VE401 Probabilistic Methods in Eng.
Solution Manual for RC 5

Chen Xiwen

April 5, 2020

University of Michigan–Shanghai Jiao Tong University Joint Institute

Assignment 3.4

A mathematics textbook has 200 pages on which typographical errors in the equations could occur. Suppose there are in fact five errors randomly dispersed among these 200 pages.

1. What is the probability that a random sample of 50 pages will contain at least one error?
2. How large must the random sample be to assure that at least three errors will be found with 90% probability? (You may use a normal approximation to the binomial distribution.)

Solution.

1. The problem is to randomly place the five errors in 200 pages, and each error has the same probability of being placed among the sampled pages.

$$\begin{aligned} P[\text{at least 1 error in 50 pages}] &= 1 - P[\text{0 error in 50 pages}] \\ &= 1 - \left(\frac{200-50}{200}\right)^5 \\ &= 76.27\%. \end{aligned}$$

2. Let the sample size be k. The number of selected errors follows a binomial distribution with

$$p = \frac{k}{200}, \qquad n = 5,$$

and thus the mean and standard deviation are given by

$$\mu = 5p = \frac{k}{40}, \qquad \sigma = \sqrt{5p(1-p)} = \sqrt{\frac{k}{40}\left(1 - \frac{k}{200}\right)}.$$

Let X be the number of errors in the sample. Then

$$P[X \geq 3] \geq 90\% \quad \Rightarrow \quad P[Y \geq 2.5] \geq 90\%,$$

where Y follows normal distribution. Transforming to standard normal variable Z, we have

$$P\left[Z \geq \frac{2.5-\mu}{\sigma}\right] \geq 0.9 \quad \Rightarrow \quad F\left[\frac{2.5-\mu}{\sigma}\right] \leq 0.1 \quad \Rightarrow \quad \frac{2.5-\mu}{\sigma} \leq -1.28,$$

which gives $k \geq 150$.
Note. Some of you may have noticed that the requirements for "good approximation" specified in lecture slides are not satisfied. However, if we calculate using $p = 0.75$ and $n = 5$ for binomial distribution,

$$P[X \geq 3] = 1 - \texttt{CDF}[\texttt{BinomialDistribution}[5, 0.75], 2] = 0.896484,$$

which is quite close to 90%. This posterior validation shows the approximation is reasonable.

About half-unit correction. If X follows a binomial distribution with parameters n and p, and Y follows a normal distribution with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. Recall normal approximation to binomial distribution,

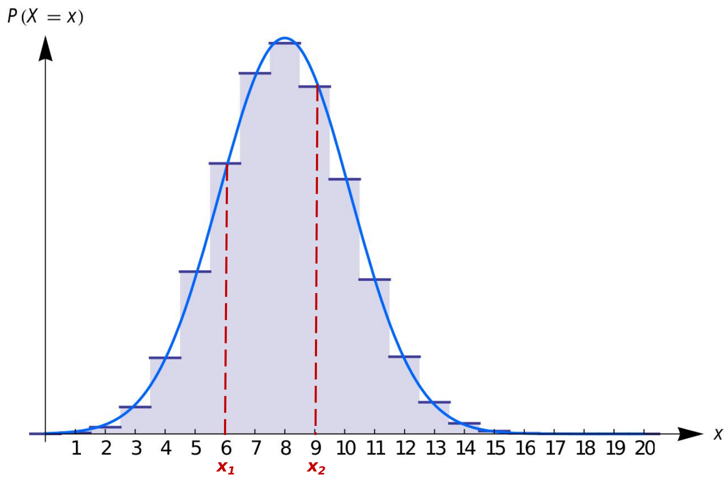
$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi\left(\frac{y + 1/2 - np}{\sqrt{np(1-p)}}\right),$$

where Φ is the cumulative distribution function for the standard normal distribution. Define

$$F(x) := \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right).$$

Then the approximation is simply $P[X \le y] \approx F(y + 1/2)$. We would like to approximate the cumulative distribution of X using the cumulative distribution of Y (denoted as Φ). Consider the following cases. What should be the corresponding approximation?

