EXAMPLES / EXERCISES

- Religious belief / Dogma

Take a set of possible events $\{A_i\}$ constituting a non-intersecting partition of Ω.
Imagine we have the following prior probability for each A_i:

$$\pi(A_i) = \begin{cases} 1 & \text{if } i = 0 \\ \emptyset & \text{if } i \neq 0 \end{cases}$$

which means we believe A_1 is absolutely true, and $A_{i \neq 1}$ are absolutely false.
Assume we acquire more information through the observation of some event B.

From Bayes theorem: $$P(A_i | B) = \frac{P(B|A_i)\,\pi(A_i)}{P(B)}$$

If $i \neq 0$: $$P(A_i | B) = 0 = P(A_i)$$

If $i = 0$: $$P(A_0|B) = \frac{P(B|A_0)\,\pi(A_0)}{\sum_i P(B|A_i)\,\pi(A_i)} = \frac{P(B|A_0)\cdot 1}{P(B|A_0)\cdot 1} = 1 = P(A_0)$$

This is what we call a dogma or religious belief, i.e. a belief that we cannot change, no matter how much new knowledge we accumulate.

The scientific method has allowed to improve our knowledge of Nature by progressingly accumulating new experimental evidences.
On the other hand, scientific progress is not possible in the presence of religious beliefs about observable facts.

• Binomial

~~Plot histograms on the same~~

~~Compute the~~ Compute the distribution of Binomial variables with $n = 150$ and

$p = 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$

Plot them all together (overlap them). Do not use any predefined method for the binomial or factorial!

Repeat for $n = 500$.

Tricks: 1) Binomial coefficient with exp(log)

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1}{k(k-1)\cdots 3\cdot 2\cdot(n-k)(n-k-1)\cdots 3\cdot 2}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 3\cdot 2}$$ ↳ works numerically up to $k \approx 170$

$$= \exp\left[\ln\left(n(n-1)\cdots(n-k+1)\right) - \ln\left(k(k-1)\cdots 3\cdot 2\right)\right]$$

$$= \exp\left[\sum_{i=n-k+1}^{n} \ln(i) - \sum_{i=2}^{k} \ln(i)\right]$$

2) Bernoulli term with exp(log) to solve situation with small p and large $(1-p)$

$$p^k (1-p)^{n-k} = \exp\left[\ln\left(p^k (1-p)^{n-k}\right)\right]$$

$$= \exp\left[k \ln(p) + (n-k)\ln(1-p)\right]$$

• ~~Poisson~~

~~Populate histograms with 10^5 random generated values~~

Ex 3

$\cancel{f(x,n) = \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)}$

- χ^2 distribution

  1) ~~Populate a histogram with 10^5 randomly generated~~

  1) Generate 10^5 sets of ~~num~~ n standard-normal random numbers x, and populate a histogram with $\chi^2 = \sum_{i=1}^{n} x^2$
  2) Repeat point (1) for $n = 1, 2, 3, 5, 10, 20$
  3) Overlap the histograms with the theoretical χ^2 distribution with the corresponding number of DOF, properly scaled to the number of generated values.

  Do not use any predefined χ^2 method.

- Central limit theorem for Gaussian random number generator

  1) Generate 10^8 sets of n uniformly distributed random numbers ~~in~~ in $[0, 1[$, with $n = 1, 2, \ldots, 10$

  2) ~~Populate For every n, populate a histogram~~

  2) For every n, populate a histogram with the 10^8 sums of ~~gener~~ the n generated random numbers
  3) Plot the histogram in linear and log-scale (on the y axis)

  Do you notice anything?

- Acceptance-rejection method
  1) Generate 10^6 points uniformly distributed in $x \in [-1, 1]$ and $y \in [-1, 1]$
  2) Compute the fraction of points falling inside a circle of radius 1
  3) Repeat for dim $= 2, 3, 4, \ldots, 15$
  4) Plot fraction of accepted points as a function of dimension
  5) Assign uncertainty to your estimate of accepted points
  6) Compare (overlap plots) with theoretical expectation of number of accepted points.

$\boxed{\text{Ex 5}}$

- Scatter plot (χ^2 fit)

1) Generate $N=11$ points with the following coordinates:

$$x_i = 0, 100, 200, \ldots, 1000$$

$$y_i = m\, x_i + q + \delta y_i \quad \text{where} \quad m = 1.5$$

$$q = 3$$

$$\delta y_i = \text{random Gaus}(\mu = 0, \sigma_i)$$

$$\sigma_i = \text{random uniform in } [10, 100[$$

To each point, assign the uncertainties:

$$\sigma_{x_i} = 0$$

$$\sigma_{y_i} = \sigma_i$$

2) Fit the scatterplot with the χ^2 fit, and retrieve the minimum χ^2 from the fitter.

↳ In ROOT:
```
TFitResultPtr resPtr = scatterplot->Fit(...);
TFitResult* res = resPtr.Get();
double chi2 = res->Chi2();
```

3) Repeat for 10^4 Toy-MC data, and plot the histogram of min-χ^2.

4) Overlap with the theoretical χ^2 distribution with the same number of DOF

↳ $nDOF = n - 2$ (because we have 2 parameters)

- Scatter plot χ^2 fit with overestimated uncertainty

Repeat the previous exercise, artificially enlarging the uncertainty assigned to y_i with respect to the one used to generate δy_i.

