Essential Mathematics for Global Leaders I

Spring 2019

Statistics

Lecture 2: 2019 April 15 -22 - 29

Xavier DAHAN
Ochanomizu Graduate Leading Promotion Center
Office:理学部2号館503
mail: dahan.xavier@ocha.ac.jp

1. Basic Probability 確率論

1.2 Probability rules 確率

Probability cast (配役)

- Experiment: a repeatable undetermined procedure.
  実験（試行）：繰り返せる、決定されていない手順。
- Sample space: Set of all possible outcomes Ω.
  標本空間：全ての可能な結果の集合。
- Event: a subset of the sample space.
  事象：標本の部分集合。
- Probability function, $P(\omega)$: gives the probability for each outcome $\omega \in \Omega$.
  それぞれの結果に確率を割り当る。
  1. Probability is between 0 and 1.
  2. Total probability of all possible outcomes is 1.

Example:

- Experiment: toss a fair coin, report heads or tails.
  実験：コイン投げ。表か裏かが出るのを報告する。

- Sample space (標本空間）: $\Omega = \{H, T\}$.

- Probability function: $P(H) = 0.5, \;\; P(T) = 0.5$.

- Another experiment: toss a coin 2 times.
- Sample space: $\Omega = \{HH, HT, TH, TT\}$
- Probability function (using table):

Outcomes	*HH*	*HT*	*TH*	*HH*
Probabilty	$1/4$	$1/4$	$1/4$	$1/4$

理解を確かめる Events, sets and words

- Experiment: toss a coin 3 times.

- Which of the following is the event “exactly two heads”?

$a.\quad A = \{THH, HTH, HHT, HHH\}$

$b.\quad B = \{THH, HTH, HHT\}$

$c.\quad C = \{HTH, THH\}$

(1) A (2) B (3) C (4)A or B

Answer:

理解を確かめる : Events, sets and words

- Experiment: toss a coin 3 times.

- Which of the following describes the event $\{THH, HTH, HHT\}$

1. “exactly one head” 表がちょうど１回
2. “exactly one tail” 裏がちょうど１回
3. “at most one tail” 裏が多くても１回
4. None of the above　どれでもない

Answer:

理解を確かめる Events, sets and words

- Experiment: toss a coin 3 times.

- The events “exactly 2 heads” and “exactly 2 tails” are disjoint.
  (。。。事象互いに素=。。。交わりを持たない)
  (1) True (2) False

- **Answer:**

- The event “at least 2 heads” implies the event “exactly two heads”.
  (1) True (2) False

- **Answer:**

Probability rules in maths notations
数学的な記号での確率法

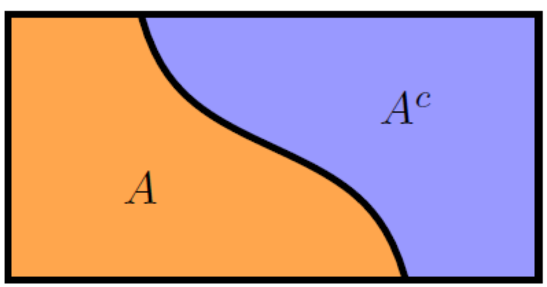
- Sample Space （標本空間）: $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$.
- Outcome (結果）: $\omega \in \Omega$.
- Probability between 0 and 1: $0 \leq P(\omega) \leq 1$

- Total probability is 1:
  $\sum_{j=1}^{n} P(\omega_j) = 1, \qquad \sum_{\omega \in \Omega} P(\omega) = 1.$

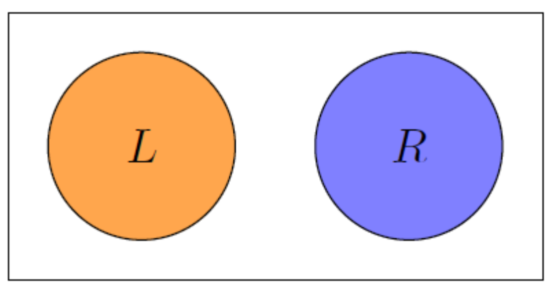
- Event (事象） $A \subset \Omega$: $P(A) = \sum_{\omega \in A} P(\omega)$.

Probability and set operations on events
確率と事象への作用

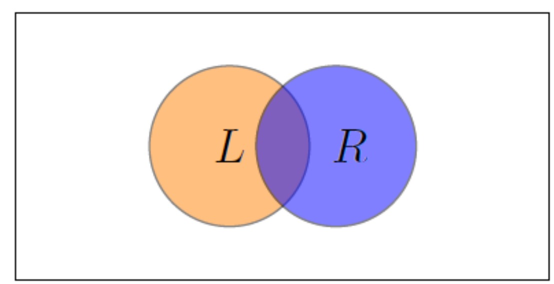
- Three events $A, L, R \subset \Omega$ (Ω: sample space)
- Rule 1: Complements （補集合） $P(A^c) = 1 - P(A)$.
- Rule 2: Disjoint events. （互いに素である事象）
  - If L and R are disjoint then
$$P(L \cup R) = P(L) + P(R).$$
- Rule 3: Inclusion-Exclusion principle （包除原理）
  - For any L and R:
$$P(L \cup R) = P(L) + P(R) - P(L \cap R).$$



Rule 1



Rule 2



Rule 3

Rule 1: example of application

- You roll a 20-sided die 9 times.
  正２０面体のさいころを9回投げる。
- Event A: there is a match among the 9 outcomes.
  出た九つの結果の中で二つ以上は一致するという事象をAとする。
- Questions:

1. For this experiment how would you define the sample space, probability function, and event?
   この実験では、標本空間、確率関数と事象をどのように定義するか？
2. Evaluate the exact probability $P(A)$ that the event A occurs.
   事象Aが起こる確率$P(A)$を評価せよ。

Answer

- Sample space Ω: Sequences of 9 numbers between 1 and 20.
  １から２０間の長さ９の数列
- Cardinal of Ω: （Ωの個数） Use the rule of product !
  $$|\Omega| = 20^9$$
- It is not very easy to count directly the nbr. of occurrences of the event A …because there might be 2, 3, 4 … or 9 rolls that are equal.
  直接事象Aの確率を評価するのは難しい。さいころの結果が２つ、３つ、…、９つまで等しいである可能性があるから。

- But it is much easier for the complement A^c.
  補集合の事象A^cを計算するほうが簡単である。
- A^c: there is no match （等しい結果が無い）
- $|A^c| = 20 \cdot 19 \cdot \cdots \cdot 12 = P_{11}^{20} = 20!/11!$.
- $P(A) = 1 - P(A^c) = 1 - \frac{|A^c|}{|\Omega|} \approx 0.881$.

