On statistical methods for searches of new physics

Aixin Tan

Department of Statistics and Actuarial Science
University of Iowa

September 2016

Outline

An example of the search for the sterile neutrino

Using confidence interval to set constraints

Using CLs to set constraints

Revisit the example

Searches for new physics often become parameter estimation problems. Findings are presented as constraints on some parameter(s) of interest.

Approaches to set constraints: CI, CLs, Bayesian credible set.

An example: the search for the sterile neutrino.

Parameter of interest: $\beta = (\sin^2 2\theta, |\Delta m^2|)$.

Here $\sin^2 2\theta = 0$ stands for no oscillation, and hence no sterile neutrino.

Example of search of a sterile neutrino: one neutrino source, two detectors (300km and 1000km away), that measure neutrino energy E^ν at 20 equally spaced bins in [1, 9] GeV.
Disappearance measurements, for $i = 1, \cdots, 20$:

$$
\begin{aligned}
N_i &\sim \text{Pois}(\mu_i) = \text{Pois}(m\pi_i) \\
\pi_i &= a_i(\eta)\, P_i + b_i(\eta) \\
P_i &= 1 - \sin^2 2\theta \cdot \sin^2\left(1.27 \cdot \Delta m^2 \frac{L}{E_i^\nu}\right)
\end{aligned}
$$

m: experiment time/size.
a_i: detector "efficiency"; b_i: background.

θ : the neutrino mixing angle;
$|\Delta m^2|$: the neutrino mass squared difference;
η : nuisance parameters, such as detector efficiency, neutrino flux from reactor, target mass...

L : neutrino traveling distance;
E_i^ν : reconstructed energy of neutrinos.

CI obtained by inverting testing procedures

A level-c CI $=$

$$\{\beta_1 : \text{the hypothesis } \beta = \beta_1 \text{ is not rejected at level } 1 - c\}$$

For eg,

$$\text{A 95\% CI} = \{\beta_1 : TS_{\beta_1}(x) \text{ has p-value} > 0.05\}$$

where $TS_\beta(\cdot)$ is a test statistic, aka, a rule to order all possible values of x.

CI obtained by inverting testing procedures

A level-c CI $=$

$$\{\beta_1 : \text{the hypothesis } \beta = \beta_1 \text{ is not rejected at level } 1 - c\}$$

For eg,

$$\text{A 95\% CI} = \{\beta_1 : TS_{\beta_1}(x) \text{ has p-value} > 0.05\}$$

where $TS_\beta(\cdot)$ is a test statistic, aka, a rule to order all possible values of x.

Property of the above procedure:
Generate X 100 times from β_1, and construct a 95% CI each time. Then about 95 of the CIs contain β_1.

CI obtained by inverting testing procedures

A level-c CI $=$

$$\{\beta_1 : \text{the hypothesis } \beta = \beta_1 \text{ is not rejected at level } 1 - c\}$$

For eg,

$$\text{A 95\% CI} = \{\beta_1 : TS_{\beta_1}(x) \text{ has p-value} > 0.05\}$$

where $TS_\beta(\cdot)$ is a test statistic, aka, a rule to order all possible values of x.

Property of the above procedure:
Generate X 100 times from β_1, and construct a 95% CI each time. Then about 95 of the CIs contain β_1.

Remarks:
Property holds if the choice of TS depend on β_1.

Suppose $\beta = (\beta[1], \beta[2])$, such as $\beta = (\sin^2 2\theta, |\Delta m^2|)$

Let $\chi^2(x, \beta_1)$ denote some measurement of deviation of x from β. For each β_1 in the parameter space,

TS	
$\Delta\chi^2$	$\chi^2(x, \beta_1) - \min_\beta \chi^2(x, \beta)$
$\Delta\chi^2_{RS}$	$\chi^2(x, \beta_1) - \min_{\beta[1]} \chi^2(x, (\beta[1], \beta_1[2]))$
ΔT	$\chi^2(x, \beta_1) - \chi^2(x, \beta_{ref})$

Suppose $\beta = (\beta[1], \beta[2])$, such as $\beta = (\sin^2 2\theta, |\Delta m^2|)$

Let $\chi^2(x, \beta_1)$ denote some measurement of deviation of x from β. For each β_1 in the parameter space,

TS	
$\Delta\chi^2$	$\chi^2(x, \beta_1) - \min_\beta \chi^2(x, \beta)$
$\Delta\chi^2_{\text{RS}}$	$\chi^2(x, \beta_1) - \min_{\beta[1]} \chi^2(x, (\beta[1], \beta_1[2]))$
ΔT	$\chi^2(x, \beta_1) - \chi^2(x, \beta_{\text{ref}})$

