Homework 2

1) ① Def: $\hat{\mu}_k = n^{-1} \sum_{i=1}^{n} X_i^k$ kth sample moment

$X_1, X_2, \ldots, X_n$ are iid random variables from population $f(x|\theta)$.

Def: $Bias(\hat{\mu}^k) = E(\hat{\mu}^k - \mu^k) = E(\hat{\mu}^k) - \mu^k$

$\mu^k = E(X^k)$

$$E(\hat{\mu}^k) = E(n^{-1} \sum_{i=1}^{n} X_i^k) = n^{-1} E(\sum_{i=1}^{n} X_i^k)$$

$$= n^{-1} \sum_{i=1}^{n} E(X_i^k) = n^{-1} \cdot n E(X_i^k) = E(X_i^k)$$

Hence, $E(\hat{\mu}^k) = E(X_i^k) = \mu^k$

$Bias(\hat{\mu}^k) = E(\hat{\mu}^k) - \mu^k = \mu^k - \mu^k = 0$

(since X_i is iid r.v. from the population, thus has same expectation as population mean.)

(2) Thm: under iid assumption, MM estimators are consistent as long as functions relating the estimates to sample moments are continuous.

Maybe restate w/ new information

Let $X_1, \ldots, X_n$ are iid r.v.s from population $f(x|\theta)$

Then, by Law of Large Numbers, for any integer k,

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k) = \mu^k \text{ as } n \to \infty.$$

Let $\hat{\mu}_k = g_k(\hat{\theta}_1, \ldots, \hat{\theta}_n)$.

$g_k(\cdot)$ is continuous since $\frac{1}{n} \sum_{i=1}^{n} X_i^k$ is the sum of iid random variables and itself is a random variable.

Hence, assuming $f(X|\theta)$ is continuous probability distribution, $g(\cdot)$ is a continuous random variable, and thus a continuous function.

Then, $\hat{\theta}_K = g^{-1}(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_K)$

$$g^{-1}(\cdot) = (g(\cdot))^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{K}\right)^{-1} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i^{K}}$$

$$= \frac{n}{\sum_{i=1}^{n} x_i^{K}}$$ which assuming $f(X|\theta)$ is a continuous probability distribution, makes $g^{-1}(\cdot)$ a continuous random variable, and a continuous function.

Thus, since g and g^{-1}, the functions relating MM estimates to sample moments are continuous, we can apply Theorem E. Hence, $\hat{\theta}_{MM}$ is consistent.

This means $\hat{\theta}_{MM} \xrightarrow{P} \theta$ as $n \to \infty$

and hence, $E(\hat{\theta}_{MM}) \to \theta$ as $n \to \infty$.

2) (1) $X_1, \dots, X_n$ are iid samples from Kumaraswamy (β)

pdf is $f(x|\beta) = \begin{cases} \beta(1-x)^{\beta-1} & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$

$$\mu_1 = E(x) = \int_0^1 \beta(1-x)^{\beta-1}\, x\, dx$$

$$\mu_1 = E(X) = \beta \int_0^1 x(1-x)^{\beta-1}$$

Hint: $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} = Beta(\alpha,\beta)$

Here, let $\alpha = 2$, $\beta = \beta$

$$E(X) = \beta\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \beta\int_0^1 x(1-x)^{\beta-1}dx$$

this integral $= Beta(2,\beta)$

$$\cancel{E(x) = \beta\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\right]_0^1}$$

where $\Gamma(.)$ is gamma function

$$\cancel{\Rightarrow E(x) = \beta\left[\frac{\Gamma(2+\beta)}{\Gamma(2)\Gamma(\beta)} x(1-x)^{\beta-1}\right]}$$

$$\mu_1 = E[X] = \beta\left[Beta(2,\beta)\right]$$

$$\mu_1 = E(X) = \beta\left[\frac{\Gamma(2)\Gamma(\beta)}{\Gamma(2+\beta)}\right]$$ since 2 is positive integer

$$\mu_1 = E(X) = \beta\left[\frac{2+\beta}{2\beta} \cdot \frac{1}{\binom{2+\beta}{2}}\right]$$

