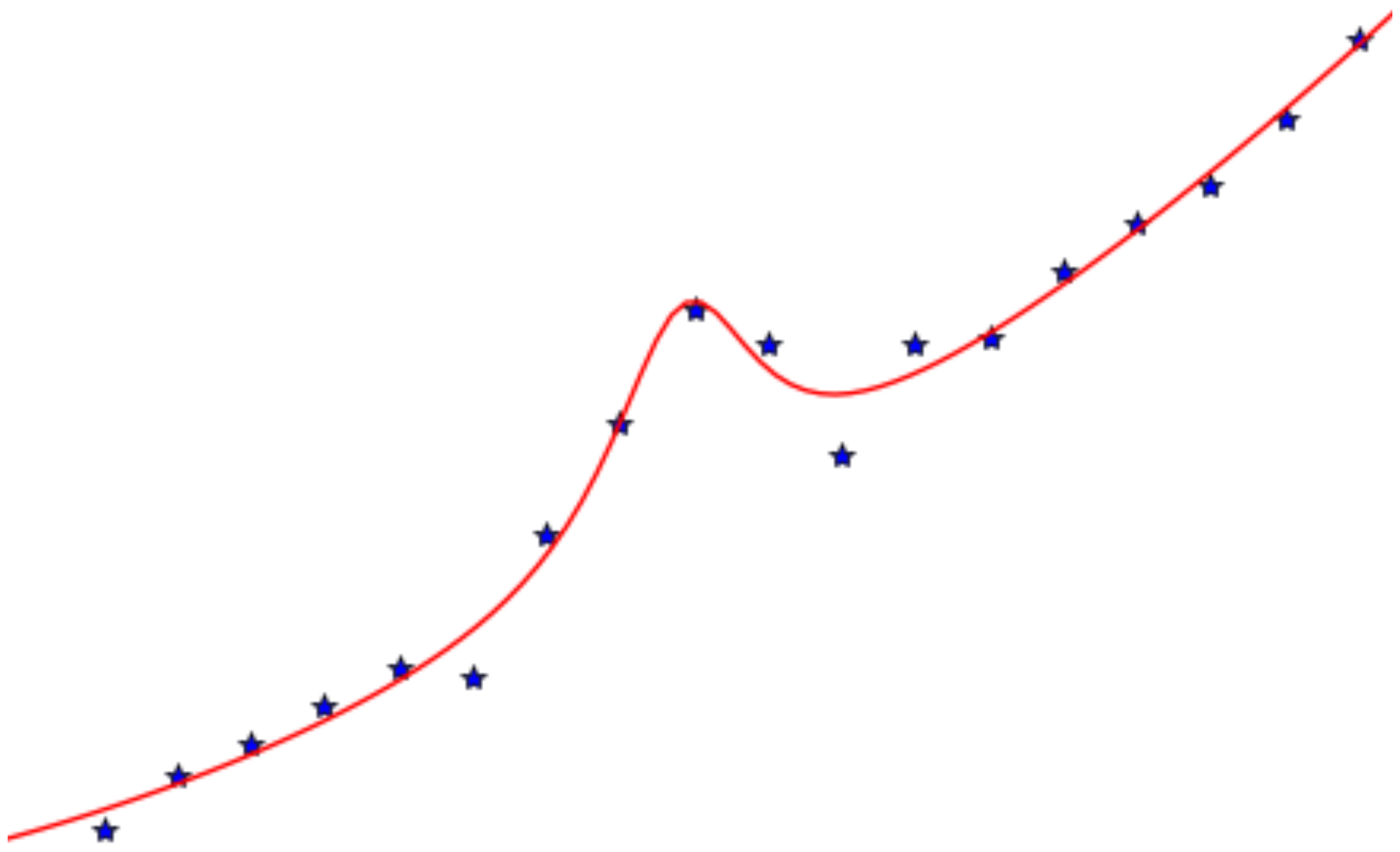
Bayesian Data Analysis I



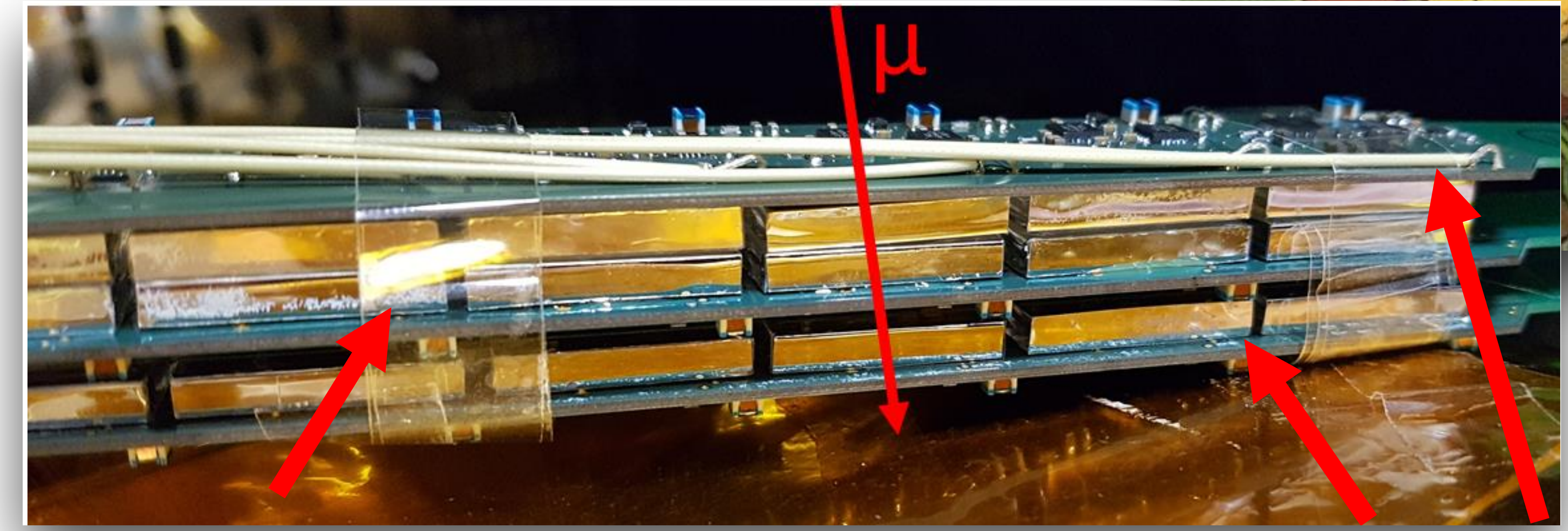
Inverted CERN School of Computing
CERN - 06.03.18

Christian Graf
Max Planck Institute for Physics, Munich

Learning Goals

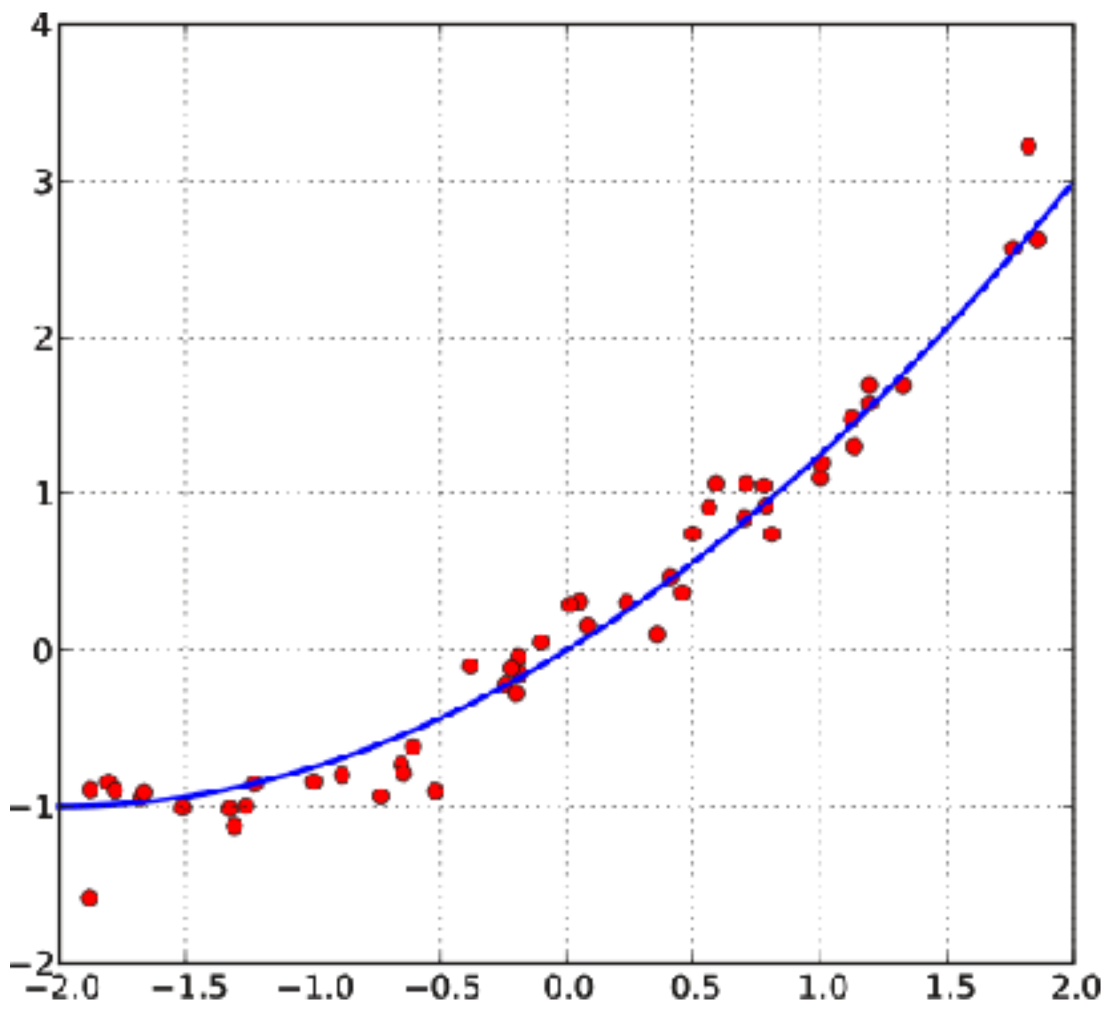
- The difference between **frequentist** and **bayesian** approach to data analysis
- How to set up a probability model and perform a simple **bayesian analysis**
- The role of the **prior** and how to choose a reasonable one
- Different ways to **present** your results
- **2nd Lecture:** Hypothesis testing, model fitting, other applications

[Andrew Walter]

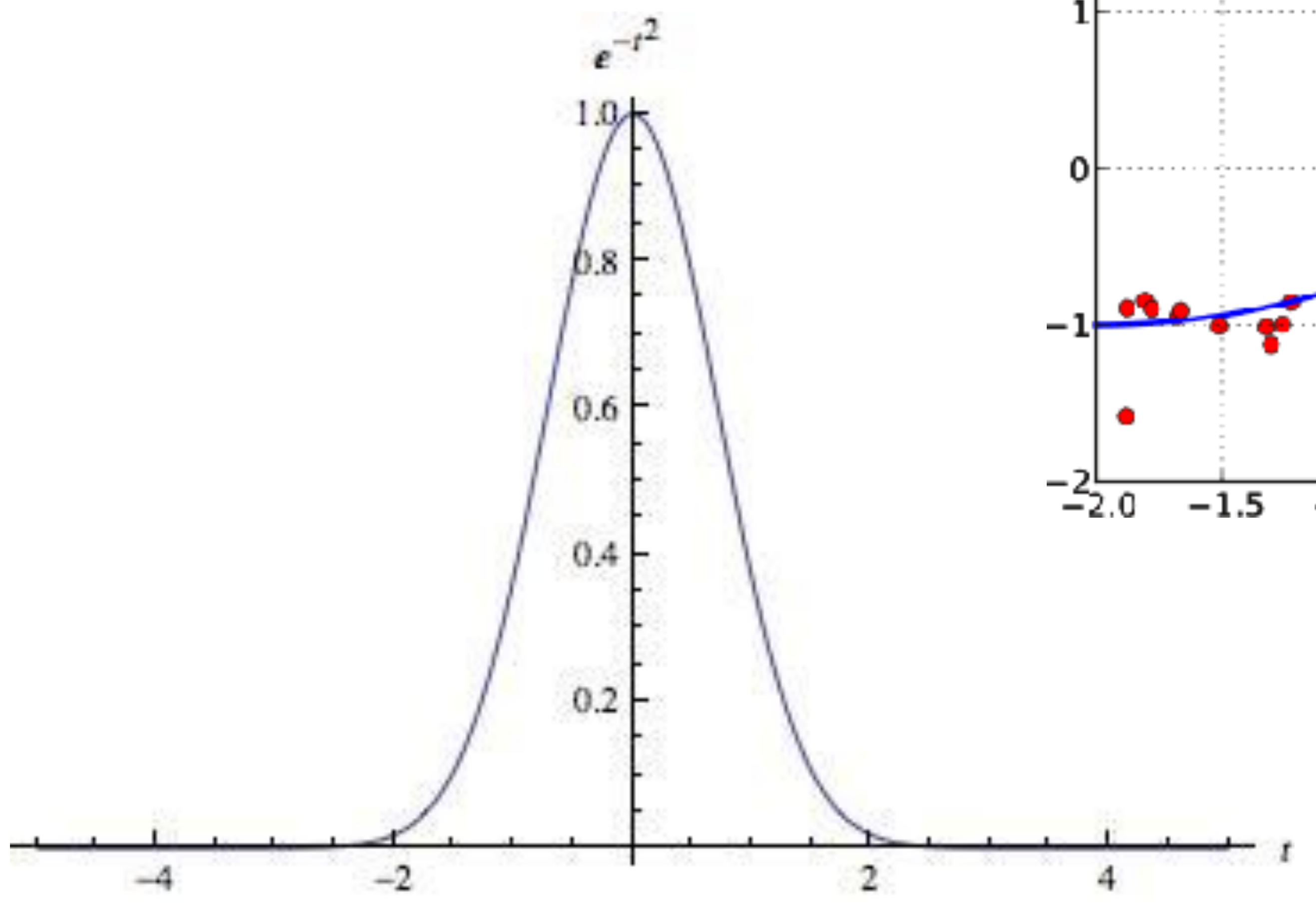


Data Analysis

*Data analysis, […], is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering **useful information**, **suggesting conclusions**, and **supporting decision-making**.*
- **Wikipedia**



- We want to learn something from the data
- **Descriptive statistics**: summarize a sample
- **Inferential statistics**: learn about the population
- **Probability theory is the foundation**



What is Probability?

What is the probability of a coin flip showing heads?

What is Probability?

What is the probability of a coin flip showing heads?

You build a model: fair coin, two sides, each side has equal probability (principle of indifference)

What is Probability?

What is the probability of a coin flip showing heads?

You build a model: fair coin, two sides, each side has equal probability (principle of indifference)

What is the probability of observing these outcomes of coin flips?

TTTTTTTTTT

THHTTTHTHH

What is Probability?

What is the probability of a coin flip showing heads?

You build a model: fair coin, two sides, each side has equal probability (principle of indifference)

What is the probability of observing these outcomes of coin flips?

Probability of sequence

TTTTTTTTTT $(1/2)^{10} = 0.001$

THHTTTHTHH $(1/2)^{10} = 0.001$

What is Probability?

What is the probability of a coin flip showing heads?

You build a model: fair coin, two sides, each side has equal probability

What is the probability of observing these outcomes of coin flips?

Binomial distribution

$$p(r, N, q) = \binom{N}{r} q^r (1-q)^{(N-r)}$$

	Probability of sequence	Probability of r Tails
TTTTTTTTTT	$(1/2)^{10} = 0.001$	$\binom{10}{10}(1/2)^{10} = 0.001$
THHTTTHTHH	$(1/2)^{10} = 0.001$	$\binom{10}{5}(1/2)^{10} = 256 * 0.001$

What is Probability?

What is the probability of a coin flip showing heads?

You build a model: fair coin, two sides, each side has equal probability

What is the probability of observing these outcomes of coin flips?

Binomial distribution

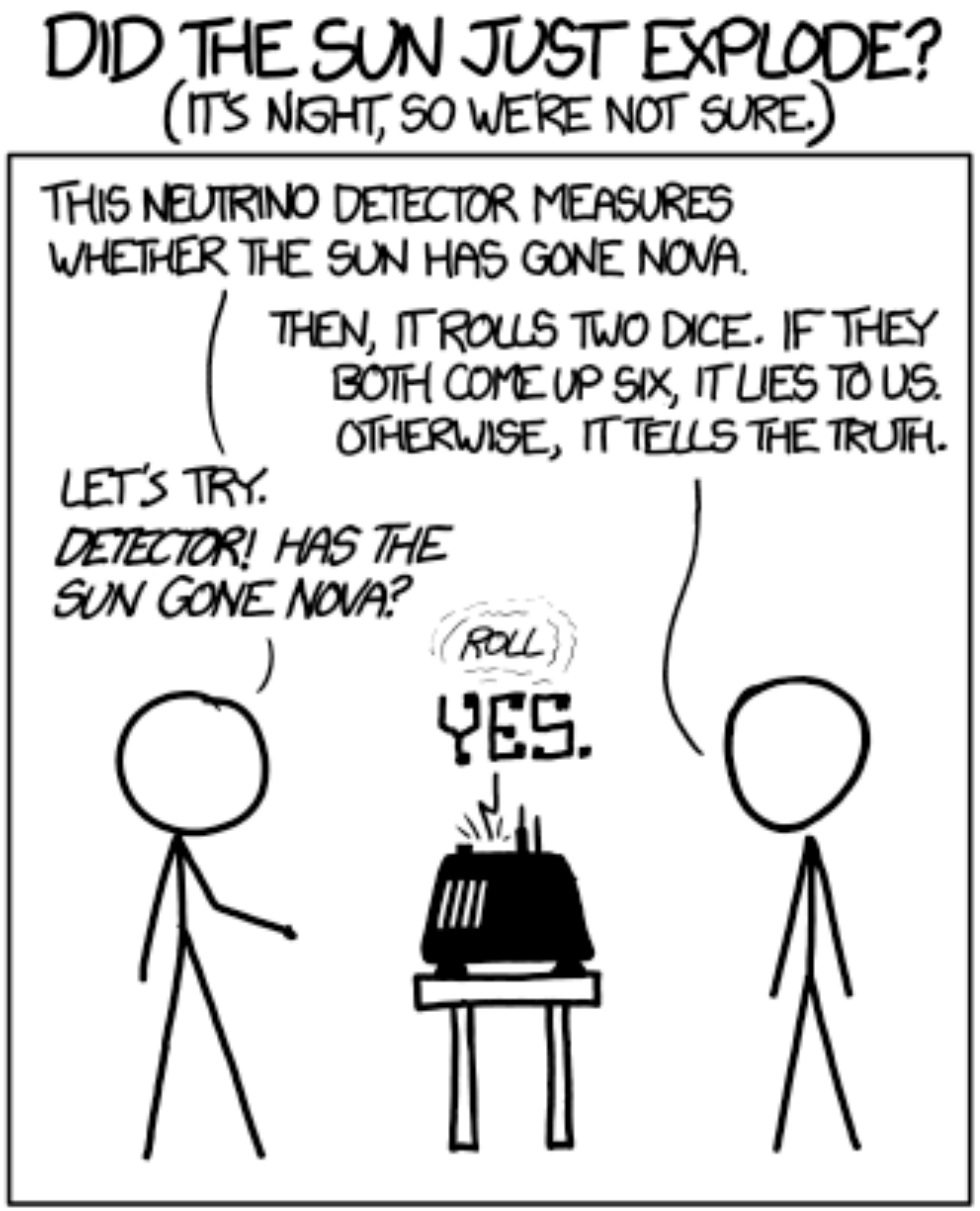
$$p(r, N, q) = \binom{N}{r} q^r \, (1-q)^{(N-r)}$$

	Probability of sequence	Probability of r Tails
TTTTTTTTTT	$(1/2)^{10} = 0.001$	$\binom{10}{10}(1/2)^{10} = 0.001$
THHTTTHTHH	$(1/2)^{10} = 0.001$	$\binom{10}{5}(1/2)^{10} = 256 * 0.001$
THTHTHTHTH	$(1/2)^{10} = 0.001$	$\binom{10}{5}(1/2)^{10} = 256 * 0.001$

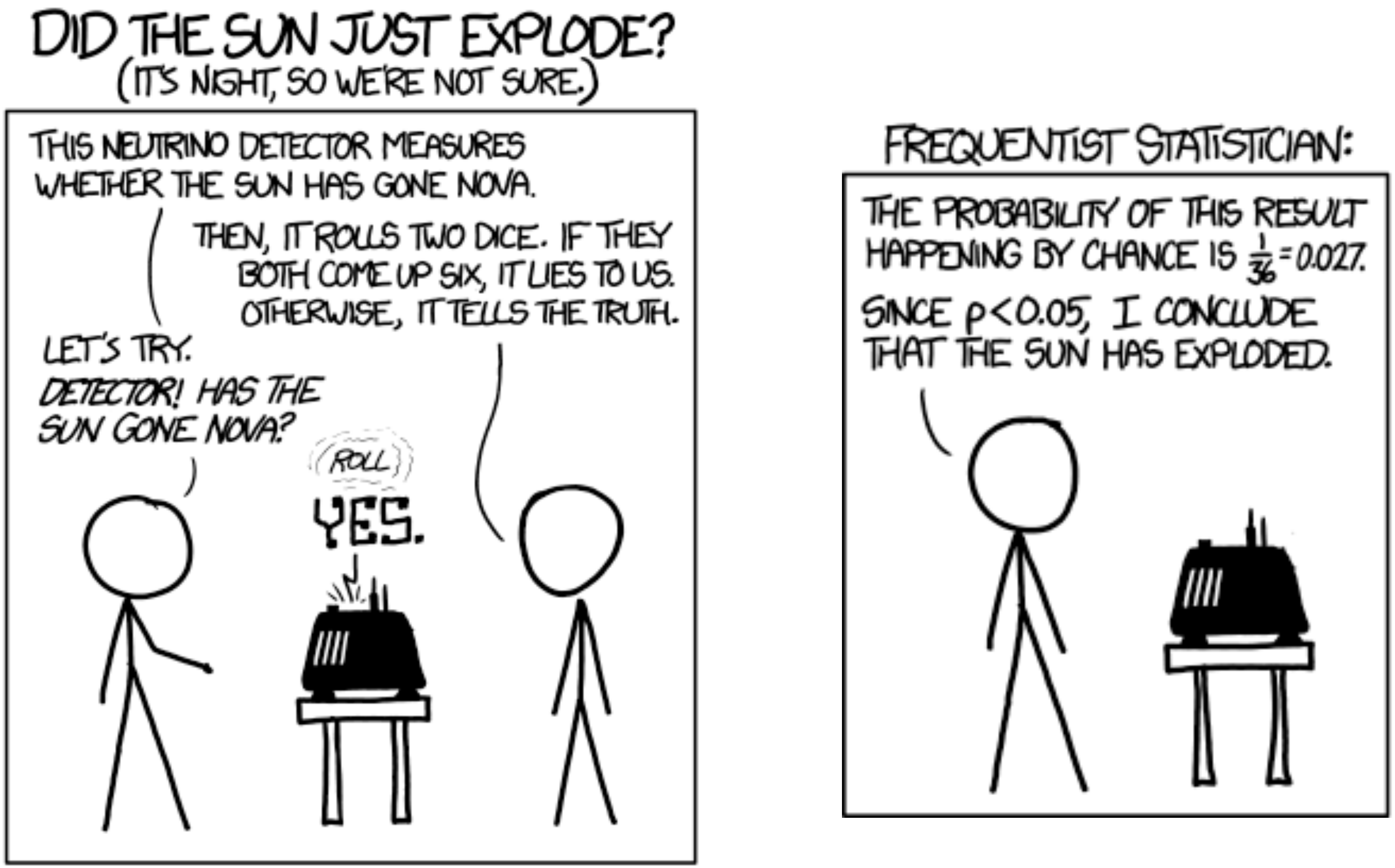
There is not an intrinsic „real" probability in a problem.

Which probability you are interested in needs to be chosen for each problem.

Frequentist vs Bayes

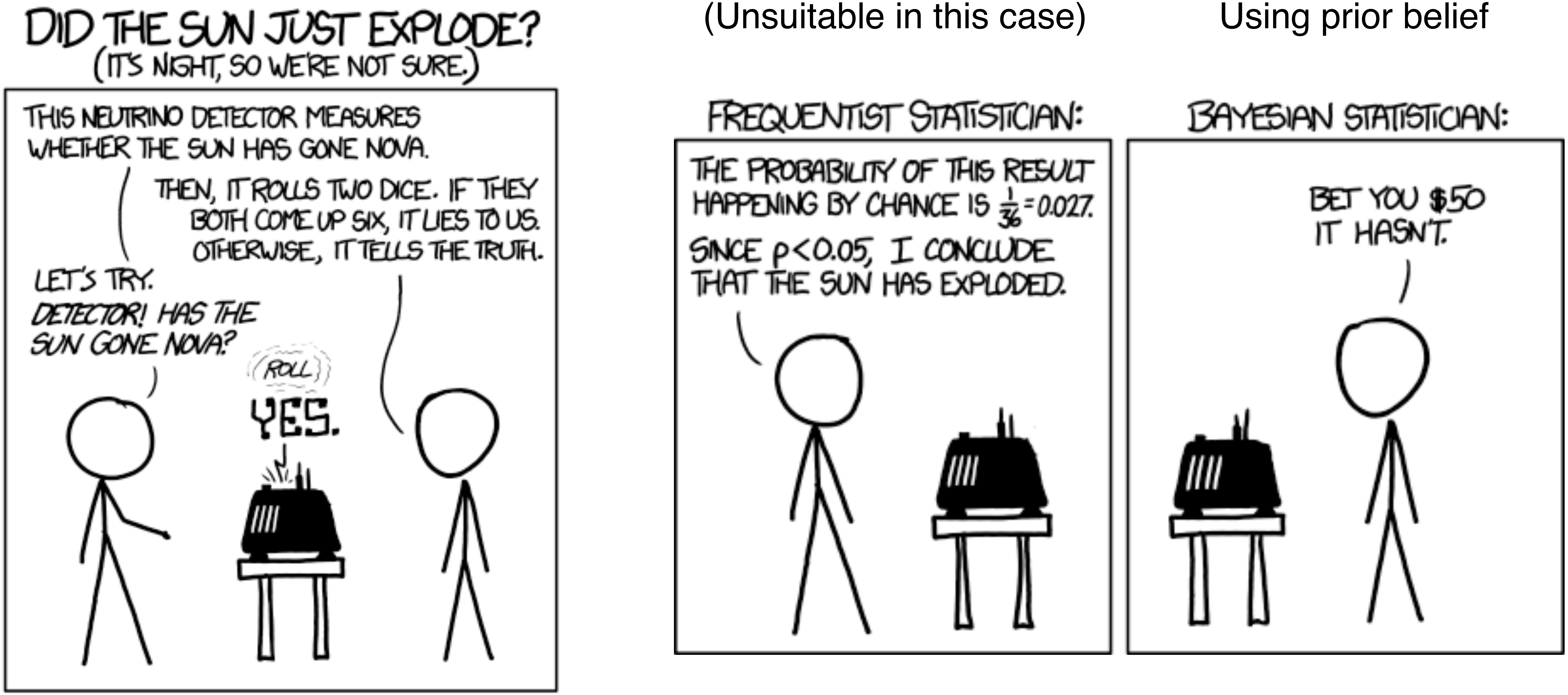


Frequentist vs Bayes



Frequentist vs Bayes

(Unsuitable in this case) Using prior belief



Frequentists vs. Bayesians

- **Frequentist probability:** Limit of its relative frequency in a large number of trials
- „After observing 58 heads out of 100 flips, I estimate the probability of observing heads to be 58%“

- **Bayesian probability:** Degree of belief
- „From previous experience we have most probable a fair coin. After observing 58 heads out of 100 flips, I update my prior belief. The most probable value now is 54%“

“The argument in the academic community is mostly esoteric tail wagging anyway. In truth most analysts out of the ivory tower don’t care that much, if at all, about Bayesian vs. Frequentist.”

