MIDTERM EXAM

Answer all four questions. Each is worth 23 points. Do not devote disproportionate time to any one question unless you have answered all the others.

(1) We are considering imposing a tax to finance a public program. The benefits from the tax are

$$B = 4(\tau - \theta\tau^2),$$

showing diminishing returns because as the tax increases the activity being taxed decreases. The deadweight loss from the tax is $C = \theta\tau^2$. We are not sure of θ, which represents the sensitivity of the activity with respect to the tax. $\theta = 1$ or $\theta = 2$ are the only two possibilities. Net benefits of the activity are of course $B - C$. Our options are restricted to setting τ to .1, .3, or .4.

(a) Calculate the pairs of losses, under $\theta = 1$ and $\theta = 2$, for each of the three τ values. [It may help to line these up in a "θ by τ" table.]

Net gains are $4\tau - 5\theta\tau^2$. Here is the table:

τ	θ	
	1	2
.1	.35	.3
.3	.75	.3
.4	.8	0

(b) Are any of the three τ choices inadmissible? Explain your answer.

$\tau = .1$ is inadmissible, since it produces lower gains than $\tau = .3$ when $\theta = 1$, and equal gains when $\theta = 2$.

(c) Which of the τ values represent Bayesian decision rules? For each of these, specify at least one probability distribution over θ that would support choice of that τ value.

The other two choices are both admissible and Bayesian. $\tau = .4$ is preferred if $P[\theta = 1] > 6/7$. Otherwise $\tau = .3$ is preferred. A possibly subtle point is that $\tau = .1$ is also Bayesian. It delivers expected gains equal to those of $\tau = .3$ in the case where $P[\tau = 2] = 1$. This illustrates the point that Bayes decision rules developed from priors that put probability zero on a set can be inadmissible through being dominated on that zero-probability set.

(2) Suppose we have an i.i.d. sample of T observations X_t, with each X_t distributed as gamma(p), i.e. with pdf $x^{p-1}e^{-x}/\Gamma(p)$.

(a) Show that the sample mean of the X_t's, $\bar{X} = \sum X_t / T$, is an unbiased estimator for p.

It was acceptable simply to point out that the mean of a standard Gamma(p,α) is p/α, so that here each X_t has mean p. Therefore $E[\bar{X}|p] = T \cdot p/T = p$, and $\bar{X}$ is unbiased for p. One could also do the integral:

$$E[X_t \mid p] = \frac{\int_0^\infty x \cdot x^{p-1} e^{-x}\, dx}{\Gamma(p)} = \frac{\Gamma(p+1)}{\Gamma(p)} = p\,,$$

where we are using the fact that $\Gamma(p) = (p-1)!$ for integer p.

(b) Display the kernel of the Bayesian posterior pdf for p when the prior pdf for p has the form $\pi(p) = \lambda^p e^{-\lambda}/\Gamma(p+1)$ over the integer values of $p = 0, 1, \ldots, \infty$, with $\lambda > 0$. (The kernel of the pdf is just any function that is proportional to the pdf, without necessarily integrating to one.) Note that here we are talking about a posterior concentrated on the non-negative integers.

The kernel is

$$\frac{\lambda^p}{\Gamma(p)^T \Gamma(p+1)} \left(\prod_{t=1}^T x_t \right)^{p-1} e^{-\sum_1^T x_t - \lambda}\,.$$

Since we are interested only in the behavior of this as a function of p, we can simplify it by removing all factors that don't depend on p, to arrive at

$$\frac{\lambda^p \left(\prod_{t=1}^T x_t \right)^{p-1}}{\Gamma(p)^{T+1} p}\,.$$

There is a glitch in the statement of this problem, though. The Gamma distribution and $\Gamma(p)$ are undefined for $p = 0$, so at best this has to be interpreted as a kernel over $p \geq 1$. $\Gamma(p)$ is defined for all $p > 0$. What I meant to state as the prior was $\pi(p) = \lambda^{p-1} e^{-\lambda}/\Gamma(p)$, for integer $p > 0$. This would have led to the posterior kernel

$$\frac{\left(\lambda \prod_{t=1}^T x_t \right)^{p-1}}{\Gamma(p)^{T+1}}$$

(c) Explain why the sample mean cannot be a Bayesian posterior mean for p, regardless of the prior distribution on p.

A reasonable answer was simply that we showed in class that an unbiased estimator cannot be a Bayesian posterior mean with any proper prior, if the estimator has finite variance. However, to make this answer complete you should have checked the finite variance property, and the question did not explicitly rule out improper priors. A direct and clean argument is simply to note that the sufficient statistic for p here is $\prod x_t$ (or equivalently $(\prod x_t)^{1/T}$, the geometric mean). Since the sample mean does not depend on the data exclusively *through the sufficient statistic, it is not determined by the likelihood function shape alone. It therefore can't be a Bayesian estimator, because Bayesian estimators, whether from proper or improper priors, and regardless of the loss function, depend on the data only through the likelihood function and (therefore) through the sufficient statistics.*

(3) Suppose we will have a single observation on a random variable X that is the sum of two independent $U(0, a)$ (uniform on the interval $(0, a)$) random variables, where $a > 0$ is unknown.

