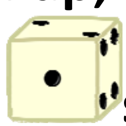
Chapter 2 Class Notes

- **Probability can be thought of in many ways, for example as a *relative frequency* of a long series of trials (e.g. flips of a coin or die)**
- **Another approach is to let an expert (such as an MD) set his/her *subjective probability*, but neither this nor the former are very practical, so we'll use the *mathematical* or *axiomatic approach* given below**
- **The theoretical relative frequency (i.e., the assumed probability distribution) for the number of dots showing in a single toss of a fair die are on p.22; whether a given die is deemed to be "fair" or "unfair" depends upon the degree to which the actual the theoretical match up; e.g., if we tossed the die 10 times and got all s, we'd think something is wrong**
- **Sometimes results fall in a "grey area" and we need a probability structure to help us decide what is truly "unlikely"; for example, the probability that *at least two students in this room have the same birthday* is 62.7%. Using R software, we get:**

  > 1-(365*364*363*362*361*360*359*358*357*356*355*354*353*352*351*350*349*348*347*346*345*344*343*342*341*340*339)/(365^27)
  [1] 0.6268593
- **If time, can do p.26 #2.4 and 2.5 in class**

Section 2.3

- **In §2.3, note we'll use set notation such as A for the set of elements {a_1, a_2,…}; the a_k are the elements of A So, for a single toss of a die, A = {1,2,3,4,5,6}, whereas for a single draw from a fair deck of cards, B = {2,3,4,5,6,7,8,9,10,J,Q,K,A of each of ♠, ♣, ♥,♦}**
- **Inclusion: $A \subset B$ if all elements from A are in B; see Figure 2.2 on p.23**
- **Union: $A \cup B$ is the set of all elements in A (alone) or B (alone) or both; see Figure 2.3 on p.23: key word OR**
- **Intersection: $A \cap B$ (or AB) is the set of all elements in both A and B; see Figure 2.4 on p.24: key word AND**
- **Complement: $\overline{A}$ (or A^c) is every element in the sample space (S) not in A: key word NOT**
- **Mutually Exclusive: A and B are M.E. if no elements are in both A and B: $A \cap B = \emptyset$ ($\emptyset$ is the empty set)**
- **DeMorgan's Laws: $\overline{A \cap B} = \overline{A} \cup \overline{B}$ and $\overline{A \cup B} = \overline{A} \cap \overline{B}$**

Section 2.4

- **Some Definitions:**
  - **an experiment is the process by which an observation is made**
  - **the outcome of an experiment is called an event (these are either simple or compound)**
  - **a simple event is an event that cannot be**

decomposed since it has only one sample point; **compound events** can be decomposed

- the **sample space** (S) associated with an experiment is the set of all possible sample points
- a **discrete sample space** is one that contains either a finite or a countable number of sample points
- an **event** in a discrete sample space S is a collection of sample points, i.e., ***any subset of S***

- Two examples are on p.29
- Key definition (p.30): Let S be the sample space associated with an experiment. To every A (subset) of S, we assign a number, P(A), called the **probability** of A, such that these axioms hold:
  - Axiom 1: $P(A) \geq 0$
  - Axiom 2: $P(S) = 1$
  - Axiom 3: if $A_1, A_2, A_3, \ldots$ are mutually exclusive events in S, then $P(A_1 \cup A_2 \cup \ldots) = \sum_{k=1}^{\infty} P(A_k)$
- Notice that these axioms are more conditions on the probability P(•) operator – they don't tell us how to assign probabilities, but only the conditions it must satisfy to be a valid probability function

Section 2.5

- Two methods to find probabilities in finite/countably infinite (denumerable) cases: sample-point and event-composition, and we discuss the sample-point now

