Bus 273:
Statistical Analysis For Business

Harald Schmidbauer

İSTANBUL BİLGİ ÜNİVERSİTESİ

About These Slides

- The present slides are not self-contained; they need to be explained and discussed. This will be done in the lectures.

- Even though being a "work in progress" and subject to revision, the slides constitute copyrighted material.
  If you want to reproduce or copy anything from the slides, please ask:

Harald Schmidbauer	**harald** at **hs-stat** dot **com**
Angi Rösch	**angi.r** at **t-online** dot **de**

- The slides were produced using LaTeX and R (the R project; www.R-project.org) on a GNU/Linux system.

- R files used for this course are available upon request.

PART II:

Probability and Stochastic Models

Chapter 5:

Stochastic Models Behind the Observations

5.1 Data and Stochastic Models

Randomness. — There is “randomness” in observed data:

- Drawing another sample will lead to a different selection.
- Many future events cannot be predicted with certainty.

Questions in this context:

- How were the data at hand produced?
- What is behind the data?

Inductive statistics. . .

- . . . is an effort to answer these questions.
- . . . needs probability.

5.1 Data and Stochastic Models

Inductive statistics.

- The paradigm of inductive statistics is:

> **Regard the observations as the outcome of a random experiment, that is, as being produced by a stochastic model.**

- Stochastic model: a mathematical model on the basis of probability.

- The object of research is then the stochastic model, rather than the observations!

5.1 Data and Stochastic Models

Example: Throwing a die once.

This is a random experiment.

- **Before** the die is thrown:

$$X = \text{number which appears}$$

  is a random variable, and $P(X = i) = 1/6$.

- **After** the die is thrown, a probability statement is no longer meaningful! But we can still see the result as being produced by a chance setup.

5.1 Data and Stochastic Models

Example: A public opinion poll.

- Q: "Do you think New Orleans should be rebuilt?"
- Define: $p =$ share of American adults who say "YES"
- For each randomly selected person, we have a random variable:

$$X = \begin{cases} 1 & \text{if the person says "YES"} \\ 0 & \text{if the person says "NO"} \end{cases}$$

- Then, $P(X = 1) = p$ — the share p can be seen as a probability!
- How can we learn about p?

5.1 Data and Stochastic Models

Example: A public opinion poll.

- Q: "Do you think New Orleans should be rebuilt?"

- There is empirical evidence:

  384 out of 609 randomly selected American adults said "YES". (According to CNN, 2005-09-08.)

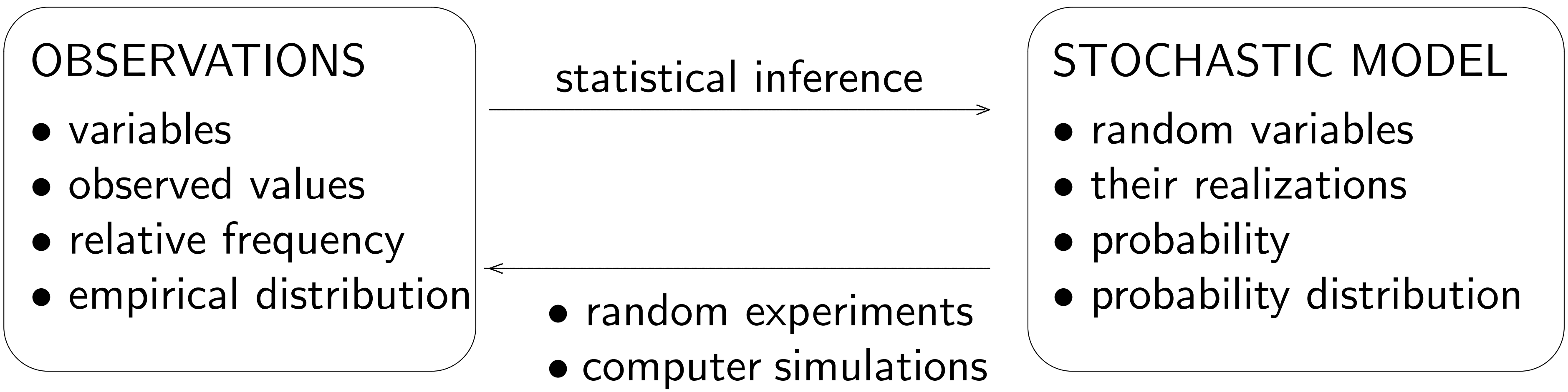
- We can then estimate p:

$$\hat{p} = \frac{384}{609} = 63\%.$$

- Observe that p and $\hat{p}$ are different objects!

