Probability

- Probability - meaning
  1) classical
  2) frequentist
  3) subjective (personal)
- Sample space, events
- Mutually exclusive, independence
- and, or, complement
- Joint, marginal, conditional probability
- Probability - rules
  1) Addition
  2) Multiplication
  3) Total probability
  4) Bayes
- Screening
  - sensitivity
  - specificity
  - predictive values

Probability

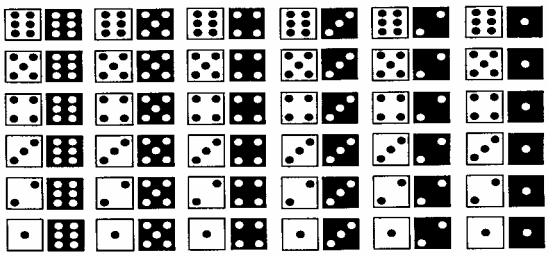
Probability provides a measure of uncertainty associated with the occurrence of events or outcomes – Models and theory

Definitions:

1. **Classical**: $P(E) = m/N$

   If an event can occur in N **mutually exclusive, equally likely** ways, and if m of these possess characteristic E, then the probability is equal to ***m/N***.

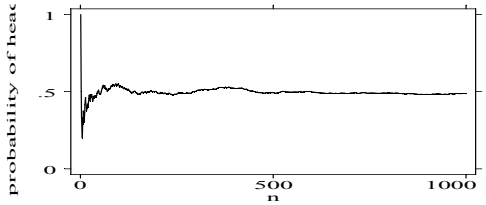
Example: What is the probability of rolling a total of 7 on two dice?



2. relative frequency: $P(E) \approx m / n$

If a process or an experiment is repeated a large number of times, *n*, and if the characteristic, E, occurs *m* times, then the relative frequency, *m/n*, of *E* will be approximately equal to the probability of *E*.

» Around 1900, the English statistician Karl Pearson heroically tossed a coin 24,000 times and recorded 12,012 heads, giving a proportion of 0.5005.



Stata: graph twoway line prob n, yscale(r(0.0,1.0)) title("Probability of heads") ylabel(0(0.5)1)

3. personal probability

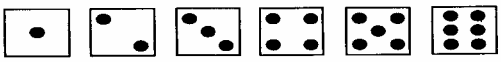
What is the probability of life on Mars?

Sample Space

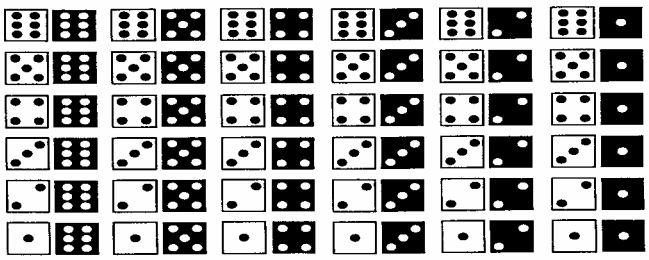
The **sample space** consists of the possible outcomes of an experiment. An **event** is an outcome or set of outcomes.

For a coin flip the sample space is (*H*,*T*).

THE SAMPLE SPACE OF THE THROW OF A *SINGLE DIE* IS A LITTLE BIGGER.

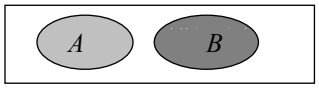


AND FOR A *PAIR* OF DICE, THE SAMPLE SPACE LOOKS LIKE THIS (WE MAKE ONE DIE WHITE AND ONE BLACK TO TELL THEM APART):



Basic Properties of Probability

1. Two events, *A* and *B*, are said to be **mutually exclusive** (disjoint) if only one or the other, but not both, can occur in a particular experiment.



