IB Mathematics Analysis and Approaches HL

Topic 4 Probability

Jiuru Lyu

April 12, 2022

Contents

1 Statistics and Probability **2**

1.1 An Introduction to Statistics . . . 2
1.2 Linear Correlation of Bivariate Data . . . 5
1.3 Probability and Expected Outcomes . . . 6
1.4 Probability Calculations . . . 7
1.5 Discrete Random Variables . . . 8
1.6 The Binomial Distribution . . . 9
1.7 The Normal Distribution and Curve . . . 10
1.8 Probability Density Function (PDF) . . . 12
1.9 The Bayes' Theorem . . . 13

1 Statistics and Probability

1.1 An Introduction to Statistics

1. Statistical inferences: when the sample data is well chosen and described, an analysis will allow us to draw conclusions about the population based on this sample.

    - **Discrete data** is data that can be counted; it gives the number of times something occurs or the number of items that exists.
    - **Continuous data** is data that is measured; however, the values of the actual data cannot be determined exactly, and the data may be limited to a range.

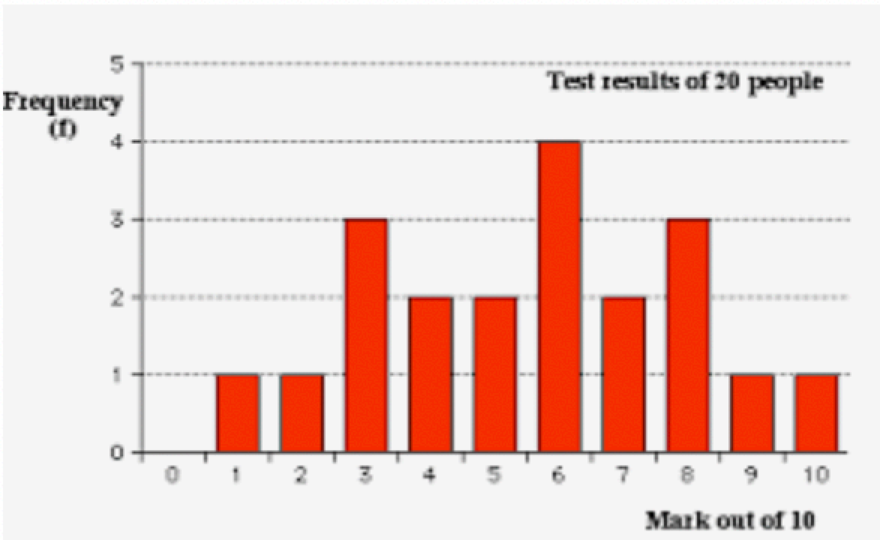
2. Reliability and Bias.

3. Data Sampling (Sampling Methods):

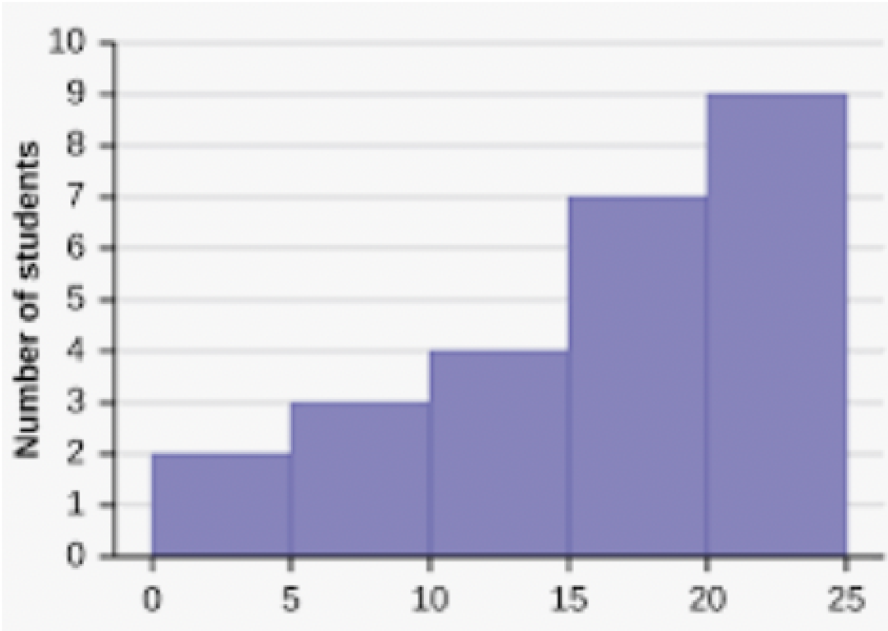
    - **Simple**: Achieving randomness by a simple, completely random process.
    - **Convenience**: Choosing a sample based on how easy it is to find the data.
    - **Systematic**: If data is listed, selecting a random starting point and then choosing the rest of the sample at a consistent interval in the list.
    - **Quota**: Choosing a sample that is only comprised of members of the population that fit certain characteristics.
    - **Satisfied**: Choosing a random sample in a way that the population of certain characteristics matches the proportion of those characteristics in the population.

