Lecture Stat 302
Introduction to Probability - Slides 6

AD

Jan. 2010

Independence

- We say that the events $\{E_i\}_{i=1}^n$ are independent if

$$P\left(\cap_{i=1}^n E_i\right) = \prod_{i=1}^n P\left(E_i\right).$$

and if for $i_1, i_2, ..., i_r$ where $i_j \neq i_k$ for $j, k \in \{1, ..., r\}$ and any $r \in \{1, ..., n\}$ we have

$$P\left(\cap \left\{E_{i_j}\right\}_{j=1,...,r}\right) = \prod_{j=1}^r P\left(E_{i_j}\right).$$

- **Example**: We say that three events E, F, G are independent if

$$P\left(E \cap F \cap G\right) = P\left(E\right) P\left(F\right) P\left(G\right)$$

and if

$$\begin{aligned} P\left(E \cap F\right) &= P\left(E\right) P\left(F\right), \; P\left(E \cap G\right) = P\left(E\right) P\left(G\right), \\ P\left(F \cap G\right) &= P\left(F\right) P\left(G\right). \end{aligned}$$

Example: Independent Trials

- An infinite sequence of independent trials is to be performed. With proba p it is a success and with proba $1-p$ this is a failure. What is the proba that 1) at least one success occurs in the first n trials 2) exactly k successes occur in the first n trials 3) all trials results in successes?
- **Answer:** 1) Let E_i event of a failure at trial i then proba of no success is

$$P\left(E_1 \cap E_2 \cap \cdots \cap E_n\right) = \prod_{i=1}^{n} P\left(E_i\right) = \left(1-p\right)^n$$

so proba of at least one success is $1-\left(1-p\right)^n$.
2) Proba of exactly k successes is

$$\binom{n}{k} p^k \left(1-p\right)^{n-k}.$$

3) Proba of all successes is
$P\left(\cap_{i=1}^{\infty} E_i^c\right) = \lim_{n\to\infty} P\left(\cap_{i=1}^{n} E_i^c\right) = \lim_{n\to\infty} p^n = 0$ for $p<1$.

Example: The problem of points

- An infinite sequence of independent trials is to be performed. With proba. p it is a success and with proba $1-p$ this is a failure. What is the proba of getting at least n successes before getting m failures?
- **Answer:** For at least n successes to occur before m failures, we need at least n successes in the first $n+m-1$ trials. Hence, we have

$$P\left(\text{at least } n \text{ successes}\right) = \sum_{k=n}^{n+m-1} P\left(k \text{ successes exactly}\right)$$
$$= \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k \left(1-p\right)^{m+n-1-k}$$

Example: Dice... again

- Independent trials consisting of rolling a pair of fair dice are performed. What is the proba that an outcome of 5 appears before an outcome of 7 when the outcome is the sum of the dice?
- **Answer**: Let $E_n =$ $\{$no 5 and 7 appears in the first $n-1$ trials and 5 appears at $n.\}$ then

$$P\left(\cup_{i=1}^{\infty} E_n\right) = \sum_{i=1}^{\infty} P\left(E_i\right).$$

where $P\left(\text{get a } 5\right) = \frac{4}{36}$, $P\left(\text{get a } 7\right) = \frac{6}{36}$

$$\begin{aligned} P\left(E_i\right) &= P\left(\text{no 5, no 7 } n-1 \text{ times}\right) P\left(\text{a 5 at } n\right) \\ &= \left(1 - \frac{10}{36}\right)^{n-1} \left(\frac{4}{36}\right) \end{aligned}$$

so finally

$$P\left(\cup_{i=1}^{\infty} E_n\right) = \frac{1}{9} \sum_{n=1}^{\infty} \left(\frac{13}{18}\right)^{n-1} = \frac{2}{5}.$$

