_3}\right)\right)$$

Akaike information criterion

$$\delta AIC = AIC(\boxed{M_4}) - AIC(\boxed{M_3})$$

Bayesian inference

Bayesian inference

- A T-rex outside the door?
- Somebody pretending to be a T-rex?

Likelihood

- A T-rex outside the door?
- Somebody pretending to be a T-rex?

Likelihood

L ( | ROAR) = *P* (ROAR | )

L ( | ROAR) = *P* (ROAR | )

Likelihood

P (ROAR | ) ≈ 1

P (ROAR | ) ≈ 1

Likelihood

$L($  $|$ ROAR $) \approx 1$

$L($  $|$ ROAR $) \approx 1$

Probability

P ( | ROAR) = ?

Bayesian inference

Thomas Bayes
(1701–1761)

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays, that his deſign at firſt in thinking on the ſubject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumſtances, upon ſuppoſition that we know nothing concerning it but that, under the ſame circum-

Ref.: Price (1763) *Philosophical Transactions of the Royal Society* 53:370–418

Bayes' theorem

Model

P ( | ROAR) =

P(ROAR | ) × P()

P (ROAR)

Bayes' theorem

Data

P ( | ROAR) = P(ROAR | ) × P() / P (ROAR)

Bayes’ theorem

Posterior probability

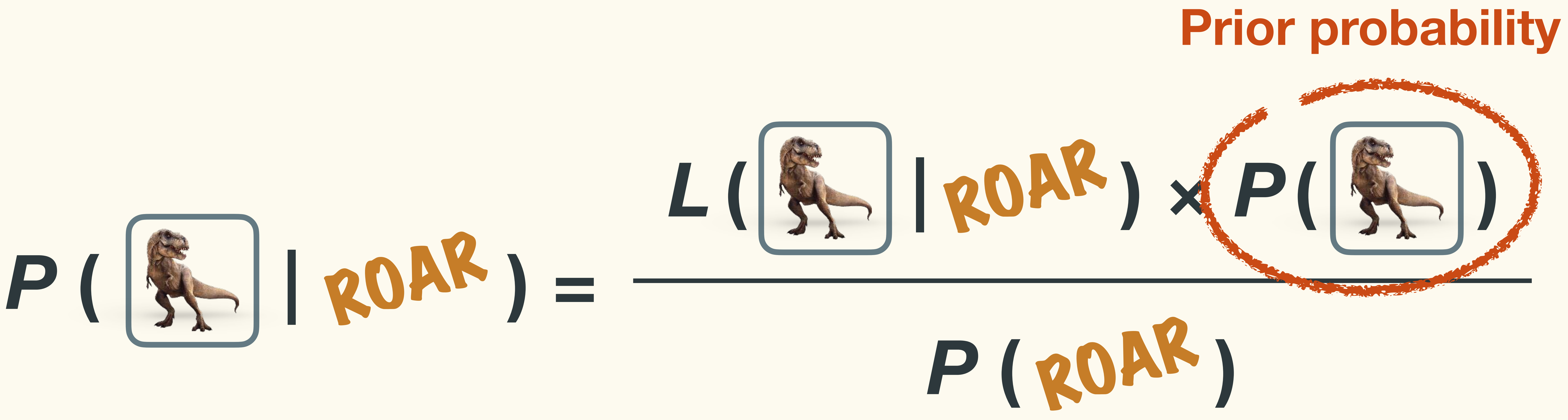
P ( | ROAR) = P(ROAR | ) × P() / P (ROAR)

Bayes’ theorem

Likelihood

P ( | ROAR) = L ( | ROAR) × P () / P (ROAR)

Bayes' theorem



Bayes' theorem

P( | ROAR) = L( | ROAR) × P() / P(ROAR)

Constant

Bayes' theorem

Bayes' theorem

≈ 1

P( | ROAR) = L( | ROAR) × P() / P(ROAR)

P( | ROAR) = L( | ROAR) × P() / P(ROAR)

Bayes' theorem

P ( | ROAR) = 1 × P() / P(ROAR)

≈ 1

P ( | ROAR) = L ( | ROAR) × P() / P(ROAR)

Bayes' theorem

P ( | ROAR) = 1 × P() / P(ROAR)

P ( | ROAR) = 1 × P() / P(ROAR)

Bayes’ theorem

P ( | ROAR) = *P* () / *P* (ROAR)

Prior probabilities

P ( | ROAR) = *P* () / *P* (ROAR)

Bayes' theorem

≈ 0.00000001

$P($  $|$ ROAR $) =$ $P($  $)$ / $P($ROAR$)$

$P($  $|$ ROAR $) =$ $P($  $)$ / $P($ROAR$)$

Bayes' theorem

P ( | ROAR) = P() / P(ROAR) $\approx$ 0.00000001 / P(ROAR)

P ( | ROAR) = P() / P(ROAR)

Bayes' theorem

P ( | ROAR) = P() / P(ROAR) ≈ 0.00000001 / P(ROAR)

P ( | ROAR) = P() / P(ROAR)

≈ 0.1

Bayes' theorem

$P($  $|$ ROAR $) =$

$P($  $)$

$P($ROAR$)$

$\approx \dfrac{0.00000001}{P(\text{ROAR})}$

$P($  $|$ ROAR $) =$

$P($  $)$

$P($ROAR$)$

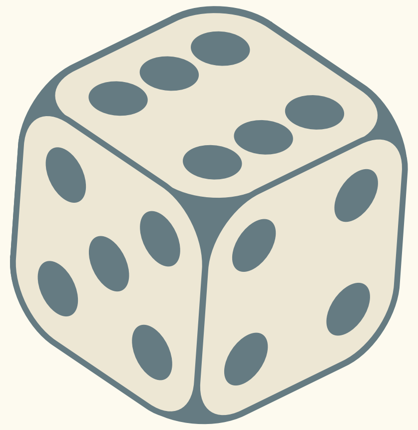
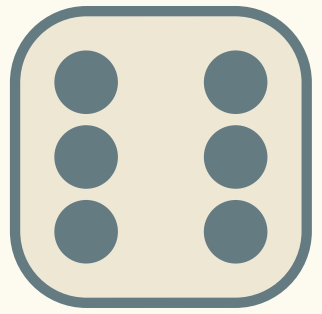
$\approx \dfrac{0.1}{P(\text{ROAR})}$

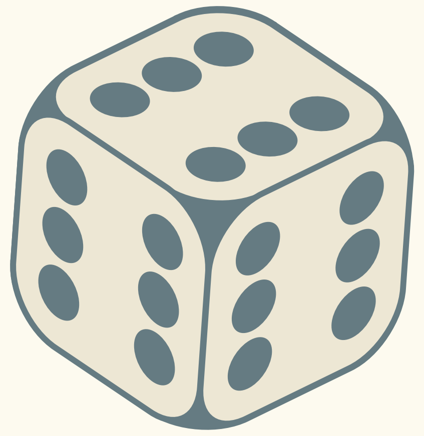
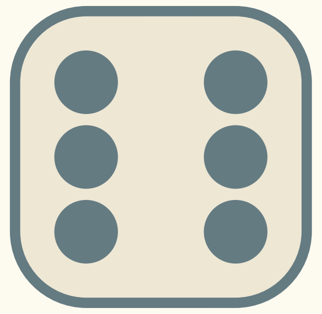
Bayes' theorem