5.1 Data and Stochastic Models

Observations and stochastic models:
analogies and their relation.



5.1 Data and Stochastic Models

Again: Why do we need stochastic models?

- The entire population is always identified with a stochastic model.
- A random sample allows us to learn about the stochastic model.
- The estimated model represents what we know about the population.

5.2 Probability Calculations

How to obtain a stochastic model. . .

Basic questions:

- Which outcomes are possible?
  Which values can the random variable take on?

- Which probabilities can be assigned to sets of possible outcomes?

5.2 Probability Calculations

Events and probabilities.

- An event is a set of possible outcomes of a random experiment.
  An event has a certain propensity (tendency) to occur.

- This propensity is expressed by a number between 0 and 1, the probability of the event.

- Probabilities must not be assigned totally arbitrarily to events! Certain rules must not be violated.

5.2 Probability Calculations

Kolmogorov's axioms.

i) Every event A has a probability $P(A) \geq 0$.

ii) $P(\Omega) = 1$, where Ω is the set of *all* possible outcomes.

iii) $P(A \cup B) = P(A) + P(B)$ for disjoint events A and B.

From this, further rules can be derived, for example:

$$
\begin{aligned}
A \subset B \quad &\Rightarrow \quad P(A) \leq P(B) \\
P(\bar{A}) \quad &= \quad 1 - P(A) \\
P(B) \quad &= \quad P(A \cap B) + P(\bar{A} \cap B) \\
P(A \cup B) \quad &= \quad P(A) + P(B) - P(A \cap B)
\end{aligned}
$$

5.2 Probability Calculations

How can we find the probability of an event?

Two important special cases. . .

- **Laplace experiments:** finite number of outcomes, each with equal probability. Then:

$$P(A) = \frac{\#\ \text{outcomes favourable for } A}{\#\ \text{possible outcomes}}$$

- **Urn models:** We are now going to look at a simple urn model. . .
  Modifications of the model will carry us very far in Chapters 6, 7 and 8.

5.3 Urn Models

Urn models: an example.

An urn contains 10 balls:

8 balls are red, 2 balls are blue.

Now suppose two balls are randomly drawn from the urn.
We want to find:

$$P(\underbrace{\text{the } 1^{\text{st}} \text{ ball is red}}_{\text{event } A} \text{ and } \underbrace{\text{the } 2^{\text{nd}} \text{ ball is blue}}_{\text{event } B})$$

