RA Math Test: Exercise 1 & Exercise 2

YILONG LIU

May 2025

1 Exercise 1, Question 1: An unbiased estimator of the variance of i.i.d random variables

Let $(Y_i)_{1\leq i\leq n}$ be n i.i.d random variables. Let $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ denote the average of these variables. Let Y be a random variable with the same distribution as the Y_is. The goal of the exercise is to show that $\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ is an unbiased estimator of $V(Y)$, the variance of the Y_is.

(1) Show that $\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{1}{n-1}\sum_{i=1}^{n} Y_i^2 - \frac{n}{n-1}(\bar{Y})^2$.

(2) Use the result in question 1) to prove that $E\left(\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right) = V(Y)$.

1.1 Question (1) Solution

We can expand the left hand side equation and have:

$$
\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2) \\
&= \frac{1}{n-1}(\sum_{i=1}^{n} Y_i^2) - \frac{2}{n-1}\bar{Y}(\sum_{i=1}^{n} Y_i) + \frac{n}{n-1}\bar{Y}^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n} Y_i^2 - \frac{2}{n-1}\bar{Y}n\bar{Y} + \frac{n}{n-1}\bar{Y}^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n} Y_i^2 - \frac{n}{n-1}\bar{Y}^2
\end{aligned}
$$

1.2 Quesiton (2) Solution

According to the result in question (1), we have the following:

$$\begin{aligned} E[\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2] &= E[\frac{1}{n-1}\sum_{i=1}^{n}Y_i^2-\frac{n}{n-1}(\bar{Y})^2] \\ &= \frac{1}{n-1}E[\sum_{i=1}^{n}Y_i^2]-\frac{n}{n-1}E[\bar{Y}^2] \\ &= E[\frac{1}{n-1}\sum_{i=1}^{n}Y_i^2]-E[\frac{n}{n-1}\bar{Y}^2] \\ &= E[Y^2]-E^2[Y] \end{aligned}$$

Note: As $\frac{1}{n-1}\sum_{i=1}^{n}Y_i^2$ *is the unbiased estimator of* $E[Y^2]$.

2 Exercise 1, Question 2: A super consistent estimator

Assume you observe an iid sample of n random variables $(Y_i)_{1\leq i\leq n}$ following the uniform distribution on $[0,\theta]$, where θ is an unknown strictly positive real number we would like to estimate. Let Y be a random variable with the same distribution as the Y_is.

(1) Compute $E(Y)$. Write θ as a function of $E(Y)$.

(2) Use question a) to propose an estimator $\hat{\theta}_{MM}$ for θ using the method of moments (reminder: that method amounts to replacing expectations by sample means).

(3) Show that $\hat{\theta}_{MM}$ is an asymptotically normal estimator of θ, and show that its asymptotic variance is $4V(Y)$.

Consider the following alternative estimator for θ: $\hat{\theta}_{ML}=\max_{1\leq i\leq n}\{Y_i\}$.

(4) Why does using $\hat{\theta}_{ML}$ to estimate θ sounds like a natural idea?

(5) Show that

$$P(\hat{\theta}_{ML}\leq x)=\begin{cases} 0 & \text{if } x<0 \\ \left(\frac{x}{\theta}\right)^n & \text{if } x\in[0,\theta] \\ 1 & \text{if } x>\theta \end{cases}$$

(6) Use the result in question (5) to show that $n\left(\frac{\theta-\hat{\theta}_{ML}}{\theta}\right)\xrightarrow{d}U$, where U follows an exponential distribution with parameter 1. *Hint:* to prove this, you need to use the definition of convergence in distribution in your lecture

notes. Also, use the fact that the cdf of an exponential distribution with parameter 1 is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - \exp(-x) & \text{if } x \geq 0 \end{cases}$$

(7) Which estimator is the best: $\hat{\theta}_{MM}$, or $\hat{\theta}_{ML}$?

(8) Illustrate this through a Monte-Carlo study. Draw 1000 iid realizations of variables following a uniform distribution on $[0, 1]$ in Stata (you need to use the "uniform()" command), compute $\hat{\theta}_{MM}$ and $\hat{\theta}_{ML}$. What is the value of θ in this example? Which estimator is the closest to θ?