Rule 2: Example of application

- A group of 50 mice　（ネズミの５０匹）
- 20 have white hairs (W), and 25 have black eyes (B).
  毛が白い２０匹 (W)、目が黒い２５匹 (B)

- For a randomly chosen mouse, what is the range of possible values $p = P(W \cup B)$?
  ランダムに選択されたネズミに対して、確率$p = P(W \cup B)$の可能な値の範囲はなにか？

  a. $p \leq .4$
  b. $.4 \leq p \leq .5$
  c. $.4 \leq p \leq .9$
  d. $.5 \leq p \leq .9$
  e. $.5 \leq p$

Answer

- **Common sense (常識を使う):** there are at least 25 mice in $W \cup B$ and at most 45. Therefore the probability is:

$$\leq p \leq$$

- **Or use the inclusion-exclusion principle:**
  また原理によると：
- $P(W \cup B) =$
- Extreme cases are:
- Giving: $\leq P(W \cup B) \leq$

More difficult example with uncertainty
不確定性を持つより複雑な例

- Lucky Lucy has a coin that you're quite sure is not fair.
  ラッキー・ルーシーはコインの一つを持って、あなたは公平だと信じない。
  - She will flip the coin twice （コインを２回投げる）
  - It's your job to bet whether the outcomes will be the same *(HH, TT)* or different *(HT, TH)*.
    結果の両方が同じ*(HH,TT)*か異なる*(HT,TH)*かに賭ける

- Which should you choose?
  a. Same
  b. Different
  c. It doesn't matter, same and different are equally likely

- Hint: Use the probability of heads p (and a little of algebra)
  表の確率をpとし、基礎的な代数を使って解ける。

Solution (解決)

Answer: a) same (same is more likely than different)

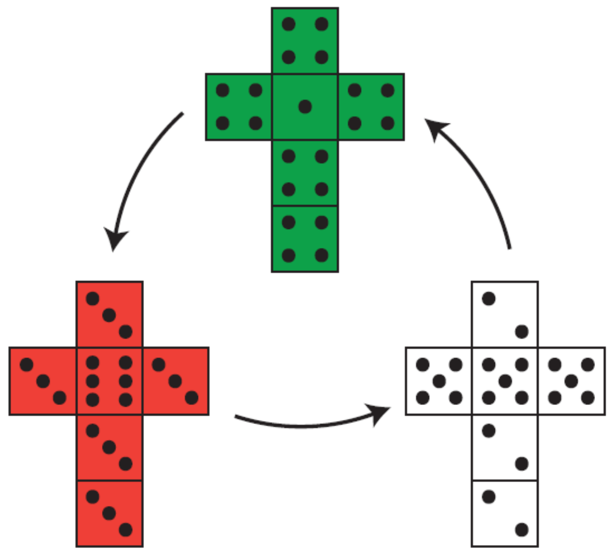
- $P(H) = p$ thus $P(T) = (1 - p) = q.$
- Since the flips are independent (Cf. Section 1.3) the probabilities multiply. This gives the following 2x2 table.
  コイントスは独立であるから、確率がかかる。以下の２行２列表を与える。

	H	T (2nd flip)
1st flip H	p^2	pq
T	pq	q^2

- $P(same) =$, $P(diff) =$.
- Algebra: $(p - q)^2 =$ $= P(same) - P(diff) > 0$
- So $P(same) > P(diff)$ if $p \neq q$.

Homework: Mari’s dice

- Mari has three six-sided dice with unusual numbering.
  マリは異常な番号付けのある６面さいころを持つ。



- A game consists of two players each choosing a dice.
  二人のプレヤーのゲームではさいころを選んで投げる。
  They roll once and the highest number wins.
  最大の結果の者が勝つ。
- **Question**: Which dice would you choose?

Hints:

A. You can proceed as follows for white and red.
   1. For red and white dice make the probability table.
      白と赤のさいころの確率表を作成する。
   2. Make a prob. table for the product sample space of red and white.
      赤の標本空間と白の標本空間の直積から確率表を作成する。
   3. What is the probability that red beats white?
      赤が白を負かす確率はなにか？

B. Then do the same for white and green dice.
C. And for the green and red dice as well.
D. Conclude which dice is the best for the game.

Answer to question A

- Red beats white is the event $(R > W)$:
- The answer is $P(R > W) = 7/12$. Find it by fulling the tables below.
- Probability table for white and red.

	Red die	
Outcomes	3	6
Probability	5/6	1/6

	White die	
Outcomes	2	5
Probability	1/2	1/2

- Both R and W outcomes in a 2x2 probability table:

		2	5	White
Red	3	5/12	5/12	
	6	1/12	1/12	

- Red entries: outcomes where red beat white.
- $P(R > W) = \frac{5}{12} + \frac{1}{12} + \frac{1}{12} \qquad = \frac{7}{12}.$

PART I. Notions of Probability 必要な確率論

1. Basic probability 確率論基礎

1.1 Counting

1.2 Probability rules

1.3 Conditional Probability, Independence, Bayes theorem

条件付き確率、独立性、ベイズの定理。

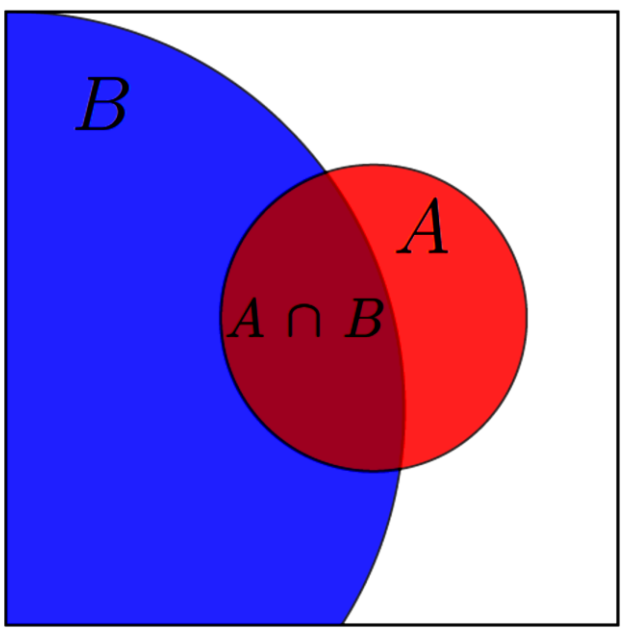
- Law of total probability 総確率の法律
- Use of trees to represent case distinction
  場合分けを表現するように木を使う
- Independence 独立性
- Bayes theorem. Base rate fallacy.　ベイズ定理。基準率無視

