Stochastic Processes: Homework 1

Martín Prado

November 2023
Universidad de los Andes – Bogotá Colombia

Exercise 3

For the following theorem, give two examples where they satisfy item 2 and two examples where they don't.

Theorem 1. Let F be a probability distribution and $\overline{F}(x) = 1 - F(x)$. Then, define

$$x_F = F^{-1}(x) = \inf\{t > 0 \mid F(t) = 1\}.$$

If these two conditions are satisfied:

- $x_F \leq \infty$,
- $\tau \in (0, \infty)$,

Then, the following items are equivalent

1. There exists a sequence $(u_n)_{n\in\mathbb{N}}$ such that

$$\lim_n n \cdot \overline{F}(u_n) = \tau.$$

2.

$$\lim_{x \to x_F-} \frac{\overline{F}(x)}{\overline{F}(x-)}.$$

Note that if F is discrete and $x_F = \infty$, then item 2. is equivalent to

$$\lim_{n\to\infty} \frac{f(n)}{\overline{F}(n)} = 0.$$

Solution Good Examples

1. **Zeta Distribution:** For this distribution, let $s \in (1, \infty)$ and $k \in \mathbb{Z}^+$. Then, for $X \sim \text{Zeta}(s)$,
$$f(k) = \frac{1}{k^s \zeta(s)},$$
$$F(k) = \mathbf{P}\{X \le k\} = \frac{\sum_{i=1}^{k} i^{-s}}{\zeta(s)} = \frac{\sum_{i=1}^{k} i^{-s}}{\sum_{i=1}^{\infty} i^{-s}}.$$
Thus,
$$\overline{F}(k) = 1 - \frac{\sum_{i=1}^{k} i^{-s}}{\sum_{i=1}^{\infty} i^{-s}} = \frac{\sum_{i=k+1}^{\infty} i^{-s}}{\sum_{i=1}^{\infty} i^{-s}} = \frac{\sum_{i=k+1}^{\infty} i^{-s}}{\zeta(s)}$$
Asymptotically,
$$\overline{F}(k) \sim \int_k^\infty x^{-s} dx = \frac{1}{s-1} k^{-s+1}.$$
Therefore,
$$\frac{f(k)}{\overline{F}(k)} \sim \frac{1}{s-1} \frac{k^{-s}}{k^{-s+1}} \sim k^{-1} \to 0$$

2. **Gauss-Kuzmin Distribution:** For this distribution,
$$f(k) = -\log_2 \left[1 - \frac{1}{(k+1)^2}\right].$$
$$\overline{F}(k) = \log_2 \left(\frac{k+2}{k+1}\right)$$
Thus,
$$\begin{aligned}
\lim_{k\to\infty} \frac{\overline{F}(k)}{\overline{F}(k-1)} &= \lim_{k\to\infty} \frac{\log(k+2) - \log(k+1)}{\log(k+1) - \log(k)} \\
&= \lim_{k\to\infty} \frac{\frac{1}{k+2} - \frac{1}{k+1}}{\frac{1}{k+1} - \frac{1}{k}} \\
&= \lim_{k\to\infty} \frac{(k+1)-(k+2)}{(k+2)\cdot(k+1)} \cdot \frac{(k+1)\cdot(k)}{(k)-(k+1)} \\
&= \lim_{k\to\infty} \frac{-k^2 - k}{-k^2 - 3k - 2} = 1
\end{aligned}$$

Solution Bad Examples

1. **Log Distribution:** For $0 < p < 1$, we define:

$$F(k) = 1 + \frac{\beta_k}{\ln(1-p)}, \quad \overline{F}(k) = \frac{\beta_k}{-\ln(1-p)},$$

$$f(k) = \frac{-p}{\ln(1-p)(1-p)},$$

where

$$\beta_k = \beta(p; k+1, 0) = \int_0^p t^k (1-t)^{-1} dt,$$

$$\frac{f(k)}{\overline{F}(k)} = \frac{p}{(1-p)} \cdot \frac{1}{\beta_k}.$$

The series expansion of β_k is the following

$$\beta_k = \frac{p^{k+1}}{k+1} + \frac{p^{k+2}}{k+2} + \frac{p^{k+3}}{k+3} + \cdots = \sum_{i=k+1}^{\infty} \frac{p^i}{i}.$$