4. Grouped Data:

    - Key terms: intervals/classes, frequency distribution table, frequency diagram.
    - Frequency diagram of discrete data:



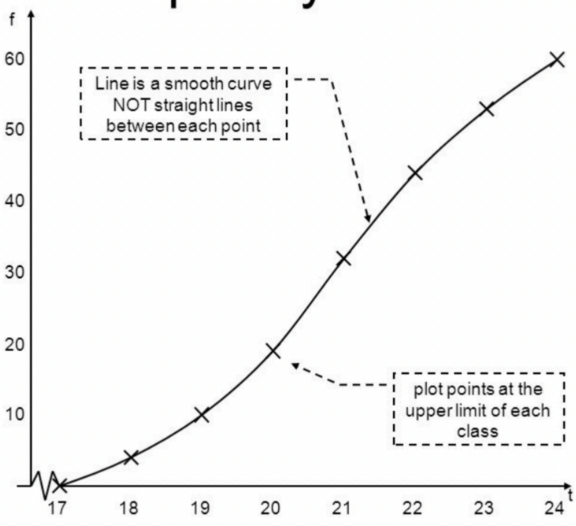
    - Frequency diagram of continuous data - **Histogram**



- Cumulative frequency table and cumulative frequency graph:

Cumulative Frequency Curve

Time (minutes)	Frequency f
17 ≤ t < 18	4
17 ≤ t < 19	11
17 ≤ t < 20	19
17 ≤ t < 21	32
17 ≤ t < 22	44
17 ≤ t < 23	53
17 ≤ t < 24	60

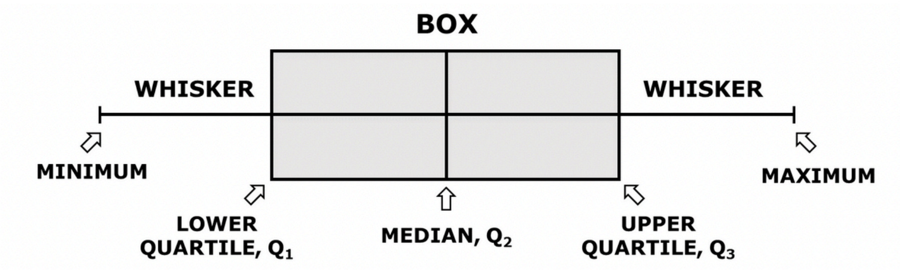


- For data grouped into intervals or classes, we can identify:
  (a) Mid-interval values
  (b) Interval width
  (c) Lower interval boundaries
  (d) Higher interval boundaries
  (e) Modal class: the class with the highest frequency

5. Quartiles:

- Minimum: the lowest value
- Q_1: the 25th percentile
- Median (Q_2): the 50th percentile

