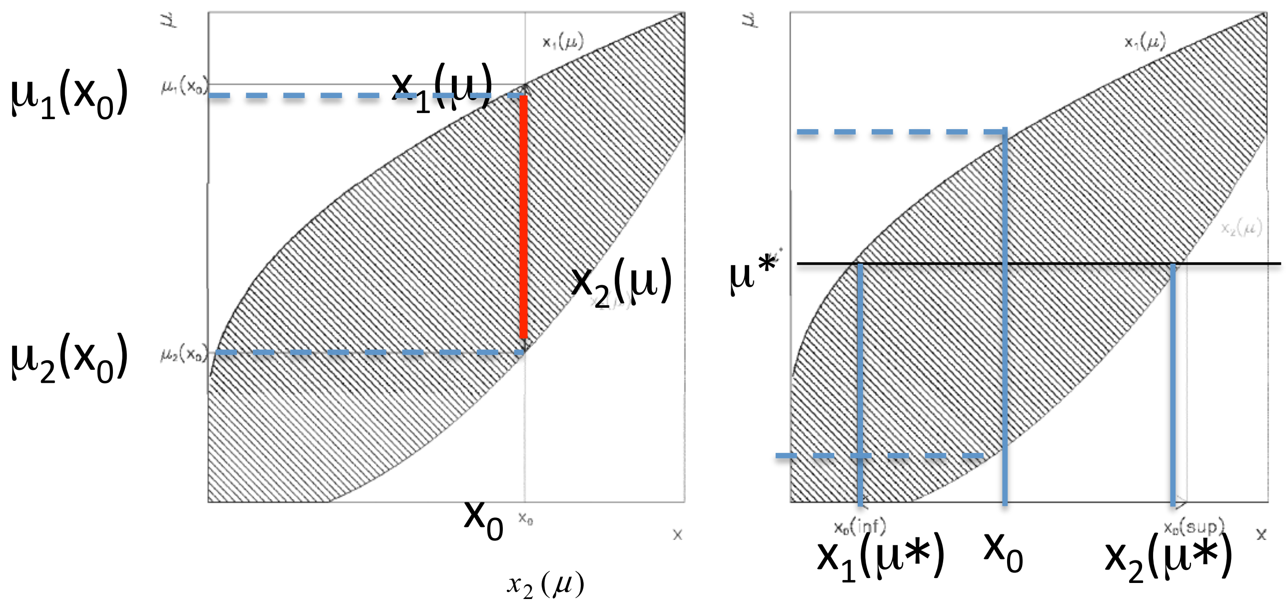
Neyman's construction



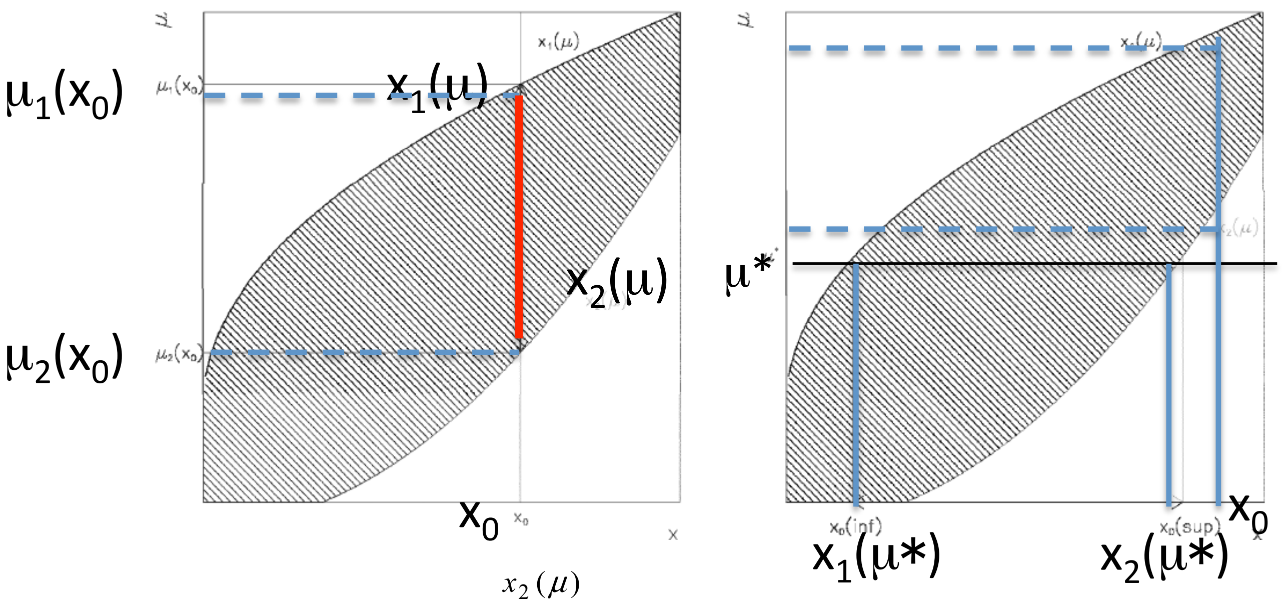
$$\int_{x_1(\mu)}^{x_2(\mu)} f(x/\mu)dx = \alpha$$

By construction the probability to measure $x_0' < x_0$ if the true value $\mu = \mu_1(x_0)$ is $(1-\alpha)/2$
$x_0' > x_0$ if the true value $\mu = \mu_2(x_0)$ is $(1-\alpha)/2$

Coverage: suppose μ^* the true value

$$P(x_1(\mu^*) < x_0 < x_2(\mu^*)) = \alpha$$

Neyman's construction



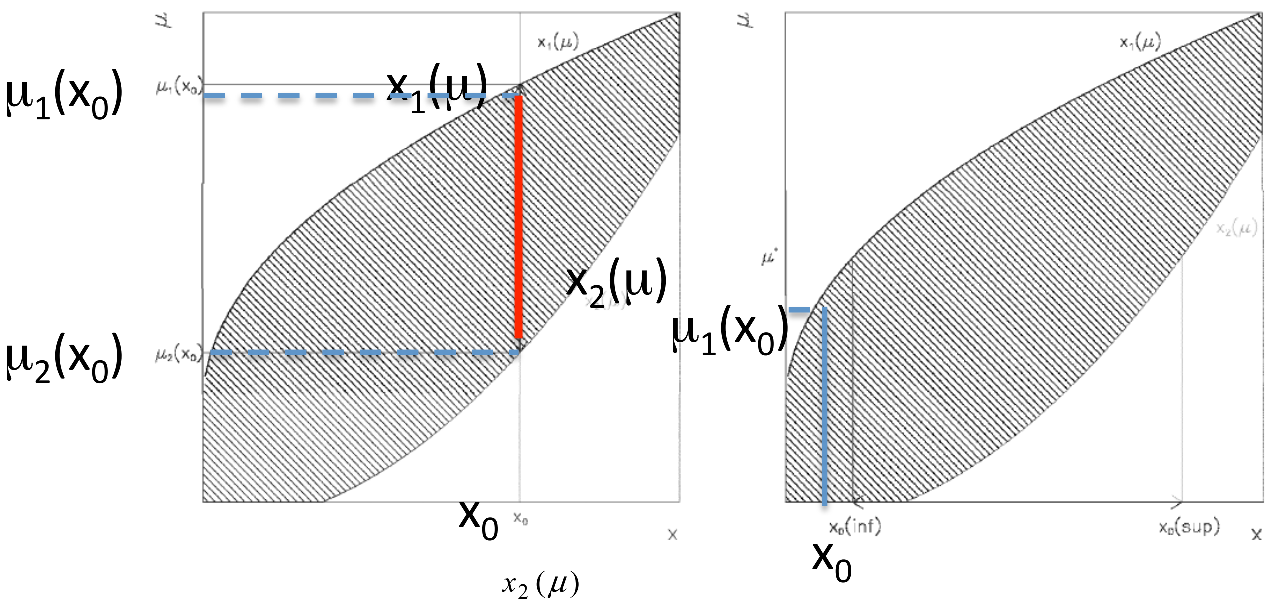
$$\int_{x_1(\mu)}^{x_2(\mu)} f(x/\mu)dx = \alpha$$