1. $P[X \ge x_1]$:
   (a) $1 - F(x_1 + 0.5)$
   (b) $1 - F(x_1 - 0.5)$
2. $P[X > x_1]$:
   (a) $1 - F(x_1 + 0.5)$
   (b) $1 - F(x_1 - 0.5)$
3. $P[X \le x_2]$:
   (a) $F(x_2 + 0.5)$
   (b) $F(x_2 - 0.5)$
4. $P[X < x_2]$:
   (a) $F(x_2 + 0.5)$
   (b) $F(x_2 - 0.5)$



Answer: 1. (b), 2. (a), 3. (a), 4. (b).

There are two ways that we can memorize it.

- We are approximating the integral of binomial density function. According to the figure, if we want $P[X \le x]$, then we are calculating the area summing up to the right of x. Similarly, if we want $P[X < x]$, we are summing up to the left of x.

- The original binomial approximation gives

$$P[X \le x] \approx F(x + 1/2),$$

which means that

$$P[X < x] = P[X \le x - 1] \approx F(x - 1 + 1/2) = F(x - 1/2).$$

With this in mind, whenever we want to calculate $P[X \ge x]$ or $P[X > x]$, we rewrite as

$$P[X \ge x] = 1 - P[X < x], \qquad P[X > x] = 1 - P[X \le x],$$

which falls in to the cases discussed above.

Assignment 3.10

Let $X = (X_1, X_2)$ be a random vector. Then we define the expectation vector and the variance-covariance matrix as follows.

$$\mathrm{E}[X] := \begin{pmatrix} \mathrm{E}[X_1] \\ \mathrm{E}[X_2] \end{pmatrix}, \qquad \mathrm{Var}\ X := \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}\ X_2 \end{pmatrix}.$$

Let A be a constant 2×2 matrix and $Y = (Y_1, Y_2) = AX$.

1. Show that $\mathrm{E}[AX] = A\mathrm{E}[X]$.

2. Show that $\mathrm{Var}(AX) = A(\mathrm{Var}\ X)A^T$.

3. Suppose X_1 and X_2 follow independent normal distributions with mean μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Show that the joint density is given by

$$f_X(x) = f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{\det \Sigma_X}} e^{-\frac{1}{2}\langle x - \mu_X, \Sigma_X^{-1}(x - \mu_X)\rangle}$$

   where $\mu_X = (\mu_1, \mu_2)$ and $\Sigma_X = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$ is the 2×2 matrix with the variances on the diagonal and all other entries vanishing.

4. Suppose that X_1 and X_2 follow independent normal distributions with means $\mu_1, \mu_2 \in \mathbb{R}$ and variances $\sigma_1^2, \sigma_2^2 > 0$, respectively. Let $Y = AX$ where A is an invertible $n \times n$ matrix. Show that

$$f_Y(y) = \frac{1}{2\pi\sqrt{|\det \Sigma_Y|}} e^{-\frac{1}{2}\langle y - \mu_Y, \Sigma_Y^{-1}(y - \mu_Y)\rangle} \tag{1}$$

   where $\mu_Y = \mathrm{E}[Y]$, $\Sigma_Y = \mathrm{Var}\ Y$ and $\langle \cdot, \cdot \rangle$ denotes the euclidean scalar product in $\mathbb{R}^2$.

5. Show that Eq. (1) can be written as

$$f_Y(y_1,y_2)=\frac{1}{2\pi\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)^2-2\rho\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)\left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right)+\left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right]}$$

where μ_{Y_i} is the mean and $\sigma_{Y_i}^2$ the variance of $Y_i, i=1,2$, and ρ is the correlation of Y_1 and Y_2.

Solution.