$P($  $\mid$ ROAR $) \gg P($  $\mid$ ROAR $)$

Likelihood

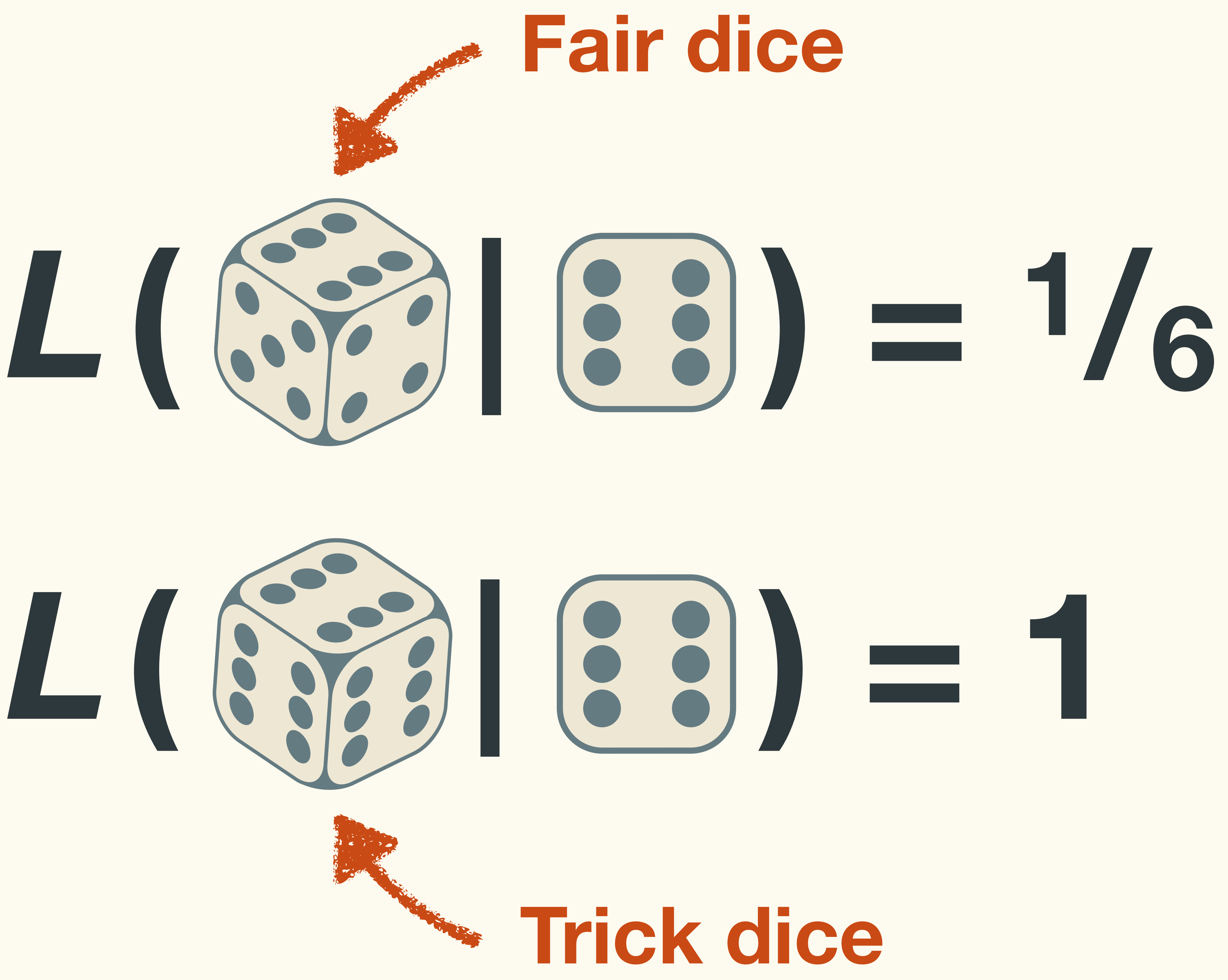
Fair dice

$L($  $|$  $) = 1/6$

$L($  $|$  $) = 1$

Trick dice

Likelihood

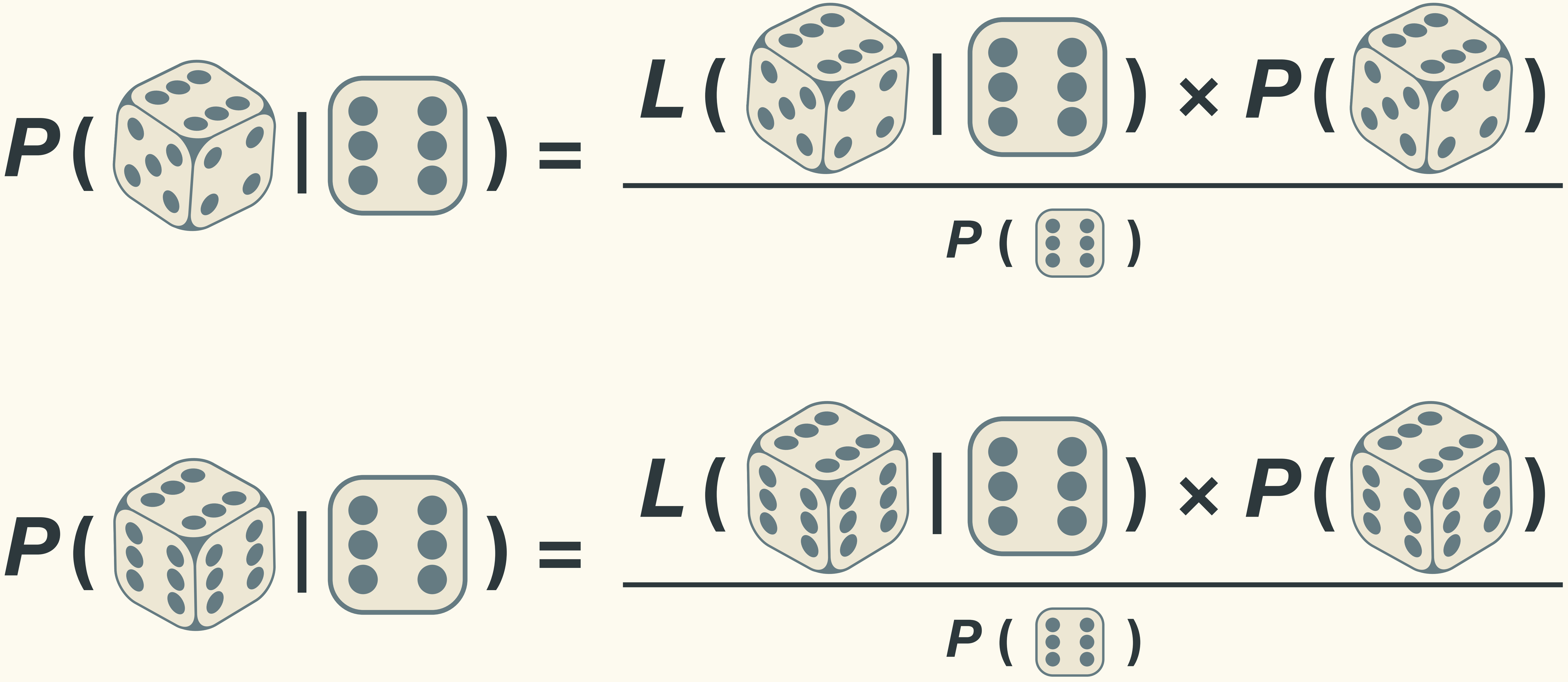


Bayes’ theorem

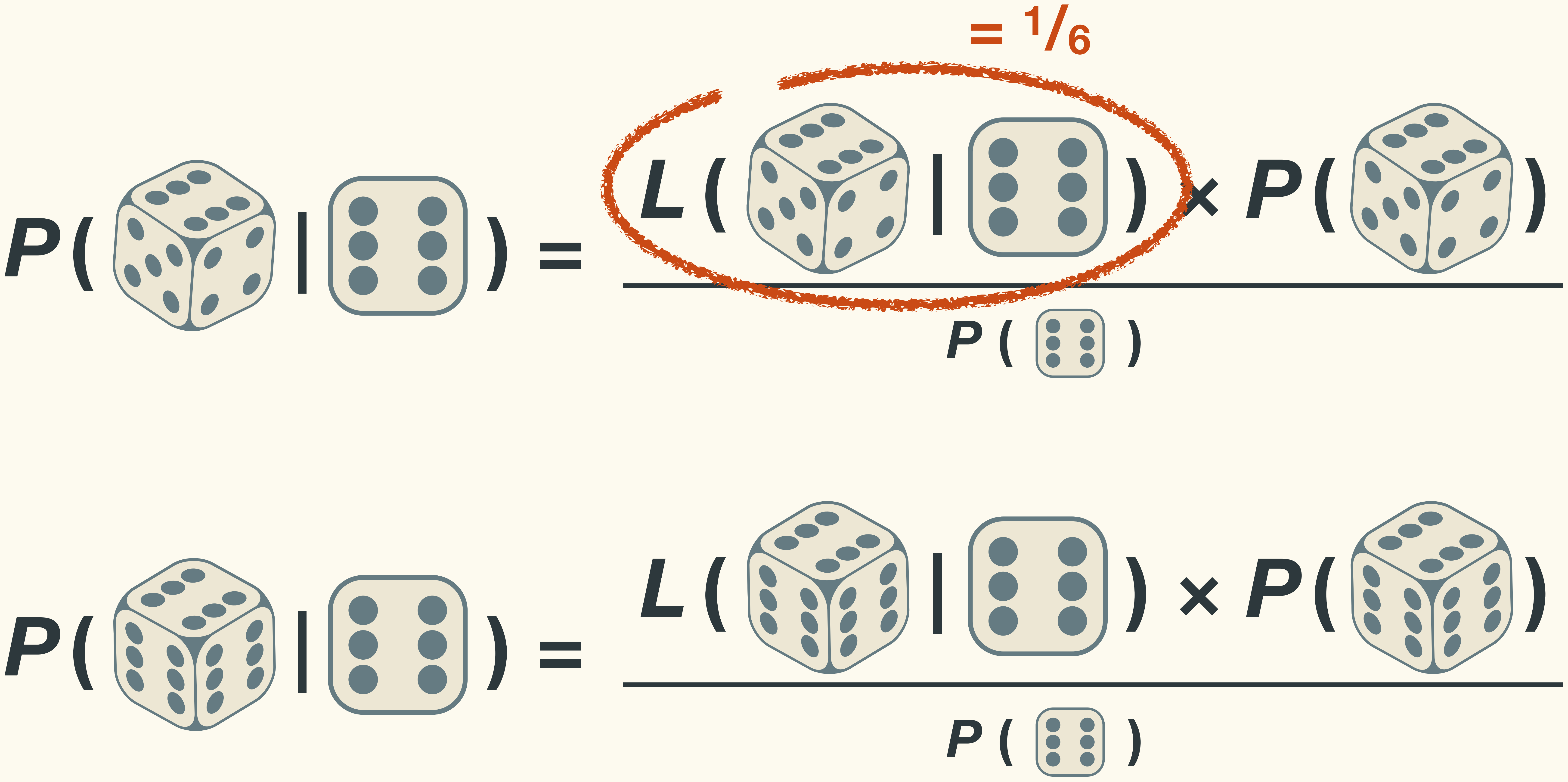
Posterior probability

Likelihood

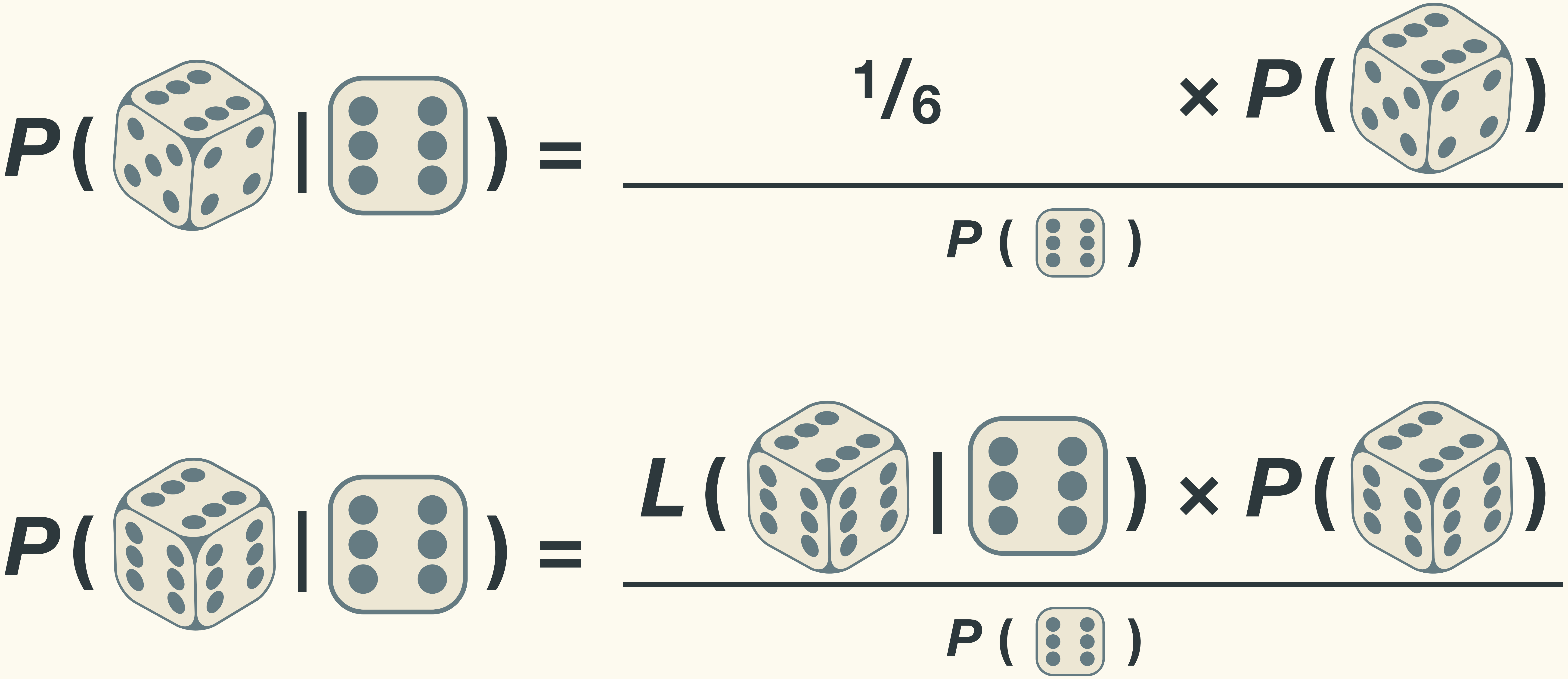
Prior probability



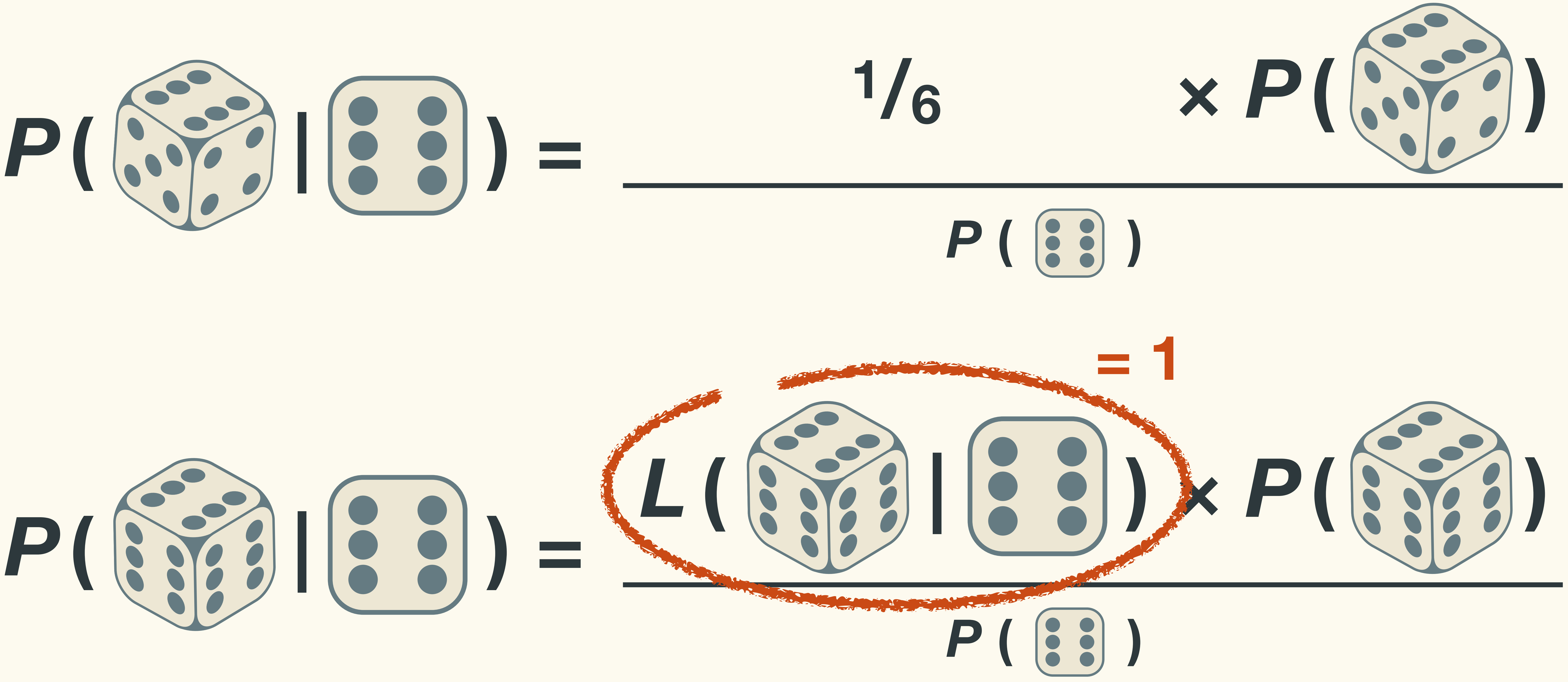
Bayes' theorem



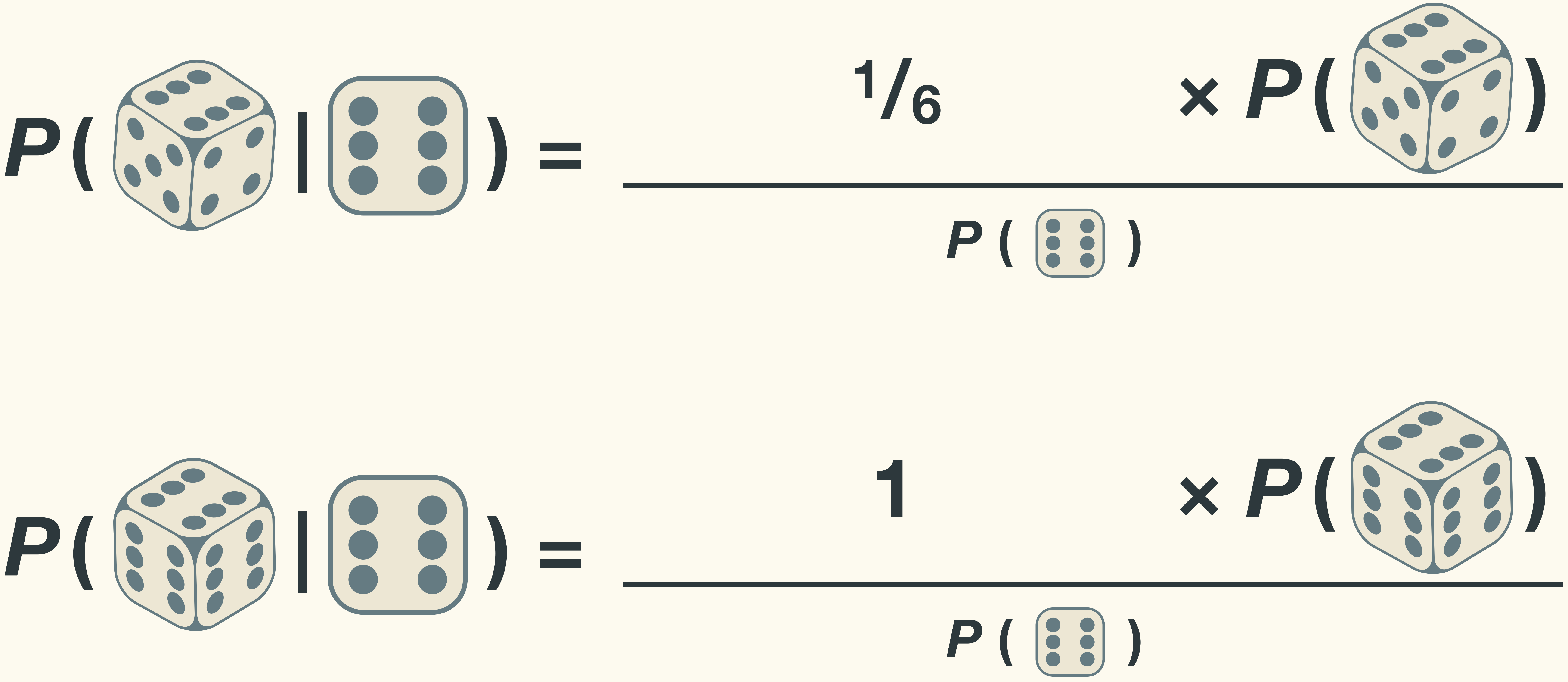
Bayes' theorem



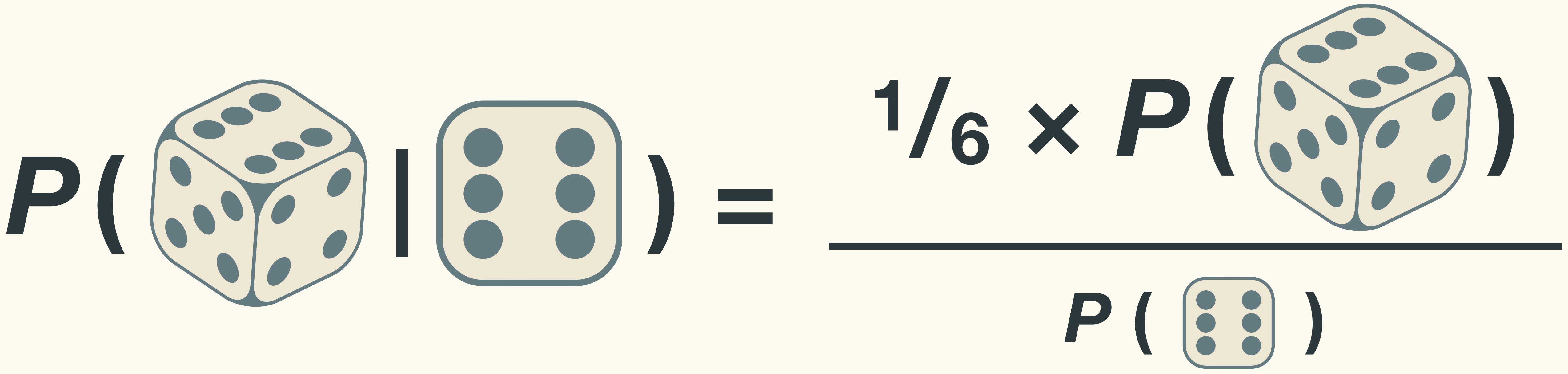
Bayes' theorem

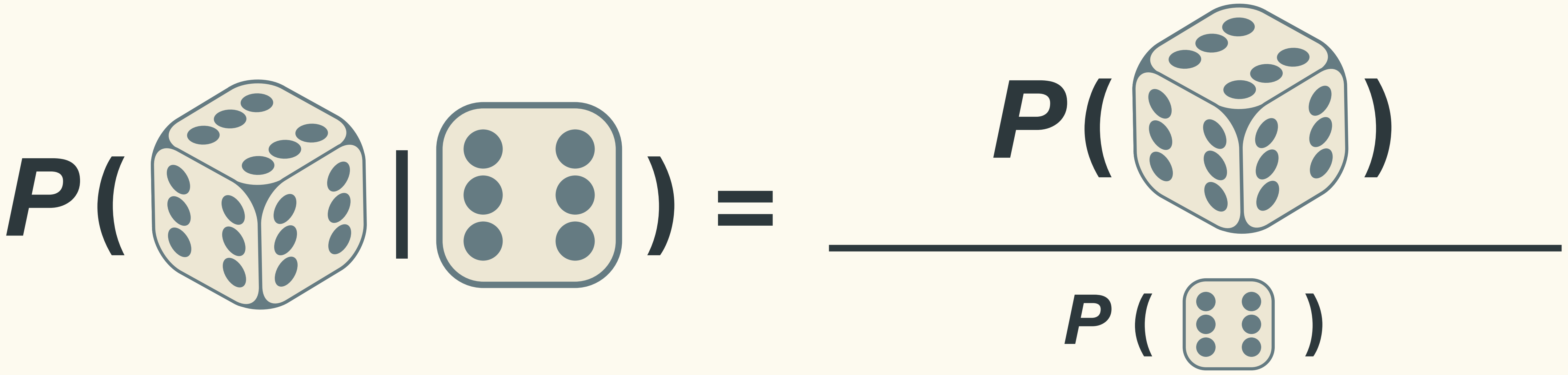


Bayes' theorem

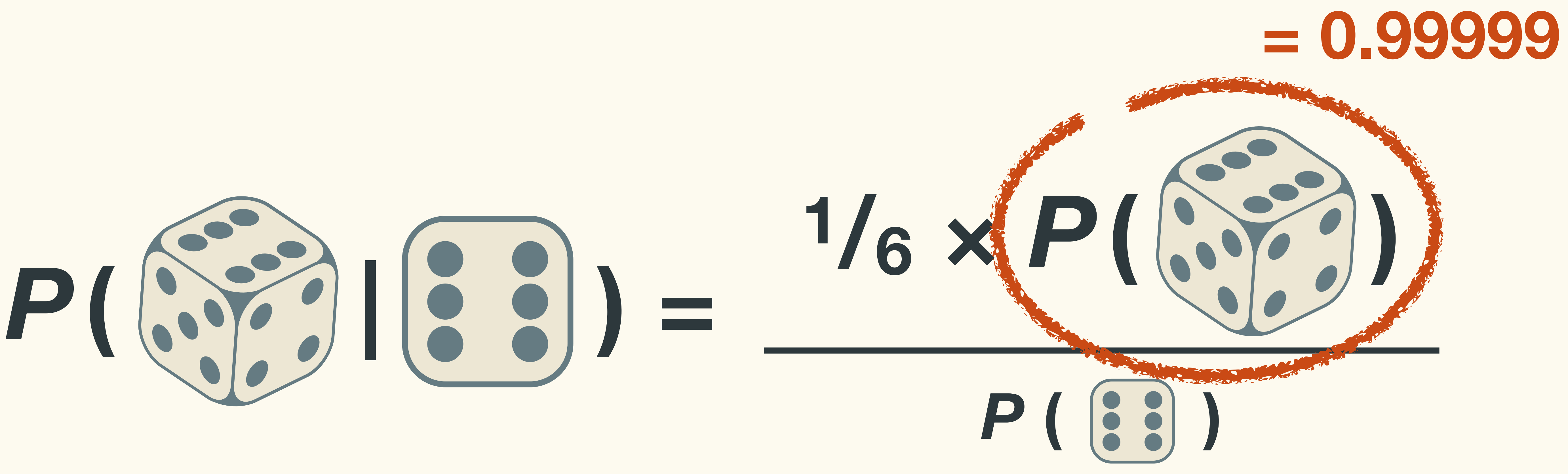


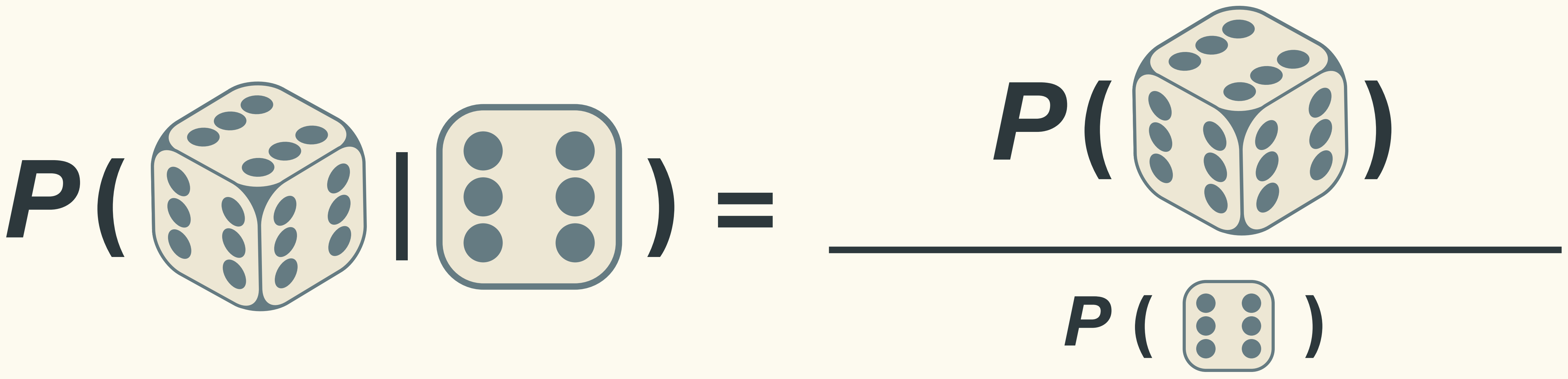
Bayes’ theorem



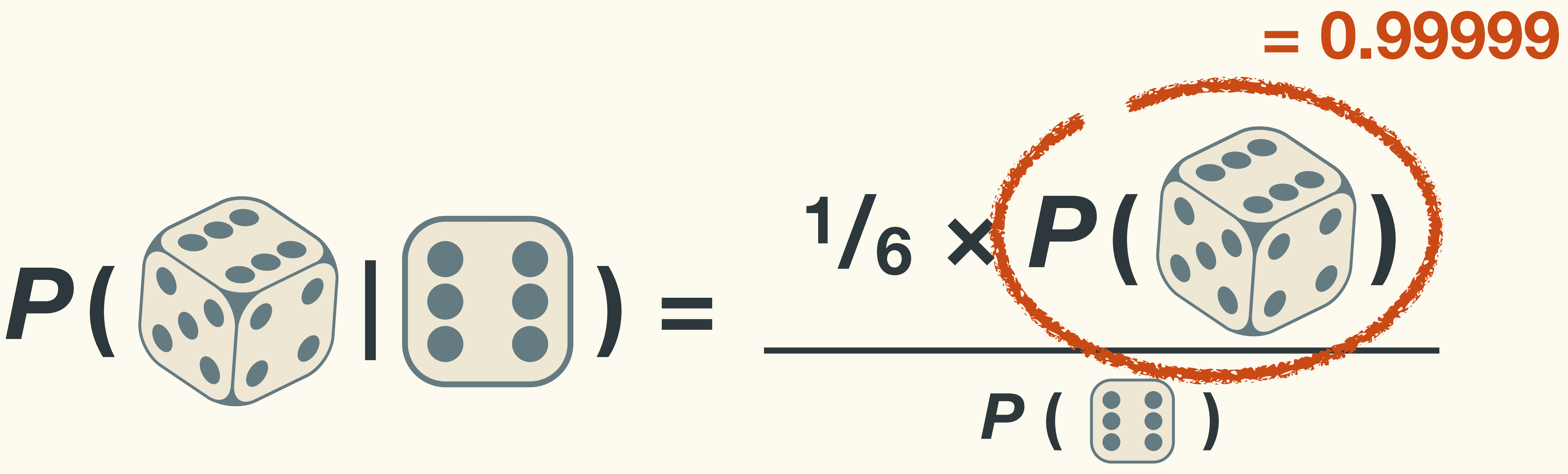


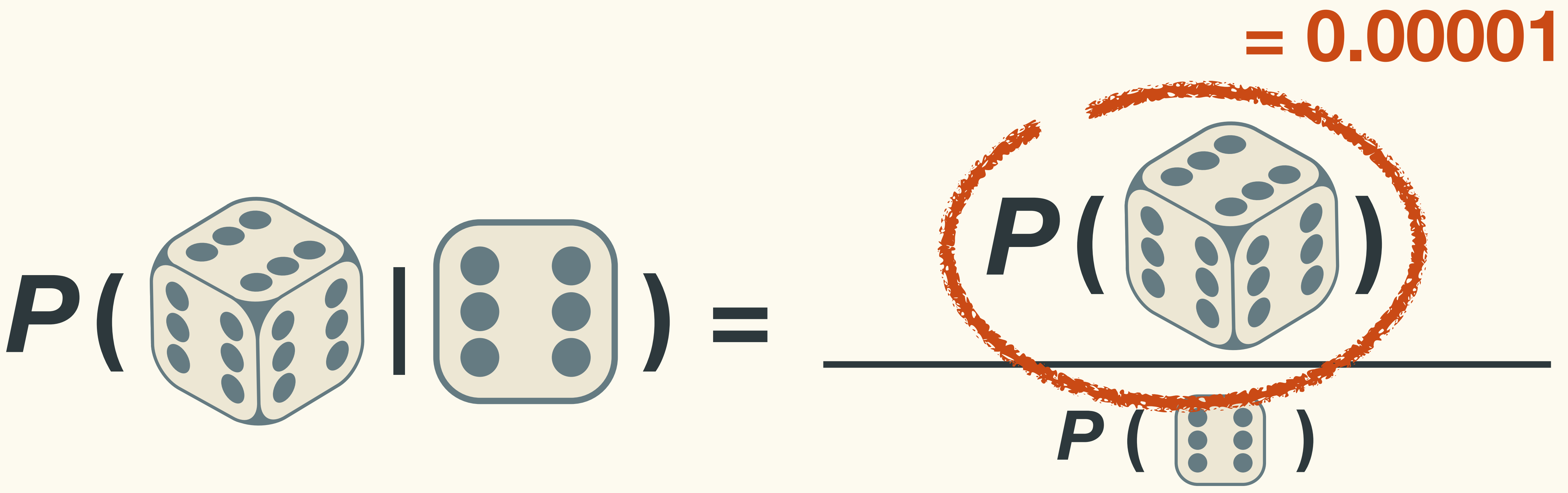
Bayes' theorem



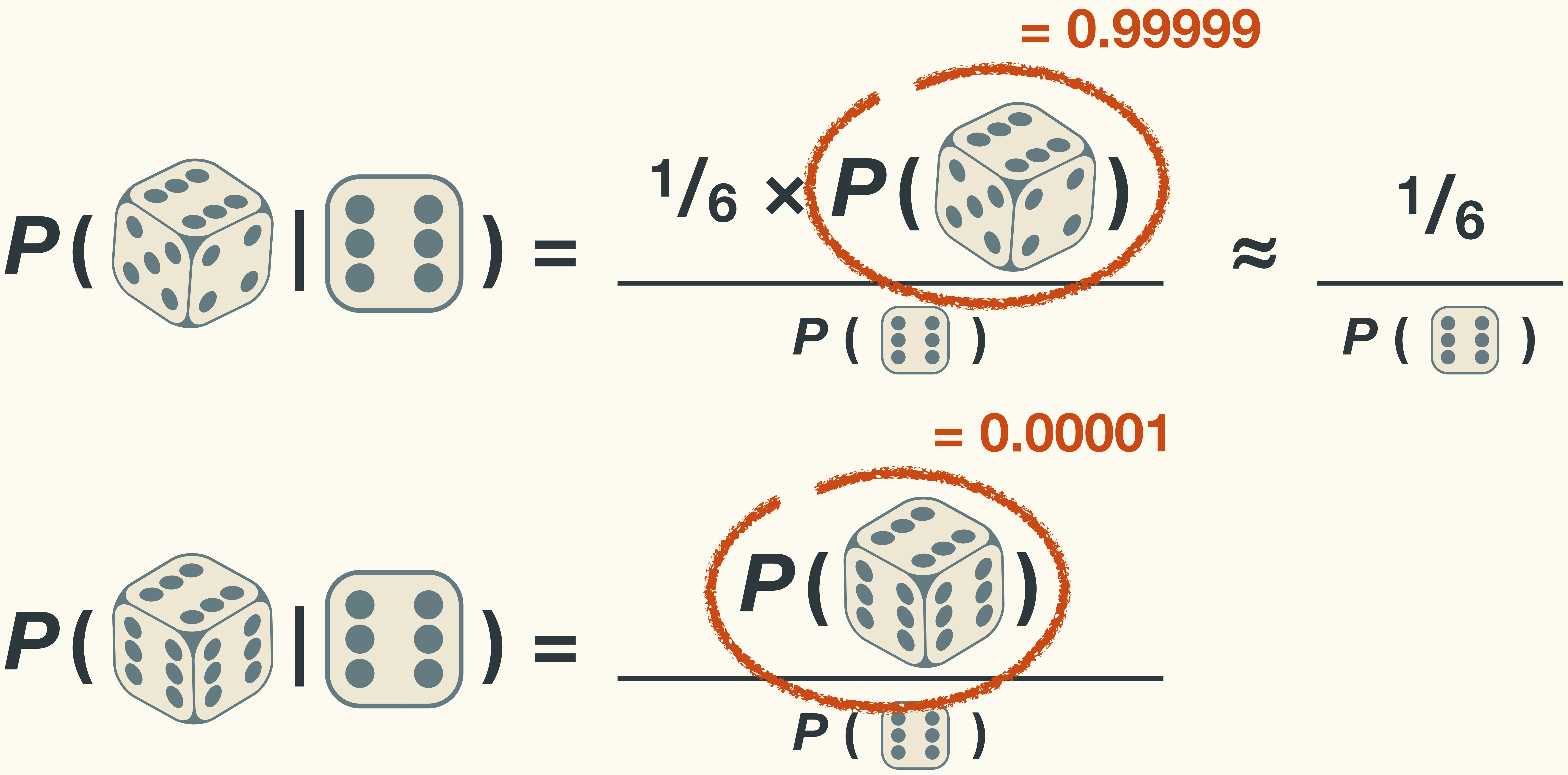


Bayes' theorem

