Topic 2: Conditional Probability and Independence

Lecture Outline

- Conditional Probability (條件機率)
  - Chain Rule
- Independence (獨立事件)
- Total Probability Theorem (全機率定理)
- Bayes Rule (貝氏定理)

Reading: Textbook 1.3, 1.4, 1.5

Motivations

Questions:

- In an experiment involving two successive rolls of a die, you are told that the sum of the two rolls is 9. How likely is it that the first roll was a 6?
- How likely is it that a person has a disease given that a medical test was negative?
- A spot shows up on a radar screen. How likely is it that it corresponds to an enemy aircraft?

我們所面對的工程系統設計問題，或是透過人工智慧處理的預估與分類問題幾乎都是在得知某些事件(如*給定某些量測訊號、資料*)的前提下作出決策判斷。這所*給定的訊號、資料*即是條件機率中的給定“條件”。

在得知其他某些事情發生後，我們對感興趣事件的了解可能會有所改變(原機率 vs. 條件機率)。系統再根據計算出的條件機率值做出對應決策。

Example

Example:

機率期中考時，混哥對某單選選擇題4選項 A、B、C、D毫無頭緒。沒唸書的混哥亂猜答對機率為1/4。

碰巧混哥坐卷姊隔壁，偷偷瞄見

在此事件發生後，對混哥而言猜對機率為？

Another Example

Let's look at a simple example before introducing the formal definition.

(Example)
Suppose that all six possible outcomes of a fair die roll are equally likely. If we are told that the outcome is even, we are left with only three possible outcomes, namely, 2, 4, and 6. What is the probability that the outcome is 6?

Then, it is reasonable to let (根據直覺得出的答案)

$$P\left(\underbrace{\text{the outcome is 6}}_{\text{Event } F\text{（感興趣之事件）}} \,\middle|\, \underbrace{\text{the outcome is even}}_{\text{Event } G\text{（觀察到的、已知事件）}}\right) = 1/3$$

Thus, an intuitive definition for ***conditional probability*** from the example is:

$$P(F \mid G) = \frac{\#\text{ of outcomes in both } F \text{ and } G}{\#\text{ of outcomes in } G}$$

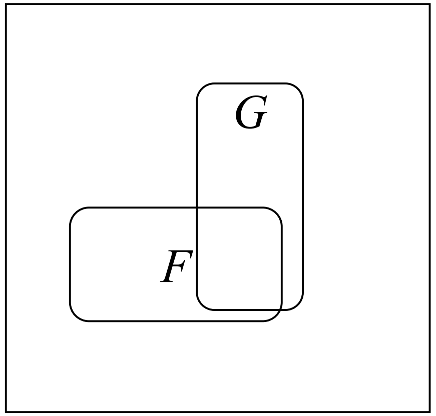
$$= \frac{n(F \cap G)}{n(G)} = \frac{n(F \cap G)/N}{n(G)/N}$$

Conditional Probability

Notation: $P(F \mid G)$

Probability of F given G occurred. G becomes the *new sample space*.

觀察到 G 之後，對 F 有了新的認識，不在 G 事件裡的 outcome 都不會發生了。



Definition:

Assume P(G)≠ 0, the conditional probability is defined by

$$\boxed{P(F \mid G) = \frac{P(F \cap G)}{P(G)}}$$

If all elementary outcomes are equally likely (古典機率), then we have

$$P(F \mid G) = \frac{\#\text{ of outcomes in both } F \text{ and } G}{\#\text{ of outcomes in } G}$$

Conditional Probabilities Specify a Probability Law

- Conditional probabilities form a legitimate probability law

We can prove that conditional probabilities satisfy the 3 axioms of probability theory, i.e.