- Q_3: the 75th percentile
- Maximum: the highest value
- $$\text{Interquartile Range (IQR)} = Q_3 - Q_1.$$
- Box-and-whisker plots: $\rightarrow$ spread of data:



6. Normal distribution, negatively skewed, and positively skewed:
7. Normal distribution: mean = mode = median
8. Positively skewed: median < mean
9. Negatively skewed: median > mean
10. Outliers:
$$\begin{aligned} &\text{Ourlier} < Q_1 - 1.5\text{IQR} \\ \text{OR} \quad &\text{Outlier} > Q_3 + 1.5\text{IQR}. \end{aligned}$$
11. Measuring central tendency:
    - **Definition 1.1.1** **Mode**: the value with the greatest frequency.
    - **Definition 1.1.2** **Median**(m): the middle value when the data is arranged in order.
    - **Definition 1.1.3** **Mean**(μ): the arithmetic mean, the sum of the numerical data divided by the number of data points.
$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}.$$
12. Measures of dispersion:
    - **Definition 1.1.4** Range:
$$\text{Range} = x_{\max} - x_{\min}.$$
    - **Definition 1.1.5** Interquartile Range (IQR):
$$\text{IQR} = Q_3 - Q_1.$$
    - **Definition 1.1.6** Variance σ^2 and standard deviation σ:
$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^{k} f_i (x_i - \mu)^2}{n} = \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \mu^2 \\ \sigma &= \sqrt{\sigma^2} \end{aligned}$$

Proof 1.1.1

$$\begin{aligned}
\sigma^2 = \frac{\sum_{i=1}^{k} f_i (x_i-\mu)^2}{n} &= \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - 2\mu\frac{\sum_{i=1}^{k} f_i x_i}{n} + \mu^2\frac{\sum_{i=1}^{k} f_i}{n} \\
&= \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - 2\mu^2 + \mu^2 \\
&= \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \mu^2
\end{aligned}$$

1.2 Linear Correlation of Bivariate Data

1. Scatter Plot
2. Linear correlation:
   - **Positive linear correlation**: the line as a general upward trend.
   - Negative correlation.
   - Quadratic trend
   - Exponential trend
3. Estimate a line of best fit:

   Theorem 1.2.1 The line must pass through the mean point of the data set $(\bar{x}, \bar{y})$, where $\bar{x}$ is the mean of x and $\bar{y}$ is the mean of y.

4. Pearson's correlation coefficient (r):
   - Range: $-1 \leq r \leq 1$
     (a) $r = -1$: perfect negative linear correlation
     (b) $r = 1$: perfect positive linear correlation
     (c) $r = 0$: no linear correlation
     (d) $0.7 < |r| \leq 1$: strong linear correlation
     (e) $0.3 < |r| \leq 0.7$: weak to moderate linear correlation
   - Formula:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

$$r = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sqrt{\sum x^2 - n\bar{x}^2}\sqrt{\sum y^2 - b\bar{y}^2}}$$

5. Prediction using the least square linear regression:
   - The dangers of extrapolation.
   - We cannot always reliably make a prediction of x from a value of y, using a y on x line.

1.3 Probability and Expected Outcomes

1. **Definition 1.3.1** **Random Events**: although we don't know the exact result in advance, we do know the set of all possible results.

   - The set of all possible results is called the **sample space** (U).
   - The number of all possible outcomes that make up the sample space is denoted as $n(U)$.
   - Any subset of a sample space is called an **event**. The event consists of one or more outcomes.
   - Each time an experiment is repeated, it is considered a **trial** of the experiment.

2. **Experimental probability** (relative frequency) is found by repeating an experiment a number of times and counting the number of times that particular outcome occurs.

3. **Definition 1.3.2** **Mutually exclusive**: two events that do not share any outcomes $\rightarrow$ They cannot occur together.

