

ROOT講習会 第3回 資料

櫻井駿介 (YMAP, OMU)
Mail: ssakurai2023@gmail.com

自己紹介

櫻井 駿介

出身：栃木県餃子市→筑波大学→東京大学大学院

→大阪公立大学（ポスドク）

研究分野：広義の宇宙線物理学

プロジェクト：MAGIC、CTA-LST、Telescope Array、FAST

専門：

- 光検出器の測定・較正
- データ取得用システム開発

プチブーム：洗濯前の衣類の漂白

一言：最近はROOTから離れて、むしろC言語

* ここ5年で一番瘦せていたころ



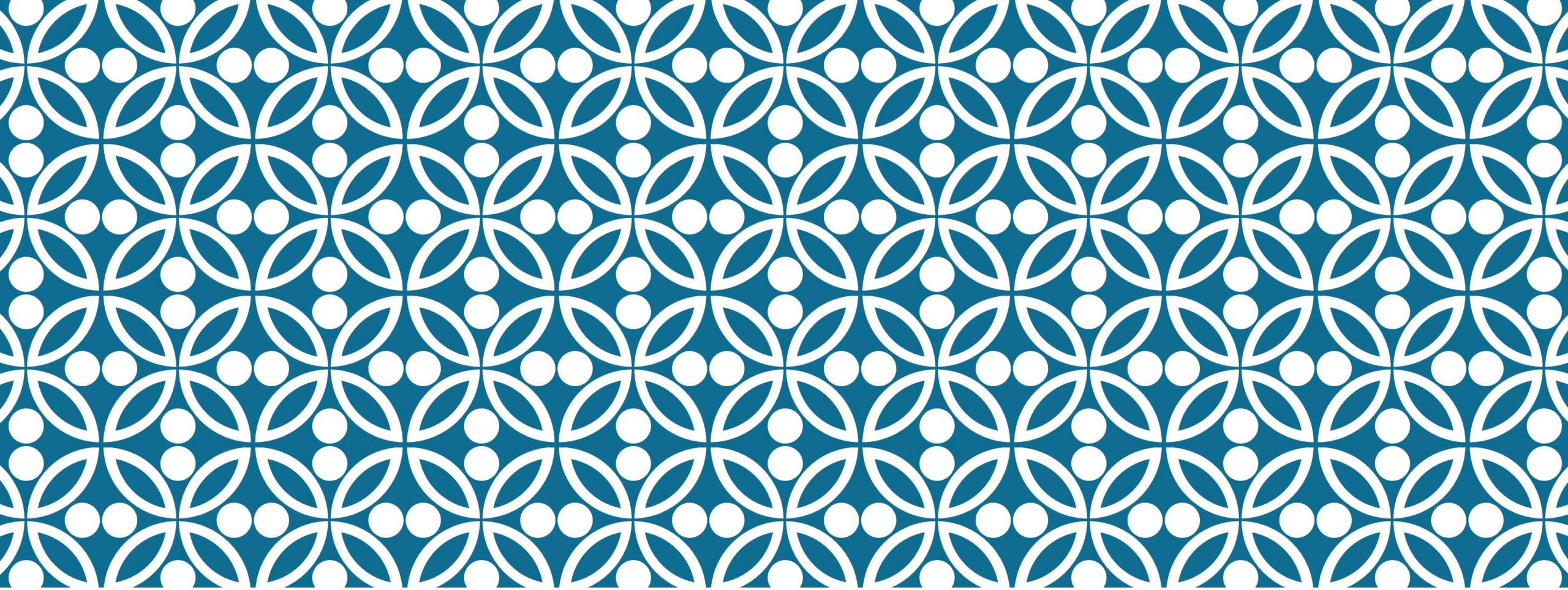
まえおき

自由に発言・質問・指摘してください

- “あなた”的ための講習会です、いつでも止めてもらって構いません。
- 質問ができるようになるまでの練習にどうぞ。
 - わからないことに気づく
 - 何がわからないかわかる
 - 何がわからないか説明できる
- その場で対応しきれない場合はSlackで対応します。
- ROOT講習会での発表は初めてなので資料にミスがあるかもしれません。

ChatGPTでもなんでも使ってください

- 作業効率が上がるならOKです。
- ただし、解釈は自分の頭を使いましょう（AIが“まだ”苦手なところ）



PART 1: REVIEW

前回までのおさらい

- 代表値（平均、標準偏差、中央値、四分位点）を学んだ
- ヒストグラムとは
- クラスの勉強
- マクロの勉強
- TH1Dの使い方

今回は、前半復習、後半はROOTのヒストグラムのより詳しい使い方

代表値いろいろ

平均 $\bar{x} = \langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i$

分散* $s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \langle x \rangle^2 = \langle x^2 \rangle - \langle x \rangle^2$

