UQF2101I Cheatsheet

for test 2, by ning

Learning Objectives

- Random variables
- Distribution of probabilities
- Taking expectation, E—mean and variance
- Normal distribution & derived distributions
- Data transformations in its distribution
- Linear combinations of random variables
- Inferring about a population from a sample
- Functions of sample data—statistics & point estimators
- Sampling distributions
- Interval estimators
- Testing the significance of sample findings
- Errors in hypothesis testing
- Reducing probability of errors in hypothesis testing

Random variables & their distributions

- Variable that can take on one or more values, each of them associated with a probability
- Can be discrete or continuous
- Discrete random variables are described by a probability mass function
- Continuous random variables are described by a probability density function
- The area under a probability distribution function (PDF) is always 1, i.e.

$$\int_{-\infty}^{\infty} P(x)\ dx = 1$$

- The mean, μ of a PDF is its central tendency; its variance is its variability, or dispersion about the mean.

  For discrete random variables,

$$\mu = \sum_{i=1}^{n} x_i \cdot f(x_i)$$

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i)$$

  For continuous random variables,

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x)\ dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\ dx$$

Expectation operator

- First, note that the mean and variance produce a single number from many outcomes using a weighted average
- Intuitively, the taking the expected value is a similar process of finding a weighted average

  Some basic properties of the expectation operator, E, for constants a, b, and random variable X,

$$\mathrm{E}(b) = b$$
$$\mathrm{E}(aX+b) = a\mathrm{E}(X) + b$$
$$\mathrm{E}(\mathrm{g}(X)) \neq \mathrm{g}(\mathrm{E}(X))$$

Normal distribution

- Typically notated as $X \sim \mathrm{N}(\mu,\ \sigma^2)$; X is a random variable that is normally distribution with mean μ and variance σ^2
- The standard normal distribution, Z is defined as

$$Z = \frac{X-\mu}{\sigma} \implies Z \sim \mathrm{N}(0,\ 1)$$

  and has a shorthand $P(Z \leq z) = \Phi(z)$

Derived distributions

The normal distribution is used as a basis to generate other important distributions.

Log-normal distribution If $Y = \ln(X)$, where $Y \sim \mathrm{N}(\lambda,\ \xi^2)$, then X is log-normally distributed with mean μ and variance σ^2,

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

$$\lambda = \ln\ \mu - \frac{1}{2}\ \xi^2$$

$$\xi^2 = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right)$$

Chi-square (χ^2)

If $X = Z^2$, where Z is the standard normal, i.e. $Z \sim \mathrm{N}(0,\ 1)$, then X is chi-square (χ^2) distributed. The χ^2 distribution has an additional parameter k, the degree of freedom.

T-distribution

If Z and χ^2_k are independent standard normal and chi-square random variables respectively, then

$$T = \frac{Z}{\sqrt{\chi^2_k / k}}$$

is t-distributed with k degrees of freedom.

Data transformations

- If the data is asymmetric, or has significant outliers, it may be useful to re-express the data in other terms
- This makes the data 'normal' and more easily understandable
- Still, when reporting results, we often report in terms of the original expression
- Some common transformations are,

Observation	Transformation
Strong positive skew	$\ln X$
Moderate positive skew	$\sqrt{X}$
Moderate negative skew	$\sqrt{K-X}$
Strong negative skew	$\ln\ (K-X)$

  where $K = \max + 1$

Linear combinations

- The linear combination of random variables is a random variable
- In particular, if $X_1 \sim \mathrm{N}(\mu_1,\ {\sigma_1}^2)$ and $X_2 \sim \mathrm{N}(\mu_2,\ {\sigma_2}^2)$; and X_1, X_2 are independent, then for $Y = a_1X_1 + a_2X_2$,

$$Y \sim \mathrm{N}(a_1\mu_1 + a_2\mu_2,\ {a_1}^2{\sigma_1}^2 + {a_2}^2{\sigma_2}^2)$$

- However, if X_1 and X_2 are not normally distributed, their linear combination Y may not always be linearly distributed
- Nonetheless, we can still obtain the mean and variance for linear combinations of random variables of any distributions,

$$\begin{aligned}
Y &= a_1X_1 + a_2X_2 \\
\mu_Y &= a_1\mathrm{E}(X_1) + a_2\mathrm{E}(X_2) \\
{\sigma_Y}^2 &= \mathrm{E}((Y-\mu_Y)^2) \\
&= {a_1}^2\mathrm{E}((X_1-\mu_q)^2) \\
&+\ {a_2}^2\mathrm{E}((X_2-\mu_2)^2) \\
&+\ 2a_1a_2\mathrm{E}(X_1-\mu_1)(X_2-\mu_2)
\end{aligned}$$