1: Basic Probability 確率論基礎

1.3 Conditional probability, independence, Bayes theorem. 条件付き確率、独立性、ベイズの定理

'The probability of A given B' (事象BにおけるAの確率)

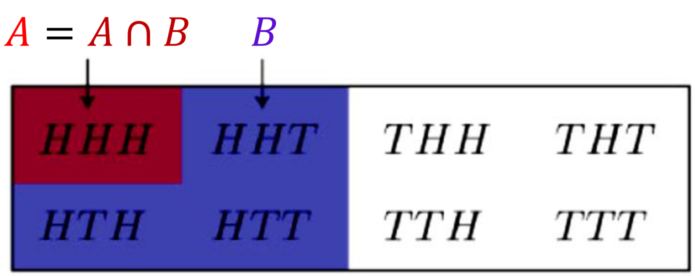
$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \text{provided(ただし)} \quad P(B) \neq 0.$$



Example of 3 coin tosses:

$B = \{HHH, HHT, HTH, HTT\}$

$A = \{HHH\}$.



Exercise

- Toss a coin 4 times. コインを４回投げる。Let the events:
  - A =‘at least three heads’ (表が少なくとも３回出る)
  - B =‘first toss is tails’ （最初の結果は裏である）

1. What is $P(A|B)$?
   a) 1/16 b) 1/8 c) 1/4 d) 1/5.

2. What is $P(B|A)$?
   a) 1/16 b) 1/8 c) 1/4 d) 1/5.

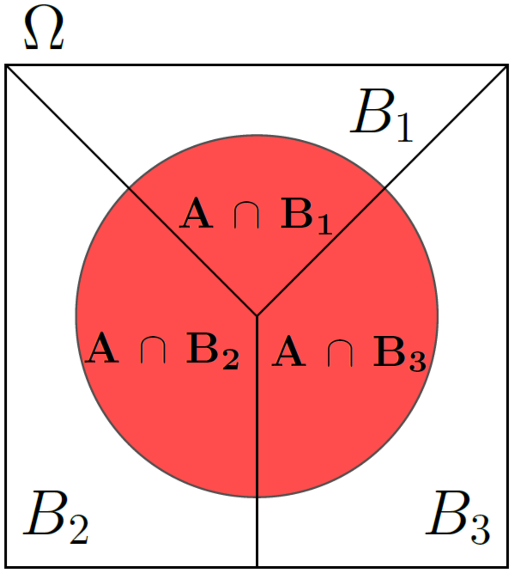
- **Answer:** 1. (b) 1/8 2. (d) 1/5.
  Counting: $|A| = 5$, $|B| = 8$ and $|A \cap B| = 1$. Since all sequences of 3 tosses are all equally likely

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|}{|B|} = \frac{1}{8}, \qquad P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{|A \cap B|}{|A|} = \frac{1}{5}.$$

Multiplication rule, Law of total probability

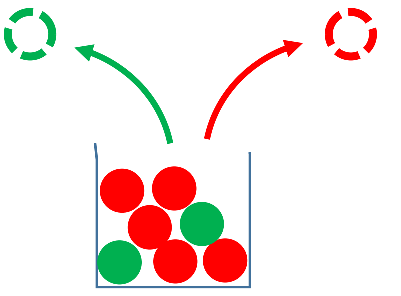
- Multiplication Rule: $P(A \cap B) = P(A|B)\,P(B)$.
- Law of Total Probability (easy case of 3 events)
  全確率の法則（事象が３つある場合）
- If B_1, B_2, B_3 is a partition （分割） of Ω then

- $P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$
  $= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$.



First example: an urn (壺) problem

- An urn contains 5 red balls and 2 green balls.
  壺に赤いボールが五つ、緑ボールが２つある。
  Two balls are drawn one after each other.
  ボールを一つづつ、２個抽出する。



- **Question:** What is the sample space Ω for this experiment?
- $\Omega = \{GG, RR, GR, RG\}$.
- Let R_1 be the event 'the 1st ball is red',
  R_2='the 2nd ball is red',
  G_1 ='the 1st ball is green' and
  G_2 ='the 2nd ball is green'.
- **Question:** What is $P(R_2)$?

Answer

- **Use intuition(but be careful)**: Every of the 7 balls is equally likely to be the second ball.
直感を使う(しかし注意せよ）：各７個のボールは同程度に２番目のボールになりうる。

$$P(R_2) = \frac{5}{7}.$$

- **Use conditional probabilities:** （条件付き確率を使う）

$$P(R_2|R_1) = \frac{4}{6}, \qquad P(R_2|G_1) = \frac{5}{6},$$

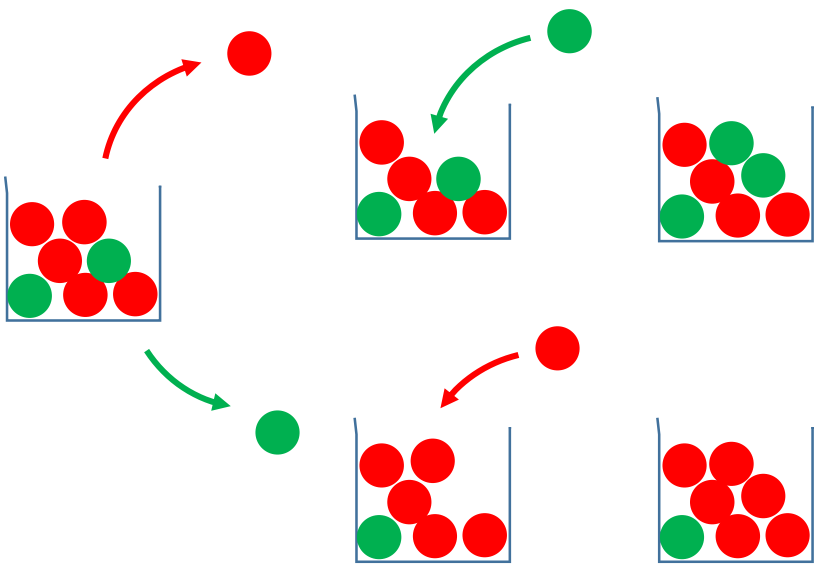
- By the total law of probability: （全確率の法則）

$$P(R_2) = P(R_2|R_1)P(R_1) + P(R_2|G_1)P(G_1)$$
$$= \frac{4}{6} \cdot \frac{5}{7} + \frac{5}{6} \cdot \frac{2}{7} = \frac{20}{42} + \frac{10}{42} = \frac{30}{42} = \frac{5}{7}.$$

