MM 225: AI and Data Science

Conditional probabilities

M P Gururajan and Hina A Gokhale

August 29, 2023

Lecture 7: Conditional probabilities (continued)

Outline

1. Recall and generalise
   - Independent trials process
   - Bayes' formula

Independent trials process

Recall

Definition

Let X_1, X_2, $\cdots$, X_n be the random variables associated with an experiment. Suppose that the sample space (i.e., the set of possible outcomes) of X_i is the set R. Then, the joint random variable $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ is defined to be the random variable whose outcomes consist of ordered n-tuples of outcomes, wth the i-th coordinate lying in the set R_i. The sample space Ω of $\mathbf{X}$ is the Cartesian product of the R_i's: $\Omega = R_1 \times R_2 \times \cdots \times R_n$. The joint distribution function of $\mathbf{X}$ is the function which gives the probability of each of the outcomes of $\mathbf{X}$.

Definition

The random variables X_1, X_2, $\cdots$, X_n are mutually independent (or, independent) if $P(X_1 = r_1, X_2 = r_2, \cdots, X_n = r_n) = P(X_1 = x_1)P(X_2 = r_2)\cdots P(X_n = r_n)$ for any choice of $r_1, r_2, \cdots r_n$.

Bernoulli trials process

- Consider a Bernoulli trials process which has a probability of success p in each experiment.
- Let $X_j(\omega) = 1$ (0) if j-th outcome is a success (failure)
- $X_1, X_2, \cdots, X_n$ is an independent trials process
- Each X_i has the same distribution function
- $m_j(0) = q$ and $m_j(1) = p$.
- If $S_n = X_1 + X_2 + \cdots + X_n$, then $P(S_n = j) = \binom{n}{j} p^j q^{n-j}$

Die roll

- Consider rolling a fair die three times
- Let X_i represent the outcome of i-th roll
- Common distribution function $m(1) = m(2) = m(3) = m(4) = m(5) = m(6) = \frac{1}{6}$
- Sample space: $R^3 = R \times R \times R$ with $R = \{1, 2, 3, 4, 5, 6\}$
- Suppose $\omega = 1, 2, 5$: $X_1(\omega) = 1$; $X_2(\omega) = 2$ and $X_3(\omega) = 5$
- $m(\omega) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}$

Independent trials process

Definition

A sequence of random variables $X_1, X_2, \cdots, X_n$ that are mutually independent and have the same distribution is called a squence of independent trials process.

- Sample space of single experiment: $R = \{r_1, r_2, \cdots, r_s\}$
- Distribution function for a single experiment: $m_j = \{p_1, p_2, \cdots, p_s\}$ for $\{r_1, r_2, \cdots, r_s\}$ respectively.
- Repeat the experiment n times. The sample space of this totoal experiment $\Omega = R \times R \times \ldots \times R$ consisting of all possible sequences $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ chosen from R
- Distribution function: product distribution $m(\omega) = m(\omega_1)m(\omega_2)\cdots, m(\omega_n)$ with $m(\omega_j) = p_k$ when $\omega_j = r_k$
- X_j: denotes the j-th coordinate of the outcome $(r_1, r_2, \cdots, r_s)$ and the random variables $X_1, X_2, \cdots, X_n$ form an independent trials process

Bayes' formula

Recall

- Two urns containing black and white balls
- Choose the urn, choose a ball
- Two stage experiment
- Given outcome in the first stage, probability of outcome in the second stage: direct
- Given an outcome in the second stage (black ball), probability of outcome in the first stage (which urn?)
- Bayes probabilities (historically called inverse probabilities): conditionality reversion
- Let us look at another example!

Example

Suppose a doctor gives a patient a test for a disease. Before the results of the test, the only evidence the doctor has is that the incidence of this disease is 1 in 1000 in the population. Experience has also shown that the test comes out positive in 99% of the cases in which the disease is present, and the test comes out negative in 95% of the cases in which it is not present. The 1% and 5% cases are known as false negatives and false positives, respectively. If the test does come positive, what probability should the doctor assign for the patient to actually being afflicted by this disease?