This series converges for every $k \in \mathbb{N}$ since $p < 1$ (ratio test). Also, the sequence β_k is decreasing because less positive terms are being summed when k is increased, and thus, $\lim_k \beta_k = 0$. Therefore,

$$\lim_k \frac{p}{(1-p)} \cdot \frac{1}{\beta_k} = \infty.$$

Failed Attempts

1. **Zipf-Mandelbrot Distribution:** Note that what played in favor for the last example was the heaviness of the tails. In this next case, something similar will happen. For $N \in \mathbb{Z}^+, q \geq 0, s > 0$, a random variable $X_{N,q,s}$ with the Zipf-Mandelbrot distribution has the following properties,

$$f(k) = f_{N,q,s}(k) = \frac{1}{(k+q)^s} \cdot \frac{1}{H_{N,q,s}},$$

where

$$H_{N,q,s} = \sum_{i=1}^{N} \frac{1}{(i+q)^s}.$$

As a matter of fact, when $q = 0$ and $N \to \infty$, $X_{N,0,s}$ converges in distribution to a zeta distribution. Now,

$$F(k) = \frac{H_{k,q,s}}{H_{N,q,s}}$$

and thus,title

$$\overline{F}(k) = \frac{\sum_{i=k+1}^{N} (i+q)^{-s}}{H_{N,q,s}}.$$

Asymptotically,

$$\overline{F}(k) \sim \int_{k}^{N} (x+q)^{-s} dx = \frac{1}{s-1} \left(k^{1-s} - N^{1-s}\right).$$

Hmmmm, it seems that it has finite support.

2. **Beta Negative Binomial:** For $\alpha, \beta > 0 \in \mathbb{R}$ and $r \in \mathbb{N}$, define

$$f(k) = \binom{r+k-1}{k} \frac{B(\alpha+r, \beta+k)}{B(\alpha, \beta)},$$

where

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Then, by Stirling's approximation we have that

$$f(k) \sim \frac{\Gamma(\alpha+r)}{\Gamma(r)B(\alpha,\beta)} \frac{k^{r-1}}{(\beta+k)^{r+\alpha}}$$

Thus, we can approximate the tail of the distribution with the following integral:

$$C = \frac{\Gamma(\alpha+r)}{\Gamma(r)B(\alpha,\beta)},$$

$$\begin{aligned}\overline{F}(n) &\sim C \int_{n}^{\infty} \frac{x^{r-1}}{(x-\beta)^{r+\alpha}} dx \\ &= C \frac{x^{r} \cdot {}_2F_1(1, 1-a; r+1; \frac{x}{b})}{br(x-\beta)^{r+a-1}}\end{aligned}$$

Honestly, I tried many things to make a direct proof of the polynomial behavior of the tails. However, I couldn't find any literature on this topic. The argument I'm trying to make is that while $f(n) \sim Kn^{1-\alpha}$, the tail should behave like a higher degree polynomial, and thus,

$$\lim_{n} \frac{f(n)}{\overline{F}(n)} = 0.$$

This argument could make sense in the case where $r = \beta = \alpha = 1$, because

$$f(k) \sim \binom{k}{k} \frac{B(2,k+1)}{B(1,1)} = \frac{1}{(k+1)(k+2)} = \mathrm{Cau}(k+1)$$

$$\overline{F(k)} \sim$$

Exercise 4

1. Simulate 50 times, graph and compare the empirical distribution of $X_{i:6}$ for $i \in \{1, \dots, 6\}$ with $X_k \sim U(0,1)$. Calculate the uniform distance $\|\cdot\|_\infty$ between the empirical and the real distributions.

2. Simulate 50 times, graph and compare the empirical distribution of $X_{i:6}$ for $i \in \{1, \dots, 6\}$ with $X_k \in F$, where

$$F = \begin{cases} 0 & x \leq 1 \\ 1 - x^{-1} & x > 1. \end{cases}$$

   Calculate the uniform distance $\|\cdot\|_\infty$ between the empirical and the real distributions. Also calculate $f_{i:6}$, $F_{i:6}$ and determine which moments are finite.

