Image credit: Marguerite Tonjes

Statistics *speed-run*

Nick Smith

Higgs Combine Tutorial

3 Feb 2023

Goal

To give you plausible familiarity with sentences like:

"An observed (expected) upper limit is placed on the signal strength μ, using the profile likelihood ratio test statistic, following the CL_s criterion, under asymptotic assumptions, and found to be …"

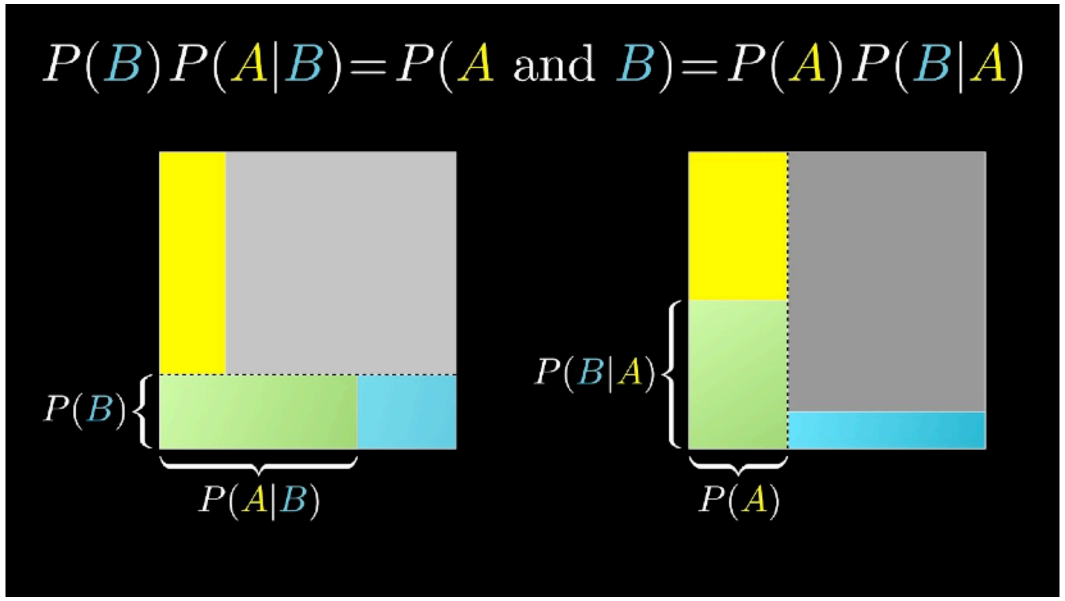
Plan:

- Probability, likelihood, and inference
  - Bayesian inference
  - Maximum likelihood point estimation
- Frequentist hypothesis tests
  - Neyman interval
  - Likelihood ratio test statistic
  - Under-fluctuation and CLs
  - Asymptotic behavior
- Adding uncertainties
  - Statistical model with auxiliary measurements
  - Profiling and marginalizing nuisance parameters

Probability, likelihood, and inference

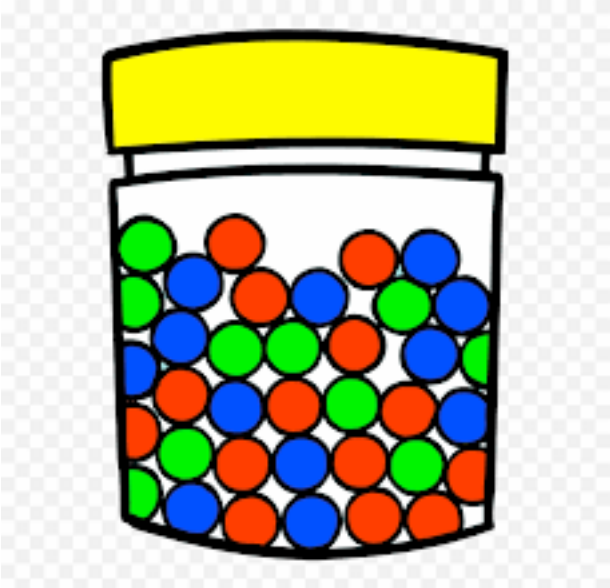
Probability

- Kolmogorov axioms: for a sample space S, we have
  - $\forall A \subset S \quad P(A) \geq 0$
  - $\forall A, B \subset S, A \cap B = \emptyset \quad P(A \cup B) = P(A) + P(B)$
  - $P(S) = 1$
- Conditional probability
  - $$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
  - Think: probability of A given *fixed* B
- Bayes' theorem
  - $$P(B \mid A) = \frac{P(A \mid B) P(B)}{P(A)}$$
- More exposition:
  - PDG review
  - 3blue1brown on YouTube

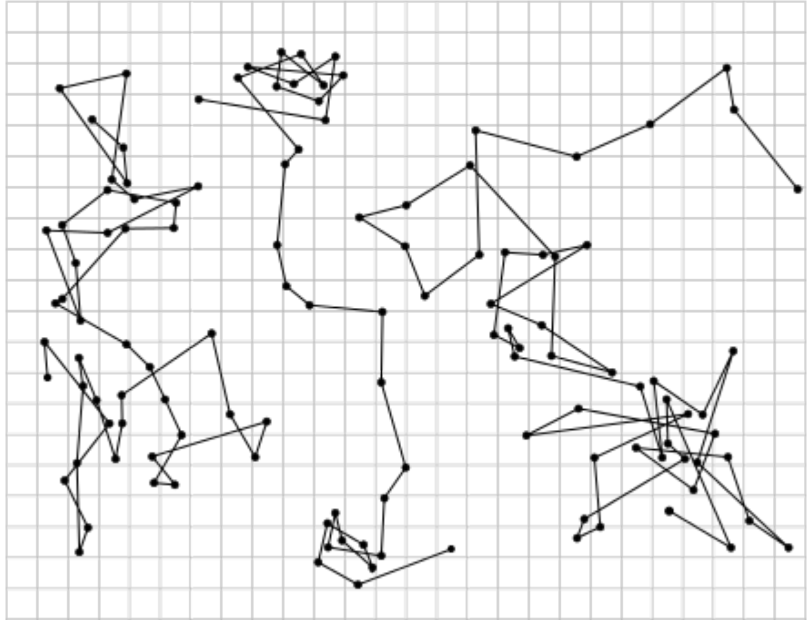


Fermilab

Probability density / mass



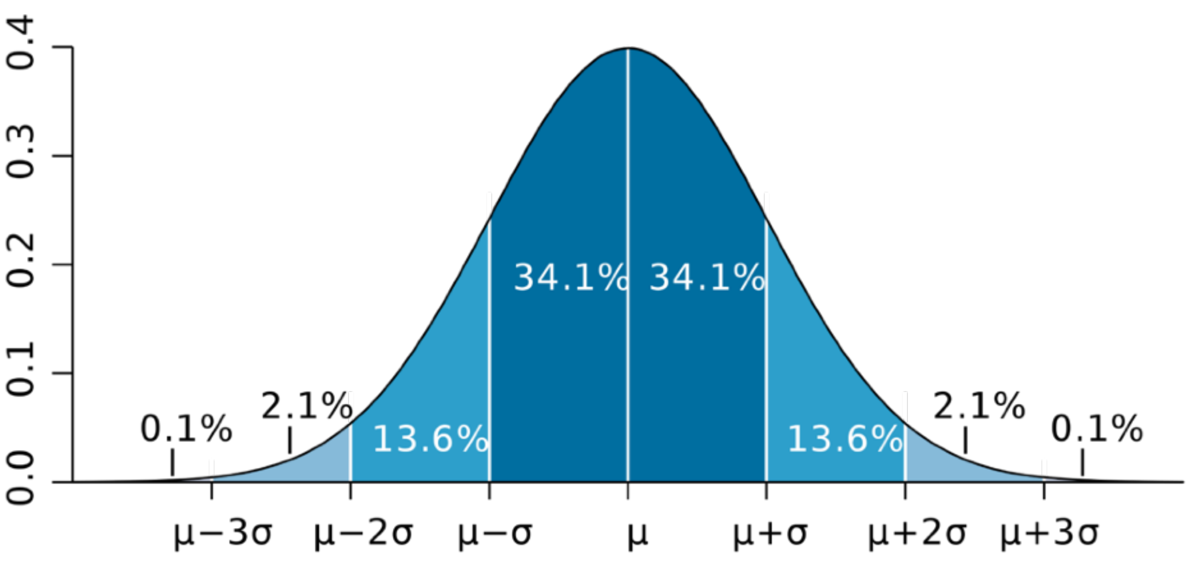
- Probability mass function (pmf)
  - probability of observing a specific outcome
- Probability density function (pdf), e.g. $P(x)$
  - $P(x)\,dx$ = *differential* probability of observing an outcome
- In both cases:
  - defined over a space of outcomes/observables/samples
  - imply a cumulative (cdf), percentile (inverse cdf), etc. in 1D
  - may be parameterized



Examples

- Marbles: P(draw 2 red, 2 green, 1 blue from jar)
  - pmf [Multivariate hypergeometric distribution]
- Brownian motion: P(displacement after some time)
  - pdf [Normal distribution]
- Counts in a particle detector after some time
  - pmf [Poisson distribution]

Distributions of interest



- Poisson distribution
  - $P(n\,|\,\lambda) = \dfrac{\lambda^n e^{-\lambda}}{n!}$
- Normal distribution
  - $P(x\,|\,\mu,\sigma) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  - Central limit theorem:
    - sums of independent random-distributed variables tend towards a Normal-distributed variable
  - Standard (Z) score:
    - Convention for interesting percentiles: "1σ" = 0.6827…, "2σ" = 0.9545…, "5σ" = 5.7e-7
      - These are 2-sided. Can also define 1-sided (half)
    - Often quote 95 %-ile
- Log-normal distribution
  - Definition: Normal in log-space (change of variables: $y = \ln(x), dy = x^{-1}dx$)
  - Corollary to central limit theorem:
    - *products* of […] tend towards a Log-normal distributed variable
    - Common model for calibration uncertainties (more later)

Poisson process

- In CMS, collision events occur at a rate $\lambda(x,t) = L(t)\,\sigma_{pp\to X}(x)\,\epsilon(x,t)$
  - Where (for now ignoring the model parameters)
    - $L(t)$ is the instantaneous luminosity
    - $\sigma_{pp\to X}$ is some cross section (differential w.r.t. observables x)
    - ϵ is our detector acceptance/efficiency (hopefully mild t-dependence!)

Poisson process

- In CMS, collision events occur at a rate $\lambda(x, t) = L(t)\, \sigma_{pp\to X}(x)\, \epsilon(x, t)$
  - Where (for now ignoring the model parameters)
    - $L(t)$ is the instantaneous luminosity
    - $\sigma_{pp\to X}$ is some cross section (differential w.r.t. observables x)
    - ϵ is our detector acceptance/efficiency (hopefully mild t-dependence!)
- Integrate over some region B (“a bin”) to get a Poisson PMF
  - in time and “observable space” (e.g. muon 4-momentum, etc.)
  - $$\Lambda_i = \int_{B_i} \lambda(x, t) dx dt, \qquad P(N_i \,|\, \Lambda_i) = \frac{\Lambda_i^{N_i} e^{-\Lambda_i}}{N_i!}$$

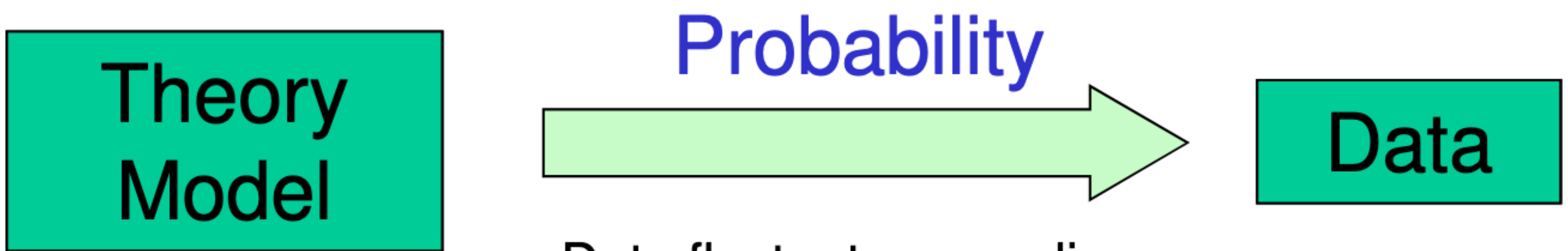
Poisson process

- In CMS, collision events occur at a rate $\lambda(x, t) = L(t)\, \sigma_{pp\to X}(x)\, \epsilon(x, t)$
  - Where (for now ignoring the model parameters)
    - $L(t)$ is the instantaneous luminosity
    - $\sigma_{pp\to X}$ is some cross section (differential w.r.t. observables x)
    - ϵ is our detector acceptance/efficiency (hopefully mild t-dependence!)
- Integrate over some region B ("a bin") to get a Poisson PMF
  - in time and "observable space" (e.g. muon 4-momentum, etc.)
  - $\Lambda_i = \int_{B_i} \lambda(x, t) dx dt, \qquad P(N_i \,|\, \Lambda_i) = \frac{\Lambda_i^{N_i} e^{-\Lambda_i}}{N_i!}$
- This is a Poisson Process
  - Binned model: overall PDF is a joint distribution (product) over disjoint regions
    - $P(\text{data} \,|\, \text{model}) = \prod_i P(N_i \,|\, \Lambda_i)$
  - Un-binned model: conditional on N, λ can be interpreted as a PDF (integrating t)
    - $P(\text{data} \,|\, \text{model}) = P(N \,|\, \Lambda) \prod_i^N \lambda(x_i) dx_i$

(actually just the limit as bin size goes to zero: R. Barlow, "Extended maximum likelihood")

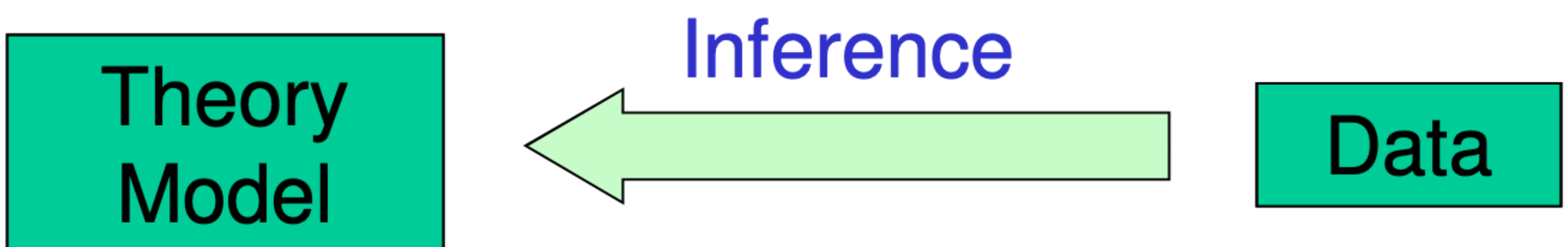
Fermilab

Inference



Data fluctuate according to process randomness

$$x_{obs} \sim P(x \mid \theta)$$

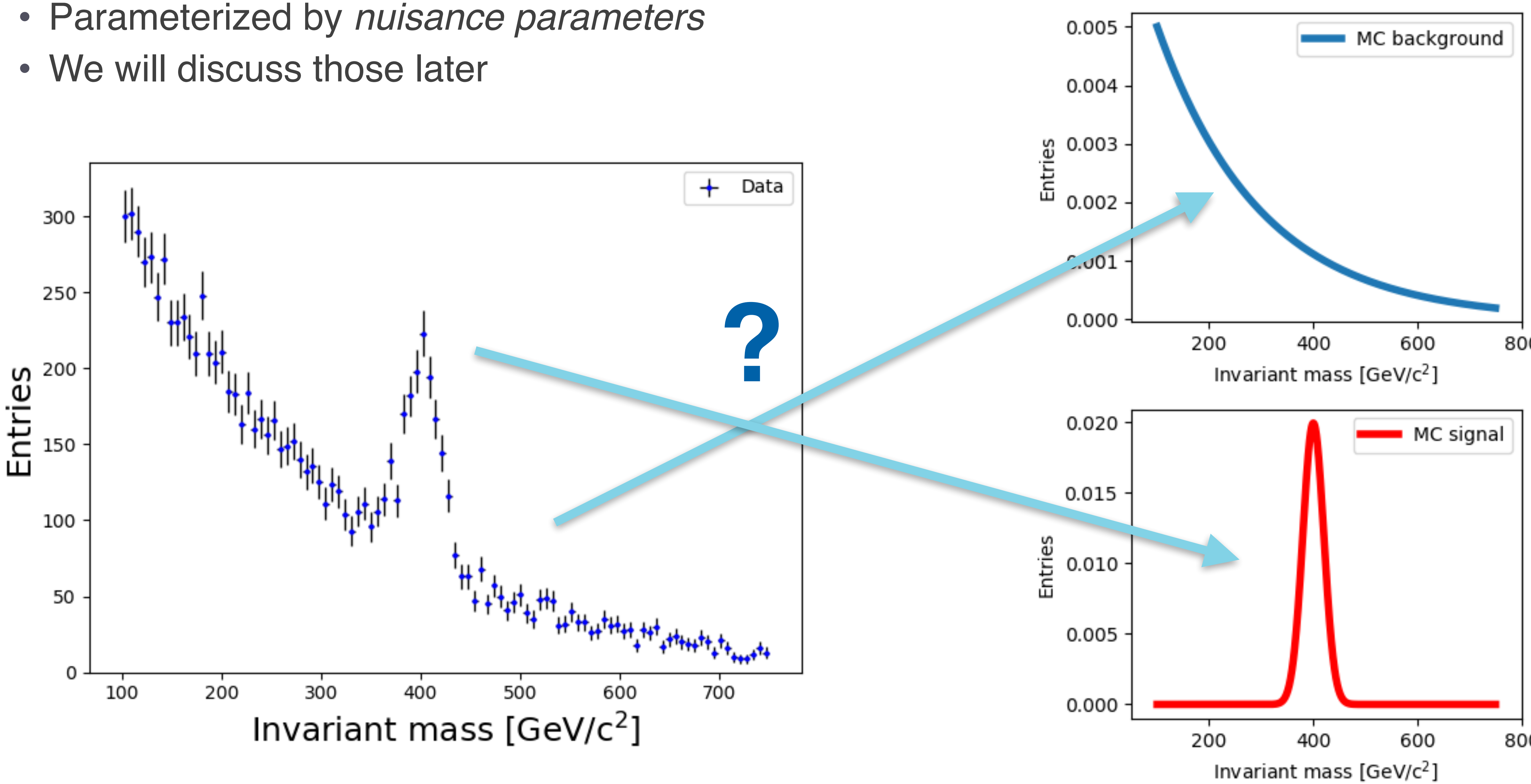


Model uncertainty due to fluctuations of the data sample

$$P(\theta \mid x_{obs})??$$

Inference example

- Given this data and a model for signal and background, I might infer:
  - The amount of signal present (a *parameter of interest*, or POI)
  - The functional form of the background, if a-priori unknown
    - Parameterized by *nuisance parameters*
    - We will discuss those later



Fermilab

Inference for full simulation

- The whole picture is more complex
  - We often cannot compute $P(x\,|\,\theta)$, but we can efficiently sample it
    - ➜ surrogate model using Monte Carlo (MC) estimates of bin yields