Disease testing as two stage experiment

- First stage: the patient has a disease or no disease
- Second stage: the test comes positive or negative
- Given an outcome in the second stage (test result), what is the probability for the outcome in the first stage (has disease?)
- Bayes!

Bayes theorem

- We have a set of events that are pairwise disjoint; $\Omega = H_1 \cup H_2 \cup \cdots \cup H_m$
- H: hypotheses
- H_1: patient has disease; H_2: patient is not afflicted
- E: event that gives information about the correctness of hypotheses
- E: evidence
- E: results of the test for disease
- Probabilities for the hypotheses before receiving evidence: *prior probabilities*
- 0.001 and 0.999 are the priors for having disease or not afflicted by it
- $P(H_i)$: known

Bayes theorem

- Conditional probabilities: known
- If we know the correct hypothesis, we know the probability for evidence
- $P(E|H_i)$ is known for all i
- In our case, the probability of evidence given afficted by disease is 0.99 and the probability of evidence given not afflicted by disease is 0.95
- Probability for hypothesis given evidence: posterior probabilities
- Notice that we are asking for stage 1 event probability given stage 2 event result
- Bayes!!

The formula

- $$P(H_i|E) = \frac{P(H_i \cap E)}{P(E)} \tag{1}$$

- Since
$$P(E|H_i) = \frac{P(E \cap H_i)}{P(H_i)} \tag{2}$$

- the numerator of Eq. (1) is
$$P(E \cap H_i) = P(H_i)P(E|H_i) \tag{3}$$

- Only one of the events H_i can occur. So
$$P(E) = \sum_i P(H_i \cap E) \tag{4}$$

- Or, using Eq. (3),
$$P(E) = \sum_i P(H_i)P(E|H_i) \tag{5}$$

The formula

- $$P(H_i|E) = \frac{P(H_i \cap E)}{P(E)} \tag{7}$$
- $$P(E \cap H_i) = P(H_i)P(E|H_i) \tag{8}$$
- $$P(E) = \sum_i P(H_i)P(E|H_i) \tag{9}$$
- $$P(H_i|E) = \frac{P(H_i)P(E|H_i)}{\sum_k P(H_k)P(E|H_k)} \tag{10}$$

Bayes' formula

$$P(H_i|E) = \frac{P(H_i)P(E|H_i)}{\sum_k P(H_k)P(E|H_k)} \tag{11}$$

- LHS: all quantities are known
- RHS can be calculated
- Recall our problem: Suppose a doctor gives a patient a test for a disease. Before the results of the test, the only evidence the doctor has is that the incidence of this disease is 1 in 1000 in the population. Experience has also shown that the test comes out positive in 99% of the cases in which the disease is present, and the test comes out negative in 95% of the cases in which it is not present. If the test does come positive, what probability should the doctor assign for the patient to actually being afflicted by this disease?
- Can you solve it?

Answer

Test comes positive. Probability that the patient is actually afflicted:

- **Link:** https://www.menti.com/al8jqveiyr9s
- **Code:** Go to menti.com and use code 3253 4188

Results

Results of the poll

Answer

- P(Test positive given Disease yes) $= 0.99$
- P(Test positive given Disease no) $= 0.01$
- P(Test negative given Disease no) $= 0.95$
- P(Test negative given Disease yes) $= 0.05$
- P(Disease yes) $= 0.001$
- P(Disease no) $= 0.999$
- P(Disease yes given test positive) $= \frac{0.99*0.001}{0.99*0.001+0.05*0.999} = 0.019$
- Prior probability of 0.001 has now increased to a posterior probability of 0.019
- Importance of understanding this result!!

Question

Recall our problem: Suppose a doctor gives a patient a test for a disease. The patient knows that the the incidence of this disease is 1 in 1000 in the population. Patient has been told that the test comes out positive in 99% of the cases in which the disease is present, and the test comes out negative in 95% of the cases in which it is not present. If the test does come negative, what probability should the patient assign to actually being afflicted by this disease?

Answer

Test comes negative. Probability that the patient is actually afflicted:

- **Link:** https://www.menti.com/alsabqaob2c2
- **Code:** Go to menti.com and use code 4369 8671

Results

▸ Results of the poll

THANK YOU!!!