

# Probability Is All You Need

---

- History
- Descriptive Statistics
- Definition of Probability
- Probability and Distribution, Expectation, Variance
- Probability and Central Limit Theorem
- Probability and Hypothesis Test
- Probability and Bayes
- Probability and Stochastic Process

## History

---

The birth of statistics

- Math: Pascal and Fermat's researches on gambling. Since 1654.
- Physics: Einstein, Perrin, Langevin, Ito's researches on Brownian Motion. Since [1905](#).
- Biology: Galton, Pearson, Fisher's researches on Eugenics. Since [1883](#).
- NLP: Markov's researches on Pushkin's poem *Eugene Onegin*. Since [1913](#).

In the math part, Bernoulli, De Moivre, Laplace, Poisson also made great contributions. In 1900, Hilbert's sixth question is about the axiomatization of probability and physics. In 1933, Kolmogorov did that.

But the true birth of statistics relied on the latter 3 parts, especially biology. **From the birth, statistics has deep connections with biology, physics and NLP.**

## Descriptive Statistics

---

In primary school, junior high school and senior high school, we learnt a lot about the so-called "descriptive statistics".

For example

- boy-girl-ratio: [France](#). [China](#).
  - In 1814, Laplace wrote in "[A Philosophical Essay on Probabilities](#)" that the ratio of male to female infants in London, Berlin, and throughout France was 104.76:100, while in Paris it was 104.17:100.
- basketball, soccer: see [kaggle](#).
- weighted average: like grades weighted by points.

But we can't just do this kind of thing, we must go beyond, build a model. Sometimes people call this "statistical inference". Here are 3 important things we need to do.

- population: just like the population of a country. (Your Model)

- sample: both a number and a random variable. (Your Data)
- $f \rightarrow p$

## Definition of Probability

---

- Just like throwing a coin
- Measure

## Probability and Distribution, Expectation, Variance

---

- Distribution: Discrete and Continuous
- E, Var: See them as Operator!
  - $E(b) = E_x(E_b(b|x))$
  - $Var(b) = E_x(Var_b(b|x)) + Var_x(E_b(b|x))$
  - [MSE = Bias + Variance + Noise](#)
  - $E(\frac{1}{n-1} \sum_{i=1}^n (X_i - \text{mean of the data})^2) = Var(X)$

Example from RL: remember the slot machine in class?

I must emphasize that the distribution has more information than E, Var. That means, when you get a data in your project, first plot its distribution!

## Probability and Central Limit Theorem

---

[Wikipedia.](#)

In the simplest form,

$$\forall i, X_i \text{ i.i.d.} \Rightarrow \sum_{i=1}^n X_i \xrightarrow{d} N(nE(X_i), nVar(X_i))$$

And this is the true reason why Gauss choose  $e^{-x^2}$  instead of  $\cos(x)$ ,  $e^{-x^4}$ ,  $(\frac{1}{1+e^{-x}})'$ ,  $(\arctan(x))'$  and so on.

There is an example from neuroscience, which I will call the core of biological neural network.

Based on Central Limit Theorem, a neuron in a big neural network should have

$\frac{\sigma}{\mu} \rightarrow 0$ , as  $n \rightarrow \infty$ , but this contradicts with the experiment result, which show finite  $\frac{\sigma}{\mu}$ .

There are 2 solutions

- balanced exhibition and inhibition
- correlated inputs

# Probability and Hypothesis Test

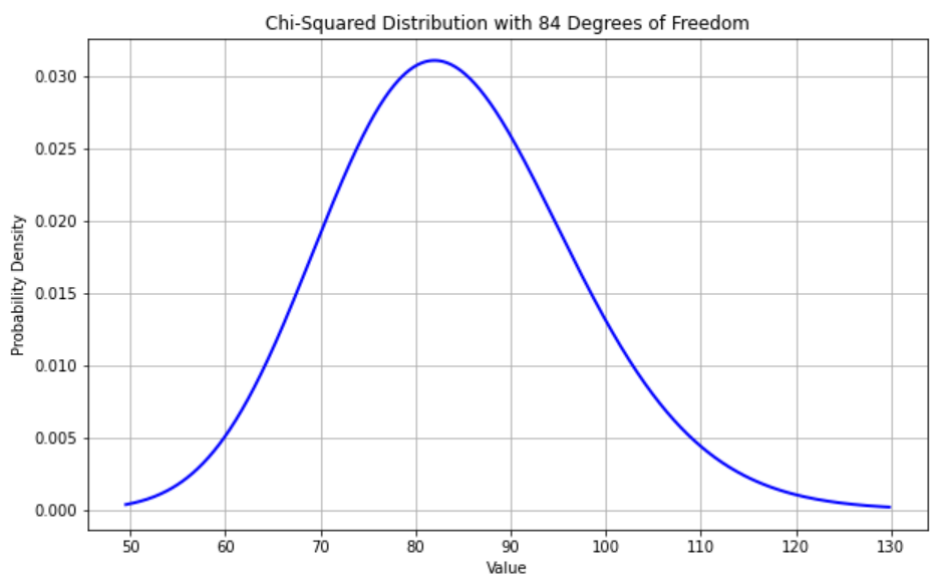
## [Lady Tasting Tea](#)