2. Given *n* mutually exclusive events, E_1, E_2, …., E_n, the probability of any event is non-negative and less than or equal to 1:

$$0 \leq P(E_i) \leq 1$$

3. The sum of the probabilities of an exhaustive collection (i.e. at least one must occur) of mutually exclusive outcomes is 1:

$$\sum_{i=1}^{n} P(E_i) = P(E_1) + P(E_2) + \ldots + P(E_n) = 1$$

4. The probability of all events other than an event A is denoted by P(A^c) [A^c stands for "*A* complement"] or $P(\bar{A})$ ["*A* bar"]. Note that

   P(A^c) = 1 – P(*A*)

Notation for Joint Probabilities

- If *A* and *B* are any two events then we write

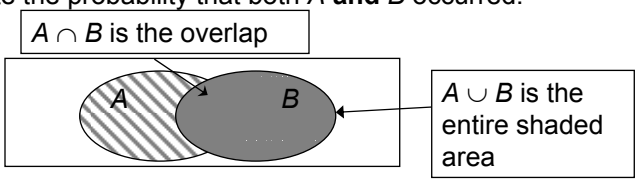
  P(*A* or *B*) *or* P($A \cup B$)

  to indicate the probability that event *A* **or** event *B* (or both) occurred.

- If *A* and *B* are any two events then we write

  P(*A* and *B*) *or* P(*AB*) *or* P($A \cap B$)

  to indicate the probability that both *A* **and** *B* occurred.



- If *A* and *B* are any two events then we write

  P(*A* given *B*) *or* P(*A*|*B*)

  to indicate the probability of *A* among the subset of cases in which *B* is known to have occurred.

Conditional Probability

The **conditional probability** of an event *A* given *B* (i.e. given that *B* has occurred) is denoted P(*A* | *B*).

		Disease Status		
		Pos.	Neg.	Total
Test Result	Pos.	9	80	89
	Neg.	1	9910	9,911
	Total	10	9990	10,000

What is P(test positive)?

What is P(test positive | disease positive)?

What is P(disease positive | test positive)?

2.6.2. The following table shows the first 1000 patients admitted to a clinic for retarded children by diagnostic classification and level of intelligence. For this group find:

(a) $P(A_3 \cap B_4)$.
(b) The probability that a patient picked at random is severely retarded.
(c) The probability that a patient picked at random is either not retarded or is borderline.
(d) The probability that a patient picked at random is profoundly retarded and has Down's syndrome.
(e) The probability that a patient is profoundly retarded, given that he has Down's syndrome.

	Level of Retardation						
Major Diagnostic Classification	A_1 Not Retarded	A_2 Pro-found	A_3 Severe	A_4 Moderate	A_5 Mild	A_6 Border-line	Total
B_1 Encephalopathies	33	38	57	114	103	55	400
B_2 Down's syndrome	2	4	34	88	27	5	160
B_3 Congenital cerebral defect	10	2	6	6	6	0	30
B_4 Mental retardation of unknown cause	0	0	9	36	62	35	142
B_5 Other	161	0	8	16	8	75	268
Total	206	44	114	260	206	170	1000

General Probability Rules

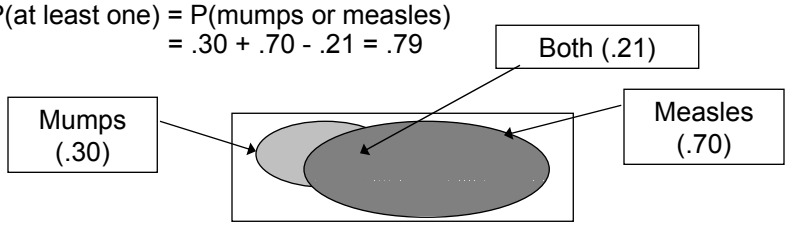
- **Addition rule**

If two events *A* and *B* are not mutually exclusive, then the probability that event *A* or event *B* occurs is:

$$P(A \text{ or } B) = P(A) + P(B) - P(AB)$$

E.g. Of the students at Anytown High school, 30% have had the mumps, 70% have had measles and 21% have had both. What is the probability that a randomly chosen student has had at least one of the above diseases?