$\hat{\mu}_1 = \bar{X}_n$

$$\bar{X}_n = \hat{\beta}_{MM}\left[\frac{2+\hat{\beta}_{MM}}{2\hat{\beta}_{MM}} \cdot \frac{1}{\frac{(2+\hat{\beta}_{MM})!}{2!(\hat{\beta}_n)!}}\right]$$

$$\bar{X}_n = \left(\frac{2+\hat{\beta}_{MM}}{2}\right)\left[\frac{2(\hat{\beta}_{MM})!}{(2+\hat{\beta}_{MM})!}\right]$$

$$\bar{X}_n = (2+\hat{\beta}_{MM})\left[\frac{\cancel{(\hat{\beta}_{MM})!}}{\cancel{(\hat{\beta}_{MM})!}(\hat{\beta}_{MM}+1)(\hat{\beta}_{MM}+2)}\right]$$

$$\bar{X}_n = \frac{\cancel{2+\hat{\beta}_{MM}}}{(\hat{\beta}_{MM}+1)\cancel{(\hat{\beta}_{MM}+2)}}$$

$$\bar{X}_n = \frac{1}{(\hat{\beta}_{MM}+1)} \qquad (\hat{\beta}_{MM}+1) = \frac{1}{\bar{X}_n}$$

$$\hat{\beta}_{MM} = \frac{1}{\bar{X}_n} - 1$$

(2) $\bar{X}_n = 0.45$.

$$\Rightarrow \hat{\beta}_{MM} = \frac{1}{0.45} - 1 = 1.222$$

By Delta method,

$$Var[g(\bar{X}_n)] \approx \frac{[g'(\mu)]^2 \sigma^2}{N}$$

$$Var\left[\frac{1}{\bar{X}_n} - 1\right] \approx \frac{\left[-\frac{1}{\mu^2}\right]^2 \sigma^2}{100}$$

where $\mu = \frac{\beta \, \Gamma(1+\frac{1}{2})\, \Gamma(\beta)}{\Gamma(1+\frac{1}{2}+\beta)}$

$M_2 = \sigma^2 + \mu^2$ where M_2 is 2nd moment of Kumaraswamy distribution this

$\sigma^2 = M_2 - \mu^2$

Γ = gamma function

Beta = Beta fn

$$\sigma^2 = \frac{\beta \Gamma(\frac{1+100}{2}) \Gamma(\beta)}{\Gamma(\frac{1+\beta+100}{2})} = \beta \, Beta\left(\frac{1+100}{2}, \beta\right)$$

Hence $SE(\hat{\beta}_{MM}) = \sqrt{\frac{\frac{1}{\mu^4} \cdot \sigma^2}{100}}$

Where μ and σ^2 are defined above.

3] WTS: $\nabla^2 \ell(\mu,\sigma^2) < 0$ where ℓ is the normal(μ,σ^2) log-likelihood.

Likelihood fn: $L(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$

log-likelihood fn: $\ell(\mu,\sigma^2) = \log(L(\mu,\sigma^2)) = \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\right)$

$$\ell(\mu,\sigma^2) = \sum_{i=1}^{n} -\tfrac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$ [result from lecture 3]

$$\frac{\partial \ell}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} -2(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$-\frac{\sigma^{-2}}{2} = -\frac{-2\sigma^{-3}}{2} = \sigma^{-3}$

$$\frac{\partial \ell}{\partial \sigma} = \sum_{i=1}^{n} +\tfrac{1}{2}\left[\frac{1}{2\pi\sigma^2}\cdot 4\pi\sigma\right] + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \ell}{\partial \sigma} = \sum_{i=1}^{n} +\frac{1}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{+n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial^2 \ell}{\partial \mu^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n} -1 = \frac{-n}{\sigma^2}$$

$-n\sigma^{-1} = n\sigma^{-2}$

$\sigma^{-3} = -3\sigma^{-4}$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = -\frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\nabla^2 \ell = \begin{bmatrix} \frac{\partial^2 \ell}{\partial \mu^2} & \frac{\partial^2 \ell}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \sigma \partial \mu} & \frac{\partial^2 \ell}{\partial \sigma^2} \end{bmatrix}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \sigma} = \frac{-2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \mu} = \frac{1}{\sigma^3}\sum_{i=1}^{n} -2(x_i-\mu) = \frac{-2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)$$

$$\nabla^2 \ell(\mu,\sigma^2) = \begin{bmatrix} -\frac{n}{\sigma^2} & \frac{-2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu) \\ \frac{-2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu) & -\frac{n}{\sigma^2}-\frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 \end{bmatrix}$$

$$\det(\nabla^2\ell) = \frac{-n}{\sigma^2}\left[\frac{n}{\sigma^2}-\frac{3}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2\right] - \left[\frac{-2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)\cdot\frac{-2}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)\right]$$

