Lecture I - Simulating an experiment

1) Opening remarks
- please, ask if unclear
- hands on, less formal, tutorial main part

1) Existing LBL experiments

\[ K2K \quad L = 295 \text{ km} \quad \nu_{\mu} \rightarrow \nu_{\mu} \]
\[ E = 1.3 \text{ PeV} \]
Superk as detector \( m_3 = 22.5 \text{ kg} \)

\[ MINOS \quad L = 733 \text{ km} \quad \text{phase I} \quad \nu_{\mu} \rightarrow \nu_{e} \]
\[ E = 3 \text{ PeV} \quad \text{phase II} \quad \nu_{\mu} \rightarrow \nu_{e} \]
5.4 \text{ kton} magnetized iron calorimeter

\[ CHOOJ \quad L = 33 \text{ km} \quad \nu_{e} \rightarrow \nu_{e} \]
\[ E = 20 \text{ peV} \]
5 at SCK CEN Opal

2) Why long baseline?

Oscillation probability

\[ P = \left| \sum_{ij} U_{ij} U_{\mu j} e^{-i \Delta m_{ij}^2 L / 2E} \right|^2 \]
- $\Delta m^2$ are given by nature
- $L$ and $E$ can, in principle, be chosen freely

Practical considerations for the choice of $E$:

Detection

CC reactions have a threshold:

- $\nu_e \rightarrow \nu_e$ @ 5 MeV
- $\nu_\mu \rightarrow \nu_\mu$ = 200 MeV
- $\nu_\tau \rightarrow \nu_\tau$ = 2 GeV

we need this to tag the flavour

Sources

- no $\nu_e$ sources (in absence of oscillation)
- no low energy ($< 100$ MeV) $\nu_e$ sources, leptonic pattern which decays into $\nu _e$ is the one?
- only low energy $\nu$-source are nuclear reactions $\nu_e(\nu_e)^{-} MeV$

which allows only $\nu_e \rightarrow \bar{\nu}_e$, i.e. no appearance measurement

$\Rightarrow$ appearance experiments need to use energies lower than 200 MeV

for $\Delta m^2 = 2.5 \times 10^{-3} eV^2 \Rightarrow L_{ee} > 200$ km

for $\Delta m^2 = 8 \times 10^{-5} eV^2 \Rightarrow L_{ee} > 6000$ km

$\Rightarrow$ only $\Delta m^2$ experiments possible

- $\nu$ will travel inside water $\Delta m^2$ results effects will play an important role

properties of source - propagation - detection
Several statements:

- Source
  - Flavor
  - Spectrum

- Propagation
  - Vector polarizations
  - More effects

- Detection
  - Energy range
  - Flavors

Mark integral for event rate calculation:

\[ \sum_{s_i} \sum_{E_{\text{min}}} \int dE \, d^2 \hat{E} \, \phi_{s_i} (E) \]

\[ \frac{\Delta A}{L^2} P_{s_i} \rightarrow \delta \left( E', \Theta_{11}, \Theta_{12}, \Theta_{13}, \delta, A_{21}, A_{23}, \Lambda, \xi, \phi(C) \right) \]

\[ \sum_{J} C \sigma_J (E) \, b_J (E, \hat{E}) \]

Detection reaction:

\[ T^J (\hat{E}) \, V^J (E', \hat{E}') \]

Detection:

- \( s_i \): initial flavor
- \( E \): true \( v \) energy
- \( s \): final (true) flavor
- \( A \): interaction type
- \( \sigma_J \): total \( x \)-section
- \( b_J \): differential energy distribution of reaction particles
- \( T^J (\hat{E}) \): detection efficiency for sec. part.
- \( V^J (E', \hat{E}') \): energy resolution factors of daughters
- \( E' \): reco. Energy
- \( s' \): reco. flavor
\[ \Phi = \{ \text{hadrons, } p, n \} \]

\[ f_{\text{hadrons}}(E_{\text{hadrons}}) = 0 \]

\[ f_{\text{protons}}(E_p) \]

\[ \sqrt{-g \cdot (E_p)} \cdot e^{-\frac{(E_p - E_p')^2}{2\sigma^2}} \]

\[ = 1 \cdot (E - E_p')^2 \]

Homework: Think about the difference for a QE

4) AEDC

(See chapter 10 & 11 of the CERN Manual)

Goals: abstract definition/description of a detector which still is realistic

- A detector maps true quantities into reconstructed quantities (observables)
- Only observables considered here: flavour and energy
- Linearity - detector response to \( N \) events is a superposition of \( N \) single events (how event rate compared to DAQ speed)
\[ p(E') = \int dE \, g(E) \, \delta(E_i, E') \]

for discrete observables like flavour \( \sum \rightarrow N(E_i, E') \rightarrow \beta_{g(E)}^i \)

\( \Rightarrow \) detector can be fully characterized by a set of mapping functions \( \Phi(E_i, E') \) and \( M_{g(E)}(E) \).

