# Empirical Processes: Maximal Inequalities and Chaining

Moulinath Banerjee

*October 7, 2010*

## 1 Maximal Inequalities

The notes here are based on Section 8.1 of Kosorok's book and Chapter 2.2 of van der Vaart and Wellner and should be read in accompaniment with these texts.

**Orlicz norms:** Let $\psi$ be a nondecreasing convex function with $\psi(0) = 0$ and $X$ a random variable. Define the Orlicz norm corresponding to $\Psi$ as:

$$\|X\|_\Psi = \inf\left\{C > 0 : \, E\,\psi\left(\frac{|X|}{C}\right) \leq 1\right\}.$$

It can be checked that this is a valid norm (on the set of random variables for which the left side of the above display is finite). Of special interest to us will be the Orlicz norms corresponding to the functions $\{\psi_p : p \geq 1\}$ where $\psi_p(x) = \exp(x^p) - 1$. Lemma 8.1 of Kosorok (2008) provides a necessary and sufficient condition for the $\psi_p$ Orlicz norm to be finite in terms of the tail-behavior of $X$. As a consequence of this lemma, if $P(|X| > x) \leq K\,\exp(-C\,x^p)$ for some constants $C, K > 0$, then

$$\|X\|_{\psi_p} \leq \left(\frac{1+K}{C}\right)^{1/p}. \tag{1.1}$$

The Orlicz norm of the maximum of finitely many random variables can be bounded by the maximum of the individual Orlicz norms. This is the content of the next lemma which will be eventually seen to play an important role in controlling the oscillations of separable stochastic processes.

**Lemma 1.1** *Let $\psi$ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and such that, for every $M > 0$, $\limsup_{x,y\to\infty} \psi(x)\,\psi(y)/\psi(cxy) < \infty$ for some constant $0 < c < \infty$. Then, for any* *random variables* $X_1, X_2, \ldots, X_m$,

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

*with the constant* $K$ *depending only on* $\psi$.

The Orlicz norms of interest to us all satisfy the conditions of the above lemma.

**Sketch of proof:** The proof starts by consider 'special' $\psi$'s for which $\psi(1) \le 1/2$ and $\psi(x)\,\psi(y) \le \psi(cxy)$ for all $x, y \ge 1$. For such $\psi$'s an argument involving some algebra (see Page 130 of Kosorok) and the definition of the Orlicz norm and the concavity of the strictly increasing function $\psi^{-1}$ shows the desired inequality with $K = 2c$. The generalization rests on the fact that for any $\psi$ in the statement of the lemma, one can find constants $0 < \sigma < 1$ and $\tau > 0$ such that $\phi(x) = \sigma\,\psi(\tau\, x)$ is a 'special' $\psi$. However the constant $c$ for the $\phi$ just constructed above could be different from the constant $c$ appearing in the limsup condition for $\psi$; however the new $c$ still continues to depend only on $\psi$.

The proof is completed by invoking the established result for the $\phi$ Orlicz norm and then translating this to what it means for the $\psi$ Orlicz norm using the inequalities in Problem 8.5.3 (a), (b) of Kosorok.

**Exercise:** Exercise 8.5.3 of Kosorok.

Our next goal is to develop a maximal inequality for the Orlicz norm of the supremum of (possibly) uncountably many increments of a separable stochastic process $X(t)$ using the above maximal inequality as a core tool. Bounds on such suprema will be developed in terms of the packing/ covering numbers of the indexing set $T$ of the stochastic process with respect to semimetrics that are at least as strong as the one induced by the $\psi$ norm. It will be seen that such bounds translate to bounds on the modulus of continuity for sub-Gaussian processes in terms of integrals of appropriate covering/packing numbers. Because the behavior of the modulus of continuity plays a central role in the convergence of processes in $l^\infty(T)$ and empirical processes are a special case, it turns out that useful sufficient conditions for the distributional convergence of the empirical process can be obtained in terms of the so-callede 'entropy integrals', which are integrals involving the logarithms of the covering numbers. To develop a bound for the Orlicz norm of the supremum in the first sentence of this paragraph, the trick is to construct a countably dense subset of $T$ such that the supremum restricted to the countably dense subset equals the actual supremum and then to approximate the Orlicz norm of this restricted supremum by those of suprema over finite increasing subsets that approximate the countable set. The suprema over the finite increasing subsets can be handled by Lemma 1.1.

Before proceeding further, two definitons.

