Statistical Methods and Data Analysis I
Lecture 5: Hypothesis Testing: Empirical Evidence

Oleg Goldshmidt
oleg.goldshmidt@post.idc.ac.il

Arison School of Business
Interdisciplinary Center (IDC)
Herzliya, Israel

March 17, 2019

Subjectivist Interpretation of Probability

Probability reflects a subjective "degree of belief".
What odds would you give and how much would you bet that event A will occur (or not)? Does empirical evidence affect your opinion?

Bayesian interpretation

1. Assign a *prior* (a.k.a. *a priori*) probability to hypothesis H.
2. Perform some measurements and gather some *evidence* E.
3. Update the probability of H to *posterior* (a.k.a. *a posteriori*) probability.

The formal methodology of such an update is called *Bayesian inference*.

Bayesian Inference

Bayesian Inference

- Let H be a *hypothesis* whose probability may be affected by *evidence* E.
  - Often there are competing hypotheses and we want to determine the most probable.

Bayesian Inference

- Let H be a *hypothesis* whose probability may be affected by *evidence* E.
  - Often there are competing hypotheses and we want to determine the most probable.
- The *prior probability* $P(H)$ is the estimate before E is observed.
  - E must not have been used to compute the prior probability $P(H)$.

Bayesian Inference

- Let H be a *hypothesis* whose probability may be affected by *evidence* E.
  - Often there are competing hypotheses and we want to determine the most probable.
- The *prior probability* $P(H)$ is the estimate before E is observed.
  - E must not have been used to compute the prior probability $P(H)$.
- The *posterior probability* $P(H|E)$ is the probability of H *after* E has been observed.
  - **This is what we want to know!**

Bayesian Inference

- Let H be a *hypothesis* whose probability may be affected by *evidence* E.
  - Often there are competing hypotheses and we want to determine the most probable.
- The *prior probability* $P(H)$ is the estimate before E is observed.
  - E must not have been used to compute the prior probability $P(H)$.
- The *posterior probability* $P(H|E)$ is the probability of H *after* E has been observed.
  - **This is what we want to know!**
- $P(E|H)$ is the *likelihood* of observing E under hypothesis H.
  - Indicates compatibility of E with H.
  - The *likelihood function* $P(E|H)$ is a function of *evidence* given the hypothesis. The *posterior probability* $P(H|E)$ is a function of the *hypothesis* given the evidence.

Bayesian Inference

- Let H be a *hypothesis* whose probability may be affected by *evidence* E.
  - Often there are competing hypotheses and we want to determine the most probable.
- The *prior probability* $P(H)$ is the estimate before E is observed.
  - E must not have been used to compute the prior probability $P(H)$.
- The *posterior probability* $P(H|E)$ is the probability of H *after* E has been observed.
  - **This is what we want to know!**
- $P(E|H)$ is the *likelihood* of observing E under hypothesis H.
  - Indicates compatibility of E with H.
  - The *likelihood function* $P(E|H)$ is a function of *evidence* given the hypothesis. The *posterior probability* $P(H|E)$ is a function of the *hypothesis* given the evidence.
- According to Bayes formula:

$$P(H|E) = \frac{P(E|H)}{P(E)} P(H)$$

where $P(E)$, the *marginal likelihood* of evidence, does not affect the *relative* probabilities of different hypotheses. $P(E|H)/P(E)$ is the *impact* of E on the probability of H.

Using Bayesian Inference (I)

Procurement scenario:

A vendor offers your company a component at a good price. However, it is known that in the past the vendor had a serious quality problem: out of the last 200 batches the defect rate in the first 128 was 44%. Then the vendor improved their QA and further 72 batches had only 15% defect rate. Upon receiving a batch of components you can afford to test one, and you need to decide whether to return the batch. A 15% defect rate in a batch is acceptable, but a 44% defect rate is not.

Using Bayesian Inference (I)

Procurement scenario:

A vendor offers your company a component at a good price. However, it is known that in the past the vendor had a serious quality problem: out of the last 200 batches the defect rate in the first 128 was 44%. Then the vendor improved their QA and further 72 batches had only 15% defect rate. Upon receiving a batch of components you can afford to test one, and you need to decide whether to return the batch. A 15% defect rate in a batch is acceptable, but a 44% defect rate is not.

