Stat 512
Emerson, Fall 15
Final Examination Key
December 14, 2015

Instructions:

- **This exam is closed book, closed notes. No use of calculators is permitted.**
- **Write answers to the following questions on separate sheets of paper, starting each problem at the top of a new page. Use only the front side of each page. Be sure to write your name on the top of each page.**
- **In order to receive full credit, you must make clear how you derived the answers to the problems.**
- **You are allowed 1 hour and 40 minutes for this exam. When time is called, you must put down your pencils**

In addition to the theorems covered in class, you may use the following facts without proof:

- The following sums

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$
$$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}$$

- If $X \sim \mathcal{E}(\lambda)$ (an exponential random variable with density $f_X(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0<x<\infty]}$ for some $\lambda > 0$), then $EX = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.

- As $n \to \infty$

$$\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a$$

1. (15 points) Provide definitions for the following terms. Be sure to make clear any notation you use.

   a. Convergence almost surely.

Ans: Let X and $X_1, X_2, \ldots$ be random variables defined on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$. The sequence $X_1, X_2, \ldots$ converges almost surely to X ($X_n \rightarrow_{a.s.} X$) if $Pr[X_n \rightarrow X] = 1$, or, more formally,

$$\mathcal{P}\left(\{\omega \in \Omega : \lim_{n\rightarrow\infty} X_n(\omega) = X(\omega)\}\right) = 1$$

   b. Convergence in probability.

Ans: Let X and $X_1, X_2, \ldots$ be random variables defined on a common probability space $(\Omega, \mathcal{A}, \mathcal{P})$. The sequence $X_1, X_2, \ldots$ converges in probability to X ($X_n \rightarrow_p X$) if $\forall \epsilon > 0 \; Pr[|X_n - X| > \epsilon] \rightarrow 0$, or, more formally,

$$\forall \epsilon > 0 \lim_{n\rightarrow\infty} \mathcal{P}\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0$$

   c. Convergence in distribution.

Ans: Let X and $X_1, X_2, \ldots$ be random variables (not necessarily defined on the same probability space) having cumulative distribution functions $F(x) = Pr[X \leq x]$ and $F_n(x) = Pr[X_n \leq x]$, respectively. The sequence $X_1, X_2, \ldots$ converges in distribution to X ($X_n \rightarrow_d X$) if $\forall x_0$ such that $F(\cdot)$ is continuous at x_0,

$$\lim_{n\rightarrow\infty} F_n(x_0) = F(x_o)$$

   d. The delta method.

Ans: Let a_n be a sequence of real numbers with $a_n \rightarrow \infty$ as $n \rightarrow \infty$, and let X and $X_1, X_2, \ldots$ be random variables which for some real θ satisfy

$$a_n(X_n - \theta) \rightarrow_d X.$$

Then for real function g that is differentiable at θ (so $g'(\theta)$ exists),

$$a_n(g(X_n) - g(\theta)) \rightarrow_d g'(\theta)X.$$

   e. Slutsky's theorem.

Ans: Let a and b be real numbers, let a_n be a sequence of real numbers with $a_n \rightarrow_p a$, let b_n be a sequence of real numbers with $b_n \rightarrow_p b$, and let X and $X_1, X_2, \ldots$ be random variables for which $X_n \rightarrow_d X$. Then

$$a_n X_n + b_n \rightarrow_d aX + b.$$

2. (15 points) State and prove a weak law of large numbers.

Ans: (WLLN) Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n = \sum_{i=1}^n X_i/n$. Then $\overline{X}_n \to_p \mu$.

Pf: By laws of expectation,

$$E\left[\overline{X}_n\right] = E\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n E\left[X_i\right] = \mu$$

and because the random variables are independent, the properties of variance dictate

$$Var\left(\overline{X}_n\right) = Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var\left(X_i\right) = \frac{\sigma^2}{n}.$$

Now using Chebyshev's inequality and the properties of probabilities, for every fixed $\epsilon > 0$

$$0 \le Pr\left[|\overline{X}_n - \mu| > \epsilon\right] \le \frac{Var(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Because $\sigma^2/(n\epsilon^2) \to 0$, we thus have

$$\forall \epsilon > 0 \lim_{n\to\infty} Pr\left[|\overline{X}_n - \mu| > \epsilon\right] = 0,$$

which is the definition of convergence of probability.

