

Homework_11

November 18, 2019

Probability for Data Science

UC Berkeley, Fall 2019

Ani Adhikari and Jim Pitman

CC BY-NC 4.0

1 Homework 11

1.0.1 Instructions

Your homeworks have two components: a written portion and a portion that also involves code. Written work should be completed on paper, and coding questions should be done in the notebook. You are welcome to LaTeX your answers to the written portions, but staff will not be able to assist you with LaTeX related issues. It is your responsibility to ensure that both components of the homework are submitted completely and properly to Gradescope. Refer to the bottom of the notebook for submission instructions.

```
[7]: from datascience import *
from prob140 import *
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
%matplotlib inline
import math
from scipy import stats
```

1.0.2 1. Functions of Exponential Random Variables

Let X and Y be independent exponential random variables with rates λ and μ respectively.

- a) Let $W = \min(X, Y)$. Find the distribution of W . Start by remembering which of the following is easiest when working with minima: the density, the cdf, or the survival function.
- b) Let c be a positive constant. Use the results about exponential random variables in Sections 16.1 and 17.2 of the Prob 140 textbook to find $P(X > cY)$ without integration.
- c) Use part (b) to find the cdf of $\frac{X}{Y}$.

1.0.3 2. Gamma Tail Bound

Let X have the gamma (r, λ) distribution.

a) Show that $P(X \geq 2E(X)) \leq (2/e)^r$.

[Remember that if h is a positive function of x , then the value of x that minimizes h is the same as the value of x that minimizes $\log(h)$.]

b) [CODE] Fix $\lambda = 1$. Display overlaid plots of the following three graphs as functions of r , for r in the interval $(0.5, 15)$:

- $P(X \geq 2E(X))$
- $(2/e)^r$
- Markov's bound on $P(X \geq 2E(X))$

The code uses `plt.plot` which is also used in Lab 9. The expression `stats.gamma.cdf(x, r, scale=1)` evaluates to the cdf of the gamma $(r, 1)$ distribution at the point x .

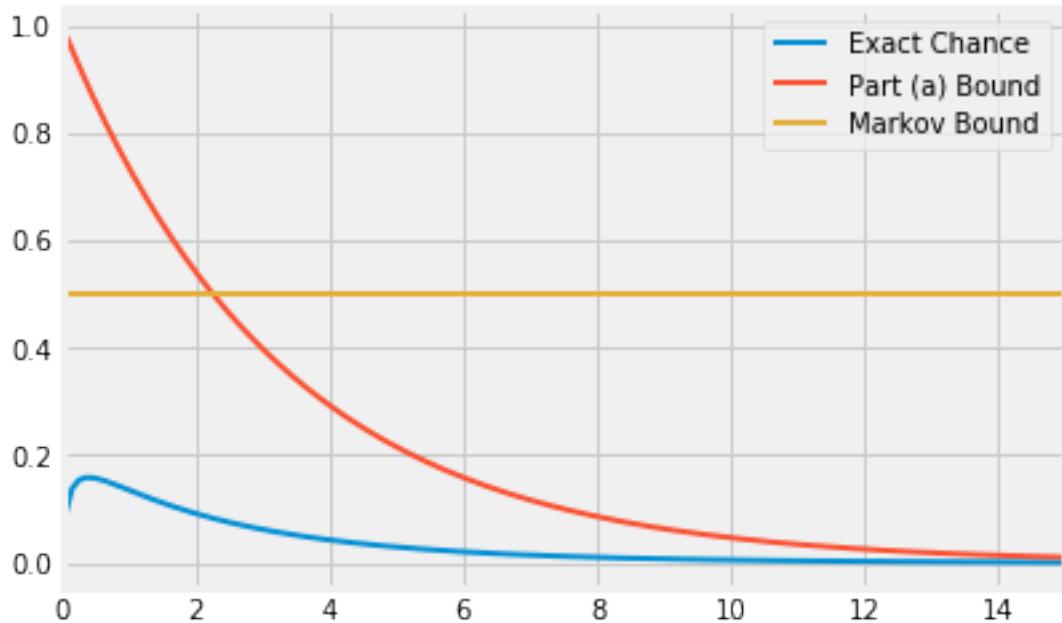
```
[8]: # Answer to 2b
r = np.arange(0.05, 15, 0.1)

markov_bound = r/(2*r)

part_a_bound = (2/np.e)**r

exact = 1 - stats.gamma.cdf(2*r,r,scale=1)
#...

plt.plot(r, exact, lw=2, label='Exact Chance')
plt.plot(r, part_a_bound, lw=2, label='Part (a) Bound')
plt.plot(r, markov_bound, lw=2, label='Markov Bound')
plt.legend()
plt.xlim(0, 15);
```