**Sub-Gaussian processes:** A stochastic process $\{X(t) : t \in T\}$ will be called *sub-Gaussian* with respect to some semi-metric $d$ on $T$ if:

$$P(|X(t) - X(s)| > x) \leq 2\, \exp(-\frac{1}{2}\, x^2/d^2(s,t))\ , \quad \text{forall } s, t \in T\,.$$

There is nothing sacred about the constants 2 and 1/2 on the right side of the inequality as the following exercise demonstrates.

**Exercise:** Suppose that $P(|X(t) - X(s)| > x) \leq K\, \exp(-C\, x^2/d^2(s,t))$ for all $s, t$, for positive constants $K$ and $C$. Then the process is sub-Gaussian for a multiple of the distance $d$. (Exercise 14 on Page 106 of VDVW)

**Separable processes:** A stochastic process $X(t)$ is called separable if there exists a countable subset $T_\star \subset T$ such that $\sup_{t \in T} \inf_{s \in T_\star} |X(s) - X(t)| = 0$ almost surely.

The next theorem gives the general maximal inequality.

**Theorem 1.1** *Let $\psi$ satisfy the conditions of Lemma 1.1 and let $\{X(t) : t \in T\}$ be a separable stochastic process such that $\|X(s) - X(t)\|_\psi \leq r\, d(s,t)$ for some constant $r$ and some semimetric $d$. Suppose also that $D(\epsilon, d) < \infty$ for every $\epsilon > 0$. Then, for any $\eta, \delta > 0$,*

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

*for a finite constant $K$ depending only on $\psi$ and $r$. Furthermore:*

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

**Proof:** Since $D(\epsilon, d)$ is finite for every positive $\epsilon$, $T$ is totally bounded with respect to $d$. We will construct a countably dense subset $T_\infty$ of $T$ such that $\sup_{s,t \in T_\infty; d(s,t) \leq \delta} |X(s) - X(t)|$ is almost surely equal to $\sup_{s,t \in T; d(s,t) \leq \delta} |X(s) - X(t)|$. Then it suffices to find a bound on $\left\| \sup_{s,t \in T_\infty\, d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi$. The set $T_\infty = \cup_0^\infty T_k$ where $T_0 \subset T_1 \subset T_2 \subset \ldots$ and $T_j$ is a finite set of points satisfying the property that $d(s,t) > \eta\, 2^{-j}$ for every $s \neq t \in T_j$. Furthermore $T_j$ is maximal in the sense that no points can be added to this set without violating the separation condition in the previous sentence. Note that the cardinality of $T_j$ cannot be larger than $D(\eta\, 2^{-j}, d)$, so that each $T_j$ is finite. We then construct the chains in the following manner: For every point $t_{j+1} \in T_{j+1}$, find a point $t_j \in T_j$ such that $d(t_j, t_{j+1}) \leq \eta\, 2^{-j}$ and 'attach' these points by a link. Then start with $t_j$ to get a $t_{j-1}$ in similar fashion and proceed all the way back up to a $t_0 \in T_0$. Note that the linkages are possible: if $t_{j+1} \in T_j$ then $t_j$ can be taken to be $t_{j+1}$ itself; if not, there has to be at least one point in $T_j$ whose $d$-distance from $t_{j+1}$ is no larger than $\eta\, 2^{-j}$, for if not, $T_j$ would not have been maximal in the first place.
Our first goal is to control, for every $k$,

$$E_k \equiv \left\| \sup_{s,t \in T_{k+1}; d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi .$$

Note that the sup over a finite set is really a max. If $s_0(s)$ and $t_0(t)$ denote the end-points of the chains starting at $s$ and $t$, it is not difficult to see that:

$$E_k \leq \left\| \sup_{s,t \in T_{k+1}; d(s,t) \leq \delta} |(X(s) - X(s_0(s)) - (X(t) - X(t_0(t))| \right\|_\psi + \left\| \sup_{s,t \in T_{k+1}; d(s,t) \leq \delta} |X(s_0(s)) - X(t_0(t))| \right\|_\psi .$$

We now focus on the second term, say $II$ with the first term being denoted $I$. Note that the number of pairs $(s_0(s), t_0(t))$ is no larger than $D^2(\eta, d)$. For each such pair, select a unique ancestor for $s_0(s)$, say $s^\star$, and a unique ancestor for $t_0(t)$, say $t^\star$ in $T_{k+1}$ (note that there can be multiple ancestors for a particular $s_0(s)$ or a $t_0(t)$) such that $d(s_\star, t_\star) < \delta$. We have:

$$|X(s_0(s)) - X(t_0(t))| \leq |(X(s_0(s)) - X(s_\star)) - (X(t_0(t)) - X(t_\star)| + |X(s_\star) - X(t_\star)| .$$

