Lecture Stat 302
Introduction to Probability - Slides 4

AD

Jan. 2010

Axioms of Probability

- Consider an experiment with sample space S. For each event E, we assume that a number $P(E)$, the *probability* of the event E, is defined and satisfies the following 3 axioms.
- **Axiom 1**

$$0 \leq P(E) \leq 1$$

- **Axiom 2**

$$P(S) = 1$$

- **Axiom 3**. For any sequence of mutually exclusive events $\{E_i\}_{i\geq 1}$, i.e. $E_i \cap E_j = \varnothing$ when $i \neq j$, then

$$P\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

- Direct consequences include $P(\varnothing) = 0$ and for mutually exclusive events $\{E_i\}_{i\geq 1}$

$$P\left(\cup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i).$$

Examples

- *Example* (coins): Assume both coins are unbiased; i.e. a head is as likely to appear as a tail, then

$$P(\{H,H\}) = P(\{H,T\}) = P(\{T,H\}) = P(\{T,T\}) = \frac{1}{4}.$$

- *Example*: A die is rolled and we assume $P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = 1/6$. Hence as a consequence from axiom 3, the probability of having an even or odd number is equal to

$$\begin{aligned} P(\{1,3,5\}) &= P(\{1\}) + P(\{3\}) + P(\{5\}) = 1/2, \\ P(\{2,4,6\}) &= P(\{2\}) + P(\{4\}) + P(\{6\}) = 1/2. \end{aligned}$$

Properties

- **Proposition**: $P(E^c) = 1 - P(E)$.
- We have $S = E \cup E^c$ and $E \cap E^c = \varnothing$ so

$$\underbrace{P(S) = 1}_{\text{axiom } 2} = \underbrace{P(E \cup E^c) = P(E) + P(E^c)}_{\text{axiom } 3}.$$

- **Proposition**: If $E \subset F$ then $P(E) \leq P(F)$.
- We have $F = E \cup (E^c \cap F)$ and $E \cap ((E^c \cap F)) = \varnothing$ so

$$P(F) = P(E) + \underbrace{P(E^c \cap F)}_{\geq 0 \text{ by axiom } 1} \geq P(E).$$

- **Proposition**: We have $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.
- We have $E \cup F = E \cup (E^c \cap F)$ and $E \cap ((E^c \cap F)) = \varnothing$ so

$$P(E \cup F) = P(E \cup (E^c \cap F)) = P(E) + P(E^c \cap F)$$

but $F = (E^c \cap F) \cup (E \cap F)$ with $(E^c \cap F) \cap (E \cap F) = \varnothing$ so

$$P(F) = P(E^c \cap F) + P(E \cap F) \Rightarrow P(E^c \cap F) = P(F) - P(E \cap F).$$

Example

- You are in a restaurant and ordering 2 dishes. With proba 0.6, you will like the first dish; with proba 0.4, you will like the second dish. With proba 0.3, you will like both of them. What is the proba. you will like neither dish?
- Let A_i the event: "You like dish i". Then the proba you like at least one is

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0.6 + 0.4 - 0.3 = 0.7.$$

- The event that you like neither dish is the complement of liking at least one, so

$$\begin{aligned} P(\text{"you will like neither dish"}) &= P\left((A_1 \cup A_2)^c\right) \\ &= 1 - P(A_1 \cup A_2) \\ &= 0.3 \end{aligned}$$

Example

- A die is thrown twice and the number on each throw is recorded. Assuming the dice is fair, what is the probability of obtaining at least one 6?
- There are clearly 6 possible outcomes for the first throw and 6 for the second throw. By the counting principle, there are 36 possible outcomes for the two throws. Let A_i the event "I have obtained a 6 for throw i". The probability we are interested in is

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \\ &= \frac{1}{6} + \frac{1}{6} - \frac{1}{36} \\ &= \frac{11}{36}. \end{aligned}$$

Inclusion-Exclusion Identity

- We have

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) \\ -P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G).$$

- Proof follows from $E \cup F \cup G = (E \cup F) \cup G$ so using $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ where $A = E \cup F$, $B = G$

$$P(E \cup F \cup G) = P(E \cup F) + P(G) - P((E \cup F) \cap G) \\ = P(E) + P(F) - P(E \cap F) + P(G) - P((E \cup F) \cap G).$$

- Now we have $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ so

$$P((E \cup F) \cap G) = P(E \cap G) + P(F \cap G) - \underbrace{P((E \cap G) \cap (F \cap G))}_{=P(E \cap F \cap G)}$$

and the result follows.