c) For which values of r is the bound in part (a) better than Markov's bound? Show the analytical calculation. Then use a code cell to find the numerical range and confirm that it is consistent with the graphs in (b).

```
[9]: # Calculation for 2c
#(2/e)**r smaller than 1/2
r_value = (np.log(1/2))/(np.log(2*np.e))
r_value
```

[9]: 2.2588913532709296

1.0.4 3. Waiting for a Random Coin to Land Heads

Let X be a random proportion. Given $X = p$, let T be the number of tosses till a p -coin lands heads.

- a) Let $P(X = 1/10) = 1/4$, $P(X = 1/7) = 1/4$, and $P(X = 1/3) = 1/2$. Find $E(T)$.
- b) Find $E(T)$ if X has the beta (r, s) density for some $r > 1$. Simplify all integrals and Gamma functions in your answer.
- c) Let X have the beta (r, s) density. For fixed $k > 0$, find the posterior density of X given $T = k$. Identify it as one of the famous ones and state its name and parameters.

1.0.5 4. Discrete and Continuous Random Selections

Fix a positive integer n , and let p be strictly between 0 and 1.

Suppose Person A picks a number uniformly in the interval $(0, n)$. Let X be the number Person A picks.

Suppose that independently of Person A, Person B picks an integer Y according to the binomial (n, p) distribution, for example by using `stats.binom.rvs(n, p, size=1)` or by tossing a p -coin n times and recording the number of heads.

Find $P(X < Y)$.

1.0.6 5. Maximum Likelihood Estimate of the Exponential Rate

For $n > 1$, let X_1, X_2, \dots, X_n be i.i.d. exponential (λ) variables.

a) Let $\hat{\lambda}_n$ be the maximum likelihood estimate (MLE) of the parameter λ . Find $\hat{\lambda}_n$ in terms of the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. The subscript n in \bar{X}_n is there to remind us that we have the average of n values. It doesn't refer to the n th sampled value X_n .

b) This part is a simulation study in which you already know λ and are examining how close your estimate comes to this known value. Take $\lambda = 0.25$. We have assigned it the name `lam` because `lambda` means something else in Python.

Repeat the following simulation 10,000 times. - Generate a sample of 400 i.i.d. exponential (`lam`) variables. Use `stats.expon.rvs(scale = 1/lam, size=400)`. - Calculate your MLE $\hat{\lambda}_{400}$ based on this sample.

