

Poisson probability: $p(N|\mu) = \frac{\mu^N e^{-\mu}}{N!}$

Upper confidence/credible limit (CL or α) on μ :

Bayesian:

$$\int_{-\infty}^{\mu_\alpha} C \mu^N e^{-\mu} \Pi(\mu) d\mu = CL$$

where

$\Pi(\mu)d\mu$ is the prior pdf for μ and the integrand is the posterior pdf.

C is a constant so that the posterior integrates to 1 over the full range of μ

The choice of $\Pi(\mu)$ is arbitrary (subjective)

A “flat” is constant over the allowed physical region $\mu \geq 0$ and 0 elsewhere: $\Pi(\mu) = \Theta(\mu)$

$$\int_0^{\mu_\alpha} C \mu^N e^{-\mu} d\mu = CL = 1 - \alpha$$

Frequentist μ_α is such that

$$\sum_{n=0}^N \frac{\mu_\alpha^n e^{-\mu_\alpha}}{n!} = 1 - CL = \alpha$$

In this simple case the answer happens to be the same.

Notation, e.g.

CL = 95%

$\alpha = 0.05$

Frequentist:

μ_α is the value of μ that corresponds to a p-value= α given that N events are observed

Solve numerically or by connecting to χ^2 distribution

With perfectly known background B:

$$p(N|\mu) = \frac{\mu^N e^{-\mu}}{N!} \rightarrow p(N|S, B) = \frac{(S + B)^N e^{-(S+B)}}{N!}$$

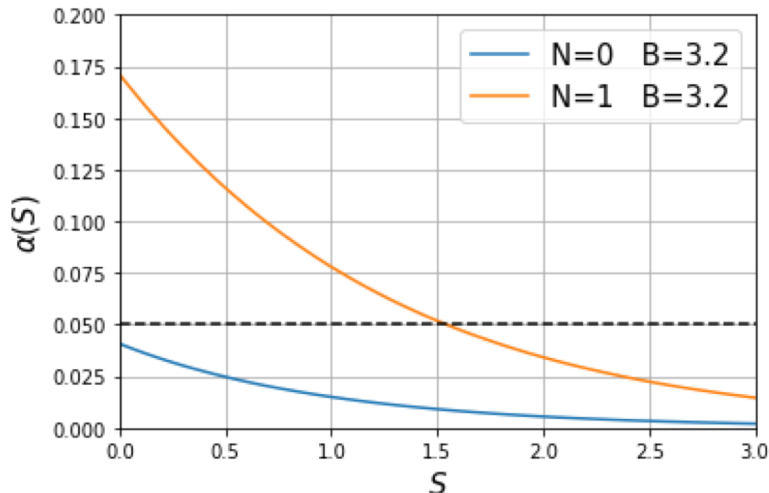
Bayesian limit with a flat prior becomes:

$$\int_0^{S_\alpha} C(S + B)^N e^{-(S+B)} dS = \text{CL} = 1 - \alpha$$

Frequentist limit becomes:

$$\sum_{n=0}^N \frac{(S_\alpha + B)^n e^{-(S_\alpha+B)}}{n!} = 1 - \text{CL} = \alpha$$

Problem, if $B > N$ in some cases that limit does not exist inside the $S \geq 0$ physical region



e.g.:

for $N=0$ $B=3.2$ $S=0$ is excluded at more than 95% CL.
This means that we also exclude the background-only hypothesis.

By definition, this will happen 5% of the times!

Enter CL_s

$$CL_s(S) = \frac{CL_{S+B}}{CL_B} = \frac{\sum_{n=0}^N (S+B)^n e^{-(S+B)} / n!}{\sum_{n=0}^N B^n e^{-B} / n!}$$

And find S such that $CL_s(S) = \alpha$.

The numerator is the same as in the previous page

The denominator is $1-pValue(B)$. By construction it is <1

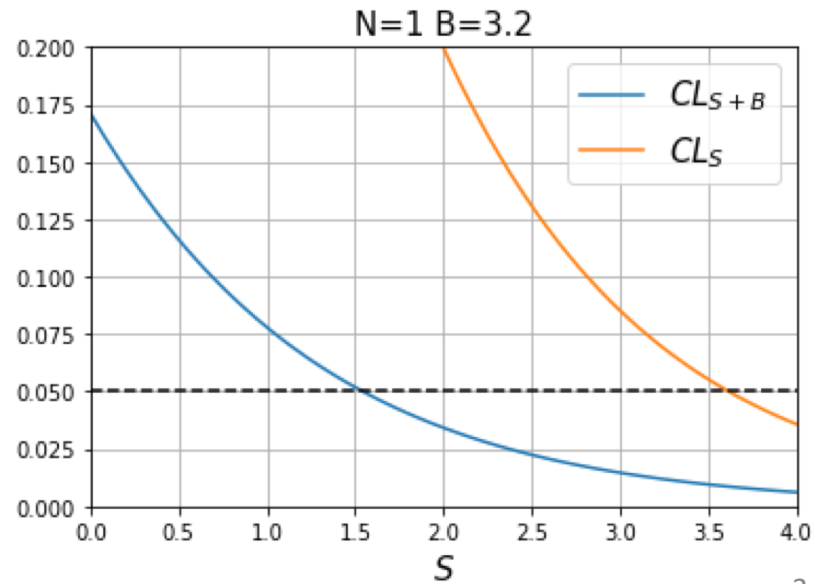
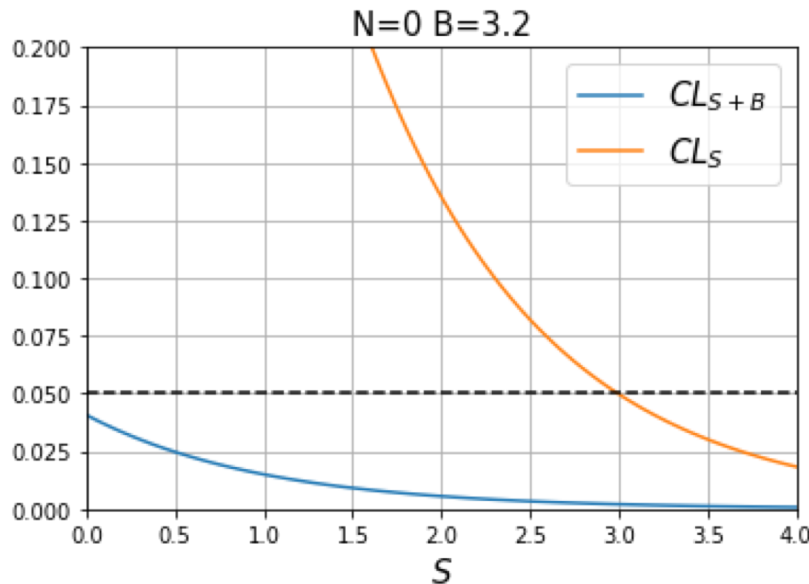
A justification is that we want to compare the S-hypothesis with the B-only hypothesis

