TEMPLATE EXAM

MODULE CODE: MA4413

SEMESTER: Autumn 2015

MODULE TITLE: Statistics for Computing

DURATION OF EXAM: 2.5 hours

LECTURER: Dr. Kevin Burke

GRADING SCHEME: 100 marks (60% of module)

INSTRUCTIONS TO CANDIDATE

- **Attempt four** of the six questions (each one carries 25 marks).
- All work must be shown *clearly and logically* using appropriate symbols and probability notation. Failure to do so will *lose marks*.
- Write down the formula you intend to use at each stage *before* filling it in with numbers.
- Formula sheets are provided at the back of this exam paper.
- Statistical tables are available from the invigilators.

Note: Exam questions are based on the examples and questions from lectures and tutorials. You have the solutions to all questions on this course. Below is the layout of the exam in terms of the lectures (you can find the corresponding tutorial questions yourself). Lec 5 (counting techniques) and Lec 17 (chi-squared test) are not examinable this year.

Question 1

a) Probability basics (Lec 3 and 4)

b) Boxplots / numerical summaries (Lec 2)

c) Hypothesis test / confidence interval for one parameter, i.e., μ or p (Lec 13, 14, 15 and 16)

Question 2

a) Histogram / numerical summaries (Lec 1 and 2)

b) Statistic and parameter / numerical summaries (Lec 1 and Lec 2)

c) Probability (law of total probability) (Lec 3 and 4)

Question 3

a) Numerical summaries (Lec 2)

b) Confidence interval based on large sample(s) (Lec 13)

c) Hypothesis test based on large sample(s) (Lec 15 and 16)

Question 4

a) Random variables (Lec 6)

b) Binomial (Lec 7)

c) Poisson and Exponential (Lec 8 and 9) but *not* queueing theory here

Question 5

a) Huffman coding and entropy (Lec 18)

b) Normal distribution (Lec 10, 11 and 12)

Question 6

a) Confidence interval or hypothesis test for difference between two means (i.e., $\mu_1 - \mu_2$) based on small samples (Lec 14 and 16)

b) Queueing theory (Lec 8 and 9)

Useful Formulae: Page 1

Histogram:

- class width $= \dfrac{\max(x) - \min(x)}{\text{number of classes}}$

Numerical Summaries:

- $\bar{x} = \dfrac{\sum x_i}{n}$
- $s^2 = \dfrac{\sum x_i^2 - n\,\bar{x}^2}{n-1}$
- Position of Q_k : $\dfrac{n+1}{4} \times k$
- $IQR = Q_3 - Q_1$
- $LF = Q_1 - 1.5 \times IQR$
- $UF = Q_3 + 1.5 \times IQR$

Probability:

- $\Pr(A^c) = 1 - \Pr(A)$
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- $\Pr(E_1 \cup E_2 \cup \cdots \cup E_k) = \Pr(E_1) + \Pr(E_2) + \cdots + \Pr(E_k)$ (if mutually exclusive)
- $\Pr(A \cap B) = \Pr(A)\ \Pr(B \mid A) = \Pr(B)\ \Pr(A \mid B)$
- $\Pr(E_1 \cap E_2 \cap \cdots \cap E_k) = \Pr(E_1)\ \Pr(E_2)\ \cdots\ \Pr(E_k)$ (if independent)
- $\Pr(A \mid B) = \dfrac{\Pr(A \cap B)}{\Pr(B)} = \dfrac{\Pr(A)\ \Pr(B \mid A)}{\Pr(B)}$
- $$\begin{aligned}\Pr(B) &= \Pr(B \cap E_1) + \Pr(B \cap E_2) + \cdots + \Pr(B \cap E_k) \\ &= \Pr(E_1)\ \Pr(B \mid E_1) + \Pr(E_2)\ \Pr(B \mid E_2) + \cdots + \Pr(E_k)\ \Pr(B \mid E_k)\end{aligned}$$

  (if $E_1, \ldots, E_k$ are mutually exclusive & exhaustive)

Useful Formulae: Page 2

Counting Techniques:

- $n\,! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1$
- $\binom{n}{k} = \dfrac{n\,!}{k\,!\,(n-k)\,!}$

Random Variables:

- $E(X) = \sum x_i \; p(x_i)$
- $E(X^2) = \sum x_i^2 \; p(x_i)$
- $Var(X) = E(X^2) - [E(X)]^2$
- $Sd(X) = \sqrt{Var(X)}$

Distributions:

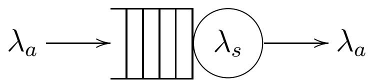
• $X \sim \text{Binomial}(n, p)$	• $X \sim \text{Poisson}(\lambda)$	• $T \sim \text{Exponential}(\lambda)$
• $\Pr(X = x) = \binom{n}{x} p^x \, (1-p)^{n-x}$	• $\Pr(X = x) = \dfrac{\lambda^x}{x\,!} \, e^{-\lambda}$	• $\Pr(T > t) = e^{-\lambda\, t}$
• $x \in \{0, 1, 2, \ldots, n\}$	• $x \in \{0, 1, 2, \ldots, \infty\}$	• $t \in [0, \, \infty)$
• $E(X) = n\, p$	• $E(X) = \lambda$	• $E(T) = \dfrac{1}{\lambda}$
• $Var(X) = n\, p\, (1-p)$	• $Var(X) = \lambda$	• $Var(T) = \dfrac{1}{\lambda^2}$