By construction the probability to measure $x_0' < x_0$ if the true value $\mu = \mu_1(x_0)$ is $(1-\alpha)/2$
$x_0' > x_0$ if the true value $\mu = \mu_2(x_0)$ is $(1-\alpha)/2$

Coverage: suppose μ^* the true value

$$P(x_1(\mu^*) < x_0 < x_2(\mu^*)) = \alpha$$

Neyman's construction



$$\int_{x_1(\mu)}^{x_2(\mu)} f(x/\mu)dx = \alpha$$

By construction the probability to measure $x_0' < x_0$ if the true value $\mu = \mu_1(x_0)$ is $(1-\alpha)/2$
$x_0' > x_0$ if the true value $\mu = \mu_2(x_0)$ is $(1-\alpha)/2$

$$P(x_1(\mu^*) < x_0 < x_2(\mu^*)) = \alpha$$

Suppose Poisson variable and n=0 is measured (no background)
=> 0±0 (freq) or 1±1 (Beyes) ?

Upper limit (lower limit =0)

By construction the probability to measure $x_0' < x_0$ if the true value $\mu = \mu_1(x_0)$ is $(1-\alpha)$ (only one limit) or the probability to measure $x_0' > x_0$ if the true value $\mu = \mu_1(x_0)$ is α

$$P(n>0/\lambda) = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} = 1 - e^{-\lambda} = \alpha$$

frequentist

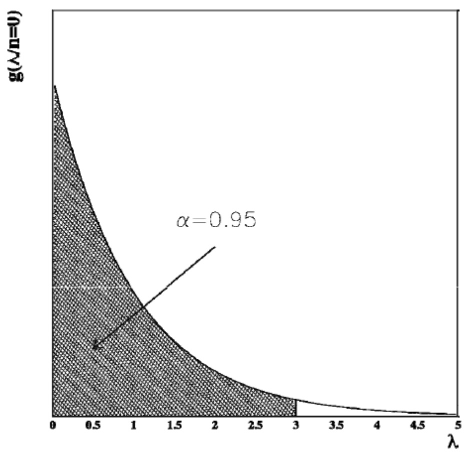
$$\bar{\lambda} = -\ln(1-\alpha)$$

$$g(\lambda/n=0) = \frac{p(n=0/\lambda) f_0(\lambda)}{\int_0^{\infty} p(n=0/\lambda) f_0(\lambda) d\lambda} = \frac{e^{-\lambda}}{\int_0^{\infty} e^{-\lambda} d\lambda} = e^{-\lambda}$$

Bayesian
(uniform prior)

$$p(\lambda < \bar{\lambda}) = \int_0^{\bar{\lambda}} e^{-\lambda} d\lambda = 1 - e^{-\bar{\lambda}} = \alpha$$

	90%	95%	99%
$\bar{\lambda}$	2.3	3.0	4.6



frequentist limits

The belt is limited on one side only, and for any result of a measurement n_0 we identify s_{up} in such a way that if $s_{true} = s_{up}$, the probability to get a counting smaller than n_0 is $1-\beta$[31]. By considering the Poisson statistics without background (b=0) we get:

$$\sum_{n=0}^{n_0} \frac{e^{-s_{up}} s_{up}^n}{n!} = 1-\beta$$

If $n_0 = 0$ we have

$$e^{-s_{up}} = 1-\beta$$

$$s_{up} = \ln \frac{1}{1-\beta}$$

from which we get the same numbers for s_{up} obtained in the bayesian case.

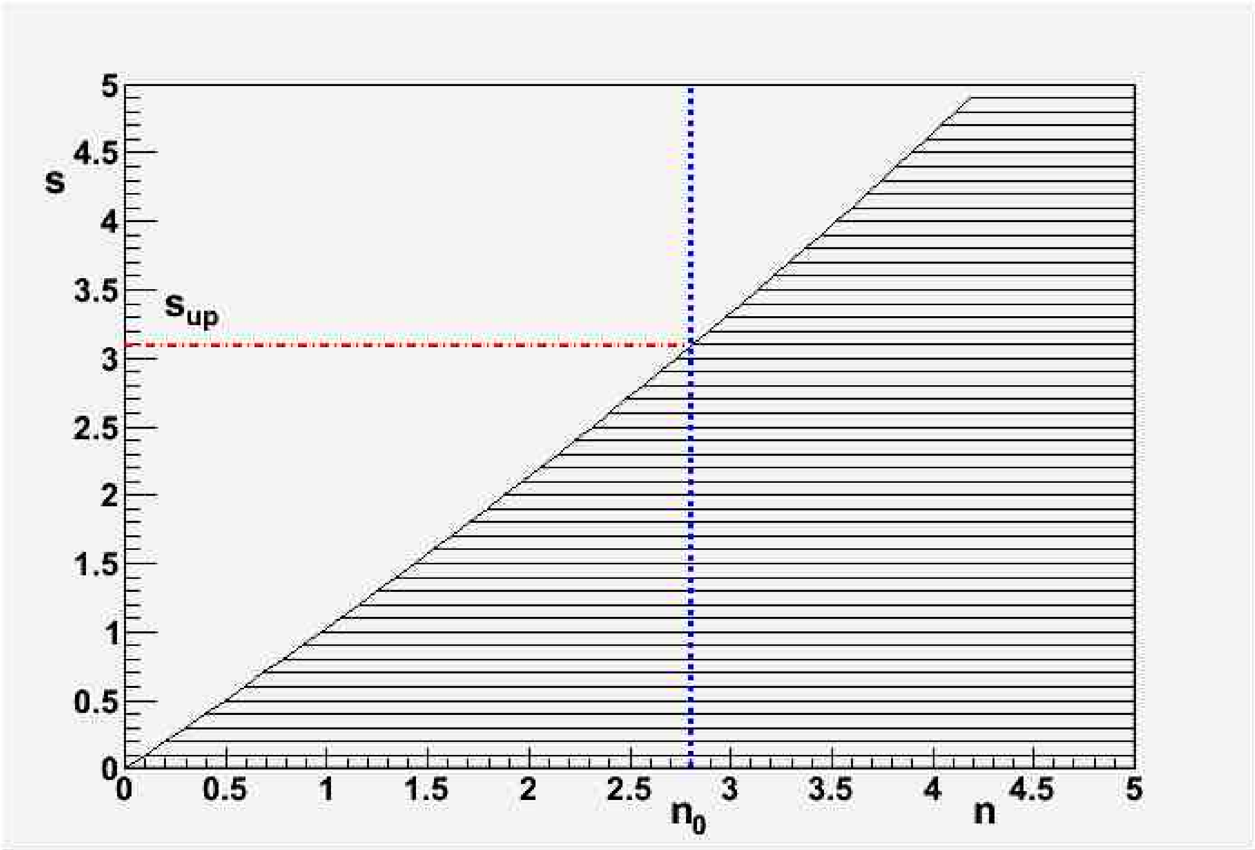


FIGURE 19. Neyman construction for the case of an upper limit. In this case a segment between $n_1(\theta)$ and ∞ is drawn for each value of the parameter θ. The segments define the confidence region. Once a value of n, n_0 is obtained, the upper limit s_{up} is found. (For simplicity the discrete variable n is considered as a real number here).

frequentist limits

If b is not equal to 0 but is known,

$$\sum_{n=0}^{n_0} \frac{e^{-(s_{up}+b)}(s_{up}+b)^n}{n!} = 1 - \beta \tag{201}$$

and from this equation upper limits can be evaluated for the different situations.

