Chapter 3: Probability

- Sample space: Set of all possible outcomes
- Event: Any set of outcomes (subset of sample space)
- Probability of event: frequency of the event over a large number of trials
- Ex.) Tuberculin skin test to detect tuberculosis

Outcome	Prob
Positive	0.1
Negative	0.7
Uncertain	0.2

Sample space: {Positive, Negative, Uncertain}

Possible events:

Event	Prob
Positive	0.1
Negative	0.7
Uncertain	0.2
{Positive, Negative}	0.8
{Positive, Uncertain}	0.3
{Negative, Uncertain}	0.9
{Positive, Negative, Uncertain}	1

- Notation: $Pr(E)$ = Probability of event E

$E = \{+, -\}$, then $Pr(E) = 0.8$

- Two events are <u>mutually exclusive</u> if they cannot both happen at the same time.

Ex) $E_1 = \{\text{Positive, Negative}\}$
$E_2 = \{\text{Uncertain}\}$
$E_3 = \{\text{Negative, Uncertain}\}$

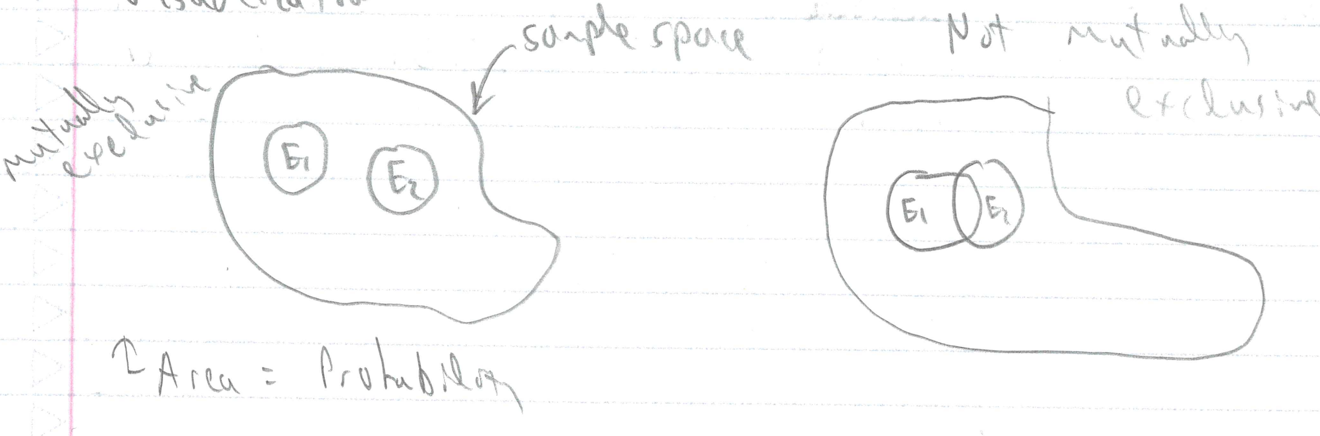
E_1 and E_2 are mutually exclusive

E_1 and E_3 can both happen if outcome is "negative"
E_1 and E_2 can both happen if outcome is "uncertain"

"or" rule

- If E_1 and E_2 are mutually exclusive, then
$$\boxed{Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2)}$$
E_1 or E_2 occurring

- Visualization

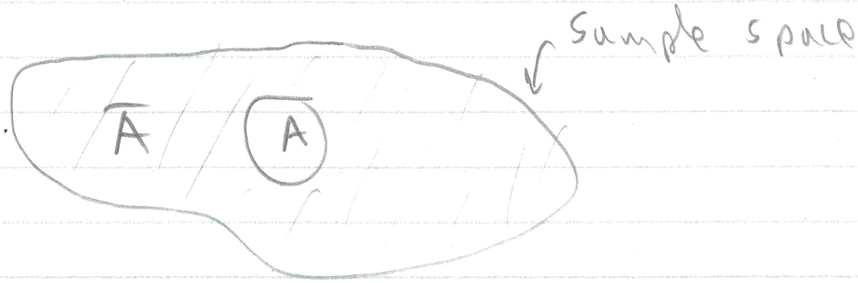


Area = Probability

- The <u>complement</u> of A, denoted $\bar{A}$, is the event that A does not occur.