This is a ratio of probability (likelihood ratio)

It does not have perfect frequentist properties (it “overcovers”)

But it protects against the pathological problem of the previous slide.

It so happens that numerically this is the same as Bayesian with a flat prior.



Systematics: $\mu = \epsilon S + \beta B$

where ϵ and β have some uncertainty

(for simplicity assume that the “central values” are $\epsilon=\beta=1$)

let $\vec{\theta} = (\epsilon, \beta)$: nuisance parameters

The Poisson probability for the observation of N is

$$p(N|S, B, \vec{\theta}) = \frac{\mu^N e^{-\mu}}{N!} = \frac{(\epsilon S + \beta B)^N e^{-(\epsilon S + \beta B)}}{N!}$$

Bayesian:

Integrate over the pdf for the nuisance parameters $p(\vec{\theta})d\vec{\theta}$

$$p(S)dS = C \left(\int \mu(S, B, \vec{\theta})^N e^{-\mu(S, B, \vec{\theta})} p(\vec{\theta})d\vec{\theta} \right) dS$$

Where I assumed a flat prior for S and C is a normalization factor. This can then be taken as a posterior probability for S from which we can extract (credible) intervals and limits in the usual way.

This process goes under the name of marginalization

Frequentist:

Given the data (ie: N) we have a p-value as function of both S and $\vec{\theta}$

Extract a p-value for any S assumption by averaging over $\vec{\theta}$

$$p\text{-value}(S) = \int p\text{-value}(S, \vec{\theta}) p(\vec{\theta}) d\vec{\theta}$$

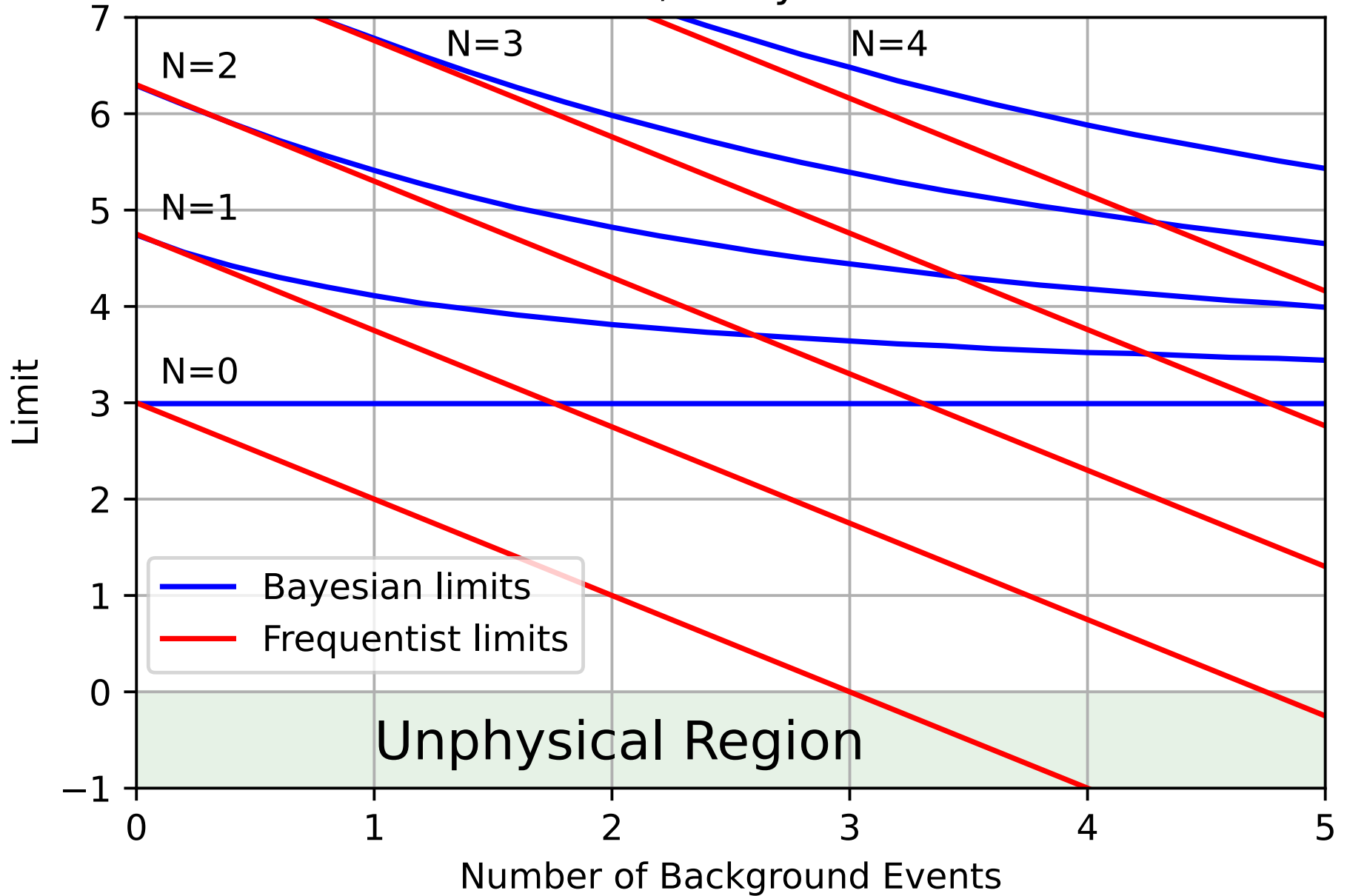
Based on this p-value can put limits on S

This includes the pdf for the true value of the nuisance parameters $p(\vec{\theta})d\vec{\theta}$

Of course the pdf of the true value is not a frequentist concept.