Use the following: $\sigma_{y_i} = k\,\sigma_i$ with $k = \{1, 1.1, 1.2, 1.3, 1.4\}$

$\boxed{\text{Ex 7}}$

• Efficiency curve fit with ~~binomial, Poisson and ?~~
binomial $\mathcal{L}$, Poisson $\mathcal{L}$ and Gaussian χ^2

1) Generate 20 points with coordinates (E_i, k_i)

$$E_i = 1, 2, 3, \ldots, 20 \text{ keV}$$

$$\mathcal{E}_i = \frac{P}{2}\left[1 + \mathrm{erf}\left(\frac{E-\mu}{\sigma}\right)\right] \quad \text{with} \quad P = 0.9, 0.95, 0.99$$
$$\mu = 3 \text{ keV}$$
$$\sigma = 1 \text{ keV}$$

$\forall \mathcal{E}_i$, generate n random values uniformly distributed in $[0, 1[$ and ~~count the set k_i = #~~ set k_i to the number of values $< \mathcal{E}_i$.

Consider the following values for n: 100, 1000, 10 000

2) Fit the data ~~assumi~~ assuming the following distributions for k_i:

a) Binomial

b) Poisson

c) χ^2 (Neyman or Pearson, your choice)

~~3) Plot the chart Repeat for all~~

3) Repeat for 10^3 toys for all combinations of n and P

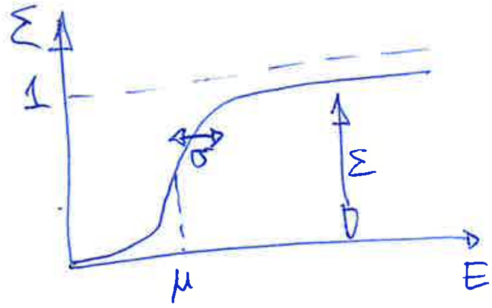
4) Plot the best fit distribution of $\hat{\mathcal{E}}$ obtained with the 3 methods, for all combinations of p and n.

This ~~example~~ is the case of a pulse generator injecting n artificial pulses into a detector, with 20 different energy values between 1 and 20 keV.

$\mathcal{E}$ is the overall trigger efficiency

μ is the trigger threshold

σ describes the rise of the efficiency curve.



Ex. 9

- Neyman belt for Binomial

Assume ~~again~~ we have a detector injected with n pulser events, and that we want to measure the trigger efficiency p.

Take: n = # of injected events = $\{0, 1, 2, \ldots, 10\}$

p = input trigger probability $= 0, 10^{-3}, 2\cdot 10^{-3}, \ldots, 1$

1) Create a 2-dim histogram for (k, p).
   Notice that $0 \le k \le n$, so k must go from 0 to 10.
2) For each value of p, populate the corresponding distribution of k with a Binomial$(k|n,p)$
3) For each value of p, extract the shortest 68% interval
4) Plot the Neyman confidence belt
5) Compute the coverage of p.

- Flip-flopping

Take a variable x with PDF: $f(x|\mu) = \frac{1}{2\pi} \exp\left(-\frac{(x-\mu)^2}{2}\right)$

Assume x can be negative, but $\mu \ge 0$.

Assume we decide to quote a central 90% interval if $x \ge 3$, or an upper 90% limit if $x < 3$.

1) Prepare a 2-dim histogram for (x, μ)
2) Populate the Neyman belt
3) Construct the 90% central confidence belt for $x \ge 3$
4) Construct the 90% lower confidence belt for $x < 3$
5) Join the two belts and plot them.
6) Compute and plot the coverage of μ.

$\boxed{\text{Ex 11}}$

- Feldman - Cousins for electron neutrino mass

In this example, we try to reproduce the ~~mass~~ limit on electron ν mass reported by KATRIN in Phys. Rev. Lett. 123 (2019) 221802.

Their limit is: $m_\nu < 1.1$ eV

obtained from: $m_\nu^2 = -1.0^{+0.9}_{-1.1}$ eV^2

~~1) Prepare a text~~ Our variables are m_ν (the parameter of the model) and ~~the~~ m^2 (the observable). Notice that m_ν can't be negative, but the observable m^2 can.

Let's approximate the PDF of m^2 with: $f(m^2 | m_\nu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(m^2 - m_\nu^2)^2}{2\sigma^2}\right)$

Let's assume $\sigma = 1$, which is close to the asymmetric uncertainties quoted above.

1) Populate the Neyman belt, using $m^2 \in [-5, 20]$ and $m_\nu \in [0, 4]$
2) Extract $\max f(m^2 | m_\nu)$ for each value of m_ν
3) Populate the likelihood ratio 2-dim histogram
4) Extract the FC confidence belt
5) Compute the coverage (using the Neyman belt)
6) Extract the central interval or limit ~~correspondi~~ on m_ν corresponding to a measured value $m^2 = -1.0$

How does it compare with the published one?

Ex 13

- Bayesian efficiency fit

1) Generate 20 points with coordinates (E_i, k_i)

$$E_i = 1, 2, 3, \ldots, 20 \text{ keV}$$

$$\varepsilon_i = \frac{p}{2}\left[1 + \mathrm{erf}\left(\frac{E-\mu}{\sigma}\right)\right] \quad \text{with} \quad \begin{aligned} p &= 0.99 \\ \mu &= 3 \text{ keV} \\ \sigma &= 1 \text{ keV} \end{aligned}$$

$\forall\ \varepsilon_i$, generate 1000 random values uniformly distributed in $[0, 1[$ and set k_i to the number of values $< \varepsilon_i$.

2) Run a Bayesian fit with Metropolis-Hastings using the following distributions for k_i:

   a) Binomial
   b) Poisson
   c) χ^2

3) Plot the posterior of p for the three fits.

- ~~Radioactive decay fit~~

Ex 15