Statistics and probability

Especially in this chapter, it is crucial that you know how to use the calculator. It will make your life so much easier!

1. Concepts of population, sample, random sample and frequency distribution of discrete and continuous data. Grouped data: mid-interval values, interval width, upper and lower interval boundaries. Mean, variance, standard deviation.
- Be able to know these underlying terms in statistics:

Population = total and full collection of target data. Statistics from population is accurate, but time consuming to gather the population

Sample = a part of population. Less accurate but time efficient.

Random sample = every data in population have equal chance of getting selected.

Frequency distribution = a table that shows how much of a certain category there is.

Discrete data = restricted to certain values, eg. Whole numbers only.

Continuous data = may take any real number.

Data can be grouped into intervals/classes.

Interval width = how wide an interval is. $10 \leq \mathrm{x} < 20$, interval width is 10.

Mid-interval value = the value halfway up on interval width. Mid-interval value for $10 \leq \mathrm{x} < 20$ is 15 because it is the value between 10 and 20.

Upper and lower interval boundaries = the boundaries stated. Be careful of $< and \leq$.

For mean, variance and standard deviation you must be able to use the formula. Standard deviation is essentially showing us "on average, how far away from the mean are the individual values?" Variance is just st.dev squared.

2. Concepts of trial, outcome, equally likely outcomes, sample space *(U)* and event. The probability of an event A as $P(A) = \frac{n(A)}{n(U)}$. The complementary events A and A' (not A). Use of Venn diagrams, tree diagrams, counting principles and tables of outcomes to solve problems.
- Be able to use basic probability concepts:

Probability = likelihood of an event to happen.

$$P(A) = probability\ of\ event\ A\ occuring = \frac{number\ of\ event\ A}{number\ of\ total\ events} = \frac{n(A)}{n(U)}$$

$$P'(A) = not\ P(A)$$

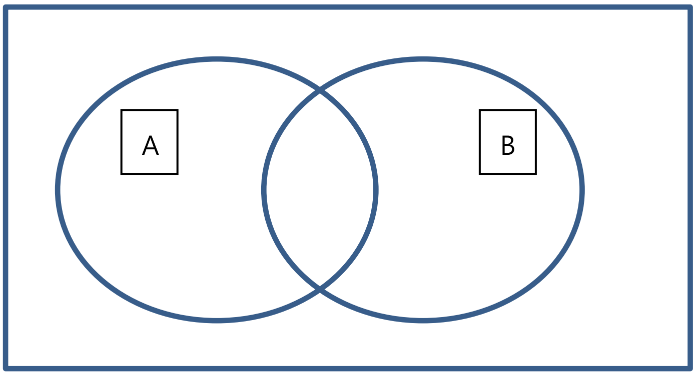
This means that $P'(A) + P(A) = 1$.

Get familiar on how to use Venn diagrams, tree diagrams, means and tables. Practice practice practice.

3. Combined events; the formula for P(A∪B). Mutually exclusive events.

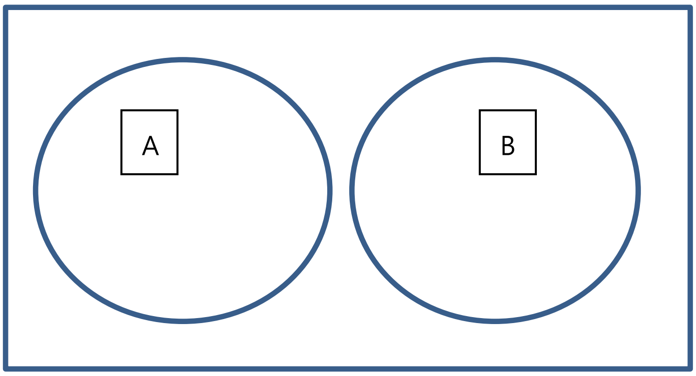
- <u>Be able to understand the difference between combined events and mutually exclusive events:</u>

Combined events overlap when drawn in Venn diagram.