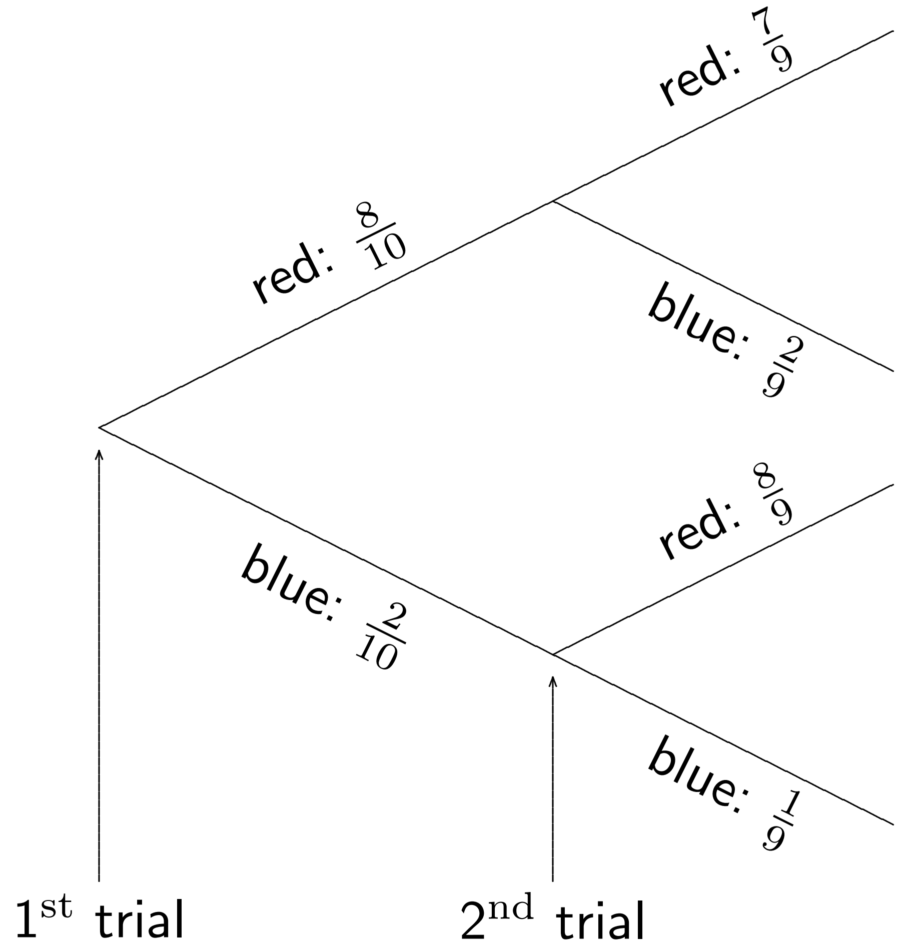
5.3 Urn Models

An urn model.

Drawing without replacement:

$$P(A \cap B) = P(A) \cdot P(B|A) = \frac{8}{10} \cdot \frac{2}{9}$$

The events A and B are dependent.



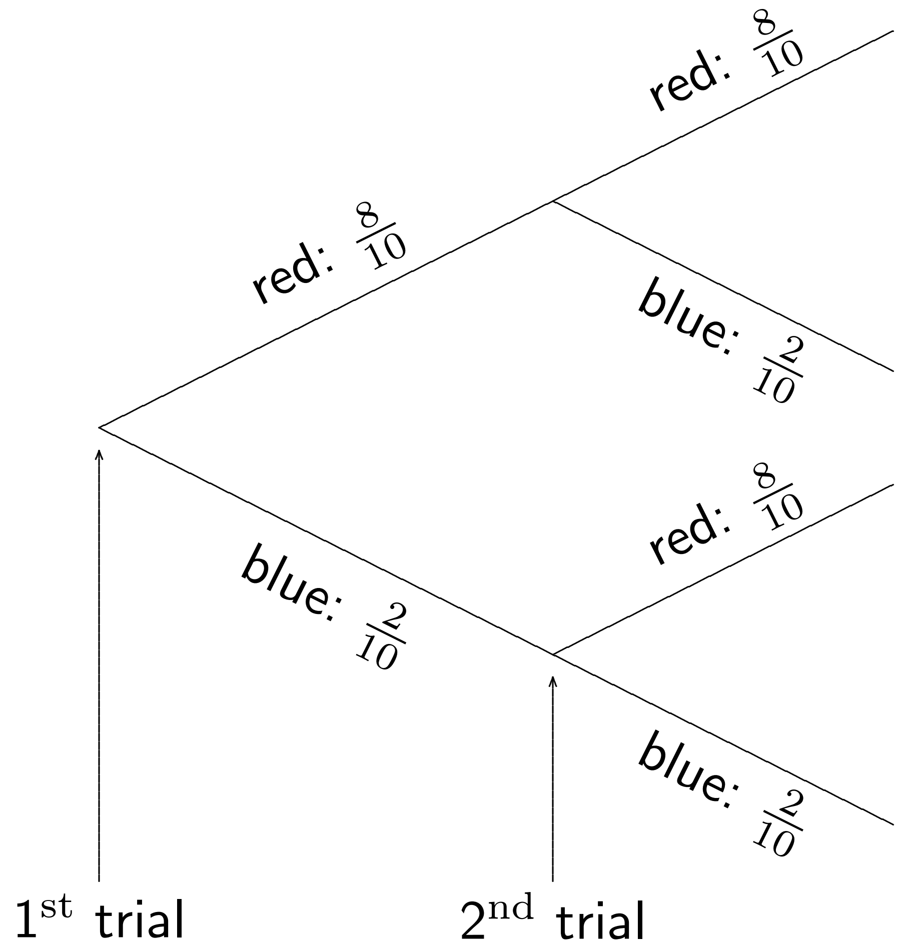
5.3 Urn Models

An urn model.