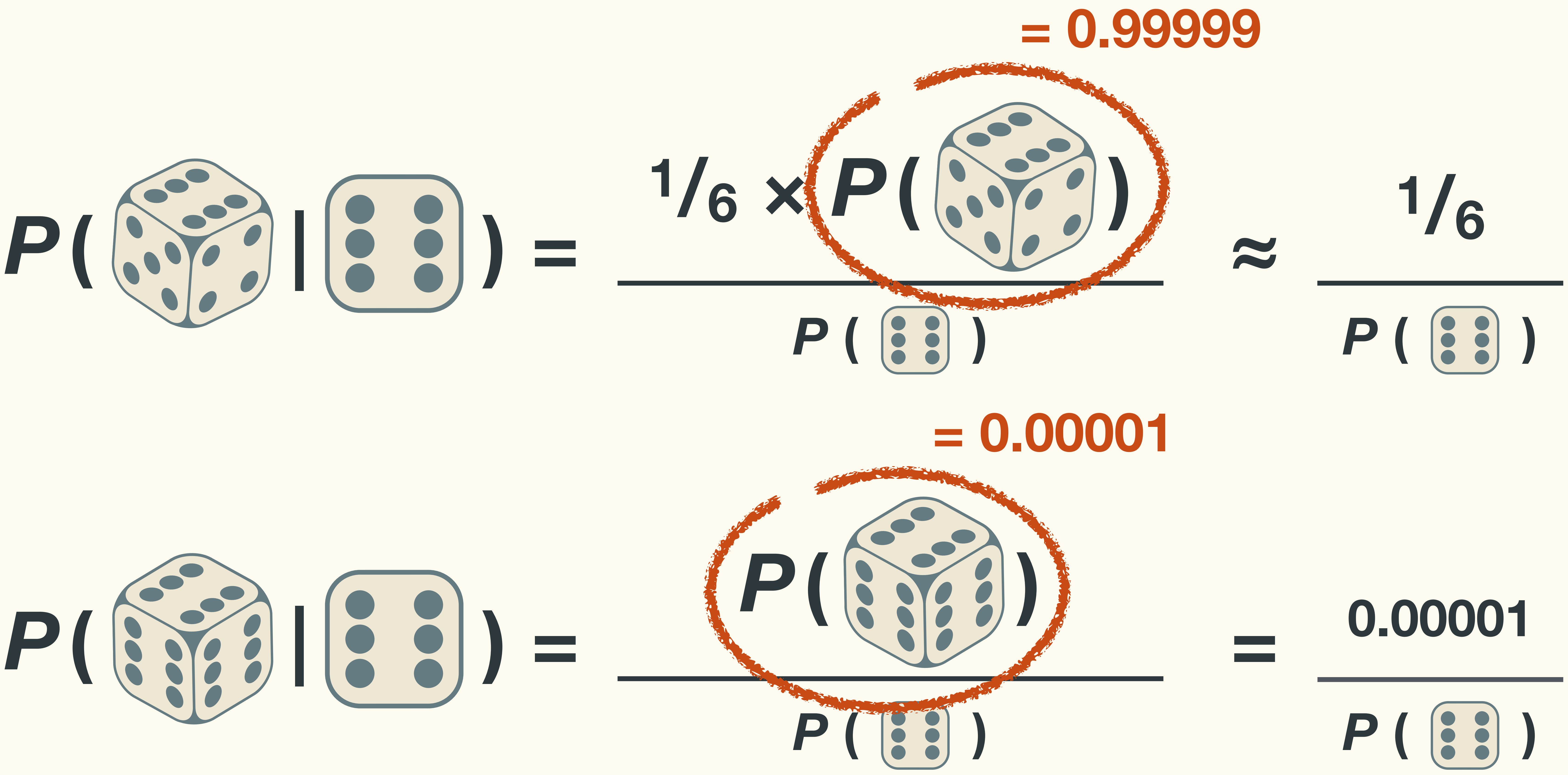




Bayes' theorem



Bayes' theorem



Estimating model parameters using Bayesian inference

MCMC

Markov-chain Monte Carlo

Monte Carlo

Monte Carlo methods

Stanisław Ulam
(1909–1984)

Ref.: Eckhardt (1987) *Los Alamos Science Special Issue*, 131–143

Monte Carlo methods

Stanisław Ulam
(1909–1984)

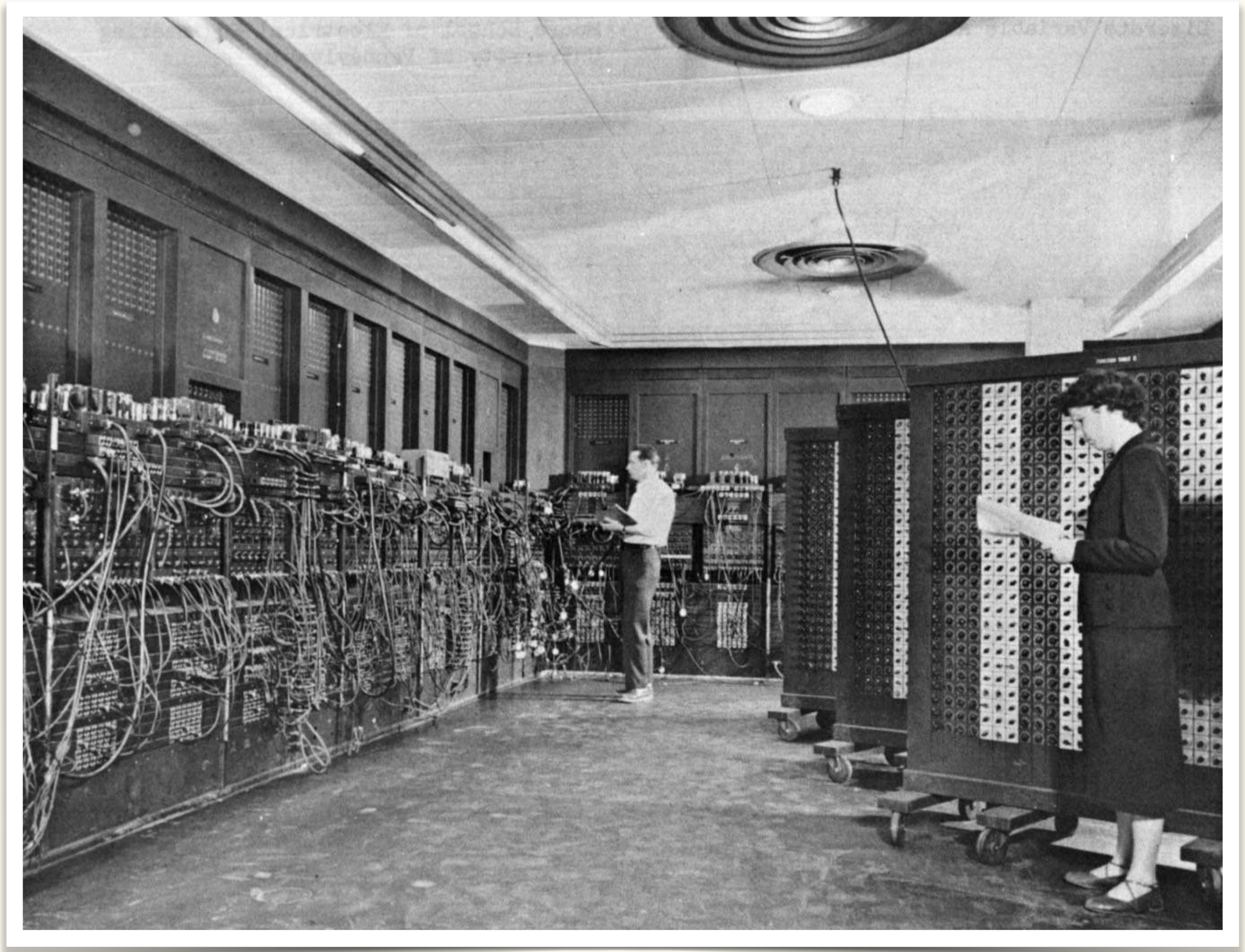
"What are the chances that a Canfield solitaire laid out with 52 cards will come out successfully?"

Stanisław Ulam, 1946

Ref.: Eckhardt (1987) *Los Alamos Science Special Issue*, 131–143

Monte Carlo methods

ENIAC
1946



Monte Carlo methods

Stanisław Ulam
(1909–1984)

"Stan had an uncle who would borrow money from relatives because he 'just had to go to Monte Carlo'."

Nicholas Metropolis

Ref.: Eckhardt (1987) *Los Alamos Science Special Issue*, 131–143

Monte Carlo

Markov chains

Andrey Markov
(1856–1922)