- ***Dr. Rob Balon***

The Raffle problem

You buy 10 tickets for a raffle, 5 of them win. What are the odds of winning?

The Raffle problem

You buy 10 tickets for a raffle, 5 of them win. What are the odds of winning?

You buy 1 ticket for a raffle, it wins. What are the odds of winning?

The Raffle problem

You buy 10 tickets for a raffle, 5 of them win. What are the odds of winning?

You buy 1 ticket for a raffle, it wins. What are the odds of winning?

*"If we be, therefore, engaged by arguments to put **trust in past experience**, and make it the standard of our **future judgement**, these arguments must be **probable only**.“*
- **David Hume**

- Bayes introduced concept of likelihood around 1750: $p(\mathrm{data}|\theta)$
- **Likelihood:** What is the probability of our observation given a certain model?

Thomas Bayes

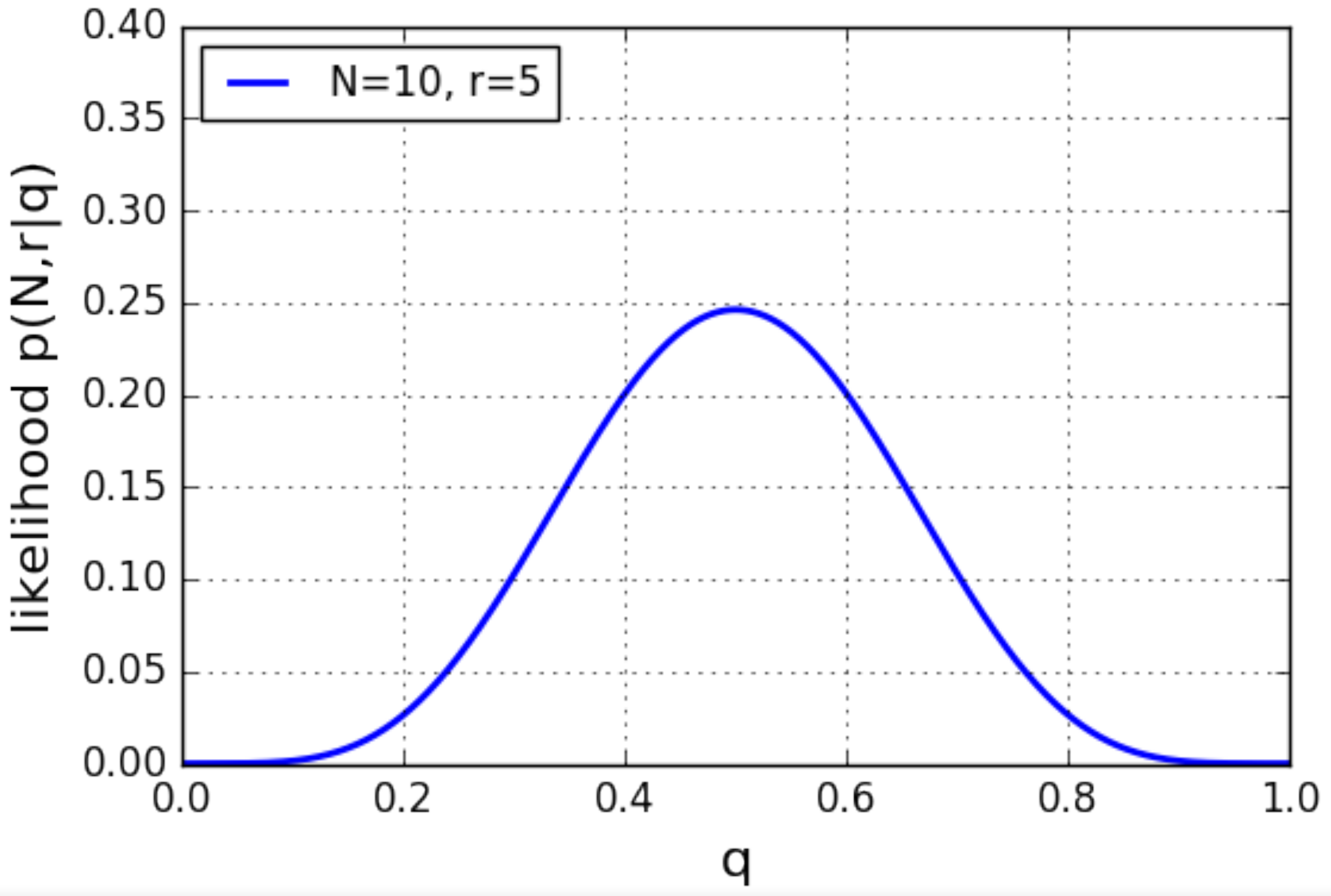
The Likelihood

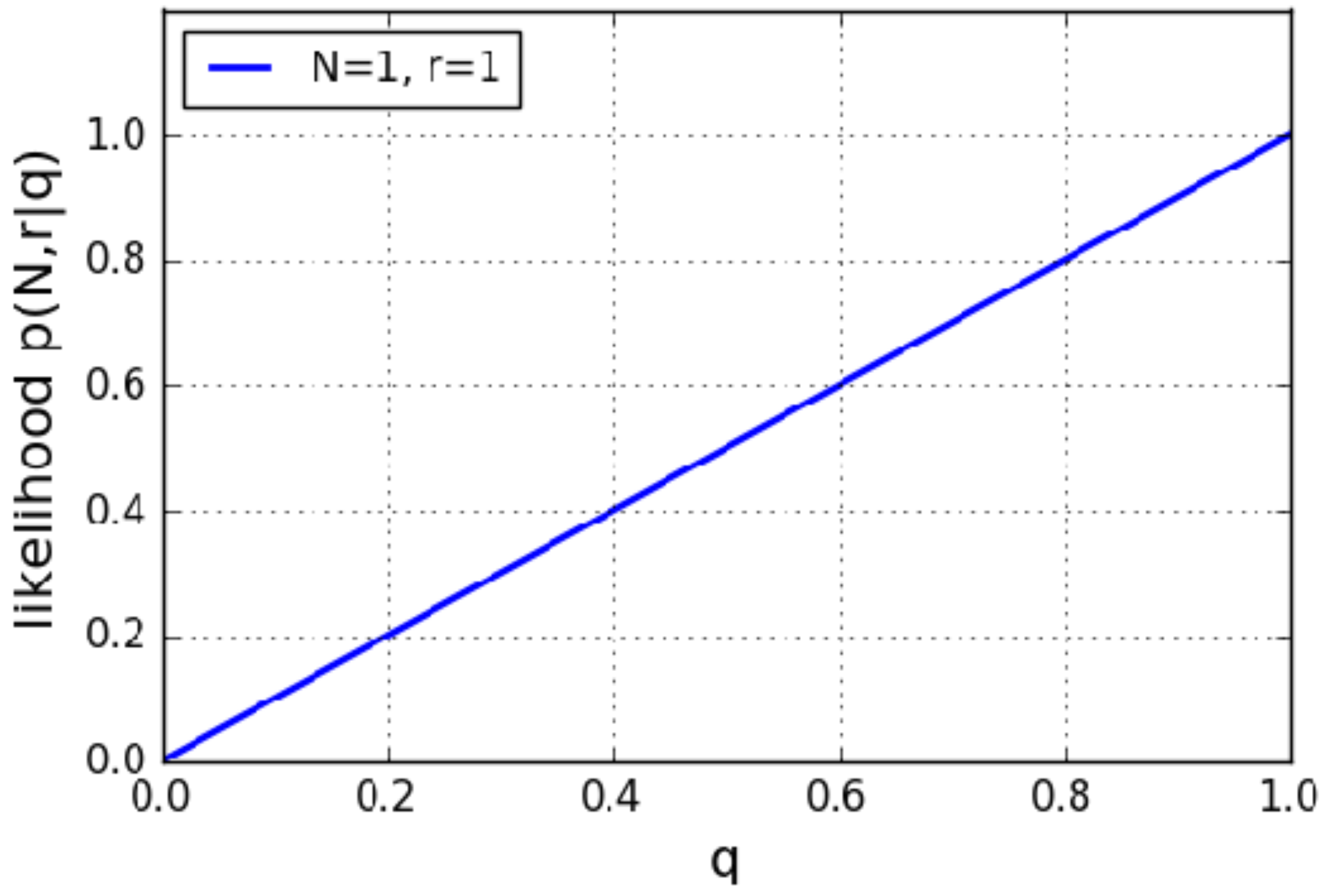
You buy 10 tickets for a raffle, 5 of them win. What are the odds of winning?

You buy 1 ticket for a raffle, it wins. What are the odds of winning?

- **Likelihood:** What is the probability of our observation given a certain model? $p(\mathrm{data}|\theta)$

$$p(N,r|q) = \binom{N}{r} q^r (1-q)^{N-r}$$





First step: define probability model

The likelihood is not normalized

The likelihood is <u>not</u> a probability distribution

The Likelihood

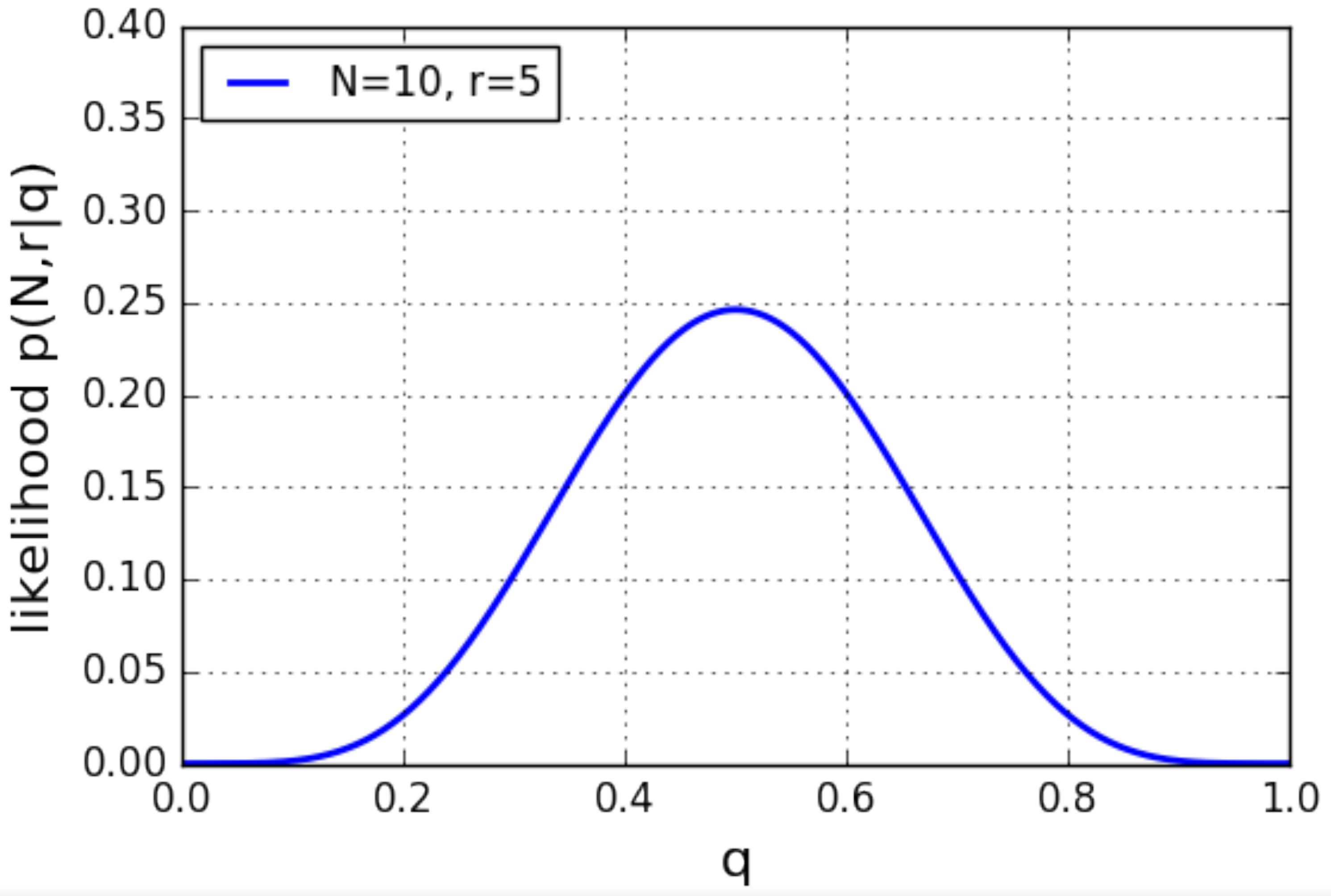
You buy 10 tickets for a raffle, 5 of them win. What are the odds of winning?

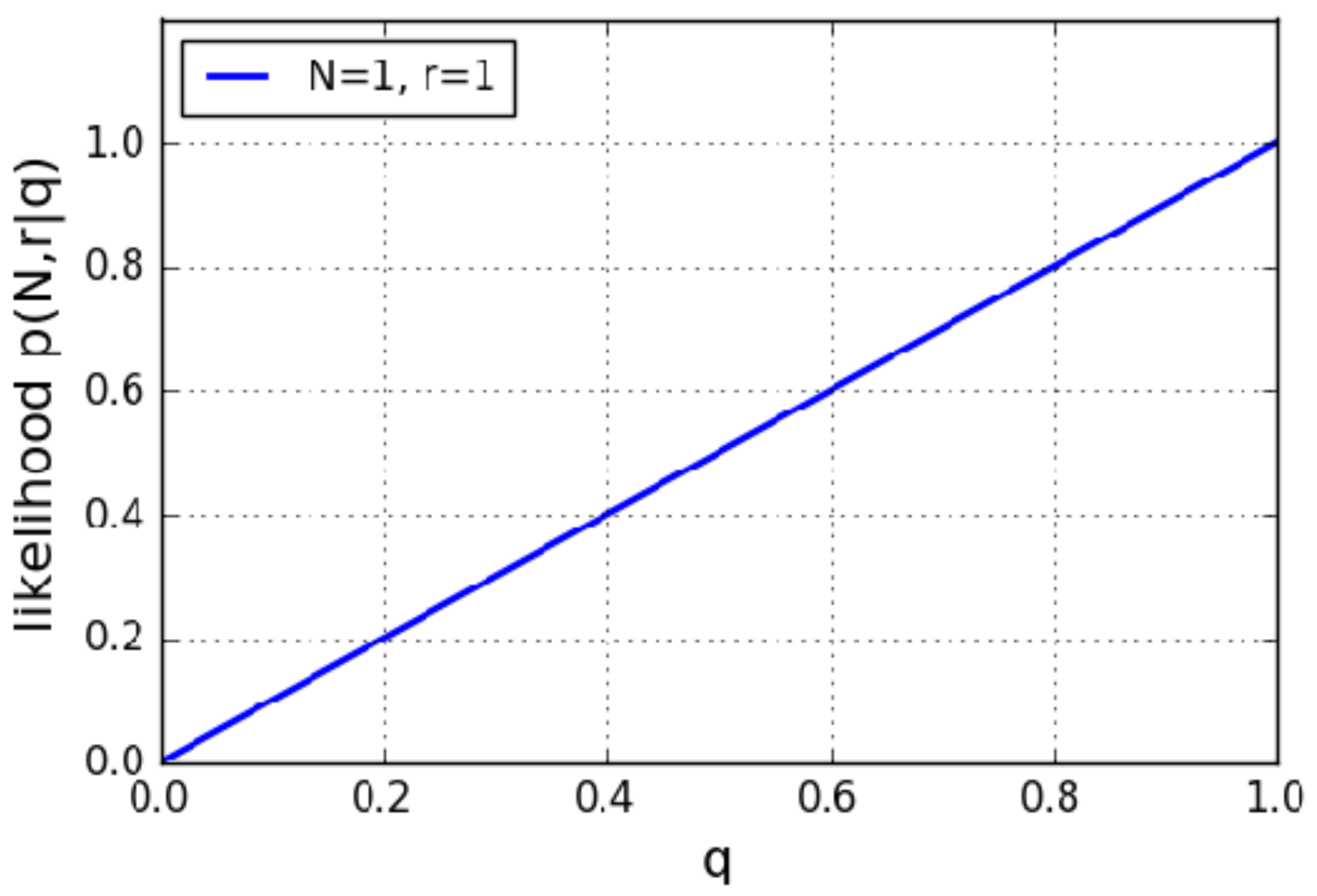
You buy 1 ticket for a raffle, it wins. What are the odds of winning?

- **Likelihood:** What is the probability of our observation given a certain model? $p(\text{data}|\theta)$

$$p(N, r|q) = \binom{N}{r} q^r (1-q)^{N-r}$$

What we really want to know: $p(\theta|\text{data})$





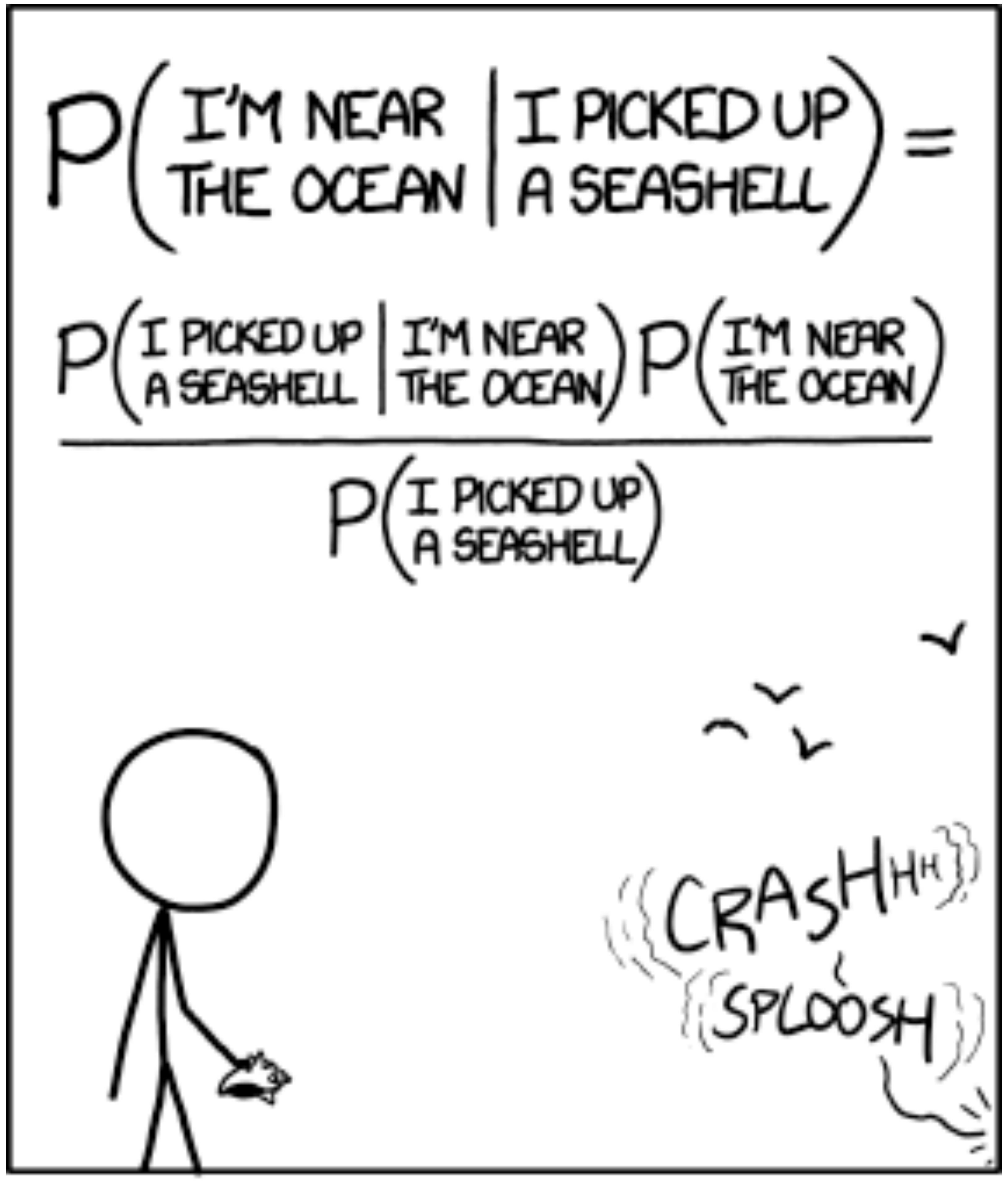
First step: define probability model

The likelihood is not normalized

The likelihood is <u>not</u> a probability distribution

Bayes Theorem

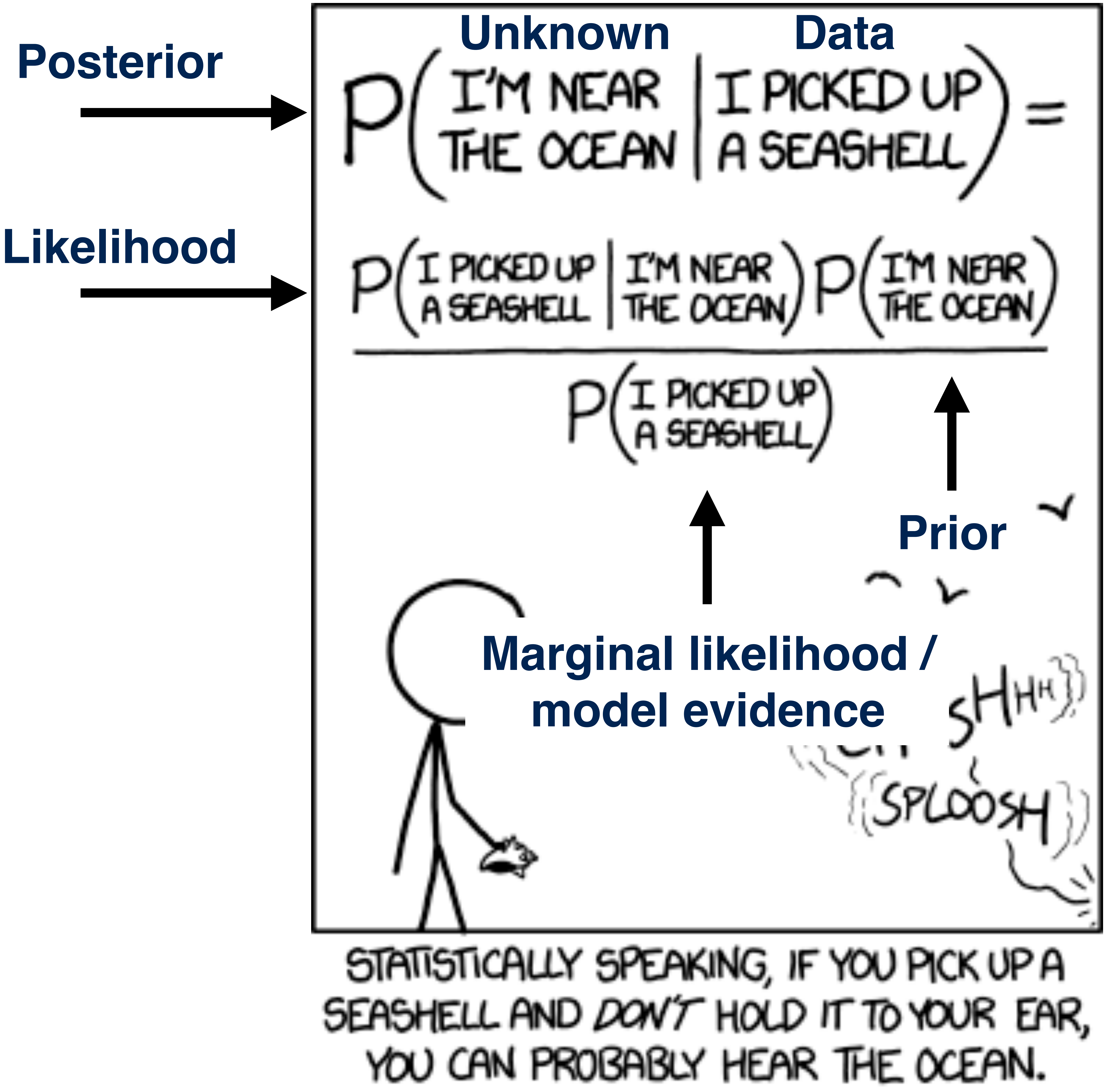
- How to get from the likelihood to a probability distribution on our unknown parameter?



