UCLA STAT 10
Introduction to Statistical Reasoning

•**Instructor:** Ivo Dinov,
Asst. Prof. In Statistics and Neurology

•**Teaching Assistants:** Yan Xiong, Will Anderson,
UCLA Statistics

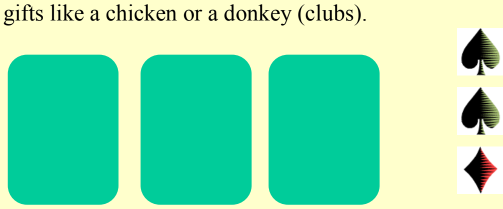
University of California, Los Angeles, Winter 2002
http://www.stat.ucla.edu/~dinov/

Part IV: Chances, Variability, Probabilities and Proportions

- Chances and chance variability
- Where do probabilities come from?
- Simple probability models
- Probability rules
- Conditional probability
- Statistical independence

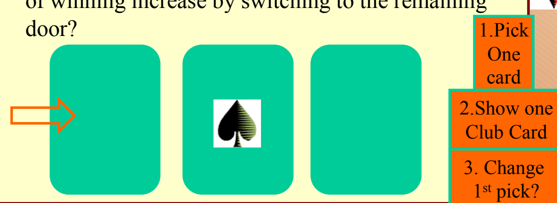
Let's Make a Deal Paradox – aka, Monty Hall 3-door problem

- This paradox is related to a popular television show in the 1970's. In the show, a contestant was given a choice of three doors/cards of which one contained a prize (diamond). The other two doors contained gag gifts like a chicken or a donkey (clubs).



Let's Make a Deal Paradox.

- After the contestant chose an initial door, the host of the show then revealed an empty door among the two unchosen doors, and asks the contestant if he or she would like to switch to the other unchosen door. The question is should the contestant switch. Do the odds of winning increase by switching to the remaining door?



Let's Make a Deal Paradox.

- The *intuition* of most people tells them that each of the doors, the chosen door and the unchosen door, are equally likely to contain the prize so that there is a 50-50 chance of winning with either selection? This, however, is **not the case**.
- The probability of winning by using the switching technique is 2/3, while the odds of winning by not switching is 1/3. The easiest way to explain this is as follows:

Let's Make a Deal Paradox

- The probability of picking the wrong door in the initial stage of the game is 2/3.
- If the contestant picks the wrong door initially, the host must reveal the remaining empty door in the second stage of the game. Thus, if the contestant switches after picking the wrong door initially, the contestant will win the prize.
- The probability of winning by switching then reduces to the probability of picking the wrong door in the initial stage which is clearly 2/3.
- StatGames.exe (Make a Deal Paradox)

Chance

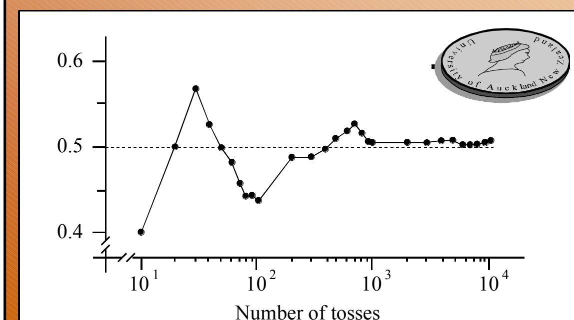
- The **chance** of something happening gives the <u>percentage of time it is expected to happen</u>, then the basic process is repeatedly performed.
- E.g., What is the chance of getting an ace (1) if we roll a regular 6-face (hexagonal) die?

- Chances are always between 0% - 100%.
- The chance of an event is equal to 100% - the chance of the opposite (complementary) event.
- E.g., Chance(getting 1) = 100 – Chance(2 or 3 or 4 or 5 or 6 turns up).