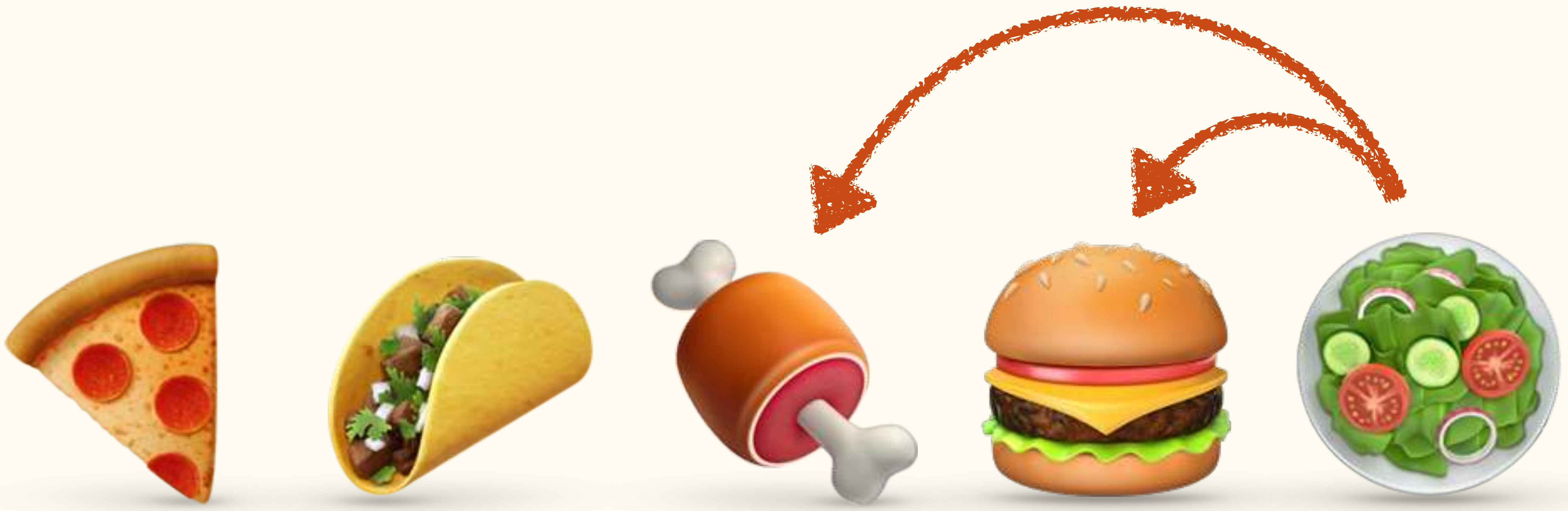
Markov chains

Markov chains

Markov chains

Markov chains

Markov chains

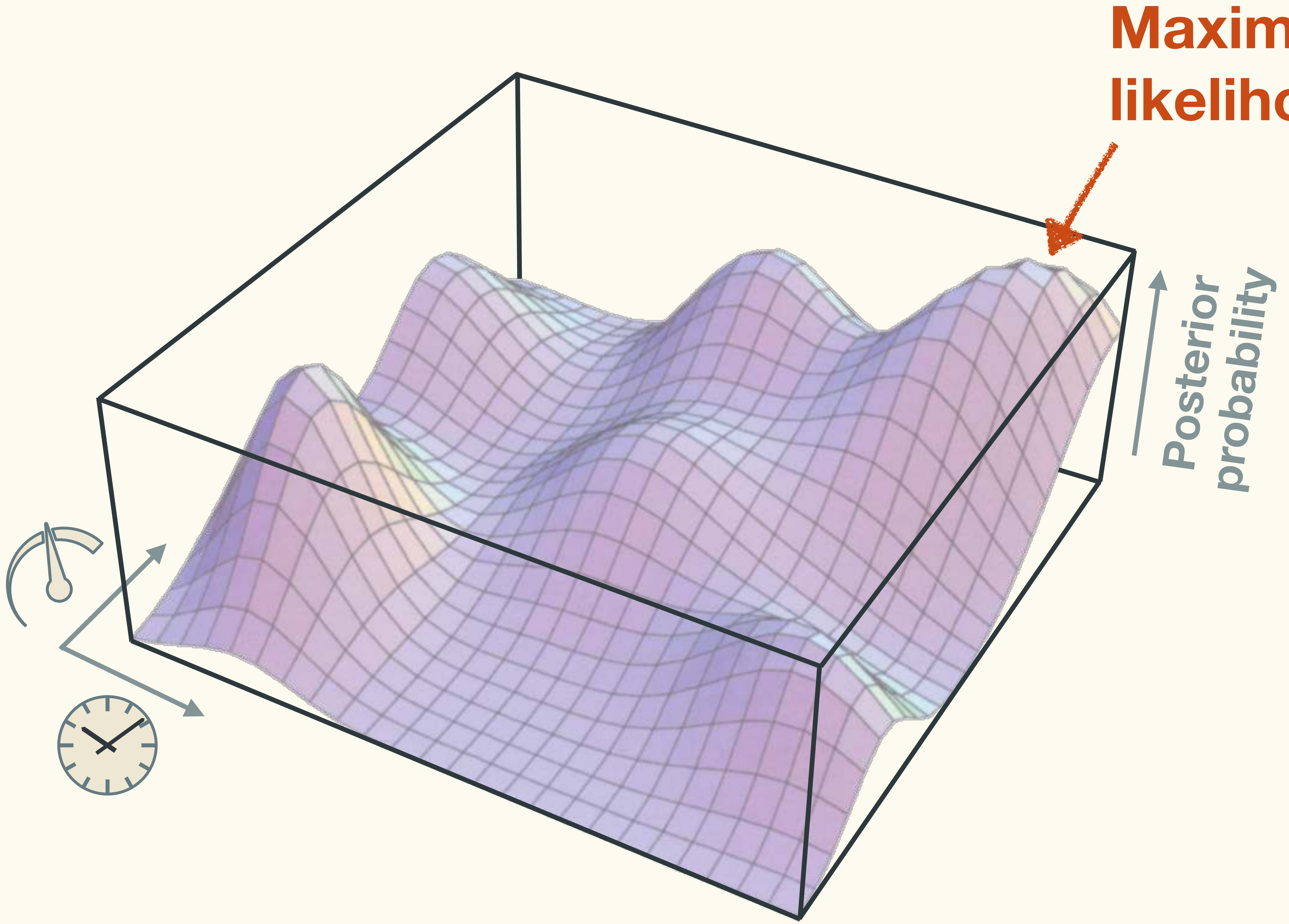


MCMC

Markov-chain Monte Carlo

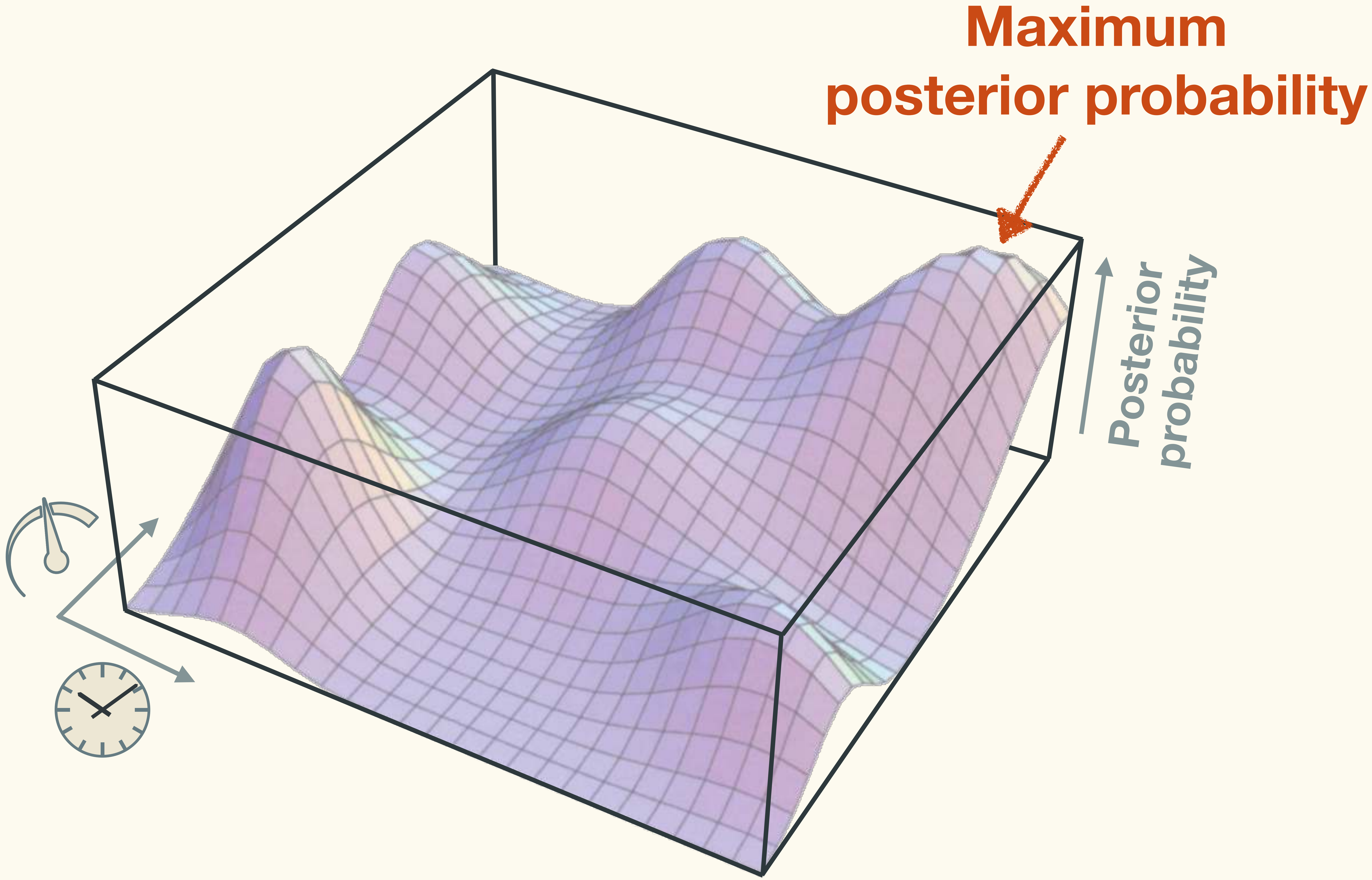
MCMC

Markov-chain Monte Carlo



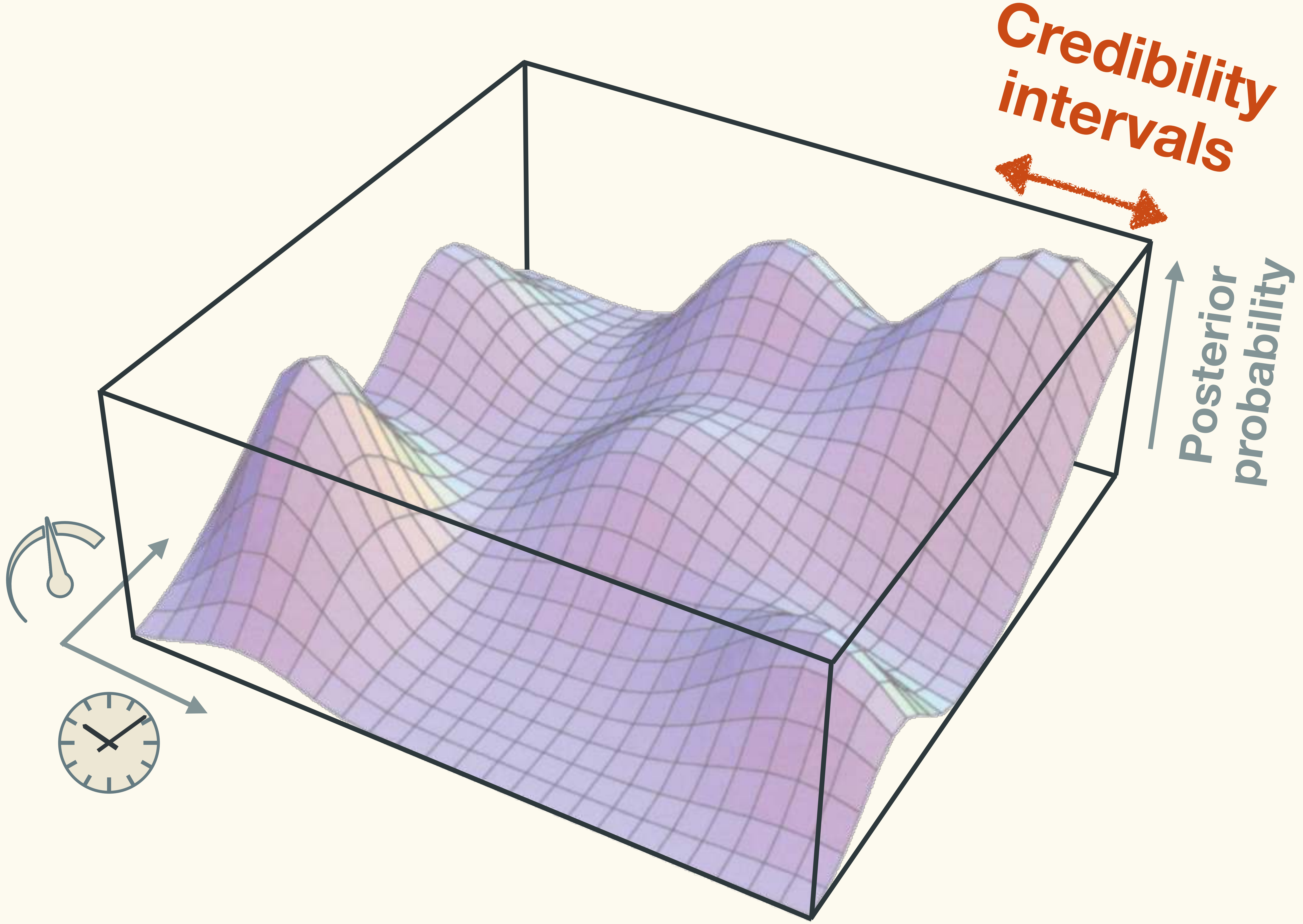
MCMC

Markov-chain Monte Carlo



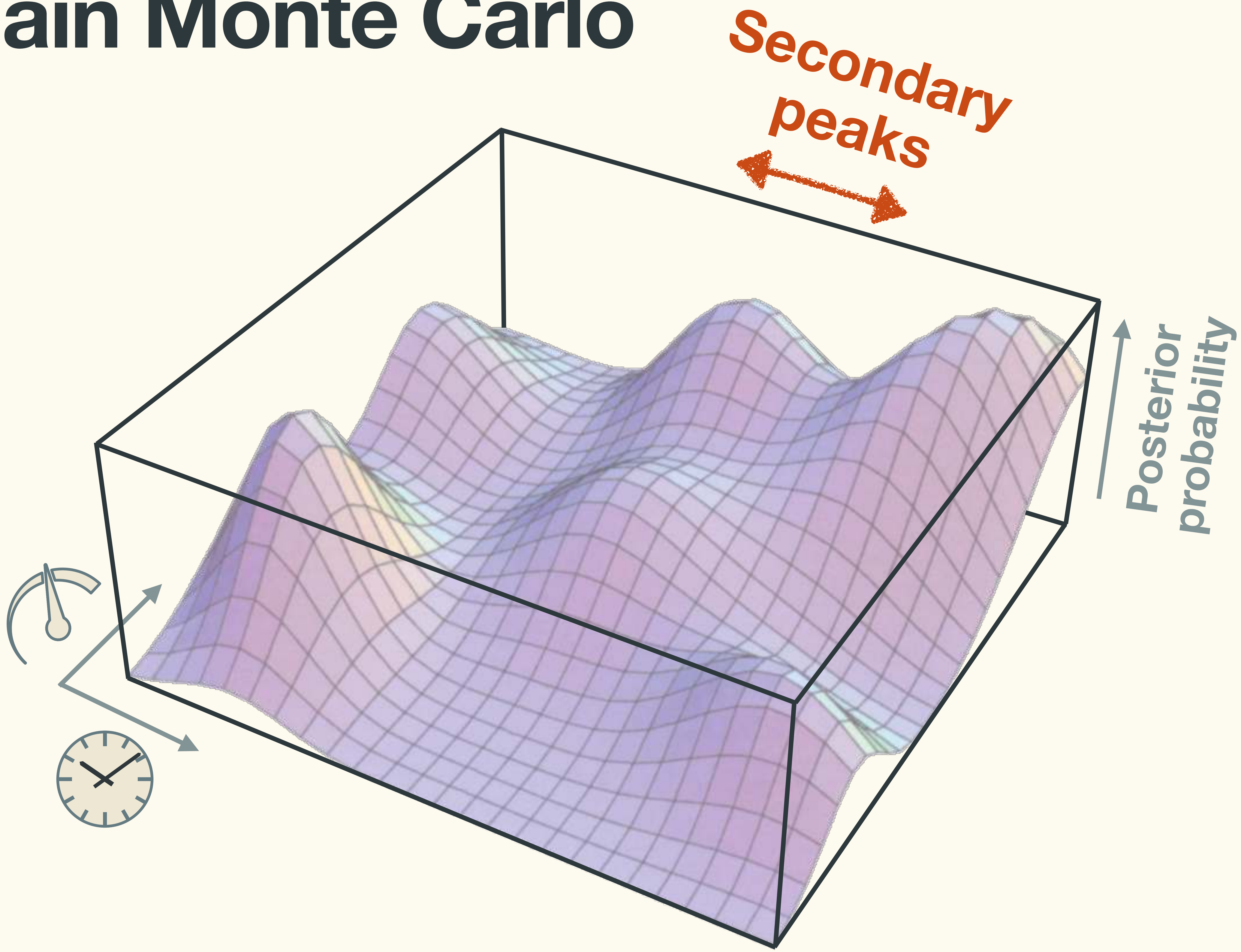
MCMC

Markov-chain Monte Carlo



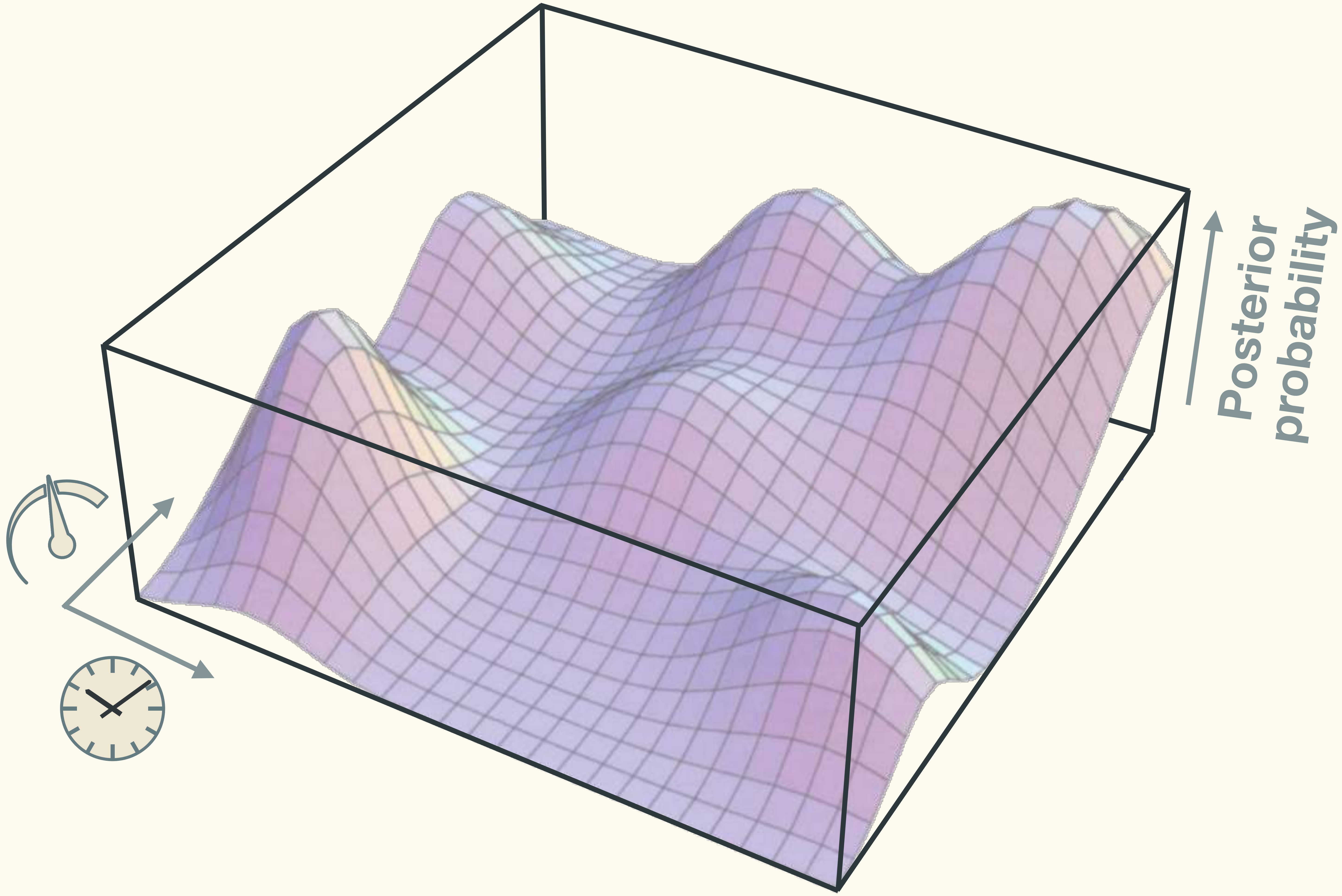
MCMC

Markov-chain Monte Carlo



