MAT 2377 Cheat Sheet

1 Chapter 1: Probabilities

The **sample space** is the set of all possible outcomes.

An **event** is a collection of outcomes in the sample space. Usually this is what we are looking to work with.

We can count items using the k stage procedure.

If we have k stages, each with n_1, n_2, n_3, ... possibilities, then the total number of possiblilites is just $n_1 \cdot n_2 \cdot n_3 \cdot ... \cdot n_k$.

1.1 Ordered Samples

If we have an ordered sample, then we see that picking $1, 2, 3$ is different than picking in a different order $1, 3, 2$.

If we draw r items from a bag of n items:

- If we replace each item after drawing, we have: $n \cdot n \cdot n \cdot ... = n^r$ possibilities
- If we do NOT replace the items, we have: $n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-r) = \frac{n!}{(n-r)!} = {}_nP_r$

1.2 Unordered Samples

This is when the order of the samples does not matter, so $1, 2, 3$ would be the same as $1, 3, 2$.

We can see the number of unordered samples possible with r draws in a sample space of size n using:

$$\frac{n!}{(n-r)!r!} = {}_nC_r$$

1.3 Probabilities

The probability of an event A with N total outcomes and a favourable outcomes is just:

$$P(A) = \frac{a}{N}$$

We can add probabilities using the following formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Any 2 events that satisfy the following expression are called **independant.**

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

1.4 Conditional Probability

We say that the probability of event B given that event A has already happened is:

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

1.5 Law of Total Probability

This basically works off of the fact that all probabilities must add up to 1.

This is the specific case to 2 events A and B:

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

This uses the fact that A and $\overline{A}$ are mutually exclusive, and exaustive (covers all of S).

So in general, if we have $A_1, A_2, ..., A_k$ and $A_1, A_2, ..., A_k$ are mutually exclusive and exaustive, then we say:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + ... + P(B|A_k)P(A_k)$$

1.6 Bayes Theorum

This is a way to get the opposite conditional probability to what we have.

If we have $P(A|B)$, among a couple other things, we can obtain $P(B|A)$ with:

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

2 Chapter 2: Discrete Random Variables

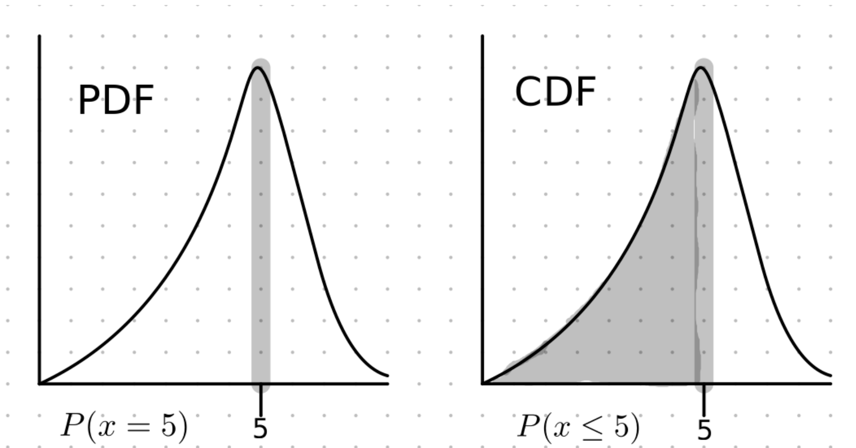
A random variable is a variable (typically a capital letter) that associates a number to every outcome of an experiment.

We have 2 functions:

- Probability Distribution Function (PDF)
- Cumulative Distribution Function (CDF)

The PDF specifies the probability of getting this specific value.

The CDF specifies the probabilitu of getting anything below this specific value.



