# Central Limit Theorem

a large sum of i.i.d. r.v., properly normalized. will have approx. a normal. distribution.

1. CLT: Suppose $X_1, X_2, \dots$ is an i.i.d. sequence of r.v. with finite $\mu$, finite $\sigma^2$. Let sum $S_n = X_1 + \dots + X_n$, and sample mean $M_n = \frac{S_n}{n}$.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\left(\frac{M_n - \mu}{\sigma}\right).$$

(Since $E(M_n) = \mu$, $Var(M_n) = \frac{\sigma^2}{n}$, gives $E(Z_n) = 0$, $Var(Z_n) = 1$).

Let $Z \sim N(0,1)$. Then probability of $Z_n$ converge to those of $Z \sim N(0,1)$.

i.e. $F_{Z_n}(z) \to F_Z(z) = \Phi(z)$. for $z \in \mathbb{R}$.

1). $\forall z \in \mathbb{R}$. $\lim_{n\to\infty} P(Z_n \leq z) = P(Z \leq z)$. i.e.

$$\hookrightarrow \lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\cdot\sigma} \leq z\right) = \lim_{n\to\infty} P(S_n \leq n\cdot\mu + \sqrt{n}\cdot\sigma z)$$

$$= \lim_{n\to\infty} P\left(\frac{S_n}{n} \leq \mu + \frac{\sigma}{\sqrt{n}}\cdot z\right) = P(Z \leq z) = \Phi(z)$$

2). $\frac{S_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} \approx Z$. and $\frac{S_n}{n} \approx \mu + \frac{\sigma}{\sqrt{n}} Z$. where $Z \sim N(0,1)$.

$\to$ converges to $\mu$. SLLN.

2. Normal Approximation: $n \to \infty$, $P\left(\frac{S_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} \leq z\right) \to \Phi(z)$.

e.g. Suppose $\{X_n\}$ are i.i.d. $X \sim$ Poisson $(4)$. Approx. for $P(X_1 + \dots + X_{900} \geq 3700)$.

providing a num. we need to apply them onto that num.

$n = 900$, which $S_n = X_1 + \dots + X_{900}$.

Since it's i.i.d. $\mu = E(X_n) = \lambda$. $\sigma^2 = Var(X_n) = \lambda \Rightarrow \sigma = \sqrt{\lambda}$.

$$P(X_1 + \dots + X_{900} \geq 3700) = P\left(\frac{S_{900} - 900\mu}{\sqrt{900}\cdot\sqrt{\lambda}} \geq \frac{3700 - 900\mu}{\sqrt{900}\cdot\sqrt{\lambda}}\right).$$

$$= P\left(\frac{S_{900} - 3600}{30\times 2} \geq \frac{3700 - 3600}{60}\right).$$

$$= P\left(Z_{900} \geq \frac{5}{3}\right) \approx P\left(Z \geq \frac{5}{3}\right) = P\left(Z \leq -\frac{5}{3}\right) = \Phi\left(-\frac{5}{3}\right)$$

$$= 0.0478.$$

3. Estimation & Confidence Interval.

1). Continuing. $\underbrace{Z_n = \frac{S_n - n\mu}{\sqrt{n}\cdot\sigma}}_{\text{estimation}}$, $P\left(-1.96 \leq \frac{S_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} \leq 1.96\right) \approx 0.95$.

$\Rightarrow P(n\cdot\mu - 1.96\cdot\sqrt{n}\cdot\sigma \leq S_n \leq n\cdot\mu + 1.96\cdot\sqrt{n}\cdot\sigma) \approx 0.95$

A 95% Confidence Interval of $S_n$ is

$[n\cdot\mu - 1.96\sqrt{n}\cdot\sigma,\ n\cdot\mu + 1.96\cdot\sqrt{n}\cdot\sigma]$.

95%: 1.96.

99%: 3.

2). For $\mu$, $P(\frac{1}{n}S_n - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \frac{1}{n}S_n + 1.96\frac{\sigma}{\sqrt{n}}) \approx 0.95$.

Estimation. A 95% C.I. of $\mu$ is $[\frac{1}{n}S_n - 1.96\frac{\sigma}{\sqrt{n}}, \frac{1}{n}S_n + 1.96\cdot\frac{\sigma}{\sqrt{n}}]$.

Also, $P(M_n - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le M_n + 1.96\frac{\sigma}{\sqrt{n}}) \approx 0.95$

e.g. $X_1, \dots, X_{500} \sim \text{Uniform}[a-1, a+1]$. Find 95% C.I. of $a$.

$X_1 + \dots + X_{500} = 29$.

$\mu = E(X_n) = \frac{a-1+a+1}{2} = a$. $\sigma^2 = Var(X_n) = \frac{(R-L)^2}{12} = \frac{4}{12} = \frac{1}{3}$. $\Rightarrow \sigma = \frac{\sqrt{3}}{3}$

$Z = \frac{S_n - n\cdot\mu}{\sqrt{n}\cdot\sigma} = \frac{S_n - n\cdot a}{\sqrt{n}\cdot\frac{1}{3}}$, $M_n = \frac{29}{500} = 0.058$.

$P(-1.96 \le \frac{S_n - n\cdot a}{\sqrt{n}\sigma} \le 1.96) = P(-1.96\cdot\sqrt{n}\cdot\sigma \le S_n - n\cdot a \le 1.96\cdot\sqrt{n}\cdot\sigma)$

$= P(-1.96\sqrt{n}\cdot\sigma - S_n \le -n\cdot a \le 1.96\cdot\sqrt{n}\cdot\sigma - S_n)$.

$= P(\frac{S_n}{n} - \frac{1.96\sigma}{\sqrt{n}} \le a \le \frac{S_n}{n} + \frac{1.96\sigma}{\sqrt{n}})$.

$= P(M_n - \frac{1.96\cdot\frac{\sqrt{3}}{3}}{\sqrt{500}} \le a \le M_n + \frac{1.96\cdot\frac{\sqrt{3}}{3}}{\sqrt{500}})$.

$\therefore a$ has a 95% C.I. at $[M_n - 0.051, M_n + 0.051]$.

$= [0.007, 0.109]$.

(Not saying $P(0.007 \le a \le 0.109) = 0.95$ X).