Maximal Inequalities

Jianqiao Wang

July 7, 2021

Overview

Overview

The main goal of Chapter 8 is to present the empirical process techniques needed to prove the Glivenko-Cantelli and Donsker theorems. The most difficult step in these proof is going from point-wise convergence to uniform convergence. Maximal inequalities are very useful tools for accomplishing this step.

Orlicz Norms and Maxima

Orlicz norms are useful for controlling the size of the maximum of a finite collection of random variables.

Definition

For a nondecreasing, nonzero convex function $\psi : [0, \infty] \mapsto [0, \infty]$, with $\psi(0) = 0$, the Orlicz norm of a real random variable X, also called ψ-norm, is

$$\|X\|_\psi \equiv \inf \left\{ c > 0 : \mathbb{E}\psi\left(\frac{|X|}{c}\right) \leq 1 \right\},$$

where the norm takes the value ∞ if no finite c exists with $\mathbb{E}\psi(|X|/c) \leq 1$.

Orlicz Norms and Maxima

With the convexity of ψ, exercise 8.5.1 verifies that $\|\cdot\|_\psi$ is indeed norm on the space of random variables with $\|X\|_\psi < \infty$.

When ψ is of the form $x \mapsto x^p$, where $p \geq 1$, the corresponding Orlicz norm is L_p-norm

$$\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$$

For maximal inequalities, Orlicz norms defined with $\psi_p(x) \equiv e^{x^p} - 1$, for $p \geq 1$, are of greater interest because of their sensitivity to behavior in the tails. Since $x^p \leq \psi_p(x)$, we have $\|X\|_p \leq \|X\|_{\psi_p}$.

By the series representation of exponentiation, $\|X\|_p \leq (p!)^{1/p} \|X\|_{\psi_1}$ for all $p \geq 1$.

Orlicz Norms and Maxima

Orlicz norms based on ψ_p relate fairly precisely to the tail probabilities:

Lemma (8.1)

For a real random variable X and any $p \in [1, \infty)$, the following are equivalent:

1. $\|X\|_{\psi_p} < \infty$.
2. *There exist constants $0 < C, K < \infty$ such that*

$$P(|X| > x) \leq K e^{-Cx^p} \tag{8.1}$$

Moreover, if either condition holds, then $K = 2$ and $C = \|X\|_{\psi_p}^{-p}$ satisfies (8.1), and , for any $C, K \in (0, \infty)$ satisfying (8.1), $\|X\|_{\psi_p} \leq \left(\frac{1+K}{C}\right)^{1/p}$.

Proof of Lemma 8.1

(1) $\Longrightarrow$ (2)
Assume (1). By Markov inequalities,

$$P(|X| > x) = P\left(\psi_p(|X|/\left\|X\right\|_{\psi_p}) \geq \psi_p(x/\left\|X\right\|_{\psi_p})\right) \leq 1 \wedge (\frac{1}{\psi_p(x/\left\|X\right\|_{\psi_p})})$$

By exercise 8.5.2, $1 \wedge (e^u - 1)^{-1} \leq 2e^u$ for all $u > 0$. Thus,

$$P(|X| > x) \leq 1 \wedge (\frac{1}{\psi_p(x/\left\|X\right\|_{\psi_p})}) \leq 2e^{-\frac{x^p}{\|X\|_{\psi_p}^p}}$$

Proof of Lemma 8.1

(2) $\Longrightarrow$ (1) Assume (2). For any $c \in (0, C)$, by Fubini's theorem,

$$\begin{aligned}
\mathbb{E}(e^{c|X|^p} - 1) =& \mathbb{E} \int_0^{|X|^p} ce^{cs} ds \\
=& \int_0^{\infty} P(|X| > s^{1/p}) ce^{cs} ds \\
\leq& \int_0^{\infty} Ke^{-Cs} ce^{cs} ds \\
=& \frac{Kc}{C - c}
\end{aligned}$$

$Kc/(C - c) \leq 1$ whenever $c \leq C/(1 + K)$, or equivalently, $c^{-1/p} \geq ((1 + K)/C)^{1/p}$. This implies $\|X\|_{\psi_p} \leq ((1 + K)/C)^{1/p} < \infty$.

Orlicz Norms and Maxima

An important use for Orlicz norms is to control the behavior of maxima.