4. **Definition 1.3.3** **Theoretical probability**: When running an experiment for N trials that results in an event A occurring $n(A)$ times, the probability of A happening, $\mathrm{P}(A)$ is

   $$\mathrm{P}(A) = \lim_{N \to \infty} \frac{n(A)}{N}.$$

   If an experiment has equally likely outcomes, the probability of event A occurring is defined as

   $$\mathrm{P}(A) = \frac{n(A)}{n(U)} = \frac{\text{Number of outcomes in which A occurs}}{\text{Total number of outcomes in the sample space}}.$$

5. Axioms:

   Axiom 1.3.1 For any event A,

   $$0 \leq \mathrm{P}(A) \leq 1.$$

   Axiom 1.3.2 The probability of nothing (o) occurring is zero:

   $$\mathrm{P}(o) = 0.$$

   The probability of one of all the outcomes in the sample space occurring is one:

   $$\mathrm{P}(U) = 1.$$

   Axiom 1.3.3 If A and B are $\in U$ and are mutually exclusive, then the probability of either A or B happening is

   $$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B).$$

   Axiom 1.3.4 If $\mathrm{P}(A')$ is the probability of event A not happening,

   $$\mathrm{P}(A') = 1 - \mathrm{P}(A),$$

   where $\mathrm{P}(A)$ is the probability of event A occurring.

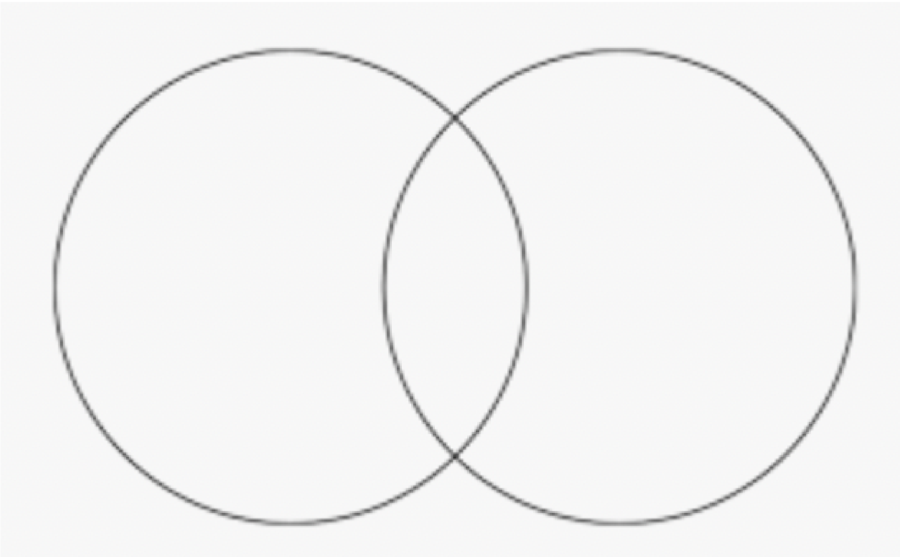
6. **Definition 1.3.4** Expectation: The formula for expected number of members of group A in a sample of size n is

$$n\mathrm{P}(A).$$

If choosing n from the population, $n\mathrm{P}(A)$ gives an approximation of how many of them would be members of group A.

1.4 Probability Calculations

1. **Definition 1.4.1** A **Venn diagram** is a model illustrating two or more sets of data using overlapping circles to show elements of each set.



- **Definition 1.4.2** The probability of event A and B occurring at the same time:

$$\mathrm{P}(A \cap B)$$

- **Definition 1.4.3** The probability of event A or B occurring (A union B):

$$\mathrm{P}(A \cup B)$$

- **Theorem 1.4.1** If $\mathrm{P}(A \cup B) = \varnothing$, then A and B are **mutually exclusive**.