3. (30 points) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. discrete random variables having probability mass function

$$p_X(x) = \frac{2x}{\theta(\theta+1)}\mathbf{1}_{[x\in\{1,2,3,\ldots,\theta\}]}$$

for some integer $\theta \ge 1$.

a. Derive a method of moments estimator for θ for sample size n. Find its bias function and show its consistency.

Ans: The easiest approach is to find the lower order moments of the distribution of X until you find a function of the moments that is equal to θ. Thus, looking first at the first moment,

$$\begin{aligned} E[X_i] &= \sum_{x=1}^{\theta} x\frac{2x}{\theta(\theta+1)} \\ &= \frac{2}{\theta(\theta+1)}\sum_{x=1}^{\theta} x^2 \\ &= \frac{2}{\theta(\theta+1)}\frac{\theta(\theta+1)(2\theta+1)}{6} = \frac{2\theta+1}{3}. \end{aligned}$$

We can thus see that $\theta = g(E[X_i])$ for $g(x) = (3x-1)/2$. Hence a method of moments estimator of θ is $\tilde{\theta}_n = (3\overline{X}_n - 1)/2$, where $\overline{X}_n = \sum_{i=1}^n X_i/n$ is the sample mean of the first n observations.

In order to find the bias function, we merely note that because the sample mean of identically distributed random variables is always an unbiased estimator of the expected value of any one of the random variables,

$$\begin{aligned} b_n(\tilde{\theta}_n, \theta) &= E[\tilde{\theta}_n] - \theta \\ &= E\left(\frac{3\overline{X}_n - 1}{2}\right) - \theta \\ &= \frac{3E\left[\overline{X}_n\right] - 1}{2} - \theta \\ &= \frac{3(2\theta+1)/3 - 1}{2} - \theta = \theta - \theta = 0, \end{aligned}$$

showing that $\tilde{\theta}_n$ is unbiased.

In order to show its consistency, we use the WLLN (either Khinchin's to avoid having to argue that $Var(X_i) < \infty$, or noting that in the next part we will show finite variance) to show that $\overline{X}_n \to_p (2\theta+1)/3$. Then, because $g(x) = (3x-1)/2$ is a continuous function, the Mann-Wald (continuous mapping) theorem dictates that

$$\tilde{\theta}_n = g(\overline{X}_n) \to_p g((2\theta+1)/3) = \theta.$$

(Many of you used Chebyshev's explicitly here, though I think it faster to invoke WLLN and Mann-Wald.)

b. Find the asymptotic distribution for your estimator in part a.

Ans: Because we know that $\tilde{\theta}_n$ is a rather straightforward function of the sample mean, the obvious approach should be to start with a central limit theorem, and then try the delta method and see if we end up with a nondegenerate (i.e., nondeterministic) distribution.

Because the random variables are i.i.d., it seems logical to first try the Levy Central Limit Theorem. In order to use the Levy Central Limit Theorem, we need to know that the variance exists and know the formula for it. This is easily derived by using the computational formula $Var(X) = E[X_i^2] - E^2[X]$. We thus find

$$\begin{aligned} E[X_i^2] &= \sum_{x=1}^{\theta} x^2 \frac{2x}{\theta(\theta+1)} \\ &= \frac{2}{\theta(\theta+1)} \sum_{x=1}^{\theta} x^3 \\ &= \frac{2}{\theta(\theta+1)} \frac{\theta^2(\theta+1)^2}{4} = \frac{\theta(\theta+1)}{2}. \end{aligned}$$