For combined events, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Mutually exclusive events do not overlap.

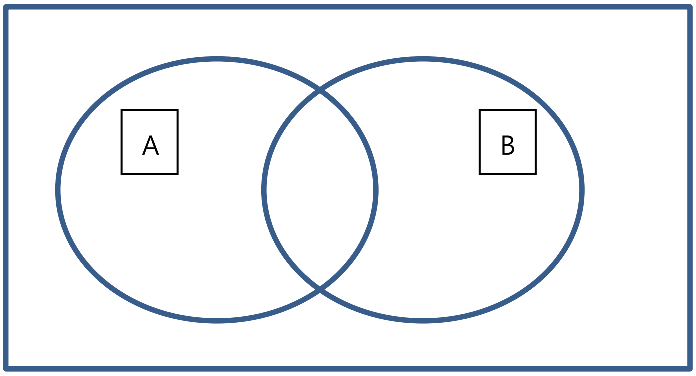


For mutually exclusive events, $P(A \cup B) = P(A) + P(B)$ because $P(A \cap B) = \emptyset$.

4. Conditional probability; the definition. Independent events; the definition. Use of Bayes' theorem for a maximum of three events.

- Be able to understand conditional probability:

Conditional probability is when you are given an additional information so that "when given B, calculate the probability of A". In other words, find the probability of having A out of B.



Mathematically, $P(A|B) = \frac{P(A \cap B)}{P(B)}$. You can see now that the denominator, the total number of events, is now the events for B. So our "total" has become B, and we try to find the events A within B.

Note that $P(A|B) \neq P(B|A)$. The denominator is different!

You should also be able to use Bayes' theorem, but I have never ever seen a question about it. Plus, it is in the formula booklet so it is not so difficult to use. Essentially, it is about combining events (maximum 3) into expressions that you can solve. The key is to identify that $P(A \cap B) = P(B \cap A)$.

Independent events do not change probability when the other event changes. To identify it,

$$P(A \cap B) = P(A) \times P(B)$$

5. Concept of discrete and continuous random variables and their probability distributions. Definition and use of probability density functions. Expected value (mean), mode, median, variance and standard deviation.

- Be able to understand probability distribution and probability density function:

Probability distribution = a table that shows the probabilities of events. The particular function that is used to find the probability of a specific event is our probability density function.

There are two types of probability distributions: discrete and continuous. Let's start with discrete.

Be able to understand discrete probability density function:

The function for discrete = $f(x) = P(X = x)$, where the lowercase x is the target event.

If $f(x) = P(X \leq x)$ then it simply becomes cumulative. This can be rewritten as $\sum_{i=0}^{n} P(X = x_i)$.

Be careful that $f(x) = P(X < x) = P(X \leq x - 1)$.

Be able to find expected value, mode, media, variance and standard deviation in terms of P(X=x), discrete variable:

$E(X) = mean = \sum_{i=1}^{n} x_i P(X = x)$, or $P(X = x)$ can be substituted as $f(x)$. The basic idea here is that the probability is fundamentally $P(X = x) = \frac{frequency}{total}$. This is then the same as the previous mean formula we learnt, which is:

$$mean = \frac{\sum_{i=1}^{k} x_i f_i}{n}$$

$$\frac{f_i}{n} = probability$$

$Mode = most\ frequently\ occuring\ x.$ This has the highest $f(x) = P(X = x)$. You simply have to plug in the numbers to density function see which one has the highest probability.

