Law of Large Numbers

Describle the behavior of the sample mean of i.i.d. as the sample size grows.

$X_1, X_2, \_\_\_, X_n$ i.i.d. with some distribution.

$\mu < \infty$. $\sigma^2 < \infty$.

$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

Theorem: (Week Law of Large Numbers).

$\forall \varepsilon > 0$, $P(|\overline{X} - \mu| > \varepsilon) \xrightarrow[\text{as } n \to \infty]{} 0$

(converge to μ in probability).

Proof: by Chebychev's inequality.

$$P(|\overline{X} - \mu| > \varepsilon) \le \frac{\sigma^2}{n\varepsilon^2} \qquad (\mathrm{Var}\,\overline{X} = \frac{\sigma^2}{n})$$

$$\lim_{n \to \infty} \frac{\sigma^2}{n\varepsilon^2} = 0$$

$\Rightarrow$ L.H.S also converges to 0.

Strong law of large numbers:

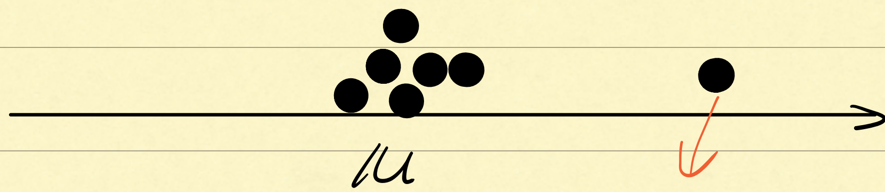
with probability 1 (wp1) or almost surely (as).

$\overline{X} \to \mu$ as $n \to \infty$

Difference between weak convergence probability and strong wp1 (as)

△ Weak Law of Large Numbers

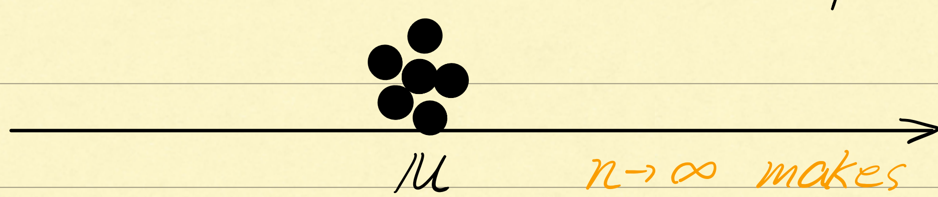
$P\{|\bar{x}-\mu|\geq\varepsilon\}\to 0$ as $n\to+\infty$ for $\forall\, \varepsilon>0$.



$n\to\infty$ counteracts the influence of outliers.

△ Strong Law of Large Numbers.

$P\{|\bar{x}-\mu|\geq\varepsilon \text{ as } n\to+\infty\}=0$ for $\forall\, \varepsilon>0$.



$n\to\infty$ makes outliers don't exist.

Central Limit Theorem (CLT)

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0,1) \text{ as } n \to +\infty.$$

Z converges in distribution to $N(0,1)$ as $n \to \infty$. i.e. $P\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\}$

$$\to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx.$$

Proof: Assume X_i's MGF exist

Assume $\mu = 0$ $\sigma^2 = 1$. we can use lineer transformations to get other situations.

MGF of X_i: $M(t) = E(e^{tX_i})$.

$$M(0) = 1, \; M'(0) = EX_i = 0, \; M''(0) = E X_i^2 = 1.$$

MGF of $\sqrt{n}\bar{X}$: $M_{new}(t) = E(e^{t\sqrt{n}\bar{X}}) = E(e^{t\frac{X_1 + X_2 + \cdots X_n}{\sqrt{n}}})$

$$= E e^{t \cdot \frac{X_1}{\sqrt{n}}} \cdot E e^{t \cdot \frac{X_2}{\sqrt{n}}} \cdots \cdots E e^{t \cdot \frac{X_n}{\sqrt{n}}}$$

$$= [M(\tfrac{t}{\sqrt{n}})]^n$$

MGF of $Z \sim N(0,1)$: $M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + xt} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2 + \frac{t^2}{2}} dx = e^{\frac{t^2}{2}}$

$$\lim_{n\to\infty} \log M_{new}(t) = \lim_{n\to\infty} n \log M\left(\frac{t}{\sqrt{n}}\right) \qquad y = \frac{1}{\sqrt{n}}$$

$$= \lim_{y\to 0} \frac{\log M(yt)}{y^2}$$

$$= \lim_{y\to 0} \frac{t M'(yt)}{2y\, M(yt)}$$

$$= \lim_{y\to 0} \frac{t^2 M''(yt)}{2M(yt) + 2yt M'(yt)}$$

$$= \frac{t^2}{2} \frac{M''(0)}{M(0)} = \frac{t^2}{2}$$

Hence $M_{new}(t) = M_Z(t)$

$$\Rightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty.$$

Chi-Square distribution.

