

**POL 602**  
**Prof. Matthew Lebo**  
**Week 2 – September 6 and 8, 2005**

**Probability Theory I**

Although people have been concerned with chance in the context of gambling games since before the dawn of recorded history, the philosophical and mathematical investigation of probability is fairly recent, beginning only in the 1660's. [For an excellent discussion of the early history of probability see Hacking's (1975) *The Emergence of Probability*.]

The early theorists of probability, whose work is now referred to as Classical Theory, were primarily concerned with games of chance involving "fair" devices such as dice, cards and coins.

A important issue for classical theory was the assumption of fairness, or *equiprobability of outcomes*, first justified by appeal to physical features of the gambling tools such as symmetry or evenly-distributed weight, and later formalized by an axiom referred to today as the Principle of Indifference (or the Principle of Insufficient Reason as Keynes put it).

*Principle of Indifference*: This principle essentially holds that alternatives are to be considered equiprobable if there is no reason to expect or prefer one over the other.

There are two problems, in the modern view, with the principle. The first concerns the circular definition of probability in terms of equiprobability (i.e. events are assigned probability based on their symmetric probability). In other words, we define how probable certain outcomes are based on assumptions about how probable the outcomes are. Rather tautological.

The second problem is that the Principle of Indifference confounds at least two different senses of prior ignorance (Smithson 1989). On the one hand, we know (given no evidence to the contrary) that all events are equally likely. However, on the other hand, we do not know how likely any of them are.

Example: You are about to throw a standard six-sided die, when the person sitting next to you whispers that the die is loaded (i.e. unbalanced in weight such that some set of sides are favored over others). By the principle of indifference, we should still regard the outcomes as equiprobable, since *we don't know in what way the die is loaded*.

Another way to think about this problem is the assignment of probabilities to outcomes more complex than simple gambling devices. We have a feeling that the outcome of a positive test result from a medical procedure, or that a friend will get married in the next 3 years, are not equiprobable with their converse (a negative result, or unmarried). But how do we assign probabilities to the proper outcome?

Throughout the history of probability theory, there have been several responses to the dilemmas of classical theory. We will examine the two dominant responses: the *relative frequency* or *frequentist* approach, and the *subjective probability* or *Bayesian* [after Rev. Thomas Bayes, some of whose results we will examine in detail later] approach.

### ***Relative Frequency or the Frequentist Approach***

Historians credit Venn, Von Mises and Reichenbach as the founders of the frequentist approach.

In simple terms, this approach holds that the probability that a specific outcome obtains can be interpreted to mean the relative frequency with which that outcome would obtain if the process were repeated a large number of times under similar conditions. (note that “large number” and “similar conditions” are rather vague terms in this simple statement of theory).

In the frequentist approach, then, we assign the probability of  $\frac{1}{2}$  to getting heads when flipping a fair coin not because of an assumption about equiprobability but because if we were to flip the coin many, many times we expect the proportion of heads to be  $\frac{1}{2}$ . This is not to say that we expect exactly  $\frac{1}{2}$  of any number ( $n > 1$ ) of flips to result in heads. In fact, we would be quite surprised if this outcome obtained consistently. This fact emphasizes the importance of a random or stochastic element to our coin flips.

Note also that the frequentist approach can be problematic in some cases, such as the probability of a friend getting married in the next three years mentioned earlier. There is no empirical way (or even a reasonable thought experiment) to have the friend relive the next three years a large number of times for us to determine the relative proportions.

### ***Bayesian Approach***

According to the subjective interpretation of probability, the probability that a person assigns to some outcome represents her own judgment of the likelihood that the outcome will occur.

Thus, if you and I have different beliefs, we will assign different *subjective probabilities* to a set of outcomes.

This approach is a powerful way to mathematically incorporate the subjectivity of scientific research into the data analysis process, but it is not without its limitations.

First, it can be difficult in many real world applications to express prior beliefs (or *priors*) in numerical terms that allow for the required calculations. Moreover, it is unlikely that all people or researchers have beliefs about the world that meet the strict requirements of probability theory. Additionally, until very recently, much Bayesian analysis was practically impossible due to computational limitations.

In this course, as in modern political science, the dominant approach to statistical inference is the Frequentist one.

It may be that in the course of our careers this may change, as recent trends in journal publications indicate, so I will try to introduce or discuss Bayesian terminology when I can.

Why might this occur? Because the assumption of the Frequentists about repeated sampling and statistical inference are so frequently broken.

Often, we are not dealing with a sample, but have the whole “universe” of cases. Such as studies of the 50 American States, or of the bills passed in the 108<sup>th</sup> Congress, or the European democracies. So what exactly are we inferring to?

The most important point to remember is that, with the exception of how probabilities are to be interpreted (and sometimes assigned), both approaches have in common the same mathematical framework of probability regarding two central questions:

1. How are the probabilities of certain events determined given the specified probabilities of each possible outcome of an experiment (i.e. after the frequentist or Bayesian assignment).
2. How are the probabilities of certain events revised or updated after additional relevant information is obtained.

It is these two questions, which require a mathematical theory of probability, that we will concern ourselves with for the remainder of today and all of next class.

The first tool in developing the theory we need is set theory, which you have already reviewed in the math course.

One terminology difference that is fairly standard across presentations of probability theory is that the universal set is now referred to as the *sample space* ( $S$ ) and is defined as the set of all possible outcomes of an experiment.