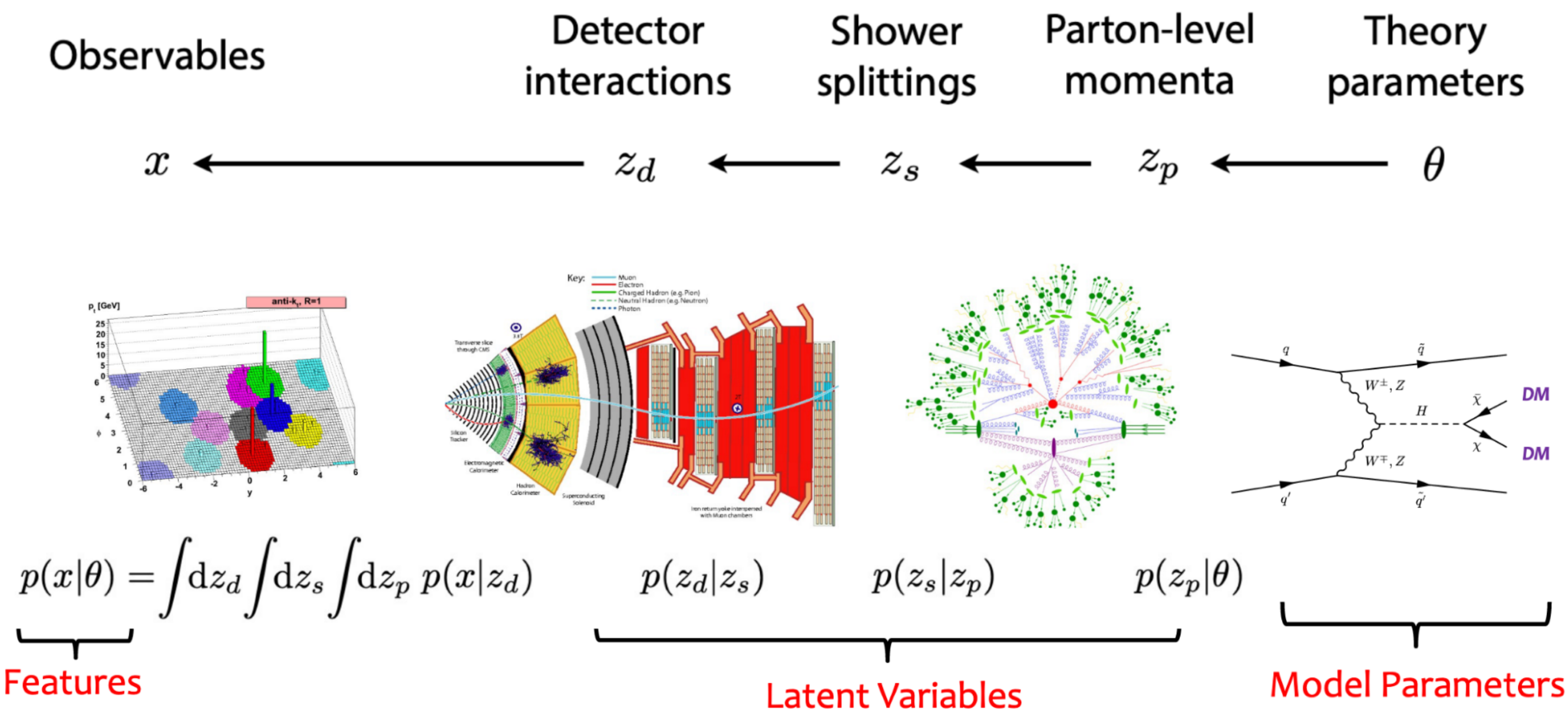
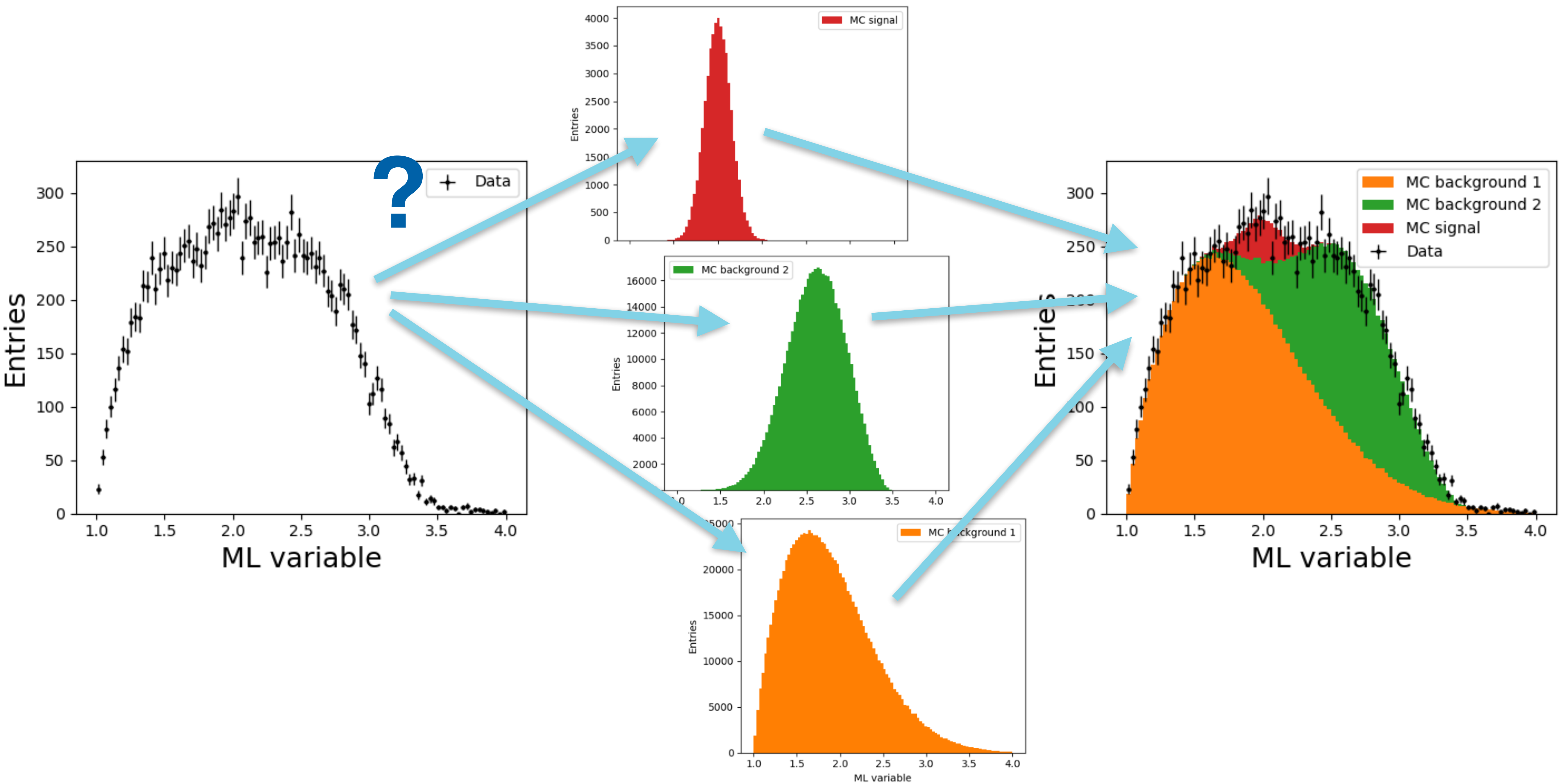


diagram: K. Cranmer

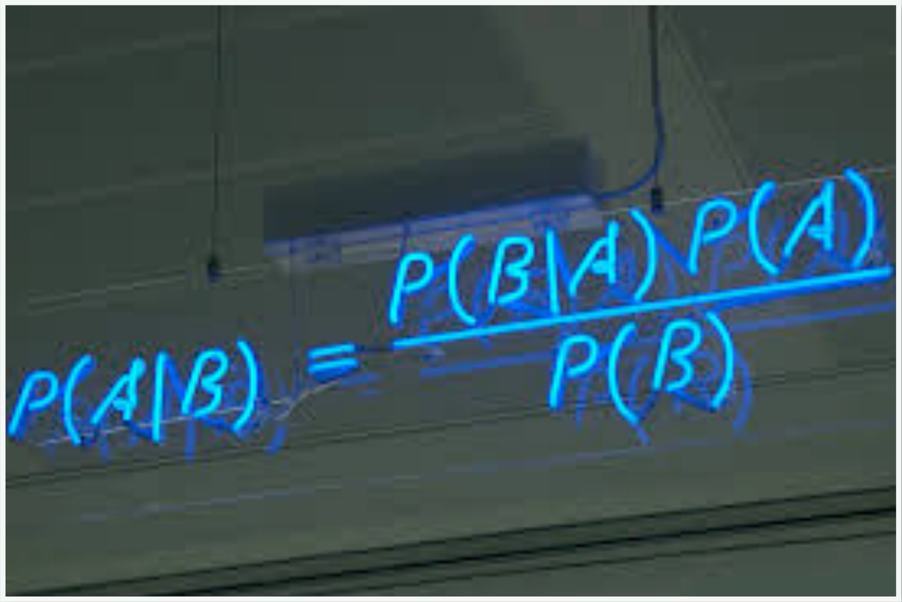
Templates

- We often build models via template histograms derived from MC
  - Typically to infer signal strength μ = normalization of signal template



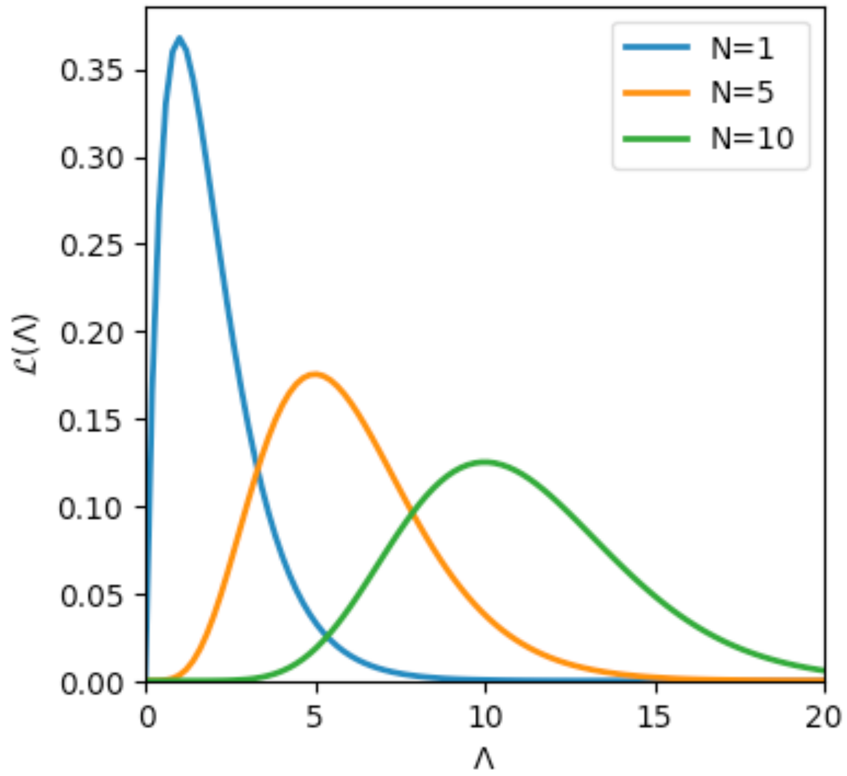
Bayesian inference

- We would like to infer $P(\theta \mid x_{obs})$
  - i.e. given our observation x, what is the probability distribution of model parameter θ?
- Bayes' theorem tells us:
  - $$P(\theta \mid x_{obs}) = \frac{P(x_{obs} \mid \theta)P(\theta)}{P(x_{obs})}$$
  - Ok, we have our model $P(x \mid \theta)$, but what about the other terms?
- $P(\theta)$ is the prior probability distribution for θ
  - Bayesian: we provide this based on our prior belief
    - Or some recipe (objective Bayes, etc.)
  - Frequentist: **no such thing!**
- $P(x_{obs})$ is the evidence
  - Here a normalization: $P(x_{obs}) = \int P(x_{obs} \mid \theta)P(\theta)d\theta$
    - In practice, can be hard to compute!
  - Better use found in Bayes' factor to compare models



Maximum likelihood point estimate

- For a fixed *observation*, can define likelihood $\mathscr{L}(\theta) = P(x_{obs} | \theta)$
  - This is a function of θ, but ***not** a probability density*
  - More exposition in [PDG 40.2]()
- $\hat{\theta}$ that maximizes this function is the *maximum likelihood estimate* (MLE)
  - $\hat{\theta} = \text{arg max}_\theta[\mathscr{L}(\theta)] = \text{arg min}_\theta[-\ln \mathscr{L}(\theta)]$
  - This is a random variable
  - We usually minimize the negative log-likelihood (NLL) numerically
    - the core job of MINUIT's MIGRAD routine
- Poisson example: $\hat{\Lambda} = N$



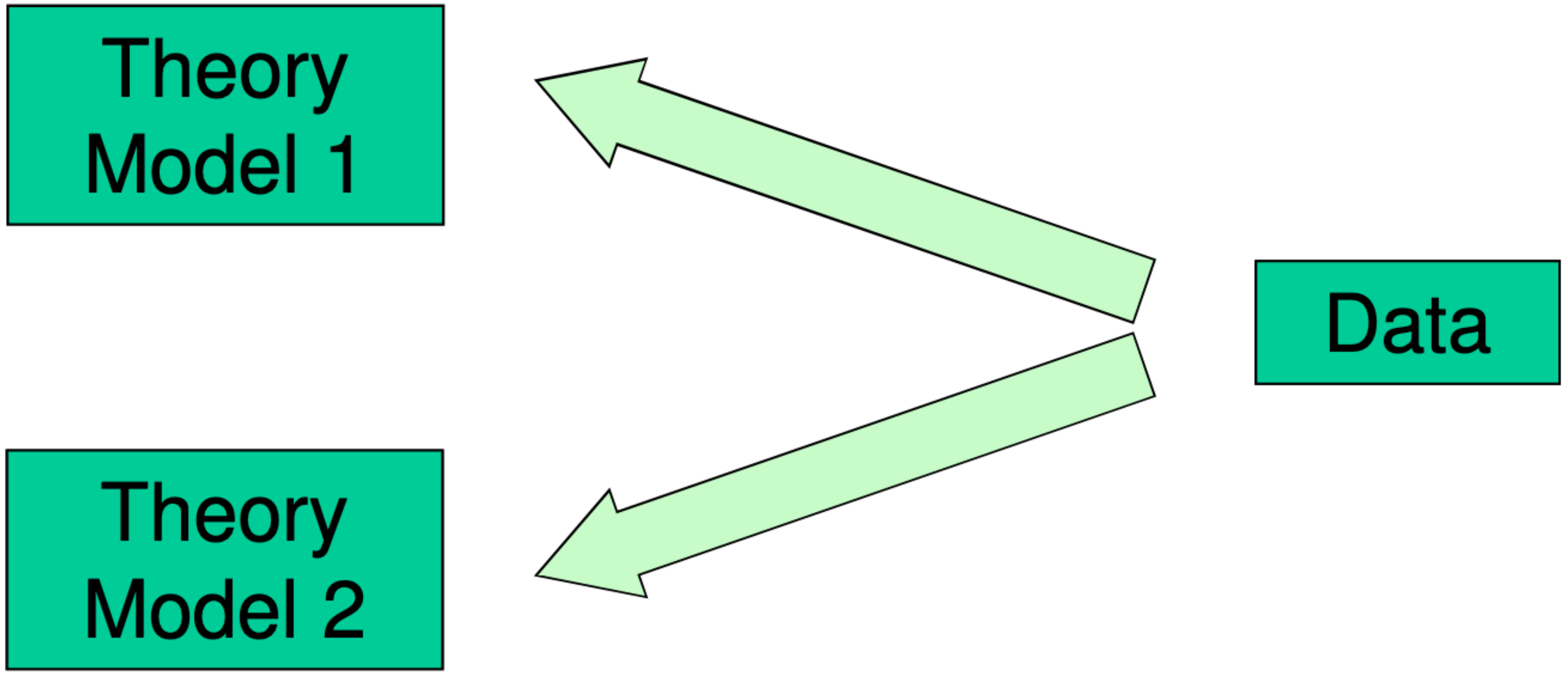
Maximum likelihood trivia

- Absolute value of $\mathscr{L}(\hat{\theta})$ is usually not meaningful
  - Can be calibrated by taking a ratio (chi-square, saturated binned model)
- The MLE has good limiting properties as sample size→∞
  - Consistent: sequence of MLEs converges to true value
  - Efficient: variance of MLE saturates the Cramér–Rao lower bound
    - Distribution of MLE approaches Normal distribution
  - Asymptotically unbiased
    - Bias can exist for finite samples, can be corrected (with increase in variance)
- The likelihood (and it's maximum) is invariant under change of variables
  - Again, it is not a PDF!

- Ok, but this is just a *point*. Can frequentists say more without a prior?

Frequentist hypothesis tests

Hypothesis tests



Which hypothesis is the most consistent with the experimental data?

Hypothesis tests

- Simple test parameterized by two probabilities: α, β
  - β with respect to an alternate hypothesis

Table of error types		**Null hypothesis (H_0) is**	
		TRUE	**FALSE**
Decision about null hypothesis (H_0)	**Don't reject**	Correct inference (true negative) (probability = $1-\alpha$)	Type II error (false negative) (probability = β)
	Reject	Type I error (false positive) (probability = α)	Correct inference (true positive) (probability = $1-\beta$)

Confidence sets

- One can ask, "assuming a value of θ, would x_{obs} be a likely outcome?"
  - This is a hypothesis test (likely or not)

Confidence sets

- One can ask, "assuming a value of θ, would x_{obs} be a likely outcome?"
  - This is a hypothesis test (likely or not)
- We have P(x|θ), so we can answer this if we:
  - Choose a significance level α of the test (e.g. 0.05)
  - Define a test statistic (ordering) of possible outcomes
  - Run pseudo-experiments (toys) for each θ to determine distribution of test statistic
  - Perform experiment, report set of θ where test statistic is below the 1-α quantile

Confidence sets

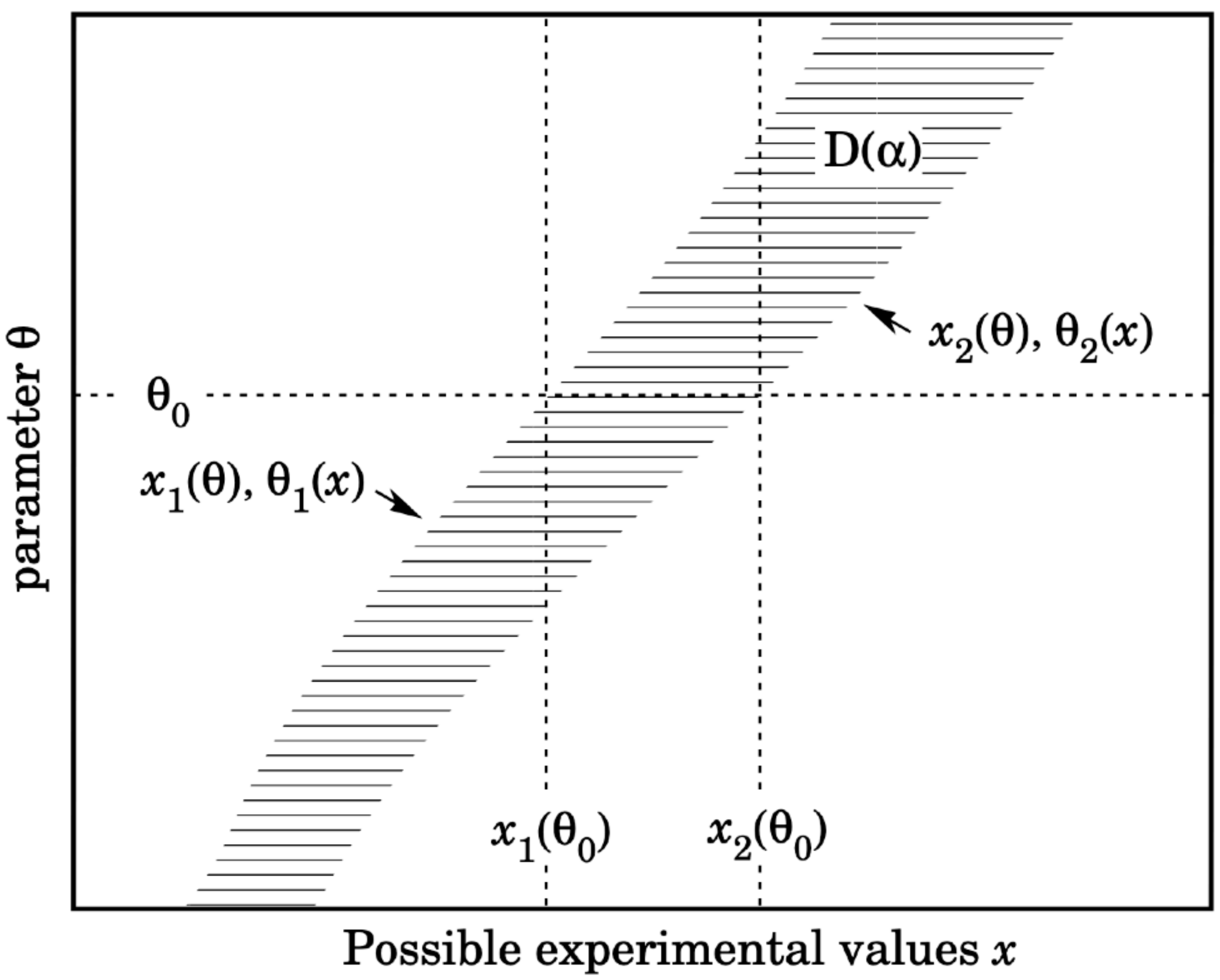
- One can ask, “assuming a value of θ, would x_{obs} be a likely outcome?”
  - This is a hypothesis test (likely or not)
- We have P(x|θ), so we can answer this if we:
  - Choose a significance level α of the test (e.g. 0.05)
  - Define a test statistic (ordering) of possible outcomes
  - Run pseudo-experiments (toys) for each θ to determine distribution of test statistic
  - Perform experiment, report set of θ where test statistic is below the 1-α quantile
- A good ordering will lead to
  - Good *coverage*: in repeated experiments the (unknown) θ_{true} will be in the set with probability at least 1-α, though it may over-cover
  - High *power* (1-β): the set does not contain θ_{alt} for some specified alternative hypothesis

Neyman interval

- Neyman construction (PDG, 40.4.2):
  - For each θ, find range $[x_1, x_2]$ s.t.