標準偏差 s

データを特徴づける良い指標だが、情報が失われていることに注意**。
分布を見ましょう。

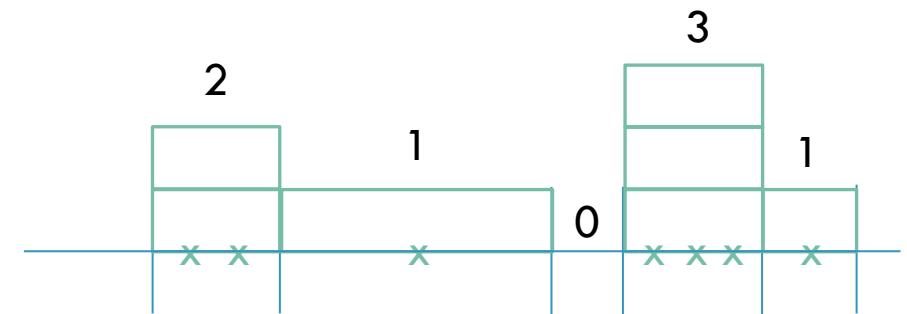
* 標本分散としておく。

**より高次の代表値もある。

ヒストグラム

データを階級ごとに分けて度数を図示したもの

- 階級：データを分ける範囲
- 度数：ある階級に属するデータの個数



メリット

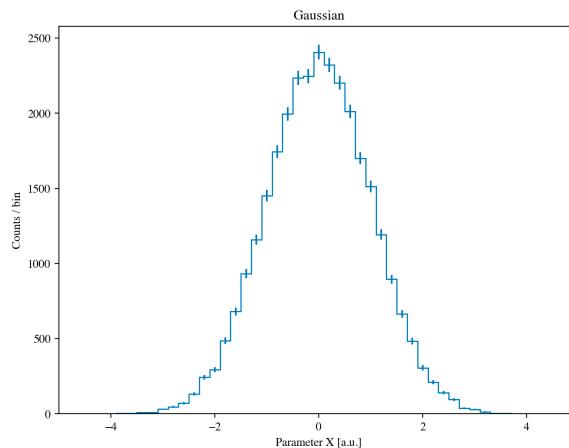
データを数値のまま眺めるよりも、視覚的な理解が簡単になる

デメリット

適切な情報を得るために適切なヒストグラムが必要。

データの分布をみることの重要性

全て
平均値=0, 標準偏差=1



共通点を探してみよう
ヒント：代表値

前回までのまとめ

ヒストグラム → データ可視化の手段

ROOT

→ データ解析向け機能がてんこもりのライブラリ

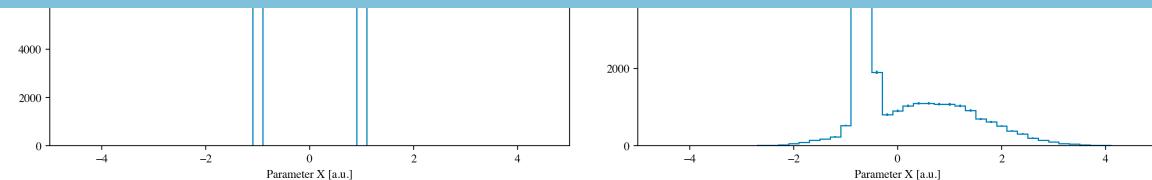
マクロ

→ ROOTを使った手続きをまとめた手順書

C++, クラス

→ ROOTを使うために必要な知識

であった...



ハンズオン1: おさらい（前回のつづき）

- ヒストグラムをROOTで図示してみよう

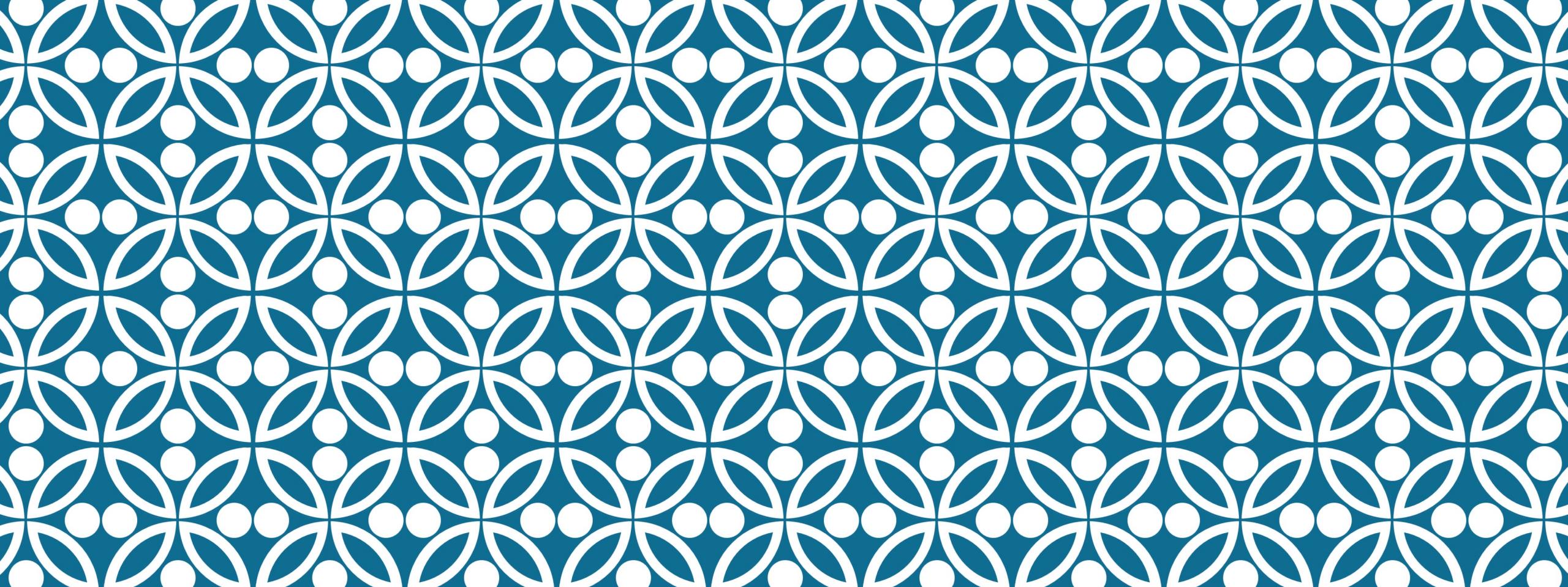
Macro:

```
/root_lecture/macros/mzks/scintillator_exp/macro_examples/draw_hist1.C  
/root_lecture/macros/mzks/scintillator_exp/macro_examples/draw_hist2.C  
/root_lecture/macros/mzks/scintillator_exp/macro_examples/draw_hists.C
```

