

Lecture notes 3

Conditional Probability and Independence

Lecture Outline

- Conditional probability
- Chain rule (multiplicative rule)
- Independence
- Total Probability
- Bayes Rule

Reading: Bertsekas & Tsitsiklis 1.3, 1.4, 1.5

Review of Probabilistic Model

- Experiment: a nondeterministic process
- Sample space: list of outcomes of an experiment (disjoint, elementary, finest grain)
- Events: Subsets of sample space
- Probability measure: An assignment $P(F)$ to every event F that satisfies three axioms:

Nonnegativity $P(F) \geq 0$ for all F .

Normalization $P(\Omega) = 1$

Additivity Given a collection $\{F_k; k = 1, 2, \dots\}$ If $F_k \cap F_j = \emptyset$ for $k \neq j$, then $P(\bigcup_k F_k) = \sum_k P(F_k)$.

Methods for Calculating Probabilities

- Counting method (for discrete uniform law, using additivity)

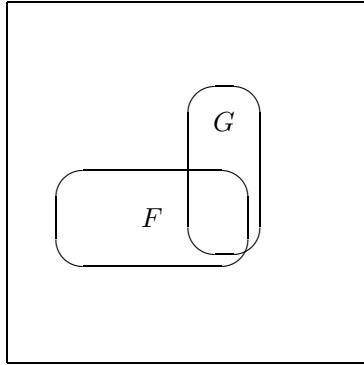
$$P(F) = \frac{\text{number of sample points in } F}{\text{total number of sample points}}$$

- Sequential method (multiplication rule)
- Divide-and-conquer method (total probability theorem + conditional probability)
- Combinations of the above three.

Conditional probabilities

- are the basis of the sequential method and the divide-and-conquer theorem.,
- allow computation and determination of probabilities based on *partial knowledge*,
- form the basis for *inference* (observe an effect and reason about the cause — next lecture).

Conditional Probability



- $P(F|G)$ = probability of F given G occurred, G becomes the new sample space

- Definition: Assuming $P(G) \neq 0$,

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

- If all elementary outcomes equally likely (begin with discrete uniform law),

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

Intuition behind definition:

Suppose discrete experiment. Being told G occurred means that only sample points in G are possible so that $P(\omega|G) = 0$ if ω is not in G .

From the additivity of probability, we must have that for any event F

$$P(F|G) = \sum_{\omega: \omega \in F} P(\omega|G)$$

This means in particular that

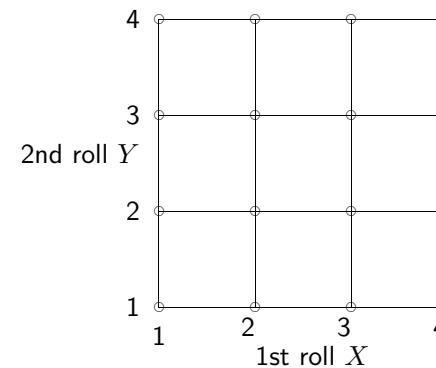
$$P(G|G) = 1 = \sum_{\omega: \omega \in G} P(\omega|G)$$

Lastly, being told that $\omega \in G$ should not change the *relative* values of probabilities. A definition with all these properties is

$$P(\omega|G) = \begin{cases} 0 & \omega \notin G \\ \frac{P(\omega)}{P(G)} & \omega \in G \end{cases}$$

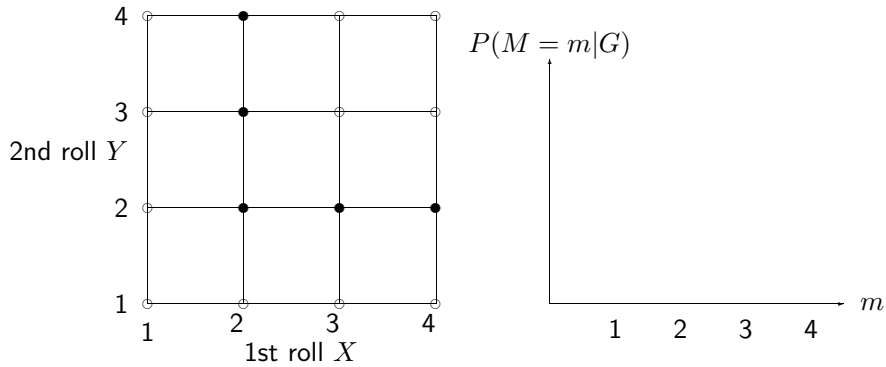
Die Roll Example

Roll a fair four-sided die twice. \Rightarrow discrete uniform law on samplespace:



All sample points have probability $1/16$.

