Mathematical Statistics II

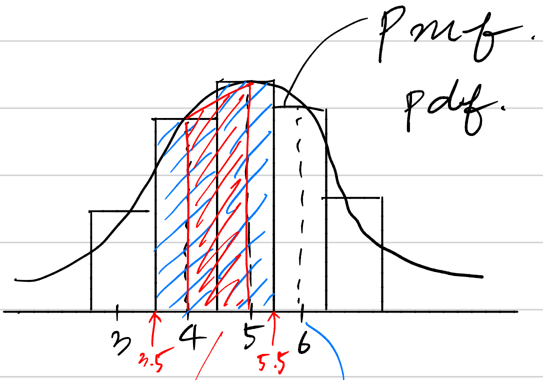
Ch.5 Distributions of Functions of Random Variables

Jungsoon Choi

jungsoonchoi@hanyang.ac.kr

Table of Contents

Ch5.7 Approximations for Discrete Distributions



cf). In discrete case,

$P(4 \le X \le 5) = P(X=4) + P(X=5)$

In continuous case,

$\underbrace{P(4 \le X \le 5)}_{>0} \ne \underbrace{P(X=4)}_{=0} + \underbrace{P(X=5)}_{=0}$.

discrete case $\xrightarrow{\uparrow}$ continuous case

adjust . prob

In discrete adjust case,

$P(4 \le X \le 5) = P(3.5 \le X \le 5.5)$ $\xrightarrow{\text{Apply CLT}}$

* When we apply the CLT in discrete case,

we need to adjust

$P(X=4) = P(3.5 \le X \le 4.5)$ $\xrightarrow{\text{apply CLT}}$

$P(4 < X \le 5) = P(4.5 \le X \le 5.5)$.

$= P(X=5)$.

* $P(X=4)$. $P(\Sigma X_i = 4)$.

CLT 바로쓰면 $P\left(\frac{\Sigma X_i - \mu}{\sigma/\sqrt{n}} = Z = 4\right) = 0$

zero가 돼버림.

Approximation of Binomial distribution

Thm.

Let $X_1, \dots, X_n \sim \mathrm{Ber}(p)$. Then $Y = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n, p)$.

$$W = \frac{Y - np}{\sqrt{np(1-p)}} = \frac{\overline{X} - p}{\sqrt{p(1-p)/n}} \to N(0,1) \quad \text{as } n \to \infty,$$

That means that $Y \sim N(np, np(1-p))$ as $n \to \infty$. ($np \geq 5$ and $n(1-p) \geq 5$)

암기 X.

pop.

Ber(p). $\rightarrow X_1 \cdots X_n$

* 포아송 → Bin 복습.

$$\sum_{i=1}^{n} X_i \sim Bin(n.p).$$

$$\frac{\Sigma X_i - np}{\sqrt{np(1-p)}} \xrightarrow{n\to\infty} N(0,1).$$

$$P(\Sigma X_i > 0).$$

$$Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$$

$$E[Y] = np, \quad Var[Y] = np(1-p)$$

$$\frac{Y - np}{\sqrt{np(1-p)}} \sim N(0,1) \text{ as } n\to\infty \text{ by CLT.}$$

$$\| $$

$$\frac{\Sigma X_i - np}{\sqrt{np(1-p)}} = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}}$$

$$Y = \sum_{i=1}^{n} X_i \sim N(np, np(1-p))$$

Note - correction for continuity

In general, $Y \sim \mathrm{Bin}(n, p)$ where n is quite large.

암기 X

$$P(Y \le k) \approx \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right)$$

$$P(Y < k) \approx \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p)}}\right)$$

Example 5.7-4

$Y \sim \mathrm{Bin}(36, 0.5)$. Compute the approximate probability of $P(Y = 20)$ and $P(12 < Y \le 18)$.

have to adjust

$X_i \sim Ber(0.5)$

$Y = \sum_{i=1}^{n} X_i \sim (np, np(1-p))$.

$(36 \times 0.5, \ 36 \times 0.5^2)$

$P(Y = 20) = P(19.5 \leq Y \leq 20.5)$.

$$= P\left(\frac{19.5 - 18}{\sqrt{36 \times 0.5^2}} \leq \frac{Y - 18}{\sqrt{36 \times 0.5^2}} \leq \frac{20.5}{\sqrt{36 \times 0.5^2}} \right)$$

$$\| \quad Z \sim N(0,1)$$

Apply CLT,

$$\approx P\left(\frac{1.5}{3} \leq Z \leq \frac{2.5}{3} \right)$$

Approximation of Poisson distribution

Example 5.7-5

Let $X_1, \ldots, X_{20} \sim \mathrm{Poi}(1)$ and $Y = \sum_{i=1}^{n} X_i$. Compute the approximate probability of $P(16 < Y \leq 21)$.

$X_1, \cdots, X_{20} \sim Poi(1) \qquad Y = \sum_{i=1}^{n} X_i$

$P(16 < Y \leq 21)$.

