Econ. 517, Econometric Theory I Fall 2001 Chris Sims

Midterm Exam Answers*

The exam consists of 9 questions. You are expected to answer all of them. They are all short. All are alloted 10 points except numbers 4 (5 points) and 9 (15 points). There are a total of 90 points, corresponding to the 90 minutes you have for the exam. Be sure to try all questions before you spend disproportionate time on any one question.

(1) (10 points) Suppose x has a Cauchy distribution, meaning it has pdf

$$\frac{1}{\pi}\frac{1}{1+x^2}\,.$$

If we construct a Taylor approximation to the log of this pdf at its peak, what is the variance of the resulting normal approximation? How does it compare to the actual variance of this pdf?

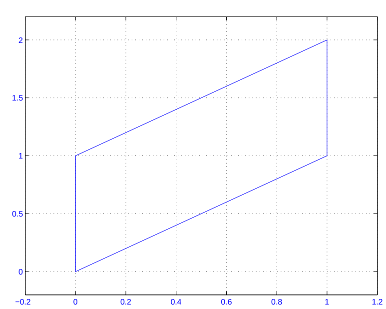
The second derivative of the log pdf is

$$\frac{-2}{1+x^2}+\frac{4x^2}{(1+x^2)^2}\,.$$

Evaluated at the peak of the pdf ($x = 0$) this becomes just -2. The second order Taylor expansion of the log pdf about the peak is thus just $-x^2$, and exponentiating this delivers e^{-x^2}. When normalized to integrate to one, this would be a $N(0, \frac{1}{2})$ pdf. The Cauchy distribution has infinite variance, which is easily seen from the fact that $x^2/(1+x^2)$ does not go to zero as $x \to \infty$. So the "comparison" requested just requires noting that the approximation has finite variance and the original pdf has infinite variance.

(2) (10 points) The marginal distribution of X has pdf $2x$ on $(0,1)$ and the conditional pdf of $Y \mid X$ is $2(y-x)$ on $(x, x+1)$. Find expressions for the conditional pdf of $X \mid Y$ and for the marginal pdf of Y. You do not need to evaluate integrals, but the integrand and limits of integration must be shown explicitly.

The joint pdf is $4x(y-x)$ on the parallelogram sketched in this figure.



The conditional pdf for $X \mid Y$ is then

$$\frac{4x(y-x)}{\int_{\max\{y-1,0\}}^{\min\{y,1\}} 4x(y-x)\,dx}\,.$$

The marginal pdf for Y is just the denominator of the expression above.

(3) (10 points) It is claimed that X is distributed so that $P[X \in [0,5)] = .5]$, $P[X \in (3,8]] = .8$ and $P[X \in [2,3]] = .25$. Show that this is impossible.

$$P\big[[0,5) - [2,3]\big] = P\big[[0,2) \cup (3,5)\big] = .25\,. \tag{A}$$

But also

$$P\big[[0,5)^c\big] = 1 - P\big[[0,5)\big] = .5 \geq P\big[[5,8]\big]$$

and therefore

$$P\big[(3,5)\big] = P\big[(3,8]\big] - P\big[[5,8]\big] \geq .8 - .5 = .3\,. \tag{C}$$

But this last inequality (C) contradicts our initial calculation (A). Here we have been using just the properties that the probability of the whole space is 1, that all probabilities are non-negative, and that the probability of a countable union of disjoint sets is the sum of their probabilities.

(4) (5 points) X is distributed with the exponential pdf e^{-x} on $[0,\infty)$. This pdf has mean 1 and variance 1. What is the mean and and variance of X conditional on the set $X > x$, where x is a real number?

Over the set $X > x$, the pdf as a function of z is proportional e^{-z}. But since its height is e^{-x} at $z = x$, we need to multiply it by e^{x} to make it integrate to one. The result is just an exponential pdf shifted to a different origin, i.e. an exponential over (x,∞) instead of over $(0,\infty)$. The variance is therefore 1 as for the exponential over $(0,\infty)$, while the mean is increased by x, so the mean is $x+1$.

(5) (10 points) X and Y are jointly normal and have the same variance, though they are not independent. Show that the pair $X+Y$ and $X-Y$ are jointly normal and independent.

