STATISTICAL ANALYSIS IN EXPERIMENTAL PARTICLE PHYSICS

Kai-Feng Chen
National Taiwan University

#1: COVARIANCE AND CORRELATION

- There is a small root tree containing 3 random variables, x, y, and z. It can be downloaded from the lecture resource webpage: http://hep1.phys.ntu.edu.tw/~kfjack/lecture/hepstat/02/exercise_01.root
- Please write a short piece of code to calculate the corresponding **3×3 covariance matrix** and the **3×3 correlation matrix**.

#2: VERIFY THE CASE OF POISSON ⊗ BINOMIAL

- In the main lecture we have discussed the case of joint Poisson and binomial distributions. Let's verify it with the following setup.
- Remember the test carried out in exercise 1:

```
{
    TRandom3 rnd;

    int count = 0;
    for(int i=0;i<100;i++)
        if (rnd.Uniform()<0.1) count++;

    cout << count << endl;
}
```

Replace the "100" times in the code with a random variable generated with Poisson μ=100.

- In this case the "count" should follow another Poisson distribution with $\mu=100\times0.1=10$. Please repeat the generation for 10,000 times and prove the variance is also $V=10$.

#3: CENTRAL LIMIT THEOREM

- According to the Central limit Theorem, summing over a sequence of independent variables will result a Gaussian distribution, regardless the original individual distributions.
- Is it really always true? Try to sum up **3 different random variables** of your own choice (*e.g. Poisson, Binomial, etc., anything you like!*) and see if the number of variables increases, the resulting distribution is getting closer to a Gaussian, ie. plot the sum of 3, 10, 100, 1000 random variables and see if this works for any random variables you have chosen!

#4: SUM/DIFF OF GAUSSIANS, RATIO OF GAUSSIANS

- Considering two random Gaussian variables, $\boldsymbol{X}$ and $\boldsymbol{Y}$, where $\boldsymbol{X}$ has $\boldsymbol{\mu=25}$ and $\boldsymbol{\sigma=4}$, $\boldsymbol{Y}$ has $\boldsymbol{\mu=15}$ and $\boldsymbol{\sigma=3}$.
- Verify the sum of these two Gaussians ($\boldsymbol{X+Y}$) results a Gaussian with $\boldsymbol{\mu=40}$ and $\boldsymbol{\sigma=5}$. You can just compare your resulting distribution with the Gaussian PDF directly.
- Verify the difference of these two Gaussians ($\boldsymbol{X-Y}$) also results another Gaussian with $\boldsymbol{\mu=10}$ and $\boldsymbol{\sigma=5}$.
- However the ratio $\boldsymbol{X/Y}$ is not a Cauchy/Breit-Wigner unless you can set the mean of $\boldsymbol{Y}$ to be zero. Verify this point (*you can just produce the $\boldsymbol{X/Y}$ distributions, and see how they look like!*)

#5: DRAW A 2D CONTOUR

- Suppose you have a likelihood function already parameterized (or coded) as following:

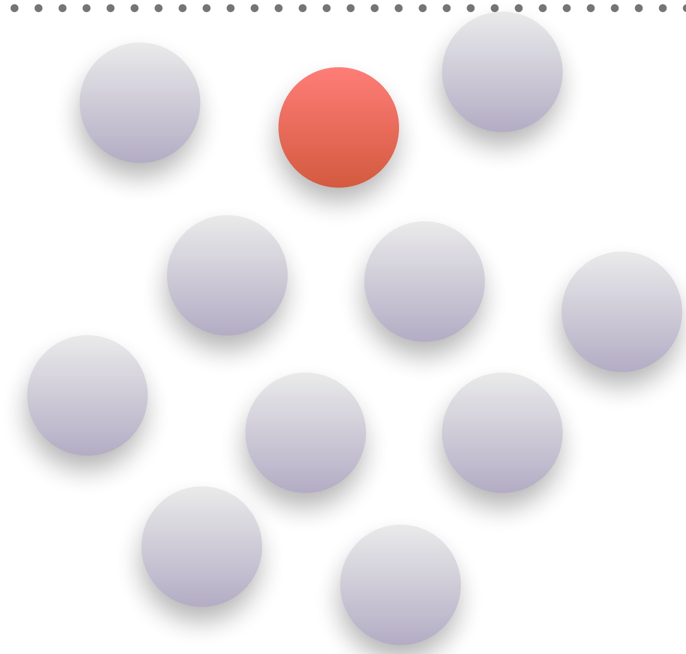
exercise_02.cc

```
double fcn(double x, double y) {
    const double data[50] = {
       0.9997,0.1629,0.2826,0.9472,0.2317,0.4850,0.9575,0.7443,0.5400,0.7400,
       0.7599,0.6586,0.3156,0.8044,0.5197,0.1686,0.4755,0.3923,0.2217,0.2132,
       0.0303,0.3335,0.1941,0.9437,0.5799,0.8983,0.6656,0.4986,0.5606,0.1823,
       0.2965,0.1174,0.0629,0.6481,0.7254,0.6371,0.7139,0.0996,0.6993,0.1078,
       0.1292,0.5024,0.2078,0.2889,0.0832,0.1281,0.5474,0.0823,0.2921,0.8916};

    double f = 0.;
    for (int i=0; i<50; i++) {
        double p = x*y*pow(1.-data[i],2)+
             (1.-x)*y*2.*data[i]*(1.-data[i])+
             (1.-y)*2.*pow(data[i],2);
        f += log(p);
    }
    return -2.*f;
}
```

- Please draw the corresponding 2D iso-probability contour in the range of **$0 \leq x \leq 1$ and $0 \leq y \leq 1$**.

#6: BUILD UP AN EXPONENTIAL



- Consider you have a bag of 100 marbles: **1 in red and 99 in white**.
- What you do is continuously picking up one marble out of the bag, check the color, and put it back.
- Count how many white marbles you get between two events of success: *pick up a red marble*.
- Plot the distribution of white marble counts, verify if your result follows an exponential distribution.

#7: CHEBYSHEV POLYNOMIALS

- As explained in the main lecture, the Chebyshev polynomials can be constructed with the following **recurrence relation**:

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$$

- Implement a simple function which takes two input, an integer n and a float point number x. It can calculate the polynomials of variable x up to order n.

- In fact, if $|x| \leq 1$, the polynomials can be expressed with the trigonometric definition:

$$T_n(x) = \cos[n \arccos(x)]$$

  Implement another function using the definition above and prove these two functions are identical when you scan over the variable x.

#8: SMEARED DISTRIBUTION

- Let's produce a "smeared" exponential distribution with the following steps:
  - Generate a random variable *x* according to the typical exponential distribution:
    $$f(x) \propto \exp(-x/1.6)$$
  - Generate another random variable *y* according to a Gaussian distribution:
    $$f(y) \propto \exp\left[-0.5\left(\frac{y-x}{0.2}\right)^2\right]$$
  - Repeat the steps above.
- See your **resulting distribution of *y*** just looks like the distribution given by the example #06 code in the main lecture?