2.10 The law of total probability and Bayes' rule

- Definition 2.11. For some positive integer k, let the sets $B_1, B_2, \ldots, B_k$ be such that
  1. $S = B_1 \cup B_2 \cup \ldots \cup B_k$,
  2. $B_i \cap B_j = \emptyset$ if $i \neq j$

  Then the collection of sets $\{B_1, B_2, \ldots, B_k\}$ is said to be a "partition" of S.
- If A is any subset of S, and $\{B_1, B_2, \ldots, B_k\}$ is a partition of S, A can be "decomposed" as:
  $A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k)\}$
  * See the Venn Diagram
- *Theorem 2.8. (The law of total probability) Assume that $\{B_1, B_2, \ldots, B_k\}$ is a partition of S such that $P(B_i) > 0$ for $i=1,\ldots,k$. Then for any event A*
  $P(A) = \Sigma_{i=1,\ldots,k} P(A|B_i)P(B_i)$
  - Proof: apply the additive law
  - Sometimes, it's easier to calculate $P(A|B_i)$ for a suitably chosen partition than to compute P(A) directly.
- *Theorem 2.9. (Bayes' Rule) Assume $\{B_1, B_2, \ldots, B_k\}$ is a partition of S such that $P(B_i) > 0$ for $i=1,2,\ldots,k$. Then*

$$P(B_j \mid A) = \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i) P(B_i)}.$$

- Example 2.23. (An electronic fuse) 5 production lines produce fuses at the same production rate, with 2% defect rate except line 1 with 5% defect rate. A customer tested three fuses and one of them failed. What is:
  P(the lot was produced in line 1|the data)=?
  P(the lot was produced in one of lines 2-4|the data)=?

- HW. Some of the exercises 2.98~116
- Keywords: the law of total probability; Bayes rule

2.11 Numerical events and random variables

- Events of major interest are "numerical events"
- Define a variable Y that is a function of the sample points in S
- {All sample points where Y=a} is the numerical event assigned to number a.
- The sample space S can be partitioned into mutually exclusive sets of points assigned to the same value of Y
- Definition 2.12. A "random variable" is a real-valued function for which the domain is a sample space.
- Convention: We let y denote an observed value of Y.
  Then $P(Y=y)= \sum\{P(E_i)$: i such that E_i is assigned to y$\}$. Formal definition comes later…
- Example 2.24. Tossing two coins. Y=# of heads. The sample points in S? $Y(E_i)$=? Sample points corresponding to {Y=y}? What is P(Y=y) for each value of y?

2.12 Random sampling

- Population vs. sample (=observations of the values of random variables)
- Sampling with/without replacement affect probabilities of outcomes
- Design of experiment is the method of sampling
- Definition 2.13. In sampling n elements from a population with N elements, if the sampling is conducted in such a way that each of NCn samples has an equal probability of being selected, the sampling is said to be "random" and the result is said to be a "random sample"
- How to do random sampling?
  - Low-tech method (e.g. drawing tickets from a jar after shaking it)
  - The random number table (Table 12)
  - Use computer (In R, run sample(1:1000, 100))
- Sometimes we don't want a completely random sample

3 Discrete Random Variables and Their Probability Distributions

3.1 Basic definition

- Definition 3.1 A random variation Y is said to be "discrete" if it can assume only a finite or countably infinite number of distinct values.
- Knowledge of the probability distributions for random variables associated with common types of experiments will eliminate the need for solving the sample problems over and over again.

3.2 The probability distribution for a discrete random variable

- You have to know the difference between Y and y.
- (Y=y) = "the set of all points in S assigned the value y by the random variable Y"
- Definition 3.2 "P(Y=y)" or "p(y)" is defined as the sum of the probabilities of all sample points in S that are assigned the value y. p(y) is sometimes called the "probability function" for Y
- Definition 3.3 The "probability distribution" for a discrete variable Y can be represented by a formula, table, or graph that provides p(y)=P(Y=y) for all y.
- Note $p(y) \geq 0$ for all y. Assume p(y)=0 for any values y not explicitly assigned a positive value.
- Example 3.1 Find the probability distribution for Y=# of women in two workers randomly selected from 3 men and 3 women workers. Interpret the result. Also derive
  table,
  graph,
  and formula
  for p(y).
- *Theorem 3.1 For any discrete probability distribution, the following holds:*
  1. $0 \leq p(y) \leq 1$ *for all y*
  2. $\sum_y p(y)=1$
- A "simulation study" would give a relative frequency histogram that is similar to the probability distribution
- HW. Some of the exercises 3.1~9
- Keywords: P(Y=y)=p(y); probability distribution

3.3 The expected value of a random variable or a function of a random variable

- Definition 3.4 For a discrete random variable Y with the probability function p(y), the "expected value" of Y, E(Y) is defined to be

  $E(Y) = \sum_y yp(y)$
- If p(y) is an accurate characterization of the population frequency distribution, then E(Y)=μ, the population mean
- This definition is consistent with the definition of the mean of a set of measurements (Definition 1.1)
- What about the mean of Y^2? The mean of $(Y-\mu)^2$?
- *Theorem 3.2 Let Y be a discrete random variable with probability function p(y) and g(y) be a real-valued function of Y. Then the expected value of g(Y) is given by*

  $E[g(Y)] = \sum_y g(y)p(y)$
  - Note this is not a definition
  - Proof: The trick is to define G=g(Y) that takes on values g1,…,gm and express $P(G=g_i)=p^*(g_i)$ in terms of $p(y_j)$
- Definition 3.5 The variance of a random variable Y is defined to be

  $V(Y) = E[(Y-\mu)^2]$.

  The "standard deviation" of Y is the positive square root of V(Y).
- If p(y) is an accurate characterization of the population frequency distribution, then V(Y) =σ^2 is the population variance and σ is the population SD.
- Example 3.2. Find the mean, variance and standard deviation of Y in the above example.
- In the following theorems, we assume Y is a discrete random variable with probability function p(y)
- *Theorem 3.3 Let Y be a discrete random variable with probability function p(y). For any constant c, E(c)=c.*
- *Theorem 3.4. Let Y be a discrete random variable with probability function p(y). For a function g(Y) of Y and a constant c, E[cg(Y)] = cE[g(Y)].*
- *Theorem 3.5. Let Y be a discrete random variable with probability function p(y). For functions $g_1(Y), g_2(Y),\ldots,g_k(Y)$ of Y,*

  $E[g_1(Y)+g_2(Y)+\ldots+g_k(Y)]=E[g_1(Y)]+E[g_2(Y)]+\ldots+E[g_k(Y)]$.
- *Theorem 3.6. Let Y be a discrete random variable with probability function p(y). Then,*

  $V(Y) = \sigma^2=E[(Y-\mu)^2] = E(Y^2)-\mu^2$.
  - This makes variance computation of example 3.2 easier.

- Example 3.4 The expected daily cost of two machines A and B.

- HW. Some of the exercises 3.10~24
- Keywords: E(Y); V(Y); E(g(Y))