UCLA STAT 251

Statistical Methods for the Life Sciences

●**Instructor**: **Ivo Dinov,**

Asst. Prof. In Statistics and Neurology

University of California, Los Angeles, Winter 2003

http://www.stat.ucla.edu/~dinov/courses_students.html

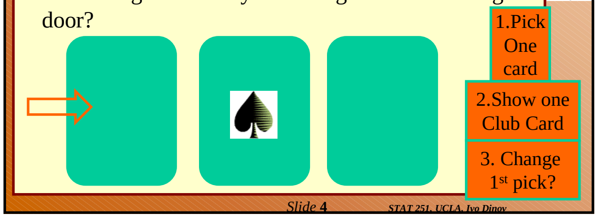
Probabilities, Bayesian Rule, Marginal and Joint PMF/PDFs

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?



1.Pick One card

2.Show one Club Card

3. Change 1st pick?

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:

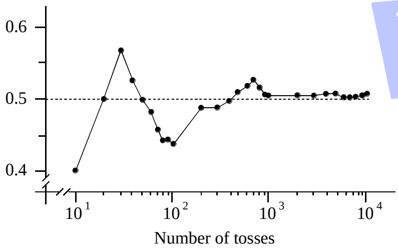
Let's Make a Deal Paradox.

- 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.

Let's Make a Deal Paradox.

- Demo: AdditionalAids.dir/StatGames.exe
  - Uncertainty→Pick a door

Long run behavior of coin tossing



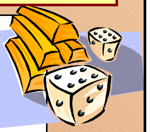


Figure 4.1.1 Proportion of heads versus number of tosses for John Kerrich's coin tossing experiment.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Definitions …

- 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 more stable as the number of tosses increases. The law of averages applies to relative frequencies not absolute counts of #H and #T.
- Two widely held misconceptions about what the law of averages 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.

Sample Spaces and Probabilities

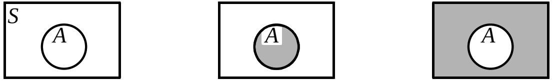
- When the relative frequency of an event in the past is used to estimate the probability that it will occur in the future, what assumption is being made?
  - The underlying process is stable over time;
  - Our relative frequencies must be taken from large numbers for us to have confidence in them as probabilities.
- All statisticians agree about how probabilities are to be combined and manipulated (in math terms), however, not all agree what probabilities should be associated with for a particular real-world event.
- When a weather forecaster says that there is a 70% chance of rain tomorrow, what do you think this statement means? (Based on our past knowledge, according to the barometric pressure, temperature, etc. of the conditions we expect tomorrow, 70% of the time it did rain under such conditions.)

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.

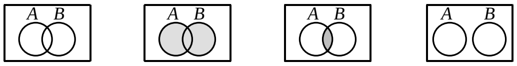


(a) Sample space containing event *A* (b) Event *A* shaded (c) $\bar{A}$ shaded

Figure 4.4.1 An event *A* in the sample space *S*.

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

Figure 4.4.2 Two events.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

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*.

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

Job losses in the US

TABLE 4.4.1 Job Losses in the US (in thousands) for 1987 to 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 cont.

	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

TABLE 4.4.2 Proportions of Job Losses from Table 4.4.1

	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?

Properties of probability distributions

- A sequence of number $\{p_1, p_2, p_3, \ldots, pn\}$ is a probability distribution for a sample space $S = \{s_1, s_2, s_3, \ldots, s_n\}$, if $pr(s_k) = p_k$, for each 1<=k<=n. The two essential properties of a probability distribution $p_1, p_2, \ldots, p_n$?

$$p_k \geq 0; \quad \sum_k p_k = 1$$

- How do we get the probability of an event from the probabilities of outcomes that make up that event?
- If all outcomes are <u>distinct</u> & <u>equally likely</u>, how do we calculate *pr(A)* ? If $A = \{a_1, a_2, a_3, \ldots, a_9\}$ and $pr(a_1)=pr(a_2)=\ldots=pr(a_9)=p;$ then

$pr(A) = 9 \text{ x } pr(a_1) = 9p$.

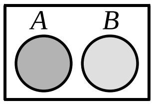
Example of probability distributions

- 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.

