(1) Let $\hat{F}_n(x) = w$, $\hat{F}_n(y) = z$, $x \neq y$

$$\text{cov}(w, z) = \mathbb{E}\left[(w - \mathbb{E}(w))(z - \mathbb{E}(z))\right]$$

$$\Rightarrow = \mathbb{E}\left[wz - w\mathbb{E}(z) - z\mathbb{E}(w) + \mathbb{E}(w)\mathbb{E}(z)\right]$$

by linearity of Expectation: $= \mathbb{E}(w)\mathbb{E}(z)$

$$\mathbb{E}[wz] - \mathbb{E}[w\mathbb{E}(z)] - \mathbb{E}[z\mathbb{E}(w)] + \mathbb{E}(w)\mathbb{E}(z)]$$

$$\Rightarrow \text{cov}(w, z) = \mathbb{E}[(w)(z)] + \mathbb{E}(w)\mathbb{E}(z)$$

notice $\mathbb{E}(\hat{F}_n(t)) = F(t) \Rightarrow \mathbb{E}(w) = \mathbb{E}(\hat{F}_n(x)) = F(x)$

$\mathbb{E}(z) = \mathbb{E}(\hat{F}_n(y)) = F(y)$

$$\Rightarrow \text{cov}(\hat{F}_n(x), \hat{F}_n(y)) = \mathbb{E}(\hat{F}_n(x) \cdot \hat{F}_n(y)) - F(x)F(y)$$

$$\mathbb{E}(\hat{F}_n(x) \cdot \hat{F}_n(y)) = \mathbb{E}\left(\frac{1}{n^2} \sum_i H(x - x_i) \sum_j H(y - x_j)\right)$$

$$\Rightarrow = \frac{1}{n^2}\left[\mathbb{E}\left(\sum_{i=j} H(x-x_i)H(y-x_i)\right) + \mathbb{E}\left(\sum_{i \neq j} H(x-x_i)H(y-x_j)\right)\right.$$

↳ n terms with average $F(\min(x,y))$ value

↳ $n(n-1)$ terms with avg $F(x) \cdot F(y)$ value

$$= \frac{1}{n^2}\left[nF(\min(x,y)) + n(n-1)F(x)F(y)\right)$$

$$\text{cov}(\hat{F}_n(x), \hat{F}_n(y)) = \frac{1}{n}F(\min(x,y)) + \frac{n-1}{n}F(x)F(y) - F(x)F(y)$$

$$\boxed{\Rightarrow \text{cov}(\hat{F}_n(x), \hat{F}_n(y)) = \frac{F(\min(x,y)) - F(x)F(y)}{n}}$$

ACM 157 PS3

(2)

$$\kappa_F = \frac{\int (x-\mu_F)^3 \, dF(x)}{\left(\int (x-\mu_F)^2 \, dF(x)\right)^{3/2}}$$

For plug in estimate: substitute $\hat{F}_n(x)$ for $F(x)$

$$\Rightarrow \frac{\int (x-\mu_F)^3 \, d\hat{F}_n(x)}{\left(\int (x-\mu_F)^2 \, d\hat{F}_n(x)\right)^{3/2}} \Rightarrow \int \overset{a(x)}{\hat{\ }} d\hat{F}_n(x) = \frac{1}{n}\sum_{n=1} a(x)$$

This gives: $$\hat{\kappa}_F = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mu}_F)^3}{\left[\frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{\mu}_F)^2\right]^{3/2}}$$

$$\Rightarrow = \frac{\left(\frac{1}{n}\right)}{\left(\frac{1}{n}\right)^{3/2}} \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^3}{\left[\sum_{i=1}^{n} (x_i - \bar{x}_n)^2\right]^{3/2}}$$

$\hat{\mu}_F = \bar{X}_n$ by same logic from above and lecture notes

$$\boxed{\Rightarrow \hat{\kappa}_F = \sqrt{n} \cdot \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^3}{\left[\sum_{i=1}^{n} (x_i - \bar{x}_n)^2\right]^{3/2}}}$$

4

(B)