This control is an extension of the following result for L_p-norms:

For any random variables $X_1, \ldots, X_m$,

$$\begin{aligned} \|\max_{1\le i\le m} X_i\|_p &= \left(\mathbb{E}\max_{1\le i\le m}|X_i|^p\right)^{1/p} \\ &\le \left(\mathbb{E}\sum_{i=1}^m |X_i|^p\right)^{1/p} \\ &\le m^{1/p}\max_{1\le i\le m}\|X_i\|_p \end{aligned}$$

Orlicz Norms and Maxima

Lemma (8.2)

Let $\psi : [0, \infty) \mapsto [0, \infty)$ *be convex, nondecreasing and nonzero, with* $\psi(0) = 0$ *and*

$$\limsup_{x,y\to\infty} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty$$

for some constant $c < \infty$. *Then, for any random variables* $X_1, \ldots, X_m$,

$$\left\| \max_{1\le i\le m} X_i \right\|_\psi \le K\psi^{-1}(m) \max_{1\le i\le m} \|X_i\|_\psi$$

where the constant K *depends only on* ψ.

Proof of Lemma 8.2

First make the stronger assumption that $\psi(1) \leq 1/2$ and that $\psi(x)\psi(y) \leq \psi(cxy)$ for all $x, y \geq 1$.

Under this assumption, $\psi(x/y) \leq \psi(cx)/\psi(y)$ for all $x \geq y \geq 1$. Hence, for any $y \geq 1$ and $k > 0$,

$$\begin{aligned}
&\max_{1\leq i\leq m} \psi\left(\frac{|X_i|}{ky}\right) \\
&\leq \max_i \left[\frac{\psi(c|X_i|/k)}{\psi(y)} 1\left\{\frac{|X_i|}{ky} \geq 1\right\} + \psi\left(\frac{|X_i|}{ky}\right) 1\left\{\frac{|X_i|}{ky} < 1\right\}\right] \\
&\leq \max_i \left[\frac{\psi(c|X_i|/k)}{\psi(y)} 1\left\{\frac{|X_i|}{ky} \geq 1\right\} + \psi(1)\right] \\
&\leq \sum_{i=1}^m \frac{\psi(c|X_i|/k)}{\psi(y)} + \psi(1)
\end{aligned}$$

Proof of Lemma 8.2

Set $k = c \max_i \|X_i\|_\psi$ and take expectations of both sides,

$$\begin{aligned}\mathbb{E}[\psi\left(\frac{\max_i |X_i|}{ky}\right)] &\leq \sum_{i=1}^m \mathbb{E}[\frac{\psi(c|X_i|/k)}{\psi(y)}] + \psi(1) \\ &\leq \frac{m}{\psi(y)} + \frac{1}{2}\end{aligned}$$

With $y = \psi^{-1}(2m)$, the right-hand-side is ≤ 1. Thus $\|\max_i |X_i|\|_\psi \leq ky = c\psi^{-1}(2m) \max_i \|X_i\|_\psi$.

Since ψ is convex and $\psi(0) = 0$, $x \mapsto \psi^{-1}(x)$ is concave and one-to-one for $x > 0$. Thus $\psi^{-1}(2m) \leq 2\psi^{-1}(m)$. The result of Lemma 8.2 follows with $K = 2c$.

Proof of Lemma 8.2

By exercise 8.5.3, for any ψ satisfying the conditions of the lemma, that there exists constants $0 < \sigma \leq 1$ and $\tau > 0$ such that $\phi(x) \equiv \sigma\psi(\tau x)$ satisfies $\phi(1) \leq 1/2$ and $\phi(x)\phi(y) \leq \phi(cxy)$ for all $x, y \geq 1$.

Furthermore, for this ϕ, $\phi^{-1}(u) \leq \psi^{-1}(u)/(\sigma\tau)$, for all $u > 0$, and for any random variable X, $\|X\|_\psi \leq \|X\|_\phi/(\sigma\tau) \leq \|X\|_\psi/\sigma$. Hence

$$\begin{aligned}\sigma\tau \left\| \max_i X_i \right\|_\psi \leq & \|\max_i X_i\|_\phi \\ \leq & 2c\phi^{-1}(m) \max_i \|X_i\|_\phi \\ \leq & \frac{2c}{\sigma}\psi^{-1}(m) \max_i \|X_i\|_\psi\end{aligned}$$

Orlicz Norms and Maxima

An important consequence of Lemma 8.2 is that maximums of random variables with bounded ψ-norm grow at the rate of $\psi^{-1}(m)$.