- Ex.) $E_1 = \{\text{Positive, Negative}\}$

  $\bar{E}_1 = \{\text{Uncertain}\}$



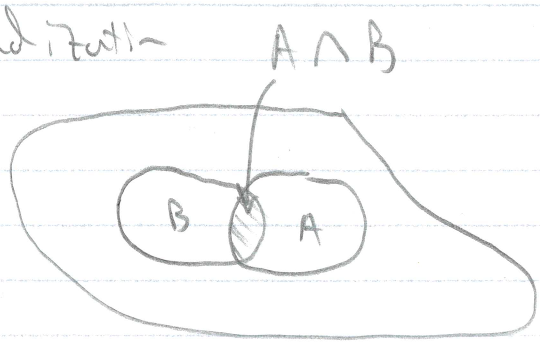
- $Pr(\{\text{Sample Space}\}) = 1$
- $\boxed{Pr(\bar{A}) = 1 - Pr(A) \iff Pr(A) = 1 - P(\bar{A})}$

Ex.) $Pr(\text{Uncertain}) = 1 - Pr(\{\text{Positive, Negative}\}) = 1 - 0.8 = 0.2$

- Two events are <u>independent</u> if

  $$\boxed{\underbrace{Pr(A \wedge B)}_{\text{A and B occurring}} = Pr(A) \cdot Pr(B)}$$

- Visualization

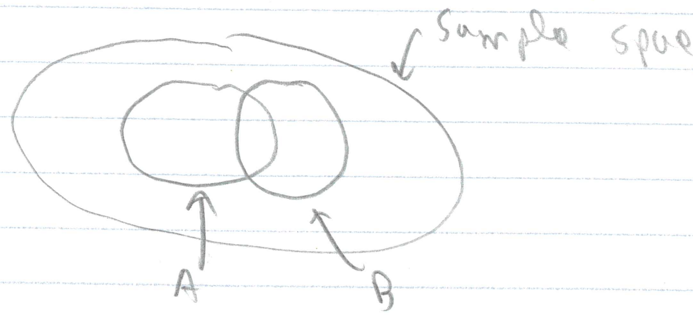


- Ex.) $\Pr(\overbrace{\text{Mother DBP} \geq 90}^{A}) = 0.1$

  $\Pr(\underbrace{\text{Father DBP} \geq 90}_{B}) = 0.2$

If independent, $\Pr(A \wedge B) = 0.1 \cdot 0.2 = 0.02$

- Why called "independent"?



A is 10% of sample space area
A is 10% of B's area

B becomes sample space if it occurs and we don't know A, so A still has a 10% chance, even if we know B

$\Rightarrow$ B occurring or not occurring does not change probability of A

- <u>Exercise:</u> A^+ = Doctor A diagnoses +
  B^+ : Doctor B diagnoses +

$\Pr(A^+) = 0.1$, $\Pr(B^+) = 0.17$, $\Pr(A^+ \wedge B^+) = 0.08$

Are the events independent?

Solution: No! $0.1 \cdot 0.17 = 0.017 \neq 0.08$

- Events $A_1, \ldots, A_n$ are <u>mutually independent</u> if

$$Pr(A_1, \ldots, A_n) = Pr(A_1) \cdots Pr(A_n) = \prod_{i=1}^{n} A_i$$

- <u>Addition Law of Prob</u>:

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

- <u>Exercise</u>: what is the probability that doctor A or B diagnose +?

Solution

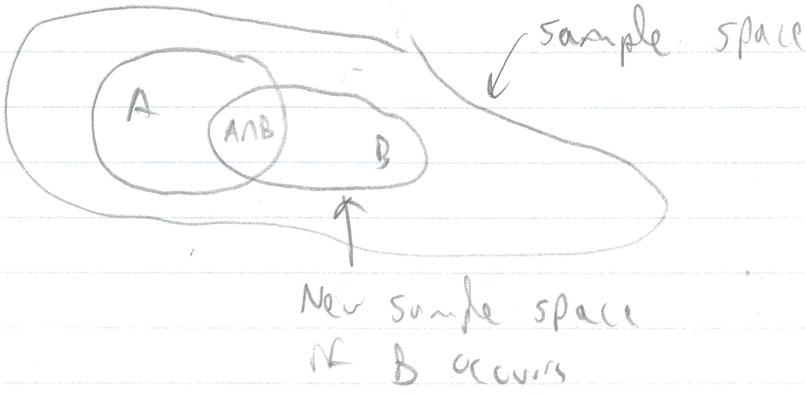
$$Pr(A^+ \cup B^+) = 0.1 + 0.17 - 0.08 = 0.19$$

