1. a)

$$
\begin{aligned}
\mathrm{E}[\hat{p}_{n,s}(x_0)] &= \mathrm{E}\left[\frac{1}{nh^{s+1}}\sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right)\right] \\
&= \frac{1}{h^{s+1}}\int K\left(\frac{u-x_0}{h}\right)p(u)du \\
&= \frac{1}{h^s}\int K(u)\,p(x_0+hu)du \\
&= \frac{1}{h^s}\int K(u)\left(\sum_{j=0}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0)\right)du \\
&= \sum_{j=0}^{l-1}\frac{h^{j-s}}{j!}p^{(j)}(x_0)\int K(u)u^j du + \frac{1}{h^s}\int K(u)\left(\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0)\right)du \\
&= p^{(s)}(x_0) + \frac{1}{h^s}\int K(u)\left(\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0)\right)du
\end{aligned}
$$

Using the Lagrange form of the remainder of a Taylor series, for some $\xi \in (x_0, x_0+hu)$,

$$
\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0) = \frac{(hu)^l}{l!}p^{(l)}(\xi)
$$

By the conditions of the Hölder class,

$$
\begin{aligned}
\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0) &\le \frac{(hu)^l}{l!}(L|\xi - x_0|^{\beta-l} + |p(x_0)|) \\
\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0) &\le \frac{(hu)^l}{l!}(L|hu|^{\beta-l} + |p(x_0)|) \\
\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0) &\le h^\beta\frac{u^l|u|^{\beta-l}L}{l!} + u^l\frac{h^l}{l!}|p(x_0)| \\
\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0) &\le h^\beta f_1(u) + u^l f_2(x_0)
\end{aligned}
$$

For $f_1(u) := \frac{u^l|u|^{\beta-l}L}{l!}$ and $f_2(x_0) := \frac{h^l}{l!}|p(x_0)|$. By similar logic, we also get

$$
\begin{aligned}
\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0) &\ge -\frac{(hu)^l}{l!}(L|\xi - x_0|^{\beta-l} + |p(x_0)|) \\
\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0) &\ge -h^\beta f_1(u) - u^l f_2(x_0)
\end{aligned}
$$

Plugging these into our expression for $\mathrm{E}[\hat{p}_{n,s}(x_0)]$, we get that

$$\begin{aligned}
\mathrm{E}[\hat{p}_{n,s}(x_0)] &= p^{(s)}(x_0) + \frac{1}{h^s}\int K(u)\left(\sum_{j=l}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0)\right)du \\
|\mathrm{E}[\hat{p}_{n,s}(x_0)] - p^{(s)}(x_0)| &\leq \frac{1}{h^s}\int K(u)(h^\beta f_1(u) + u^l f_2(x_0))du \\
&\leq h^{\beta-s}\int K(u)f_1(u)du + \frac{f_2(x_0)}{h^s}\int K(u)u^l du \\
&\leq h^{\beta-s}c
\end{aligned}$$

For $c := \int K(u)f_1(u)du$, showing the bias is bounded as desired. To bound the variance term, we have that

$$\begin{aligned}
\mathrm{Var}\left(\hat{p}_{n,s}(x_0)\right) &= \frac{1}{nh^{2s+2}}\mathrm{Var}\left(K\left(\frac{X_i - x_0}{h}\right)\right) \\
&= \frac{1}{nh^{2s+2}}\left(\mathrm{E}\left[K\left(\frac{X_i - x_0}{h}\right)^2\right] - \mathrm{E}\left[K\left(\frac{X_i - x_0}{h}\right)\right]^2\right) \\
&\leq \frac{1}{nh^{2s+2}}\left(h\int K^2(u)\left(\sum_{j=0}^{\infty}\frac{(hu)^j}{j!}p^{(j)}(x_0)\right)du - \left(h^{s+1}p^{(s)}(x_0) + h^{\beta+1}c\right)^2\right) \\
&\leq \frac{p(x_0)}{nh^{2s+1}}\int K^2(u)du + O\left(\frac{1}{n}\right)
\end{aligned}$$

So $\mathrm{Var}\left(\hat{p}_{n,s}(x_0)\right)$ is bounded by $\frac{c'}{nh^{2s+1}}$ for $c' > 0$.

b) Combining the bounds on bias squared and variance from part a) to get the risk, we get

$$\begin{aligned}
\max(\text{Ł}^2(\hat{P}_{n,s}(x))) &= c^2h^{2(\beta-s)} + \frac{c'}{nh^{2s+1}} \\
\frac{d}{dh}\max((\text{Ł}^2(\hat{P}_{n,s}(x))) &= 2(\beta-s)c^2h^{2(\beta-s)-1} - \frac{(2s+1)c'}{nh^{2s+2}}
\end{aligned}$$

