IC105: Probability and Statistics **2021-22-M**

Lecture 3: Conditional Probability

Instructor: Dr. Kuldeep Kumar Kataria *Scribe:*

Equally Likely Probability Models for Finite Sample Space:

Suppose that the sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$ is finite (has k elements). Here singletons $\{\omega_i\}$ are called elementary events and $\Omega = \bigcup_{i=1}^{k}\{\omega_i\}$. Suppose that

$$P(\{\omega_i\}) = \frac{1}{k},\ i = 1, 2, \ldots, k \ \text{ (each elementary event is equally likely).}$$

For any event $E \subseteq \Omega$, we have $E = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_r}\}$, for some $i_1, i_2, \ldots, i_r \in \{1, 2, \ldots, k\}$, $1 \le r \le k$. Then, $E = \bigcup_{j=1}^{r}\{\omega_{i_j}\}$ and

$$\begin{aligned} P(E) = P\left(\bigcup_{j=1}^{r}\{\omega_{i_j}\}\right) &= \sum_{j=1}^{r} P\left(\{\omega_{i_j}\}\right) \\ &= \sum_{j=1}^{r}\frac{1}{k} = \frac{r}{k} = \frac{\text{number of ways favourable to event } E}{\text{total number of ways in which the random experiment can terminate}}. \end{aligned}$$

Here the assumption of equally likely $\left(P(\{\omega_i\}) = \frac{1}{k},\ i = 1, 2, \ldots, k\right)$ is a part of probability modelling.

"*At random*": In a random experiment with finite sample space Ω, whenever we say that the experiment has been performed at random it means that all the outcomes in the sample space are equally likely.

Example 3.1 (Birthday Problem)**.** *Suppose that a college has n students, including you. Each of them were born on non-leap years.*

(a) Find the probability that at least two of them have the same birthday. For what values of n this probability is more than 0.5, 0.8, 0.95?

(b) For what value of n the probability that you will find someone who shares your birthday is 0.5.

Solution: Required probability = $1 - P(\text{all of them have different birthdays}) = 1 - \dfrac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$

Required probability = $1 - P(\text{no one shares the same birthday as mine}) = 1 - \dfrac{364^{n-1}}{365^{n-1}}$.

For $1 - \dfrac{364^{n-1}}{365^{n-1}} \approx 0.5$, $n \approx 253$.

Example 3.2. *Five cards are drawn at random and without replacement from a deck of 52 cards. Find the probability that* *(i) each card is spade (event E_1),*

(ii) at least one card is spade (event E_2),

(iii) exactly three cards are king and two cards are queen (event E_3),

(iv) exactly two kings, two queens and one jack are drawn (event E_4).

Solution: (i) $P(E_1) = \frac{\binom{13}{5}}{\binom{52}{5}}$,

(ii) $P(E_2) = 1 - P(E_2^c) = 1 - P(\text{no card is spade}) = 1 - \frac{\binom{39}{5}}{\binom{52}{5}}$,

(iii) $P(E_3) = \frac{\binom{4}{3}\binom{4}{2}}{\binom{52}{5}}$,

(iv) $P(E_4) = \frac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{\binom{52}{5}}$.

Example 3.3 (Capture/Recapture Method for Estimating Population Size)**.** *In a wildlife population suppose that the population size n is unknown. To estimate the population size n, 20 animals are captured, tagged and then released back. Thereafter 40 animals are captured at random and it is found that 8 of them are tagged. Find an estimate of the population size n based on the given data.*

Solution: We have

$$\begin{aligned} n &= \text{total number of animals}, \\ 20 &= \text{number of tagged animals in the population}, \\ n - 20 &= \text{number of untagged animals in the population}. \end{aligned}$$

Data: Sample of 40 animals yield

$$\begin{aligned} &\text{number of tagged animals} = 8, \\ &\text{number of untagged animals} = 32. \end{aligned}$$

The probability of obtaining this data is

$$l(n) = \frac{\binom{20}{8}\binom{n-20}{32}}{\binom{n}{40}}, \;\; n \geq 52.$$

$$\begin{aligned} l(n+1) > l(n) &\iff \frac{\binom{n-19}{32}}{\binom{n+1}{40}} > \frac{\binom{n-20}{32}}{\binom{n}{40}} \\ &\iff \frac{n-19}{(n-51)(n+1)} > 1 \\ &\iff n < 99. \end{aligned}$$

Similarly $l(n+1) < l(n) \iff n > 99$. Thus l is maximized at $n = 99$, that is, for $n = 99$, the observe data (among the captured animals 8 are tagged and 32 are untagged) is most probable.