General Inclusion-Exclusion Identity

- We have

$$
\begin{aligned}
P\left(E_1 \cup E_2 \cup \cdots \cup E_n\right) &= \sum_{i=1}^{n} P\left(E_i\right) - \sum_{i_1 < i_2}^{n} P\left(E_{i_1} \cap E_{i_2}\right) + \cdots \\
&\quad + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r}^{n} P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right) \\
&\quad + \cdots + (-1)^{n+1} P\left(E_1 \cap \cdots \cap E_n\right) \\
&= \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r}^{n} P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right)
\end{aligned}
$$

- This can be proven by induction.

Example: Matching problem

- You have n letters and n envelopes and randomly stuff the letters in the envelopes. What is the probability that at least one letter will match its intended envelope?
- The sample space is the space of permutations of $\{1, 2, ..., n\}$ and thus has $n!$ outcomes.
- Let $E_i =$"letter i matches its intended envelope". We are interested in $P\left(E_1 \cup E_2 \cup \cdots \cup E_n\right)$.
- Consider the event $E_{i_1} \cap \cdots \cap E_{i_r}$ the event that each of the r letters $i_1, ..., i_r$ match their intended envelopes. There are $(n-r)(n-r-1)\cdots 1$ such outcomes corresponding to the number of ways the remaining r envelopes can be matched. Assuming all outcomes equiprobable, we have

$$P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right) = \frac{(n-r)(n-r-1)\cdots 1}{n!} = \frac{(n-r)!}{n!}$$

Example: Matching problem (cont.)

- We want to compute

$$P\left(E_1 \cup E_2 \cup \cdots \cup E_n\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r}^{n} P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right)$$

- We have $\binom{n}{r}$ terms of the form $P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right)$. Moreover,

$$\binom{n}{r} \underbrace{P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right)}_{=\frac{(n-r)!}{n!}} = \frac{1}{r!}.$$

- It follows that

$$\begin{aligned} P\left(E_1 \cup E_2 \cup \cdots \cup E_n\right) &= \sum_{r=1}^{n} (-1)^{r+1} \frac{1}{r!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{n+1}}{n!} \\ &\underset{n\to\infty}{\longrightarrow} 1 - e^{-1} \end{aligned}$$

Sample Spaces with Equally Likely Outcomes

- Assume $S = \{1, 2, ..., N\}$ then it is often natural to assume $P(\{i\}) = 1/N$ so, for any event E,

$$P(E) = \frac{\#\text{ outcomes in } E}{\#\text{ outcomes in } S}.$$

- **Example**: If 5 balls are "randomly drawn" from a bowl containing 10 white and 7 black balls, what is the probability that 3 of the balls are white and the 2 other black?
- **Answer**: There are $\binom{17}{5}$ possible outcomes of the experiment which are assumed equally likely. We have $\binom{10}{3}$ ways to select 3 white balls among 10 and $\binom{7}{2}$ ways to select 2 black balls among 7. So the probability is given by

$$\binom{10}{3}\binom{7}{2} / \binom{17}{5} = 0.4072$$

Example: Lottery

- In the 6/49 Lottery, there are 49 numbered balls, and 6 of these are selected at random. What is the probability that exactly 3 of the 6 numbers we select are drawn? What is the probability of having at least 3 numbers?
- The number of possible draws (the number of different sets of 6 numbers) is $\binom{49}{6} = 13,983,816$. The number of possible draws with exactly 3 "good" numbers is $\binom{6}{3}\binom{43}{3} = 246,820$. So the probability is $246,820/13,983,816 \approx 0.0177$.
- The number of possible draws with at least 3 "good" numbers is

$$\binom{6}{3}\binom{43}{3} + \binom{6}{4}\binom{43}{2} + \binom{6}{5}\binom{43}{1} + \binom{6}{6}\binom{43}{0}$$
$$= 246,820 + 13,545 + 258 + 1 = 260,624$$

So the probability is $260,624/13,983,816 \approx 0.0186$.

Example: Witness Identification

- A line-up of 10 men is conducted in order that a witness can identify 3 suspects. Suppose that 3 people in the line-up actually committed the crime in question. If the witness does not recognise any of the suspects, but simply chooses three men at random, what is the probability that the three guilty men are selected? What is the probability that the witness selects three innocent men?
- There are $\begin{pmatrix} 10 \\ 3 \end{pmatrix} = 120$ ways to select 3 men out of ten. So there is a probability $\frac{1}{120} \approx 0.0083$ to pick the 3 guilty men randomly.
- There are $\begin{pmatrix} 7 \\ 3 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 35$ ways to pick 3 innocent men. So there is a probability $\frac{35}{120} \approx 0.2917$ to pick the 3 innocent men randomly.