Trees for conditional probability
条件付き確率を表現するための木

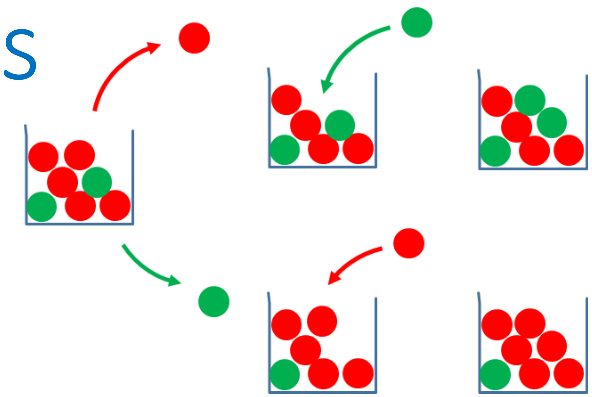
- Organize computations 計算をまとめる。
- Compute total probabilities 全確率を計算するため
- Compute Bayes formula

- **Example:** Game: 5 red and 2 greens in an urn (壺)



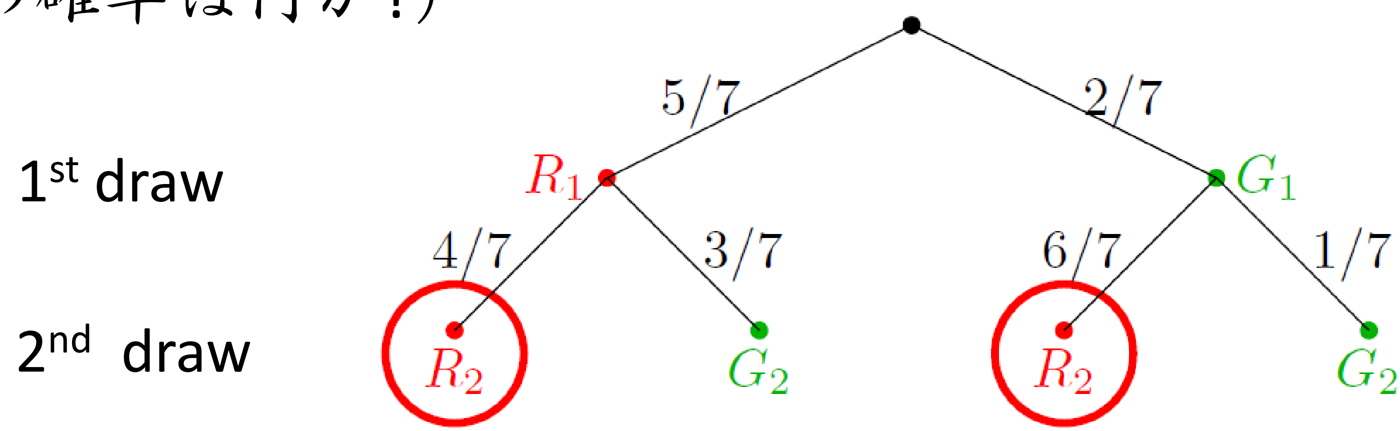
1. A random ball is selected ランダムにボールが選択され、
2. It is replaced by a ball of the other color 他の色のボールに取り換える。

Trees



1. A random ball is selected
   ランダムにボールが選択され、
2. It is replaced by a ball of the other color
   他の色のボールに取り換える。

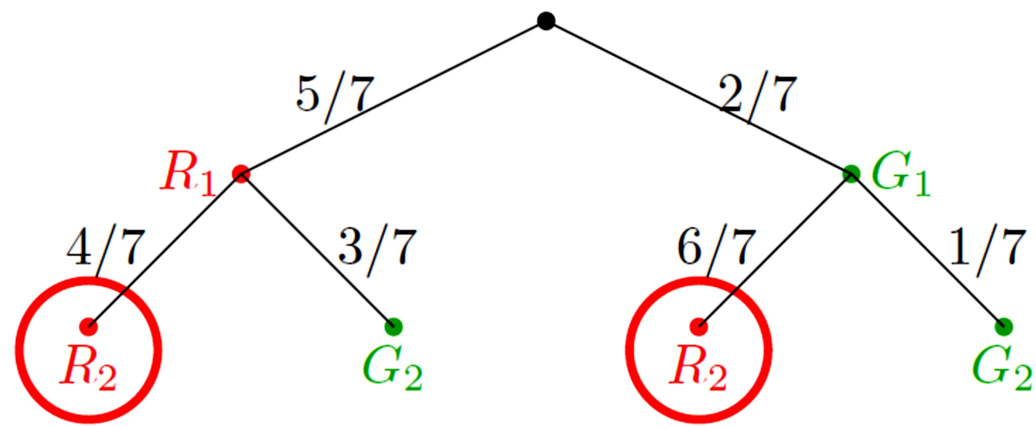
1. What is the probability that the 2nd draw （取出し） is red?
2. What is the probability that the 1st draw was red given the 2nd draw was red? （２番目の取出しが赤い場合における１番目の取出しが赤いという事象の確率は何か?）



1. What is the probability that the 2^{nd} draw（取出し）is red?
2. What is the probability that the 1^{st} draw was red given the 2^{nd} draw was red? （２番目の取出しが赤い場合における１番目の取出しが赤いという事象の確率は何か?）