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

$$\det(\nabla^2\ell) = \frac{n}{\sigma^4} + \frac{3n-4}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2$$

$\forall \mu \in \mathbb{R}$ and $\sigma > 0$,
$\det(\nabla^2\ell) > 0$.

~~when does~~

$$\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)\frac{3n-4}{\sigma^6} = \frac{n}{\sigma^4}$$

$$\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)\frac{3n-4}{\sigma^2} = n$$

$$\frac{1}{n}\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)\frac{3n-4}{\sigma^2} = 1$$

$$\sigma^2 \cdot \frac{3n-4}{\sigma^2} = 1$$

$$3n-4 = 1$$

$$3n = 5$$

$$n = \frac{5}{3}$$

From IBM Theorem 3.1 [linked in homework question],
For A $n \times n$ symmetric matrix,
$Q(x) = x^T A x$, the following are equivalent:

(i) $Q(x)$ is negative definite

⋮

(iv) The determinants, $(-1)^k \det(A_k) > 0$ for $1 \le k \le n$.

From week 2 discussion on bcourses, $A = (a_{ij})_{2\times 2}$ is semi-negative definite $\iff a_{11} < 0$, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$.

[alternative statement of (iv)].

Thus, $\nabla^2 \ell(\mu, \sigma^2)$ is semi-negative definite since it satisfies both conditions:

- $\nabla^2 \ell(\mu, \sigma^2)_{11} = -\frac{n}{\sigma^2}$ which is < 0 always since $n > 0$ + $\sigma^2 > 0$.

- $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$ as proven above.

This means by the IBM Theorem 3.1, $Q(x)$ is negative definite, were $Q(x)$ is the related quadratic form to $\nabla^2 \ell(\mu, \sigma^2)$.

Hence, $\nabla^2 \ell(\mu, \sigma^2) < 0$, thus proving the solution is a maximum point.

4) $L(\mu,\lambda) = \prod_{i=1}^{n} \left(\frac{\lambda}{2\pi x_i^3}\right)^{\frac{1}{2}} e^{\left(\frac{-\lambda(x_i-\mu)^2}{2\mu^2 x_i}\right)}$

$$\ell(\mu,\lambda) = \sum_{i=1}^{n} \log\left[\left(\frac{\lambda}{2\pi x_i^3}\right)^{\frac{1}{2}} e^{\left(\frac{-\lambda(x_i-\mu)^2}{2\mu^2 x_i}\right)}\right]$$

$$\ell(\mu,\lambda) = \frac{n}{2}\log(\lambda) - \sum_{i=1}^{n}\log(2\pi x_i^3) - \sum_{i=1}^{n}\frac{\lambda(x_i-\mu)^2}{2\mu^2 x_i}$$

$$\frac{\partial \ell}{\partial \mu} = -\sum_{i=1}^{n}\left[\frac{-2\lambda(x_i-\mu)2\mu^2 x_i - \lambda(x_i-\mu)^2 4\mu x_i}{4\mu^4 x_i^2}\right]$$

$x_i^2+\mu^2-2x_i\mu$

$$\frac{\partial \ell}{\partial \mu} = -\sum_{i=1}^{n}\frac{-\lambda(x_i-\mu)\mu - \lambda(x_i-\mu)^2}{\mu^3 x_i}$$

$$\frac{\partial \ell}{\partial \mu} = \sum_{i=1}^{n}\frac{-\lambda x_i\mu + \lambda\mu^2 - \lambda x_i^2 - \lambda\mu^2 + 2\lambda x_i\mu}{\mu^3 x_i}$$

$$\frac{\partial \ell}{\partial \mu} = \sum_{i=1}^{n}\frac{\lambda x_i\mu - \lambda x_i^2}{\mu^3 x_i} \qquad \frac{\partial \ell}{\partial \mu} = \sum_{i=1}^{n}\frac{\lambda x_i(\mu - x_i)}{\mu^3 x_i}$$