Then,

$$Var(X_i) = \frac{\theta(\theta+1)}{2} - \left(\frac{2\theta+1}{3}\right)^2 = \frac{\theta^2+\theta-2}{18} < \infty.$$

So from the Levy Central Limit Theorem we have

$$\sqrt{n}\left(\overline{X}_n - \frac{2\theta+1}{3}\right) \to_d \mathcal{N}\left(0, \frac{\theta^2+\theta-2}{18}\right).$$

Now, we use the δ-method for $\tilde{\theta}_n = g(\overline{X}_n)$, with $g(x) = (3x-1)/2$ and $g'(x) = 3/2$. Thus

$$\sqrt{n}\left(\tilde{\theta}_n - \theta\right) \to_d \frac{3}{2}\mathcal{N}\left(0, \frac{\theta^2+\theta-2}{18}\right) = \mathcal{N}\left(0, \frac{\theta^2+\theta-2}{8}\right).$$

(When using this asymptotic distribution as an approximation in finite samples, we would thus use

$$\tilde{\theta}_n \dot{\sim} \mathcal{N}\left(\theta, \frac{\theta^2+\theta-2}{8n}\right)$$

in order to compute frequentist statistics such as P values and confidence intervals.)

c. Find a maximum likelihood estimator for θ for sample size n.

Ans: We find the likelihood function as

$$\begin{aligned} L(\theta \,|\, \vec{X}_n) &= \prod_{i=1}^n \frac{2X_i}{\theta(\theta+1)} \mathbf{1}_{[X_i \in \{1,2,\ldots,\theta\}]} \\ &= \frac{2^n}{\theta^n(\theta+1)^n} \prod_{i=1}^n X_i \mathbf{1}_{[X_{(n)} \in \{1,2,\ldots,\theta\}]}, \end{aligned}$$

where $X_{(n)}$ is the nth order statistic in the sample $\{X_1, \ldots, X_n\}$.

As the support of the distribution of the X_i's varies with θ, it is generally easier to find the MLE by noting the following properties of the likelihood function:

- The value of $2^n/(\theta^n(1+\theta)^n)$ is decreasing as θ increases.
- The likelihood function is zero for $\theta < X_{(n)}$ and positive, otherwise.

It therefore follows that $L(\theta)$ is maximized at the lowest value of θ for which $L(\theta) > 0$. This occurs at MLE $\hat{\theta} = X_{(n)}$.

(Many of you looked at the derivative with respect to θ of the log likelihood function. Note, however, that because θ is discrete, the function is not truly differentiable. Nevertheless, we can show that an extended function defined on a real valued θ and that is equal to the log likelihood at all integer θ is differentiable for

all $\theta > X_{(n)}$, and that extended function is decreasing in θ for $\theta > X_{(n)}$. Hence, in this case the discrete parameter space is not really a problem.)

(I did not ask you to show the bias or consistency of this estimator, nor did I ask for the asymptotic distribution. Because the estimator does not involve a sum, it is most likely that such properties would have to be derived by brute force using the definitions of expectation, convergence in probability, and convergence in distribution.)

4. (40 points) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. exponential random variables having the density $f_X(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0<x<\infty]}$.

a. Find a method of moments estimator for λ and give its asymptotic distribution.

Ans: Again, we start by finding the lowest order moments. From the facts that I allowed you to use without derivation, we know that $E[X_i] = 1/\lambda$. Hence the easiest method of moment estimator to use is $\tilde{\lambda}_n = 1/\overline{X}_n$, where $\overline{X}_n$ is the sample mean of the first n random variables.

In order to find its asymptotic distribution, we note that the MME $\tilde{\lambda}_n$ is a continuous function of the sample mean. This then suggests an approach based on using a central limit theorem and then the δ-method.

