H2 Further Mathematics 9649

Statistics

Zhuo Zhuzhen

Hwa Chong Institution

Preface

One of my most vivid memory from JC is how I was contemplating my life decisions after finding out that my friend who dropped FM for computing was getting easy As while I was stuck with my Ds no matter how hard I tried. Indeed, I don't think many can deny the fact that FM is a difficult subject, and I would argue its one of the most difficult A level subject there is. Perhaps you are feeling the same struggle - and maybe even a bit of helplessness - as I once did.

Stats, for many, is much easier than pure mathematics, and I am certainly one of them. I probably don't have to say this, but don't be complacent and neglect it. Make sure not to make careless mistakes, memorise the formulas well and follow the format that the teachers gave diligently.

I created this set of notes as a convenient reference for myself for everything that I needed to know about FM and as a tool for memorising the dozens of formulas. Now that it has served its purpose, I hope it will be as useful to you in achieving your desired goals, just as I was able to achieve mine.

All the best for all your future exams!

Acknowledgements

This entire set of notes is an adapted and summarised version of the Further Mathematics notes created by HCI Math Department. Most figures are taken directly from the source. Special thanks must also be given to my FM teachers, Mr Ng Say Tiong and Miss Ho Jia Yuan, for working so hard and being so selfless.

Table of Contents

CHAPTER 1

Discrete Random Variable

1.1 Poisson Distribution

Poisson random variable have the following characteristics:

1. Events occur *randomly* in a **fixed interval** and **independent** of each other.
2. The mean rate of occurrence of the event is **constant** in the given interval.
3. The probability of *more than one* event occurring within a short interval is **negligible**.

Note that Poisson distribution can also be used to estimate binomial distribution when $n \to \infty$ and $p \to 0$ (as the mean $\mathrm{E}(X) = np$ is kept constant).

If we let X be the number of occurrences in a fixed interval, and the mean number of occurrence of an event in the given interval be λ

$$X \sim \mathrm{Po}(\lambda)$$

While the probability of obtaining x success in the given interval (can be found in MF 26) is:

$$\mathrm{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, 2, 3, \cdots$$

Note:

1. The Poisson random variable X ranges over an infinite number of integer values (starting from 0).
2. The value of λ is proportional to the length of the interval.
3. Whenever the mean or the interval length changes, a new Poisson variable must be defined.
4. The probability distribution of X becomes more symmetrical as λ increases.
5. When $\lambda \to \infty$, X follows a normal distribution approximately.

1.1.1 Expectation, Variance and Mode of Poisson Distribution

$$\text{If } X \sim \mathrm{Po}(\lambda), \text{ then } \mathrm{E}(X) = \mathrm{Var}(X) = \lambda$$

Note that this can be found in MF 26. However, one may need a rough idea of the proof (though questions may provide clues, but the starting point is the same).

Proof. If we let $X \sim \text{Po}(\lambda)$,

$$\begin{aligned}
\text{E}(X) &= \sum_{r=0}^{\infty} rP(X = r) \\
&= \sum_{r=0}^{\infty} re^{-\lambda}\frac{\lambda^r}{r!} \\
&= \lambda e^{-\lambda}\sum_{r=1}^{\infty}\frac{\lambda^{r-1}}{(r-1)!} \\
&= \lambda e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right) \\
&= \lambda e^{-\lambda}\left(e^{\lambda}\right) \\
&= \lambda
\end{aligned}$$

$$\begin{aligned}
\text{E}(X^2) &= \sum_{r=0}^{\infty} r^2P(X = r) \\
&= \sum_{r=0}^{\infty} r^2e^{-\lambda}\frac{\lambda^r}{r!} \\
&= e^{-\lambda}\sum_{r=1}^{\infty} r\frac{\lambda^r}{r!} + e^{-\lambda}\sum_{r=1}^{\infty} r(r-1)\frac{\lambda^r}{r!} \\
&= \lambda + \lambda^2e^{-\lambda}\sum_{r=2}^{\infty}\frac{\lambda^{r-2}}{(r-2)!} \\
&= \lambda + \lambda^2e^{-\lambda}\sum_{r=1}^{\infty}\frac{\lambda^{r-1}}{(r-1)!} \\
&= \lambda + \lambda^2e^{-\lambda}(e^{\lambda}) \\
&= \lambda + \lambda^2
\end{aligned}$$