Based on exercise 8.5.4, $\psi_p = e^{x^p} - 1$ satisfies the conditions of Lemma 8.2 with $c = 1$, for any $p \in [1, \infty)$. The growth of maxima is at most logarithmic, since $\psi_p^{-1}(m) = (\log(1 + m))^{1/p}$.

We now present an inequality for collections $X_1, \ldots, X_m$ of random variables which satisfy

$$P(|X_i| > x) \leq 2e^{-\frac{1}{2}\frac{x^2}{b+ax}} \tag{8.3}$$

for all $x > 0$, $i = 1, \ldots, m$ and some $a, b \geq 0$. This setting will arise later in the development of a Donsker theorem based on bracketing entropy.

Orlicz Norm and Maxima

Lemma (8.3)

Let $X_1, \ldots, X_m$ *be random variables that satisfy the tail bound*

$$P(|X_i| > x) \leq 2e^{-\frac{1}{2}\frac{x^2}{b+ax}} \tag{8.3}$$

for $1 \leq i \leq m$ *and some* $a, b \geq 0$. *Then*

$$\|\max_{1\leq i\leq m} |X_i|\|_{\psi_1} \leq K\left\{a\log(1+m) + \sqrt{b}\sqrt{\log(1+m)}\right\}$$

where the constant K *is universal, in the sense that it does not depend on* a, b *or on the random variables.*

Proof of Lemma 8.3

Assume for now that $a, b > 0$. The condition implies for all $x \leq b/a$ the upper bound $2\exp(-x^2/(4b))$ for $P(|X_i| > x)$, since in this case $b + ax \leq 2b$. For all $x > b/a$ the conditions implies an upper bound of $2\exp(-x/(4a))$, since $b/a + x \leq 2a$ in this case. This implies that

$$P(|X_i|1\{|X_i| \leq b/a\} > x) \leq 2\exp(-x^2/(4b))$$
$$P(|X_i|1\{|X_i| > b/a\} > x) \leq 2\exp(-x/(4a))$$

for all $x > 0$. By Lemma 8.1, the Orlicz norms $\||X_i|1\{|X_i| \leq b/a\}\|_{\psi_2}$ and $\||X_i|1\{|X_i| > b/a\}\|_{\psi_1}$ are bounded by $\sqrt{12b}$ and $12a$ respectively. Since ψ_p-norms increase in p,

$$\|\max_i |X_i|\|_{\psi_1} \leq \|\max_i |X_i|1\{|X_i| > b/a\}\|_{\psi_1} + \|\max_i |X_i|1\{|X_i| \leq b/a\}\|_{\psi_2}$$

Proof of Lemma 8.3

Result of Lemma 8.3 follows by Lemma 8.2 combined with above inequality.

Suppose now that $a > 0$ but $b = 0$. Then the tail bound 8.3 holds for all $b > 0$, and the result of Lemma 8.3 follows by letting $b \downarrow 0$.

A similar argument will verify that the results holds when $a = 0$ and $b > 0$.

When $a = b = 0$, $X_i = 0$ almost surely for $i = 1, \ldots, m$.

Maximal Inequalities for Processes

The goals of this section are to first establish a general maximal inequality for *separable* stochastic processes and then specialize to *sub-Gaussian* processes.

A stochastic process $\{X(t), t \in T\}$ is *separable* when there exists a countable subset $T_* \subset T$ such that

$$\sup_{t \in T} \inf_{s \in T_*} |X(t) - X(s)| = 0$$

almost surely.

In statistical applications, the separability of certain processes is hidden in other conditions and its verification is seldom required.

Maximal Inequalities for Processes

A stochastic process is sub-Gaussian when for all $s, t \in T, x > 0$

$$P(|X(t) - X(s)| > x) \leq 2e^{-\frac{1}{2}x^2/d^2(s,t)}$$

for a semimetric d on T. In this case, we say that X is sub-Gaussian with respect to d.

Examples of sub-Gaussian processes include Rademacher process and Brownian motion on [0, 1].

The conclusion of Lemma 8.2 is not immediately useful for maximizing $X(t)$ over $t \in T$ since a potentially infinite number of random variables is involved.

Maximal Inequalities for Processes

For an arbitrary semimetric space (T, d), the covering number $N(\epsilon, T, d)$ or $N(\epsilon, d)$ is the minimal number of closed $d-$balls of radius ϵ required to cover T.

The packing number $D(\epsilon, T, d)$ or $D(\epsilon, d)$ is the maximal number of points that can fit in T while maintaining a distance greater than ϵ between all points.