It has been pointed out that the use of eq.201 gives rise to some problems. In particular negative values of s_{up} can be obtained using directly the formula[32]. This doesn't happen in the bayesian context where the condition $s > 0$ is put directly by using the proper prior.

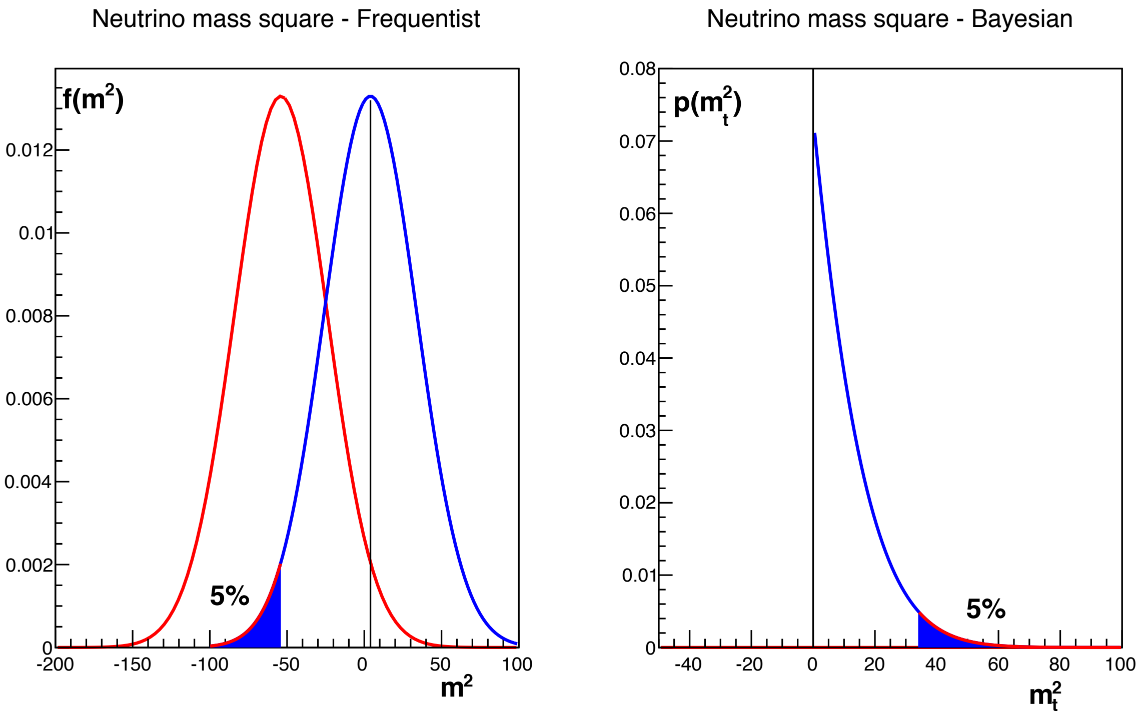
[32]A rate is a positive-definite quantity. However, if a rate is 0 or very small with respect to the experimental sensitivity, the probability that n_0 is larger than b is exactly equal to the probability that n_0 is lower than b. This implies that a negative rate naturally comes out from an experimental analysis based on a difference between two counts. The acceptance of such results is a sort of "philosophical" question and is controversial.

Flip-flop problem

$$\overline{m}^2 = -54 \pm 30 \ \ \text{eV}^2$$

$$p(m_t^2/\overline{m}^2) = \frac{L(\overline{m}^2/m_t^2)\pi(m_t^2)}{\int dm_t^2 L(\overline{m}^2/m_t^2)}$$

$$m^2 < 4.6 \ \text{eV}^2$$



$$m_t^2 < 34 \ \text{eV}^2$$

FIGURE 21. Example of the square neutrino mass. Construction of the upper limit in the frequentist approach (left plot) and in the bayesian approach (right plot). (left) The red gaussian is the experimental likelihood, the blue gaussian corresponds to the 95% CL upper limit that leaves 5% of possible the experiment outcomes below the present experimental average. (right) The blue curve is the result of the Bayes theorem when a prior forcing to positive values is applied (eq.202).

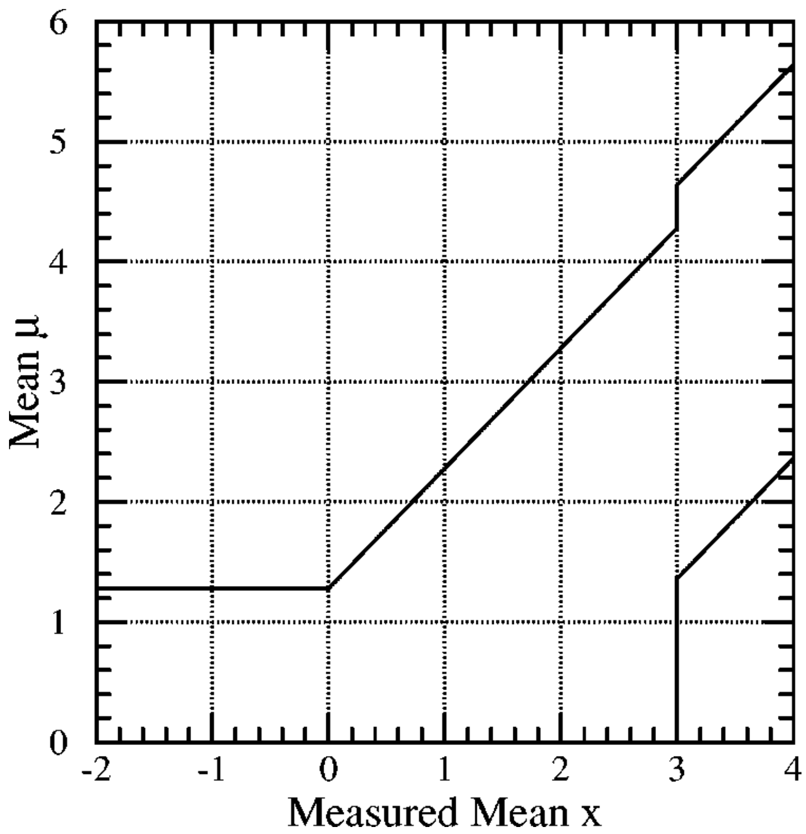


FIG. 4. Plot of confidence belts implicitly used for 90% C.L. confidence intervals (vertical intervals between the belts) quoted by flip-flopping physicist X, described in the text. They are not valid confidence belts, since they can cover the true value at a frequency less than the stated confidence level. For $1.36<\mu<4.28$, the coverage (probability contained in the horizontal acceptance interval) is 85%.