Using the fact that $\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2$,

$$\begin{aligned}
\text{Var}(X) &= \lambda + \lambda^2 - \lambda^2 \\
&= \lambda
\end{aligned}$$

As for the mode, it is usually close to the mean, $\text{E}(X)$, and the mode can take 2 values of X.

1.1.2 Additive Property of Poisson Random Variables

If $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\mu)$, where X and Y are independent, then $X + Y \sim \text{Po}(\lambda + \mu)$.

1.2 Geometric Distribution

As the geometric distribution involves Bernoulli trials like the binomial distribution, the conditions are similar to those for the binomial distribution:

1. Independent trials are carried out.
2. The outcome is either a success or failure (Bernoulli trials).
3. The probability of a successful outcome is the same for each trial.

If we let X be the number of trials needed to obtain the first successful outcome, and the probability of success of a trial, p,

$$X \sim \text{Geo}(p)$$

The probability of obtaining the first success at the xth attempt (can be found in MF26) is:

$$\text{P}(X = x) = (1-p)^{x-1}p,\ x = 1, 2, 3, \cdots$$

Note:

1. X cannot take the value of 0.
2. The number of trials can be infinite.
3. The probability to failure is usually defined as: $\text{P(failure)} = q = 1 - p$

1.2.1 Expectation, Variance and Mode of Geometric Distribution

$$\text{If } X \sim \text{Geo}(p), \text{ then } \text{E}(X) = \frac{1}{p}, \text{ and } \text{Var}(X) = \frac{1-p}{p^2}$$

The formulae for these can be found in MF 26, though the proof of this may be required in a question (hints will be given).

Proof. If we let $X \sim \text{Geo}(p)$,

$$\begin{aligned}
\text{E}(X) &= \sum_{r=1}^{\infty} rP(X = r) \\
&= \sum_{r=1}^{\infty} rq^{r-1}p \\
&= p\frac{\text{d}}{\text{d}q}\sum_{r=0}^{\infty} q^r \\
&= p\frac{\text{d}}{\text{d}q}\left(\frac{1}{1-q}\right) \\
&= p(1-q)^{-2} \\
&= p\frac{1}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

$$\begin{aligned}
\text{E}(X^2) &= \sum_{r=1}^{\infty} r^2P(X = r) \\
&= \sum_{r=1}^{\infty} r^2q^{r-1}p \\
&= p\sum_{r=1}^{\infty} rq^{r-1} + p\sum_{r=1}^{\infty} r(r-1)q^{r-1} \\
&= \frac{1}{p} + pq\sum_{r=1}^{\infty} r(r-1)q^{r-2} \\
&= \frac{1}{p} + pq\frac{\text{d}^2}{\text{d}q^2}\sum_{r=0}^{\infty} q^r \\
&= \frac{1}{p} + pq\frac{\text{d}^2}{\text{d}q^2}\left(\frac{1}{1-q}\right) \\
&= \frac{1}{p} + 2pq(1-q)^{-3} \\
&= \frac{2-p}{p^2}
\end{aligned}$$

Hence, using the formula to find variance,

$$\begin{aligned}
\text{Var}(X) &= \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 \\
&= \frac{1-p}{p^2}
\end{aligned}$$

As for the mode, it is always 1 as each subsequent x value means there was one more unsuccessful trial, whose probability has an extra term of $(1-p)$ multiplied as compared to the previous x.