Hence,

$$II \leq \left\| \sup_{s,t \in T_{k+1}; d(s,t) \leq \delta} |(X(s_0(s)) - X(s_\star)) - (X(t_0(t)) - X(t_\star))| \right\|_\psi + \left\| \sup_{s,t \in T_{k+1}; d(s,t) \leq \delta} |X(s_\star) - X(t_\star)| \right\|_\psi .$$

Using Lemma 1.1, the second term in the above display can be bounded by a $\psi$-dependent constant $C$ times $\psi^{-1}(D^2(\eta, d))\, \delta$, using the fact that the number of variables involved in the supremum in the second term is at most $D^2(\eta, d)$ and the fact that $\|X(s_\star) - X(t_\star)\|_\psi$ is atmost $r\,\delta$. Both the first term in the above display and $I$ are bounded by

$$\left\| \sup_{s,t \in T_{k+1}} |(X(s) - X(s_0(s)) - (X(t) - X(t_0(t))| \right\|_\psi$$

and this term can be shown to be dominated by $4\,K_0 \int_0^\eta \psi^{-1}(D(\epsilon, d))\,d\epsilon$. It follows that:

$$E_k \equiv \left\| \sup_{s,t \in T_{k+1}; d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi \leq 8\,K_0 \int_0^\eta \psi^{-1}(D(\epsilon, d))\,d\epsilon + C\,\delta\,\psi^{-1}(D^2(\eta, d))\,.$$

Note that the right side does not depend on $k$. Since $\sup_{s,t \in T_{k+1}; d(s,t) \leq \delta} |X(s) - X(t)|$ increases to $\sup_{s,t \in T_{\inf}; d(s,t) \leq \delta} |X(s) - X(t)|$, by monotone convergence for Orlicz norms (if random variables $Y_n$'s increase to $Y$, the Orlicz norm of $Y_n$ increases to that of $Y$; see problem 6 on Page 105 of VDVW) we conclude that:

$$E_\infty \equiv \left\| \sup_{s,t \in T_\infty; d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi \leq 8\,K_0 \int_0^\eta \psi^{-1}(D(\epsilon, d))\,d\epsilon + C\,\delta\,\psi^{-1}(D^2(\eta, d))\,,$$

which readily implies the first conclusion of the theorem. The second conclusion of the theorem follows by choosing $\delta = \eta = \text{diam}\ T$ and noting that then $D(\eta, d) = D^2(\eta, d) = 1$ and using the monotonicity of $\psi^{-1}$. The argument justifying the equality of the supremum over $T$ and $T_\infty$ is the content of the following lemma. However, for some stochastic processes the equality follows from a more direct argument. For example, if you *know* that the paths of $X$ are almost surely continuous with respect to $d$, the equality is immediate. □

**Lemma 1.2**

$$\left\| \sup_{s,t \in T_\infty; d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi = \left\| \sup_{s,t \in T; d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi \quad a.s.$$

This lemma, as well as Lemma 1.3 below depend on Exercise 8.5.5. of Kosorok stated below.

**Exercise 8.5.5 (Kosorok):** Let $\psi$ satisfy the conditions of Lemma 1.1. Show that for any sequence of random variables $X_n$, $\|X_n\|_\psi \to 0$ implies that $X_n \to_P 0$. (Hint: $\liminf_{x \to \infty} x^{-1} \psi(x) > 0$.)

