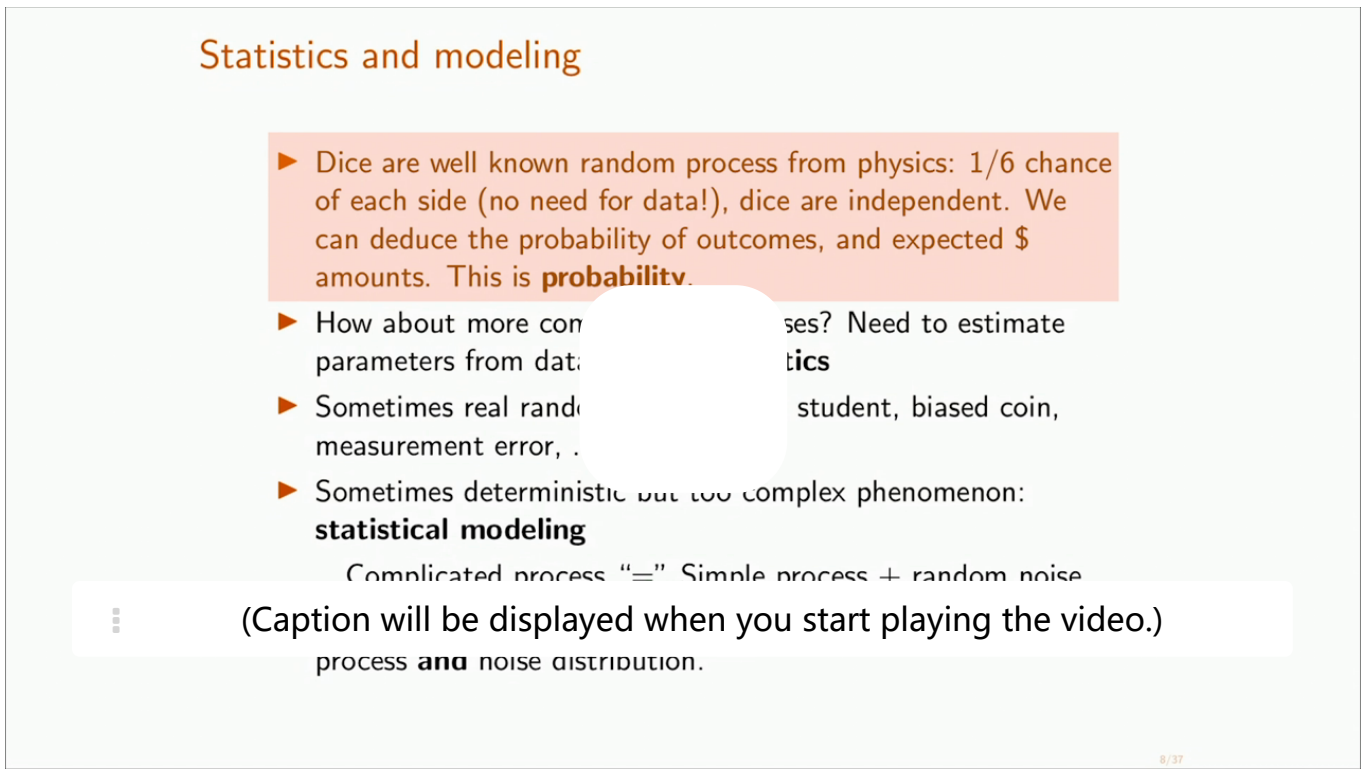
6. Statistics and modelling
Statistical modelling and the central dogma of statistics and probability



(Caption will be displayed when you start playing the video.)

Start of transcript. Skip to the end.

OK, so now that we've--

so if we understand what probability is,

what probability is, I give you a particular description

of a random process and I ask you things about the outcome,

maybe an expectation, maybe a probability.

The question is, where does statistics fits in all of this,

right?

视频
下载视频文件

字幕
下载 SubRip (.srt) file
下载 Text (.txt) file

Probability or statistics 1

2/2 points (graded)
Determine whether each of the problems below is a probabilistic or a statistical problem. (You are **not** asked to solve them.)

1. Assume we have a population consisting of two subpopulations, A and B. A particular drug has a different chance of treatment success depending on the subpopulation, namely 70% for group A and 50% for group B.