1. Following properties for expectation, we have

$$\mathrm{E}[AX]=\mathrm{E}\left[\begin{pmatrix}a_{11}X_1+a_{12}X_2\\ a_{21}X_1+a_{22}X_2\end{pmatrix}\right]=\begin{pmatrix}a_{11}\mathrm{E}[X_1]+a_{12}\mathrm{E}[X_2]\\ a_{21}\mathrm{E}[X_1]+a_{22}\mathrm{E}[X_2]\end{pmatrix}=A\mathrm{E}[X].$$

2. By definition, we have

$$\begin{aligned}\mathrm{Var}(AX)&=\mathrm{Var}\begin{pmatrix}a_{11}X_1+a_{12}X_2\\ a_{21}X_1+a_{22}X_2\end{pmatrix}\\ &=\begin{pmatrix}\mathrm{Var}(a_{11}X_1+a_{12}) & \mathrm{Cov}(a_{11}X_1+a_{12}X_2,a_{21}X_1+a_{22}X_2)\\ \mathrm{Cov}(a_{21}X_1+a_{22}X_2,a_{11}X_1+a_{12}X_2) & \mathrm{Var}(a_{21}X_1+a_{22}X_2)\end{pmatrix}.\end{aligned}$$

From the properties of variance and covariance, we have

$$\begin{aligned}\mathrm{Var}(a_{11}X_1+a_{12}X_2)&=a_{11}^2\mathrm{Var}X_1+a_{12}^2\mathrm{Var}X_2+2a_{11}a_{12}\mathrm{Cov}(X_1,X_2),\\ \mathrm{Var}(a_{21}X_1+a_{22}X_2)&=a_{21}^2\mathrm{Var}X_1+a_{22}^2\mathrm{Var}X_2+2a_{21}a_{22}\mathrm{Cov}(X_1,X_2),\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(a_{11}X_1+a_{12}X_2,a_{21}X_1+a_{22}X_2)&=\mathrm{Cov}(a_{21}X_1+a_{22}X_2,a_{11}X_1+a_{12}X_2)\\ &=a_{11}a_{21}\mathrm{Var}(X_1)+a_{12}a_{22}\mathrm{Var}(X_2)+\\ &\qquad+(a_{11}a_{22}+a_{12}a_{21})\mathrm{Cov}(X_1,X_2).\end{aligned}$$

Therefore,

$$\begin{aligned}A(\mathrm{Var}\ X)A^T&=\begin{pmatrix}a_{11}&a_{12}\\ a_{21}&a_{22}\end{pmatrix}\begin{pmatrix}a_{11}\mathrm{Var}X_1+a_{12}\mathrm{Cov}(X_1,X_2) & a_{21}\mathrm{Var}X_1+a_{22}\mathrm{Cov}(X_1,X_2)\\ a_{11}\mathrm{Cov}(X_2,X_1)+a_{12}\mathrm{Var}X_2 & a_{21}\mathrm{Cov}(X_1,X_2)+a_{22}\mathrm{Var}X_2\end{pmatrix}\\ &=\mathrm{Var}(AX).\end{aligned}$$

3. We have

$$\sqrt{\det \Sigma_X} = \sigma_1\sigma_2, \qquad \Sigma_X^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}.$$

   and

$$\langle x - \mu_X, \Sigma_X^{-1}(x - \mu_X)\rangle = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}.$$

   Since X_1 and X_2 are independent,

$$\begin{aligned} f_X(x) &= f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \\ &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} \\ &= \frac{1}{2\pi\sqrt{\det \Sigma_X}} e^{-\frac{1}{2}\langle x-\mu_X, \Sigma_X^{-1}(x-\mu_X)\rangle}. \end{aligned}$$

4. Since $Y = AX$, from (1) and (2) we know that

$$\begin{aligned} \mu_Y = \mathrm{E}[AX] = A\mu_X, \qquad \Sigma_Y = A\Sigma_X A^T \quad &\Rightarrow \quad \Sigma_Y^{-1} = (A^T)^{-1}\Sigma_X^{-1}A^{-1}, \\ &\Rightarrow \quad \det \Sigma_Y = (\det A)^2 \det \Sigma_X. \end{aligned}$$