Examples:

- $$\chi^2(x, \beta) = -2 \log p(x|\beta) + c(x)$$

Say, in observing $x = \{N_i\}_{i=1}^n$ where $N_i \sim \text{Poisson}(\mu_i(\beta))$.

$$\chi^2(x, \beta) = \sum_{i=1}^{n} 2 \left(\mu_i(\beta) - N_i + N_i \log \frac{N_i}{\mu_i(\beta)} \right) .$$

The corresponding TS are called likelihood ratio test (LRT) statistics. In HEP, the unified approach (Feldman and Cousins, 1998) to construct CIs uses the LRT statistic $\Delta\chi^2$.

Suppose $\beta = (\beta[1], \beta[2])$, such as $\beta = (\sin^2 2\theta, |\Delta m^2|)$

Let $\chi^2(x, \beta_1)$ denote some measurement of deviation of x from β. For each β_1 in the parameter space,

TS	
$\Delta\chi^2$	$\chi^2(x, \beta_1) - \min_\beta \chi^2(x, \beta)$
$\Delta\chi^2_{\mathrm{RS}}$	$\chi^2(x, \beta_1) - \min_{\beta[1]} \chi^2(x, (\beta[1], \beta_1[2]))$
ΔT	$\chi^2(x, \beta_1) - \chi^2(x, \beta_{\mathrm{ref}})$

Examples:

• Pearson's Chi-square statistic:

$$\chi^2(x, \beta) = \sum_{i=1}^{n} \frac{(N_i - \mu_i(\beta))^2}{\mu_i(\beta)}$$

Suppose $\beta = (\beta[1], \beta[2])$, such as $\beta = (\sin^2 2\theta, |\Delta m^2|)$

Let $\chi^2(x, \beta_1)$ denote some measurement of deviation of x from β. For each β_1 in the parameter space,

TS	
$\Delta\chi^2$	$\chi^2(x, \beta_1) - \min_\beta \chi^2(x, \beta)$
$\Delta\chi^2_{\text{RS}}$	$\chi^2(x, \beta_1) - \min_{\beta[1]} \chi^2(x, (\beta[1], \beta_1[2]))$
ΔT	$\chi^2(x, \beta_1) - \chi^2(x, \beta_{\text{ref}})$

With nuisance parameters η and additional constraints:

- $$\chi^2(x, \beta) = \min_\eta \left\{ \sum_{i=1}^{n} 2 \left(\mu_i(\beta, \eta) - N_i + N_i \log \frac{N_i}{\mu_i(\beta, \eta)} \right) + \chi^2_{\text{penalty}}(\eta) \right\}$$

or

- $$\chi^2(x, \beta) = \min_\eta \left\{ \sum_{i=1}^{n} \frac{(N_i - \mu_i(\beta, \eta))^2}{\mu_i(\beta, \eta)} + \chi^2_{\text{penalty}}(\eta) \right\}$$

where $\chi^2_{\text{penalty}}(\eta) = (\eta - \eta_0)^T V_\eta^{-1} (\eta - \eta_0)$

For each β_1 in the parameter space,

TS		natural for H_0	H_a	nested
$\Delta\chi^2$	$\chi^2(x, \beta_1) - \min_\beta \chi^2(x, \beta)$	$\beta = \beta_1$	$\beta \neq \beta_1$	yes
$\Delta\chi^2_{RS}$	$\chi^2(x, \beta_1) - \min_{\beta[1]} \chi^2$	$\beta[1] = \beta_1[1]$ $\beta[2] = \beta_1[2]$	$\beta[1] \neq \beta_1[1]$ $\beta[2] = \beta_1[2]$	yes
ΔT	$\chi^2(x, \beta_1) - \chi^2(x, \beta_{ref})$	$\beta = \beta_1$	$\beta = \beta_{ref}$	no

Recall with a certain choice of *TS*,

$$\text{A 95\% CI} = \{\beta_1 : TS(x, \beta_1) \text{ has p-value} > 0.05\}$$

Hence, need parent/sampling distn. of $TS(X, \beta_1)$ for $X \sim \beta_1$.