Solution Part 1

We use the letter F for the uniform distribution, $F_{i:n}$ for the order distribution of the $i-th$ element of the sample, and $F^*_{i:n}$ for the empirical distribution obtained from sampling m times the $i-th$ order statistic. In fact, since we are working with the uniform distribution,

$$F_{i:n}(x) = \mathrm{Beta}_{(i,n-i+1)}(x).$$

The formula for the empirical distribution is the following:

$$F^*_{i:n} = \frac{1}{m}\sum_{k=1}^{m} \mathbb{1}\{X^{(k)}_{i:n} < x\}.$$

Finally, since $F_{i:n}(x)$ is monotonically increasing and $F^*_{i:n}(x)$ is piecewise constant, the formula for the uniform distance can be approximated by calculating M times the following formula

$$\|F^*_{i:n} - F_{i:n}\|_\infty \approx \max_{k\in\{1,\dots,M\}} |F_{i:n}(k/m) - F^*_{i:n}(k/m)|.$$

The exercise required $n = 6$ and $m = 50$, and I personally used $M = 5000$ for graphing and calculating the uniform distance.

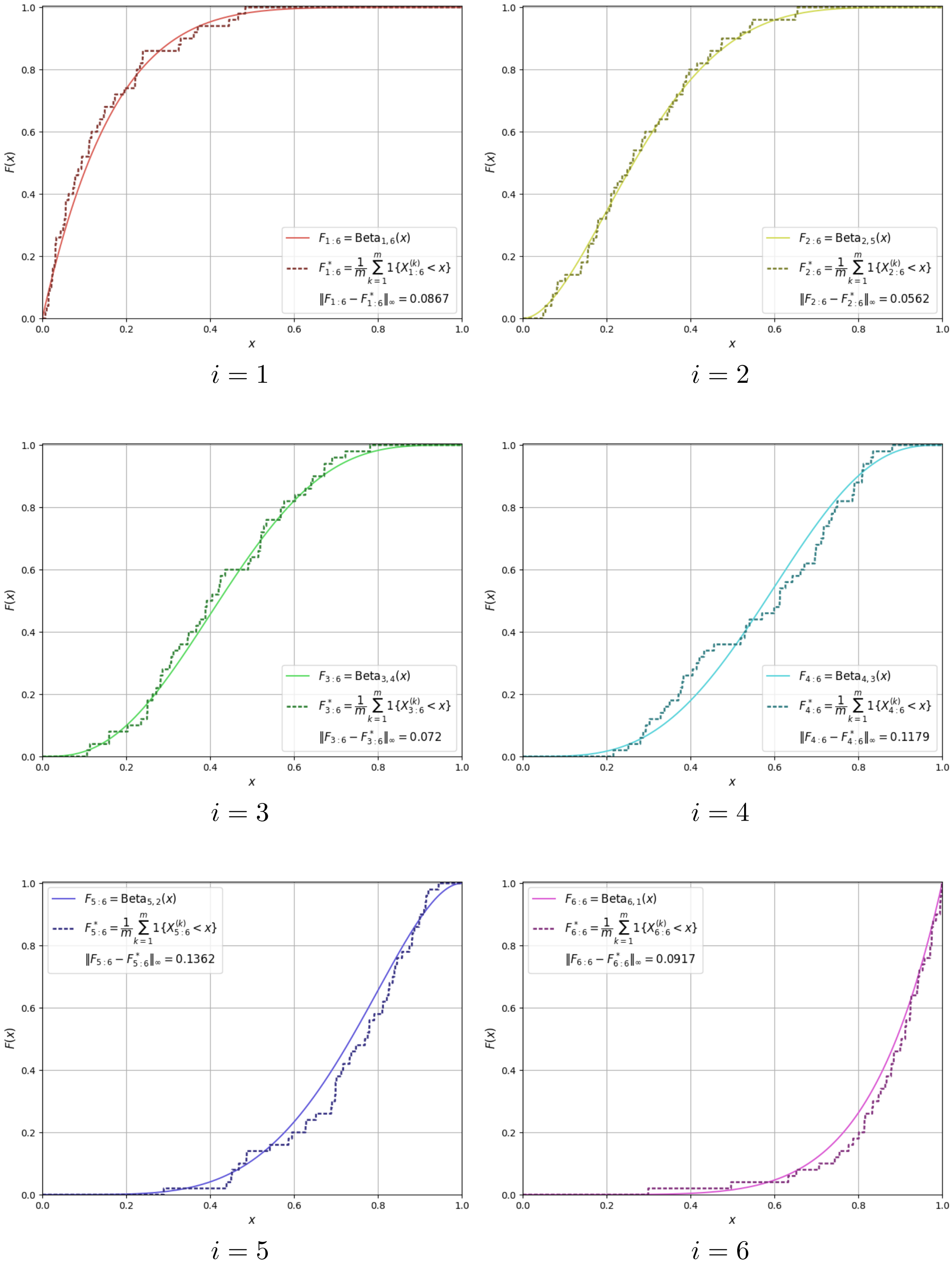

Figure 1: Simulation of the six order statistics for 6 samples of the Uniform Distribution

Solution Part 2