   Using transformation of variables,

$$\begin{aligned} f_Y(y) &= f_Y \circ (A^{-1}y) \cdot |\det A^{-1}| \\ &= \frac{1}{2\pi\sqrt{\det \Sigma_X}} e^{-\frac{1}{2}\langle A^{-1}y - A^{-1}\mu_Y, \Sigma_X^{-1}(A^{-1}y - A^{-1}\mu_Y)\rangle} \cdot \frac{1}{|\det A|} \\ &= \frac{1}{2\pi\sqrt{\det \Sigma_X \cdot (\det A)^2}} e^{-\frac{1}{2}\langle y-\mu_Y, \Sigma_Y^{-1}(y-\mu_Y)\rangle} \\ &= \frac{1}{2\pi\sqrt{|\det \Sigma_Y|}} e^{-\frac{1}{2}\langle y-\mu_Y, \Sigma_Y^{-1}(y-\mu_Y)\rangle}. \end{aligned}$$

5. Rewriting $\sqrt{|\det \Sigma_Y|}$ as

$$\begin{aligned} \sqrt{|\det \Sigma_Y|} &= \sqrt{\sigma_{Y_1}^2\sigma_{Y_2}^2 - \mathrm{Cov}^2(Y_1, Y_2)} \\ &= \sigma_{Y_1}\sigma_{Y_2}\sqrt{1 - \left(\frac{\mathrm{Cov}(Y_1, Y_2)}{\sigma_{Y_1}\sigma_{Y_2}}\right)^2} \\ &= \sigma_{Y_1}\sigma_{Y_2}\sqrt{1 - \rho^2}, \end{aligned}$$

and

$$\Sigma_Y = \begin{pmatrix} \sigma_{Y_1}^2 & \rho\sigma_{Y_1}\sigma_{Y_2} \\ \rho\sigma_{Y_1}\sigma_{Y_2} & \sigma_{Y_2}^2 \end{pmatrix} \quad \Rightarrow \quad \Sigma_Y^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_{Y_1}^2} & -\frac{\rho}{\sigma_{Y_1}\sigma_{Y_2}} \\ -\frac{\rho}{\sigma_{Y_1}\sigma_{Y_2}} & \frac{1}{\sigma_{Y_2}^2} \end{pmatrix},$$

we have

$$\begin{aligned}\langle y-\mu_Y, \Sigma_Y(y-\mu_Y)\rangle &= \frac{1}{1-\rho^2}\langle \begin{pmatrix} y_1-\mu_{Y_1} \\ y_2-\mu_{Y_2} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sigma_{Y_1}^2} & -\frac{\rho}{\sigma_{Y_1}\sigma_{Y_2}} \\ -\frac{\rho}{\sigma_{Y_1}\sigma_{Y_2}} & \frac{1}{\sigma_{Y_2}^2} \end{pmatrix}\rangle \\ &= \frac{1}{1-\rho^2}\left[\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)^2 - 2\rho\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)\left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right) + \left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right].\end{aligned}$$

Therefore, Eq. (1) can be written as

$$f_Y(y_1,y_2) = \frac{1}{2\pi\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)^2 - 2\rho\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)\left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right) + \left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right]}.$$

Assignment 3.11

A system consists of two independent components connected in series. The life span (in hours) of the component follows a Weibull distribution with $\alpha = 0.006$ and $\beta = 0.5$; the second has a lifespan in hours follows the exponential distribution with $\beta = 1/25000$.

1. Find the reliability of the system at 2500 hours.
2. Find the probability that the system will fail before 2000 hours.
3. If the two components are connected in parallel, what is the system reliability at 2500 hours?

Solution.