Rules for manipulating Probability Distributions

For mutually exclusive events,

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



From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Review

- If *A* and *B* are mutually exclusive, what is the probability that both occur? (0) What is the probability that at least one occurs? (sum of probabilities)
- If we have two or more mutually exclusive events, how do we find the probability that at least one of them occurs? (sum of probabilities)
- Why is it sometimes easier to compute *pr(A)* from $pr(A) = 1 - pr(\bar{A})$? (The complement of the even may be easer to find or may have a known probability. E.g., a random number between 1 and 10 is drawn. Let A ={a number less than or equal to 9 appears}. Find pr(A) = 1 – pr($\bar{A}$)). probability of $\bar{A}$ is pr({10 appears}) = 1/10 = 0.1. Also Monty Hall 3 door example!

Melanoma – type of skin cancer – an example of laws of conditional probabilities

TABLE 4.6.1: 400 Melanoma Patients by Type and Site

	Site			
Type	Head and Neck	Trunk	Extremities	Row Totals
Hutchinson's melanomic freckle	22	2	10	34
Superficial	16	54	115	185
Nodular	19	33	73	125
Indeterminant	11	17	28	56
Column Totals	68	106	226	400

Contingency table based on Melanoma histological type and its location

Conditional Probability

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

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

Suppose we select one out of the 400 patients in the study and we want to find the probability that the cancer is on the *extremities* *given that* it is of type *nodular*: P = 73/125 = P(C. on Extremities | Nodular)

$$\frac{\#\text{nodular patients with cancer on extremities}}{\#\text{nodular patients}}$$

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

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

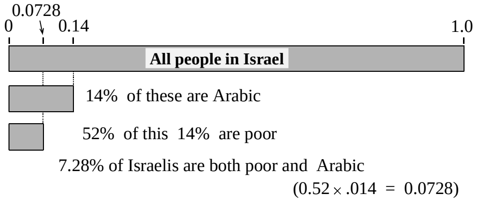


Figure 4.6.1 Illustration of the multiplication rule.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

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

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

	Jews	Arabs	Total
Poor	0.0772	0.0728	**Assume 0.15**
Not-poor	0.7828	0.0672	0.85
Total	0.86	0.14	1.0

A tree diagram for computing conditional probabilities

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*}) =
4/7 x 3/6 + 4/6 x 3/7 = 4/7.

Mutually exclusive

A tree diagram

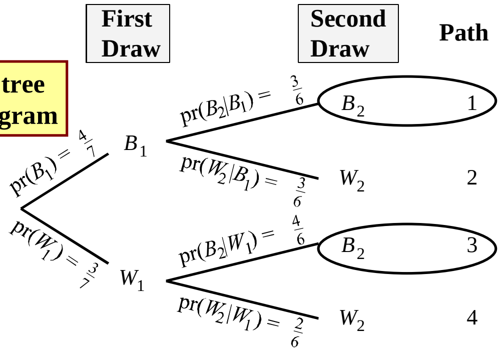
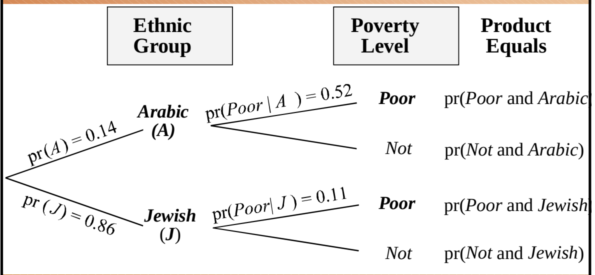


Figure 4.6.2 Tree diagram for a sampling problem.

Tree diagram for poverty in Israel



2-way table for poverty in Israel

pr(*Poor* and *Arabic*) =
pr(*Poor*|*Arabic*) × pr(*Arabic*)
[= 52% of 14%]

pr(*Poor* and *Jewish*) =
pr(*Poor*|*Jewish*) × pr(*Jewish*)
[= 11% of 86%]

		Ethnicity		
		Arabic	Jewish	Total
Poverty	Poor	.52 × .14	.11 × .86	?
	Not poor	?	?	?
	Total	.14	.86	1.00

pr(*Arabic*) = .14

pr(*Jewish*) = .86

Figure 4.6.4 Proportions by Ethnicity and Poverty.