**Proof of lemma:** Since $X$ was assumed separable, there exists a countable susbet of $T$, say $T_\star$, such that on a set $\Omega_\star$ with $P(\Omega_\star) = 1$, $\sup_{t \in T} \inf_{s \in T_\star} |X(s) - X(t)| = 0$. Now, fix a point in $t$. If $d(t, t_n)$ goes to 0, the conditions of the theorem guarantee that $\|X(t) - X(t_n)\|_\psi$ goes to 0. By the exercise above $X(t_n)$ converges to $X(t)$ almost surely along a subsequence. Now, consider the (countably dense) subset $T_\infty$ of $T$ that we used in the chaining argument. For each $t \in T_\star$, there is a ($t$–dependent) sequence $t_n$ in $T_\infty$converging to $t$, which implies that for some subsequence $\{t_{n_k}\}$, $X_{t_{n_k}}$ converges to $X(t)$ on a set $\Omega_t$ with $P(\Omega_t) = 1$. Let $\tilde{\Omega} = \Omega_\star \cap (\cap_{t \in T_\star} \Omega_t)$. This set has probability 1 and on this set , $\sup_{t \in T} \inf_{s \in T_\infty} |X(s) - X(t)| = 0$. To show this last inequality, it suffices to show that for any fixed $t \in T$, , $\inf_{s \in T_\infty} |X(s) - X(t)| = 0$ for $\omega \in \tilde{\Omega}$. Consider such an $\omega$. Given any pre-assigned positive $\epsilon$, there exists $t_\star \in T_\star$ such that $|X(t) - X(t_\star| < \epsilon/2$. For this $t_\star$, there exists a $t_\infty \in T_{\inf}$ such that $|X(t_\star) - X(t_\infty)| < \epsilon/2$, by virtue of the fact that $\omega \in \Omega_{t_\star}$. It follows that $|X(t) - X(t_\infty)| < \epsilon$, which implies what we sought to show. $\Box$

**Addendum to proof of Theorem :** We show that:

$$\left\| \sup_{s,t \in T_{k+1}} |(X(s) - X(s_0(s)) - (X(t) - X(t_0(t))| \right\|_\psi \leq 4\, K_0 \int_0^\eta \psi^{-1}(D(\epsilon, d))\, d\epsilon\,.$$

In what follows we abbreviate $s_0(s)$ and $t_0(t)$ to $s_0$ and $t_0$ respectively. Now, for any $s, t \in T_{k+1}$,

$$\begin{aligned} |(X(s) - X(s_0)) - (X(t) - X(t_0))| &= \left| \sum_{j=0}^{k} (X(s_{j+1}) - X(s_j)) - \sum_{j=0}^{k} (X(t_{j+1}) - X(t_j)) \right| \\ &\leq 2 \sum_{j=0}^{k} \max_{(u,v) \in \text{Links}(j+1,j)} |X(u) - X(v)|\,; \end{aligned}$$

note that the right side of the above display continues to remain an upper bound if we insert a supremum over all $(s,t) \in T_{k+1}$ before the left side. Hence,

$$A_k \equiv \left\| \sup_{s,t \in T_{k+1}} |(X(s) - X(s_0(s)) - (X(t) - X(t_0(t))| \right\|_\psi \leq 2 \sum_{j=0}^{k} \left\| \max_{(u,v) \in \text{Links}(j+1,j)} |X(u) - X(v)| \right\|_\Psi .$$

Each term inside the sum on the right side, being the Orlicz norm of the maximum of finitely many random variables can be bounded using Lemma 1.1; for the $j$'th term in the sum there are at most $D(\eta\, 2^{-(j+1)})$ links involved and the Orlicz norm of each $|X(u) - X(v)|$ is at most $r\eta 2^{-j}$. Conclude that for some constant $K_0$,

$$\begin{aligned} A_k &\leq K_0 \sum_{j=0}^{k} \psi^{-1}(D(\eta\, 2^{-(j+1)}, d))\, \eta\, 2^{-j} \\ &= 4K_0 \sum_{j=0}^{k} \psi^{-1}(D(\eta\, 2^{-(j+1)}, d))\, \eta\, 2^{-j+2} \\ &= 4K_0 \; \frac{\eta}{4}\psi^{-1}(D(\eta/2, d)) + \frac{\eta}{8}\psi^{-1}(D(\eta/4, d)) + \frac{\eta}{16}\psi^{-1}(D(\eta/8, d)) + \ldots\Big] \\ &\leq 4K_0\eta \left[\frac{1}{2}\psi^{-1}(D(\eta/2, d)) + \frac{1}{4}\psi^{-1}(D(\eta/4, d)) + \frac{1}{8}\psi^{-1}(D(\eta/8, d)) + \ldots\right] \end{aligned}$$

$$\begin{aligned}
&\leq \; 4K_0\eta \int_0^1 \psi^{-1}(D(\eta u, d))\, du \\
&= \; 4K_0 \int_0^1 \psi^{-1}(D(\epsilon, d))\, d\epsilon\,,
\end{aligned}$$

and we are done. □

An important corollary of the above theorem is Corollary 2.2.8. of VDVW (Corollary 8.5 of Kosorok) which constructs a bound on the expected modulus of continuity of a separable sub-Gaussian process $X$ in terms of the logarithms of its packing numbers. The corollary says that:

$$E\left(\sup_{s,t\in T_\infty; d(s,t)\leq\delta} |X(s)-X(t)|\right) \leq K \int_0^\delta \sqrt{\log\, D(\epsilon,d)}\, d\epsilon\,.$$

