Math 1112

Unit 8: Quantification of uncertainty in health

8.1 What is uncertainty?

1. Uncertainty is not an object
2. Phenomena randomness and variability
3. Probability as a quantification of uncertainty
4. Describe scenarios where uncertainty is present
5. What do probabilistic statements mean?

8.2 From games to ‘real life’

6. Link between probabilities in games of chance to risk in human health
7. Exercise 1: coin toss as simulation of disease occurrence
8. Calculate probabilities in non-games contexts

8.3 Contingency tables as thinking tool in health research context

9. Exercises with risk, probability and rates in contingency tables
10. Introduction to Relative Risk
11. continued
12. Reverse engineering contingency tables

8.4 Games of chance as simulations of health care context: random events

13. Tree diagrams, screening, and diagnosis.
14. continued
15. Random distribution of disease

8.5 Conditional probabilities with contingency table as thinking tool

16. Conditional probability with contingency tables

Uncertainty is not an object

Those of us who try to make sense of the world, or strive to help ourselves/others get healthier, don't like uncertainty. Uncertainty is another way of saying, I don't know, and in the world we live in, that is not acceptable. The Covid-19 pandemic has presented us with a wave of uncertainties, and has reminded us scientific knowledge is complex, and that a desire for a simple solution (and certainty) can be dangerous to us and others.

Humans, at times, mistake the confidence of the messenger, or simplicity of result for certainty and truth. We often accept simple, or even simplistic explanations when faced with complex phenomena with unpredictable outcomes.

Uncertainty is not a concrete object, like a cup of tea or a frog, **nor is it an abstract object**, like the number 17, money, legal contracts, or property ownership. Uncertainty is a phenomenon, like love, that can best be conceptualized as a feeling (of angst?) or a state of mind, that stems from an awareness of one's lack of knowledge, lack of control, and perhaps a loss of identity (who am I?) especially at times when decisions or actions are required. A rational mathematico-scientific approach to uncertainty tries to set out useful ways to think about the above and avoid the despair that causes us to become immobilized.

There have been and still are many ways that humans deal with uncertainty: some of us try to *influence or sway* external forces (gods and fates) that control events (like weather) where outcomes are uncertain, while others study the phenomenon itself in order to better *predict* those outcomes. Mathematicians, and other thinkers from the scientific worldview have taken another approach - they look for ways to make sense of uncertainty as a phenomenon of nature by quantifying it.

Uncertainty needs to be about something in the world. If it is about a moral, or ethical dilemna, or other phenomena that are not quantifiable (should I where my fancy shirt today?), then probability won't help, if it is about predicting measurable or countable phenomena (especially when making inferences about the future or about large populations) then probability can be helpful.

Through science and mathematics we have learned a lot about the workings of our bodies, and we can often control our health, or at least symptoms of ill health with medicines, surgeries and other interventions. Nevertheless, uncertainty remains a force in many aspects of public and personal health contexts. A few examples:

- Public: Hospitals as institutions must consider uncertainty about the economy, demographics, disease occurrence (pandemics), politics, law, governance, technological changes and others in planning for the future.
- Personal: Uncertainty about effects of treatments. Individuals must consider the inability of the medical system to be 100% certain of outcomes (including side effects) of many treatments of disease, and in recommendations for prevention of disease.

Types of uncertainty: Though there is no consensus, it seems that there is general agreement that the types of uncertainty in health care come from 3 sources:

- probability as a quantity (i.e. 20% risk of XXX);
- ambiguity in statements of probability (i.e. between 10 and 30% risk of XXX)
- complexity (i.e. probabilities associated with climate, or our autoimmune systems that involve feedback loops and other seemingly chaotic phenomena)

Phenomena, randomness, and tools to think about them: probability and risk

<u>Phenomena:</u> are occurrences, or *patterns of events* that we observe around us… or things that happen (flowers bloom, people fall in love ...). Some of these phenomena are predictable and well understood - the sun will rise about 7 in the morning in Toronto in early fall; there will be borscht in my mother's fridge when I visit her - while other phenomena – like my 17 year old's mood tomorrow morning, or will this medication help with my chronic pain – are ones that we have more uncertainty about.

