Conditional Probability

Mark Huiskes, LIACS
mark.huiskes@liacs.nl

Introduction

- This class will be about conditional probability.
- I'm sure you've seen the notation before: P(E|F) with the straight bar. What is the probability of event E given that you already know event F occurred?
- So today we are going to see, in some detail, how this works: for both the discrete and the continuous probability spaces.
- These conditional probabilities are important because they allow us to quantify how different probabilistic processes interact and, for instance, to define what it means for events or random variables to be independent.
- Other topics are joint and marginal distributions.
- So, quite some new and useful theory: let's get started.

Discrete Conditional Probability

- Suppose $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is a discrete sample space, with distribution function $m(\omega)$
- We learn that event E has occurred. What is our new distribution function $m(\omega|E)$?
  1. If $\omega \notin E$ then $m(\omega|E) = 0$
  2. For $\omega \in E$ the relative magnitudes of the outcome probabilities should stay the same: $m(\omega|E) = c \times m(\omega)$
- We can compute c from $\sum_{\omega \in \Omega} m(\omega|E) = c \sum_{\omega \in E} m(\omega) = 1$, giving

$$c = 1/\sum_{E} m(\omega) = 1/P(E)$$

- So, we define the conditional distribution given E as

$$m(\omega|E) = \frac{m(\omega)}{P(E)}, \text{ for } \omega \in E, \text{ and } 0 \text{ elsewhere.}$$

Universiteit Leiden

Conditional Probability of Events

- The probability of an event F is defined by

$$P(F) = \sum_{\omega \in F} m(\omega)$$

- The probability of an event F given E (given that the event E has occurred) is then:

$$P(F|E) = \sum_{\omega \in F} m(\omega|E) = \sum_{\omega \in F \cap E} \frac{m(\omega)}{P(E)} = \frac{P(F \cap E)}{P(E)}$$

- $P(F|E)$ is called the conditional probability of F given E

- Note that this means

$$P(F \cap E) = P(F|E)P(E) = P(E|F)P(F)$$

Example

Assignment 4. (a)

Card drawn randomly from full deck of (52) cards. What is the probability that it is a heart given that it is red. (Should be quite obvious that it is one half; but let's compute it anyway).

P(heart|red) = P(heart and red) / P(red)

P(red) = 26/52 = ½

P(heart and red) = P(heart) = ¼

P(heart|red) = ½ / ¼ = ½

Universiteit Leiden

Using Conditional Probability

- Urn 1 contains 4 red and 6 green balls while urn 2 contains 7 red and 3 green balls. An urn is chosen at random and then a ball is chosen from the selected urn.

  (a) Find the probability that the ball is green.

  (b) Given that the ball is green, find the conditional probability that urn 1 was selected.

Using Conditional Probability

- (From the second lecture: Theorem 1.3)
- Let $H_1, \ldots, H_n$ be pairwise disjoint events with $\Omega = H_1 \cup \ldots \cup H_n$ and let E be any event. Then

$$P(E) = \sum_{i=1}^{n} P(E \cap H_i)$$

- This means that

$$P(E) = \sum_{i=1}^{n} P(E|H_i)P(H_i)$$

- This decomposing a probability over different disjoint events is called conditioning.

Using Conditional Probability: Bayes' Formula

- Again $H_1, \ldots, H_n$, $\Omega = H_1 \cup \ldots \cup H_n$ disjoint events: hypotheses and E an event: the evidence.
- Suppose we know:

$P(E|H_i)$: the probability of the evidence given the hypothesis

$P(H_i)$: the prior probabilities (of the hypothesis: before evidence)

- We want to know: $P(H_i|E)$ the posterior probabilities

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

- With $P(E) = \sum_{i=1}^{n} P(E|H_i)P(H_i)$ based on conditioning,

we get Bayes's formula: $P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^{n} P(E|H_j)P(H_j)}$