P(A & B) = P(A | B) x P(B),
P(A | B) = P(A & B) / P(B)
P(A & B) = P(B & A) = P(B | A) x P(A).
P(A | B) = [P(B | A) x P(A)] / P(B).

2-way table for poverty in Israel cont.

pr(*Poor* and *Arabic*) =
pr(*Poor*|*Arabic*) × pr(*Arabic*)
[= 52% of 14%]

pr(*Poor* and *Jewish*) =
pr(*Poor*|*Jewish*) × pr(*Jewish*)
[= 11% of 86%]

		Ethnicity		
		Arabic	Jewish	Total
Poverty	Poor	.52 × .14	.11 × .86	?
	Not poor	?	?	?
	Total	.14	.86	1.00

pr(*Arabic*) = .14

pr(*Jewish*) = .86

TABLE 4.6.3 Proportions by Ethnicity and Poverty

		Ethnicity		
		Arabic	**Jewish**	**Total**
Poverty	Poor	.0728	.0946	.1674
	Not Poor	.0672	.7654	.8326
	Total	.14	.86	1.00

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)

TABLE 4.6.5 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
2 - 2.99	73		2	
3 - 3.99	15		7	
4 - 4.99	3		7	
5 - 5.99	2		15	
6 -11.99	2		36	
12+	0		21	
Total	297		88	

Test cut-off

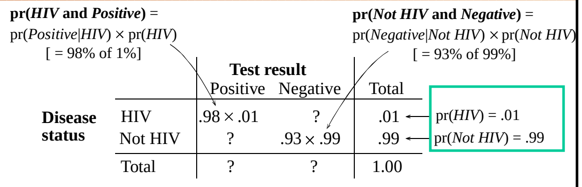
False-Negatives (β)

False-Positives (α)

Power of a test is: (1- β)
1-P(FNE)= 1-P(Neg|HIV) ~ 0.976

Adapted from Weiss et al.[1985]

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%]

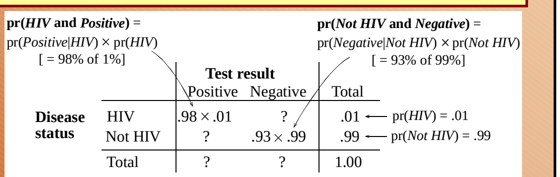
		Test result		
		Positive	Negative	Total
Disease status	HIV	.98 × .01	?	.01
	Not HIV	?	.93 × .99	.99
	Total	?	?	1.00

pr(*HIV*) = .01
pr(*Not HIV*) = .99

Figure 4.6.6 Putting HIV information into the table.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

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%]

		Test result		
		Positive	Negative	Total
Disease status	HIV	.98 × .01	?	.01
	Not HIV	?	.93 × .99	.99
	Total	?	?	1.00

pr(*HIV*) = .01
pr(*Not HIV*) = .99

TABLE 4.6.6 Proportions by Disease Status and Test Result

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

Proportions of HIV infections by country

TABLE 4.6.7 Proportions Infected with HIV

Country	**No. AIDS Cases**	**Population (millions)**	**pr(HIV)**	Having | Test **pr(HIV | Positive)**
United States	218,301	252.7	0.00864	0.109
Canada	6,116	26.7	0.00229	0.031
Australia	3,238	16.8	0.00193	0.026
New Zealand	323	3.4	0.00095	0.013
United Kingdom	5,451	57.3	0.00095	0.013
Ireland	142	3.6	0.00039	0.005

Statistical independence

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

  i.e. if **pr(*A* | *B*) = pr(*A*)**

- Similarly, P(*B* | *A*) = P(*B*), since

  $P(B|A)=P(B \& A)/P(A) = P(\cancel{A|}B)P(B)/\cancel{P(A})= P(B)$

- If *A* and *B* are *statistically independent*, then

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

People vs. Collins

TABLE 4.7.2 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.

Summary

- What does it mean for two events *A* and *B* to be *statistically independent*?
- Why is the working rule under independence, *P(A and B) = P(A) P(B)* , just a special case of the multiplication rule *P(A & B) = P(A | B) P(B)* ?
- *Mutual independence* of events $A_1, A_2, A_3, \ldots, A_n$ *if and only* if $P(A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n) = P(A_1)P(A_2)\ldots P(A_n)$
- What do we mean when we say two human characteristics are *positively associated*? *negatively associated*? (blond hair – blue eyes, pos.; black hair – blue eyes, neg.assoc.)