Then, this risk bound is minimized when its derivative is at 0, so for the optimal h,

$$\begin{aligned}
0 &= 2(\beta-s)c^2h^{2(\beta-s)-1} - \frac{(2s+1)c'}{nh^{2s+2}} \\
\frac{(2s+1)c'}{2(\beta-s)c^2}\frac{1}{n} &= h^{2\beta+1} \\
\left(\frac{(2s+1)c'}{2(\beta-s)c^2}\right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}} &= h
\end{aligned}$$

Plugging this into the original risk function, and defining a new constant $c_1 = \left(\frac{(2s+1)c'}{2(\beta-s)c^2}\right)^{\frac{1}{2\beta+1}}$, we get

$$\begin{aligned}
\max(\text{Ł}^2(\hat{P}_{n,s}(x))) &= c^2c_1^{2(\beta-s)}n^{\frac{2(\beta-s)}{2\beta+1}} - \frac{c'}{c^{2s+1}}n^{\frac{2s+1-2\beta+1}{2\beta+1}} \\
\max(\text{Ł}^2(\hat{P}_{n,s}(x))) &= \left(c^2c_1^{2(\beta-s)} + \frac{c'}{c^{2s+1}}\right)n^{\frac{2(\beta-s)}{2\beta+1}}
\end{aligned}$$

So the maximum risk is of order $O\left(n^{\frac{2(\beta-s)}{2\beta+1}}\right)$.

2. a)

$$
\begin{aligned}
L_F(x) &= \lim_{\epsilon\to 0} \frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon} \\
L_F(x) &= \lim_{\epsilon\to 0} \frac{((1-\epsilon)F(b) + \epsilon\delta_x(b)) - ((1-\epsilon)F(a) + \epsilon\delta_x(a)) - (F(b) - F(a))}{\epsilon} \\
L_F(x) &= \lim_{\epsilon\to 0} \frac{\epsilon(I_{[x,\infty)}(b) - I_{[x,\infty)}(a)) - \epsilon(F(b) - F(a))}{\epsilon} \\
L_F(x) &= I_{[a,b)}(x) - T(F)
\end{aligned}
$$

b)

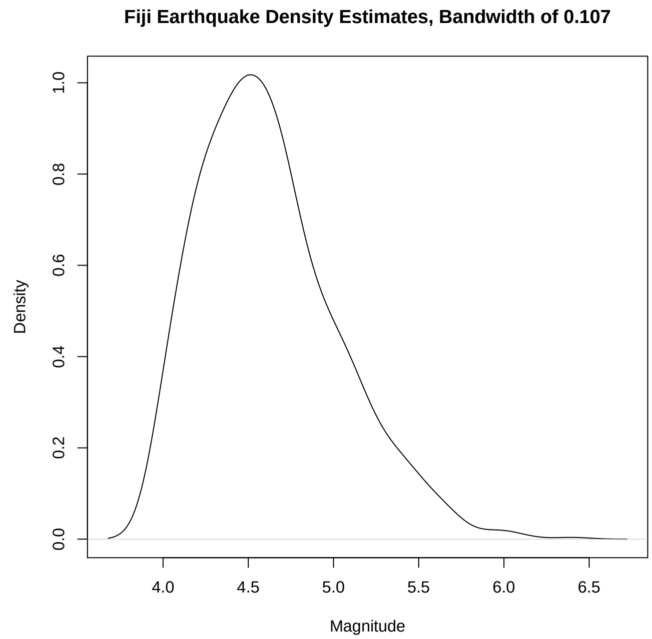
$$
\begin{aligned}
\hat{se}(\hat{\theta}) &= \frac{\hat{\tau}}{\sqrt{n}} \\
\hat{se}(\hat{\theta}) &= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} (I_{[a,b)}(X_i) - T(\hat{F}_n))^2 \\
\hat{se}(\hat{\theta}) &= \frac{1}{n^{\frac{3}{2}}} \sum_{i=1}^{n} (I_{[a,b)}(X_i) - \hat{F}_n(b) + \hat{F}_n(a))^2 \\
\hat{se}(\hat{\theta}) &= \frac{1}{\sqrt{n}} \Big((\hat{F}_n(b) - \hat{F}_n(a))(1 - \hat{F}_n(b) + \hat{F}_n(a))^2 + (1 - \hat{F}_n(b) + \hat{F}_n(a))(\hat{F}_n(b) - \hat{F}_n(a))^2 \Big) \\
\hat{se}(\hat{\theta}) &= \frac{1}{\sqrt{n}} \Big((\hat{F}_n(b) - \hat{F}_n(a) + 1 - \hat{F}_n(b) + \hat{F}_n(a))(\hat{F}_n(b) - \hat{F}_n(a))(1 - \hat{F}_n(b) + \hat{F}_n(a)) \Big) \\
\hat{se}(\hat{\theta}) &= \frac{1}{\sqrt{n}} \Big((\hat{F}_n(b) - \hat{F}_n(a))(1 - \hat{F}_n(b) + \hat{F}_n(a)) \Big)
\end{aligned}
$$