Hybrid Frequentist-Bayesian

95% limit, no systematics



Bayesian flat prior

N	B	no syst	dbG=20% dEff=0%	dbG=40% dEff=0%	dbG=0% dEff=10%	dbG=0% dEff=20%	dbG=40% dEff=20%
0	0.5	3.0	3.0	3.0	3.1	3.2	3.2
2	0.5	5.8	5.8	5.8	6.0	6.3	6.3
0	2.0	3.0	3.0	3.0	3.1	3.2	3.2
2	2.0	4.8	4.9	4.9	4.9	5.2	5.3
0	3.5	3.0	3.0	3.0	3.1	3.2	3.2
2	3.5	4.3	4.4	4.5	4.4	4.6	4.8

Bayesian $\frac{1}{\sqrt{S}}$ prior

N	B	no syst	dbG=20% dEff=0%	dbG=40% dEff=0%	dbG=0% dEff=10%	dbG=0% dEff=20%	dbG=40% dEff=20%
0	0.5	1.6	1.6	1.6	1.6	1.7	1.7
2	0.5	4.7	4.7	4.7	4.8	5.0	5.0
0	2.0	1.6	1.6	1.6	1.6	1.7	1.7
2	2.0	3.2	3.2	3.3	3.2	3.3	3.5
0	3.5	1.6	1.6	1.6	1.6	1.7	1.7
2	3.5	2.6	2.7	2.8	2.6	2.7	2.9

Classic Frequentist

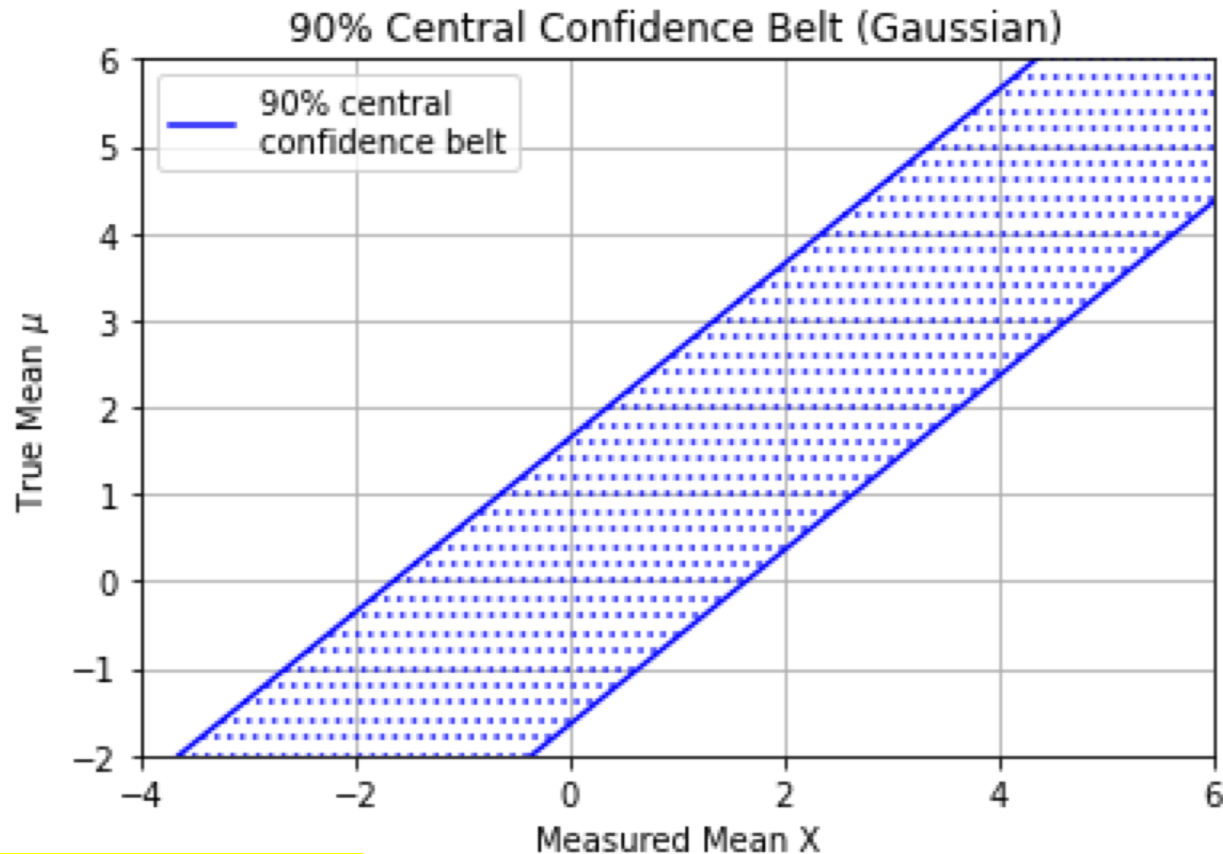
N	B	no syst	dbG=20% dEff=0%	dbG=40% dEff=0%	dbG=0% dEff=10%	dbG=0% dEff=20%	dbG=40% dEff=20%
0	0.5	2.5	2.5	2.5	2.5	2.6	2.6
2	0.5	5.8	5.8	5.8	5.9	6.1	6.1
0	2.0	1.0	1.0	1.1	1.0	1.0	1.1
2	2.0	4.3	4.3	4.3	4.3	4.5	4.5
0	3.5	N/A	N/A	N/A	N/A	N/A	N/A
2	3.5	2.8	2.9	3.0	2.8	2.9	3.1