Coin toss experiments (Head vs. Tail)

- The law of averages about the behavior of coin tosses – the relative proportion (relative frequency) of heads-to-tails in a coin toss experiment becomes more and <u>more stable</u> as the <u>number of tosses increases</u>. The law of averages applies to relative frequencies not absolute counts of #H and #T.
- Two widely held misconceptions about what the <u>law of averages</u> about coin tosses:
  - Differences between the actual numbers of heads & tails becomes more and more variable with increase of the number of tosses – a seq. of 10 heads doesn't increase the chance of a tail on the next trial.
  - Coin toss results are fair, but behavior is still unpredictable.

Long run behavior of coin tossing – proportion of heads vs. number of tosses



Coin Toss Models

- Is the coin tossing model adequate for describing the sex order of children in families?
  - This is a *rough model* which is not exact. In most countries <u>rates of B/G is different</u>; from 48% … to 52%, usually. Birth rates of boys in some places are higher than girls, however, female population seems to be about 51%.
  - <u>Independence</u>, if a second child is born the chance it has the same gender (as the first child) is slightly bigger.

Two die throw example

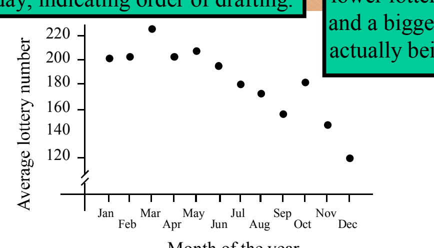
- What is the chance that the sum of the numbers, turning up when 2 dice are rolled, is equal to 8?
  - Do the HTML Java-applet: **clt.htm**

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Data from a "random" draw

366 cylinders (for each day in the year) for the US Vietnam war draft. The N-th drawn number, corr. to one B-day, indicating order of drafting.

So, people born later in the year tend to have lower lottery numbers and a bigger chance of actually being drafted.



Types of Probability

- Probability models have two essential components (*sample space*, the space of all possible outcomes from an experiment; and a list of *probabilities* for each event in the sample space). Where do the outcomes and the probabilities come from?
- Probabilities from models – say mathematical/physical description of the sample space and the chance of each event. Construct a fair die tossing game.
- Probabilities from data – data observations determine our probability distribution. Say we toss a coin 100 times and the observed Head/Tail counts are used as probabilities.
- Subjective Probabilities – combining data and psychological factors to design a reasonable probability table (e.g., gambling, stock market).

CA State Lottery – Supper Lotto *Plus*

- California Lotto, chose 5 out of 47 and choose one Mega from [1 : 27], fee $1, your odds are 1 in 41,416,353! Why?
- 47-choose-5 = [47!]/[(47-5)!(5!)] ➔

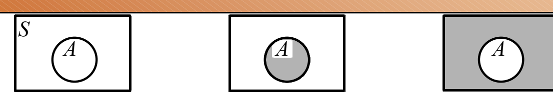
  47-choose-5 x 27 = 1,533,939 x 27 = 41,416,353

Sample spaces and events

- A ***sample space***, *S*, for a random experiment is the set of all possible outcomes of the experiment.
- An ***event*** is a *collection* of outcomes.
- An event ***occurs*** if any outcome making up that event occurs.

The complement of an event

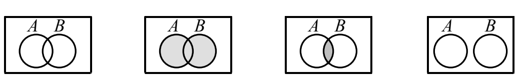
- The complement of an event *A*, denoted $\bar{A}$, occurs *if and only if A* does not occur.



(a) Sample space containing event *A* (b) Event *A* shaded (c) $\bar{A}$ shaded

Combining events – *all statisticians agree on*

- ***"A* or *B"*** contains all outcomes in *A* **or** *B* (or both).
- ***"A* and *B"*** contains all outcomes which are **in both** *A* and *B*.



(a) Events *A* and *B* (b) "*A* or *B*" shaded (c) "*A* and *B*" shaded (d) Mutually exclusive events

Mutually exclusive events cannot occur at the same time.