Historical Example: Chevalier de Méré, Fermat and Pascal

- **Example**: Chevalier de Méré, who was a big gambler, realized empirically that he could earn money by betting that, if you throw a fair dice 4 times, at least one "6" would appear. What is the probability of this event?
- **Answer**: The complement event is to never obtain a "6" when you throw the dice 4 times, this has a probability $\left(\frac{5}{6}\right)^4$. Hence the probability of interest is given by $1-\left(\frac{5}{6}\right)^4 \approx 0.5177$.
- **Example**: Chevalier de Méré decided to extend this "trick" to two dices and conjectured that he would still make a money by betting that, if you throw two dices 24 times, at least one double "6" would appear. He started losing money and asked Fermat and Pascal to explain him why.
- **Answer**: The complement event is to never obtain a double "6" when you throw the dice 24 times, it has a probability $\left(\frac{35}{36}\right)^{24}$. Hence the probability of interest is given by $1-\left(\frac{35}{36}\right)^{24} \approx 0.4914$.

Example: Portable Music Player

- **Example**: You are having 4,000 songs on your portable music player and are using an option that picks the songs randomly (after each song, it picks a new song, each with proba $1/4000$). You then realize that the 75th song played by your player is the same as the 30th (or the 12th or whatever) you have listened to. Should you complain to the manufacturer?
- **Answer:** If the songs are picked randomly, then the probability of not listening twice to the same song among the 75 first songs is

$$\frac{4000.3999 \cdot \cdots \cdot (4000 - 75 + 1)}{4000^{75}} \approx 0.4975.$$

- Your portable music player works just fine.

Example: Birthday Problem

- **Example**: If n people are present in a room, what is the probability that at least two of them celebrate their birthday on the same day of the year? (We assume that there are 365 days and proba of being born a given day is $1/365$).
- **Answer:** The complementary event is that no two of them celebrate their birthday on the same day which is given by

$$\overline{P}_n = \frac{365 \cdot 364 \cdot \cdots \cdot (365 - n + 1)}{365^n}$$

so the probability is $P_n = 1 - \overline{P}_n$.

- For $n = 23$, we have $P_n \geq 0.5$ and for $n = 70$, we have $P_n \approx 0.9992$.
- This might appear suprising but there are $\binom{n}{2}$ possible pairs of people. For $n = 23$, $\binom{23}{2} = 253$ and for $n = 70$, $\binom{70}{2} = 2415$.

Example: Three children with the same birthday

- A recent news story in the UK featured a family whose three children had all been born on the same day. But is this so remarkable?
- The sample space is $S = ((i, j, k); i \in \{1, ..., 365\}, j \in \{1, ..., 365\}, j \in \{1, ..., 365\})$ so assuming each day is equally likely, the proba the three days coincides is
$$\frac{365}{365 \times 365 \times 365} \approx \frac{7.5}{1,000,000},$$
this is quite small but much higher that winning at the lottery. [The proba of the 3 birthdays being on a very specific day, e.g. Christmas, is $\frac{1}{365 \times 365 \times 365} \approx 2.06.10^{-8}$]
- There are 24,000,000 households in the UK, and 1,000,000 of them are made up of a couple and 3 or more dependent children. Therefore we would expect around 7 or 8 families in Britain to have three children all born on the same day, and so this family is unlikely to be unique in this country.

Probability as a Continuous Set Function

- A sequence of events $\{E_n\}_{n\geq 1}$ is said to be increasing if

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

and we define a new event, denoted $\lim_{n\to\infty} E_n$, by

$$\lim_{n\to\infty} E_n = \cup_{i=1}^{\infty} E_i$$

- Similarly, $\{E_n\}_{n\geq 1}$ is said to be decreasing if

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

and we define a new event, denoted $\lim_{n\to\infty} E_n$, by

$$\lim_{n\to\infty} E_n = \cap_{i=1}^{\infty} E_i$$

- **Proposition**: If $\{E_n\}_{n\geq 1}$ is either an increasing or decreasing sequence of events, then

$$\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right).$$

Proof

- We prove it for a sequence of increasing events. The idea is to use Axiom 3 to establish the result.
- We have as $\{E_n\}_{n\geq 1}$ is increasing

$$\cup_{i=1}^{n} E_i = E_n$$

and we have the following "donuts" decomposition

$$E_n = \cup_{i=1}^{n} F_i$$

with

$$F_1 = E_1 \text{ and } F_n = E_n \cap E_{n-1}^c.$$

- It is easy to check that

$$F_i \cap F_j = \varnothing \text{ and } \cup_{i=1}^{n} E_i = \cup_{i=1}^{n} F_i \text{ for all } n \geq 1$$

Proof (cont.)