- The *prior* probabilities of a bad/good batch:

$$P(B) = 128/200 = 0.64, \;\; P(G) = 72/200 = 0.36$$

.

Using Bayesian Inference (I)

Procurement scenario:

A vendor offers your company a component at a good price. However, it is known that in the past the vendor had a serious quality problem: out of the last 200 batches the defect rate in the first 128 was 44%. Then the vendor improved their QA and further 72 batches had only 15% defect rate. Upon receiving a batch of components you can afford to test one, and you need to decide whether to return the batch. A 15% defect rate in a batch is acceptable, but a 44% defect rate is not.

- The *prior* probabilities of a bad/good batch:

$$P(B) = 128/200 = 0.64, \;\; P(G) = 72/200 = 0.36$$

.

- The likelihoods of the sampled component being defective or OK:

$$P(D|B) = 0.44, \;\; P(D|G) = 0.15, \;\; P(OK|B) = 0.56, \;\; P(OK|G) = 0.85$$

Using Bayesian Inference (II)

Using Bayesian Inference (II)

- Overall probability of the sampled component being defective is

$$P(D) = P(D|B)P(B) + P(D|G)P(G) = 0.44 \times 0.64 + 0.15 \times 0.36 = 0.3356$$

Using Bayesian Inference (II)

- Overall probability of the sampled component being defective is

$$P(D) = P(D|B)P(B) + P(D|G)P(G) = 0.44 \times 0.64 + 0.15 \times 0.36 = 0.3356$$

- Overall probability of the sampled component being OK is

$$P(OK) = P(OK|B)P(B) + P(OK|G)P(G) = 0.56 \times 0.64 + 0.85 \times 0.36 = 0.6644$$

Using Bayesian Inference (II)

- Overall probability of the sampled component being defective is

$$P(D) = P(D|B)P(B) + P(D|G)P(G) = 0.44 \times 0.64 + 0.15 \times 0.36 = 0.3356$$

- Overall probability of the sampled component being OK is

$$P(OK) = P(OK|B)P(B) + P(OK|G)P(G) = 0.56 \times 0.64 + 0.85 \times 0.36 = 0.6644$$

- If the sampled component is defective, the *posterior* probability that the batch is bad is

$$P(B|D) = \frac{P(D|B)}{P(D)}P(B) = \frac{0.44}{0.3356} \times 0.64 \approx 0.84$$

Using Bayesian Inference (II)

- Overall probability of the sampled component being defective is

$$P(D) = P(D|B)P(B) + P(D|G)P(G) = 0.44 \times 0.64 + 0.15 \times 0.36 = 0.3356$$

- Overall probability of the sampled component being OK is

$$P(OK) = P(OK|B)P(B) + P(OK|G)P(G) = 0.56 \times 0.64 + 0.85 \times 0.36 = 0.6644$$

- If the sampled component is defective, the *posterior* probability that the batch is bad is

$$P(B|D) = \frac{P(D|B)}{P(D)}P(B) = \frac{0.44}{0.3356} \times 0.64 \approx 0.84$$

- If the sampled component is OK, the *posterior* probability that the batch is good is

$$P(G|OK) = \frac{P(OK|G)}{P(OK)}P(G) = \frac{0.85}{0.6644} \times 0.36 \approx 0.46$$

Using Bayesian Inference (III)

Using Bayesian Inference (III)

- *What if you can test another component from the same (large) batch?*

Using Bayesian Inference (III)

- *What if you can test another component from the same (large) batch?*
- Your *prior* and *marginal* probabilities (but not *likelihoods* — in a *large* batch) have changed:
  - If the first sampled component was defective:

$$P(B) = 0.84, \quad P(G) = 0.16$$

$$P(OK) = 0.56 \times 0.84 + 0.85 \times 0.16 \approx 0.6064, \quad P(D) = 0.44 \times 0.84 + 0.15 \times 0.16 = 0.3936$$