- Let G be the event $\min(X, Y) = 2$.
- Let $M = \max(X, Y)$.
- Compute $P(M = m|G)$.



Dark circles denote event G .

Conditional Probability Models

Before: Probability law \Rightarrow conditional probabilities

Can do the reverse: Conditional probabilities \Rightarrow probability law

Use *chain rule* (also called *multiplication rule*):

Definition of conditional probability $\Rightarrow P(F \cap G) = P(F|G)P(G)$.

Suppose that F_1, F_2, \dots, F_n are events. Then

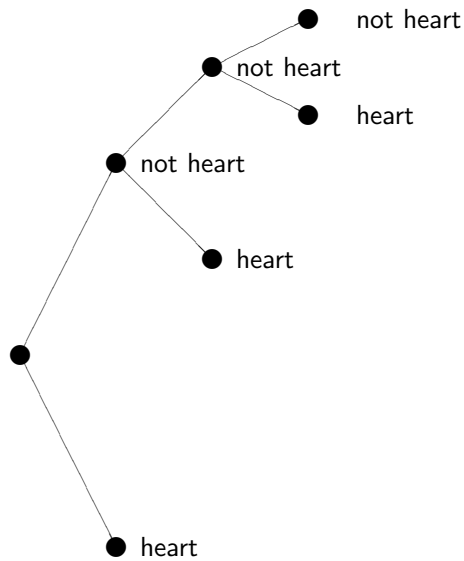
$$\begin{aligned}
 &P(F_1 \cap F_2 \cap F_3 \cdots \cap F_n) \\
 &= P(F_n | F_1 \cap F_2 \cap F_3 \cdots \cap F_{n-1}) \times P(F_1 \cap F_2 \cap F_3 \cdots \cap F_{n-1}) \\
 &= P(F_n | F_1 \cap F_2 \cap F_3 \cdots \cap F_{n-1}) \\
 &\quad \times P(F_{n-1} | F_1 \cap F_2 \cap F_3 \cdots \cap F_{n-2}) \\
 &\quad \times P(F_1 \cap F_2 \cap F_3 \cdots \cap F_{n-2})
 \end{aligned}$$

$$\begin{aligned}
 &\vdots \\
 &= P(F_1) \times P(F_2 | F_1) \times P(F_3 | F_1 \cap F_2) \cdots \\
 &\quad \times P(F_n | F_1 \cap F_2 \cap F_3 \cdots \cap F_{n-1})
 \end{aligned}$$

Sequential Calculation of Probabilities

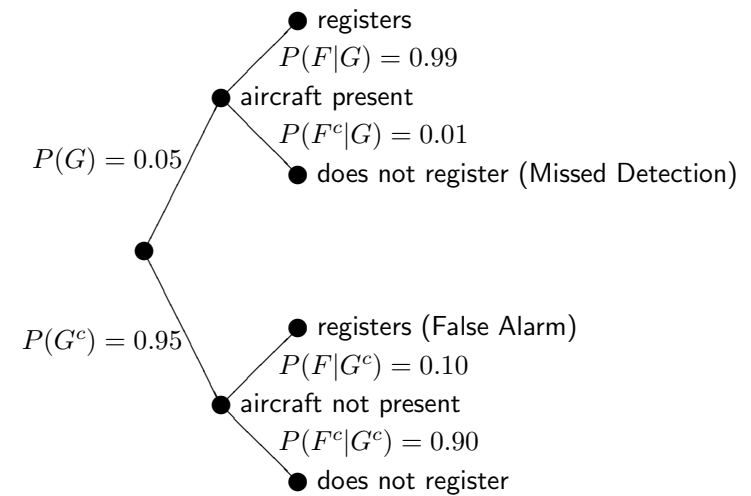
1. Construct a sequential description of the sample space
2. Record the corresponding conditional probabilities along the branches of the tree
3. Obtain the probability of any one outcome by using the multiplication rule

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



Another Example

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



What is the probability of

- missed detection? $P(G \cap F^c)$
- a false alarm? $P(F \cap G^c)$

Note that both of these can be interpreted as *error probabilities* since the signal processor makes a mistake.