Drawing with replacement:

$$
\begin{aligned}
P(A \cap B) &= P(A) \cdot P(B|A) \\
&= P(A) \cdot P(B) \\
&= \frac{8}{10} \cdot \frac{2}{10}
\end{aligned}
$$

The events A and B are independent.



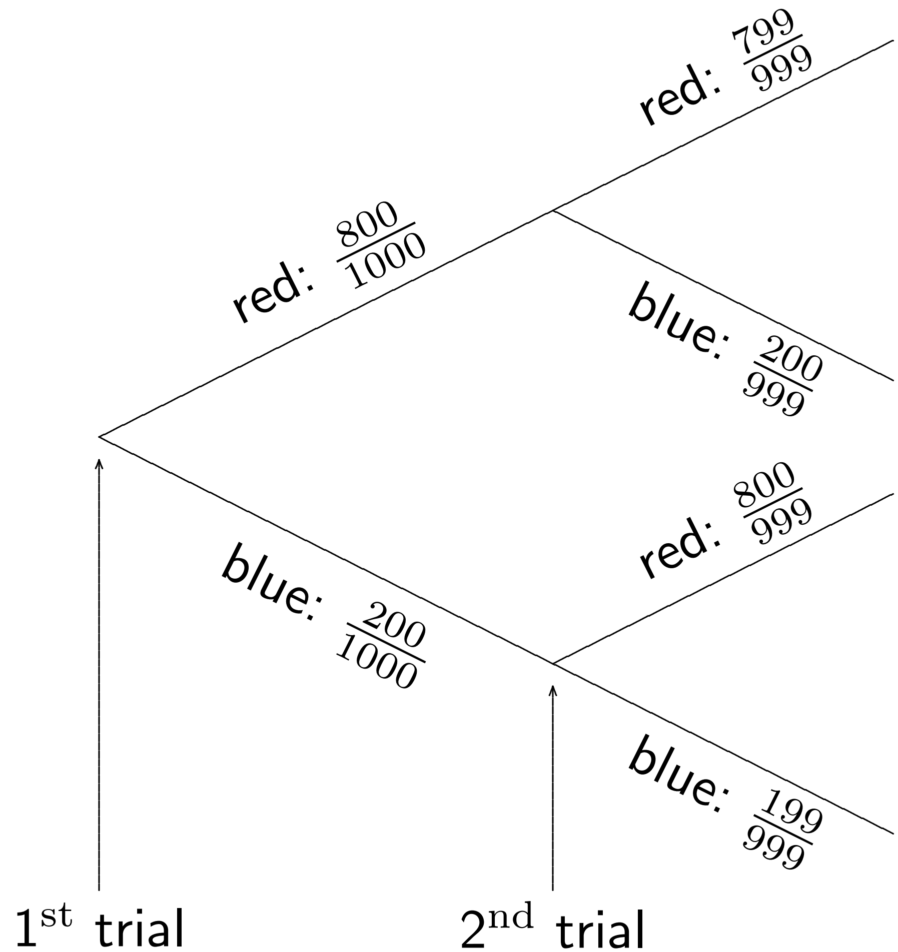
5.3 Urn Models

An urn model.

Drawing without replacement, but large number of balls:

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B|A) \\ &\approx P(A) \cdot P(B) \\ &= \frac{8}{10} \cdot \frac{2}{10} \end{aligned}$$

The events A and B are (almost) independent.