2.1 Expectation

2.2 Variance

2.3 Binomial Distribution

2.4 Geometric Distribution

2.5 Poisson Distribution

3 Chapter 3: Continuous Random Variables

3.1 Expectation

3.2 Normal Distribution

3.3 Exponential Distribution

3.4 Gamma Distribution

3.5 Joint Distributions

4 Chapter 4: Descriptive Statistics and Sampling

To describe a dataset, we have measures of central tendancy such as mean and median, and measures of spread such as standard deviation, quartiles, and inter-quartile-range.

The median is just the middle value of a **sorted** dataset. If there are 2 middle values, we just take the mean of both of those.

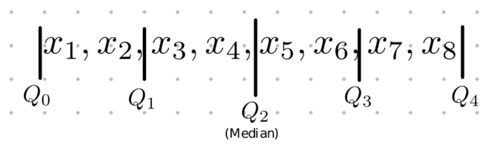
The mean is just:

$$MEAN = \overline{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

The median is often used since it is not heavily influenced by outliers unlike mean.

4.1 Quartiles

The quartile is like taking the median of the lower half of the data (under the true median).



We call the **Inter Quartile Range (IQR)** as the difference between the third and first quartile $IQR = Q_3 - Q_1$

We identify a datapoint x as an outlier if:

$$x < Q_1 - 1.5IQR \qquad \text{or} \qquad x > Q_3 + 1.5IQR$$

4.2 Sample Statistics

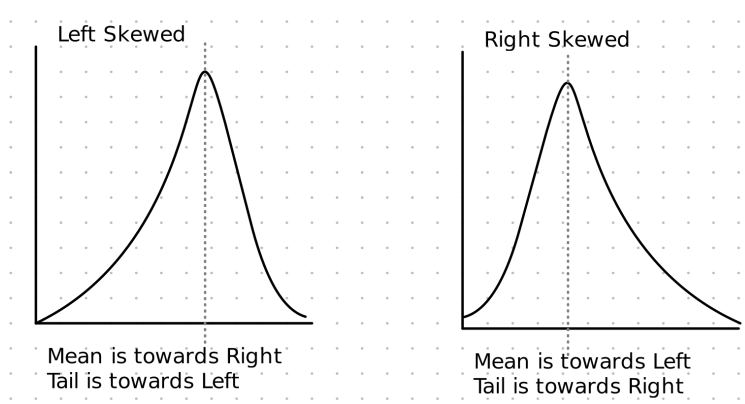
If we do not know the variance of a whole dataset (such as the entire earths population) then we can consider a sample of this population to estimate the population.

We have **sample standard deviation** s and **sample variance** s^2. These estimage the standard deviation σ and variance σ^2 respectively.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

4.3 Skewness

We call a dataset **left skewed** of the tail of the data is to the left (outliers on the left) or **right skewed** if the tail of the data is on the right (outliers on the right).



4.4 Independant and Identically Distributed (IID) Case

When all variables are independant and identically distributed we say the expected value and variance of the entire set is just the number of variables times the variance/expected value of one item.

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu \qquad \text{Var}\left[\sum_{i=1}^{n} X_i\right] = n\sigma^2$$

Then we say that if we are considering a sample of these, we have:

$$\mathbb{E}[\overline{X}] = \mu \qquad \text{Var}[\overline{X}] = \frac{\sigma^2}{n}$$

We can also use the normal distribution if the population is normally distributed to model $\sum_{i=1}^{n} X_i$ or $\overline{X}$.

4.5 Central Limit Theorum

This states that as the number of runs of an experiment, it will start to reach a normal distribution. This is regardless of whether or not the experiment is normal or not.

4.6 Difference between 2 Means

We can work with 2 variables $X_1, X_2, ..., X_n$ with μ_1, σ_1^2, and $Y_1, Y_2, ..., Y_m$ with μ_2, σ_2^2 using the following formula:

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

4.7 Other Distributions

We have 2 other main distributions. The Chi squared (χ^2) distribution, and Student's t distribution. Student's t distribution is used when the population variance is unknown, and we have to approximate using the sample variance (standard deviation).

Both of these distributions have the degrees of freedom which is just $n - 1$.