Two types of error can have significantly different effects or *costs*.

Same issue arises in medical diagnoses, e.g., detecting cancer.

Missing a cancer that is there (missed detection or *false negative*)

False diagnosing a cancer (false alarm or *false positive*)

Independent Events

If F and G are “independent” in an intuitive sense, then the occurrence of G should have no effect on the probability of F , that is, $P(F|G) = P(F)$ or (assuming $P(G) \neq 0$)

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

or

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

So *define* two events F and G to be independent if

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

Note: This definition is more general than the first one since there is no requirement that $P(G) \neq 0$

Question Are disjoint events independent?

Answer No, unless one of the events has zero probability.

Think about it, being disjoint is a very strong form of dependence.

Another example

A colorblind man selects one coin from a box containing one red coin and four white coins. He knows that for a red coin, $P(H) = 2/3$ and for a white coin $P(H) = 1/2$.

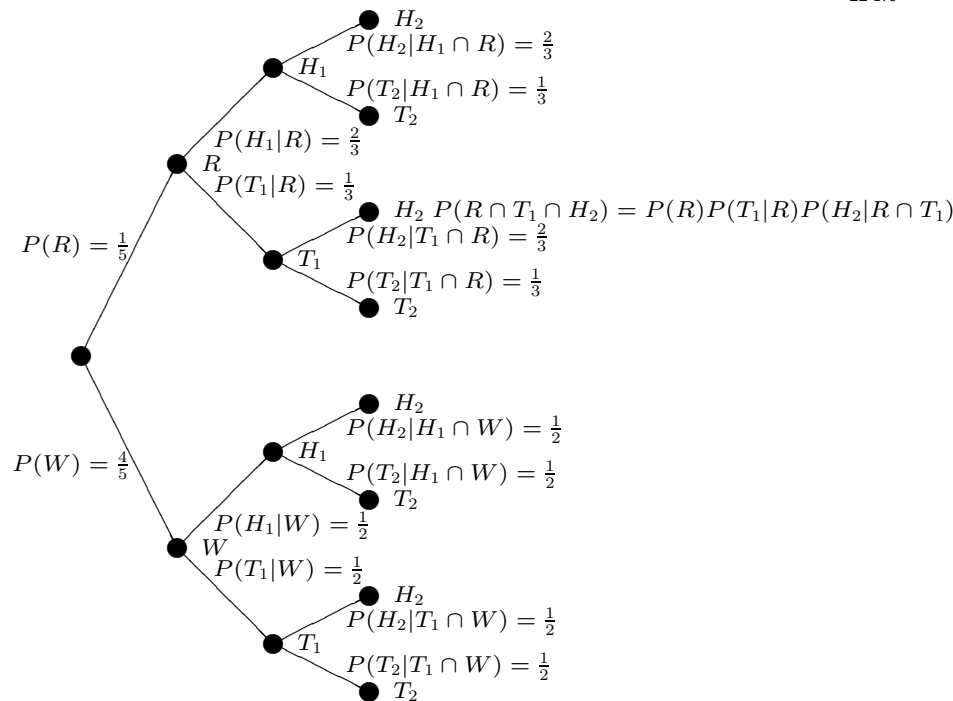
1. If after two flips of the coin he selected he has observed the sequence HT, what is the probability that one more flip of this coin will result in a head? (An inference problem)
2. If after observing the HT sequence he pockets the first coin and draws another from the box, what is the probability that the first two flips of the new coin will result in the sequence HH?
3. Is the event H_2 (the second flip is a head) independent of H_1 (the first flip is a head)?
4. Is the event H_2 independent of H_1 in a conditional sample space where the color is given?