The distribution F from part 2 is a type I Pareto distribution

$$\text{Pareto}_{(\alpha,\sigma)}(x) = \begin{cases} 0 & x \leq \sigma \\ 1 - \left(\frac{\sigma}{x}\right)^{\alpha} & x > \sigma. \end{cases}$$

with parameters $\alpha = \sigma = 1$. Therefore, the condition required for it to have the k-moment is that $k < \alpha$. Since $\alpha = 1$, it doesn't have any finite moments. Using the formula for the i-th statistic, we obtain

$$F_{i:n}(x) = \sum_{j=i}^{n} \binom{n}{j} F^{j}(x)\overline{F}^{n-j}(x)$$

$$= \sum_{j=i}^{n} \binom{n}{j} (1 - x^{-1})^{j} x^{-n+j}.$$

Also,

$$f_{i:n}(x) = \frac{1}{i} \cdot \binom{n}{i} f(x)F^{i-1}(x)\overline{F}^{n-i}(x)$$
$$= \frac{1}{i} \cdot \binom{n}{i} x^{-2}(1 - x^{-1})^{i-1} x^{i-n}$$

Thus, the formula for the k-moment is the following

$$E[X^k] = \frac{1}{i} \cdot \binom{n}{i} \int_1^{\infty} \underbrace{x^{i+k}x^{-2-n}}_{\text{dominant}} (1 - x^{-1})^{i-1} dx$$

This last integral converges if $n + 2 - i - k > 1$, thus, the only moments available are when $k < n - 1 - i$.

For the uniform norm I did the same as the previous part.