1^{st} draw

2^{nd} draw



1. Total law of probability gives

$$P(R_2) = P(R_2 \cap R_1) + P(R_2 \cap G_1)$$

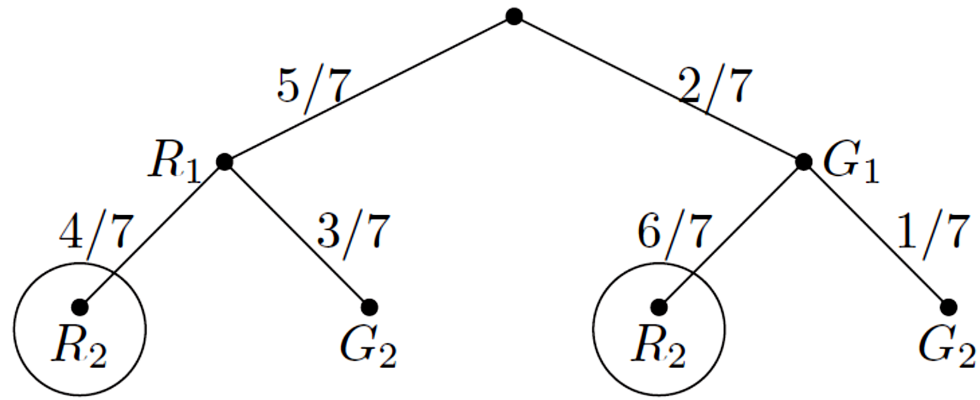
$$= P(R_2|R_1)P(R_1) + P(R_2|G_1)P(G_1) = \frac{5}{7} \cdot \frac{4}{7} + \frac{2}{7} \cdot \frac{6}{7} = \frac{32}{49}$$

2. Bayes rule gives:

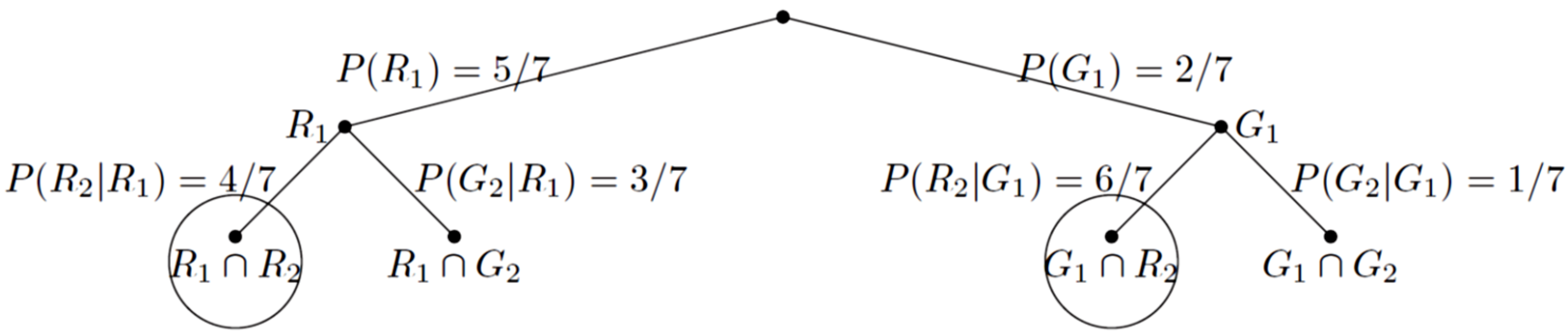
$$P(R_1|R_2) = \frac{P(R_1 \cap R_2)}{P(R_2)} = \frac{P(R_2|R_1)P(R_1)}{P(R_2)} = \frac{20/49}{32/49} = \frac{20}{32}$$

Shorthand notation for tree（略記法）

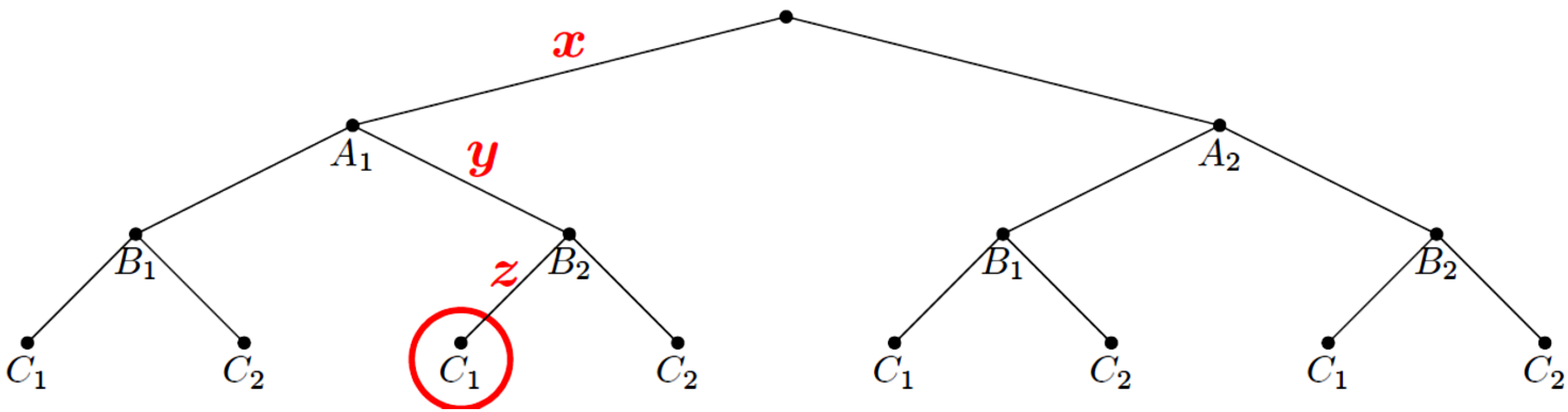
- In the previous example, the tree is a shorthand notation (略記法)



and the following tree is the complete notation.