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

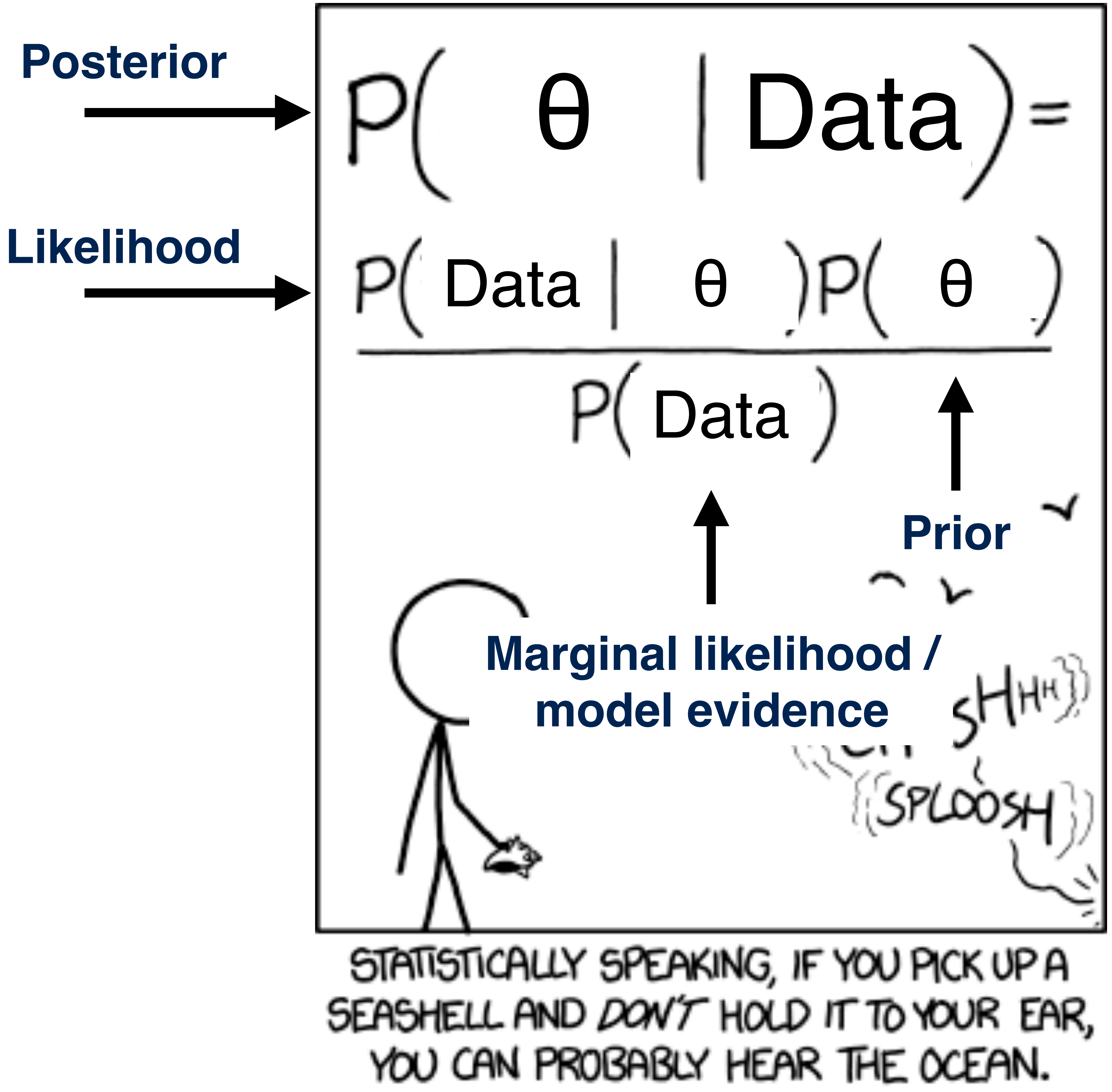
Bayes Theorem

- How to get from the likelihood to a probability distribution on our unknown parameter?



Bayes Theorem

- How to get from the likelihood to a probability distribution on our unknown parameter?



Bayes Theorem

Derivation

Definition of conditional probability:

$$p(A|B) = \frac{p(A \cap B)}{p(B)}, \text{ if } p(B) \neq 0$$

$$p(B|A) = \frac{p(B \cap A)}{p(A)}, \text{ if } p(A) \neq 0$$

$$p(A \cap B) = p(B \cap A)$$

$$\boxed{p(A|B) = \frac{p(B|A)\, p(A)}{p(B)}, \text{ if } p(B) \neq 0}$$

Posterior

Likelihood

$$P(\theta \mid \text{Data}) = \frac{P(\text{Data} \mid \theta)\, P(\theta)}{P(\text{Data})}$$

Prior

Marginal likelihood / model evidence

Mongolian Swamp Fewer

- Rare disease: Mongolian swamp fever **(1 in 10 000 infected)**
- Luckily there is a test:
  - If you **have MSF**, the test will report ***positive*** with 99.9% probability
  - If you **don’t have MSF**, the test will report ***positive*** with 0.5% probability
- You are taking the test and it reports: ***positive***. Should you be worried?
- What is the probability you indeed have MSF?

[Andrew Walter]

Mongolian Swamp Fewer

- Rare disease: Mongolian swamp fever **(1 in 10 000 infected)**
- Luckily there is a test:
  - If you **have MSF**, the test will report ***positive*** with 99.9% probability
  - If you **don’t have MSF**, the test will report ***positive*** with 0.5% probability
- You are taking the test and it reports: ***positive***. Should you be worried?
- What is the probability you indeed have MSF?

Law of total probability

$$p(A) = \sum p(A|C_i)p(C_i)$$

$$p(\mathrm{MSF}|\mathrm{test\ pos}) = \frac{p(\mathrm{test\ pos}|\mathrm{MSF})\, p(\mathrm{MSF})}{p(\mathrm{test\ pos})}$$

$$p(\mathrm{test\ pos}|\mathrm{MSF}) = 0.999$$

$$p(\mathrm{MSF}) = 10^{-4}$$

Law of total probability

$$\begin{aligned} p(\mathrm{test\ pos}) &= p(\mathrm{test\ pos}|\mathrm{MSF})p(\mathrm{MSF}) + p(\mathrm{test\ pos}|\mathrm{no\ MSF})p(\mathrm{no\ MSF}) \\ &= 0.999 * 10^{-4} + 0.005 * (1 - 10^{-4}) \\ &= 0.005 \end{aligned}$$

$$p(\mathrm{MSF}|\mathrm{test\ pos}) = 0.02$$

Only 2% probability of having MSF with positive test

Intuition on the Prior

Somebody is throwing a coin 10 times and reports it to the game master

For each head (H), the other person gets 1$, for each tail (T), you get 1$

- You play with a good friend: He reports 2T, 8H. Is he lying?
- You play with a stranger: He reports 2T, 8H. Is he lying?

Intuition on the Prior

Somebody is throwing a coin 10 times and reports it to the game master

For each head (H), the other person gets 1$, for each tail (T), you get 1$

- You play with a good friend: He reports 2T, 8H. Is he lying?
- You play with a stranger: He reports 2T, 8H. Is he lying?
- You play with a stranger: He reports 8T, 2H. Is he lying?

Prior may significantly influence your conclusions

Intuition on the Prior

Somebody is throwing a coin 10 times and reports it to the game master

For each head (H), the other person gets 1$, for each tail (T), you get 1$

- You play with a good friend: He reports 2T, 8H. Is he lying?
- You play with a stranger: He reports 2T, 8H. Is he lying?
- You play with a stranger: He reports 8T, 2H. Is he lying?

Prior may significantly influence your conclusions

Good data beats the prior

If I give you a coin. You should test the probability of H. What prior would you use?

Flat prior, gaus around 0.5, max at 0 and 1?

Intuition on the Prior

Somebody is throwing a coin 10 times and reports it to the game master

For each head (H), the other person gets 1$, for each tail (T), you get 1$

- You play with a good friend: He reports 2T, 8H. Is he lying?
- You play with a stranger: He reports 2T, 8H. Is he lying?
- You play with a stranger: He reports 8T, 2H. Is he lying?

Prior may significantly influence your conclusions

Good data beats the prior

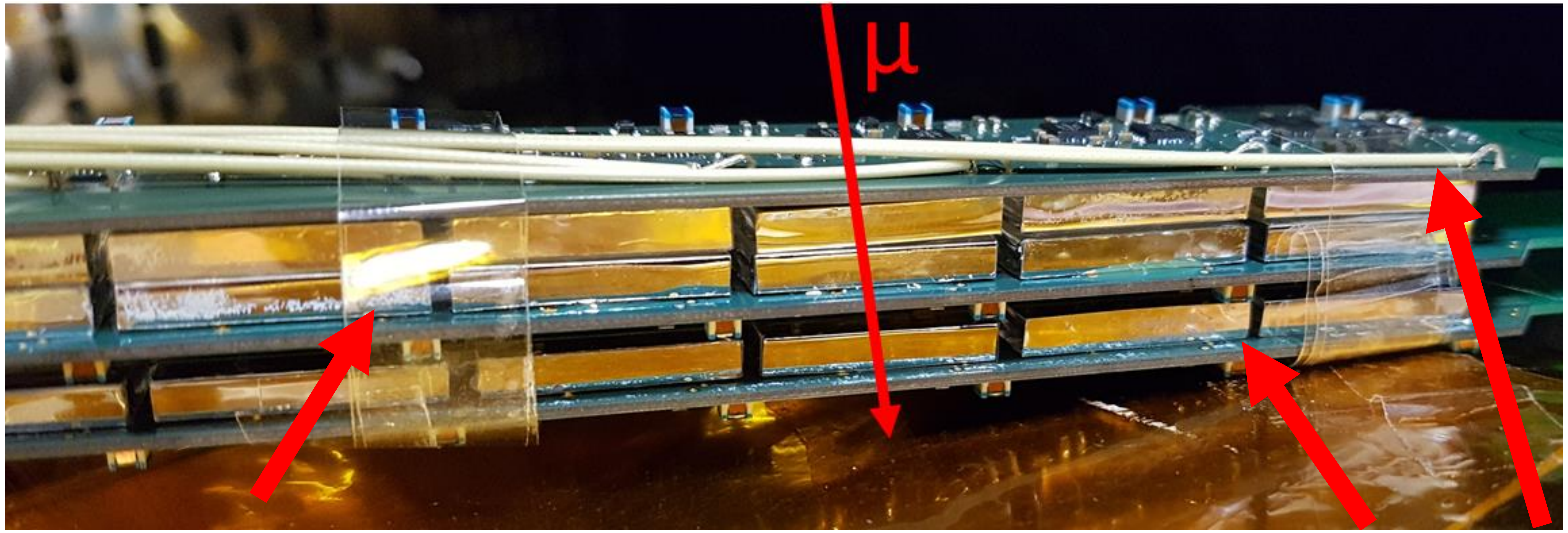
If I give you a coin. You should test the probability of H. What prior would you use?

Flat prior, gaus around 0.5, max at 0 and 1?

Prior may be subjective

How to Perform a Bayesian Analysis?

Example: Efficiency of a Detector



device under test trigger

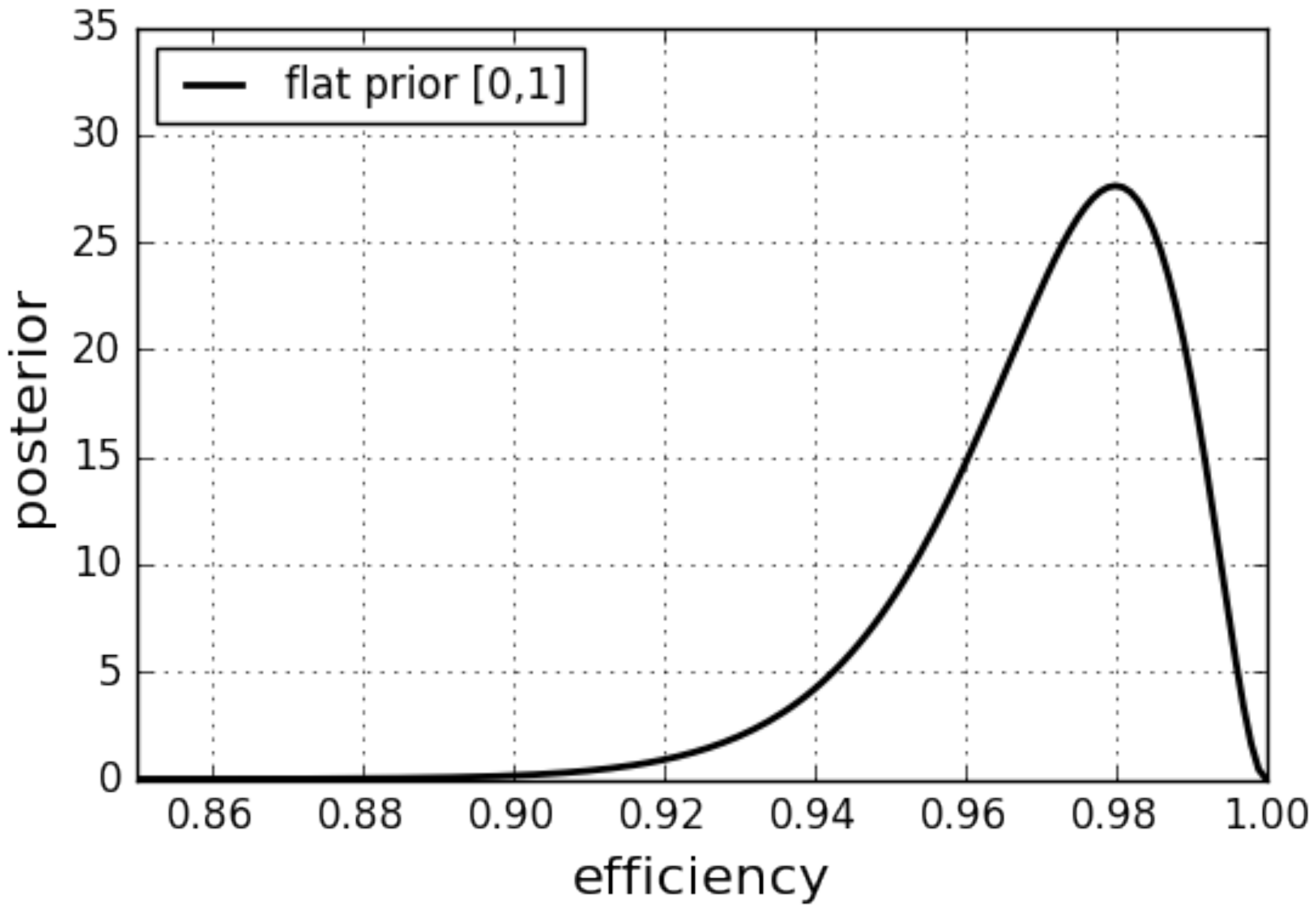
- Measure the efficiency of a device
- N = 100 trials, r = 98 successes

Efficiency of a Detector: Flat Prior

- N = 100 trials, r = 98 successes
- 1. Define a probability model: binomial

$$p(N,r|\epsilon) = \frac{N!}{(N-r)!\,r!}\,\epsilon^r(1-\epsilon)^{N-r}$$ **Likelihood**

- 2. Select a prior: flat prior —> analytical solution
- 3. Calculate posterior



Frequentist only gives point estimate

Bayes Theorem

$$p(\epsilon|N,r) = \frac{p(\epsilon)p(N,r|\epsilon)}{p(N,r)}$$

Flat prior: $p(\epsilon) = 1$

Law of total probability

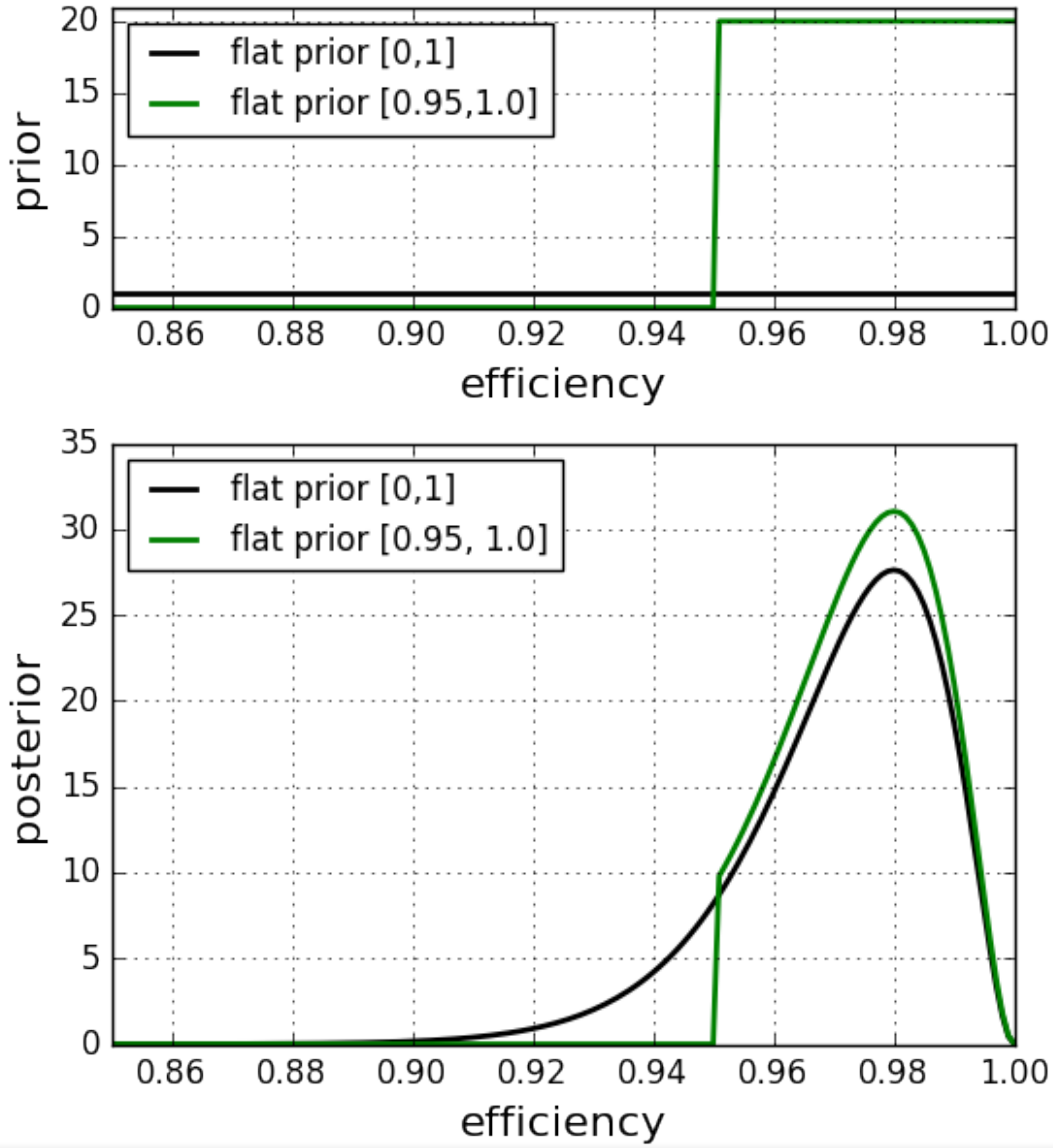
$$p(N,r) = \int_0^1 \binom{N}{r}\epsilon^r(1-\epsilon)^{N-r}\mathrm{d}\epsilon$$

$$p(\epsilon|N,r) = \frac{\binom{N}{r}\epsilon^r(1-\epsilon)^{N-r}}{\int_0^1\binom{N}{r}\epsilon^r(1-\epsilon)^{N-r}\mathrm{d}\epsilon}$$

$$p(\epsilon|N,r) = \frac{(N+1)!}{r!(N-r)!}\epsilon^r(1-\epsilon)^{N-r}$$

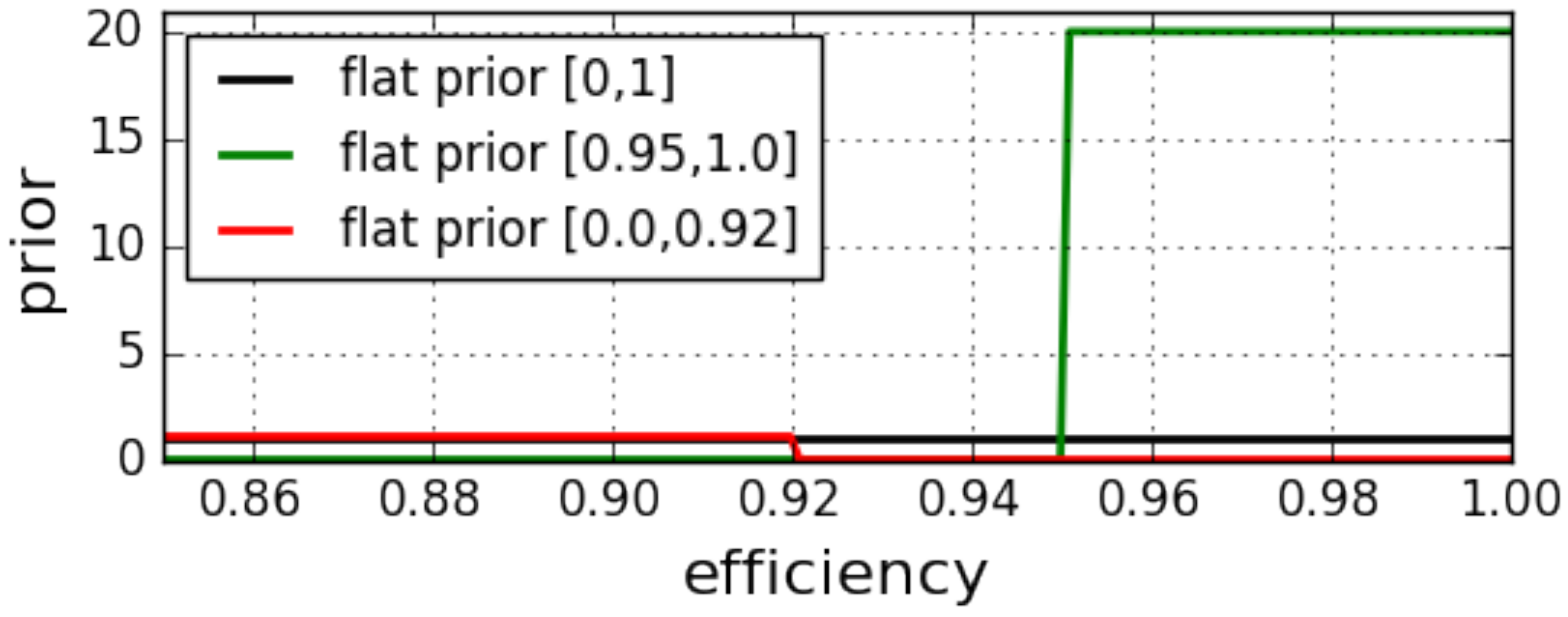
Efficiency of a Detector: Flat Prior

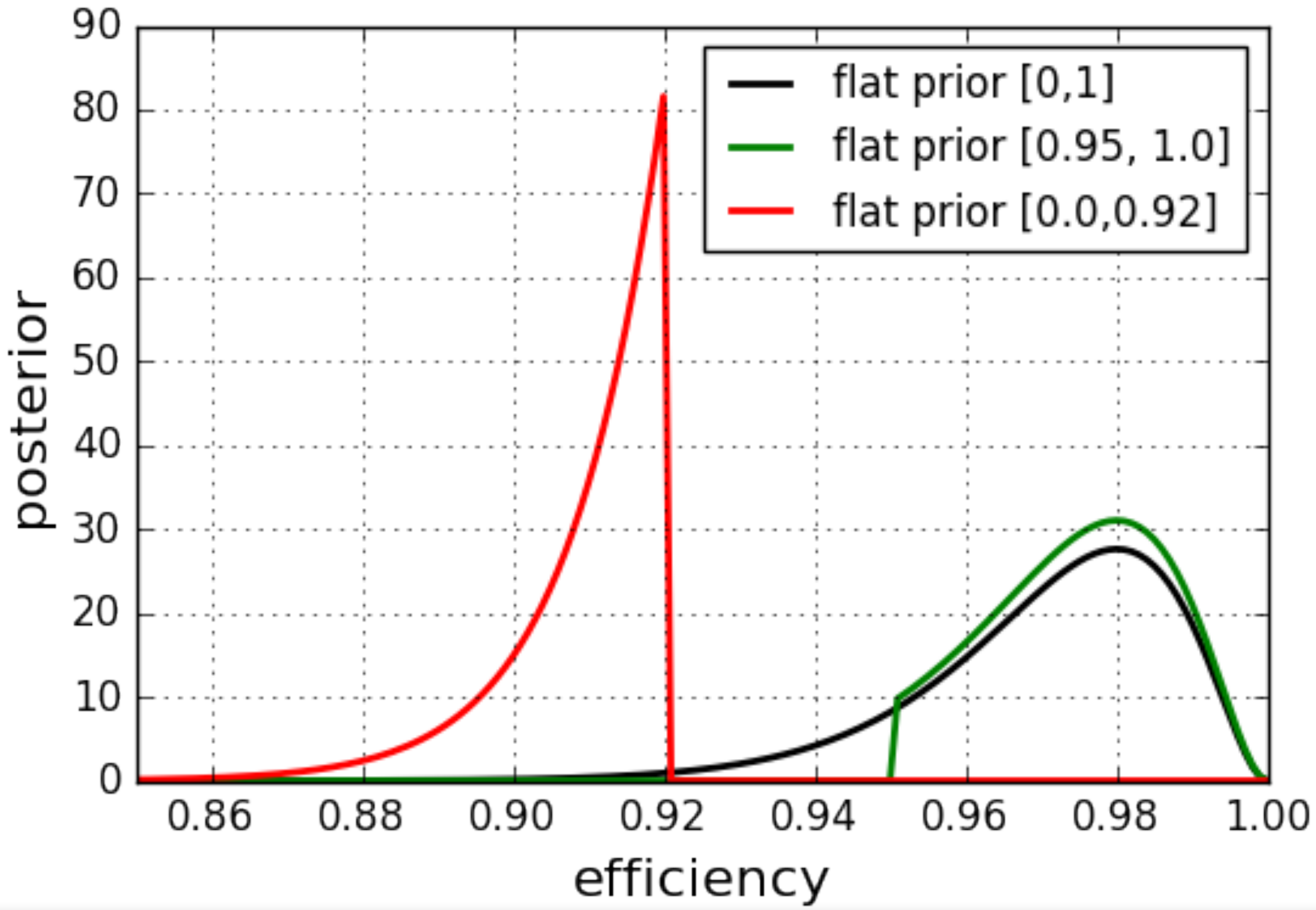
- *What happens if we change the prior?*
- The producer excludes efficiencies below 95%
—> Flat prior [0.95, 1]