- 思い出してほしいこと
 - TH1Dの使い方
 - マクロの実行方法

時間が余る人向け

- 画像として保存できるようにしよう：前回の講義資料p40
- データを追加しよう：前回の講義資料p37
- 外部変数を受け取れるようにしよう：



PART 2: 1D HISTOGRAM

ヒストグラムの使い道

データを可視化するだけでなく、解析にも使う

数える行為を伴えば、そこにはヒストグラム

- ・粒子の反応数
- ・粒子の検出数
- ・読み出した電流、電圧の値の回数...

今日の目標

ヒストグラム同士の比較やデータへのアクセスができるようにする。

ヒストグラムのデータへのアクセス

```
TH1D* hist = new TH1D("hist", "hist", MCACCh, HistMin-0.5, HistMax-0.5);
double BufferValue;
ifstream ifs("../data/data1.txt");
while(ifs >> BufferValue){
    hist->Fill(BufferValue);
}

hist->Rebin(8);

cout << "Sample Size      : " << hist->GetEntries() << endl;
cout << "Maximum Value   : " << hist->GetMaximum() << endl;
cout << "Maximum Bin     : " << hist->GetMaximumBin() << endl;
cout << "Mean             : " << hist->GetMean() << endl;
cout << "Standard Dev    : " << hist->GetStdDev() << endl;
cout << "Mean Error       : " << hist->GetMeanError() << endl;

cout << "7th Bin content: " << hist->GetBinContent(7) << endl;
```

```
root [0] TH1D::GetSize
GetSkewness
GetStatOverflows
GetStats
GetStdDev
GetStdDevError
GetSum
GetSumOfWeights
GetSumw2
GetSumw2N
root [0] TH1D::GetS■
```

ここでTab

一覧が

プログラミングに慣れていると混乱するが、最初のビンは1

ヒストグラムの比較

基本的には前回のマクロを参考に

Macro: /root_lecture/macros/mzks/scintillator_exp/macro_examples/draw_hists.C

目的に応じて使いわける必要がある

- 形状を比較したい場合→高さを合わせる（規格化）
 - 確率をみたいならエントリ数で合わせる
- 頻度を比較したい場合
 - 測定条件などを知る必要がある

ハンズオン2: ヒストグラムの比較と計算

Macro:

`/root_lecture/macros/mzks/scintillator_exp/macro_examples/draw_hists.C`
`/root_lecture/macros/mzks/scintillator_exp/macro_examples/use_methods.C`

