MSE201 Revision notes

1 Probability

1.1 Sample Space

The collection of all possible outcomes of an experiment is called the *sample space* of the experiment. We will denote the sample space by S.

1.1.1 The empty set

The empty set, denoted ϕ, is defined by the subset of S that contains no outcomes. Note that every set contains the empty set and so $\phi \subset A \subset S$.

1.1.2 Complements

The complement of an event A, denoted $\overline{A}$ is defined to be event that contains all outcomes in the sample space S which *do not* belong to A.

1.1.3 Union

If A and B are any 2 events, then the *union* of A and B is defined to be the event containing all outcomes that belong to A alone, to B alone or to both A and B. We denote the union of A and B by $A \cup B$.

1.1.4 Intersection

If A and B are any 2 events, then the *intersection* of A and B is defined to be the event containing all outcomes that belong both to A and to B. We denote the intersection of A and B by $A \cap B$.

1.1.5 Disjoint events

Two events A and B are disjoint if they contain no outcomes in common, *i.e.* $A \cap B = \phi$.

1.1.6 Further results

a) DeMorgan's Laws: for any 2 events A and B we have

$$\boxed{\overline{(A \cup B)} = \overline{A} \cap \overline{B}}$$

$$\boxed{\overline{(A \cap B)} = \overline{A} \cup \overline{B}}$$

b) For any 3 events A, B and C we have:

$$\boxed{A \cap (B \cup C) = (A \cap B) \cup (A \cap C)}$$

$$\boxed{A \cup (B \cap C) = (A \cup B) \cap (A \cup C)}$$

1.2 Axioms of Probability

Axiom 1: for any event A, $\boxed{\mathrm{P}(A) \geq 0}$

Axiom 2: $\boxed{\mathrm{P}(S) = 1}$

Axiom 3: For any sequence of *disjoint* events $A_1, A_2, A_3, \cdots$

$$\boxed{\mathrm{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)}$$

Definition: A *probability distribution*, or simply a *probability* on a sample space S is a specification of numbers $\mathrm{P}(A)$ which satisfy Axioms 1–3.
Properties of probability:

1. $\boxed{\mathrm{P}(\phi) = 0}$
2. For any event A, $\boxed{\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)}$
3. For any event A, $\boxed{0 \leq \mathrm{P}(A) \leq 1}$
4. **The addition law of probability:** For any 2 events A and B:

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

1.2.1 Conditional probability

Definition: if A and B are any 2 events with $\mathrm{P}(B) > 0$, then

$$\boxed{\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}}$$

Giving **the multiplication law of probability**:

$$\boxed{\mathrm{P}(A \cap B) = \mathrm{P}(A \mid B)\mathrm{P}(B) = \mathrm{P}(B \mid A)\mathrm{P}(A)}$$

1.2.2 Independence

Two events A and B are said to be *independent* if

$$\boxed{\mathrm{P}(A \mid B) = \mathrm{P}(A)}$$

So if A and B are independent we have

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

1.2.3 Bayes Theorem

Let S denote the sample space of some experiment and consider k events $A_1, \cdots, A_k$ in S such that $A_1, \cdots, A_k$ are *disjoint* and $\cup_{i=1}^{n} A_i = S$. Such a set of events is said to form a *partition* of S. Then,

Theorem of total probability:

$$\boxed{\mathrm{P}(B) = \sum_{j=1}^{k} \mathrm{P}(A_j)\mathrm{P}(B \mid A_j)}$$

Bayes Theorem:

$$\boxed{\mathrm{P}(A_i \mid B) = \frac{\mathrm{P}(B \mid A_i)\mathrm{P}(A_i)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B \mid A_i)\mathrm{P}(A_i)}{\sum_{j=1}^{k} \mathrm{P}(A_j)\mathrm{P}(B \mid A_j)}}$$

2 Random Variables and their distributions

2.1 Discrete Distributions

Let $p_i = \mathrm{P}(X = x_i),\ i = 1, 2, 3, ...$ Then any set of p_i's such that