Efficiency of a Detector: Flat Prior

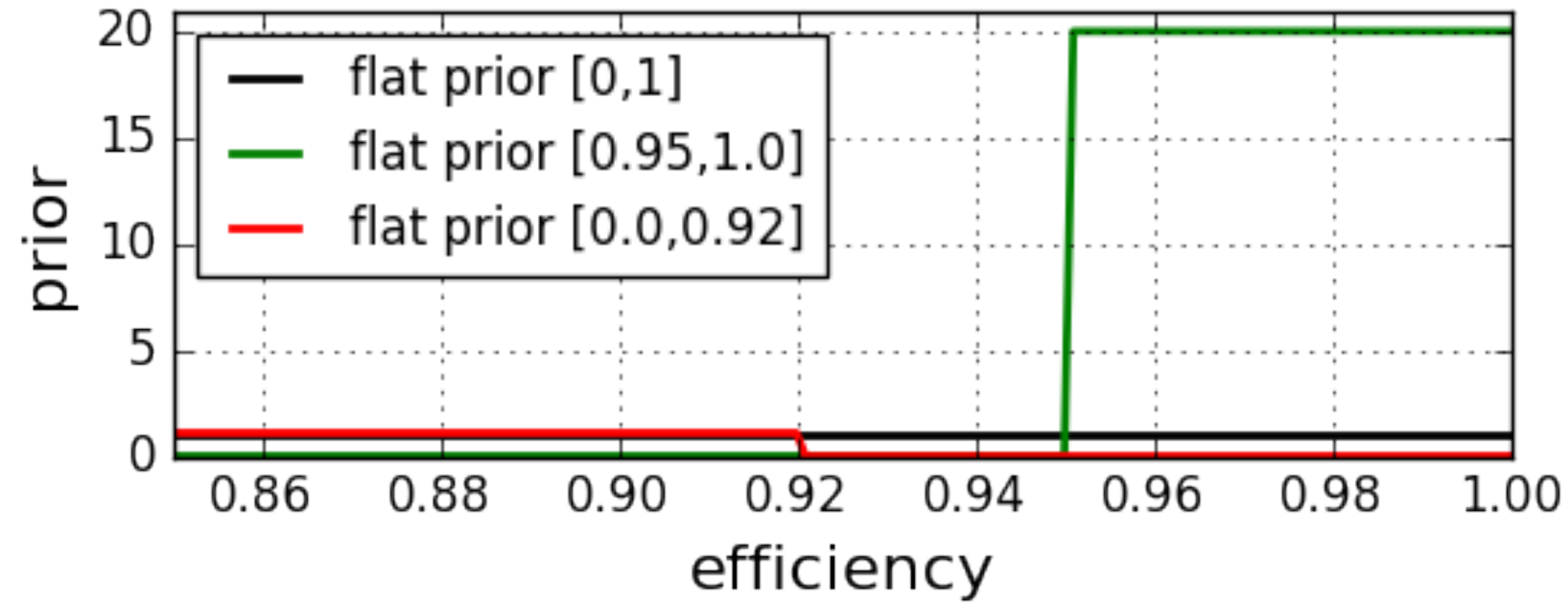
- *What happens if we change the prior?*
- The producer excludes efficiencies below 95%
  —> Flat prior [0.95, 1]
- The producer excludes efficiencies above 92%
  —> Flat prior [0, 0.92]

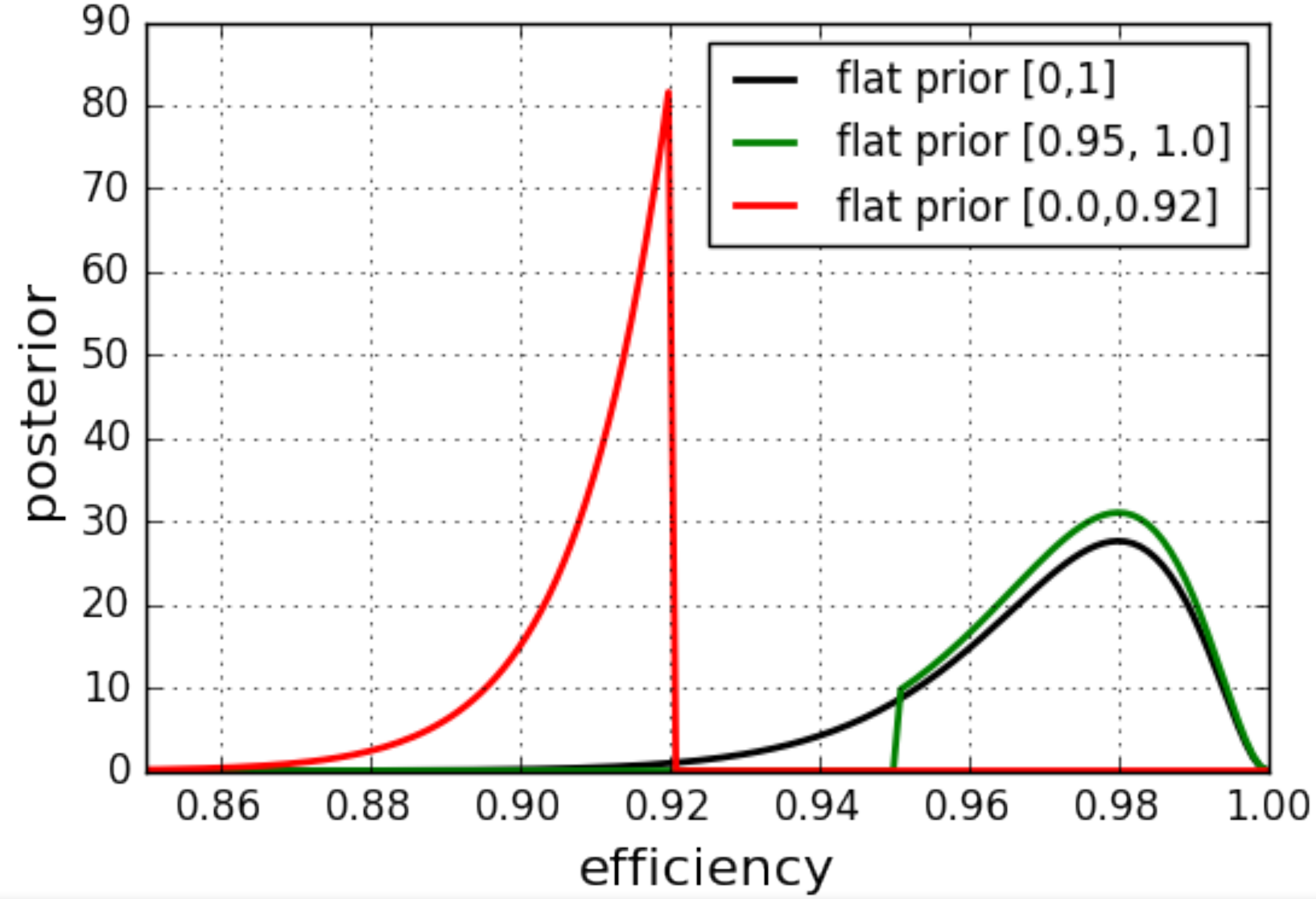




Efficiency of a Detector: Flat Prior

- *What happens if we change the prior?*
- The producer excludes efficiencies below 95%
  —> Flat prior [0.95, 1]
- The producer excludes efficiencies above 92%
  —> Flat prior [0, 0.92]



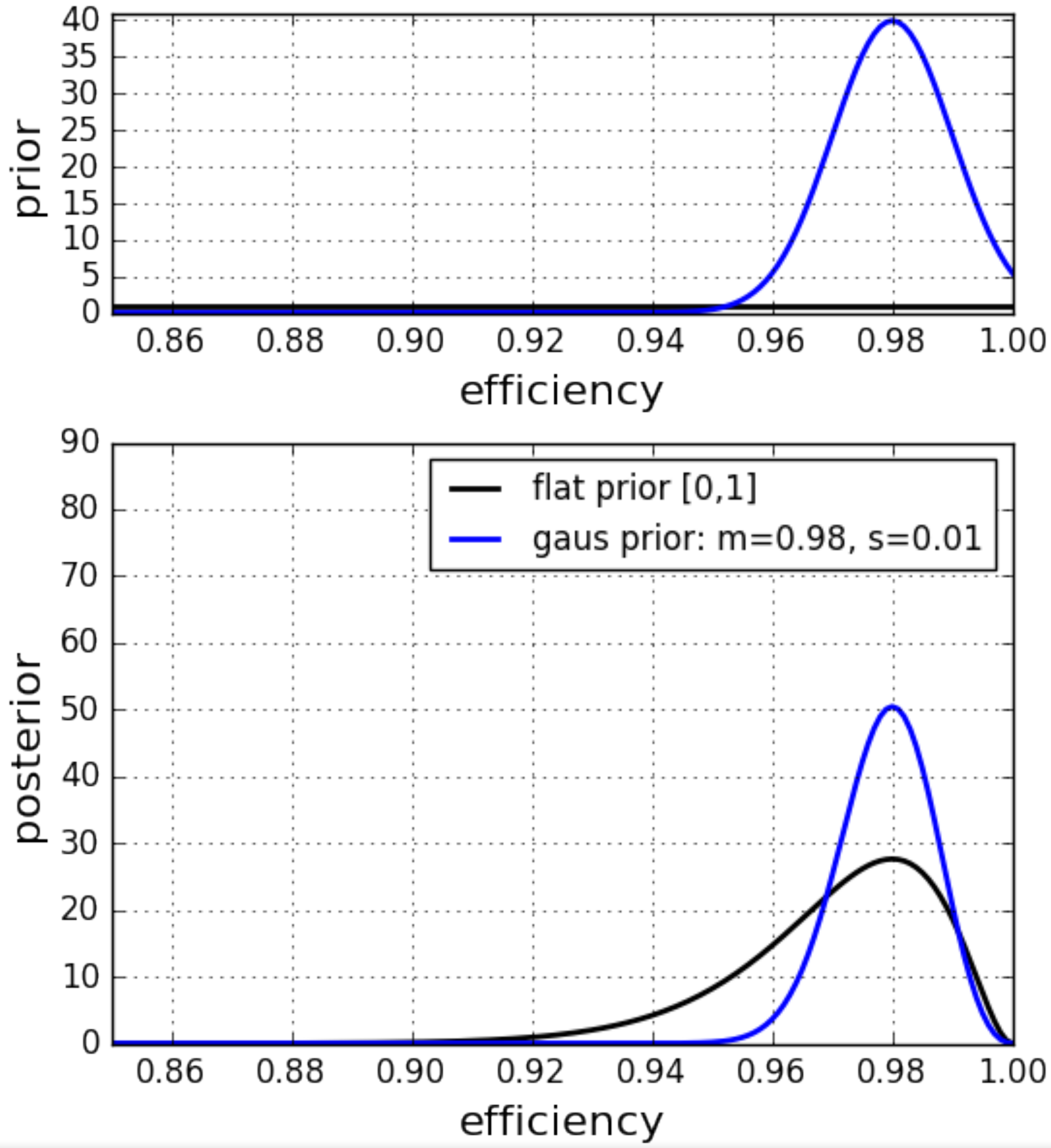


Choosing the wrong prior may have catastrophic consequences for the posterior

Evaluate the consistency between the priors, probability model and posterior

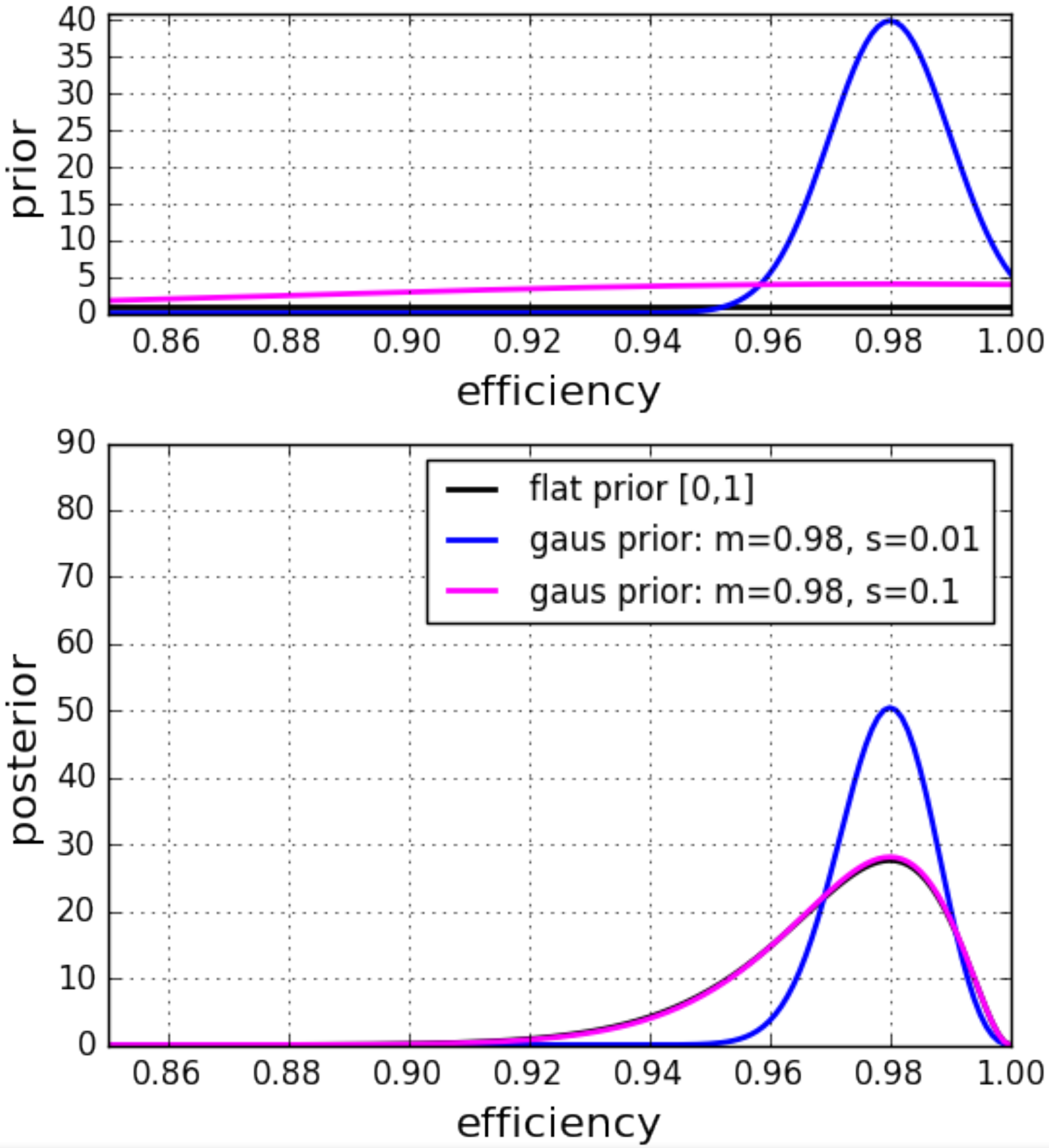
Efficiency of a Detector: Gaussian Prior

- *What happens if we change the prior?*
- A very precise and accurate prior **strengthens the conclusions of the data**.



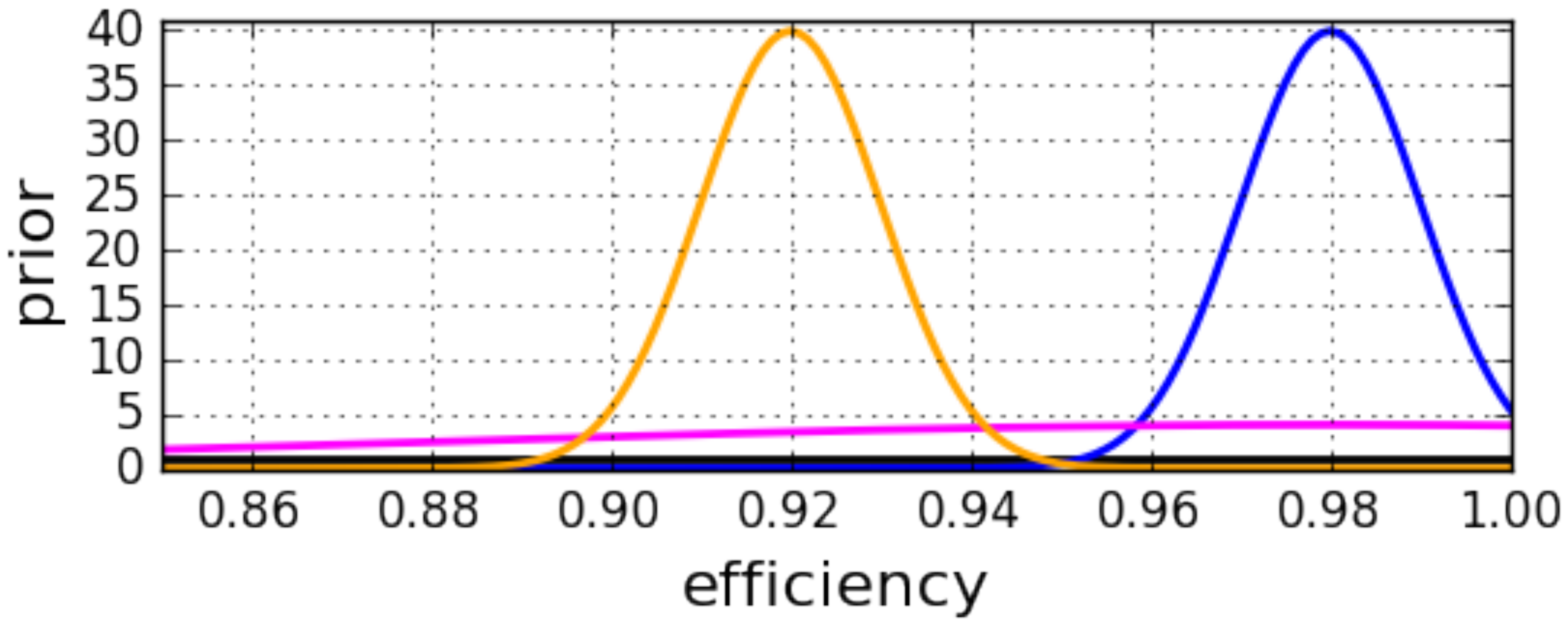
Efficiency of a Detector: Gaussian Prior

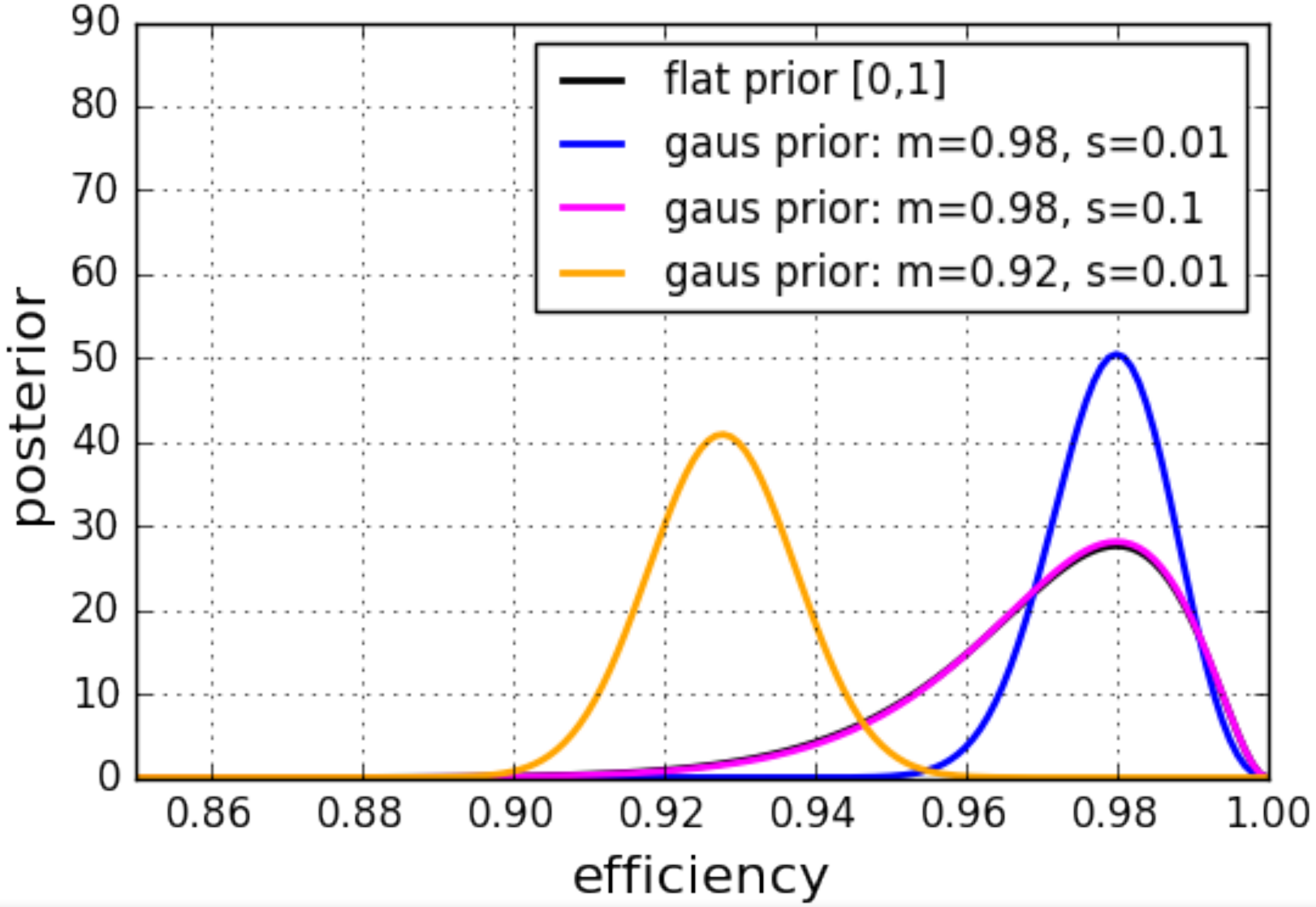
- *What happens if we change the prior?*
- A very precise and accurate prior **strengthens the conclusions of the data**.
- An imprecise prior **does not add much information**



Efficiency of a Detector: Gaussian Prior

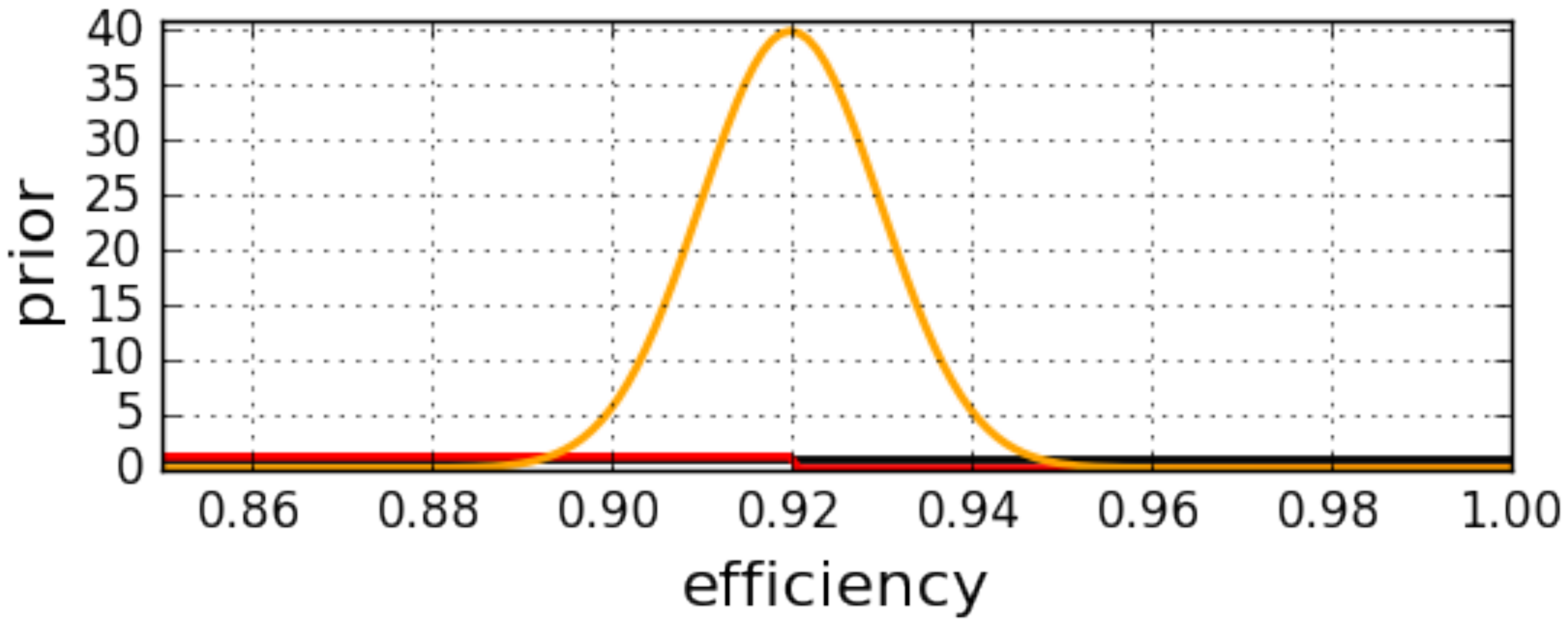
- *What happens if we change the prior?*
- A very precise and accurate prior **strengthens the conclusions of the data**.
- An imprecise prior **does not add much information**
- A precise but inaccurate prior **distorts the conclusions of the data**.