を元にして。

二つのヒストグラムの高さを変更してみる。

時間が余る人向け

- ピークのまわりだけ積分してみる
 - `GetBinContent`, `GetBinCenter`
 - 人によってはまったく使わないかも

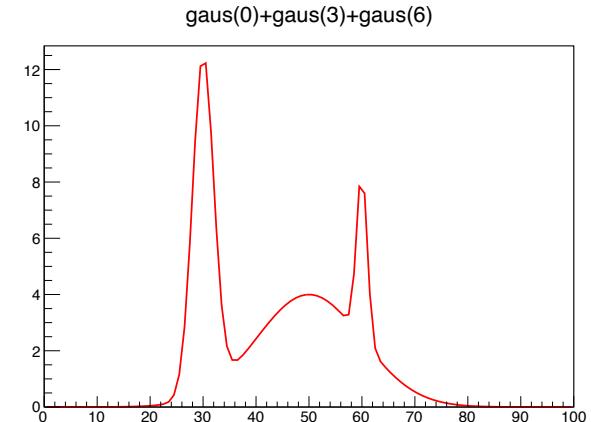
ハンズオン3: 統計学の定理を体感しよう

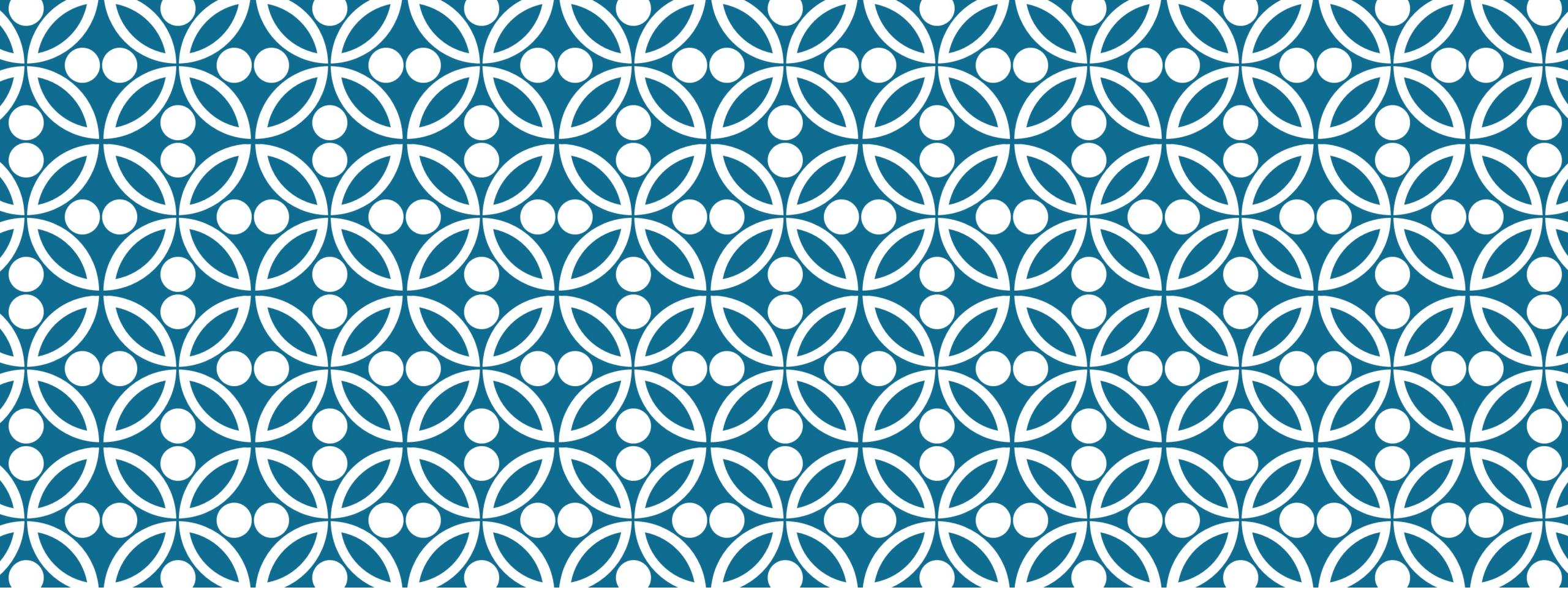
中心極限定理

“ごくおおまかにいえば、母集団分布が何であっても、和 $X_1+...+X_n$ の確率分布の形は、nが大なるときは、大略正規分布と考えて良い” 「統計学入門」 東京大学出版会

Template: //root_lecture/macros/sskr/hist_1d/macro_examples/draw_random_minimum.C

1. 雛形のマクロを完成させ、分布関数の形状を確認する。
2. 適当な数の乱数を生成し、ヒストグラムを作成する。
上で確認した関数の形状に従っているか確認してみる。
3. 生成した乱数をN個平均することを適当な数くり返し、ヒストグラムを作成する。
Nを増やした時のふるまいをヒストグラムを重ね書きして比較する。
4. 時間が余る人：振る舞いを解析解と比較してみよう。