- We have

$$\begin{aligned} P\left(\cup_{i=1}^{\infty} E_i\right) &= P\left(\cup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P\left(F_i\right) \text{ (axiom 3)} \\ &= \lim_{n\rightarrow\infty} \sum_{i=1}^{n} P\left(F_i\right) \\ &= \lim_{n\rightarrow\infty} P(\underbrace{\cup_{i=1}^{n} F_i}_{=\cup_{i=1}^{n} E_i}) \\ &= \lim_{n\rightarrow\infty} P\left(E_n\right) \end{aligned}$$

- For decreasing events $\{E_n\}_{n\geq 1}$, use the fact that $\{E_n^c\}_{n\geq 1}$ is increasing.

Example: A Paradox

- You possess an infinitely large urn and an infinite collection of balls labelled number 1,2,3 etc. At time 0, balls 1 to 10 are placed in the urn and ball 10 withdrawn. After $1/2$ minute, balls 11 to 20 are placed in the urn and ball 20 withdrawn. After one more $1/4$ minute, balls 21 to 30 are places in the urn and ball 30 withdrawn etc. How many balls do we have in the urn after one minute?
- We clearly have an infinite number as, at time

$$\frac{1}{2} + ... + \frac{1}{2^n}$$

we have $9\left(n+1\right)$ balls after the $n+1$th withdrawal; i.e. balls 1 to 9, 11 to 19,..., $10n+1$ to $10n+9$.
- We now change the experiment so that at time 0, balls 1 to 10 are placed in the urn and ball 1 is withdrawn. After $1/2$ minute, balls 11 to 20 are placed in the urn and ball 2 withdrawn. After $1/4$ more minute, balls 21 to 30 are places in the urn and ball 3 withdrawn etc. How many balls do we have in the urn after one minute?

Example: A Paradox (cont.)

- The urn is empty as ball number n is removed at time 0 for $n=1$ and time
$$\frac{1}{2}+\ldots+\frac{1}{2^{n-1}}<1 \text{ for } n>1.$$
- Consider now the case where, whenever a ball is to be withdrawn, that ball is randomly selected from among those present; i.e. at time 0 balls 1 to 10 are placed in the urn and one is randomly selected and withdrawn, and so on. How many balls will you have in the urn after one minute?
- Consider ball number 1. Define E_n the event that this ball is still in the urn after n withdrawals. We have
$$P\left(E_n\right)=\frac{9\cdot 18\cdot 27\cdot\cdots\cdot 9n}{10\cdot 19\cdot 28\cdot\cdots\cdot(9n+1)}=\prod_{i=1}^{n}\frac{9i}{9i+1}$$
- Being in the urn at time 1 is the event $\cap_{i=1}^{\infty}E_i$ and $\left\{E_i\right\}_{i\geq 1}$ is increasing so $P\left(\cap_{i=1}^{\infty}E_i\right)=\lim_{n\to\infty}P\left(E_n\right).$

Example: A Paradox (cont.)

- We have

$$\frac{1}{P(E_n)} = \prod_{i=1}^{n}\left(1+\frac{1}{9i}\right) \geq \frac{1}{9}+\frac{1}{18}+\cdots+\frac{1}{9n} \underset{n\to\infty}{\to} \infty$$

so

$$P\left(\cap_{i=1}^{\infty} E_i\right) = 0.$$

- This reasoning can be extended to any ball number k. Define $p(k)$ the positive integer such that $k = 10\, p(k) + d(k)$ where $0 \leq d(k) \leq 9$ and define E_n the event that ball k is still in the urn after n $(n \geq p(k)+1)$ withdrawals then similarly

$$P(E_n) = \prod_{i=p(k)+1}^{n} \frac{9i}{9i+1} \underset{n\to\infty}{\to} \infty$$

- Hence, the urn will be empty at time 1.

Interpretation of Probability

- Consider an event E of the sample space S. Assume you replicate the experiment n times, then it is tempting to define "practically"

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

where $n(E)$ is the number of times the event E occurred in the n experiments.
- This is known as the frequentist approach: you should repeat an infinite number of times an experiment and the probabilities corresponds to the limiting frequencies.
- *Problem*. How do you attribute a probability to the following event "There will be a major earthquake in Tokyo on the 27th April 2013"?
- In many scenarios, probabilities are measures of the individual's degree of belief: this is *subjective*.
- This does not have any impact on the mathematical "machinery" as long as you define the axioms 1,2 and 3 are satisfied.