(a) Given $\theta = \min\{x : F(x) = 1\}$

$\hat{\theta}_n = \min\{x : \hat{F}_n(x) = 1\}$

for data $x_1 \dots x_n$ $\hat{F}_n(x_n) = 1$ and thus is the minimum value since $\hat{F}_n(x_n - \varepsilon) < 1$

thus we conclude

$\boxed{\hat{\theta}_n = X_n}$ where X_n is the nth order statistic in the data sample of size n which is essentially the max of the dataset

(b) $\text{Bias}(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta = \frac{n\theta}{n+1} - \theta = \boxed{\frac{-\theta}{n+1}}$

(c) Subtracting estimated bias from $\hat{\theta}_n$ via jackknife estimation yields

$$\hat{\theta}_n^J = \hat{\theta}_n - \hat{\mathbb{B}}_J[\hat{\theta}_n] = n\hat{\theta}_n - (n-1)\bar{\theta}_n^J$$

$$= nX_n - (n-1)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{-i}\right)$$

$$= nX_n - (n-1)\left[\frac{1}{n}\left[(n-1)X_n + X_{n-1}\right]\right]$$

$$= nX_n - \frac{(n-1)^2}{n}(X_n) - \left(\frac{n-1}{n}\right)(X_{n-1})$$

$$= \frac{n^2 - (n-1)^2}{n}(X_n) - \left(\frac{n-1}{n}\right)(X_{n-1})$$

$$\boxed{\hat{\theta}_n^J = \frac{2n-1(X_n) - (n-1)(X_{n-1})}{n}}$$

(d) $\mathbb{B}[\hat{\theta}_n^J] =$

$= E[\quad E$

$E[\hat{\theta}_n^J] = E[$

by linearity:

$E[\hat{\theta}_n^J] = 2n-1$

$E(X_n) = \frac{n\theta}{n+1}$

$E[\hat{\theta}_n^J] = (2n$

$\Rightarrow = (2n^2-n)$

$= \frac{n^2+n-2}{n^2+n}$

$\mathbb{B}[\hat{\theta}_n^J] -$

$\mathbb{B}[\hat{\theta}_n^J$

(d) $\mathbb{B}[\hat{\theta}_n] = \mathbb{E}[\hat{\theta}_n] - \theta$

$= \cancel{\mathbb{E}[\hat{\theta}_n]}\ \mathbb{E}[\hat{\theta}_n] \Rightarrow$ from (c)

$$\mathbb{E}[\hat{\theta}_n] = \mathbb{E}\left[\frac{2n+2(X_n) + (n-2)(X_{n-2})}{n}\right]$$

by linearity:

$$\mathbb{E}[\hat{\theta}_n] = \frac{2n+1(\mathbb{E}[X_n]) + (n-1)\mathbb{E}(X_{n-2})}{n}$$

$$\mathbb{E}(X_n) = \frac{n\theta}{n+2} \qquad \mathbb{E}(X_{n-2}) = \frac{(n-2)\theta}{n+2}$$

$$\mathbb{E}[\hat{\theta}_n] = \frac{(2n+2)\left(\frac{n\theta}{n+2}\right) + (n-1)\left(\frac{(n-1)\theta}{n+2}\right)}{n}$$

$$\Rightarrow = \frac{(2n^2+n)(\theta) + (n^2-2n+2)(\theta)}{n^2+n}$$

n^2+n-2

$$= \left(\frac{\overset{n^2+n-2}{\cancel{3n^2-3n+2}}}{n^2+n}\right)\theta \Rightarrow \mathbb{B}[\hat{\theta}_n] = \left[\frac{\cancel{3n^2-3n+2}}{n^2+n}\right]\theta - \left[\frac{n^2+n}{n^2+n}\right]\theta$$

$$\mathbb{B}[\hat{\theta}_n] = \left(\frac{\overset{n^2+n-2-n^2-n}{\cancel{2n^2-2n+2}}}{n^2+n}\right)\theta = \frac{-\theta}{n^2+n}$$

$$\boxed{\mathbb{B}[\hat{\theta}_n] = \frac{-\theta}{n^2+n}}$$

(5)(a)

$$B[\hat{\theta}_n] = \mathbb{E}[\hat{\theta}_n] - \theta = \mathbb{E}[\hat{\theta}_n] - e^{\mu}$$