  Note that the last term is the covariance, and the covariance = 0 if X_1 and X_2 are independent

Central limit theorem

- The sum of a large number of identical and independent random variables has an approximately normal distribution
- Rule of thumb for 'large number': $n \geq 30$

From sample to population

- Population: the totality of observations that we are interested; Sample: a subset of the population obtained
- In reality, it is often practically impossible to collect enough data to definitively draw conclusions and make decisions about the population
- What we can do, however, is to sample a portion of the population, and infer about he population from having analysed the sample
- However, while we can be certain about our conclusions about the sample group, there is going to be some ambiguity when we apply sample conclusions to unobserved data points in the population
- Using probability as an analogy,

Sample-population	Probability
Population	Sample space
Sample	Event

Point estimators

- Suppose we are trying to obtain some parameters of the population distribution, e.g. mean, μ or variance, σ^2
- One way to estimate these parameters is to apply some function to our sample to condense them to a single number (that estimates the parameter we are trying to obtain)
- This function is known as a point estimator, $\hat{\Theta}$; it produces a point estimate, or statistic $\hat{\theta}$, which estimates the population parameter θ

Unbiased estimators

- An estimator is unbiased if $\mathrm{E}(\hat{\Theta}) = \theta$
- Otherwise, the bias of an estimator is bias $= \mathrm{E}(\hat{\Theta}) - \theta$
- Intuitively, the expectation of $\hat{\Theta}$ is the average value of that estimator over many outcomes
- Note that the sample mean, $\bar{X}$ and the sample variance, S^2 are unbiased estimators for the population mean, μ, and population variance, σ^2 respectively

Estimator variance

- Many estimators can be biased; we can further rate estimators by their variance (precision)
- Typically presented as the standard error, which is the standard deviation of the estimator

Mean square error

- Overall, we can quantify the goodness of an estimator by its bias and standard error. The mean square error combines the two into a single quantity, the mean square error:

$$\begin{aligned}
\mathrm{MSE}(\hat{\Theta}) &= (\text{Std error})^2 + (\mathrm{E}(\hat{\Theta}) - \theta)^2 \\
&= \mathrm{Var}(\hat{\Theta}) + \text{bias}^2
\end{aligned}$$

  Note: an unbiased estimator has MSE = Var

Sampling distributions

- Point estimators, $\hat{\Theta}$, are random variables
- Therefore, they have a mean and variance (above)
- But also a probability distribution
- The probability distribution of a point estimator is known as the sampling distribution—the type of distribution depends on the nature of the underlying population, sample size n, the estimator itself

Sample mean

Under conditions of the central limit theorem, the sample mean, $\bar{X}$ is normally distributed,

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

Sample variance

The sample variance is χ^2 distributed,

$$\frac{(n-1)\ S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Sample mean, unknown σ^2

If the population variance is unknown, it is substituted by the sample variance, a χ^2-distributed random variable. Hence, the sample mean takes the form of a t-distribution,

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

Interval estimates

- The confidence of an interval is not the probability that a calculated interval around the sample mean includes the population mean; that probability is either 0 or 1, since the interval is already calculated—"the die has been cast"
- Rather, it can be understood as the proportion of calculated intervals that will contain the population mean

Hypothesis testing

- X, a sample, can be defined as
$$X = \mu + \epsilon$$
where ϵ is a random variable representing 'random disturbance' due to any sources of variability in the sampling method or population
- Therefore, there is an uncertainty that comes with estimating μ with $\bar{X}$—is the value obtained, $\bar{x}$ the population mean, μ? If not, how close are $\bar{x}$ and μ?
- Since sample variance is practically non-zero, then we are unable to make the correct conclusions every time
- We can only hope to 'be correct most of the time'
- The aim of hypothesis testing is 'to know if the mean of the unknown population has a value of μ_0, with only a single small sample available
- i.e. to make a general conclusion about the population based on specific observations from the sample
- H_0 must be specific, H_1 is usually non-specific
- The significance level α, determines the values of $\bar{x}$ where H_0 is rejected
- The p-value is the probability of obtaining a result equal to or more severe than what was obtained, given that H_0 is true
- For a two-tail test, the p-value is $2 \times \mathrm{P}(\bar{X} \leq \bar{x}$ or $2 \times \mathrm{P}(\bar{X} \geq \bar{x}$, whichever is smaller