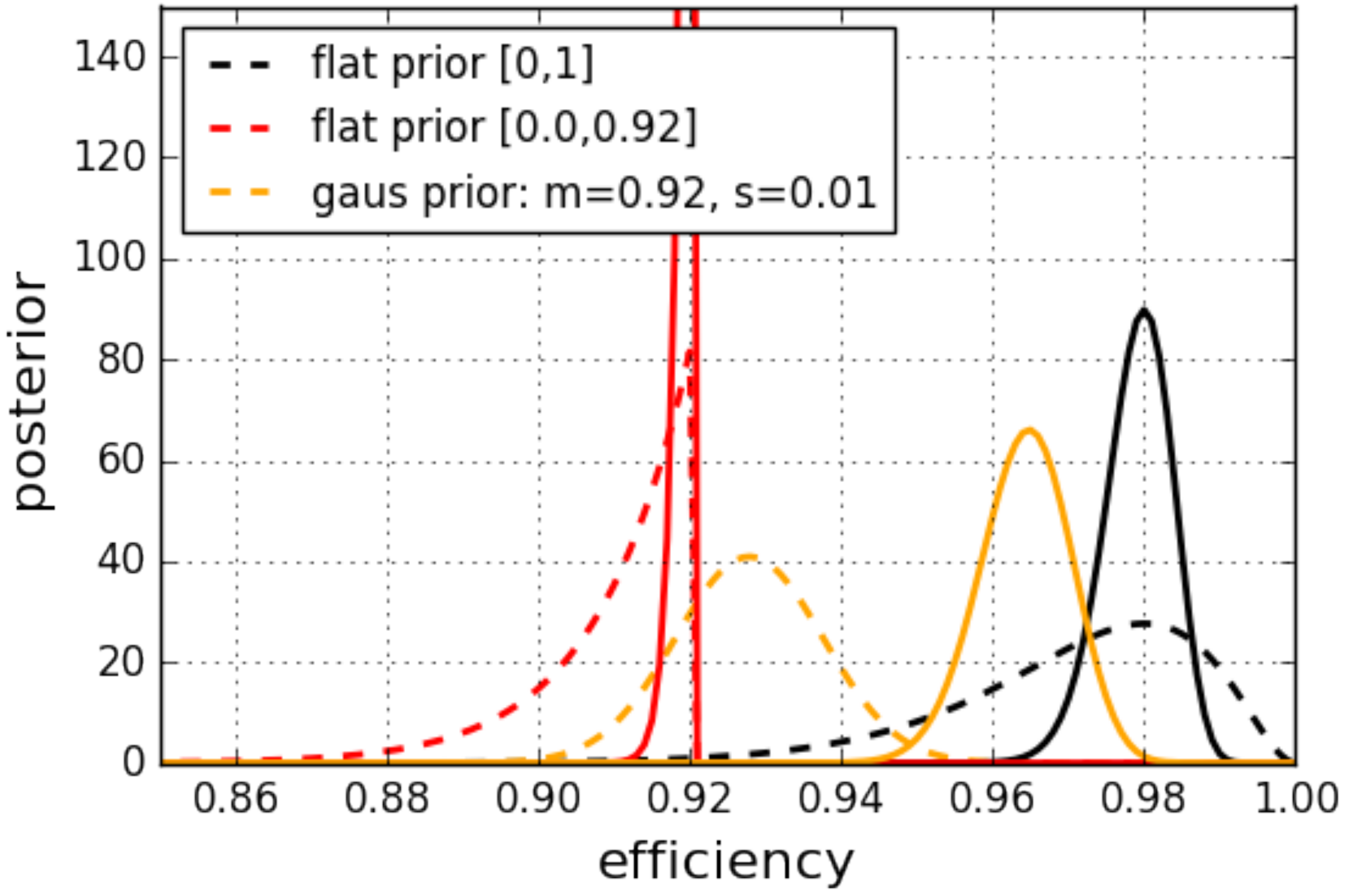




Efficiency of a Detector: Gaussian Prior

- Suppose we have better data:
  n = 1000, r = 980

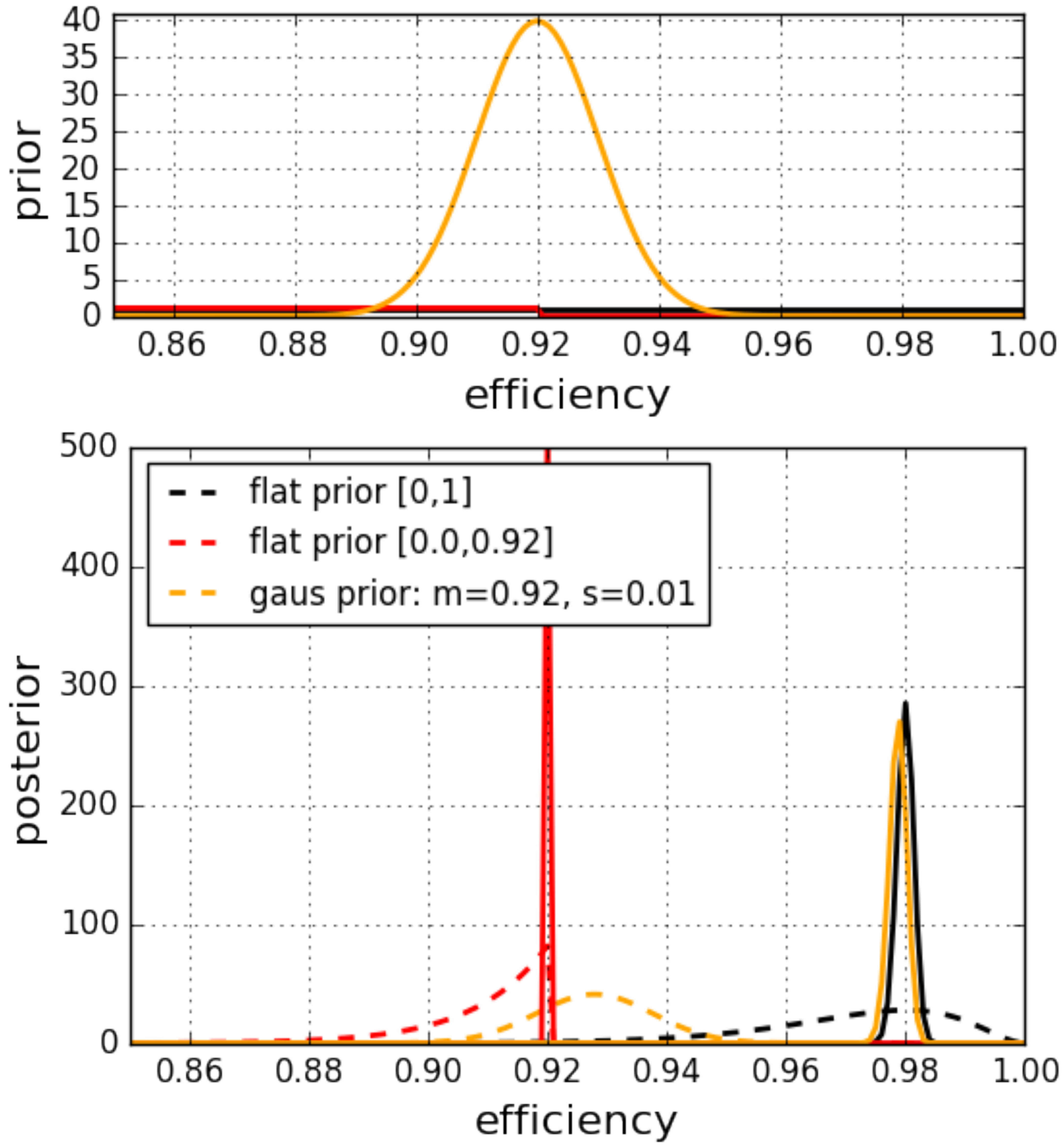




Good data may improve a bad prior

Efficiency of a Detector: Gaussian Prior

- Suppose we have better data:
  n = 10000, r = 9800



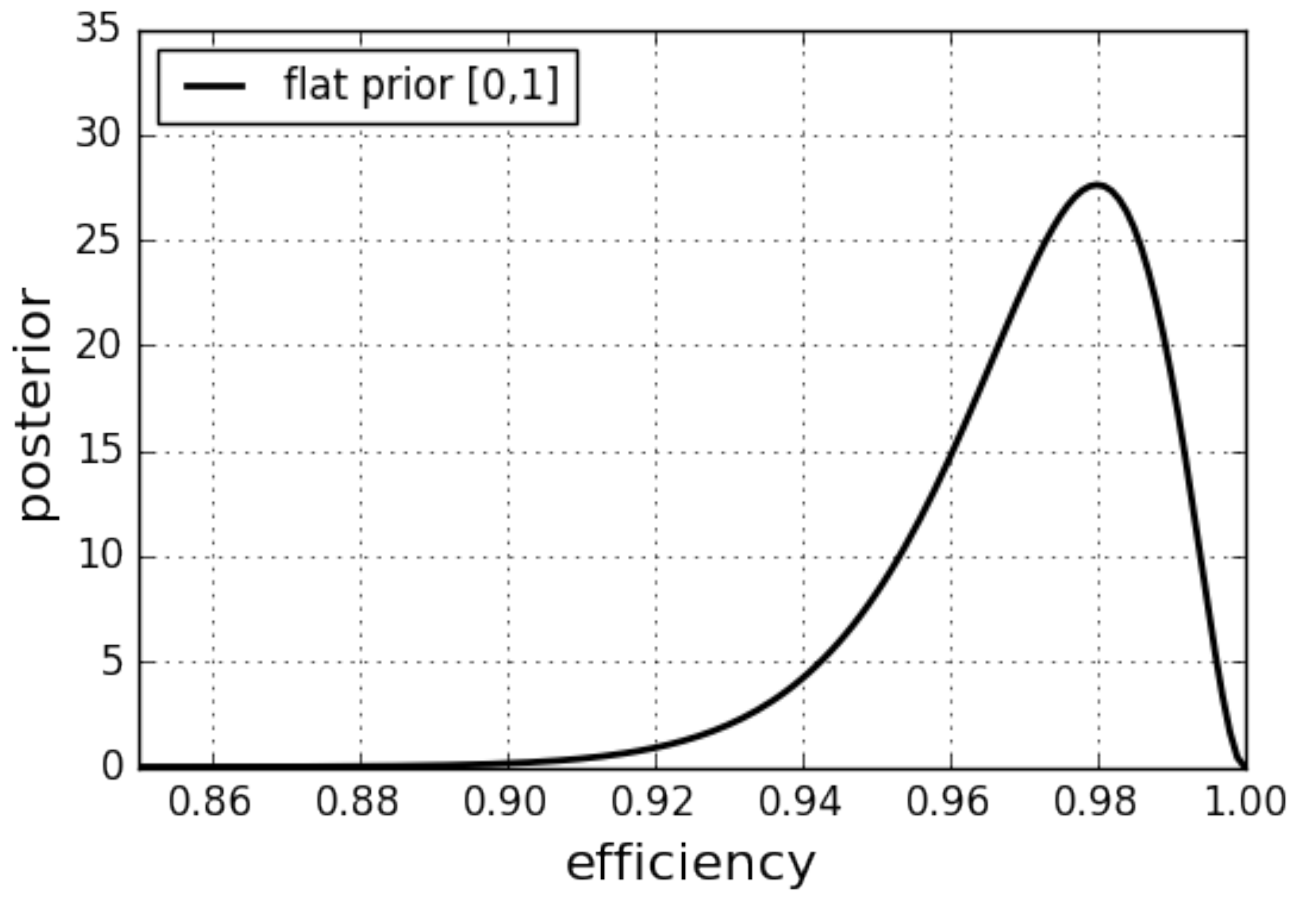
Good data may improve a bad prior

Efficiency of a Detector: Second Measurement

- You are a new student and want to repeat the measurement with more data. You find the old results in a PhD thesis: $N_1 = 100$, $r_1 = 98$
- You measure: $N_2 = 300$, $r_2 = 289$

Bayes Theorem

$$p(\epsilon|N,r) = \frac{p(\epsilon)p(N,r|\epsilon)}{p(N,r)}$$

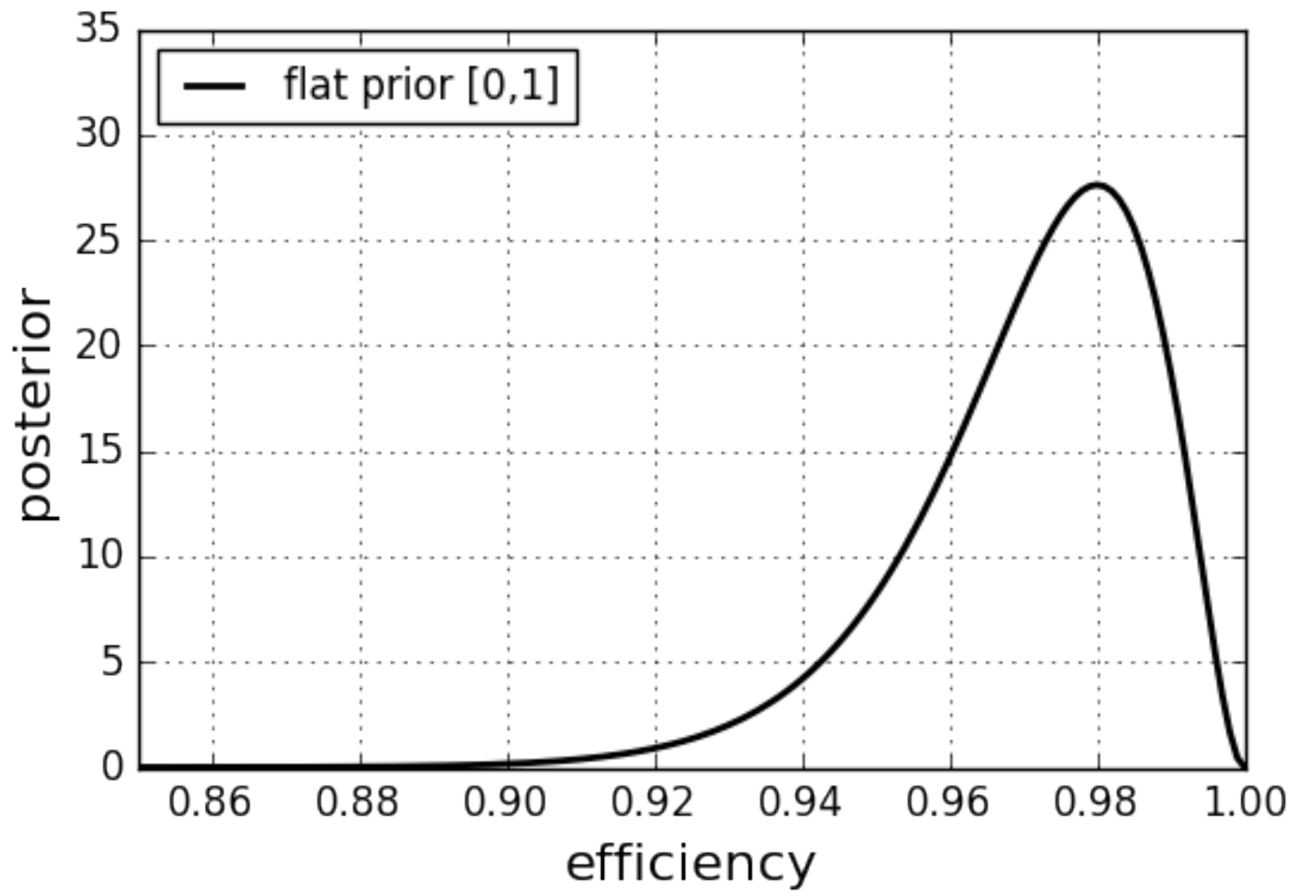


Efficiency of a Detector: Second Measurement

- You are a new student and want to repeat the measurement with more data. You find the old results in a PhD thesis: $N_1 = 100$, $r_1 = 98$
- You measure: $N_2 = 300$, $r_2 = 289$

Bayes Theorem

$$p(\epsilon|N,r) = \frac{p(\epsilon)p(N,r|\epsilon)}{p(N,r)}$$



Previous result: $$p(\epsilon|N_1,r_1) = \frac{(N_1+1)!}{(N_1-r_1)!r_1!}\epsilon^{r_1}(1-\epsilon)^{N_1-r_1}$$

Likelihood: $$p(N_2,r_2|\epsilon) = \frac{N_2!}{(N_2-r_2)!\,r_2!}\,\epsilon^{r_2}(1-\epsilon)^{N_2-r_2}$$

Posterior: $$p(\epsilon|N_2,N_1,r_2,r_1) = \frac{p(r_2|N_2,\epsilon)\,p(\epsilon|N_1,r_1)}{\int p(r_2|N_2,\epsilon)\,p(\epsilon|N_1,r_1)\mathrm{d}\epsilon}$$

$$p(\epsilon|N_2,N_1,r_2,r_1) = \frac{(N_1+N_2+1)!}{(r_1+r_2)!(N_1+N_2-r_1-r_2)!}\,\epsilon^{r_1+r_2}(1-\epsilon)^{N_1+N_2-r_1-r_2}$$

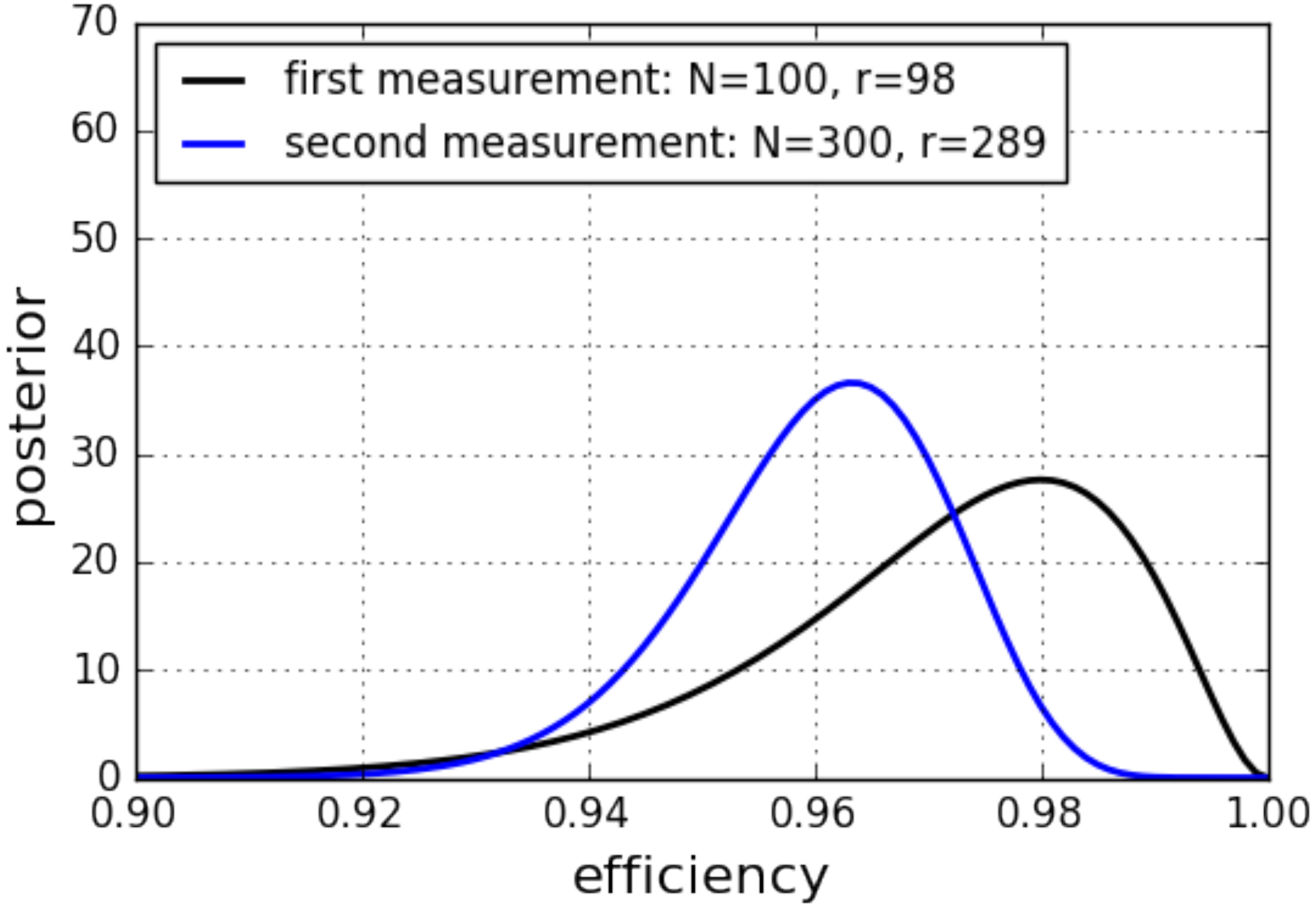
$$N = N_2 + N_1$$
$$r = r_2 + r_1$$

$$p(\epsilon|N,r) = \frac{(N+1)!}{(N-r)!r!}\epsilon^{r}(1-\epsilon)^{N-r}$$

Two consecutive measurements give the same results as one combined measurement

Efficiency of a Detector: Second Measurement

- You are a new student and want to repeat the measurement with more data. You find the old results in a PhD thesis: $N_1 = 100$, $r_1 = 98$
- You measure: $N_2 = 300$, $r_2 = 289$



Two consecutive measurements give the same results as one combined measurement

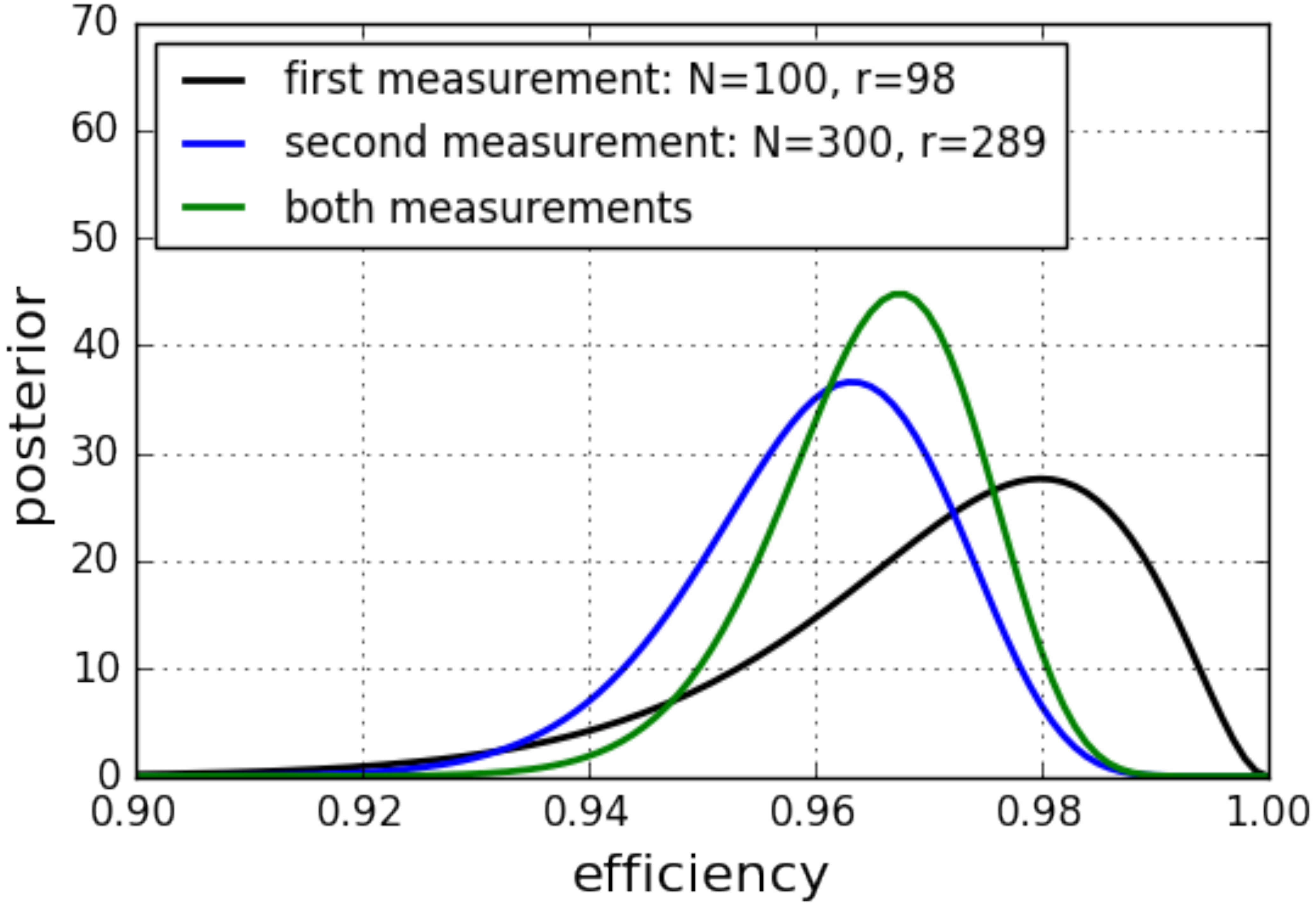
Efficiency of a Detector: Second Measurement

- You are a new student and want to repeat the measurement with more data. You find the old results in a PhD thesis: $N_1 = 100$, $r_1 = 98$
- You measure: $N_2 = 300$, $r_2 = 289$

A Bayesian is one who, vaguely expecting a horse,

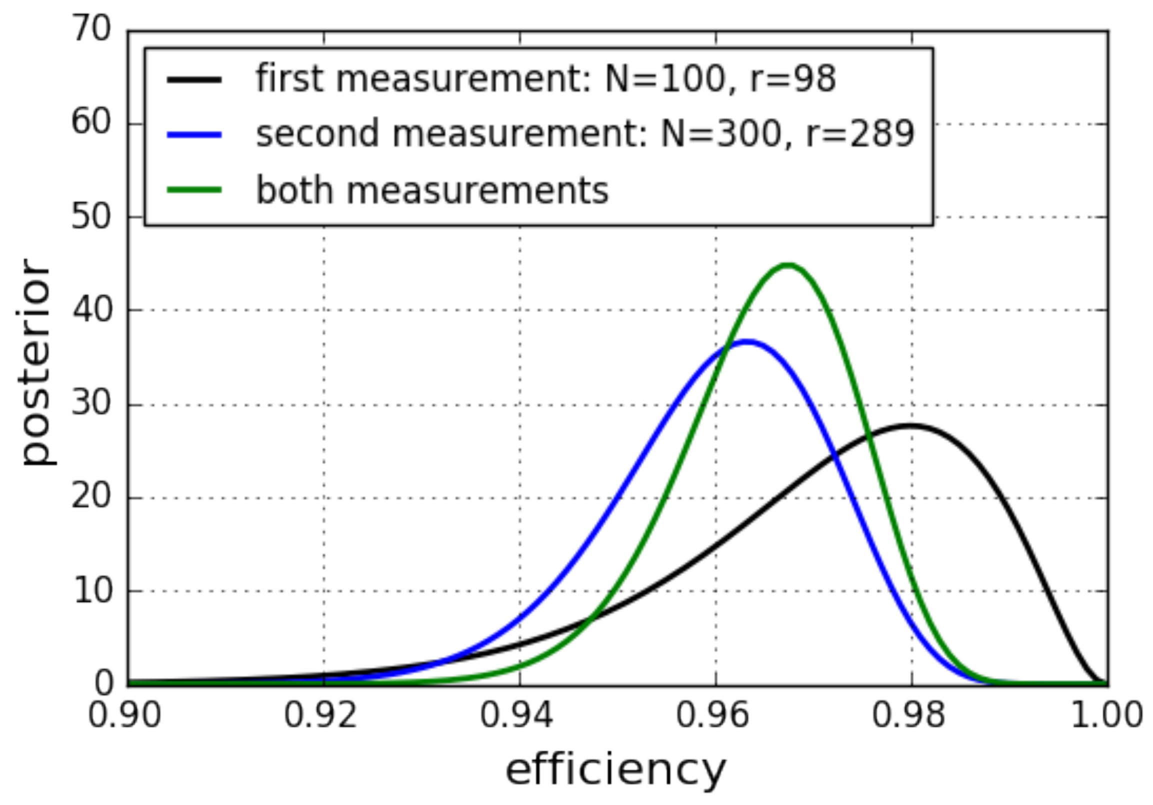
and catching a glimpse of a donkey,

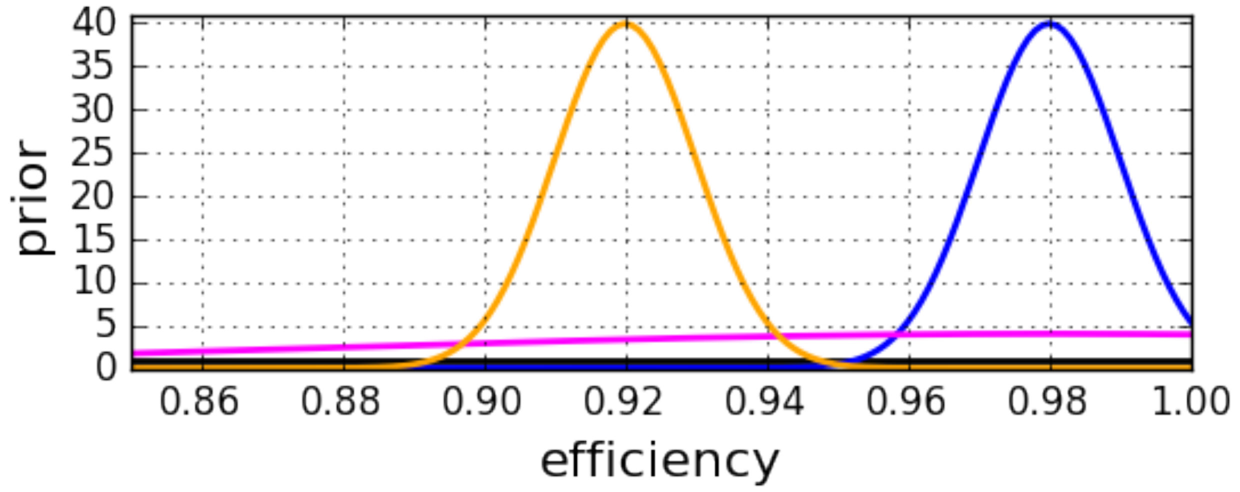
strongly believes he has seen a mule.