TABLE IV. 90% C.L. intervals for the Poisson signal mean μ, for total events observed n_0, for known mean background b ranging from 0 to 5.

$n_0\backslash b$	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0	0.00, 2.44	0.00, 1.94	0.00, 1.61	0.00, 1.33	0.00, 1.26	0.00, 1.18	0.00, 1.08	0.00, 1.06	0.00, 1.01	*0.00, 0.98*
1	0.11, 4.36	0.00, 3.86	0.00, 3.36	0.00, 2.91	0.00, 2.53	0.00, 2.19	0.00, 1.88	0.00, 1.59	0.00, 1.39	0.00, 1.22
2	0.53, 5.91	0.03, 5.41	0.00, 4.91	0.00, 4.41	0.00, 3.91	0.00, 3.45	0.00, 3.04	0.00, 2.67	0.00, 2.33	0.00, 1.73
3	1.10, 7.42	0.60, 6.92	0.10, 6.42	0.00, 5.92	0.00, 5.42	0.00, 4.92	0.00, 4.42	0.00, 3.95	0.00, 3.53	0.00, 2.78
4	1.47, 8.60	1.17, 8.10	0.74, 7.60	0.24, 7.10	0.00, 6.60	0.00, 6.10	0.00, 5.60	0.00, 5.10	0.00, 4.60	0.00, 3.60
5	1.84, 9.99	1.53, 9.49	1.25, 8.99	0.93, 8.49	0.43, 7.99	0.00, 7.49	0.00, 6.99	0.00, 6.49	0.00, 5.99	0.00, 4.99
6	2.21,11.47	1.90,10.97	1.61,10.47	1.33, 9.97	1.08, 9.47	0.65, 8.97	0.15, 8.47	0.00, 7.97	0.00, 7.47	0.00, 6.47
7	3.56,12.53	3.06,12.03	2.56,11.53	2.09,11.03	1.59,10.53	1.18,10.03	0.89, 9.53	0.39, 9.03	0.00, 8.53	0.00, 7.53
8	3.96,13.99	3.46,13.49	2.96,12.99	2.51,12.49	2.14,11.99	1.81,11.49	1.51,10.99	1.06,10.49	0.66, 9.99	0.00, 8.99
9	4.36,15.30	3.86,14.80	3.36,14.30	2.91,13.80	2.53,13.30	2.19,12.80	1.88,12.30	1.59,11.80	1.33,11.30	0.43,10.30
10	5.50,16.50	5.00,16.00	4.50,15.50	4.00,15.00	3.50,14.50	3.04,14.00	2.63,13.50	2.27,13.00	1.94,12.50	1.19,11.50
11	5.91,17.81	5.41,17.31	4.91,16.81	4.41,16.31	3.91,15.81	3.45,15.31	3.04,14.81	2.67,14.31	2.33,13.81	1.73,12.81
12	7.01,19.00	6.51,18.50	6.01,18.00	5.51,17.50	5.01,17.00	4.51,16.50	4.01,16.00	3.54,15.50	3.12,15.00	2.38,14.00
13	7.42,20.05	6.92,19.55	6.42,19.05	5.92,18.55	5.42,18.05	4.92,17.55	4.42,17.05	3.95,16.55	3.53,16.05	2.78,15.05
14	8.50,21.50	8.00,21.00	7.50,20.50	7.00,20.00	6.50,19.50	6.00,19.00	5.50,18.50	5.00,18.00	4.50,17.50	3.59,16.50
15	9.48,22.52	8.98,22.02	8.48,21.52	7.98,21.02	7.48,20.52	6.98,20.02	6.48,19.52	5.98,19.02	5.48,18.52	4.48,17.52
16	9.99,23.99	9.49,23.49	8.99,22.99	8.49,22.49	7.99,21.99	7.49,21.49	6.99,20.99	6.49,20.49	5.99,19.99	4.99,18.99
17	11.04,25.02	10.54,24.52	10.04,24.02	9.54,23.52	9.04,23.02	8.54,22.52	8.04,22.02	7.54,21.52	7.04,21.02	6.04,20.02
18	11.47,26.16	10.97,25.66	10.47,25.16	9.97,24.66	9.47,24.16	8.97,23.66	8.47,23.16	7.97,22.66	7.47,22.16	6.47,21.16
19	12.51,27.51	12.01,27.01	11.51,26.51	11.01,26.01	10.51,25.51	10.01,25.01	9.51,24.51	9.01,24.01	8.51,23.51	7.51,22.51
20	13.55,28.52	13.05,28.02	12.55,27.52	12.05,27.02	11.55,26.52	11.05,26.02	10.55,25.52	10.05,25.02	9.55,24.52	8.55,23.52

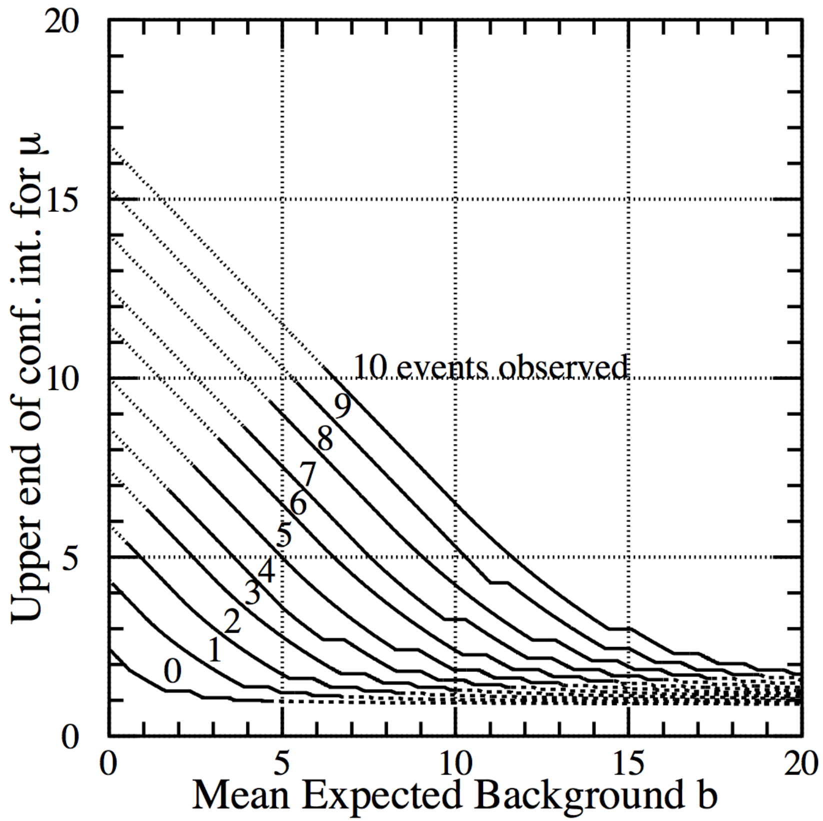


FIGURE 20. 90% limit s_{up} (Upper end of confidence interval for μ in the figure) vs. b for different values of n_0. These are the upper limits resulting from a frequentist treatment in the framework of the so called "Unified approach". The dotted portions of the lines correspond to configuration where central intervals rather than upper limits should be given. The dashed portions of the lines correspond to very unlikely configuration (very small n_0 when b is quite large, so that p(n_0) is below 1%). (taken from G.Feldmann, R.Cousins, Phys.Rev.D57 (1998) 3873)

