條件机率 Conditional Probability & Bayes' Rule

new information

- **Q**: Should the following probabilities be different?
  - Event = a pitcher wins at least 16 games in a season (same event)
    - changed
    - P=?? in the beginning of the season
    - P=?? in the middle of the season
    - P=?? close to the end of the season
  - Event = rain tomorrow (same event) — Q: how to evaluate them?
    - changed
    - P=?? if no information about where you are staying
    - P=?? if you are staying in a desert
    - P=?? if a typhoon will hit the place you stay tomorrow
- **Q**: What causes the differences?
  - For an event, *new information* (i.e., some other event has occurred) could change its probability
  - We call the altered probability a conditional probability — Recall. subjective probability

- Mathematical Definition: If A and B are two events in a sample space Ω and $P(A)>0$, then (event occurred: A; event of interest: B)

reasonable to define for A with P(A)=0?

$$P(B|A) \equiv \frac{P(A\cap B)}{P(A)}$$

is called the *conditional probability* of B given A.

codtioned on continuous random variables (future lecture)

- In the classical probability, (symmetric outcomes in Ω)

$$P(A) = \#A/\#\Omega \quad \text{and} \quad P(A\cap B)=\#(A\cap B)/\#\Omega$$

$$\Rightarrow P(B|A) = \frac{\#(A\cap B)/\cancel{\#\Omega}}{\#A/\cancel{\#\Omega}} = \frac{\#(A\cap B)}{\#A} \quad \text{cf.} \quad P(B) = \frac{\#B}{\#\Omega}$$

~~$A\cap B$~~

$P(B) \xrightarrow{\text{update}} P(B|A)$

- Example: A family is known to have 2 children, *at least one of whom is a girl* (A). **Q**: Probability that the other is a boy=?? ($P(B|A)$)
  - $\Omega = \{bb, bg, gb, gg\}$ — symmetric outcomes
  - $A=\{bg, gb, gg\}$ and $B=\{bb, bg, gb\}$ — at least one boy — $P(B)=3/4$
  - $P(B|A) = \#(A\cap B)/\#A = 2/3$ — update
- Note: $\#\Omega$ is reduced to $\#A$.

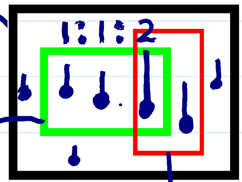
- In effect by conditioning,
  - we are restricting the sample space from Ω to A, i.e.,

$$\Omega \rightarrow A,$$

  - and, for an arbitrary event B (in Ω) to occur when A has occurred, we need that both A and B occur together, i.e.,

$$B \rightarrow B \cap A.$$

Ω (original sample space)

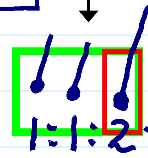


A

B

new information A occurs

A (new sample space)



$A \cap B$

- The division by $P(A)$ in the definition above rescales all probabilities from the entire sample space Ω to being relative to the new sample space A ($\because$ P(sample space) = 1)

- $P(B|A)$ is a probability measure defined on B, $\boxed{\text{but not } A}$. —(*)

a subset of Ω

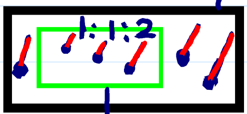
It can be regarded as a probability measure defined on 2^Ω

  - $P(\cdot|A)$ satisfies the 3 axioms of probability (exercise)

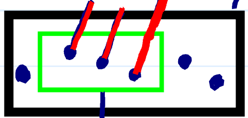
can substitute any events of Ω

check textbook Sec. 3.5 for details

$P(\cdot)$



$P(\cdot|A)$



  - Any propositions developed in Chapter 2 for probability measures can be applied on $P(\cdot|A)$. (For example, $P(B^c|A)=1-P(B|A)$.) cf. $P(B|A)+P(B|A^c) \neq 1$ in general ((*) in LN p. 4-3)

$\because$ $P(B|A)$ & $P(B|A^c)$ might be easier to evaluate

- 3 Useful Formulas for Calculating Probabilities (for 2-events case)

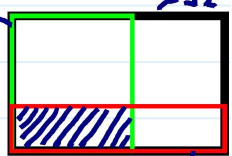
  1. If $P(A)>0$, then $P(A \cap B) = P(A)\, P(B|A)$. —①

order not matter $= P(B)P(A|B)$ if $P(B)>0$.

to garantee the cond. prob. well defined

proof. By the definition of cond. prob., $P(B|A) = P(A\cap B)/P(A)$.

meaning?