- $P(F|G)$ is nonnegative
- $P(\Omega|G)=1$
- $P\left(F_1 \cup F_2|G\right)=P(F_1|G)+P(F_2|G)$ for mutually exclusive F_1 and F_2

Chain Rule

Motivation:

It is often convenient to calculate unconditional probabilities (of the intersection of events) using conditional probabilities

Chain Rule (乘法法則)

It's just a restatement of definition of conditional probability $P(F \cap G) = P(F \mid G)P(G)$

Multiplication rule (chain rule):

可以延伸到

$$P(F_1 \cap F_2 \cap F_3) = ?$$

$$P(F_1 \cap F_2 \cap \ldots \cap F_n) = ?$$

Sequential Calculation of Probabilities

Example:

Three cards are drawn (without replacement) from an ordinary deck. What is the probability of *not* drawing any heart?

Let A_i= {抽出的第 i 張不是紅心} i=1, 2, 3

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

Radar Detection Example

If an aircraft is present in a certain area, a radar detects and generates an alarm signal with probability 0.99. If it is not present, the radar generates an alarm with probability 0.10. We assume that an aircraft is present with probability 0.05.

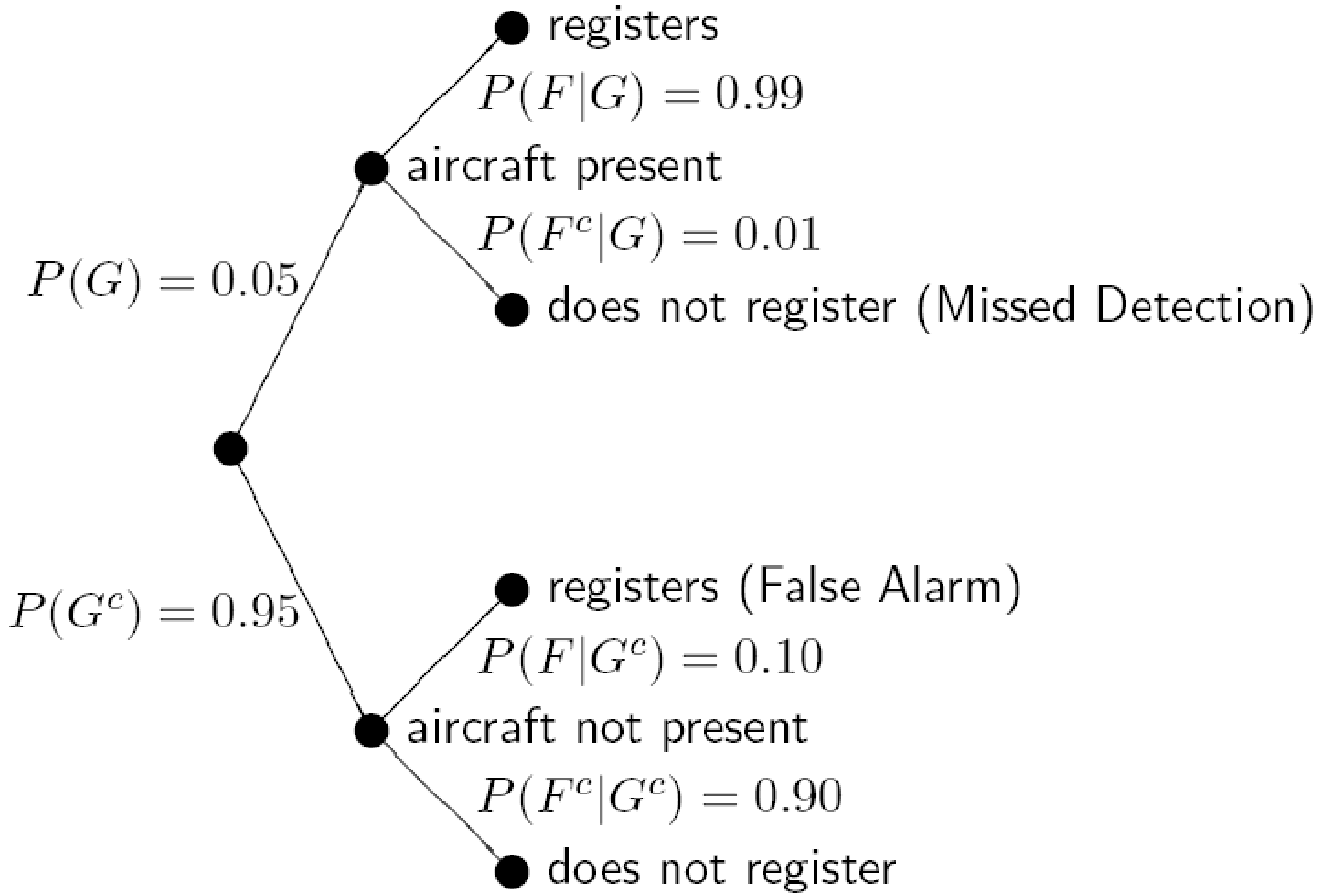
1. What is the probability of no aircraft presence and false alarm?
2. What is the probability of aircraft presence and no detection?