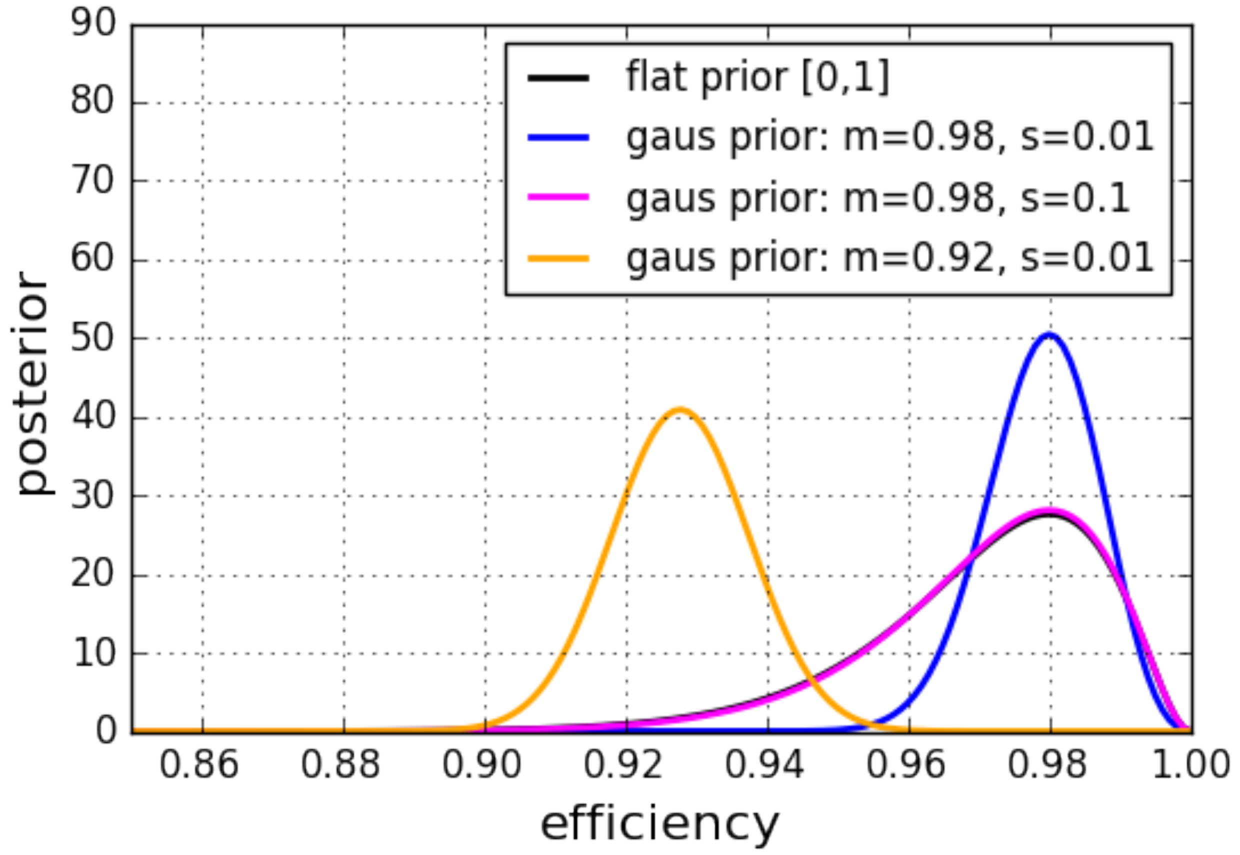


Summary: Prior

- **How to choose a prior?**
- Use past experience if available
- Exclude unphysical regions
- Be careful to correctly implement your degree of uncertainty
- No information: Uninformative prior
  - Flat prior
  - Jeffrey's prior $p_{\text{Jeffrey}}(\theta) \propto \sqrt{\det \mathcal{I}(\theta)}$

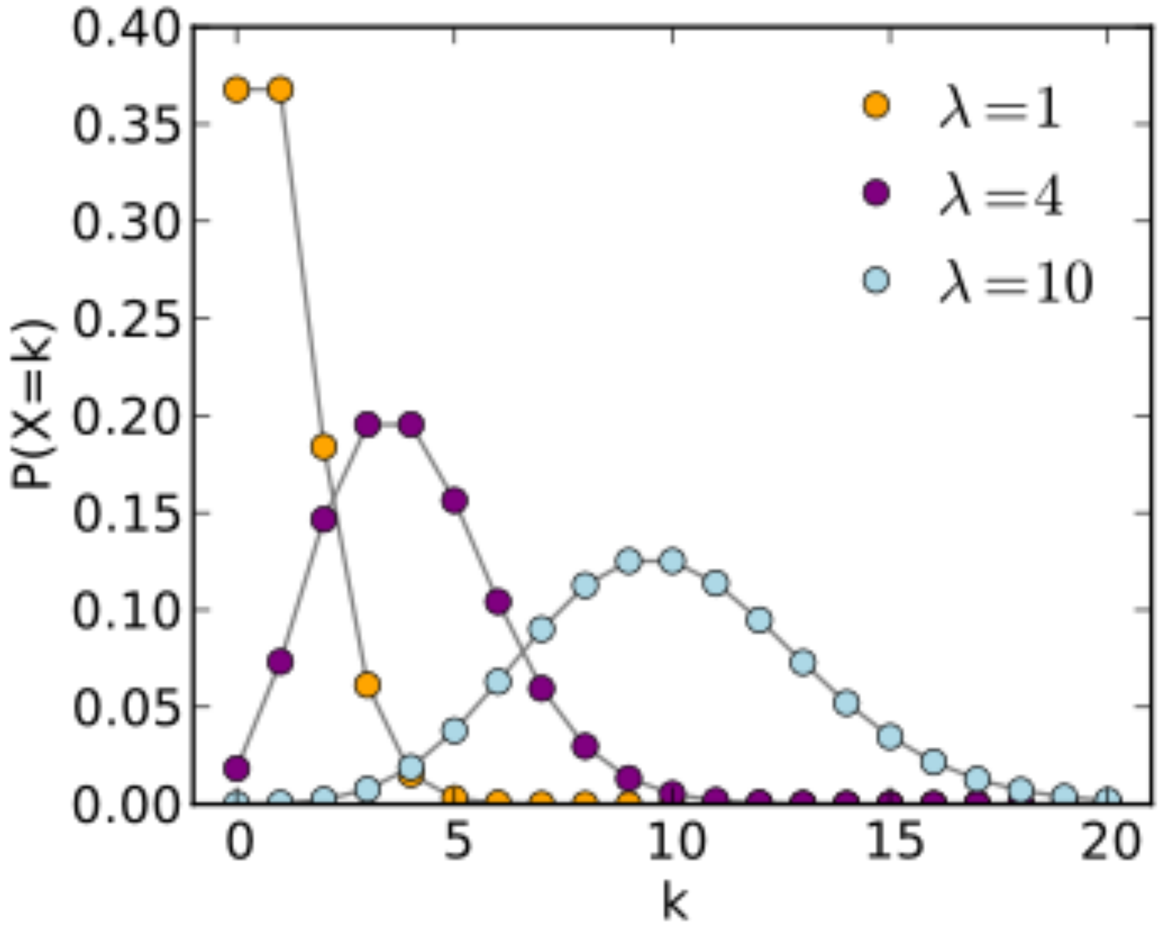




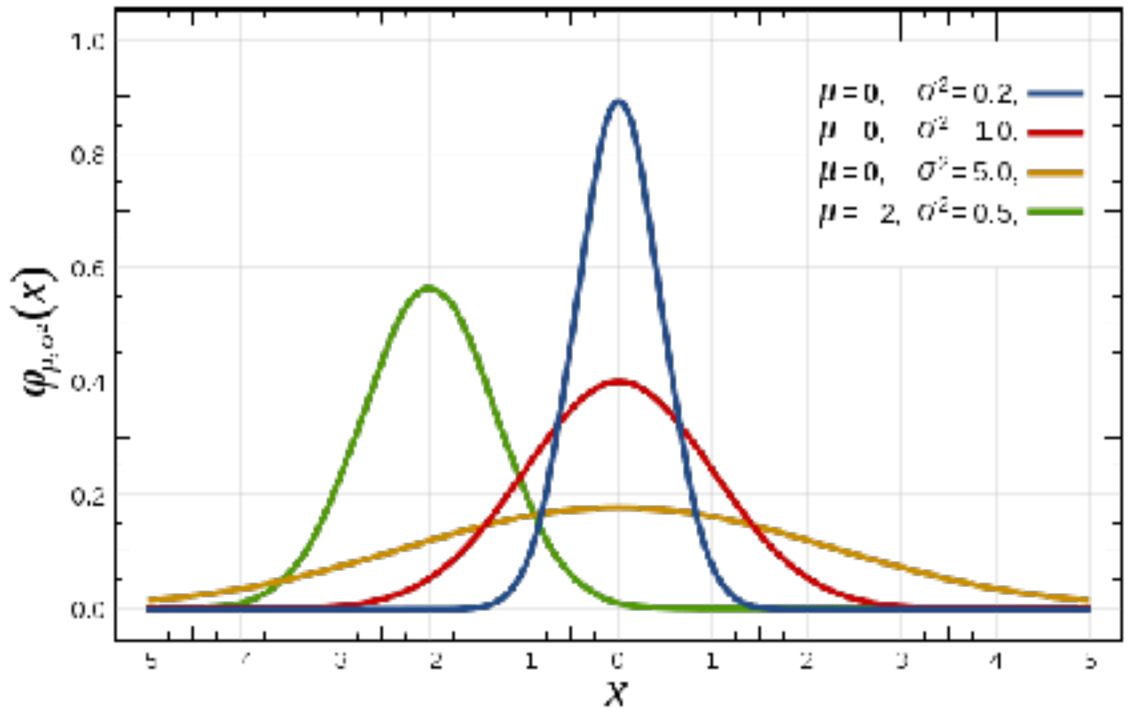


Different Probability Models for the Likelihood

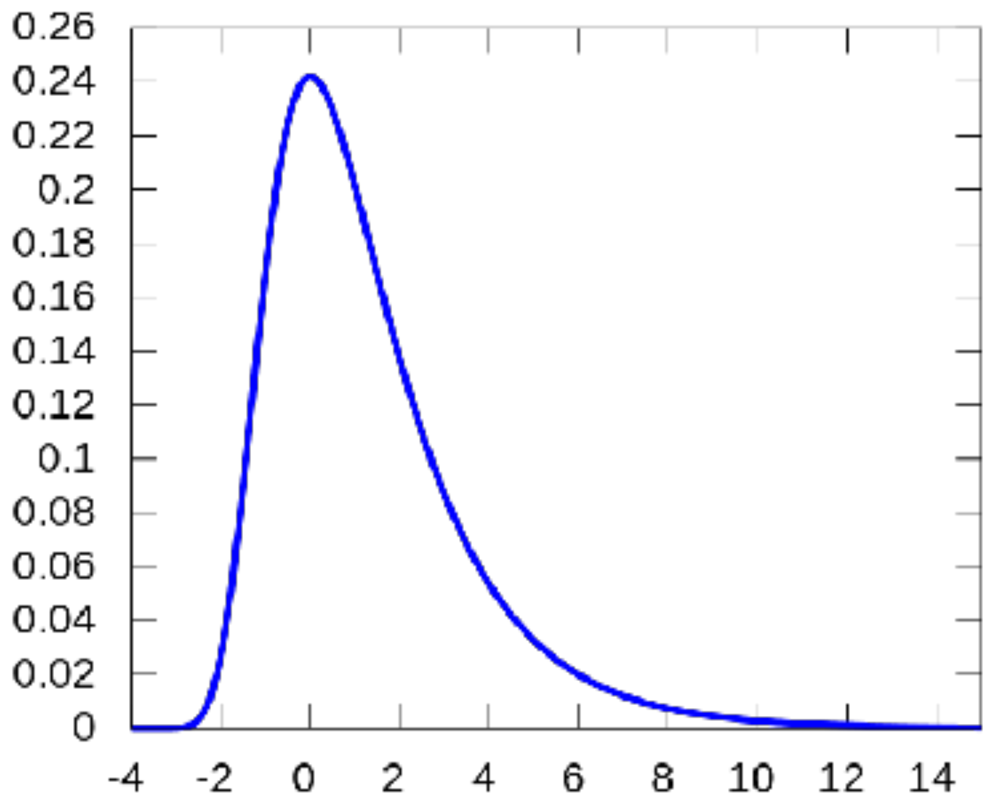
- **How to choose a probability model?**
- Counting *r* "successes" out of *N* trials with a fixed probability: **Binomial**
- Counting „events" with a fixed probability: **Poisson**
- **Normal distribution** (Central Limit Theorem)
- Underlying physics, e.g.:
  - Energy deposition of a minimal ionizing particle in a thin absorber: **Landau**
  - Resonances in high-energy-physics: **Breit-Wigner**
  - **…**

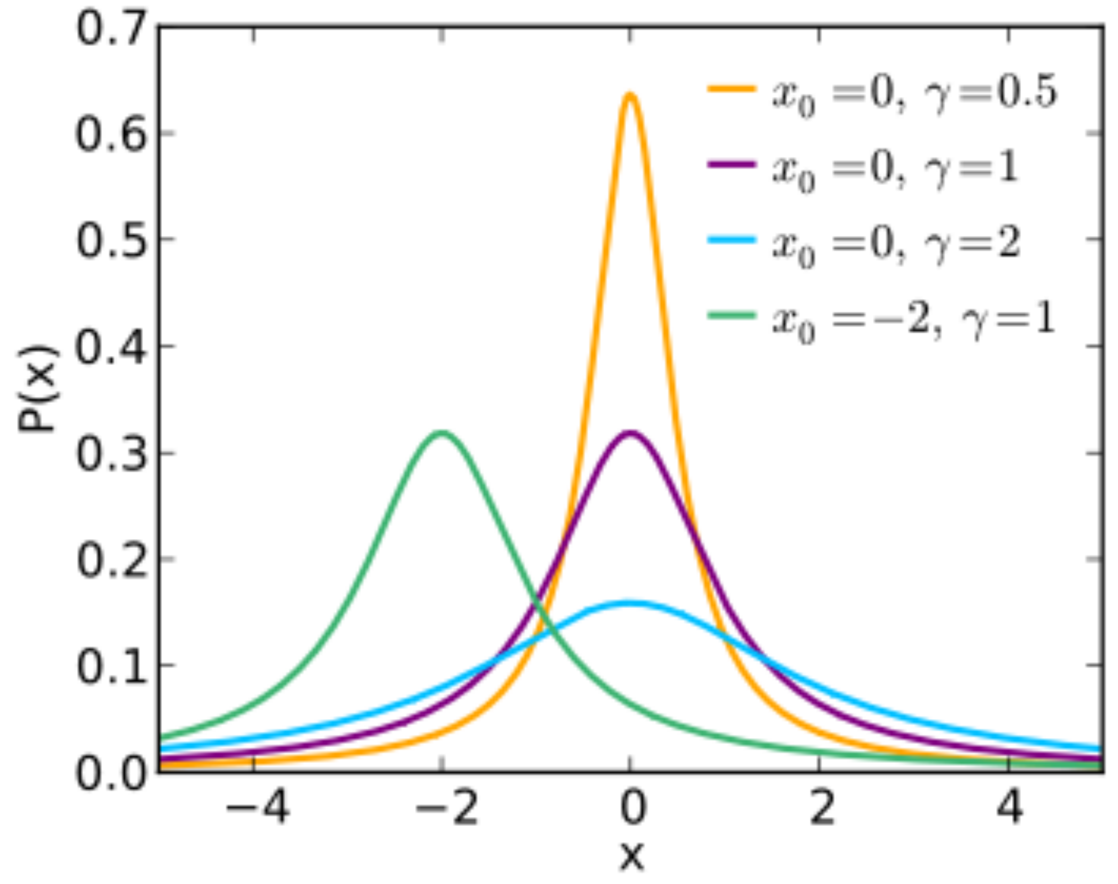


[Poisson Distribution, Wikipedia]



[Normal Distribution, Wikipedia]


[Landau Distribution, Wikipedia]

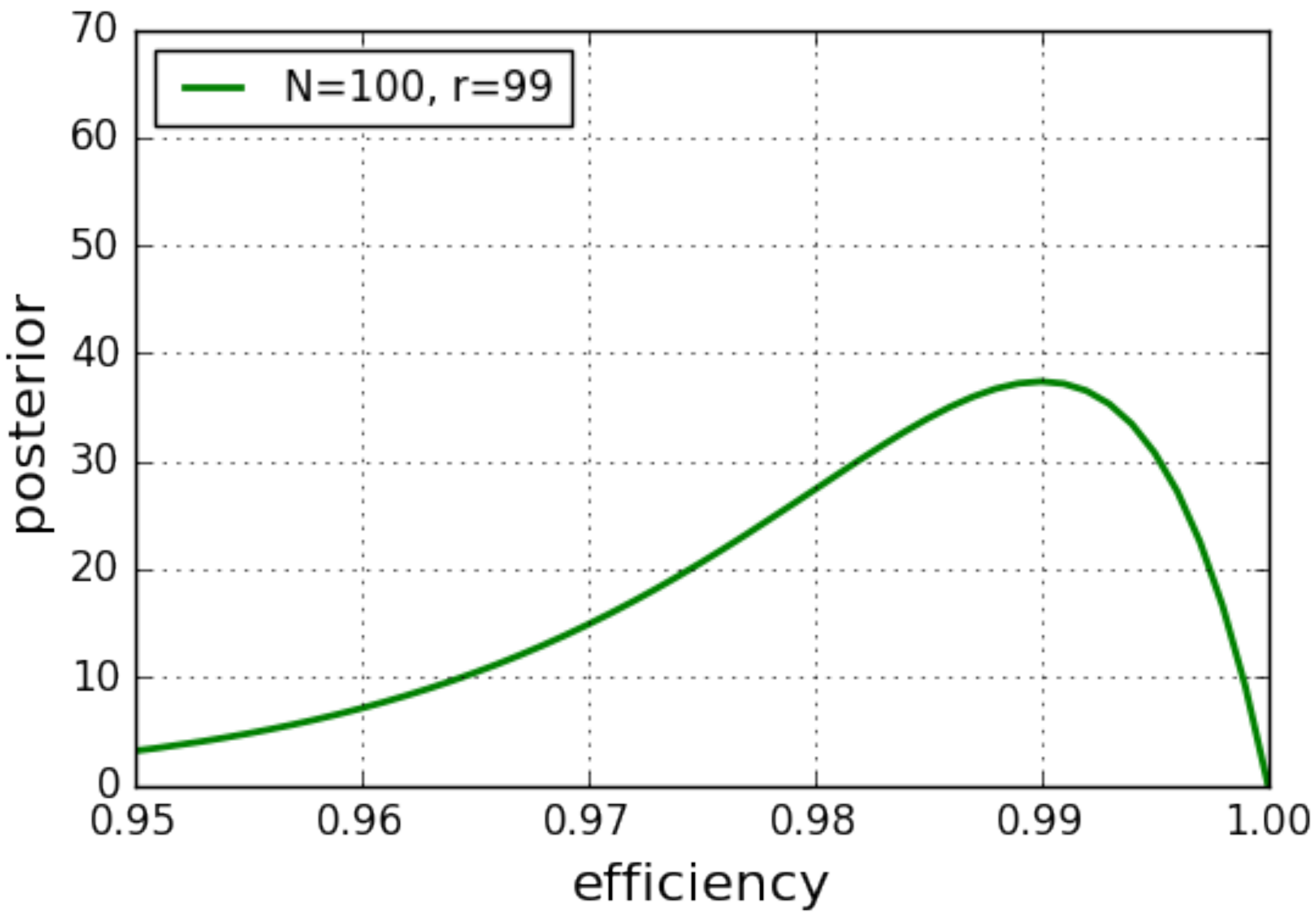


[Breit-Wigner Distribution, Wikipedia]

How to present your Results?

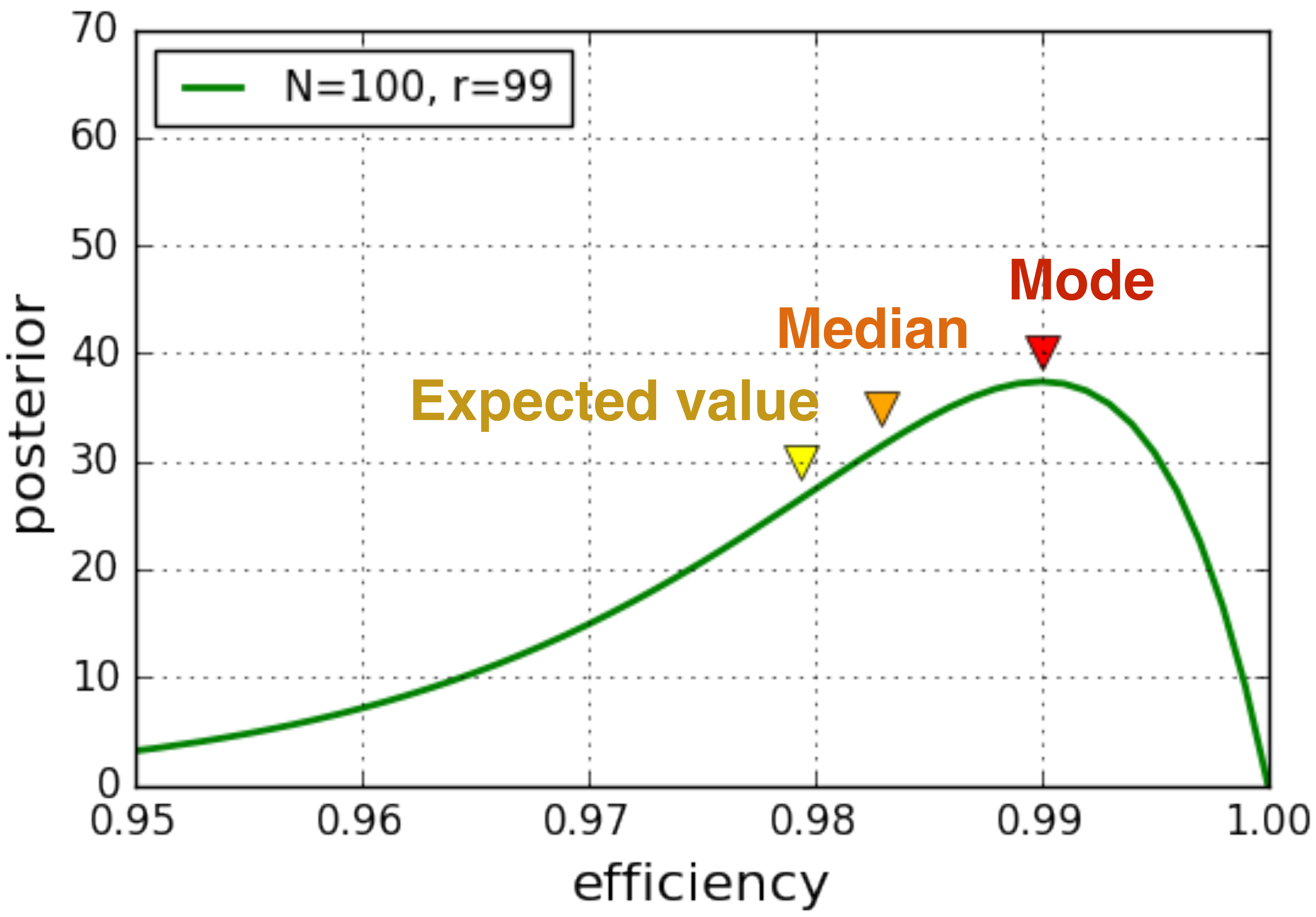
How to Present your Results?

- **Ideally**: Show the whole posterior —> full information
- But how to summarize the posterior?
- Give a point estimate and an interval that includes a certain amount of probability



How to Present your Results?