Probability distributions

- Probabilities always lie between 0 and 1 and they sum up to 1 (across all simple events) .
- ***pr(A)*** can be obtained by adding up the probabilities of all the outcomes in *A*.

$$pr(A) = \sum_{E \text{ outcome in event } A} pr(E)$$

Job losses in the US in $1,000, 1987-1991

	Reason for Job Loss			
	Workplace moved/closed	Slack work	Position abolished	Total
Male	1,703	1,196	548	3,447
Female	1,210	564	363	2,137
Total	2,913	1,760	911	5,584

Job losses raw-data vs. proportions

	Workplace moved/closed	Slack work	Position abolished	Total
Male	1,703	1,196	548	3,447
Female	1,210	564	363	2,137
Total	2,913	1,760	911	5,584

	Reason for Job Loss			
	Workplace moved/closed	Slack work	Position abolished	Row totals
Male	.305	.214	.098	.617
Female	.217	.101	.065	.383
Column totals	.552	.315	.163	1.000

Review

- What is a sample space? What are the <u>two essential criteria</u> that must be satisfied by a possible sample space? (completeness – every outcome is represented; and uniqueness – no outcome is represented more than once.
- What is an event? (collection of outcomes)
- If A is an event, what do we mean by its complement, $\bar{A}$? When does $\bar{A}$ occur?
- If *A* and *B* are events, when does ***A* or *B*** occur? When does ***A* and *B*** occur?

Example of probability distributions

- Tossing a coin twice. *Sample space* S={HH, HT, TH, TT}, for a fair coin each outcome is equally likely, so the probabilities of the 4 possible outcomes should be identical, *p*. Since, p(HH)=p(HT)=p(TH)=p(TT)=p and
$$p_k \geq 0; \quad \sum_k p_k = 1$$
- p = ¼ = 0.25.

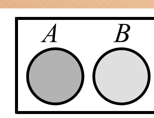
Proportion vs. Probability

- How do the concepts of a proportion and a probability <u>differ</u>? A proportion is a <u>partial description</u> of a real population. The probabilities give us the <u>chance</u> of something happening in a random experiment. Sometimes, proportions are <u>identical</u> to probabilities (e.g., in a real population under the experiment *choose-a-unit-at-random)*.
- See the ***two-way table of counts*** (*contingency table*) on Table 4.4.1, slide 19. E.g., *choose-a-person-at-random* from the ones laid off, and compute the chance that the person would be a <u>male</u>, laid off due to <u>position-closing</u>. We can apply the same rules for manipulating probabilities to proportions, in the case where these two are identical.

Rules for manipulating Probabilities

For mutually exclusive events,

pr(*A* or *B*) = pr(*A*) + pr(*B*)



Paradox of the Cavalier De Mere

- Betting on the event E={In 4 die rolls at least 1 ace turns up }. B={In 24 rolls of a pair of dice, at least one double-ace shows up }.
- Claim: P(E) = P(B)?!?
- Reasoning E: 1 roll gives a chance 1/6 for an ace! So, in 4 rolls we have 4 x 1/6 = 2/3 to get at least 1 ace!

  B: In one roll of a pair of dice, chance of a double-ace is 1/36. So in 24 rolls we have 24 x 1/36 = 2/3 chance.
- Experience showed P(E) > P(B)!!!
- What's wrong? Well, extrapolating these arguments we get that the chance of getting 1 ace in 6 rolls is 6x1/6=1? Obviously, incorrect!

Paradox of the Cavalier De Mere

- The chance of winning (getting at least one ace) is hard to compute, but can we calculate the chance of loosing – the complement event?!?
- Than *chance-of-winning* = 1-*chance-of-loosing*.
- E^, complement of E, = {none of 4 rolls shows an ace}.
- In one roll, chance of loosing is 5/6, no ace turns up.
- 2 die rolls are independent, hence we can use the multiplication rule, Chance of no ace in two rolls is $(5/6)^2$. Similarly, chance of 4 rolls with no ace, the probability P(E^) = $(5/6)^4 \sim 0.482$.
- Game 2: Pair-of-dice: Chance of no-ace in 1 roll is 35/36. Hence, P({no-Ace in 24 rolls})=$(35/36)^{24} \sim 0.509$.
- P({at-least-1-ace-in-4-rolls})=1-0.482=0.518 >>>>

  P({at-least-1-double-ace-in-24-rolls})=1-0.509=0.491.