TS	natural for H_0	H_a	nested	Approx $X \sim \beta_1$	Distn $X \sim \beta_2(\beta_{ref}$
$\Delta\chi^2(X, \beta_1)$	$\beta = \beta_1$	$\beta \neq \beta_1$	yes	Chisq*	
$\Delta\chi^2_{RS}(X, \beta_1)$	$\beta[1] = \beta_1[1]$ $\beta[2] = \beta_1[2]$	$\beta[1] \neq \beta_1[1]$ $\beta[2] = \beta_1[2]$	yes	Chisq*	
$\Delta T(X, \beta_1)$	$\beta = \beta_1$	$\beta = \beta_{ref}$	no		

Requires large data size *m*, and

* cont. space of (β, η) (or $(\beta[1], \eta)$ in RS); likelihood smooth in them; true para. value not on boundary [Wilks 1938]

Recall with a certain choice of *TS*,

$$\text{A 95\% CI} = \{\beta_1 : TS(x, \beta_1) \text{ has p-value} > 0.05\}$$

Hence, need parent/sampling distn. of $TS(X, \beta_1)$ for $X \sim \beta_1$.

TS	natural for H_0	natural for H_a	nested	Approx $X \sim \beta_1$	Distn $X \sim \beta_2(\beta_{ref}$
$\Delta\chi^2(X, \beta_1)$	$\beta = \beta_1$	$\beta \neq \beta_1$	yes	Chisq*	nc Chisq**
$\Delta\chi^2_{RS}(X, \beta_1)$	$\beta[1] = \beta_1[1]$ $\beta[2] = \beta_1[2]$	$\beta[1] \neq \beta_1[1]$ $\beta[2] = \beta_1[2]$	yes	Chisq*	nc Chisq**
$\Delta T(X, \beta_1)$	$\beta = \beta_1$	$\beta = \beta_{ref}$	no		

Requires large data size *m*, and

* cont. space of (β, η) (or $(\beta[1], \eta)$ in RS); likelihood smooth in them;
true para. value not on boundary [Wilks 1938]
** and $|\pi(\beta_1, \eta_1^*) - \pi(\beta_2, \eta_2)| = O(m^{-\frac{1}{2}})$ [Wald 1943]. $\pi = \mu/m$

Recall with a certain choice of *TS*,

$$\text{A 95\% CI} = \{\beta_1 : TS(x, \beta_1) \text{ has p-value} > 0.05\}$$

Hence, need parent/sampling distn. of $TS(X, \beta_1)$ for $X \sim \beta_1$.

TS	natural for H_0	H_a	nested	Approx $X \sim \beta_1$	Distn $X \sim \beta_2(\beta_{ref}$
$\Delta\chi^2(X, \beta_1)$	$\beta = \beta_1$	$\beta \neq \beta_1$	yes	Chisq*	nc Chisq**
$\Delta\chi^2_{RS}(X, \beta_1)$	$\beta[1] = \beta_1[1]$ $\beta[2] = \beta_1[2]$	$\beta[1] \neq \beta_1[1]$ $\beta[2] = \beta_1[2]$	yes	Chisq*	nc Chisq**
$\Delta T(X, \beta_1)$	$\beta = \beta_1$	$\beta = \beta_{ref}$	no	Gaus$^{\#}$	Gaus$^{\#}$

Requires large data size *m*, and

* cont. space of (β, η) (or $(\beta[1], \eta)$ in RS); likelihood smooth in them; true para. value not on boundary [Wilks 1938]
** and $|\pi(\beta_1, \eta_1^*) - \pi(\beta_2, \eta_2)| = O(m^{-\frac{1}{2}})$ [Wald 1943]. $\pi = \mu/m$

$\#$ cont. space of η; likelihood smooth in η; plus diff kinds of conditions [Cowan et.al. 2011, Blennow et.al. arXiv1311.1822, Qian et.al. 2016]

* cont. space of (β, η) (or $(\beta[1], \eta)$ in RS); likelihood smooth in them; true para. value not on boundary [Wilks 1938]

$\#$ cont. space of η; likelihood smooth in η.

For a given $\beta_1 = (\sin^2 2\theta, |\Delta m^2|)$:
(1) if $\sin^2 2\theta = 0$, * does not hold, but $\#$ does.

(2) if $\sin^2 2\theta$ is near zero, * holds, but required sample size can be very large.

(3) if $\sin^2 2\theta$ is further away from 0, * holds.

In case of (1) and (2), no simple approx for computing p-value associated w. $\Delta\chi^2$ and $\Delta\chi^2_{\mathrm{RS}}$ in general. But yes to that of $\Delta T(X, \beta_1)$, using Gaus($\overline{\Delta T}, 4\overline{\Delta T}$) [next page].

