L09 Strong consistent estimator and SLLN

1. Properties of almost sure convergence

   In math $X_n \to X \iff g(X_n) \to g(X)\,\forall$ continuous $g(\cdot)$, $X_n \to X \iff X_{n_k} \to X\,\forall\, X_{n_k}$; and $X_n \to X \in R^p \iff (X_n)_k \to X_k\,\forall\, k = 1,..,p$. Similar results hold for the convergence with probability 1.

   (1) $X_n \stackrel{wp1}{\longrightarrow} X \iff g(X_n) \stackrel{wp1}{\longrightarrow} g(X)$ for all continuous $g(\cdot)$.

   (2) $X_n \stackrel{wp1}{\longrightarrow} X \iff X_{n_k} \stackrel{wp1}{\longrightarrow} X$ for all subsequence X_{n_k}.

   (3) $X_n \stackrel{wp1}{\longrightarrow} X \in R^p \iff (X_n)_k \stackrel{wp1}{\longrightarrow} X_k$ for all $k = 1,..,p$ and $(X_n)_k$, $k = 1,..,p$, are defined in the same probability space.

   Proof: $\Rightarrow$**:** There exists $A \subset \Omega$ with $P(A) = 1$. On A $X_n \longrightarrow X$. Hence on A
   (1) $g(X_n) \longrightarrow g(X)$; (2) $X_{n_k} \longrightarrow X$; and (3) $(X_n)_k \longrightarrow X_k$.
   $\Leftarrow$**:** (1) Take continuous $g(x) = x$. (2) Take $X_{n_k} = X_n$.
   (3) There are $A_i \subset \Omega_i$ with $P(A_i) = 1$. On A_i $(X_n)_i \longrightarrow X_i$.
   But $\Omega_1 = \cdots = \Omega_p = \Omega$. So $A = \cap A_i \subset \Omega$ with $P(A) = 1$. On A, $X_n \longrightarrow X$.

   Comment: On $A \subset \Omega$, the convergence $X_n(w) \longrightarrow X(w)$ is in the ordinary mathematical sense by $\epsilon - N$ language. Whether or not this convergence is wp1 depends on if $P(A) = 1$.

2. Strong consistent estimator

   (1) Definition
   Estimator η_n from a sample of size n for parameter η is a strong consistent estimator if $\widehat{\eta}_n \stackrel{a.s.}{\longrightarrow} \eta$.

   (2) Comment
   In $X_n \stackrel{a.s.}{\longrightarrow} X$, generally X is random. But in $\widehat{\eta}_n \stackrel{a.s.}{\longrightarrow} \eta$, η is non-random and unknown.

3. Strong law of large numbers (SLLN)

   (1) Lemma 1
   Let $\overline{X}_n$ be the mean of a sample of size n from $X \sim (\mu,\, \sigma^2)$. Then $\overline{X}_{n^2} \stackrel{wp1}{\longrightarrow} \mu$.

   Proof: Consider continuous X with pdf $f(x)$ be the pdf of X.

   $$\begin{aligned} P(|X| > \epsilon) &= P(X^2 > \epsilon^2) = \int_{x^2>\epsilon^2} 1 \cdot f(x)dx \le \int_{x^2>\epsilon^2} \tfrac{x^2}{\epsilon^2} f(x)dx \\ &\le \int_x \tfrac{x^2}{\epsilon^2}\, f(x)\, dx = \tfrac{E(X^2)}{\epsilon^2}. \end{aligned}$$

   $P(|X| > \epsilon) \le \frac{E(X^2)}{\epsilon^2}$ is the Chebyshev's inequality. by Chebyshev's inequality

   $$P(|\overline{X}_{n^2} - \mu| > \epsilon) \le \frac{E(\overline{X}_{n^2}-\mu)^2}{\epsilon^2} = \frac{\text{var}(\overline{X}_{n^2})}{\epsilon} = \frac{\sigma^2}{n^2\epsilon^2}.$$

   So $\sum_{n=1}^{\infty} P\left(|\overline{X}_{n^2} - \mu| > \epsilon\right) \le \frac{\sigma^2}{\epsilon^2}\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty$.
   By Borel-Cantelli lemma, $\overline{X}_{n^2} \stackrel{wp1}{\longrightarrow} \mu$. □

   (2) Lemma 2
   Let $\overline{X}_n$ be the mean of a sample of size n from $X \sim (\mu,\, \sigma^2)$ with $X \ge 0$. Then $\overline{X}_n \stackrel{wp1}{\longrightarrow} \mu$.