2. The tree diagram

3. **Theorem 1.4.2** The probability of a union:

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B).$$

Proof 1.4.1

$$\begin{aligned} n(A \cup B) &= n(A) + n(B) - n(A \cap B), \\ \mathrm{P}(A \cup B) = \frac{n(A \cup B)}{n(U)} &= \frac{n(A) + n(B) - n(A \cap B)}{n(U)} \\ &= \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B). \end{aligned}$$

4. Conditional probability:

   - Conditional probability shrinks the sample space and, therefore, increases the probability of an event occurring, unless the given information renders the event impossible.
   - **Theorem 1.4.3**
   $$\mathrm{P}(A|B) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)},$$
   where $\mathrm{P}(A|B)$ is read as probability of event A occurring, given event B occurring.

   Proof 1.4.2 If we know that B has occurred, the sample space now contains all the elements of B but no more. Now, we select events from the sample space that falls in A:
   $$\begin{aligned}\mathrm{P}(A|B) &= \frac{n(A\cap B)}{n(B)} \\ &= \frac{\frac{n(A\cap B)}{n(U)}}{\frac{n(B)}{n(U)}} = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}.\end{aligned}$$

5. The probability of independent events:

   - **Definition 1.4.4** A and B are **independent events** if the occurrence of A has no effect on the probability of B occurring.
   $$\mathrm{P}(B|A) = \mathrm{P}(B) \quad \text{OR} \quad \mathrm{P}(A|B) = \mathrm{P}(A).$$
   - **Theorem 1.4.4** A and B are independent if and only if
   $$\mathrm{P}(A\cap B) = \mathrm{P}(A)\times\mathrm{P}(B).$$

6. **Theorem 1.4.5** Probability of union when A and $B$are mutually exclusive:
   $$\mathrm{P}(A\cup B) = \mathrm{P}(A)+\mathrm{P}(B).$$

1.5 Discrete Random Variables

1. **Definition 1.5.1** If a random variable can take exactly N values, each of which corresponds to a unique outcome in the sample space of an experiment, then this variable is a **discrete random variable**.

   - The probability that x takes on any of the N values is written as $\mathrm{P}(X=x)$
   - A **probability distribution** is a combination of the sample space of a random experiment with the probabilities of each of the events in the sample space.
   - PDF: probability distribution function
   - **Axiom 1.5.1** For each value $\{x_i\}$ of the random variable X, we have that
   $$0 \leq \mathrm{P}(X=x_i) \leq 1.$$

- **Axiom 1.5.2** As X may take on any of the N values of the sample space, the sum of the probabilities of each of them as an outcome of the experiment must equal one:

$$\sum_{i=1}^{N} \mathrm{P}(X = x_i) = \mathrm{P}(X = x_1) + \mathrm{P}(X = x_2) + \mathrm{P}(X = x_3) + \cdots + \mathrm{P}(X = x_N) = 1.$$

- **Definition 1.5.2** **Well-defined probability distribution**: a probability distribution that satisfies both

$$\begin{cases} 0 \leq \mathrm{P}(X = x_i) \leq 1 \\ \sum_{i=1}^{N} \mathrm{P}(X = x_i) = 1 \end{cases}$$

2. Calculating expected value:

- **Definition 1.5.3** The **expected value** (expected mean) is the value you would expect to obtain on average if you performed an experiment many times.

- The expected value of a random variable X with a probability distribution function $\mathrm{P}(X = x)$ is written as $\mathrm{E}(X)$ or μ.

- Formula:

$$\begin{aligned} \mathrm{E}(X) = \mu = \sum_{i=1}^{N} x_i \mathrm{P}(X = x_i) \\ = x_1 \mathrm{P}(X = x_1) + x_2 \mathrm{P}(X = x_2) + x_3 \mathrm{P}(X = x_3) + \cdots + x_N \mathrm{P}(X = x_N) \end{aligned}$$

- Application: fairness of a game
A fair game is perfectly balanced so that the expected mean pay-out is 0 for both players.

1.6 The Binomial Distribution

1. The binomial distribution discrete the probability distribution of different outcomes of repeated binary events.

- Since there are only two outcomes, we use p to represent the probability of success, meaning the probability that the event we are looking for occurs.

