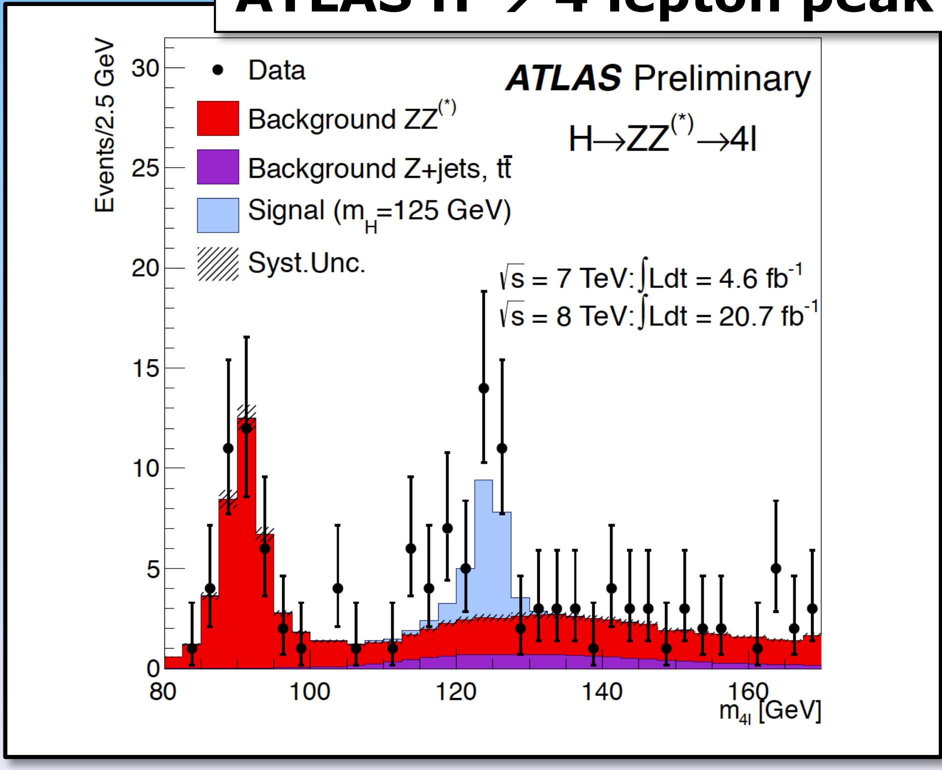
Why does RooFit put asymmetric errors on data points ?

Ivo van Vulpen
(UvA/Nikhef)

