Sample Questions Statistics and Econometrics

European University Institute

Tobias Broer

These questions are meant to give incoming students an idea about the level of difficulty in the introductory part of the Statistics and Econometrics Sequence. You are not meant to be able to answer all questions. But if you feel you don't know how you could approach the majority of the questions (without necessarily getting the right answer straight away), you should consult some introductory text in Statistics and Econometrics, such as Goldberger, A.S., *A Course in Econometrics*, Cambridge and London: HUP 1991.

Question 1: Probability set theory

Using set algebra (union, intersection, complementation), show that

- $P(C1 \cap C2) \leq P(C1) \leq P(C1 \cup C2) \leq P(C1) + P(C2)$
- $P(C1 \cap C2) \geq P(C1) + P(C2) - 1$ (Bonferroni's Inequality)

Question 2: Probability spaces, combinatorics, and simple random variables

A box contains three balls, numbered from 1 to 3. The probability of drawing a particular ball is the same for all balls in the box.

Consider the random experiment that draws a single ball from the box. What is the sample space Ω (the set of all possible outcomes) of the experiment? Define in your own words the meaning of the word "event". Write down the set of all possible events (also called a $\sigma - algebra$ Σ of Ω, defined here as the set of all possible events including the "sure" and "impossible" events).

Consider now the random experiment: a first ball is drawn from the box and its number written down. Then, the ball is replaced and another ball is drawn, its number written down. What is the sample space associated with this new experiment? Give a general formula for the number of outcomes (equal to the elements of the sample space) when from a box containing $n > r$ numbered balls r balls are drawn sequentially (i.e. the numbers of drawn balls are written down in the order of draws from 1 to r), and with replacement of the balls after each draw.

Consider now the random experiment: 2 balls are drawn from the box sequentially, without replacement of the first ball drawn. What is the sample space now? What is a general formula for the number of possible outcomes when drawing r from $n > r$ balls without replacement?

For the rest of this question, consider two draws from a box with three balls (n=3, r=2), ordered, with replacement. What is the probability of the following events

- A: "the ball with number 1 is drawn first"
- B: "the ball with number 1 is drawn twice"
- C: "the ball with number 1 is drawn once"
- D: "the ball with number 1 is drawn at least once"
- E: "the ball with number 2 is not drawn"

What is the probability of "not C"? What is the probability of the event "A or E or both". What is the probability of event A conditional on event C occurring? What is the probability of event B conditional on event D occurring?

Suppose now you calculated, after each trial of the experiment, the sum of the numbers shown on the balls. This establishes a mapping, call it S for "sum", from the sample space Ω of the random experiment into the set of natural numbers $\mathbb{N}$. What is the probability of S=3? What is the (cumulative) probability distribution function of S? What is its expectation and what is its variance?

Question 3: Distribution functions

For each of the following, find the constant c so that f(x) satisfies the conditions of a probability density or probability mass function of a random variable

- $f(x) = c(\frac{2}{3})^x$ for x=1,2,3,... , and $f(x) = 0$ elsewhere
- $f(x) = cxe^{-x}$ for $0 < x < \infty$, and $f(x) = 0$ elsewhere

Question 4: Moments of distribution functions

For each of the following probability density or probability mass functions, derive the expectation and variance of the random variable X

- **The uniform distribution** $f_X(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, $f_X(x) = 0$ elsewhere
- **The Poisson distribution** $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 1, 2, 3, ...$, $f_X(x) = 0$ elsewhere. (Difficult! The answer is E(X)=Var(X)=λ. Hint: To show this use the definition of the exponential function from its MacLaurin Series - its Taylor expansion around zero - as $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$)

Question 5: Decomposition of variance

Consider two jointly distributed random variables X and Y. Using the simple law of iterated expectations, i.e. $E_X(E_Y(Y|X = x)) = E_{YX}(Y)$ where $E_Z()$ takes expectations of a random variable with respect to Z, show that
$Var(Y) = E_X(Var(Y|X = x)) + Var_X(E_Y(Y|X = x))$.

Question 6: Statistics as estimators of moments of distributions

Suppose you have n independent observations $x_1, ..., x_n$ on a random variable X. Consider the mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ as an estimator of the expectation E(X). Is $\overline{x}$ an *unbiased* estimator for the expectation E(X)? Can you derive the variance of it's sampling distribution in terms of the variance of X and the number of observations n?
Now consider the simple variance of the observations $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$
Is s_x^2 unbiased for the variance var(X)? (Difficult.)

Question 7: Classical linear regression model and OLS sampling distribution

Suppose you have n observations on Y and X and estimate by Ordinary Least Squares (OLS) the coefficients α and β in the model $Y = \alpha + X\beta + \epsilon$, where ϵ is a random variable with finite variance, alpha is a scalar, beta a k x 1 vector of coefficients, and X a 1 x k vector of regressors.

Under what assumptions are the OLS estimates *unbiased* for α and β?
Under what assumptions do the OLS estimates have minimum variance among

all linear estimators for α and β (i.e. OLS is BLUE in the jargon of econometricians)?

What assumptions or qualifications do you have to add for the following statement to make sense, and to be true: "The t-statistic of β_1 is 2, so the first variable x_1 is significant."