Thus an estimate of n is $\hat{n} = 99$ (Maximum likelihood estimator).

3.1. Conditional Probability

Consider a probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is finite and

$$P(\{\omega_i\}) = \frac{1}{n},\ i = 1, 2, \ldots, n\ \text{ (equally likely probability model).}$$

Then, for any $A \in \mathcal{F}$

$$P(A) = \frac{\text{number of cases favourable to } A}{\text{total number of cases}} = \frac{|A|}{|\Omega|} = \frac{|A|}{n}.$$

Now suppose it is known a priori that event A has occured (*i.e.* outcome of the experiment is an element of A), where $|A| \geq 1$ (so that $P(A) = |A|/n > 0$). Given this prior information (that the event A has occured) we want to define probability function say $P(B|A)$ on the event space $\mathcal{F}$. A natural way to define $P(B|A)$ is

$$P(B|A) = \frac{|A \cap B|}{|A|} = \frac{|A \cap B|/n}{|A|/n} = \frac{P(A \cap B)}{P(A)},\ B \in \mathcal{F}.$$

Definition 3.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$ be such that $P(A) > 0$. Then*

$$P(B|A) = \frac{P(A \cap B)}{P(A)},\ \ B \in \mathcal{F},$$

is called the conditional probabilty of event B given the event A.

Remark 3.5. *(a) In the above definition the event A (with $P(A) > 0$) is fixed and for this fixed $A \in \mathcal{F}$, $P(\cdot|A)$ is a set function defined on $\mathcal{F}$. Is it a probability function/ measure?*

(b) $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ for A, $B \in \mathcal{F}$.

Theorem 3.6. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$ be such that $P(A) > 0$ be fixed. Then $P(\cdot|A) : \mathcal{F} \to \mathbb{R}$ is a probability function (called the conditional probabilty function) on $\mathcal{F}$ (so that $(\Omega, \mathcal{F}, P(\cdot|A))$ is a probability space).*

Proof. Note that $P(B|A) = \dfrac{P(A \cap B)}{P(A)} \geq 0$ for all $B \in \mathcal{F}$ and $P(\Omega|A) = \dfrac{P(A \cap \Omega)}{P(A)} = 1.$

Let $\{B_n\}_{n \geq 1}$ be a sequence of disjoint events in $\mathcal{F}$. Then,

$$P\left(\bigcup_{n=1}^{\infty} B_n \Big| A\right) = \frac{P\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \cap A\right)}{P(A)} = \frac{P\left(\bigcup_{n=1}^{\infty} (B_n \cap A)\right)}{P(A)}.$$

Since $\{B_n\}_{n \geq 1}$ are disjoint then subsets $\{B_n \cap A\}_{n \geq 1}$ are also disjoint. Since $P(\cdot)$ is a probability measure, we get

$$P\left(\bigcup_{n=1}^{\infty} B_n \mid A\right) = \frac{\sum_{n=1}^{\infty} P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} \frac{P(B_n \cap A)}{P(A)} = \sum_{n=1}^{\infty} P(B_n|A).$$

It follows that $P(\cdot|A)$ is a probabilty function on $\mathcal{F}$ for any fixed $A \in \mathcal{F}$ with $P(A) > 0$. $\square$

Example 3.7. *Five cards are drawn at random (without replacement) from a deck of 52 cards. Define events*

$$B : \textit{all spade in hand and } \ A : \textit{at least 4 spade in hand}.$$

Find $P(B|A)$.