(9) For any $q \in (0, 1)$, let t_q denote the q^{th} quantile of the exp(1) distribution: $t_q = F^{-1}(q)$. Show that

$$IC(\alpha) = \left[\hat{\theta}_{ML}, \hat{\theta}_{ML} + \hat{\theta}_{ML}\frac{t_{1-\alpha}}{n}\right]$$

is a confidence interval for θ with asymptotic coverage $1-\alpha$. You should use the result from the previous question and the Slutsky lemma. You can use without proving it the fact that $\hat{\theta}_{ML} \xrightarrow{P} \theta$ (actually, that directly follows from the fact $\hat{\theta}_{ML}$ is an n-consistent estimator of θ).

2.1 Question (1) Solution

By definition:

$$E[Y] = \int_0^\theta y \cdot f(y)dy$$

As it is uniform distribution: $f(y) = \frac{1}{\theta}$ Then we have the follows:

$$\begin{aligned} E[Y] &= \int_0^\theta \frac{y}{\theta}dy \\ &= \frac{1}{2\theta} \cdot y^2|_0^\theta \\ &= \frac{\theta}{2} \end{aligned}$$

Therefore, $E[Y] = \frac{\theta}{2}$, $\theta = 2E[Y]$.

2.2 Question (2) Solution

From (1), $\theta = 2E[Y]$, we propose that the estimator $\hat{\theta}_{MM}$ for θ to be:

$$\hat{\theta}_{MM} = \frac{2}{n}\sum_{i=1}^{n} Y_i$$

based on n random variables $(Y_i)_{1 \leq i \leq n}$.

2.3 Question (3) Solution

From (1), $\theta = 2E[Y]$, we propose that the estimator $\hat{\theta}_{MM}$ for θ to be:

$$\hat{\theta}_{MM} = \frac{2}{n}\sum_{i=1}^{n} Y_i$$

based on n random variables $(Y_i)_{1\leq i\leq n}$.
By the Central Limit Theorem, we have:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - E[Y]\right) \xrightarrow{d} \mathcal{N}(0, V(Y))$$

which implies that

$$\sqrt{n}(\hat{\theta}_{MM} - \theta) = \sqrt{n}\left(2\bar{Y} - 2E[Y]\right) = 2\sqrt{n}(\bar{Y} - E[Y]) \xrightarrow{d} \mathcal{N}(0, 4V(Y)).$$

Therefore, $\hat{\theta}_{MM}$ is an asymptotically normal estimator of θ, and its asymptotic variance is $4V(Y)$.

2.4 Question (4) Solution

It is because that the random variables $(Y_i)_{1\leq i\leq n}$ follows a uniform distribution on $[0, \theta]$. Therefore for any i such that $1 \leq i \leq n$, we have $Y_i \in [0, \theta]$, and it is then a natural idea to assume the key parameter $\hat{\theta}_{ML} = max_{1\leq i\leq n} Y_i$.

2.5 Question (5) Solution

By definition, $P(\hat{\theta}_{ML} \leq x) = P(max_{1\leq i\leq n}\{Y_i\} \leq x)$, then:
if $x < 0$, then the probility is 0 as $Y_i \in [0, \theta]$ for any i.
if $x \in [0, \theta]$, then:

$$\begin{aligned} P(max_{1\leq i\leq n}\{Y_i\} \leq x) &= \Pi_{i=1}^{n} P(Y_i \leq x) \\ &= \Pi_{i=1}^{n} F(X) \\ &= (\frac{x-0}{\theta-0})^n \\ &= (\frac{x}{\theta})^n \end{aligned}$$

if $x > \theta$, then the probability is 1 as $Y_i \in [0, \theta]$ for any i.

2.6 Question (6) Solution

According to the description, we have the following:

$$n\left(\frac{\theta - \hat{\theta}_{ML}}{\theta}\right) = n\left(1 - \frac{\hat{\theta}_{ML}}{\theta}\right)$$

Then, we have:

$$\begin{aligned}
F\left(n\left(1-\frac{\hat{\theta}}{\theta}\right)<x\right) &= F\left(\frac{\hat{\theta}}{\theta}>1-\frac{x}{n}\right) \\
&= 1-F\left(\frac{\hat{\theta}}{\theta}<1-\frac{x}{n}\right) \\
&= 1-\left(1-\frac{x}{n}\right)^{n}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lim_{n\to\infty} F(x) &= \lim_{n\to\infty}\left[1-\left(1-\frac{x}{n}\right)^{n}\right] \\
&= 1-\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^{n} \\
&= 1-\lim_{n\to\infty}\exp\left[n\ln\left(1-\frac{x}{n}\right)\right] \\
&= 1-\lim_{h\to 0}\exp\left[\frac{-x\ln(1+h)}{h}\right] \\
&= 1-\exp(-x)
\end{aligned}$$