<u>Random variation:</u> Is random variation a mysterious force?... well… sort of… it mostly just means that we can't predict (nor control) which of the possible outcomes of a game, or consequences of exposure to treatment, or moods of a teenager etc will happen. When considering a random phenomenon, we can study patterns of outcomes from controlled events to help predict what will happen in the future, or for other individuals. The more random variation there is in our experience of a phenomenon, the more uncertain we are about its future characteristics.

<u>Random does not just mean unexpected, or surprising it means strictly due to chance</u> *(Random or Sort of Random?)* Results of sports games and your attendance at class can be thought of as random events, but it is hard to argue that the outcome (win, loss or tie; or attendance yes vs no) is purely due to chance. Mathematicians (and bettors) have modeled almost every event in which we don't know the outcome ahead of time as if they are random, but true randomness is an elusive concept. Uncertainty and randomness are…. sort of… the same thing.

<u>Slow thinking as a thinking tool:</u> thinking about phenomena, and possible outcomes of events can be automatic (and fast) when we know the context well, but often will require some slow and disciplined thinking. Using 'sets' as a thinking tool will help you slow down your thinking.

<u>Number and Algebra as thinking tools:</u> All of the thinking you did about numbers and algebra will help you in this section, though you will not be doing a lot of complicated calculations. Once you understand the set of possible and desired outcomes, you will have to decide which thinking tool will best answer the question posed. Crucial to this thinking is a strong intuitive understanding of rates and ratios. Contingency tables (when looking at 2 events simultaneously), tree diagrams and Venn diagrams will be useful thinking tools as well.

Probability as a measure (quantification) of uncertainty

A cup is a physical tool with which we can measure the volume of flour you need for baking bread.

A probability (small p) is a thinking tool a measure of uncertainty in context of a specific phenomenon (should I bet on red? Should I take this medication?) by quantifying it on a scale between 0 and 1 and thus making our decision clearer. (note: both 0 and 1 represent certainty: 0 means that the event in question definitely will not happen, and 1 means that the event under consideration will happen for sure.).

Probability (capital P) is a series of mathematical and logical thinking tools that we use to produce small p probabilities. Probability theory is a subset of mathematics related to statistics and used to help us make sense of, not eliminate, uncertainty.

Historical background: Probability (capital P) or *probabilis* (Latin for plausible, or probable), has been used for over 2000 years. Games of chance, which have served as a model to help study uncertainty, have been around in all human cultures for longer. It was not until the 1600s that a systematic (mathematical) approach to quantifying phenomena where uncertainty is present was developed through the creation of a new set of mathematical thinking tools called Probability (capital P).

Studying games of chance provided a good model for mathematicians studying uncertainty as they were able to assess whether the mathematics provided good-enough predictions before delving into more complex human contexts.

The goals were initially selfish for some mathematicians and scientist gamblers, who in the 1600s reverse engineered games of chance to gain advantage over, and win money from their rivals.

Expressing uncertainty about our health: In the doctor/patient relation uncertainty about outcomes after treatments, or of disease are expressed as rates and ratios but can be called likelihood, probability, chance or course risk. These will be used interchangeably in this course. You will look at 2 ways risk is communicated in health care

- ***Direct expression of risk:*** A person with your test results, age and family background has a 50% chance of getting disease X in the next 5 – 10 years.
- ***Comparing risk:*** A person with 'high' blood pressure is 1.78 times more likely to get disease X in the next 5 – 10 years. (what is unstated here is that there is a comparator group – those with 'normal' blood pressure).

Each of these comparisons sound simple, but to be valid, they must come from a series of research studies often using complex mathematical models in their analysis. Throughout this course and the statistics courses to come, you will be gaining insight into how these assessments are developed. For the rest of this unit you will practice with deciphering probabilistic statements in games of chance, health care, and other more concrete contexts.

Describe scenarios where uncertainty is present

<u>Exercise 1:</u> For each scenario described below state which of the following description of likelihood makes most sense. Justify your choice for each.

i. impossible ii. possible, but unlikely iii. probable iv. certain

a. You flip a coin, and it turns into a bird and flies away.

b. You answer a True/False test without even thinking about the questions and pass.

c. You can see dark clouds gathering in the distance and hear the sound of distant thunder – rain is coming.

d. You will drop your cell phone into a toilet in the next 10 years.