1.2.2 Memoryless Property

If the probability of events happening in the future is *independent* of what went on before, the random variable is memoryless, and the two distributions which are are:

1. Exponential distribution of non-negative real numbers.
2. Geometric distribution of non-negative integers.

This property can be expressed in numerous ways, such as:

$$\begin{aligned}
\mathrm{P}(X > a + b \mid X > a) &= \mathrm{P}(X > b) \\
\mathrm{P}(X = a + b \mid X > a) &= \mathrm{P}(X = b) \\
\mathrm{P}(X \geq a + b \mid X \geq a) &= \mathrm{P}(X > b)
\end{aligned}$$

Which when combined with $\mathrm{P}(A \mid B) = \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$ gives:

$$\begin{aligned}
\mathrm{P}(X > a + b) &= \mathrm{P}(X > a)\mathrm{P}(X > b) \\
\mathrm{P}(X = a + b) &= \mathrm{P}(X > a)\mathrm{P}(X = b) \\
\mathrm{P}(X \geq a + b) &= \mathrm{P}(X \geq a)\mathrm{P}(X > b)
\end{aligned}$$

Note:

1. We have to deal with the inequality signs carefully in the case of discrete random variables.
2. The memoryless property implies that the probability of an outcome does not depend on previous trials.

CHAPTER 2

Continuous Random Variable

The probability of a CRV taking a specific value is always 0. A *simplistic* way of looking at this is $\frac{\mathrm{n}(x)}{\mathrm{n}(S)} = \frac{1}{\infty} = 0$, where S is the sample space of X, and X is a CRV. From this, it can also be inferred that inequalities involving CRV does not matter if the end values are included or excluded.

2.1 Probability Density Function

The pdf of a CRV is such that:

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

It should also be noted that $\int_{-\infty}^{\infty} \mathrm{f}(x)\,\mathrm{d}(x) = 1$

2.2 Expectation and Variance

The calculation for expectation and variance is similar to DRV, except for the use of integration:

$$\begin{aligned}
\mathrm{E}(X) &= \int_{\text{all } x} x\,\mathrm{f}(x)\,\mathrm{d}x \\
\mathrm{E}(X^2) &= \int_{\text{all } x} x^2\,\mathrm{f}(x)\,\mathrm{d}x \\
\mathrm{Var}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 \\
\mathrm{E}(g(X)) &= \int_{\text{all } x} g(x)\,\mathrm{f}(x)\,\mathrm{d}x
\end{aligned}$$

The following properties are also useful:

$$\begin{aligned}
\mathrm{E}(aX + b) &= a\mathrm{E}(X) + b \\
\mathrm{Var}(aX + b) &= a^2\mathrm{Var}(X)
\end{aligned}$$

2.3 Cumulative Distribution Function

If X is a CRV with pdf $\mathrm{f}(x)$, then the cdf $\mathrm{F}(x)$ is:

$$\mathrm{F}(x) = \mathrm{P}(X \leq x) = \int_{-\infty}^{x} \mathrm{f}(x)\,\mathrm{d}x$$

With the following properties:

1. The distribution function is a non-decreasing function. i.e. $x_1 < x_2 \Rightarrow \mathrm{F}(x_1) \leq \mathrm{F}(x_2)$.
2. $\lim\limits_{x\to-\infty} \mathrm{F}(x) = 0$ and $\lim\limits_{x\to+\infty} \mathrm{F}(x) = 1$

3. $\mathrm{P}(x_1 \leq X \leq x_2) = \mathrm{F}(x_2) - \mathrm{F}(x_1)$

4. As $\mathrm{F}(z) = \int_{-\infty}^{z} \mathrm{f}(x)$, $\dfrac{\mathrm{d}}{\mathrm{d}z}\mathrm{F}(z) = \dfrac{\mathrm{d}}{\mathrm{d}z}\displaystyle\int_{-\infty}^{z} \mathrm{f}(x)\,\mathrm{d}(x) = \mathrm{f}(z)$.

5. If X has a median m, then $\mathrm{F}(m) = \mathrm{P}(X \leq m) = \mathrm{P}(X > m) = \frac{1}{2}$. For a CRV, the median is always unique.