Using $\Delta T(X, \beta_1)$ for CI has some problems [next section], an alternative is CLs.

[Goal] Approx the distr. of $\Delta T(X) = \chi^2(X, \beta_1) - \chi^2(X, \beta_{ref})$.

Denote per unit time expected count by π_1 and π_{ref} ($\pi = \mu/m$)

• Classical statistics literatures often assume

$$\pi_1^* - \pi_{ref} = O(m^{-\frac{1}{2}})$$

in order that the limiting distr. of $\chi^2(X, \beta_1)$ under β_{ref} exists (nc Chi-sq). [Wald 1943]
Also, $\chi^2(X, \beta_{ref})$ under β_{ref} follows Chi-sq. [Wilks 1938]

• Qian et. al. (2016) consider the case

$$\pi_1^* - \pi_{ref} = O(1)$$

Limiting distn of $\Delta T(X)$ under β_{ref} does not exist, yet an approximation at large m (omitting some O(1) terms) is Gaussian($\overline{\Delta T}, 4\overline{\Delta T} + ms$), where

$$\overline{\Delta T} = \min_{\eta} \sum_i \frac{(\mu_{ref,i}^{\text{true}} - \mu_{1,i}(\eta))^2}{\mu_{1,i}(\eta)} = \Delta T(x_{H_0}^{\text{Asimov}}, \beta_1)$$

$$s = \sum_i \frac{(\pi_{ref,i}^{\text{true}} - \pi_{1,i}^*)^3}{(\pi_{1,i}^*)^2}$$

• Qian et. al. (2016) approx distr. of $\Delta T(X)$, allowing for $\pi_1 - \pi_{ref} = O(1)$.

Conditions: (1) large m, (2) under both β_1 and β_{ref}, space of nuisance para. η cont., likelihood func. smooth in η

Approximate the distn of $\Delta T(X)$ under β_{ref} with Gaussian$(\overline{\Delta T}, 4\overline{\Delta T} + ms)$ where

$$\overline{\Delta T} = \min_{\eta} \sum_i \frac{(\mu^{\text{true}}_{ref,i} - \mu_{1,i}(\eta))^2}{\mu_{1,i}(\eta)} = \Delta T(x^{\text{Asimov}}_{H_0}, \beta_1)$$

$$s = \sum_i \frac{(\pi^{\text{true}}_{ref,i} - \pi^*_{1,i})^3}{(\pi^*_{1,i})^2}$$

Remarks:

• Qian et. al. (2016) approx distr. of $\Delta T(X)$, allowing for $\pi_1 - \pi_{ref} = O(1)$.

Conditions: (1) large m, (2) under both β_1 and β_{ref}, space of nuisance para. η cont., likelihood func. smooth in η

Approximate the distn of $\Delta T(X)$ under β_{ref} with Gaussian($\overline{\Delta T}, 4\overline{\Delta T} + ms$) where

$$\overline{\Delta T} = \min_{\eta} \sum_i \frac{(\mu^{\text{true}}_{ref,i} - \mu_{1,i}(\eta))^2}{\mu_{1,i}(\eta)} = \Delta T(x^{\text{Asimov}}_{H_0}, \beta_1)$$

$$s = \sum_i \frac{(\pi^{\text{true}}_{ref,i} - \pi^*_{1,i})^3}{(\pi^*_{1,i})^2}$$

Remarks:

Result was proved for $\chi^2(x,\beta) = \min_\eta \left\{ \sum_i \frac{(N_i - \mu_i(\beta,\eta))^2}{\mu_i(\beta,\eta)} \right\}$.

If further assume $|\pi^*_1 - \pi_{ref}| << \pi^*_1 \sim \pi_{ref}$ (often the case in search of new physics through precision measurements), then,

- ms small compared to $\overline{\Delta T}$, can omit in practice.
- Approx extends to other versions of $\chi^2(x,\beta)$

CI and CLs for comparing non-nested hypotheses

$$\Delta T(x) = \chi^2(x, \beta_1) - \chi^2(x, \beta_{\text{ref}})$$

can be used to form both CIs and CL$_s$, differ in how p-values are used. (Note that both are easy to carry out with the Gaussian approximation for the distribution of $\Delta T(x)$.)