<u>Exercise 2:</u> For each scenario below describe the uncertainty, list (or describe) all possible outcomes, and finally think about whether the scenario involves a truly random process, sort of random (i.e. it can be modeled as if it was random variability) or not random at all (i.e. cannot be thought of by using probability/risk/likelihood or chance).

a. The weight of cheese (kgs) in 30 of your friends' fridges.

b. Finding one person in a classroom of 30 individuals with the same birthday as you.

c. Types of pets owned by members of your extended family.

What do probabilistic statements mean? Restate (or sketch) the following scenarios to make the probabilistic statements in a way that can be understandable to a 9-year-old (hint: focus on possible outcomes).

1. The probability of winning a lottery is advertised as 0.07.

2. The probability of a person getting hit by lightning in their lifetime is 1 in 3500.

3. The probability of rain tomorrow in Toronto is 30%

4. A 9 year-old child you know and care about comes to you and expresses fear about getting cancer because she heard that a child at school got 'disease X' and may possibly die from it. You know that 5 out of 100,000 children will get 'disease X' in Canada. Draw her a sketch in the space below to show how unlikely she is to get the disease.

5. The probability of getting heads on one flip of a coin is ½. If you flip a coin 5 times and get 5 heads in a row, the probability of getting heads on the 6^{th} flip is still ½ .

Link between probabilities in games of chance to risk in human health

- Games of chance provide us with situations that our minds can grasp and easily develop expectations about. Then we can run many trials of the games (flip coins or roll dice) and verify that the mathematical models work correctly. Other phenomena, (e.g. genetics, or buying 2 baby geckos) mimic these games closely, but others (e.g. probabilities of side effects due to medications) look very different.

- Probability (capital P) in health care (biostatistics) starts with research focused on effects of various exposures on probability (small p) of disease.

- In games of chance we study events, in health care we study events and characteristics of individuals coded as variables.

- Probabilities (small p) in health care are calculated from data collected by researchers. The data is organized into exposures and outcomes, or… independent and dependent variables.

- The independence of events in games is established a priori, while in health, the degree of dependence of variables is one of the primary questions we are trying to answer.

- *In games of chance* we ask questions like: what is the probability of rolling a 'sum of 8' when rolling 2 dice. We do not ask if the first die gets different results from the second die.
- *In health care research* we use the probability of outcomes of disease (or risk of disease) to compare people who are exposed (e.g. smoking) to those not exposed.

Exercise: Coin toss as simulation of random distribution of disease:

Use coins to simulate random distributions of disease in humans

Step 1: Take 2 coins that are different (e.g. coin1= nickel and coin2=quarter)

Variable 1: (smokers vs non-smokers) will be represented by 'type of coin'
coin1 = smokers, coin2 = non-smokers.
Variable 2: (disease yes vs no) will be represented by the event 'result of coin flip':
Heads means disease = yes; Tails means disease = no.

Step 2: Each coin flip will represent one person's diagnosis.

Flip coin 1(smokers) 20 times: record #of heads. ___ out of 20 smokers have disease

Flip coin2 (non-smokers) 20 times: record #of heads. ___ out of 20 non-smokers have disease

Step 3: use the results from Step 2 to fill in the table and answer the following.

		Disease status		Total
		Yes (H)	No (T)	
Smoker status	Smoker (coin1)			
	Non-smoker (coin2)			
Total				

P(smoker has disease) = P(non-smoker has disease) =

Q1: If the disease is distributed between smokers and non-smokers randomly by a coin toss, then we would expect the following:

P(smoker has disease)= P(non-smoker has disease) =

Q2: Did your results come up 50:50? If not, what is going on?

Q3: Describe whether you think coin flipping is a good model for thinking about rates of disease in smokers vs non-smokers? If not, what game might be a better model?

Calculate probability in contexts similar to games of chance.

Now that you have mastered probabilities in games of chance, play around with some scenarios in which you will use your knowledge of games of chance to calculate probabilities in other contexts.

Exercise 1: You are purchasing two baby geckos and really want to get a male and female, but it is nearly impossible to know their sex until they grow into adulthood. If you get two males then 'you lose' as they will fight and you will have to separate them into different cages. List the set of all possible outcomes for the sex of the two geckos.

P(2males) =

Which game of chance is concern about the sex of a purchased gecko similar to?

Exercise 2: Imagine you are randomly choosing student from a classroom with 13 males and 7 females.