1. $\boxed{p_i \geq 0}$ and
2. $\boxed{\sum_{i=1}^{\infty} p_i = \mathrm{P}(X \in S) = 1}$

forms a *probability distribution* over $x_1, x_2, x_3, ...$

The *distribution function* $F(x)$ of a *discrete random variable* is given by

$$\boxed{F(x_j) = \mathrm{P}(X \leq x_j) = \sum_{i=1}^{j} p_i = p_1 + p_2 + ... + p_j}$$

2.1.1 The Uniform Distribution

If X has a uniform distribution on $1, 2, \ldots, k$, then the probability distribution of X is given by:

$$\boxed{\mathrm{P}(X = x) = p_i = \begin{cases} \frac{1}{k} & \text{for } x = 1, 2, ..., k, \\ 0 & \text{otherwise.} \end{cases}}$$

The *distribution function* is given by

$$\boxed{F(j) = \mathrm{P}(X \leq j) = \sum_{i=1}^{j} p_i = p_1 + p_2 + ... + p_j = \tfrac{j}{k} \quad \text{for } j = 1, 2, ..., k}$$

2.1.2 The Binomial Distribution

If $X \sim \mathrm{Bin}(n, p)$, the probability distribution of X is given by:

$$\boxed{\mathrm{P}(X = x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} & \text{for } x = 0, 1, 2, ..., n, \\ 0 & \text{otherwise.} \end{cases}}$$

where

$$\boxed{\begin{pmatrix} n \\ x \end{pmatrix} = \frac{n!}{x!(n-x)!}}$$

2.1.3 The Poisson Distribution

If $X \sim \text{Poisson}(\mu)$, the probability distribution of X is given by:

$$\boxed{\mathrm{P}(X = x) = \begin{cases} \dfrac{e^{-\mu}\mu^x}{x!} & \text{for } x = 0, 1, 2, ..., \\ \\ 0 & \text{otherwise.} \end{cases}}$$

2.2 Continuous Distributions

For a continuous random variable X we have a function f, called the *probability density function* (pdf). Every probability density function must satisfy:

1. $\boxed{f(x) \geq 0}$ and

2. $\boxed{\int_{-\infty}^{\infty} f(x)dx = 1}$

For any interval A we have

$$\boxed{\mathrm{P}(X \in A) = \int_A f(x)dx}$$

The *distribution function* is given by:

$$\boxed{F(x_0) = \mathrm{P}(X \leq x_0) = \int_{-\infty}^{x_0} f(x)dx}$$

2.2.1 The Uniform Distribution

Suppose X is equally likely to occur anywhere within the range $[a, b]$ ($X \sim \mathrm{U}[a, b]$). Then the probability density function of X is given by:

$$\boxed{f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \leq x \leq b, \\ \\ 0 & \text{otherwise.} \end{cases}}$$

2.2.2 The Exponential Distribution

If $T \sim \exp(\lambda)$, the probability density function of X is given by:

$$\boxed{f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t > 0, \\ \\ 0 & \text{otherwise.} \end{cases}}$$

The *distribution function* is given by:

$$\boxed{F(t) = 1 - e^{-\lambda t} \quad \text{for } t > 0}$$

2.2.3 The Normal Distribution

If $X \sim N(\mu, \sigma^2)$, the probability density function of X is given by:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad -\infty < x < \infty}$$

Important result: if $X \sim N(\mu, \sigma^2)$ then the random variable Z defined by

$$\boxed{Z = \frac{X-\mu}{\sigma}}$$

is distributed as $Z \sim N(0, 1)$.

