MH3500 Statistics

Tutorial 3 AY2022/23 Semester 2

Problem 3.1

For each $n \in \{10, 100, 1000, 10000\}$, use CLT to approximate the following probabilities and compare the approximations with the exact probabilities.

a) $\Pr(\sum_{i=1}^{n} X_i \leq n)$ where $X_1, \dots, X_n \sim \text{Gamma}(1,1)$
b) $\Pr\left(\sum_{i=1}^{n} X_i \leq \frac{101n}{200}\right)$ where $X_1, \dots, X_n \sim \text{Bernoulli}(0.5)$

Hint: first, we need to identify the distribution of $\sum_{i=1}^{n} X_i$.

CLT $X_1, \dots, X_n$ i.i.d. RV. $E(X_i) = \mu$, $Var(X_i) = \sigma^2$ $\bar{X} = \frac{1}{n}\sum X_i$

$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ converges in distribution to $Z \sim N(0,1)$

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq t\right) - P(Z \leq t) \right| = 0$$

$X_1, \dots, X_n \sim \text{Gamma}(1,1)$ $n \in \{10, 10^2, 10^3, 10^4\}$

Check by MGF $\Rightarrow \sum X_i \sim \text{Gamma}(n,1)$. Let $Y = \sum X_i$

$$P(Y \leq n) = \int_0^n (\text{PDF of } Y)\, dy = \underline{}$$

exact probability.

n	$P(Y \leq n)$
10	0.542
10^2	0.513
10^3	0.504
10^4	0.501

$\downarrow$

0.5

Approximate $P(\sum X_i \le n)$ by CLT

$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ converges in distribution to $Z \sim N(0,1)$

$$P\left(\sum X_i \le n\right) = P\left(\frac{1}{n}\sum X_i \le 1\right) = P(\bar{X} \le 1)$$

$X_i \sim \text{Gamma}(1,1)$

$\mu = E[X_i] = 1$

$\sigma^2 = Var[X_i] = 1$

$$= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le \frac{\sqrt{n}(\overbrace{1-\mu}^{=0})}{\sigma}\right)$$

$$= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le 0\right)$$

$$\approx P(Z \le 0) = 0.5$$

b) $\Pr\left(\sum_{i=1}^{n} X_i \le \frac{101n}{200}\right)$ where $X_1, \dots, X_n \sim \text{Bernoulli}(0.5)$

Bern $(0.5) \sim$ Binomial $(1, 0.5) \sim X_i$

Check by MGF : $Y = \sum X_i \sim \text{Binom}(n, 0.5)$

$P(Y=k) = \binom{n}{k} 0.5^k \, 0.5^{n-k}$ $\leftarrow$ p. mass function.

$n=10$: $P\left(Y \le \frac{101n}{200}\right) = P(Y \le 5.05) = P(Y \le 5)$

$$= P(Y=0) + P(Y=1) + \dots + P(Y=5)$$

$$= \sum_{k=0}^{5} \binom{10}{k} 0.5^k \, 0.5^{10-k} = \underline{\hspace{3em}}$$

$\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ converges in distribution to $Z \sim N(0,1)$

$$P\left(\sum X_i \le \frac{101n}{200}\right) = P\left(\underbrace{\frac{1}{n}\sum X_i}_{=\bar{X}} \le 0.505\right)$$

$X_i \sim Bern(p)$

$$= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le \frac{\sqrt{n}(0.505-\overset{0.5}{\overset{\shortparallel}{\mu}})}{\sigma = 0.5}\right)$$

$\mu = p \quad = 0.5$

$\sigma^2 = p(1-p) = 0.5^2$

$$= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le 0.01\sqrt{n}\right)$$

$$(CLT) \approx P(Z \le 0.01\sqrt{n})$$

Problem 3.2

A battery manufacturer claims that the run time of a new “ultra strong” laptop battery has a population mean of 40 hours with a standard deviation of 5 hours. A random sample $X_1, \dots, X_n$ of size $n = 100$ is taken.

a) Find an approximation to $Pr(\bar{X} \leq 36.7)$. ≈ 0
b) If the claim of the manufacturer were true, would an average run time of 36.7 among 100 batteries be unusually short?
c) If you observed an average run time of 36.7 among 100 batteries, would you find the claim of the manufacturer plausible?
d) Answer parts a) to c) with 36.7 replaced by 39.8. $\longrightarrow$ (a) 0.345.

(a) By CLT.