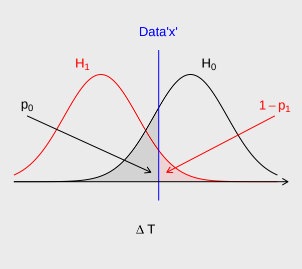


Figure: The distribution of $\Delta T(X) = \chi^2(X, \beta_1) - \chi^2(X, \beta_{ref})$, where $X \sim H_0 : \beta = \beta_{ref}$ (black) and $H_1 : \beta = \beta_1$ (red), respectively.

Large values of ΔT favors H_0 over H_1.

- CI excludes β_1 if $(1 - p_1)$ is small.

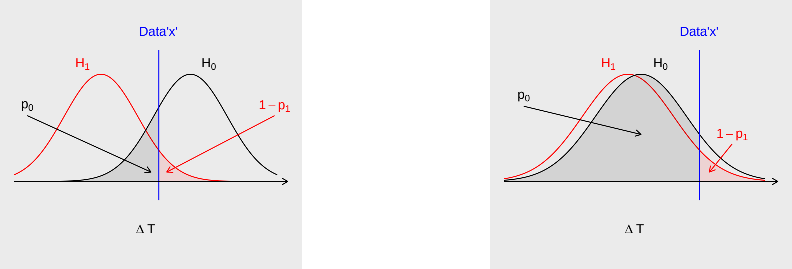


Figure: The distribution of $\Delta T(X) = \chi^2(X, \beta_1) - \chi^2(X, \beta_{ref})$, where $X \sim H_0 : \beta = \beta_{ref}$ (black) and $H_1 : \beta = \beta_1$ (red), respectively.

Large values of ΔT favors H_0 over H_1.

- CI excludes β_1 if $(1 - p_1)$ is small.

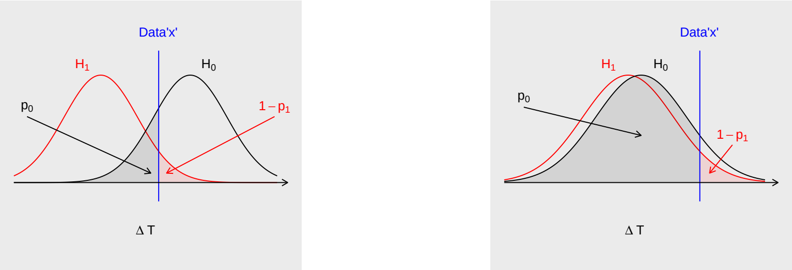


Figure: The distribution of $\Delta T(X) = \chi^2(X, \beta_1) - \chi^2(X, \beta_{ref})$, where $X \sim H_0 : \beta = \beta_{ref}$ (black) and $H_1 : \beta = \beta_1$ (red), respectively.

Large values of ΔT favors H_0 over H_1.

- CI excludes β_1 if $(1 - p_1)$ is small.

- CLs excludes β_1 if $(1 - p_1)$ is small **relative** to $(1 - p_0)$. Specifically,

$$\mathrm{CL}_s(x) = \frac{1 - p_1}{1 - p_0}.$$

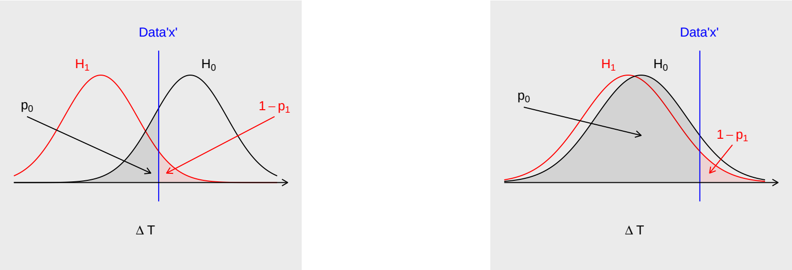


Figure: The distribution of $\Delta T(X) = \chi^2(X, \beta_1) - \chi^2(X, \beta_{ref})$, where $X \sim H_0 : \beta = \beta_{ref}$ (black) and $H_1 : \beta = \beta_1$ (red), respectively.

Large values of ΔT favors H_0 over H_1.

- CI excludes β_1 if $(1 - p_1)$ is small.

- CLs excludes β_1 if $(1 - p_1)$ is small **relative** to $(1 - p_0)$.

Specifically,

$$\mathrm{CL}_s(x) = \frac{1 - p_1}{1 - p_0}.$$

- Possible alternative: exclude β_1 if $(1 - p_1) < \alpha_1$ AND $(1 - p_0) > \alpha_0$, for some prespecified α_1 and α_0. [van Dyk, 2014]

$$\mathrm{CL}_s(x) = \frac{1-p_1}{1-p_0}.$$

In searching for new physics beyond SM, $\beta_{ref} = 0$,
level $(1-\alpha)$ exclusion set $= \{\beta : CL_s(x, \beta) < 1-\alpha\}$.