CLs

N	B	no syst	dbG=20% dEff=0%	dbG=40% dEff=0%	dbG=0% dEff=10%	dbG=0% dEff=20%	dbG=40% dEff=20%
0	0.5	3.0	3.0	3.0	3.0	3.1	3.1
2	0.5	5.8	5.8	5.8	5.9	6.1	6.1
0	2.0	3.0	3.0	3.0	3.0	3.1	3.1
2	2.0	4.8	4.8	4.9	4.9	5.0	5.1
0	3.5	3.0	3.0	3.0	3.0	3.1	3.1
2	3.5	4.3	4.3	4.3	4.4	4.5	4.5

Feldman Cousins

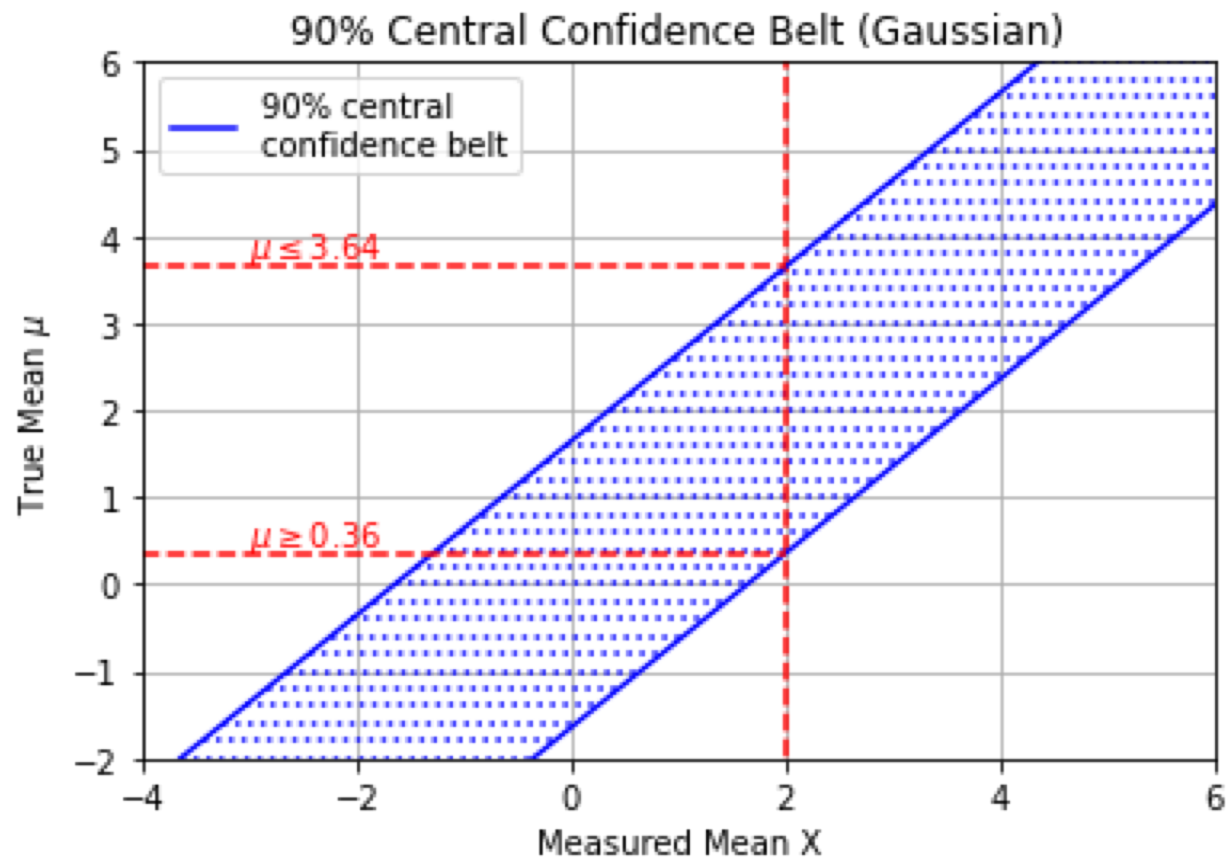
Reminder of frequentist construction



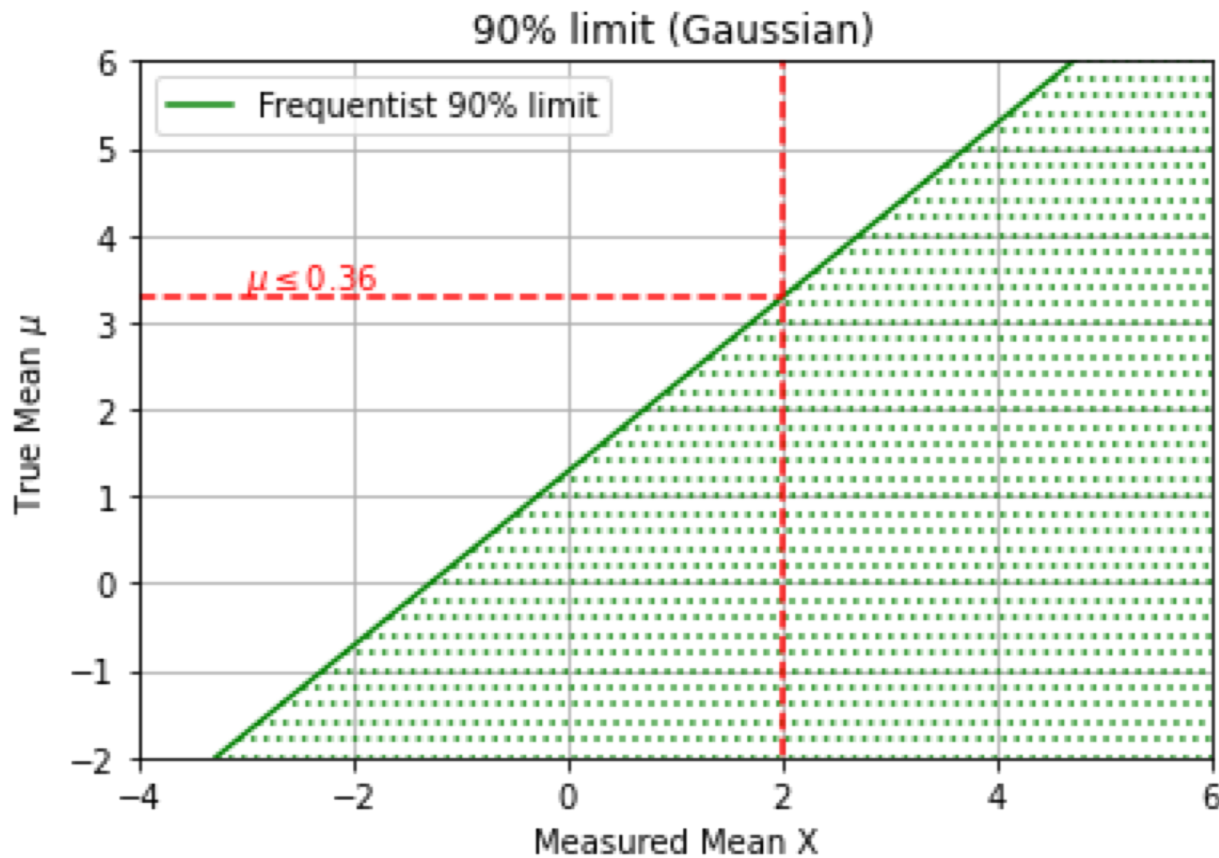
For this example:

- unit gaussian
- physical region: $\mu \geq 0$

Assume I measure $x=2$

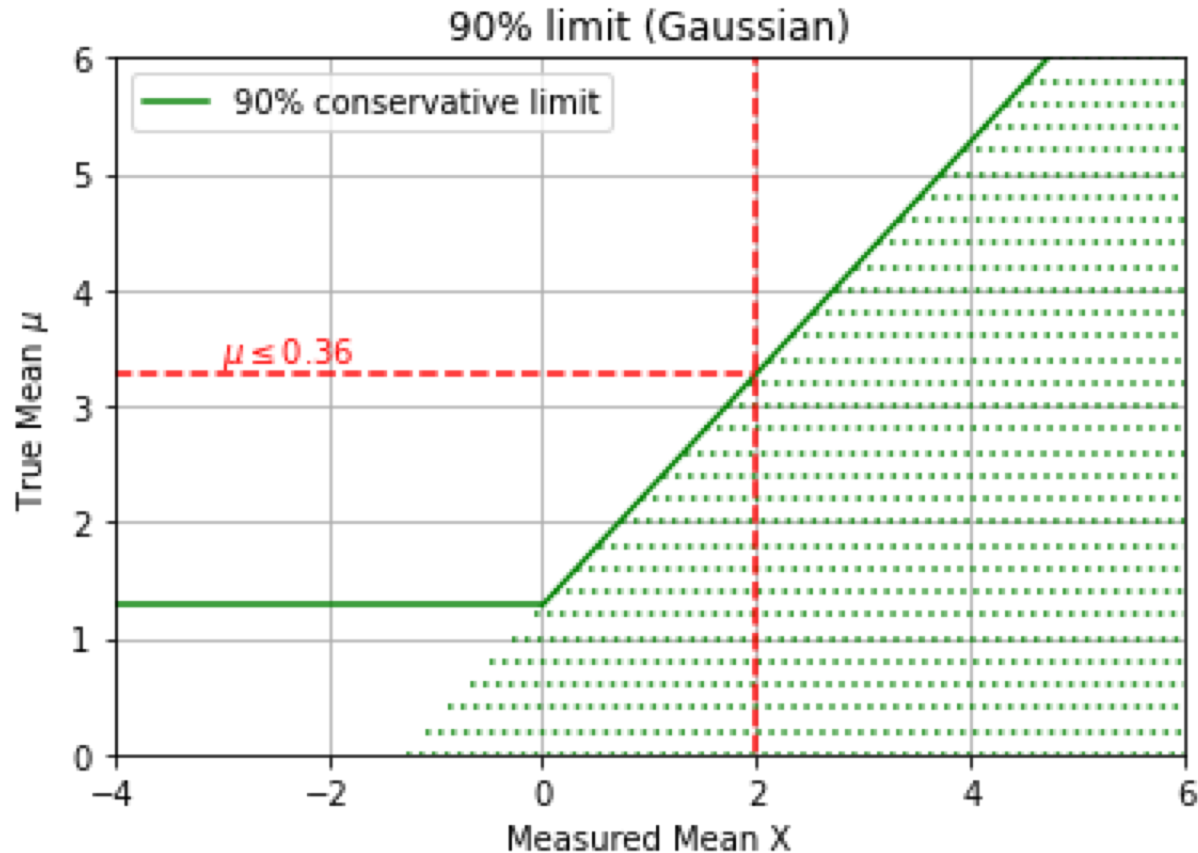


If I instead want the 90% upper limit



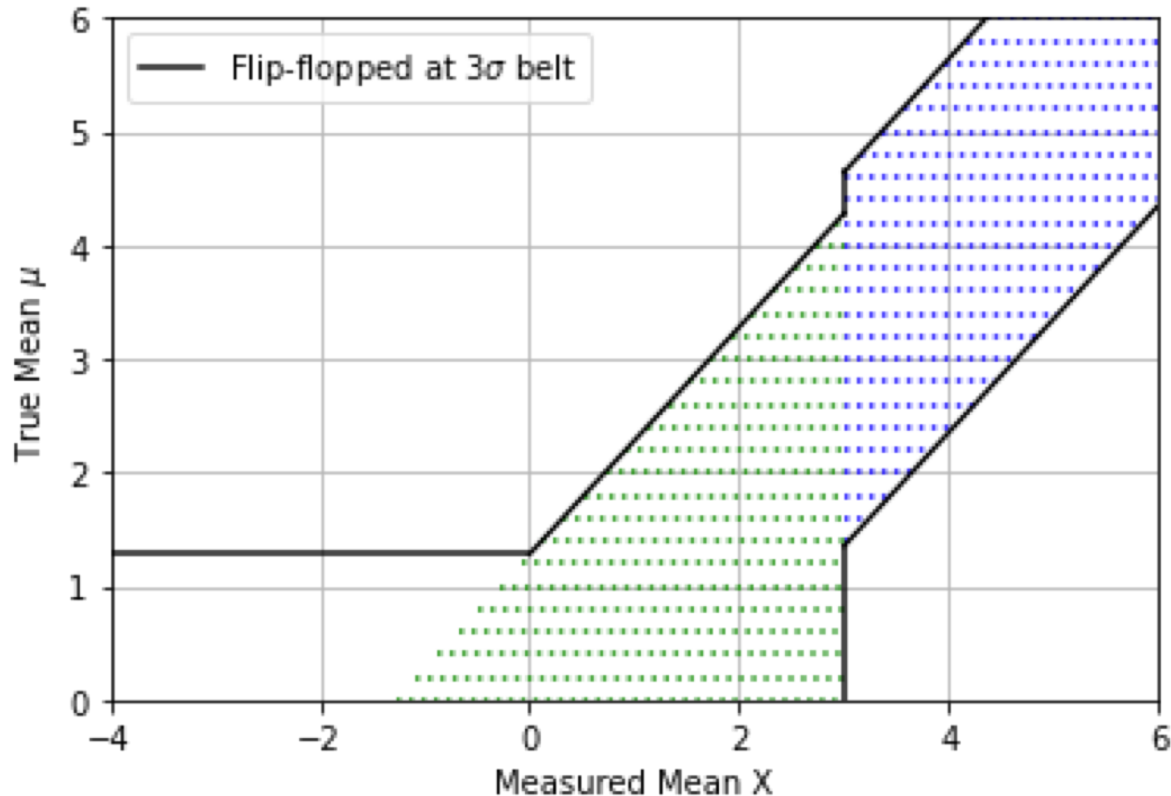
Note: if X is too negative the 90% limit is in the unphysical region

“Conservative” limit



The coverage at the low end is not 90% anymore (conservative)

Flip flop issue



Suppose an experimentalist decides:

- If I have a $> 3\sigma$ effect, I will quote a two-sided interval
- Otherwise I will quote an upper limit

The resulting belt has a kink, and in the kink region the coverage is $< 90\%$

Feldman-Cousins interval construction

- Goal is to construct intervals that
 1. Avoid the flip flopping problem.