Again using the facts given at the start of the exam, we know $Var(X_i) = 1/\lambda^2$, so the Levy CLT tells us

$$\sqrt{n}\left(\overline{X}_n - \frac{1}{\lambda}\right) \rightarrow_d \mathcal{N}\left(0, \frac{1}{\lambda^2}\right).$$

Now, $\tilde{\lambda}_n = g(\overline{X}_n)$ for $g(x) = 1/x$, with $g'(x) = -1/x^2$ existing for all $x > 0$. We then note that $g(E[X_i]) = \lambda$ and $g'(E[X_i]) = \lambda^2$, so by the δ-method we find

$$\sqrt{n}\left(\tilde{\lambda}_n - \lambda\right) \rightarrow_d \lambda^2\mathcal{N}\left(0, \frac{1}{\lambda^2}\right) = \mathcal{N}\left(0, \lambda^2\right).$$

b. Find a maximum likelihood estimator for λ and give its asymptotic distribution.

Ans: The likelihood function for $\lambda > 0$ is

$$\begin{aligned} L(\lambda \,|\, \vec{X}_n) &= \prod_{i=1}^{n} \lambda e^{-\lambda X_i}\mathbf{1}_{[X_i \in (0,\infty)]} \\ &= \lambda^n e^{-\lambda n \overline{X}_n}. \end{aligned}$$

Because the support of the distribution of X_i is independent of λ, this is a setting in which the MLE is usually most easily found by maximizing the log likelihood. That maximization is in turn most easily effected using differentiation. So the log likelihood is

$$\mathcal{L}(\lambda \,|\, \vec{X}) = n\log(\lambda) - n\lambda\overline{X}_n,$$

and the derivative with respect to λ is

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{n}{\lambda} - n\overline{X}_n.$$

In trying to find the MLE, we consider the values $\hat{\lambda}$ which satisfy

$$\left.\frac{\partial \mathcal{L}}{\partial \lambda}\right|_{\lambda=\hat{\lambda}_n} = 0,$$

which yields $\hat{\lambda}_n = 1/\overline{X}_n$. Examining the second derivative of the log likelihood gives

$$\frac{\partial^2 \mathcal{L}}{\partial \lambda^2} = -\frac{n}{\lambda^2}.$$

Because that second derivative is negative at $\hat{\lambda}_n$, we know that $\hat{\lambda}_n$ is the MLE.

Of course, the MLE $\hat{\lambda}_n$ is the same estimator as the MME $\tilde{\lambda}_n$ considered in part a, so the asymptotic distribution is the same as that derived above.

c. Find a maximum likelihood estimator for $\theta = Pr[X_1 > 1]$ and give its asymptotic distribution.

Ans: First, we find the value of θ as a function of λ. We can then use the fact that the MLE of a function of λ is the function of the MLE of λ.

By integrating the density, we find

$$\theta = Pr[X_1 > 1] = \int_1^\infty \lambda e^{-\lambda x}\, dx = e^{-\lambda}.$$

Thus, the MLE $\hat{\theta}_n$ for θ is

$$\hat{\theta}_n = e^{-1/\hat{\lambda}_n}.$$

We find the asymptotic distribution using the δ-method with $g(x) = e^{-x}$ and $g'(x) = -e^{-x}$. Hence, from the asymptotic distribution of $\hat{\lambda}_n$ we find

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \rightarrow_d \mathcal{N}\left(0, \lambda^2 e^{-2\lambda}\right) = \mathcal{N}\left(0, \theta^2 \log^2(\theta)\right).$$

d. Let $Y_i = \mathbf{1}_{[X_i > 1]}$. Find a maximum likelihood estimator for θ based on the Y_i's and give its asymptotic distribution.