For jointly normal random variables independence is equivalent to zero covariance. The covariance of $X+Y$ with $X-Y$ is

$$\begin{aligned} &E[(X+Y)(X-Y)] - E[X+Y]E[X-Y] \\ &\qquad = E[X^2] - (EX)^2 - E[Y^2] + (EY)^2 = \mathrm{Var}(X) - \mathrm{Var}(Y)\,. \end{aligned}$$

But clearly this is zero when X and Y have the same variance.

(6) (10 points) Use Jensen's inequality (f concave implies $E[f(X)] \leq f(E[X])$) to show that if $\hat{\beta}$ is an unbiased estimator for β, $1/\hat{\beta}$ cannot be an unbiased estimator of $1/\beta$.

This was a bad question because the assertion isn't true without qualification. The function $f(\beta) = -1/\beta$ is convex over $(0, \infty)$, but not over the whole real line. Thus if $\hat{\beta}$ takes on only positivevalues, $E[1/\hat{\beta}] \geq 1/E[\hat{\beta}]$. If this latter inequality is strict, then it implies that $\hat{\beta}$ can't be an unbiased estimator for β at the same time that $1/\hat{\beta}$ is an unbiased estimator for $1/\beta$. The inequality is indeed strict whenever $\hat{\beta}$ has a non-degenerate distribution. However, the problem statement left out the two key qualifications, that the parameter space had to contain only positive β's, and that the estimator had to be non-degenerate (which follows from unbiasedness if the parameter space contains more than a single point.) It also should have given you the fact that Jensen's inequality is strict for strictly concave functions and non-degenerate random variables.

(7) (10 points) If $X_i \mid \beta \sim N(\beta, 1)$ for each $i = 1, \ldots, n$ and β can be any element of $\mathbb{R}$, then it is a standard result that

$$\bar{X} \pm \frac{1.96}{\sqrt{n}}$$

is a 95% non-Bayesian confidence interval for β. But now suppose that we know for sure, before looking at the sample, that β has to be between 0 and 1. If we replace our original interval with

$$\left\{ \bar{X} \pm \frac{1.96}{\sqrt{n}} \right\} \cap (0, 1)\,,$$

(in other words, use just the part of the original interval that lies within $(0, 1)$) will it still be a 95% confidence interval? Why or why not?

It is still a 95% confidence interval. If for every real β it has a 95% chance of containing β when β is the parameter value generating the data, then it certainly has this property for every β in (0,1). Since we know that β is in (0,1), we don't affect the probability that the random interval contains the true β by truncating it to (0,1).

(8) (10 minutes) The two-dimensional random vector $(X, Y) \sim N(0, \Sigma)$. The two-dimensional random vector (U, V) is defined by $U = 1/(X + Y)$, $V = Y$. Display the joint pdf for U and V. Is the conditional distribution of $V \mid U$ normal?

The absolute value of the deerminant of the Jacobian of the transformation is

$$\left| \frac{\partial(u, v)}{\partial(x, y)} \right| = \left| \begin{bmatrix} \frac{-1}{(x+y)^2} & \frac{-1}{(x+y)^2} \\ 0 & 1 \end{bmatrix} \right| = u^2\,.$$

Therefore the joint pdf is

$$\frac{1}{2\pi u^2} \left|\Sigma\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left(\tfrac{1}{u} - v \quad v \right) \Sigma^{-1} \left(\tfrac{1}{u} - v \quad v \right)' \right)\,.$$

Though this pdf is certainly not normal, with u held fixed it is e with an exponent quadratic in v, and it therefore will be, when normalized to integrate to one, a normal pdf. So the conditional distribution of $V \mid U$ is indeed normal.

Another way to answer the second part of the question was to observe that the σ-field generated by U is the same as the σ-field generated by $X+Y$, and $X+Y$, being a linear combination of X and Y, is jointly normally distributed with Y. But if two random variables are jointly normal, we know that the conditional distribution of one given the other is normal. Thus the conditional distribution of Y given $X+Y$ is normal, and since $X+Y$ and U generate the same σ-field, the conditional distribution of Y given U must also be normal. (And of course V and Y are the same variable.)