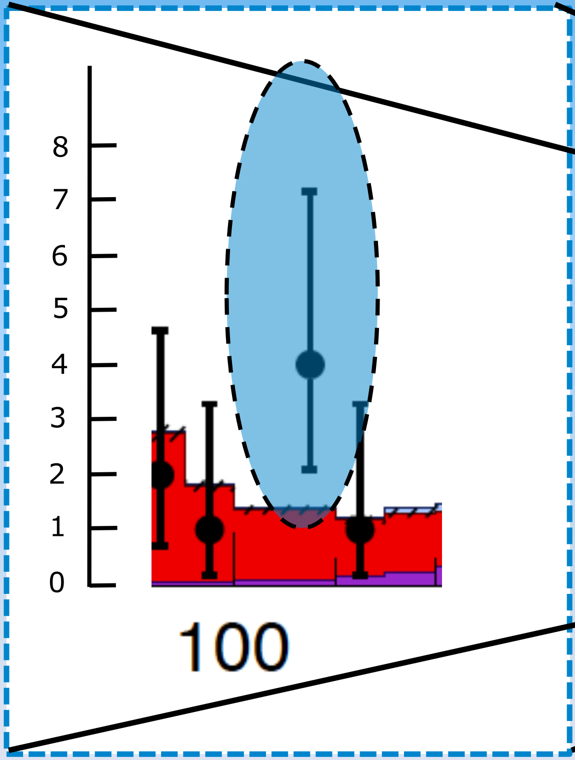
ATLAS H → 4 lepton peak

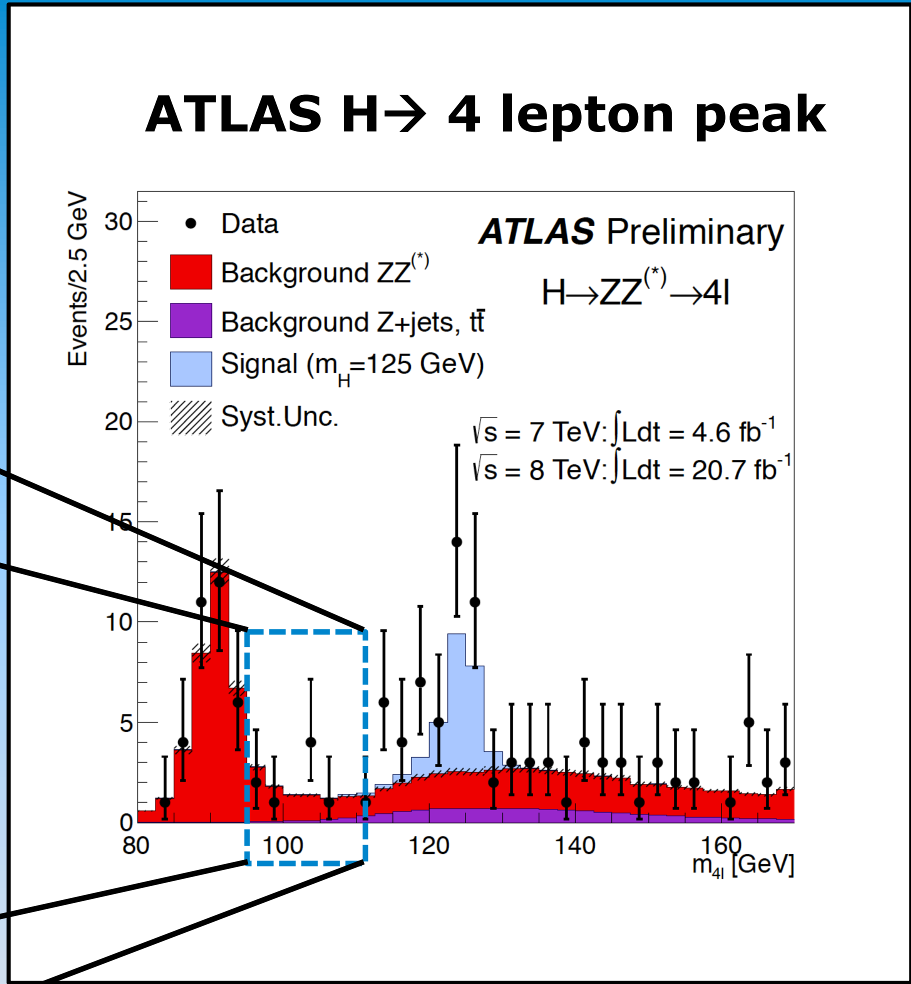


10 slides on a 'too easy' topic that I hope confuse, but do not irritate you

http://www.nikhef.nl/~ivov/Statistics/PoissonError/BobCousins_Poisson.pdf

Why put an error on a data-point anyway ?





- Summarize measurement
- Make statement on underlying true value

I'll present 5 options. You tell me which one you prefer

Known λ (Poisson)

Binomial with n→∞, p →0 en np=λ

$$P(n\,|\,\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

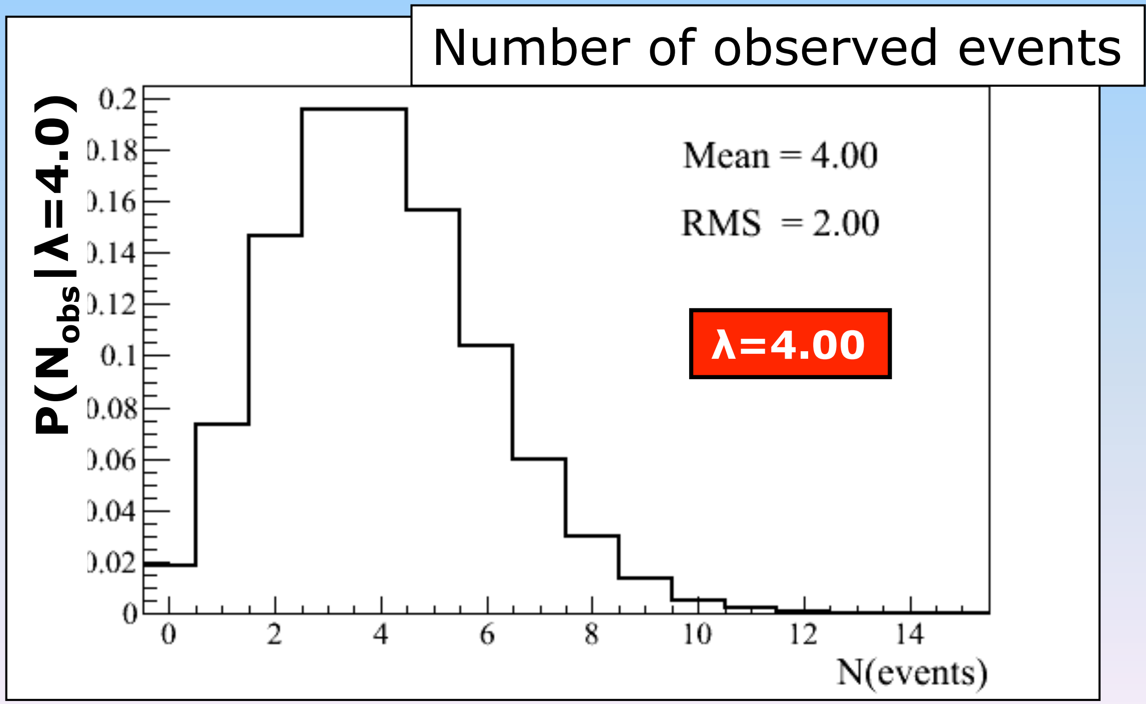
Poisson distribution

Probability to observe n events when λ are expected

$$P(0\,|\,4.0) = 0.01832$$
$$P(2\,|\,4.0) = 0.14653 \quad !$$
$$P(3\,|\,4.0) = 0.19537$$
$$P(4\,|\,4.0) = 0.19537$$
$$P(6\,|\,4.0) = 0.10420 \quad !$$

#observed → **varying**

λ hypothesis → **fixed**



Known λ (Poisson)

