

STATISTICAL ANALYSIS IN EXPERIMENTAL PARTICLE PHYSICS

.....
Kai-Feng Chen
National Taiwan University

EXERCISE 1

#1: BINOMIAL DISTRIBUTION FROM PRINCIPLE

- Let's generate a binomial distribution based on the “principle” of binomial itself, e.g. accept event with a fixed probability $p=10\%$ for a given $N=100$ trials:

```
{
    TRandom3 rnd;

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

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

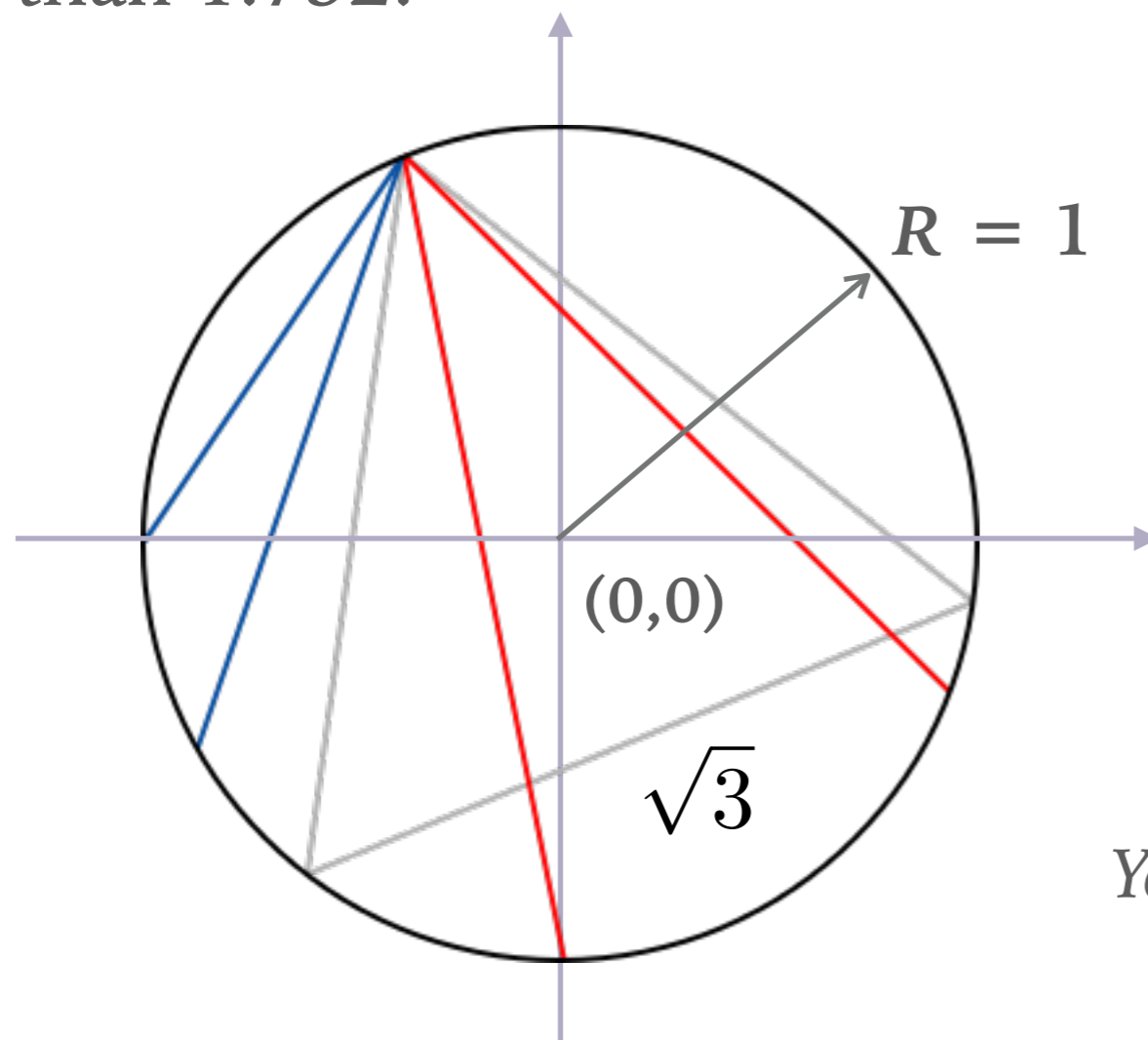
- The code above will give one single “count” which follows the binomial distribution. Please repeat the generation for 10,000 times and prove the variance follows $Np(1-p) = 9$.

#2: SOME OF TWO BINOMIAL VARIABLE

- ▶ Based on the previous exercise, generate two random binomial variables:
 - $N = 100, p = 10\%$
 - $N' = 200, p = 25\%$
- ▶ Verify the variance of the sum of the two random binomial variables is smaller than an averaged binomial distribution:
 - $N = 300, p = 20\%$

#3: BERTRAND'S PARADOX

- Consider a circle, centered at $(0,0)$ with radius 1. Randomly choose two points on the circumference of the circle and calculate the distance between them. Draw the distribution for the distance of the chords, and verify $1/3$ of the chord is longer than 1.732.



What if the you choose random two points within the circle and produce a chord? What would be the fraction in this case?

You can try the other two proposed methods.

#4: FREQUENTIST VS BAYESIAN

- Find **10 different probabilities in daily life** and see if they can be defined either by **frequentist** or by **Bayesian**.
- Just like the example given in the main lecture, the “raining probability” can be only defined by Bayesian.



Thomas Bayes

#5: CONDITIONAL PROBABILITY

- Consider someone rolls **two fair dice**, and we must predict the outcome (the sum of the two upward faces), which should be within $[2, 12]$.
- Let D_1 and D_2 are the value rolled on die 1 and die 2, respectively.
- Suppose event A is $D_1 + D_2 > 8$, and event B is $D_1 = D_2$, show this example fulfill the condition below:

$$P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$$

#6: BAYERS THEOREM FOR DISCRETE EVENTS

- The entire output of a factory is produced on three machines. The three machines account for different amounts of the factory output, namely 25%, 35%, and 40%.
- The fraction of defective items produced is this: for the first machine, 4%; for the second machine, 2%; for the third machine, 1%.
- If an item is chosen at random from the total output and is found to be defective, what is the probability that it was produced by the third machine?

#7: PRACTICE WITH PDF

- Consider a physics process which produce a probability distribution proportional to $1+t^2+\exp(-t)$, while t is in the range of $[0,1]$.
- Derive the **probability density function** for this physics process and produce a plot of it.
- Derive the **Cumulative distribution** for this PDF and plot it.

#8: BAYERS THEOREM FOR CONTINUOUS VARIABLE

- Consider a particle experiment which can measure the mass of an unknown new particle. Suppose the true mass of the particle is denoted by M , which is known to be within the range of 1-3 GeV with uniform probability density.
- The detector resolution has been estimated to be 0.1 GeV, and can be perfectly modeled by a Gaussian distribution.
- Suppose the measured mass is 1.5 GeV, plot **the posterior probability density distribution for a given true mass M , ie. $P(M|1.5 \text{ GeV})$** .
- If the prior density is not uniform but also a Gaussian of mean 2 GeV and width 0.5 GeV (e.g. constrained by other studies), what is the posterior density distribution in this case?