Ans: The Y_i's are i.i.d. random variables with $Y_i \sim \mathcal{B}(1, \theta)$. Hence, the likelihood for θ based on the observed Y_i's is

$$L(\theta \,|\, \vec{Y}_n = \prod_{i=1}^n \theta^{Y_i}(1-\theta)^{1-Y_i} = \theta^{n\overline{Y}_n}(1-\theta)^{n-n\overline{Y}_n}.$$

Because the support of Y_i is independent of $\theta \in (0,1)$, we again first try to find the MLE of θ by differentiating the log likelihood. The log likelihood is given by

$$\mathcal{L}(\theta \,|\, \vec{Y}_n) = n\overline{Y}_n \log(\theta) + (n - n\overline{Y}_n)\log(1-\theta),$$

and the first derivative with respect to θ is

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{n\overline{Y}_n}{\theta} - \frac{n - n\overline{Y}_n}{1-\theta} = \frac{n(\overline{Y}_n - \theta)}{\theta(1-\theta)}.$$

In trying to find the MLE, we consider the values $\tilde{\theta}$ which satisfy

$$\left.\frac{\partial \mathcal{L}}{\partial \theta}\right|_{\theta=\tilde{\theta}_n} = 0,$$

which yields $\tilde{\theta} = \overline{Y}_n$. Examining the second derivative of the log likelihood gives

$$\frac{\partial^2 \mathcal{L}}{\partial \theta^2} = -n\frac{\overline{Y}_n(1-\theta)^2 + (1-\overline{Y}_n)\theta^2}{\theta^2(1-\theta)^2}.$$

Because that second derivative is negative at $\tilde{\theta}_n = \overline{Y}_n$ (and in fact for all $\theta \in (0,1)$, because $0 \le \overline{Y}_n \le 1$), we know that $\tilde{\theta}_n$ is the MLE.

Because our estimator is a sample mean of i.i.d. random variables having finite mean, we can find the asymptotic distribution directly from the Levy CLT. Since $Var(Y_i) = \theta(1-\theta)$, the asymptotic distribution is thus

$$\sqrt{(n)}\left(\tilde{\theta}_n - \theta\right) \to_d \mathcal{N}\left(0, \theta(1-\theta)\right) = \mathcal{N}\left(0, e^{-\lambda} - e^{-2\lambda}\right).$$

e. Discuss the relative advantages and disadvantages of your estimators in parts c and d.

Ans: We can examine the optimality of the estimators with respect to bias, consistency, and asymptotic variance. (If one of the estimators is biased, we might want to consider mean squared error, but this will be a bit more involved.)

$\tilde{\theta}_n$ is the sample mean of random variables having expectation θ. Hence, we can immediately see that $\tilde{\theta}$ is unbiased. On the other hand, the bias of $\hat{\theta}_n = e^{-1/\overline{X}_n}$ is difficult to evaluate, and I did not expect you to evaluate this during the examination. Given the extreme nonlinear transformation of the the sample mean, we would expect it to be a biased estimator. (Note that if the transformation were strictly convex or concave, we could use Jensen's inequality to establish the bias of this estimator. Unfortunately, this function is neither convex or concave.)

With regard to consistency, we can use the WLLN to establish that $\tilde{\theta}_n$ is consistent for θ. Similarly, we can use the WLLN to establish that $\overline{X}_n$ is consistent for

$1/\lambda$, and then use the Mann-Wald (continuous mapping) theorem to show the consistency of $\hat{\theta}_n$.

In comparing the asymptotic variances, we can examine the asymptotic variance of $\tilde{\theta}_n$ minus the asymptotic variance of $\hat{\theta}_n$ and compare that difference to 0, or, alternatively, we can examine the ratio of those asymptotic variances and compare the ratio to 1. As the ratio does not immediately simplify, I consider the difference. Noting that $\lambda = -\log(\theta)$, we then define

$$h^*(\theta) = \frac{\theta(1-\theta)}{n} - \frac{\theta^2 \log^2(\theta)}{n}.$$

If $h^*(\theta)$ is always positive or always negative, then one of the estimators has smaller variance. Because $0 \le \theta \le 1$ and $n > 0$, we can consider the sign of

$$h(\theta) = \frac{n}{\theta} h^*(\theta) = 1 - \theta - \theta \log^2(\theta).$$