- **Ideally**: Show the whole posterior —> full information
- But how to summarize the posterior?
- Give a point estimate and an interval that includes a certain amount of probability



Point estimate:

Expected value: $E[X] = \int xp(x)\mathrm{d}x$

Median: $p(X \leq m) \geq \frac{1}{2} \text{ and } p(X \geq m) \geq \frac{1}{2}$

Mode: $\arg\max\ p(x)$

Credible Interval Construction

- Construct an interval to show the width of the distribution
- 1. **Define an α** s.t the interval contains 1 - α of the probability
  popular choices: α = 0.32, α = 0.1, α = 0.05
- 2. **Choose an interval**,
  e.g., +- std. dev., central interval, smallest interval

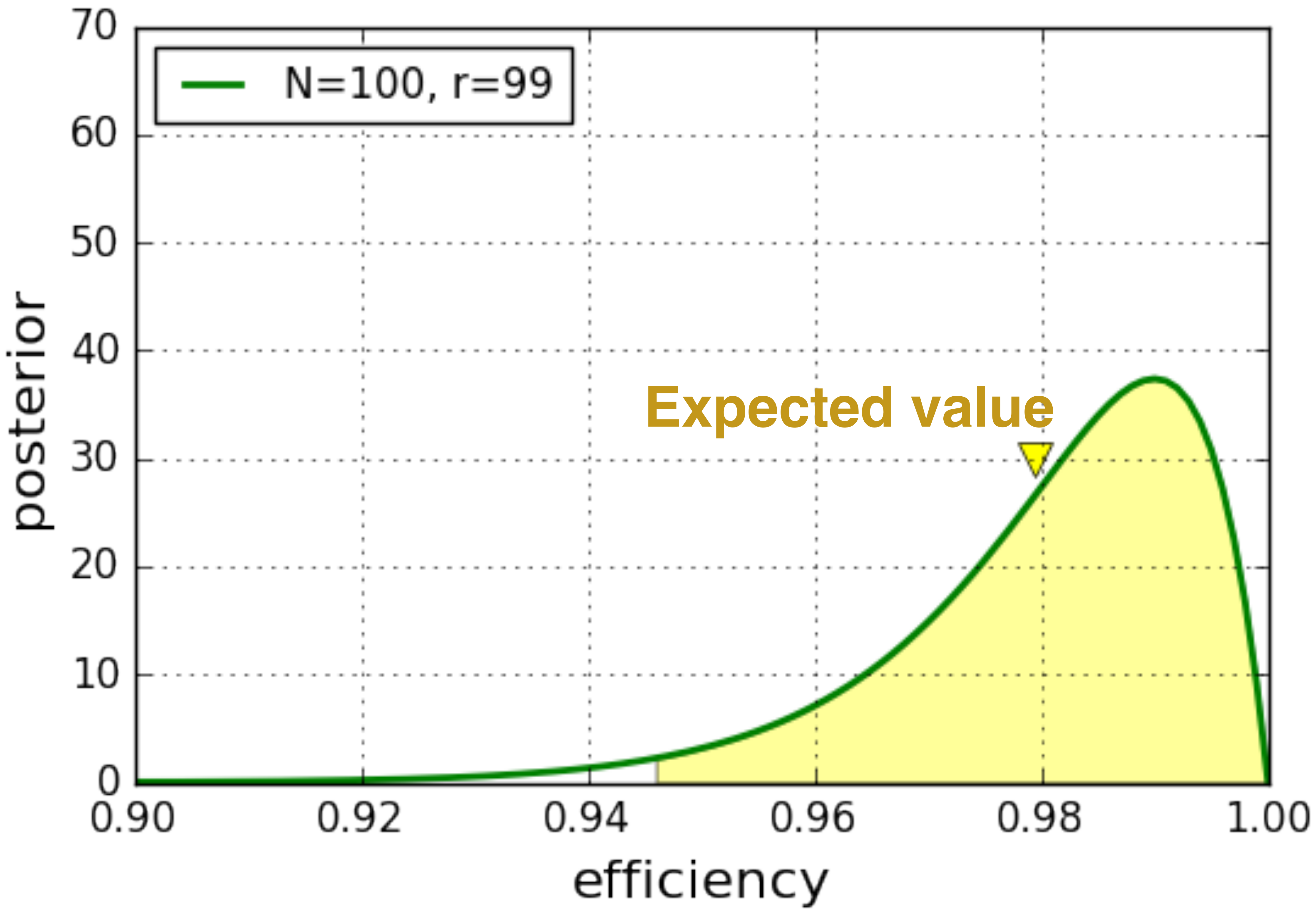
Expected Value & Standard Deviation

- Construct an interval to show the width of the distribution
- 1. **Define an α** s.t the interval contains 1 - α of the probability

  popular choices: $\alpha = 0.32$, $\alpha = 0.1$, $\alpha = 0.05$
- 2. **Choose an interval,**
  e.g., +- std. dev., central interval, smallest interval

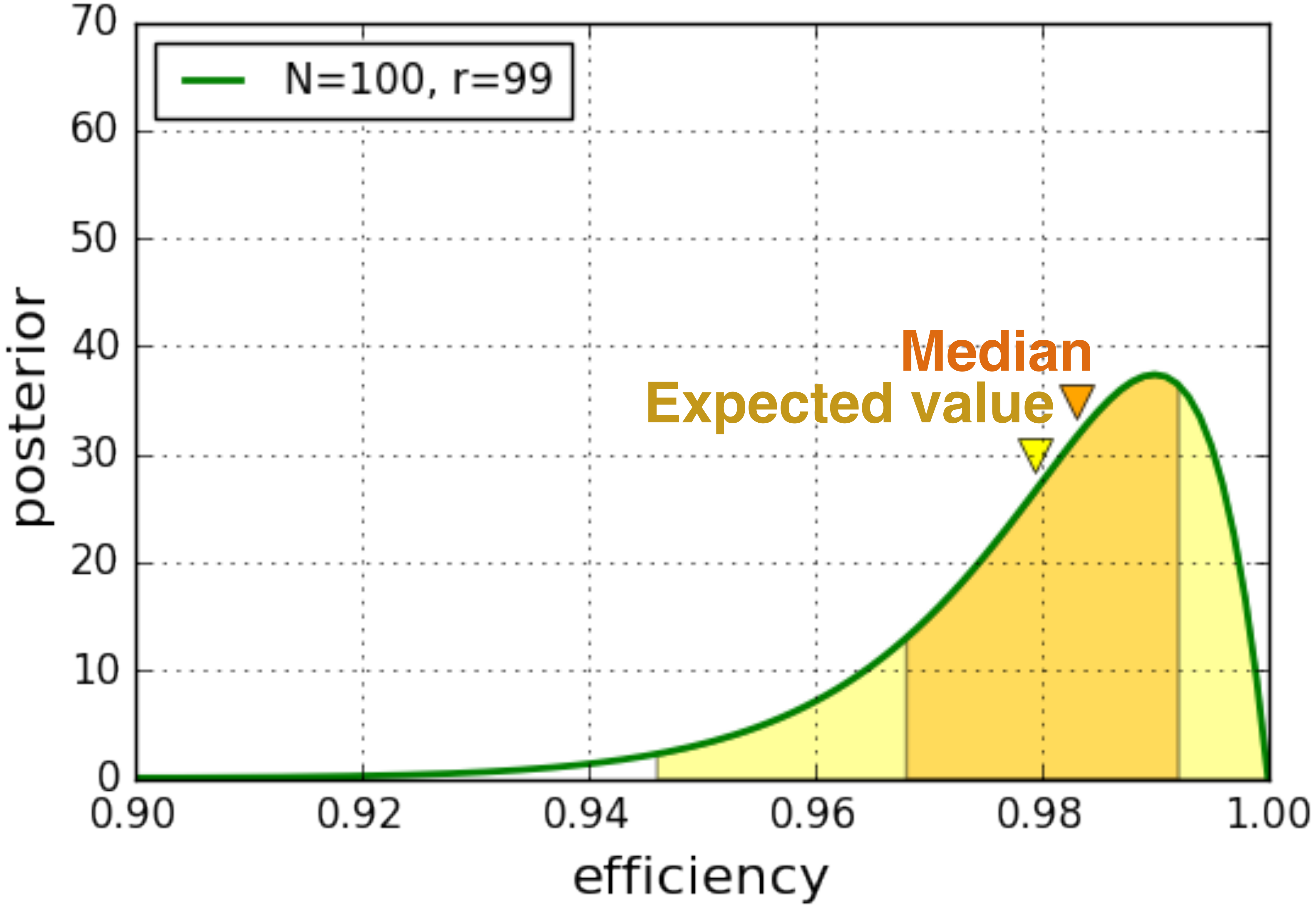
Expected value & std. dev:

$$\sigma = \sqrt{E[X^2] - (E[X])^2}$$



Median & Central Interval

- Construct an interval to show the width of the distribution
- 1. **Define an α** s.t the interval contains 1 - α of the probability
  popular choices: α = 0.32, α = 0.1, α = 0.05
- 2. **Choose an interval**,
  e.g., +- std. dev., central interval, smallest interval



Median & central interval:

Choose smallest x_{min} and largest x_{max} s.t.:

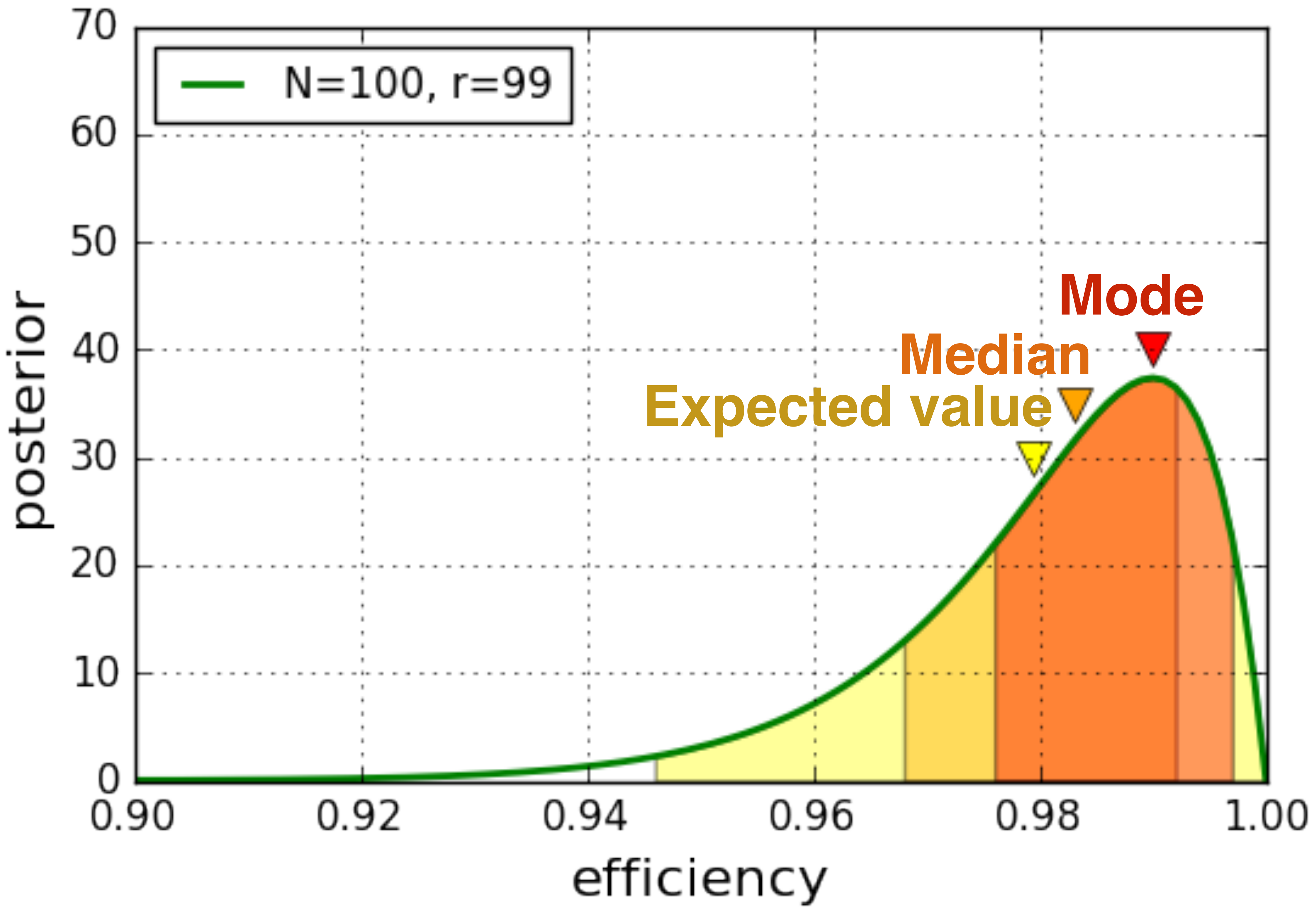
$$p(x < x_{\min}) \leq \alpha/2 \qquad p(x > x_{\max}) \leq \alpha/2$$

While maximizing these probabilities

Mode & Smallest Interval

- Construct an interval to show the width of the distribution
- 1. **Define an α** s.t the interval contains 1 - **α** of the probability

  popular choices: **α** = 0.32, α = 0.1, α = 0.05
- 2. **Choose an interval**,
  e.g., +- std. dev., central interval, smallest interval



Mode & smallest interval:

Choose x_{min} and x_{max} s.t.:

$$p(x_{\mathrm{min}} > x > x_{\mathrm{max}}) \geq 1 - \alpha$$

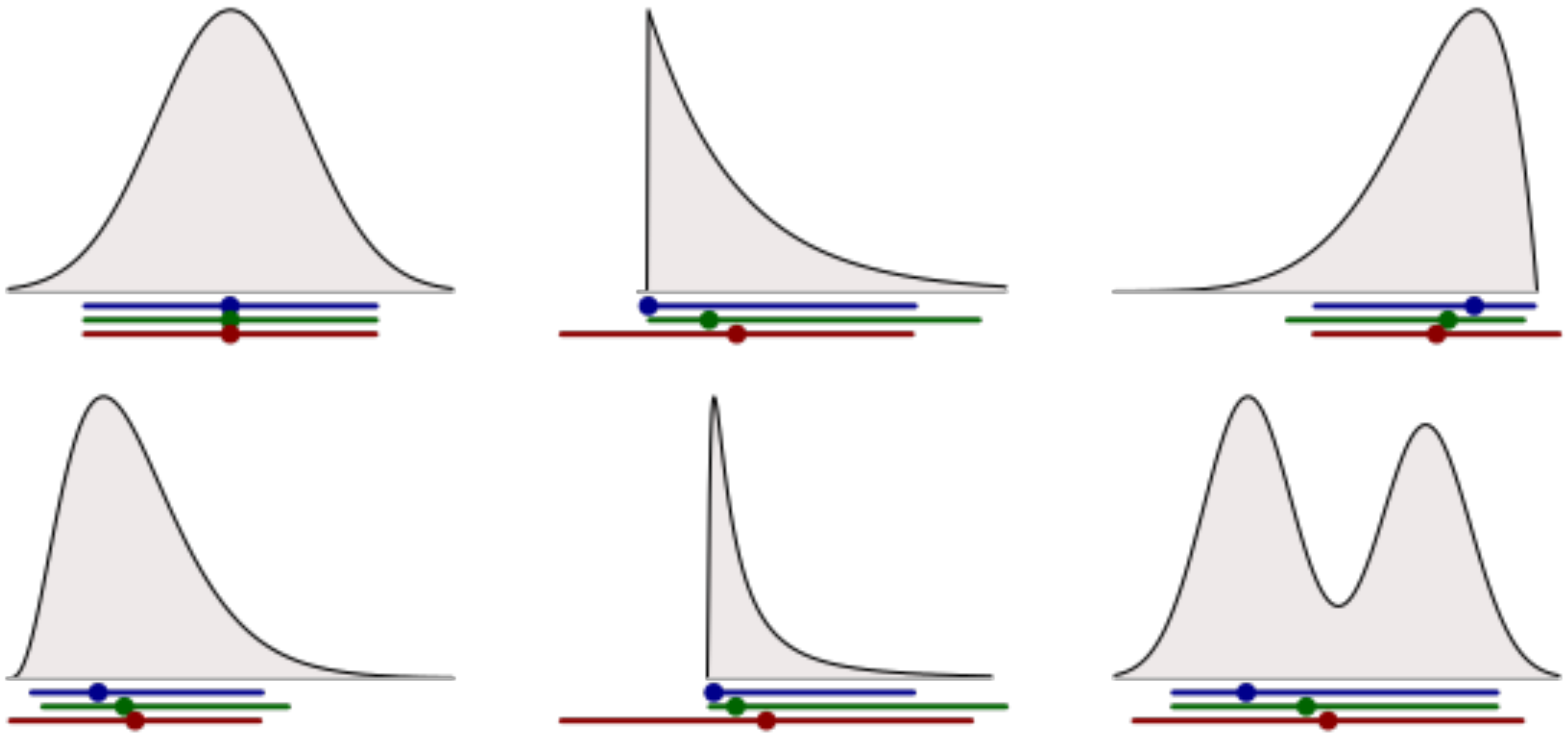
With $p(x_{\mathrm{min}}) = p(x_{\mathrm{max}})$

Interval Construction

Mode & smallest int.

Median & central int.

Mean & std. dev.

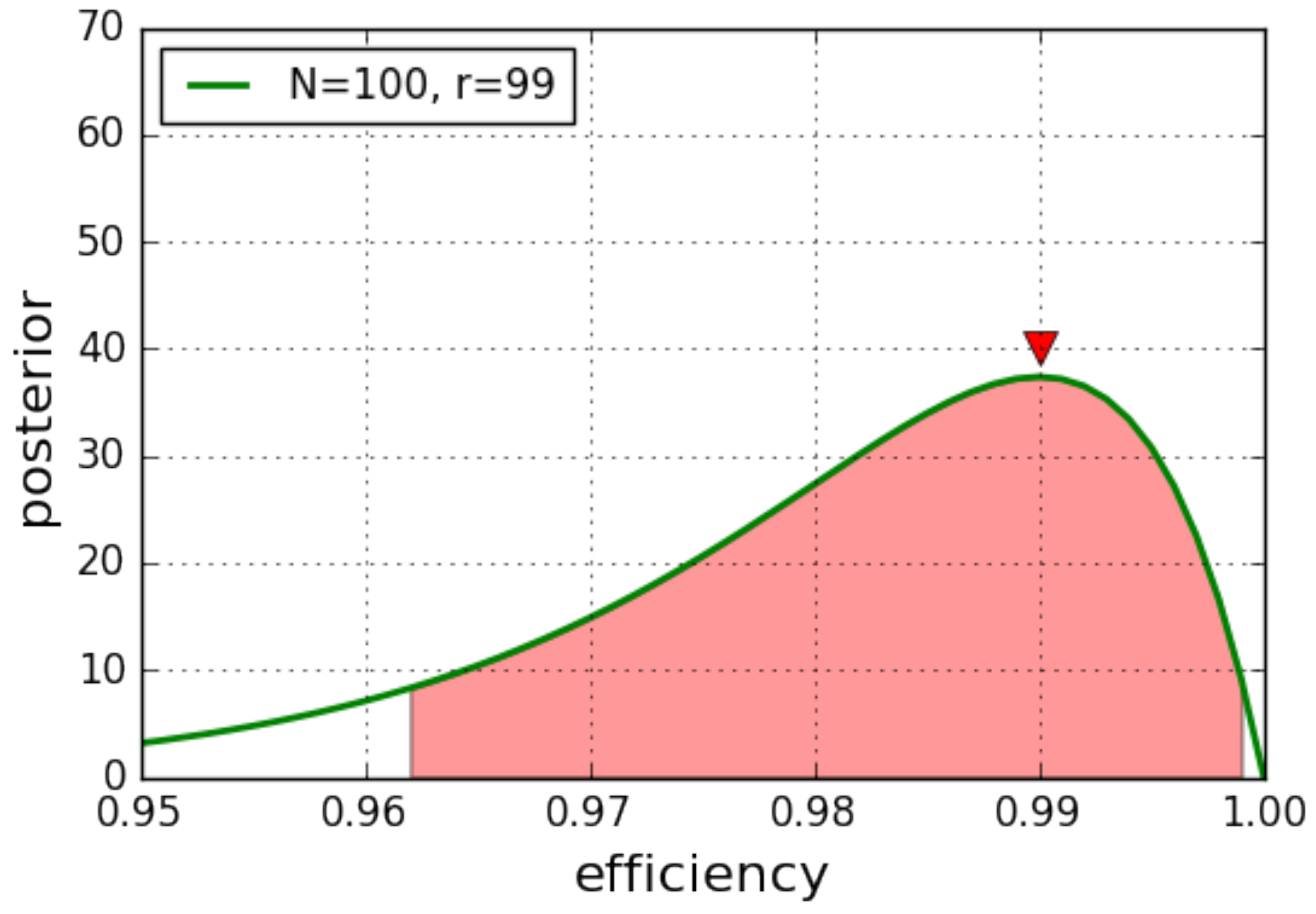


[Rasmus Bååth, www.sumsar.net]

- All three variables and intervals are the same in the gaussian case
- Mean and std. dev. are misleading for skewed distributions
- Central interval is always >= smallest interval
- Nothing really works well in the last case

Comparison to Frequentist Confidence Interval

- **Bayesian view:** Posterior is a probability distribution of the unknown parameter
- For a 90% credible interval, the unknown parameter lies within *this interval* with a probability of 90%



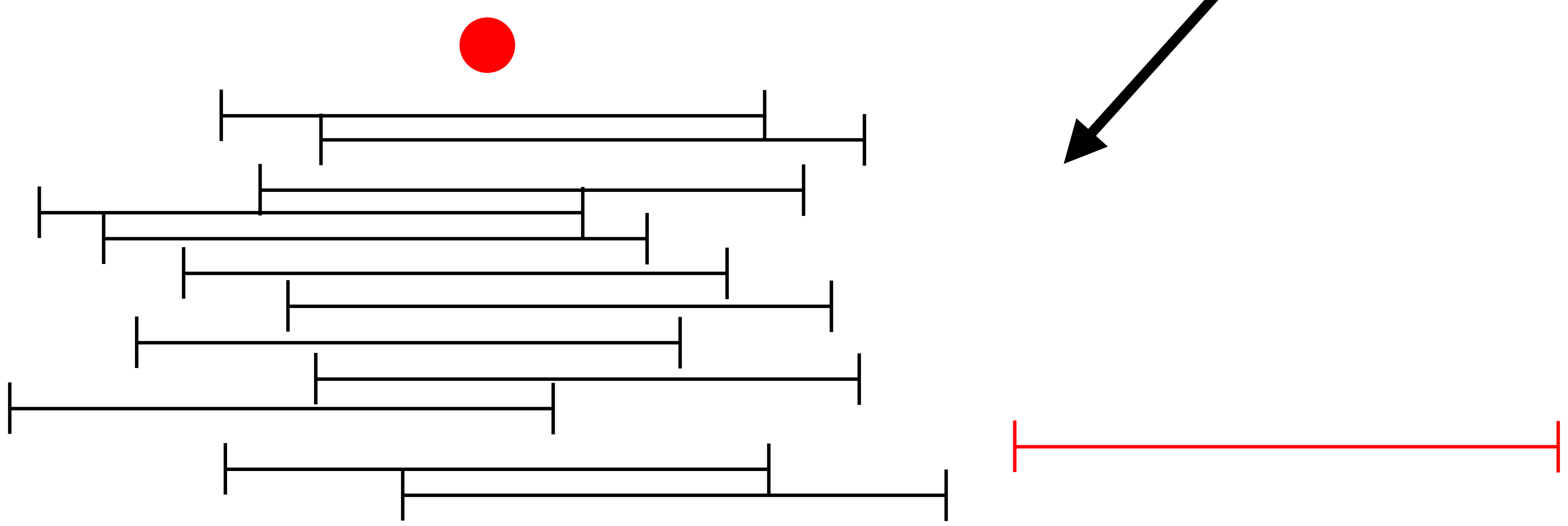
Credible interval, bayesian

Comparison to Frequentist Confidence Interval

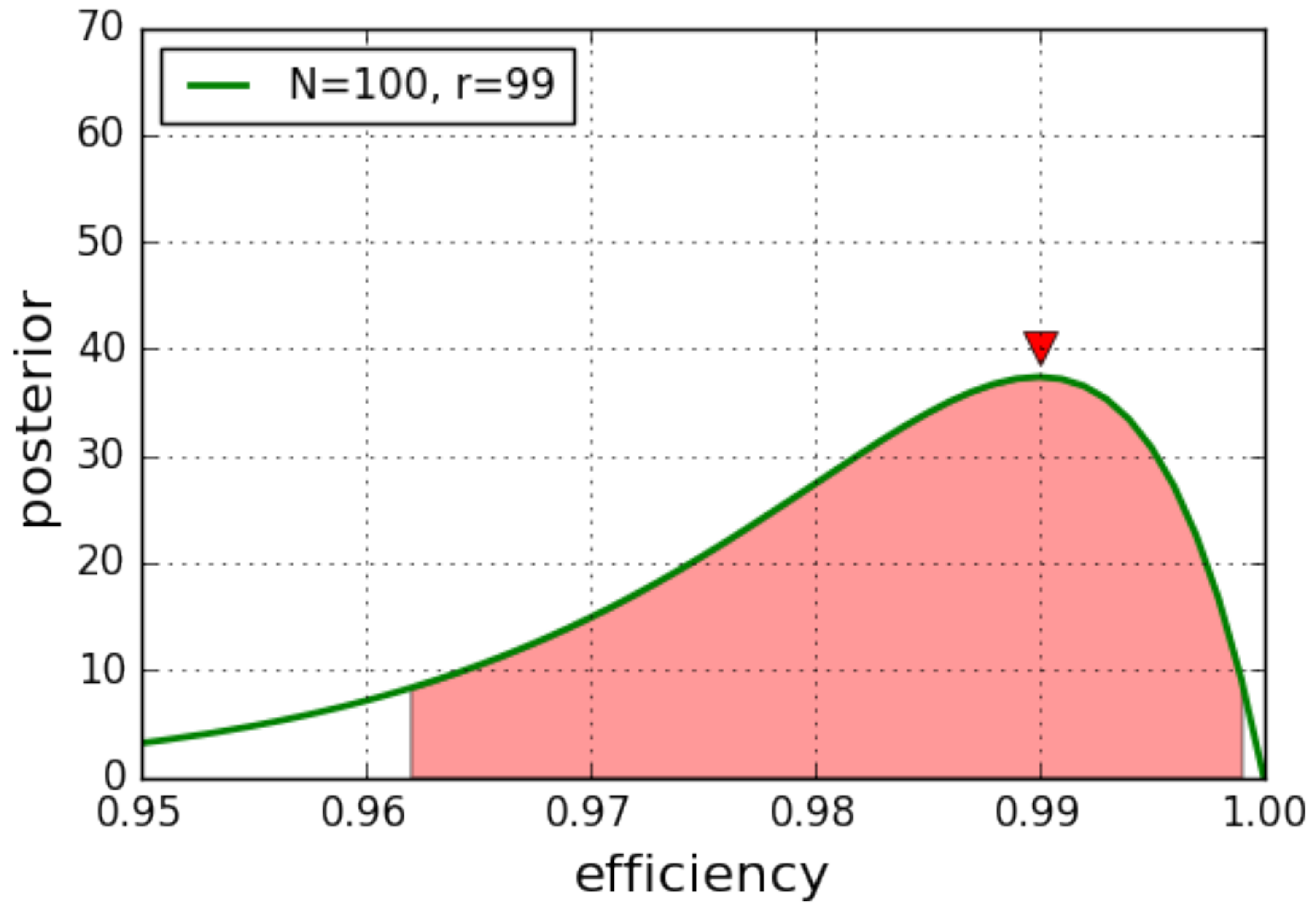
- **Bayesian view:** Posterior is a probability distribution of the unknown parameter
- For a 90% credible interval, the unknown parameter lies within *this interval* with a probability of 90%

- **Frequentist view:** Estimate of the unknown parameter is fix
- 90% of the constructed confidence intervals include *this estimate* of the unknown parameter

Pay attention how to interpret!