Initial thoughts:

Told: $P(R) = 1/5$, $P(W) = 4/5$. Given R , then subsequent coin flips are independent with $P(H|R) = 2/3$. Given W , then subsequent coin flips are independent with $P(H|W) = 1/2$.

Now draw a tree for the experiment to show the sequential sample space.

To find the probability of a node, multiply the branch probabilities.



Question 4: The question is does $P(H_1 \cap H_2|R) = P(H_1|R)P(H_2|R)$ and $P(H_1 \cap H_2|W) = P(H_1|W)P(H_2|W)$?

The answer is yes because this is built in to the model. In particular, using the chain rule

$$\begin{aligned} P(H_1 \cap H_2|R) &= \frac{P(H_1 \cap H_2 \cap R)}{P(R)} \\ &= \frac{P(H_2|H_1 \cap R)P(H_1|R)P(R)}{P(R)} \\ &= \frac{2}{3} \times \frac{2}{3} \\ &= P(H_2|R)P(H_1|R) \end{aligned}$$

Note that the chain rule could be applied to the conditional probabilities to save a little work, i.e., $P(H_1 \cap H_2|R) = P(H_2|H_1 \cap R)P(H_1|R)$. This is the short cut that combined with bad notation caused confusion in class. The same idea works for W and the details are left as an exercise.

Question 3: Is $P(H_1 \cap H_2) = P(H_1)P(H_2)$? Use total probability and the chain rule to write

$$\begin{aligned} P(H_1 \cap H_2) &= P(H_1 \cap H_2 \cap R) + P(H_1 \cap H_2 \cap W) \\ &= P(H_2|H_1 \cap R)P(H_1|R)P(R) + P(H_2|H_1 \cap W)P(H_1|W)P(W) \\ &= \frac{2}{3} \times \frac{2}{3} \times \frac{1}{5} + \frac{1}{2} \times \frac{1}{2} \times \frac{4}{5} = \frac{13}{45} \\ P(H_1) &= P(H_1 \cap R) + P(H_1 \cap W) \\ &= P(H_1|R)P(R) + P(H_1|W)P(W) \\ &= \frac{2}{3} \times \frac{1}{5} + \frac{1}{2} \times \frac{4}{5} = \frac{8}{15} \\ &= P(H_2) \end{aligned}$$

where the last equation follows from an identical calculation for H_2 .

Since

$$\frac{13}{45} \neq \frac{64}{225}$$

the answer is **no**

Question 1: The question asks for $P(H_3|H_1 \cap T_2)$.

Again combine total probability and the chain rule.

$$\begin{aligned} P(H_3|H_1 \cap T_2) &= \frac{P(H_1 \cap T_2 \cap H_3)}{P(H_1 \cap T_2)} \\ &= \frac{P(R \cap H_1 \cap T_2 \cap H_3) + P(W \cap H_1 \cap T_2 \cap H_3)}{P(R \cap H_1 \cap T_2) + P(W \cap H_1 \cap T_2)} \\ &= \frac{P(H_2|R \cap H_1 \cap T_2)P(T_2|R \cap H_1)P(H_1|R)P(R)}{P(T_2|R \cap H_1)P(H_1|R)P(R) + P(T_2|W \cap H_1)P(H_1|W)P(W)} \\ &\quad + \frac{P(H_2|W \cap H_1 \cap T_2)P(T_2|W \cap H_1)P(H_1|W)P(W)}{P(T_2|R \cap H_1)P(H_1|R)P(R) + P(T_2|W \cap H_1)P(H_1|W)P(W)} \\ &= \frac{\frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{5} + (\frac{1}{2})^3 \times \frac{4}{5}}{\frac{1}{3} \times \frac{2}{3} \times \frac{1}{5} + (\frac{1}{2})^2 \times \frac{4}{5}} = \frac{\frac{47}{370}}{\frac{2}{11}} \approx .53 \end{aligned}$$

Question 2: This question extends the tree and the picture will get complicated unless we do some pruning.