Note: $h=\frac{-x}{n}$

2.7 Question (7) and (8) Solution

To evaluate which estimator is better, we consider their performance across 1000 simulations:

Estimator	Mean Estimate	Standard Deviation	Min – Max
$\hat{\theta}_{MM}$	1.000846	0.0177101	0.9523674 – 1.052426
$\hat{\theta}_{ML}$	0.998992	0.0009990	0.9935493 – 0.9999998

Although both estimators are consistent, their statistical properties differ:
- The method of moments estimator $\hat{\theta}_{MM}$ is unbiased and its average estimate is closer to the true value $\theta = 1$.
- The maximum likelihood estimator $\hat{\theta}_{ML}$ is biased (slightly underestimates θ on average), but it has a much smaller variance.

Therefore, in terms of bias, $\hat{\theta}_{MM}$ performs better. In terms of efficiency (i.e., lower variance), $\hat{\theta}_{ML}$ is superior.
The choice of the "best" estimator depends on the criterion:
- If minimizing mean squared error (bias^2 + variance), $\hat{\theta}_{ML}$ is preferred because its variance is much smaller.
- If unbiasedness is prioritized, then $\hat{\theta}_{MM}$ is preferred.

Overall, for large n, the bias of $\hat{\theta}_{ML}$ becomes negligible, and its lower variance makes it the better estimator in most practical contexts.

We draw a sample of size $n = 1000$ from the uniform distribution on $[0, 1]$ using Stata's 'uniform()' command.
The true value of the parameter is:

$$\theta = 1.$$

From the Stata output of one simulation:

$$\hat{\theta}_{MM} = 0.98998555, \quad \hat{\theta}_{ML} = 0.99975073.$$

In this sample, $\hat{\theta}_{ML}$ is closer to the true value $\theta = 1$ than $\hat{\theta}_{MM}$. This aligns with the simulation results in (6), which show that although $\hat{\theta}_{MM}$ is on average closer to θ, its individual estimates fluctuate more due to higher variance. $\hat{\theta}_{ML}$, while slightly biased, tends to provide estimates with much lower dispersion.

2.8 Question (9) Solution

For any $q \in (0, 1)$, let t_q denote the q^{th} quantile of the exp(1) distribution, i.e. $t_q = F^{-1}(q)$. From question (6), we have shown that

$$n\left(\frac{\theta - \hat{\theta}_{ML}}{\theta}\right) \xrightarrow{d} \text{Exp}(1),$$

which implies that

$$P\left(n\left(\frac{\theta - \hat{\theta}_{ML}}{\theta}\right) \leq t_{1-\alpha}\right) \to 1 - \alpha.$$

This is equivalent to:

$$P\left(\theta \leq \hat{\theta}_{ML} + \hat{\theta}_{ML} \cdot \frac{t_{1-\alpha}}{n}\right) \to 1 - \alpha.$$

Hence, the asymptotic $(1 - \alpha)$ confidence interval for θ is given by:

$$IC(\alpha) = \left[\hat{\theta}_{ML},\ \hat{\theta}_{ML} + \hat{\theta}_{ML} \cdot \frac{t_{1-\alpha}}{n}\right].$$

This result follows by applying the convergence in distribution from (7) and the Slutsky lemma. Note that $\hat{\theta}_{ML} \xrightarrow{p} \theta$ since $\hat{\theta}_{ML}$ is an n-consistent estimator.

3 Exercise 1, Question 3

Assume you observe two sequences of random variables $(U_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$. Assume that $U_n \xrightarrow{P} l$ and $V_n \xrightarrow{P} l'$, where l and l' are two real numbers.

1. Use the *continuous mapping theorem* to prove that $U_n \times V_n \xrightarrow{P} l \times l'$.

2. Use the *Slutsky lemma* to prove that $U_n \times V_n \xrightarrow{P} l \times l'$. You need to use the two following facts:

   1. If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$ (convergence in probability implies convergence in distribution)

   2. If $X_n \xrightarrow{d} x$ and x is a real number, then $X_n \xrightarrow{P} x$ (convergence in distribution **towards a real number** implies convergence in probability towards that real number)

3.1 Question (1) Solution

To prove that $U_n \times V_n \xrightarrow{P} l \times l'$ using the *continuous mapping theorem*, define the continuous function $g(x, y) = xy$ on $\mathbb{R}^2$. Since $U_n \xrightarrow{P} l$ and $V_n \xrightarrow{P} l'$, we have:

$$(U_n, V_n) \xrightarrow{P} (l, l').$$

Then, by the continuous mapping theorem:

$$g(U_n, V_n) = U_n \cdot V_n \xrightarrow{P} g(l, l') = l \cdot l'.$$

3.2 Question (2) Solution

To prove the same result using the *Slutsky lemma*, we proceed in steps:

- Since $U_n \xrightarrow{P} l$ and $V_n \xrightarrow{P} l'$, from Fact 1, we know that:

$$U_n \xrightarrow{d} l, \quad V_n \xrightarrow{d} l'.$$

- Then, using the known result that if $X_n \xrightarrow{d} x$ and $Y_n \xrightarrow{d} y$ with $x, y \in \mathbb{R}$, then $X_n Y_n \xrightarrow{d} xy$, we get:

$$U_n V_n \xrightarrow{d} ll'.$$

- Now, since ll' is a real number, we apply Fact 2:

$$U_n V_n \xrightarrow{P} ll'.$$

Asymptotic variance of the estimator of the variance of iid random variables. Let $(Y_i)_{1\leq i\leq n}$ be an iid sample of n random variables with a 4th moment and with a strictly positive variance. Let Y denote a random variable with the same distribution as the Y_is. Let $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$, and let $\bar{Y}^2 = \frac{1}{n}\sum_{i=1}^n Y_i^2$. Let $\hat{V}(Y) = \bar{Y}^2 - (\bar{Y})^2$ be an estimator of $V(Y)$.

4 Exercise 2 Question 1

Show that

$$\sqrt{n}\left(\begin{pmatrix}\overline{Y^2}\\ \overline{Y}\end{pmatrix} - \begin{pmatrix}E(Y^2)\\ E(Y)\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, V_0),$$

where

$$V_0 = \begin{pmatrix} V(Y^2) & \mathrm{cov}(Y^2, Y)\\ \mathrm{cov}(Y^2, Y) & V(Y)\end{pmatrix}.$$

4.1 Question 1 Solution

This result follows directly from the multivariate Central Limit Theorem (CLT), which states that for a vector of sample means of i.i.d. variables with finite second moments, the scaled difference between the sample means and their expectations converges in distribution to a multivariate normal distribution.

Let $X_i = (Y_i^2, Y_i)'$ be a 2-dimensional random vector. Then by the multivariate CLT,

$$\sqrt{n}\left(\begin{pmatrix}\overline{Y^2}\\ \overline{Y}\end{pmatrix} - \begin{pmatrix}E(Y^2)\\ E(Y)\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, V_0),$$

where V_0 is the covariance matrix of X_i.

5 Exercise 2, Question 2

Use the previous question and a well-known theorem to prove that

$$\sqrt{n}\left(\widehat{V}(Y) - V(Y)\right) \xrightarrow{d} \mathcal{N}(0, V_1),$$

where

$$V_1 = (1, -2E(Y))V_0(1, -2E(Y))'.$$

5.1 Question 2 Solution

We know that the sample variance is defined as:

$$\widehat{V}(Y) = \overline{Y^2} - \overline{Y}^2 \quad \text{and} \quad V(Y) = E(Y^2) - (E(Y))^2$$

Define the function $g(a,b) = a - b^2$. This is a differentiable function from $\mathbb{R}^2 \to \mathbb{R}$. By the delta method, since

$$\sqrt{n}\begin{pmatrix}\overline{Y^2} - E(Y^2) \\ \overline{Y} - E(Y)\end{pmatrix} \xrightarrow{d} \mathcal{N}(0, V_0),$$

we apply the delta method with gradient $\nabla g(a,b) = (1, -2b)$. Evaluated at $(a,b) = (E(Y^2), E(Y))$, this becomes $(1, -2E(Y))$.
Hence,

$$\sqrt{n}\left(\widehat{V}(Y) - V(Y)\right) \xrightarrow{d} \mathcal{N}\left(0, (1, -2E(Y))V_0(1, -2E(Y))'\right),$$

which proves the result.

6 Exercise 2, Question 3

In this question, we focus on the special case where the Y_is are binary random variables. To alleviate the notation, let $p = E(Y)$. We have $V(Y) = p(1-p)$. Moreover, in this special case $\widehat{V}(Y) = \overline{Y}(1-\overline{Y})$.

(a) Use the results from questions 1 and 2 to prove that $V_1 = p(1-p)(1-2p)^2$.

(b) For $p = \frac{1}{2}$, $p(1-p) = \frac{1}{4}$, and $p(1-p)(1-2p)^2 = 0$. Therefore, if $p = \frac{1}{2}$, what is the asymptotic distribution of $\sqrt{n}\left(\overline{Y}(1-\overline{Y}) - \frac{1}{4}\right)$?