- **Sample-point method steps** (see p.36):
  - Define the experiment and clearly determine how to describe on simple event
  - List the simple events (and make sure simple): this defines S [Note: sometimes this is tough.]
  - Assign 'reasonable' probabilities making sure these are non-negative and sum to 1
  - Define event of interest A as a specific collection of simple events – make sure you have them all
  - Find P(A) by summing the respective probabilities
- **Clearly work through examples 2.2 (p.36), 2.3 (p.37) and 2.4 (p.38), and if time, p.40 #2.31**

Section 2.6

- **Permutations and combinations permit us to count sample points more quickly and accurately**
- ***Theorem 2.1*** **states that given m elements $a_1, a_2, \ldots a_m$ and n elements $b_1, b_2, \ldots b_n$, then the total number of pairs is mn; this result extends to three or more factors**
- **Examples are: tossing a die twice (36 sample points) or k times (6^k sample points); tossing a coin 3 times (2^3 = 8 sample points); calculating birthday probabilities in Ex. 2.7: room of 20 people with 365 distinct birth 'scores', so each sample point is a 20-tuple of 1 to 365, all birthdays are different in $n_a = 365*364*\ldots*346$ ways, so probability is $n_a / 365^{20} = 0.5886$**

- **With help from R:**

```
permm=function(n,r) factorial(r)*choose(n,r)
> permm(365,20)/(365^20)
[1] 0.5885616
```

- **Notice for the last example, it would be virtually impossible to *list* all the elements in S, so we clearly need this "mn rule"**
- **An *ordered* arrangement of r distinct objects is called a permutation, and the number of ways of ordering n distinct objects taken r at a time is denoted P_r^n**
- ***Theorem 2.2* (p.43) states that $P_r^n = \frac{n!}{(n-r)!}$**
- **Example 2.8: in an office of 30 employees, 3 are chosen – without replacement – and the first wins \$100, 2nd wins \$50, and 3rd wins \$25; there are then $P_3^{30} = 30 \times 29 \times 28 = 24360$ sample points in S**
- **It can also be the case that r = n as in Ex.2.9 p.44 (machine parts in assembly), and note that $P_n^n = n!$**
- ***Theorem 2.3*: the number of ways of partitioning n distinct objects into k distinct groups containing n_1, n_2, … n_k objects respectively (such that $n_1 + n_2 + \ldots + n_k = n$ and so each object appears in exactly one group) is**

$$N = \binom{n}{n_1\, n_2 \ldots n_k} = \frac{n!}{n_1!\, n_2! \ldots n_k!}$$

- **It's instructive to understand why, in this formula, the *n!* is divided by the $n_1!\, n_2! \ldots n_k!$ – see the proof on p.44**
- **Note the technique used in <u>Ex.2.10 on p.45</u>: the 20 laborers are divided into jobs of 6, 4, 5 and 5 laborers in jobs I, II, III and IV respectively, and I is the very undesired of these jobs. So the number of sample points in total is $N = \binom{20}{6\ 4\ 5\ 5} = \frac{20!}{6!4!5!5!}$ As it turns out, all 4 people of a certain group were placed into job I; if we assume that this placement was at random, then the probability of this event A occurring is these number of sample points (n_a) times 1/N. Also, since $n_a = \binom{16}{2\ 4\ 5\ 5} = \frac{16!}{2!4!5!5!}$, then P(A) = n_a/N = 0.0031**
- **Since this is indeed very small, we can guess (or hypothesize) that perhaps the assignment of these four individuals was not done at random**
- **Some help from R:**

```
>N=factorial(20)/(factorial(6)*factorial(4)*factorial(5)*
factorial(5))
>na=factorial(16)/(factorial(2)*factorial(4)*factorial(5)
*factorial(5))
> na/N
[1] 0.003095975
```

- **In the case of permutations, order matters, but in some cases, such as playing cards, {AK} and {KA} are** really just one Ace and one King - that is, order doesn't matter. This brings us to combinations instead of permutations