P(female) =

Which game of chance is choosing a student similar to?

Exercise 3: You will be getting a mark based on choosing a test randomly from a set of test papers with 7A's, 5B's, 4C's, 5D's, and 3F's

P(grade $\geq$ B) =

Which game of chance is this form of assigning grades similar to?

Exercise 4: A store has 180 halogen lightbulbs and 10 are defective. You buy one of them.

P(bulb is defective) =

Which game of chance is buying lightbulbs similar to?

Exercise 5: There is a new family moving into your apartment. You heard that they have two children, and that one of them is a girl. (assume that the probability of boy and girl is equal and that those are the only options).

P(both are girls) =

What happened here… this scenario is a bit different?

Contingency Tables and the Health Research Context: probabilities from data

***<u>Exercise 1</u>*:** 670 randomly sampled Canadians responded to following questions: 'Do you have diabetes?' 'Have you ever smoked daily?'. The data was tabulated into a contingency table and the results are below. Data from the Canadian Community Health Survey

Ever smoked daily * Has diabetes Crosstabulation

Count

		Has diabetes		
		YES	NO	Total
Ever smoked daily	YES	59	532	591
	NO	4	75	79
Total		63	607	670

a) Find P(respondent had diabetes)

b) Find P(respondent had 'ever smoked daily')

c) Find P(respondent had diabetes given that he/she 'ever smoked daily')

d) Find P(respondent had diabetes given that he/she did not 'ever smoke daily)

e) Are respondents who 'ever smoked daily' more likely to have diabetes?

f) How many times more likely are those who ever smoked to have diabetes?

Health Research Context: introduction to Relative risk

Contingency tables in Health Sciences are often used to compare rates (of disease in two types of communities for example). One way of doing this easily is by calculating a ratio of probabilities. This measure is most often called Relative Risk (RR), or likelihood ratio.

Exercise 2: 1998 Canadians were asked if their province of residence and whether they consulted an alternative health care provider (Naturopath, Homeopath, etc.) The results are in the chart below

Consulted altern. health care provider * que.vs.rest Crosstabulation

Count

				Total
		Quebec	Rest of Canada	
Consulted altern. health care provider	YES	51	186	237
	NO	367	1394	1761
Total		418	1580	1998

a) Find P(resident of Quebec consulted an alternative health care provider)

b) Find P(resident of Rest of Canada consulted an alternative health care provider)

c) Are Quebecers different from the rest of Canadians in this regard?

d) Quebecers are _____ times more likely to consult an alternative health care provider than other Canadians.

e) Make a statement of comparing rates starting with ‘residents of the Rest of Canada are

Health Research Context: introduction to Relative risk continued

<u>Exercise 3:</u> 727 randomly sampled Canadians were asked how much alcohol they consumed, and whether they used a seatbelt when they were in a car. The data was tabulated into a contingency table and the results are below. Answer the questions below the table by extracting the information from the table.

		high alcohol consumption		Total
		YES	NO	
Using the seatbelt	YES	30	580	610
	NO	67	50	117
Total		97	630	727

How many times more likely are Canadians who don't wear their seatbelt to have high alcohol consumption?

<u>Exercise 4:</u> 999 urban vs rural Canadians were asked whether they were happy. It was thought that rural Canadians would have a higher rate of responding yes.

		happy		Total
		YES	NO	
residence	rural	356	221	577
	urban	301	121	422
Total		657	342	999

Who has a higher probability of being happy, rural Canadians or urban Canadians? How many times more likely to be happy are they?

Reverse Engineering contingency tables

<u>Exercises:</u>

1. a. A study was conducted on 500 adults. One group of individuals who smoke had a 7% risk of CVD and they were twice as likely to get CVD as non-smokers. What was the chance of getting CVD in non-smokers?

   b. additional information: there are 300 smokers and 200 non-smokers in the study. Fill in the contingency table below using all given information.

				Total
Total				

2. a. Adults who excessively ate candy as children were just as likely to end up convicted of violent crimes one study showed. If the probability of ending up convicted for violent crime in excessive candy eaters was 0.003, what is the rate of conviction for violent crime in those adults who did not each candy excessively as children?

   b. additional information: there are 350 excessive candy eaters and 3000 non-excessive candy-eaters in the study. Fill in the contingency table below using all given information.