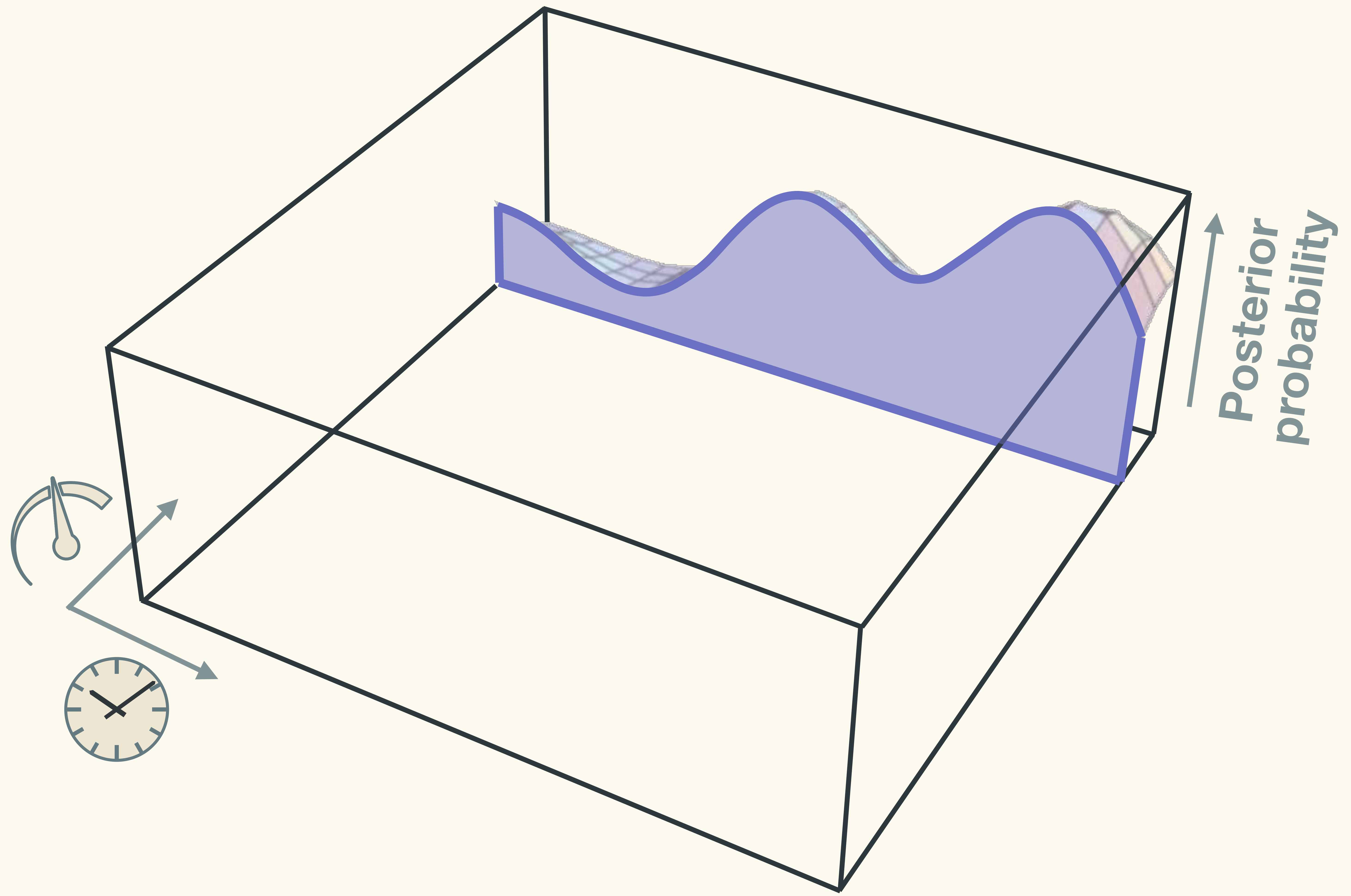
MCMC

Markov-chain Monte Carlo



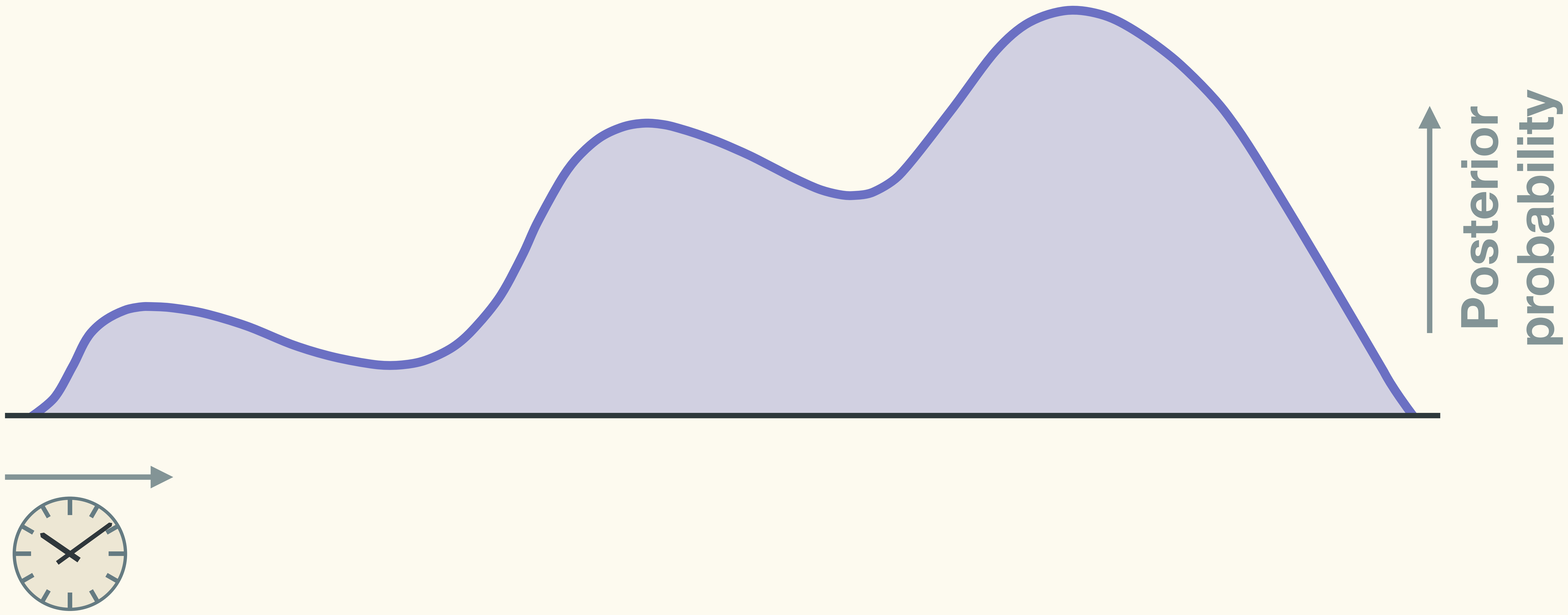
MCMC

Markov-chain Monte Carlo

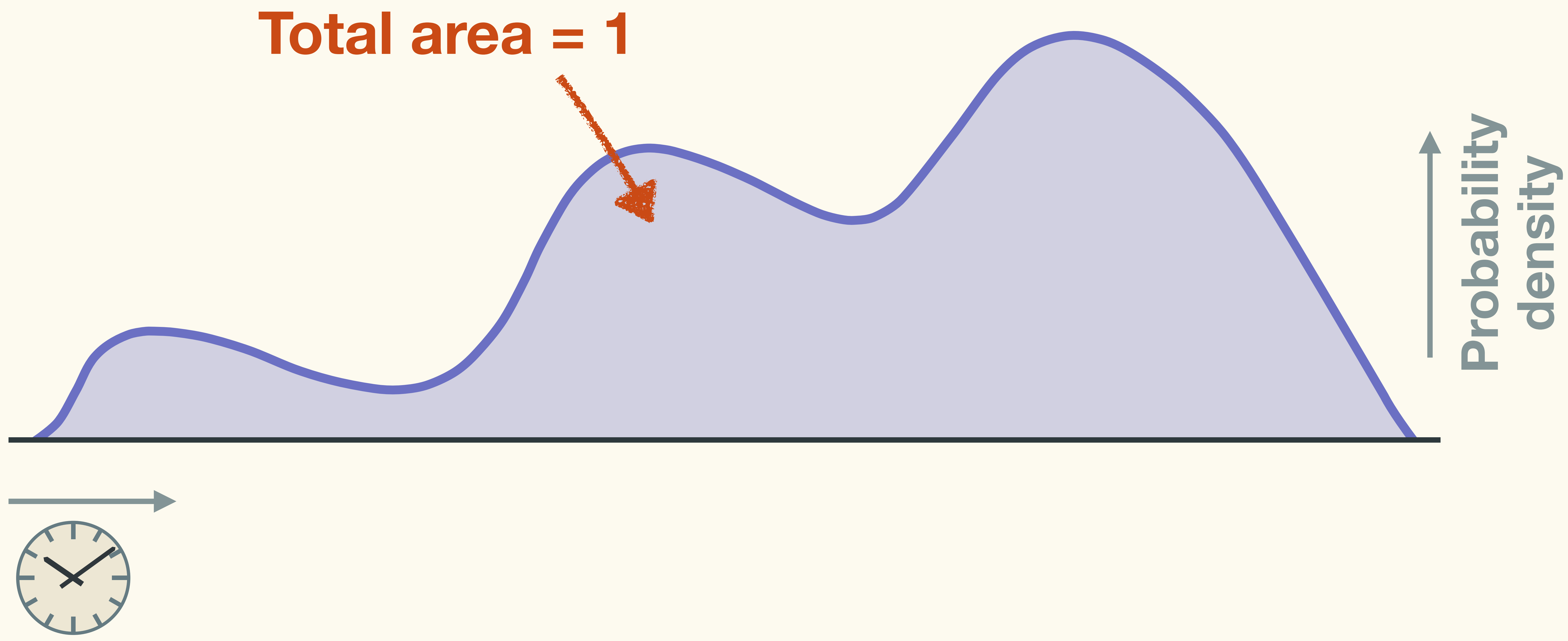


MCMC

Markov-chain Monte Carlo

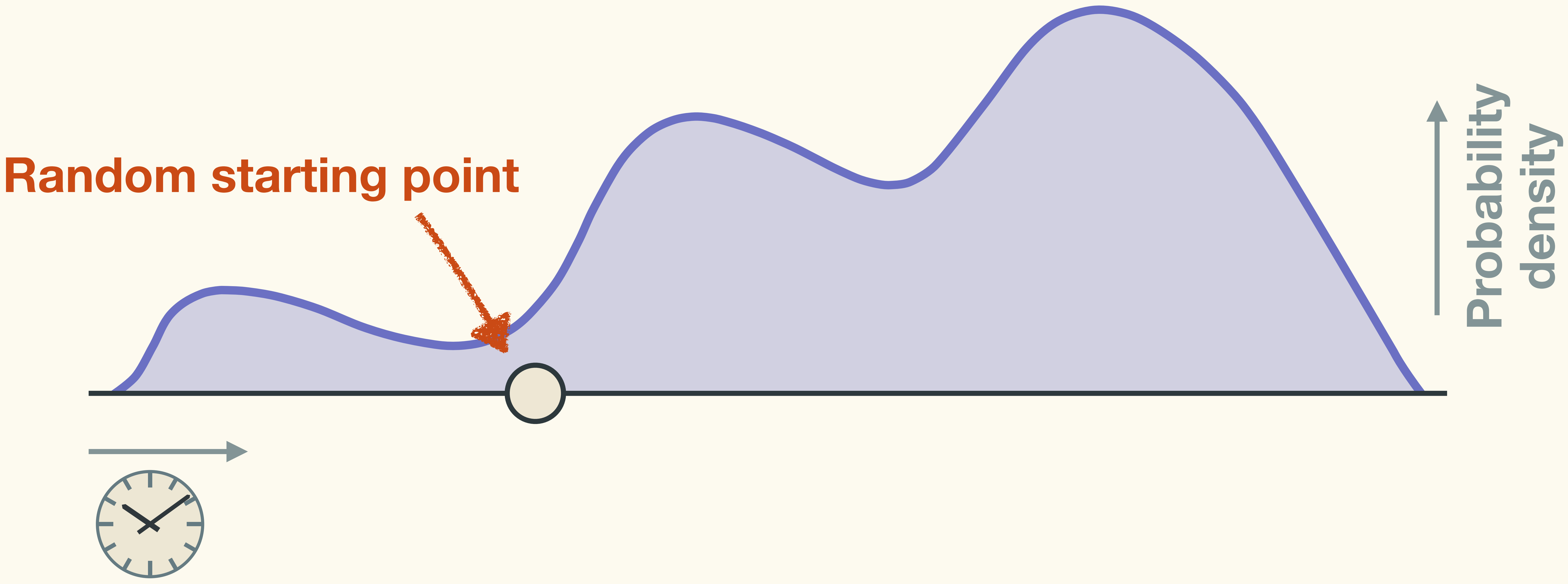


MCMC
Markov-chain Monte Carlo



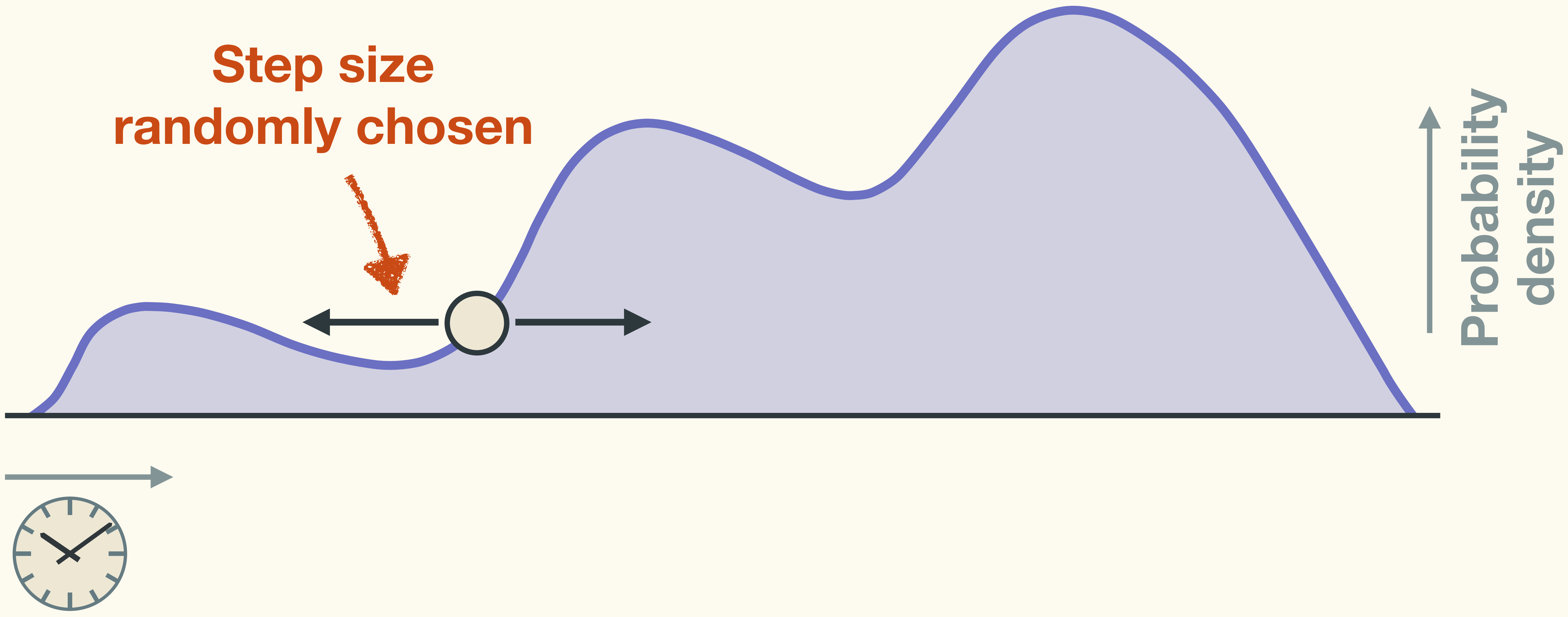
MCMC

Markov-chain Monte Carlo



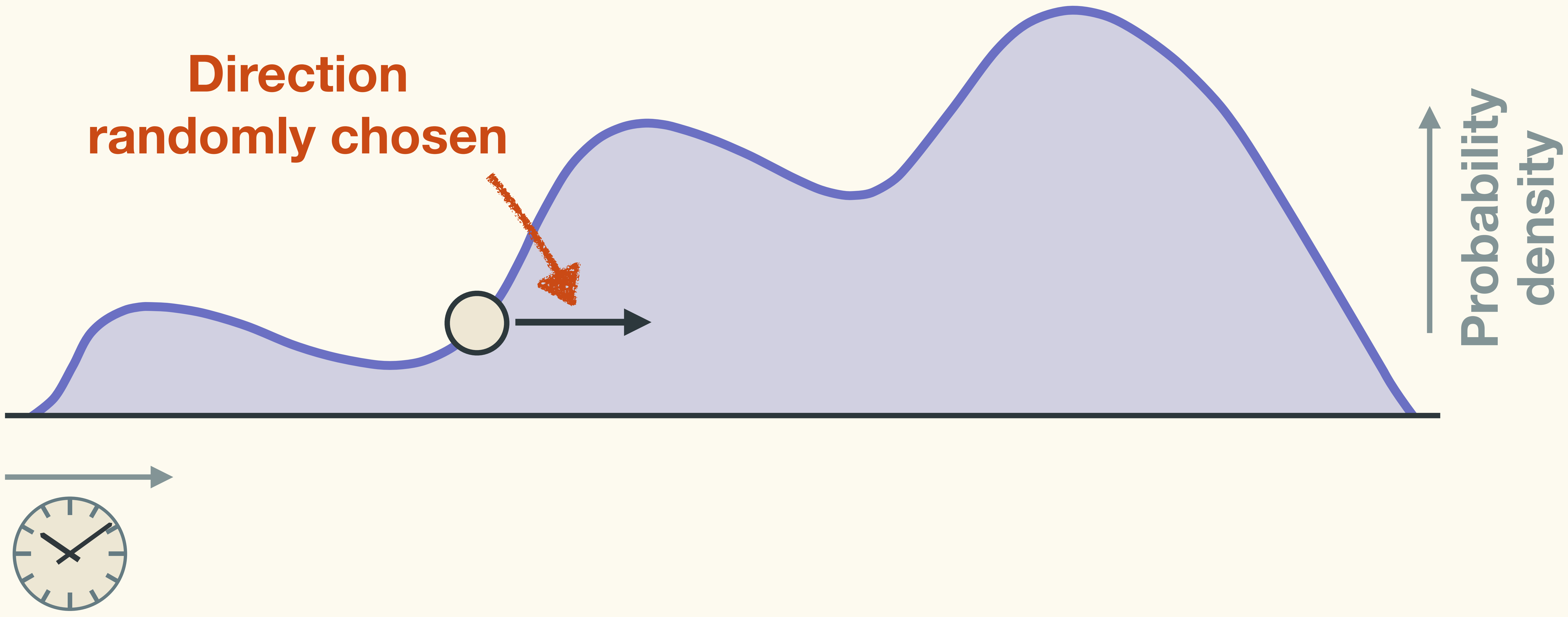
MCMC

Markov-chain Monte Carlo



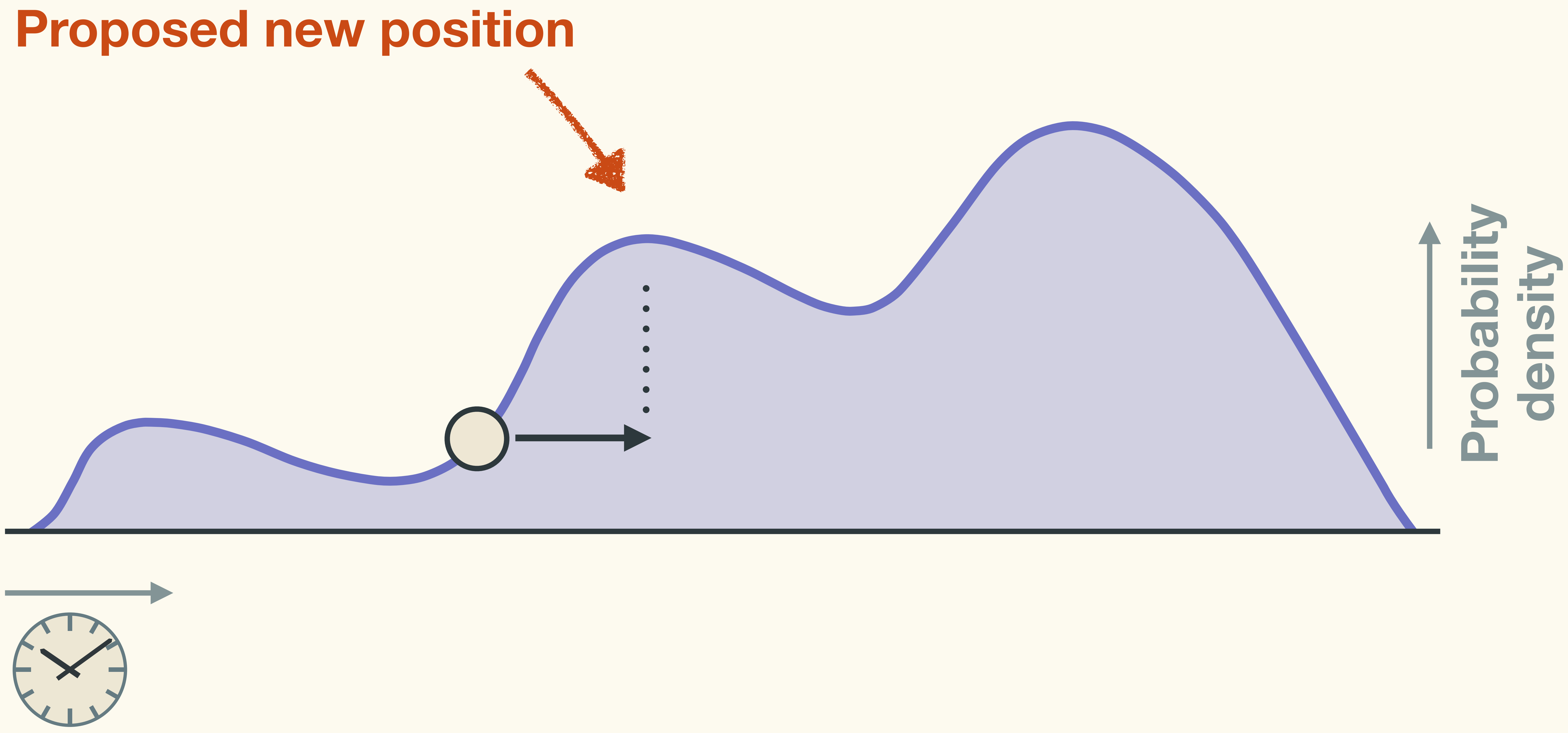
MCMC

Markov-chain Monte Carlo