Below, we provide an an argument. Since the process $X$ is sub-Gaussian with respect to $d$, we can apply inequality (1.1) to conclude that $\|X(s)-X(t)\}_{\psi_2} \lesssim d(s,t)$ and we can apply the Theorem above with $\psi(x)=\psi_2(x)=e^{x^2}-1$. Also choose $\eta=\delta$. Note that $\psi_2^{-1}(m)=\sqrt{\log(1+m)}$. We thus obtain:

$$\left\|\sup_{s,t\in T_\infty; d(s,t)\leq\delta} |X(s)-X(t)|\right\|_{\psi_2} \lesssim \int_0^\delta \sqrt{\log(1+D(\epsilon,d))} d\epsilon + \delta\,\sqrt{\log(1+D^2(\delta,d))}\,.$$

First consider the case that $\delta < \mathrm{diam}(T)$. In this case $D(\epsilon,d)\geq 2$ for all $\epsilon\leq\delta$. For $m\geq 2$, we have $1+m\leq m^2$ and hence $\sqrt{\log(1+m)}\leq\sqrt{2}\,\sqrt{\log m}$. Using this fact, it is easy to see that at the expense of introducing a larger (universal) constant, we have:

$$\left\|\sup_{s,t\in T_\infty; d(s,t)\leq\delta} |X(s)-X(t)|\right\|_{\psi_2} \lesssim \int_0^\delta \sqrt{\log D(\epsilon,d)} d\epsilon + \delta\,\sqrt{\log D^(\delta,d)} \lesssim \int_0^\delta \sqrt{\log D(\epsilon,d)} d\epsilon\,,$$

since $D(\epsilon,d)$ is decreasing in $\epsilon$. When $\delta=\mathrm{diam}(T)$, the first term in the display preceding the last can still be dominated, up to a universal constant, by the extreme right expression of the above display (when $\epsilon=\mathrm{diam}(T)$, $D(\epsilon,d)=1$ but this is at a single point and does not cause a problem). The second term in the above display is simply $\sqrt{\log 2}\,\mathrm{diam}(T)$ which is no larger than $\int_0^T \sqrt{\log D(\epsilon,d)} d\epsilon$, since for $\epsilon<\mathrm{diam}(T), D(\epsilon,d)\geq 2$. Finally, one uses the fact that the $L_2$ norm is less than a constant times the $\psi_2$ norm; see, for example, Problem 4 on page 105 of VDVW. Also note that the above inequality can also be written in terms of an integral involving the entropy number (the log of the covering number) for a tweaked universal constant. If for some $\delta=\delta_0$, the integral on the right side of the above display converges, the DCT shows that the expected modulus of continuity of $X$ goes to 0 as $\delta$ goes to 0.

The above inequality will play an important role in establishing a Donsker theorem later on where a sufficient condition for a class of functions to be Donsker will be developed in terms of the finiteness of an entropy integral. The technique will involve bounding the modulus of conitnuity of the empirical process $\mathbb{G}_n$ as $n$ increases. By a symmetrization technique to be discussed shortly, the expected modulus of continuity can be dominated by an analogous quantity corresponding to a symmetrized version of the empirical process using independent Rademacher variables. This latter quantity can be controlled well using the above inequality as the symmetrized version of the empirical process is sub-Gaussian for the $L_2(\mathbb{P}_n)$ semimetric.

The next lemma provides a sufficient condition for the process $\{X(t) : t \in T\}$ to induce a tight Borel probability measure on $l_\infty(T)$ in terms of the packing numbers $D(\epsilon, d)$.