CLs method

7.4. **A modified frequentist approach: the CL_s method.** Now we consider a method, developed in the last years and applied in many analyses especially from LHC experiments, including the search for the Higgs boson. It is the **modified frequentist** approach to the problem of setting upper/lower limits in search experiments.

n_i events and expected events $y_i = \mu s_i + b_i$

Signal strength $\mu = \frac{\sigma}{\sigma_{th}}$ Theory expectation μ=1

$$L(\underline{n}/\mu, \underline{\theta}) = \prod_{i=1}^{M} \frac{(\mu s_i + b_i)^{n_i} e^{-(\mu s_i + b_i)}}{n_i!}$$

CLs method

Add histogram of control regions, mj , background enriched

$E[m_j] = u_j(\underline{\theta})$ depending on the nuisance parameters (and not on μ)

$$L(\underline{n}/\mu, \underline{\theta}) = \prod_{i=1}^{M} \frac{(\mu s_i + b_i)^{n_i} e^{-(\mu s_i + b_i)}}{n_i!} \prod_{j=1}^{K} \frac{u_j^{m_j} e^{-u_j}}{m_j!}$$

CLs method

Define the test statistics

$$q_\mu = -2\ln\frac{L(\mu,\hat{\hat{\underline{\theta}}})}{L(\hat{\mu},\hat{\underline{\theta}})}$$

profile likelihood ratio

symbols: $\hat{\mu}$ and $\hat{\underline{\theta}}$ are the best values of the parameters obtained by maximizing L; $\hat{\hat{\underline{\theta}}}$ are the values of the nuisance parameters obtained by maximizing L at μ fixed. The test

CLs method

7.4.2. *Discovery.* In order to falsify a null hypothesis H_0 we need to test the background-only hypothesis. This can be done by using the test statistics q_0, that is eq. 207 for $\mu = 0$

$$q_0 = -2 \ln \frac{L(0, \hat{\hat{\underline{\theta}}})}{L(\hat{\mu}, \hat{\underline{\theta}})} \tag{210}$$

If we call q_0^{obs} the value of q_0 obtained using the data, we can easily define a p-value

$$p_0 = \int_{q_0^{obs}}^{\infty} f(q_0/0) dq_0 \tag{211}$$

that, for what we have seen in the previous paragraph, is essentially a χ^2 test. If p_0 is below the defined limit we falsify the hypothesis and we have done the discovery.

CLs method

7.4.3. *Signal exclusion: CL_{s+b}.* We consider now how the test statistics shown in eq. 207 can be used for the exclusion of a given theory. Eq. 207 is rewritten with $\mu = 1$[36]

$$q_1 = -2 \ln \frac{L(1, \underline{\hat{\hat{\theta}}})}{L(\hat{\mu}, \underline{\hat{\theta}})} \tag{212}$$

The lower is q_1, the more compatible the data are with the theory, and the less compatible the data are with the pure background expectations. The pdf of q_1 can be evaluated starting from MC samples, either generated with $\mu = 1$ or for samples of pure background events generated with $\mu = 0$. We call respectively $f(q_1/1)$ and $f(q_1/0)$ the two pdf's. A graphical example of these pdf's is shown in Figure 22. The separation between the two pdf's determines the capability to discriminate the searched model with respect to the background[37].

CLs method

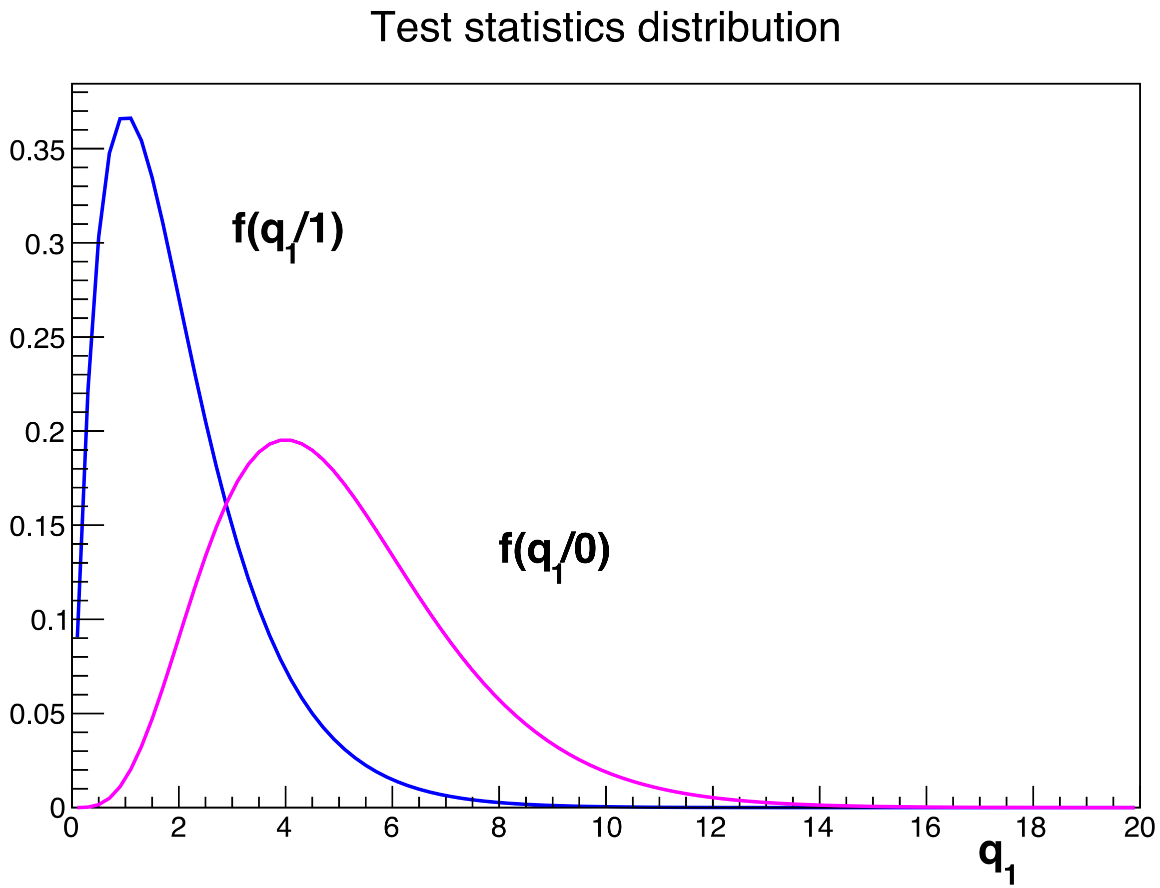


FIGURE 22. Example of q_1 distributions in the two hypotheses, namely $\mu = 1$ and $\mu = 0$. The separation between the two distributions indicate the capability to discriminate the two hypotheses.