$$\frac{\partial \ell}{\partial \mu} = \sum_{i=1}^{n}\frac{\lambda(\mu - x_i)}{\mu^3} = \frac{n\lambda}{\mu^3}(\bar{x}-\mu)$$

$$\frac{n\lambda}{\mu^3}(\bar{X}-\mu) = 0$$

$$\bar{x}-\mu = 0 \qquad \hat{\mu}_{MLE} = \bar{X}_n$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{2\lambda} - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\mu^2 x_i}$$

$$\frac{n}{2\lambda} - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\mu^2 x_i} = 0$$

$$\frac{n}{2\lambda} = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\mu^2 x_i} \qquad \lambda = \frac{n}{\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\mu^2 x_i}}$$

$$\lambda = \frac{n}{\frac{1}{\bar{X}_n^2}\sum_{i=1}^{n}\left(X_i - 2\bar{X}_n + \frac{\bar{X}_n}{X_i}\right)}$$

$$\lambda = \frac{n}{\frac{1}{\bar{X}_n^2}\left(n\bar{X}_n - 2n\bar{X}_n + \bar{X}_n\sum_{i=1}^{n}\frac{1}{X_i}\right)}$$

$$\lambda = \frac{n}{\frac{1}{\bar{X}_n^2}\left(-n\bar{X}_n + \bar{X}_n\sum_{i=1}^{n}\frac{1}{X_i}\right)}$$

$$\lambda = \frac{n}{\frac{1}{\bar{X}_n}\left(-n + \sum_{i=1}^{n}\frac{1}{X_i}\right)} \qquad \frac{-n}{\bar{X}_n} = \sum_{i=1}^{n} -\frac{1}{\bar{X}_n}$$

$$\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}_n}\right)}$$

5) $f(x_i|\theta) = \begin{cases} 2\theta & x_i=1 \quad n_1 \text{ times} \\ \theta & x_i=2 \quad n_2 \text{ times} \\ 3\theta & x_i=3 \quad n_3 \text{ times} \\ 1-6\theta & x_i=4 \quad n_4 \text{ times} \end{cases}$ $n = n_1+n_2+n_3+n_4$

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = P(x_i=1;\theta)^{n_1} P(x_i=2;\theta)^{n_2} P(x_i=3;\theta)^{n_3} P(x_i=4;\theta)^{n_4}$$

$$L(\theta) = [(2\theta)^{n_1}][\theta^{n_2}][(3\theta)^{n_3}][(1-6\theta)^{n_4}]$$

$$\ell(\theta) = n_1\log(2\theta) + n_2\log(\theta) + n_3\log(3\theta) + n_4\log(1-6\theta)$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = \left(\frac{n_1}{2\theta}\cdot 2\right) + \left(\frac{n_2}{\theta}\right) + \left(\frac{n_3}{3\theta}\cdot 3\right) + \left(\frac{n_4}{1-6\theta}\cdot -6\right) = 0$$

$$\frac{\cancel{2}n_1}{\cancel{2}\theta} + \frac{n_2}{\theta} + \frac{\cancel{3}n_3}{\cancel{3}\theta} - \frac{6n_4}{1-6\theta} = 0$$

$$\frac{n_1+n_2+n_3}{\theta} - \frac{6n_4}{1-6\theta} = 0$$

$$\frac{n_1+n_2+n_3}{\theta} = \frac{6n_4}{1-6\theta}$$

$$(n_1+n_2+n_3)(1-6\theta) = 6n_4\theta$$

$$n_1+n_2+n_3-6n_1\theta-6n_2\theta-6n_3\theta = 6n_4\theta$$

$$n_1+n_2+n_3 = 6\theta(n_1+n_2+n_3+n_4)$$

$$\hat{\theta}_{MLE} = \frac{n_1+n_2+n_3}{6(n_1+n_2+n_3+n_4)}$$

6) (1) From lecture 3, Gamma(α, β)

$\mu = \frac{\alpha}{\beta}, \sigma^2 = \frac{\alpha}{\beta^2}$

Hence,

$\hat{\alpha}_{MM} = \frac{0.9537^2}{0.4103} \approx 2.217$ $\Rightarrow \hat{\alpha}_n = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$

$\hat{\beta}_{MM} = \frac{0.9537}{0.4103} \approx 2.324$ $\hat{\beta}_n = \frac{\bar{X}_n}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$

(2) We assume Gamma(2.217, 2.324) is population and generate simulations.