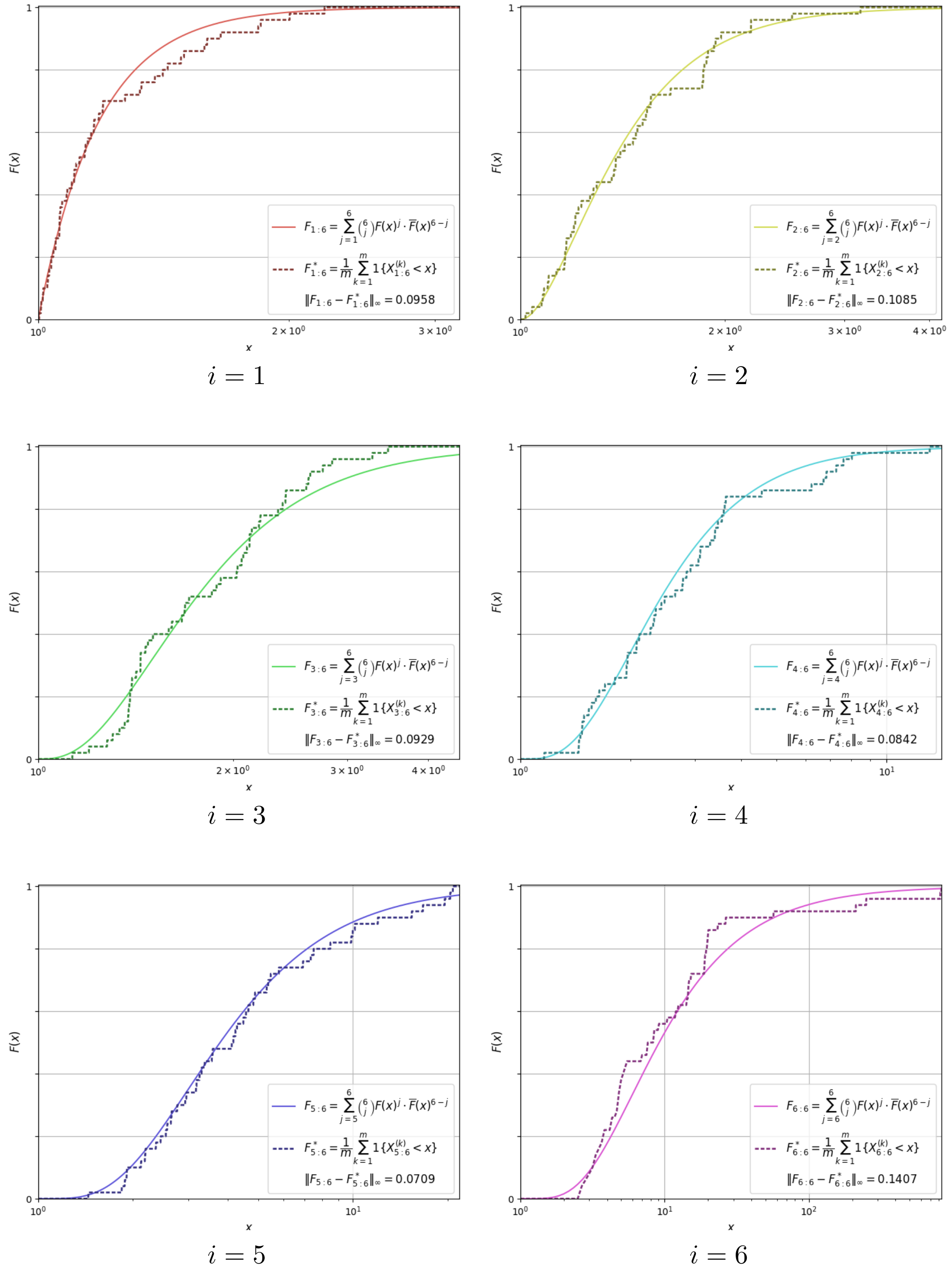

Figure 2: Simulation of the six order statistics for 6 samples of the Pareto Distribution

Exercise 1

Show that, for a sequence $(X_n)_{n\in\mathbb{N}}$ of Bernoulli random variables, with $X_n \sim \mathrm{Be}(n, p_n)$ and $p_n \to 0$. If

$$\mathbf{P}\left\{X_n = k\right\} \overset{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!}e^{-\lambda}$$

then,

$$\lim_{n\to\infty} n \cdot p_n = \lambda$$

Solution

Using what we obtained from the other direction of the proof,

$$\mathbf{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \binom{n}{k}\frac{1}{n^k}(n\cdot p_n)^k(1-\frac{n\cdot p_n}{n})^{n-k}.$$

For $k = 0$, we have

$$\lim_n \mathbf{P}(X_n = 0) = \binom{n}{0}\frac{1}{n^0}(n\cdot p_n)^0\left(1-\frac{n\cdot p_n}{n}\right)^n = (1-p_n)^n = e^{-\lambda}.$$

Taking logarithms at both side will leave us with

$$\lim_n n\ln(1-p_n) = -\lambda.$$

Using Taylor's expansion, we can assert that

$$-\ln(1-p_n) = p_n + o(p_n).$$

Finally,

$$-\lim_n n\ln(1-p_n) = \lim_n n(p_n + o(p_n)) = \lambda.$$

Since p_n dominates over $o(p_n)$, we can conclude that

$$\lim_n n\cdot p_n = \lambda.$$

Exercise 2

Prove the following theorem for $\tau = 0$ and $\tau = \infty$.