CLs method

evaluate the **sensitivity** of the experiment.

define $\tilde{q}_1$ as the **median** of the $f(q_1/0)$ function[3]

expected

$$CL_{s+b}^{exp} = \int_{\tilde{q}_1}^{\infty} f(q_1/1)dq_1$$

CLs method

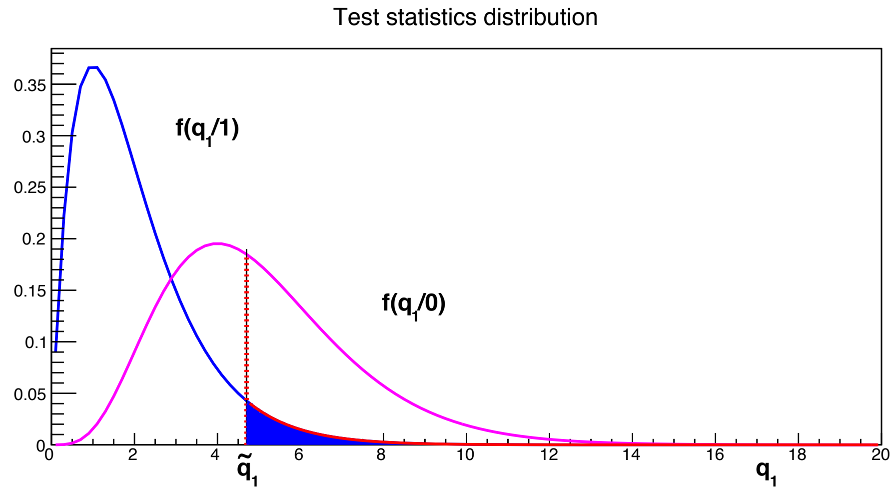


FIGURE 23. For the same example of alternative hypotheses shown in Fig. 22: construction of CL_{s+b}^{exp} (upper plot) and of CL_{s+b}^{obs} (lower plot). In both cases the CL is given by the blue area. In the upper plot the median q_1 from background experiments is indicated as $\tilde{q}_1$; in the lower plot the q_1 obtained by data is indicated as q_1^{obs}.

CLs method

However, we have determined the median CL only. In actual background-only experiments, we will have background fluctuations, in such a way that q_1 values will be obtained distributed according to $f(q_1/0)$. So we can evaluate an interval of confidence levels, by repeating the procedure above for two positions of q_1 , $\tilde{q}_1^{(1)}$ and $\tilde{q}_1^{(2)}$ such that respectively:

$$\int_{-\infty}^{\tilde{q}_1^{(1)}} f(q_1/0)dq_1 = \frac{1-\beta}{2} \tag{214}$$

$$\int_{-\infty}^{\tilde{q}_1^{(2)}} f(q_1/0)dq_1 = \frac{1+\beta}{2} \tag{215}$$

with e.g. $\beta = 68.3\%$ to have a gaussian one-std.deviation interval. Confidence levels are then evaluated applying eq. 213 to $\tilde{q}_1^{(1)}$ and $\tilde{q}_1^{(2)}$.

CLs method

Observation

$$CL_{s+b}^{obs} = \int_{q_1^{obs}}^{\infty} f(q_1/1)dq_1 \tag{216}$$

and this is the **observed** confidence level. If it is below, say 5% we exclude the signal at 95% CL.

CLs method

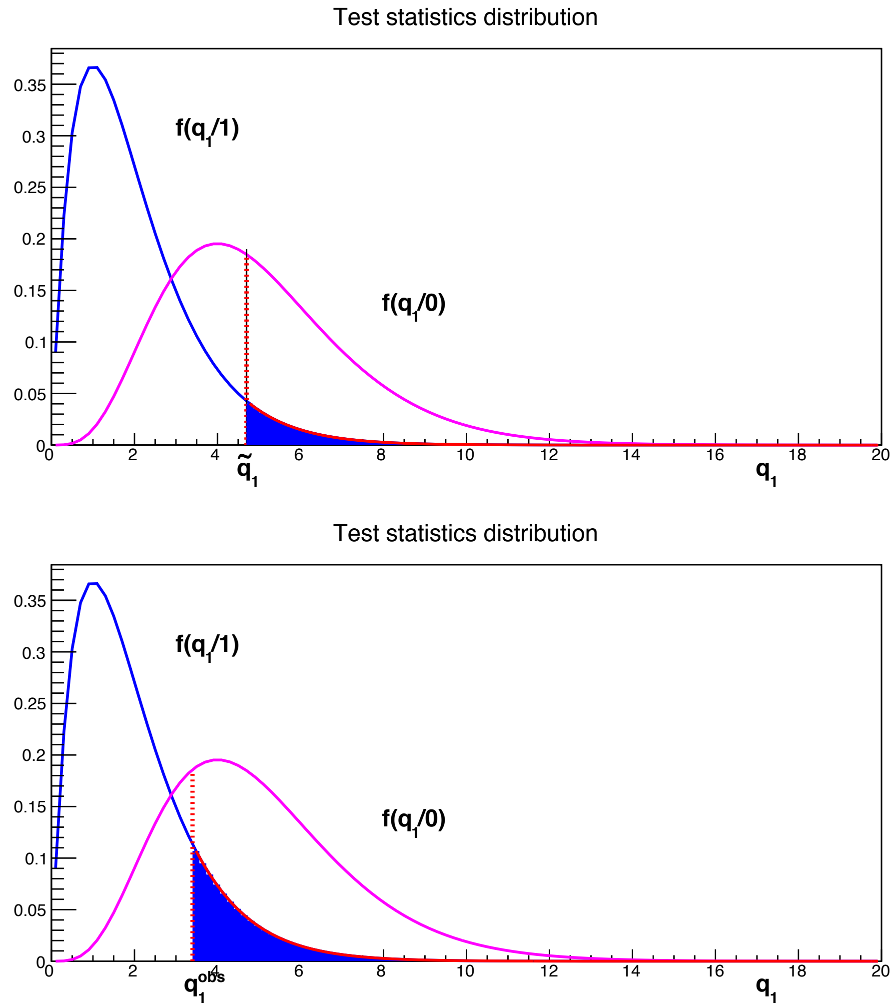


FIGURE 23. For the same example of alternative hypotheses shown in Fig. 22: construction of CL_{s+b}^{exp} (upper plot) and of CL_{s+b}^{obs} (lower plot). In both cases the CL is given by the blue area. In the upper plot the median q_1 from background experiments is indicated as $\tilde{q}_1$; in the lower plot the q_1 obtained by data is indicated as q_1^{obs}.

CLs method

7.4.4. *Signal exclusion: CL_s.* A problem in the procedure outlined in the previous section has been put in evidence, and a correction to it, the so called modified frequentist approach has been proposed. We discuss now this method, also called CL_s method that is now widely employed for exclusion of new physics signals.

Let's consider the situation shown in Figure 24 where the two pdf's $f(q_1/0)$ and $f(q_1/1)$ have a large overlap signaling a small sensitivity. If we evaluate in this situation CL_{s+b}^{exp} we find a large value, so that we do not expect to be sensitive to exclusion. However what happens if q_1^{obs} is the one shown in the same Figure ? The observed CL_{s+b}^{obs} is well below 5% and the signal has to be excluded at 95% CL. But, are we sure that we have to exclude it ? In the same Figure the quantity CL_b^{obs} is reported:

$$CL_b^{obs} = \int_{q_1^{obs}}^{\infty} f(q_1/0)dq_1 \tag{217}$$

that is also very small in this case. Apparently the signal is small and the background "under-fluctuates", so that q_1^{obs} is scarcely compatible with the signal+background hypothesis but also with the background-only hypothesis. So, we are excluding the signal, essentially because the background has fluctuated.