2.4 Uniform Distribution

A uniform distribution is a distribution where all intervals of the same length on the distribution's support are equally probable, and is defined by two parameters, a and b, which are the minimum and maximum values of the CRV respectively.

If we let X follow a uniform distribution,

$$X \sim \mathrm{U}(a, b)$$

With the pdf:

$$\mathrm{f}(x) = \frac{1}{b-a}\ ,\ a \leq x \leq b$$

2.5 Exponential Distribution

If we let X be the wait time for a specific event to first occur in a Poisson process, and the rate of occurrence being λ,

$$X \sim \exp(\lambda)$$

The probability of the wait time for the first occurrence at the after x (can be found in MF26) is:

$$\mathrm{f}(x) = \lambda e^{-\lambda x},\ x \geq 0$$

2.5.1 Expectation and Variance

If $X \sim \exp(\lambda)$, $\mathrm{E}(X) = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

The formulae for these can be found in MF 26, though the prove of this may be required in a question (hints will be given).

Proof. If we let $X \sim \exp(\lambda)$,

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^\infty x(\lambda e^{-\lambda x})\,\mathrm{d}x \\
&= \left[x\left(-e^{-\lambda x}\right)\right]_0^\infty - \int_0^\infty \left(-e^{-\lambda x}\right)\,\mathrm{d}x \\
&= -\frac{1}{\lambda}\left[e^{-\lambda x}\right]_0^\infty \\
&= \frac{1}{\lambda}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(X^2) &= \int_0^\infty x^2(\lambda e^{-\lambda x})\,\mathrm{d}x \\
&= \left[x^2\left(-e^{-\lambda x}\right)\right]_0^\infty - \int_0^\infty \left(-2xe^{-\lambda x}\right)\,\mathrm{d}x \\
&= \frac{2}{\lambda}\int_0^\infty x(\lambda e^{-\lambda x})\,\mathrm{d}x \\
&= \frac{2}{\lambda^2}
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 \\
&= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\
&= \frac{1}{\lambda^2}
\end{aligned}$$

Note:

1. For all exponential distributions, the mean is larger than the median.
2. Mode = 0.
3. All exponential distributions are memoryless.

CHAPTER 3

Confidence Interval

A **confidence interval** for a population parameter θ with confidence interval level $100(1-\alpha)\%$, where α is a small non-negative number close to zero, is an interval with random variable confidence limits a, b such that for whatever the true value of θ, $\mathrm{P}(a < \theta < b) = 1 - \alpha$.

Note: The random variable is the interval (a, b) and not the fixed value of θ, even though θ is unknown. Hence, it in *incorrect* to say that the probability of θ lying in the interval (a, b) is $(1-\alpha)$. Instead, it is "the probability of the confidence interval (a, b) containing θ is $(1-\alpha)$.

3.1 Confidence Interval for Population Mean μ

Sample Size	σ **known**	σ **unknown**	**Assumptions Required**
$n < 30$	$\overline{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\overline{X} \pm t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}$	Population is normally distributed.
$n \geq 30$		$\overline{X} \pm z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$	As CLT is used, no assumptions is needed.

Note:

1. If the sample size n is fixed, then increasing the level of confidence increases the width of the interval.
2. If the confidence level is fixed, then increasing the sample size reduces the interval width.
3. If the interval width is fixed, the sample size has to increase to increase the level of confidence.
4. If any assumptions needs to be stated, write down what the population is instead of simply X.

Given the pettiness of examiners and markers, what has to be written is demonstrated in the following examples. Note that the actual population and the values of the parameter have to be written.

Example 1. Given a population that follows a normal distribution, and the value of σ^2 is given, calculate a symmetric $100(1-\alpha)\%$ confidence interval for the population mean.

(1) Let X be the *population*, μ be the population mean, and σ^2 be the population variance of X.