P(at least one) = P(mumps or measles)
= .30 + .70 - .21 = .79



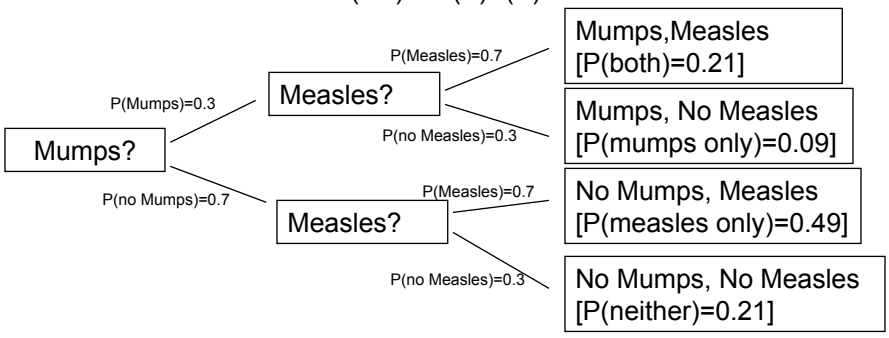
What is P(neither)? Identify the area for "neither"

General Probability Rules

- **Multiplication rule** (special case – **independence**)

If two events, *A* and *B*, are "independent" (probability of one does not depend on whether the other occurred) then

$$P(AB) = P(A)P(B)$$



Easy to extend for independent events $A,B,C,\ldots$:

$$P(ABC\ldots) = P(A)P(B)P(C)\ldots$$

General Probability Rules

- **Multiplication rule (general)**

More generally, however, *A* and *B* may not be independent. The probability that one event occurs may depend on the other event. This brings us back to conditional probability. The general formula for the probability that both *A* and *B* will occur is

$$\mathrm{P}(AB) = \mathrm{P}(A \mid B)\mathrm{P}(B) = \mathrm{P}(B \mid A)\mathrm{P}(A)$$

Two events *A* and *B* are said to be **independent** if and only if

P(*A*|*B*) = P(*A*) or

P(*B*|*A*) = P(*B*) or

P(*AB*) = P(*A*)P(*B*).

(Note: If any one holds then all three hold)

E.g. Suppose P(mumps) = .7, P(measles) = .3 and P(both) = .25. Are the two events independent?

No, because P(mumps and measles) = .25 while
P(mumps)P(measles) = .21

- **Total probability rule**

If $A_1,\ldots A_n$ are mutually exclusive, exhaustive events, then

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(B \text{ and } A_i) = \sum_{i=1}^{n} \mathrm{P}(B \mid A_i)\mathrm{P}(A_i)$$

E.g. The following table gives the estimated proportion of individuals with Alzhiemer's disease (AD) by age group. It also gives the proportion of the general population that are expected to fall in the age group in 2030. What proportion of the population in 2030 will have Alzhiemer's disease?

		Proportion with AD	Proportion population
Age group	< 65	.00	.80
	$65 - 75$	.01	.11
	$75 - 85$	.07	.07
	> 85	.25	.02

P(AD in 2030) = 0*.8 + .01*.11 + .07*.07 + .25*.02 = .011

• **Bayes rule** (combine multiplication rule with total probability rule)

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}$$

We will only apply this to the situation where A and B have two levels each, say, A and $\overline{A}$, B and $\overline{B}$. The formula becomes

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})}$$

12

Screening - an Application of Bayes Rule

Suppose we have a random sample of a population...

		Disease Status		
		Pos.	Neg.	Total
Test Result	Pos.	90	30	120
	Neg.	10	970	980
	Total	100	1000	1100

D = disease pos.
T = test pos.

Prevalence = P(D) = 100/1100 = .091

Sensitivity = P(T | D) = 90/100 = .9

Specificity = P($\overline{T}$ | $\overline{D}$) = 970/1000 = .97

PPV = Predictive Value of a Positive Test

= P(D | T) = 90/120 = .75

NPV = Predictive Value of a Negative Test

= P($\overline{D}$ | $\overline{T}$) = 970/980 = .99

13

Screening - an Application of Bayes Rule

Now suppose we have taken a sample of 100 disease positive and 100 disease negative individuals (e.g. case-control design)

		Disease Status		
		Pos.	Neg.	Total
Test Result	Pos.	90	3	93
	Neg.	10	97	107
	Total	100	100	200

D = disease pos.
T = test pos.

