FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF MATHEMATICS AND STATISTICS

END OF SEMESTER EXAMINATION

MODULE CODE: MA4413

SEMESTER: Autumn 2013

MODULE TITLE: Statistics for Computing

DURATION OF EXAM: 2.5 hours

LECTURER: Kevin O'Brien

GRADING SCHEME: 100 marks
70% of total module marks

EXTERNAL EXAMINER: Prof. Brendan Murphy

INSTRUCTIONS TO CANDIDATES

This paper is comprised of five questions, each worth 25 marks. Attempt any four questions.
Scientific calculators approved by the University of Limerick can be used.
Formula sheet and statistical tables are provided.

Formulae

Probability

- Conditional probability:
$$P(B|A) = \frac{P\left(A \text{ and } B\right)}{P\left(A\right)}.$$

- Bayes' Theorem:
$$P(B|A) = \frac{P\left(A|B\right) \times P(B)}{P\left(A\right)}.$$

- Binomial probability distribution:
$$P(X = k) =^n C_k \times p^k \times (1-p)^{n-k} \qquad \left(\text{where} \qquad {}^nC_k = \frac{n!}{k!\,(n-k)!}.\right)$$

- Poisson probability distribution:
$$P(X = k) = \frac{m^k \mathrm{e}^{-m}}{k!}.$$

Information Theory

- $I(p) = -log_2(p) = log_2(1/p)$

- $I(pq) = I(p) + I(q)$

- $H = -\sum_{i=1}^{m} p_i \, log_2(p_i)$

- $E(L) = \sum_{i=1}^{m} l_i p_i$

- Efficiency $= H/E(L)$

- $I(X;Y) = H(X) - H(X|Y)$

- $P(C[r]) = \sum_{j=1}^{m} P(C[r]|Y = d_j)P(Y = d_j)$
- $R = rH(X)$ (b/second)

Confidence Intervals

One sample

$$
\begin{aligned}
S.E.(\bar{X}) &= \frac{\sigma}{\sqrt{n}}. \\
S.E.(\hat{P}) &= \sqrt{\frac{\hat{p} \times (100 - \hat{p})}{n}}.
\end{aligned}
$$

Two samples

$$
\begin{aligned}
S.E.(\bar{X}_1 - \bar{X}_2) &= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \\
S.E.(\hat{P}_1 - \hat{P}_2) &= \sqrt{\frac{\hat{p}_1 \times (100 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \times (100 - \hat{p}_2)}{n_2}}.
\end{aligned}
$$

Hypothesis tests

One sample

$$
\begin{aligned}
S.E.(\bar{X}) &= \frac{\sigma}{\sqrt{n}}. \\
S.E.(\pi) &= \sqrt{\frac{\pi \times (100 - \pi)}{n}}
\end{aligned}
$$

Two large independent samples

$$
\begin{aligned}
S.E.(\bar{X}_1 - \bar{X}_2) &= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \\
S.E.(\hat{P}_1 - \hat{P}_2) &= \sqrt{(\bar{p} \times (100 - \bar{p}))\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.
\end{aligned}
$$

Two small independent samples

$$
\begin{aligned}
S.E.(\bar{X}_1 - \bar{X}_2) &= \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \\
s_p^2 &= \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}.
\end{aligned}
$$

Paired sample

$$
S.E.(\bar{d}) = \frac{s_d}{\sqrt{n}}.
$$

Standard deviation of case-wise differences

$$s_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}.$$

Binary Classification

- $F = 2 \times \frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$