Now $h(1) = 0$. So if we can show that $h(\theta)$ is strictly decreasing on $\theta \in (0,1)$, we will know that $h(\theta) > 0$ for $\theta \in (0,1)$ and $\tilde{\theta}_n$ has greater asymptotic variance. If we can show that $h(\theta)$ is strictly increasing, then we will know that $h(\theta) < 0$ for $\theta \in (0,1)$ and $\hat{\theta}_n$ will have the larger asymptotic variance. (If $h(\theta)$ is neither strictly increasing nor strictly decreasing, we won't know anything, and will have to consider something else.) Taking the first derivative with respect to θ,

$$\begin{aligned} h'(\theta) &= -1 - \log^2(\theta) - \frac{2\theta \log(\theta)}{\theta} \\ &= -(1 + 2\log(\theta) + \log^2(\theta)) = -(1 + \log(\theta))^2 < 0. \end{aligned}$$

(We could have also worked on the scale of λ instead of θ.) Thus $\hat{\theta}_n$ is the more efficient estimator. (This, of course, makes intuitive sense. Otherwise we would never bother to keep complete data, and would just dichotomize all our data.)

Lastly, we note that the estimator $\tilde{\theta}_n$ consistently and unbiasedly estimates $\theta = Pr(X_i > 1)$ for every probability distribution, while the optimality properties of $\hat{\theta}_n$ are heavily dependent upon the data truly having an exponential distribution.

5. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having mean μ and variance σ^2, and let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. random variables having mean ν and variance τ^2. Further suppose that X_i and Y_j are independent if $i \ne j$, but that the correlation between X_i and Y_j is ρ if $i = j$.

   a. (10 points) Find a method of moments estimator for $\mu + \nu$ and derive its asymptotic distribution. Is your estimator unbiased? Consistent?

Ans: Let $\vec{U}_i = (X_i, Y_i)$. Then $E[\vec{U}_i] = (\mu, \nu)$, and a method of moments estimator will be $\overline{X}_n + \overline{Y}_n$. By laws of expectation, this is unbiased, and by the WLLN, this is consistent. The asymptotic distribution, can be most easily found by noting that if we define $V_i = X_i + Y_i$, then our MME is merely $\overline{V}_n$. As the V_i's are independent, the Levy Central Limit Theorem provides that $\overline{V}_n$ will be asymptotically normally distributed. We find

$$Var(V_i) = Var(X_i + Y_i) = Var(X_i) + Var(Y_i) + 2Cov(X_i, Y_i) = \sigma^2 + \tau^2 + 2\rho\sigma\tau,$$

so

$$\sqrt{n}\left((\overline{X}_n + \overline{Y}_n) - (\mu + \nu)\right) \to_d \mathcal{N}(0, \sigma^2 + \tau^2 + 2\rho\sigma\tau).$$

b. (10 points) Define $W_i = X_i^2 + Y_i^2$. Let $\theta = E[W_i]$. Find an unbiased estimator of θ. Is your estimator consistent?

Ans: Because every sample mean of identically distributed data is an unbiased estimator of the expectation of a random variable, the obvious choice is $\hat{\theta}_n = \overline{W}_n$. Because the W_i's are i.i.d, we can use Khinchin's WLLN to show that $\hat{\theta}$ is consistent for θ, i.e., $\hat{\theta}_n \to_p \theta$.

(Note that the use of Khinchin's avoided the need to find the variance of the W_i's.)

(Many of you found an expression for $\theta = \sigma^2 + \mu^2 + \tau^2 + \nu^2$, and then substituted unbiased estimates of the means and variances into that equation. This would of course work fine for a consistent estimate, and that approach would would end up being asymptotically equivalent to the far simpler approach I gave in the answer. You would have to be very careful in using the squares of the unbiased estimates of the means, however, if our goal is an unbiased estimator. Not all of you who tried this approach considered this issue.)

c. (10 points) Show that the estimator $\tilde{\theta} = \overline{X}_n^2 + \overline{Y}_n^2$ is a biased estimator for θ. Find the bias function.