(c) Show that $n\left(\overline{Y}(1-\overline{Y}) - \frac{1}{4}\right) = \left(\sqrt{n}\left(\overline{Y} - \frac{1}{2}\right)\right)^2$, and use this equality to derive the asymptotic distribution of $n\left(\overline{Y}(1-\overline{Y}) - \frac{1}{4}\right)$.

(d) The previous questions show that the asymptotic distribution of $\sqrt{n}\left(\widehat{V}(Y) - V(Y)\right)$ depends on the value of p, which is unknown. Then, how could you build a confidence interval for $V(Y)$?

6.1 Question (a) Solution

From question 2, we know that the asymptotic variance is given by:

$$V_1 = (1, -2E(Y))V_0(1, -2E(Y))'$$

In the binary case, $Y_i \in \{0, 1\}$:

$$E(Y) = p, \quad E(Y^2) = E(Y) = p, \quad \Rightarrow V(Y) = p(1-p)$$

$$V(Y^2) = p(1-p), \quad \text{cov}(Y^2, Y) = \text{cov}(Y, Y) = V(Y) = p(1-p)$$

So the matrix V_0 becomes:

$$V_0 = \begin{pmatrix} p(1-p) & p(1-p) \\ p(1-p) & p(1-p) \end{pmatrix}$$

Now compute:

$$V_1 = \begin{pmatrix} 1 & -2p \end{pmatrix} \begin{pmatrix} p(1-p) & p(1-p) \\ p(1-p) & p(1-p) \end{pmatrix} \begin{pmatrix} 1 \\ -2p \end{pmatrix}$$

Performing the matrix multiplication:

$$(1, -2p)\left(p(1-p)(1-2p) + (-2p)p(1-p)\right) = p(1-p)(1-2p)^2$$

6.2 Question (b) Solution

When $p = \frac{1}{2}$:

$$p(1-p) = \frac{1}{4}, \quad (1-2p)^2 = 0$$

From part (a), $V_1 = p(1-p)(1-2p)^2 = 0$, so the asymptotic variance is zero. Therefore:

$$\sqrt{n}\left(\overline{Y}(1-\overline{Y}) - \frac{1}{4}\right) \xrightarrow{P} 0$$

That is, the difference converges to zero faster than $1/\sqrt{n}$ — the asymptotic distribution is degenerate at 0.

6.3 Question (c) Solution

We expand:

$$\overline{Y}(1-\overline{Y}) = \overline{Y} - \overline{Y}^2 = \frac{1}{4} + \left(\overline{Y} - \frac{1}{2}\right) - \left(\overline{Y} - \frac{1}{2}\right)^2$$

This simplifies to:

$$\overline{Y}(1-\overline{Y}) - \frac{1}{4} = -\left(\overline{Y} - \frac{1}{2}\right)^2 \Rightarrow n\left(\overline{Y}(1-\overline{Y}) - \frac{1}{4}\right) = -n\left(\overline{Y} - \frac{1}{2}\right)^2 = -\left(\sqrt{n}\left(\overline{Y} - \frac{1}{2}\right)\right)^2$$

From the Central Limit Theorem:

$$\sqrt{n}\left(\overline{Y} - \frac{1}{2}\right) \xrightarrow{d} \mathcal{N}(0, \tfrac{1}{4}) \Rightarrow \left(\sqrt{n}\left(\overline{Y} - \frac{1}{2}\right)\right)^2 \xrightarrow{d} \frac{1}{4}\chi_1^2$$

Thus:

$$n\left(\overline{Y}(1-\overline{Y}) - \frac{1}{4}\right) \xrightarrow{d} -\frac{1}{4}\chi_1^2$$

6.4 Question (d) Solution

The asymptotic distribution of $\sqrt{n}(\widehat{V}(Y) - V(Y))$ depends on p, which is unknown. Since

$$V_1 = p(1-p)(1-2p)^2 = V(Y)(1-2p)^2,$$

we can estimate p by $\overline{Y}$ and substitute it into the asymptotic variance expression. Therefore, an approximate $(1-\alpha)$ confidence interval for $V(Y)$ is:

$$\left[\widehat{V}(Y) \pm z_{1-\alpha/2} \cdot \frac{\sqrt{\widehat{V}(Y)}|1-2\overline{Y}|}{\sqrt{n}}\right]$$

where $z_{1-\alpha/2}$ is the standard normal quantile.