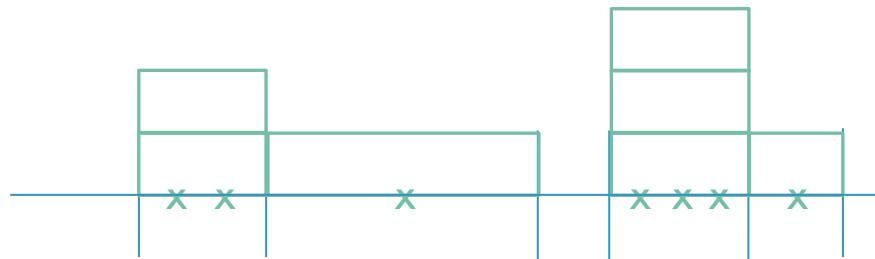




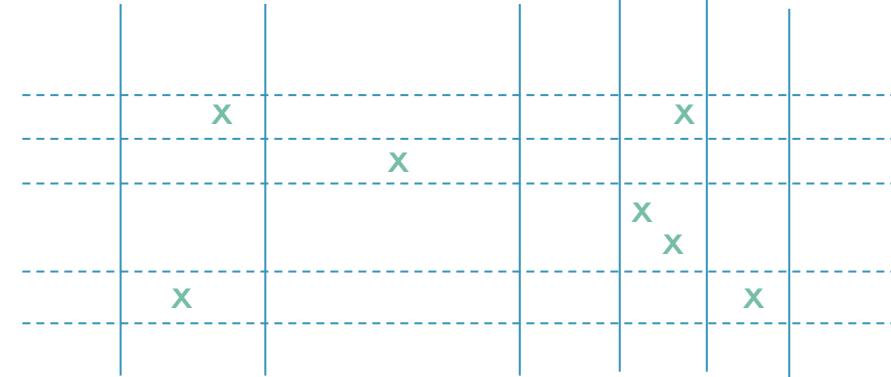
PART 3: 2D HISTOGRAM

二次元ヒストグラム

これまで:1D



ここから:2D



ある範囲内の数を図示するという点で同じ

用途はさまざま

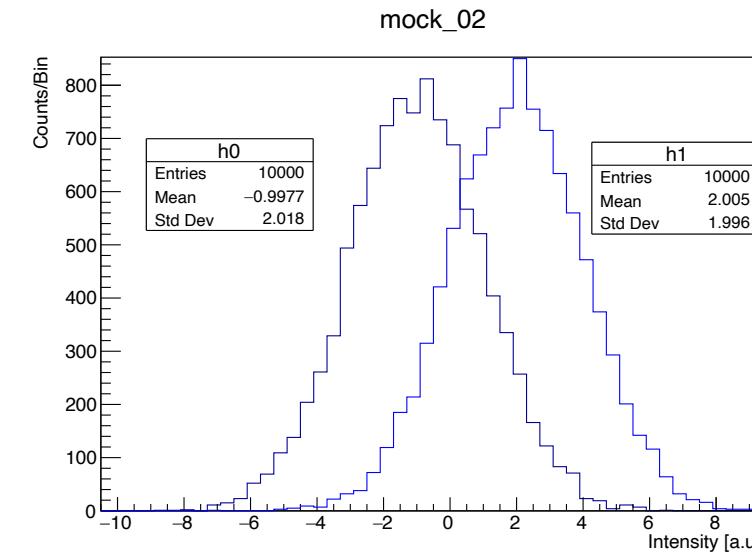
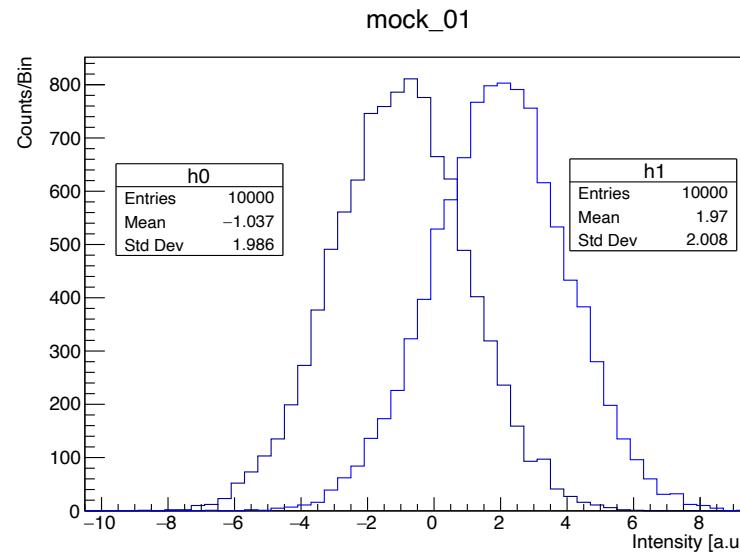
- ・検出器の入力と出力の関係
- ・宇宙線の到来方向ごとの頻度

散布図ではだめ？ → データ数が少ないうちはどうにかなる

頻度を高さにして3Dにする場合もあるがあまりみない（みにくい）

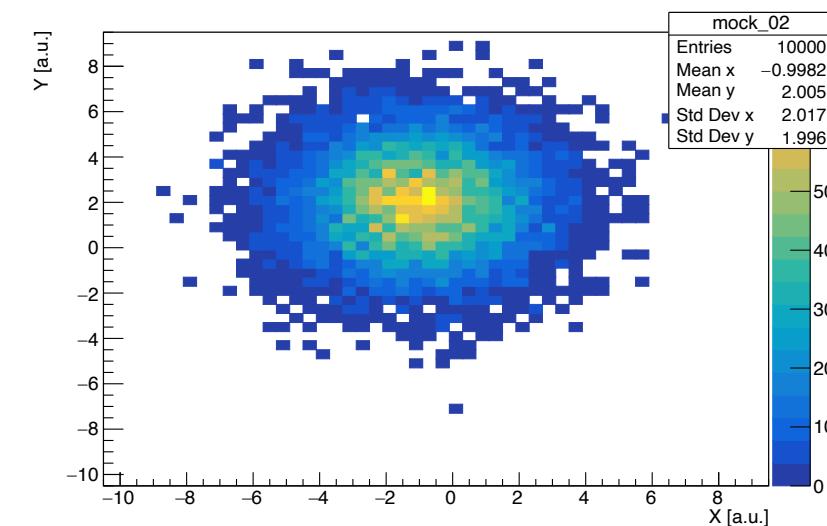
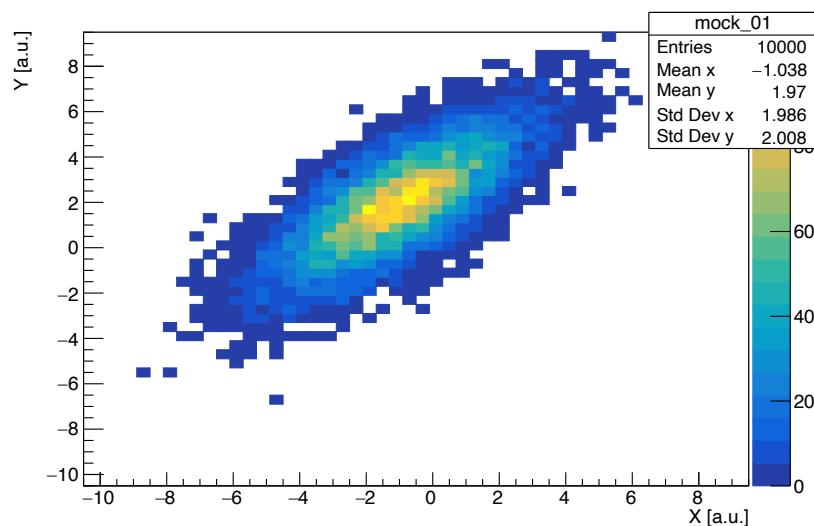
データを見ることの大切さ（再掲！）

二つの変数の組 (x, Y) を一万回測定したデータセットが二つ。
どちらも変数ごとの分布は同じ。これらのデータセットは同一視できる？



データを見ることの大切さ（再掲！）

- 二次元で見てみると振る舞いが異なる！この場合は相関の強さが違った。
- データが持つ性質を様々な視点で見るのが大切。



ROOTで二次元ヒストグラム

```
const int MCACh = 50;
const double HistMin = -10.0;
const double HistMax = 10.0;

TH2D *hist = new TH2D(filename, filename,
                      MCACh, HistMin - 0.5, HistMax - 0.5,
                      MCACh, HistMin - 0.5, HistMax - 0.5);

const TString data_dir = "../data/";
double BufferValueX;
double BufferValueY;
std::ifstream ifs(data_dir + filename + ".txt");
while (ifs >> BufferValueX >> BufferValueY)
{
    hist->Fill(BufferValueX, BufferValueY);
}
return hist;
```