\( K \) and \( M \) will be determined from detector Monte Carlo.

AEDC building blocks

- Energy module (\( E \))
- K-sochi
- \( V/\overline{V} \)
- Si, Si

\[ \text{flux} \rightarrow \text{channel} \rightarrow \text{event rate} \]
Lecture 2

1. We have data (also called observations) $x = (x_1, x_2, \ldots, x_b)$
   \# number of bins

2. We now want to infer some information about $f$.

   a) point estimation - which value of $f$ fits the data best?
   b) interval estimation - what are the limits of the answer?
   c) goodness of fit - how well is the data described by our model?

We would like to be able to know the probability distribution function (pdf) of $f$ once we have observed $x$.

Excursion

A pdf assigns a probability density to each point in its domain, with scaling property:

\[
\rho(x) \geq 0 \quad x \in \Omega
\]

\[
\int_{\Omega} \rho(x) \, dx = 1
\]

Conditional probability: given $y$ has happened (or is true)

\[
\rho(x | y) \rightarrow \int \rho(x | y) \, dx = 1
\]

Bayes' Theorem

\[
\rho(x | y) \rho(y) = \rho(y | x) \rho(x)
\]

End (excursion)
\[ \rho(\mathbf{x} | \mathbf{z}) = \frac{\rho_{\mathbf{z}}(\mathbf{z}) \rho_{\mathbf{y}}(\mathbf{y})}{\rho_{\mathbf{z}}(\mathbf{z})} \]

\[ \text{determined from } \int \rho(\mathbf{z} | \mathbf{z}) \, d \mathbf{z} = 1 \]

\[ \rho(\mathbf{z}) \text{- so called prior knowledge on } \mathbf{z} \text{ or state vector, i.e. what we know about } \mathbf{z} \text{ without priori knowledge.} \]

\[ \text{is uniquely defined?} \]

\[ \rho(\mathbf{z} | \mathbf{x}) \text{- so called likelihood, since it is made pdf for } \mathbf{z}? \]

\[ \mathcal{L}(\mathbf{z}) = \rho(\mathbf{z} | \mathbf{x}) \]

The likelihood contains all the information about the experimental setup and the data.

\[-2 \frac{1}{\sigma_z^2} \mathcal{L}_{\text{min}} = \sum \frac{x_i - \mathbb{E}(\mathbf{x})}{\sigma_z} + \frac{x_i}{\sigma_z} \log \frac{x_i}{\mathbb{E}(\mathbf{x})} \]

Values of \( \mathbf{z} \) which have a large \( \mathcal{L} \) make it more probable to observe \( \mathbf{x} \), i.e. in some sense the fit the data better. Hence the best fit estimate for \( \mathbf{z} \) is the one which has the largest value of \( \mathcal{L} \).

\[ \mathcal{L}(\hat{\mathbf{z}}) > \mathcal{L}(\mathbf{z}) \quad \forall \mathbf{z} \in \mathcal{D} \]

This is called the 'maximum likelihood estimate' of \( \mathbf{z} \).

This solves \( 1 \). \( \Box \)

Since \( \mathcal{L} \) is the likelihood and a probability for \( \mathbf{z} \), we can define a normalized version

\[ T = \frac{\mathcal{L}(\mathbf{z})}{\mathcal{L}(\hat{\mathbf{z}})} \]

which compares the likelihood of \( \mathbf{z} \) with respect to \( \hat{\mathbf{z}} \).
In designing experiments or a so-called allowed region, we require that points $\hat{a}$ are added to $\alpha$ in increasing order $\pm T(\alpha)$.

An example:

$$
\begin{array}{c}
\text{T}_{\alpha} \\
\bullet
\end{array}
\begin{array}{c}
\text{T}_{\alpha} \\
\bullet
\end{array}
\begin{array}{c}
\text{T}_{\alpha} \\
\bullet
\end{array}
$$

How do we determine $T_{\alpha}$?

If we repeat the same experiment many times, we would get a distribution $P(\alpha)$ at each point $T_{\alpha}$ such that

$$
\int_{T_{\alpha}} P(\alpha) \, d\alpha = \alpha \quad \text{where } \alpha \text{ is the so-called confidence level, often expressed in percentage standard deviation (\sigma)}
$$

Theorem of Wilks

If $P(z \mid x_1, \ldots, x_d)$ and we fit $x_1 = x_1 \ldots x_d = x_d$ then $-2 \ln L \alpha$ follows a $\chi^2$ distribution with $d - \text{dof}$.