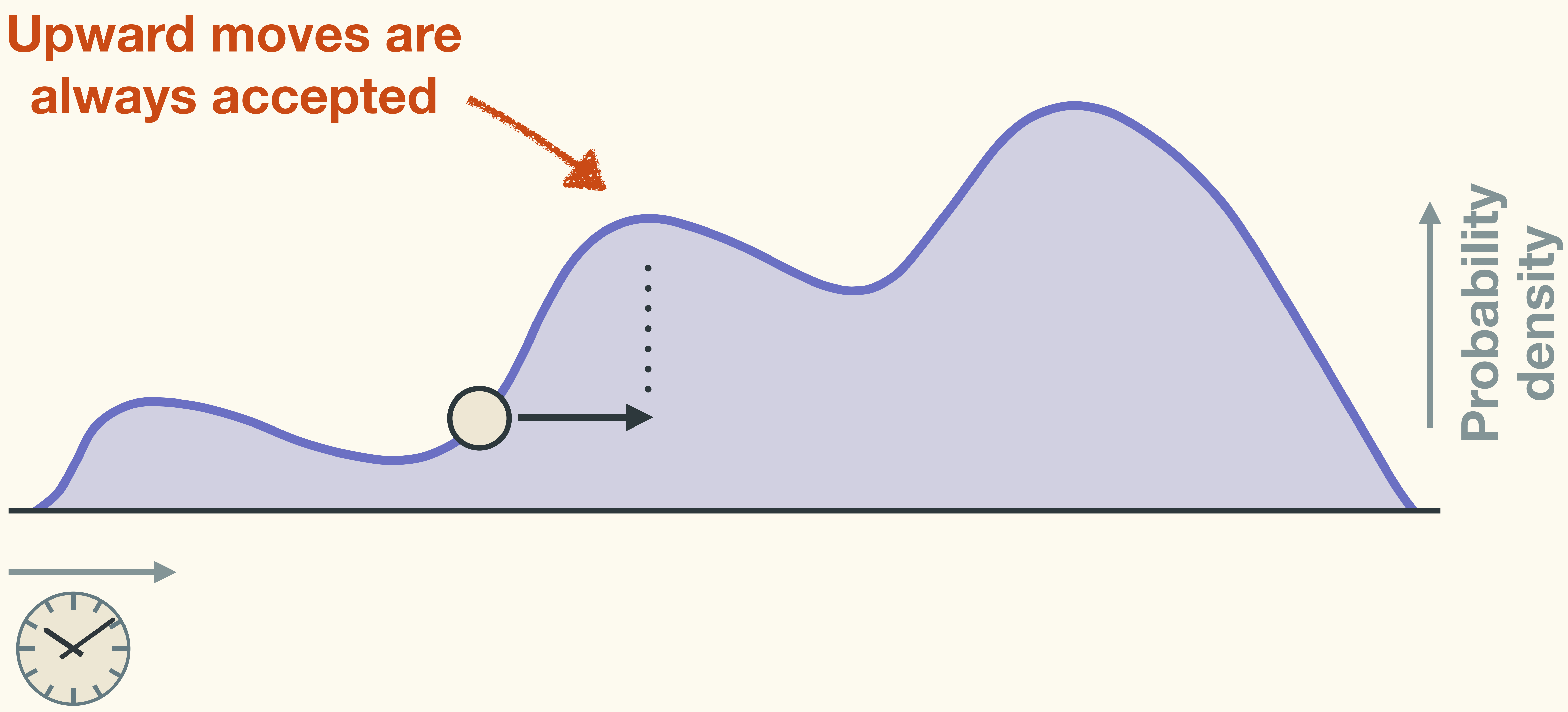
MCMC

Markov-chain Monte Carlo

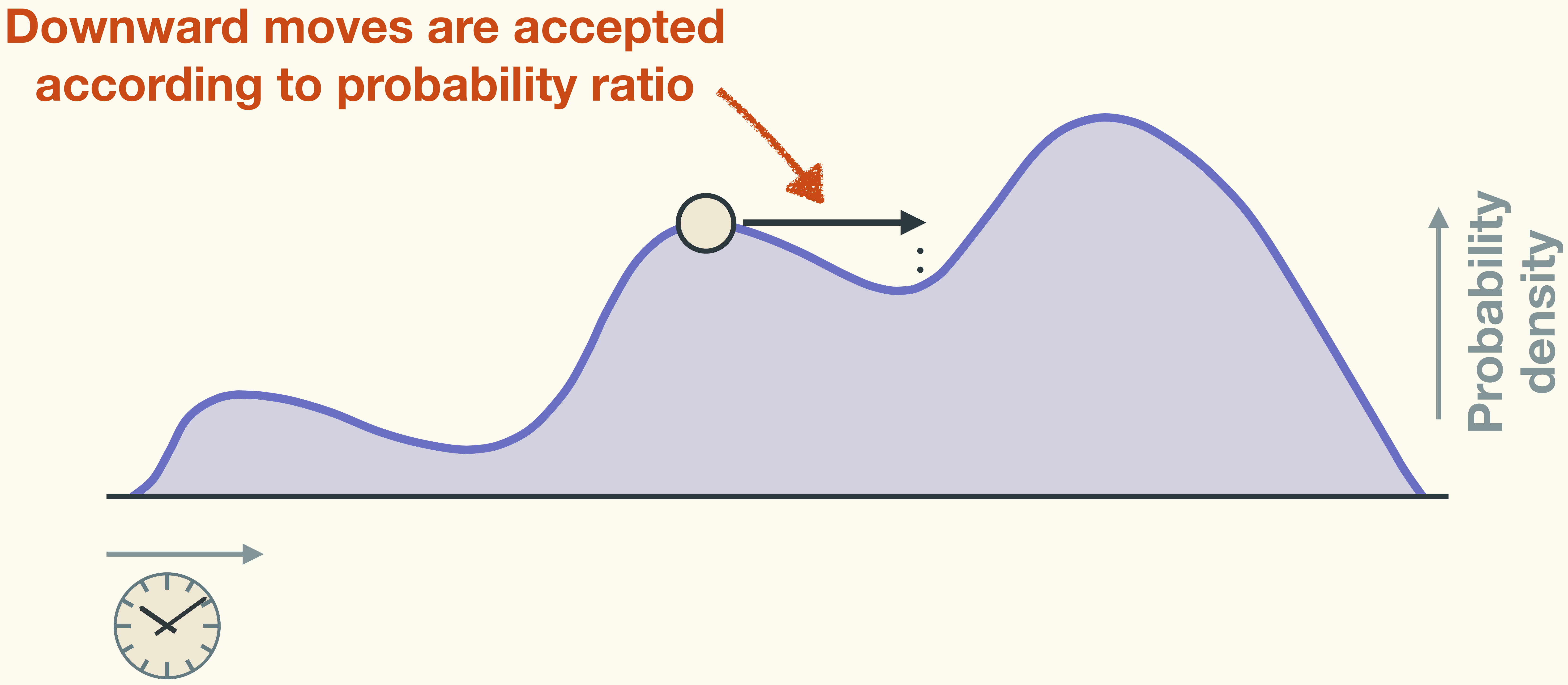


MCMC

Markov-chain Monte Carlo

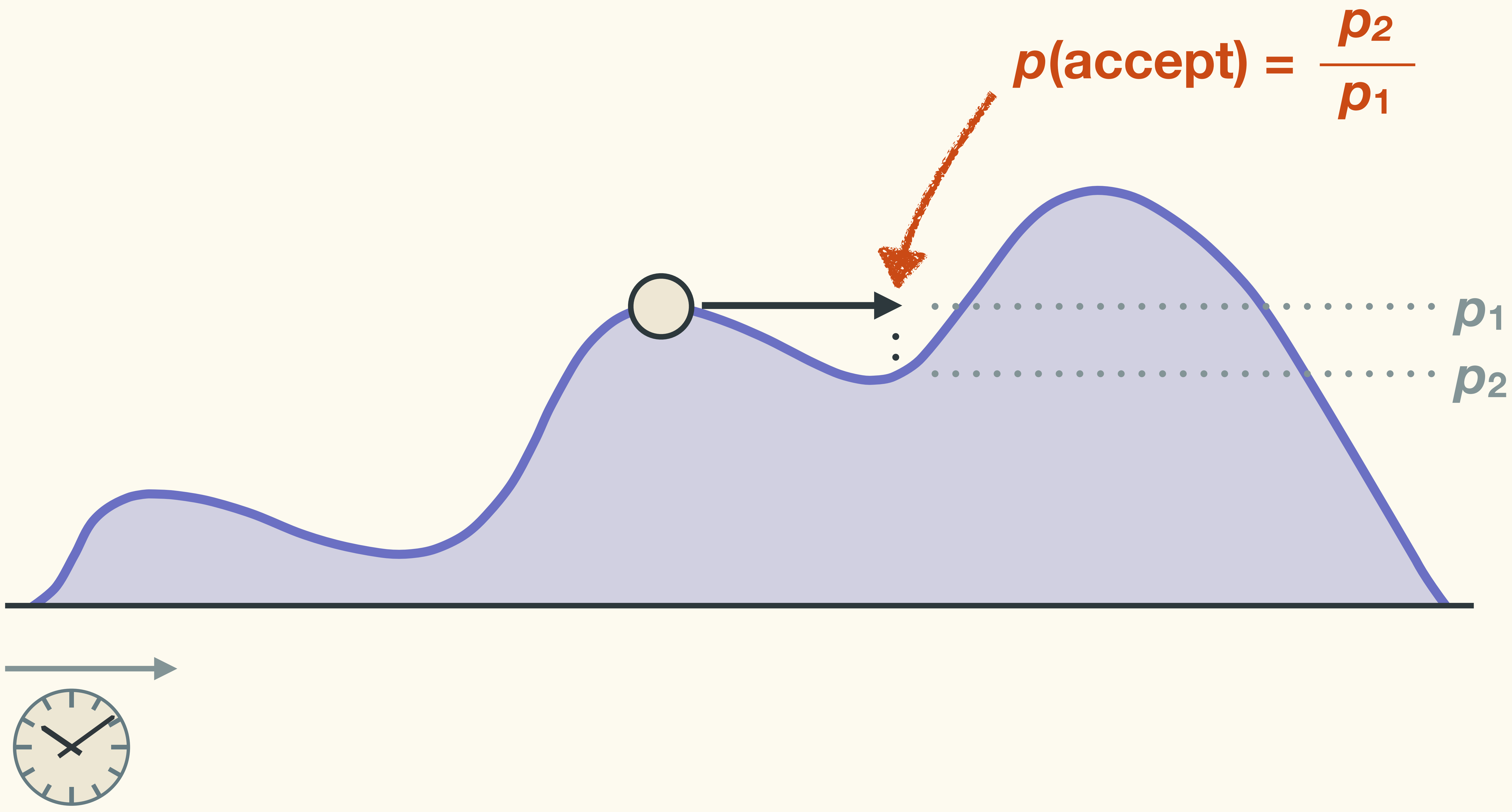


MCMC
Markov-chain Monte Carlo

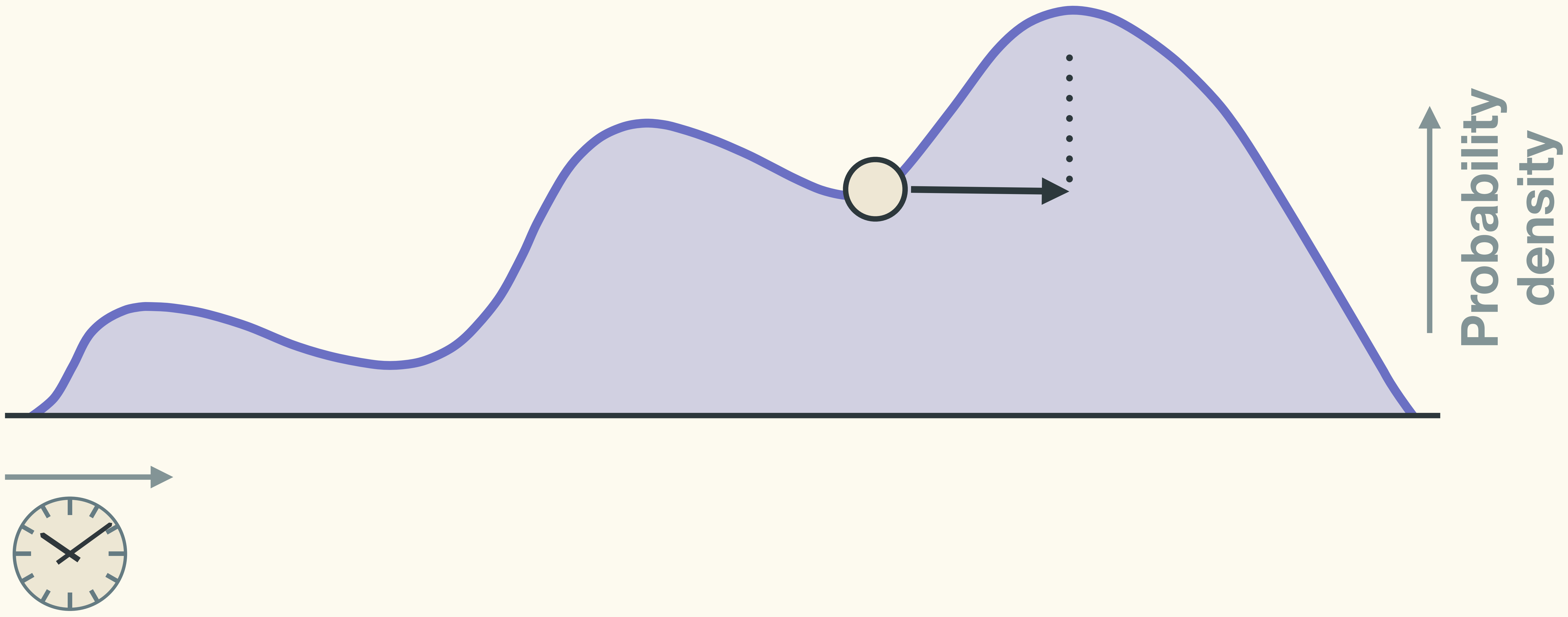


MCMC

Markov-chain Monte Carlo

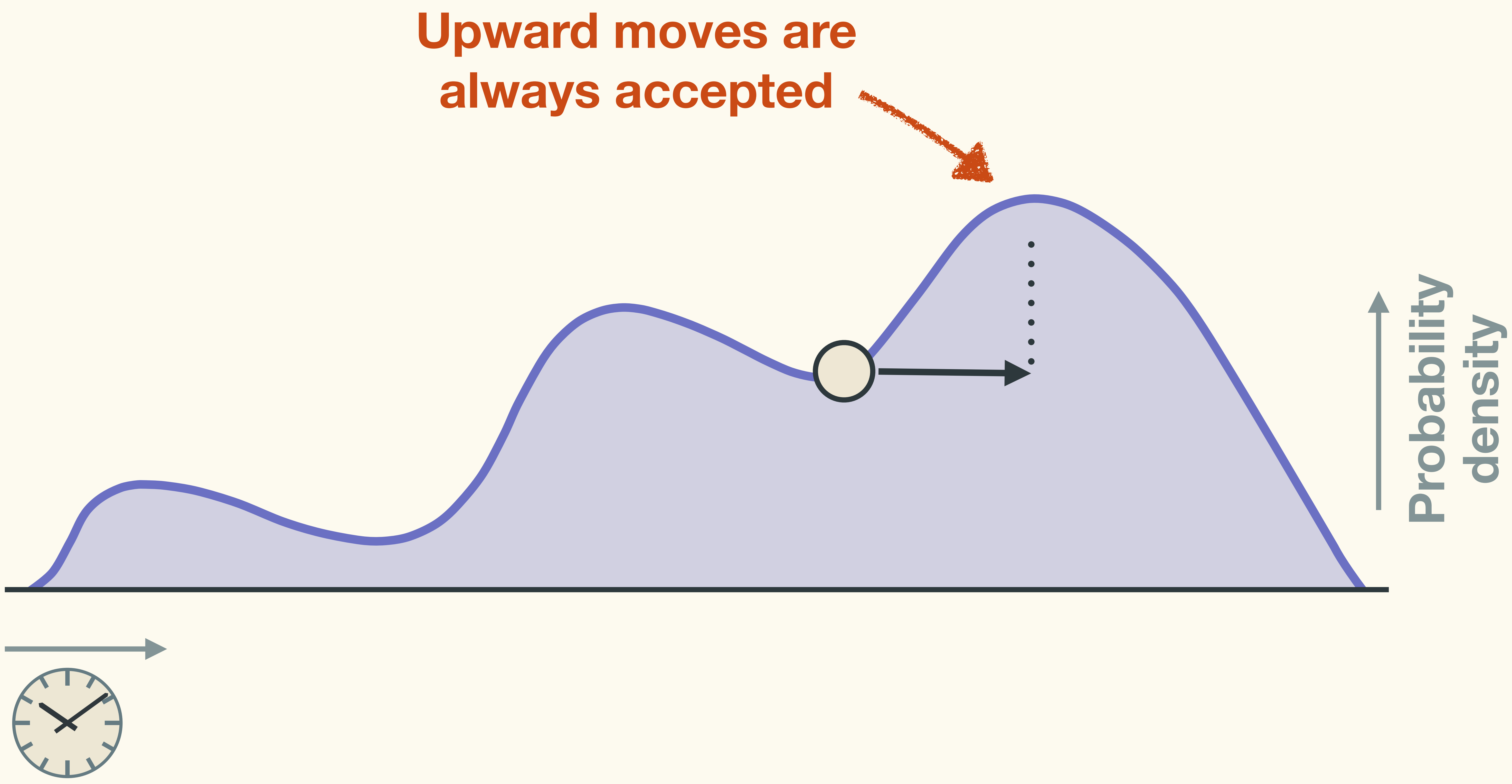


MCMC
Markov-chain Monte Carlo



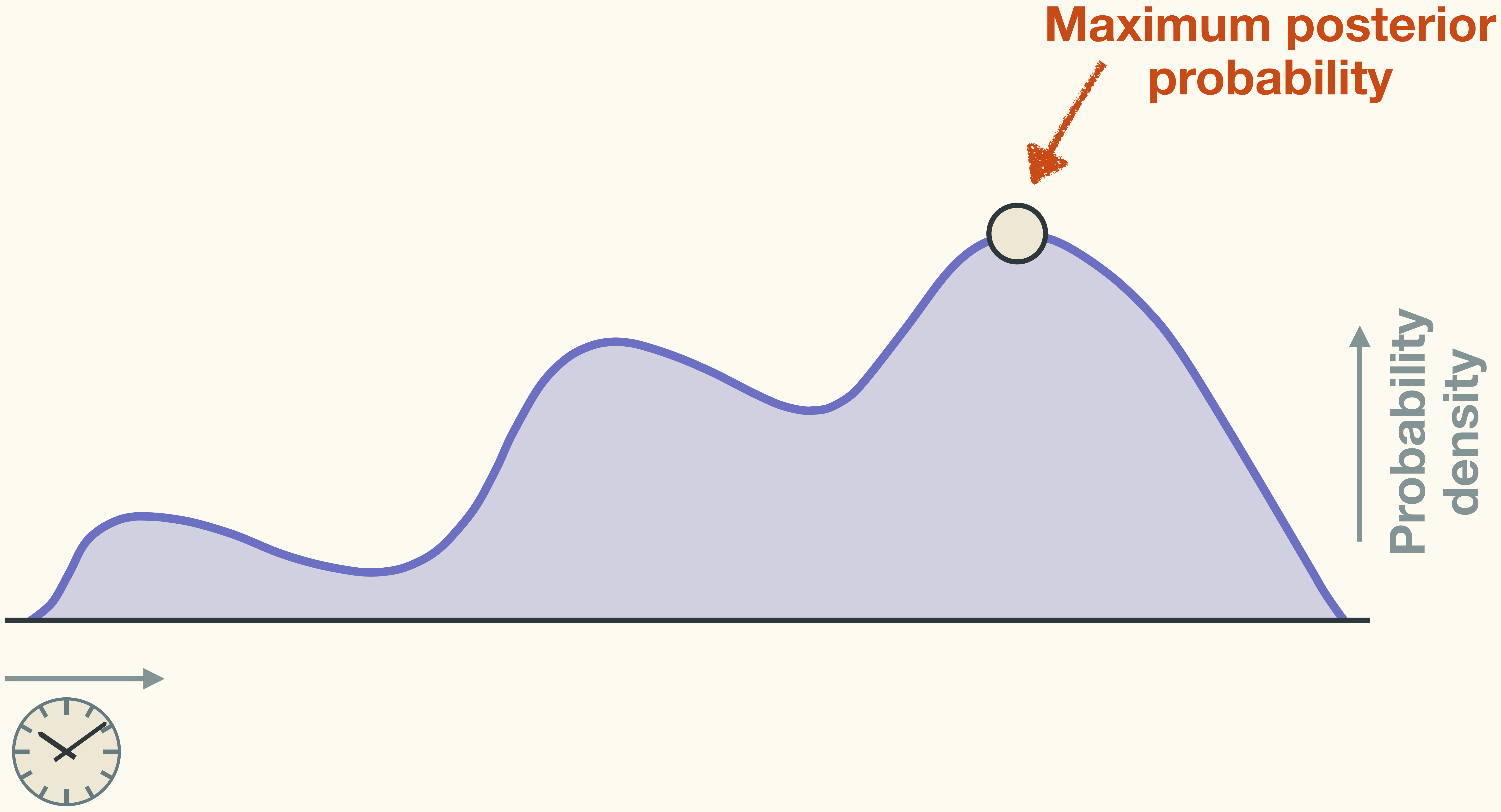
MCMC

Markov-chain Monte Carlo



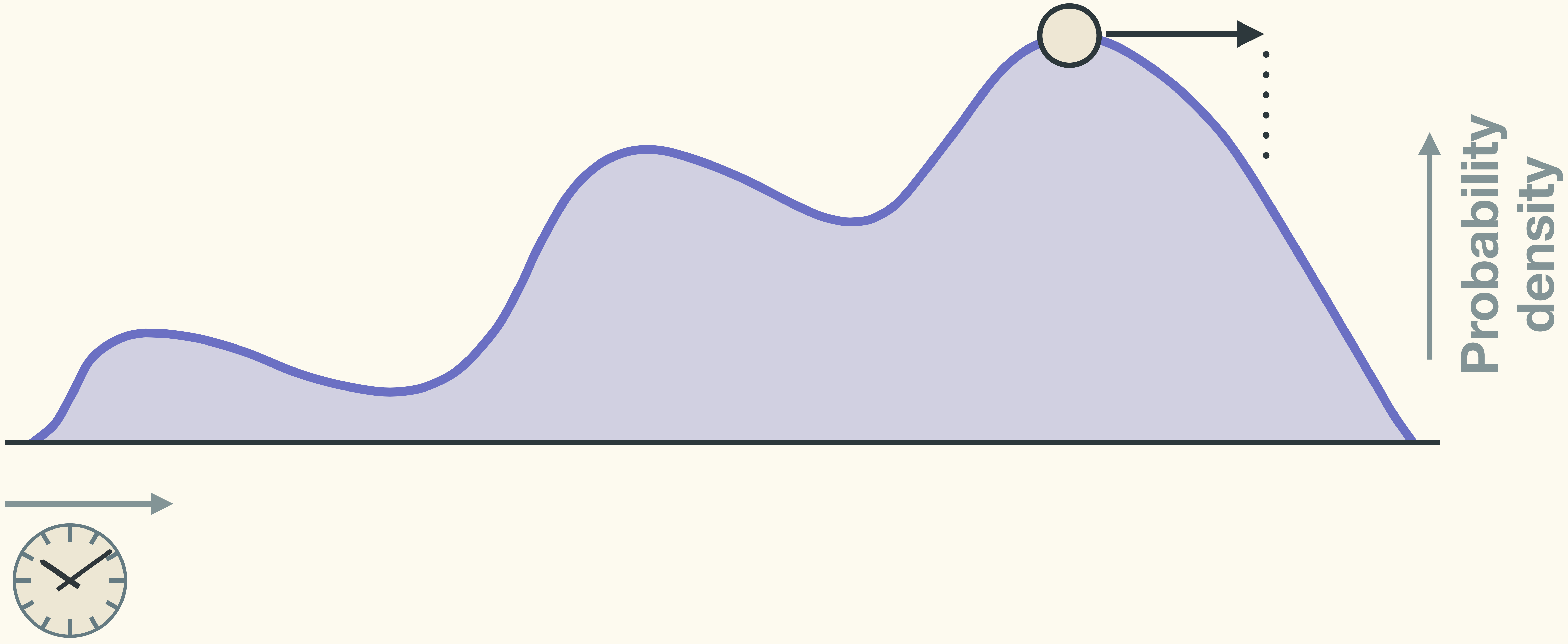
MCMC