$\text{pmf of } Poi(\lambda) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$

$\text{mgf of } Poi(\lambda) = \exp\{\lambda(e^t - 1)\}$

$$M_Y(t) = E[e^{tY}] = E[e^{t\sum X_i}]$$
$$= E[e^{tX_1 + tX_2 + \cdots + tX_n}]$$
$$= \prod_{i=1}^{n} E[e^{tX_i}] = \prod_{i=1}^{n} M_{X_i}(t) = [M_X(t)]^n$$
$$= [\exp\{\lambda(e^t - 1)\}]^n$$
$$= \exp\{n\lambda(e^t - 1)\}$$
$$: \text{mgf of } Poi(n\lambda).$$

$\therefore \; Y \sim Poi(20)$.

$$P(16 < Y \leq 21) = P(17 \leq Y \leq 21) = P(16.5 \leq Y \leq 21.5)$$
$E[Y] = 20, \quad Var[Y] = 20.$

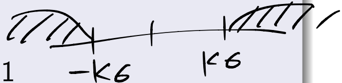
$$P\left(\frac{16.5 - 20}{\sqrt{20}} \leq Z \leq \frac{21.5 - 20}{\sqrt{20}}\right)$$
$$= P\left(\frac{-3.5}{\sqrt{20}} \leq Z \leq \frac{1.5}{\sqrt{20}}\right).$$
$Z \sim N(0,1)$.

Ch5.8 Chebyshev's Inequality

Chebyschev's Inequality

Thm. 5.8-1 (Chebyschev's Inequality)

If the random variable X has a mean μ and variance σ^2, then for every $k \geq 1$, ≥ 0



$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Corollary 5.8-1

If $\epsilon = k\sigma$, then

$(X-\mu)^2 \geq \varepsilon^2$

$k\sigma = \varepsilon$

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

$k = \frac{\varepsilon}{\sigma}$

$\frac{1}{k^2} = \frac{\sigma^2}{\varepsilon^2}$ $k^2 = \frac{\varepsilon^2}{\sigma^2}$

In computing a probability

i) if $dist^n$ is known, we can obtain the exact one.

ii) if " unknown. but n is large,
$P(\bar{X})$ or $P(\Sigma X_i)$ can be approximately calculated (by CLT).

iii) if " unknown & n is small or single r.v is contained in prob.

$P(X > 3.5)$
$X \sim Z(3, 2^2)$.

$X \sim (\mu, \sigma^2)$.

$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

pf). Assume that X is discrete r.v, following $f_X(x)$ pmf.

$\sigma^2 = Var(X) = E[(X-\mu)^2] = \sum_{x \in \mathcal{X}} (x-\mu)^2 f_X(x)$

$= \underbrace{\sum_{x \in A} \underbrace{(x-\mu)^2}_{+} \underbrace{f_X(x)}_{+}}_{\geq 0} + \underbrace{\sum_{x \in A^c} \underbrace{(x-\mu)^2}_{+} \underbrace{f_X(x)}_{+}}_{\geq 0}$

$\geq \sum_{x \in A} (x-\mu)^2 f_X(x) \quad \cdots ①$

where $A = \{x \mid |x-\mu| \geq k\sigma\}$.

Within A,

$(x-\mu)^2 \geq (k\sigma)^2$, for $\forall x \in A$ //

$\Rightarrow (x-\mu)^2 f_X(x) \geq k^2\sigma^2 f_X(x)$ for $\forall x \in A$

$\Rightarrow \sum_{x \in A} (x-\mu)^2 f_X(x) \geq \sum_{x \in A} k^2\sigma^2 f_X(x) = k^2\sigma^2 \underbrace{P(X \in A)}_{\sum_{x \in A} f(x)} \cdots ②$

By ① & ②

$\sigma^2 \geq \sum_{x \in A} (x-\mu)^2 f_X(x) \geq k^2\sigma^2 P(X \in A)$

$\therefore P(X \in A) = P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$ ▨

$P(X \in A) \leq \frac{1}{k^2}$

$\varepsilon = \overset{>0}{k}\overset{>0}{\sigma} \Rightarrow k = \frac{\varepsilon}{\sigma}$

$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$

Example 5.8-1

Let X have a mean of 25 and a variance of 16. Find the lower bound for $P(17 < X < 33)$.

$(25, 4^2)$. $\sigma = 4$.

$P(|X-25| \geq 4 \cdot k) \leq \frac{1}{k^2}$ ...... ①.

$P(17 < X < 33)$

$= P(-8 < X-25 < 8) = P(|X-25| < 8)$

by ① $P(|X-25| \geq 8) \leq \frac{1}{4}$
$P(|X-25| < 8) \geq \frac{3}{4}$

$= P(|X-25| < 2 \cdot \sigma) \geq 1 - \frac{1}{4} = \frac{3}{4} = 0.75$

$P(|X-25| \geq 12)$
$= P(|X-\mu| \geq 3 \cdot \sigma) \leq \frac{1}{3^2} = \frac{1}{9}$

$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

if $\varepsilon = k\sigma$, $P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$

$X \sim (25, 4^2)$

$P(17 < X < 33) = P(17-25 < X-25 < 33-25)$

$= P(|X-25| < 8) = 1 - \underline{P(|X-25| \geq 8)} \leq \frac{1}{4}$

$\therefore P(|X-25| < 8) \geq 1 - \frac{1}{4} = \frac{3}{4}$

[Definition] Convergence in probability

For any $\epsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1$$

Then, X_n converges in probability to X.

$$X_n \xrightarrow{P} X \iff \lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

$\overline{X}_n$: sample mean from $X_i \sim (\mu, \sigma^2)$

$$\overline{X}_n \sim \,?\,(\mu, \tfrac{\sigma^2}{n})$$

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

$$0 \leq \lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\cdot\varepsilon^2} = 0$$

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) = 0$$

$$\iff \lim_{n\to\infty} P(|\overline{X} - \mu| < \varepsilon) = 1.$$