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} P(x\,|\,\theta)\,dx \geq 1 - \alpha$$

  - Perform experiment
  - Report confidence interval: $[\theta_1, \theta_2]$ where $x_{obs} \in [x_1(\theta), x_2(\theta)]$ for all $\theta \in [\theta_1, \theta_2]$



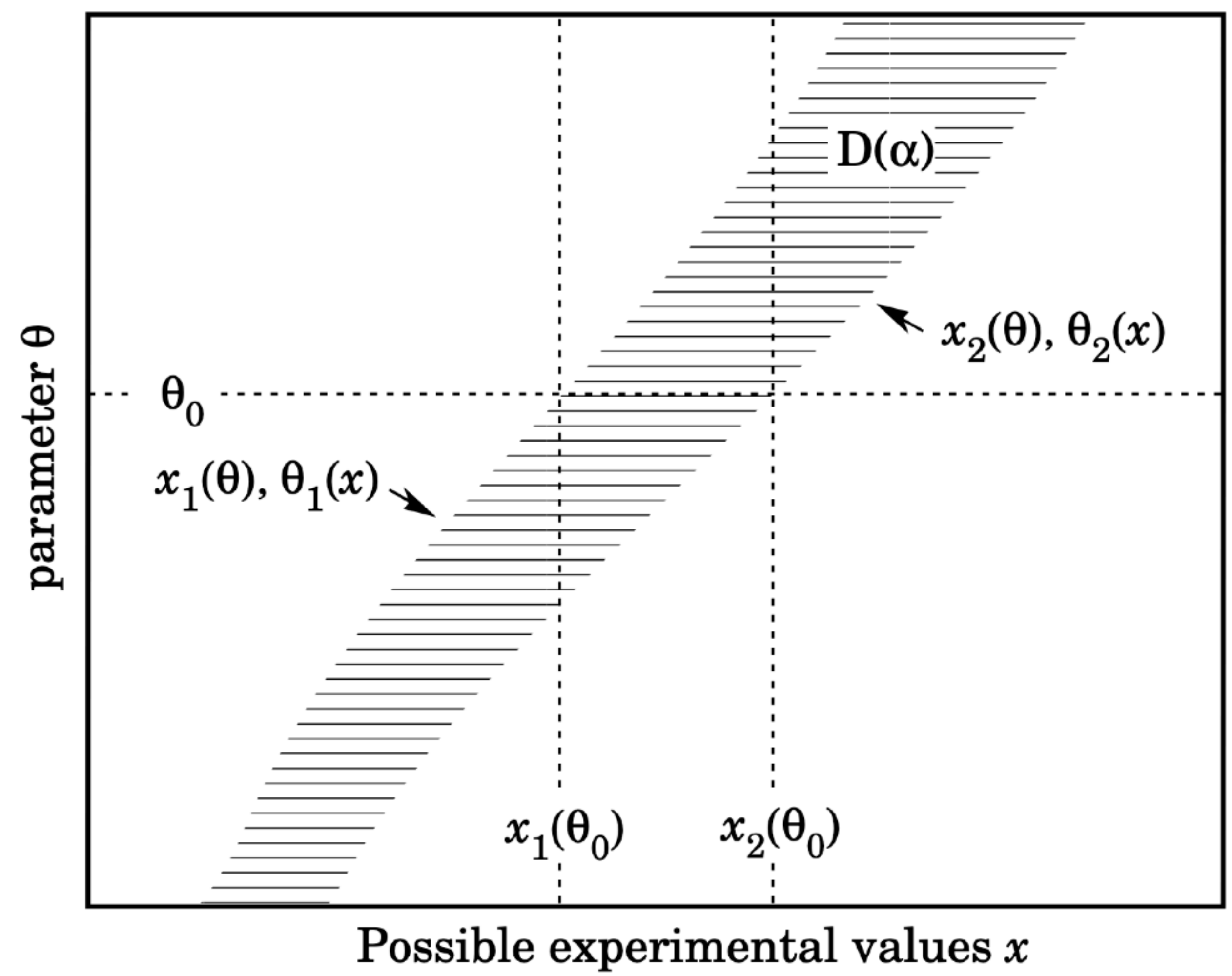
Fermilab

Neyman interval

- Neyman construction (PDG, 40.4.2):
  - For each θ, find range $[x_1, x_2]$ s.t.

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} P(x\,|\,\theta)\,dx \geq 1 - \alpha$$

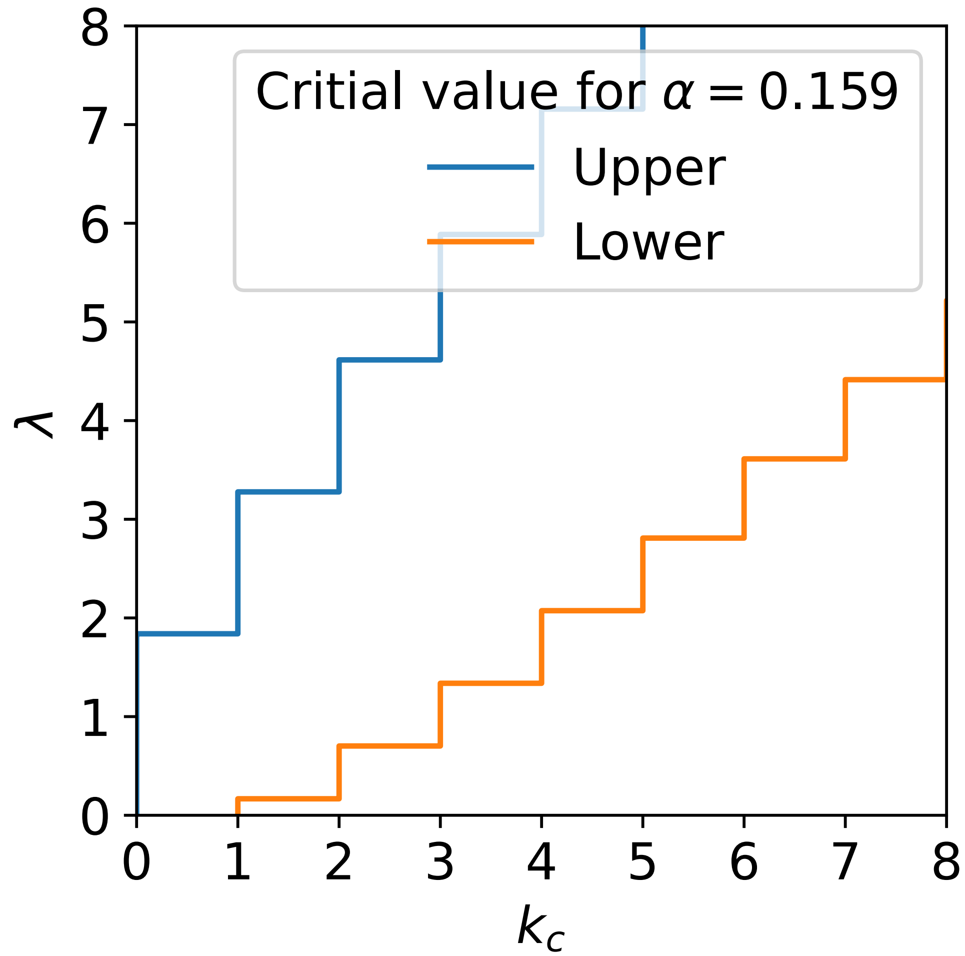
  - Perform experiment
  - Report confidence interval: $[\theta_1, \theta_2]$ where $x_{obs} \in [x_1(\theta), x_2(\theta)]$ for all $\theta \in [\theta_1, \theta_2]$

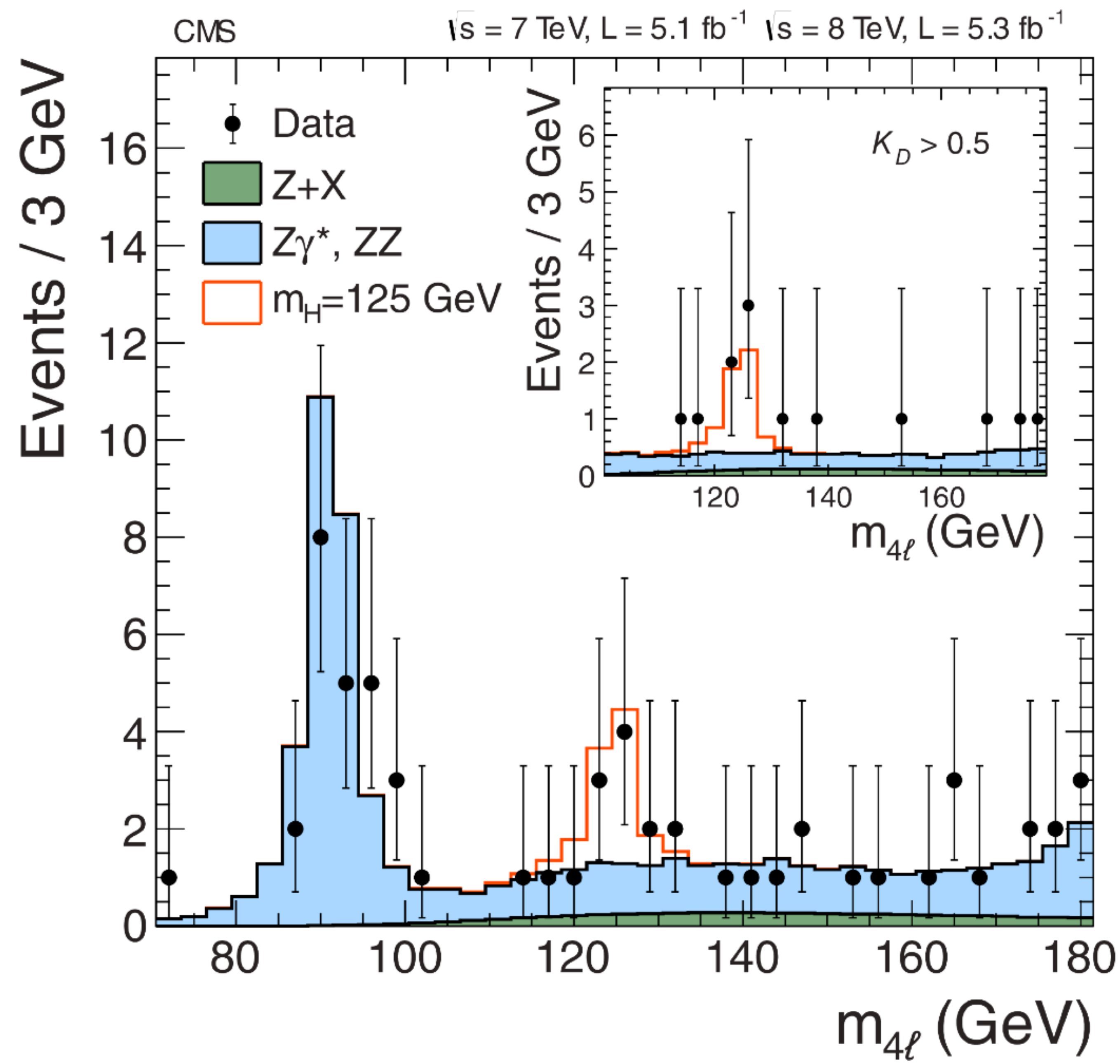


- Interval has coverage $1-\alpha$.
  - For an ensemble of experiments, the interval $[\theta_1, \theta_2]$ will contain (unknown) θ_{true} with probability $1-\alpha$. This is a statement about the distribution of θ_1 and θ_2, NOT θ_{true}.

Neyman interval

- You have all seen these: error bars on data points are the Neyman intervals for a Poisson distribution
  - With $\alpha = 0.159\ldots$
  - Also referred to as Garwood intervals





Likelihood ratio test

1. Define $t_\theta(x) = -2 \ln \dfrac{\mathscr{L}(\theta)}{\mathscr{L}(\hat{\theta})}$
2. Compute associated pdf (change of variables)

$$P(t_\theta \,|\, \theta') = \int \delta(t_\theta - t_\theta(x)) P(x \,|\, \theta') dx$$

3. For each θ, find the critical value $t_{\theta,c}$ s.t.

$$\int_{t_{\theta,c}}^{\infty} P(t_\theta \,|\, \theta')\, dt_\theta \leq \alpha$$

4. Perform experiment, get x_{obs}, report confidence set $\{\theta \,|\, t_\theta(x_{obs}) < t_{\theta,c}\}$

Again, the set is the random variate, and will contain (unknown) θ_{true} with probability 1-α.

Dimension of θ and x are arbitrary.

If θ is 1-d and t_θ is monotone, can make a Feldman-Cousins interval.

Likelihood ratio examples

$$t_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

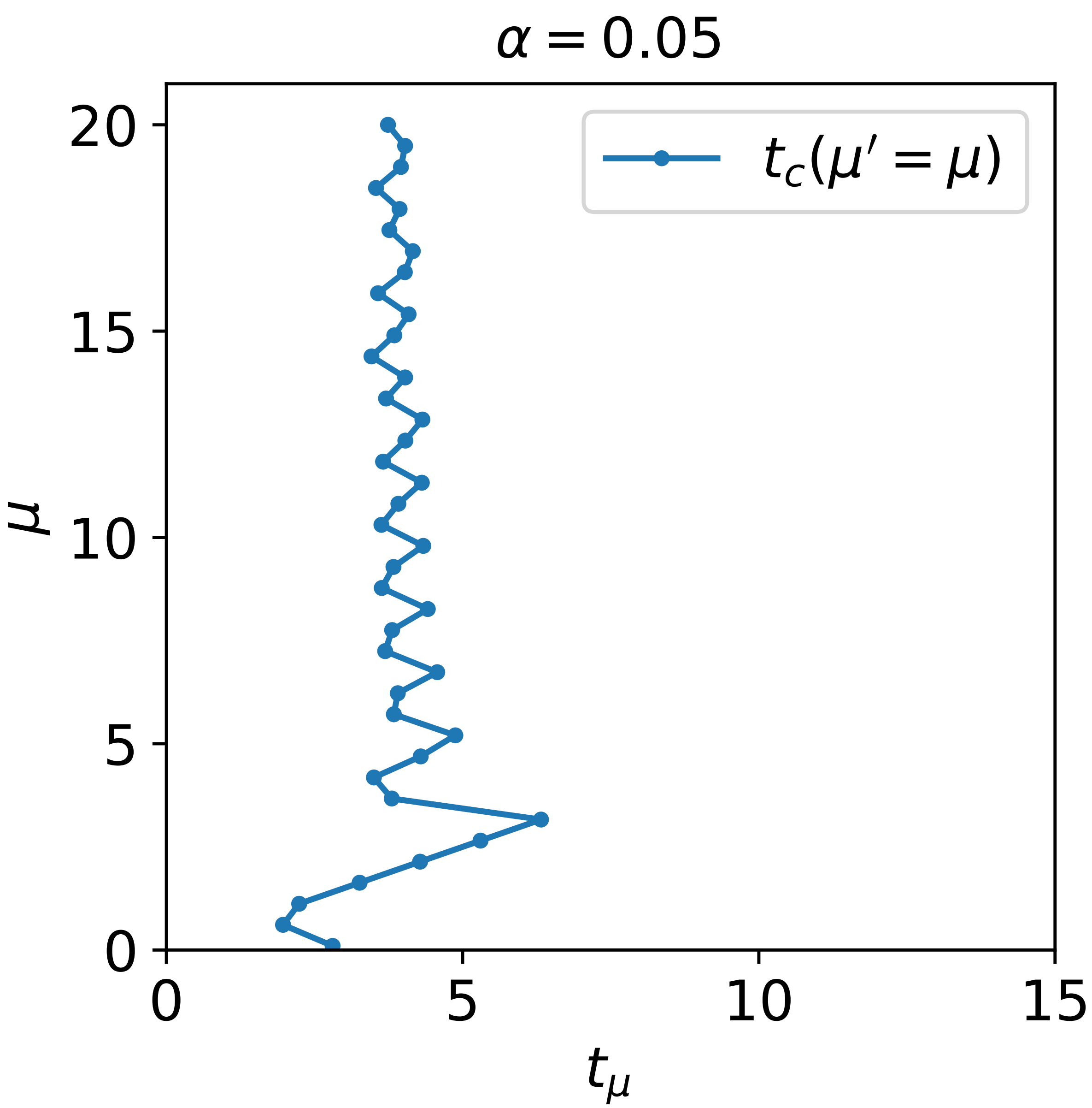
Likelihood ratio examples

- Poisson example
  - $P(x \mid \mu) = \dfrac{\mu^x e^{-\mu}}{x!}$

$$t_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Likelihood ratio examples

- Poisson example
  - $P(x\,|\,\mu) = \dfrac{\mu^x e^{-\mu}}{x!}$
- For each μ:
  - Throw 10k toys
  - Compute $\hat{\mu}$, t_μ
  - Find 0.95 quantile in distribution

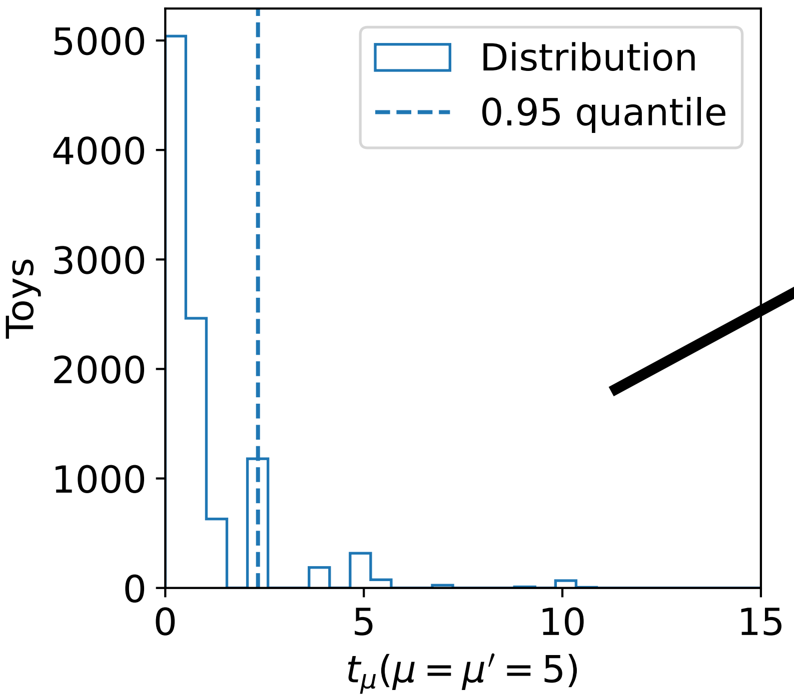


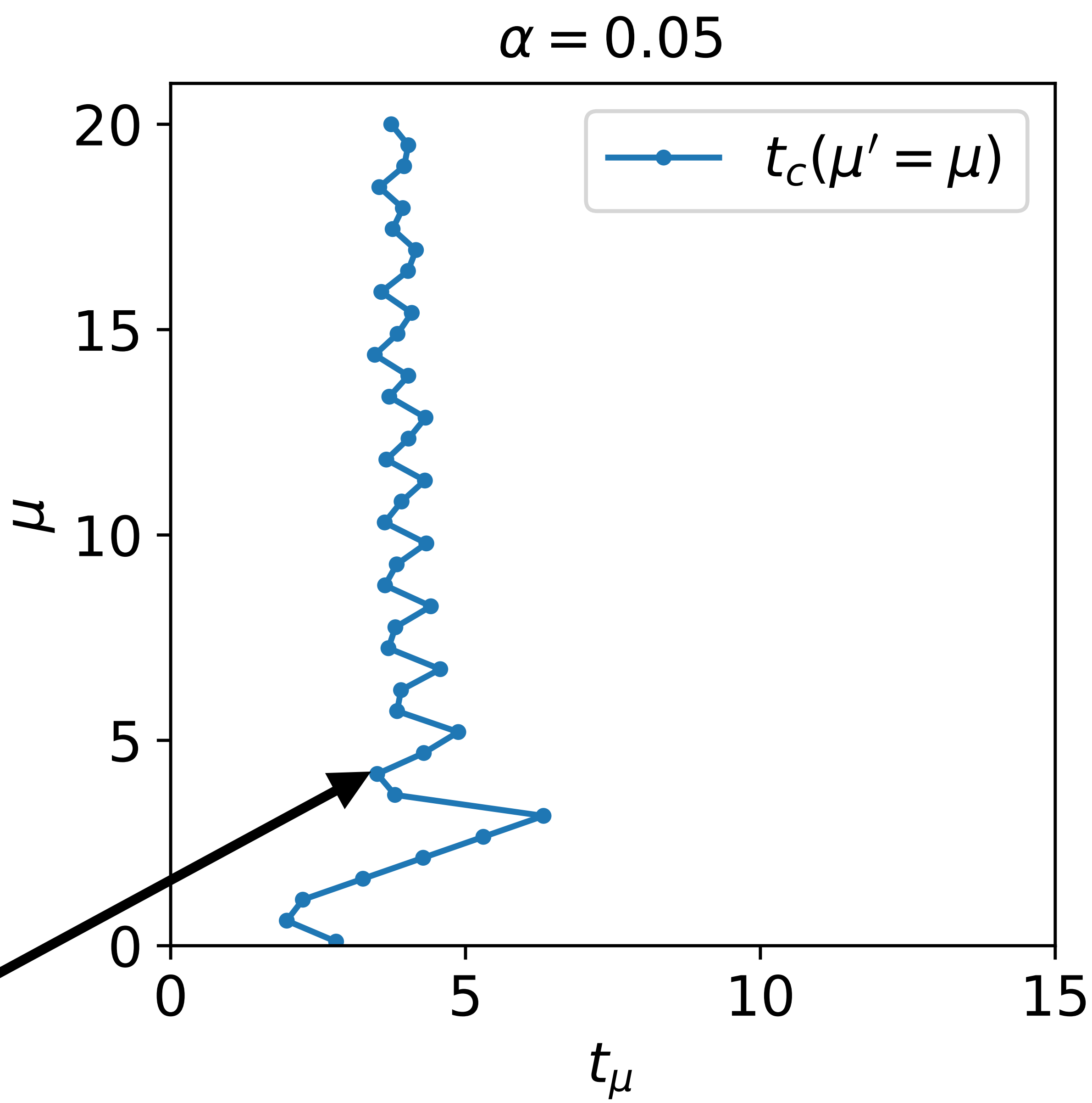
$$t_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Likelihood ratio examples

- Poisson example
  - $P(x \mid \mu) = \dfrac{\mu^x e^{-\mu}}{x!}$
- For each μ:
  - Throw 10k toys
  - Compute $\hat{\mu}$, t_μ
  - Find 0.95 quantile in distribution