$\mathbb{E}[\hat{\theta}_n]$ given $\hat{\theta}_n = e^{\bar{X}_n}$, since

all X_i are ~~...~~ i.i.d we know $\bar{X}_n$ follows a normal distribution as well with center/mean $= \mu$ and $\mathbb{V}[\bar{X}_n] = \mathbb{V}[\frac{1}{n}\sum X_i]$, since X_i are i.i.d ~~then~~ from the same distribution $\sigma^2 = 1$, we know

$$\mathbb{V}[\tfrac{1}{n}\sum X_i] = \frac{1}{n^2}\sum \mathbb{V}[X_i] = \frac{1}{n^2}n(\sigma^2) = \frac{n}{n^2} = \frac{1}{n}$$

thus $\bar{X}_n \sim N(\mu, \frac{1}{n})$, and $\hat{\theta}_n = e^{\bar{X}_n}$ follows a log-normal distribution $\ln N(\mu, \frac{1}{n})$ and thus:

$$\mathbb{E}(e^{\bar{X}_n}) = e^{\mu + \frac{1}{2n}}$$

The Taylor Expansion of $e^{\mu + \frac{1}{2n}}$ around μ is given by:

$$e^{\mu}\sum_{k=0}^{\infty}\frac{(\frac{1}{2n})^k}{k!} = e^{\mu}\left(1 + \frac{1}{2n} + \frac{1}{8n^2} + O\left(\frac{1}{n^3}\right)\right)$$

Thus $B[\hat{\theta}_n] = \mathbb{E}(e^{\bar{X}_n}) - \theta = e^{\mu}\left(1 + \frac{1}{2n} + \frac{1}{8n^2} + O\left(\frac{1}{n^3}\right)\right) - e^{\mu}$

$\Rightarrow = \dfrac{\frac{1}{2}e^{\mu}}{n} + \dfrac{\frac{1}{8}e^{\mu}}{n^2} + O\left(\frac{1}{n^3}\right)$, thus we've shown

$B[\hat{\theta}_n] = \dfrac{\frac{1}{2}e^{\mu}}{n} + \dfrac{\frac{1}{8}e^{\mu}}{n^2} + O\left(\frac{1}{n^3}\right)$ completing our proof that the Jackknife assumption is held with $a = \frac{1}{2}e^{\mu}$, $b = \frac{1}{8}e^{\mu}$

```
clc; clear; close all;
```

Problem 3

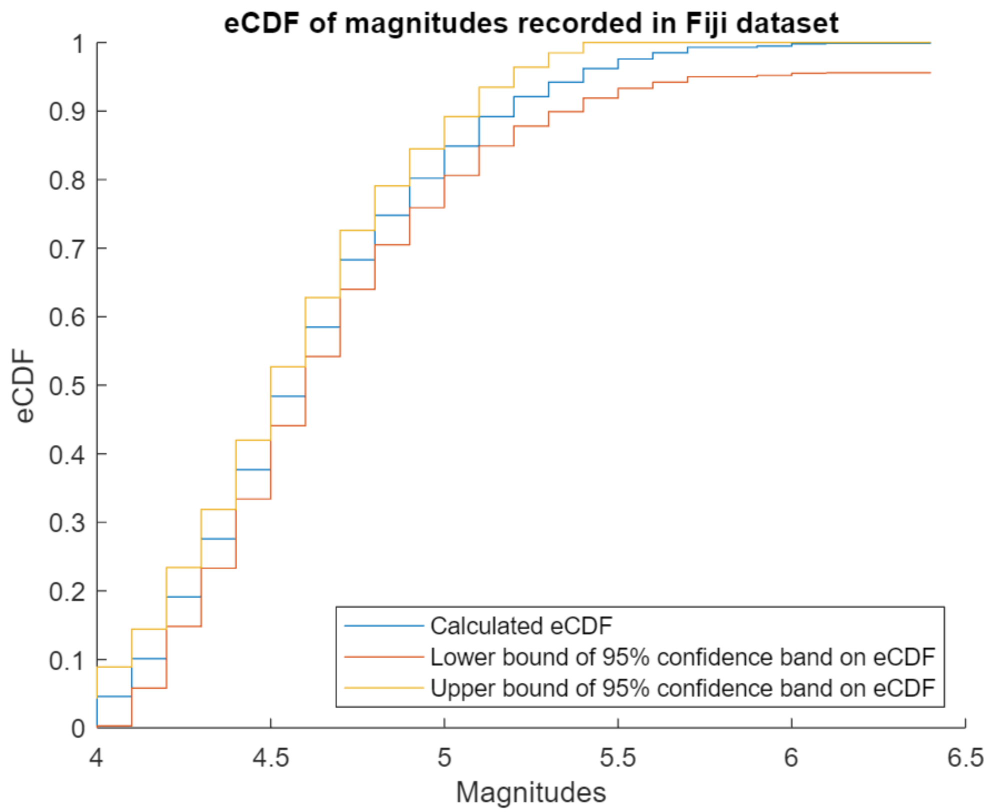
Loading in data

```
data = textread("C:\Users\sayuj\OneDrive - California Institute of
Technology\ACM_157\fiji.txt");
magnitudes = data(:,5);
```