Binomial with n→∞, p →0 en np=λ

$$P(n\,|\,\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

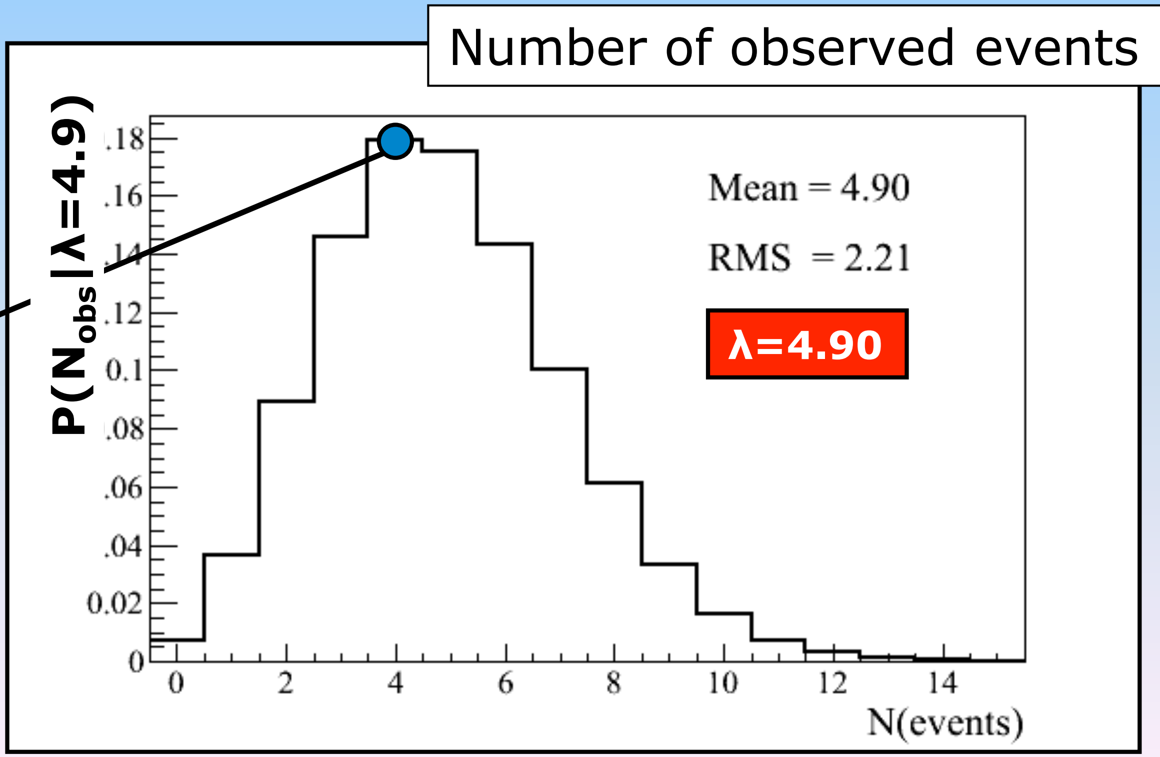
Poisson distribution

Probability to observe n events when λ are expected

$$P(0\,|\,4.9) = 0.00745$$
$$P(2\,|\,4.9) = 0.08940$$
$$P(3\,|\,4.9) = 0.14601$$
$$P(4\,|\,4.9) = 0.17887$$

#observed — **varying**

λ hypothesis — **fixed**



Option 1: Poisson spread for fixed λ

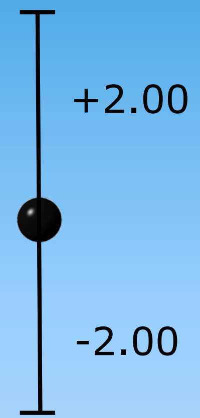
the famous √N

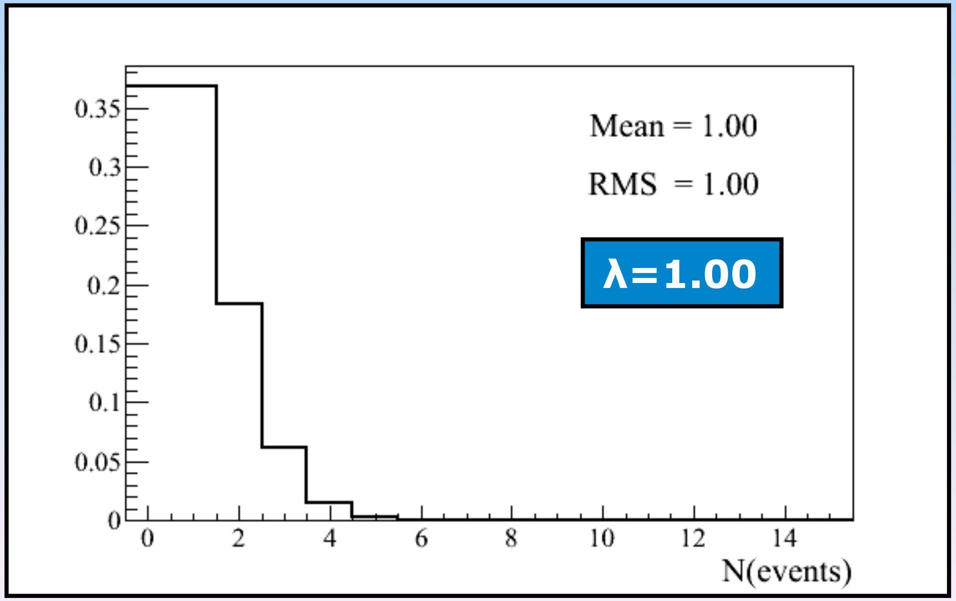
properties

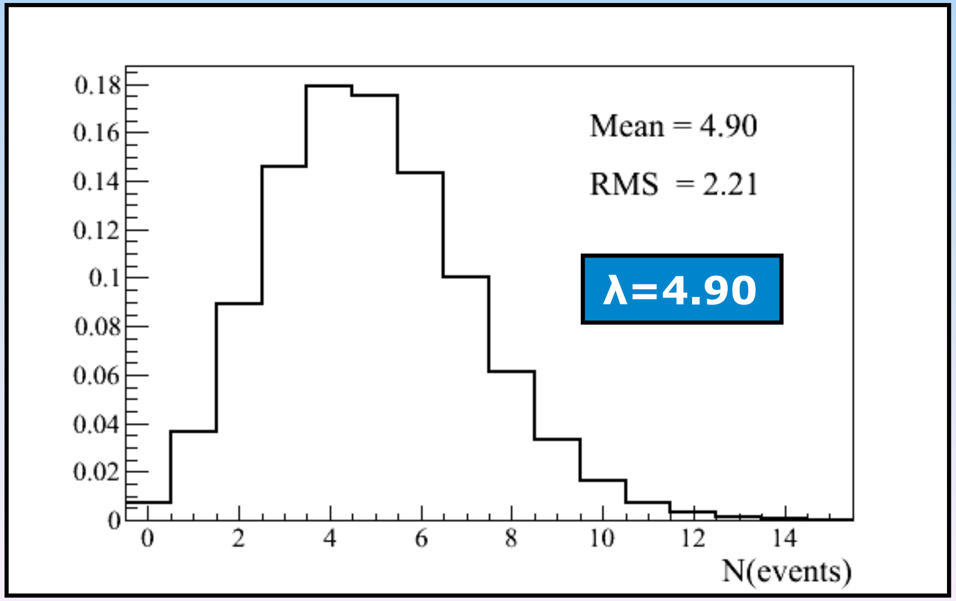
(1) Mean: $\langle n \rangle = \lambda$

(2) Variance: $\langle (n - \langle n \rangle)^2 \rangle = \lambda$

(3) Most likely: first integer ≤ λ







Treating it like a normal measurement

What you have:

$P(N_{obs} | \lambda)$

1) construct Likelihood
   λ as free parameter

↓

2) Find value of λ that
   maximizes Likelihood

↓

3) Determine error interval:
   $\Delta(-2\text{Log}(\text{lik})) = +1$

Likelihood (ratio)

$$L(n \mid \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

Likelihood

Probability to observe n events when λ are expected

$$P(4 \mid 0) = 0.00000$$

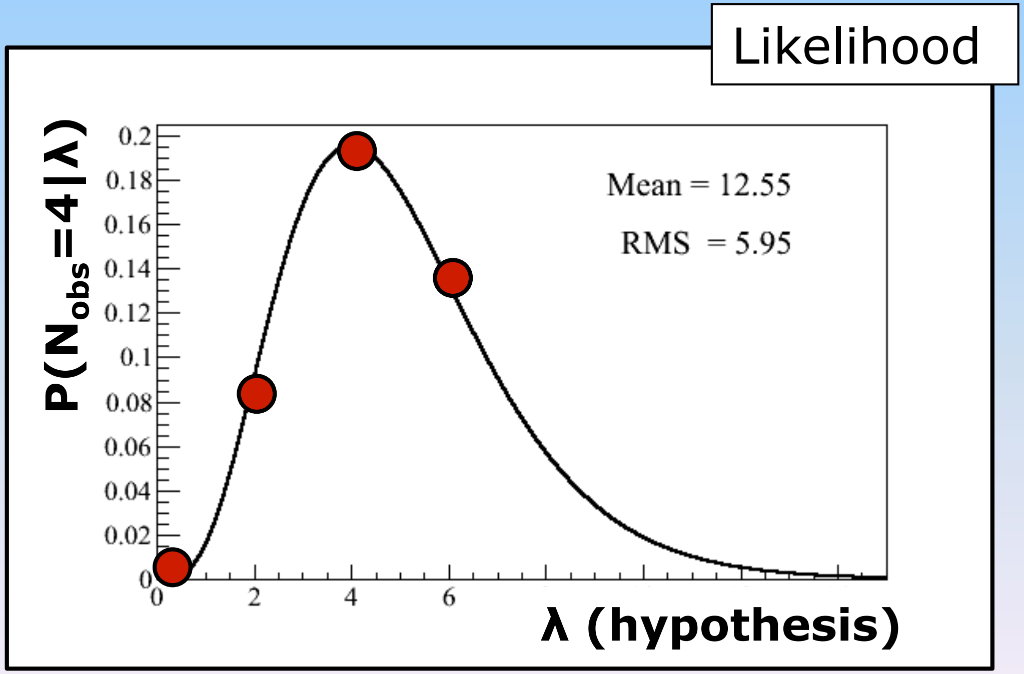
$$P(4 \mid 2) = 0.09022 \quad !$$

$$P(4 \mid 4) = 0.19537$$

$$P(4 \mid 6) = 0.13385 \quad !$$

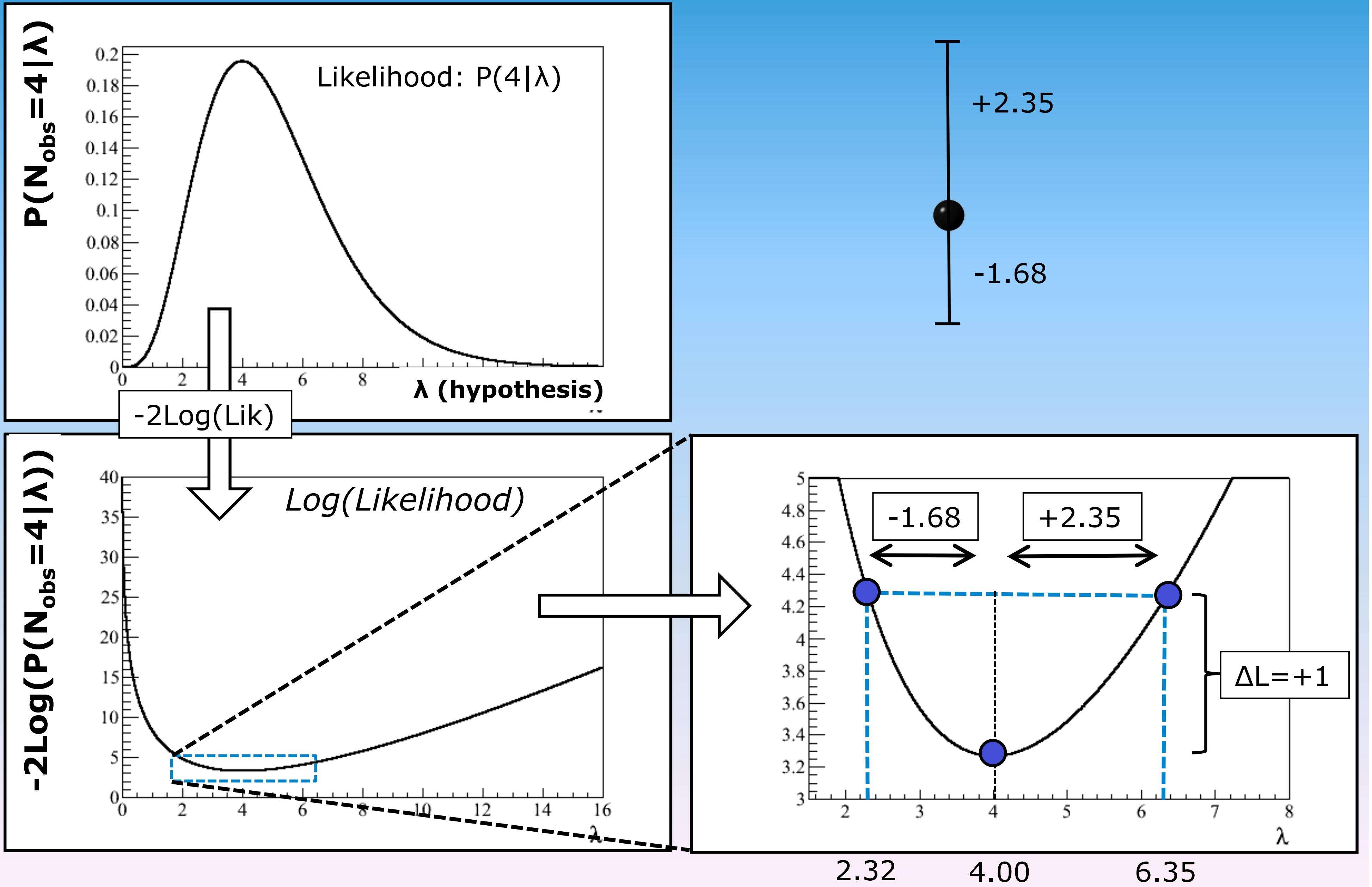
#observed — **fixed**

λ hypothesis — **varying**



Note: normally you use -2log(Lik)