Note: jagged behavior is due to discrete nature of t, not limited toy statistics



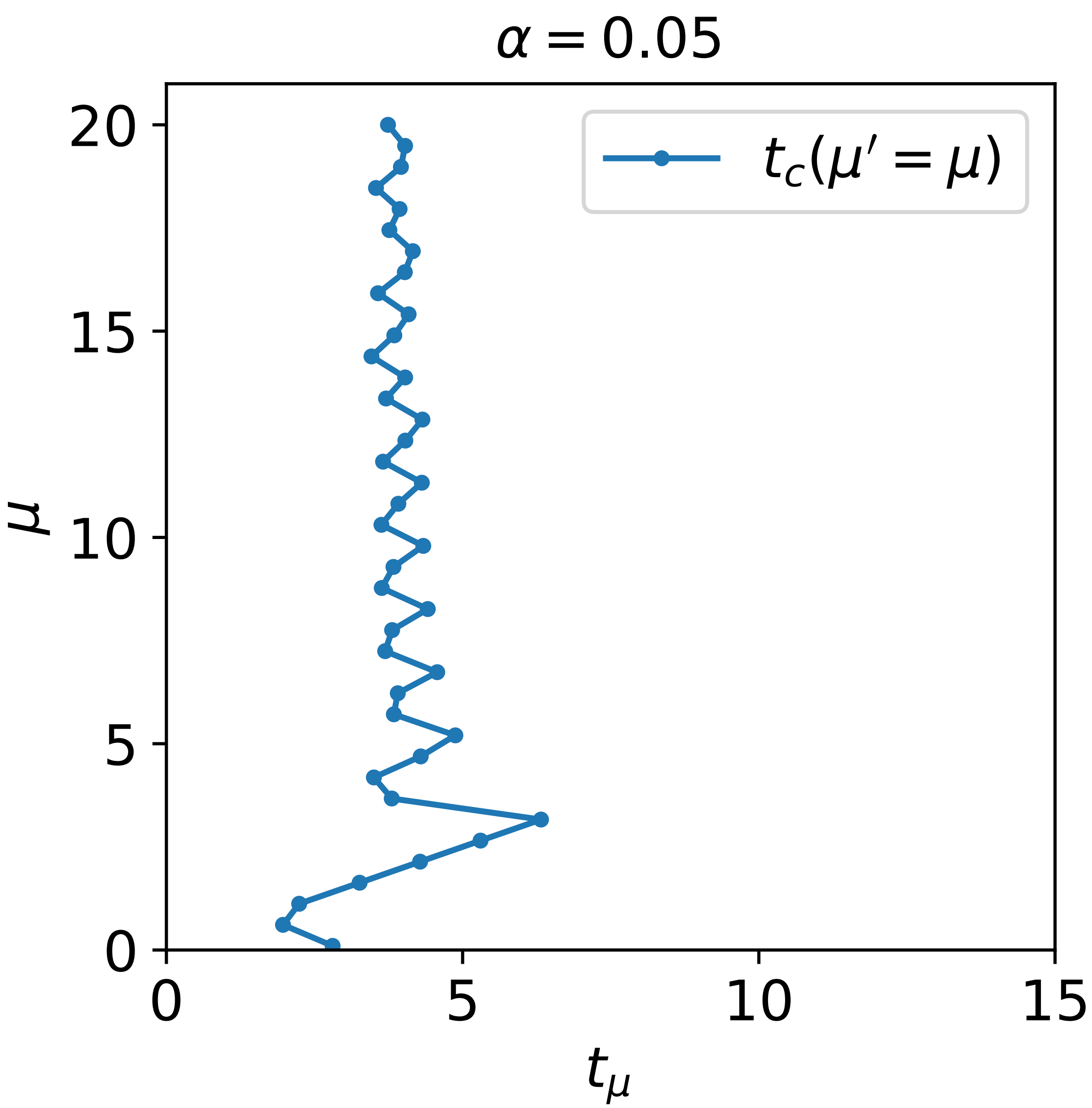


$$t_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Fermilab

Likelihood ratio examples

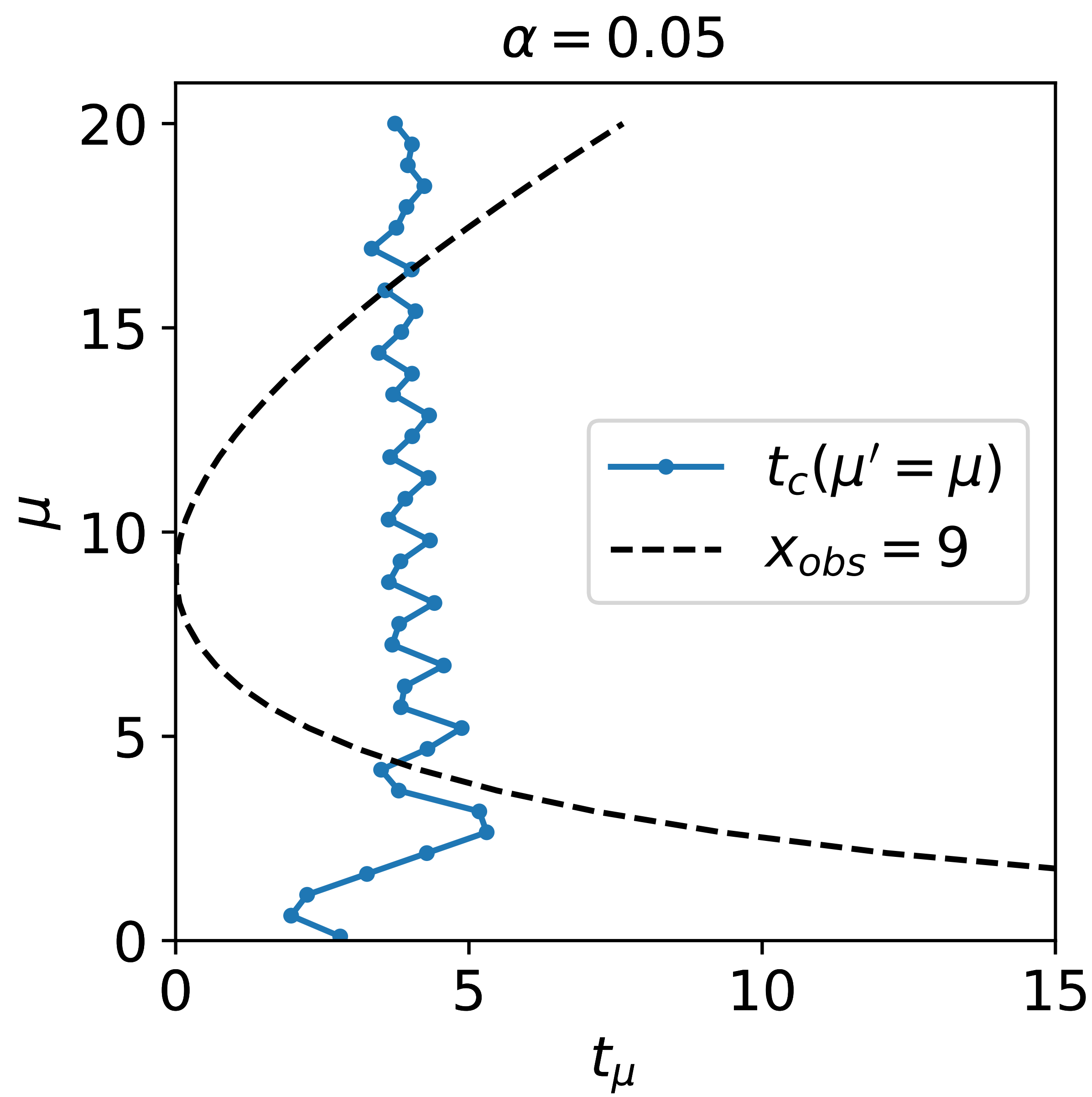
- Poisson example
  - $P(x\,|\,\mu) = \dfrac{\mu^x e^{-\mu}}{x!}$
- For each μ:
  - Throw 10k toys
  - Compute $\hat{\mu}$, t_μ
  - Find 0.95 quantile in distribution



$$t_\mu = -2\ln\frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Likelihood ratio examples

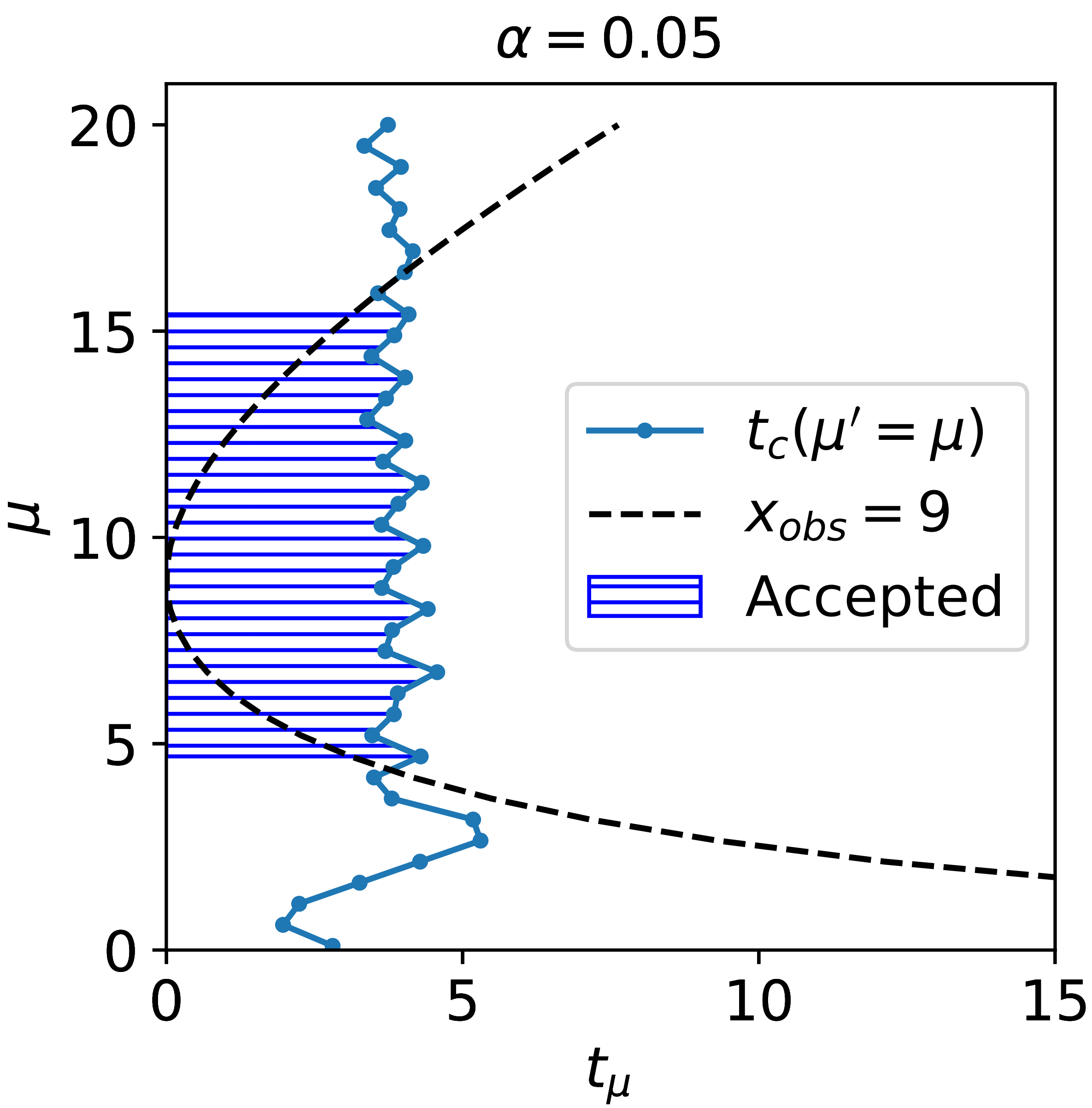
- Poisson example
  - $P(x\,|\,\mu) = \dfrac{\mu^x e^{-\mu}}{x!}$
- For each μ:
  - Throw 10k toys
  - Compute $\hat{\mu}$, t_μ
  - Find 0.95 quantile in distribution
- Draw $t_\mu(x_{obs})$ contour



$$t_\mu = -2\ln\frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Likelihood ratio examples

- Poisson example
  - $P(x\,|\,\mu) = \dfrac{\mu^x e^{-\mu}}{x!}$
- For each μ:
  - Throw 10k toys
  - Compute $\hat{\mu}$, t_μ
  - Find 0.95 quantile in distribution
- Draw $t_\mu(x_{obs})$ contour
- Accept $t_\mu(x_{obs}) < t_{\mu,c}$



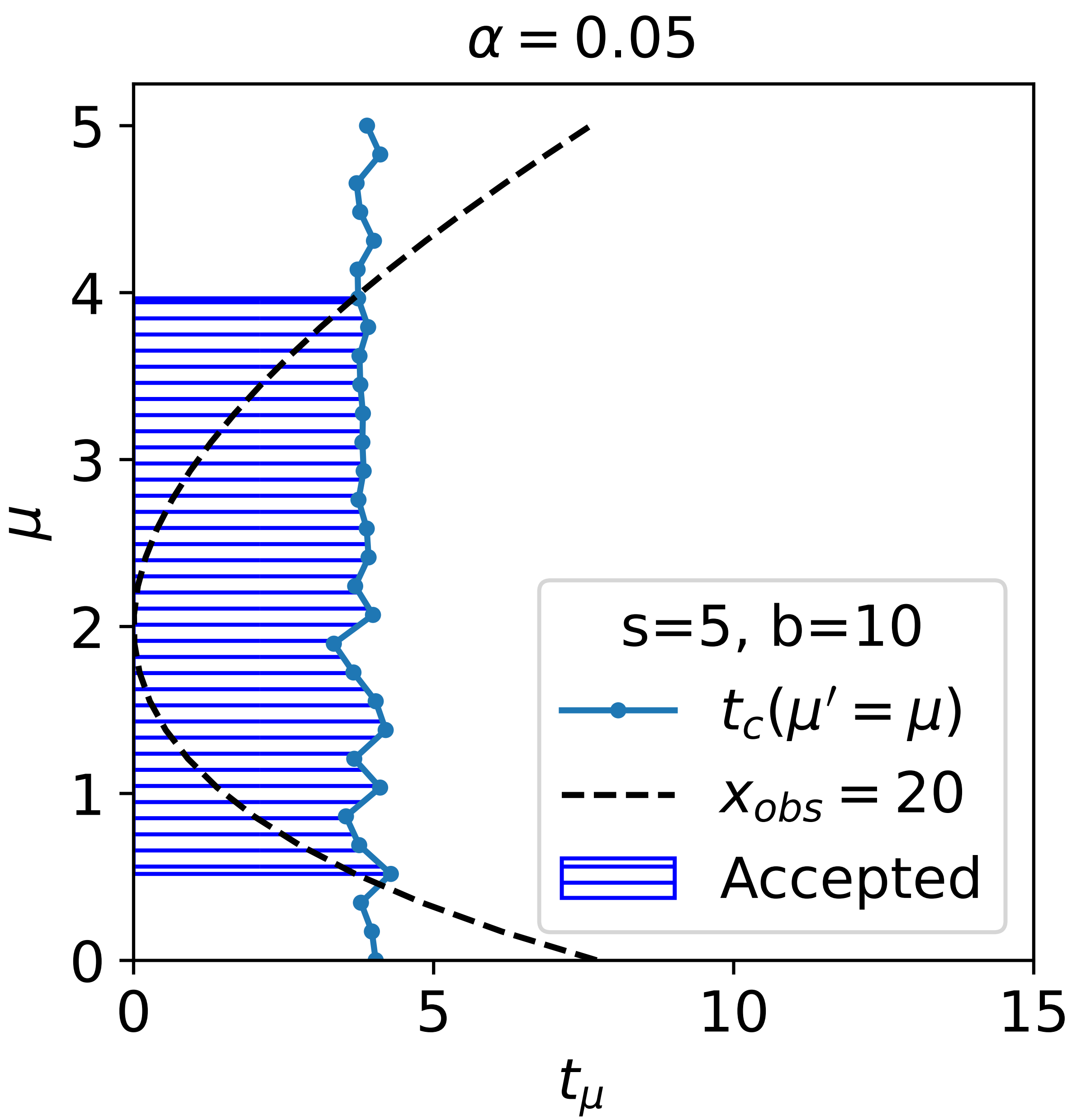
$$t_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Likelihood ratio examples

- Poisson with background example
  - $$P(x \mid \mu s + b) = \frac{(\mu s + b)^x e^{-(\mu s + b)}}{x!}$$
  - s=5, b=10 fixed, x=20

Likelihood ratio examples

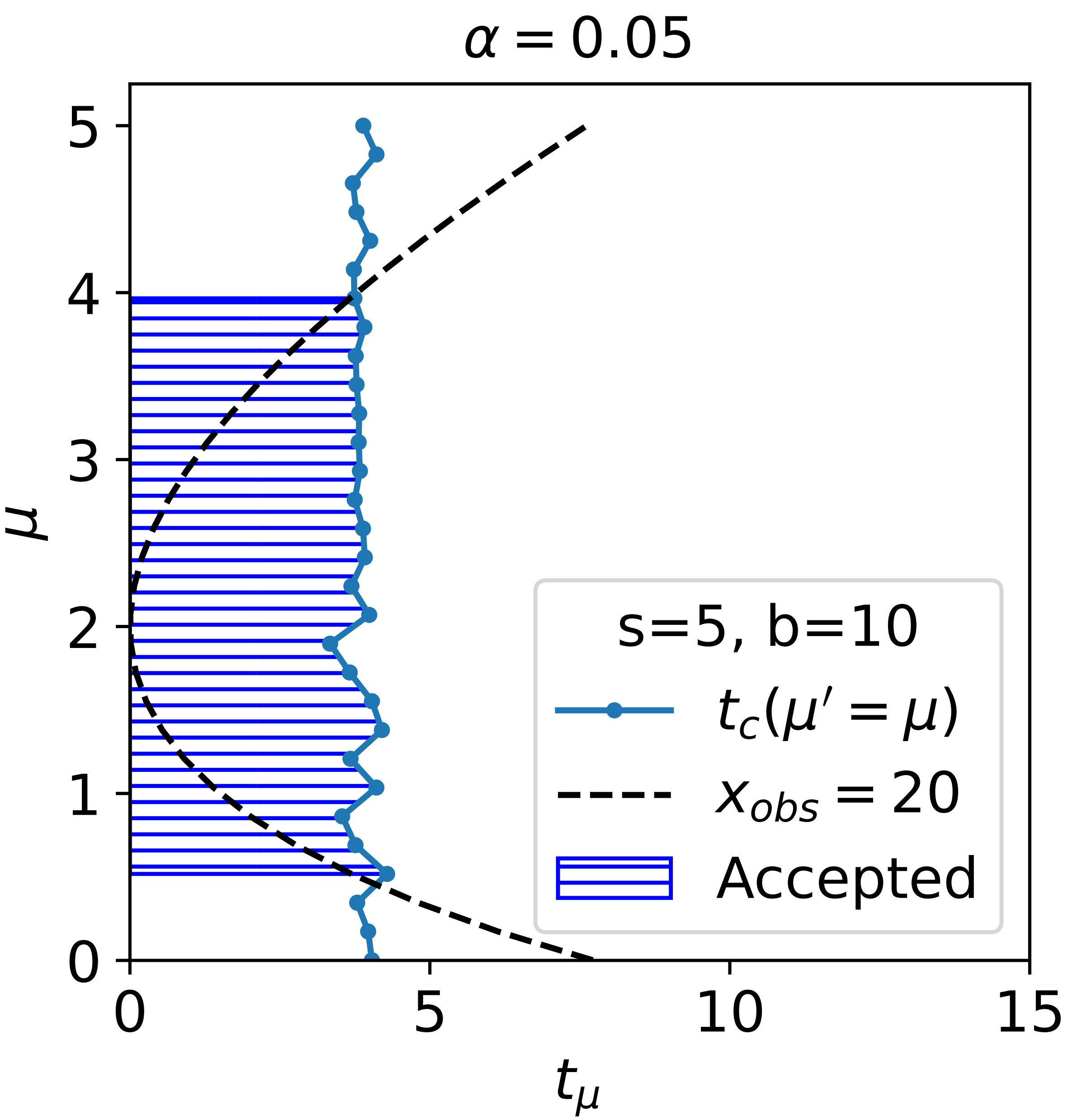
- Poisson with background example
  - $P(x\,|\,\mu s + b) = \dfrac{(\mu s + b)^x e^{-(\mu s + b)}}{x!}$
  - s=5, b=10 fixed, x=20
- Plan: set upper limit on μ



$$t_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\hat{\mu})}$$

Likelihood ratio examples

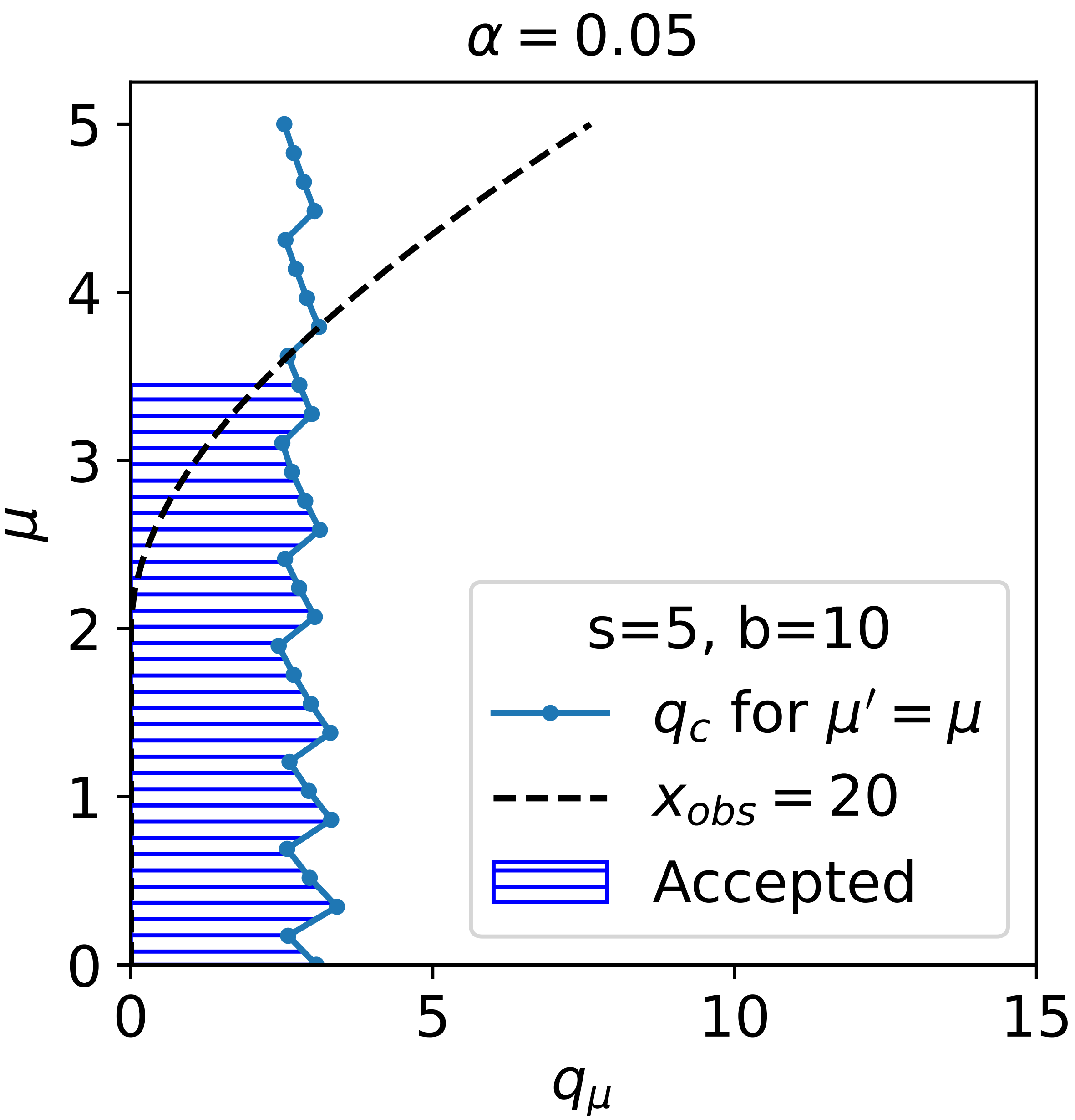
- Poisson with background example
  - $P(x \mid \mu s + b) = \dfrac{(\mu s + b)^x e^{-(\mu s + b)}}{x!}$
  - s=5, b=10 fixed, x=20
- Plan: set upper limit on μ
- Problem: two-sided region
  - We should not consider $\hat{\mu} > \mu$ to indicate less compatibility with a model that assumes a rate μ.