$X \sim N(\mu, \sigma^2) \Rightarrow$ (2) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

A $100(1-\alpha)\%$ confidence interval is given by: (3) $(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$,

Where: (4) $\bar{\sigma^2} = \sigma_1^2$, $n = n_1$, $z_{\frac{\alpha}{2}} = z_{\frac{\alpha}{2},1}$, $\bar{x} = \bar{x}_1$

Hence, the $100(1-\alpha)\%$ confidence interval is: (5) $(\gamma_{\text{lower bound}}, \gamma_{\text{upper bound}})$

Example 2. Given a population with a large sample size, and the sample variance and sample mean given, find a symmetric $100(1-\alpha)\%$ confidence interval for the population mean.

(1) Let X be the *population*, μ be the population mean, and σ^2 be the population variance of X.

(2) Since sample size $n = n_1$ is large, by CLT $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ **approx**
$\Rightarrow Z = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ **approx**

Hence, $100(1-\alpha)\%$ confidence interval: (3) $(\bar{x} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}})$

Where: (4) $s^2 = s_1^2$, $z_{\frac{\alpha}{2}} = z_{\frac{\alpha}{2},1}$, $\bar{x} = \bar{x}_1$.

Hence, the $100(1-\alpha)\%$ confidence interval is: (5) $(\gamma_{\text{lower bound}}, \gamma_{\text{upper bound}})$

Example 3. Given a population normally distributed with a small sample size, and the sample variance and sample mean given, find a symmetric $100(1-\alpha)\%$ confidence interval for the population mean.

(1) Let X be the *population*, μ be the population mean, and σ^2 be the population variance of X.

(2) Since sample size $n = n_1$ is small, $T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$.

Hence, $100(1-\alpha)\%$ confidence interval: (3) $(\bar{x} - t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}})$

Where: (4) $s^2 = s_1^2$, $t_{\frac{\alpha}{2}}(n-1) = t_{\frac{\alpha}{2},1}(n-1)$, $\bar{x} = \bar{x}_1$.

Hence, the $100(1-\alpha)\%$ confidence interval is: (5) $(\gamma_{\text{lower bound}}, \gamma_{\text{upper bound}})$

3.2 Confidence Interval for Population Proportion p

In the syllabus, only a large sample size will be tested, where the upper and lower bounds of the $100(1-\alpha)$ confidence interval is:

$$P_s \pm z_{\frac{\alpha}{2}} \sqrt{\frac{P_s(1-P_s)}{n}}$$

Example 4. Given a population where the sample size is n, that is large, with a proportion p, calculate a symmetric $100(1-\alpha)\%$ confidence interval for the population mean.

(1) Let P be the *population*, and $q = 1 - p$.

(2) Since sample size $n = n_1$ is large, by CLT $\overline{P} \sim N(p, \frac{pq}{n})$ **approx**
$\Rightarrow Z = \frac{P_s - p}{\sqrt{\frac{pq}{n}}}$ **approx**

Hence, $100(1-\alpha)\%$ confidence interval: (3) $(P_s - z_{\frac{\alpha}{2}}\sqrt{\frac{P_s(1-P_s)}{n}}, P_s + z_{\frac{\alpha}{2}}\sqrt{\frac{P_s(1-P_s)}{n}})$

Where: (4) $z_{\frac{\alpha}{2}} = z'_{\frac{\alpha}{2}}$.

Hence, the $100(1-\alpha)\%$ confidence interval is: (5) $(\gamma_{\text{lower bound}}, \gamma_{\text{upper bound}})$

CHAPTER 4

Hypothesis Testing

In general, a similar working to the following working can be used for any hypothesis test question:

Let X be the *population*.
Let μ be the population mean of X.
Let σ^2 be the population variance of X.

To test H_0: $\mu = \ldots$ against
H_1: $\mu < \ldots$ at $\gamma\%$ level of sig.

Test stat: ...

Using GC,
$\bar{x} = \ldots$,
...
p-value $= \ldots > / < \ldots$

$\therefore$ H_0 is rejected/ not rejected at $\gamma\%$ level of sig., i.e. there is sufficient/ insufficient evidence to conclude that ...