(9) (15 minutes) We are considering how to model the distribution of income across the population. This can be thought of as a random variable whose value is the income of a randomly chosen individual. We are contemplating using the Pareto distribution, which has pdf proportional to (with x representing income)

$$\frac{p-1}{(1+x)^p}$$

on $(0,\infty)$, with p lying in the interval $(1,\infty)$. Suppose that the candidate models have been boiled down to the Pareto with $p=3$ and the exponential, with pdf e^{-x}. It is proposed to reject the Pareto as null hypothesis if any one of a set of 5 people drawn at random from the population has an income higher than $y_3(.01)$, where

$$\int_{y_3(.01)}^{\infty} \frac{3-1}{(1+x)^3}\,dx = .01\,.$$

What is the significance level of this test? Explain why, with the alternative being exponential, there are better forms for a test of this null hypothesis, even if we maintain the constraint that we use only a random draw of 5 people as data. Discuss what such a better test might look like. It might be useful to know that $y_3(.01)$ is 4.6 and that $e^{-x} = (p-1)/(1+x)^p$ at approximately $x=.5$ and $x=4.3$. [This problem is of course too simplified to be realistic. Your discussion should nonetheless assume that one of these two assumed distributions is the true distribution.]

One component of the hints for this question was wrong. The value of $y_3(.01)$ is not 4.6, but rather 9, as is fairly easily checked analytically. I can't see how I arrived at 4.6, except possibly by doing the computation numerically at the same time I numerically computed the points where e^{-x} and $2/(1+x)^3$ are equal. In that case, I probably truncated the range of x too tightly, which would have given a low number for $y_3(.01)$. For a qualitative answer, all that mattered was that $y_3(.01) > 4.3$, so that the density of the Pareto is above the density of the exponential in the entire rejection region, implying that the probability of rejection is higher under the Pareto null than under the exponential alternative. But since it was fairly easy to compute analytically the true value of $y_3(.01)$, the hint could have been confusing.

The test has been constructed so that the probability of rejection of H_0 is $1 - .99^5$, so that is the test's significance level. The test is not good against an exponential alternative because the likelihood ratio for the Pareto over the exponential is not monotone, and indeed strongly favors the Pareto when the observed X values are large. Thus rejecting when the observed X's are really large is perverse. The test is biased, meaning that it rejects more often when the null is true than when the alternative is true. Because the Pareto density declines as one over a polynomial, while the exponential declines much faster than that, it is clear that for large enough x values, the Pareto density will always be higher. You were given that the densities are equal at just two points, with the larger being $x = 4.3$. Thus the Pareto density must be higher for all $x > 4.3$, and any set of the form (a, ∞) with $a > 4.3$ therefore has higher probability under the Pareto.

A better test would be based on a rejection region that has higher probability under the null than under the alternative. For example, the probability under the Pareto null that all 5 observations will be less than the median of the Pareto (which is .41) is .0625, so with this rejection region we have a .0625 level test. The probability of rejection under the exponential alternative is $e^{-5 \cdot .41} = .13$, which is higher than the significance level and makes the test unbiased.

A better test still could be based on the likelihood ratio, rejecting when the log likelihood ratio

$$5 \log 2 - 3 \sum_{j=1}^{5} \log(1 + x_i) + \sum_{j=1}^{5} x_i \tag{$*$}$$

is below some critical value. Determining critical values and rejection regions for such a test requires numerical work that could not be done on an exam. Such a test has power greater than 50% with a significance level of 5%.

Of course a better decision procedure might be not to use a classical test at all, but instead to use a posterior odds ratio. With equal prior probabilities on the two distributions, the posterior odds ratio would be the likelihood ratio, i.e $(*)$ exponentiated, and the ratio would favor the Pareto when this ratio exceeded one. This observation was not asked for in the test question, though.

Of course this answer gives much more detail than was expected on the exam. A perfect score answer would have recognized that the probability of large X values is higher under the Pareto than under the exponential, so that the rejection region should ideally not include very high values of X, which the proposed test does. It would have also showed that the proposed test is biased, though just noting that rejecting for large X's is perverse would already have gained substantial credit.