  - If the first sampled component was OK:

$$P(B) = 0.54, \quad P(G) = 0.46$$

$$P(OK) = 0.56 \times 0.54 + 0.85 \times 0.46 \approx 0.6934, \quad P(D) = 0.44 \times 0.84 + 0.15 \times 0.46 \approx 0.3066$$

IDC HERZLIYA

Using Bayesian Inference (III)

- *What if you can test another component from the same (large) batch?*
- Your *prior* and *marginal* probabilities (but not *likelihoods* — in a *large* batch) have changed:
  - If the first sampled component was defective:

$$P(B) = 0.84, \;\; P(G) = 0.16$$

$$P(OK) = 0.56 \times 0.84 + 0.85 \times 0.16 \approx 0.6064, \;\; P(D) = 0.44 \times 0.84 + 0.15 \times 0.16 = 0.3936$$

  - If the first sampled component was OK:

$$P(B) = 0.54, \;\; P(G) = 0.46$$

$$P(OK) = 0.56 \times 0.54 + 0.85 \times 0.46 \approx 0.6934, \;\; P(D) = 0.44 \times 0.84 + 0.15 \times 0.46 \approx 0.3066$$

- Both samples defective:

$$P(B|D) = \frac{P(D|B)}{P(D)}P(B) = \frac{0.44 \cdot 0.84}{0.3936} \approx 0.94$$

- Both samples OK:

$$P(G|OK) = \frac{P(OK|G)}{P(OK)}P(G) = \frac{0.85 \cdot 0.46}{0.6934} \approx 0.56$$

Probability Inversion

Note that along the way we performed what is known as *"probability inversion"*:

Probability Inversion

Note that along the way we performed what is known as *"probability inversion"*:

- We started from conditional probabilities of having a defective/OK component in a "bad" batch and in a "good" batch:

$$P(D|B), \;\; P(D|G), \;\; P(OK|B), \;\; P(OK|G)$$

- We ended up with computing conditional probabilities of having a "bad" batch or a "good" batch given that tested components were defective or OK:

$$P(B|D), \;\; P(G|D), \;\; P(B|OK), \;\; P(G|OK)$$

Example: Breast Cancer Screening

Disclaimer: numbers below are approximate and very difficult to estimate.

- About 1.1% of women in the US have breast cancer
  (Source: American Cancer Society.)
- Mammogram efficacy:
  - *Sensitivity* of mammograms is 87% (i.e., the *false negative* rate is 13%).
  - The *false positive* rate of a mammogram is between 7% and 12% — we will assume 10%.

  (Source: Susan G. Komen Foundation.)

Hypothesis	Positive test	Negative test
Cancer	0.87	0.13
No cancer	0.1	0.9

Inverting Cancer Probabilities: One Test

A patient undergoes breast cancer screening. There are two competing hypotheses:

1. The patient has cancer (C).
2. The patient does not have cancer (N).

Inverting Cancer Probabilities: One Test

A patient undergoes breast cancer screening. There are two competing hypotheses:

1. The patient has cancer (C).
2. The patient does not have cancer (N).

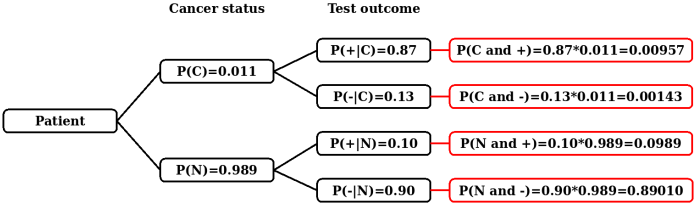
If a mammogram yields a positive result what is the probability that the patient has cancer?

Inverting Cancer Probabilities: One Test

A patient undergoes breast cancer screening. There are two competing hypotheses:

1. The patient has cancer (C).
2. The patient does not have cancer (N).

If a mammogram yields a positive result what is the probability that the patient has cancer?