2.3 Mean and variance

For a discrete random variable X the *expected value* (or *mean*) is defined as

$$\boxed{E(X) = \textstyle\sum_{i=1}^{\infty} x_i \mathrm{P}(X = x_i)}$$

For a continuous random variable with probability density function $f(x)$ the *expected value* (or *mean*) is defined as

$$\boxed{E(X) = \textstyle\int_{-\infty}^{\infty} x f(x) dx}$$

For constants $a_1, ..., a_k$ and b and random variables $X_1, ..., X_k$ we have

$$\boxed{E(a_1X_1 + ... + a_kX_k + b) = a_1E(X_1) + ... + a_kE(X_k) + b}$$

The *variance* of a random variable X is defined as

$$\boxed{\mathrm{var}(X) = E[(X-\mu)^2]}$$

where

$$\mu = E(X).$$

So the variance is the mean of the squared distances from the mean.
The *standard deviation* is defined as

$$\boxed{\mathrm{sd}(X) = \sqrt{\mathrm{var}(X)}}$$

For constants $a_1, ..., a_k$ and b and *independent* random variables $X_1, ..., X_k$ we have

$$\boxed{\mathrm{var}(a_1X_1 + ... + a_kX_k + b) = a_1^2\mathrm{var}(X_1) + +... + a_k^2\mathrm{var}(X_k)}$$

Important result: If $X_1, ..., X_n$ are independent random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ and $a_1, ..., a_n$ are constants then the random variable defined by

$$\boxed{Z = \textstyle\sum_{i=1}^{n} a_iX_i}$$

has a normal distribution with mean

$$\boxed{E(Z) = E\left(\textstyle\sum_{i=1}^{n} a_iX_i\right) = \textstyle\sum_{i=1}^{n} a_i\mu_i}$$

and variance

$$\boxed{\mathrm{var}(Z) = \mathrm{var}\left(\textstyle\sum_{i=1}^{n} a_iX_i\right) = \textstyle\sum_{i=1}^{n} a_i^2\sigma_i^2}$$

2.4 Sampling distribution of a normal mean

Definition: a *random sample* of size n from a density $f(x)$ is a set of independent and identically distributed random variables $X_1, ..., X_n$ each with the density $f(x)$.

If $X_1, ..., X_n$ are a *random sample* from the normal density $N(\mu, \sigma^2)$ then the random variable

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

has a normal distribution with:

$$\textit{mean},\ \boxed{E(\bar{X}) = \mu}\ \text{ and } \textit{variance},\ \boxed{\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}}$$

3 Statistical Analysis

3.1 Point estimation

Definition: any real-valued function $T = g(x_1, ..., x_n)$ of the observations in the random sample is called a *statistic*.

T is known as an *estimator* if we use it as a guess of the value of some unknown parameter, θ.

(i) **The average value**: the estimator $T = g(\underline{X})$ is said to be *unbiased* if

$$\boxed{E[T] = \theta}$$

(ii) **The variance**: if we have two estimators that are unbiased then we would rather use the one which has the smaller variance.

Result: if we have a random sample of size n from the normal distribution $N(\mu, \sigma^2)$ with μ and σ^2 both unknown then the estimator

$$\boxed{s = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

is an unbiased estimator of σ^2.

3.2 Confidence Interval: KNOWN variance

The $\boxed{100(1-\alpha)\%}$ confidence interval for the mean μ of a normal distribution when the variance σ^2 is known is given by:

$$\boxed{\left(\bar{x} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}}, \bar{x} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right)}$$

where $z_{\alpha/2}$ is that value such that

$$\boxed{1 - \alpha = P(-z_{\alpha/2} < Z < z_{\alpha/2})}$$

where $Z \sim N(0, 1)$.

3.3 Confidence Interval: UNKNOWN variance

The $\boxed{100(1-\alpha)\%}$ confidence interval for the mean μ of a normal distribution when the variance, σ^2 is unknown and estimated by s given by:

$$\boxed{\left(\bar{x} - \frac{t_{\alpha/2}s}{\sqrt{n}}, \bar{x} + \frac{t_{\alpha/2}s}{\sqrt{n}}\right)}$$

where $t_{\alpha/2}$ is that value such that

$$\boxed{1 - \alpha = P(-t_{\alpha/2} < T < t_{\alpha/2})}$$

where T has a t-distribution with $n - 1$ degrees of freedom.