基本的な扱いからはTH1と同じ

Y軸が増えたので
BINの設定も増えた

データも一列増えているので注意

Fillも引数が増えた

ハンズオン4: 二次元ヒストグラムを作ろう

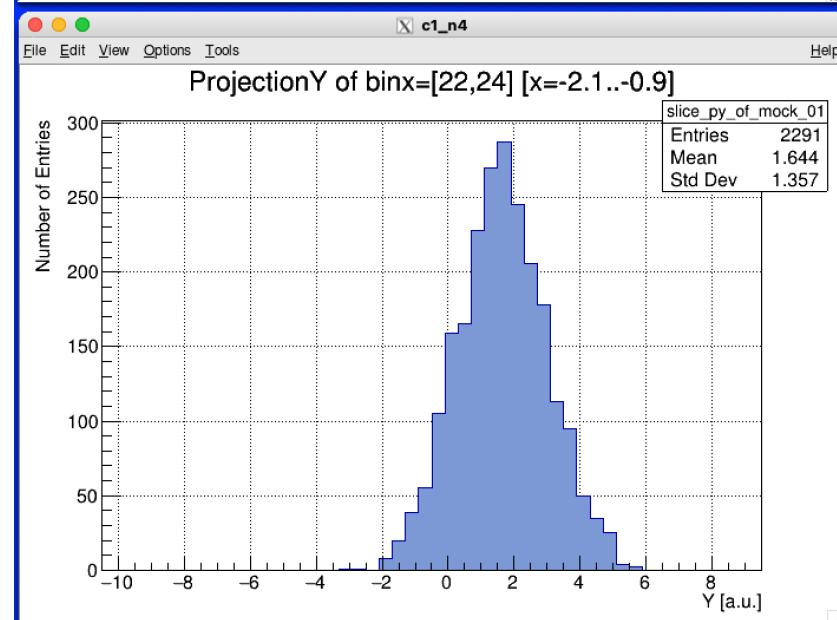
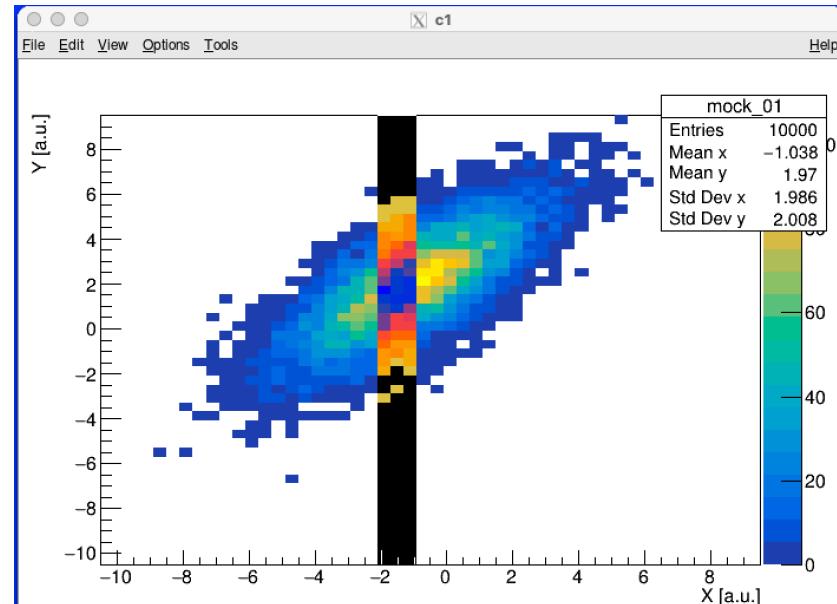
```
Data   : //root_lecture/macros/sskr/hist2d/data/mock_01.txt  
Macro  : //root_lecture/macros/sskr/hist2d/macro_example/draw_2dhist.C
```

時間が余った人向け

- 色（頻度）の範囲を変えてみる。
- 色の見え方は人により異なる。より多くの人に正しい情報を伝えるにはどうすれば良いか。

データの一部を見る

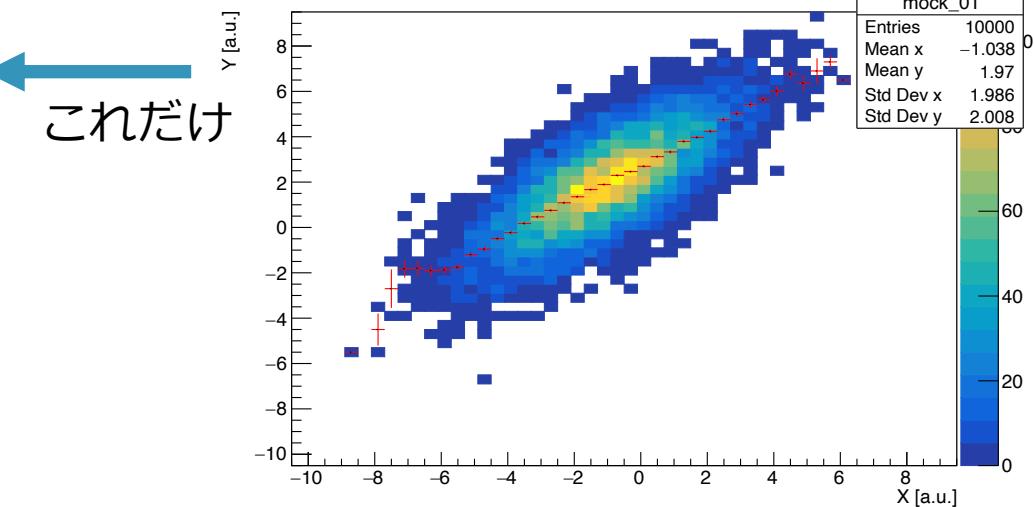
- 2D ヒストограмの表示中に
[右クリック] → [SetShowProjectionX/Y] で
X/Y軸への射影が表示できる。
- より高次元の場合は TTree や TCut を使用する
ほうが良い。