Goal: evaluate

$$P(H_3 \cap H_4|H_1 \cap T_2) = \frac{P(H_1 \cap T_2 \cap H_3 \cap H_4)}{P(H_1 \cap T_2)}$$

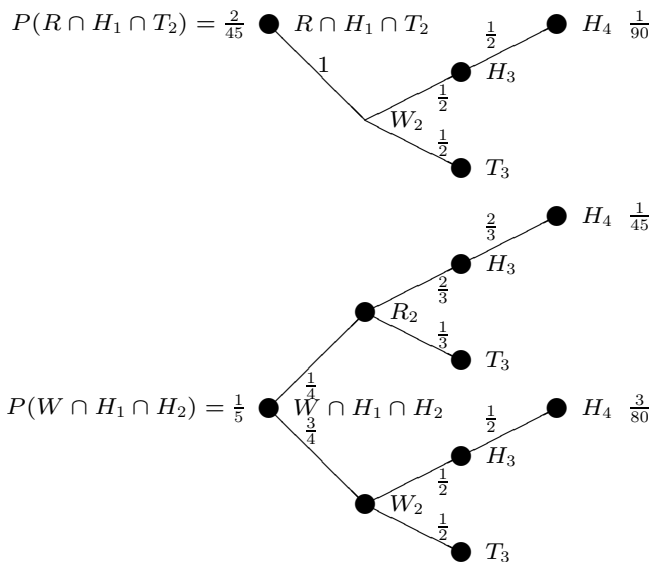
where we have already found $P(H_1 \cap T_2) = 11/45$

Only two nodes of the previous tree are retained when we condition on $H_1 \cap T_2$, these are

- $R \cap H_1 \cap T_2$ with probability $P(T_2|R \cap H_1)P(H_1|R)P(R) = (1/3)(2/3)(1/5) = 2/45$
- $W \cap H_1 \cap T_2$ with probability $P(T_2|W \cap H_1)P(H_1|W)P(W) = (1/2)^2(4/5) = 1/5$

Consider two ways of doing the problem, one will just extend the tree from these nodes and not draw irrelevant stuff. Second way will construct a conditional tree.

First way:



Or, writing it out:

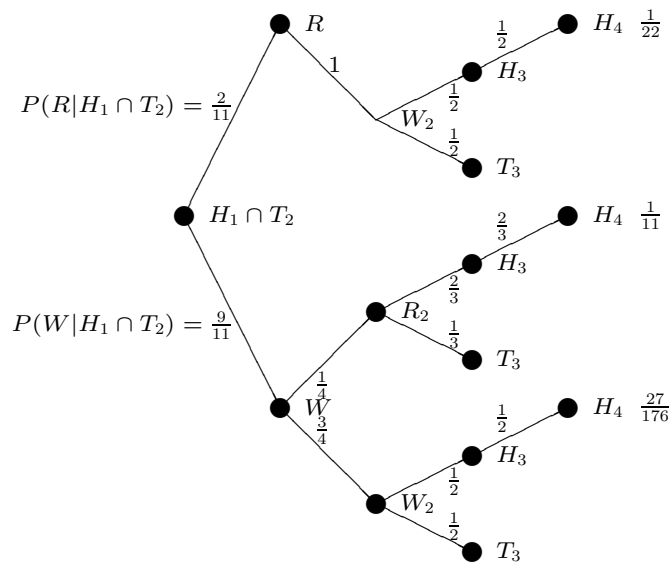
$$\begin{aligned}
 &P(H_1 \cap T_2 \cap H_3 \cap H_4) \\
 &= P(R \cap H_1 \cap T_2 \cap H_3 \cap H_4) + P(W \cap H_1 \cap T_2 \cap H_3 \cap H_4) \\
 &= P(R \cap H_1 \cap T_2 \cap W_2 \cap H_3 \cap H_4) + \\
 &\quad P(W \cap H_1 \cap T_2 \cap R_2 \cap H_3 \cap H_4) + P(W \cap H_1 \cap T_2 \cap W_2 \cap H_3 \cap H_4)
 \end{aligned}$$

So

$$\frac{P(H_1 \cap T_2 \cap H_3 \cap H_4)}{P(H_1 \cap T_2)} = \frac{\frac{1}{90} + \frac{1}{45} + \frac{3}{80}}{\frac{11}{45}} \approx .29$$