c) An approximate $1-\alpha$ confidence interval is given by

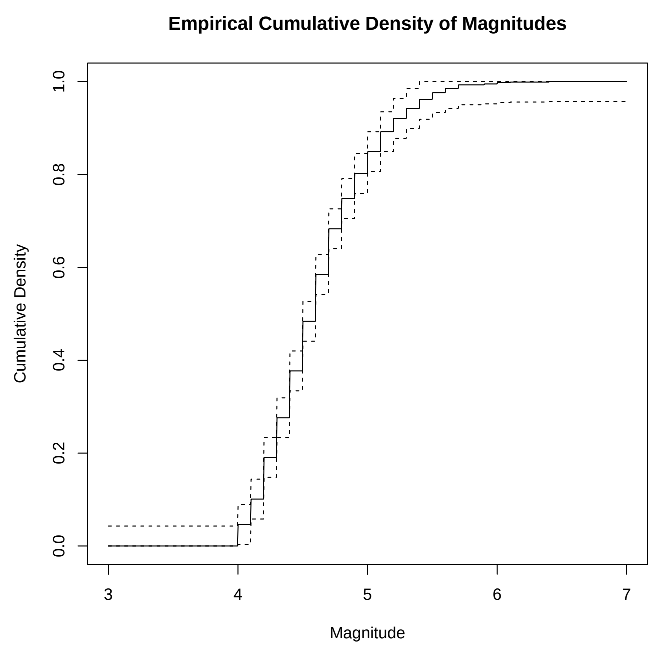
$$
\hat{\theta} \pm z_{\frac{\alpha}{2}} \hat{se}(\hat{\theta}) = F_n(b) - \hat{F}_n(a) \pm z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n}} \Big((\hat{F}_n(b) - \hat{F}_n(a))(1 - \hat{F}_n(b) + \hat{F}_n(a)) \Big)
$$

d) To estimate $se(\hat{\theta})$, one would draw a sample, with replacement, of n from the X_is. For each of these samples, $\hat{\theta}$ would be calculated (this is the portion of the sample in $[a, b)$. Then, $se(\hat{\theta})$ is the standard deviation of $\hat{\theta}$ among all the samples.

3. a) I fit a kernel density estimate with a Gaussian kernel and bandwidth chosen by the Normal Reference Rule. I didn't use cross validation, since the rounding in the magnitudes made it so that cross validation would default to an arbitrarily small bandwidth (as we discovered in our last homework).



b) Below I have plotted the estimated CDF along with a 95% confidence interval. This band is generated from $\hat{F}(x) \pm \sqrt{\frac{1}{2n} \log\left(\frac{2}{.05}\right)}$, winsorized above at 1 and below at 0.



b) Using the density from problem a, we get estimate for $F(4.6) - F(4.3)$ of

$$\int_{4.3}^{4.6} f(x)dx = .533$$

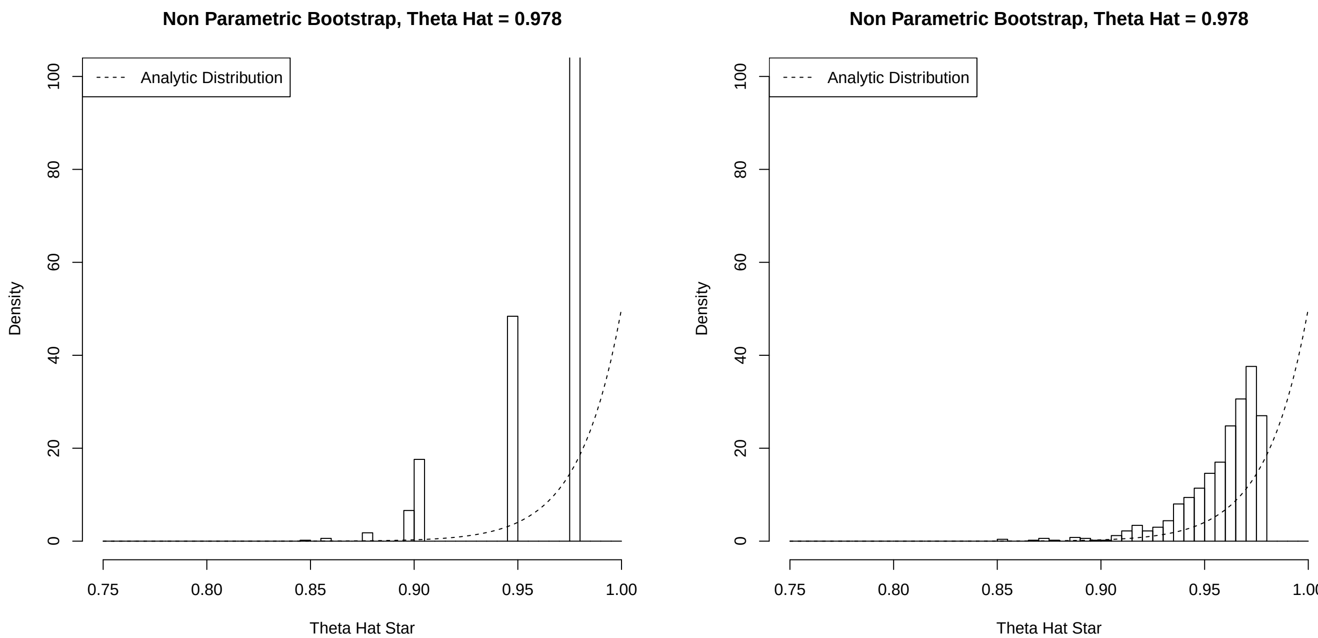
Using the plug-in estimator, we get $\hat{F}(4.6) - \hat{F}(4.3) = .526$. Then, using the result from problem 2, we get an approximate confidence interval of $.526 \pm z_{.025} \frac{.526(1-.526)}{\sqrt{1000}} = (.522, .530)$.