Thus the sample space for the roll of a six-sided die is:

$$S = \{1, 2, 3, 4, 5, 6\}$$

I should note also that the textbook uses the bar notation for complements ( $\bar{A}$ ).

Other results from set theory that the text presents that you may not be familiar with are *De Morgan's Laws*:

$$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$$

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$$

A few more definitions are in order before proceeding to the axioms of probability.

First, an *experiment* is any process by which an observation is made.

*Events* are the possible outcomes of an experiment.

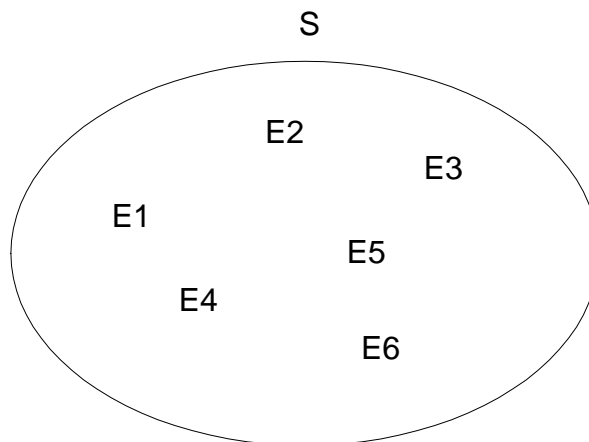
*Simple events* ( $E$ ) are those which can not be decomposed.

They correspond to one *sample point*. The sample space,  $S$ , is the set of all possible sample points.

A *discrete sample space* is one that contains either a finite or denumerably infinite (countable) number of distinct sample points.

*Compound events* are sets of sample points or unions of simple events.

Example: Venn diagram of the die-tossing experiment



Simple Events:

Rolling a 1 =  $\{E_1\}$

Rolling a 2 =  $\{E_2\}$  etc...

Compound Events:

Rolling an odd number, event  $A = \{E_1, E_3, E_5\}$

Rolling  $>4$ , event  $B = \{E_5, E_6\}$

We can now present the axioms of probability. (Note that there are several orders of increasing mathematical sophistication with which these axioms can be presented):

Suppose  $S$  is a sample space associated with an experiment. To every event  $A$  in  $S$  ( $A \subset S$ ), we assign a number  $P(A)$  [often written  $\text{Pr}(A)$ ] called the probability of  $A$ , such that:

1.  $P(A) \geq 0$
2.  $P(S) = 1$
3.  $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$  for all disjoint events  $A_1, A_2, \dots$

Thus a probability distribution on a sample space  $S$  is a specification of numbers  $P(A_i)$  which satisfy axioms 1, 2, and 3.

From these simple statements, all the other properties of the probability function can be derived.

Note that this does not tell us how to assign particular probabilities to simple events (this is the role of classical, frequentist or Bayesian theory).

What are some of the other basic properties of Probabilities? We'll come back to some of these in more detail.

$P(\bar{A}) = 1 - P(A)$  the probability of not- $A$  is equal to 1 minus the probability of  $A$ .

$P(\emptyset) = 0$  the probability of an event not in the sample space (the null set) is 0.

If  $A \subset B$ , then  $P(A) \leq P(B)$  if  $A$  is a component of  $B$ , the probability of  $A$  is less than or equal to the probability of  $B$ .

For any event  $A$ ,  $P(A) \leq 1$ .

Another expression of rule 3 above:

Let  $A_1, A_2, \dots, A_n$  be events defined over  $S$ .

$$\text{If: } A_i \cap A_j = \emptyset \text{ for } i \neq j, \text{ then } P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

An example using some of these rules:

Biff decides to apply to two colleges,  $X$  and  $Y$ . Based on how his friends have fared, he estimates that his probability of being accepted at  $X$  is 0.7 and at  $Y$  is 0.4. He also suspects that there is a 75% chance that at least one of his applications will be rejected. What is the probability that he gets at least one acceptance?

Let  $A$  be the event "School  $X$  accepts him" and  $B$ , the event that "School  $Y$  accepts him."

We are given that  $P(A) = 0.7$ ,  $P(B) = 0.4$ , and  $P(\bar{A} \cup \bar{B}) = 0.75$ .

We are looking for  $P(A \cup B)$ .

Which is:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

De Morgan's laws tell us that  $\overline{A \cap B} = \overline{A} \cup \overline{B}$ , so

$$P(A \cap B) = 1 - P(\overline{A \cap B}) = 1 - 0.75 = 0.25$$

We now have the 3 components we need – Biff's chances are pretty good with an 85% chance of getting in somewhere:

$$P(A \cup B) = 0.7 + 0.4 - 0.25 = 0.85.$$

**The Sample-Point Method:**

A “cookbook” method (see pg. 34):

1. Define the experiment and clearly determine how to describe one simple event.
2. List the simple events associated with the experiment and test each to make certain it cannot be decomposed. This defines sample space S.
3. Assign reasonable probabilities to the sample points in S in accordance with the 1<sup>st</sup> and 2<sup>nd</sup> axioms of probability.
4. Define the event of interest, A, as a collection of sample points.
5. Find P(A) by summing the probabilities of sample points in A (axiom 3).

Example (2.1) in the text (pg. 30)

Note that not all sample spaces are discrete (an experiment testing the response time of subjects to stimulus after viewing a negative political ad, for example). These will be discussed in a few weeks. For now, we will focus on methods of calculating the probabilities of events in discrete sample spaces.

We have 5 seemingly identical computer terminals that can be shipped – but two are defective.