- <u>Note</u>: $Pr(A \cap B) = 0$ if A and B are mutually exclusive

- <u>Conditional Probability</u>:

$$\boxed{Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)}} = \text{"Probability of A given B"}$$

↑ what proportion of B is also A



- If $A \perp\!\!\!\perp B$ then $Pr(A|B) = Pr(A)$

  Independent

proof: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A)\,\cancel{Pr(B)}}{\cancel{Pr(B)}} = Pr(A)$

- The relative risk of B given A: $\frac{Pr(B|A)}{Pr(B|\bar{A})}$

  Measure of dependence between A and B

  $=1 \Rightarrow$ Independent
  $<1 \Rightarrow$ B more likely given $\bar{A}$
  $>1 \Rightarrow$ B more likely given A

- Exercise: $Pr(A^+) = 0.1$, $Pr(B^+) = 0.17$, $Pr(A^+ \cap B^+) = 0.08$

Q: $Pr(A^+|B^+)$, $Pr(B^+|A^+)$, RR B^+ given A^+?

$Pr(A^+|B^+) = \frac{0.08}{0.17}$ $\qquad Pr(B^+|A^+) = \frac{0.08}{0.1}$

$Pr(B^+|\overline{A^+}) = \frac{Pr(B^+ \cap \overline{A^+})}{Pr(\overline{A^+})} = \frac{0.09}{1-0.1}$

$$\begin{array}{ccccc} Pr(B^+) & = & Pr(B^+ \cap A^+) & + & Pr(B^+ \cap \overline{A^+}) \\ \downarrow & & \downarrow & & \downarrow \\ 0.17 & & 0.08 & & ? \end{array}$$

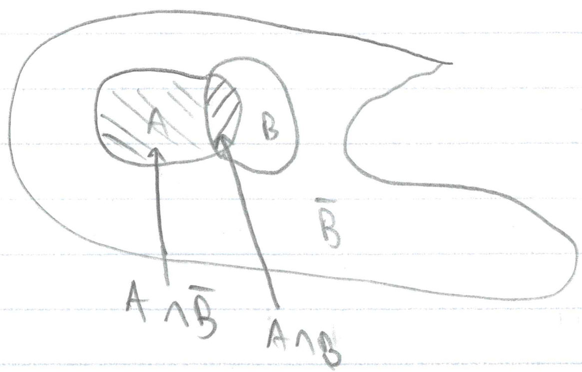
$\Rightarrow Pr(B^+ \cap \overline{A^+}) = 0.09$

Note: In biostat, typically write $\overline{A^+} = A^-$

- Law of Tutatal Probability

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

$$= Pr(A|B)\,Pr(B) + Pr(A|\bar{B})\,Pr(\bar{B})$$



- Exercise: $A = \{\text{mamogram } +\}$

$B = \{\text{Breast cancer } +\}$

$Pr(B|\bar{A}) = 0.0002$
$Pr(B|A) = 0.1$
$Pr(A) = 0.07$

Q: What is $Pr(B)$?

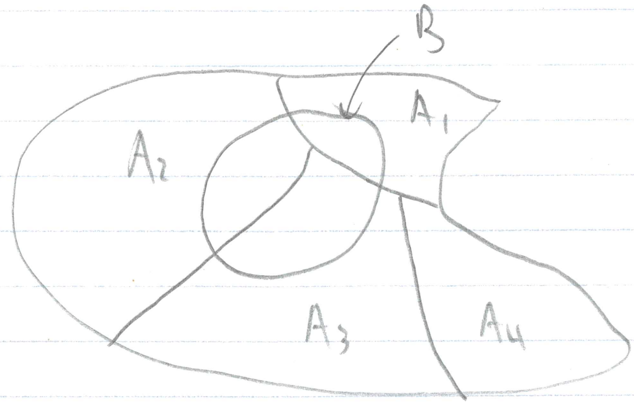
$$Pr(B) = Pr(B|\bar{A})\,Pr(\bar{A}) + Pr(B|A)\,Pr(A)$$

$$= 0.0002 \cdot 0.93 + 0.1 \cdot 0.07$$

mutually exclusive
exhaustive

- More generally if $A_1, A_2, \ldots, A_k$ partitions the sample space, then

$$Pr(B) = \sum_{h=1}^{k} Pr(B \mid A_h)\, Pr(A_h)$$



Ex.)	$Pr(A_h)$	$Pr(B \mid A_h)$
$A_1 = \{60-64\}$	0.45	0.024
$A_2 = \{65-69\}$	0.28	0.046
$A_3 = \{70-74\}$	0.2	0.088
$A_4 = \{75+\}$	0.07	0.153