  2. If $0<P(A)<1$, then

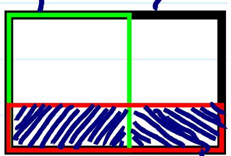
weighed average

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c). \quad ②$$

proof. $B = B\cap\Omega$
$= B\cap(A\cup A^c)$
$= (B\cap A)\cup(B\cap A^c)$

mutually exclusive

$\| \; P(B\cap A)$ by ① $P(B\cap A^c) \; \|$



  3. If $0<P(A)<1$ and $P(B)>0$, then

$P(\cdot) \xrightarrow[\text{B occurs}]{\text{update}} P(\cdot|B)$

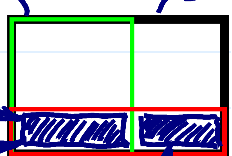
$$③ \quad P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}.$$

proof. $\| $

$P(A\cap B)$ — by ① → 分子

$P(B)$ — by ② → 分母

Note. In ② & ③, A & A^c : a partition of Ω



➤ Example (Urn Problem) ← prototype of many applications

- The Story. n balls sequentially and randomly chosen, *without replacement*, from an urn containing R red and $N-R$ white balls ($n \le N$). **Q**: Given that k of the n balls are red (#A∩B) ($k \le R$), probability that the 1st ball chosen is red = ??

do not put the drawn balls back into the urn

- Let $A=\{k \text{ of the } n \text{ balls are red}\}$ (#A)

  $B=\{1^{\text{st}} \text{ ball chosen is red}\}$

  symmetric outcomes

  ball 1 (r), ball R (r), ball R+1 (w), ball N (w); 1st draw, …, nth draw (#Ω)

  1st draw, nth draw: {□, …, □} k positions for red balls

  ~~1st draw~~ 2nd draw, nth draw: {□, …, □} k-1 positions for red balls

- Method 1: $P(B|A) = \binom{n-1}{k-1} \Big/ \binom{n}{k} = k/n$

  meaning? **Q**: Why different #A?

- Method 2: $\dfrac{\#A\cap B}{\#A}$

  hyper-geometric distribution

  ① $P(A) = \left[\binom{R}{k} \times \binom{N-R}{n-k}\right] \Big/ \binom{N}{n}$

  "#A/#Ω" $= \binom{n}{k} \times [(R)_k \times (N-R)_{n-k}] / (N)_n$

  $= \dfrac{\binom{n}{k} \times \binom{R}{k}\cancel{k!} \times \binom{N-R}{n-k}\cancel{(n-k)!}}{\binom{N}{n}\cancel{n!}}$

  ball 1 (r), ball R (r), ball R+1 (w), ball N (w) (#Ω, symmetric outcomes): {□, …, □} n positions for n balls

  #A: 1st draw, 2nd draw, nth draw: {□, …, □} k positions for red balls; ball 1 (r), ball R (r): (r_1th draw, …, r_kth draw); ball R+1 (w), ball N (w): (w_1th draw, …, w_{n-k}th draw)

  #A: ball 1 (r), ball R (r): {□, …, □} k positions; ball R+1 (w), ball N (w): {□, …, □} n-k positions

---

9/28

□ $P(A\cap B) \underset{\text{by ①}}{=} P(B)P(A|B) = \dfrac{R}{N} \times \boxed{\dfrac{\binom{R-1}{k-1} \times \binom{N-R}{n-k}}{\binom{N-1}{n-1}}}$

easier to evaluate — urn containing R-1 Ⓡ, N-R Ⓦ, n-1 balls drawn, k-1 Ⓡ, n-k Ⓦ

□ $P(B|A) = P(A\cap B)/P(A) = k/n$

➤ Example (Sampling Experiments): An urn contains R red balls and $N-R$ white balls. Sample 2 balls from the urn.

- $A = \{\text{red on the first draw}\}$

  $B = \{\text{red on the second draw}\}$

  ball 1 (r), ball R (r), ball R+1 (w), ball N (w): (□, □) 1st draw, 2nd draw — Ω ← symmetric outcomes

  #Ω = N(N-1)

(r, w)	(w, w)
(r, r)	(w, r)

- Sampling *Without Replacement*:

$P(A\cap\Omega) = P(A) = \dfrac{\boxed{R}\boxed{(N-1)}}{\boxed{N}\boxed{(N-1)}} = \dfrac{R}{N}, \quad P(A\cap B) \underset{\text{by ①}}{=} \dfrac{\boxed{R}\boxed{(R-1)}}{\boxed{N}\boxed{(N-1)}} = P(A)P(B|A)$

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

easier to evaluate

$P(B|A) = \dfrac{P(A\cap B)}{P(A)} = \dfrac{(R(R-1))/(N(N-1))}{R/N} = \dfrac{R-1}{N-1}.$

Similarly, $P(B|A^c) = \dfrac{R}{N-1}$. (exercise) ⇒

meaning

Note: $P(B|A) + P(B|A^c) = \dfrac{R-1}{N-1} + \dfrac{R}{N-1} \ne 1$ in general