Mathematical Formulation:

- G: Aircraft is flying above
- F: The radar generates an alarm

Equivalent to finding the probability of

1. missed detection? $P(G \cap F^c)$
2. false alarm? $P(F \cap G^C)$


registers
$P(F|G) = 0.99$
aircraft present
$P(F^c|G) = 0.01$
$P(G) = 0.05$
does not register (Missed Detection)
$P(G^c) = 0.95$
registers (False Alarm)
$P(F|G^c) = 0.10$
aircraft not present
$P(F^c|G^c) = 0.90$
does not register

More on Radar Detection:

1. Both cases can be interpreted as ***error probabilities*** since the signal processor (the radar) makes a mistake (either missed detection or false alarm)

2. Two types of error can have significantly different effects or costs
   - False alarm: not comfortable, but we can survive (=false positive 假警報、偽陽性)
   - Missed detection: not tolerable (=false negative, 偽陰性)

3. Same issue arises in medical diagnoses, e.g., detecting cancer
   - False diagnosing a cancer (referred to as false alarm or false positive)
   - Missing a cancer that is there (missed detection or false negative)

Independent Events (獨立事件)

If F and G are independent, then the occurrence of G should have no effect on the probability of F, that is, (assuming $P(G) \neq 0$)

$$P(F \mid G) = P(F) \qquad (1)$$

Or, equivalently, the above can be rewritten as

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

Definition

So, a formal definition for two events F and G to be independent is

$$P(F \cap G) = P(F)P(G) \qquad (2)$$

Remark

The definition in (2) is more general than def. (1) since there is no requirement for $P(G) \neq 0$. Nonetheless, def. (1) reveals the essence (physical interpretation, 物理意義) of independence. (條件機率=原機率，表示給定的條件無關緊要)

Disjoint and Independent

Question*:*

Are disjoint events independent?

Conditional Independence

Definition:

Given an event C with P(C)>0, the events *A* and *B* are called **conditionally independent** if

$$\mathbf{P}(A \cap B \mid C) = \mathbf{P}(A \mid C)\mathbf{P}(B \mid C)$$

Equivalent Condition:

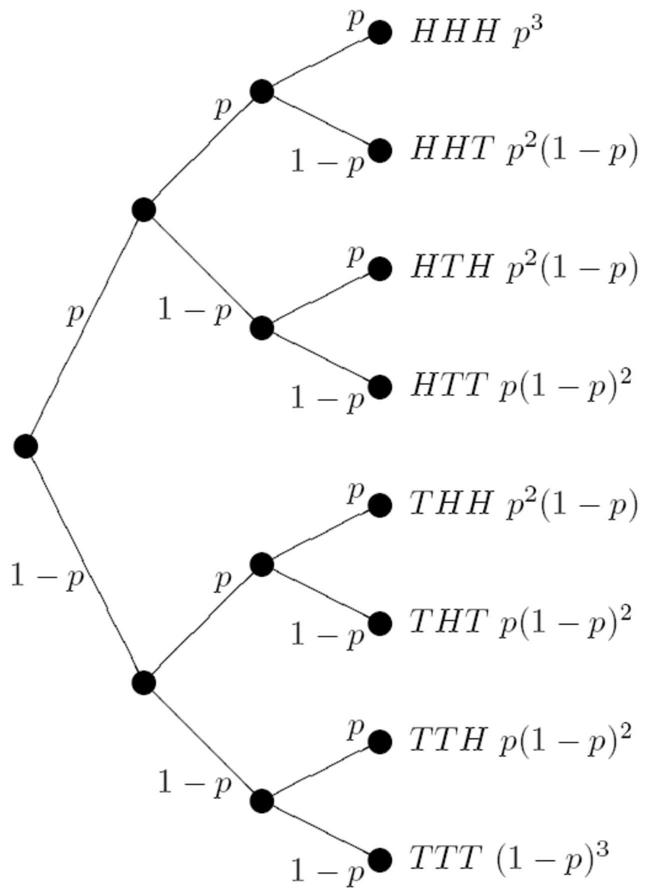
$$\mathbf{P}(A \mid B \cap C) = \mathbf{P}(A \mid C)$$

Physical Interpretation:

This relation states that if *C* is known to have occurred, the additional knowledge that *B* also occurred does not change the probability of *A.* In other words, the knowledge that *B* contains about *A* is totally included in *C*

Independence makes the sequential approach (multiplication rule) easier:

Example:
Independent tosses of a biased coin with P(Head)=p and P(Tail)=1-p



In a sequential description of an experiment where a collection of outcomes are mutually independent, we say that the experiment is a *sequence* of *independent trials*.

If the outcome is binary, we say we have a sequence of *Bernoulli trials*.

Suppose we have a sequence of Bernoulli trials with outcomes $F_1, F_2, \dots, F_n$ such that $F_i = \{$ toss i of a biased coin is a head $\}$ with $P(H) = p$.

What is $\Pr($ k heads in n tosses $)$? (or, in a more general language, of k successes in n Bernoulli trials)

Note that for $n = 4$ and $k = 2$,

$$\begin{aligned}
\Pr(k \text{ heads in } n \text{ tosses}) &= P(HHTT) + P(HTHT) + P(HTTH) \\
&\quad + P(THHT) + P(THTH) + P(TTHH) \\
&= 6p^2(1-p)^2 \\
&= \binom{4}{2} p^2(1-p)^2
\end{aligned}$$

Binomial probability law:

In general:

$$\Pr(k \text{ heads in } n \text{ tosses}) = \binom{n}{k} p^k (1-p)^{n\text{-}k}; \; x = 0, 1, 2, \ldots, n$$

are binomial coefficients

Bernoulli probability law

Bernoulli law is a special case of Binomial law in which $n=1$:

$$\Pr(k \text{ heads in } 1 \text{ toss}) = p^k (1-p)^{1\text{-}k} = \begin{cases} p & k = 1 \\ 1-p & k = 0 \end{cases}$$

Total Probability Theorem (TPT)

Recall the elementary property of probability we called "total probability":

If $F_1, F_2, \ldots, F_K$ is a finite partition of Ω, i.e., $F_i \cap F_k = \phi$ when $i \neq k$ and $\bigcup_{i=1} F_i = \Omega$, then

$$P(G) = \sum_{i=1}^{K} P(G \cap F_i)$$

If $P(F_i) > 0$ for all i, then using conditional probability we can write

$$\boxed{P(G) = \sum_{i=1}^{K} P(G \mid F_i) P(F_i)}$$

Key point: 將複雜的事件G分解成較簡單且彼此互斥的事件$\{G \cap F_1\}$、$\{G \cap F_2\}$、⋯、 $\{G \cap F_K\}$ 之聯集

再次提及，這是一個非常重要的結果!