If $d \alpha = d \text{dof}$, valid only if $\hat{a}$ is unbiased and of variance $\sigma$.

Simulated data

Typically, we will be talking about future experiments, hence we have to compute these. It is straightforward to compute $\hat{a}$ for any choice of $\hat{a}$ (see Lecture 2).
There is some confusion in the literature about how to add effective charges to amorphous $\langle \hat{r}_e(\hat{r}_0) \rangle$ in an ensemble. Let us consider one particular example of this problem. In that case, we add different effective charges at the N ion and average all results with respect to this ensemble. This is of course very CPU intensive and for most applications impossible.

- Use $\langle \hat{r}_e(\hat{r}_0) \rangle$ instead and perform analysis for these data (once).

Let us call a result for an $X^2$ analysis $r$, the test issue can be rephrased as

$$r(\langle \hat{r}_e \rangle) = \langle r(\hat{r}_e) \rangle?$$

In a recent paper by T. Schinka it was shown that this assumption is valid with reasonable accuracy.

As a result $L(\hat{r}) = \lambda$?

---

Systematic errors

Examples: on-site correlations are only known up to some finite accuracy:
- Statistical volume determinants
- Background size

How to include these effects in the analysis? A pull approach.

Suggestion: Parameterize the effect.
Example: Statistical volume $\hat{V}(\xi(2)) \rightarrow (1+\xi) \xi(2)$
**Step 2** add a constant to all nuisance parameters

**Example:** \[ C(\tilde{e}) = \left( \frac{\tilde{e}}{\sigma_2} \right)^2 \] \( \sigma_2 \) is fiducial value equal to 10

Result: \( \chi^2 \)

\[ \chi^2_{\text{sys}} (\tilde{e}; \tilde{\tilde{e}}) = \chi^2(\tilde{e}; \tilde{\tilde{e}}) + C(\tilde{e}) \]

In our example,

\[ \chi^2_{\text{sys}} (\tilde{e}; \tilde{\tilde{e}}) = \begin{bmatrix} \sum_{i=1}^{N} i \end{bmatrix} \left\{ \frac{\xi_i^0 - \xi_i(\tilde{e}; \tilde{\tilde{e}})}{\sigma_i^2(\tilde{e}; \tilde{\tilde{e}})} \right\}^2 + \left( \frac{\tilde{e}}{\sigma_2} \right)^2 \]

**Step 3** get rid of the nuisance parameters

minimize \( \chi^2_{\text{sys}} \) with respect to \( \tilde{\tilde{e}} \)

\[ \chi^2_{\text{sys}} (\tilde{\tilde{e}}) = \min_{\tilde{\tilde{e}}} \chi^2_{\text{nuisance}} (\tilde{e}; \tilde{\tilde{e}}) \]

also called 'marginalization'

In general, the most difficult part is to choose a special one to be taken to

1. prevent identify possible sources of systematic errors
2. do fit a proper parametrization
3. determine the few most important ones
The 90% allowed region describes all combinations of \( \{ \theta_{13}, \theta_{23} \} \) which fits the data at the 90% CL. This is we are looking at the simultaneous distribution of \( \theta_{13} \) and \( \theta_{23} \), hence we have to use 2 dof.

Here we look at the 90% allowed region or intevally i.e. all values of \( \theta_{13} \) which are consistent with the data irrespective of the value \( \theta_{23} \) hence we use 2 dof.

This 90% error includes the effect of any correlation with \( \theta_{23} \). In the following work always use the number of fitted parameters shown on the axis i.e. if no fitted point is shown use 1 dof.

\( \theta_{13} \) – sensitivity

Assume: \( \theta_{13} = 0 \) independent of \( \delta \) since for \( \theta_{13} = 0 \), \( \delta \) is a physical

important to keep \( \delta \) free in the fit, since if you expand

is compatible with \( \theta_{13} = 0 \) there is no information about \( \delta \)
Assume $\theta_{33} \neq 0$ (now we need also to pick a value for $\delta^0$)

![Graph showing function $N^2$ vs $\theta_{33}$]

If $N^2(\theta_{33}=0) > N_c$, $\theta_{33}^0$ can be discovered at some CL.

Since we had to pick $\delta^0$ as well it makes usually sense to have the following plot

![Plot showing discovery possible]

CP violation discovery

Assume $\delta^0$ is CP violating, i.e. $\delta^0 \neq 0$ or $\delta^0 \neq \pi$ (need to pick $\theta_{33}^0$ as well)

![Graph showing function $N^2$ vs $\theta_{33}^0$]

If $\min (N^2(\delta=0), N^2(\delta=\pi)) \geq N_c$, we can discover CPV

Since we had to pick $\theta_{33}^0$, it makes sense to have following plot

![Plot showing discovery of CPV]