Solution: We have

$$\begin{aligned}P(B|A) &= \frac{P(A\cap B)}{P(A)} = \frac{P(B)}{P(A)} \ \ (\text{since } \ B\subseteq A)\\ &= \frac{\binom{13}{5}/\binom{52}{5}}{\left[\binom{13}{4}\binom{39}{1}+\binom{13}{5}\right]/\binom{52}{5}} = 0.441.\end{aligned}$$

Remark 3.8 (Multiplication Law)**.** *(i)* $P(A\cap B) = P(A)P(B|A)$*, if* $P(A) > 0$*.*

(ii) $P(A\cap B\cap C) = P(A\cap B)P(C|A\cap B) = P(A)P(B|A)P(C|A\cap B)$*, provided* $P(A\cap B) > 0$ *(which ensures that* $P(A) > 0$ *as* $A\cap B\subseteq A$*).*

(iii) Using principle of mathematical induction, we have

$$\begin{aligned}P\left(\bigcap_{i=1}^{n} C_i\right) &= P(C_1\cap C_2\cap\cdots\cap C_n)\\ &= P(C_1\cap C_2\cap\cdots\cap C_{n-1})P(C_n|C_1\cap C_2\cap\cdots\cap C_{n-1})\\ &= P(C_1\cap C_2\cap\cdots\cap C_{n-2})P(C_{n-1}|C_1\cap C_2\cap\cdots\cap C_{n-2})P(C_n|C_1\cap C_2\cap\cdots\cap C_{n-1})\\ &\ \ \vdots\\ &= P(C_1)P(C_2|C_1)P(C_3|C_1\cap C_2)\dots P(C_n|C_1\cap C_2\cap\cdots\cap C_{n-1})\end{aligned}$$

provided $P(C_1\cap C_2\cap\cdots\cap C_{n-1}) > 0$ *(which also ensures that* $P(C_1\cap C_2\cap\cdots\cap C_i) > 0$*,* $i = 1, 2, \dots, n-2$*).*

Due to symmetry, if $(\alpha_1, \alpha_2, \dots, \alpha_n)$ *is a permutation of* $(1, 2, \dots, n)$*, then*

$$\begin{aligned}P\left(\bigcap_{i=1}^{n} C_i\right) &= P(C_{\alpha_1}\cap C_{\alpha_2}\cap\cdots\cap C_{\alpha_n})\\ &= P(C_{\alpha_1})P(C_{\alpha_2}|C_{\alpha_1})P(C_{\alpha_3}|C_{\alpha_1}\cap C_{\alpha_2})\dots P(C_{\alpha_n}|C_{\alpha_1}\cap C_{\alpha_2}\cap\cdots\cap C_{\alpha_{n-1}})\end{aligned}$$

provided $P(C_{\alpha_1}\cap C_{\alpha_2}\cap\cdots\cap C_{\alpha_{n-1}}) > 0$ *(which also ensures that* $P(C_{\alpha_1}\cap C_{\alpha_2}\cap\cdots\cap C_{\alpha_i}) > 0$*,* $i = 1, 2, \dots, n-2$*).*

Example 3.9. *A bowl contains 3 red and 5 blue chips. All chips that are of the same colour are identical. Two chips are drawn successively at random and without replacement. Define events*

$$\begin{aligned}&A : \textit{first draw resulted in a red chip},\\ &B : \textit{second draw resulted in a blue chip}.\end{aligned}$$

Find $P(A\cap B)$*,* $P(A)$ *and* $P(B)$*.*

Solution: $P(A) = \dfrac{3}{8}$, $P(B|A) = \dfrac{5}{7}$ and

$$P(B) = P(A\cap B) + P(A^c\cap B) = P(B|A)P(A) + P(B|A^c)P(A^c) = \frac{5}{7}\times\frac{3}{8} + \frac{4}{7}\times\frac{5}{8} = \frac{35}{56}.$$