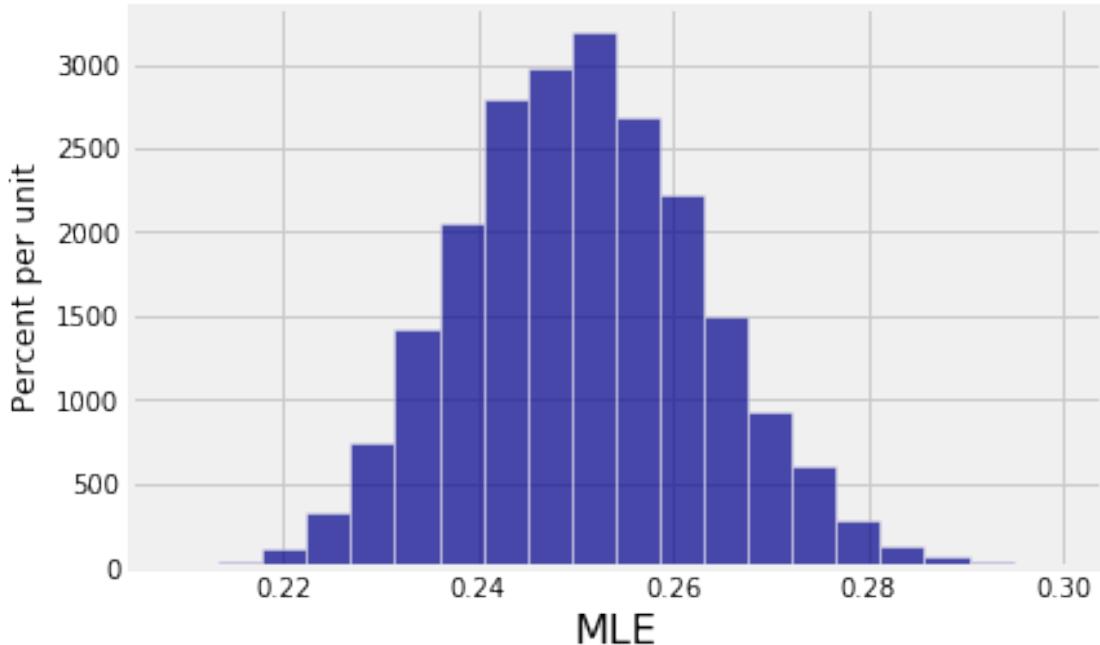
Then draw the histogram and find the mean of your 10,000 MLEs.

```
[10]: # Your answer to 5b
lam = 0.25

# Define a function that returns the MLE based on
# one i.i.d. sample of size 400
# from the exponential distribution with lambda = 0.25
def mle_from_one_sample():
    sample = stats.expon.rvs(scale = 1/lam, size=400)
    sample_mean = np.mean(sample)
    return 1/sample_mean

# Construct an array of 10,000 MLEs
mle_400 = make_array()
for i in range(10000):
    one_value = mle_from_one_sample()
    mle_400 = np.append(mle_400, one_value)

# Plot a histogram of the 10,000 MLEs
Table().with_column('MLE', mle_400).hist(bins=20)
```



```
[11]: # The mean of your 10000 mle's
bootstrap_mean = np.mean(mle_400)
bootstrap_mean
```

[11]: 0.25070524038503783

c) Use facts about sums and linear transformations to find the distribution of \bar{X}_n with little or no calculation. Recognize it as one of the famous ones and provide its name and parameters. Use it to find $E(\hat{\lambda}_n)$.

d) Is $\hat{\lambda}_n$ an unbiased estimate of λ ? If it is biased, does it overestimate on average, or does it underestimate? Is it asymptotically unbiased? That is, does $E(\hat{\lambda}_n)$ converge to λ as $n \rightarrow \infty$?

e) Your simulation in b) should show an approximately normal distribution for $\hat{\lambda}_n$. The Central Limit Theorem says the distribution of \bar{X}_n should be approximately normal, but is $\hat{\lambda}_n$ a linear function of \bar{X}_n ? If not, why is its distribution approximately normal?

[This one needs a little thought but no calculation. Think about what a smooth function looks like over a small interval in its domain, and then think about why small intervals are relevant here.]

1.1 Submission Instructions

Many assignments throughout the course will have a written portion and a code portion. Please follow the directions below to properly submit both portions.

1.1.1 Written Portion

- Scan all the pages into a PDF. You can use any scanner or a phone using an application. Please **DO NOT** simply take pictures using your phone.
- Please start a new page for each question. If you have already written multiple questions on the same page, you can crop the image or fold your page over (the old-fashioned way). This helps expedite grading.
- It is your responsibility to check that all the work on all the scanned pages is legible.

1.1.2 Code Portion

- Save your notebook using File > Save and Checkpoint.
- Generate a PDF file using File > Download as > PDF via LaTeX. This might take a few seconds and will automatically download a PDF version of this notebook.
 - If you have issues, please make a follow-up post on the general HW 11 Piazza thread.

1.1.3 Submitting

- Combine the PDFs from the written and code portions into one PDF. [Here](#) is a useful tool for doing so.
- Submit the assignment to Homework 11 on Gradescope.
- **Make sure to assign each page of your pdf to the correct question.**
- **It is your responsibility to verify that all of your work shows up in your final PDF submission.**

1.1.4 We will not grade assignments which do not have pages selected for each question.

hw 11

1.

a.