 2. Transition smoothly away from the non-physical region
- The prescription to construct an interval for a given μ ,
 - Instead of
 - Setting upper/lower limit by including/excluding all x greater or smaller than a given value until some sum-of-prob value is reached
 - Or
 - Selecting a central region with certain equal prob of x falling above or below region
 - Do the following:
 - Start from x which maximizes some ranking criteria R , eg, $R=p(x|\mu)$. Continue to add values of x ranked by R until the total probability is CL (e.g., 90%)

Feldman Cousins Ranking

$$R = \frac{p(x|\mu)}{p(x|\mu_0)}$$

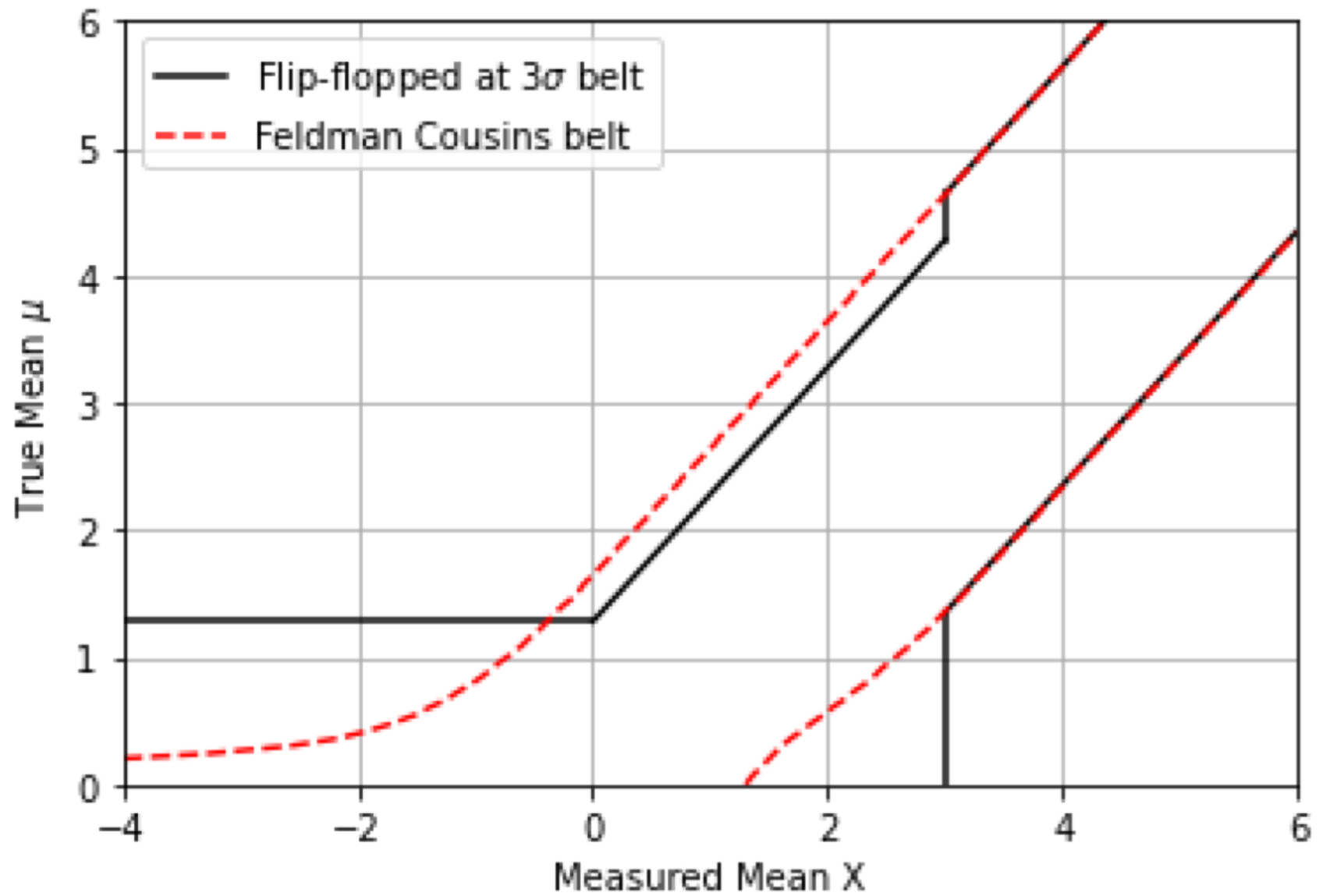
Where μ_0 is the value of μ that maximizes $p(x|\mu)$ and is in the physical region.