Conditional Independence

- Given an event E and events $T_1, T_2, \ldots, T_n$, the probability of E given $I_n = T_1 \cap T_2 \cap \ldots \cap T_n$

$$P(E|I_n) = \frac{P(I_n|E)P(E)}{P(I_n|E)P(E) + P(I_n|E^c)P(E^c)}$$

- We say that the events $T_1, T_2, \ldots, T_n$ are conditionally independent given E (respectively E^c) if

$$\begin{aligned} P(I_n|E) &= P(T_1 \cap T_2 \cap \cdots \cap T_n|E) = \prod_{i=1}^{n} P(T_i|E), \\ P(I_n|E^c) &= P(T_1 \cap T_2 \cap \cdots \cap T_n|E^c) = \prod_{i=1}^{n} P(T_i|E^c). \end{aligned}$$

Application to Spam Emails Detection

- **Example**. When you receive an email, your spam filter uses Bayes rule to decide whether it is spam or not. Basic spam filters check whether some pre-specified words appear in the email you have received; e.g. {diplomat,lottery,money,politics,president,sincerely,huge,work...}. That is to each email is associated n events T_i telling us whether the ith pre-specified word is in the email or not. Let $E = \{\text{email received is spam}\}$.
- Based on many (humanly) labelled training data, i.e. non-spam and spam emails, we estimate

$$P(E) = \text{proba. email is a spam},$$

$$\begin{aligned} P(\text{word } i \text{ in}| E) &= \text{proba. } i^{\text{th}} \text{ word in email given it is spam}, \\ P(\text{word } i \text{ in}| E^c) &= \text{proba. } i^{\text{th}} \text{ word in email given it is not spam}. \end{aligned}$$

- The spam filter makes the assumption that the events T_i are conditionally independent to compute $P(E| I_n)$ and decide whether it is spam or not.

Updating Information Sequentially

- **Example**. A patient came to a clinic. We want to assess whether he has a nasty disease. For this we need to run not one but a battery of $n > 1$ tests. Let introduce the event $E = \{\text{patient sick}\}$ and the events $T_i = \{\text{result test } i\}$. Assuming for example that we have n tests and observed the events $T_1, T_2, ..., T_n$ then we are interested in

$$P(A|I_n) = \frac{P(I_n|A)P(A)}{P(I_n|A)P(A) + P(I_n|A^c)P(A^c)}.$$

where $I_n = T_1 \cap T_2 \cap \cdots \cap T_n$ is the result of all the n tests

- These tests might be very expensive so instead of running directly two tests, we might want to first compute $P(A|I_1) = P(A|T_1)$ after the 1st test which is given by

$$P(A|I_1) = \frac{P(T_1|A)P(A)}{P(T_1|A)P(A) + P(T_1|A^c)P(A^c)},$$

then only running the 2nd test if $P(A|I_1)$ is not close enough to 0 or 1 so that we cannot decide whether the patient is healthy or sick.

Updating Information Sequentially

- Assume we run the 2nd test and thus observe T_2, then we are now interested in $P(A|\, I_2)$ given by

$$P(A|\, I_2) = \frac{P(I_2|\, A)\, P(A)}{P(I_2|\, A)\, P(A) + P(I_2|\, A^c)\, P(A^c)}$$

  If $P(A|\, I_2)$ is not close enough to 0 or 1 so that we cannot decide whether the patient is healthy or sick, then we run the 3rd test and compute $P(A|\, I_3)$ and so on.
- In many situations, we have conditional independence of the events

$$\begin{aligned} P(I_n|\, A) &= P(T_1 \cap T_2 \cap \cdots \cap T_n|\, A) = \prod_{i=1}^{n} P(T_i|\, A), \\ P(I_n|\, A^c) &= P(T_1 \cap T_2 \cap \cdots \cap T_n|\, A^c) = \prod_{i=1}^{n} P(T_i|\, A^c). \end{aligned}$$

- In this case, we can compute $P(A|\, I_k)$ as a function of $P(A|\, I_{k-1})$.