4.1 t-Test for Population Mean

When given a small population size with unknown variance, the test statistic is given by:

$$T = \frac{\overline{X} - \mu_0}{\sqrt{\frac{S^2}{n}}} \sim t(n-1)$$

4.2 Two-Sample Tests for Testing Difference between Independent Population Means

In the syllabus, we will only be concerned with random samples taken from two samples X_1, X_2, where $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. For populations which follows a normal distribution:

Sample Size	σ_1, σ_2 **known**	σ_1, σ_2 **unknown**
$n < 30$	$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$
$n \geq 30$		$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$

Note that for the t-test, $\sigma_1^2 = \sigma_2^2$, and S_p is given by $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ (this form is not given in MF 26). Moreover, as the tests usually consists of the case when $\mu_1 = \mu_2$, the expression $\mu_1 - \mu_2 = 0$, and hence can be ignored. When using G.C. to calculate the 2 sample t-test, ensure that "Pooled" is selected to be "Yes".

When the population does not follow a normal distribution, use CLT to approximate the sample to follow a normal distribution, and apply the formulas accordingly.

4.3 Two-Sample Tests for Testing Difference between Paired Population Means

If observations of an object in one sample is paired with observations of the **same** object in the other sample, or when two different items are subjected to the **same** unique conditions, the two samples are not independent.

In this case, the pairs have to be matched, and we define $D = X_1 - X_2$. *Assuming that D is normally distributed*, we can use the test stat:

$$T = \frac{\overline{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \sim t(n-1)$$

Then, a similar procedure is done to the one-sample t-test. Note that $\mu_D = \mu_1 - \mu_2$, and $S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$.

4.4 CI and Hypo Testing

A hypothesis test at a $\alpha\%$ level of significance with $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$ is equivalent to constructing a $100(1-\alpha)\%$ CI for μ. Hence, if μ_0 lies *outside* the CI, we *do not* reject H_0, and vice versa.

CHAPTER 5

Chi-Square Tests

The χ^2 distributions have the following properties (not really needed):

1. Total area = 1.
2. Each χ^2 curve (other than that of degree of freedom 1 and 2) begin at 0 on the horizontal axis, increases to the peak, and then approaches the horizontal axis asymptotically from above.
3. As the degree of freedom increases, the curve becomes more and more symmetrical and looks more like a normal curve, with the peak shifting rightwards.
4. If $Z \sim N(0,1)$, then $Z^2 \sim \chi_1^2$. Furthermore, if $Z_1, Z_2, \ldots Z_n$ are independent observations, then $Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi_n^2$.
5. For $\nu > 2$, $\mathrm{E}(\chi_\nu^2) = \nu$ and $\mathrm{Var}(\chi_\nu^2) = 2\nu$, mode $= \nu - 2$.
6. If X_1 and X_2 are independent, and $X_1 \sim \chi_n^2$, $X_2 \sim \chi_m^2$, then $X_1 + X_2 \sim \chi_{n+m}^2$.

The χ^2 statistic is calculated using:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad \text{Where } O = \text{observed frequency and } E = \text{expected frequency}$$

It is a discrete distribution that approximates the continuous distribution. When there is close agreement between O and E, the χ^2 value is small and hence the null hypothesis will not be rejected and vice versa.

5.1 Test for Goodness of Fit

(1) State the hypothesis

To test H_0: ... against
H_1: ..., at $\gamma\%$ level of sig.

(2) Find the expected frequency (Don't have to give individual contribution unless needed).

Under H_0, X $\cdots$,

X	α	$\cdots$
Observed Frequency	O_1	$\cdots$
Expected Frequency	E_1	$\cdots$
Contribution	$\frac{(O-E)^2}{E}$	$\cdots$

(3) Merges/ changes

If the expected frequency of an entry is below 5, it needs to be merged with another column to ensure all entries have an expected frequency of more than 5. Moreover, also ensure that $\sum O = \sum E$. If not, minus away small values from one of the E to ensure that the total remains the same.