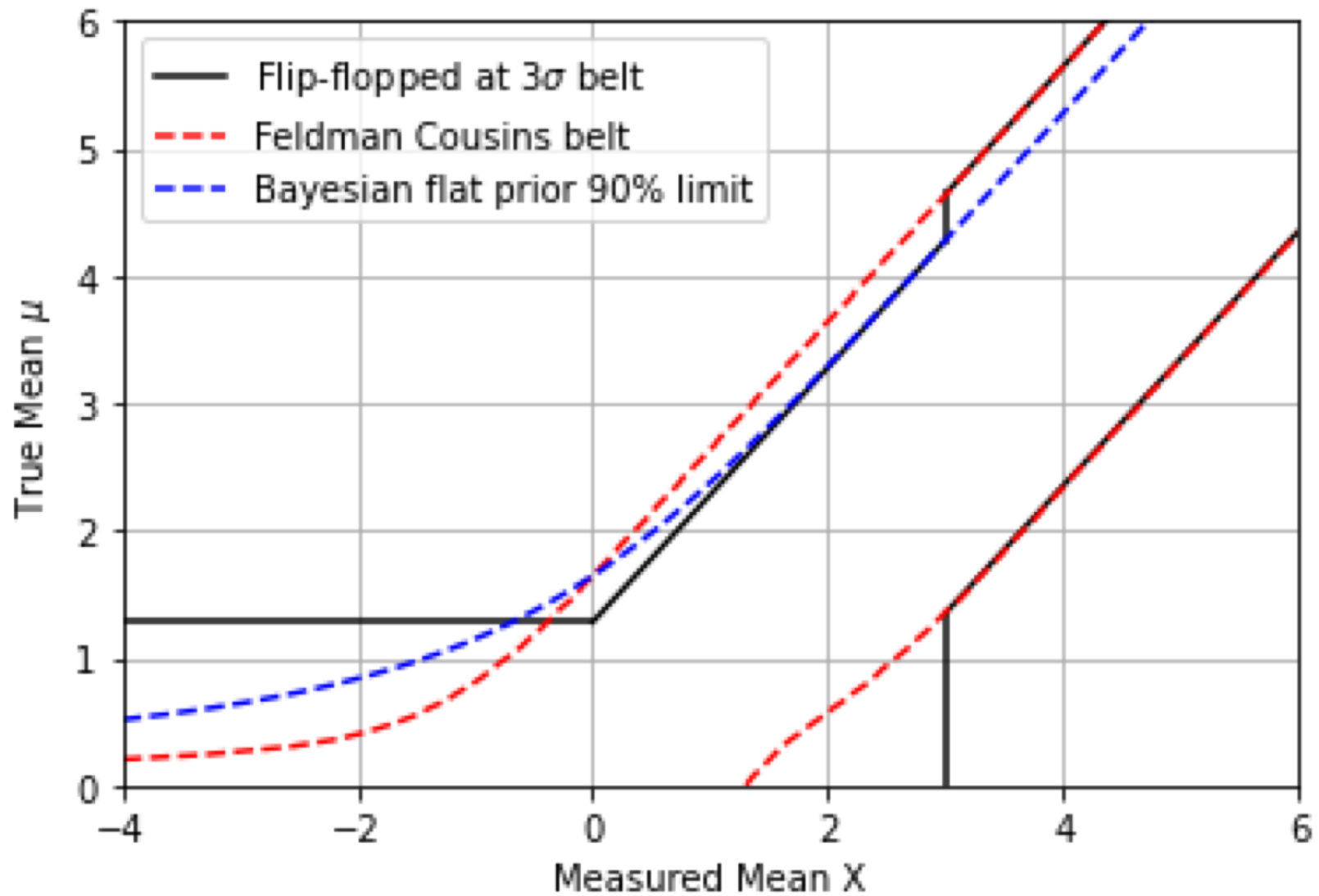
In the (unit) Gaussian case:

- If $x \geq 0$: $R = e^{-(x-\mu)^2/2}$
- If $x < 0$: $R = e^{-(x-\mu)^2/2} / e^{-x^2/2}$

