McGill University
Department of Economics
Economics 257D

December examination, 2011

John W. Galbraith — Associate Examiner: V. Zinde-Walsh

You have three hours in which to complete this exam. If you find any question ambiguous, explain the difficulty, make a reasonable assumption about the intent, and go on. **Show your calculations and reasoning in detail.**

I will come to the examination room once, so please read the exam carefully right away so that if you have any questions, you can ask me while I am there.

Calculators which do not store text or formulae are OK.

Attached you will find tables of the standard normal, t- and χ^2 distributions. If you need to interpolate, do it roughly and indicate where you obtained your numbers, but I don't expect great accuracy given the time constraint.

1. (25 marks) **Define, explain and discuss** each of the following:
   (a) mean squared error, bias and variance of an estimator
   (b) the relationship among the quantities in (a)
   (c) Type I error
   (d) the Tchebychev (or Chebychev...) inequality
   (e) symmetry and skewness

2. (5+5+5+10 marks)

A random variable X provides an observable current indicator of whether or not the economy has entered recession. When $X > 0$, the economy is in recession 5% of the time; when $X < 0$ the economy is in recession 60% of the time. X is never exactly zero. Overall the economy is in recession 15% of the time.

a) Find the unconditional probability that X is negative.

b) Find the probability that X is positive if the economy is in recession (note: not 5%).

c) Find the probability that X is negative if the economy is in recession.

d) Find the probabilities that X is negative or positive if the economy is not in recession. Interpret the results.

3. (5+5+10+5 marks) You are given data on 40 years of monthly claim payouts (480 sample points) from an insurance company, having a sample mean of 10.2 and a sample variance of 19.0. (The data are measured in millions of constant dollars,

adjusted for inflation.) All of the observations are positive. You do not know the form of the distribution, but it is skewed right (long upper tail).

a) Give the approximate distribution of the sample mean of the payouts and give a 99% confidence interval for the true mean.

b) On these 480 monthly data points (corresponding with 120 three-month periods or 'quarters', so the company has to report quarterly results 120 times over this period), the company takes a loss 20 times out of 120. Give an approximate 99% confidence interval for the proportion of the times the company takes a loss. Would the hypothesis that the proportion of three-month periods showing a loss is 25% or more be rejected at the 99% level? (You don't have to give a p-value, just explain whether it would be rejected or not.)

c) Someone suggests that a χ^2_{10} distribution provides a reasonable fit to these data. If that is true, approximately what proportion of the time should monthly claim payouts exceed \$20 (million)? (Since you will not find the exact value that you need in a table, interpolate roughly to give an answer–again, I don't expect great accuracy).

d) Another insurance company has the same distribution of claim payouts as the first, but operates in a distant geographical area, so that its payouts are independent. If these two companies had merged 40 years ago, what would the sample variance of payouts of the larger joint company have been, other things equal? Why isn't this smaller than the single-company variance? Interpret your answer in terms of risk sharing.

4. (5+10+10 marks) A health economist is interested in whether prescription drug use depends on whether an individual has a health plan that pays drug costs, or not. She has a sample of 400 people who have a health plan that pays these costs (I'll call these the 'haves'), and an independent sample of 324 people who do not (I'll call these the 'havenots').

In the sample of haves, mean drug expenditure per annum was \$420 and the standard error of expenditure was \$150; in the sample of havenots, mean expenditure was \$340 and the standard error of expenditure was \$160.

(a) Give 95 per cent confidence intervals for the mean expenditure on prescription drugs for both the haves and the havenots.

(b) Give a 95 percent confidence interval for the difference in mean expenditure between the two populations (ie haves and havenots).

(c) Test the hypothesis that mean drug expenditure is the same in have and havenot populations (give an approximate p-value). Interpret the result and explain the relation to the result of (b).

McGill University
Department of Economics
Economics 257: Honours Statistics
December examination 2010

You have 3 hours (+ 10 minutes reading time) in which to complete this exam. If you find any question ambiguous, explain the difficulty, make a reasonable assumption about the intent, and go on. All questions 1-5 have equal weight (i.e. 20%).

Please define notation clearly and show your reasoning in detail.

1. Define, explain and discuss the importance of each of the following:
   (a) the Chebychev inequality
   (b) mean squared error
   (c) standardization
   (d) the fourth moment

2. The number of alcoholic drinks consumed per week by undergrads at South Beach U has a mean of 22. and std deviation of 4. Treat the population distribution as Normal. A random sample of students is taken.

a) find the probability of obtaining the misleading result that the (sample) mean number of drinks is less than or equal to 14 per week if one samples (i) 1, (ii) 4, (iii) 25 students.

b) explain why the three answers differ, and illustrate with a graph.

c) if you did not know that the population was normal, explain conditions under which you could give approximate answers to each part of question (a). Compute your approximate answers.