$$t_\mu = -2 \ln \frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})}$$

Likelihood ratio examples

- Poisson with background example
  - $P(x \mid \mu s + b) = \dfrac{(\mu s + b)^x e^{-(\mu s+b)}}{x!}$
  - s=5, b=10 fixed, x=20
- Plan: set upper limit on μ
- Problem: two-sided region
  - We should not consider $\hat{\mu} > \mu$ to indicate less compatibility with a model that assumes a rate μ.
- Solution: modify test statistic
  - Define $q_\mu = -2 \ln \dfrac{\mathscr{L}(\mu)}{\mathscr{L}(\min(\mu, \hat{\mu}))}$
  - i.e. over-fluctuations are "not extreme"

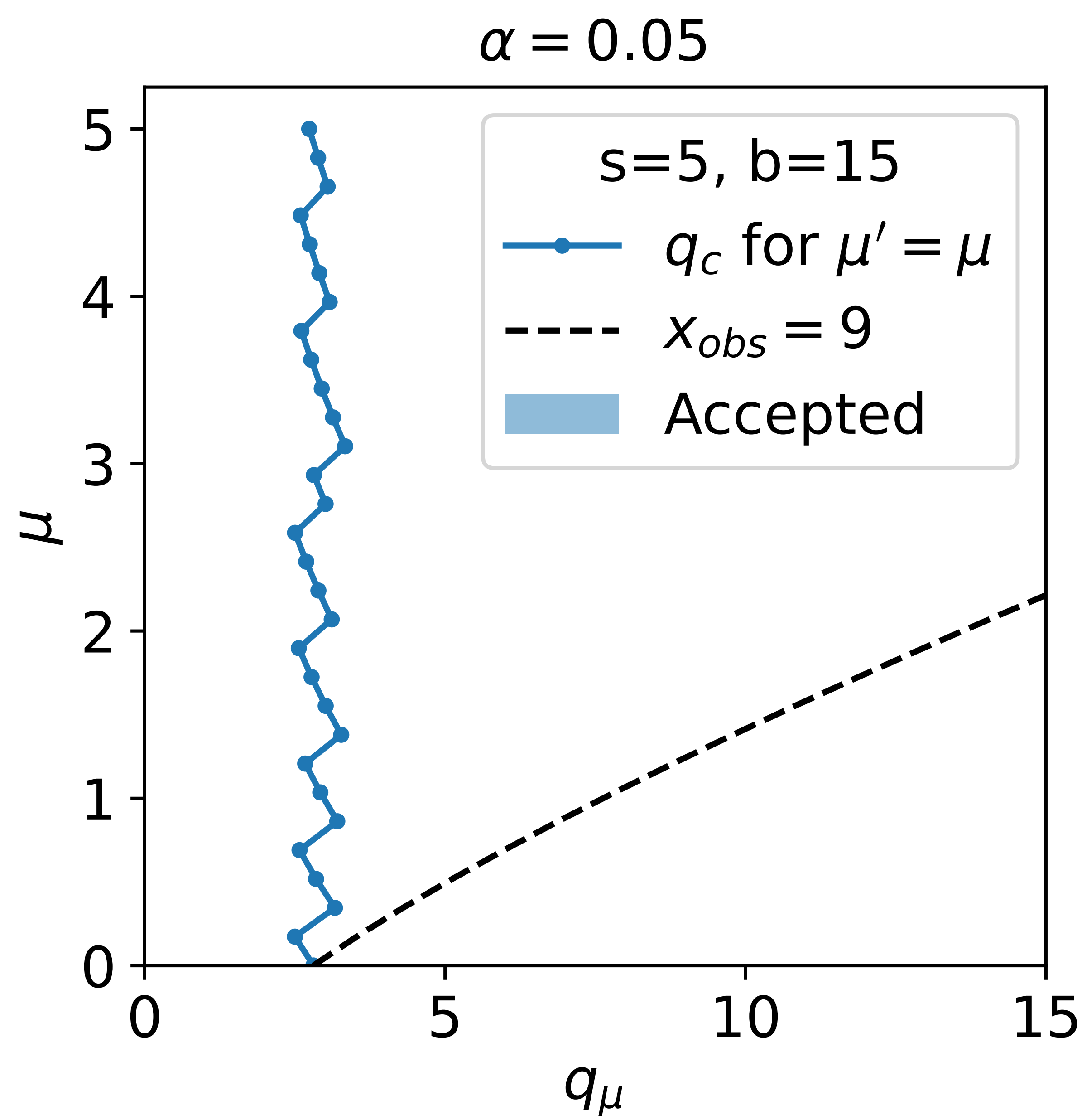


Fermilab

Likelihood ratio examples

- Same example as before, but b=15, x=9

- Problem: under fluctuation
  - No values accepted!
  - Possible but unsatisfying outcome of frequentist test



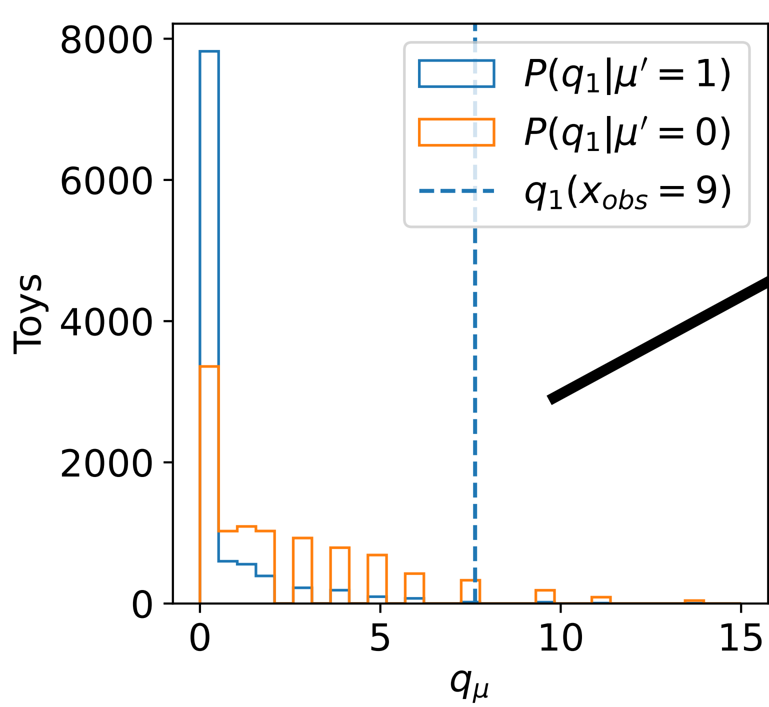
$$q_\mu = -2\ln\frac{\mathscr{L}(\mu)}{\mathscr{L}(\min(\mu,\hat{\mu}))}$$

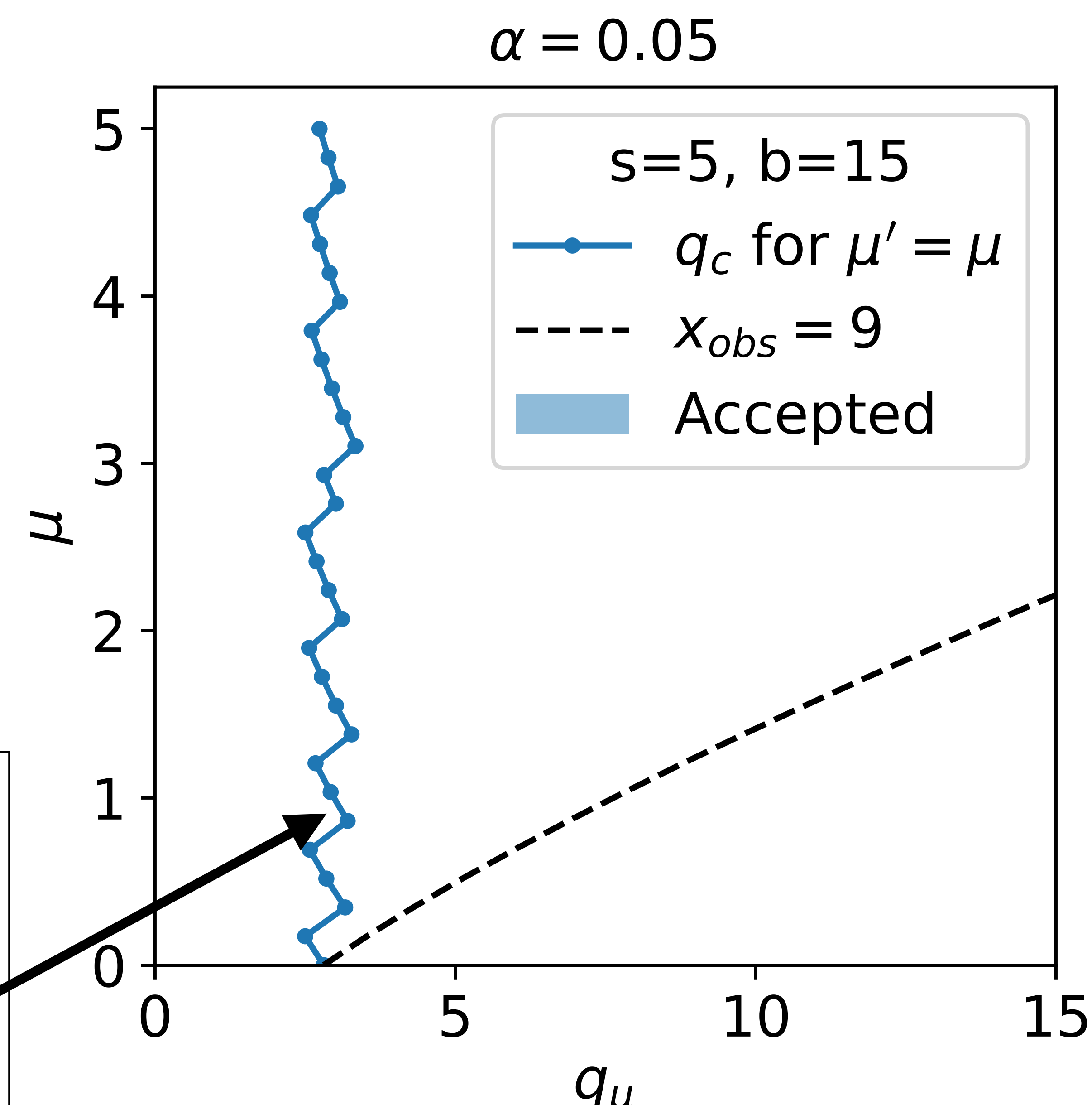
Likelihood ratio examples

- Same example as before, but b=15, x=9

- Problem: under fluctuation
  - No values accepted!
  - Possible but unsatisfying outcome of frequentist test

The result x=9 is rare for both S+B and B-only hypotheses.



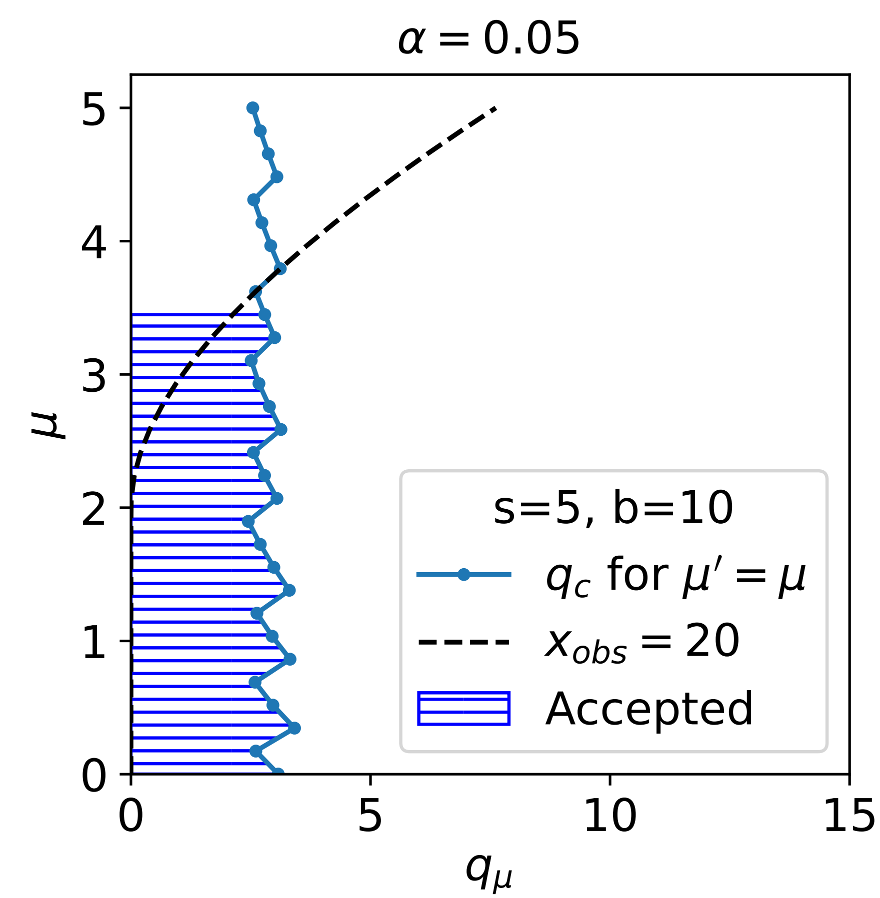


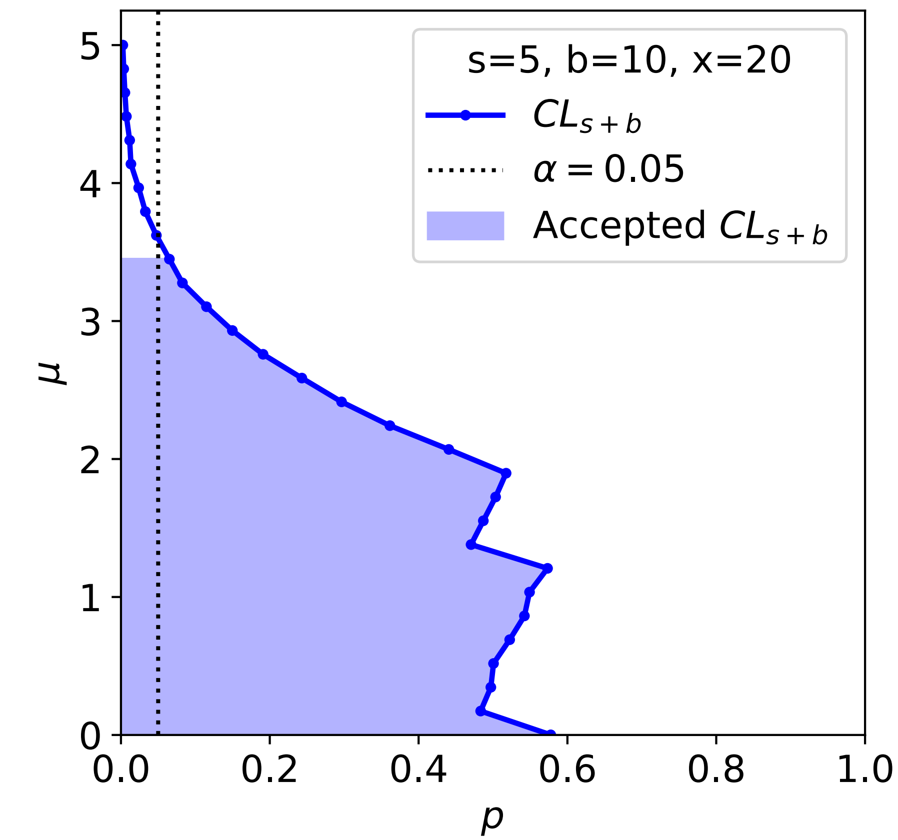
$$q_\mu = -2 \ln \frac{\mathscr{L}(\mu)}{\mathscr{L}(\min(\mu, \hat{\mu}))}$$

CLs

- CL_s criterion departs from purely frequentist CL to ameliorate the null set problem (among others)
  - Original expositions by A. Read, T. Junk
  - See also PDG 40.4.2.4

- First we reformulate our old test:
  - Define $CL_{s+b} = p_\mu = \int_{t_\mu(x_{obs})}^{\infty} P(t_\mu \,|\, \mu' = \mu)\, dt_\mu$
    - This is a p-value
  - Then we accept the region $CL_{s+b} > \alpha$
    - Right: initial S+B example reformulated



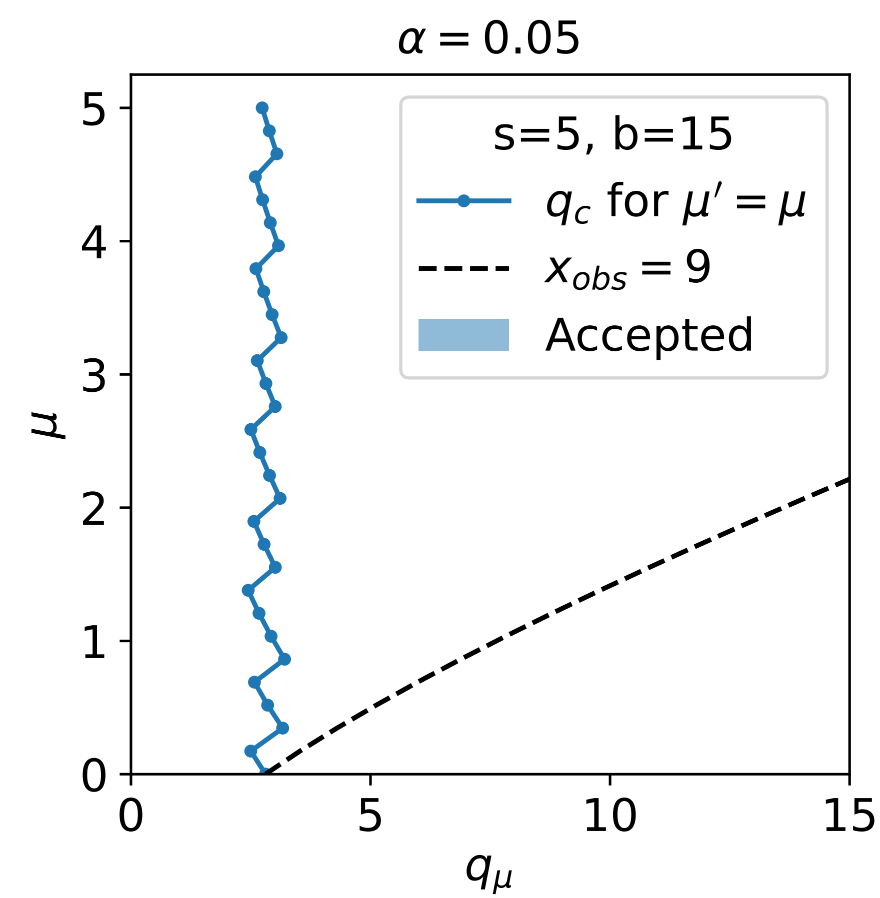


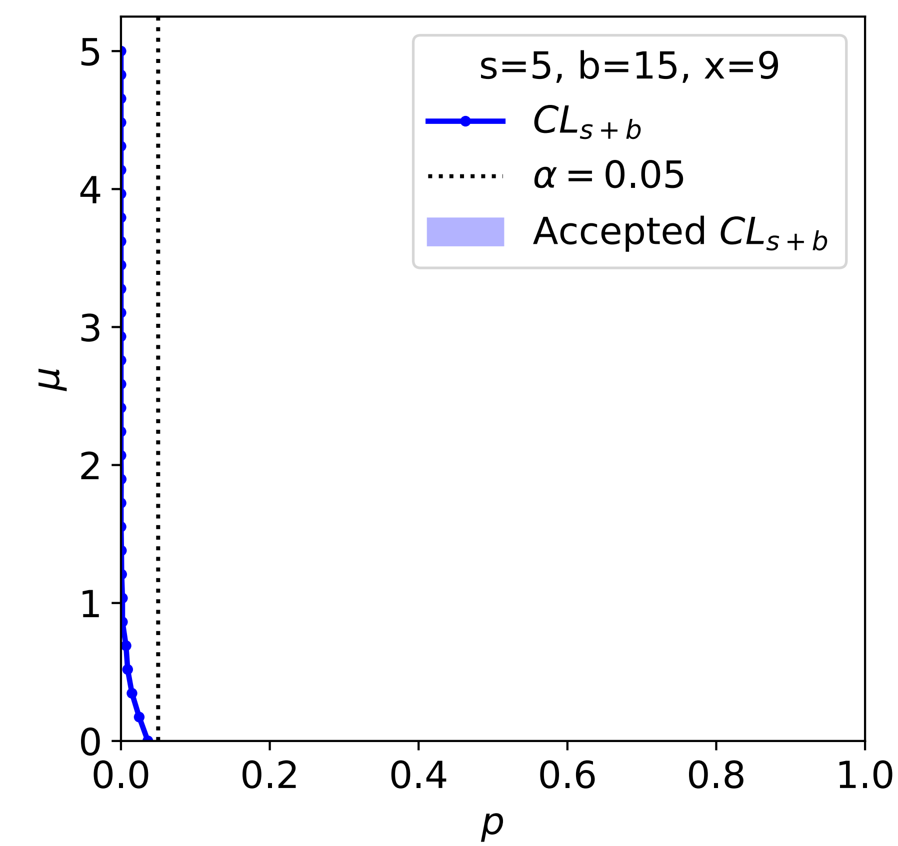
Fermilab

CLs

- CL_s criterion departs from purely frequentist CL to ameliorate the null set problem (among others)
  - Original expositions by A. Read, T. Junk
  - See also PDG 40.4.2.4