Understanding tree 1



1. The probability x represents

a) $P(A_1)$

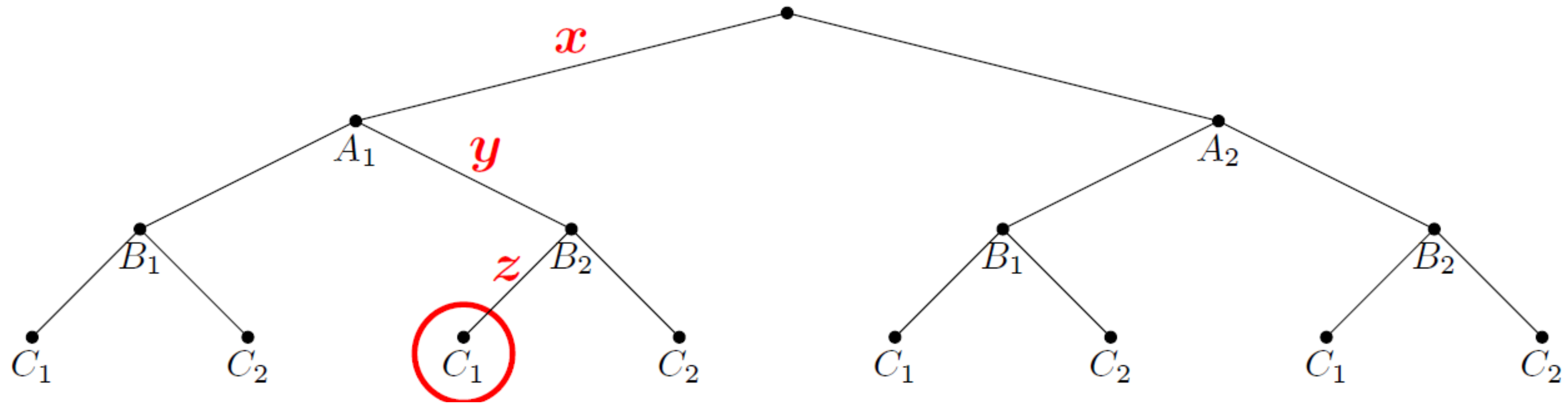
b) $P(A_1|B_2)$

c) $P(B_2|A_1)$

d) $P(C_1|B_2 \cap A_1)$

- <u>answer:</u> (a) $P(A_1)$

Understanding tree 2



2. The probability y represents

a) $P(B_2)$

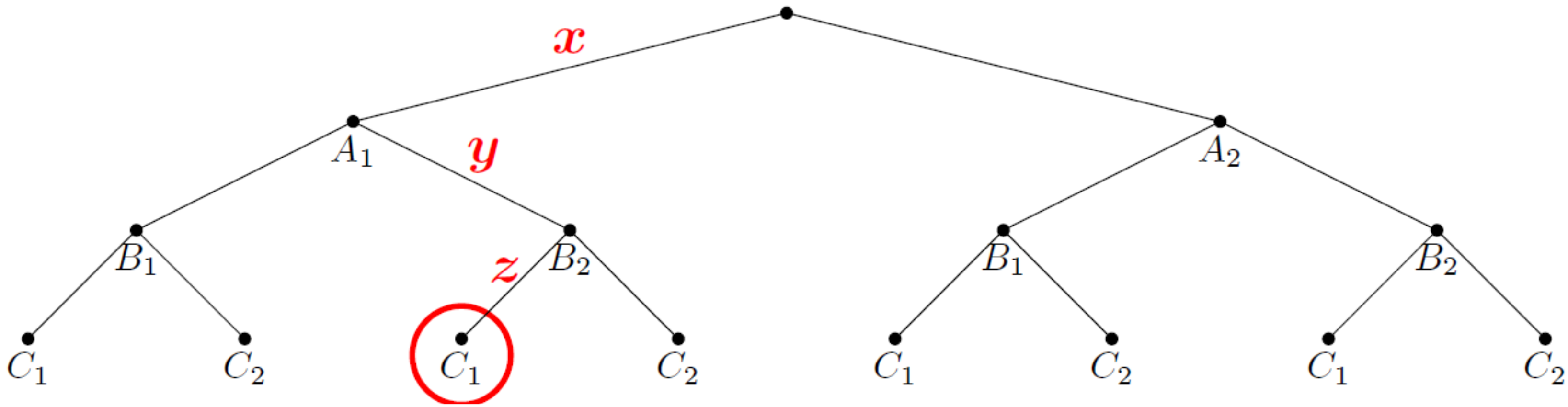
b) $P(A_1|B_2)$

c) $P(B_2|A_1)$

d) $P(C_1|B_2 \cap A_1)$

- <u>answer:</u> (c) $P(B_2|A_1)$

Understanding tree 3



3. The probability z represents

a) $P(C_1)$

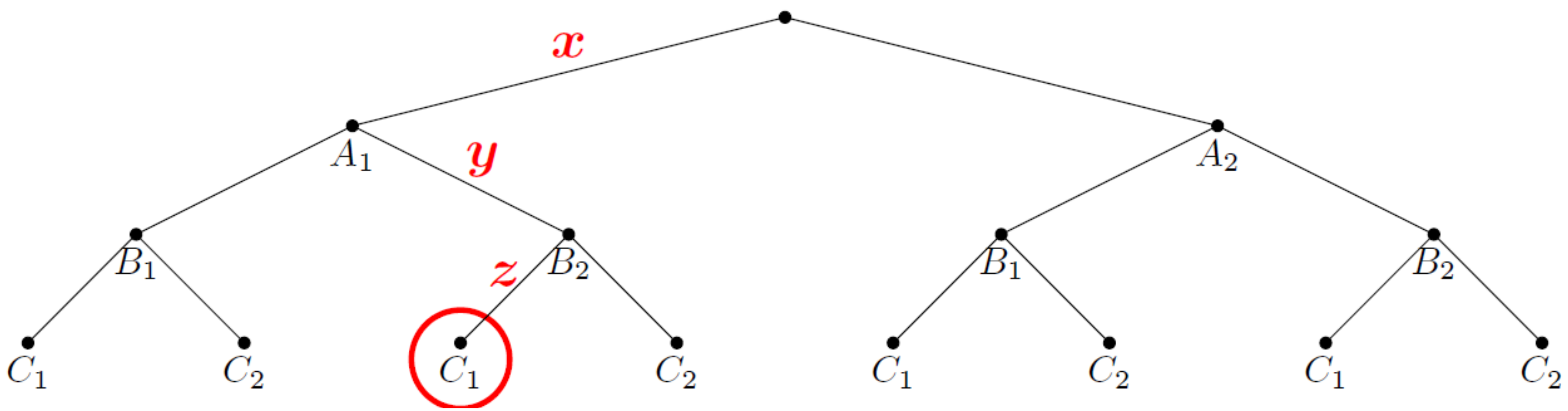
b) $P(B_2|C_1)$

c) $P(C_1|B_2)$

d) $P(C_1|B_2 \cap A_1)$

- <u>answer:</u> (d) $P(C_1|B_2 \cap A_1)$

Understanding tree 4



4. The circles node represents the event

a) C_1

b) $B_2 \cap C_1$

c) $A_1 \cap B_2 \cap C_1$

d) $C_1 | B_2 \cap A_1$

- <u>answer:</u> c) $A_1 \cap B_2 \cap C_1$

Example: Monty Hall problem

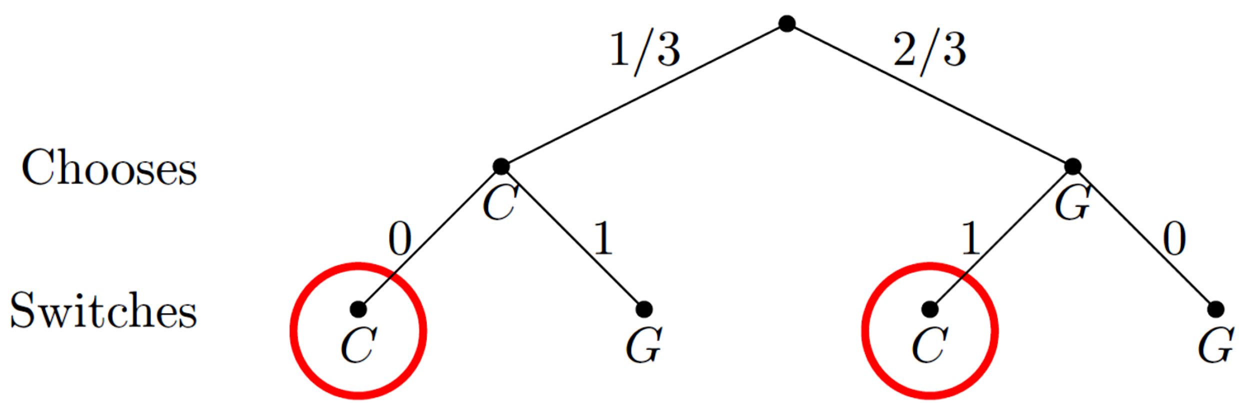
- Monty Hall is an American TV host in the 1970s and presented a TV game show called “Let’s make a deal”.

1. One door hides a car,
   two doors hide a goat.
   ひとつのドアの後ろに車があり、
   他の二つのドアにヤギがいる。
2. Contestant chooses a door.
   出場者はドアを選ぶ。
3. Monty (who knows where is the car) opens a different door with a goat.
   Montyは他の二つのドアの中、ヤギのあるドアを開く
4. Contestant can switch doors or keep the original choice.
   出場者は選んだドアを変えるか、そのままにするか。

Question:
To win a car: (a) switch, (b) don’t switch, (c) doesn’t matter?

Monty Hall problem

- Organize the Monty Hall problem into a tree and compute the probability of winning if you always switch.
木を用いてこの問題を解決し、ドアを変えるときに車を得る確率を評価せよ。
- You can use the events:
C='you chose the car'
G='he shows a goat' *switch*='Switch door'
- **Answer:** Switch! $P(C|switch) = 2/3$.



- $P(C|swi$…

Independence (独立性)