- The number of combinations of n objects taken r at a time, denoted C_r^n or $\binom{n}{r}$, is the number of subsets, each of size r, that can be formed from the n objects
- *Theorem 2.4* points out that $\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$
- Aside: Just as the term $\binom{n}{n_1\ n_2\ \ldots n_k}$ is the coefficient of $y_1^{n_1}y_2^{n_2}\ldots y_k^{n_k}$ in the expansion of $(y_1 + y_2 + \cdots + y_k)^n$, so too the term $\binom{n}{r} = \binom{n}{r\quad n-r}$ is the coefficient of $x^r y^{n-r}$ in the expansion of $(x+y)^n$ – the former of these is called a multinomial coefficient and the latter is called a binomial coefficient
- Example 2.11 – the number of ways of choosing 2 individuals from a group of 5 is $\binom{5}{2}$ = 10 just as on p.36
- Example 2.12 – five people are rank in terms of test scores as I, II, III, IV, V; we choose two of these at random, and let A be the event that we choose exactly one of I or II. The number of elements in A is $\binom{2}{1} \times \binom{3}{1}$ = 2*3 = 6 since we want exactly 1 of the 2 in {I,II} and exactly 1 of the 3 in {III,IV,V} – gotten by multiplying

the results by the above 'mn rule'. By the last example, it follows that P(A) = 6/10 = 0.60

- **Example 2.13** **– a company orders supplies from M distributors and places n orders (with n < M). With no restrictions on the orders, let's find the probability that distributor 1 gets exactly k orders (with k ≤ n): since each of the M distributors could get up to n orders, the number of different ways the orders could be assigned is M^n. The number of ways the n orders can result in k orders to distributor 1 is $\binom{n}{k}$, and there remains (n-k) orders of the remaining (M-1) vendors, resulting in $(M-1)^{n-k}$ ways, so the number of successful outcomes is $\binom{n}{k} \times (M-1)^{n-k}$. Putting this together, the desired probability is $\frac{\binom{n}{k}\times(M-1)^{n-k}}{M^n}$**
- **Some help from R:**

```
> ex2p13=function(M,n,k) {
+ top=choose(n,k)*(M-1)^(n-k)
+ bot=M^n
+ top/bot }
> ex2p13(6,4,0)
[1] 0.4822531
> ex2p13(6,4,1)
[1] 0.3858025
> ex2p13(6,4,2)
```

```
[1] 0.1157407
> ex2p13(6,4,3)
[1] 0.0154321
> ex2p13(6,4,4)
[1] 0.0007716049
```

Section 2.7

- **Illustration: In a single role of a fair die, the probability of a "1", 1/6, is an *unconditional probability*. Had we been given some prior knowledge such as that the toss was an odd number, then we would have then calculated the *conditional probability* of 1/3**
- **To make the previous example more formal, let A = {1} and B = {die is odd} = {1,3,5}, then the conditional probability of A given that B has occurred, denoted P{A|B}, is equal to $\frac{P(A \cap B)}{P(B)}$ provided that P{B} ≠ 0**
- **In the die example, it turns out that $A \cap B$ = A = {1} – this is usually not the case – so we get (as above):**

$$P(A|B) = \frac{1/6}{1/2} = {}^{1}/_{3}$$

- **Events A and B are called *independent* if any of these [equivalent] conditions is met:**
  - **P(A|B) = P(A)**
  - **P(B|A) = P(B)**
  - **P($A \cap B$) = P(A) × P(B)**

- Events that are not independent are ***dependent***
- To illustrate using p.55 ex.2.73, P(at least one R) = ¾, P(at least one r) = ¾, P(one r|Red flowers) = ⅔; since P(one r) = ½ ≠ ⅔, we conclude recessive allele and flower color are ***dependent*** (no surprise there)

Section 2.8

- **Multiplication Law of Probability** – for any two events A and B:

$$P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(A|B)$$

- Above extends to:

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times \ldots \times P(A_k|A_1 \cap A_2 \cap \ldots \cap A_{k-1})$$