Other hypothesis test:  $\chi^2$ , t, F, permutation test, rank, KS, ...

Further Reading

- [Do you believe that certain zodiac signs are more compatible with each other?](#)
- [The Misuse of p-value in the Senior High School](#)
- [p-hacking: What it is and how to avoid it!](#)
- Example from Neuroscience: [Sen Song's work](#)

## Fisher vs Mendel



## Common Hypothesis Tests

表 6.2.1: 一样本正态总体 $N(\mu, \sigma^2)$ .

检验对象	检验统计量	分布	拒绝域†
$\mu$ ( $\sigma^2$ 已知)	$U = \sqrt{n}(\bar{X} - \mu_0)/\sigma$	$N(0, 1)$	$\begin{cases}  U  > u_{\alpha/2} \\ U > u_{\alpha} \\ U < -u_{\alpha} \end{cases}$
$\mu$ ( $\sigma^2$ 未知)	$T = \sqrt{n}(\bar{X} - \mu_0)/S$	$t_{n-1}$	$\begin{cases}  T  > t_{n-1}(\alpha/2) \\ T > t_{n-1}(\alpha) \\ T < -t_{n-1}(\alpha) \end{cases}$
$\sigma^2$ ( $\mu$ 已知)	$\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu)^2$	$\chi_n^2$	$\begin{cases} \chi^2 > \chi_n^2(\alpha/2) \text{ 或者 } \chi^2 < \chi_n^2(1 - \alpha/2) \\ \chi^2 > \chi_n^2(\alpha) \\ \chi^2 < \chi_n^2(1 - \alpha) \end{cases}$
$\sigma^2$ ( $\mu$ 未知)	$\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2$	$\chi_{n-1}^2$	$\begin{cases} \chi^2 > \chi_{n-1}^2(\alpha/2) \text{ 或者 } \chi^2 < \chi_{n-1}^2(1 - \alpha/2) \\ \chi^2 > \chi_{n-1}^2(\alpha) \\ \chi^2 < \chi_{n-1}^2(1 - \alpha) \end{cases}$

†有关均值的检验: 对立假设分别为 $\mu \neq \mu_0$ ,  $\mu > \mu_0$  和  $\mu < \mu_0$ . 有关方差的检验: 对立假设分别为 $\sigma^2 \neq \sigma_0^2$ ,  $\sigma^2 > \sigma_0^2$  和  $\sigma^2 < \sigma_0^2$ .

表 6.2.2: 两样本正态总体的假设检验

检验对象	检验统计量	分布	拒绝域 <sup>†</sup>
均值(方差已知)	$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$	$N(0, 1)$	$\begin{cases}  U  > u(\alpha/2) \\ U > u(\alpha) \\ U < -u(\alpha) \end{cases}$
均值(方差未知) <sup>‡</sup>	$T = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{\frac{1}{m} + \frac{1}{n}}}$	$t_{m+n-2}$	$\begin{cases}  T  > t_{m+n-2}(\alpha/2) \\ T > t_{m+n-2}(\alpha) \\ T < -t_{m+n-2}(\alpha) \end{cases}$
方差(均值已知)	$F = \frac{\sum_{i=1}^m (X_i - \mu_1)^2 / m}{\sum_{i=1}^n (X_i - \mu_2)^2 / n}$	$F_{m,n}$	$\begin{cases} F > F_{m,n}(\alpha/2) \text{ 或 } F < \frac{1}{F_{n,m}(\alpha/2)} \\ F > F_{m,n}(\alpha) \\ F < \frac{1}{F_{n,m}(\alpha)} \end{cases}$
方差(均值未知)	$F = \frac{S_1^2}{S_2^2}$	$F_{m-1,n-1}$	$\begin{cases} F > F_{m-1,n-1}(\alpha/2) \text{ 或 } F < \frac{1}{F_{n-1,m-1}(\alpha/2)} \\ F > F_{m-1,n-1}(\alpha) \\ F < \frac{1}{F_{n-1,m-1}(\alpha)} \end{cases}$