Since $\mathrm{CL}_s(x) = \frac{1-p_1}{1-p_0} \geq (1-p_1) =$ the p-value at β_1,
the set of para. values not excluded by CLs contains CI of the same level.
So region retained by a $(1-\alpha)$ level CLs covers true β with prob. over $1-\alpha$.

After all, CI and CL_s serve diff. purposes. The CL_s value appears more reasonable when the choice of TS has similar distr. under H_0 and H_1. i.e., when available data size is unlikely to carry enough info to differentiate them.

An example of the search for the sterile neutrino

Using confidence interval to set constraints

Using CLs to set constraints

Revisit the example

- Performance of various approximations
- CI vs CLs based on ΔT
- The Gaussian CL_s method vs. the MC CI method vs. the Raster-Scan MC CI method

Example of search of a sterile neutrino: one neutrino source, near and far detectors (300km and 1000km away), that measure neutrino energy E^{ν} at 20 equally spaced bins in [1, 9] GeV, with mean no. of neutrino events 10k and .9k per bin without oscillation. Background assumed linear in E^{ν}, w. 130 events in 1st bin to 73 for 20th bin.

Disappearance measurements, for $i = 1, \cdots, 20$, at $j =$near, far:

$$N_i^j \sim \text{Pois}(\mu_i^j) = \text{Pois}(m\,\pi_i^j)$$
$$\pi_i^j = a_i^j(\eta)\, P_{dis,i}^j + b_i^j(\eta)$$
$$P_{dis,i}^j = 1 - \sin^2 2\theta \cdot \sin^2\left(1.27 \cdot \Delta m^2 \frac{L_j}{E_i^{\nu}}\right)$$

Appearance measurements are similarly modeled, with $P_{app,i}^j = 1 - P_{dis,i}^j$.

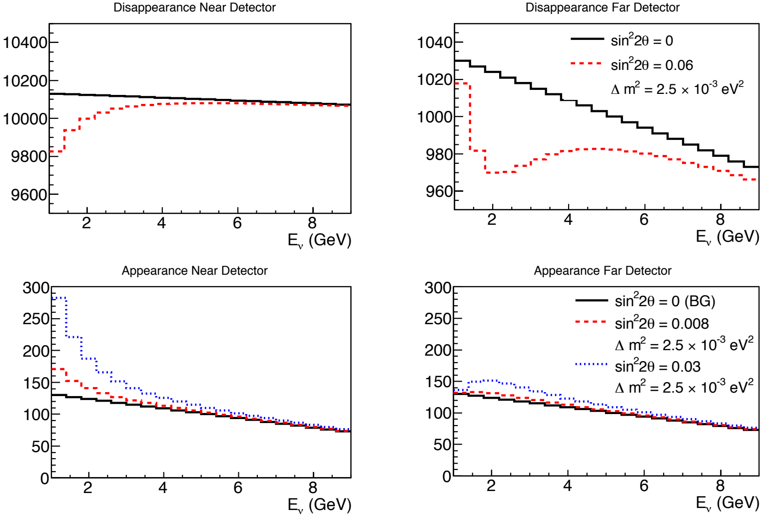


Figure: Mean no. of events seen at the near and the far detectors, in a disappearance (top) and an appearance (bottom) experiment.

Mean count of events in i-th bin of j-th detector,
$\mu_i^j(\epsilon, \eta_j, \sin^2 2\theta, |\Delta m^2|)$.

Para. of interest, $\beta = (\sin^2 2\theta, |\Delta m^2|)$

Nuisance para., $\eta = (\epsilon, \eta_n, \eta_f)$:
ϵ: detector efficiency and neutrino flux, w. 5% uncertainty.
η_n, η_f : background normalization factors for the near and the far detectors, with 2% uncertainty, uncorrelated between the two.