Prevalence = ???? (not .5!)

Sensitivity = P($T \mid D$) = 90/100 = .9

Specificity = P($\overline{T} \mid \overline{D}$) = 97/100 = .97

PPV = P($D \mid T$) = 90/93 **NO!**

NPV = P($\overline{D} \mid \overline{T}$) = 97/107 **NO!**

Screening - an Application of Bayes Rule

D = disease pos.
T = test pos.

Assume we know, from external sources, that P(D) = 100/1100.

$$\text{PPV} = \text{P}(D \mid T) = \frac{\text{P}(T \mid D)\text{P}(D)}{\text{P}(T)}$$

$$= \frac{\text{P}(T \mid D)\text{P}(D)}{\text{P}(T \mid D)\text{P}(D) + \text{P}(T \mid \overline{D})\text{P}(\overline{D})}$$

$$= \frac{.9 \times {}^{100}\!/_{1100}}{.9 \times {}^{100}\!/_{1100} + .03 \times {}^{1000}\!/_{1100}} = .75$$

$$\text{NPV} = \text{P}(\overline{D} \mid \overline{T}) =$$

Screening for Alzheimer's Disease - an Application of Bayes Rule

The presence of a particular genotype, APOE, has been considered as a screen for people who develop AD. The sensitivity and specificity of the test are 60% and 70%, respectively.

So, there is clearly an association between APOE and AD. Those who develop AD are twice as likely to have the genotype as those who don't develop AD. (To see this, compare P(APOE+|AD+) to P(APOE+|AD-).)

But is this a useful screen for AD?

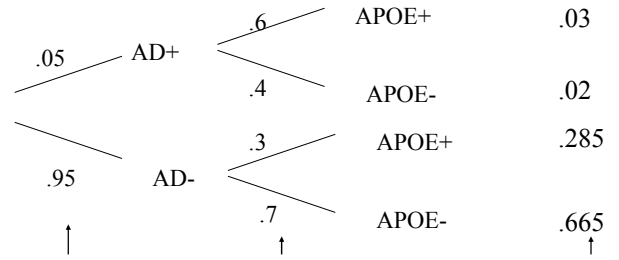
Compare 2 elderly populations:

- the general elderly population where the chance of developing AD is 5%
- an elderly population with memory complaints where the chance of developing AD is 70%

16

Screening for Alzheimer's Disease - an Application of Bayes Rule

Suppose we are looking at an elderly population where the chance of developing memory complaints is 5%.

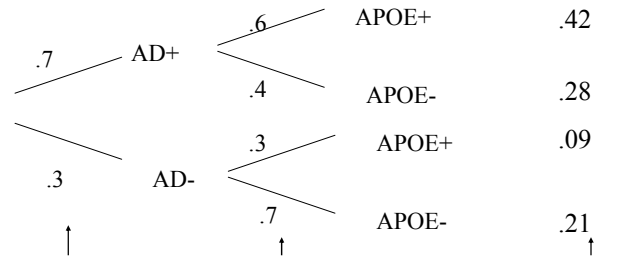


$$\text{PPV} = \frac{.03}{.03 + .285} = .095 \qquad \text{NPV} = \frac{.665}{.665 + .02} = .97$$

17

Screening for Alzheimer's Disease - an Application of Bayes Rule

Suppose we are looking at an elderly population with memory problems where the chance of developing memory complaints is 70%.



$$PPV = \frac{.42}{.42 + .08} = .82 \qquad NPV = \frac{.21}{.21 + .28} = .43$$

Screening for Alzheimer's Disease - an Application of Bayes Rule Comments

We found that a test with given sensitivity and specificity will have different predictive values in different populations. So the usefulness of the test will depend on the population context and the implications of incorrect or missed diagnoses

- If disease prevalence is low, high specificity is particularly important. Otherwise almost all positive tests will be false positives and these diagnoses will burden the health care system and patients with unnecessary follow-up.
- If consequences of false positives are very bad (e.g. unnecessary high-risk surgery), high specificity is also important.
- If disease prevalence is high and/or it is very important to find true positive cases, then you would want a test with high sensitivity.