3.4 Hypothesis Testing

A statistical test consists of the following 4 elements:

1. The **null hypothesis** H_0 about one or more unknown parameters.
2. The **alternative hypothesis** H_1 with which the null hypothesis is being compared.
3. The **test statistic** which is computed from the data.
4. The **rejection region** which indicates which values of the test statistic will lead to rejection of the null hypothesis.

Definition: rejecting the null hypothesis if it is true is known as a **type I error**. We denote by α the probability of making a type I error, *i.e.*

$$\mathrm{P}(\text{reject } \ H_0 \mid H_0 \ \text{ true}) = \alpha.$$

3.4.1 One- and two-tailed tests

Suppose we have data $X_1,\ldots,X_n$ which are assumed to be normally distributed with mean μ. Consider the null hypothesis

$$H_0 : \mu = \mu_0 \ .$$

There are various alternative hypotheses we may consider:

1. $H_1 : \mu \neq \mu_0$ – the general alternative.
2. $H_1 : \mu > \mu_0$.
3. $H_1 : \mu < \mu_0$.

3.4.2 Hypothesis testing: KNOWN variance

If σ^2 is *known*, the test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

If H_0 is true then $Z \sim N(0,1)$.
The rejection regions for the three alternative hypotheses are:

1. $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$,
2. $Z > z_\alpha$,
3. $Z < -z_\alpha$.

3.4.3 Hypothesis testing: UNKNOWN variance

If σ^2 is *unknown*, the test statistic is

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}.$$

where,

$$s = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

If H_0 is true then $T \sim t^{n-1}$ (t- distribution on $n-1$ degrees of freedom).
The rejection regions for the three alternative hypotheses are:

1. $T < -t_{\alpha/2}^{n-1}$ or $T > t_{\alpha/2}^{n-1}$,
2. $T > t_{\alpha}^{n-1}$,
3. $T < -t_{\alpha}^{n-1}$.

3.5 Hypothesis testing: Two normal means

Suppose we have a random sample, $X_1, \ldots X_{n_1-1}$, from the normal distribution $N(\mu_1, \sigma_1^2)$ and a second random sample, $Y_1, \ldots Y_{n_2-1}$, from a normal distribution $N(\mu_2, \sigma_2^2)$. Further suppose we are interested in testing the hypothesis

$$H_0 : \mu_1 = \mu_2,$$

Again, there are various alternative hypotheses we may consider:

1. $H_1 : \mu_1 \neq \mu_2$ – the general alternative.
2. $H_1 : \mu_1 > \mu_2$.
3. $H_1 : \mu_1 < \mu_2$.

3.5.1 Hypothesis testing: Two normal means: KNOWN variance

If σ^2 is *known*, the test statistic is

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

If H_0 is true then $Z \sim N(0,1)$.
The rejection regions for the three alternative hypotheses are:

1. $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$,
2. $Z > z_{\alpha}$,
3. $Z < -z_{\alpha}$.

3.5.2 Hypothesis testing: Two normal means: UNKNOWN variance

If σ^2 is *unknown*, the test statistic is

$$T = \frac{\bar{X} - \bar{Y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Where the variance is estimated by,

$$s = \left[\frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_1 + n_2 - 2}\right]^{1/2} = \left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right]^{1/2}$$

where

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$$

and

$$s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2.$$

If H_0 is true then $T \sim t^{n_1+n_2-2}$ (t- distribution on $n_1 + n_2 - 2$ degrees of freedom). The rejection regions for the three alternative hypotheses are:

1. $T < -t_{\alpha/2}^{n_2+n_2-2}$ or $T > t_{\alpha/2}^{n_1+n_2-2}$,
2. $T > t_{\alpha}^{n_1+n_2-2}$,
3. $T < -t_{\alpha}^{n_1+n_2-2}$.