Ans: Recalling that for any random variable Z, $E[Z^2] = Var(Z) + E^2[Z]$, we easily find that

$$\theta = E[W_i] = E[X_i^2] + E[Y_i^2] = \sigma^2 + \mu^2 + \tau^2 + \nu^2.$$

On the other hand, we also know that because the X_i's are i.i.d., $E[\overline{X}_n] = \mu$ and $Var(\overline{X}_n) = \sigma^2/n$ yielding $E[\overline{X}_n^2] = \sigma^2/n + \mu^2$. Similarly, the Y_i's are i.i.d., so $E[\overline{Y}_n] = \nu$, $Var(\overline{Y}_n) = \tau^2/n$, and $E[\overline{Y}_n^2] = \tau^2/n + \nu^2$. (Of course, $\overline{X}_n$ and $\overline{Y}_n$ are correlated, but that need not concern us when we are just finding the expectation of $\tilde{\theta}_n$.) We thus find

$$E[\tilde{\theta}_n] = \frac{\sigma^2}{n} + \mu^2 + \frac{\tau^2}{n} + \nu^2,$$

and the bias function is

$$b(\tilde{\theta}_n, \theta) = \frac{1-n}{n}\left(\sigma^2 + \tau^2\right) \neq 0.$$

d. (10 points) Show that $\tilde{\theta}_n$ is not consistent for θ.

Ans: By the WLLN, $\overline{X}_n \to_p \mu$ and $\overline{Y}_n \to_p \nu$. Thus by Mann-Wald, $\overline{X}_n^2 \to_p \mu^2$ and $\overline{Y}_n^2 \to_p \nu^2$. Then by the properties of convergence in probability,

$$\overline{X}_n^2 + \overline{Y}_n^2 \to_p \mu^2 + \nu^2.$$

So $\tilde{\theta}_n$ is not consistent for $\theta = \sigma^2 + \mu^2 + \tau^2 + \nu^2$.

e. (Bonus: 20 points) Find a method of moments estimator for $\mu^2 + \nu^2$ and derive its asymptotic distribution.

Ans: Using the results of part (a), the obvious choice is $\tilde{\theta} = \overline{X}_n^2 + \overline{Y}_n^2$ as defined in part (c). Clearly, this estimator is a function of sample means, so the best approach will be to use the multivariate CLT and then the multivariate δ-method.

By the multivariate CLT,

$$\sqrt{n}\left(\begin{pmatrix}\overline{X}_n \\ \overline{Y}_n\end{pmatrix} - \begin{pmatrix}\mu \\ \nu\end{pmatrix}\right) \to_d \mathcal{N}_2\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2\end{pmatrix}\right).$$

Now, for $g(\vec{x}) = x_1^2 + x_2^2$, $\nabla g(\vec{x}) = (2x_1, 2x_2)^T$. Thus, by the multivariate δ-method

$$\sqrt{n}\left(g\begin{pmatrix}\overline{X}_n \\ \overline{Y}_n\end{pmatrix} - g\begin{pmatrix}\mu \\ \nu\end{pmatrix}\right) \to_d \nabla g\begin{pmatrix}\mu \\ \nu\end{pmatrix} \cdot \mathcal{N}_2\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2\end{pmatrix}\right).$$

This then yields

$$\sqrt{n}\left((\overline{X}_n^2 + \overline{Y}_n^2) - (\mu^2 + \nu^2)\right) \to_d \mathcal{N}\left(0, \begin{pmatrix}2\mu & 2\nu\end{pmatrix}\begin{pmatrix}\sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2\end{pmatrix}\begin{pmatrix}2\mu \\ 2\nu\end{pmatrix}\right),$$

so performing the matrix multiplication we obtain

$$\sqrt{n}\left((\overline{X}_n^2 + \overline{Y}_n^2) - (\mu^2 + \nu^2)\right) \to_d \mathcal{N}\left(0, 4\mu^2\sigma^2 + 8\rho\mu\nu\sigma\tau + 4\nu^2\tau^2\right).$$