1. 我們可透過將G拆解，分項討論$P(G|F_i)$後再綜合起來計算複雜事件G的機率值。
2. 後續談到隨機變數時會再次使用TPT，而會有四個重要的變形!

Example

- Chess tournament, 3 types of opponents
  - P(Type 1) = 0.5, P(Win|Type 1) = 0.3
  - P(Type 2) = 0.25, P(Win|Type 2) = 0.4
  - P(Type 3) = 0.25, P(Win|Type 3) = 0.5

- What is probability of Win?

  {贏棋}是複雜事件，必須分成不同類別的對手分項討論!

- Let W= Win, F_i = Type i :

$$
\begin{aligned}
P(W) &= P(W \mid F_1)P(F_1) + P(W \mid F_2)P(F_2) \\
&\quad + P(W \mid F_3)P(F_3) \\
&= 0.5 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.5 \\
&= 0.375
\end{aligned}
$$

Bayes' Rule (貝式定理)

- Bayes' rule states the relation between $P(F_i|G)$ and $P(G|F_i)$

- Plug the result of total probability for $P(G)$ into the denominator below

$$P(F_i \mid G) = \frac{P(F_i \cap G)}{P(G)} = \frac{P(G \mid F_i)P(F_i)}{P(G)}$$

$$= \frac{P(G \mid F_i)P(F_i)}{\sum_j P(G \mid F_j)P(F_j)}$$

- **貝氏定理在工程上的應用:**

工程系統設計(如手機通訊演算法)，或是透過人工智慧處理的預估與分類問題幾乎都是在得知某些事件(如*已知量測訊號、蒐集到的資料)*的前提下作出**決策判斷**。這所*已知的訊號、資料*即是貝氏定理中的給定“條件”(上式的G事件)。

在得知G事件後，我們對感興趣事件的了解可能會有所改變(原機率 vs. 條件機率)。系統再根據計算出的條件機率值做出對應決策。如:

手機接收到訊號，手機晶片該執行何種運算(基頻演算法)，用以偵測傳送訊號是0或1?

Applications of Bayes' Rule

Consider a more generic problem in statistical inference (統計推論)

Given observed "effect" or "result" (event G，*結果*), infer the unobserved "cause" (one of the events $F_1, F_2, \ldots, F_n$，*原因*).

Procedure in Bayesian statistical inference

- Assume we know the "prior" or "*a priori*" probabilities $P(F_i)$, for all i, and the conditional probabilities $P(G|F_i)$ （ $P(F_i)$ 和 $P(G|F_i)$ 較容易估算，但我們的目的是 $P(F_i|G)$ ）
- *Compute* $P(F_i|G)$, for all i *(i.e., all possible causes $F_1 \ldots F_n$)*, i.e., the "posterior" probabilities (事後機率) using Bayes' rule
- *Compare* posterior probabilities for all i *to infer which cause is the most likely to* have led to G
  - *Choose the cause F_i that has the maximum posterior probability $P(F_i|G)$ among $P(F_1|G)$, $P(F_2|G)$, … $P(F_n|G)$*

上述做法稱做 *maximum a posterior probability* (MAP) 判斷法則。

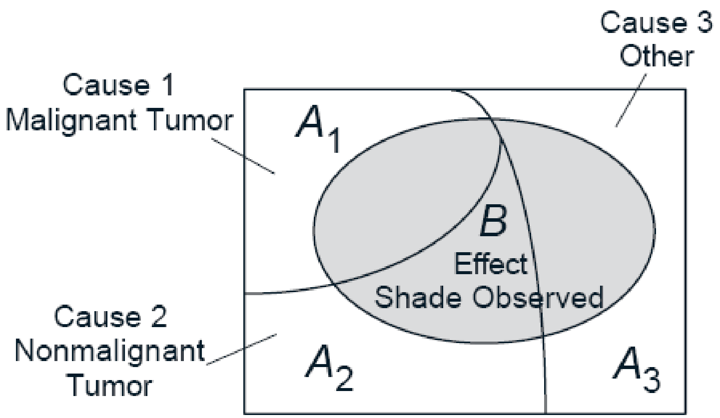
Why is it so special?