   Assume that subpopulation A is 10% of the entire subpopulation and subgroup B is 90%. What is the chance of a successful treatment if we pick a random person from the entire population?

   This is a

   - statistical problem
   - probabilistic problem ✔

2. Now, consider the scenario where we do not know the true composition of the population, which may be different from the previous setup. Among 1000 randomly chosen patients, we observe that the treatment was successful in 700 of them. What is a good estimate of the composition of the population?

   This is a

   - statistical problem ✔

probabilistic problem

Solution:

1. The first one is a probabilistic problem, because we are given all relevant parameters and are trying to compute corresponding derived probabilities. In particular, if we denote the population of a randomly selected person by $X \in \{A, B\}$ and the treatment outcome of drug D by $Y \in \{\textbf{sucess}, \textbf{failure}\}$, we are given $\mathbf{P}(X = A), \mathbf{P}(X = B), \mathbf{P}(Y = \textbf{success}|X = A)$, and $\mathbf{P}(Y = \textbf{success}|X = B)$ and are asked to compute $\mathbf{P}(Y = \textbf{success})$.

2. The second one is a statistical problem, because we are trying to estimate an underlying probabilistic parameter from data. More explicitly, we have 1000 i.i.d. draws from the Bernoulli random variable Y, 700 of which correspond to $Y = \textbf{success}$, and are now being asked to draw conclusions about $\mathbf{P}(Y = \textbf{success})$ and from there about $\mathbf{P}(X = A)$ and $\mathbf{P}(X = B)$.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Probability or statistics 2

2/2 points (graded)

John Arbuthnot wrote a paper in 1710 entitled ‘An Argument for Divine Providence', where he studied, based on the Christening records in London for 1629-1710, the chances that a randomly chosen baby born is a girl or a boy. Is this a statistical problem, or a probabilistic problem?

A statistical problem. ✔

A probabilistic problem.

Next, you read Arbuthnot's paper, and went to a gyneacology facility, in which there are 10 babies whom are expected to be born on the day you arrived, and you are interested in, what are the odds that 6 of those will be a boy, and the remaining will be a girl. Is this a statistical problem, or a probabilistic problem?

A statistical problem.

A probabilistic problem. ✔

Solution:

The first one is a statistical problem, and the second one is a probabilistic problem. To see this, suppose that each newborn baby is a boy with probability p, and a girl with probability $1 - p$, independent of each other. The data that Arbuthnot analyzed simply corresponds to realizations of this Bernoulli variable; and from that knowledge, we simply want to extract the underlying parameter, p. This is an example of a statistical problem.
You take Arbuthnot's finding, and assume that this is the 'true' probability for the aforementioned birth process; and want to compute a probability, which is simply,

$$\binom{10}{4} p_A^6 (1 - p_A)^4,$$

where, p_A is the value that Arbuthnot has reported (namely, you are computing the probability that a certain Binomial random variable is equal to 6.

提交 你已经尝试了2次（总共可以尝试3次）

Answers are displayed within the problem

Probability or statistics 3

2/2 points (graded)
A doctor realizes that there is an allergy medicine which is effective in treating seasonal allergies with probability at least 90%. From here, he claims:

- Out of 100 patients admitted to clinic with seasonal allergies, this drug will cure 90 patients, on average.
- At least 70 patients will be cured, with 99.99% chances.

Does he rely on statistics, or probability?

- ○ Statistics
- ◉ Probability ✔

Now, a newly-hired scientist at a pharmacology company deduces that, “I am 95% confident that if we repeat this experiment, then the drug will be effective on between 85% and 95% patients." Does he rely on statistics, or probability?

- ◉ Statistics ✔
- ○ Probability

Solution:

- The doctor relies on probability. The point of discussion is about averages and that the odds that at least 70 patients will be cured.
- The scientist relies on statistics. This is hinted at, because he discusses confidence regions.

知不知道truth是什么。如果知道就是probability，不知道就是statistics，因为第二个里面有推断置信区间，所以是statistics。第一个里面的猜测是概率，是truth，没有涉及到置信度。

提交 你已经尝试了2次（总共可以尝试3次）

ℹ Answers are displayed within the problem

讨论

显示讨论

主题： Unit 1 Introduction to statistics:Lecture 1: What is statistics / 6. Statistics and modelling

认证证书是什么?