In order to avoid this somehow unmotivated exclusion, the CL_s procedure has been defined. The idea is to use, as confidence level, the CL_s quantity, either expected or

CLs method

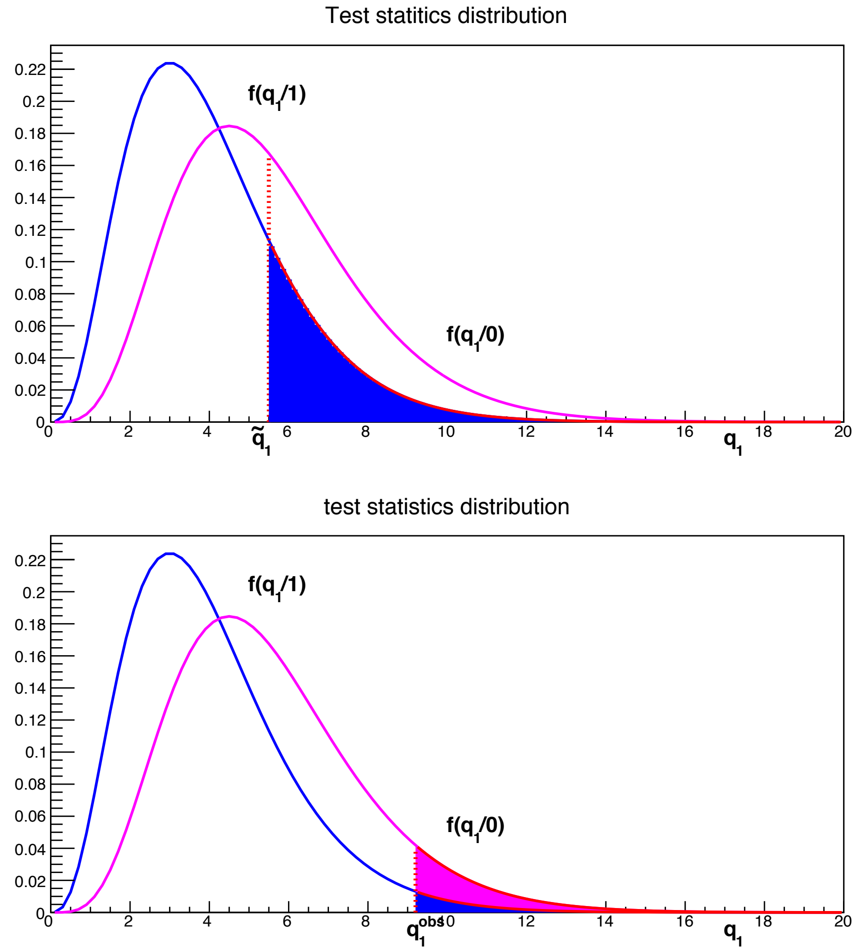


FIGURE 24. Same construction of Fig. 23 for a situation where the discrimination between the two hypotheses is particularly poor and the overlap between the two distributions is high. The CL_{s+b}^{exp} is high (upper plot) but for a particular experiment with a under fluctuation of the background the CL_{s+b}^{obs} can be small in such a way to reject the signal hypothesis (lower plot). In the lower plot the magenta area shows CL_b^{obs} from which CL_s is built. In this case using the CL_s prescription rather than the CL_{s+b} one the signal is not rejected.

CLs method

$$CL_s = \frac{CL_{s+b}}{CL_b}$$

The CL_s method is also said **modified frequentist** approach. In fact, the confidence interval obtained in this way has not the coverage properties required by the ”orthodox” frequentist paradigm. So if we build a confidence interval with a CL_s of α, the coverage is in general larger than α, so that the Type-I errors are less than $1 - \alpha$.

Basic Definitions: POWER

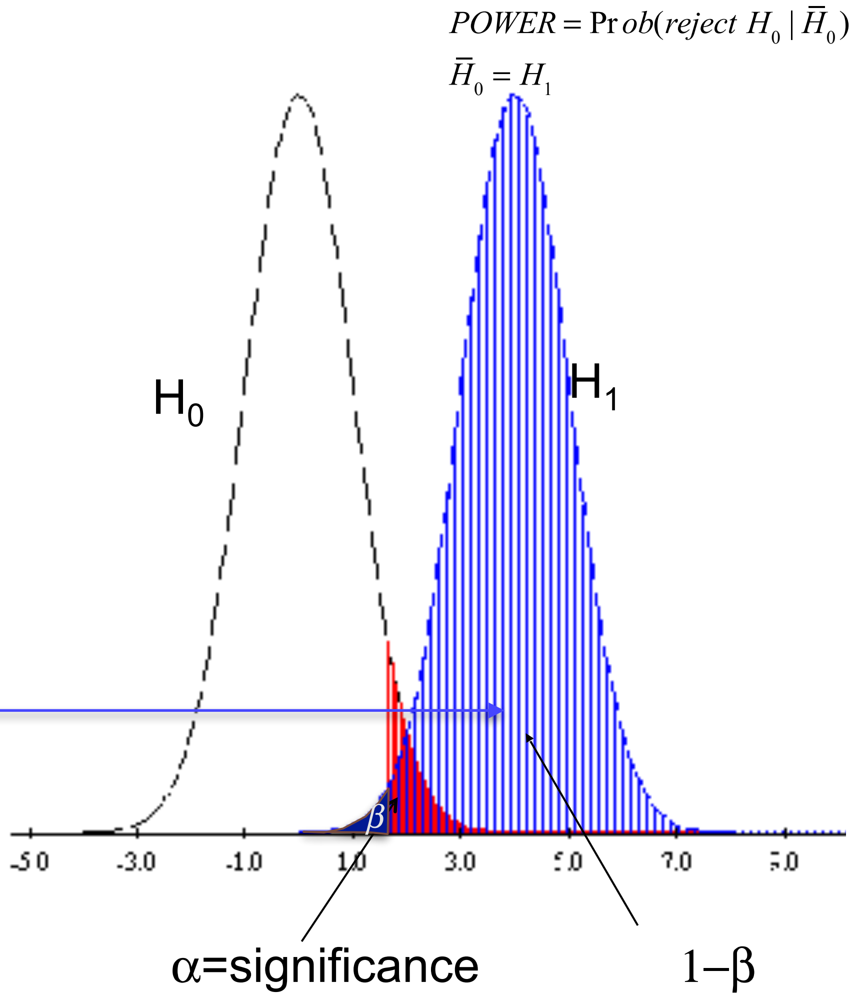
- $\alpha = \Pr ob(reject\ H_0 \mid H_0)$
- The POWER of an hypothesis test is the probability to reject the null hypothesis when it is indeed wrong (the alternate analysis is true)
- $POWER = \Pr ob(reject\ H_0 \mid \bar{H}_0)$