If they are chosen randomly, what is the probability that a shipment of 2 will contain no defective terminals.

List the sample space: combinations of terminals that can be either D1, D2, G1, G2, or G3.

There are ten combinations.

A is the event that two are chosen without either being defective.

A is the set of  $E_8 = \{G_1, G_2\}$ ,  $E_9 = \{G_1, G_3\}$ ,  $E_{10} = \{G_2, G_3\}$

If they are each selected randomly the probability of each is  $1/10$  and the probability of A is  $1/10 + 1/10 + 1/10 = 3/10$ .

Example 2.3 on page 35 – tossing a fair coin 3 times. What is probability of 2/3 heads?

The probability is  $3/8$ .

Here's a digression from the textbook.

For an experiment where there are only two possible outcomes, we can use a shortcut to see how many possible combinations there can be and what the distribution of those combinations can be in terms of the number of sample points in each combination.

### Pascal's Triangle

Begin with the number 1 – if we flip a coin 0 times, there is only one outcome, the null set.

1

Add one to the unseen zero to left of one and again to the right of 1 to get the possible combinations of one flip.

1            1  
1            2            1

Then add the numbers in pairs, beginning with the unseen zero to the left of the first one in the second column.

1            1            1  
1            2            3            1

And so on....

1            1            3            6            10            15            21            28            36            45            55            66            78            91            105            120            136            153            171            190            210            231            253            276            300            325            351            378            406            435            465            496            528            561            596            632            669            707            747            788            830            873            917            962            1008            1056            1106            1157            1209            1262            1317            1373            1431            1490            1550            1611            1673            1736            1800            1866            1933            2001            2070            2140            2211            2283            2356            2430            2505            2581            2658            2736            2815            2895            2976            3058            3141            3225            3310            3396            3483            3571            3660            3750            3841            3933            4026            4120            4215            4311            4408            4506            4605            4705            4806            4908            5011            5115            5220            5326            5433            5541            5650            5760            5871            5983            6096            6210            6325            6441            6558            6676            6795            6915            7036            7157            7279            7402            7526            7651            7777            7904            8032            8161            8291            8422            8554            8687            8821            8956            9092            9229            9367            9506            9646            9787            9929            10072            10216            10361            10507            10654            10802            10951            11101            11252            11404            11557            11711            11866            12022            12179            12337            12496            12656            12817            12979            13142            13306            13471            13637            13804            13972            14141            14311            14482            14654            14827            15001            15176            15352            15529            15707            15886            16066            16247            16429            16612            16796            16981            17167            17354            17542            17731            17921            18112            18304            18497            18691            18886            19082            19279            19477            19676            19876            20077            20279            20482            20686            20891            21097            21304            21512            21721            21931            22142            22354            22567            22781            22996            23212            23429            23647            23866            24086            24307            24529            24752            24976            25201            25427            25654            25882            26111            26341            26572            26804            27037            27271            27506            27742            27979            28217            28456            28696            28937            29179            29422            29666            29911            30157            30404            30652            30901            31151            31402            31654            31907            32161            32416            32672            32929            33187            33446            33706            33967            34229            34492            34756            35021            35287            35554            35822            36091            36361            36632            36904            37177            37451            37726            38002            38279            38557            38836            39116            39397            39679            39962            40246            40531            40817            41104            41392            41681            41971            42262            42554            42847            43141            43436            43732            44029            44327            44626            44926            45227            45529            45832            46136            46441            46747            47054            47362            47671            47981            48292            48604            48917            49231            49546            49862            50179            50497            50816            51136            51457            51779            52102            52426            52751            53077            53404            53732            54061            54391            54722            55054            55387            55721            56056            56392            56729            57067            57406            57746            58087            58429            58772            59116            59461            59807            60154            60502            60851            61201            61552            61904            62257            62611            62966            63322            63679            64037            64396            64756            65117            65479            65842            66206            66571            66937            67304            67672            68041            68411            68782            69154            69527            69901            70276            70652            71029            71407            71786            72166            72547            72929            73312            73696            74081            74467            74854            75242            75631            76021            76412            76804            77197            77591            77986            78382            78779            79177            79576            79976            80377            80779            81182            81586            81991            82397            82804            83212            83621            84031            84442            84854            85267            85681            86096            86512            86929            87347            87766            88186            88607            89029            89452            89876            90301            90727            91154            91582            92011            92441            92872            93304            93737            94171            94606            95042            95479            95917            96356            96796            97237            97679            98122            98566            99011            99457            99904            100352            100801            101251            101702            102154            102607            103061            103516            103972            104429            104887            105346            105806            106267            106729            107192            107656            108121            108587            109054            109522            109991            110461            110932            111404            111877            112351            112826            113302            113779            114257            114736            115216            115697            116179            116662            117146            117631            118117            118604            119092            119581            120071            120562            121054            121547            122041            122536            123032            123529            124027            124526            125026            125527            126029            126532            127036            127541            128047            128554            129062            129571            130081            130592            131104            131617            132131            132646            133162            133679            134197            134716            135236            135757            136279            136802            137326            137851            138377            138904            139432            139961            140491            141022            141554            142087            142621            143156            143692            144229            144767            145306            145846            146387            146929            147472            148016            148561            149107            149654            150202            150751            151301            151852            152404            152957            153511            154066            154622            155179            155737            156296            156856            157417            157979            158542            159106            159671            160237            160804            161372            161941            162511            163082            163654            164227            164801            165376            165952            166529            167107            167686            168266            168847            169429            170012            170596            171181            171767            172354            172942            173531            174121            174712            175304            175897            176491            177086            177682            178279            178877            179476            180076            180677            181279            181882            182486            183091            183697            184304            184912            185521            186131            186742            187354            