Paradox of the Cavalier De Mere

- The chance of winning (getting at least one ace) is hard to compute, but can we calculate the chance of loosing – the complement event?!?
- Than *chance-of-winning* = 1-*chance-of-loosing*.
- E^, complement of E, = {none of 4 rolls shows an ace}.
- In one roll, chance of loosing is 5/6, no ace turns up.
- 2 die rolls are independent, hence we can use the multiplication rule, Chance of no ace in two rolls is $(5/6)^2$. Similarly, chance of 4 rolls with no ace, the probability P(E^) = $(5/6)^4 \sim 0.482$.
- Game 2: Pair-of-dice: Chance of no-ace in 1 roll is 35/36. Hence, P({no-Ace in 24 rolls})=$(35/36)^{24} \sim 0.509$.
- P({at-least-1-ace-in-4-rolls})=1-0.482=0.518 >>>>

  P({at-least-1-double-ace-in-24-rolls})=1-0.509=0.491.

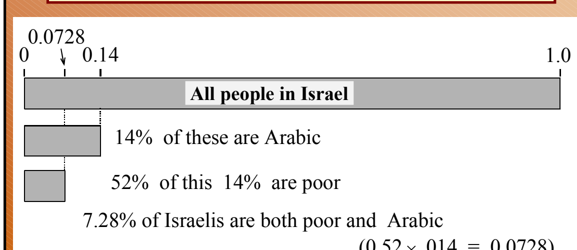
Conditional Probability

The ***conditional probability*** of ***A*** occurring ***given*** that *B* occurs is given by

$$\text{pr}(A \mid B) = \frac{\text{pr}(A \text{ and } B)}{\text{pr}(B)}$$

Multiplication rule- what's the percentage of Israelis that are poor and Arabic?

$$\text{pr}(P \text{ and } A) = \text{pr}(P \mid A)\text{pr}(A) = \text{pr}(A \mid P)\text{pr}(P)$$



Basic Rules for computing Probabilities

$$\text{pr}(A \text{ and } B) = \text{pr}(A \mid B)\text{pr}(B) = \text{pr}(B \mid A)\text{pr}(A)$$

$$\text{pr}(A) = 1 - \text{pr}(\bar{A})$$

Properties of probabilities.

$$\{p_k\}_{k=1}^{N} \text{ define probabilities} \Leftrightarrow p_k \ge 0; \ \sum_k p_k = 1$$

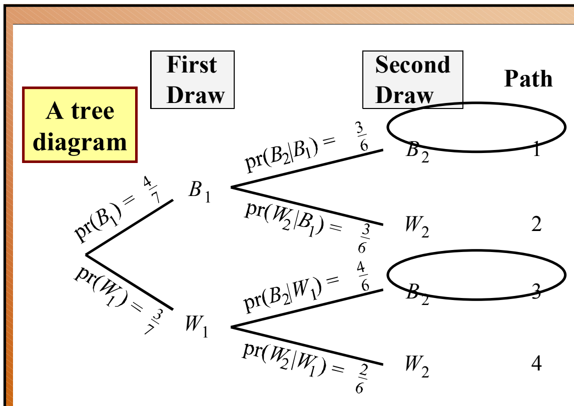
A tree diagram for computing conditional probabilities

$$\text{pr}(A \text{ and } B) = \text{pr}(A \mid B)\text{pr}(B) = \text{pr}(B \mid A)\text{pr}(A)$$

Suppose we draw 2 balls at random one at a time *without replacement* from an urn containing **4 black** and **3 white** balls, otherwise identical. What is the probability that the *second ball is black*? Sample Spc?