Second method: conditional sample space:

Consider conditional tree with root node $H_1 \cap T_2$ with initial branches labeled by conditional probabilities $P(R|H_1 \cap T_2) = (2/45)/(2/45 + 1/5) = 2/11$ and $P(W|H_1 \cap T_2) = (1/5)/(2/45 + 1/5) = 9/11$:



Use total probability and chain rule conditioned on $H_1 \cap T_2$)

$$\begin{aligned}
 &P(H_3 \cap H_4|H_1 \cap T_2) \\
 &= P(R \cap H_3 \cap H_4|H_1 \cap T_2) + P(W \cap H_3 \cap H_4|H_1 \cap T_2) \\
 &= P(R \cap W_2 \cap H_3 \cap H_4|H_1 \cap T_2) + \\
 &\quad P(W \cap R_2 \cap H_3 \cap H_4|H_1 \cap T_2) + P(W \cap W_2 \cap H_3 \cap H_4|H_1 \cap T_2)
 \end{aligned}$$

So

$$P(H_3 \cap H_4|H_1 \cap T_2) = \frac{1}{22} + \frac{1}{11} + \frac{27}{176} \approx .29$$

Mutual Independence

A collection of events F_1, F_2, \dots, F_n is said to be *mutually independent* if for any collection of distinct indices $i_k; k = 1, \dots, m$ with $i_k \in \{1, 2, \dots, n\}$ and $1 \leq m \leq n$,

$$P(\cap_{k=1}^m F_{i_k}) = \prod_{k=1}^m P(F_{i_k})$$

If all the sets have nonzero probability, this is equivalent to

$$P(F_i | F_j \cap F_k \cap \dots \cap F_m) = P(F_i), \text{ all } i \text{ and } j, k, \dots, m \neq i,$$

This implies $P(F_1 \cap F_2 \cap \dots \cap F_n) = P(F_1)P(F_2) \dots P(F_n)$, but not vice versa.

Pairwise independence does not imply mutual independence. E.g., two independent fair coin tosses.

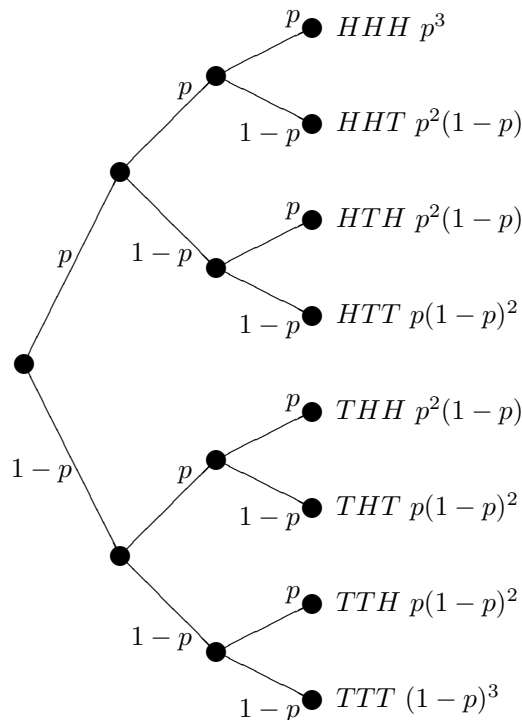
- F : First toss is Head
- G : Second toss is Head
- H : First and second toss have different outcome

F and G are independent, F and H are independent, G and H are independent.

Are F, G, H mutually independent? E.g., does $P(F) = P(F|G, H)$?

Independence makes the sequential approach (multiplication rule) easier:

Example: Independent tosses of a biased coin $P(H) = p, P(T) = 1 - p$.



In a sequential description of an experiment where a collection of outcomes are mutually independent, say that the experiment is a *sequence of independent trials*. If binary as above, we say we have a sequence of *Bernoulli trials*.

Suppose have sequence $F_1, F_2, \dots, F_i, \dots, F_n$ of k events such as $F_i = \{ \text{toss } i \text{ of a coin is a head} \}$. (Or item i on an assembly line is defective, etc.)

What is $\Pr(k \text{ heads in } n \text{ tosses})$?

