NYU MathUA-233 Theory of Probability (Summer 2017 Session II)

(Week 2) Mon July 10- Thurs July 13, 2017

1 Week 2 Topics

1.1 Chapter 3: Conditional Probability

Probability: The quantitative study of randomness and uncertainty. Conditioning: The spirit of probability and statistics.

1. **Independence: means that the occurrence of one event does not affect the probability of the other**:
$$P(AB) = P(A)P(B)$$
   Independence *is different* from disjoint (or mutually exclusive).

   If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

   Independence of $E_1, E_2, \ldots, E_n$ events: if *for any subset* $E_{i_1}, \ldots, E_{i_r}$ of them: $P(E_{i_1}E_{i_2}\ldots E_{i_r}) = P(E_{i_1})P(E_{i_2})\ldots P(E_{i_r})$.

2. **Conditional probability**: $P(A|B) = \frac{P(AB)}{P(B)}$ (assuming $P(B) \neq 0$). This gives the probability of A *given that* B *occurred.* Think of this as *updating* the (unconditional) probability of A when *new evidence* B becomes available.

3. Prove that $P(A|B) = \frac{P(AB)}{P(B)}$ is a probability measure.

4. **Multiplication Rule**: $P(E_1E_2\ldots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\ldots P(E_n|E_1E_2\ldots E_{n-1})$.

5. **The law of total probability**: $P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$ (conditioning on $F_1, F_2, \ldots, F_n$, where F_i are mutually exclusive events whose union is the entire sample space). In particular
$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$
   Think of *divide and conquer*. Divide the problem into easier subproblems then solve.

   How do we choose the events $\{F_i\}$ on which to condition? Choose them based on what you wish you knew in order to find the desired probability.

6. **Probability that an event E occurs first** (assuming independence in the sense that E or F or $(E \cup F)^c$ are the only possible occurrences of the experiment, and that E and F are mutually exclusive): Conditioning on these outcomes leads to $P(E \text{ occurs first}) = \frac{P(E)}{P(E)+P(F)}$.

7. **Bayes's formula**: Relates $P(A|B)$ to $P(B|A)$ (the whole useful (and controversial) Bayesian statistics is based on this formula.)
$$P(B|A) = P(A|B)\frac{P(B)}{P(A)}$$

8. This can be concluded from Bayes's formula: $P(F_j|A) = P(A|F_j)\frac{P(F_j)}{\sum_{i=1}^{n} P(A|F_i)P(F_i)}$ where F_j is a partition of the sample space S.

9. **Odds of an event** A: $\frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)}$. Also we have $\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)P(E|H)}{P(H^c)P(E|H^c)}$.

10. **Avoid the following common mistakes**:
    - Confusing disjoint events with independent events.
    - Confusing $P(A|B)$ with $P(B|A)$ and drawing conclusions based on that (accuracy of medical tests, prosecutor's fallacy).
    - Confusing $P(A)$ and $P(A|B)$.
    - Confusing independence with conditional independence (neither implies the other).

11. Famous examples worked out in class:
    (a) **Casino**
    (b) **Accuracy of medical tests**
    (c) **Prosecutor's fallacy**
    (d) **Three door game**
    (e) **Simpson's paradox**
    (f) **Random walk: Gambler's ruin** Conditioning on the first step.

2 Reading assignment

- Chapter 3 sections: 3.1, 3.2, 3.3, 3.4 and 3.5.

3 Homework Assignments

1. on Mon July 10 lecture: Homework from Lecture on Thursday July 6.

2. On Tues July 11 lecture: page 98 (6, 13, 15, 28, 33, 58) and page 106 (1, 2, 5).

3. on Wed July 12 lecture: page 98 (6, 13, 15, 28, 33, 58) and page 106 (1, 2, 5).

4. on Thurs July 13 lecture:
    (a) Give an example of 3 events A, B, C which are pairwise independent but not independent. (Hint: find an example where whether C occurs is completely determined if we know whether A occurred and whether B occurred, but completely undetermined if we know only one of these things.)
    (b) A crime is committed by one of two suspects, A and B. Initially, there is equal evidence against both of them. In further investigation at the crime scene, it is found that the guilty party had a blood type found in 10% of the population. Suspect A does match this blood type, whereas the blood type of Suspect B is unknown.
        i. Given this new information, what is the probability that A is the guilty party?
        ii. Given this new information, what is the probability that B?s blood type matches that found at the crime scene?

(c) (Gambler's ruin) A gambler repeatedly plays a game where in each round, he wins a dollar with probability 1/3 and loses a dollar with probability 2/3. His strategy is "quit when he is ahead by \$2," though some suspect he is a gambling addict anyway. Suppose that he starts with a million dollars. Show that the probability that he?ll ever be ahead by \$2 is less than 1/4.

4 Test 2 on Thursday July 13

Name:

1. Compute the probability that a hand of 13 cards contains:
   (a) The ace and king of at least one suit.
   (b) All 4 of at least one of the 13 denominations.

2. A certain family has 6 children, consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?

3. A city with 6 districts has 6 robberies in a particular week. Assume the robberies are located randomly, with all possibilities for which robbery occurred where equally likely. What is the probability that some district had more than 1 robbery?

4. *Independence*: Is it always true that if A and B are independent events, then A^c and B^c are independent events? Show that it is, or give a counterexample.

5. In recent years, much has been written about the possible link between cigarette smoking and lung cancer. Suppose that in a large medical center, of all the smokers who were suspected of having lung cancer, 90% of them did, while only 5% of the nonsmokers who were also suspected of having lung cancer actually did. If the proportion of smokers is 0.45, what is the probability that a lung cancer patient who is randomly selected is a smoker?

6. *Prosecutor's fallacy*: A crime-scene DNA sample is compared against a database of 20,000 men. A match is found, that man is accused and at his trial, it is testified that the probability that two DNA profiles match by chance is only 1 in 10,000. This does not mean the probability that the suspect is innocent is 1 in 10,000.

   Show that even if none of the men in the database left the crime-scene DNA, a match by chance to an innocent is almost 86%.

7. *Simpson's paradox*: Suppose two people, Lisa and Bart, each edit articles for two weeks. In the first week, Lisa fails to improve the only article she edited, and Bart improves 1 of the 4 articles he edited. In the second week, Lisa improves 3 of 4 articles she edited, while Bart improves the only article he edited. Explain how this results in a Simpson's paradox when the totals for the two weeks are added together.