- Events A and B are independent IFF

$$P(A \cap B) = P(A) \times P(B)$$

- The **Addition Law of Probability** states that for two events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Above extends:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- Further, for mutually exclusive events, $P\left(\cup_{j=1}^{k} A_j\right) = \sum_{j=1}^{k} P(A_j)$ - e.g., $P(A \cup B) = P(A) + P(B)$

- **For any event A, $P(\bar{A}) = 1 - P(A)$**
- **Example: p.60 ex.2.93 – wins with HHH or HHM or MHH, (which are obviously mutually exclusive) so add the probabilities: (0.70)(0.40)(0.70) + (0.70)(0.40)(0.30) + (0.30)(0.40)(0.70) = 0.364**
- **Another Example: p.61 ex.2.99**

Section 2.9

- **Event-composition method: decomposes an event into unions and/or intersections of other events; the steps, given on p.64, are: (1) define experiment, (2) visualize the sample points, (3) express the event A as union and/or intersection composition, (4) apply the additive and/or multiplication rules to get the answer**
- **Seven examples are given in the section – some are:**
- **Ex. 2.19 (p.64): A, the event that exactly one of the two best applicants is chosen, is decomposed into $B \cup C$ with B representing the best and one of the 3 poorest and C being the second best and one of the 3 poorest. Note that these two events are mutually exclusive, so P(A) = P($B \cup C$) = P(B) + P(C). We next argue that P(B) = 2 × ⅕ × ¾ = 3/10 since we want both the best applicant chosen on the first draw (B_1) and one of the 3 poorest applicants chosen on the second draw (B_2) or vice versa; 'vice versa' here means getting**

one of the 3 poorest on the first draw (B_3) and the best on the second draw (B_4). Note that $P(B_1 \cap B_2)$) = $P(B_1) \times P(B_2|B_1)$ = ⅕ × ¾ = 3/20 and $P(B_3 \cap B_4)$) = $P(B_3) \times P(B_4|B_3)$ = ⅗ × ¼ = 3/20. A very similar argument shows P(C) = 3/10, so P(A) = ⅗.

- Ex. 2.20 (p.65): Let R be the event that a patient with the disease responds. In the 3 patients, the probability that at least one responds is 1 – probability that none responds = $1 - P(\bar{R}_1 \cap \bar{R}_2 \cap \bar{R}_3)$ = (by independence) $1 - [P(\bar{R})]^3$ =1 – $(0.10)^3$ = 0.999
- Ex. 2.22 (p.67): A_1 is the event that the color matches on the red box; A_2 and A_3 correspond to the black and white boxes. There are 3! = 6 ways to arrange the 3 balls; two of these match, so $P(A_1) = P(A_2) = P(A_3)$ = 2/6 = ⅓. All pairwise intersections and $A_1 \cap A_2 \cap A_3$ are equivalent to *all 3 balls matching*, so $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = P(A_1 \cap A_2 \cap A_3)$ = ⅙, so the probability of *at least one match* is $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$ =1 – 3(⅙) + ⅙ = ⅔. (a) It follow the probability of *no matches* is 1 - ⅔ = ⅓. (b) Notice that exactly one match is {at least one match} – {at least two matches}, and by the above {at least two matches} = {all three match}, so the probability here is ⅔ - ⅙ = ½.

- Ex. 2.21 (p.66) IFF enough time (geometric).

Section 2.10

- **Law of Total Probability**: Let the sets B_1, B_2 … B_k be such that they are pairwise mutually exclusive (each $B_i \cap B_j = \emptyset$) and $S = B_1 \cup B_2 \cup \ldots \cup B_k$, then {$B_1$, B_2 … B_k} is called a partition of S.
- For any subset A of S and partition {B_1, B_2 … B_k}, A can be decomposed as

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots (A \cap B_k)$$

- Let {B_1, B_2 … B_k} be a partition of S such that for all i = 1, 2 … k, $P(B_i) > 0$, then for any event A