  $\beta = Prob(accept\ H_0 \mid \bar{H}_0)$

  $1 - \beta = Prob(reject\ H_0 \mid \bar{H}_0)$

  $\bar{H}_0 = H_1$

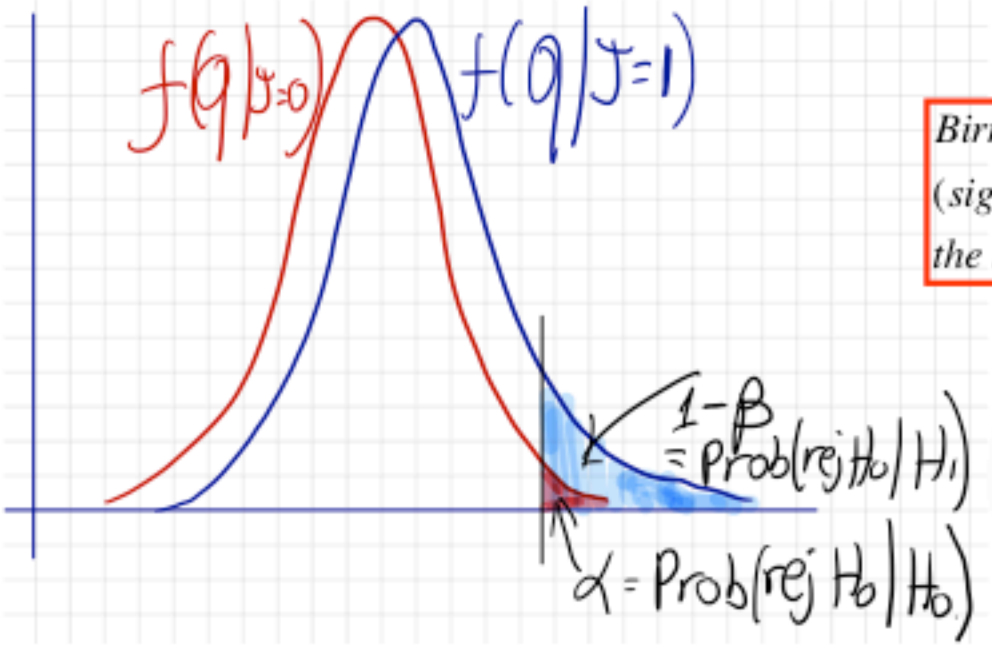
  $1 - \beta = Prob(reject\ H_0 \mid H_1)$
- The power of a test increases as the rate of type II error decreases



CLs

Birnbaum (1977)

"A concept of statistical evidence is not plausible unless it finds 'strong evidence for H_1 as against H_0' with small probability (α) when H_0 is true, and with much larger probability ($1-\beta$) when H_1 is true."



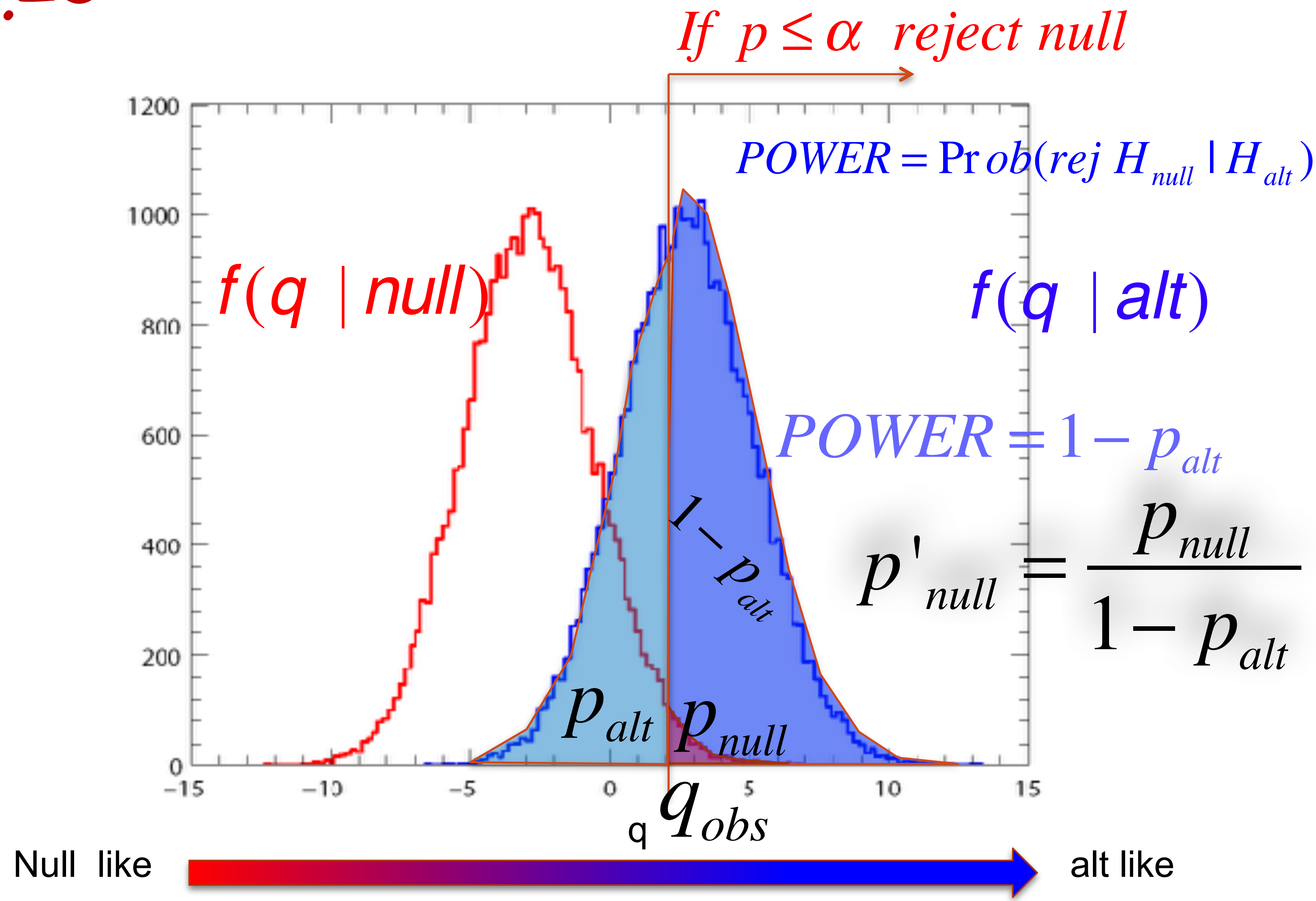
Birnbaum (1962) *suggested that* $\alpha / 1-\beta$ (*significance / power*) *should be used as a measure of the strength of a statistical test, rather than* α *alone*

$$p = 5\% \rightarrow p' = 5\% / 0.155 = 32\%$$

$$p' \equiv CL_S$$

$$p'_\mu = \frac{p_\mu}{1-p_0}$$

CLs



If $p \le \alpha$ reject null
$POWER = \Pr ob(rej\ H_{null} \mid H_{alt})$
$f(q \mid null)$
$f(q \mid alt)$
$POWER = 1 - p_{alt}$
$1 - p_{alt}$
$p'_{null} = \frac{p_{null}}{1 - p_{alt}}$
p_{alt}
p_{null}
1200
1000
800
600
400
200
0
−15
−10
−5
0
5
10
15
q
q_{obs}
Null like
alt like

p-value – testing the null hypothesis

When testing the b hypotheis (null=b), it is custom to set

$\alpha = 2.9\ 10^{-7}$

→ if $p_b < 2.9\ 10^{-7}$ the b hypothesis is rejected

→Discovery

When testing the s+b hypothesis (null=s+b), set α =5%
if p_{s+b}<5% the signal hypothesis is rejected at the 95% Confidence Level (CL)

→ Exclusion