Confidence interval, frequentist



Credible interval, bayesian

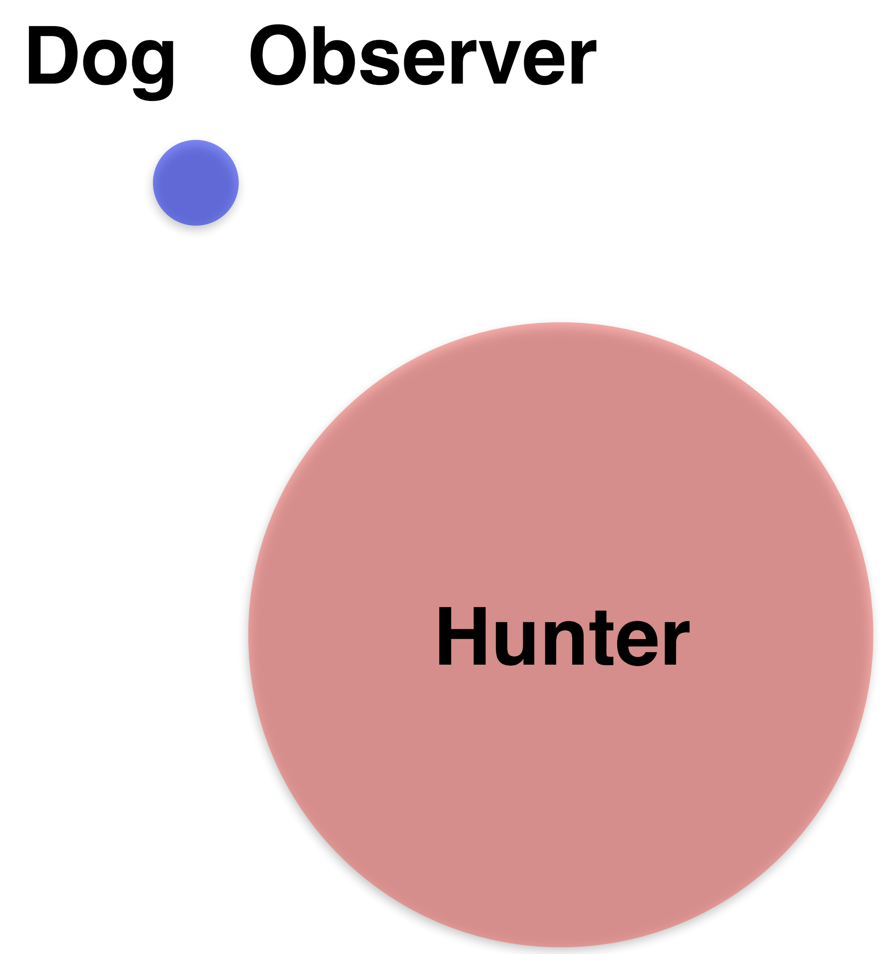
Comparison to Frequentist Confidence Interval

- The dog has a 90% probability of being 100m from the hunter $p(\mathrm{data}|H)$
- If we observe the dog, what can we say about the hunter?
- **Analogy:** Hunter <—> true value; Dog <—> observable
- *"The hunter is, with 90% probability, within 100m of the dog“* $p(H|\mathrm{data})$

[Example adapted from: Giulio D’Agostini, Bayesian Reasoning in Data Analysis]

Comparison to Frequentist Confidence Interval

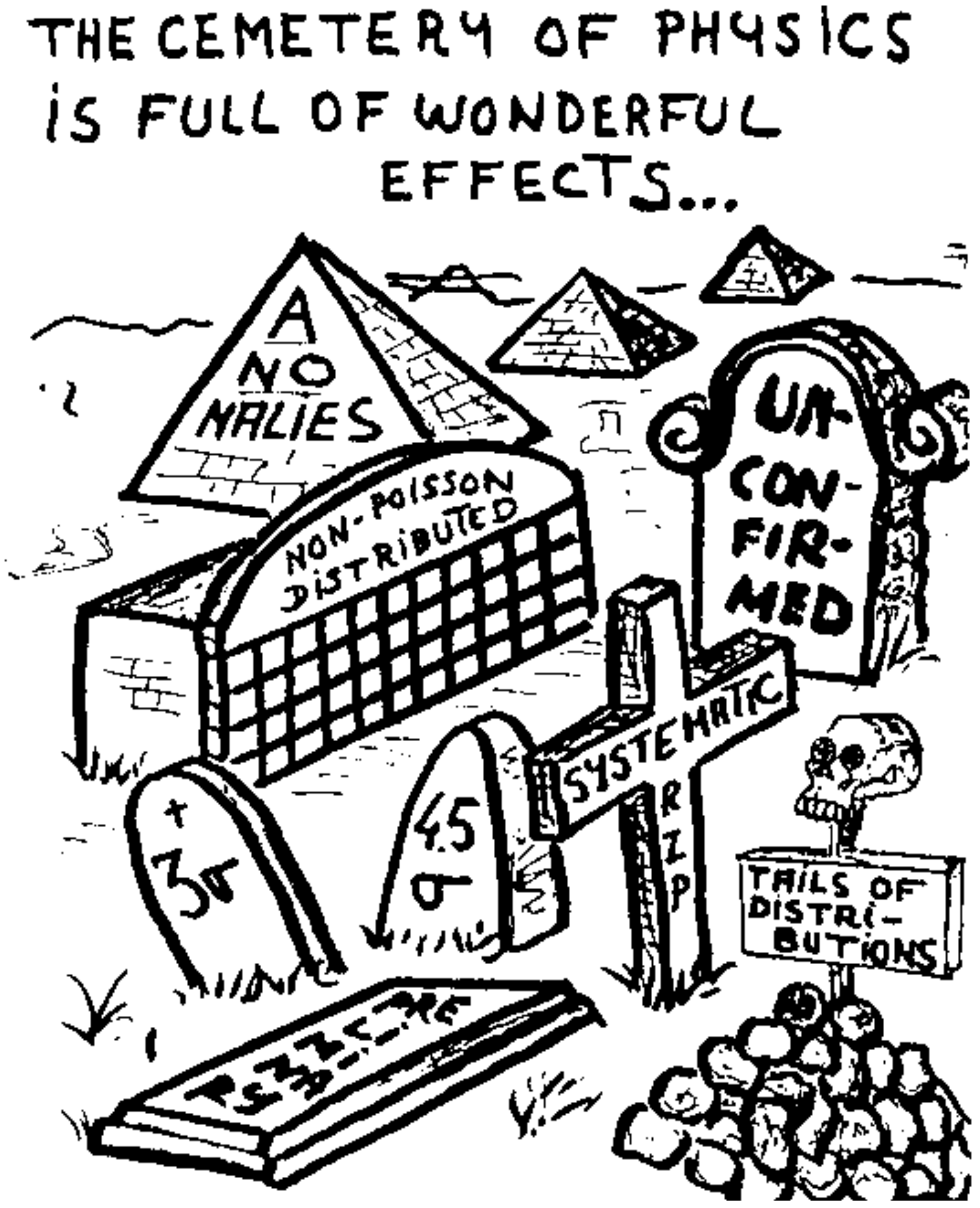
- The dog has a 90% probability of being 100m from the hunter $p(\mathrm{data}|H)$
- If we observe the dog, what can we say about the hunter?
- **Analogy:** Hunter <—> true value; Dog <—> observable
- *"The hunter is, with 90% probability, within 100m of the dog“* $p(H|\mathrm{data})$
- **Only works if*:***
  - The hunter can be anywhere around the dog
  - The dog has no preferred direction of arrival at the point where we observe him



[Example adapted from: Giulio D’Agostini, Bayesian Reasoning in Data Analysis]

Comparison to Frequentist Confidence Interval

- The dog has a 90% probability of being 100m from the hunter $p(\text{data}|H)$
- If we observe the dog, what can we say about the hunter?
- **Analogy:** Hunter <—> true value; Dog <—> observable
- *"The hunter is, with 90% probability, within 100m of the dog"* $p(H|\text{data})$
- **Only works if:**
  - The hunter can be anywhere around the dog
  - The dog has no preferred direction of arrival at the point where we observe him
- **Real world "deal-breakers":**
  - Closeness to physical boarders
  - Asymmetries in physical quantity or observable
  - Prior knowledge

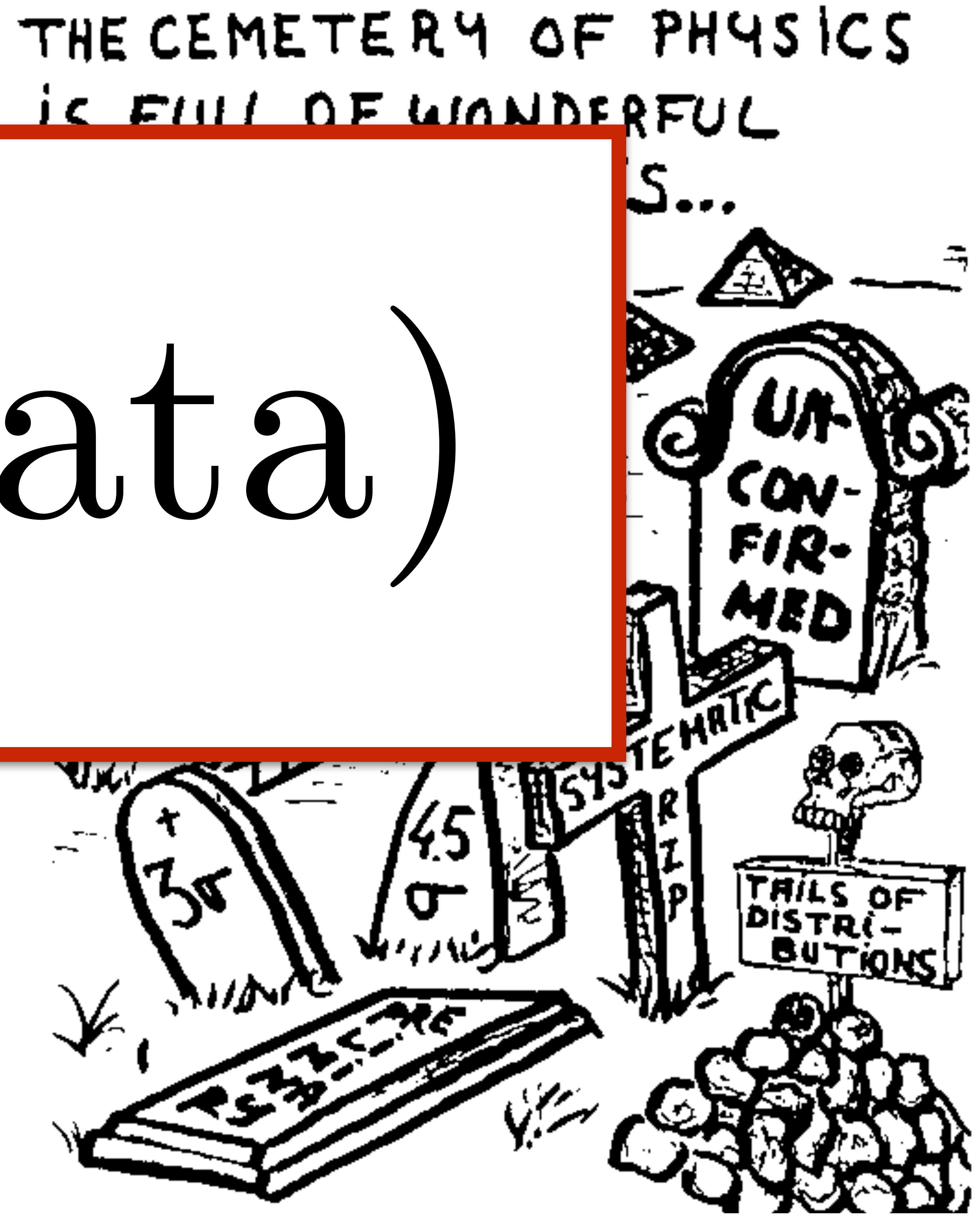


[De Rujula, Snapshot of the 1985 high energy physics panorama]

Comparison to Frequentist Confidence Interval

- The dog has a 90% probability of being 100m from the hunter $p(\text{data}|H)$
- If we observe the dog, what can we say about the hunter?
- **Analogy:** Hunter <—> true value; Dog <—> observable
- *"The hunter is, with 90% probability, within 100m of the dog“* $p(H|\text{data})$
- **Only w**
  - The
  - The
    wher

$$p(\text{data}|H) \neq p(H|\text{data})$$



- **Real world "deal-breakers“:**
  - Closeness to physical boarders
  - Asymmetries in physical quantity or observable
  - Prior knowledge

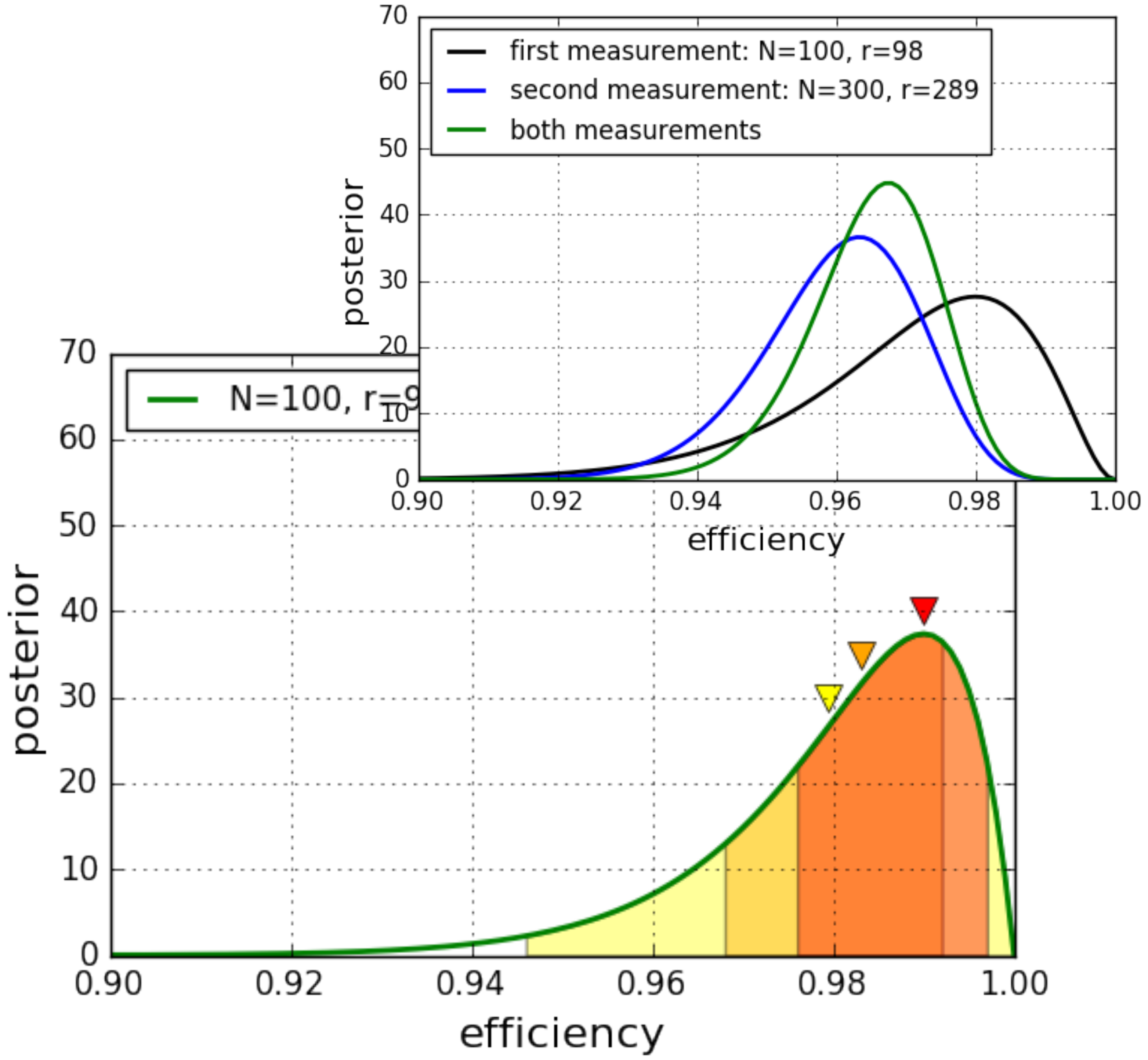
[De Rujula, Snapshot of the 1985 high energy physics panorama]

Summary

- **Frequentist probability:** Limit of its relative frequency in a large number of trails
- **Bayesian probability:** Degree of belief
- Bayesian analysis gives access to full probability distribution of unknown quantity
- Prior is important and may introduce subjectivity
- Frequentist definition of probability may cause misinterpretations

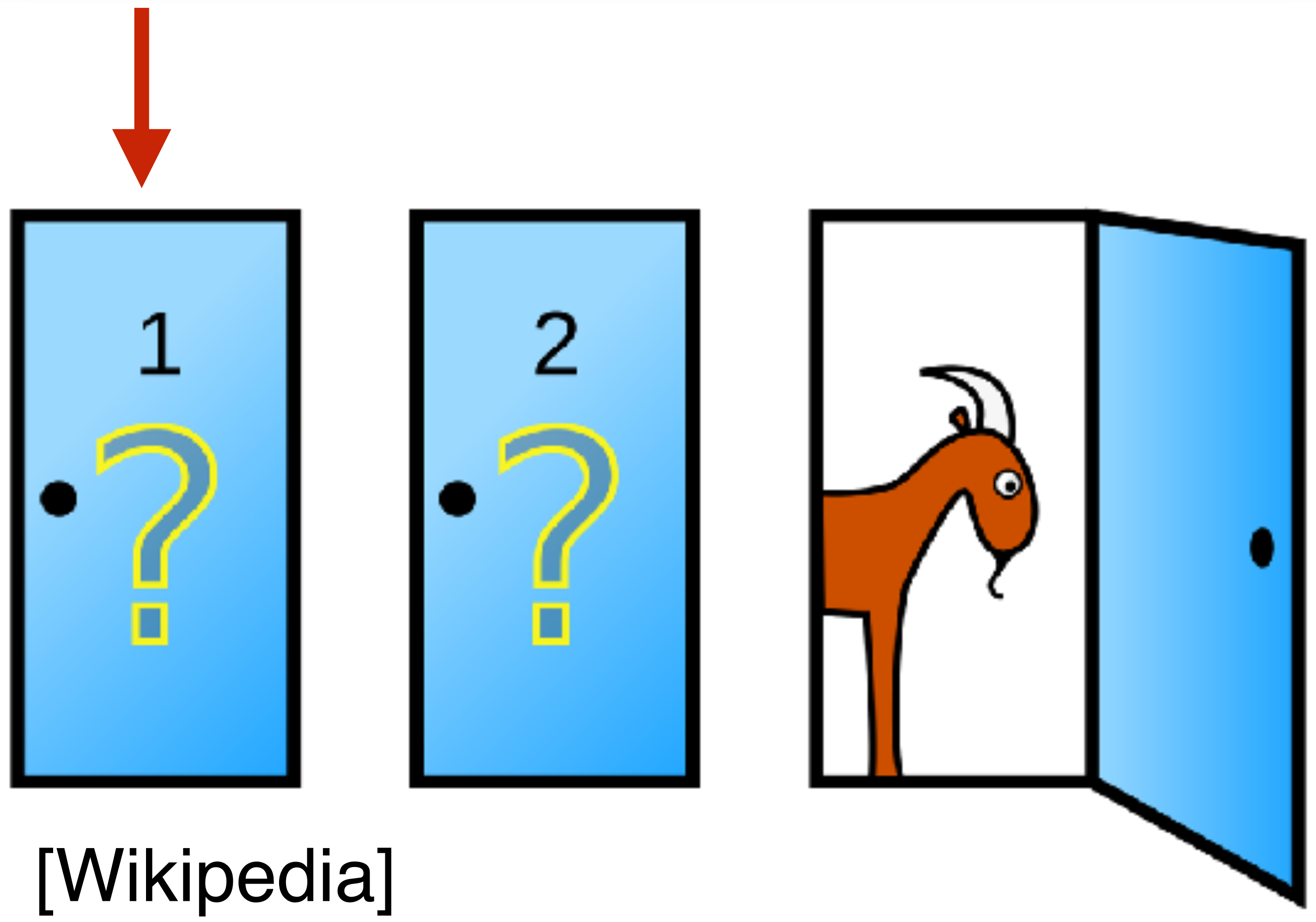
$$p(\mathrm{data}|H) \neq p(H|\mathrm{data})$$

$$p(H|\mathrm{data}) = \frac{p(\mathrm{data}|H)\,p(H)}{p(\mathrm{data})}$$



Monty Hall Problem

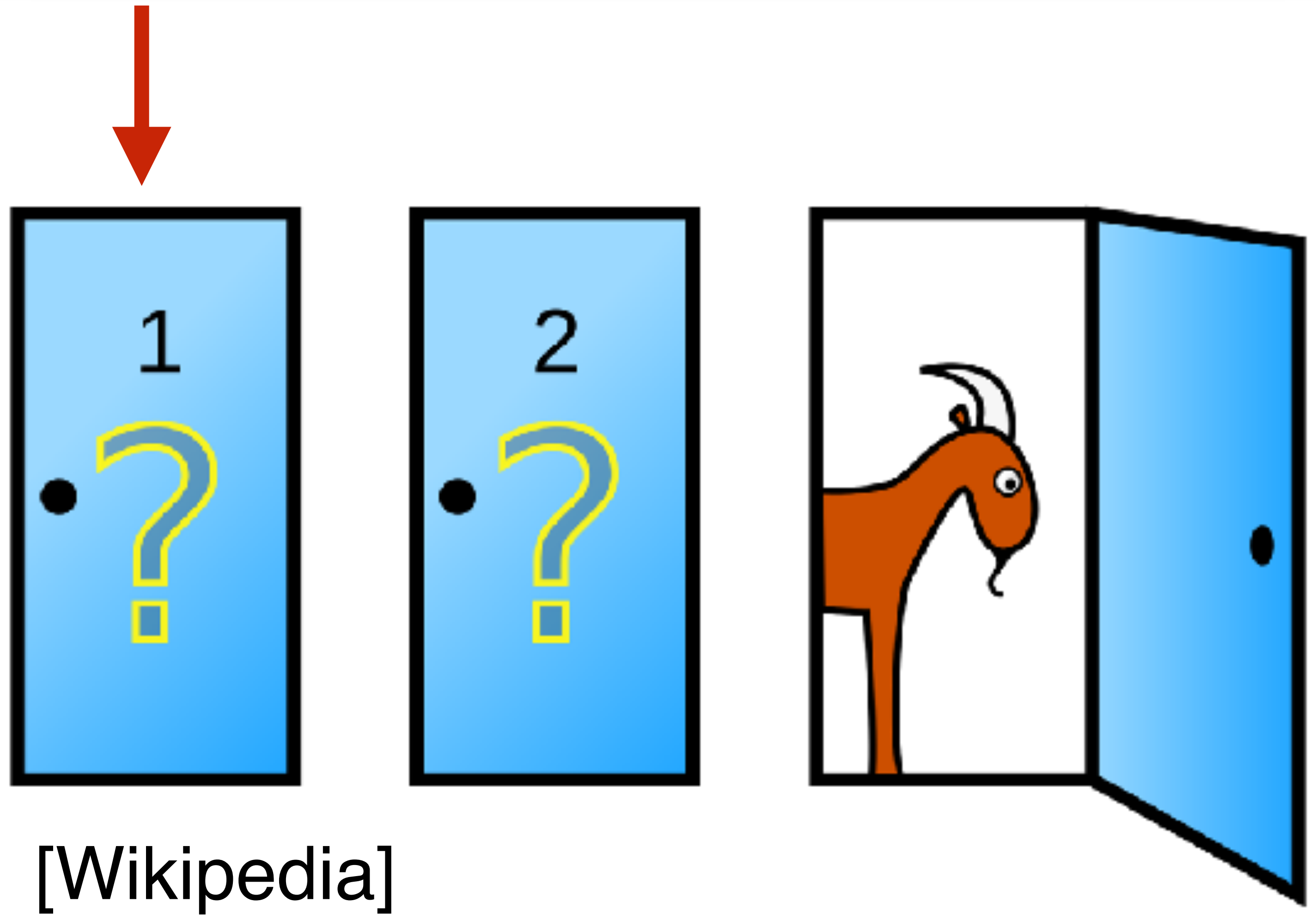
- You are in a game show.
- You need to pick **one of three doors**
- Behind one of the doors there is a **car**, the other two doors give you a **goat**
- After picking a door, the game master **opens a door that has a goat** behind it (no matter what door you chose)
- He offers you to **change your choice** and to take the other unopened door
- Should you take the deal?
- Bob argues: *„After the host has opened door 3, the probability of the car being behind door 1 or door 2 is equal (p = 1/2) and thus it does not matter which door to choose“*



[Wikipedia]

Monty Hall Problem

- You are in a game show.
- You need to pick **one of three doors**
- Behind one of the doors there is a **car**, the other two doors give you a **goat**
- After picking a door, the game master **opens a door that has a goat** behind it (no matter what door you chose)
- He offers you to **change your choice** and to take the other unopened door
- Should you take the deal?
- Bob argues: „*After the host has opened door 3, the probability of the car being behind door 1 or door 2 is equal (p = 1/2) and thus it does not matter which door to choose*“



[Wikipedia]

Next lecture: hypothesis testing, model fitting, other applications

Thank you