**Lemma 1.3** *Consider the stochastic process $X$ as defined in Theorem 1. Also, let $\psi$ be as above and let $D(\epsilon, d)$ be finite for all positive $\epsilon$. Further, assume that $\int_0^{\eta_0} \psi^{-1}(D(\epsilon, d))\, d\epsilon < \infty$ for some $\eta_0 > 0$. Then, the process $X$ induces a tight Borel probability measure on $l(T)$.*

**Proof:** By the last (assumed) condition, it follows that $\int_0^{\eta} \psi^{-1}(D(\epsilon, d))\, d\epsilon \to 0$ as $\eta \to 0$. By the first conclusion of Theorem 1 it then follows that as $\delta \to 0$, $\left\|\sup_{d(s,t)\le\delta} |X(s) - X(t)|\right\|_\psi \to 0$ (why?) and therefore, by Exercise 8.5.5 of Kosorok, $\sup_{d(s,t)\le\delta} |X(s) - X(t)| \to_P^\star 0$. By Exercise 17 on Page 106 of VDVW, conclude that almost all sample paths of $X$ are uniformly continuous with respect to $d$. Sinc $D(\epsilon, d)$ is finite for all $d$, $(T, d)$ is totally bounded (and diam $T$ is finite). It is easy to see (using the total boundedness of $T$ and the almost sure uniform continuity of the sample paths of $X$) that almost all sample paths of $X(t)$ are bounded. Then, by Proposition 0 of the previous set of notes, the conclusion of the lemma follows directly.

**Solution to Exercise 8.5.5. of Kosorok (2008):** Since $\psi$ is an increasing convex non-zero function with $\psi(0) = 0$, there exists $x_0 > 0$ such that $\phi(x_0) > 0$. Take any $x > x_0$. By convexity:

$$\frac{\psi(x_0)}{x_0} = \frac{\psi(x_0) - \psi(0)}{x_0 - 0} \le \frac{\psi(x) - \psi(x_0)}{x - x_0},$$

which readily implies that

$$\frac{\psi(x)}{x} \ge \frac{x - x_0}{x} \times \frac{\psi(x_0)}{x_0}.$$

Taking liminfs on either side then readily yields the conclusion that $\liminf_{x\to\infty} \psi(x)/x > 0$. This proves the hint. Without loss of generality we take $X_n$'s to be positive. Now, let $c_n$ denote the $\psi$-Orlicz norm of $X_n$. We have, for any $M > 0$,

$$1 \geq E[\psi(X_n/c_n)] = E(\psi(X_n/c_n)\,1(X_n/c_n \leq M)) + E(\psi(X_n/c_n)\,1(X_n/c_n > M))\,.$$

Now, choose $M$ so large that $u > M$ implies that $\psi(x)/x \geq \lambda$ for some $\lambda > 0$. By the statement in the hint, this can be achieved. Conclude that:

$$1 \geq E(\lambda\,(X_n/c_n)\,1(X_n > c_n M))$$

and therefore

$$c_n \geq \lambda\,E(X_n\,1(X_n > c_n M))\,,$$

showing that the term on the right side goes to 0. Now

$$E(X_n) = E(X_n\,1(X_n \leq c_n\,M)) + E(X_n\,1(X_n > c_n\,M)) \leq c_n\,M + E(X_n\,1(X_n > c_n\,M)) \to 0\,,$$

which implies convergence in probability. □

**Partial solution to Problem 8.5.3 of Kosorok (2008) – going from $\psi$ to $\phi$:** By assumption there exist $K, K' > 1$ such that for all $x, y \geq K$, $\psi(x)\psi(y) \leq K^{'}\,\psi(cxy)$. Note that if this holds for some pair $(K, K')$ it will hold for any other pair that is component-wise larger than this pair. Thus, for all $x, y \geq 1$,

$$\psi(Kx)\,\psi(Ky) \leq K^{'}\,\psi(cK^2xy)\,.$$

Now, we seek $\sigma, \tau > 0$ and $\sigma \leq 1$ such that $\phi(x) = \sigma\,\psi(\tau x)$ satisfies $\phi(1) \leq 1/2$ and $\phi(x)\phi(y) \leq \phi(\tilde{c}xy)$ for all $x, y \geq 1$, where $\tilde{c} > 0$ depends solely on $\psi$. This last condition, after some simple algebra, translates to:

$$\psi(\tau x)\psi(\tau y) \leq \frac{1}{\sigma}\,\psi(\tilde{c}\tau xy)\,.$$

We also need $\sigma\,\psi(\tau) \leq 1/2$. So choose $\tau = K$, $\sigma$ strictly smaller than the minimum of $1/2\,\psi(K)$ and $1/K'$ and $\tilde{c} = c\,K$. These choices immediately ensure that the desired conditions on $\phi$ are met and note that $\tilde{c}$ still depends completely on the function $\psi$.