Proof. For n there exists m such that

$$\begin{aligned} m^2 \le n \le (m+1)^2 &\implies X_1+\cdots+X_{m^2} \le X_1+\cdots+X_n \le X_1+\cdots+X_{(m+1)^2} \\ &\implies \tfrac{X_1+\cdots+X_{m^2}}{(m+1)^2} \le \tfrac{X_1+\cdots+X_n}{n} \le \tfrac{X_1+\cdots+X_{(m+1)^2}}{m^2} \\ &\iff \overline{X}_{m^2}\tfrac{m^2}{(m+1)^2} \le \overline{X}_n \le \overline{X}_{(m+1)^2}\tfrac{(m+1)^2}{m^2}. \end{aligned}$$

By (1) $\overline{X}_{m^2} \xrightarrow{wp1} \mu$. So $\exists\, A \subset \Omega$ with $P(A)=1$ on A,

$$\overline{X}_{m^2} \longrightarrow \mu \Longrightarrow \overline{X}_{m^2}\tfrac{m^2}{(m+1)^2} \longrightarrow \mu.$$

By (1) $\overline{X}_{(m+1)^2} \xrightarrow{wp1} \mu$. So $\exists\, B \subset \Omega$ with $P(B)=1$ on B,

$$\overline{X}_{(m+1)^2} \longrightarrow \mu \Longrightarrow \overline{X}_{(m+1)^2}\tfrac{(m+1)^2}{m^2} \longrightarrow \mu.$$

Thus on $A\cap B$ with $P(A\cap B)=1$,

$$\overline{X}_{m^2}\tfrac{m^2}{(m+1)^2} \le \overline{X}_n \le \overline{X}_{(m+1)^2}\tfrac{(m+1)^2}{m^2} \Longrightarrow \overline{X}_n \longrightarrow \mu.$$

So $\overline{X}_n \xrightarrow{a.s.} \mu$. $\square$

(3) Lemma 3

Let $\overline{X}_n$ be the mean of a sample of size n from $X \sim (\mu,\, \sigma^2)$ with $X \le 0$. Then $\overline{X}_n \xrightarrow{wp1} \mu$.

Proof. $X \sim (\mu,\, \sigma^2) \Longrightarrow -X \sim (-\mu,\, \sigma^2) \overset{(2)}{\Longrightarrow} -\overline{X}_n \xrightarrow{a.s.} -\mu \Longrightarrow \overline{X}_n \xrightarrow{a.s.} \mu$. $\square$

(4) Theorem

Strong law of large numbers (SLLN), Univariate version

Let $\overline{X}_n$ be the mean of a sample of size n from $X \sim (\mu,\, \sigma^2)$. Then $\overline{X}_n \xrightarrow{wp1} \mu$.

Proof. Let $(\overline{X}_n)_+ = \max[0,\, \overline{X}_n]$, $(\overline{X}_n)_- = \min[0,\, \overline{X}_n]$, $X_+ = \max[0,\, X]$ and $X_- = \min[0,\, X]$. Then $(\overline{X}_n)_+$ is the mean of a random sample of size n from $X_+ \sim (\mu_+,\, \sigma_+^2)$ where $\mu_+ = \int_0^\infty xf(x)dx$, and $(\overline{X}_n)_-$ is the mean of a random sample of size n from $X_- \sim (\mu_-,\, \sigma_-^2)$ where $\mu_- = \int_{-\infty}^0 xf(x)dx$.