Summary of ideas cont.

- An ***event*** is a collection of outcomes
- An event ***occurs*** if any outcome making up that event occurs
- The probability of event A can be obtained by adding up the probabilities of all the outcomes in A
- If all outcomes are equally likely,

$$\text{pr}(A) = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}}$$

Summary of ideas cont.

- The ***complement*** of an event *A*, denoted $\overline{A}$, occurs if *A* does not occur
- It is useful to represent events diagrammatically using ***Venn diagrams***
- ***A union*** of events, ***A* or *B*** contains all outcomes in *A* or *B* (including those in both). It occurs if at least one of *A* or *B* occurs
- ***An intersection*** of events, ***A* and *B*** contains all outcomes which are in ***both*** *A* and *B*. It occurs only if both *A* and *B* occur
- ***Mutually exclusive*** events cannot occur at the same time

Summary of ideas cont.

- The ***conditional probability*** of *A* occurring ***given*** that *B* occurs is given by
$$\text{pr}(A \mid B) = \frac{\text{pr}(A \text{ and } B)}{\text{pr}(B)}$$
- Events *A* and *B* are ***statistically independent*** if knowing whether *B* has occurred gives no new information about the chances of *A* occurring, i.e. if $P(A \mid B) = P(A) \rightarrow P(B|A)=P(B)$.
- If events are physically independent, then, under any sensible probability model, they are also statistically independent
- Assuming that events are independent when in reality they are not can often lead to answers that are grossly too big or grossly too small

Formula summary cont.

- $\text{pr}(S) = 1$
- $\text{pr}(\overline{A}) = 1 - \text{pr}(A)$
- If *A* and *B* are mutually exclusive events, then $\text{pr}(A \text{ or } B) = \text{pr}(A) + \text{pr}(B)$
  (here "or" is used in the inclusive sense)
- If $A_1, A_2, \ldots, A_k$ are mutually exclusive events, then
  $\text{pr}(A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_k) = \text{pr}(A_1)+\text{pr}(A_2)+\ldots+\text{pr}(A_k)$

Formula summary cont.

Conditional probability

- Definition:

$$\mathrm{pr}(A\,|\,B) \;=\; \frac{\mathrm{pr}(A \text{ and } B)}{\mathrm{pr}(B)}$$

- Multiplication formula:

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

Formula summary cont.

Multiplication Rule under independence:

- If A and B are independent events, then

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

- If $A_1, A_2, \dots, A_n$ are mutually independent,

$\mathrm{pr}(A_1 \text{ and } A_2 \text{ and } \dots \text{ and } A_n) = \mathrm{pr}(A_1)\ \mathrm{pr}(A_2) \dots \mathrm{pr}(A_n)$

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)=1-1/365 ~ 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(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 …

Marginal vs. Joint Bivariate Distributions

- The joint density, P{***X,Y***}, of the number of minutes waiting to catch the first fish, ***X*** , and the number of minutes waiting to catch the second fish, ***Y***, is given below.

P {*X* = i,*Y* = k }			k		
		1	**2**	**3**	**Row Sum P{ *X* = i }**
	1	0.01	0.02	0.08	0.11
i	2	0.01	0.02	0.08	0.11
	3	0.07	0.08	0.63	0.78
Column Sum					
P {*Y* =k }		**0.09**	**0.12**	**0.79**	**1.00**

The (joint) chance of **waiting 3 minutes to catch the first fish and 3 minutes to catch the second fish** is:

The (marginal) chance of **waiting 3 minutes to catch the first fish** is:

The (marginal) chance of **waiting 2 minutes to catch the first fish** is:

Marginal vs. Joint Bivariate Distributions

P {*X* = i,*Y* = k }			k		
		1	2	3	Row Sum P{ *X* = i }
	1	0.01	0.02	0.08	0.11
i	2	0.01	0.02	0.08	0.11
	3	0.07	0.08	0.63	0.78
Column Sum					
P {*Y* =k }		**0.09**	**0.12**	**0.79**	**1.00**