4. a) To get the analytical distribution of $\hat{\theta}$, its easiest to start by calculating the distribution function. For

any $x \in [0, \theta]$

$$P(\hat{\theta} \leq x) = P(X_1 \leq x \cap ... \cap X_n \leq x)$$
$$P(\hat{\theta} \leq x) = \prod_{i=1}^{n} P(X_i \leq x)$$
$$P(\hat{\theta} \leq x) = \prod_{i=1}^{n} \left(\frac{x}{\theta}\right)^n$$

Differentiating both sides, we get the distribution function $f(x) = \frac{nx^{n-1}}{\theta^n}$. Below, I drew a sample of 50 from a uniform with $\theta = 1$, and then preformed a parametric and nonparametric bootstrap of $\hat{\theta}$:



As you can see, the parametric bootstrap gives a distribution fairly close to the analytic one (just shifted a little to the left), while the nonparametric bootstrap fails to do so, having the majority of the point mass at the max from the original sample.

b Since in the parametric case $X_i^* \sim \text{Uniform}(0, \hat{\theta})$,

$$\begin{aligned} P(\hat{\theta}^* = \hat{\theta}) &= P(\exists i, X_i^* = \hat{\theta} \cap \forall j, X_j^* \leq \hat{\theta}) \\ &\leq P(\exists i, X_i^* = \hat{\theta}) \\ &\leq \prod_{i=1}^{n} (1 - P(X_i^* = \hat{\theta})) \\ &\leq 0 \end{aligned}$$

On the other hand, in the nonparametric case,

$$\begin{aligned} P(\hat{\theta}^* = \hat{\theta}) &= P(\exists i, X_i^* = \hat{\theta} \cap \forall j, X_j^* \leq \hat{\theta}) \\ &= P(\exists i, X_i^* = \hat{\theta}) \\ &= 1 - P(\forall i, X_i^* \neq \hat{\theta}) \\ &= 1 - \left(1 - \frac{1}{n}\right)^n \end{aligned}$$

For $n = 50$, this gives us .636, but in the limiting case, we get

$$\begin{aligned}
\lim_{n\to\infty} P(\hat{\theta}^* = \hat{\theta}) &= 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n \\
&= 1 - \lim_{n\to\infty} \left(\frac{n}{n-1}\right)^{-n} \\
&= 1 - \lim_{n\to\infty} \left(1 + \frac{1}{n-1}\right)^{-n} \\
&= 1 - \lim_{n\to\infty} \left(1 + \frac{1}{n-1}\right) \left(1 + \frac{1}{n-1}\right)^{-n+1} \\
&= 1 - \frac{1}{e} \\
&\approx .632
\end{aligned}$$

5. a) For this model, the bootstrapped 95% confidence interval for θ I got was $(.240, .282)$, and the "true" θ, which was the result of running the experiment for $n = 500$, was .264. Thus, the confidence interval included the "true" parameter, and also the real $\theta = .25$.

   b) For this model, my confidence interval was $-.280, .120)$ and the "true" θ was .023. Thus, the confidence interval covered the "true" parameter and also the real parameter $\theta = 0$.

   c) For this model, my confidence interval was $(.282, .482)$ with the "true" $\theta = .865$. Thus, my confidence interval was far below the "true" θ, and even farther below the actual θ of 0.