Markov-chain Monte Carlo



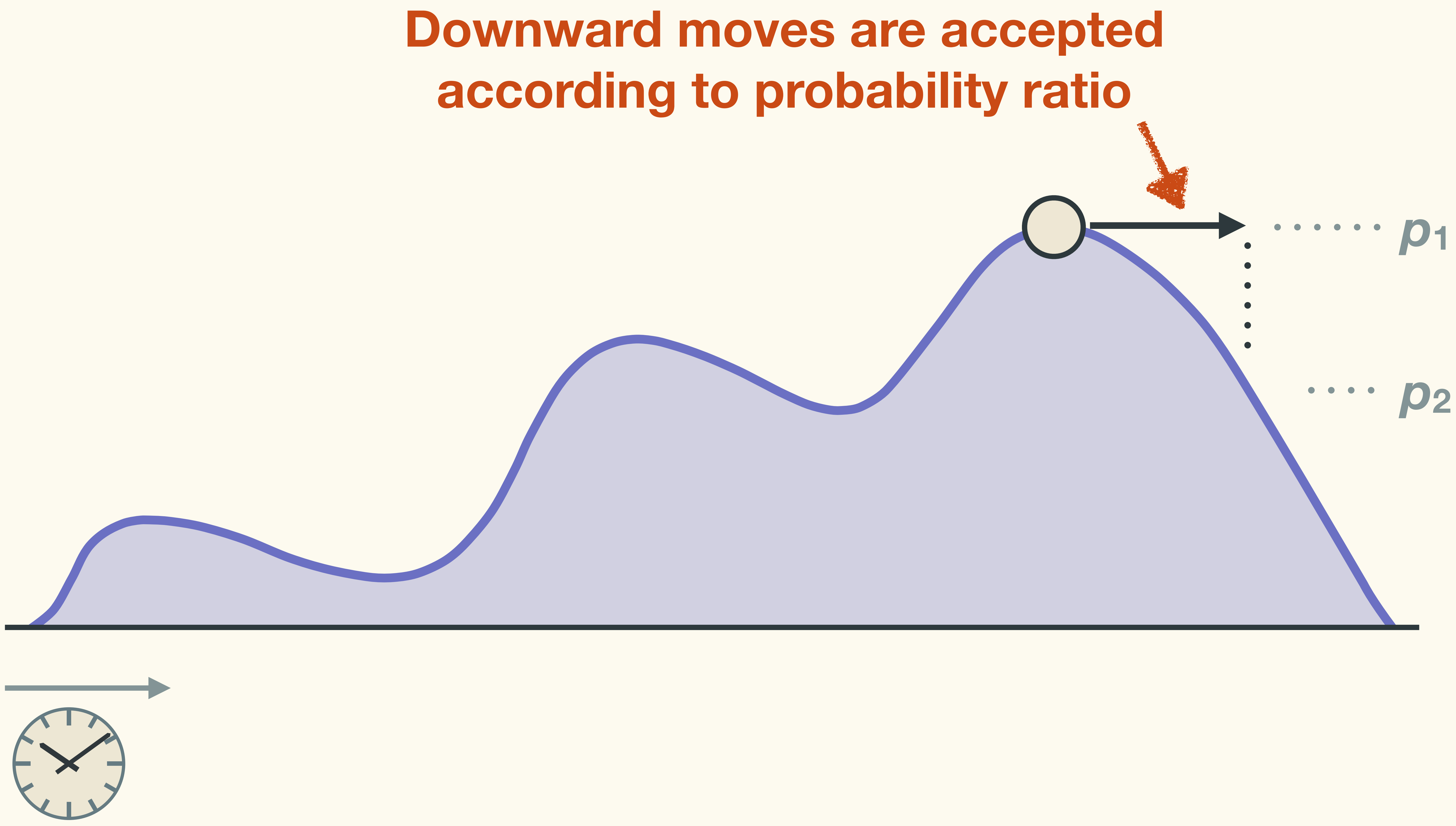
MCMC

Markov-chain Monte Carlo



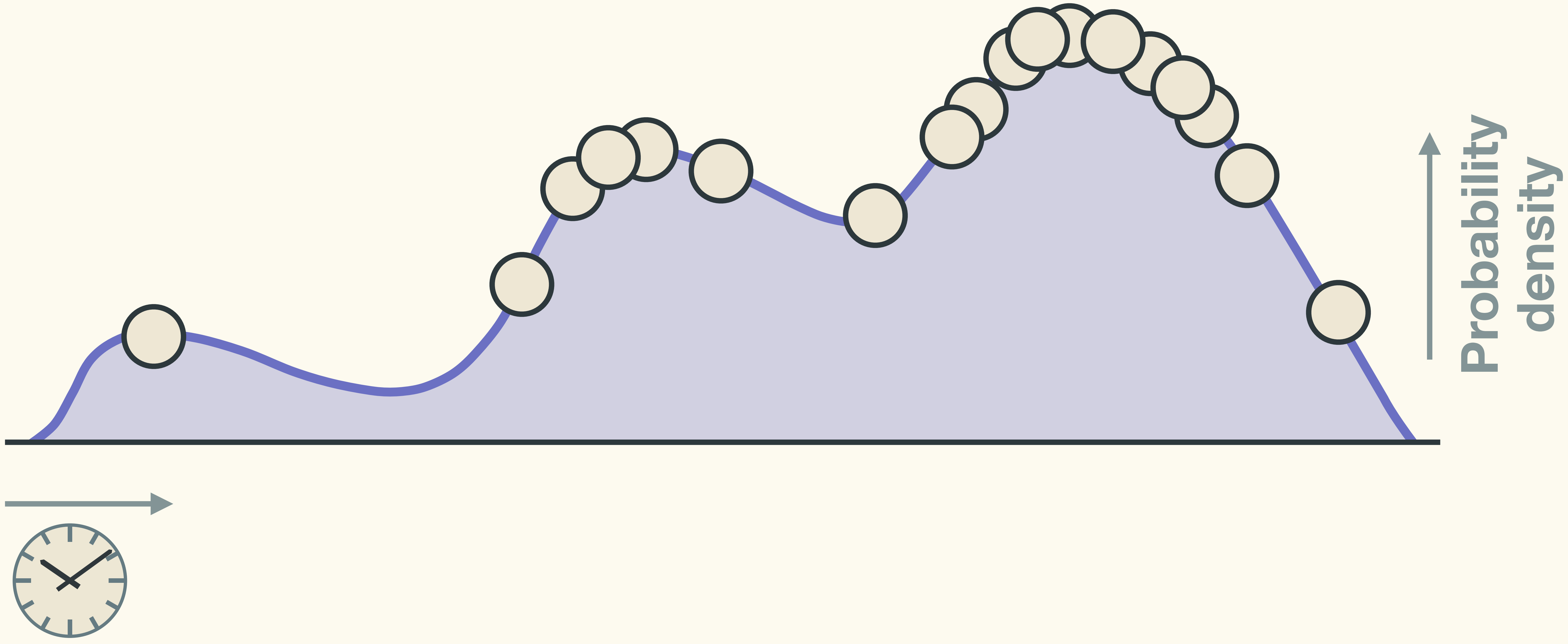
MCMC

Markov-chain Monte Carlo



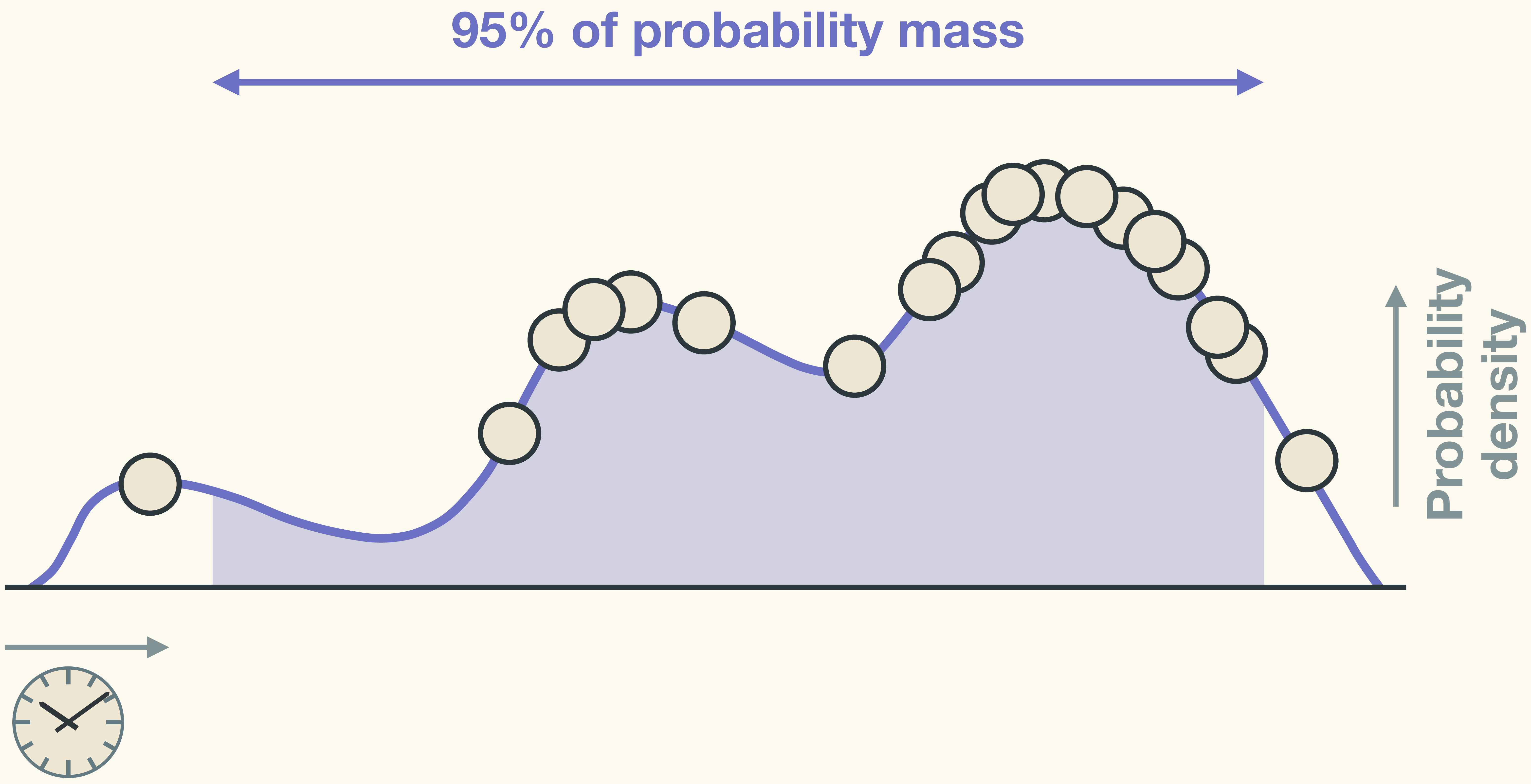
MCMC

Markov-chain Monte Carlo



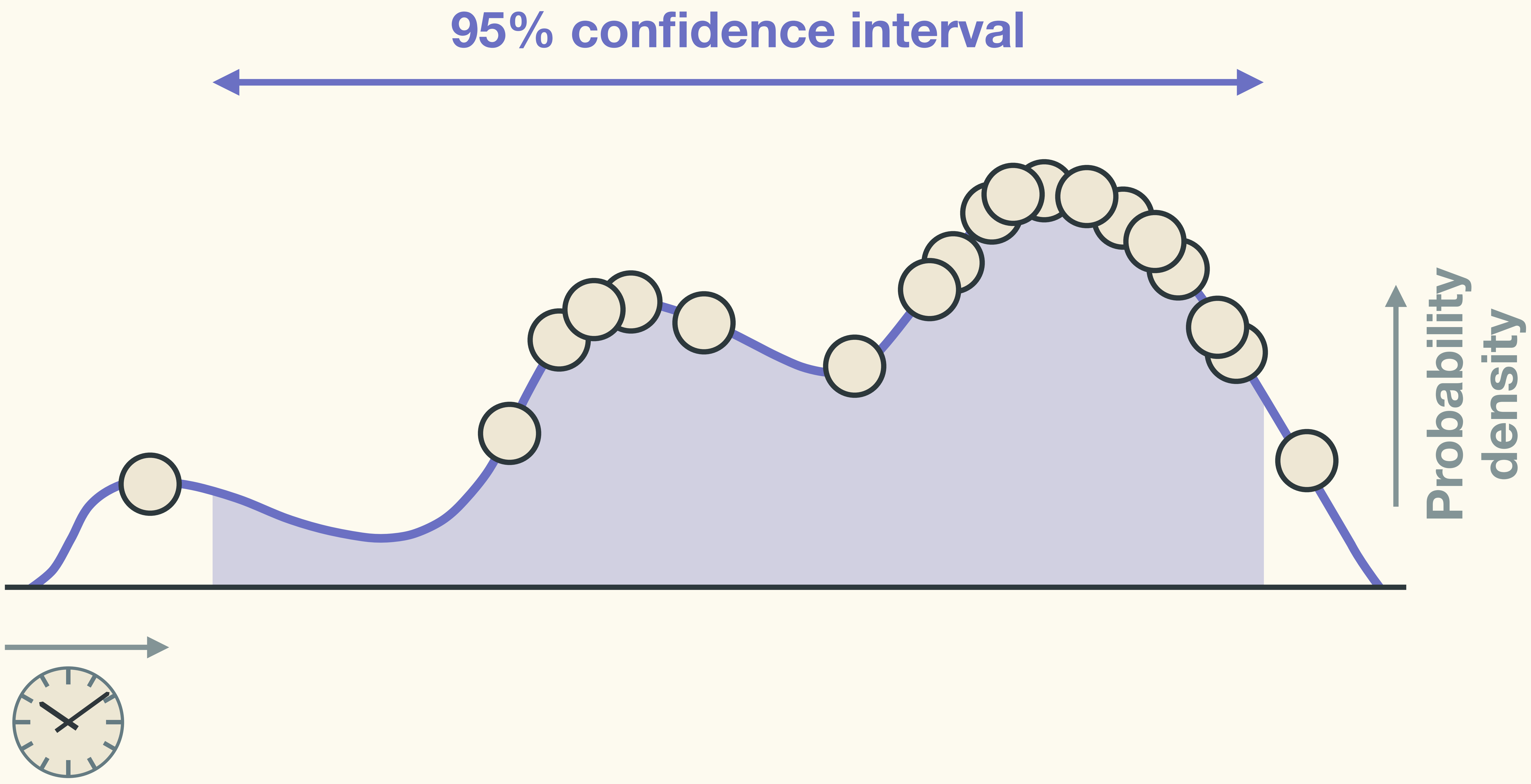
MCMC

Markov-chain Monte Carlo

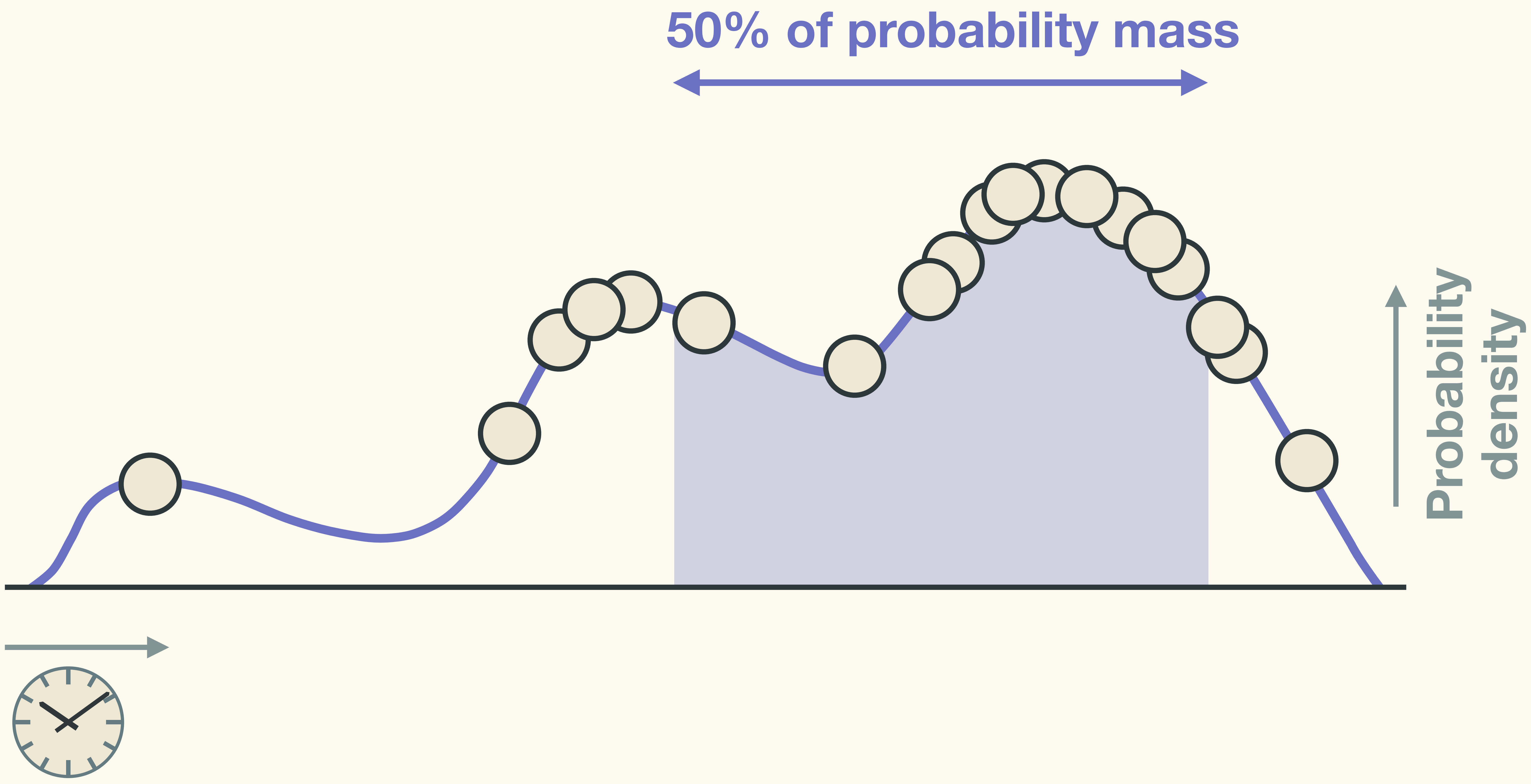


MCMC

Markov-chain Monte Carlo

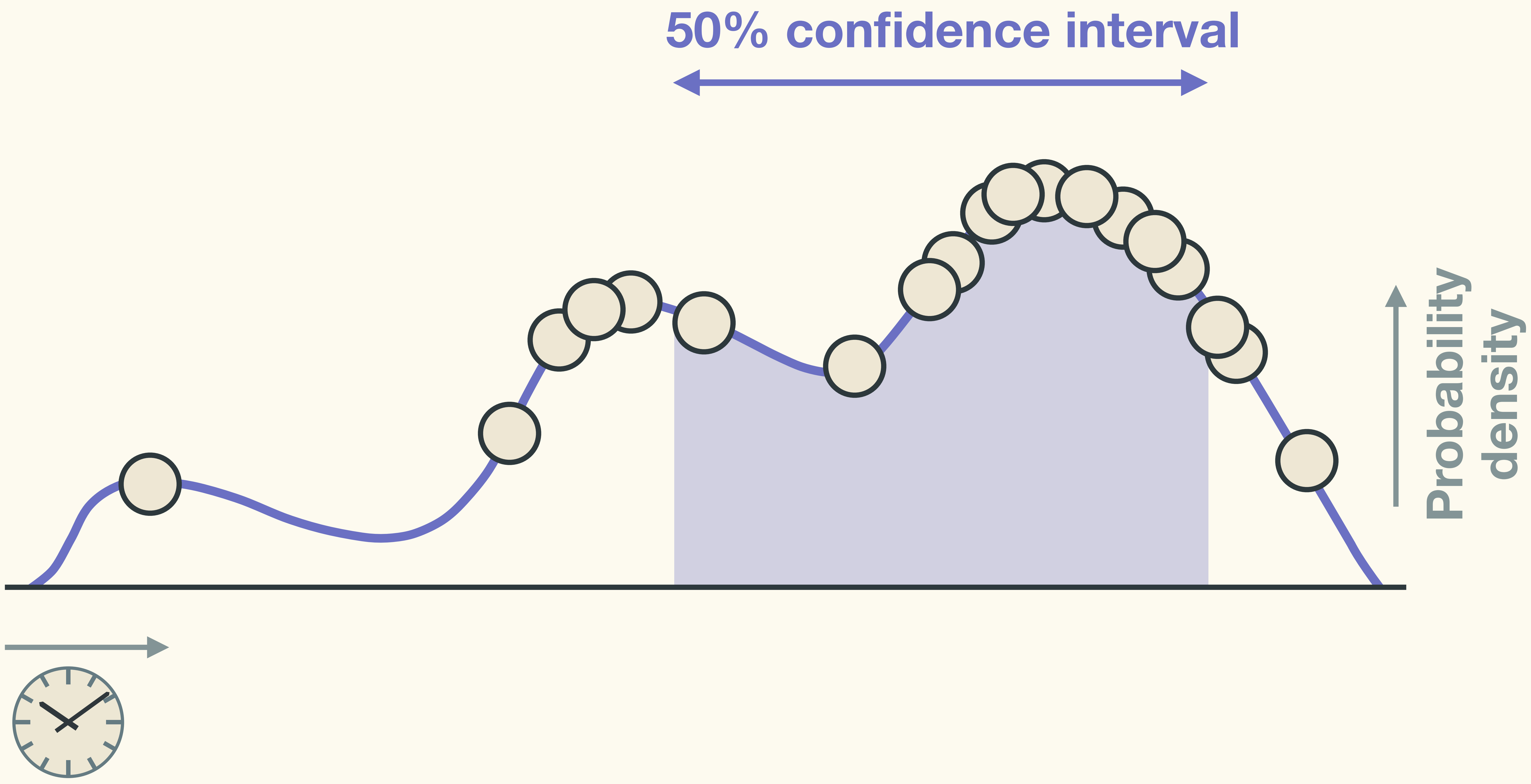


MCMC
Markov-chain Monte Carlo



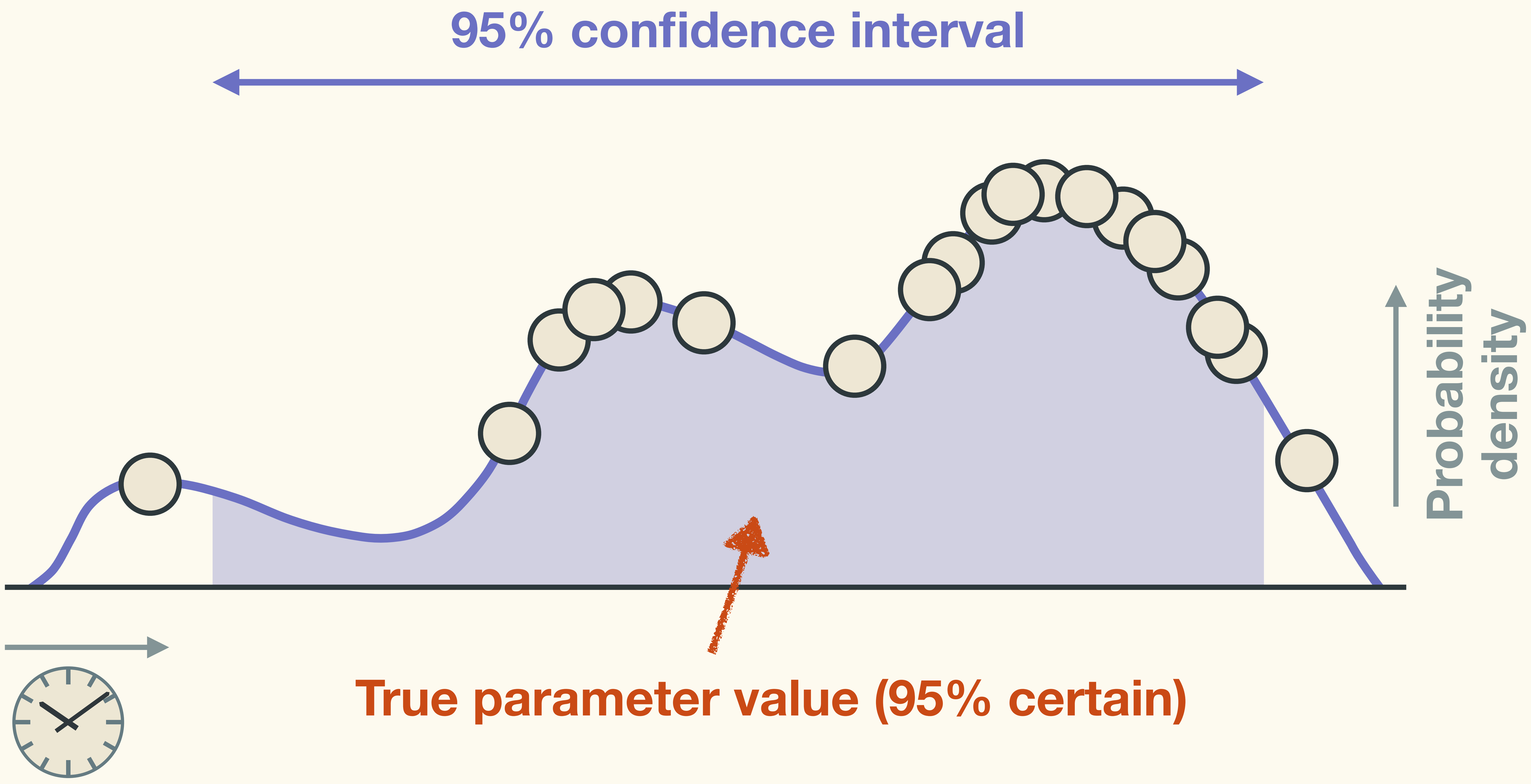