				Total
Total				

3. Create a contingency table for the following statement. Males are 7 times more likely to crack their knuckles than females. {more than one possible answer}

				Total
Total				

Probability Tree and contingency tables as thinking tools (effectiveness of screening)

<u>Exercise 1A</u>: set up a tree diagram for the following scenario using given probabilities. Screening tests aim to catch diseases before they can be diagnosed. A good test is one in which every person who screens 'positive' for diseaseX, also ends up with a positive diagnosis of the disease.

A screening test has been designed for disease X and the following information is known about it.

P(screening is positive) = 3/4; P(diagnosis is positive) = 2/3

a) Sketch a probability tree to represent the scenario in the space below assuming that screening and diagnosis are independent events 😊 (hint: similar to marbles in bag game).

SS =

Calculate the following:

P(both positive) P(diagnosis is positive given that screening was positive)

b) Use the results of the tree diagram to fill in the contingency blank below. Note that the total individuals in data set is 120.

		Diagnosisstatus		Total
		positive	negative	
Screening status	positive			
	negative			
Total				120

P(both tests are positive) = P(diagnosis is positive given that screening was positive) =

P(diagnosis is positive given that screening was negative) = RR=

c) In this setup screening is presented as independent of diagnosis. How can we recognize this in the tree diagram and/or contingency table?

d) What is the difference between P(both are positive) and P(diagnosis + given that screen +)

From Probability Tree to contingency table in screening (continued)

***Exercise 1B*:**

a) The given probabilities stay the same: P(screening is positive) = ¾ and P(diagnosis is positive) = 2/3; but now we also are told that 70 of those who screened positive were diagnosed as having the disease.
Fill in the contingency table below

		Diagnosis status		Total
		positive	negative	
Screening status	positive			
	negative			
Total				120

P(both tests are positive) =

P(diagnosis is positive given that screening was positive) =

P(diagnosis is positive given that screening was negative) =

Those that were screened positive are ____ times more likely to get diagnosed positive

b) The above is a non-random distribution – but is it one that we would expect in research on effectiveness of screening tests? (recall that one characteristic of an excellent screening test is one in which the rate of positive diagnosis in those who screen positive will be 100%)

c) Discuss how the idea of randomness (think of the behaviour choosing a marble from a 2 bags) can help understand what is going on between screening and diagnosis of disease.

Random Distribution of disease:

***Exercise 1*: ** Let's look at data from the Canadian Community Health Survey again. 670 randomly sampled Canadians responded to following questions: 'Do you have diabetes?' 'Have you ever smoked daily?'.

Ever smoked daily * Has diabetes Crosstabulation

Count

		Has diabetes		Total
		YES	NO	
Ever smoked daily	YES	59	532	591
	NO	4	75	79
Total		63	607	670

Recall that those who ever smoked daily were 2 times more likely to have diabetes.

a) Fill in a contingency table to demonstrate what you would expect a random distribution of diabetes to look like. (hint: use choosing marble from a bag as a model) We call these the expected results!

		Diabetes		Total
		YES	NO	
Ever smoked	yes			591
	no			79
Total		63	607	670

b) Compare the two contingency tables (and the relative risks of diabetes in each). Is disease distributed sort of randomly in the actual data in the first contingency table?

Contingency tables and conditional probability

Use contingency tables as as thinking tool to resolve the challenges given below (try answering them firs without the tables):

1. In a group of 100 HIM students we know that 60 previously went to university, 70 passed math 1112. We also know that only 20 of those who did not go to university passed math 1112. What is the probability of a member of this student group passing given that we know they previously went to university?

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2"></td><td>Total</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td rowspan="2"></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td></td><td></td><td></td><td></td></tr>
</table>

2. Breast cancer screening has been conducted using mammography in a certain region. You know the following information about the women in this region:
   - The probability that a woman has breast cancer is 1% (prevalence)
   - If a woman has breast cancer, the probability that she screens positive is 90%
   - If a woman does not have breast cancer, the probability that she nevertheless screens positive is 9%

   A woman screens positive. She wants to know from you whether that means that she has breast cancer for sure, or what the chances are. How do you respond?

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2"></td><td>Total</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td rowspan="2"></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td></td><td></td><td></td><td></td></tr>
</table>