Two events A and B are independent if the probability that one occurred is not affected by knowledge that the other occurred.
もしも事象AとBの間、A(同様にB)の起こる確率はB(同様にA)が起こったかどうかを知っても知らなくても影響を受けないとき、

AとBは独立という。

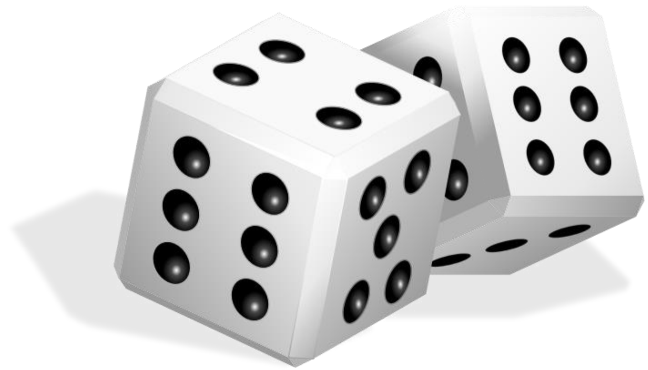
同様な定義：

$$P(A \cap B) = P(A)P(B)$$
$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Independence: example

Roll two dice.

- A = ‘first dice is 3’
- B = ‘sum of two dice is 6’



- A and B are independent
  a) True b) False

- $P(A) = 1/6$ $P(A|B) = 1/5$
- $P(B) = 5/36$ $P(B|A) = 1/6$
- $P(A \cap B) = 1/36$ $P(A)P(B) = 5/216$

Bayes theorem（ベイズの定理）

- Also called "Bayes rule" or "Bayes formula"
- Allows to find $P(A|B)$ from $P(B|A)$ if $P(B) \neq 0$.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- The denominator $P(B)$ is often computed using the total of probability
  分母$P(B)$を計算するにあたり、全確率の法則がよく使われている。

$$P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$$

Example: False positive in a screening test

- Screening test for a disease.
  疾患を検出するスクリーニングテスト
- Frequency of the disease in the population (base rate): 0.5%
  人口における疾患の頻度：0.5%
- Test is quite accurate: 5% of false positive
  10% of false negative
  テストはかなり正確である：偽陽性率は5%、
  偽陰性率は10%.

Question: You take the test and it is positive. What is the probability that you have the disease?
テストを受けて結果は陽性である。疾患を持つ確率は何か?

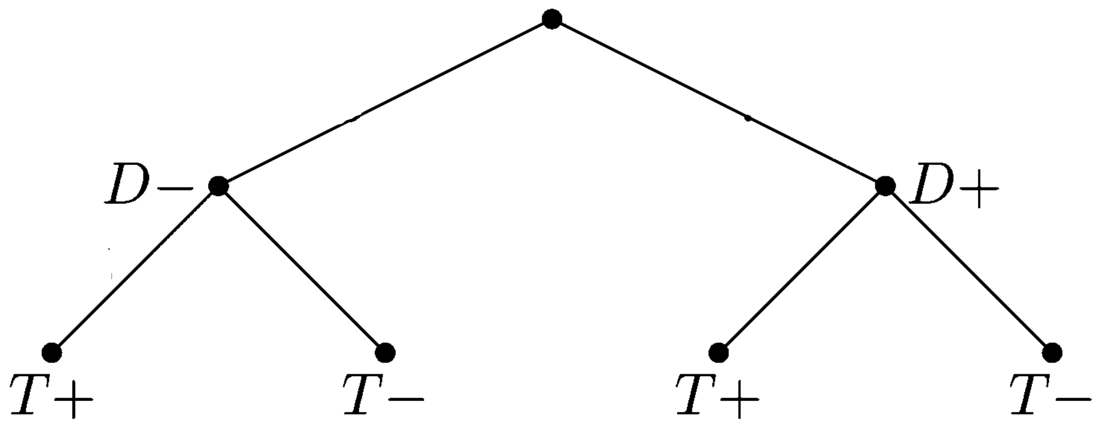
Example: False positive in a screening test

- Use the following events and build a tree:
  以下の事象を使って、関連の木を立てよ。
  - D+ = 'you have the disease'　（疾患を持つ）
  - D− = 'you do not have the disease'（疾患を持たない）
  - T+ = 'you tested positive'　（テスト陽性）
  - T− = 'you tested negative'.　（テスト陰性）

- What is the probability that we must compute ?