データの傾向を抜き出す

- TProfileを使うとTH2の軸に沿って binごとの平均値が抜き出せる。
- 便利だが、ヒストグラムに依存するので注意

```
TH2D *hist = make_hist(fname);
TProfile *prof = (TProfile *)hist->ProfileX();
hist->SetTitle("X [a.u.];Y [a.u.]");
hist->Draw("colz");
prof->SetLineColor(kRed);
prof->Draw("same");
return 0;
```



まとめ

- ✓ データの分布を見ることはとても大切。
- ✓ ROOTでヒストグラムが作れるようになった。
- ✓ 次はデータのモデル化（フィッティング）です。

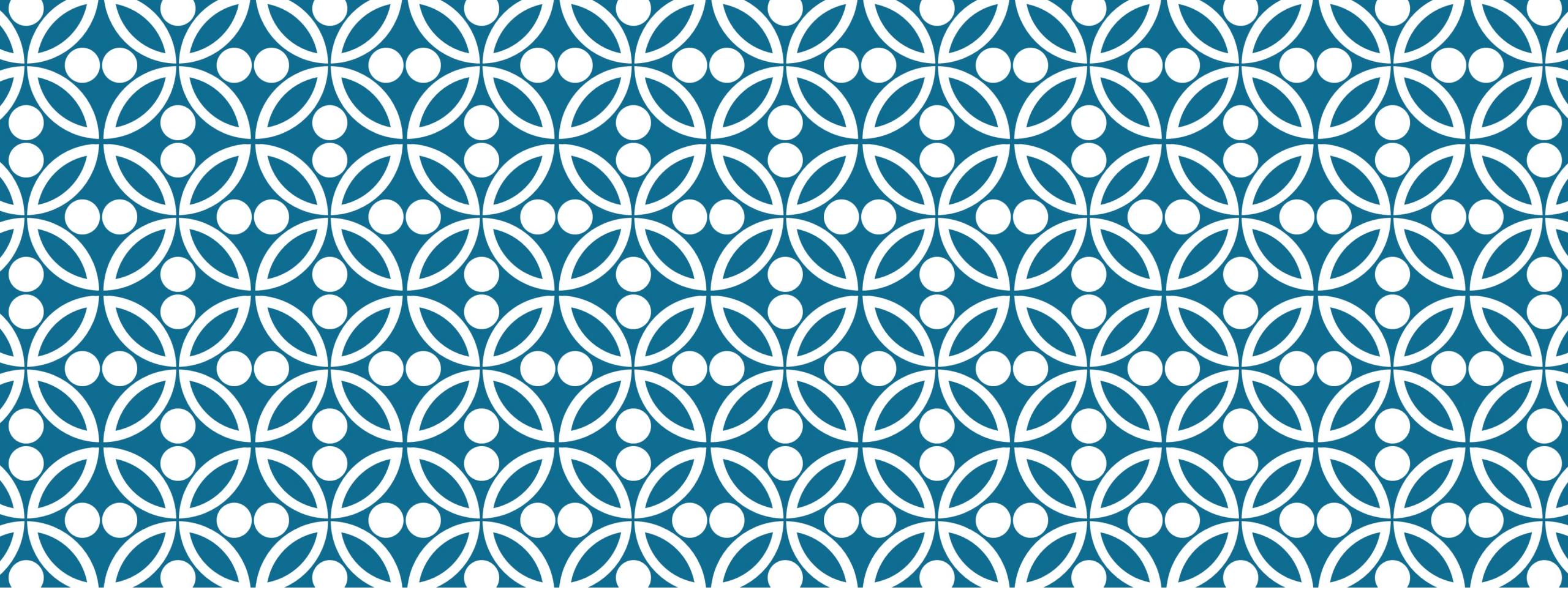
宿題: 高エネルギー宇宙線の観測データを見よう

Pierre Auger 観測所はWebで検出した宇宙線の一部について、その詳細な情報を公開している。

URL: <https://opendata.auger.org/data.php>

- どの公開データでも良いので、一次元または二次元ヒストグラムを作り、変数どうしの関係について考察してください。
- 様式も提出も自由です。公開データ中の各変数の意味についてはSlackで質問してください。

一例



PART 4: FITTING (予習)

フィッティング

物理実験におけるフィッティングとは

- データにモデル関数をあてはめ、特徴量を得る操作
 - フィッティングが合う≠データの真の分布がモデル関数に従う
- モデル関数は、慣習やセンスによる
 - しばしばガウス関数やローレンツ関数が登場
 - 研究者の意図が入る
- フィッティングは万能な手法では無い
 - 初期値によってはうまく計算が収束しない場合も
 - わりと経験が必要