Independence of Events

- Let E and F be events. They are independent if:
  1. P(F|E) = P(F), and P(E|F) = P(E) or:
  2. At least one of the events has probability 0

- Mention that P(F|E) = P(F), also means P(E|F) = P(F) [so if E doesn't tell us about F, F also doesn't tell us about E]: we will see that in a moment (from P(F and E) = P(E)P(F))

- Two events are independent if and only if ($\Leftrightarrow$)

$$P(E \cap F) = P(E)P(F)$$

- Proof: true for zero-event probabilities; else P(F|E) = P(F), or P(F and E) / P(E) = P(F), so P(F and E) = P(E) P(F); the other way around directly follows from this as well.
- Mention extension to more events: Definition 4.2. Explain with last expression: only one way.

Example

- Two coin tosses. A = {First toss is a head}; B = {two outcomes are the same}. Are these events independent?

- P(B|A) = P(A and B) / P(B) = P({HH})/P({HH,TT}) =1/2 = P(B)

Joint Distribution Functions

- Let $X_1, \ldots, X_n$ random variables with sample space R_i. The joint random variable $X = (X_1, \ldots, X_n)$ has sample space Omega = $R_1 \times R_2 \times \ldots R_n$.

- The joint distribution function of X is the function that gives the probability of each of the outcomes of X.

Example

- X2/X1

X2/X1	1	2	3
1	0.05	0.1	0.1
2	0.05	0.3	0.05
3	0.05	0.1	0.2

Marginal Distributions

- The probability distributions of the individual variables in a joint distribution.

- They can be obtained by summing:

$$m_{X_1}(\omega_1) = \sum_{\omega_2} m_X(\omega_1, \omega_2)$$

Independence of Random Variables

- The random variables X_1, …, X_n are (mutually) independent if

$$P(X_1 = r_1, X_2 = r_2, \ldots, X_n = r_n = P(X_1 = r_1)P(X_2 = r_2) \cdots P(X_n = r_n)$$

for any choice of $r_1, r_2, \ldots, r_n$.

- So if X_1, …, X_n are independent then their joint distribution function is the product of the individual distribution functions.

Continuous Conditional Probability

- The conditional density function is defined by:

$$f(x|E) = \begin{cases} f(x)/P(E) & \text{if} \quad x \in E \\ 0 & \text{if} \quad x \notin E \end{cases}$$

- Probability of an event:

$$P(F) = \int_{x \in F} f(x)\, dx$$

- Probability of an event F, given event E:

$$P(F|E) = \int_{x \in F} f(x|E)\, dx = \int_{x \in F \cap E} \frac{f(x)}{P(E)}\, dx = \frac{P(F \cap E)}{P(E)}$$

Independence

- Independence is defined exactly as for discrete sample spaces.

Joint Density Functions

- Theorem 4.2: Let X_1, X_2, …., X_n be continuous random variables with density functions f_1(x), f_2(x), …., f_n(x) and joint density f(x). Then these variables are independent if and only if:

$$f(x_1, \ldots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$$

for any choice $x_1, x_2, \ldots, x_n$

Binomial distribution

- Given *n* Bernouilli trials with probability *p* of success on each experiment, the probability of exactly *k* successes is:

$$b(n,p,k) = \binom{n}{k} p^k q^{n-k}$$

- Explanation:

$$P(E) \quad \text{with} \quad E = \{\omega | \omega \,\text{has}\, k \,\text{successes}\}\,.$$

$$P(E) = \sum_{\omega \in E} m(\omega)$$

Using the tree: every path with *k* successes and *n-k* failures:

$$m(\{k \,\text{successes}, n-k \,\text{failures}\}) = p^k q^{n-k}$$

How many such paths are there? n possible trials, k should be successes: $\binom{n}{k}$

- If B is a random variable counting the number of successes in a Bernouilli trials process with parameters n and p. Then the distribution m(k) = b(n,p,k) is called the Binomial distribution.