3. A pollster surveys n randomly sampled women, and another n randomly sampled men, asking them if they're going to vote yes or no, or have not decided, on a referendum to require a balanced government budget. Among women, 48% say 'yes'; among men 54% say 'yes'. Assume that all respondents give answers that genuinely reflect their intentions.

a) Find 95% confidence intervals for the proportions of men and women who will vote yes.

b) Assume that there are equal numbers of registered male and female voters and that an equal proportion will vote. Find a 95% confidence interval for the overall proportion of yes votes.

c) Test the hypothesis that the proportions of male and female voters who will vote yes is in fact the same. Give a p-value. Interpret the result.

4. You have set up your own insurance company to offer insurance to computer owners. You find that every year in Montreal, 1.5% of computer users have their computers stolen. They have a mean insured value of \$2000; you therefore decide to set premiums to make a small profit, at \$35 per \$2000. of insured value. You plan to take on 10,000 customers, which you are able to do, so your premium income is \$350,000.

(a) When making your business plan, you incorrectly assumed that your total claim payouts would be a normally distributed random variable with a mean of \$300,000 ($.015 \times 2000 \times 10{,}000$ customers) and a standard deviation of 24,000. What did you *think* the probability of a loss was, given that your premium income was to be \$350,000?

(b) A client who takes careful precautions against theft has a theft probability of 1%; one who does not do so has a theft probability of 2%. Half of your clients are of each type, but you cannot tell which is which. A particular client takes out insurance, and after one year reports no theft; what is the probability that this is a client who takes careful precautions to avoid being robbed?

5. Among a group of 200 laid-off factory workers, 100 are randomly selected (ie there is no self-selection bias) to participate in an online course to upgrade their basic job skills. Within one year, all have found new jobs. Among those who participated, the distribution of earnings in the new jobs has a (sample) mean of \$38,000, a standard error of earnings of \$3000, and a positive third moment. Among those who did not participate, the (sample) mean is \$34,000, standard error of earnings is \$2800, and again the third moment is positive. The workers have varying lengths of working experience; assume that their length of working experience comes from a population distribution with a mean of 18 years and a standard deviation of 4 years; treat this length-of-experience variable as normally distributed.

a) test the hypothesis that mean earnings are in fact unaffected by the skill-upgrading course. Give a p-value and indicate clearly the method you are using.

b) Give a 95% confidence interval for the population standard deviation of years of working experience for each of the groups of workers.

McGill University
Department of Economics
Economics 257D

December examination 2006

John W. Galbraith — Associate Examiner: D. Sutthiphisal

You have three hours in which to complete this exam. If you find any question ambiguous, explain the difficulty, make a reasonable assumption about the intent, and go on. All questions have equal weight.

Calculators which do not store text or formulae are OK.

Attached you will find tables of the standard normal, t- and χ^2 distributions. If you need to interpolate, do it roughly and indicate where you obtained your numbers, but I don't expect great precision given the time constraint.

1. Define, explain and discuss each of the following:
   (a) permutations vs. combinations
   (b) percentiles and median
   (c) the exponential distribution
   (d) loss function of an estimator, with specific examples
   (e) the weak law of large numbers

2. An international development agency introduces a micro-credit program to encourage the growth of new businesses in a developing country. There are six hundred applicants. One third of applicants are under 30, half are 31-50, and the remaining sixth are over 50 years of age. From among the applicants, candidates are chosen to receive loans as follows: of applicants under 30, 25% are randomly selected to receive loans; of applicants 30-50, 40% are randomly selected to receive loans; of those over 50, 20% are randomly selected to receive loans. Of those who receive loans, 72% established, and continued for at least two years, a business which employed at least one other person (we will refer to this condition as 'successfully started a business' hereafter). Of those who did not receive the loans, 66% successfully started a business.

Despite the planners' apparent ideas, people in each age group are equally able to start a business successfully given the same conditions (eg those who get a loan and are under 30, and those who get a loan and are over 50, are equally likely to succeed).

After the two years, prizes are given for the three most successful businesses.