187967            188581            189196            189812            190429            191047            191666            192286            192907            193529            194152            194776            195401            196027            196654            197282            197911            198541            199172            199804            200437            201071            201706            202342            202979            203617            204256            204896            205537            206179            206822            207466            208111            208757            209404            210052            210701            211351            212002            212654            213307            213961            214616            215272            215929            216587            217246            217906            218567            219229            219892            220556            221221            221887            222554            223222            223891            224561            225232            225904            226577            227251            227926            228602            229279            229957            230636            231316            231997            232679            233362            234046            234731            235417            236104            236792            237481            238171            238862            239554            240247            240941            241636            242332            243029            243727            244426            245126            245827            246529            247232            247936            248641            249347            250054            250762            251471            252181            252892            253604            254317            255031            255746            256462            257179            257897            258616            259336            260057            260779            261502            262226            262951            263677            264404            265132            265861            266591            267322            268054            268787            269521            270256            270992            271729            272467            273206            273946            274687            275429            276172            276916            277661            278407            279154            279902            280651            281401            282152            282904            283657            284411            285166            285922            286679            287437            288196            288956            289717            290479            291242            292006            292771            293537            294304            295072            295841            296611            297382            298154            298927            299701            300476            301252            302029            302807            303586            304366            305147            305929            306712            307496            308281            309067            309854            310642            311431            312221            313012            313804            314597            315391            316186            316982            317779            318577            319376            320176            320977            321779            322582            323386            324191            325007            325824            326642            327461            328281            329102            329924            330747            331571            332396            333222            334049            334877            335706            336536            337367            338199            339032            339866            340701            341537            342374            343212            344051            344891            345732            346574            347417            348261            349106            349952            350799            351647            352496            353346            354197            355049            355902            356756            357611            358467            359324            360182            361041            361901            362762            363624            364487            365351            366216            367082            367949            368817            369686            370556            371427            372299            373172            374046            374921            375797            376674            377552            378431            379311            380192            381074            381957            382841            383726            384612            385501            386391            387282            388174            389067            389961            390856            391752            392649            393547            394446            395346            396247            397149            398052            398956            399861            400767            401674            402582            403491            404401            405312            406224            407137            408051            408966            409882            410799            411717            412636            413556            414477            415399            416322            417246            418171            419097            420024            420952            421881            422811            423742            424674            425607            426541            427476            428412            429349            430287            431226            432166            433107            434049            434992            435936            436881            437827            438774            439722            440671            441621            442572            443524            444477            445431            446386            447342            448299            449257            450216            451176            452137            453099            454062            455026            455991            456957            457924            458892            459861            460831            461802            462774            463747            464721            465696            466672            467649            468627            469606            470586            471567            472549            473532            474516            475501            476487            477474            478462            479451            480441            481432            482424            483417            484411            485406            486402            487401            488401            489402            490404            491407            492411            493416            494422            495429            496437            497446            498456            499467            500479            501492            502506            503521            504537            505554            506572            507591            508611            509632            510654            511677            512701            513726            514752            515779            516807            517836            518866            519897            520929            521962            522996            524031            525067            526104            527142            528181            529221            530262            531304            532347            533391            534436            535482            536529            537577            538626            539676            540727            541779            542832            543886            544941            546007            547074            548142            549211            550281            551352            552424            553497            554571            555646            556722            557801            558881            559962            561044            562127            563211            564296            565382            566469            567557            568646            569736            570827            571919            573012            574106            575201            576297            577394            578492            579591            580691            581792            582894            583997            585101            586206            587312            588419            589527            590636            591746            592857            593969            595082            596196            597311            598427            599544            600662            601781            602901            604022            605144            606267            607391            608516            609642            610769            611897            613026            614156            615287            616419            617552            618686            619821            620957            622094            623232            624371            625511            626652            627794            628937            630081            631226            632372            633519            634667            635816            636966            638117            639269            640422            641576            642731            643887            645044            646202            647361            648521            649682            650844            652007            653171            654336            655502            656669            657837            659006            660176            661347            662519            663692            664866            666041            667217            668394            669572            670751            671931            673112            674294            675477            676661            677846            679032            680219            681407            682596            683786            684977            686169            687362            688556            689751            690947            692144            693342            694541            695741            696942            698144            699347            700551            701756            702962            704169            705377            706586            707796            709007            710219            711432            712646            713861            715077            716294            717512            718731            719951            721172            722394            723617            724841            726066            727292            728519            729747            730976            732206            733437            734669            735902            737136            738371            739607            740844            742082            743321            744561            745802            747044            748287            749531            750776            752022            753269            754517            755766            757016            75

The sample-point method is mechanical but can be tedious and almost impossible to use in the case of large numbers of events or sample points. Accordingly, we borrow elementary results from *combinatorial analysis* or the theory of counting to assist us in analyzing sample spaces.

### Tools for Counting Sample Points

The first is the so-called multiplication rule or *mn rule*. With  $m$  elements  $a_1, a_2, \dots, a_m$  and  $n$  elements  $b_1, b_2, \dots, b_n$ , it is possible to form  $m \times n$  pairs containing one element from each group. (Proof given is by inspection of an  $a \times b$  table). For example, there are 6 times 6 = 36 different ways to roll 2 dice.