$B = \{\text{cataract}\}$

$Pr(B) = \ldots$

- Predictive value Positive (PV^+) (also called Precision)

  $Pr(\text{disease} \mid \text{test}^+)$

- Predictive value negative (PV^-)

  $Pr(\text{no disease} \mid \text{test}^-)$

- Confusion Matrix

		Test	
		Positive (PP)	Negative (PN)
Truth	Positive (P)	True Positive (TP)	False Negative (FN) Type II error
	Negative (N)	False Positive (FP) Type I error	True Negative (TN)

$TPR = \frac{TP}{P}$ (sensitivity) (power)

$FNR = \frac{FN}{P}$

$FPR = \frac{FP}{N}$

$TNR = \frac{TN}{N}$ (specificity)

$PV^+ = \frac{TP}{PP}$

$FOR = \frac{FN}{PN}$

$FDR = \frac{FP}{PP}$

$PV^- = \frac{TN}{PN}$

- Ex.) $A = \{\text{Mamogram}^+\}$ $\quad Pr(B|A) = 0.1$
  $B = \{\text{Breast cancer} +\}$ $\quad Pr(B|\bar{A}) = 0.0002$

PV^+? PV^-?

$PV^+ = Pr(B|A) = 0.1$

$PV^- = Pr(\bar{B}|\bar{A}) = 1 - Pr(B|\bar{A}) = 1 - 0.0002 = 0.9998$

- __Sensitivity__ = Pr(test + | disease +) = true positive rate
- __Specificity__ = Pr(test − | disease −) = True negative rate
- Note: "test" can also refer to other items, like symptoms
- Ex.) 90% of folks with lung cancer are smokers
  30% of folks without lung cancer are smokers

  Q.) What are sensitivity and specificity of smoking as a screening test?

  A = smoker, B = lung cancer

  Sensitivity $= Pr(A|B) = 0.9$
  specificity $= Pr(\bar{A}|\bar{B}) = 1 - Pr(A|\bar{B}) = 1 - 0.3 = 0.7$

- Bayes Rule

$$\underbrace{Pr(B|A)}_{\text{"Posterior"}} = \frac{Pr(A|B)\overbrace{Pr(B)}^{\text{"Prior"}}}{Pr(A)}$$

- You typically get $Pr(A)$ by the law of total probability

$$Pr(A) = Pr(A|B)\,Pr(B) + Pr(A|\bar{B})\,Pr(\bar{B})$$

- Motivation: You want $Pr(\text{disease}^+ | \text{test}^+) = PV^+$ but in practice you typically only have $\underbrace{Pr(\text{test}^+ | \text{disease}^+)}_{\text{sensitivity}}$ and $\underbrace{Pr(\text{test}^- | \text{disease}^-)}_{\text{specificity}}$.

$$PV^+ = \frac{\text{Sensitivity} \cdot x}{\text{Sensitivity} \cdot x + \text{specificity} \cdot (1-x)}$$

$x = Pr(B)$
$=$ prevalence of disease

- Example: 84% of hypertensives and 24% of normotensives are classified as hypertensive by a mall machine. Prevalence is 20%.

Q.) PV^+ and PV^-?

$A = \{\text{test}^+\}$, $B = \{\text{hypertensive}\}$

wvrt

$Pr(B|A)$ and $Pr(\bar{B}|\bar{A})$

$PV^+ = Pr(B|A) = \dfrac{Pr(A|B)\ Pr(B)}{Pr(A)}$, $PV^- = Pr(\bar{B}|\bar{A}) = \dfrac{Pr(\bar{A}|\bar{B})\ Pr(\bar{B})}{Pr(\bar{A})}$