Note that here the outcomes of second draw is dependent on outcome of first draw ($P(B|A)\neq P(B)$). Also,

$$P(A\cap B) = P(A)P(B|A) = \frac{3}{8}\times\frac{5}{7} = 0.2679.$$

Theorem 3.10 (Theorem of Total Probability)**.** *For a countable set* Δ *(that is elements of* Δ *can either be put in 1-1 correspondence with* $\mathbb{N} = \{1, 2, \dots\}$ *or with* $\{1, 2, \dots, n\}$ *for some* $n\in\mathbb{N}$*), let* $\{E_\alpha : \alpha\in\Delta\}$ *be a countable* *collection of mutually exclusive (i.e., $E_\alpha \cap E_\beta = \phi,\ \forall\ \alpha \neq \beta$) and exhaustive (i.e., $P\left(\bigcup_{\alpha\in\Delta} E_\alpha\right) = 1$) events. Then, for any $E \in \mathcal{F}$,*

$$P(E) = \sum_{\alpha\in\Delta} P(E\cap E_\alpha) = \sum_{\substack{\alpha\in\Delta\\ P(E_\alpha)>0}} P(E|E_\alpha)P(E_\alpha).$$

Proof. Since $P\left(\bigcup_{\alpha\in\Delta} E_\alpha\right) = 1$, we have

$$\begin{aligned} P(E) &= P\left(E\bigcap\left(\bigcup_{\alpha\in\Delta} E_\alpha\right)\right) = P\left(\bigcup_{\alpha\in\Delta}(E\cap E_\alpha)\right) \\ &= \sum_{\alpha\in\Delta} P(E\cap E_\alpha),\ \ (E_\alpha\text{'s are disjoint} \implies \text{their subsets } (E\cap E_\alpha)\text{'s are disjoint}) \\ &= \sum_{\substack{\alpha\in\Delta\\ P(E_\alpha)>0}} P(E\cap E_\alpha),\ \ (P(E_\alpha)=0 \implies P(E\cap E_\alpha)=0,\ \alpha\in\Delta) \\ &= \sum_{\substack{\alpha\in\Delta\\ P(E_\alpha)>0}} P(E|E_\alpha)P(E_\alpha). \end{aligned}$$

This completes the proof. □

Example 3.11. *A population comprises of 40% female and 60% male. Suppose that 15% of female and 30% of male in the population smoke. A person is selected at random from the population.*

(a) Find the probability that he/she is a smoker.

(b) Given that the selected person is smoker, find the probability that he is male.

Solution: Define the events

$$\begin{aligned} M &: \text{selected person is a male,} \\ F = M^c &: \text{selected person is a female,} \\ S &: \text{selected person is a smoker,} \\ T = S^c &: \text{selected person is a non-smoker.} \end{aligned}$$

We have $P(F) = 0.4,\ P(M) = 0.6,\ P(F\cup M) = P(F)+P(M) = 1,\ P(S|F) = 0.15,\ P(T|F) = 0.85,\ P(S|M) = 0.30,\ P(T|M) = 0.70.$

(a) By using Theorem of total probability, we get

$$P(S) = P(S\cap F) + P(S\cap M) = P(S|F)P(F) + P(S|M)P(M) = 0.15\times 0.4 + 0.30\times 0.6 = 0.24.$$

(b)

$$P(M|S) = \frac{P(M\cap S)}{P(S)} = \frac{P(S|M)P(M)}{P(S)} = \frac{0.30\times 0.60}{0.24} = \frac{3}{4}.$$

Theorem 3.12 (Bayes' Theorem)**.** *Let $\{E_\alpha : \alpha\in\Delta\}$ be a countable collection of mutually exclusive and exhaustive events and let E be any event $P(E) > 0$. Then, for $j\in\Delta$ with $P(E_j) > 0$,*

$$P(E_j|E) = \frac{P(E|E_j)P(E_j)}{\sum\limits_{\substack{\alpha\in\Delta\\ P(E_\alpha)>0}} P(E|E_\alpha)P(E_\alpha)}.$$