Errors in hypothesis testing

- Type I error: rejecting H_0 when H_0 is true
- The probability of committing a type I error is the significance level, α of the test
- Type II error: failing to reject H_0 when H_1 is true
- Power: probability of not committing a type II error, i.e.
$$\text{Power} = (1 - \beta)$$
Power is also the probability of correctly rejecting a false null hypothesis
- Power is a measure of specificity—the ability of a test to detect differences
- Then, there is a specific difference to detect; i.e. the difference between $\mu_0 = 50; \mu_1 = 52$ is harder to detect than the difference between $\mu_0 = 50; \mu_1 = 60$
- Therefore, to calculate the probability of a type II error, one must specify what is the magnitude of difference to detect
- Fixing the sample size, n, decreasing the probability of one type of error increases the probability of the other

Reducing probability of errors in hypothesis testing

- To increase the power of a test, i.e. decrease the probability of a type II error, define a larger difference that you would like to reliability detect
- Alternatively, increase the sample size
- The probability of a type I error is usually defined by the tester as the level of significance, α

Examples

Calculating the 95% confidence interval for sample mean

$$P\left(\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Calculating the 95% confidence interval for sample mean with unknown σ^2,

$$P\left(\bar{X} - t_{0.025,n-1} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{0.025,n-1} \cdot \frac{s}{\sqrt{n}}\right) = 0.95$$

Hypothesis testing on the mean

1. Identify the parameter of interest from the problem context
2. State the null hypothesis, H_0
$$\mathrm{H}_0 : \mu = \mu_0 = 60$$
3. State an appropriate alternative hypothesis, H_1
$$\mathrm{H}_1 : \mu \neq 60$$
4. Choose a significance level, α
$$\alpha = 0.05$$
5. State an appropriate test statistic; usually Z if population variance is known, otherwise t
$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
6. State and compute the rejection region for the statistic, e.g. for $\sigma^2 = 100$ and $n = 10$
$$P(z_{\alpha=0.025} \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha=0.975}) = 0.95$$
$$P(-1.96 \cdot \frac{10}{\sqrt{10}} + 60 \leq \bar{X} \leq 1.96 \cdot \frac{10}{\sqrt{10}} + 60) = 0.95$$
$$P(53.8 \leq \bar{X} \leq 66.2) = 0.95$$
7. Decide if H_0 should be rejected, and report it in the problem context

The amount of time for a lab to process samples is a RV with mean 9.2 mins and standard deviation 1.8 mins. Suppose a random sample of $n = 48$ is collected. Find the probability that the average processing time for these samples is (a) less than 10 mins; (b) between 9 and 10 mins.

let $\bar{X} = \frac{1}{48}\sum_{i=1}^{48} X_i$,

then, by the central limit theorem,

$\bar{X} \sim \mathrm{N}(9.2,\ 1.8^2/48)$

$$\begin{aligned}
&\mathrm{P}(\bar{X} \leq 10) \\
&= \mathrm{P}\left(\frac{\bar{X} - 9.2}{\sqrt{1.8^2/48}} \leq \frac{10 - 9.2}{\sqrt{1.8^2/48}}\right) \\
&= \Phi\left(\frac{10 - 9.2}{\sqrt{1.8^2/48}}\right) \\
&= \Phi(3.07) \\
&= 0.99893 \approx 0.999
\end{aligned}$$

$$\begin{aligned}
&\mathrm{P}(9 \leq \bar{X} \leq 10) \\
&= \Phi\left(\frac{10 - 9.2}{\sqrt{1.8^2/48}}\right) - \Phi\left(\frac{9 - 9.2}{\sqrt{1.8^2/48}}\right) \\
&= \Phi(3.07) - \Phi(-0.7698) \\
&= 0.99893 - 0.220650 \\
&= 0.77828 \approx 0.778
\end{aligned}$$

Find the mean square error for $\hat{\Theta} = 1/2 \cdot (2X_1 - X_6 + X_4)$, for a random sample $X_1, X_2, \cdots, X_6$

$$\begin{aligned}
\text{MSE} &= \text{Std. Error}^2 + \text{Bias}^2 \\
&= \mathrm{Var}(X_1 + 1/4 \cdot X_6 + 1/4 \cdot X_4) + \\
&\quad (1/2 \cdot \mathrm{E}(2X_1 - X_6 + X_4) - \mu)^2 \\
&= 3/2 \cdot \sigma^2 + (1/2 \cdot (2\mu) - \mu)^2 \\
&= 3/2 \cdot \sigma^2
\end{aligned}$$