Theorem 2. For any given X with F_X and $\overline{F}_X(x) = 1 - F_X(x)$. Also let $M_n = X_{n:n}$. For $x \in \mathbb{R}$, if

- $\tau \in [0, \infty]$
- $(u_n)_{n\in\mathbb{N}}$ a non-decreasing sequence,

then the following items are equivalent,

1. $\lim_{n\to\infty} \mathbf{P}(M_n \le u_n) = e^{-\tau}$
2. $\lim_{n\to\infty} n\overline{F}_X(u_n) = \tau$.

Solution $\tau = 0$

- (2) $\implies$ (1) If $n \cdot \overline{F}(u_n) \to 0$, then $\overline{F}(u_n) = o(1/n)$. Therefore,

$$\begin{aligned}\lim_n \mathbf{P}(M_n \le u_n) &= \lim_n F(u_n)^n = \lim_n \left(1 - \overline{F}(u_n)\right)^n \\ &= \lim_n \left(1 - o\left(\tfrac{1}{n}\right)\right)^n\end{aligned}$$

  What this means is that for every $\varepsilon > 0$, we will eventually have that

$$\tfrac{-\varepsilon}{n} \le -\overline{F}(u_n) \le \tfrac{\varepsilon}{n}.$$

  In particular, for every $\varepsilon > 0$,

$$e^{-\varepsilon} \le \lim_n (1 - \tfrac{\varepsilon}{n})^n \le \lim_n \left(1 - \overline{F}(u_n)\right)^n \le \lim_n (1 + \tfrac{\varepsilon}{n})^n = e^{\varepsilon}.$$

  Therefore, by making ε go to 0 we would obtain,

$$1 \le \lim_n \left(1 - \overline{F}(u_n)\right)^n \le 1.$$

- (1) $\implies$ (2) Now, the hypothesis says that

$$\mathbf{P}(M_n \le u_n) = \lim_n \left(1 - \overline{F}(u_n)\right)^n = 1.$$

  To prove that $\overline{F}(u_n) \to 0$, we use the same argument from the original proof. If $\liminf_n \overline{F}(u_n) = \alpha > 0$, then there exists a subsequence $u_{n_k} \subset u_n$ such that

$$1 = \lim_k \left(1 - \overline{F}(u_{n_k})\right)^{n_k} \le \lim_k (1 - \alpha)^{n_k} = 0.$$

  With that in mind, we take logarithm at both sides to obtain

$$\lim_n n \cdot \ln((1 - \overline{F}(u_n))) = 0.$$

  From Taylor's formula, $-\ln(1 - x) = x + o(x)_{x\to 0+}$. Thus,

$$-0 = -\lim_n n \cdot \ln((1 - \overline{F}(u_n))) = \lim_n n\overline{F}(u_n)$$

Solution $\tau = \infty$

- (2) $\implies$ (1) $n \cdot \overline{F}(u_n) \to \infty$ is equivalent to

$$\lim_n \frac{1/\overline{F}(u_n)}{n} = 0.$$

Which by definition means that $\overline{F}(u_n)^{-1} = o(\frac{1}{n})$. This implies that for every $\varepsilon > 0$,

$$\overline{F}(u_n)^{-1} \leq \varepsilon n, \text{ (eventually)}$$
$$\implies -\overline{F}(u_n) \leq -\frac{\varepsilon'}{n}$$

$$\lim_n \mathbf{P}(M_n \leq u_n) = \lim_n F(u_n)^n = \lim_n (1 - \overline{F}(u_n))^n$$
$$\leq \lim_n (1 - \tfrac{\varepsilon}{n})^n = e^{-\varepsilon}.$$

By making $\varepsilon \to \infty$, we can conclude that

$$\lim_n \mathbf{P}(M_n \leq u_n) = 0.$$

- (1) $\implies$ (2) Now, note that

$$\lim_n n \cdot \ln((1 - \overline{F}(u_n))) = -\infty.$$

Using Taylor's polynomial like we did in the case $\tau = 0$, will give us that

$$\lim_n n \cdot (\overline{F}(u_n) + o(\overline{F}(u_n))) = \infty.$$

Since, by definition, $\overline{F}(u_n)$ dominates over $o(\overline{F}(u_n))$, we can conclude that

$$\lim_n n \cdot \overline{F}(u_n) = \infty.$$