Updating Information Sequentially

- We have

$$P(A|I_k) = \frac{P(T_k|A)\,P(A|I_{k-1})}{P(T_k|A)\,P(A|I_{k-1}) + P(T_k|A^c)\,P(A^c|I_{k-1})}.$$

- The proof follows from

$$P(A|I_k) = \frac{P(I_k|A)\,P(A)}{P(I_k|A)\,P(A) + P(I_k|A^c)\,P(A^c)}$$

where $P(I_k|A) = P(T_k|A)\,P(I_{k-1}|A)$ and $P(I_k|A^c) = P(T_k|A^c)\,P(I_{k-1}|A^c)$. So we have

$$\begin{aligned} P(A|I_k) &= \frac{P(T_k|A)P(I_{k-1}|A)P(A)}{P(T_k|A)P(I_{k-1}|A)P(A)+P(T_k|A^c)P(I_{k-1}|A^c)P(A^c)} \\ &= \frac{P(T_k|A)P(I_{k-1}|A)P(A)/P(I_{k-1})}{P(T_k|A)P(I_{k-1}|A)P(A)/P(I_{k-1})+P(T_k|A^c)P(I_{k-1}|A^c)P(A^c)/P(I_{k-1})} \end{aligned}$$

and the result follows.

Example: Conditional Independence and HIV Test

- Data from joint United Nation Programs on HIV/AIDS (2006): We have a test such that $P\left(T^{+}\middle|HIV^{+}\right)=0.99$ (sensitivity) and $P\left(T^{-}\middle|HIV^{-}\right)=0.99$ (specificity). The prevalence of HIV in East Asia is 0.001. Given the first test is positive, event T_1^{+}, what is the probability of being HIV positive? The policy for a positive HIV test is a follow-up confirmatory test. Given the 2nd test is positive, event T_2^{+} , what is the proba of being HIV positive?
- **Answer.** We have

$$P\left(HIV^{+}\middle|T_1^{+}\right)=\frac{P\left(T_1^{+}|HIV^{+}\right)P(HIV^{+})}{P\left(T_1^{+}|HIV^{+}\right)P(HIV^{+})+P\left(T_1^{+}|HIV^{-}\right)P(HIV^{-})}$$
$$=\frac{0.99\times0.001}{0.99\times0.001+0.01\times0.999}=0.0901$$

If the 2nd test is positive, we need to compute

$$P\left(HIV^{+}\middle|T_1^{+}\cap T_2^{+}\right)=\frac{P\left(T_1^{+}\cap T_2^{+}|HIV^{+}\right)P(HIV^{+})}{P\left(T_1^{+}\cap T_2^{+}|HIV^{+}\right)P(HIV^{+})+P\left(T_1^{+}\cap T_2^{+}|HIV^{-}\right)P(HIV^{-})}$$

Example: Conditional Independence and HIV Test

- Given the information we have been given, we can only compute this proba if we assume that

$$
\begin{aligned}
P\left(T_1^+ \cap T_2^+ | HIV^+\right) &= P\left(T_1^+ | HIV^+\right) P\left(T_2^+ | HIV^+\right), \\
P\left(T_1^+ \cap T_2^+ | HIV^-\right) &= P\left(T_1^+ | HIV^-\right) P\left(T_2^+ | HIV^-\right)
\end{aligned}
$$

In this case we have

$$
P\left(HIV^+ | T_1^+ \cap T_2^+\right) = \frac{0.99 \times 0.99 \times 0.001}{0.99 \times 0.99 \times 0.001 + 0.01 \times 0.01 \times 0.999} = 0.980
$$

- Alternatively, we can use the sequential updating rule

$$
P\left(HIV^+ | T_1^+ \cap T_2^+\right) = \frac{P\left(T_2^+ | HIV^+\right) P\left(HIV^+ | T_1^+\right)}{P\left(T_2^+ | HIV^+\right) P\left(HIV^+ | T_1^+\right) + P\left(T_2^+ | HIV^-\right) P\left(HIV^- | T_1^+\right)}
$$

$$
= \frac{0.99 \times 0.0901}{0.99 \times 0.0901 + 0.01 \times 0.020} = 0.980
$$