$Median = the\ \mathrm{x}value\ that\ gives\ cumulative\ value\ of\ exactly\ 0.5.$ This would be most often possible with continuous distribution.

$$Variance = \sigma^2$$

Be able to understand continuous probability density function:

The function for continuous = $f(x) = P(a \leq x \leq b)$. Why? For continuous values, we use integrals because that is the only way to calculate the area under continuous functions (area under = the probability in this case), and in integrals cannot be in one value. Having an integral of one value would mean an infinitesimal thin line, which gives value of 0. So we write continuous probability density function as:

$\int_a^b f(x)dx = F(b) - F(a)$

Thus $f(x) = PDF\ and\ F(x) = CDF, and\ F'(x) = f(x)$

Be able to find expected value, mode, media, variance and standard deviation in terms of P(a≤x≤b), continuous variable:

The idea is exactly the same as discrete, only that we now replace the sigma with integral.

$$E(X) = mean = \int_{i=0}^{i=max} x_i f(x)dx$$

$Mode = most\ frequently\ occuring\ x$. In continuous, we find the mode by simply finding the maximum point. In other words, mode is when $f'(x) = 0\ and\ f''(x) < 0$.

$Median = the$ x $value\ that\ gives\ cumulative\ value\ of\ exactly\ 0.5$. Thus: $\int_{i=0}^{m} f(x)dx = 0.5$ and you find out the "m" which is the median.

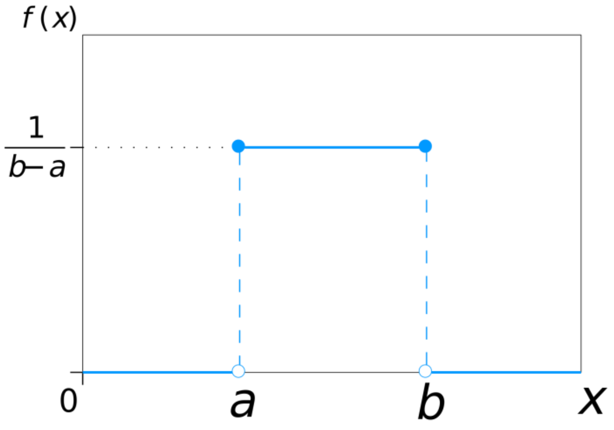
Also, we use the same idea to find first quartile and third quartile. For first quartile we set the density function as 0.25 and third quartile as 0.75.

$Variance = \sigma^2$, and this can expressed as (according to formula booklet)

$$\int_{i=0}^{i=max} (x_i - \mu)^2 f(x)dx = E(X^2) - E(X)^2$$

Be able to understand continuous uniform distribution:

Continuous uniform distribution = a distribution that has an equal probability across the interval.



Since $\int_a^b f(x)dx = 1$, the area under that square is 1. This is obvious since maximum probability is 1. Therefore the height of the square must be $\frac{1}{b-a}$ because $\frac{1}{b-a} \times (b-a) = 1$. And as we know from earlier, the height here is the probability so there is a uniform probability of $\frac{1}{b-a}$ across all possible events (x values within given boundaries). This makes integral easier since the probability is constant and hence can be brought out from integral sign.

6. Binomial distribution, its mean and variance. Poisson distribution, its mean and variance.

- Be able to use binomial distribution:

Binomial distribution = a model to calculate the outcome of a True/False result, hence the name bi-nomial. A very simple example is the distribution of flipping a coin.

$$P(X = r) = \binom{n}{r} P_{sucess}^{r} \times P_{failure}^{n-r}$$

Where $P_{success} + P_{failure} = 1, r = number\ of\ success\ \ n = number\ of\ trials,$ $P = probability$

Its parameter is written as $X \sim B(n, p)$.

Then you should be able to find out $E(X)\ and\ Var(X)$ of binomial distributions. The formula is written in the booklet!

Its cumulative would simply be the sum of the distribution just like all cumulative results. We would therefore write is as $P(X \leq r)$, where "r" is the specific number of success.

$$\sum_{r=a}^{r} \binom{n}{r} P_{sucess}^{r} \times P_{failure}^{n-r}$$

Be able to use Poisson distribution:

Poisson distribution = a model to calculate expected rate. It extremely important to note the word "rate" here. Think of the Poisson as a fraction.

$$P(X = x) = \frac{m^x e^{-m}}{x!}\ or\ \frac{\lambda^x e^{-\lambda}}{x!}$$

$$\lambda = \frac{average\ occurence}{space\ or\ time}$$

Thus, when solving Poisson distribution, make sure that the units of λ is the same as in the question. Ex, a car breaks 8 times per week. The λ here would be 8. However, if a question instead asks the probability of breaking a car 4 times **a day**, you adjust the rate. λ would now be simply $\frac{8}{7}$ because you have divided the per week by 7 to make it into per day. Then you plug into the formula and solve!