Option 2: likelihood (ratio)



Bayesian: statement on true value of λ

What you have: $P(N_{obs} \mid \lambda)$

What you want: $P(\lambda \mid N_{obs})$

$$P(\lambda \mid N_{obs}) = P(N_{obs} \mid \lambda) P(\lambda)$$

Likelihood: Poisson distribution

"what can I say about the **measurement** (number of observed events) given a theory expectation ?"

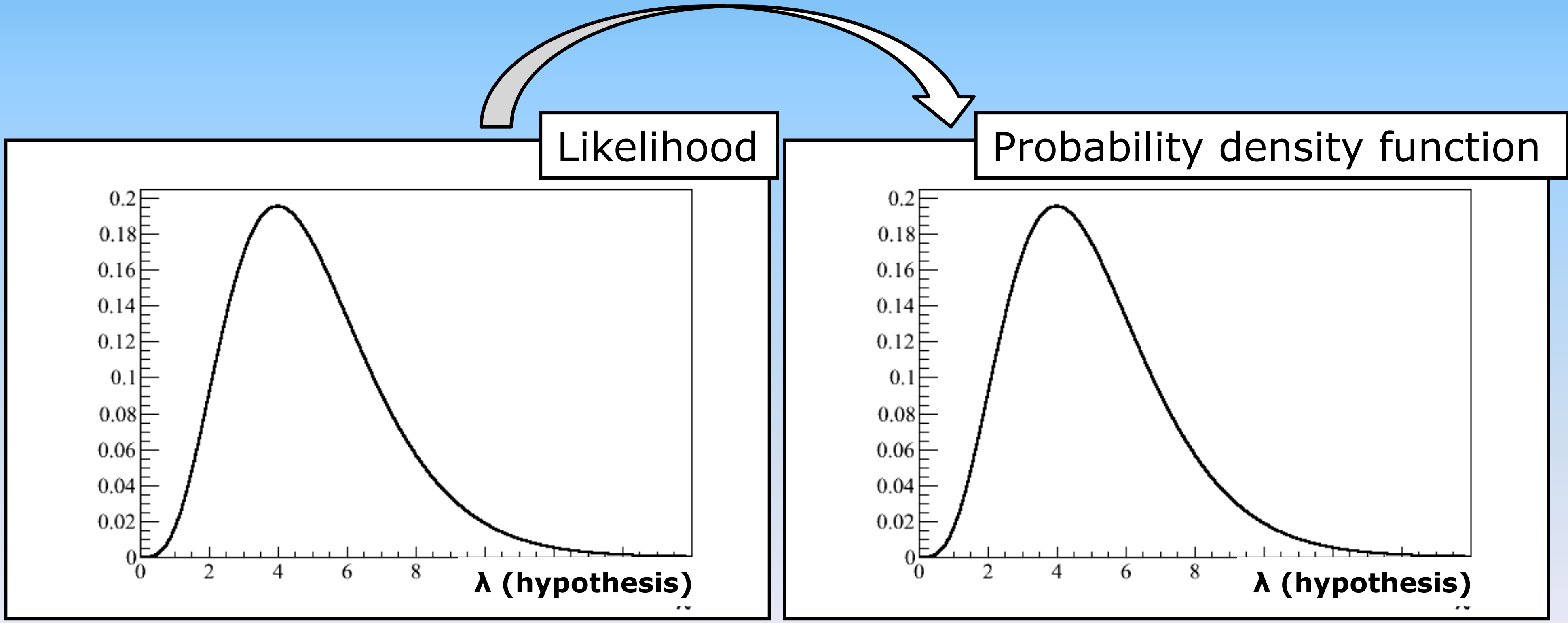
Posterior pdf for λ:

"what can I say about the underlying **theory** (true value of λ) given that I have observed of 4 events ?"

Bayesian: statement on true value of λ

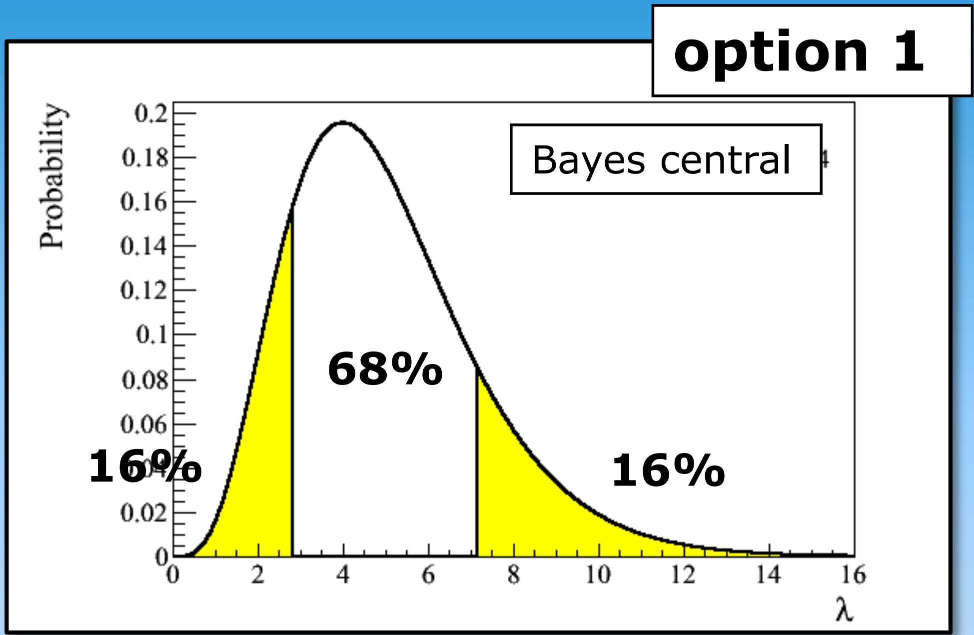
$$P(\lambda \mid N_{obs}) = P(N_{obs} \mid \lambda) P(\lambda)$$