This rule can, of course be extended to any number of sets and can thus be used to count the sample points of any experiment that is reducible to a series of steps or parts.

Example—in the coin tossing case above,  $m=n=p=2$ , therefore the sample space has  $2^3=8$  points. That is, 2 possibilities for each of 3 events.

The second tool is the *permutation*.

Consider an experiment in which we select three cards from a deck, one-at-a-time, without replacing the cards already selected (this is called sampling without replacement).

Given a standard 52-card deck, the size of the sample space is given by  $52(52-1)(52-2)$ . There are thus 132,600 permutations or ways to draw three cards. More generally, if we want to find the number of permutations of  $n$  elements taken  $r$  at a time (so 52, 3 at a time in the example), the formula is:

$$P_r^n = n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!}$$

Note that this implies sampling without replacement and that  $n!$  (read  $n$  factorial) is defined as  $n(n-1)(n-2)\dots 1$ , and that  $0! = 1$ .

Book's definition here: An ordered arrangement of  $r$  distinct objects is called a permutation. The number of ways of ordering  $n$  distinct objects taken  $r$  at a time is designated by the symbol  $P_r^n$ . Sometimes this is shown as  ${}_n P_r$ .

Example: Suppose a club consists of 25 members. How many ways can a president and secretary be chosen (assuming only one office per person)?

$$\text{Answer: } P_2^{25} = \frac{25!}{(25-2)!} = \frac{25!}{23!} = \frac{(25)(24)(23!)}{23!} = (25)(24) = 600$$

Example: How many ways can one arrange six books on a shelf?

$$\text{Answer: } P_6^6 = \frac{6!}{(6-6)! \cdot 0!} = \frac{6!}{0!} = 6! = 720$$

Note on sampling *with* replacement: the number of sample points is simply  $n^r$ .

For example, think of drawing 1 ball from 10 balls numbered 1..10 in an urn, noting the number and replacing it, a total of 5 times.

What is the probability of getting the sequence 1,2,3,4,5?

The number of sample points is  $n^r=10^5$ .

With one choice we have a vector of  $n$  objects.

With two, we have a grid that is  $n$  by  $n$ .

Three, a cube, and so on.

Assuming independent random draws,  $P(A)=1/10^5$ .

### **Obtaining Different Numbers**

If  $r > n$  it is impossible for all of the selected balls to be different.

So suppose that  $r \leq n$ .

The number of outcomes in the event E is the number of vectors for which all  $r$  components are different.

This equals  $P_{n,r}$  since the first component  $x_1$  of each vector can have  $n$  possible values, the second component  $x_2$  can have any of the other  $n - 1$  values and so on.

Since S is a simple space containing  $n^r$  vectors, the probability  $p$  that  $r$  different numbers will be selected is:

$$p = \frac{P_r^n}{n^r} = \frac{n!}{(n-r)!n^r}$$

### Example The Birthday Problem:

What is the probability,  $p$ , that at least two people in a group of  $r$  [2,365] people will have the same birthday? Assuming no twins, etc.

The sample space is given by  $365^r$  and the possible permutations wherein each person has a different birthday by  $P_r^{365}$ . Thus the probability of each person having a different

birthday is  $\frac{P_r^{365}}{365^r}$  and, by the second axiom of probability, the probability of each person not having a different birthday (i.e. at least two with the same birthday) is given by:

$$1 - \frac{P_r^{365}}{365^r} = 1 - \frac{365!}{(365-r)!365^r}$$

In a class where  $r = 15$ ,  $p = .253$ . There is about a 25% chance that at least two of us share the same birthday.

*Combinations*: the number of combinations of  $n$  objects taken  $r$  at a time (or the number of subsets  $r$  that can be formed by  $n$  objects) is denoted by:

$$nC_r, C_r^n \text{ or } \binom{n}{r}$$

The latter is referred to as a *binomial coefficient*.

The number of unordered subsets of size  $r$  chosen (without replacement) from  $n$  objects is:

$$nC_r, C_r^n \text{ or } \binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$$

Example: Suppose a class contains 15 boys and 30 girls and that 10 students are to be selected at random for a special assignment. What is the probability that exactly 3 boys are selected?

Answer: The number of different combinations (assumed to be equally probable) is:

$$\binom{45}{10}.$$

The ways that 3 boys can be selected are  $\binom{15}{3}$  and the way 7 girls can be selected is

$$\binom{30}{7}.$$

The answer is thus: 
$$p = \frac{\binom{15}{3}\binom{30}{7}}{\binom{45}{10}} = .2904$$

**Suggested Homework Problems:** 1.6, 1.8, 1.11, 1.19, 1.23, \*1.30, 2.8, 2.9, 2.16, 2.21, 2.31, 2.35, 2.43, 2.45, 2.55

## Probability Theory II

Last time we discussed the philosophy of probability, the axioms of probability and some ways to count sample points (some combinatorics).

Before we move on, I want to introduce one more result from combinatorial theory: the *multinomial coefficient*.

$$N = \binom{n}{n_1 n_2 \dots n_r}$$

where  $N$  is the number of arrangements of  $n$  items in  $n_r$  groups—order doesn't matter.