5 Chapter 5: Point and Interval Estimation

This chapter is mostly about confidence intervals (CI).

We have 2 main confidence intervals that we use. Each of them has an α value where if we say the n percent interval, we have $\alpha = 1 - n$.

So for the 95% confidence interval, $\alpha = 0.05$.

The 2 main confidence intervals are the 95 percent, and 99 percent.

5.1 CI When σ is known

We can use a normal distribution to model this since we know σ, n, and $\overline{X}$.

We use the equation:

$$CI = \overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Using the normal table we can get $Z_{\alpha/2}$. For $\alpha = 0.05$ we have $Z_{0.025} = 1.96$ and for $\alpha = 0.01$ we have $Z_{0.005} = 2.575$.

5.2 CI When σ is unknown

Here we have to find the sample variance s and we know n, and $\overline{X}$.

We use Student's t distribution with the equation:

$$CI = \overline{X} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

Recall that $n-1$ is the degrees of freedom for the t distribution.

5.3 CI For a Proportion

When we are dealing with a proportion for a binomial distributions (2 options, either success of failure), we say that P is the probability of success.

We can model this using the normal distribution using:

$$CI = P \pm z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

6 Chapter 6: Hypothesis Testing

This chapter is about hypotheses. We create 2 hypotheses. The first one, H_1 is the alternative hypothesis. We test it against the null hypothesis H_0. We do the test for H_0 and we either **reject** the null hypothesis in favour of H_1, or **fail to reject** the null hypothesis. We reject the null if evidence against the null is **strong**.

We commit a **Type 1 error** if we reject H_0 when H_0 is actually true. The probability of a type 1 error is:

$$\alpha = P(\text{reject } H_0 | H_0 \text{ is True})$$

We commit a **Type 2 error** if we fail to reject H_0 when H_0 is actually false. The probability of a type 2 error is:

$$\beta = P(\text{fail to reject } H_0 | H_0 \text{ is False})$$

6.1 Types of Hypotheses

Typically we are testing if the mean is the same as we expect (null), or if it differs (alternative).

We say that:

$$
\begin{aligned}
&H_0 : \mu = \mu_0 \\
&H_1 : \mu \neq \mu_0 \quad \text{OR} \quad \mu < \mu_0 \quad \text{OR} \quad \mu > \mu_0
\end{aligned}
$$

6.2 Test for Mean with Known Variance

If the population is normal, or it has a large number of samples (n is large), then we can use a normal distribution to approximate the test.

We can get a value from the normal distribution using the variance, mean, and number of samples.

6.3 Test for Mean with Unknown Variance

If we do not know the variance, we can use Student's t distribution.

6.4 Two Sample Test

7 Chapter 7: Linear Regresion

7.1 Correlation Coefficient

The coefficient of correlation ρ between 2 variables x and y is:

$$\rho_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

We can also wright these as:

$$
\begin{aligned}
S_{xy} &= \sum x_i y_i - n\overline{xy} \\
S_{xx} &= \sum x_i^2 - n\overline{x}^2 \\
S_{yy} &= \sum y_i^2 - n\overline{y}^2
\end{aligned}
$$

7.2 Linear Regression

We can get a line of best fit using the equation of:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Here ϵ is the error term which is often omitted.

The β_1 can be found using the following equation:

$$\beta_1 = \frac{S_{xy}}{S_{xx}}$$

The β_0 can be found by subbing in a known values for the equation of $\overline{y} = \beta_0 + \beta_1\overline{x}$ adn solving for β_0.

We can also estimate the variance by doing:

$$\hat{\sigma}^2 = \frac{S_{yy} - \beta_1 S_{xy}}{n - 2w}$$

7.3 Hypothesis Testing

We can do hypothesis testing using these β values such as where:

$$H_0 : \beta_0 = \beta_{0,0} \qquad H_1 : \beta_0 \neq \beta_{0,0}$$

Similarly to chapter 6, this result is often normally distributed so can be easily calculated using the normal table.