最小二乗法

データとモデルの“距離”が最小になるようなモデルを探す手法

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - f(x_i))^2}{\sigma_i^2}$$

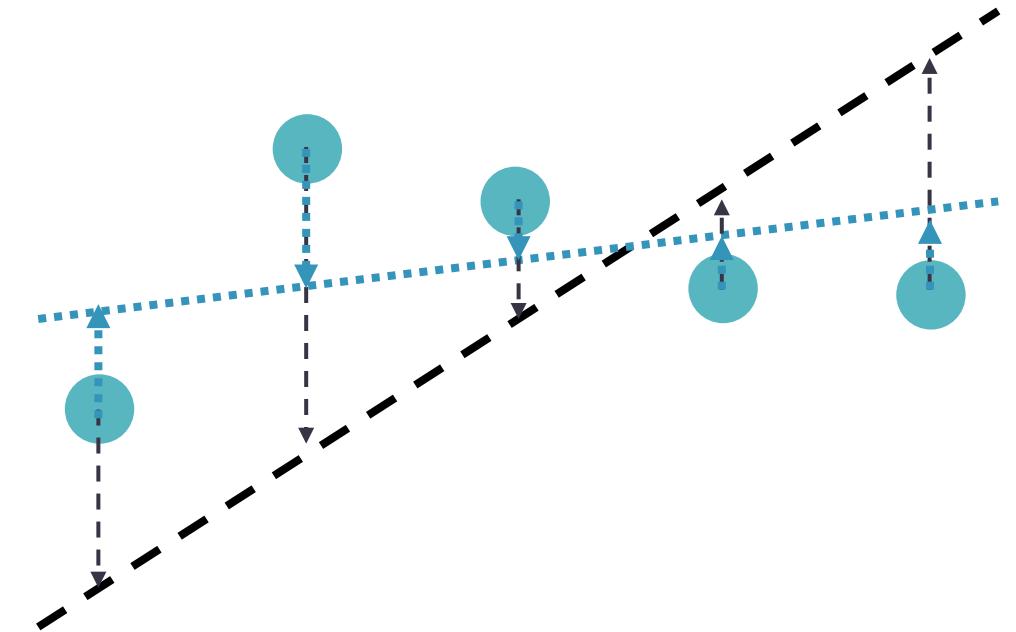
- 物理実験屋さんが最小二乗法という時、それはカイ二乗を最小することを意図していたりする。
- ある分野ではカイ二乗はすでにポアソン分布が仮定されていたりする
- ある分野では誤差は無いものとして扱ったりする。
- 気をつけましょう

FITTINGの気持ち

黒より青の方がなんだか良さそう

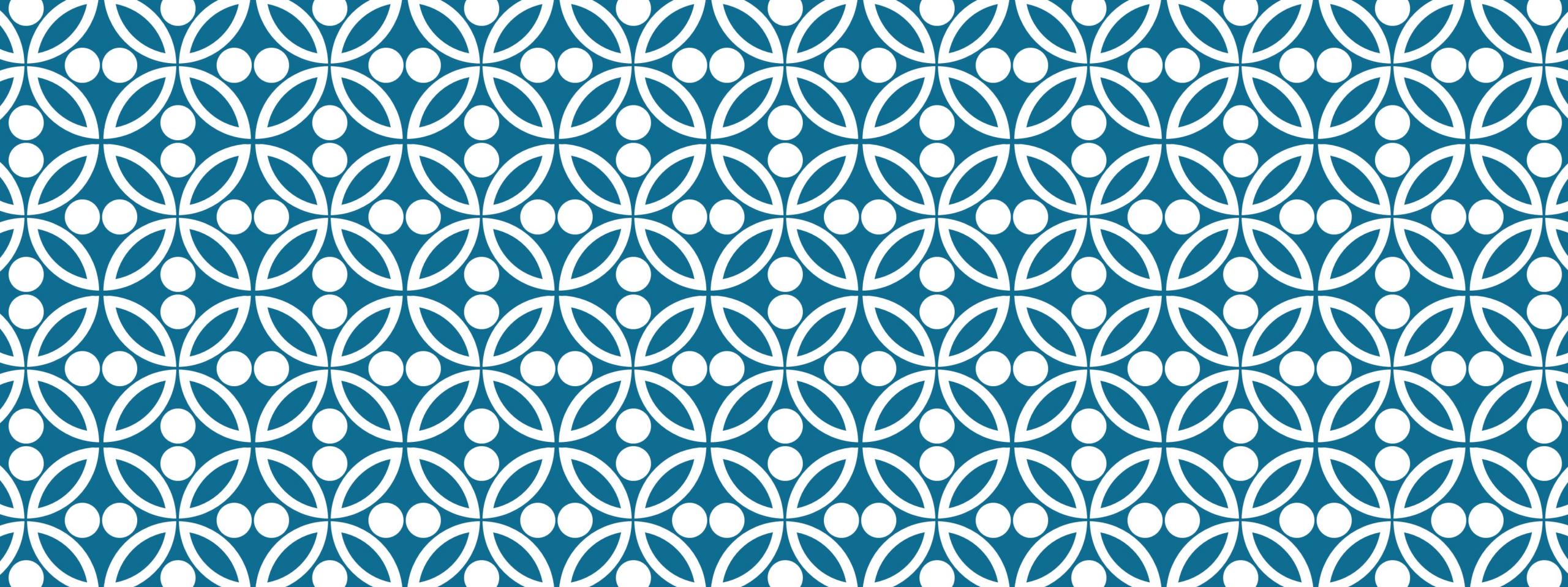
矢印の大きさの合計が一番小さくなるパラメーターを探す

物理実験には誤差がついてる
誤差が大きいときはどうする？



ハンズオン4：宿題

- 先週の宿題をやろう
- Step up!!
 - 一次関数の時の最小二乗法は解析解が求められるので導出してみよう
 - 参考文献



おまけスライド

前提知識

ROOTの使い方を学びたい

Linux 基礎コマンド

標準偏差

ROOTマクロ

ヒストグラム