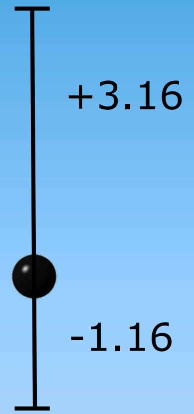
Choice of prior P(λ):
Assume all values for λ are equally likely ("I know nothing")

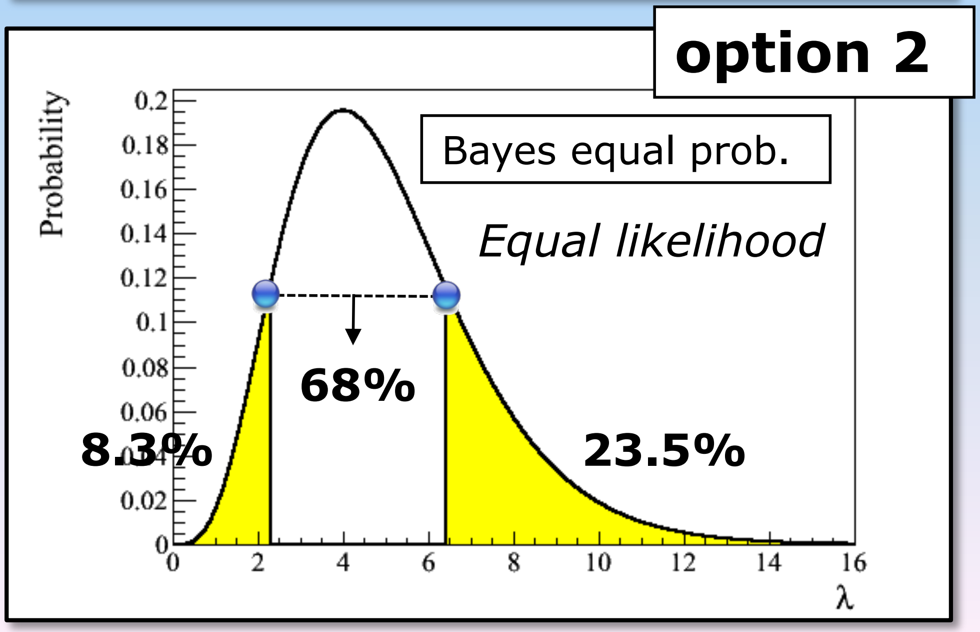


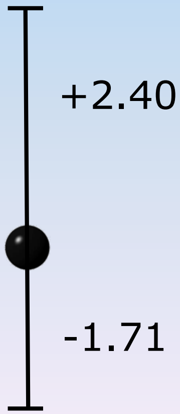
Posterior PDF for λ
→ Integrate to get confidence interval