- We use $X \sim$ to denote a random variable X that is distributed in a certain way.

- $X \sim \mathrm{B}(n, p)$ represents the binomial distribution,
where *mathrmB* stands for binomial, n represents the number of trials in the binomial experiment, and p is the probability of success.

- The probability of failure is q:

$$q = 1 - p.$$

- **Theorem 1.6.1** Expected value of a binomial distribution:

$$\mathrm{E}(X) = np.$$

- **Theorem 1.6.2** Variance of a binomial distribution:

$$\mathrm{Var}(X) = np(1-p) = npq.$$

2. Probabilities within the binomial distribution:

   Theorem 1.6.3 A random variable that is binomially distributed and takes on the value of the number of success in n trials is written as $X \sim \mathrm{B}(n,p)$, where p is the probability of success in one trial. Then, the probability that X takes on an actual value of x successes is given by:

$$X \sim \mathrm{B}(n,p) \Rightarrow \mathrm{P}(X=x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0,1,2,3,\cdots,n.$$

3. Finding binomial probabilities using GDC:

   - **Binomial PDF** returns a specific value of $\mathrm{P}(X=x)$
   - **Binomial CDP** (cumulative distribution function) gives the value of

$$\mathrm{P}(0 \leq X \leq \text{upper bound}).$$

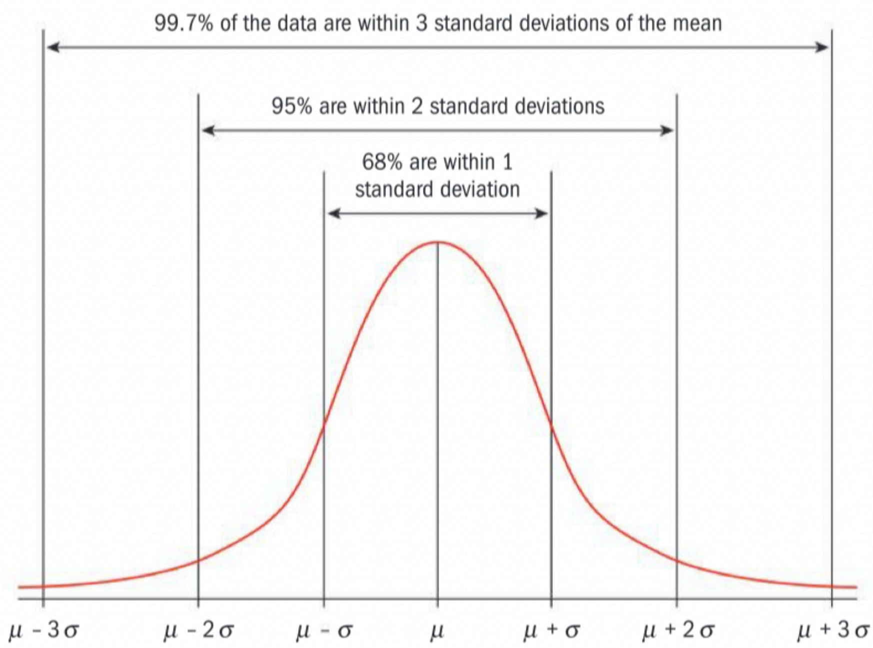
   - **Theorem 1.6.4**

$$\mathrm{P}(a \leq X \leq b) = \mathrm{P}(0 \leq X \leq b) - \mathrm{P}(0 \leq X \leq a-1).$$

1.7 The Normal Distribution and Curve

1. The normal (bell) curve:

   - The normal distribution is modeled graphically with a **normal curve**:



- The normal curve always has its highest point in the center, and that point is the mean (μ), median, and mode.
- The axis is often labeled using the standard deviation σ.
- The normal distribution deals with continuous random variables.

2. Expressing the normal distribution algebraically:

   - Normal distribution is denoted as

   $$X \sim \mathrm{N}(\mu, \sigma^2),$$

   where N means normal distribution, μ is the mean, and σ is the standard deviation.