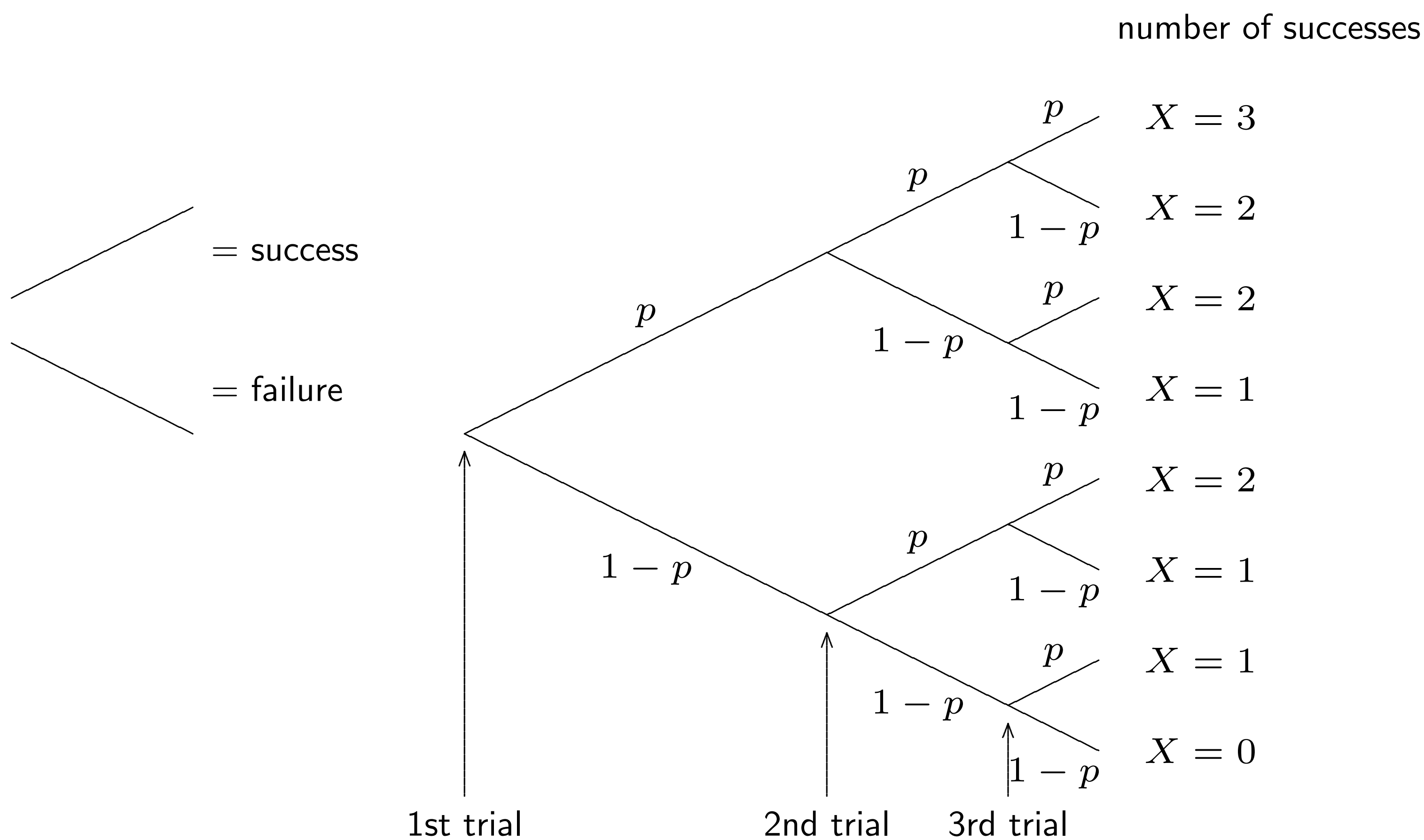
5.3 Urn Models

Consequences for sampling (with a binary variable).

- Sampling with replacement is like complete random sampling.
- Sampling without replacement reduces the uncertainty if the population is small.
- Sampling without replacement from a large population: We may act as if the sample was obtained by complete random sampling.

5.3 Urn Models

A setup with three independent trials:



5.3 Urn Models

Bernoulli trials; the binomial distribution. . .

- Combining branches for each event $\{X = i\}$:

$$\begin{aligned} P(X=0) &= (1-p)^3 \\ P(X=1) &= 3p(1-p)^2 \\ P(X=2) &= 3p^2(1-p) \\ P(X=3) &= p^3 \end{aligned}$$

5.3 Urn Models

Bernoulli trials; the binomial distribution. . .