Option 3 and 4: Bayesian



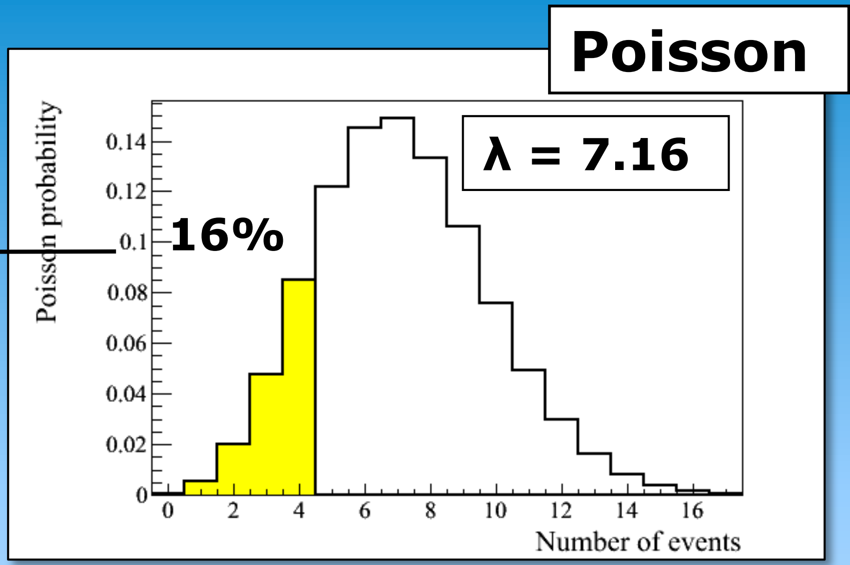






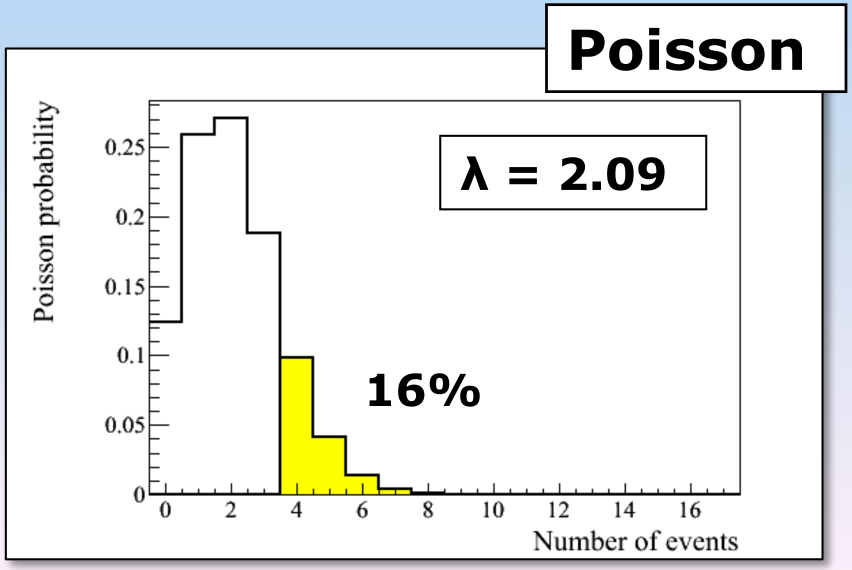
Option 4: Frequentist

If λ < 7.16 then probability to observe 4 events (or less) <16%



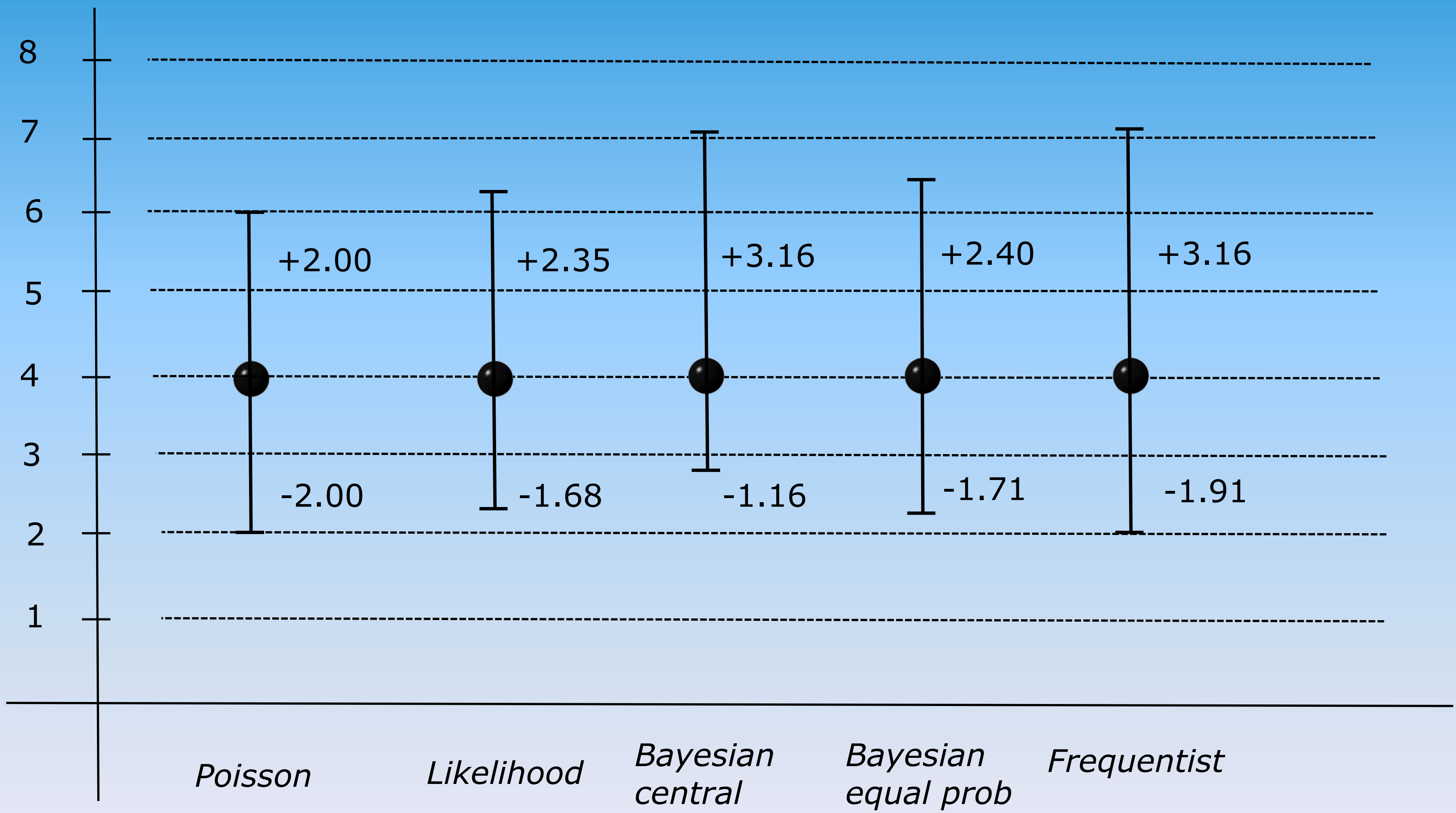
+3.16 → smallest λ for which $P(n \leq n_{obs} | \lambda) = 0.159$