By (2) $(\overline{X}_n)_+ \xrightarrow{a.s.} \mu_+$, i.e., $\exists\, A \subset \Omega$ with $P(A)=1$ on A $(\overline{X}_n)_+ \longrightarrow \mu_+$.
By (3) $(\overline{X}_n)_- \xrightarrow{a.s.} \mu_-$, i.e., $\exists\, B \subset \Omega$ with $P(B)=1$ on B $(\overline{X}_n)_- \longrightarrow \mu_-$.
Thus on $A\cap B$ with $P(A\cap B)=1$,

$$\begin{aligned} \overline{X}_n = (\overline{X}_n)_+ + (\overline{X}_n)_- \longrightarrow \mu_+ + \mu_- &= \int_0^\infty xf(x)dx + \int_{-\infty}^0 xf(x)dx \\ &= \int_{-\infty}^\infty xf(x)dx = \mu. \end{aligned}$$

Thus $\overline{X}_n \xrightarrow{a.s.} \mu$. $\square$

(5) Theorem SLLN: Multivariate version

Let $\overline{X}_n \in R^p$ be the mean of a sample of size n from $X \sim (\mu,\, \Sigma)$. Then $\overline{X}_n \xrightarrow{wp1} \mu \in R^p$.

Proof. $\overline{X}_n = \begin{pmatrix} (\overline{X}_n)_1 \\ \vdots \\ (\overline{X}_n)_k \end{pmatrix}$ is the mean of a random sample of size n from

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \sim \left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix},\, \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_p^2 \end{pmatrix} \right).$$

So $(\overline{X}_n)_i$ is the mean of a random sample of size n from $X_i \sim (\mu_i,\, \sigma_i^2)$, $i = 1, ..., p$.
By (4), $(\overline{X}_n)_i \xrightarrow{wp1} \mu_i$ for all $i = 1, ..., p$. By (3) in 1, $\overline{X}_n \xrightarrow{wp1} \mu$. $\square$

L10 Convergence in probabilities

1. Convergence in probabilities

   (1) Definition
   Let $A(\epsilon, n) = \{w \in \Omega : \|X_n(w) - X(w)\| > \epsilon\} \subset \Omega$. We say that X_n converges to X in probability if $\lim_n P(A(\epsilon, n)) = 0$ for all $\epsilon > 0$.
   This convergence is denoted by "$X_n \longrightarrow X$ in probability", $X_n \xrightarrow{p} X$ or $\text{plim}_n X_n = X$.
   Clearly
   $$\begin{aligned} X_n \xrightarrow{p} X &\iff \lim_n P(\|X_n - X\| > \epsilon) = 0 \ \forall \epsilon \\ &\iff \lim_n P(\|X_n - X\| \leq \epsilon) = 1 \ \forall \epsilon > 0 \\ &\iff \lim_n P\left(\|X_n - X\| > \tfrac{1}{k}\right) = 0 \ \forall k > 0 \\ &\iff \lim_n P\left(\|X_n - X\| \leq \tfrac{1}{k}\right) = 1 \ \forall k > 0. \end{aligned}$$
   In the above expressions "$>$" in $> \epsilon$ and $> \frac{1}{k}$ can be replaced by "$\geq$".
   "$\leq$" in $\leq \epsilon$ and $\leq \frac{1}{k}$ can be replaced by "$<$".

   (2) Other expressions
   $X_n \xrightarrow{p} X \iff X_n - X \xrightarrow{p} 0 \iff \|X_n - X\| \xrightarrow{p} 0.$
   Proof: In $P(\|X_n - X\| > \epsilon) = 0$, $\|X_n - X\| = \|(X_n - X) - 0\| = |\|X_n - X\| - 0|$.
   So $X_n \xrightarrow{p} X \iff X_n - X \xrightarrow{p} 0 \iff \|X_n - X\| \xrightarrow{p} 0.$

2. Three iff properties

   (1) $X_n \xrightarrow{p} X \iff g(X_n) \xrightarrow{p} g(X)$ for all continuous $g(\cdot)$.
   Proof. $\Leftarrow$**:** With continuous $g(x) = (x)$, $g(X_n) \xrightarrow{p} g(X) \Longrightarrow X_n \xrightarrow{p} X$.
   $\Rightarrow$**:** By the definition of the continuity of $g(\cdot)$,
   $\forall \epsilon > 0, \exists \delta > 0,$ such that $\|X_n - X\| < \delta \Longrightarrow \|g(X_n) - g(X)\| < \epsilon$.
   So $[w \in \Omega : \|X_n(w) - X(w)\| < \delta] \subset [w \in \Omega : \|g(X_n(w)) - g(X(w))\| < \epsilon]$.
   Thus $P(\|X_n - X\| < \delta) \leq P(\|g(X_n) - g(X)\| < \epsilon)$.
   But $P(\|X_n - X\| < \delta) \longrightarrow 1$ as $n \longrightarrow \infty$ since $X_n \xrightarrow{p} X$.
   So $P(\|g(X_n) - g(X)\| < \epsilon) \longrightarrow 1$ as $n \longrightarrow \infty$. Hence $g(X_n) \xrightarrow{p} g(X)$. $\square$