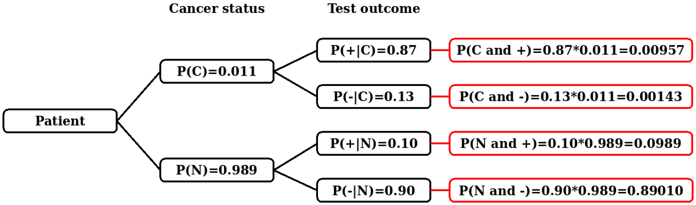


Inverting Cancer Probabilities: One Test

A patient undergoes breast cancer screening. There are two competing hypotheses:

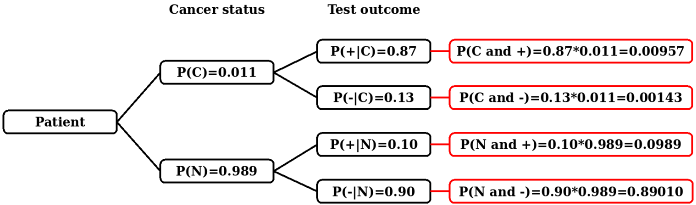
1. The patient has cancer (C).
2. The patient does not have cancer (N).

If a mammogram yields a positive result what is the probability that the patient has cancer?

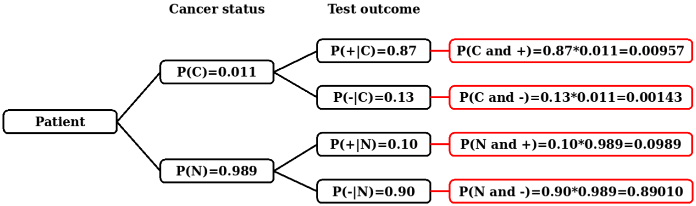


$$P(C|+) = \frac{P(C \; and \; +)}{P(+)} = \frac{P(C \; and \; +)}{P(C \; and \; +) + P(N \; and \; +)} = \frac{0.00957}{0.00957 + 0.0989} \approx 0.088$$

Applying Bayesian Inference



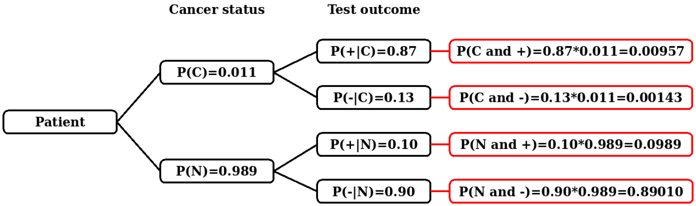
Applying Bayesian Inference



The patient obtained a positive mammogram result and wants to be tested again. To analyze the results of the **second test**, *what should we assume to be the probability that* **this particular patient** *has cancer?*

1. 1.1%
2. 8.8%
3. 87%
4. 0.957%

Applying Bayesian Inference



The patient obtained a positive mammogram result and wants to be tested again. To analyze the results of the **second test**, *what should we assume to be the probability that* **this particular patient** *has cancer?*

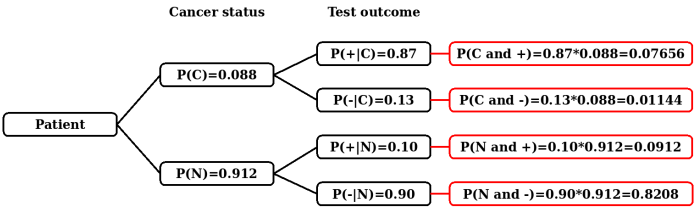
1. 1.1%
2. 8.8%
3. 87%
4. 0.957%

Inverting Cancer Probabilities: Second Test

What is the probability that our patient has cancer if the **second mammogram** *also yields a positive result?*

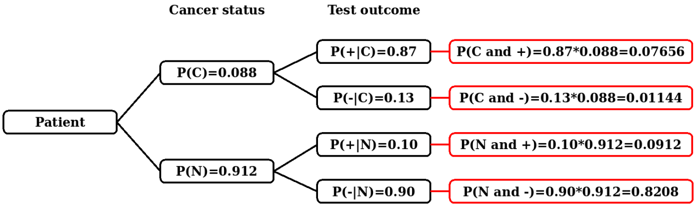
Inverting Cancer Probabilities: Second Test

What is the probability that our patient has cancer if the **second mammogram** *also yields a positive result?*