(4) Calculating chi-square test stat

Either use the formula for χ^2 test or use the GC to from a list for the expected values and another for the observed values and use the χ^2 -GOF Test. Note that the degree of freedom is (no. of columns -1) $-$ (no. of parameters estimated).

$$\text{Test stat: } \chi^2 = \frac{\sum(O-E)^2}{E} \sim \chi^2_\nu$$

By G.C.,
$\chi^2 = \cdots$
p-value $= \cdots$

(5) Arrive at conclusion

Use the p-value or critical region to reach a conclusion, i.e. since the p-value is $\ldots > / < \ldots$, we (do not) reject H_0 at $\alpha\%$ level of sig, i.e. there is insufficient/ sufficient evidence to conclude that ....

5.2 Test of Homogeneity/ Independence

To test for these, χ^2 test is used in combination with the contingency table. The way to solve these type of question is shown:

(1) State the hypothesis

To test H_0: There is no difference/ it is independent of ..., against
H_1: There is a difference/ it is not independent of ..., at $\gamma\%$ level of sig.

(2) Find the expected frequency

Find the frequency of each entry using:

$$\begin{aligned}\text{Expected Frequency} &= \mathrm{P}(A)\mathrm{P}(X) \times \text{grand total} \\ &= \frac{\text{row total} \times \text{column total}}{\text{grand total}}\end{aligned}$$

And write it next to each entry in brackets. i.e.

	X	Y	$\cdots$	Total
A	α_1 (γ_1)	α_2 (γ_2)	$\cdots$	$\sum \alpha$
B	α_1' (γ_1')	α_2' (γ_2')	$\cdots$	$\sum \alpha'$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Total	$\alpha_1 + \alpha_1' + \cdots$	$\alpha_2 + \alpha_2' + \cdots$	$\cdots$	$\sum \alpha + \sum \alpha' + \cdots$

(3) Merges/ changes

If the expected frequency of an entry is below 5, it needs to be merged with another row or column to ensure all entries have an expected frequency of more than 5. Moreover, also ensure that $\gamma_1 + \gamma_2 + \cdots = \sum \alpha = \alpha_1 + \alpha_2$. If not, minus away small values from one or more of the γ to ensure that the total remains the same.

(4) Calculating chi-square test stat

Either use the formula for χ^2 test or use the GC to from a matrix for the expected values and another for the observed values and use the χ^2 - Test. Note that the degree of freedom of any contingency table is (no. of rows -1) $\times$ (no. of columns -1).

$$\begin{aligned}\text{degree of freedom} &= (r-1)(c-1) \\ &= \nu\end{aligned}$$

$$\text{Test stat: } \chi^2 = \frac{\sum (O-E)^2}{E} \sim \chi^2_\nu$$

By G.C.,
$\chi^2 = \cdots$
p-value $= \cdots$

(5) Arrive at conclusion

Use the p-value or critical region to reach a conclusion, i.e. since the p-value is $\ldots > / < \ldots$, we (do not) reject H_0 at $\alpha\%$ level of sig, i.e. there is insufficient/ sufficient evidence to conclude that ....

CHAPTER 6

Non-Parametric Tests

A non-parametric test or 'distribution free' test is a hypothesis that does not assume any knowledge of the distribution or the value of any parameters of the population involved.

The non-parametric test tests for the median instead of mean.

6.1 Sign Test

The sign test is highly appropriate when the sample size is small and the data have outliers/ heavily skewed. However, it is less powerful than t-test when the population is close to normally distributed as it does not use all of the information given (mean, variance).

For matched-pair problems, the sign test only looks at the signs of the differences of the data pairs and not the magnitude of the differences, hence is less precise. If it is the same, it is assigned a 0.

Requires the assumption that the sample is drawn randomly from a continuous distribution.