MCMC

Markov-chain Monte Carlo



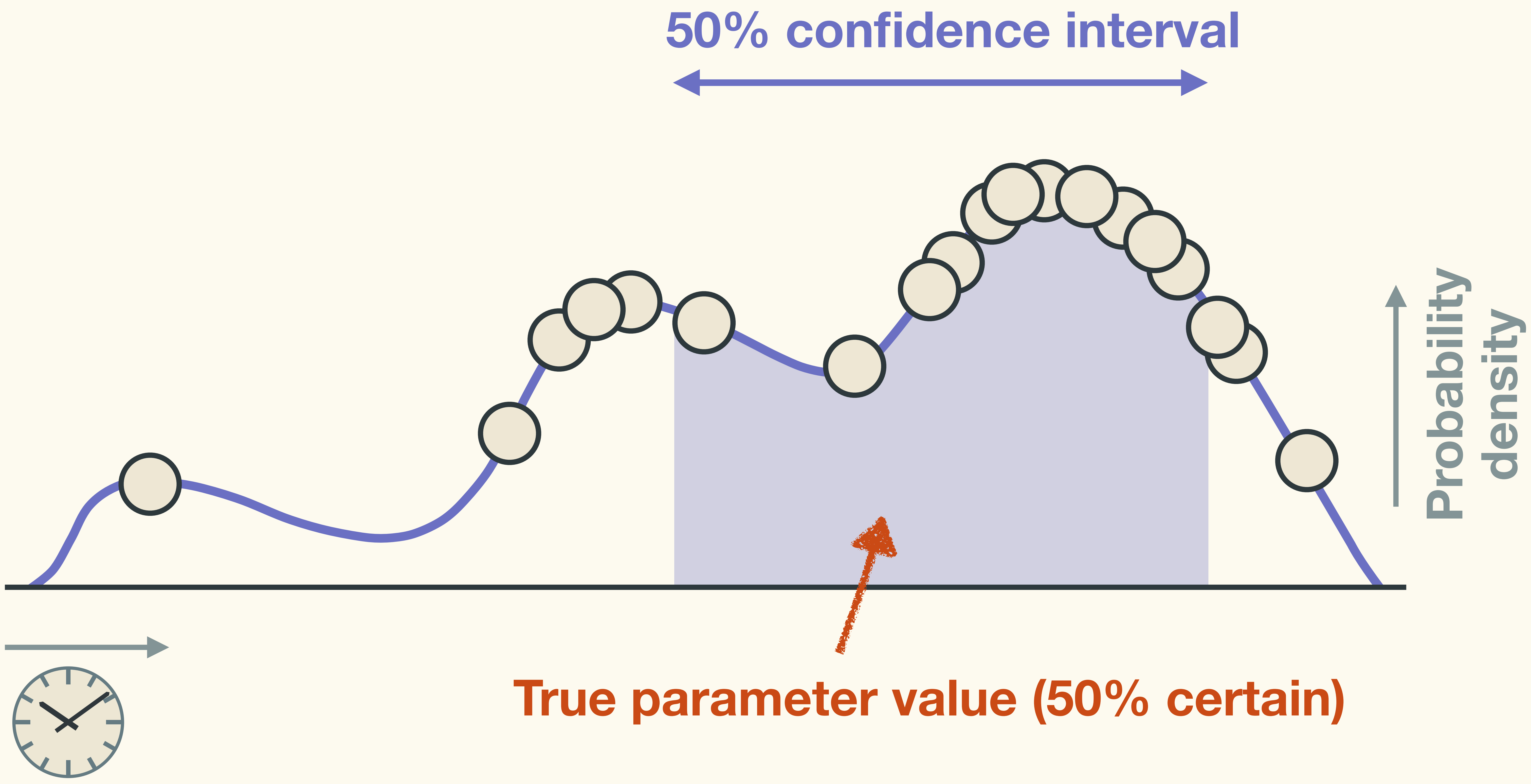
MCMC

Markov-chain Monte Carlo



MCMC

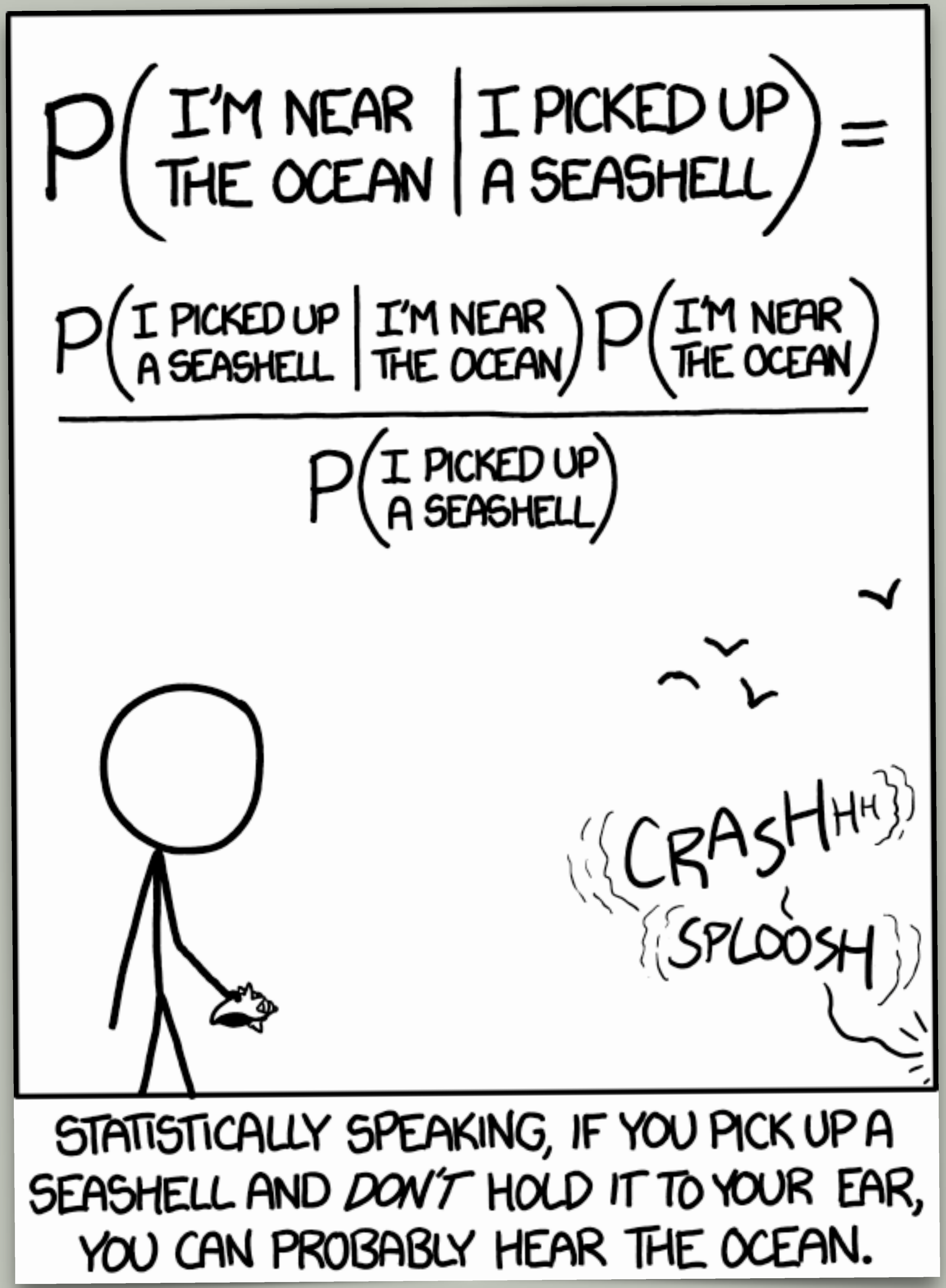
Markov-chain Monte Carlo



MCMC Robot

Thanks

Thanks


P(I'M NEAR THE OCEAN | I PICKED UP A SEASHELL) =
P(I PICKED UP A SEASHELL | I'M NEAR THE OCEAN) P(I'M NEAR THE OCEAN)
P(I PICKED UP A SEASHELL)
CRASHHH
SPLOOSH
STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND DON'T HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.