Probability- Class 2

January 9, 2014 Debdeep Pati

Sample space, events and probability

1. Sample space: the set of all possible outcomes of an experiment.
2. Flipping a coin, $S = \{H, T\}$
3. Flipping two coins, $S = \{(H, H), (H, T), (T, H), (T, T)\}$
4. Breast cancer, $S = \{Y, N\}$; Two women $(Y, Y), (Y, N), (N, Y), (N, N)$.
5. Cholesterol level, $S = [0, \infty)$
6. Events: An event is a set of outcomes, E, of the sample space S.
7. Flipping two coins, $E = \{(H, H), (H, T)\}$ is the event that a head appears on the first coin.
8. For breast cancer $E = \{(Y, N), (N, Y)\}$ is the event that one of them get cancer.
9. For cholesterol level, $E = [100, 200]$ is the event that blood pressure is between 100 and 200.
10. Probability of an event: Relative frequency of the set of outcomes of the event over an indefinitely large number of trials.

Axioms of probability

1. Probability of an event A, $P(A)$ satisfies $0 \leq P(A) \leq 1$
2. Probability of the sample space $P(S) = 1$
3. If A and B are two events that cannot both happen at the same time, then P(A or B occurs) = P(A) + P(B)

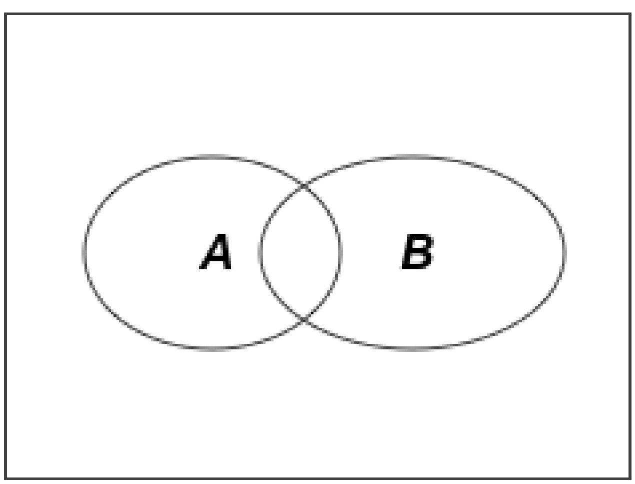
Operations on Events

1. Union: $A \cup B$, Union of A and B: event that either A or B occurs.
2. $A = \{Y\}$, $B = \{N\}$, $A \cup B = \{Y, N\}.A = \{X < 90\}, B = \{90 \leq X < 95\}, A \cup B = \{X < 95\}$

3. Intersection: $A \cap B$
4. Intersection of A and B: event that both A and B occur $A = \{(H, H), (H, T), (T, H)\}, B = \{(H, T), (T, H), (T, T)\}, A \cap B = \{(H, T), (T, H)\}$
5. Commutative laws $A \cup B = B \cup A, A \cap B = B \cap A$
6. Associative laws $A \cup (B \cup C) = (A \cup B) \cup C$
7. Distributive laws $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
8. If A and B have no outcomes in common, then we say A and B are mutually exclusive and denoted by $A \cap B = \phi$
9. Complement of a set E is denoted by E^c or $\bar{E}$ and represents the outcomes in S that are not in E.
10. $E = \{(H, H), (H, T), (T, H)\}, E^c = \{(T, T)\}$.
11. E is contained in F, denoted by $E \subset F$, meaning all of the outcomes in E are also in F. $F = \{(H, H), (H, T), (T, H)\}, E = \{(H, T), (T, H)\}$.

Venn diagram

Venn diagram is a graphical representation for illustrating logical relations among events i.e., union, intersection, complement, containment.



Conditional probability

1. A small business with 120 workers and 2 employee benefits, A and B. (Venn diagram in Figure)

	A	A^c	
B	64	34	98
B^c	7	15	22
	71	49	120

2. Calculate the prob. that an employee chosen at random has the retirement benefit (A) given that she takes the health-care benefit (B). Restrict our attention to the rst row: 64/98 If the sample space S is finite, and all outcomes are equally likely, it is often convenient to compute $P(E \mid F)$ by using F as the sample space as

$$P(E \mid F) = |E \cap F|/|F|$$

   Count $|E \cap F|$ in the reduced sample space F. (Note that $E \cap F$ is an event in F since $E \cap F \subset F$.)