The associated entropy numbers are the respective logarithms of the covering and packing numbers. These concepts define metric entropy.

Metric Entropy

For a semimetric space (T, d) and each $\epsilon > 0$,

$$N(\epsilon, d) \leq D(\epsilon, d) \leq N(\epsilon/2, d)$$

There exists a minimal subset $T_\epsilon \subset T$ such that the cardinality of $T_\epsilon = D(\epsilon, d)$ and the minimum distance between distinct points in T_ϵ is $> \epsilon$.

If we now place closed ϵ-balls around each point in T_ϵ, we have a covering of T. Thus, $N(\epsilon, d) \leq D(\epsilon, d)$.

No ball of radius $\leq \epsilon/2$ can cover more than one point in T_ϵ, and thus at least $D(\epsilon, d)$ closed $\epsilon/2$-balls are needed to cover T_ϵ. Hence $D(\epsilon, d) \leq N(\epsilon/2, d)$.

Above discussion reveals that covering and packing numbers are essentially equivalent in behavior as $\epsilon \downarrow 0$.

Maximal Inequalities for Processes

Theorem (General maximal inequality)

Let ψ satisfy the conditions of Lemma 8.2, and let $\{X(t), t \in T\}$ be a separable stochastic process with $\|X(s) - X(t)\|_\psi \leq rd(s,t)$, for all $s, t \in T$, some semimetric d on T, and a constant $r < \infty$.
Then for any $\eta, \delta > 0$,

$$\left\| \sup_{s,t \in T: d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi \leq K \left[\int_0^\eta \psi^{-1}(D(\epsilon, d)) d\epsilon + \delta \psi^{-1}(D^2(\eta, d)) \right]$$

for a constant $K < \infty$ which depends only on ψ and r. Moreover,

$$\left\| \sup_{s,t \in T} |X(s) - X(t)| \right\|_\psi \leq 2K \int_0^{diam T} \psi^{-1}(D(\epsilon, d)) d\epsilon$$

where $diam T \equiv \sup_{s,t \in T} d(s,t)$ is the diameter of T.

Maximal Inequalities for Processes

An important application of above theorem is to sub-Gaussian processes. Suppose $\{X(t), t \in T\}$ is a sub-Gaussian process wit respect to d. By Lemma 8.1,

$$\|X(t) - X(s)\|_{\psi_2} \leq \sqrt{6} d(s,t)$$

Corollary (8.5)

Let $\{X(t), t \in T\}$ be a separable sub-Gaussian process with respect to d. Then for all $\delta > 0$,

$$\mathbb{E}\left(\sup_{s,t \in T : d(s,t) \leq \delta} |X(s) - X(t)| \right) \leq K \int_0^{\delta} \sqrt{\log D(\epsilon, d)} d\epsilon$$

where K is a universal constant. Also for any $t_0 \in T$,

$$\mathbb{E}\left(\sup_{t \in T} |X(t)| \right) \leq \mathbb{E}|X(t_0)| + K \int_0^{diam T} \sqrt{\log D(\epsilon, d)} d\epsilon$$

Proof of Corollary 8.5

Apply general maximal inequality with $\psi = \psi_2$ and $\eta = \delta$. Because $\psi_2^{-1}(m) = \sqrt{\log(1+m)} \leq \sqrt{2}\sqrt{\log(1+\sqrt{m})}, \psi_2^{-1}(D^2(\delta, d)) \leq \sqrt{2}\psi_2^{-1}(D(\delta, d))$. Hence the second term of general maximal inequality can be replaced by

$$\sqrt{2}\delta\psi_2^{-1}(D(\delta, d)) \leq \sqrt{2}\int_0^\delta \psi_2^{-1}(D(\epsilon, d))d\epsilon$$

and we obtain

$$\| \sup_{d(s,t)\leq\delta} |X(s) - X(t)| \|_{\psi_2} \leq K \int_0^\delta \sqrt{\log(1 + D(\epsilon, d))}d\epsilon$$

Proof of Corollary 8.5

Note $D(\epsilon, d) \geq 2$ for all ϵ strictly less than $\operatorname{diam} T$. Since $(1 + m) \leq m^2$ for all $m \geq 2$, the 1 inside of the logarithm can be removed at the cost of increasing K, whenever $\delta < \operatorname{diam} T$.

The second conclusion is an consequence of the first conclusion with triangle inequality.