(1) State the hypothesis

To Test: $H_0 : m = m_0$ against
$H_1 : m < / \neq / > m_0$ at $\alpha\%$ level of sig.

(2) Assign Sign

A '−' is assigned if the value is below the median while a '+' is assigned if the value is above the median.
If it is a 2 sample test, simply compare the sign of the difference of the 2 samples with 0.

Person	A	$\cdots$
Sign of Difference	$+/-$	$\cdots$

(3) Perform the Test

Using sign test, test stat $S = S_+/S_-/\min(S_+, S_-)$.
The opposite of the null hypothesis is chosen for the sake of consistency

Adjusted sample size $= n'$ *(if the value is assigned 0, it is ignored.)*
$S_+ = \cdots$, $S_- = \cdots$

Let X $B(n, \frac{1}{2})$

p-value $= P(X \leq S) = \cdots < / > \alpha$

(4) Come to a conclusion

For **large** sample sizes (> 25), X $B(n, \frac{1}{2}) = X$ $N(np, npq)$ approx, where $p = q = \frac{1}{2}$. Then, use the normal distribution to perform the test, though remember to add/ subtract 0.5 to S due to continuous correction (written as $P(X \leq S) \xrightarrow{c.c.} P(X \leq 30.5)$).

6.2 Wilcoxon Signed-Rank Test

Requires the assumptions that

1. The sample of differences is randomly and independently selected from the population of differences.

2. The probability distribution from which the sample of paired differences is drawn is random and continuous.

(1) State the hypothesis

To Test: $H_0 : m = m_0$ against
$H_1 : m < / \neq / > m_0$ at $\alpha\%$ level of sig.

(2) Process Data

1. Calculate the observed values difference D.
2. Discard zero differences and obtain modified sample size n.
3. Assign ranks by arranging $|D|$ in ascending orders (starting from 1)
   (a) Non-zero differences which has the same $|D|$ will be assigned ranks they would receive if they are unequal but occur in successive order (i.e. if its a tie for 5 and 6, order = 5.5).
4. Record the corresponding signs.

D	α	$\cdots$
$\|D\|$	$\|\alpha\|$	$\cdots$
Rank	β	$\cdots$
Sign	$+/-$	$\cdots$

(3) Perform the Test

Using Wilcoxon signed rank test, test stat $T = T_+/T_-/\min(T_+, T_-)$.
The opposite of the null hypothesis is chosen due to limitations to MF26.

Adjusted sample size $= n$, $t_+ = \cdots$, $t_- = \cdots$

Reject H_0 if $T \leq \zeta$

(4) Come to a conclusion

Since $t_-/t_+ \leq / > \zeta$, ...

Note that $T_+ + T_- = \frac{n}{2}(n+1)$ can be used for checking purposes.

For large sample size ($n > 25$), it can be approximated such that $T \sim N(\frac{1}{4}n(n+1), \frac{1}{24}n(n+1)(2n+1)$ (given in MF26). Then, find the p-value by finding $P(T \leq t)$ (remmeber to times 2 for 2-tailed test).

6.3 Comparing Parametric and Non-Parametric Test

	Parametric Test	Non-Parametric Test
Advantages	1. More reliable and accurate since every data is taken into account. 2. Higher statistical power efficiency (as to detect any given effect at a specified significance level, a smaller sample size is required.)	1. Makes fewer assumptions, allowing it to be used in more situations (when population isn't normally distributed). 2. Can be used to test hypotheses that do not have any population parameters. 3. Can be used for data that are nominal or ordinal, or those with outliers.
Disadvantages	1. More restrictive (need sample size to be large or distribution to be normal when sample size is small). 2. Computations are more demanding.	1. Less accurate and reliable (does not make full use of data). 2. Less sensitive and efficient. 3. Difficult to compute manually if sample size is large.

If no measurement of things where each singular value may not represent something, for instance "tastiness" (5 is very tasty, 1 is bad), then one should not use parametric test as variance does not mean anything as the numbers are subjective.