The number of ways of partitioning  $N$  is:

$$N = \binom{n}{n_1 n_2 \dots n_r} = \frac{n!}{n_1! n_2! \dots n_r!}$$

Example: How many ways can one place 100 experimental subjects into 1 group of 50 and two groups of 25?

$$N = \binom{100}{50 \ 25 \ 25} = \left( \frac{100!}{50! 25! 25!} \right) \approx 1.2 \times 10^{43}$$

Example II:

A deck of cards contains 13 hearts. The cards are shuffled and dealt among four players, A, B, C, D as they are in bridge so that each player receives 13 cards.

What is the probability  $p$  that player A gets 6 hearts, B gets 4 hearts, C gets two hearts and D gets one heart?

The total number  $N$  of different ways in which the 52 cards can be distributed among the four players so that each player receives 13 cards is:

$$N = \binom{52}{13, 13, 13, 13} = \frac{52!}{(13!)^4}$$

We can assume that each of these ways is equally probable.

Now calculate the number of  $M$  ways of distributing the cards so that each player gets the specified number of hearts.

The number of different ways they can be distributed and give 6,4,2, and 1 hearts to A,B,C, and D, respectively is:

$$\binom{13}{6,4,2,1} = \frac{13!}{6!4!2!1!}$$

and the number of different ways in which the other 39 cards can then be distributed to the four players so that each will have a total of 13 cards is:

$$\binom{39}{7,9,11,12} = \frac{39!}{7!9!11!12!}$$

$$M = \frac{13!}{6!4!2!1!} * \frac{39!}{7!9!11!12!}$$

And the required probability is:

$$p = \frac{M}{N} = \frac{13!39!(13!)^4}{6!4!2!1!7!9!11!12!52!} = 0.00196$$

So, multinomial coefficients generalize binomial coefficients.

The coefficient  $\binom{n}{n_1, \dots, n_r}$  is the number of ways to partition a set of  $n$  items into distinguishable subsets of sizes  $n_1, \dots, n_r$  where  $n_1, \dots, n_r = n$ .

It is also the number of arrangements of  $n$  items of  $r$  different types for which  $n_i$  are of type  $i$  for  $i=1, \dots, r$ .

### Conditional Probability

The probability of an event can often be dependent on other random variables.

Another way of saying this is that our priors about the probability of an event may be different if we have additional information.

For example, what is the probability of a certain bill passing the Senate, *given* a Republican majority?

Or what is the probability that G.W. Bush will win a general election given a series of independent variables?

What is the probability that a voter will respond to a telephone survey given that she is White? Black? A Pacific Islander?

These are all examples of *conditional probabilities*.

More formally: 
$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

where  $P(A|B)$  is read as “the probability of A given B,” or “A conditional on B,” or “A conditioned on B.”

Example: imagine an experiment is repeated N times where N is a very large number.

There are two outcomes of interest, A and B, and their frequencies are given by the following table:

	A	$\bar{A}$
B	$n_{11}$	$n_{12}$
$\bar{B}$	$n_{21}$	$n_{22}$

So  $(A \cap B) = n_{11}$  and  $P(A \cap B) = \frac{n_{11}}{N}$ , etc.

By inspection,  $P(A) = \frac{n_{11} + n_{21}}{N}$ ,  $P(B) = \frac{n_{11} + n_{12}}{N}$ .

What is  $P(A|B)$ ?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{n_{11} / N}{(n_{11} + n_{12}) / N} = \frac{n_{11}}{n_{11} + n_{12}}$$

As another example, given a balanced die, what is the probability of rolling a 1 given that you roll an odd number?

$$P(1|odd) = \frac{P(1 \cap odd)}{P(odd)} = \frac{1/6}{1/2} = \frac{1}{3}$$

[Note: why 1/6? Since “odd” contains the event “roll a 1,” their intersection is equal to the probability of rolling a 1:

$$A \cap B = A$$

What if the probability of event A is not affected by the occurrence of event B:

$P(\text{Kerry Wins} | \text{New York's average annual rainfall} = 7)$

In this case, the events are said to be *independent*. More formally:

Two events, A and B, are independent if:

$$P(A|B) = P(A) \quad \text{or}$$

$$P(B|A) = P(B) \quad \text{or}$$

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

Note that if any one of these conditions hold the others will hold as well.

So, showing that one holds (and it is sometimes easier to show one or another) is sufficient proof of independence.

Do not confuse independence with mutually exclusive. The latter means that two things cannot occur simultaneously, implying that outcomes may be related.

### Multiplicative Law

We can generalize the third condition for independence with the multiplicative law of probability.

Again given two events A and B,

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

$$= P(B)P(A|B)$$

where  $P(A \cap B)_{\text{independent}} = P(A)P(B)$  is a special case of this law.

Note that this law generalizes further to  $k$  events:

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

We also need laws for the probability of the union of events that are not necessarily disjoint (thus a generalization of the third axiom):

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

Proof:

Draw this out as you go...

Assume  $S = \{A, B\}$

By inspection,

$$A \cup B = A \cup (\bar{A} \cap B)$$

$$B = (\bar{A} \cap B) \cup (A \cap B)$$

By Axiom 3,

$$P(A \cup B) = P(A) + P(\bar{A} \cap B)$$

$$P(B) = P(\bar{A} \cap B) + P(A \cap B)$$

Rearranging the second equation:

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

Substitute this into the first probability statement above:

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

Q.E.D.