8 tests: $SE(\hat{\alpha}_{MM}) = 0.31$

50 tests: $SE(\hat{\alpha}_{MM}) = 0.29$

100 tests: $SE(\hat{\alpha}_{MM}) = 0.28$

400 tests: $SE(\hat{\alpha}_{MM}) = 0.26$

800 tests: $SE(\hat{\alpha}_{MM}) = 0.26$

1000 tests: $SE(\hat{\alpha}_{MM}) = 0.26$

(3) As seen in the Rshiny app and above in (2), SE stabilizes after 400.

Hence, $SE(\hat{\alpha}_{MM}) = 0.26$

[7] $\hat{\alpha}_{MLE} = 2.175879$ ← computed from attached code file

$\hat{\beta}_{MLE} = 2.186432$

95% bootstrap CI's for α & β separately using MLE

α : [1.777758, 2.574000] ← computed from code file

β : [1.816150, 2.756714]

8] (1) see attached code file for histogram and observation generation.

(2) $\hat{\mu}_{MM} = \bar{X}_n$ $\hat{\sigma}_{MM} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$

$\hat{\mu}_{MM} = 1.793418$ ⟵ see code attached for observations that generated the mean and SD of sample.

$\hat{\sigma}_{MM} = \sqrt{3.624037}$

≈ 1.90369

[3] (a) $\hat{\sigma}^2 = 3.624037$ (from attached code)

$SE(\hat{\mu}_{MM}) = \frac{\sigma}{\sqrt{n}} = \frac{1.90369}{\sqrt{150}} \approx 0.1554$

(b) start from $N(1.793418, 3.624037)$
attached code generates simulations and statistics

approximate SE's:

$SE(\hat{\mu}_{MM}) = 0.1495415$

$SE(\hat{\sigma}_{MM}) = 0.130177$

(C) please see attached code for marked histograms.

They are both in surprising places, as they're both to the right of the center of their respective histograms.

(d) Comparing (a)'s $SE(\hat{\mu}_{MM})$ of 0.1554 and (b)'s $SE(\hat{\mu}_{MM})$ of 0.1995, they are relatively close to each other, with (b) producing a higher estimated standard error.

9] Def: statistic $T(X_n)$ is sufficient for population parameter θ if $X_n | T(X_n)$ is independent of θ.

(1) ~~the sufficient statistic for θ is the maximum likelihood estimate of θ.~~

$X_1, \dots, X_n$ are random sample from pdf

$$f(x|\theta) = \theta x^{-2}, \quad 0 < \theta \leq x < \infty$$

$$f(x_1, \dots, x_n | \theta) = (\theta x^{-2})^n$$

$$f(x_1, \dots, x_n | \theta) = \theta^n x^{-2n}$$

Let $\bar{x} = \frac{1}{n} \sum x_i$ be the potential sufficient statistic.

$$f(x_1, \dots, x_n | \theta) = \theta^n \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^{-2n}$$

since we can factor the joint pdf into two components

$$g[T(X_n), \theta] = \theta^n \left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)^{-2n}$$

and

$$h(X_n) = 1$$

By factorization theorem, the sample mean $(\bar{x})$ is the sufficient statistic for θ.

$$f(x|\theta) = \theta x^{-2} \qquad 0 < \theta \le x < \infty$$

(2) $\mu_1 = E(X) = \int_\theta^\infty x\,\theta x^{-2}\,dx = \theta \int_\theta^\infty \frac{1}{x}\,dx$

$= \theta \left[\log(x)\right]_\theta^\infty \leftarrow$ this integral is divergent...