Then

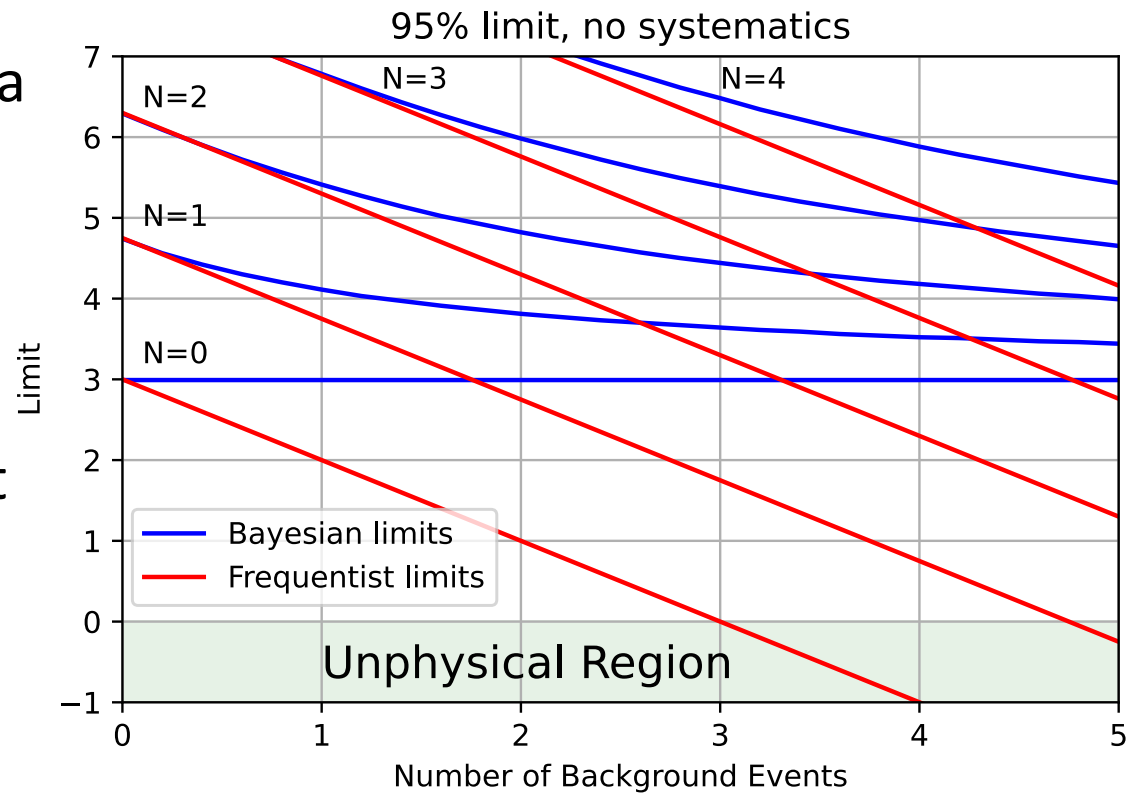
- Whether $\mu=0$ is (limit on μ) or is not (measurement of μ) in the 90% (or whatever) interval is “automatic”
- Solves the issue with unphysical region





Criticism of frequentist approach

- A not-so-good experiment that gets “lucky” can claim a better limit than a good experiment
- Compare (N=0 B=0.1) and (N=0 B=2.5) curves.
- Effect also for Bayesian, but less pronounced, and absent for N=0
- It is not fair!



The uneasiness comes from the fact that we unconsciously interpret limits as probabilistic statements on the mean (“ μ is $< xx$ with 95% probability”)

(And why else would we do the experiment????)

But this is a Bayesian statement (“the probability of the true value...”)

A frequentist analysis is after a different question....

Enter “expected limits”

- It has now become customary to also quote the “expected limit” (mean? median?) and also the 1σ and 2σ intervals
 - “If I was to repeat this experiment many times, in the absence of a signal, I will on average get limit = $\langle\mu\rangle$ and a limit in the interval $[\mu_1 \mu_2]$ 68% of the times”, etc.
- For a one-bin Poisson experiment this does not make sense
 - Results quantized: $N=0, 1, 2, \dots$ with probs. $p = p_0, p_1, p_2, \dots$
 - Resulting limits quantized $\mu=\mu_0, \mu_1, \mu_2, \dots$
- $\langle\mu\rangle$ that I quote is a limit that in most case I will not get
 - OK, I can live with that
- But cannot build intervals $[\mu_1 \mu_2]$ that cover exactly 68%..of the cases
 - Again, unless we are numerically lucky

Expected limits...what is actually done then?

IMHO, a cheat

- Pretend that the nuisances will change within uncertainties from experiment to experiment
- Then (say) all experiments with $N=0$ would not yield a unique limit μ_0 but a set of different limits clustered more or less tightly around μ_0 .
- You then have an expected continuous distribution of μ and it makes sense to talk about 68% intervals $[\mu_1 \mu_2]$ etc.
- This does not correspond to reality
 - Maybe defensible before some of the nuisances are constrained by auxiliary measurements, eg, control regions.
 - Not defensible for theoretical uncertainties, for example
- Be careful about these pre-packaged calculations of expected intervals in low stats cases (not just single channels)
 - In addition to the questionable procedure, I have seen standard tools give crazy results in some cases.

It gets worse (still IMHO)

- The standard procedure to quote expected limits at the LHC is such that the **expected limit depends on the observation**
- Thought process:
 - Suppose I calculate $B=1.5\pm 0.5$ before opening the cookie-jar (aka look at the signal region)
 - I observe $N=2$.
 - Therefore it is likely that B is a bit larger (say $B=1.7\pm 0.4$... I made this number up, it would come from a fit, but you get the idea).
 - Therefore I calculate the expected limit based on $B=1.7\pm 0.4$

LHC stats gurus, ca 2010:

“Silly you, it depends on what you mean by expected”