Let $V = Z_1^2 + \cdots + Z_n^2$ where Z_i are i.i.d. $N(0,1)$

$V \sim \chi^2_{(n)}$ is said to have Chi-Square distribution with n degrees of freedom.

Theorem: $\chi^2_{(n)}$ is the same as $\text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$.

Proof: $F_{Z^2}(x) = P(Z^2 \leq x)$

$$= P(-\sqrt{x} \leq Z \leq \sqrt{x})$$

$$= 2\Phi(\sqrt{x}) - 1$$

$$f_{Z^2}(x) = \frac{\phi(\sqrt{x})}{\sqrt{x}} = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \longrightarrow \text{p.d.f of Gamma}(\tfrac{1}{2}, \tfrac{1}{2})$$

Hence $\chi^2_{(n)} = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \underline{\text{Gamma}(\frac{n}{2}, \frac{1}{2})}$

$$E\,\chi^2_{(n)} = n \cdot E Z^2 = n \qquad = \frac{\frac{n}{2}}{\frac{1}{2}}$$

$$\text{Var}\,\chi^2_{(n)} = n \cdot \text{Var}\, Z^2 = n \cdot (E Z^4 - E^2 Z^2) \qquad = \frac{\frac{n}{2}}{(\frac{1}{2})^2}$$

$$= n \cdot (3 - 1)$$

$$= 2n.$$

$$M_{\chi^2_{(n)}}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}, \quad t < \frac{1}{2}.$$

Example: $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})^2$$

Fact: $\dfrac{S_n^2 (n-1)}{\sigma^2} \sim \chi^2_{(n-1)}$

Assume $Z_i \sim N(0,1)$

Proof: $\displaystyle\sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2 = \sum_{i=1}^{n} (Z_i - \bar{Z})^2$

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} (Z_i - \bar{Z})^2 + n\bar{Z}^2$$

Hence, MGF of $\sum_{i=1}^{n}(Z_i-\bar{Z})^2+n\bar{Z}^2$ is $\left(\frac{1}{1-2t}\right)^{\frac{n}{2}}$

Since $\sum_{i=1}^{n}(Z_i-\bar{Z})^2$ is independent of $\bar{Z}$

and $\bar{Z}\sim(0,\frac{1}{n}) \Longrightarrow n\bar{Z}^2\sim\chi^2_{(1)}$

we know $M_{\sum(Z_i-\bar{Z})^2}(t)\cdot M_{n\bar{Z}^2}(t)=M_{\chi^2_{(n)}}(t)$

$$M_{\sum(Z_i-\bar{Z})^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{n}{2}} \Big/ \left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$$

$$=\left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}}$$

$\Longrightarrow \sum(Z_i-\bar{Z})^2\sim\chi^2_{(n-1)}$

$\Longrightarrow \frac{S_n^2(n-1)}{\sigma^2}\sim\chi^2_{(n-1)}$

Define: t-distribution

$$T=\frac{Z}{\sqrt{V/n}}, \quad Z\sim N(0,1) \quad V\sim\chi^2_{(n)}$$

T is called the t-distribution with n degrees of freedom.

$$f_T(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

Properties: 1. Symmetry: if $T\sim t_n$, $-T\sim t_n$

2. t_1 is the same as Cauchy distribution.

3. As $n \to \infty$, the t_n distribution approaches the standard Normal distribution.

Proof: ① $T = \dfrac{Z}{\sqrt{\frac{V}{n}}}$ $\quad -T = \dfrac{-Z}{\sqrt{\frac{V}{n}}}$, $-Z \sim N(0,1)$.

②. when $n=1$ $\quad T = \dfrac{Z}{V}$, $Z \sim N(0,1)$
$V \sim N(0,1)$

③ $n \to \infty$ $\quad \dfrac{V}{n} = \dfrac{Z_1^2 + \cdots + Z_n^2}{n} \longrightarrow E(Z_i^2) = 1$.

$\Longrightarrow T \longrightarrow Z \sim N(0,1)$.