Maximal Inequalities for Processes

Modulus of continuity of a stochastic process $\{X(t) : t \in T\}$, where (T, d) is a semimetric space, is defined as

$$m_X(\delta) = \sup_{s,t \in T : d(s,t) \leq \delta} |X(s) - X(t)|$$

Corollary (8.6)

Assume the conditions of corollary 8.5. Also assume there exists a differentiable function $\delta \mapsto h(\delta)$, with derivative $h'(\delta)$, satisfying $h(\delta) \geq \sqrt{\log D(\delta, d)}$ for all $\delta > 0$ small enough and $\lim_{\delta \downarrow 0}[\delta h'(\delta)/h(\delta)] = 0$. Then

$$\lim_{M \to \infty} \limsup_{\delta \downarrow 0} P(\frac{m_X(\delta)}{\delta h(\delta)} > M) = 0$$

Maximal Inequalities for Processes

Theorem (General maximal inequality)

Let ψ satisfy the conditions of Lemma 8.2, and let $\{X(t), t \in T\}$ be a separable stochastic process with $\|X(s) - X(t)\|_\psi \le r d(s,t)$, for all $s, t \in T$, some semimetric d on T, and a constant $r < \infty$.
Then for any $\eta, \delta > 0$,

$$\left\| \sup_{s,t \in T : d(s,t) \le \delta} |X(s) - X(t)| \right\|_\psi \le K \left[\int_0^\eta \psi^{-1}(D(\epsilon, d)) d\epsilon + \delta \psi^{-1}(D^2(\eta, d)) \right]$$

for a constant $K < \infty$ which depends only on ψ and r. Moreover,

$$\left\| \sup_{s,t \in T} |X(s) - X(t)| \right\|_\psi \le 2K \int_0^{diam T} \psi^{-1}(D(\epsilon, d)) d\epsilon$$

where $diam T \equiv \sup_{s,t \in T} d(s,t)$ is the diameter of T.

Proof of General Maximal Inequality

Note that if the first integral were infinite, the inequalities would be trivially true. Hence, without loss of generality, assume that the packing numbers and associated integral are bounded.

Construct a sequence of finite nested set $T_0 \subset T_1 \subset \cdots \subset T$ such that for each T_j,

- $d(s,t) > \eta 2^{-j}$ for every distinct $s, t \in T_j$, and
- each T_j is "maximal" in the sense that no additional points can be added to T_j without violating the inequality.

Thus, the number of points in T_j is bounded by $D(\eta 2^{-j}, d)$.

Proof of General Maximal Inequality

Do the chaining part of the proof.

Begin by "linking" every point $t_{j+1} \in T_{j+1}$ to one and only one $t_j \in T_j$ such that $d(t_j, t_{j+1}) \leq \eta 2^{-j}$, for all points in T_{j+1}.

Continue this process to link all points in T_j with points in T_{j-1}, and so on, for every $t_{j+1} \in T_{j+1}$, to obtain a chain $t_{j+1}, t_j, t_{j-1}, \ldots, t_0$ that connects to a point in T_0.

Proof of General Maximal Inequality

For any integer $k \geq 0$ and arbitrary points $s_{k+1}, t_{k+1} \in T_{k+1}$, the difference in increments along their respective chains connecting to s_0, t_0 can be bounded as follows:

$$
\begin{aligned}
&|(X(s_{k+1}) - X(t_{k+1})) - (X(s_0) - X(t_0))| \\
=&|\sum_{j=0}^{k}\{X(s_{j+1}) - X(s_j)\} - \sum_{j=0}^{k}\{X(t_{j+1}) - X(t_j)\}| \\
\leq&2\sum_{j=0}^{k} \max |X(u) - X(v)|
\end{aligned}
$$

where for fixed j the maximum is taken over all links (u, v) from T_{j+1} to T_j. Hence, the jth maximum is taken over at most the cardinality of T_{j+1} links, with each link having $\|X(u) - X(v)\|_\psi$ is bounded by $rd(u, v) \leq r\eta 2^{-j}$.