\Rightarrow let $W = \min(X, Y) \rightarrow$ for minima, use survival-func.

$$\Rightarrow P(W > w) = P(\min(X, Y) > w)$$

$\Rightarrow \min(X, Y) > w \rightarrow$ other func $>$ smaller func $> w$

$\Rightarrow P(W > w) = P(X > w, Y > w) \text{ & independent } X, Y$

$$= P(X > w) * P(Y > w)$$

\Rightarrow since $X \& Y$ are exponential dist

$$P(X > w) = e^{-\lambda w}$$

$$P(Y > w) = e^{-\mu w}$$

$$\Rightarrow P(X > w, Y > w) = P(X > w) * P(Y > w)$$

$$= e^{-\lambda w} * e^{-\mu w}$$

$$= e^{-(\lambda + \mu)w}$$

$$\Rightarrow P(W = w) = \text{survival-func} = e^{-(\lambda + \mu)w}$$

\Rightarrow exponential $(\lambda + \mu)$ dist

\Rightarrow thus, the dist of W is exponential $(\lambda + \mu)$ dist.

b.

$$\Rightarrow P(X > Y) = \frac{\mu}{\lambda + \mu} \quad \leftarrow \text{section 17.2}$$

$\Rightarrow cY$ is linear transformation of Y

$$\begin{aligned}\Rightarrow E(cY) &= c * E(Y) = c * \frac{\mu}{\lambda} \\ &= \frac{c}{\lambda}.\end{aligned}$$

$$\begin{aligned}\Rightarrow cY &= \text{exponential} \left(\frac{1}{\frac{c}{\lambda}} \right) \quad \leftarrow \text{section 16.1} \\ &= \text{exponential} \left(\frac{\lambda}{c} \right) \text{ dist}\end{aligned}$$

$$\begin{aligned}\Rightarrow P(X > cY) &= \frac{\mu}{c} / (\lambda + \frac{\mu}{c}) \\ &= \frac{\mu}{c\lambda + \mu}\end{aligned}$$

$$\Rightarrow \text{thus, } P(X > cY) = \frac{\mu}{c\lambda + \mu}$$

c.

$$\Rightarrow \text{let } \frac{X}{Y} = k$$

$$\Rightarrow \text{cdf} = P\left(\frac{X}{Y} < k\right)$$

$$= P(X < kY) \quad \leftarrow \text{non-negative } Y.$$

$$\Rightarrow P(X < Y) = P(Y > X) = \frac{\lambda}{\lambda + \mu} \quad \leftarrow \text{section 16.1}$$

$$\Rightarrow kY = \text{exponential} \left(\frac{\mu}{k} \right) \text{ dist}$$

$$\Rightarrow P(X < kY) = P(kY > X) = \lambda / (\lambda + \frac{\mu}{k}).$$

$$= \frac{\lambda k}{\lambda k + \mu} \leftarrow \text{part b.}$$

\Rightarrow thus, the cdf of $\frac{X}{Y}$ is

$$P\left(\frac{X}{Y} < k\right) = \frac{\lambda k}{\lambda k + \mu}.$$

2.

a.

$$\Rightarrow P(X \geq c) \leq \min_{t>0} \left(\frac{M_X(t)}{e^{tc}} \right).$$

\Rightarrow for gamma(r, λ) dist

$$E(X) = \frac{r}{\lambda}$$

$$\Rightarrow 2E(X) = \frac{2r}{\lambda}.$$

$$\Rightarrow M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} \frac{\lambda^r}{I(r)} x^{r-1} e^{-\lambda x} dx \\ = \left(\frac{\lambda}{\lambda-t}\right)^r$$

$$\Rightarrow P(X \geq 2E(X)) \leq \min_{t>0} \left(\frac{M_X(t)}{e^{t+2E(X)}} \right).$$

$$\Rightarrow P(X \geq \frac{2r}{\lambda}) \leq \min_{t>0} \left(\frac{\left(\frac{\lambda}{\lambda-t}\right)^r}{e^{\frac{2rt}{\lambda}}} \right)$$