Note: the normal distribution is shown on the next page

Useful Formulae: Page 3

Queueing Theory:

- $E(N) = \lambda_a \, E(T)$
- $\rho = \dfrac{\lambda_a}{\lambda_s}$
- $M/M/1$ System:



$$\Rightarrow \; T \sim \text{Exponential}(\lambda_s - \lambda_a)$$

(where T is the total time in the system)

Normal Distribution:

- $X \sim \text{Normal}(\mu, \sigma)$
- $E(X) = \mu$
- $Var(X) = \sigma^2$
- $(1-\alpha)100\%$ of the Normal(μ, σ) distribution lies in $\mu \pm z_{\alpha/2}\, \sigma$
- $\Pr(X > x) = \Pr\left(Z > \dfrac{x-\mu}{\sigma}\right)$
- $\Pr(Z < -z) = \Pr(Z > z)$
- $\Pr(Z > -z) = \Pr(Z < z) = 1 - \Pr(Z > z)$
- If $X_1 \sim \text{Normal}(\mu_1, \sigma_1)$ and $X_2 \sim \text{Normal}(\mu_2, \sigma_2)$

$$\Rightarrow \quad \text{Sum:} \quad X_1 + X_2 \sim \text{Normal}\left(\mu_1 + \mu_2,\, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

$$\Rightarrow \quad \text{Difference:} \quad X_1 - X_2 \sim \text{Normal}\left(\mu_1 - \mu_2,\, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

- For $X_1, \ldots, X_n \sim$ any distribution with $\mu = E(X)$ and $\sigma = Sd(X) = \sqrt{Var(X)}$

$$\Rightarrow \quad \text{Sample mean:} \quad \overline{X} \sim \text{Normal}\left(\mu,\, \frac{\sigma}{\sqrt{n}}\right) \quad \text{if } n > 30$$

Useful Formulae: Page 4

Statistics and Standard Errors:

Parameter	Statistic	Standard Error	Samples	Details
μ	$\bar{x}$	$\dfrac{s}{\sqrt{n}}$	large / small	$\nu = n - 1$
p	$\hat{p}$	$\sqrt{\dfrac{\hat{p}\,(1-\hat{p})}{n}}$	large	confidence interval
		$\sqrt{\dfrac{p_0\,(1-p_0)}{n}}$	large	hypothesis test
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$	large / small	$\nu = \dfrac{(a+b)^2}{\frac{a^2}{n_1-1} + \frac{b^2}{n_2-1}}$ $a = \dfrac{s_1^2}{n_1}, \quad b = \dfrac{s_2^2}{n_2}$
		$\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}$ where $s_p^2 = \dfrac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1+n_2-2}$	small	$\nu = n_1 + n_2 - 2$ assuming $\sigma_1^2 = \sigma_2^2$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\dfrac{\hat{p}_1\,(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2\,(1-\hat{p}_2)}{n_2}}$	large	confidence interval
		$\sqrt{\dfrac{\hat{p}_c\,(1-\hat{p}_c)}{n_1} + \dfrac{\hat{p}_c\,(1-\hat{p}_c)}{n_2}}$ where $\hat{p}_c = \dfrac{x_1+x_2}{n_1+n_2}$	large	hypothesis test

Confidence Intervals:

- Large sample: statistic $\pm\ z_{\alpha/2} \times$ standard error
- Small sample: statistic $\pm\ t_{\nu,\alpha/2} \times$ standard error

Useful Formulae: Page 5

Hypothesis Testing:

- $$z = \frac{\text{statistic} - \text{hypothesised value}}{\text{standard error}}$$

- $$\text{p-value} = \begin{cases} 2 \times \Pr(Z > |z|) & \text{if } H_a : \mu \neq \mu_0 \\ \Pr(Z < z) & \text{if } H_a : \mu < \mu_0 \\ \Pr(Z > z) & \text{if } H_a : \mu > \mu_0 \end{cases}$$

- $$F = \frac{\text{larger variance}}{\text{smaller variance}} = \frac{s^2_{\text{larger}}}{s^2_{\text{smaller}}}$$

  $$\nu_1 = n_{\text{top}} - 1, \quad \nu_2 = n_{\text{bottom}} - 1$$

- $$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

Goodness-of-fit: $e_i = \text{total} \times p(x_i), \qquad \nu = n_f - 1 - k$

Independence: $e_{ij} = \dfrac{r_i \times c_j}{\text{total}}, \qquad \nu = (n_r - 1) \times (n_c - 1)$

Information Theory:

- $$h(x) = -\log_2[p(x)]$$

- $$H(X) = E[h(X)] = \sum h(x_i)\, p(x_i)$$

- $l(x_i)$ = code-length for character x_i

- $$E(L) = \sum l(x_i)\, p(x_i)$$

- $$e = \frac{H(X)}{E(L)}$$

- $$\sum 2^{-l(x_i)} \leq 1$$