<sup>†</sup>有关均值的检验: 对立假设分别为  $\mu_1 \neq \mu_2$ ,  $\mu_1 > \mu_2$  和  $\mu_1 < \mu_2$ . 有关方差的检验: 对立假设分别为  $\sigma_1^2 \neq \sigma_2^2$ ,  $\sigma_1^2 > \sigma_2^2$  和  $\sigma_1^2 < \sigma_2^2$ .

<sup>‡</sup>假定方差相等

## Probability and Bayes

- Hypothesis Test
- [Three door problem](#)
- Three choice problem

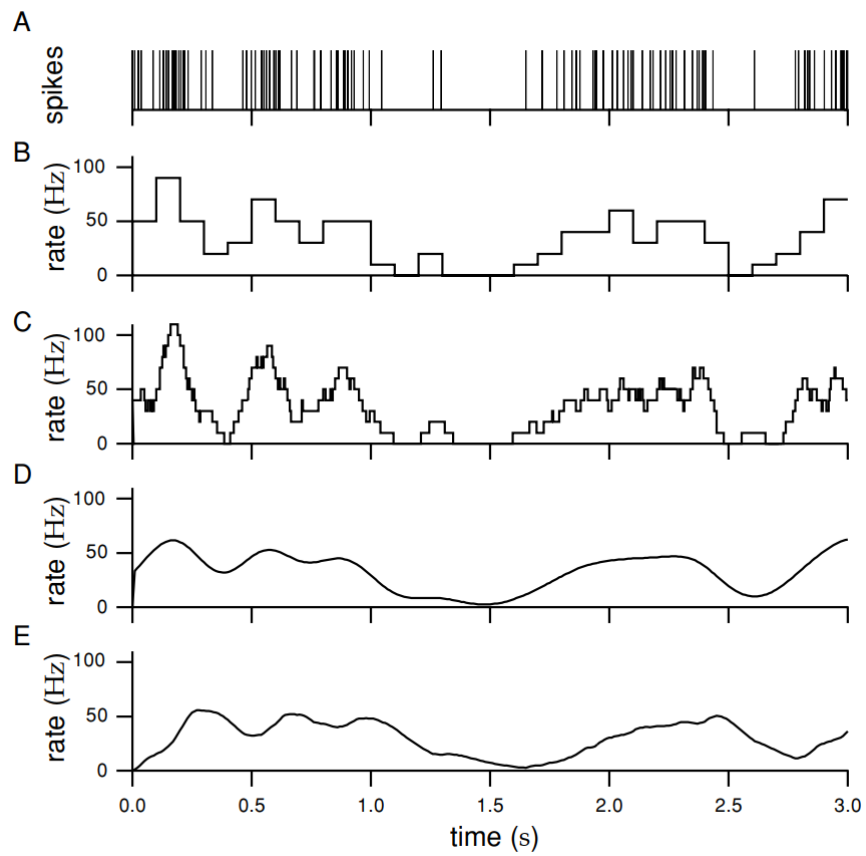
## Probability and Stochastic Process

### Basic

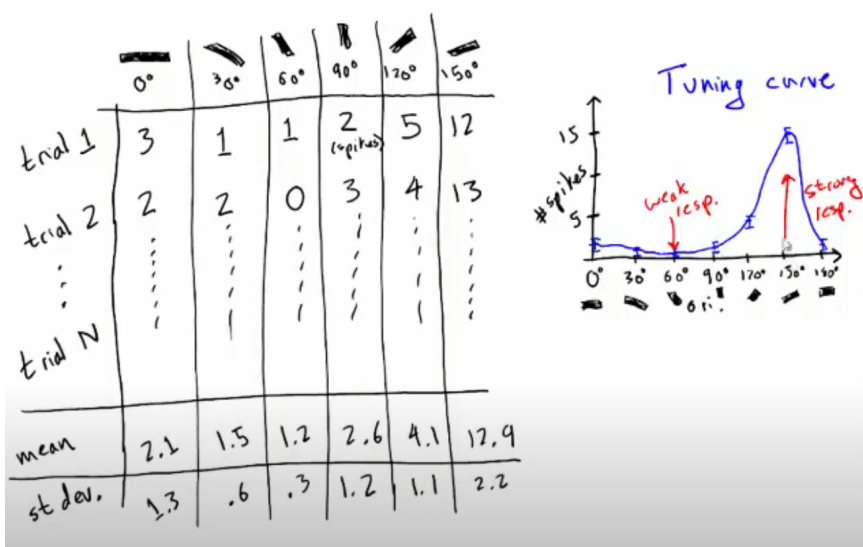
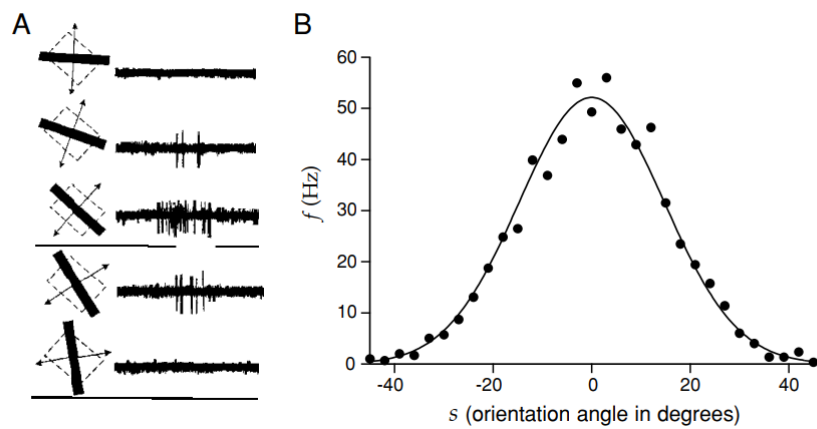
- Markov Process
- Poisson Process
- Steady Process

## Stochastic Process and Neuroscience

### Count

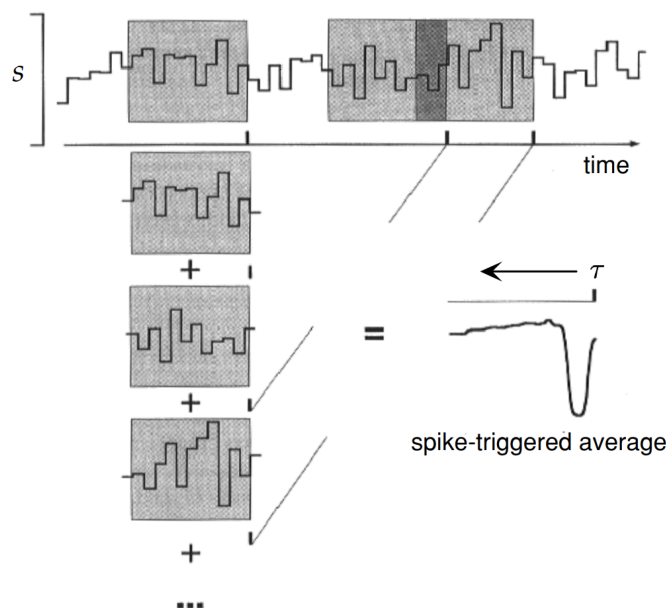


## Error-Bar



(figure from the Internet)

## The-stimulus-before-spikes



## Homework

1. Search the boy-girl-ratio of Japan, China, India, Turkey and Israel, write down what you see.
2. prove
  - $E(b) = E_x(E_b(b|x))$
  - $Var(b) = E_x(Var_b(b|x)) + Var_x(E_b(b|x))$
  - $MSE = Bias + Variance + Noise$
  - $E(\frac{1}{n-1} \sum_{i=1}^n (X_i - \text{mean of the data})^2) = Var(X)$
3. In the lady-testing-tea-problem, calculate  $P(\text{more than 8 cups are right} | \text{no distinguishing ability})$
4. In the 3-door-problem, calculate  $P(H_A|data)$  and  $P(H_B|data)$
5. In the 3-choice-problem, calculate  $P(H_A|data)$  and  $P(H_B|data)$
6. Suppose that you are Andrey Markov, who recently read a long poem of Alexander Pushkin, how to test if the appearance of the vowel and the consonant has Markov property or not.
7. Suppose that you take the video of a moving object and want to test if the trajectory of it has Markov property or not, how to figure that out? This [reference](#) maybe helpful.
8. When testing if [the data of all FIFA world cup](#) follows the Poisson process, which is better, concatenate 2 games or not? Write down your reasons.

Reward:

People who solve 3 problems of 8 can get bonus(2-3) in the final score.

People who solve problem 7 can get bonus(2-3) in the final score.