P({*2-nd ball is black*}) =
P({*2-nd is black*} &{*1-st is black*}) +
P({2-nd is black} &{*1-st is white*}) =
3/6 x 4/7+ 4/6 x 3/7 = 4/7.

Mutually exclusive



Conditional probabilities and 2-way tables

- Many problems involving conditional probabilities can be solved by constructing two-way tables
- This includes *reversing the order of conditioning*

P(A & B) = P(A | B) x P(B) = P(B | A) x P(A)

Number of Individuals Having a Given Mean Absorbance Ratio (MAR) in the ELISA for HIV Antibodies

MAR	Healthy Donor		HIV patients	
<2	202	} 275	0	} 2 False-Negatives (FNE)
2 - 2.99	73	Test cut-off	2	
3 - 3.99	15	False-positives	7	Power of a test is:
4 - 4.99	3		7	1-P(FNE)=
5 - 5.99	2		15	1-P(Neg|HIV)
6 -11.99	2		36	~ 0.976
12+	0		21	
Total	297		88	

HIV cont.

pr(*HIV* and *Positive*) = pr(*Positive*|*HIV*) × pr(*HIV*) [= 98% of 1%]

pr(*Not HIV* and *Negative*) = pr(*Negative*|*Not HIV*) × pr(*Not HIV*) [= 93% of 99%]

Disease status		Test result: Positive	Negative	Total	
	HIV	.98 × .01	?	.01	⟵ pr(*HIV*) = .01
	Not HIV	?	.93 × .99	.99	⟵ pr(*Not HIV*) = .99
	Total	?	?	1.00	

86/88 ~0.98 **275/297 ~ 0.93**

HIV – reconstructing the contingency table

pr(*HIV* and *Positive*) = pr(*Positive*|*HIV*) × pr(*HIV*) [= 98% of 1%]

pr(*Not HIV* and *Negative*) = pr(*Negative*|*Not HIV*) × pr(*Not HIV*) [= 93% of 99%]

Disease status		Test result: Positive	Negative	Total	
	HIV	.98 × .01	?	.01	⟵ pr(*HIV*) = .01
	Not HIV	?	.93 × .99	.99	⟵ pr(*Not HIV*) = .99
	Total	?	?	1.00	

		Test Result: Positive	Negative	Total
Disease Status	HIV	.0098	.0002	.01
	Not HIV	.0693	.9207	.99
	Total	.0791	.9209	1.00

Statistical independence

- Events A and B are *statistically independent* if knowing whether B has occurred gives <u>no new information</u> about the chances of A occurring,

  i.e. if **pr(A | B) = pr(A)**
- Similarly, P(B | A) = P(B), since

  P(B|A)=P(B & A)/P(A) = P(A|B)P(B)/P(A) = P(B)
- If A and B are *statistically independent*, then

$$\text{pr}(\ A \text{ and } B) = \text{pr}(A) \times \text{pr}(B)$$

People vs. Collins

Frequencies assumed by the prosecution:

Yellow car	$\frac{1}{10}$	Girl with blond hair	$\frac{1}{3}$
Man with mustache	$\frac{1}{4}$	Black man with beard	$\frac{1}{10}$
Girl with ponytail	$\frac{1}{10}$	Interracial couple in car	$\frac{1}{1000}$

- The first occasion where a conviction was made in an American court of law, largely on statistical evidence, 1964. A woman was mugged and the offender was described as a wearing dark cloths, with blond hair in a pony tail who got into a yellow car driven by a black male accomplice with mustache and beard. The suspect brought to trial were picked out in a line-up and fit all of the descriptions. Using the *product rule for probabilities* an expert witness computed the chance that a random couple meets these characteristics, as 1:12,000,000.