$\Rightarrow \left(\frac{\lambda}{\lambda-t}\right)^r$ & $e^{\frac{2rt}{\lambda}}$ are increasing func

so $\frac{\left(\frac{\lambda}{\lambda-t}\right)^r}{e^{\frac{2rt}{\lambda}}}$ is increasing func

$$\begin{aligned}
\Rightarrow \log\left(\frac{\left(\frac{\lambda}{\lambda-t}\right)^r}{e^{\frac{2rt}{\lambda}}}\right) &= \ln\left(\left(\frac{\lambda}{\lambda-t}\right)^r\right) - \ln(e^{\frac{2rt}{\lambda}}) \\
&= r * \ln\left(\frac{\lambda}{\lambda-t}\right) - \frac{2rt}{\lambda} \\
&= r * \left(\ln\left(\frac{\lambda}{\lambda-t}\right) - \frac{2t}{\lambda}\right). \\
&= r * \ln(\lambda) - r * \ln(\lambda-t) - 2rt\lambda^{-1}
\end{aligned}$$

\Rightarrow minimum at $f'(x)=0$.

$$\begin{aligned}
\Rightarrow f'(x) &= 0 - r * \frac{1}{\lambda-t} * -\lambda - 2r\lambda^{-1} \\
&= r * \frac{1}{\lambda-t} - \frac{2r}{\lambda} \\
\Rightarrow \frac{r}{\lambda-t} - \frac{2r}{\lambda} &= 0 \rightarrow \frac{1}{\lambda-t} = \frac{2}{\lambda} \rightarrow \lambda = 2\lambda - 2t
\end{aligned}$$

$$\Rightarrow t = \frac{\lambda}{2}$$

\Rightarrow so the entire func is minimum at $t = \frac{\lambda}{2}$

plug it back to the original func

$$\begin{aligned}
\Rightarrow \frac{\left(\frac{\lambda}{\lambda-t}\right)^r}{e^{\frac{2rt}{\lambda}}} &= \left(\frac{\frac{\lambda}{\lambda-\frac{\lambda}{2}}}{e^{\frac{2r\frac{\lambda}{2}}{\lambda}}}\right)^r \\
&= \left(\frac{\frac{2}{1}}{e}\right)^r
\end{aligned}$$

\Rightarrow the minimum of the func is $\left(\frac{2}{e}\right)^r$ at $t = \frac{\lambda}{2}$.

\Rightarrow thus, the chernoff bound is

$$P(X \geq 2E(X)) \leq \left(\frac{2}{e}\right)^r.$$

c.

\Rightarrow for part (a) bound better than markov's bound.

chernoff bound - exact < markov bound - exact

\Rightarrow chernoff bound < markov bound.

\Rightarrow chernoff bound = $(\frac{2}{e})^r \leftarrow$ part a

markov bound = $\frac{\mathbb{E}(X)}{2\mathbb{E}(X)} = \frac{1}{2}$.

$\Rightarrow (\frac{2}{e})^r < \frac{1}{2}$

$\Rightarrow r \ln(\frac{2}{e}) < \ln(\frac{1}{2})$

$r > \frac{\ln(\frac{1}{2})}{\ln(\frac{2}{e})} \leftarrow \ln(\frac{2}{e}) < 0$

$\Rightarrow r > 2.2589$.

\Rightarrow thus, for bound in part a. to be better than markov's

bound, $r > 2.2589$

3.

a.

\Rightarrow at $x = \frac{1}{10}$

$$\mathbb{E}(\tau | x = \frac{1}{10}) = \frac{1}{\frac{1}{10}} = 10 \quad \leftarrow \text{chapter 9}$$

\Rightarrow at $x = \frac{1}{7}$

$$E(T | x = \frac{1}{7}) = \frac{1}{\frac{1}{7}} = 7$$

at $x = \frac{1}{3}$

$$E(T | x = \frac{1}{3}) = \frac{1}{\frac{1}{3}} = 3$$

$$\Rightarrow E(T) = 10 * \frac{1}{4} + 7 * \frac{1}{4} + 3 * \frac{1}{2}$$

$$= \frac{23}{4}$$

\Rightarrow thus, $E(T) = \frac{23}{4}$ under given condition.

b.

$$\begin{aligned}\Rightarrow E(T) &= \int_0^1 \frac{1}{x} * \frac{(r+s-1)!}{(r-1)!(s-1)!} x^{r-1} (1-x)^{s-1} dx \\ &= \frac{(r+s-1)!}{(r-1)!(s-1)!} \int_0^1 x^{r-2} (1-x)^{s-1} dx. \leftarrow \text{still density.} \\ &= \frac{(r+s-1)!}{(r-1)!(s-1)!} * \frac{(r-2)!(s-1)!}{(r+s-2)!} \\ &= \frac{r+s-1}{r-1}\end{aligned}$$

\Rightarrow thus, for beta dist (r,s) of x

$$E(T) = \frac{r+s-1}{r-1}$$

c.