(or of k successes in n Bernoulli trials)

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

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

In general:

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

binomial probability law

$\binom{n}{k}$ are *binomial coefficients* $n!/k!(n-k)!$

Special case: $n = 1$:

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

Bernoulli probability law

Total Probability

Recall the elementary property of probability we called “total probability”:

If $\{F_i; i = 1, 2, \dots, K\}$ is a finite partition of Ω , i.e., if $F_i \cap F_k = \emptyset$ when $i \neq k$ and $\bigcup_{i=1}^K F_i = \Omega$, then

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

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

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

Comment: This still makes sense even if $P(F_i) = 0$ for some i no matter how we define $P(G|F_i)$, so long as it is a probability.

“Divide and conquer” by carving up complicated event into many simpler events.

Chess Example

- Chess tournament, 3 types of opponents
 - $P(\text{Type 1}) = 0.5$, $P(\text{Win}|\text{Type 1}) = 0.3$
 - $P(\text{Type 2}) = 0.25$, $P(\text{Win}|\text{Type 2}) = 0.4$
 - $P(\text{Type 3}) = 0.25$, $P(\text{Win}|\text{Type 3}) = 0.5$
- What is probability of Win?

Let $W = \text{Win}$, $F_i = \text{Type } i$:

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

Bayes' Rule

- *Objective:* Inference instead of prediction or observation
- Given observed “effect” or “result” (event G), infer the unobserved “cause” (one of the events F_1, F_2, \dots, F_n).
- We assume we know the “prior” or “a priori” probabilities $P(F_i)$ and the conditional probabilities $P(G|F_i)$.
- Wish to compute $P(F_i|G)$, the “posterior” or “a posteriori” probabilities.

$$\begin{aligned} P(F_i|G) &= \frac{P(F_i \cap G)}{P(G)} = \frac{P(G|F_i)P(F_i)}{P(G)} \\ &= \frac{P(G|F_i)P(F_i)}{\sum_j P(G|F_j)P(F_j)} \end{aligned}$$

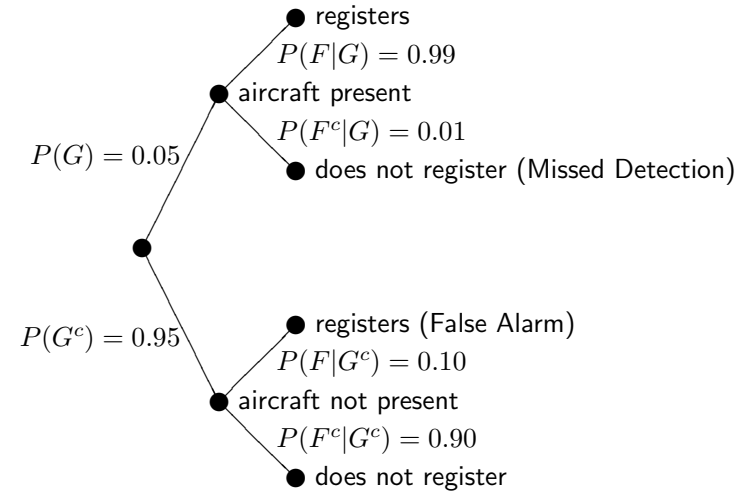
Chess Example Revisited

- Chess tournament, 3 types of opponents
 - $P(\text{Type 1}) = 0.5$, $P(\text{Win}|\text{Type 1}) = 0.3$
 - $P(\text{Type 2}) = 0.25$, $P(\text{Win}|\text{Type 2}) = 0.4$
 - $P(\text{Type 3}) = 0.25$, $P(\text{Win}|\text{Type 3}) = 0.5$
- Given that I win, what is the probability I had a Type 1 opponent?
- Using Bayes' rule,

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

Radar Example

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



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

$$\begin{aligned}
 P(\text{airplane}|\text{register}) &= P(G|F) \\
 &= \frac{P(F|G)P(G)}{P(F|G)P(G) + P(F|G^c)P(G^c)} \\
 &= \frac{0.05 \times 0.99}{0.05 \times 0.99 + 0.95 \times 0.1} \\
 &\approx 0.3426
 \end{aligned}$$