- This can be written more elegantly:

$$P(X = i) = \binom{3}{i} p^i (1-p)^{3-i}, \quad i = 0, \ldots, 3.$$

- This is a special case of the binomial distribution:

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, \ldots, n.$$

5.4 Conditional Probability

Definition of conditional probability.

- Let A be an event such that $P(A) > 0$. Then,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

is called the conditional probability of B on condition A.

- With this:

$$\begin{aligned} P(A \cap B) &= P(B|A) \cdot P(A) \\ P(B) &= P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A}) \end{aligned}$$

5.4 Conditional Probability

Examples.

- Rolling a die once. Let $A = \{4, 5, 6\}, B = \{4, 6\}$.
  Then, $P(B|A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{1/3}{1/2} = \dfrac{2}{3}$.

- What is $P(B)$ in the urn example (drawing without replacement, page 17)?

$$P(B|A) = \frac{2}{9}, \quad P(B|\bar{A}) = \frac{1}{9};$$

therefore,

$$P(B) = \frac{2}{9} \cdot \frac{8}{10} + \frac{1}{9} \cdot \frac{2}{10} = \frac{2}{10}.$$

5.5 The Bayes Theorem

The Bayes theorem and its purpose.

- The Bayes theorem states that

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}$$

- $P(A)$: prior probability of A
  (before we know if B occurs or not)

- $P(A|B)$: posterior probability of A
  (after we know B has occurred)

- The prior $P(A)$ is augmented to the posterior $P(A|B)$, using the empirical evidence that B has occurred.

5.5 The Bayes Theorem

Example 1: Sampling from boxes.

- Two boxes with 5 balls each:
  - Box I: 2 white, 3 black.
  - Box II: 4 white, 1 black.
- A ball is drawn from one of the boxes. This ball is black. It is not known from which box the ball was drawn.
- Probability that the drawn ball comes from Box I?

5.5 The Bayes Theorem

Example 1: Sampling from boxes.

- Define events:

$$\begin{aligned} A_1 &= \text{Box I was chosen,} \\ A_2 &= \text{Box II was chosen,} \\ B &= \text{the drawn ball is black.} \end{aligned}$$

- Problem: Find $P(A_1|B)$!

- Solution: Apply

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2)}$$

5.5 The Bayes Theorem

Example 1: Sampling from boxes.

- To apply

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2)},$$

  we need the prior probabilities: $P(A_1)$, $P(A_2)$.

- We set: $P(A_1) = P(A_2) = \frac{1}{2}$ (reflecting ignorance!)