This rule can also be extended to more than 2 events. In the case of 3, for example,

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

Finally, before we move onto some examples, recall also that the 2<sup>nd</sup> axiom allows us to say:

$$P(A) = 1 - P(\bar{A})$$

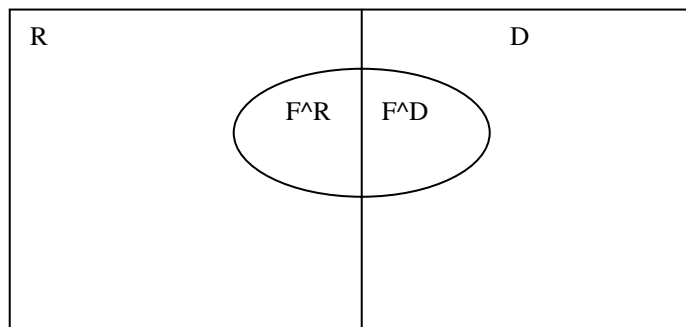
Example (2.17): In a given city, 40% of the voters are Republicans and 60% Dems.

Of the R's, 70% favor a bond issue. Of the D's, 80% are in favor of the issue.

If a voter is selected at random, what is the probability that he favors the bond issue?

Solution: Let F be the event that the issue is favored.

The sample space looks like:



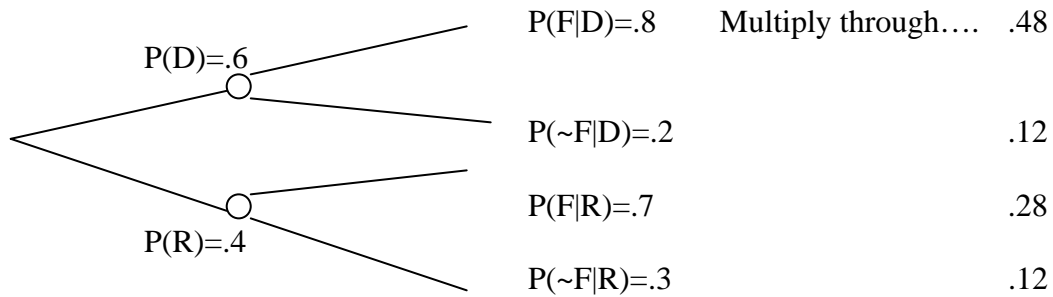
$$P(R) = .4, P(D) = .6, P(F|R) = .7, P(F|D) = .8$$

$$P(F) = P(F \cap R) + P(F \cap D)$$

$$P(F) = P(F | R)P(R) + P(F | D)P(D)$$

$$P(F) = (.7)(.4) + (.8)(.6) = .76$$

This can also be demonstrated using a tree diagram.



And add .48+.28=.76

Example: A man has  $n$  keys on a key ring, one of which opens the door to his apartment. Having celebrated a bit too much one evening, he returns home only to find himself unable to distinguish one key from another. Resourceful, he decides to choose a key at random and try it. If it fails to open the door, he will discard it and choose at random one of the remaining  $n-1$  keys, and so on. Clearly, the probability that he gains entrance with the first key he selects is  $1/n$ . Show that the probability the door opens with the *third* key he tries is also  $1/n$ . (Hint: what has to happen before he gets to the third key?)

Let  $K_i$  be the event that the  $i$ th key tried opens the door,  $i=1,2,\dots,n$ . Then  $P(\text{door opens first time with 3<sup>rd</sup> key}) =$

$$P(\bar{K}_1 \cap \bar{K}_2 \cap K_3) = P(\bar{K}_1) * P(\bar{K}_2 | \bar{K}_1) * P(K_3 | \bar{K}_1 \cap \bar{K}_2)$$

What is the probability that the first key *doesn't* open the door?

$$1 - \frac{1}{n} = \frac{n}{n} - \frac{1}{n} = \frac{n-1}{n}$$

The probability that the second key opens the door, given that the first didn't is  $\frac{1}{n-1}$ .

The complement of that is:

$$1 - \frac{1}{n-1} = \frac{n-1}{n-1} - \frac{1}{n-1} = \frac{n-1-1}{n-1} = \frac{n-2}{n-1}$$

And,  $p$  that the third key opens the door given that the first two didn't is  $\frac{1}{n-2}$ .

$$\text{So, } P(\bar{K}_1 \cap \bar{K}_2 \cap K_3) = P(\bar{K}_1) * P(\bar{K}_2 | \bar{K}_1) * P(K_3 | \bar{K}_1 \cap \bar{K}_2)$$

$$= \frac{n-1}{n} * \frac{n-2}{n-1} * \frac{1}{n-2} = \frac{1}{n}$$

### Event Composition Approach

The text generalizes the procedure used to solve this example into what it calls the *event composition approach*, which differs from the sample point approach in that it is not necessary to count all sample points.

1. Define the experiment
2. Visualize sample points
3. Write an equation for the event of interest as a function of known quantities.
4. Apply the multiplicative and additive laws to find the probability of interest.

Example (2.20): The probability of a patient responding to treatment is .9. If 3 patients are treated independently, what is the probability that at least one will respond?

Solution: Let A be the event of interest (at least one responds) and  $B_n$  be the event that patient  $n$  does not respond.

$$P(A) = 1 - P(\bar{A})$$

$$P(A) = 1 - P(B_1 \cap B_2 \cap B_3)$$

$$P(A) = 1 - P(B_1)P(B_2 | B_1)P(B_3 | B_1 \cap B_2)$$

$$P(A) = 1 - P(B_1)P(B_2)P(B_3)$$

$$P(A) = 1 - .1^3 = .999$$

We can go from line 3 to line 4 because the patients are independent from each other.