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \quad \text{conv in dist.} \quad Z \sim N(0,1)$$

$$P(\bar{X} \leq 36.7) = P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq \frac{\sqrt{n}(36.7-\mu)}{\sigma}\right)$$

$n = 100$
$\mu = 40$
$\sigma = 5$

$$= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq -6.6\right)$$

$$\approx P(Z \leq -6.6) \approx 2 \times 10^{-11}$$

$$\parallel$$

$$\int_{-\infty}^{-6.6} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

Problem 3.3

Let $X_1, \dots, X_n$ be an i.i.d. random sample drawn from $U(a,\ b)$ (uniform distribution on the interval $(a,\ b)$), where a and b are unknown parameters.

a) Find the method of moments estimators for a and b.
b) The following observations for $x_1, \dots, x_{30}$ for $X_1, \dots, X_{30}$ are given with $n = 30$. What are the estimates for a and b that you get from the method of moments estimators for these data?

15.4, 71.4, 13.8, 16.4, 56.9,	18.7, 83.6, 83.6, 75.0, 23.5,
69.4, 56.7, 97.6, 68.4, 82.0,	50.8, 48.9, 84.3, 17.5, 22.0,
25.6, 45.2, 84.8, 82.3, 15.4,	45.9, 57.4, 47.5, 69.1, 66.5

[In fact, these observations are drawn from $U(10,\ \ 100)$]

Weak Law of Large Numbers.

$S_1 = \bar{X} = \frac{1}{n}\sum X_i$ converges to $\mu_1 = E(X)$ in probability.

$\forall \varepsilon > 0$, $\lim_{n\to\infty} P(|\bar{X} - \mu| > \varepsilon) = 0$

$S_m = \frac{1}{n}\sum_{i=1}^{n} X_i^m$ conv in probability to $\mu_m = E(X^m)$

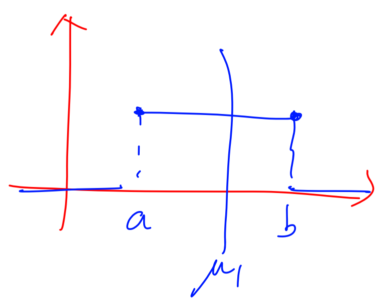
$X_1 \sim \text{Unif}(a, b)$

$\mu_1 = E(X_1) = \frac{a+b}{2}$ — (1)

pdf of

$\mu_2 = E(X_1^2) = \int_a^b x^2 \left(\frac{1}{b-a}\right) dx$

$= \frac{a^2 + ab + b^2}{3}$ — (2)



(1) : $b = 2\mu_1 - a$ Substitute into (2).

$$\mu_2 = \frac{a^2 + a(2\mu_1 - a) + (2\mu_1 - a)^2}{3}$$

$$a^2 - 2\mu_1 a + 4\mu_1^2 - 3\mu_2 = 0$$

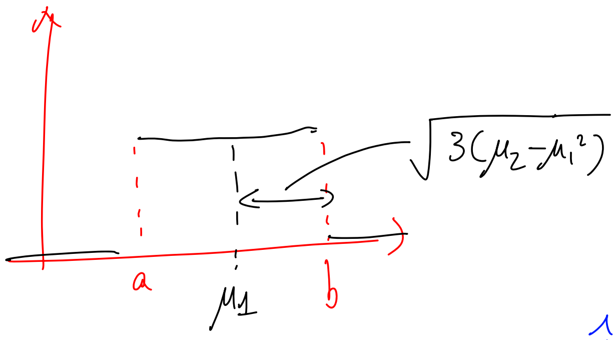
$$a = \mu_1 \pm \sqrt{\mu_1^2 - (4\mu_1^2 - 3\mu_2)} = \mu_1 \pm \sqrt{3(\mu_2 - \mu_1^2)}$$

$$b = 2\mu_1 - a = 2\mu_1 - \left(\mu_1 \pm \sqrt{3(\mu_2 - \mu_1^2)}\right)$$

$$= \mu_1 \mp \sqrt{3(\mu_2 - \mu_1^2)}$$

$X_i \sim \text{Unif}(a, b)$ $\quad a < b \Rightarrow$ $a = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}$

$$b = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)}$$



Weak LLN:

$S_1 \longrightarrow \mu_1$ $\qquad S_1 = \frac{1}{n}\sum X_i$

$S_2 \longrightarrow \mu_2$ $\qquad S_2 = \frac{1}{n}\sum X_i^2$

$\hat{a} = S_1 - \sqrt{3(S_2 - S_1^2)}$ $\qquad \hat{b} = S_1 + \sqrt{3(S_2 - S_1^2)}$

b) The following observations for $x_1, \ldots, x_{30}$ for $X_1, \ldots, X_{30}$ are given with $n = 30$. What are the estimates for a and b that you get from the method of moments estimators for these data?