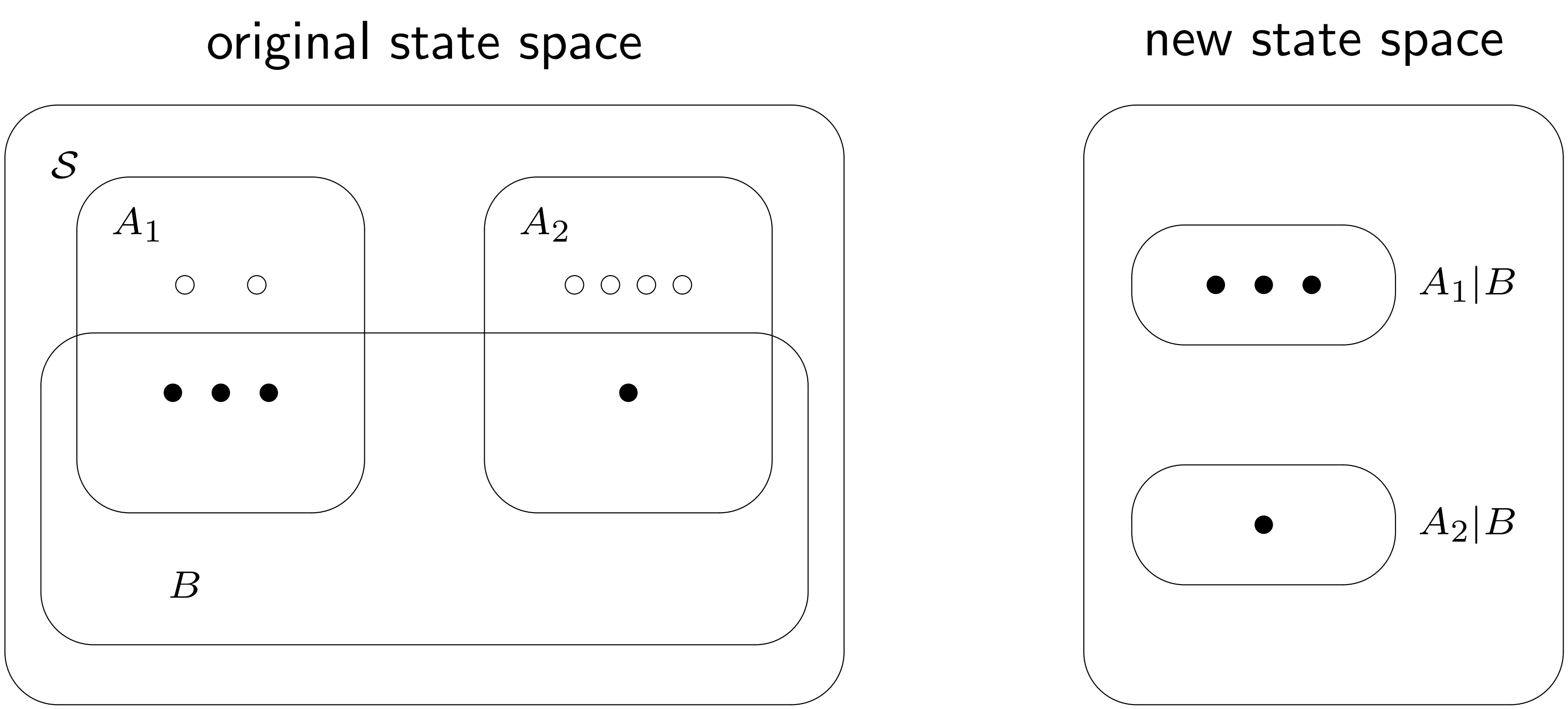
- Then,

$$P(A_1|B) = \frac{\frac{3}{5} \cdot 0.5}{\frac{3}{5} \cdot 0.5 + \frac{1}{5} \cdot 0.5} = 0.75.$$

5.5 The Bayes Theorem

Example 1: Sampling from boxes.

Illustration of Bayes's theorem:



5.5 The Bayes Theorem

Example 2: HIV tests.

- No HIV test is 100% accurate:
  It may produce a false positive or a false negative result.

- A person is tested. Define events:

  A: person is HIV infected;
  $\bar{A}$: person is not HIV infected;

  B: test result is positive;
  $\bar{B}$: test result is negative.

5.5 The Bayes Theorem

Example 2: HIV tests.

- Probabilities and their names in epidemiology:

$$\begin{aligned} \text{sensitivity} &= P(B|A), \\ \text{specificity} &= P(\bar{B}|\bar{A}), \\ \text{prevalence} &= P(A), \end{aligned} \qquad \begin{aligned} \text{PPV} &= P(A|B), \\ \text{NPV} &= P(\bar{A}|\bar{B}). \end{aligned}$$

- Suppose a person has tested positive. What is $P(A|B)$?

- Bayes:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}$$

5.5 The Bayes Theorem

Example 2: HIV tests.

- For a numerical example, let's assume:
  $P(B|A) = 0.98, \;\; P(\bar{B}|\bar{A}) = 0.99.$

- Scenario 1: Low HIV prevalence: $P(A) = 0.001$. Then:

$$P(A|B) = \frac{0.98 \cdot 0.001}{0.98 \cdot 0.001 + 0.01 \cdot 0.999} = 0.089.$$

  91% false positives! Most probably no infection in spite of positive test.

- Scenario 2: High HIV prevalence: $P(A) = 0.1$. Then:

$$P(A|B) = \frac{0.98 \cdot 0.1}{0.98 \cdot 0.1 + 0.01 \cdot 0.9} = 0.916.$$

  8.4% false positives. Less uncertainty, PPV is much higher.

5.5 The Bayes Theorem

Example 2: HIV tests.

- What we can learn from the test result depends on what we already know or assume to know.
- The Bayes theorem processes the information contained in the experiment (the test).
- A positive test result need not indicate infection.
- A negative test result need not indicate the absence of infection.