Inverting Cancer Probabilities: Second Test

What is the probability that our patient has cancer if the **second mammogram** *also yields a positive result?*



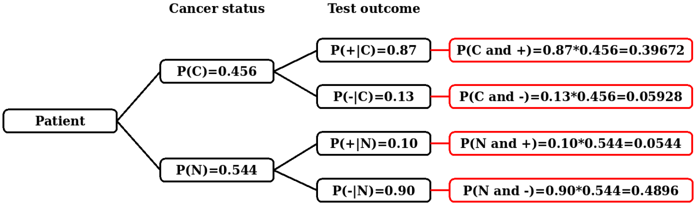
$$P(C|+) = \frac{P(C\ and\ +)}{P(+)} = \frac{P(C\ and\ +)}{P(C\ and\ +) + P(N\ and\ +)} = \frac{0.07656}{0.07656 + 0.0912} \approx 0.456$$

Inverting Cancer Probabilities: Third Test

What is the probability that our patient has cancer if the ***third mammogram*** *is also positive?*

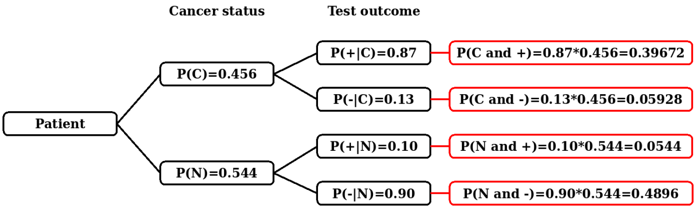
Inverting Cancer Probabilities: Third Test

What is the probability that our patient has cancer if the **third mammogram** *is also positive?*



Inverting Cancer Probabilities: Third Test

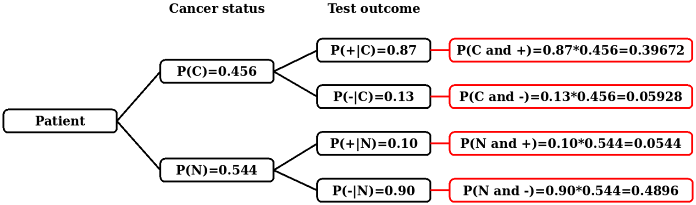
What is the probability that our patient has cancer if the **third mammogram** *is also positive?*



$$P(C|+) = \frac{P(C \text{ and } +)}{P(+)} = \frac{P(C \text{ and } +)}{P(C \text{ and } +) + P(N \text{ and } +)} = \frac{0.39672}{0.39672 + 0.0544} \approx 0.88$$

Inverting Cancer Probabilities: Third Test

What is the probability that our patient has cancer if the **third mammogram** *is also positive?*

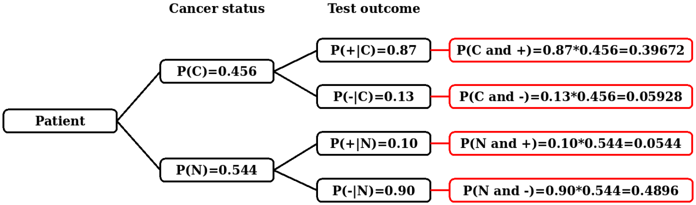


$$P(C|+) = \frac{P(C \text{ and } +)}{P(+)} = \frac{P(C \text{ and } +)}{P(C \text{ and } +) + P(N \text{ and } +)} = \frac{0.39672}{0.39672 + 0.0544} \approx 0.88$$

We get progressively more certain that the patient has cancer.

Inverting Cancer Probabilities: Third Test

What is the probability that our patient has cancer if the **third mammogram** *is also positive?*



$$P(C|+) = \frac{P(C \text{ } and \text{ } +)}{P(+)} = \frac{P(C \text{ } and \text{ } +)}{P(C \text{ } and \text{ } +) + P(N \text{ } and \text{ } +)} = \frac{0.39672}{0.39672 + 0.0544} \approx 0.88$$

We get progressively more certain that the patient has cancer.

Aside: we needed 3 tests to get to 88%. Not sure how many physicians think that the 1st test yields the same certainty (87%)...