$Pr(B) = 0.2 \Rightarrow Pr(\bar{B}) = 0.8$

$Pr(A|B) = 0.84$

$Pr(A|\bar{B}) = 0.23 \Rightarrow Pr(\bar{A}|\bar{B}) = 1 - 0.23 = 0.77$

$Pr(A) = Pr(A|B)\,Pr(B) + Pr(A|\bar{B})\,Pr(\bar{B})$

$= 0.84 \cdot 0.2 + 0.23 \cdot 0.8$

$= 0.352$

$\Rightarrow Pr(\bar{A}) = 0.648$

$PV^+ = \dfrac{0.84 \cdot 0.2}{0.352} = 0.48$

$PV^- = \dfrac{0.77 \cdot 0.8}{0.648} = 0.95$

- Recall: Law of Total Probability may include more than 2 events

$$Pr(A) = Pr(A|B_1)Pr(B_1) + Pr(A|B_2)Pr(B_2) + \dots + Pr(A|B_k)Pr(B_k)$$

- <u>Prevalence</u> = Pr(disease+)

  ↕ proportion population that has disease

- <u>Cumulative Incidence</u> = Pr(will get disease | never had it)

Exercise 3.53 – 3.63 of Rosner

M^+ = mother smoker
F^+ = Father smoker
O^+ = offspring has asthma

Scenario	$Pr(O^+ \mid \text{Scenario})$
M^+, F^+	0.15
M^+, F^-	0.13
M^-, F^+	0.05
M^-, F^-	0.04

1.) $Pr(M^+) = 0.4$, $Pr(F^+) = 0.5$, $M^+ \perp\!\!\!\perp F^+$, what is $Pr(F^+, M^+)$

A: $0.4 \cdot 0.5 = 0.2$

2.) What is $Pr(F^+ \mid M^-)$?

A: 0.5 b/c independence

3.) Suppose $Pr(M^+ \mid F^+) = 0.6$
$Pr(M^+ \mid F^-) = 0.2$
$Pr(F^+) = 0.5$

what is $Pr(F^+ \wedge M^-)$

A: $Pr(F^+ \wedge M^-) = Pr(M^- \mid F^+) \, Pr(F^+) = (1 - 0.6) \, 0.5 = 0.5 * 0.5 = 0.2$

4.) Is $F^+ \perp\!\!\!\perp M^+$?

A: $Pr(M^+) = Pr(M^+ \mid F^+) \, Pr(F^+) + Pr(M^+ \mid F^-) \, Pr(F^-)$
$= 0.6 \cdot 0.5 + 0.2 \cdot 0.5 = 0.4$

No

$Pr(M^-) \cdot Pr(F^+) = 0.6 \cdot 0.5 = 0.3$

Not equal

5.) Find $Pr(O^+)$ $\qquad$ $Pr(M^+|F^+)\,Pr(F^+)$ etc...

$$\begin{aligned} Pr(O^+) &= Pr(O^+ \mid M^+ F^+)\,Pr(M^+ F^+) \\ &+ Pr(O^+ \mid M^+, F^-)\,Pr(M^+, F^-) \\ &+ Pr(O^+ \mid M^-, F^+)\,Pr(M^-, F^+) \\ &+ Pr(O^+ \mid M^-, F^-)\,Pr(M^-, F^-) \end{aligned}$$

$$\begin{aligned} &= 0.15 \cdot 0.6 \cdot 0.5 \\ &+ 0.13 \cdot 0.2 \cdot 0.5 \\ &+ 0.05 \cdot 0.4 \cdot 0.5 \\ &+ 0.04 \cdot 0.8 \cdot 0.5 = 0.084 \end{aligned}$$

6.) $Pr(F^+ \mid O^+) = \dfrac{Pr(F^+, O^+)}{Pr(O^+)}$

$$= \frac{Pr(O^+ \mid F^+, M^+)\,Pr(F^+, M^+) + Pr(O^+ \mid F^+, M^-)\,Pr(F^+, M^-)}{Pr(O^+)}$$

$$= \frac{0.15 \cdot 0.6 \cdot 0.5 + 0.05 \cdot 0.4 \cdot 0.5}{0.084}$$

$$= 0.65$$

7.) $Pr(M^+ \mid O^+) = \dfrac{Pr(O^+, M^+)}{Pr(O^+)}$

$$= \frac{Pr(O^+ \mid M^+, F^+)\,Pr(M^+, F^+) + Pr(O^+ \mid M^+, F^-)\,Pr(M^+, F^-)}{Pr(O^+)}$$

etc...