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

- See the diagram of this result in Figure 2.12, p.71
- **Bayes' Rule**: Let {B_1, B_2 … B_k} be a partition of S such that for all i = 1, 2 … k, $P(B_i) > 0$ and A is any event, then

$$P\left(B_j|A\right) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}$$

- Ex. 2.23 (p.71): B is the event that a fuse came from line 1 (so $\bar{B}$ from lines 2 – 5), and A is the event that a fuse is defective. Then P(B) = 0.20, P(A|B) = $3(0.05)(0.95)^2$ = 0.135375 and P(A|$\bar{B}$) = $3(0.02)(0.98)^2$ = 0.057624. Since $\{B, \bar{B}\}$ is a partition of S, P(A) = $P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$

= (0.135375)(0.2) + (0.057624)(0.8) = 0.0731742,

and therefore by Bayes' Rule,

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{(0.135375)(0.2)}{0.0731742} = 0.37$$

Also, $P(\overline{B}|A)$ = 1 - 0.37 = **0.63**.

Section 2.11 Numerical Events and Random Variables

- A ***Random Variable*** is a real-valued function for which the domain is a sample space (S)
- Example 2.24 (p.76): Tossing a coin twice and letting Y be the number of HEADS (H), then
  - {Y = 0} corresponds to {TT} = E_4
  - {Y = 1} corresponds to {HT,TH} = $E_2 \cup E_3$
  - {Y = 2} corresponds to {HH} = E_1
- Example 2.25 (p.77): In the last example, it follows that
  - P(0) = P(Y = 0) = ¼
  - P(1) = P(Y = 1) = ¼ + ¼ = ½
  - P(2) = P(Y = 2) = ¼
- Another example is the monkey example, Ex.2.22 on p.67 – here, Y is the number of matches of ball and box color, and the values of Y and probabilities are
  - P(0) = P(Y = 0) = ⅓
  - P(1) = P(Y = 1) = ½

- **P(2) = P(Y = 2) = 0**
- **P(3) = P(Y = 3) = ⅙**

- **As we'll see in the next chapter, sometimes the assignment of probabilities to a random variable may be by a real-valued function. One such example is given in Ex.2.21 (p.66), where for r = 1, 2, ….**

$$P(Y = r) = \left({}^{5}/_{6}\right)^{r-1}\left({}^{1}/_{6}\right)$$

Section 2.12 Random Sampling

- **In Statistics, we draw a sample from a population, and use what we observe in the sample to make inferences regarding the population**
- **From a population of N = 5 elements, a sample of n = 2 is taken; if sampling is done without replacement, then each pair has a 1/10 = 0.10 chance of being selected, whereas if sampling is done with replacement, then each pair has a 2/25 = 0.08 chance of being selected**
- **The specific selection method of the sample is called the design of (the) experiment**
- **For N and n representing the population and sample size resp., if the sampling is conducted in such a way that each of the $\binom{N}{n}$ samples is equally likely, then the sampling is called simple random sampling (SRS)**

- **SRS is tough to achieve in practice – although a table of random digits can help (see discussion on pp.78-9) – and we are often lead to use other types of sampling methods such as cluster sampling, stratified sampling, hierarchical sampling, etc. Best to learn more about these by taking a Sampling course.**

Additional Examples

- **p.81, ex.2.151**

```
world.series.4.games = function(p) p^4+(1-p)^4
world.series.5.games =
     function(p) 4*( (p^4)*(1-p) + ((1-p)^4)*(p) )
world.series.6.games =
     function(p) 10*( (p^4)*((1-p)^2) + ((1-p)^4)*(p^2) )
world.series.7.games =
     function(p) 20*( (p^4)*((1-p)^3) + ((1-p)^4)*(p^3) )
world.series.4.games(0.5)
0.125
world.series.5.games(0.5)
0.25
world.series.6.games(0.5)
0.3125
world.series.7.games(0.5)
[0.3125
```

- **p.80, ex.2.146**
- **p.81, ex.2.153**