   - Normal probability density function:
     (a) A probability density function is the equation of a curve that has the probabilities as the area underneath.
     (b) **Definition 1.7.1** For $X \sim \mathrm{N}(\mu, \sigma^2)$,

   $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ where } -\infty < x < \infty.$$

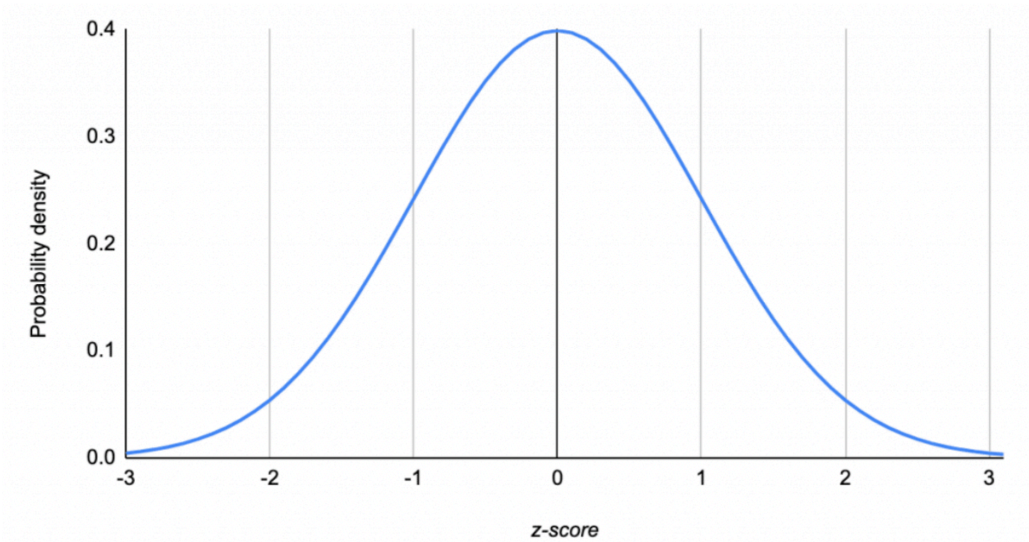
   $$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x = 1.$$

3. Calculating normal distributed probabilities with GDC:

   - Normal PDF and Normal CDF.
   - Inverse Normal Function.

4. Standard Normal Distribution:

   - Transform $X \sim \mathrm{N}(\mu, \sigma^2)$ into $Z \sim \mathrm{N}(0, 1)$:



   - Z-score:

   $$z = \frac{x-\mu}{\sigma}.$$

- Finding probabilities with z-scores:
  (a)
  $$P(-1 < z < 1) \approx 0.68$$
  $$P(-2 < z < 2) \approx 0.95$$
  $$P(-3 < z < 3) \approx 0.997$$
  (b) Normal CDP on GDC
- Using z-score to find μ and σ.

1.8 Probability Density Function (PDF)

1. For a **continuous random variable** X with **probability density function** $f(x)$, it holds

$$\begin{cases} f(x) \geq 0 \quad \text{i.e., the function is non-negative} \\ \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1 \quad \text{i.e., the total area under the curve is 1} \end{cases}.$$

2. The probability that X takes values between a and b is

$$P(a \leq X \leq b) = \int_a^b f(x)\mathrm{d}x$$

N.B.: $P(a \leq X \leq b) = P(a < X < b)$ because $P(X = a) = 0$.