- First we reformulate our old test:
  - Define $CL_{s+b} = p_\mu = \int_{t_\mu(x_{obs})}^{\infty} P(t_\mu \,|\, \mu' = \mu)\, dt_\mu$
    - This is a p-value
  - Then we accept the region $CL_{s+b} > \alpha$
    - Right: under-fluctuation S+B example reformulated





Fermilab

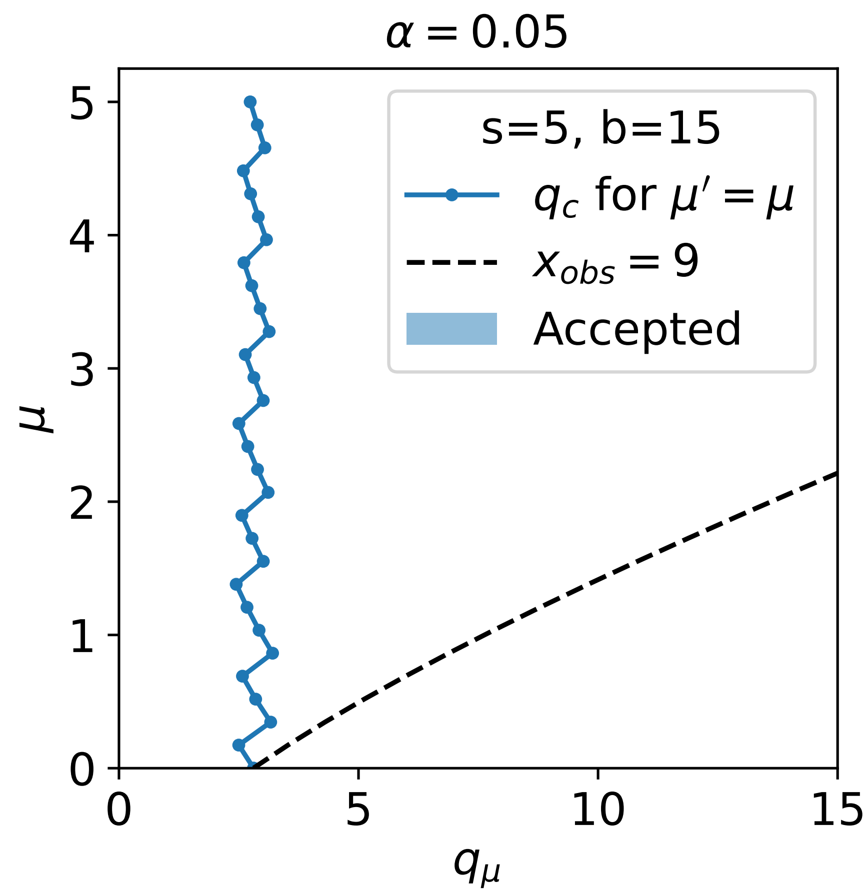
CLs

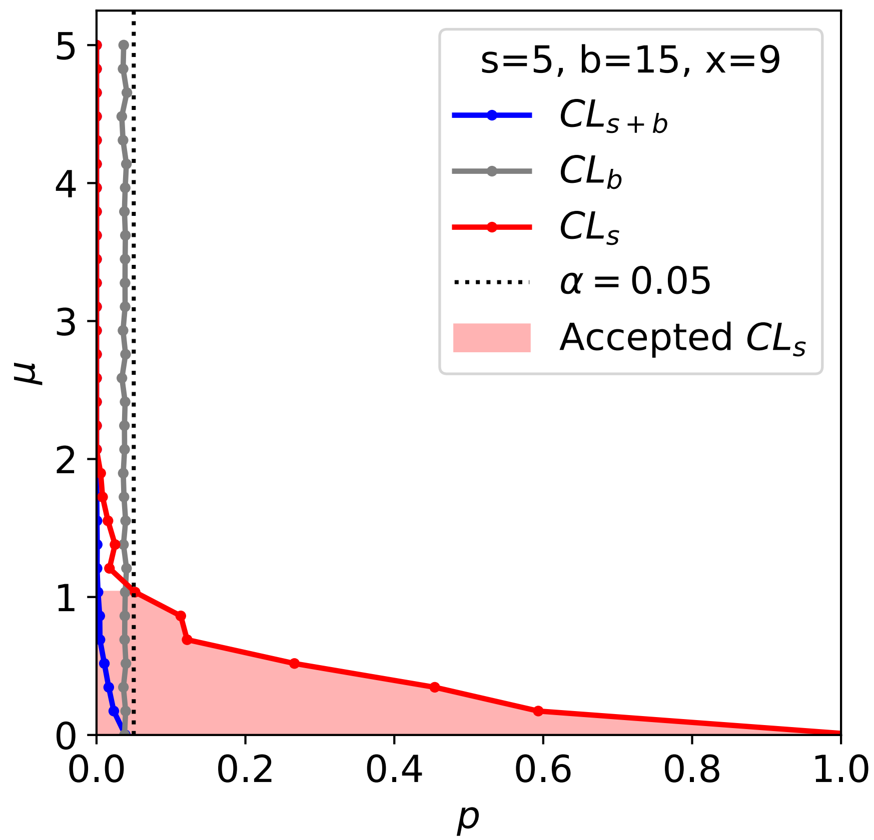
- CLs criterion departs from purely frequentist CL to ameliorate the null set problem (among others)
  - Original expositions by A. Read, T. Junk
  - See also PDG 40.4.2.4

- First we reformulate our old test:
  - Define $CL_{s+b} = p_\mu = \int_{t_\mu(x_{obs})}^{\infty} P(t_\mu \,|\, \mu' = \mu)\, dt_\mu$

- Now define background-only p-value
  - $CL_b = 1 - p_b = \int_{t_\mu(x_{obs})}^{\infty} P(t_\mu \,|\, \mu' = 0)\, dt_\mu$

- Accept instead $CL_s = CL_{s+b}/CL_b > \alpha$





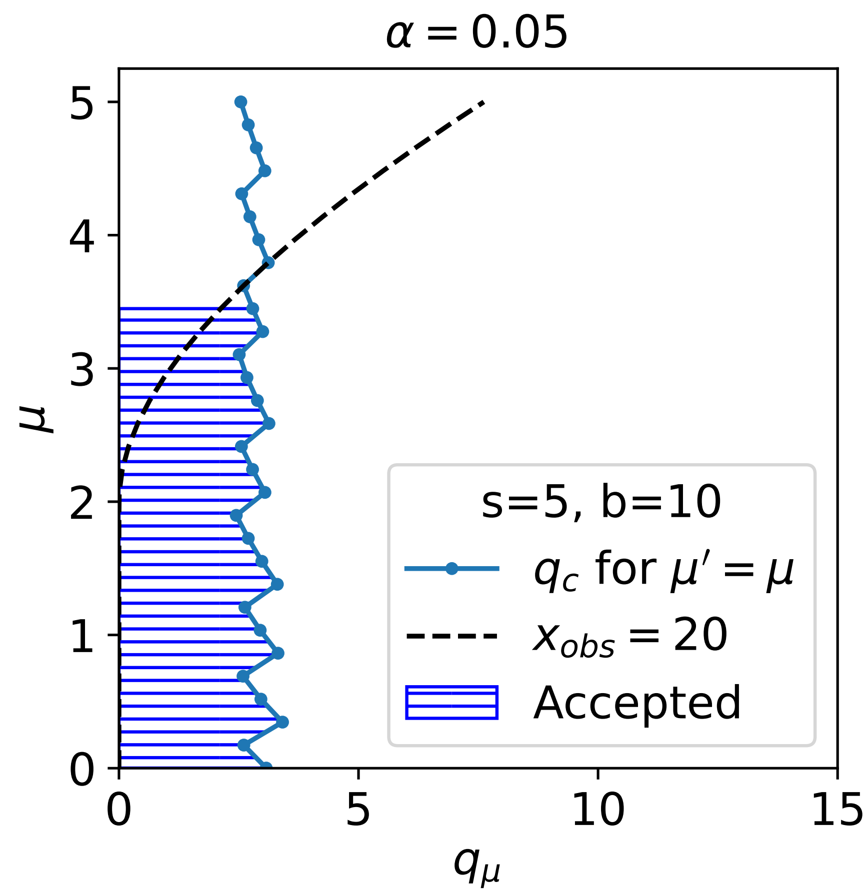
CLs

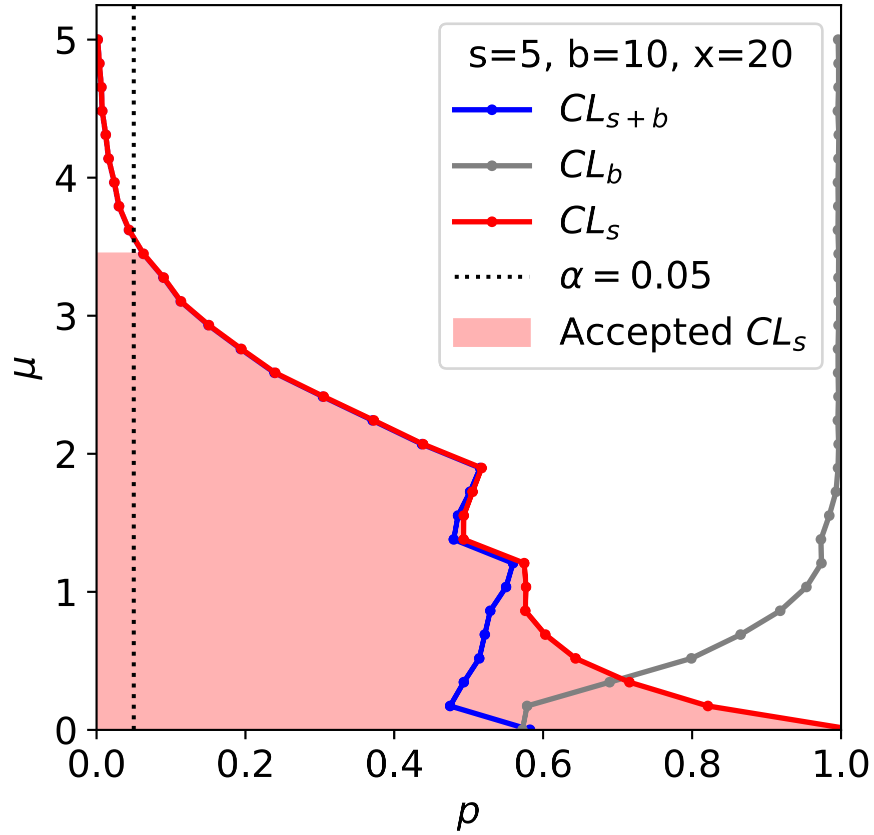
- CLs criterion departs from purely frequentist CL to ameliorate the null set problem (among others)
  - Original expositions by A. Read, T. Junk
  - See also PDG 40.4.2.4

- First we reformulate our old test:
  - Define $CL_{s+b} = p_\mu = \int_{t_\mu(x_{obs})}^{\infty} P(t_\mu \,|\, \mu' = \mu)\, dt_\mu$

- Now define background-only p-value
  - $CL_b = 1 - p_b = \int_{t_\mu(x_{obs})}^{\infty} P(t_\mu \,|\, \mu' = 0)\, dt_\mu$

- Accept instead $CL_s = CL_{s+b}/CL_b > \alpha$
  - No effect in the first example

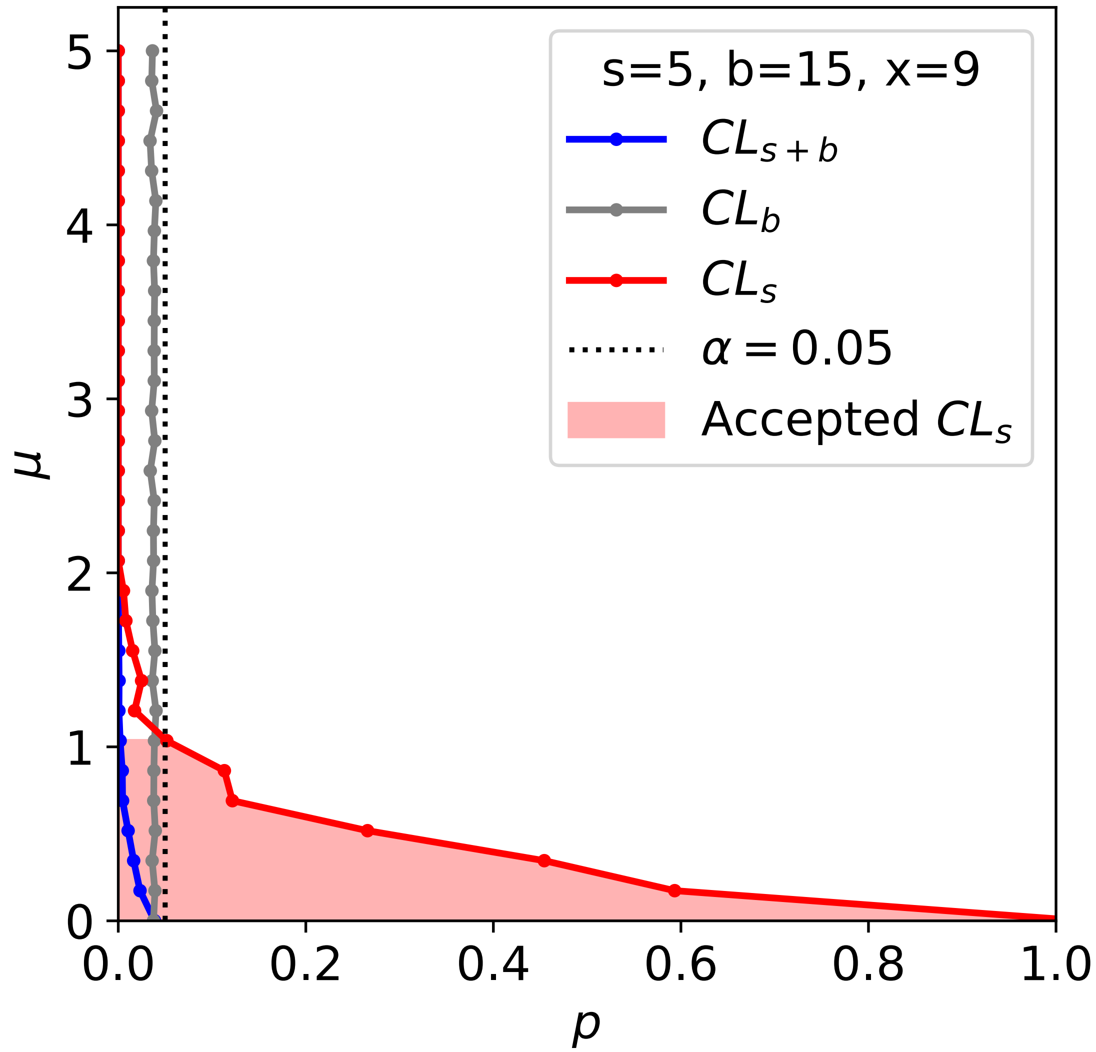




Fermilab

Finally, a combine command

```
$ cat card.txt
imax 1
jmax 1
kmax 0
---------------
bin           bin1
observation      9
---------------
bin               bin1  bin1
process            sig   bkg
process              0     1
rate                 5    15
$ combine -M HybridNew --LHCmode LHC-limits card.txt
…
Limit: r < 1.10036 +/- 0.0189137 @ 95% CL
```

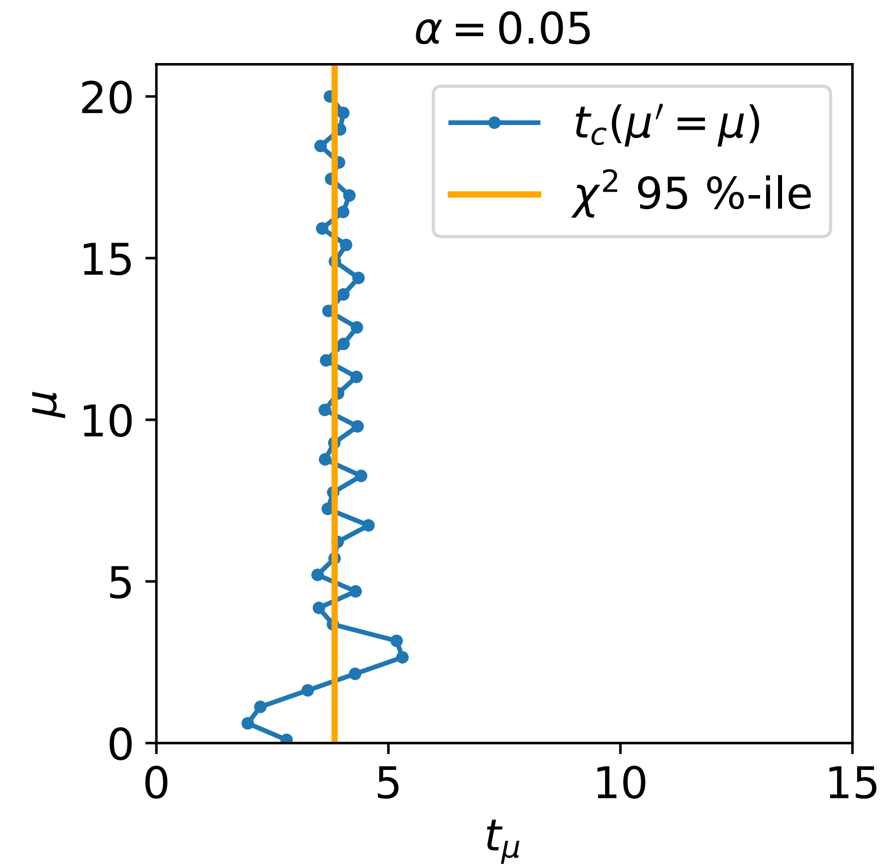


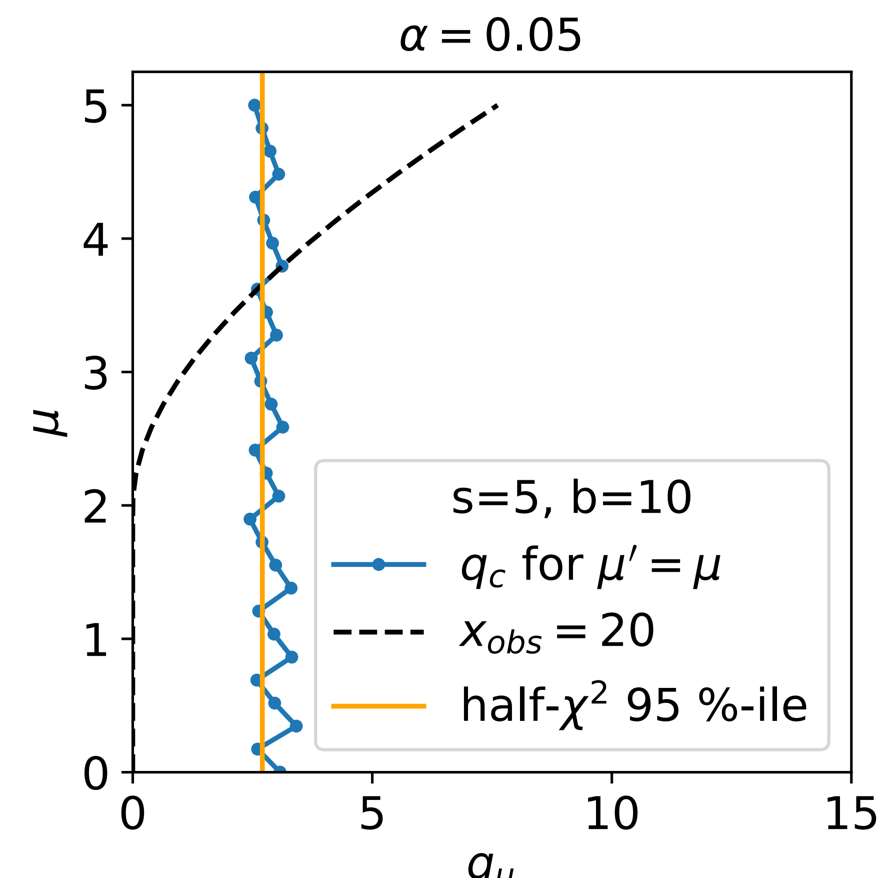
Asymptotic behavior

- Notice how $t_{\mu,c}$ and $q_{\mu,c}$ tend towards a constant?