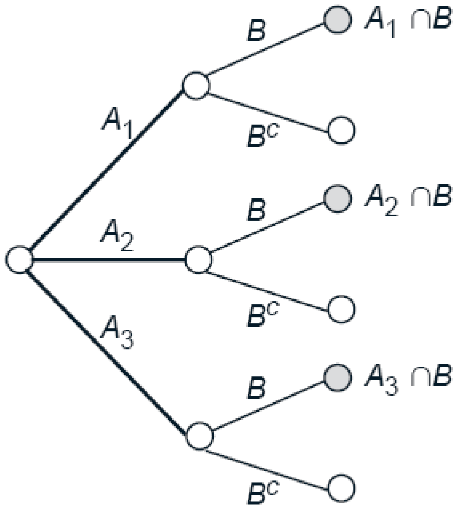
An Example of Statistical Inference

Observe a shade in a person's X-ray (this is event B, the "effect").
We want to estimate the *likelihood* of three mutually exclusive and collectively exhaustive potential causes:

1. cause 1 (event A_1) is that there is a malignant tumor
2. cause 2 (event A_2) is that there is a nonmalignant tumor
3. cause 3 (event A_3) corresponds to reasons other than a tumor

What interests us are the *posterior probabilities* $P(A_i|B)$ for $i=1,2,3$.





Chess Example Again

- Chess tournament, 3 types of opponents
  - P(Type 1) = 0.5, P(Win|Type 1) = 0.3
  - P(Type 2) = 0.25, P(Win|Type 2) = 0.4
  - P(Type 3) = 0.25, P(Win|Type 3) = 0.5

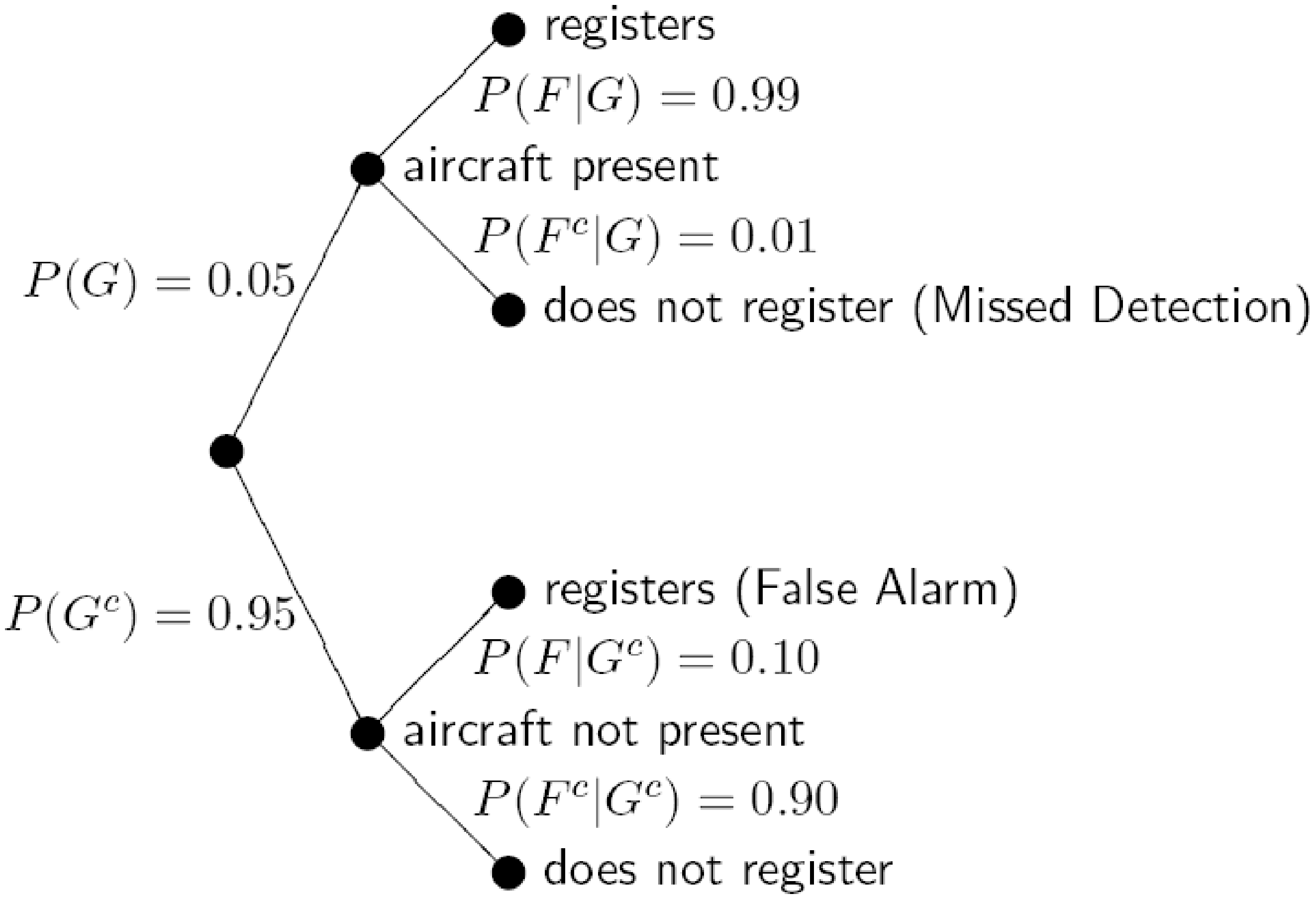
- Given that I win, what is the probability I had a Type 1 opponent?