(a) Find the pdf of $X \mid a$.

There are several ways to get the answer, which is that the pdf of X is $1/a - \left|(x-a)/a^2\right|$ *for* $x \in (0, 2a)$, *0 elsewhere, i.e. a triangle on* $(0, 2a)$ *with peak height* $1/a$ *at a. One way to get this result is to recall that the sum of two independent random variables is the convolution of their pdfs, so here we have, for* $x \in (0, a)$

$$\int_0^x \frac{1}{a^2}\, dy \;=\; \frac{x}{a^2}\,.$$

It should be clear by the symmetry, that this linear behavior for $x < a$ *just gets reversed as x proceeds above a. One could also think directly about the 2-dimensional geometry. If* U, V *are the two independent uniforms that sum to X, they have a joint density spread evenly over the unit square in* $\mathbb{R}^2$. *Transforming from* u, v *coordinates to* x, v *coordinates has a Jacobian of 1, but we have to recognize that for a fixed value of x, v can range only between* $\max(0, x-a)$ *and* $\min(x, a)$. *When we then integrate w.r.t. y, we get the same result.*

(b) Show that X is an unbiased estimator for a.

This is obvious, since the pdf is symmetric about a and bounded, so its expectation exists.

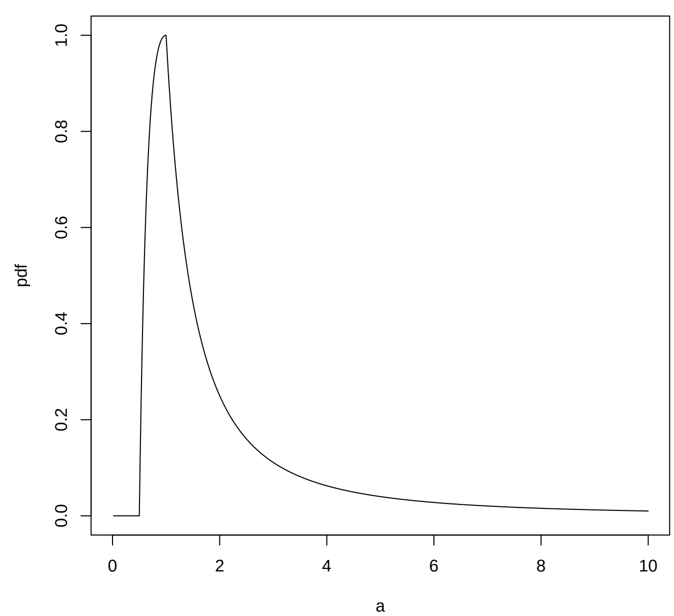
(c) Construct a two-tail test of size .04 for the null hypothesis $H_0 : a = a_0$, for an arbitrary $a_0 > 0$.

We want to find an $\bar{x}$ *such that* $\int_0^{\bar{x}} (x/a_0^2)\, dx = .02$. *This can easily be seen to be* $\bar{x} = .2a_0$. *Therefore the symmetric interval that, conditional on* a_0, *has probability .96, is* $(.2a_0, 1.8a_0)$. *The test should therefore reject whenever X is outside this interval.*

(d) Construct a 96% confidence interval for a based on the family of tests you contructed in (3c)

For the confidence interval, we just collect all values of a that are not rejected for a given observation on X. These are all a's such that $X > .2a$ *or* $X < 1.8a$, *or in other words all a's in the interval* $(\frac{5}{9}X, 5X)$.

(e) Calculate the likelihood function and sketch a plot of it against a. Will the posterior mean under a flat prior be X? Is it likely that a decision maker with loss function $(a - \hat{a})^2$ will choose $\hat{a} = X$? Why or why not.
The likelihood is x/a^2 for $a > x$, $2/a - x/a^2$ for $a \in (x/2, x)$. See the figure for a "sketch". Since the pdf for large a is x/a^2, the flat-prior posterior pdf, though integrable, has infinite first moment, so certainly X is not the posterior mean. A decision-maker should be using a proper prior, so the fact that the flat-prior posterior implies infinite expected loss under the loss function given does not mean the decision-maker would see infinite expected loss. But if the decision-maker's prior were not sharply concentrated relative to the likelihood, the heavy skewness to the right of the likelihood would pull the posterior mean (which is the optimal choice under a quadratic loss function) well above X.
As an aside, (you were not asked for this) the flat-prior 96% equal-tail interval for a is $(.59X, 36X)$, much more skewed to the right than the confidence region.