Set

$$\chi^2(x, \beta, \eta) = \sum_{j=n,f} \sum_{i=1}^{20} 2\left(\mu_i^j - N_i^j + N_i^j \log \frac{N_i^j}{\mu_i^j}\right) + \frac{\epsilon^2}{0.05^2} + \frac{\eta_n^2}{0.02^2} + \frac{\eta_f^2}{0.02^2}.$$

Compare different ways to set constraints to β:

T		CI	CLs
$\Delta\chi^2$	$\chi^2(x, \beta_1) - \min_\beta \chi^2(x, \beta)$	*	
$\Delta\chi^2_{\mathrm{RS}}$	$\chi^2(x, \beta_1) - \min_{\beta[1]} \chi^2(x, (\beta[1], \beta_1[2]))$	*	
ΔT	$\chi^2(x, \beta_1) - \chi^2(x, \beta_{\mathrm{ref}})$	*	*

Invalid Chi-square approx to $\Delta\chi^2$ at $\sin^2 2\theta = 0$

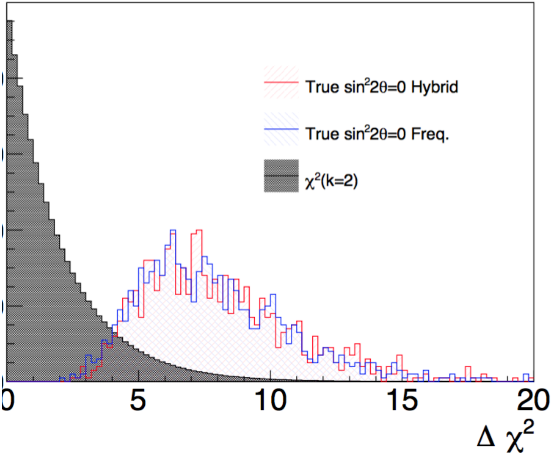
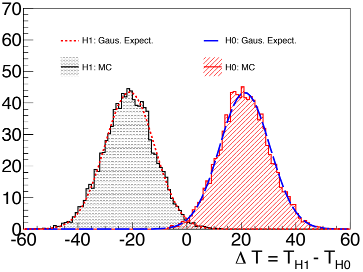


Figure: Distribution of $\Delta\chi^2 = \chi^2(\sin^2 2\theta = 0, |\Delta m^2|) - \min_\beta \chi^2(\beta)$ based on MCs with true $\sin^2 2\theta = 0$. Compared to the Chi-square(2) distribution.

Gaussian approx to ΔT at $\sin^2 2\theta = 0$ and elsewhere

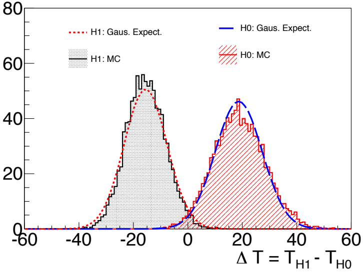
Disappearance measurements.

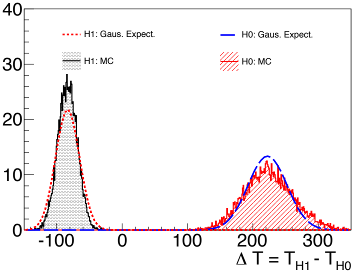


H_0 : $\sin^2 2\theta = 0$, vs
H_1 : $\sin^2 2\theta = 0.06$ and $|\Delta m^2| = 2.5 \times 10^{-3}$ eV^2.

Gaussian approx to ΔT (cont.)

Appearance measurements.





H_0 : $\sin^2 2\theta = 0$, vs
Left Panel H_1 : $\sin^2 2\theta = 0.008$ and $|\Delta m^2| = 2.5 \times 10^{-3}$ eV^2.
Right Panel: H_1 : $\sin^2 2\theta = 0.03$ and $|\Delta m^2| = 2.5 \times 10^{-3}$ eV^2.

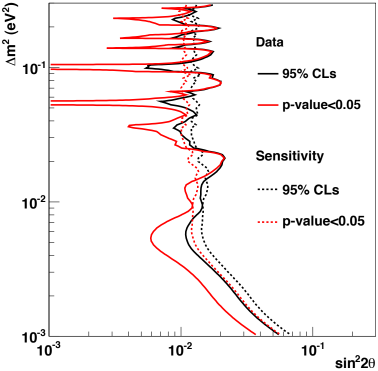
Cond. $|\mu_1 - \mu_{ref}| << \mu_1 \sim \mu_{ref}$ more likely to be violated for appearance measurements. Use $4\overline{\Delta_T} + ms$ for var. helps a bit.