3. While Biopsy is accurate for breast cancer diagnosis, but expensive and invasive. An alternative is to use mammogram which is noninvasive and less expensive. The idea is to have mammogram first and then decide whether to take biopsy based on mammogram result. A = {mammogram+}, B = {breast cancer} Conditional probability of B given A: $P(B \mid A) = P(A \cap B)/P(A)$. Relative risk of B given A: $P(B \mid A)/P(B \mid A^c)$. (You'll see later that If two events A and B are independent, then the relative risk is 1).

   Suppose that among 100,000 women with negative mammogram 20 will have breast cancer diagnosed within 2 years, or $P(B \mid A^c) = 20/10^5 = 0.0002$, whereas 1 woman in 10 with positive mammograms will have breast cancer diagnosed within 2 years, or $P(B \mid A) = 0.1$. The two events A and B are highly dependent, since $P(B \mid A)/P(B|A^c) = .1/.0002 = 500$.

4. Total Probability Rule: For any events A and B, $P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$
   Example. Cancer A: mammogram positive
   B: developing breast cancer in next 2 years
   Suppose that 7% of the general population of women will have a positive mammogram. What is the probability of developing breast cancer over the next 2 years among women in the general population?
   P(breast cancer | mammogram+) = .1

P(breast cancer | mammogram-) = .0002
P(B) = P(breast cancer)
= P(breast cancer | mammogram+)P(mammogram+) + P(breast cancer | mammogram-)P(mammogram-) = .1(.07) + .0002(.93) = 0.00719

Independence

Two events, E and F, are independent if knowing that one had occurred gave us no information about whether the other had or had not occurred. If

$$P(E) = P(E \mid F),\ P(F) = P(F \mid E)$$

then E is independent on F. If $P(E) \neq P(E \mid F)$, then E is dependent on F.

Definition 1 *Two events E and F are independent if $P(E \cap F) = P(E)P(F)$. Two events E and F that are not independent are said to be dependent.*

1. A card is randomly selected from a deck. Let E denote the event that the card is an ace and F the event that it is a spade. Are E and F independent? Knowing that the card is an ace gives no information about its suit. $P(E) = 4/52, P(F) = 13/52$, and $P(E \cap F) = 1/52$ $P(E \cap F) = P(E)P(F)$

2. Suppose we are conducting a hypertension-screening program in the home. Consider all possible diastolic blood-pressure measurements from a mother and her first-born child. Let A = {mothers DBP $>$ 95}, B = {first-born childs DBP $>$ 80} Suppose $P(A \cap B) = 0.05, P(A) = .1, P(B) = .2$ Then $P(A \cap B) = .05 > P(A) \times P(B) = .02$, and the events A, B would be dependent.

3. Multiplication law of probability: If $A_1, A_2, \ldots, A_k$ are mutually independent events, then

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2) \cdots P(A_k)$$

4. Addition law of probability: In the previous example, what is $P(A \cup B)$, the probability that either mother's DBP $>$ 95 or the first born child's DBP $>$ 95? $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .1 + .2 - .05 = .25$

Predictive values

Predictive value positive (PV+): P(disease | test+)
Predictive value negative (PV-): P(no disease | test-)
In the last example, PV+ = P(breast cancer|mammogram+) = .1 PV- = P(no breast cancer | mammogram-) = 1- P(breast cancer | mammogram-) = 1-.0002 = .9998

Sensitivity and specificity

Sensitivity of a symptom is the probability that the symptom is present given that the person has a disease = P(symptom | disease)
Specificity of a symptom is the probability that the symptom is not present given that the person does not have a disease = P(no symptom | no disease)
A false negative is defined as a person who tests out as negative but who is actually positive.
A false positive is defined as a person who tests out at positive but who is actually negative.

Bayes' Rule

Let A = symptom and B = disease. Then

$$PV+ = P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

This can be written as

$$PV+ = \frac{sensitivity \times x}{sensitivity \times x + (1 - specificity) \times (1 - x)}$$

where $x = P(B)$ = probability of disease in the reference population.
Example: Cancer Suppose the disease is lung cancer and the symptom is cigarette smoking. If we assume 90% of people with lung cancer and 30 % of people without lung cancer are smokers, What is the sensitivity and specificity? Symptom: smoking, Disease: lung cancer
Sensitivity = P(symptom | disease) = .9
Specificity = P(no symptom | no disease) = 1- P(symptom | no disease) = .7