Part A - eCDF with confidence bands

The step-wise plot using stair functions of eCDF estimates of the Fiji dataset magnitudes data is provided below:

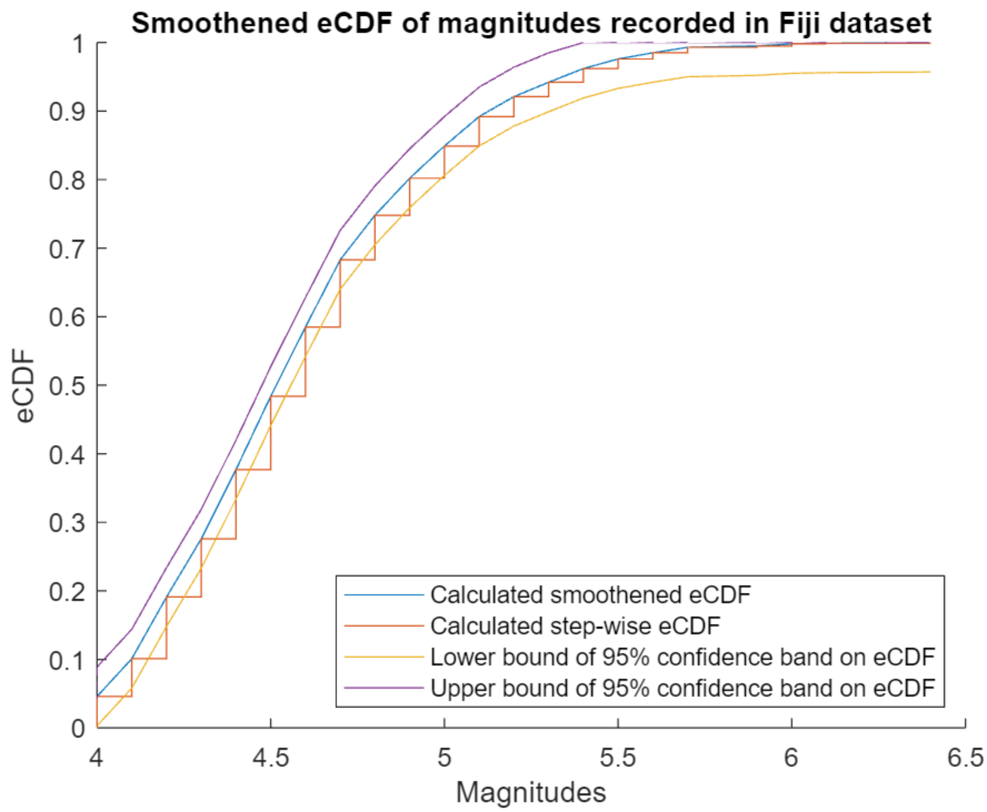
```
[f_estimates, sorted_magnitudes] = ecdf(magnitudes);

alpha_level = .05;
epsilon = sqrt((1 / (2*length(magnitudes))) * log(2 / alpha_level));

lower_bound = max(f_estimates - epsilon, 0);
upper_bound = min(f_estimates + epsilon, 1);

figure;
hold on;
stairs(sorted_magnitudes, f_estimates)
stairs(sorted_magnitudes, lower_bound)
stairs(sorted_magnitudes, upper_bound)
xlabel('Magnitudes')
ylabel('eCDF')
title('eCDF of magnitudes recorded in Fiji dataset')
legend('Calculated eCDF', 'Lower bound of 95% confidence band on eCDF', 'Upper
bound of 95% confidence band on eCDF', 'Location', 'southeast')
hold off;
```