   (2) $X_n \xrightarrow{p} X \iff X_{n_k} \xrightarrow{p} X$ for all subsequence X_{n_k}.
   Proof. $\Leftarrow$**:** Take X_n as its own subsequence. Then $X_n \xrightarrow{p} X$.
   $\Rightarrow$**:** For $\epsilon > 0$, $p_n = P(\|X_n - X\| < \epsilon) \longrightarrow 1$ since $X_n \xrightarrow{p} X$.
   But $p_n \longrightarrow 1 \Longrightarrow p_{n_k} \longrightarrow 1$, i.e., $P(\|X_{n_k} - X\| < \epsilon) \longrightarrow 1$.
   Therefore $X_{n_k} \xrightarrow{p} X$. $\square$

   (3) $X_n = \begin{pmatrix} X_{1n} \\ \vdots \\ X_{kn} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} = X \iff$ In $X_n = \begin{pmatrix} X_{1n} \\ \vdots \\ X_{kn} \end{pmatrix}$ $X_{in} \xrightarrow{p} X_i$ for all $i = 1, ..., k$.
   Proof. $\Rightarrow$**:** With continuous $g_i(X) = e_i' X = X_i$, by (1)
   $$X_{in} = g_i(X_n) \xrightarrow{p} g_i(X) = X_i \text{ for all } i = 1, ..., k.$$
   $\Leftarrow$**:** If $k = 1$, then $\Leftarrow$ holds. By induction on k, assume when $k = t$ $\Leftarrow$ is true. With
   $k = t + 1$, write $X_n = \begin{pmatrix} X_{1n} \\ \vdots \\ X_{tn} \\ X_{t+1,n} \end{pmatrix} = \begin{pmatrix} X_{In} \\ X_{t+1,n} \end{pmatrix}$. and $X = \begin{pmatrix} X_1 \\ \vdots \\ X_t \\ X_{t+1} \end{pmatrix} = \begin{pmatrix} X_I \\ X_{t+1} \end{pmatrix}$.

Let $A_1 = \left(\|X_{In} - X_I\|^2 < \frac{\epsilon^2}{2}\right)$, $A_2 = \left(|X_{t+1,n} - X_{t+1}|^2 < \frac{\epsilon^2}{2}\right)$ and $A = \left(\|X_n - X\|^2 < \epsilon^2\right)$. Then $A_1 \cap A_2 \subset A$. So $P(A_1 \cap A_2) \leq P(A)$. Thus $P(A_1) + P(A_2) - P(A_1 \cup A_2) \leq P(A)$. Let $n \to \infty$. Then $lim_n P(A) \geq 1 + 1 - \lim_n P(A_1 \cup A_2) \geq 1$. So $P(\|X_n - X\|^2 < \epsilon) \longrightarrow 1$. Therefore $X_n \xrightarrow{p} X$. $\square$

3. Consistent estimators, asymptotically UE estimators and asymptotically efficient estimators

(1) A relation: $X_n \xrightarrow{as} X \Rightarrow X_n \xrightarrow{p} X$

Proof. Recall that

$$\begin{aligned}
X_n \xrightarrow{a.s.} X &\overset{def}{\iff} P\left(\cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} \left[\|X_n - X\| < \tfrac{1}{k}\right]\right) = 1 \\
&\iff P\left(\cup_{k=1}^{\infty} \cap_{N=1}^{\infty} \cup_{n=N}^{\infty} \left[\|X_n - X\| \geq \tfrac{1}{k}\right]\right) = 0 \\
&\iff \forall\, k > 0,\ P\left(\cap_{N=1}^{\infty} \cup_{n=N}^{\infty} \left[\|X_n - X\| \geq \tfrac{1}{k}\right]\right) = 0 \\
&\iff \forall\, k > 0,\ P\left(\lim_{N\to\infty} \cup_{n=N}^{\infty} \left[\|X_n - X\| \geq \tfrac{1}{k}\right]\right) = 0 \\
&\iff \forall\, k > 0,\ \lim_{N\to\infty} P\left(\cup_{n=N}^{\infty} \left[\|X_n - X\| \geq \tfrac{1}{k}\right]\right) = 0.
\end{aligned}$$