$\Rightarrow T = \# \text{ of toss till head}$

$= I_1 + I_2 + \dots \text{ & } I_1 = I_2 = I_n = 0 \text{ & each } I_i = \text{bernoulli}(p).$

\Rightarrow in the total k toss \rightarrow 1 head & $k-1$ tail.

$$\Rightarrow f(x|T=k)(p) \propto \binom{k}{0} (1-p)^k * p^0 * \frac{(r+s-1)!}{(r-1)! (s-1)!} p^{r-1} (1-p)^{s-1}$$

$$\propto \text{constant} * p^{r-1} (1-p)^{s+k-1}$$

$\Rightarrow f(x|T=k)(p) = \text{beta}(r, s+k) \text{ density} \leftarrow \text{chapter 21.}$

$$\Rightarrow f(x|T=k)(p) = \frac{(r+s+k-1)!}{(r-1)! (s+k-1)!} p^{r-1} (1-p)^{s+k-1}$$

\Rightarrow thus, posterior dist of X (given $T=k$) is

beta($r, s+k$) dist, and it's density

$$f(x|T=k)(p) = \frac{(r+s+k-1)!}{(r-1)! (s+k-1)!} p^{r-1} (1-p)^{s+k-1}.$$

4.

$$\Rightarrow P(X < Y) = \sum_0^n P(X < Y | Y=y) * P(Y=y) \leftarrow \text{chapter 20.}$$

$\Rightarrow X = \text{uniform}(0, n)$ dist

$$P(X < k) = \frac{k}{n}.$$

$\Rightarrow Y = \text{binomial}(n, p)$ dist

$$P(Y=y) = \binom{n}{y} p^y * (1-p)^{n-y}$$

$$\Rightarrow P(X < Y) = \sum_0^n \left(\frac{y}{n}\right) * P(Y=y)$$

$$= \frac{1}{n} * \sum_0^n y * P(Y=y)$$

$$\Rightarrow \mathbb{E}(Y) = \sum_0^n y * P(Y=y) \leftarrow \text{value} * \text{probability.}$$

$$= n * p$$

$$\Rightarrow P(X < Y) = \frac{1}{n} * n * p$$

$$= p$$

$$\Rightarrow \text{thus, } P(X < Y) = p.$$

5.

a.

$$\Rightarrow \text{joint density of } \lambda = \lambda * e^{-\lambda x_i}$$

$$\Rightarrow L(\lambda) = \prod_{i=1}^n (\lambda * e^{-\lambda x_i})$$

$$= \lambda^n * e^{-\lambda T} \quad \& \quad T = \sum_{i=1}^n x_i$$

$$\Rightarrow \text{for MLE } (\lambda) \rightarrow \text{need } \max(L(\lambda))$$

$$\max_{x \geq 0} (f(x)) = \max_{x \geq 0} (\log(f(x)))$$

$$\Rightarrow L(\lambda) = n \ln(\lambda) + (-\lambda T)$$

$$\Rightarrow f(x) \text{ max at } f'(x) = 0.$$

$$\Rightarrow f'(\lambda) = n * \frac{1}{\lambda} - T = 0.$$

$$T = \frac{n}{\lambda}$$

$$\frac{n}{\lambda} = \sum_{i=1}^n x_i$$

$$\lambda = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\sum_{i=1}^n x_i/n}$$

$$\lambda = \frac{1}{\bar{x}_n}$$

$$\Rightarrow \hat{\lambda}_n = \frac{1}{\bar{x}_n}$$

\Rightarrow thus, MLE for parameter λ

$$\therefore \hat{\lambda}_n = \frac{1}{\bar{x}_n}$$

c.