-1.91 → largest λ fow which $P(n \geq n_{obs} | \lambda) = 0.159$



Note: also using data that you did **not** observe

The options

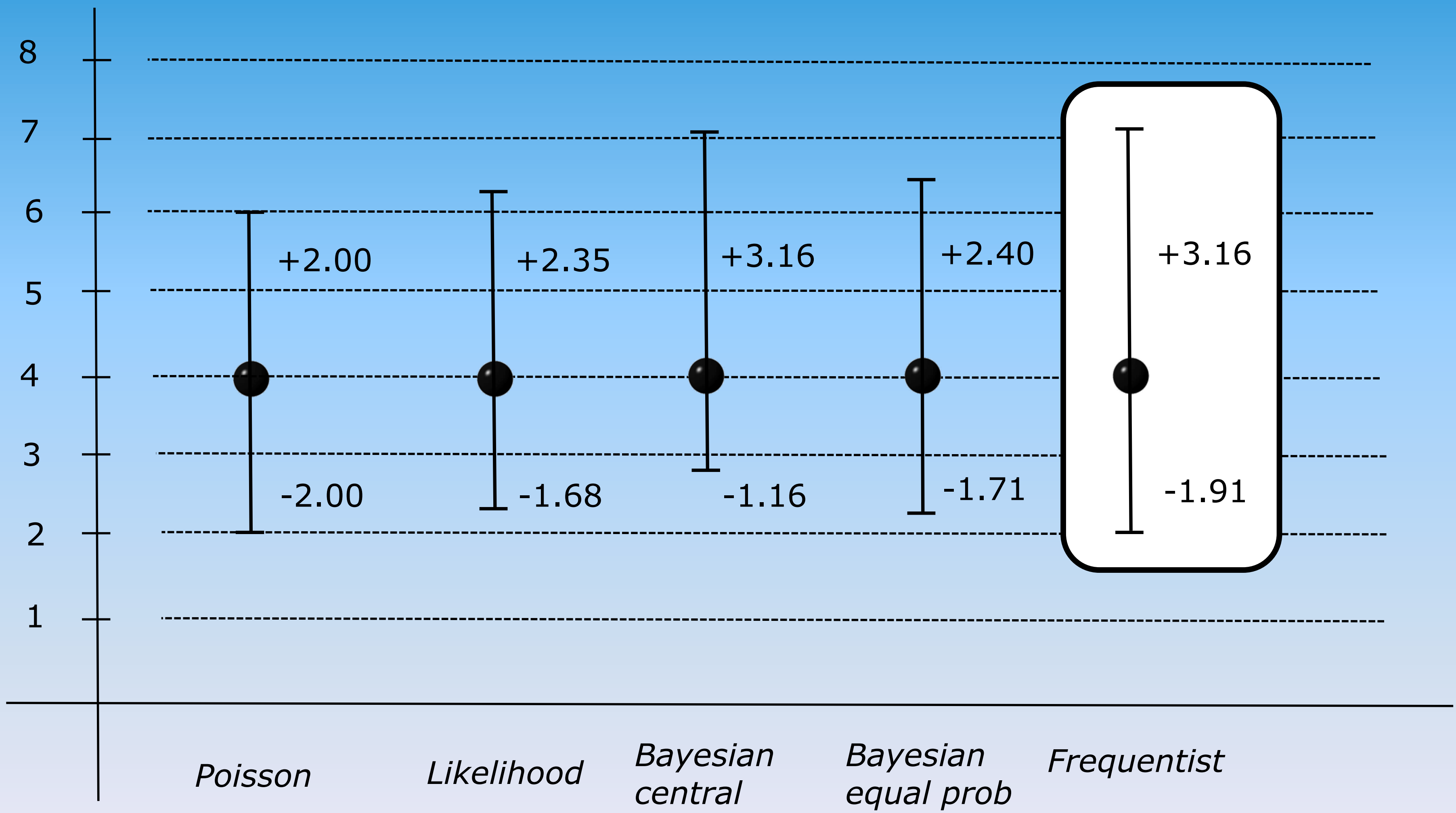


Δλ:	4.00	4.03	4.32	4.11	5.07

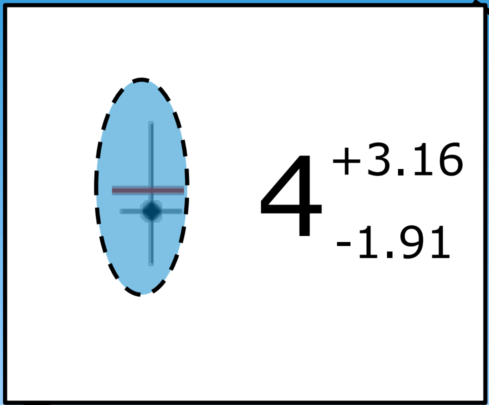
Think about it and discuss with your colleagues tonight

I'll give the answer tomorrow

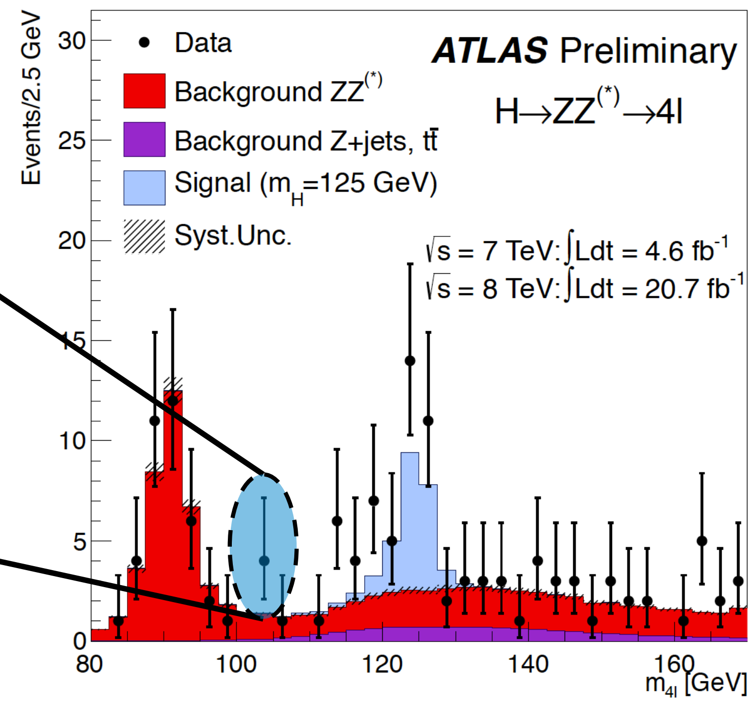
The options



Δλ:	4.00	4.03	4.32	4.11	5.07



ATLAS H→ 4 lepton peak



Conclusion:

- Now you know what RooFit uses
- Hope you are a bit confused

http://www.nikhef.nl/~ivov/Statistics/PoissonError/

BobCousins_Poisson.pdf → Paper with details
PoissonError.C → Implementation options shown here
Ivo_Analytic_Poisson.pdf → Analytic properties (inverted) Poisson