3. The **mean** μ, or the **expected value** $E(X)$, is defined by

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x.$$

4. The **Variance** $\mathrm{Var}(X)$ is defined by

$$\mathrm{Var}(X) = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x.$$

A equivalent and more practical definition is

$$\mathrm{Var}(X) = E(X^2) - \mu^2, \text{ where } E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x.$$

5. Mode: the value of x where $f(x)$ has its maximum.

6. Median: The value of m where $P(X \leq m) = 0.5$

Example 1.8.1

$$\text{Let } f(x) = \begin{cases} f_1(x), a \leq x \leq b \\ f_2(x), b < x \leq c \end{cases}, \text{ we check } \int_a^b f_1(x)\mathrm{d}x = A:$$

(a) If $A > 0.5$, the median $m \in [a,b]$, solve

$$\int_a^m f_1(x)\mathrm{d}x = 0.5.$$

(b) If $A < 0.5$, the median $m \in [b,c]$, solve

$$\int_m^c f_2(x)\mathrm{d}x = 0.5.$$

7. Quartiles:

   - The **lower quartile Q_1** is defined by $\mathrm{P}(X \leq Q_1) = 0.25$, i.e., solve

$$\int_{-\infty}^{Q_1} f(x)\mathrm{d}x = 0.25.$$

   - The **upper quartile Q_3** is defined by $\mathrm{P}(X \leq Q_3) = 0.75$, i.e., solve

$$\int_{-\infty}^{Q_3} f(x)\mathrm{d}x = 0.75.$$

8. The effects of linear transformations on the random variable X:

   -
$$\mathrm{E}(aX+b) = a\mathrm{E}(X)+b.$$

   -
$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X).$$

1.9 The Bayes' Theorem

1. Bayes' Theorem:

   Theorem 1.9.1 Bayes' Theorem

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A)\mathrm{P}(B|A)}{\mathrm{P}(B)}.$$

   Proof 1.9.1

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)} \;\Rightarrow\; \mathrm{P}(A\cap B) = \mathrm{P}(A|B)P(B)$$

$$P(B|A) = \frac{\mathrm{P}(B\cap A)}{\mathrm{P}(A)} \;\Rightarrow\; \mathrm{P}(B\cap A) = \mathrm{P}(B|A)P(A)$$

$$\because \mathrm{P}(A\cap B) = \mathrm{P}(B\cap B),$$

$$\therefore \mathrm{P}(A|B)P(B) = \mathrm{P}(B|A)P(A).$$

$$\therefore \mathrm{P}(A|B) = \frac{\mathrm{P}(A)\mathrm{P}(B|A)}{\mathrm{P}(B)}.$$

2. Other versions of the Bayes' theorem:

- $$\mathrm{P}(B)=\mathrm{P}(A\cap B)+\mathrm{P}(A'\cap B)=\mathrm{P}(A)\mathrm{P}(B|A)+\mathrm{P}(A')\mathrm{P}(B|A').$$

- Substitute $\mathrm{P}(B)$:
$$\mathrm{P}(A|B)=\frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(A)\mathrm{P}(B|A)+\mathrm{P}(A')\mathrm{P}(B|A')};$$
$$\mathrm{P}(B|A)=\frac{\mathrm{P}(A|B)\mathrm{P}(B)}{\mathrm{P}(B)\mathrm{P}(A|B)+\mathrm{P}(B')\mathrm{P}(A|B')}.$$

3. Bayes' theorem for three events:

$$\begin{aligned}\mathrm{P}(B)&=\mathrm{P}(A_1\cap B)+\mathrm{P}(A_2\cap B)+\mathrm{P}(A_3\cap B)\\&=\mathrm{P}(A_1)\mathrm{P}(B|A_1)+\mathrm{P}(A_2)\mathrm{P}(B|A_2)+\mathrm{P}(A_3)\mathrm{P}(B|A_3).\end{aligned}$$

$$\begin{aligned}\therefore \mathrm{P}(A_i|B)&=\frac{\mathrm{P}(A_i)\mathrm{P}(B|A_i)}{\mathrm{P}(B)}\\&=\frac{\mathrm{P}(A_i)\mathrm{P}(B|A_i)}{\mathrm{P}(A_i)\mathrm{P}(B|A_i)+\mathrm{P}(A_2)\mathrm{P}(B|A_2)+\mathrm{P}(A_3)\mathrm{P}(B|A_3)}.\end{aligned}$$