(f) Suppose we want to test $H_0 : a < 1$. Consider the following procedure: Reject H_0 if the confidence interval you constructed in (3d) is disjoint from $(0, 1)$ (i.e., does not overlap it at all). What is the size of this test? Is it unbiased?
When $a = 1$, this test rejects if and only if X is in the right tail of the rejection region for the part (3c) test of a as null hypothesis, so the rejection probability

is .02. The test rejects whenever X exceeds 5. The probability of such an open-ended-to-the-right interval, conditional on a, is strictly increasing in a. Thus for a's less than one the rejection probability is smaller, so the size of the test is .02, and for a's greater than one the rejection probability is larger, so the test is unbiased.

(4) We have data for a standard normal linear model, $\{Y \mid X\} \sim N(X\beta, \sigma^2 I)$. There are 10 observations and

$$X'X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \quad (X'X)^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \quad \hat{\beta}_{OLS} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

$$X'Y = \begin{bmatrix} 6 \\ 11 \\ 14 \end{bmatrix}, \quad Y'Y = 80\,.$$

(a) Find a 95% two-sided confidence interval for $\beta_2 + \beta_3$. The .025 tail for a $t_7(0,1)$ distribution is at 2.36.
First we need the usual estimate of residual variance, which here can be found as $(Y'Y - \hat{\beta}'_{OLS} X'Y)/(T-k) = 10/7$. We are interested in the linear combination of β's with weights $w = (0, 1, 1)$. We can therefore treat $\hat{\beta}_{OLS,2} - \beta_2 + \hat{\beta}_{OLS,3} - \beta_3$, conditional on β, as t_7 with variance parameter

$$(10/7) \cdot w(X'X)^{-1} w' = 10/7\,.$$

This makes the requested interval $5 \pm (2.36 * 1.2) = (2.2, 7.8)$.

(b) Explain how you would calculate a 95% HPD set for $\beta_1/(\beta_1 + \beta_2)$ under a $d\sigma/\sigma$ prior, without actually trying to carry out the calculation. You should explain as if giving intructions to an undergraduate RA well trained in algebra and calculus and computer programming, but who knows nothing about probability or statistics.
There are two possible approaches: Jacobian or Monte Carlo simulation.

Jacobian: *If we set $\gamma = \beta_1/(\beta_1 + \beta_2)$, the Jacobian is*

$$\left| \frac{\partial(\beta_1, \beta_2)}{\partial(\gamma, \beta_2)} \right| = \frac{\beta_2}{(1-\gamma)^2}\,.$$

Substitute $\beta_1 = \gamma\beta_2/(1-\gamma)$ for β_1 in the multivariate t_7 density for β, and multiply the result by the Jacobian. Numerically integrate this expression with respect to β_2 at a finely spaced set of values of γ, chosen to cover the region where the integrand is large. Take special care with $\gamma = 1$ because the Jacobian suggests the pdf will show an integrable discontinuity there; it may be necessary to use a finer grid in that neighborhood. Normalize the resulting vector of marginal pdf heights so they *sum to one. For each point on this grid of γ values, calculate the marginal pdf height $h(\gamma)$ and the cdf value $F(\gamma)$(found as the sum of pdf values over γ's less than or equal to the current one). Once this is done, examine all pairs γ_1, γ_2 such that $F(\gamma_1) + F(\gamma_2) - 1 \doteq .05$ Among these, pick the pair that most nearly satisfies $h(\gamma_1) = h(\gamma_2)$. Note that, though $1/(1-\gamma)^2$ is not integrable over domains that include $\gamma = 1$, the joint and marginal pdf's we are dealing with here must be integrable, because they are pdf's for functions of β, which has an integrable likelihood for $T > k$. As $\gamma \to 1$ the Jacobian goes to infinity, but at the same time $\beta_1 = \gamma\beta_2/(1-\gamma_1)$ goes rapidly to infinity, for fixed β_2, and this expression appears squared in the denominator of the t pdf.*

Monte Carlo: *You can imagine that there is a program to make i.i.d. pseudorandom draws from a multivariate t distribution. If this were available, you would just instruct the RA to make a few thousand such draws, say 4000, with the degrees of freedom 7 and the covariance matrix $(10/7)(X'X)^{-1}$. From these draws, form the vector of values for $\gamma = \beta_1/(\beta_1 + \beta_2)$. Sort these values in increasing order. An equal-tail posterior probability interval is then approximately given by the range from the 200'th to the 3800'th value of γ in these draws. This is not the HPD interval, however, unless it happens that a good estimate of the pdf at these points matches. One could locate an estimate of the HPD interval by looking at the 41 intervals defined by, say, the 180'th to the 220'th draws paired with the corresponding 3780'th to 3820'th draws. The shortest of these intervals is likely a good estimate of the HPD, unless the shortest interval turns out very close to the (220,3820) or (180,3780) pair. In that case, the search should be extended further to the right or left (respectively).*

Since a multivariate t random generator is not in the R base package, a still better answer might have explained how to generate a multivariate t by drawing first from the Gamma(T) posterior on $\nu = 1/\sigma^2$, then from the $N\left(\hat{\beta}_{OLS}, (\nu X'X)^{-1}\right)$ for β.