Note that for all $k > 0$, $\left[\|X_N - X\| \geq \frac{1}{k}\right] \subset \cup_{n=N}^{\infty} \left[\|X_n - X\| \geq \frac{1}{k}\right]$.
So $P\left(\|X_N - X\| \geq \frac{1}{k}\right) \leq P\left(\cup_{n=N}^{\infty} \left[\|X_n - X\| \geq \frac{1}{k}\right]\right)$.
Thus $\lim_{N\to\infty} P\left(\|X_N - X\| \geq \frac{1}{k}\right) \leq \lim_{N\to\infty} P\left(\cup_{n=N}^{\infty} \left[\|X_n - X\| \geq \frac{1}{k}\right]\right) = 0$.
Hence $\lim_{N\to\infty} P\left(\|X_N - X\| \geq \frac{1}{k}\right) = 0$. Therefore $X_n \xrightarrow{p} X$. $\square$

(2) Weak law of large numbers (WLLN)
Let $\overline{X}_n$ be the mean of a sample of size n from $X \sim N(\mu, \Sigma)$. Then $\overline{X}_n \xrightarrow{p} \mu$.

(3) Definitions of consistent estimator and etc
Suppose $\widehat{\eta}_n$ is an estimator for η based on a random sample of size n.

(i) Consistent estimator

$\widehat{\eta}_n$ is a consistent estimator $\overset{def}{\iff} \widehat{\eta}_n \xrightarrow{p} \eta$.

(ii) Asymptotically unbiased estimator

$\widehat{\eta}_n$ is an asymptotically unbiased estimator $\overset{def}{\iff} E(\widehat{\eta}_n) \longrightarrow \eta$.

(iii) Asymptotically efficient estimator

$\widehat{\eta}_n$ is an asymptotically efficient estimator $\overset{def}{\iff} \widehat{\eta}_n \xrightarrow{p} \eta$ and
$n\left[\mathrm{Cov}(\widehat{\eta}_n) - \mathrm{CRLB}(\eta)\right] \to 0$.

Ex: Let $\widehat{\mu}_n = \frac{n}{n-1}\overline{X}_n$ be an estimator for μ in $X \sim N(\mu, \sigma^2)$.

(i) By SLLN, $\overline{X}_n \xrightarrow{a.s.} \mu$. Thus $\exists A \subset \Omega$ with $P(A) = 1$ on A $\overline{X}_n \longrightarrow \mu \Longrightarrow \widehat{\mu}_n \longrightarrow \mu$.
So $\widehat{\mu}_n \xrightarrow{a.s.} \mu \Longrightarrow \widehat{\mu}_n \xrightarrow{p} \mu$. Thus $\widehat{\mu}_n$ is a consistent estimator for μ.

(ii) $\widehat{\mu}_n \sim \left(\frac{n}{n-1}\mu,\ \frac{n}{(n-1)^2}\sigma^2\right)$. Here $E(\widehat{\mu}_n) \neq \mu$, but $E(\widehat{\mu}_n) \longrightarrow \mu$. So $\widehat{\mu}_n$ is a biased estimator, but asymptotically unbiased estimator for μ.

(iii) $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$, $I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$, $\mathrm{CRLB}(\mu) = \frac{\sigma^2}{n}$ and $\mathrm{CRLB}\left(\frac{n}{n-1}\mu\right) = \frac{n}{(n-1)^2}\sigma^2$.

(iv) With n fixed, $\mathrm{var}(\widehat{\mu}_n) = \mathrm{CRLB}\left(\frac{n}{n-1}\mu\right)$. So $\widehat{\mu}_n$ is an efficient statistic, a MVUE for $\frac{n}{n-1}\mu$, but not for μ.
$n\left[\mathrm{var}(\widehat{\mu}_n) - \mathrm{CRL}(\mu)\right] = \frac{n^2}{(n-1)^2}\sigma^2 - \sigma^2 \longrightarrow 0$ along with $\widehat{\mu}_n \xrightarrow{p} \mu$, $\widehat{\mu}_n$ is an asymptotically efficient estimator for μ.