- This is Wilk's theorem in action
  - Statement: as sample size grows, the distribution of the likelihood ratio P(t_θ|θ') approaches a χ^2 distribution
    - With df = dim(θ)
  - Hence we can approximate by just evaluating $t_\theta(x_{obs})$!

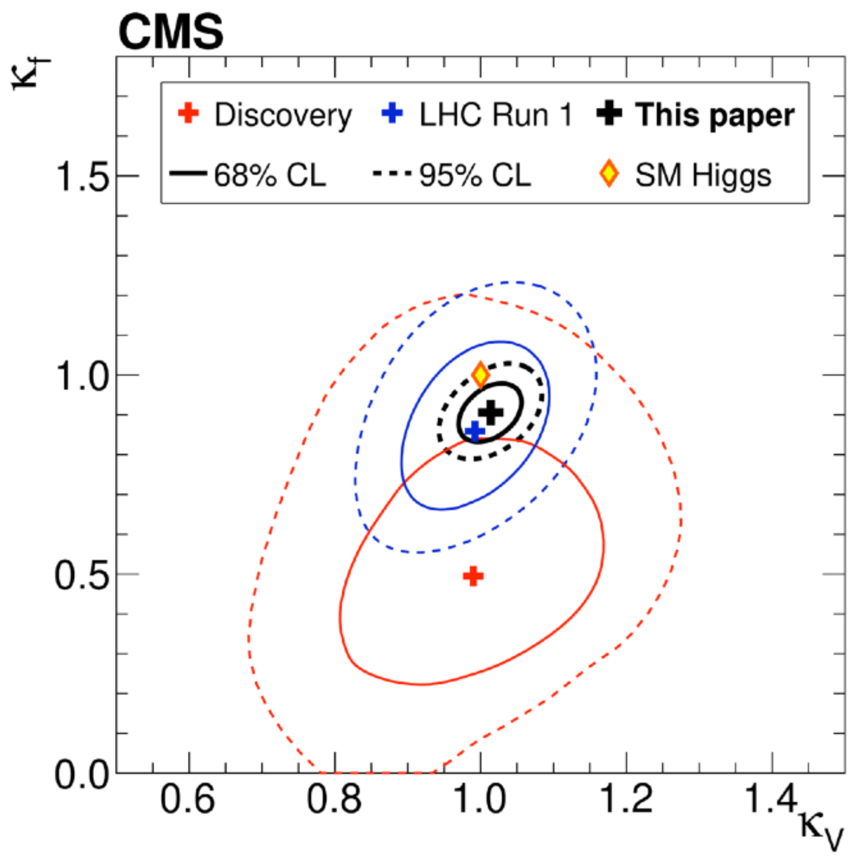
- For q, formulas slightly more complex
  - CCGV provide the recipe: non-central half-χ^2
  - The non-centrality is found using the *Asimov* dataset
    - A special x_μ for a given μ such that $\hat{\mu}(x) = \mu$
    - Note for Poisson data, it may be non-integral!
    - This dataset produces the median expected limit

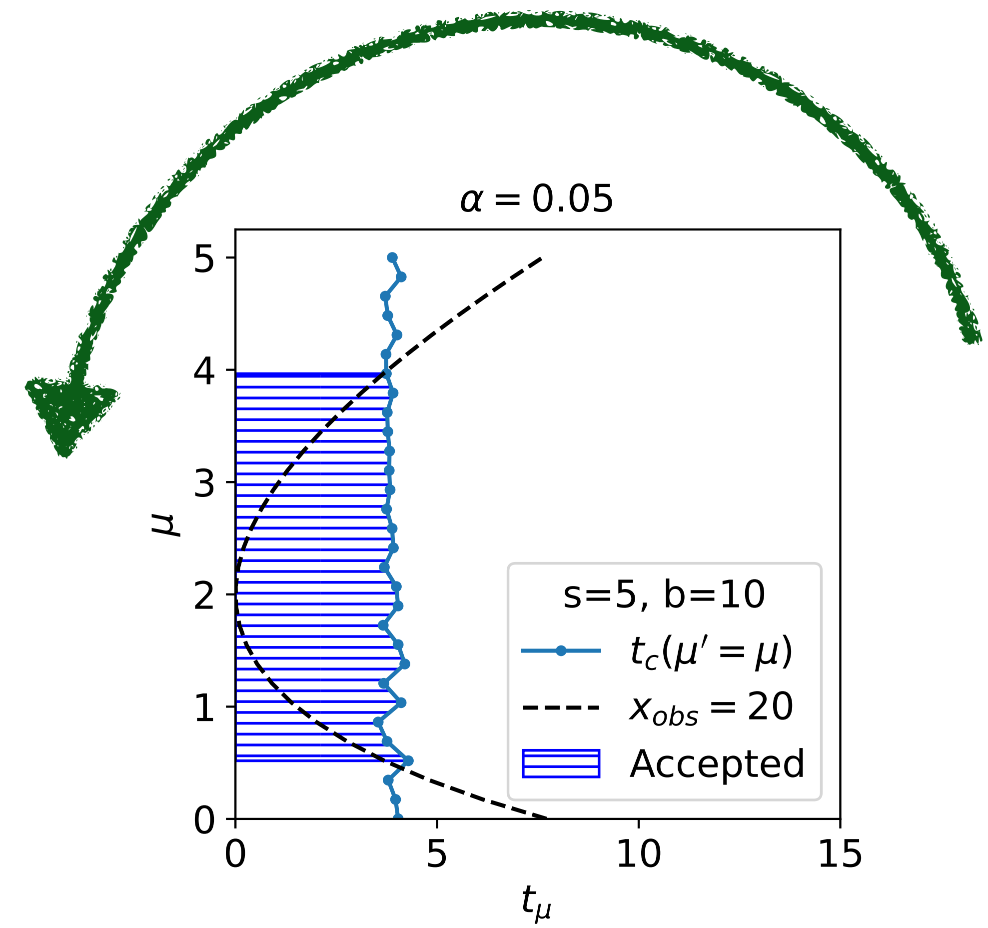


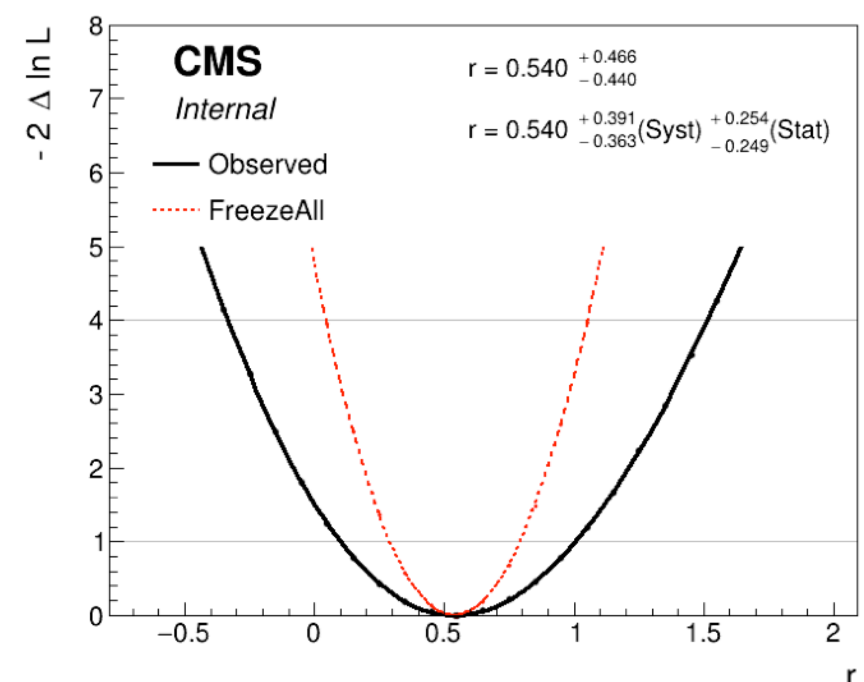


Asymptotic behavior

- This is how we make deltaNLL contours







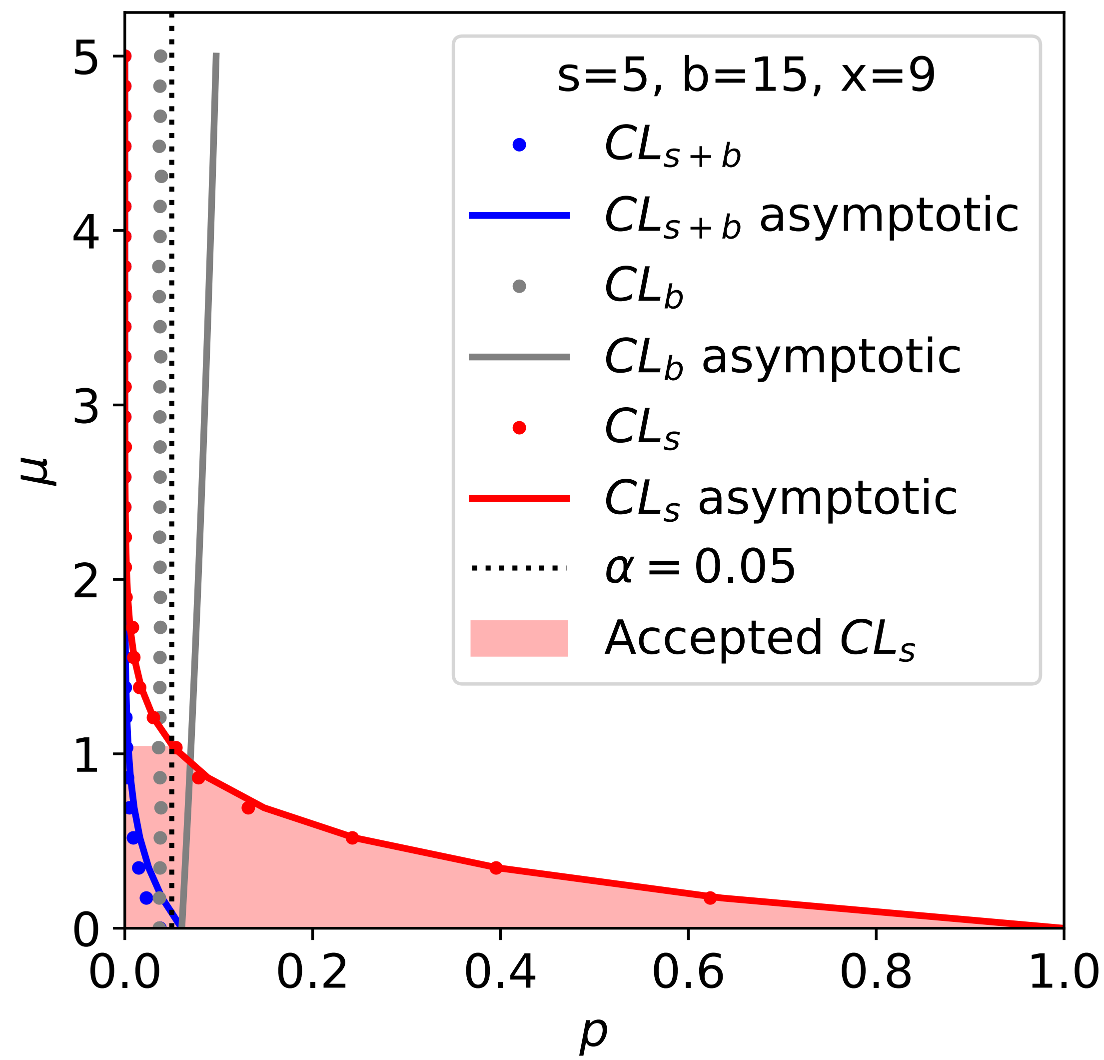

```
scipy.stats.chi2.ppf(q, df)
```

Quantile	t_c	
	df=1	df=2
0.68	0.989	2.279
1σ (0.6827…)	1	2.296
0.95	3.841	5.991
2σ (0.9545…)	4	6.180

Fermilab

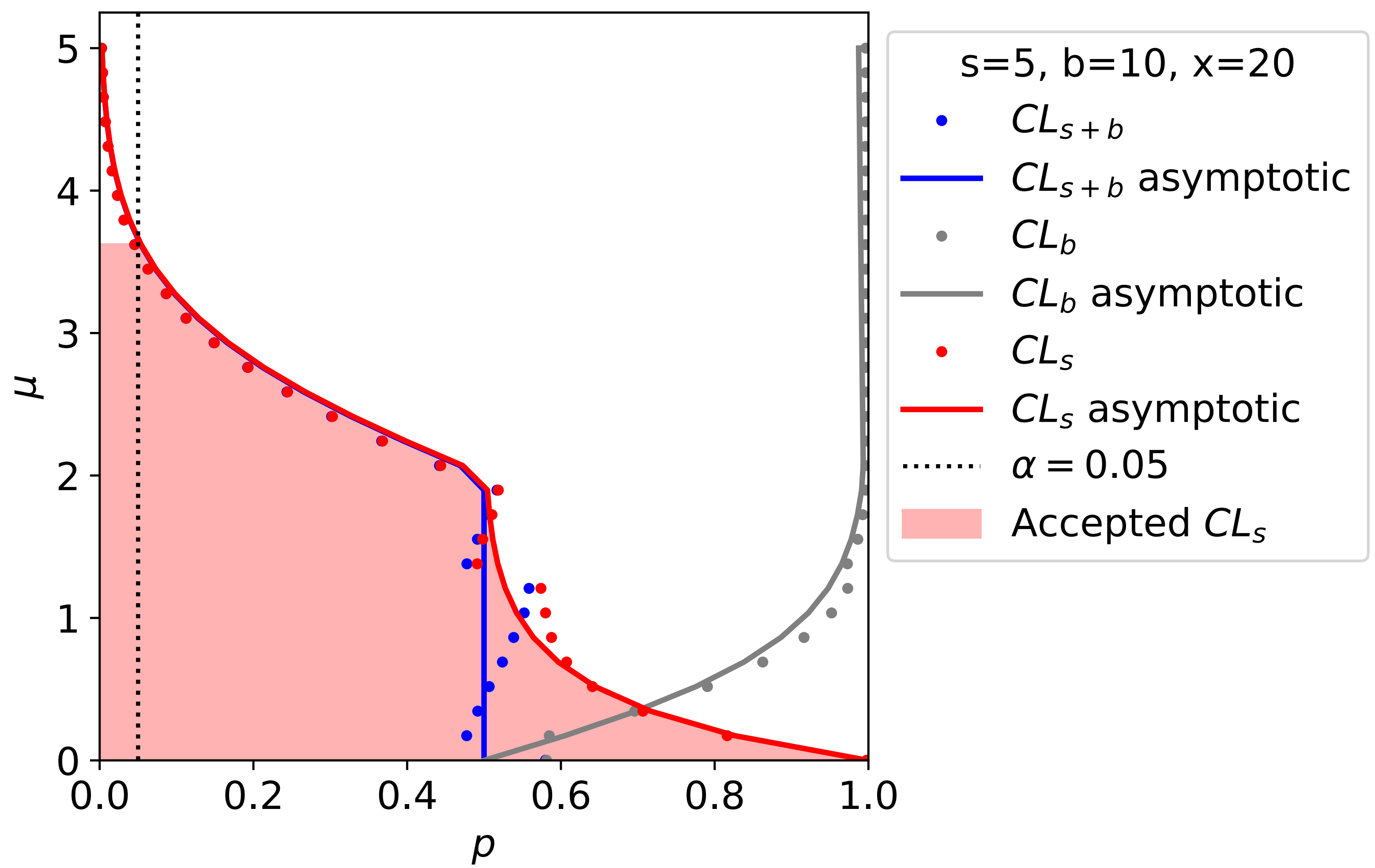
AsymptoticLimits in combine

```
$ cat card.txt
imax 1
jmax 1
kmax 0
---------------
bin           bin1
observation     9
---------------
bin                bin1  bin1
process             sig   bkg
process               0     1
rate                  5    15
$ combine -M AsymptoticLimits card.txt
…
 -- AsymptoticLimits ( CLs ) --
Observed Limit: r < 1.0502
```



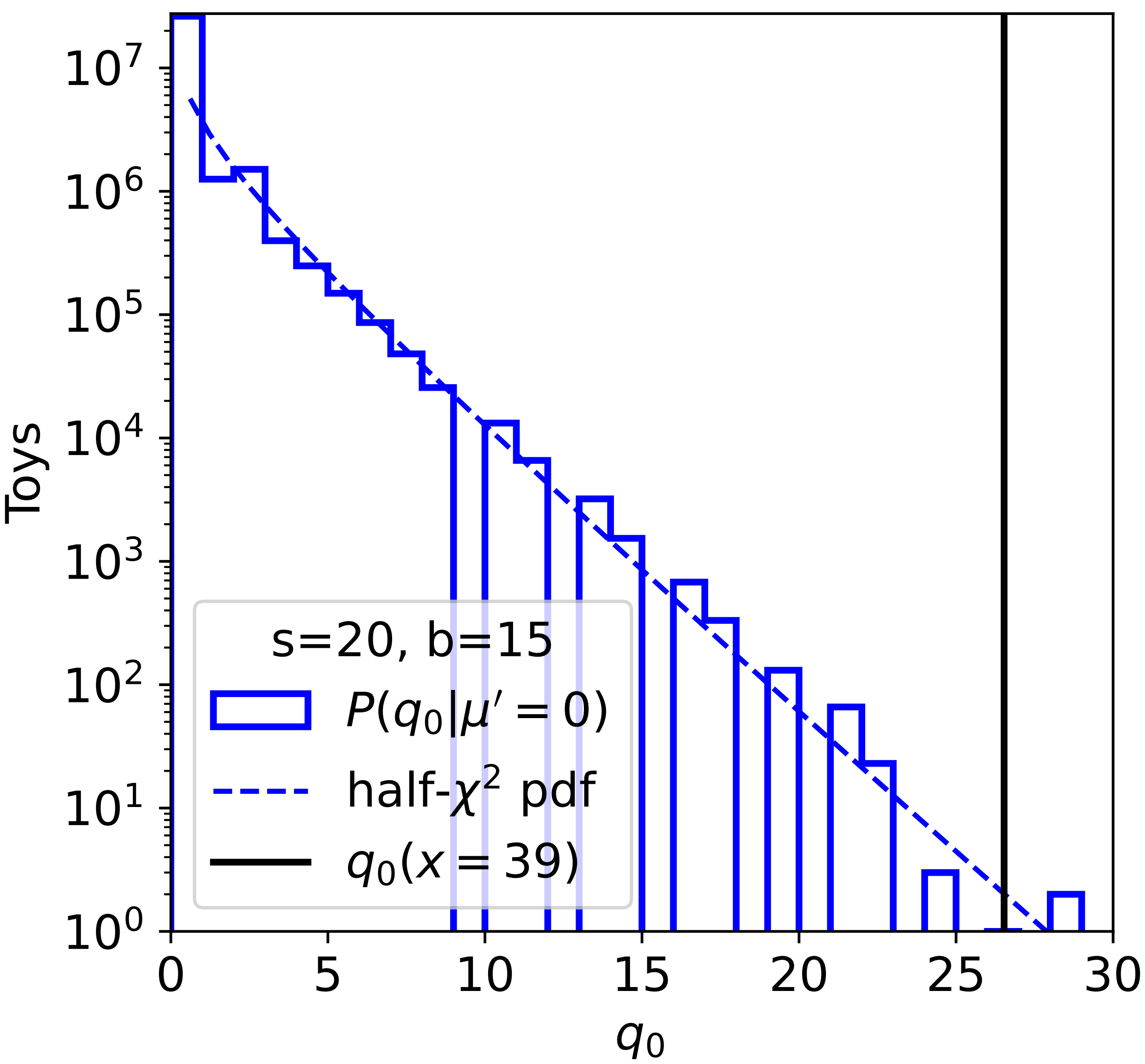
AsymptoticLimits in combine

```
$ cat card.txt
imax 1
jmax 1
kmax 0
---------------
bin             bin1
observation      20
---------------
bin                 bin1   bin1
process              sig    bkg
process                0      1
rate                   5     10
$ combine -M AsymptoticLimits card.txt
…
 -- AsymptoticLimits ( CLs ) --
Observed Limit: r < 3.6595
```



Test statistic for discovery

- Poisson with background example
  - $P(x \,|\, \mu s + b) = \dfrac{(\mu s + b)^x e^{-(\mu s + b)}}{x!}$
  - s=20, b=15 fixed, x=39
- Cannot use t_μ:
  - Severe under-fluctuation would count as discovery! Certainly something was discovered, but not an excess over background. Disallow in test statistic:
    - Define $q_0 = -2 \ln \dfrac{\mathscr{L}(0)}{\mathscr{L}(\max(0,\hat{\mu}))}$
    - i.e. under-fluctuations are “not extreme”
  - Deceptively simple result: $Z = \sqrt{q_0(x_{obs})}$
    - Only true if one POI



Fermilab

Adding uncertainties

Also changing notation (sorry): θ is changing

An uncertain background

- Adding some background uncertainty to our model
  - Let $\mathscr{P}(n\,|\,\lambda)$ stand for Poisson pdf, and $\mathcal{N}(x\,|\,\mu,\sigma)$ for Normal pdf

An uncertain background

- Adding some background uncertainty to our model
  - Let $\mathscr{P}(n\,|\,\lambda)$ stand for Poisson pdf, and $\mathscr{N}(x\,|\,\mu,\sigma)$ for Normal pdf

- Option 1:
  - $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b(1+\delta\theta_b))\,\mathscr{N}(\theta_0\,|\,\theta_b,1)$
  - (for some δ close to 0) Not great: b should not go negative. In combine: `trG`

An uncertain background

- Adding some background uncertainty to our model
  - Let $\mathscr{P}(n\,|\,\lambda)$ stand for Poisson pdf, and $\mathscr{N}(x\,|\,\mu,\sigma)$ for Normal pdf