Examples

- Two coins are given. One is fair (P(H)=0.5) and the other is biased with P(H)=2/3. <u>One of the coins is tossed once, resulting in H</u>. The other is <u>tossed three times, resulting in two heads</u>. Which coin is more likely to be the biased one?
- We won't look for the probability of the first or the second coin being the biased one, rather we look for the probability of the given outcomes in two different cases: the first coin being the fair one, and the second--the biased one, and vice versa.
- If we <u>assume that the first coin is fair</u>, then the probability of the heads is 1/2. The second coin must be the biased one, and the probability of it coming up with 2 heads and 1 tail in three tosses is 3*2/3*2/3*1/3 = 4/9. Note that there are three ways to get 2 heads: HHT, HTH, THH, the probability of each being 4/27. Thus, the probability of both coins coming up with the given results is 2/9.
- If, on the other hand, the <u>first coin is the biased one</u>, and the second coin is fair the probability of them resulting in the combination given in the problem is (2/3)*(3* 1/2*1/2*1/2) = 1/4, or 2/8 > 2/9. Therefore, it is more probable that the first coin is the biased one.

Examples

- Let X be the sum of spots from rolling 4 fair dice. Determine the <u>expected value</u>, the <u>variance</u>, and the <u>standard deviation</u> of the random variable X.
- Let X = X1 + X2 + X3 + X4, where X1, X2, X3, X4 be the random variables for the spots showing on the 1st, 2nd, 3rd, and 4th dice, respectively. [E(X) = Sum(X_value times P(x)).]

$$E(Y) = \sum_y y \times P(Y = y)$$

- For each die, E(X1)=1x1/6+2x1/6+3x1/6+4x1/6+5x1/6+6x1/6=3.5
- E(X) = E(X1)+E(X2)+E(X3)+E(X4) =3.5 + 3.5 + 3.5 + 3.5 = 14

Examples

- Let X be the sum of spots from rolling 4 fair dice. Determine the expected value, the variance, and the standard deviation of the random variable X. X = X1 + X2 + X3 + X4.

$$E(Y) = \sum_y y \times P(Y = y) \quad ; \quad Var(Y) = \frac{1}{N-1} \sum_y (y - E(Y))^2$$

- Because the rolls are independent:
- Var(X) = Var(X1) + Var(X2) + Var(X3) + Var(X4) .
- The variance for any single roll is: $(1/5)*(1-3.5)^2 + (1/5)*(2-3.5)^2 + (1/5)*(3-3.5)^2 + (1/5)*(4-3.5)^2 + (1/5)*(5-3.5)^2 + (1/5)*(6-3.5)^2 = 3.5$.
- So, Var(X) = 4 x 3.5 = 14. SD(X) = sqrt(Var(X)) = 3.74.
- So, from 4 dice, the expected value (Sum) is 14, with a SE of 3.74

Examples – Birthday Paradox

- The Birthday Paradox: In a random group of N people, what is the change that at least two people have the same birthday?
- E.x., if N=23, P>0.5. Main confusion arises from the fact that in real life we rarely meet people having the same birthday as us, and we meet more than 23 people.
- The reason for such high probability is that any of the 23 people can compare their birthday with any other one, not just you comparing your birthday to anybody else's.
- There are N-Choose-2 = 20*19/2 ways to select a pair or people. Assume there are 365 days in a year, P(one-particular-pair-same-B-day)=1/365, and
- P(one-particular-pair-failure)=1-1/355~0.99726.
- For N=20, 20-Choose-2 = 190. E={No 2 people have the same birthday is the event all 190 pairs fail (have different birthdays)}, then $P(E) = P(\text{failure})^{190} = 0.99726^{190} = 0.59$.
- Hence, P(at-least-one-success)=1-0.59=0.41, quite high.
- Note: for N=42 ➔ P>0.9 …

Summary

- What does it mean for two events A and B to be *statistically independent*?
- Why is the working rule under independence, $P(A \text{ and } B) = P(A)\,P(B)$, just a special case of the multiplication rule $P(A \,\&\, B) = P(A \mid B)\,P(B)$?
- *Mutual independence* of events $A_1, A_2, A_3, \ldots, A_n$ *if and only* if $P(A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n) = P(A_1)P(A_2)\ldots P(A_n)$