$$\Rightarrow \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

\Rightarrow since each $x_i = \text{exponential}(\lambda) \text{ dist} = \text{gamma}(1, \lambda) \text{ dist}$

$$\sum_{i=1}^n x_i = \text{gamma}(n, \lambda) \text{ dist} \leftarrow \text{chapter 18.2}$$

$$\Rightarrow \frac{1}{n} * \sum_{i=1}^n x_i = \text{gamma}\left(n, \frac{\lambda}{n}\right) \text{ dist} \leftarrow \text{chapter 18.2}$$

$$\Rightarrow \bar{x}_n = \text{gamma}(n, n\lambda) \text{ dist}$$

$$\begin{aligned} \Rightarrow E(\hat{\lambda}_n) &= E\left(\frac{1}{\bar{x}_n}\right) = \int_0^\infty x^{-1} * \frac{(n\lambda)^n}{(n-1)!} x^{n-1} e^{-(n\lambda)x} dx \\ &= \int_0^\infty \frac{(n\lambda)^n}{(n-1)!} x^{n-2} e^{-(n\lambda)x} dx \\ &= \int_0^\infty \frac{(n\lambda)^n}{(n-1)!} x^{n-2} e^{-(n\lambda)x} * \frac{(n-1)}{(n-1)} dx \\ &= \frac{n\lambda}{(n-1)} * \int_0^\infty \frac{(n\lambda)^{n-1}}{(n-2)!} x^{n-2} e^{-(n\lambda)x} dx \\ &= \frac{n\lambda}{(n-1)} * \text{area under gamma}(n-1, n\lambda) \text{ dist} \end{aligned}$$

$$= \frac{n\lambda}{n-1}$$

$$= \frac{n\lambda}{n-1}$$

$$\Rightarrow E(\hat{\lambda}_n) = E\left(\frac{1}{\bar{x}_n}\right) = \frac{n\lambda}{n-1}$$

\Rightarrow thus, the dist of \bar{x}_n is gamma($n, n\lambda$) dist

$$\text{and } E(\hat{\lambda}_n) = \frac{n\lambda}{n-1}$$

d.

$$\Rightarrow E(\hat{\lambda}_n) = \frac{n}{n-1} * \lambda \neq \lambda$$

$$\Rightarrow \text{since } \frac{n}{n-1} < 0$$

$\Rightarrow \hat{\lambda}_n$ bias & over-estimate.

\Rightarrow as $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} E(\hat{\lambda}_n) = \lim_{n \rightarrow \infty} \frac{n}{n-1} * \lambda$$

$$= \lim_{n \rightarrow \infty} \left(\frac{1}{1 - \frac{1}{n}} * \lambda \right)$$

$$= \lambda$$

$\Rightarrow \lim_{n \rightarrow \infty} E(\hat{\lambda}_n) = \lambda$ & asymptotically unbiased.

\Rightarrow thus, $\hat{\lambda}_n$ is biased & over-estimate,

and $\hat{\lambda}_n$ is asymptotically unbiased.

e. $\rightarrow CLT = \bar{X}_n$ normal dist
 $\rightarrow \hat{\lambda}_n$ linear func of X_n & why normal.
 $\Rightarrow \hat{\lambda}_n = \frac{1}{\bar{X}_n} = (\bar{X}_n)^{-1}$

$\Rightarrow \hat{\lambda}_n$ is not linear func of X_n

$$\hat{\lambda}_n = \frac{1}{\sum X_i} = n \left(\frac{1}{\sum x_i} \right)$$

\Rightarrow however, at really small-interval

$\frac{1}{X}$ is approximately linear

\Rightarrow so $X \rightarrow \frac{1}{X}$ is approximately linear function (transformation).

\Rightarrow dist of linear transformation of a normal dist variable

is also normal dist

$\Rightarrow \hat{\lambda}_n = \frac{1}{\bar{X}_n}$ is linear transformation & normal dist at
really small interval

$\Rightarrow \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow$ as sample size $n \uparrow$

$\text{Var}(\bar{X}_n) \downarrow$ & $\text{SD}(\bar{X}_n) \downarrow$

\Rightarrow as Variance & SD \downarrow , the interval size \downarrow

\Rightarrow in part b, $n=400 \rightarrow$ large sample size & large SD.

so the function is transform in really small intervals.

\Rightarrow thus, $\hat{\lambda}_n = \frac{1}{\bar{X}_n}$ is approximately linear function (transformation)

so $\hat{\lambda}_n$ have approximately normal dist.