$\mu_2 = E(X^2) = \int_\theta^\infty x^2(\theta x^{-2})\,dx = \theta \int_\theta^\infty dx$

$= \theta \left[x\right]_\theta^\infty \leftarrow$ divergent

$\rightarrow \mu_{1/2} = E(\sqrt{x}) = \int_\theta^\infty \sqrt{x}\,\theta x^{-2}\,dx$

from week 2 discussion, bcourses

$= \theta \int_\theta^\infty x^{-\frac{3}{2}}\,dx$

$= \theta \left[\frac{x^{-\frac{3}{2}+1}}{-\frac{3}{2}+1}\right]_\theta^\infty = \theta \left[\frac{x^{-\frac{1}{2}}}{-\frac{1}{2}}\right]_\theta^\infty$

$\mu_{1/2} = \theta \left[\frac{-2}{\sqrt{x}}\right]_\theta^\infty = \theta \left[\frac{2}{\sqrt{\theta}}\right]$

$\mu_{1/2} = 2\sqrt{\theta}$

$\overline{\sqrt{x}} = 2\sqrt{\hat{\theta}_{MM}} \Rightarrow \hat{\theta}_{MM} = \frac{\bar{x}}{4}$

(3) $L(\theta) = \prod_{i=1}^{n} f(x_i | \theta) = \prod_{i=1}^{n} \theta x_i^{-2}$

$\cancel{\ell(\theta) = \sum_{i=1}^{n} \log(\theta x_i^{-2}) = n\log(\theta) - 2\sum_{i=1}^{n} \log(x_i)}$

$\cancel{\frac{\partial \ell(\theta)}{\partial \theta} = \frac{n}{\theta}}$

Notice that since $L(\theta)$ is an increasing function with respect to θ, $L(\theta)$ is maximized when θ is at its maximum.

(using $0 < \theta \leq x < \infty$ condition)

Thus, $\hat{\theta}_{MLE} = \min(x_1, \dots, x_n)$

$$P(X=x|\theta) = \theta^x(1-\theta)^{1-x} \Rightarrow X \sim bin(1,\theta)$$

10) (1) $\mu_1 = E(X) = \theta$

$$\hat{\theta}_{MM} = \bar{X}$$

$$L(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$\ell(\theta) = \sum_{i=1}^{n} \log(\theta^{x_i}(1-\theta)^{1-x_i})$$

$$\ell(\theta) = \sum_{i=1}^{n} x_i \log(\theta) + \sum_{i=1}^{n}(1-x_i)\log(1-\theta)$$

$$\ell(\theta) = \log(\theta)\sum_{i=1}^{n} x_i + \log(1-\theta)\sum_{i=1}^{n}(1-x_i)$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{\sum_{i=1}^{n}(1-x_i)}{1-\theta} = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{\theta} = \frac{\sum_{i=1}^{n}(1-x_i)}{1-\theta}$$

$$\theta\sum_{i=1}^{n}(1-x_i) = \sum_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i$$

$$\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n}(1-x_i) + \sum_{i=1}^{n} x_i} = \frac{n\bar{x}}{n(1-\bar{x}) + n\bar{x}} = \frac{n\bar{x}}{n} = \bar{x}$$

$$\hat{\theta}_{MLE} = \bar{x}$$

under iid, $E(\hat{\theta}_{MM}) = E(\bar{x}) = \theta$

(2) $MSE(\hat{\theta}_{MM}) = var(\hat{\theta}_{MM}) + [E(\hat{\theta}_{MM}) - \theta]^2$ ← definition from lec1

$MSE(\hat{\theta}_{MM}) = var(\hat{\theta}_{MM}) + [\theta - \theta]^2$

$MSE(\hat{\theta}_{MM}) = Var(\bar{x}) = \frac{\theta(1-\theta)}{n}$ ← $var(\bar{x}) = \frac{\sigma^2}{n}$, bernoulli variance is $p(1-p)$

$MSE(\hat{\theta}_{MLE}) = Var(\hat{\theta}_{MLE}) + [E(\hat{\theta}_{MLE}) - \theta]^2$

$= var(\bar{X}) + [\theta - \theta]^2$

$MSE(\hat{\theta}_{MLE}) = \frac{\theta(1-\theta)}{n}$

(3) I prefer the MLE estimator to the MM estimator!

- under assumptions that hold here, since the MLE estimator has asymptotic normality, the MLE is asymptotically efficient, which is better than the MM estimator's guarantee of consistency
- MLE always achieves Cramer-Rao lower bound asymptotically, whereas MM doesn't guarantee this.
- MM estimators give unrealistic estimates for small sample sizes.
- MM estimators sometimes ignore relevant information in sample.