Find each of the following probabilities, using a formal notation to indicate the method that you used:

a) probability that a randomly-selected applicant receives a loan
b) probability that an applicant who successfully started a business received a loan
c) probability that an applicant who did not successfully start a business received a loan
d) probability that each of the three most successful businesses was begun by someone over 50

3. You are playing a game at a casino which has the feature that your probability of winning is 0.49, of losing is 0.51 (no draws). Each time you bet \$1, you get \$2 if you win, \$0. if you lose (so your net gain is \$ +1 or -1). After you have bet n times, we will say that you are a net winner if your total (or average) net gain is positive (ie you win more than 50% of the time), net loser if your total net gain is negative. (e.g. if you play three times, win twice and lose once, your net gain is +1+1-1 = 1 >0, so you are a net winner).

a) Find the mean and variance of your net gain, or 'payoff', on each bet.

b) Find the mean and variance of your *average* payoff on n bets.

c) State the asymptotic distribution of your average payoff.

d) Using c), find the approximate probability of being a net winner after playing 5, 20 and 500 times.

e) Interpret d) using the weak law of large numbers.

4. You sample 400 people who own cars and estimate that their average annual expenditure on fuel is \$1200. The standard error that you estimate is \$300. You don't know the form of the distribution of fuel expenditure.

a) Treat the estimates of the mean and standard error as if they are equal to the true values for this part. What proportion of people can you conclude spend between \$600 and \$1800 per year?

b) What are 90% and 99% confidence intervals for average fuel expenditure?

c) You read somewhere that average fuel consumption among car owners is \$1100. Does this sound reasonable given your sample? Why or why not?

d) If the individual observations on fuel expenditure could be treated as normally distributed, what would be a 90% confidence interval for the true variance of the data?

McGill University
Department of Economics
Economics 154-257D

December examination

John W. Galbraith December 2004

You have three hours in which to complete this exam. If you find any question ambiguous, explain the difficulty, make a reasonable assumption about the intent, and go on. The weight of each question is given out of 100.

Calculators which do not store text or formulae are OK; however, it is acceptable to leave answers in unsimplified form if you have no calculator.

Attached you will find tables of the standard normal, $t-$ and χ^2 distributions. If you need to interpolate, do it roughly and indicate where you obtained your numbers, but I don't expect tremendous accuracy given the time constraint.

1. [25 marks] Define, explain and discuss each of the following:

(a) an unbiased estimator

(b) mean squared error of an estimator

(c) the relation between bias and mean squared error of an estimator [this does not require a definition, just a statement and discussion]

(d) binomial distribution

(e) the weak law of large numbers

2. [35 marks] You have set up your own insurance company to offer insurance to computer owners. You find that every year in Montreal, 1.5% of computer users have their computers stolen. They have a mean insured value of $2000; you therefore decide to set premiums to make a small profit, at $35 per $2000. of insured value. You plan to take on 10,000 customers, which you are able to do, so your premium income is $350,000.

(a) In your first year, 1.5% of your clients report a theft. However, to your surprise, your claims (payouts) exceed 1.5% of average insured value, and you take a loss. What must be true of the distribution of values of stolen computers versus the overall distribution of computer values? Why might this arise?

(b) When making your business plan, you incorrectly assumed that your total claim payouts would be a normally distributed random variable with a mean of $300,000 (.015 × 2000 × 10,000 customers) and a standard deviation of 24,000. What did you *think* the probability of a loss was, given that your premium income was to be $350,000?

(c) A client who takes careful precautions against theft has a theft probability of 1%; one who does not do so has a theft probability of 2%. Half of your clients are of each type, but you cannot tell which is which. A particular client takes out insurance, and after one year reports no theft; what is the probability that this is a client who takes careful precautions to avoid being robbed?

d) Explain verbally how you might use statistical measures to adjust each individual's premium each year in order to increase your profits.

3. [40 marks] You sample 100 (currently-employed) people randomly, and ask them (i) how much they spend on employment-related expenses, such as transportation to work, each year (not re-imbursed by the employer), and (ii) whether they have ever been unemployed for a continuous period of eight months or more. You find (i) an average of $675. (std. dev. 200) and (ii) 31 respondents have had such an unemployment spell, in your sample, and you find that employment expenses are distributed asymmetrically, with a long upper (right) tail.

(a) What can you conclude about the proportion of this population having employment expenses between $275 and $1075?

(b) Give a 95 per cent confidence interval for the *population mean* of these employment expenses.

(c) Give a 95 per cent confidence interval for the *proportion of the population* that has had an unemployment spell of eight months or more.

(d) You feel that these confidence intervals are too wide to be useful. How many people would you need to sample in order to have confidence intervals one-quarter as wide as those in b) and c) above, assuming that the next people sampled have roughly the same characteristics as those of the first hundred? Explain.