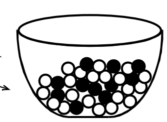
Binomial Distribution

The two-color urn model

N balls in an urn, of which there are

M black balls

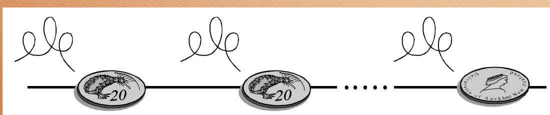
$N - M$ white balls



Sample n balls and count X = # black balls in sample

We will compute the probability distribution of the R.V. X

The biased-coin tossing model



toss 1 pr(H) = p | toss 2 pr(H) = p | toss n pr(H) = p

Perform n tosses and count X = # heads

We also want to compute the probability distribution of this R.V. X!
Are the two-color urn and the biased-coin models <u>related</u>? How do we present the models in <u>mathematical terms</u>?

The answer is: Binomial distribution

- The distribution of the number of heads in n tosses of a biased coin is called the ***Binomial distribution***.

Binomial(N, p) – the probability distribution of the number of Heads in an N-toss coin experiment, where the probability for Head occurring in each trial is p.
E.g., Binomial(6, 0.7)

	x	0	1	2	3	4	5	6
Individual	**pr($X = x$)**	0.001	0.010	0.060	0.185	0.324	0.303	0.118
Cumulative	**pr($X \le x$)**	0.001	0.011	0.070	0.256	0.580	0.882	1.000

For example $P(X=0) = P(\text{all 6 tosses are Tails}) =$

$$(1-0.7)^6 = 0.3^6 = 0.001$$

Binary random process

The ***biased-coin tossing model*** is a physical model for situations which can be characterized as a series of trials where:

- each trial has only two outcomes: *success* or *failure*;
- p = P(*success*) is the same for every trial; and
- trials are independent.

- The distribution of X = number of successes (heads) in N such trials is

Binomial(N, p)

Sampling from a finite population – Binomial Approximation

If we take a sample of size n

- from a much larger population (of size N)
- in which a proportion p have a characteristic of interest, then the distribution of X, the number in the sample with that characteristic,
- is *approximately* Binomial(n, p).
  - (Operating Rule: Approximation is adequate if $n / N < 0.1$.)
- Example, polling the US population to see what proportion is/has-been married.

Review

- For what types of situation is the urn-sampling model useful? For modeling binary random processes. When sampling with replacement, Binomial distribution is exact, where as, in sampling without replacement Binomial distribution is an approximation.
- For what types of situation is the biased-coin sampling model useful? Defective parts. Approval poll of cloning for medicinal purposes. Number of Boys in 151 presidential children (90).
- Give the three essential conditions for its applicability. (two outcomes; same p for every trial; independence)

The Expected value

- The expected value:

$$E(X) = \sum_{\text{all } x} x \, P(x)$$

- = Sum of (value times probability of value)

Example

A couple wants to have children, but they insist on stopping when they have at least one of each gender or at most 3 children example, where X ={number of Girls} we have:

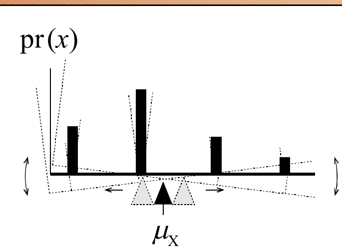
X	0	1	2	3
pr(x)	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$E(X) = \sum_x x\,P(x)$$

$$= 0\times\frac{1}{8} + 1\times\frac{5}{8} + 2\times\frac{1}{8} + 3\times\frac{1}{8}$$

$$= 1.25$$

The expected value as the point of balance



The mean μ_X is the balance point.

Expected values

- The game of chance: cost to play:\$1.50; Prices {\$1, \$2, \$3}, probabilities of winning each price are {0.6, 0.3, 0.1}, respectively.
- Should we play the game? What are our chances of winning/loosing?