- What probability corresponds to: $P(D+\mid T+)$
  - Disease frequency 0.5%:
  - False positive 5%:
  - False negative 10%:

$$P(D+) = 0.005$$
$$P(T+|D-) = 0.05$$
$$P(T-|D+) = 0.1$$



- Fill all the edges with probabilities.
  辺の全てを確率で入力せよ。

- Use Bayes formula (and the total law of probability) to compute the answer!

$$P(D+|T+) = \frac{P(T+|D+) \cdot P(D+)}{P(T+)}$$

- $P(T+) = P(T+|D-)P(D-) + P(T+|D+)P(D+)$
  $= .995 \times .05 + .005 \times .9 = .05425$

- $P(D+|T+) = \frac{.005 \times .9}{.05425} = 0.082949 \approx 8.3\%$

Base rate fallacy　基準率無視

- The test is accurate but you have only 8.3% of chances to have the disease if the test is positive.
テストは正確だが、陽性ならば疾患を持つ可能性は8.3%だけである。

95% of accuracy doesn't mean that that you have 95% to have the disease.
95%の正確さだと、疾患を持つ可能性が95%だとは言えないだろう。

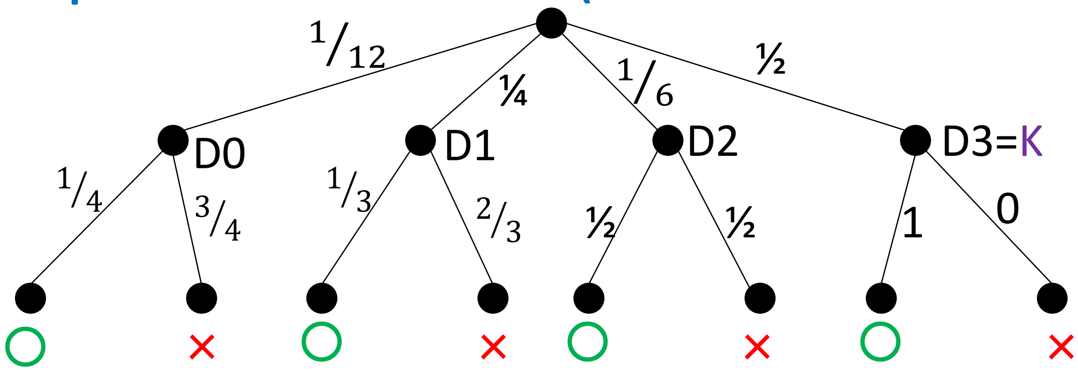
- This is because the disease frequency 0.5% (base rate) is so low that the overwhelming of people taking the test are healthy, so even a small rate of false positive represents a lot of people.
これは、疾患の頻度0.5%（基準率）がとても低いためにテストを受ける人が圧倒的に健康だという理由によるもので、偽陽性率が小さくても大勢の人を含めてしまう。

It is at the heart of many statistical misunderstandings.
これは統計に関する様々な誤解の由来である。

Tree: Example 3: MCQ (多肢選択問題)

- 4 choices per question. 問いごとに選択肢が4つある。
- Probability that you know the answer is ½.
  正答がわかる可能性は1/2.
- If you don’t know the correct answer, there are 1 chance out of 4 that you can discard one wrong answers, 1 chance out of 6 that you can discard two wrong answers.
  正答がわからない場合は、不正解の選択肢一つを捨てられる可能性は1/4で、不正解の2つを捨てられる可能性は1/6である。
- **Question:** What is the probability that you know the answer given that you answer correctly.
  正確に答えたことにおける正答がわかった条件付き確率はなにか。

Example 3: MCQ (多肢選択問題)



- D0, D1,D2… = 'I can discard 0,1,2… wrong answers'
- K=D3='I know the answer' （正答がわかる=他の３つの選択肢は不正解だとわかる）
- ○、 ×='correct answer', 'wrong answer'
- Question: $P(K|○)$?
- $P(○|K) = 1.$
- $P(○) = \frac{1}{12}\frac{1}{4} + \frac{1}{3}\frac{1}{4} + \frac{1}{2}\frac{1}{6} + 1\frac{1}{2} = \frac{1}{48} + \frac{1}{12} + \frac{1}{12} + \frac{1}{2} = \frac{33}{48}$
- $P(K|○) = {}^{1}/_{2}/{}^{33}/_{48} = 24/33.$

Homework (about Section 1.2)

1. Finish the problem “Mari’s dice” page 12 & 13
   B. Then do the same for white and green dice.
   C. And for the green and red dice as well.
   D. Conclude which dice is the best for the game.

2. Boys and girls “paradox”
   a) Ms. Jones has two children. The older child is a girl. What is the probability that both children are girls?
   Jonesさんは子どもが二人、長子は女の子である。子どもの二人とも女の子である確率は何か？
   b) Ms. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?
   Smithさんは子どもが二人、少なくとも一人の男の子がいる。子どもの二人とも男の子である確率は何か？

Homework (about Section 1.3)

3. Base rate fallacy (基準率無視): The blue taxi

- There are 100 taxis in a city : 99 Green, and 1 Blue.
- A witness observes a hit-and-run by a taxi at night.
  He recalls that the taxi is blue.
  夜にひき逃げ事故を起こしたタクシーを目撃した人がいる。記憶によると、目撃者はタクシーがBlueだと思う。
- The blue taxi driver claims his innocence.
  Blueタクシーの運転手は無実を主張する。
- Experimental data suggest that the witness sees a blue taxi as blue 99% of time, and a green taxi as blue 2% of the time.
  実験データによると、目撃者が夜にBlueタクシーをBlueのように見る確率は99%で、GreenタクシーをBlueのように見てしまう確率は2%.

→ Can we say that the driver of blue taxi is guilty?
Blueタクシーの運転手は有罪と言えるか？

Homework (about Section 1.3)

Independence

4. Suppose that $P(A) = 0.4,\ P(B) = 0.3$ and $P\big((A \cup B)^c\big) = 0.42$. Are A and B independent ?

5. Suppose that the events A, B, C are *mutually independent (互いに独立)* with:
   $P(A) = .3, P(B) = .4, P(C) = .5.$
   Compute (Use a Venn diagram!)
   i. $P(A \cap B \cap C^c)$
   ii. $P(A \cap B^c \cap C)$
   iii. $P(A^c \cap B \cap C)$