This feels like "wth" in the beginning but after about 5 questions, it will be the easiest distribution you have done.

Its cumulative would be the sum again.

$$P(x \leq n) = P(x = 0) + P(x = 1) + P(x = 2) \ldots P(x = n) = \sum_{x=0}^{n} \frac{\lambda^x e^{-\lambda}}{x!}$$

An important characteristic of Poisson is that $E(X)\ and\ Var(X)$ are both equal to λ. Please realize that different distributions have different characteristics, so use them wisely. A question may ask whether a certain distribution can be expressed into another distribution. In those cases, $E(X)\ and\ Var(X)$ will come handy.

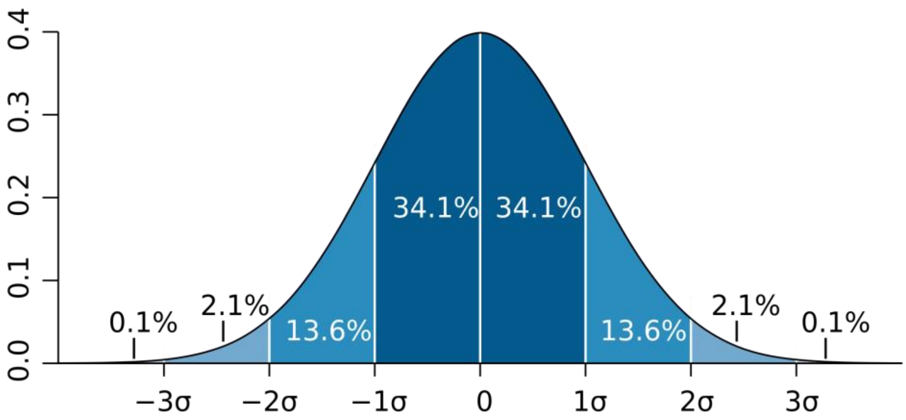
7. Normal distribution. Properties of the normal distribution. Standardization of normal variables.

- Be able to use normal/Gaussian distribution:

Normal distribution = most commonly occurring distribution.

$$P(X = x) = complicated\ formula, but\ it\ is\ in\ your\ booklet$$

Its parameters are written as $X \sim N(\mu, \sigma^2)$, where μ is mean and σ^2 is variance.



Within one standard deviation, we find about 0.68% of the values. Within two standard deviations, we find about 0.95% of the values. Within three standard deviations, almost all values are within that boundary.

Thus the characteristics are:

1. $\int_{-3\sigma}^{3\sigma} f(x)dx \approx 1$

2. It is symmetrical from μ. The graph above is standardized so μ is 0. You will learn more about standardization below.

3. It is bell-shaped.

4. Maximum value is $\frac{1}{\sigma\sqrt{2\pi}}$. This means that if there is high standard deviation, the values are spread out so mean is likely to be low. It is the opposite for low standard deviation.

Be able to standardize normal distribution:

The whole idea is to convert $X \sim N(\mu, \sigma^2)$ to $Z \sim N(0,1)$. This gets rid of the units and makes all normal distributions comparable, hence the term "standardizing" the distribution.

$$\mathrm{Z} = \frac{x - \mu}{\sigma}$$

This also feels like alien language but after a couple of questions, it will be simple.

8. Use of calculators!!!

- I cannot emphasize this enough. If you know how to use your calculator, not only in statistics but in all topics, your life will be so much easier.

SO PLEASE GRAB YOUR TEACHER AND MAKE HIM TEACH YOU HOW TO USE THE GDC.