Prize (\$)	x	1	2	3	
Probability	pr(x)	0.6	0.3	0.1	
What we would "expect" from 100 games					*add across row*
Number of games won		0.6×100	0.3×100	0.1×100	↙
\$ won		$1 \times 0.6 \times 100$	$2 \times 0.3 \times 100$	$3 \times 0.1 \times 100$	Sum

Total prize money = Sum; ***Average prize money*** = Sum/100 $= 1 \times 0.6 + 2 \times 0.3 + 3 \times 0.1 = 1.5$

Theoretically Fair Game: price to play EQ the expected return!

Binomial Probabilities - ?!?!?!

- Suppose X ~ Binomial(n, p), then the probability

$$P(X=x) = \binom{n}{x} p^x (1-p)^{(n-x)}, \quad 0 \le x \le n$$

- Where the binomial coefficients are defined by

$$\binom{n}{x} = \frac{n!}{(n-x)!\ x!}, \quad n! = 1 \times 2 \times 3 \times \ldots \times (n-1) \times n$$

n-factorial

Binomial Formula - examples

- Rolling a pair of dice 4 times. What's the chance of getting at least 2 but less than 4 aces (1-1)?

$$\sum_x P(X=x) = \sum_x \binom{n}{x} p^x (1-p)^{(n-x)}$$

- Let X = {Event an ace turns up}, p=P(success)=1/36

$$P(2 \le X < 4) = \sum_{x=2}^{3} \binom{4}{x} p^x (1-p)^{(4-x)} =$$

$$\binom{4}{2} p^2 (1-p)^2 + \binom{4}{3} p^3 (1-p)^1 =$$

$$6 \times \left(\frac{1}{36}\right)^2 \times \left(\frac{35}{36}\right)^2 + 4 \times \left(\frac{1}{36}\right)^3 \times \left(\frac{35}{36}\right)^1 = 0.005$$

For the Binomial distribution . . . Mean/SD

$E(X) = n\,p$

$SD(X) = \sqrt{np(1-p)}$

Example

- According to the data base of a large multinational airline company, 35% of all its 475 pilots are over 40 years of age. The company is about to purchase the latest model *Boeing* and is planning to select a random sample of 25 pilots to receive training in flying this new plane.

  Let X be the number of pilots over 40 years of age in this sample.

  - Discuss the validity of using the Binomial distribution in this situation. (Hint: each trial has only two outcomes: p = P(*success*) is the same for every trial; and trials are independent. If n/N <0.1 binomial approximation is justified! 25/475 ~ 0.05).

Examples

- 35% of all its 475 pilots are over 40. The company is to select a random sample of 25 pilots. X = # pilots over 40 years in sample.
  - State the value of the parameter(s) of this distribution.
    - Binomial(25, 0.35)
  - Assuming that the Binomial distribution you have described above is an appropriate model for X, Find the probability that:

(i) more than 7 of the pilots selected are over 40 years of age.
P(X>7)=1-P([0,6])=0.173, From the online table, but need to know how to compute by hand

(ii) 5 or 6 of the pilots selected are over 40 years of age.

$$P(X=5) = \binom{25}{5} 0.35^5 \times 0.65^{20} + \binom{25}{6} 0.35^6 \times 0.65^{19}$$

$$= 0.051 + 0.091 = 0.142$$

(iii) between 13 and 18 (inclusive) of the pilots selected are over 40 years of age. P(13<=X<=18) = 0.06, From the online table.

Examples

- 35% of all its 475 pilots are over 40 years of age. The company is to select a random sample of 25 pilots. *X* = # pilots over 40 years in this sample.
  - How many of the pilots selected would you expect to be over 40 years of age? What is the standard deviation of X?
    - $E(X) = n \times p = 25 \times 0.35 = 8.75$
    - $SD^2(X) = Var(X) = n \times p \times (1-p) = 5.6875$
    - $SD(X) = 2.385$
  - Why would the airline company be interested in the variable "pilots over 40 years of age"? Suggest another variable the company may be interested in measuring? Briefly justify your answer. (GO TO Ch. 21, CI's)