Part B - smoother results using piecewise linear functions

The smoothened plot using piecewise linear functions between eCDF estimates from part (A) is provided below:

```
figure;
hold on;
plot(sorted_magnitudes, f_estimates)
stairs(sorted_magnitudes, f_estimates)
plot(sorted_magnitudes, lower_bound)
plot(sorted_magnitudes, upper_bound)
xlabel('Magnitudes')
ylabel('eCDF')
title('Smoothened eCDF of magnitudes recorded in Fiji dataset')
legend('Calculated smoothened eCDF', 'Calculated step-wise eCDF', 'Lower bound of
95% confidence band on eCDF', 'Upper bound of 95% confidence band on eCDF',
'Location', 'southeast')
hold off;
```



Problem 5

Part A

The results below show the exact bias calculated using both the Taylor expansion simplified to the jackknife assumption requirement with the O(1/n^3) term removed, as well as the actual exact estimate using the expectation of a log-normal distribution.

```
n = 100;
mu = 5;
exact_bias = exp(mu + (.5 * (1/n))) - exp(mu)
```

```
exact_bias = 0.7439
```

```
exact_bias_taylor = exp(mu) * (.5 / n + .125 / (n^2))
```

```
exact_bias_taylor = 0.7439
```

Part B

Running a snigle jackknife estimate for comparison against exact value as well as average of multiple estimates. It can be noticed that on average the jackknife bias estimate is pretty close to the true exact bias.

```
estimates = zeros(1, 1000);

for k = 1:1000
```

```
    estimated_bias = jackknife(100, randn(1, 100) + 5);
    estimates(k) = estimated_bias;
end

single_jackknife_estimate_example = jackknife(100, randn(1, 100) + 5)
```

```
single_jackknife_estimate_example = 0.7030
```

```
average_jackknife_estimate = mean(estimates)
```

```
average_jackknife_estimate = 0.7503
```

Part C

Below is the experiment for the calculated bias on 10^4 iterations of the simple plug-in bias versus the jackknife bias-corrected estimate bias, we can see that the plug-in bias (b1) averages to around the exact bias it was expected to have while the jackknife bias-corrected estimate b2 has a significantly reduced bias compared to b1.

```
b1 = 0;
b2 = 0;


for r = 1:10000
    sample = randn(1, 100) + 5;
    b1 = b1 + (exp(mean(sample)) - exp(5));
    b2 = b2 + (jackknife_correct(100, sample) - exp(5));
end

b1 = b1 / 10000
```

```
b1 = 0.7778
```

```
b2 = b2 / 10000
```

```
b2 = 0.0331
```

Function below for running a jackknife bias estimate based on

```
function jackknife_bias = jackknife(n, sample)

    jackknife_estimate = 0;

    for i = 1: n
        jackknife_sample = sample;
        jackknife_sample(i) = [];
        jackknife_estimate = jackknife_estimate + exp(mean(jackknife_sample));
    end

    jackknife_bias = (n - 1) * ((jackknife_estimate / n) - exp(mean(sample)));
end
```

Function below for running a jackknife bias-corrected estimate

```
function jackknife_estimate = jackknife_correct(n, sample)
    bias = jackknife(n, sample);
    estimate = exp(mean(sample));
    jackknife_estimate = estimate - bias;
end
```