### Law of Total Probability

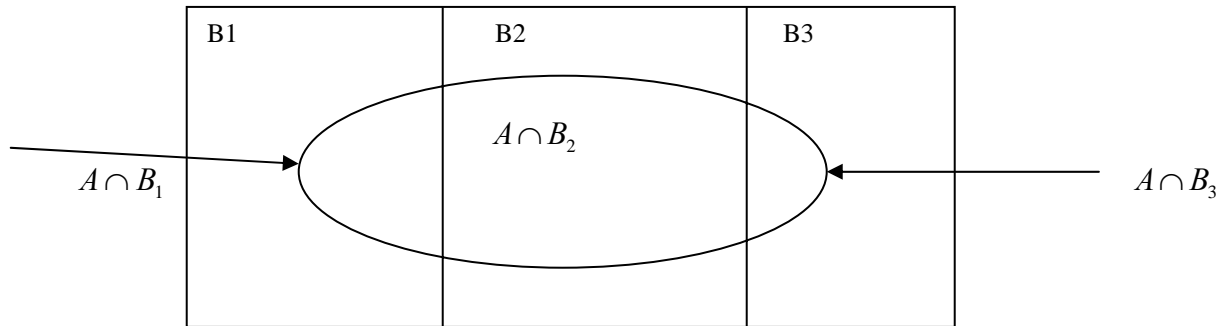
Sometimes the ECA is easier if S is thought of as a union of disjoint subsets. The law of total probability allows for this partitioning:

For a positive integer  $k$ , let  $B_1, B_2, \dots, B_k$  be defined so that:

$$1. S = \bigcup_{n=1}^k B_n$$

$$2. B_i \cap B_j = \emptyset \quad \forall i \neq j$$

Thus  $\{ B_1, B_2, \dots, B_k \}$  forms a partition of  $S$ . An event  $A$  can thus be decomposed ( $k=3$ ):



Generally,

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

(So long as condition 2 above holds.)

This allows for the following result:

if  $\{ B_1, B_2, \dots, B_k \}$  is a partition of  $S$  such that  $P(B_i) > 0$ , then

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

[proof given in text]

### Bayes' Rule

The law of total probability allows us to state Bayes' Rule:

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

Proof: By the definition of conditional probability:

$$P(B_j | A) = \frac{P(A \cap B_j)}{P(A)}$$

Rewriting the numerator using the multiplicative law:

$$P(B_j | A) = \frac{P(A | B_j) P(B_j)}{P(A)}$$

[note this is a common form of presentation for Bayes Rule]

Then, by the law of total probability, the denominator is rewritten:

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

Or, in its simplest form:  $P(B | A) = \frac{P(A | B)P(B)}{P(A)}$

Since  $P(A) = P(A | B)P(B) + P(A | \bar{B})P(\bar{B})$

We can expand this to:  $P(B | A) = \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | \bar{B})P(\bar{B})}$

Bayes' Rule allows us to formalize the idea of "updating" or revising a belief about a probability given new evidence. In the language of Bayesian statistics,  $P(B_j)$  is called the prior probability and  $P(B_j|A)$  is a posterior probability.

Example: Suppose that a given medical test is 90% reliable (i.e. if you have the disease there is .9 chance of a proper positive response and if you don't have the disease there is a .1 chance you will have a false positive). Data from a large sample yields that 1/10,000 in the population actually have the disease [prior].

If you test positive, what is the probability that you have the disease [posterior]?

Solution: if A = you test positive; B= you have the disease

$$P(B | A) = \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | \bar{B})P(\bar{B})}$$

$$P(B | A) = \frac{.9(.0001)}{.9(.0001) + .1(.9999)} = .0009 = \frac{9}{10000}$$

Another way to think about this is 1/10000 people have the disease, but the test gives a false positive result for 1/10 people who take it. So, the number of positive results is about 1000 times greater than the number of people with the disease.

## Example II

Recently the U.S. Senate Committee on Labor and Public Welfare investigated the feasibility of setting up a national screening program to detect child abuse. A team of consultants estimated the following probabilities: (1) 1 child in 90 is abused, (2) a physician can detect an abused child 90% of the time, and (3) a screening program would incorrectly label 3% of all nonabused children as abused. What is the probability that a child is actually abused given that the screening program diagnoses him as such? How does this probability change if the incidence of abuse is 1 in 1000? 1 in 50?

Let  $B$  denote the event that the program diagnoses the child as abused, and let  $A$  be the event that the child *is* abused. Then  $P(A) = 1/90$ ,  $P(B|A)=0.90$ , and  $P(B|\bar{A})=0.03$ , so

$$P(A|B) = \frac{(0.90)(1/90)}{(0.90)(1/90) + (0.03)(89/90)} = 0.25$$

If  $P(A)=1/1000$ ,  $P(A|B)=0.029$ ; if  $P(A)=1/50$ ,  $P(A|B)=0.38$ .

Two final definitions that we can now state with more precision:

1. A *random variable* is a real-valued function for which the domain is a sample space. If  $y$  is an observed value of random variable  $Y$ , then  $P(Y=y)$  is the sum of the probabilities of the sample points assigned value  $y$ .

2. A *random sample* is a sample selected such that each of the  $\binom{N}{n}$  possible samples has an equal probability of selection.

Practice Homework Problems: 2.57, 2.58, 2.70, 2.77, 2.78, 2.96, 2.101, 2.103, 2.109, 2.110, 2.112, 2.114.