Compare different ways to set constraints to β:

T		CI	CLs
$\Delta\chi^2$	$\chi^2(x, \beta_1) - \min_\beta \chi^2(x, \beta)$	*	
$\Delta\chi^2_{\mathrm{RS}}$	$\chi^2(x, \beta_1) - \min_{\beta[1]} \chi^2(x, (\beta[1], \beta_1[2]))$	*	
ΔT	$\chi^2(x, \beta_1) - \chi^2(x, \beta_{\mathrm{ref}})$	*	*

CI vs CLs based on $\Delta T = \chi^2(x, \beta_1) - \chi^2(x, \beta_{ref})$:

CI (red) based on ΔT can exclude β_1 that are not much less compatible with x than β_{ref} is. CLs (black) based on ΔT avoid this.



True value of $\sin^2 2\theta$ set to 0. For the CI (CL_s) method, para. values to the right of the red (black) line have p-values (CL_s values) below 0.05. Median sensitivity curves are based on MC: at each $|\Delta m^2|$, 50% of the MC samples have better exclusion limit than the sensitivity curve.

An example of the search for the sterile neutrino

Using confidence interval to set constraints

Using CLs to set constraints

Revisit the example

- Performance of various approximations
- CI vs CLs based on ΔT
- The Gaussian CL_s method vs. the MC CI method vs. the Raster-Scan MC CI method

Gaussian CL_s vs. the MC CI vs. Raster-Scan MC CI

Constraints produced by CLs and CI serve different purposes. Can still compare them, see for e.g., Read (2002); Cousins & Highland (1992).

We look at:

- CI based on $\Delta\chi^2$
- CI based on $\Delta\chi^2_{RS}$ (raster scan)
- CLs based on ΔT

Chi-square distribution does not well approximate the distribution of $\Delta\chi^2$ or $\Delta\chi^2_{RS}$ when the true $\sin^2 2\theta$ is near 0. Monte Carlo needed for both CI methods.

Gaussian approx for CLs.

MC CI based on $\Delta\chi^2$, MC CI based on $\Delta\chi^2_{RS}$, Gaussian CL_s, if true $\sin^2 2\theta = 0$:

Median sensitivity of the three methods are similar.

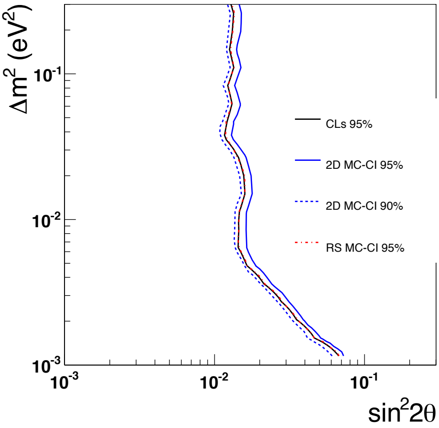


Figure: Set $\sin^2 2\theta = 0$ in generating MC. At each $|\Delta m^2|$, 50% of MC samples have a tighter exclusion limit than the sensitivity curve.

MC CI based on $\Delta\chi^2$, MC CI based on $\Delta\chi^2_{RS}$, Gaussian CL_s, if true $\sin^2 2\theta > 0$:

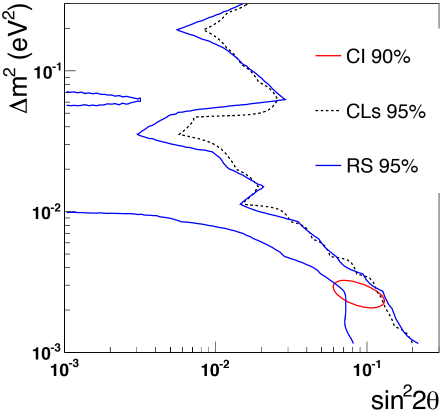


Figure: True para. value is $\sin^2 2\theta = 0.1$ and $\Delta m^2 = 2.5 \times 10^{-3}$ eV^2.

Use CI with $\Delta\chi^2$ if calculation of its parent distribution under each β_1 is affordable (Wilks, or in case of cond. violation, MC).

Otherwise, can adopt ΔT for the simplicity of its Gaussian approx given cond. satisfied, and pair it with CLs.
Gaussian CLs is also easy to use for combining multiple independent experimental results (Qian's talk this afternoon): the CL_s value at each hypothesis $\beta = \beta_1$ from experiments 1 to M can be calculated with

$$\begin{aligned} \Delta T &= \sum_{k=1}^{M} \Delta T(x_k, \beta_1), \\ \overline{\Delta T_{H_1}} &= \sum_{k=1}^{M} \overline{\Delta T_{H_1}^{(k)}} \\ \overline{\Delta T_{H_0}} &= \sum_{k=1}^{M} \overline{\Delta T_{H_0}^{(k)}}. \end{aligned}$$