Proof of General Maximal Inequality

By Lemma 8.2, for a constant $K_0 < \infty$ depending only on ψ and r,

$$\begin{aligned}
&\left\| \max_{s,t \in T_{k+1}} \left| \{X(s) - X(s_0)\} - \{X(t) - X(t_0)\} \right| \right\|_\psi \\
\leq& 2 \sum_{j=0}^{k} \|\max |X(u) - X(v)|\|_\psi \\
\leq& K_0 \sum_{j=0}^{k} \psi^{-1}(D(\eta 2^{-j-1}, d)) \eta 2^{-j} \\
=& 4K_0 \sum_{j=0}^{k} \psi^{-1}(D(\eta 2^{-k+j-1}, d)) \eta 2^{-k+j-2} \\
\leq& 4\eta K_0 \int_0^1 \psi^{-1}(D(\eta u, d)) du = 4K_0 \int_0^\eta \psi^{-1}(D(\epsilon, d)) d\epsilon
\end{aligned}$$

Proof of General Maximal Inequality

In this bound, s_0 and t_0 depend on s and t in that they are the endpoints of the chains starting at s and t respectively.

The maximum of the increments $|X(s_{k+1}) - X(t_{k+1})|$ over all s_{k+1} and t_{k+1} in T_{k+1} with $d(s_{k+1}, t_{k+1}) < \delta$ is bounded by

- the left-hand-side of above inequality
- plus the maximum of the discrepancies at the ends of the chains $|X(s_0) - X(t_0)|$ for those points in T_{k+1} which are less than δ apart.

Proof of General Maximal Inequality

For every such pair of endpoints s_0, t_0 of chains starting at two points in T_{k+1} within distance δ of each other, choose one and only one pair s_{k+1}, t_{k+1} in T_{k+1}, with $d(s_{k+1}, t_{k+1}) < \delta$, whose chains end at s_0, t_0. By the definition of T_0, this results in at most $D^2(\eta, d)$ pairs.

$$
\begin{aligned}
|X(s_0) - X(t_0)| \leq & |\{X(s_0) - X(s_{k+1})\} - \{X(t_0) - X(t_{k+1})\}| \\
& + |X(s_{k+1}) - X(t_{k+1})|
\end{aligned}
$$

Take the maximum over all pairs of endpoints s_0, t_0. The maximum of the second term of the right-hand-side is the maximum of $D^2(\eta, d)$ terms with ψ-norm bounded by $r\delta$. By Lemma 8.2, this maximum is bounded by some constant C times $\delta\psi^{-1}(D^2(\eta, d))$.

Proof of General Maximal Inequality

Combining above inequalities,

$$\begin{aligned}
&\|\max_{s,t\in T_{k+1}:d(s,t)<\delta}|X(s)-X(t)|\|_\psi \\
&\qquad \leq 8K_0\int_0^\eta \psi^{-1}(D(\epsilon,d))d\epsilon + C\delta\psi^{-1}(D^2(\eta,d))
\end{aligned}$$

By the fact that the right-hand-side does non depend on k, we can replace T_{k+1} with $T_\infty = \cup_{j=1}^{\infty} T_j$ by the monotone convergence theorem.

Proof of General Maximal Inequality

If we can verify that taking the supremum over T_∞ is equivalent to taking the supremum over T, then the first conclusion of the theorem follows with $K = (8K_0) \vee C$.

Since X is separable, there exists a countable subset $T_* \subset T$ such that $\sup_{t \in T} \inf_{s \in T_*} |X(t) - X(s)| = 0$ almost surely.

Let Ω_* denote the subset of the sample space of X for which this supremeum is zero. $P(\Omega_*) = 1$.

Proof of General Maximal Inequality

Now for any point t and sequence t_n in T, exercise 8.5.5 shows that $d(t, t_n) \to 0$ implies $|X(t) - X(t_n)| \to 0$ almost surely.

For each $t \in T_*$, let Ω_t be the subset of the sample space of X for which $\inf_{s \in T_\infty} |X(s) - X(t)| = 0$.

Since T_∞ is a dense subset of the semimetric space (T, d), $P(\Omega_t) = 1$.

Let $\tilde{\Omega} = \Omega_* \cap (\cap_{t \in T_*} \Omega_t)$, $P(\tilde{\Omega}) = 1$.

Combined with the fact that

$$\sup_{t \in T} \inf_{s \in T_\infty} |X(t) - X(s)| \leq \sup_{t \in T} \inf_{s \in T_*} |X(t) - X(s)| + \sup_{t \in T_*} \inf_{s \in T_\infty} |X(s) - X(t)|$$

implies that $\sup_{t \in T} \inf_{s \in T_\infty} |X(t) - X(s)| = 0$ almost surely.

Thus taking the supremum over T is equivalent to taking the supremum over T_∞.

Proof of General Maximal Inequality

The second conclusion of general maximal inequality follows from previous result by setting $\delta = \eta = \operatorname{diam} T$ and noting $D(\eta, d) = 1$.