15.4, 71.4, 13.8, 16.4, 56.9, $\quad$ 18.7, 83.6, 83.6, 75.0, 23.5,
69.4, 56.7, 97.6, 68.4, 82.0, $\quad$ 50.8, 48.9, 84.3, 17.5, 22.0,
25.6, 45.2, 84.8, 82.3, 15.4, $\quad$ 45.9, 57.4, 47.5, 69.1, 66.5

[In fact, these observations are drawn from $U(10, \ 100)$]

→ Compute S_1, S_2

→ find $\hat{a}, \hat{b}$

$\parallel \quad \parallel$

$a \quad b$

$\hat{a} = 8.18$

$\hat{b} = 98.19$

MH3500 Statistics

Tutorial 3 AY2022/23 Semester 2

Problem 3.1

For each $n \in \{10, 100, 1000, 10000\}$, use CLT to approximate the following probabilities and compare the approximations with the exact probabilities.

a) $\Pr(\sum_{i=1}^{n} X_i \leq n)$ where $X_1, \dots, X_n \sim \text{Gamma}(1,1)$
b) $\Pr\left(\sum_{i=1}^{n} X_i \leq \frac{101n}{200}\right)$ where $X_1, \dots, X_n \sim \text{Bernoulli}(0.5)$

Hint: first, we need to identify the distribution of $\sum_{i=1}^{n} X_i$.

Problem 3.2

A battery manufacturer claims that the run time of a new "ultra strong" laptop battery has a population mean of 40 hours with a standard deviation of 5 hours. A random sample $X_1, \dots, X_n$ of size $n = 100$ is taken.

a) Find an approximation to $Pr(\bar{X} \leq 36.7)$.
b) If the claim of the manufacturer were true, would an average run time of 36.7 among 100 batteries be unusually short?
c) If you observed an average run time of 36.7 among 100 batteries, would you find the claim of the manufacturer plausible?
d) Answer parts a) to c) with 36.7 replaced by 39.8.

Problem 3.3

Let $X_1, \dots, X_n$ be an i.i.d. random sample drawn from $U(a,\ b)$ (uniform distribution on the interval $(a,\ b)$), where a and b are unknown parameters.

a) Find the method of moments estimators for a and b.
b) The following observations for $x_1, \dots, x_{30}$ for $X_1, \dots, X_{30}$ are given with $n = 30$. What are the estimates for a and b that you get from the method of moments estimators for these data?

$$\begin{array}{ll} 15.4, 71.4, 13.8, 16.4, 56.9, & 18.7, 83.6, 83.6, 75.0, 23.5, \\ 69.4, 56.7, 97.6, 68.4, 82.0, & 50.8, 48.9, 84.3, 17.5, 22.0, \\ 25.6, 45.2, 84.8, 82.3, 15.4, & 45.9, 57.4, 47.5, 69.1, 66.5 \end{array}$$

[In fact, these observations are drawn from $U(10,\ \ 100)$]

$\mu = 95$
$n = 50$

$\bar{X} = 92.25$, $s = 10$

$H_0 : \mu = 95$

$H_1 : \mu < 95$

By Central Limit Theorem, $\dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$$Z = \frac{92.25 - 95}{\frac{10}{\sqrt{50}}} = -1.7678$$

Critical value, $Z_c \Rightarrow P(Z \leq Z_c) = 0.05 \Rightarrow Z_c = -1.645$

Since $|Z| > |Z_c| \Rightarrow H_0$ is rejected $\Rightarrow$ Claim by statement ($\mu = 95$) is invalid

X = Years

Y = Premium

$\Sigma x = 90$ $\qquad n = 8$

$\Sigma x^2 = 1396$

$\Sigma y = 474$

$\Sigma xy = 4739$

$Y = a + bX$

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{235194}{3068} = 76.66$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{-4748}{3068} = -1.5476.$$

Correlation coefficient, $r = \dfrac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}$

$$= \frac{8(4739) - (90)(474)}{\sqrt{8 \cdot 1396 - 90^2} \cdot \sqrt{8 \cdot 28642 - 474^2}}$$

$$= \frac{-4748}{6182.82} = -0.767$$