- Using Bayes' rule,

$$
\begin{aligned}
P(F_1 \mid G) &= \frac{P(G \mid F_1)P(F_1)}{\sum_{j=1}^{3} P(G \mid F_j)P(F_j)} \\
&= \frac{0.5 \times 0.3}{0.5 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.5} \\
&= 0.4
\end{aligned}
$$

Radar Example

- G: Airpcraft is flying above
- F: Something registers on radar screen



What is the probability that an aircraft is actually there given that the radar indicates a detection?

柯南的推理

經過調查，柯南已確信琴酒有60%的可能偷走了園子姐姐的鑽石。現在，新的證據顯示真正的犯人是個左撇子。

已知在米花市左撇子人數占該市總人口數比例為20%。柯南認為，琴酒如果是一個左撇子，他就*幾乎確認*琴酒會是真正的犯人。

請問，柯南的『幾乎確認』該如何用科學方法換算出真正的機率值?

令事件 G={琴酒是真犯}、F={琴酒是左撇子}

$$
\begin{aligned}
P(G|F) &= \frac{P(F|G)P(G)}{P(F|G)P(G) + P(F|G^c)P(G^c)} \\
&= \frac{1 \times 0.6}{1 \times 0.6 + 0.2 \times (1 - 0.6)} \\
&\approx 0.882
\end{aligned}
$$

(Example 3f, A First Course in Probability, S. Ross, 8th edition)

三個事件的獨立性質

If we have a collection of three events, A_1, A_2, and A_3, independence amounts to satisfying the ***four*** conditions

1. $P(A_1 \cap A_2) = P(A_1)P(A_2)$
2. $P(A_1 \cap A_3) = P(A_1)P(A_3)$
3. $P(A_2 \cap A_3) = P(A_2)P(A_3)$
4. $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

- The first three conditions are known as **pairwise independence**

- The 4 conditions must be satisfied *simultaneously* for us to claim indep. of , A_1, A_2 and A_3
  - Pairwise independence does NOT imply the 4th condition
  - Conversely, the 4th condition does NOT imply pairwise independence (See **examples 1.22** and **1.23**)