- Option 1:
  - $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b(1+\delta\theta_b))\,\mathscr{N}(\theta_0\,|\,\theta_b,1)$
  - (for some δ close to 0) Not great: b should not go negative. In combine: `trG`
- Option 2:
  - $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b\kappa^{\theta_b})\,\mathscr{N}(\theta_0\,|\,\theta_b,1)$
  - (for some κ close to 1) Better: log-normal. In combine: `lnN`

An uncertain background

- Adding some background uncertainty to our model
  - Let $\mathscr{P}(n\,|\,\lambda)$ stand for Poisson pdf, and $\mathscr{N}(x\,|\,\mu,\sigma)$ for Normal pdf

- Option 1:
  - $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b(1+\delta\theta_b))\,\mathscr{N}(\theta_0\,|\,\theta_b, 1)$
  - (for some δ close to 0) Not great: b should not go negative. In combine: `trG`
- Option 2:
  - $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b\kappa^{\theta_b})\,\mathscr{N}(\theta_0\,|\,\theta_b, 1)$
  - (for some κ close to 1) Better: log-normal. In combine: `lnN`
- Option 3:
  - $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b\theta_b)\,\mathscr{P}(n_{CR}\,|\,b_{CR}\theta_b)$
  - If such a background-pure control region can be constructed
  - In combine: `gmN` or `rateParam`

An uncertain background

- Adding some background uncertainty to our model
  - Let $\mathcal{P}(n\,|\,\lambda)$ stand for Poisson pdf, and $\mathcal{N}(x\,|\,\mu,\sigma)$ for Normal pdf

- Option 1:
  - $P(n\,|\,\mu,\theta_b) = \mathcal{P}(n\,|\,\mu s + b(1+\delta\theta_b))\,\mathcal{N}(\theta_0\,|\,\theta_b,1)$
  - (for some δ close to 0) Not great: b should not go negative. In combine: `trG`
- Option 2:
  - $P(n\,|\,\mu,\theta_b) = \mathcal{P}(n\,|\,\mu s + b\kappa^{\theta_b})\,\mathcal{N}(\theta_0\,|\,\theta_b,1)$
  - (for some κ close to 1) Better: log-normal. In combine: `lnN`
- Option 3:
  - $P(n\,|\,\mu,\theta_b) = \mathcal{P}(n\,|\,\mu s + b\theta_b)\,\mathcal{P}(n_{CR}\,|\,b_{CR}\theta_b)$
  - If such a background-pure control region can be constructed
  - In combine: `gmN` or `rateParam`
- …and many more
  - In all cases we now have a new observable (θ_0, n_{CR}), a new parameter θ_b, and several new constants (δ, κ, b_{CR}) to compute (e.g. from simulation)

Fermilab

Full statistical model

- Split likelihood parameters into *parameters of interest* (POIs) μ and nuisance parameters θ, and define *auxiliary measurements* y that target the latter
  - Then the pdf factorizes $P(x, y \,|\, \mu, \theta) = P(x \,|\, \mu, \theta) P(y, \theta)$
  - N.B. *y* are *global observables* in root

Full statistical model

- Split likelihood parameters into *parameters of interest* (POIs) μ and nuisance parameters θ, and define *auxiliary measurements* y that target the latter
  - Then the pdf factorizes $P(x, y \,|\, \mu, \theta) = P(x \,|\, \mu, \theta) P(y, \theta)$
  - N.B. *y* are *global observables* in root

- To frequentists, P can be used as a likelihood in μ,θ
  - *Profile* away θ-dependence: $\mathscr{L}_p(\mu; x, y) = \max_\theta \mathscr{L}(\mu, \theta; x, y)$
  - With L_p we can do all of what was shown before (in approximation)

Fermilab

Full statistical model

- Split likelihood parameters into *parameters of interest* (POIs) μ and nuisance parameters θ, and define *auxiliary measurements* y that target the latter
  - Then the pdf factorizes $P(x, y \mid \mu, \theta) = P(x \mid \mu, \theta) P(y, \theta)$
  - N.B. *y* are *global observables* in root

- To frequentists, P can be used as a likelihood in μ,θ
  - *Profile* away θ-dependence: $\mathscr{L}_p(\mu; x, y) = \max_{\theta} \mathscr{L}(\mu, \theta; x, y)$
  - With L_p we can do all of what was shown before (in approximation)

- Bayesians can insert a *ur-prior* $P(\theta)$ and use Bayes' theorem to get $P(\theta \mid y)$
  - Then *marginalize* out θ: $P(x \mid \mu) = \int P(x, y \mid \mu, \theta) P(\theta \mid y)\, d\theta = \int P(x, y \mid \mu, \theta) \frac{P(y \mid \theta) P(\theta)}{P(y)}$
  - Proceed as before with $P(x \mid \mu)$

Full statistical model

- Split likelihood parameters into *parameters of interest* (POIs) μ and nuisance parameters θ, and define *auxiliary measurements* y that target the latter
  - Then the pdf factorizes $P(x, y \mid \mu, \theta) = P(x \mid \mu, \theta) P(y, \theta)$
  - N.B. *y* are *global observables* in root

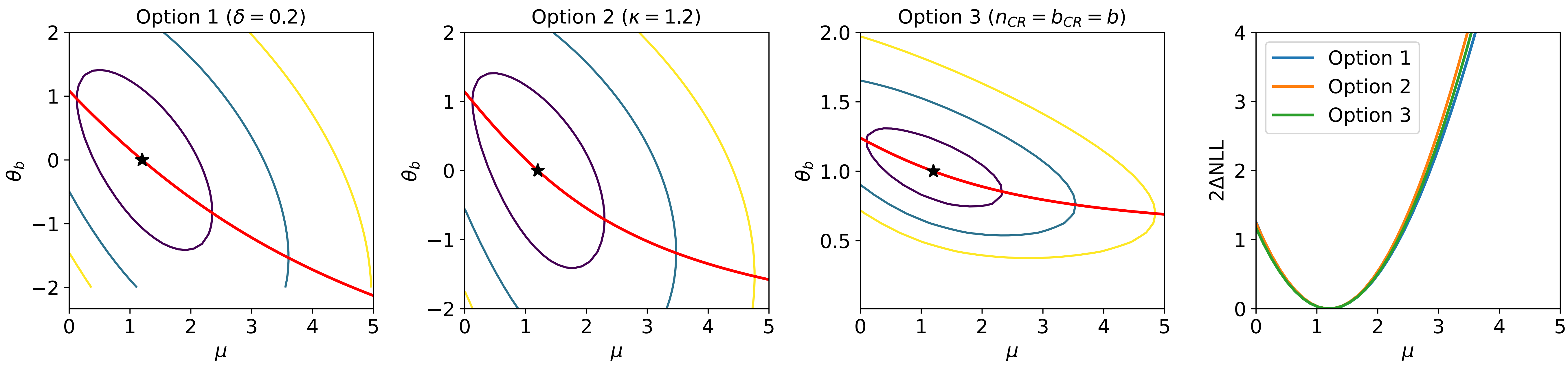
- To frequentists, P can be used as a likelihood in μ,θ
  - *Profile* away θ-dependence: $\mathscr{L}_p(\mu; x, y) = \max_\theta \mathscr{L}(\mu, \theta; x, y)$
  - With L_p we can do all of what was shown before (in approximation)

- Bayesians can insert a *ur-prior* $P(\theta)$ and use Bayes' theorem to get $P(\theta \mid y)$
  - Then *marginalize* out θ: $P(x \mid \mu) = \int P(x, y \mid \mu, \theta) P(\theta \mid y)\, d\theta = \int P(x, y \mid \mu, \theta) \frac{P(y \mid \theta) P(\theta)}{P(y)}$
  - Proceed as before with $P(x \mid \mu)$

- Renewed interest in publishing such statistical models: arxiv:2109.04981
  - Enables recasting in either language

Profiling example

- Profile likelihood for the single-bin background uncertainty example
  - Option 1: $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b(1+\delta\theta_b))\,\mathscr{N}(\theta_0\,|\,\theta_b,1)$
  - Option 2: $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b\kappa^{\theta_b})\,\mathscr{N}(\theta_0\,|\,\theta_b,1)$
  - Option 3: $P(n\,|\,\mu,\theta_b) = \mathscr{P}(n\,|\,\mu s + b\theta_b)\,\mathscr{P}(n_{CR}\,|\,b_{CR}\theta_b)$
  - s=10, b=25, x=37

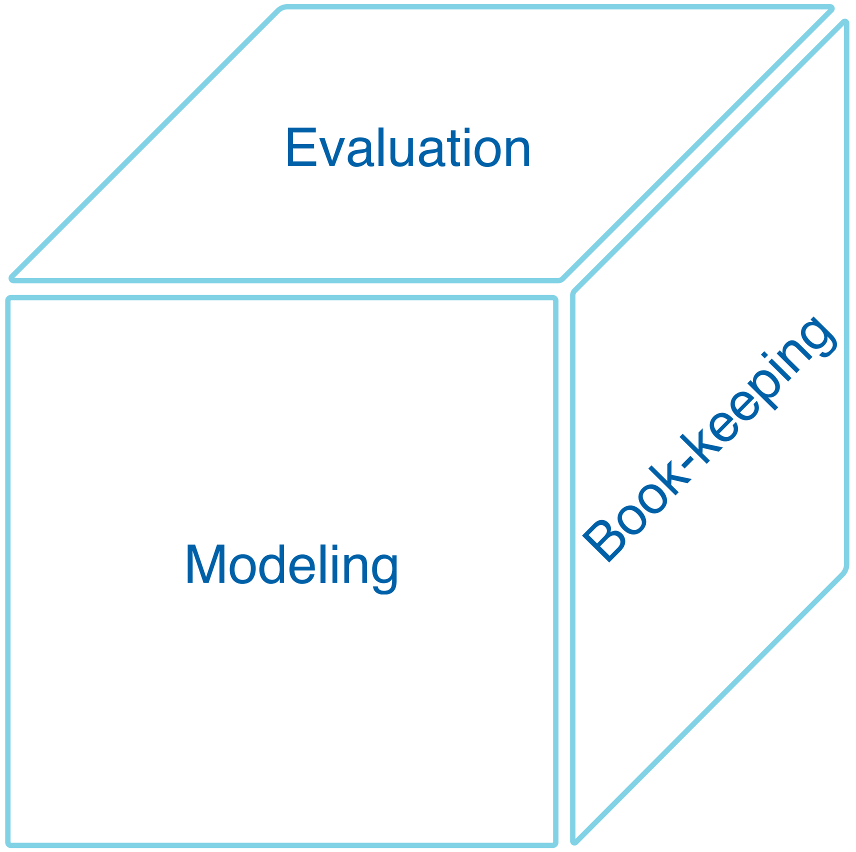


Typical tasks in enumerating systematics

- Enumerate effects to get dimension of θ
  - Don't forget anything! Unknown unknowns?
- Choose a parameterization
  - e.g. the options 1-3 from before
- Evaluate the constants
  - In practice: interpolate between shifted or weighted MC

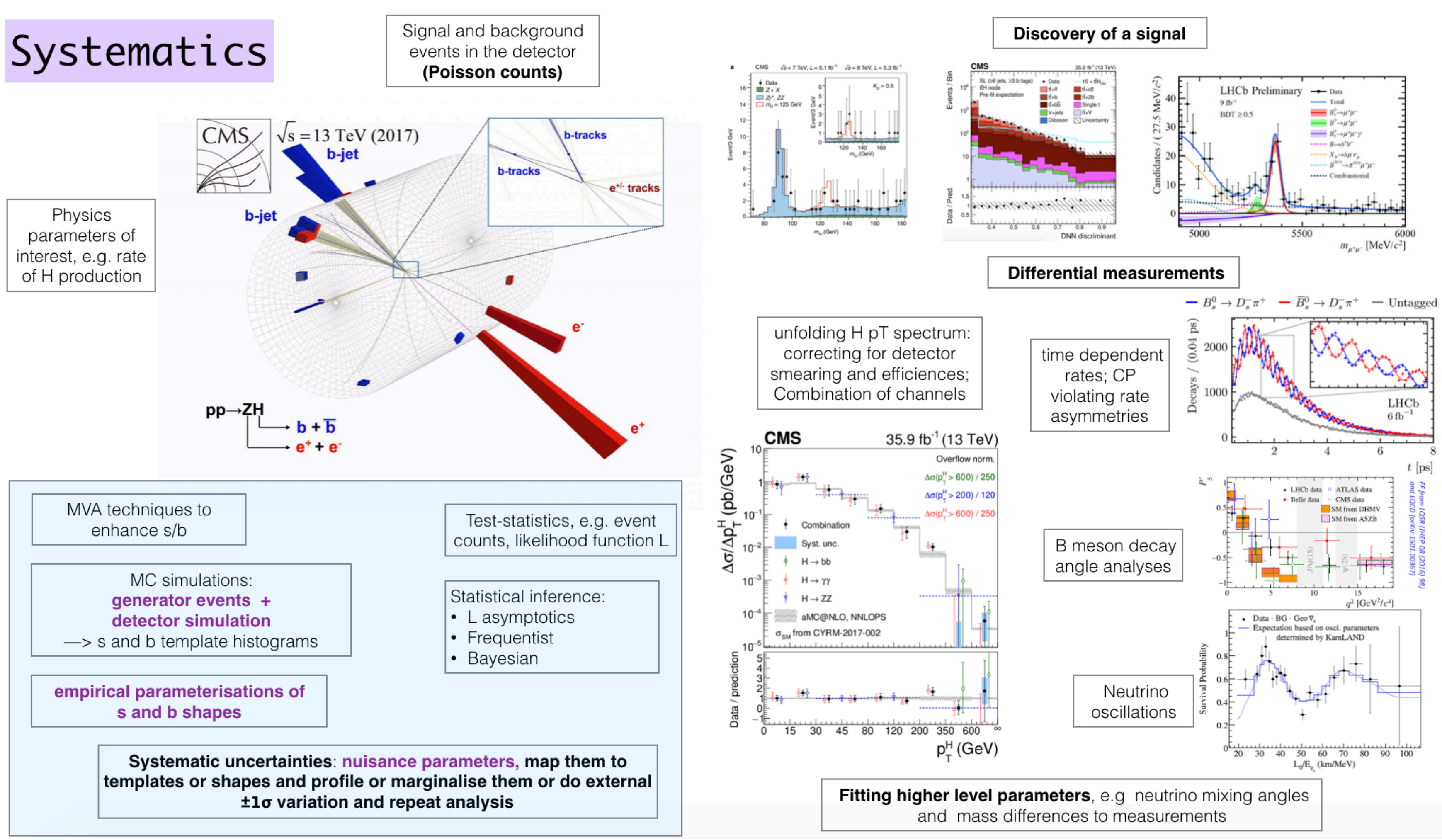
- Iterate
  - Compromise: fidelity/computability/practicality
    - Prune low-impact effects
  - Initial model might not fit observed y well

- A core feature of *Combine* is to simplify these tasks



Modeling techniques

- Can be a whole workshop
  - Was: PHYSTAT-Systematics 2021
  - Excellent presentations covering a wide range of techniques
  - A one-slide overview was produced (more on types than techniques)



Nuisance pars:

Luminosity

Detector:
- Acceptance
- Efficiency for specific particles
- Energy scales
- Resolutions

Signal process template:
- Theory modelling uncertainties
- Limited MC statistics

Background processes template:
- Theory total cross section uncertainty
- Theory modelling uncertainties
- Limited MC statistics

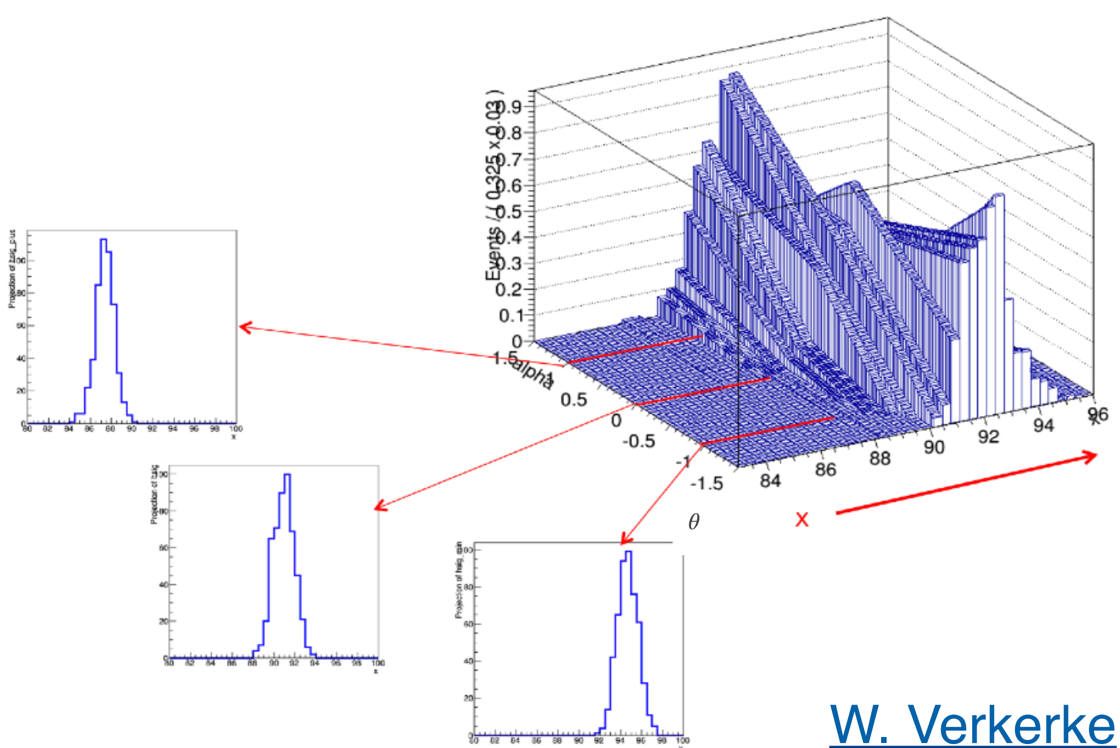
Empirical s and b shape modelling;
- Parameterisations
- Non-parametric
- smoothing and morphing of MC templates

Nuisance parameters can be constrained from:
- Detector calibration data
- Control samples with different event selection
- from the data distributions
- measurements from other experiments
- theory calculations

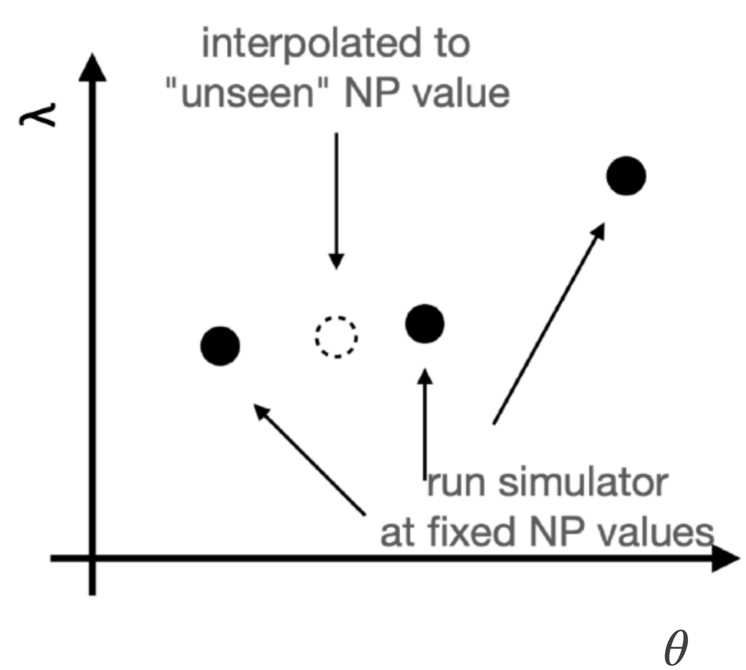
Modeling techniques

- **Rich** set of interpolation/extrapolation techniques at end-stage
  - Morphing: vertical, horizontal, moment; splines; gaussian process; asymmetric shift interpolation; additive/multiplicative effects; MC stat uncertainty, BB-lite; …
  - i.e. what is done in RooFit/pyhf/zfit/iMinuit/combine/etc.
    - What features do each of these tools offer? Nobody has it all!

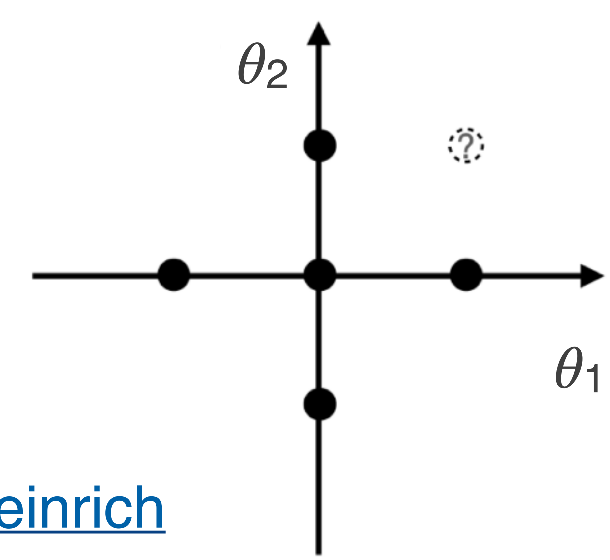
Visualization of bin-by-bin linear interpolation of distribution



W. Verkerke



Combining effects



L. Heinrich

The end

Hopefully you have some idea now what this means

"An observed (expected) upper limit is placed on the signal strength μ, using the profile likelihood ratio test statistic, following the CL_S criterion, under asymptotic assumptions, and found to be …"

Additional references

- Procedure for LHC Higgs combination http://cdsweb.cern.ch/record/1379837
- R. Cousins, Statistics in Theory https://arxiv.org/abs/1807.05996
- Asymptotic formulae for likelihood-based tests "CCGV" https://arxiv.org/abs/1007.1727
- Publishing statistical models https://arxiv.org/abs/2109.04981