The chance of **waiting at least two minutes to catch the first fish** is

P {X =2 } = 0 .11 +0 .78 =0 .89

P {X =2 ,Y =1 } +P{X =2 ,Y =2 } + P{X =2 ,Y =3 } + P{X =3 ,Y =1 } + P{X =3 ,Y =2 } + P{X =3 ,Y =3 } =

0 .01 +0 .02 + 0 .08 +0 .07 +0 .08 +0 .63 = 0 .89

$p_X(2)$ = p(2 ,1) + p(2 ,2) + p 2 ,3) = 0 .01 + 0 .02 + 0 .08 = 0 .11

$p_X(2)$ = Sum_{y :p (2 ,y)>0} [p (2 ,y)] = p(2 ,1) + p(2 ,2) + p(2 ,3) = 0 .01 +0 .02 + 0 .08 = 0 .11

Marginal vs. Joint Bivariate Distributions

P {*X* = i,*Y* = k }			k		
		1	2	3	Row Sum P{ *X* = i }
	1	0.01	0.02	0.08	0.11
i	2	0.01	0.02	0.08	0.11
	3	0.07	0.08	0.63	0.78
Column Sum					
P {*Y* =k }		**0.09**	**0.12**	**0.79**	**1.00**

The chance of **waiting at most two minutes to catch the first fish and at most two minutes to catch the second fish** is

P {X =2 ,Y =2 } = 0 .06

P {X =1 ,Y =1 } + P{X =1 ,Y =2 } + P{X =2 ,Y =1 } + P{X = 2 ,Y =2 } = 0 .01 + 0 .02 + 0 .01 + 0 .02 = 0 .06

F (2 ,2) = P{X <=2 ,Y<=2 } = 0 .06

$F_X(2)$ = F(2 , 8) = F(2 ,3) = 0 .22

Law of Total Probability & Bayesian Rule

A useful identity (sometimes called the law of total probability) is

$$P(E) = P(E | F) P(F) + P(E | F^c) P(F^c)$$

A version of Bayes Formula is

$$P(F | E) = P(E | F) P(F) / P(E) \quad \leftarrow \rightarrow$$

$$P(F | E) = P(E | F) P(F) / [P(E | F) P(F) + P(E | F^c) P(F^c)]$$

Example: Given that probability of survival given genetic pre-disposure is 0.3, what is the probability that someone has the genetic disorder given that the subject died from the disease? Assume this genetic disorder occurs in 1% of population and P(dying | no genetic disorder) = 0.1%.

$P(G|S^c) = P(S^c|G) P(G)/[P(S^c|G)P(G) + P(S^c|G^c)P(G^c)] =$

$0.7*0.01 / [0.7*0.01 + 0.1*0.99] = 0.07$

Hints for Prob. 1 in Project I

- Suppose a scientific experiment (e.g., reading DNA sequences) has exactly 4 possible outcomes (encoded by: ***A, C, T, G***). Assume all 4 possible outcomes are independent and equally likely at each observation. A composite experiment is carried out until one of the following stopping criteria occurs:
- A maximum of 4 observations are made, or
- 3 ***T***-bases occur in any order, or
- A ***TAA*** triplet ***(stop codon)*** appears in this order, or
- A ***G*** occurs on position 2 or 3.

Let ***X*** be the random variable representing the total number of ***T***-bases observed in this composite experiment. Identify a meaningful probability distribution of the random variable ***X*** . What are the *mean* and the *standard deviation* of ***X*** ?

Hints for Prob. 1 in Project I

X = the total number of T-bases observed in the composite experiment. X is a discrete random variable which its values will take on {0, 1, 2, 3}

Ignoring the stopping criteria, the entire sample space will be …. = 256 total possible permutations. It would be rather time consuming and tedious to list all 256 permutations. So we will create a new R.V., Y.

Y = the total number of bases observed until a stopping criterion is met. Y is a discrete R.V. which its values will take on {2, 3, 4}. Conditioning X on Y, now we can calculate their joint probabilities.

Y = 2 $X = 0 \rightarrow P(X = 0 \mid Y = 2) = \ldots$

$X = 1 \rightarrow P(X = 1 \mid Y = 2) = \ldots$

P(X=0) = Sum_{all y values} $P(X = 0 \mid Y = y)$, etc.