Proof. For $j \in \Delta$,

$$P(E_j|E) = \frac{P(E_j \cap E)}{P(E)} = \frac{P(E|E_j)P(E_j)}{\sum\limits_{\substack{\alpha \in \Delta \\ P(E_\alpha)>0}} P(E|E_\alpha)P(E_\alpha)}, \text{ (using Theorem of total probability).}$$

This completes the proof. $\square$

Remark 3.13. *(a) Suppose that occurrence of any of the mutually exclusive and exhaustive events $\{E_\alpha : \alpha \in \Delta\}$ (where Δ is a countable set) may cause the occurrence of an event E. Given that the event E has occurred (i.e., given the effect), Bayes' Theorem provides the conditional probability that the event E (effect) is caused by occurrence of event $E_j,\ j \in \Delta$.*

(b) In Bayes' Theorem $\{P(E_j) : j \in \Delta\}$ are called prior probabilities and $\{P(E_j|E) : j \in \Delta\}$ are called posterior probabilities.

Example 3.14. *Bowl C_1 contains 3 red and 7 blue chips. Bowl C_2 contains 8 red and 2 blue chips. Bowl C_3 contains 5 red and 5 blue chips. All chips of the same colour are identical.*

A die is cast and a bowl is selected as per the following schemes:

Bowl C_1 is selected if 5 or 6 spots show on the upper side,
Bowl C_2 is selected if 2,3 or 4 spots show on the upper side,
Bowl C_3 is selected if 6 spots show on the upper side.

The selected bowl is handed over to another person who drawns two chips at random from this bowl. Find the probability that:

(a) Two red chips are drawn.

(b) Given that drawn chips are both red, find the probability that it came from bowl C_3.

Solution: Define the events

A_i: selected bowl is C_i, $i = 1, 2, 3$, and R: the chips drawn from the selected bowl are both red.

Then $P(A_1) = \frac{2}{6} = \frac{1}{3}$, $P(A_2) = \frac{3}{6} = \frac{1}{2}$, $P(A_3) = \frac{1}{6}$. Note that $\{A_1, A_2, A_3\}$ are mutually exclusive and exhaustive.

(a)

$$P(R) = P(R|A_1)P(A_1) + P(R|A_2)P(A_2) + P(R|A_3)P(A_3) = \frac{\binom{3}{2}}{\binom{10}{2}} \times \frac{1}{3} + \frac{\binom{8}{2}}{\binom{10}{2}} \times \frac{1}{2} + \frac{\binom{5}{2}}{\binom{10}{2}} \times \frac{1}{6} = \frac{10}{27}.$$

(b)

$$P(A_3|R) = \frac{P(R|A_3)P(A_3)}{P(R)} = \frac{\frac{\binom{5}{2}}{\binom{10}{2}} \times \frac{1}{6}}{\frac{10}{27}} = \frac{1}{10}.$$

Remark 3.15. *In the above example,*

$$
\begin{aligned}
P(A_1|R) &= \frac{P(R|A_1)P(A_1)}{P(R)} = \frac{\frac{\binom{3}{2}}{\binom{10}{2}} \times \frac{1}{3}}{\frac{10}{27}} = \frac{3}{50}, \\
P(A_2|R) &= \frac{P(R|A_2)P(A_2)}{P(R)} = \frac{\frac{\binom{8}{2}}{\binom{10}{2}} \times \frac{1}{2}}{\frac{10}{27}} = \frac{21}{25}, \\
P(A_1|R) &= \frac{3}{50} < \frac{1}{3} = P(A_1) \iff P(A_1 \cap R) < P(A_1)P(R) \longleftrightarrow \textit{R has negative information about } A_1, \\
P(A_2|R) &= \frac{21}{25} > \frac{1}{2} = P(A_2) \iff P(A_2 \cap R) > P(A_2)P(R) \longleftrightarrow \textit{R has positive information about } A_2, \\
P(A_3|R) &= \frac{1}{10} < \frac{1}{6} = P(A_3) \iff P(A_3 \cap R) < P(A_3)P(R) \longleftrightarrow \textit{R has negative information about } A_3.
\end{aligned}
$$

Note that proportion of red chips in C_2 > proportion of red chips in C_i, $i = 1, 3$.