1. The reliability function for the two components are given by

$$R_1(t) = e^{-\alpha_1 t^{\beta_1}}, \qquad \begin{aligned} R_2(t) &= 1 - \int_0^t f_{T_2}(x)\mathrm{d}x \\ &= 1 + e^{-\beta_2 x}\Big|_0^t = e^{-\beta_2 t}. \end{aligned}$$

Therefore, the reliability of the system at $t = 2500$ is given by

$$R(t) = R_1(t)\cdot R_2(t) \quad \Rightarrow \quad R(2500) = 0.7408 \times 0.9048 = 0.6703.$$

2. The probability that the system fail before 2000h is given by

$$P[T < 2000] = 1 - R(2000) = 1 - 0.7059 = 0.2941.$$

3. The reliability at $t = 2500$ for the parallel system is

$$R(t) = 1 - (1 - R_1(t))(1 - R_2(t)) \quad \Rightarrow \quad R(2500) = 0.9753.$$

Assignment 4.2

Let $X_1, \ldots, X_n$ be a random sample of size n from a random variable with variance σ^2. We have seen that the sample variance

$$S_{n-1}^2 := \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2$$

is an unbiased estimator for σ^2. It can be shown that

$$\mathrm{Var}(S_{n-1}^2) = \mathrm{MSE}(S_{n-1}^2) = \frac{1}{n} \left(\mathrm{E}[(X - \overline{X})^4] - \frac{n-3}{n-1}\sigma^4 \right) = \frac{1}{n} \left(\gamma_2 + \frac{2n}{n-1} \right) \sigma^4 \tag{2}$$

where $\gamma_2 := \mathrm{E}[(X - \mu)^4]/\sigma^4 - 3$ is called the *excess kurtosis* of a distribution.

1. Show that if X follows a normal distribution with mean μ and variance σ^2,

$$\mathrm{MSE}(S_{n-1}^2) = \frac{2}{n-1}\sigma^4.$$

2. For $a > 0$ set

$$S_a^2 := \frac{n-1}{a} S_{n-1}^2.$$

Find $\mathrm{MSE}(S_a^2)$ and show that the mean square error is minimized for

$$a = n + 1 + \frac{n-1}{n}\gamma_2.$$

In the case of a normal distribution with mean μ and variance σ^2, show that this reduces to $a = n + 1$.

Solution.

1. We know that,

$$\mathrm{MSE}[S_{n-1}^2] = \mathrm{Var}[S_{n-1}^2] + \mathrm{bias}^2 = \mathrm{Var}[S_{n-1}^2].$$

Since X follows a normal distribution,

$$\chi^2_{n-1} = \frac{(n-1)S_{n-1}}{\sigma^2}$$

follows Chi-squared distribution with $n-1$ degrees of freedom. The variance is

$$\mathrm{Var}[\chi^2_{n-1}] = 2(n-1) \quad \Rightarrow \quad \mathrm{Var}[S^2_{n-1}] = \frac{\sigma^4}{(n-1)^2} \cdot 2(n-1) = \frac{2\sigma^4}{n-1}.$$

2. By definition, the MSE for S_a^2 is given by

$$\begin{aligned}\mathrm{MSE}[S_a^2] &= \mathrm{E}[S_a^4 - 2\sigma^2 S_a^2 + \sigma^4] \\ &= \mathrm{E}[S_a^4] - 2\sigma^2 \cdot \frac{n-1}{a}\sigma^2 + \sigma^4 \\ &= \mathrm{E}[S_a^4] + \left(1 - \frac{2(n-1)}{a}\right)\sigma^4.\end{aligned}$$

Using Eq. (2) and property for variance, we have

$$\begin{aligned}\mathrm{E}[S_a^4] &= \mathrm{Var}[S_a^2] + \mathrm{E}[S_a^2]^2 \\ &= \frac{(n-1)^2}{a^2}\mathrm{Var}[S^2_{n-1}] + \frac{(n-1)^2}{a^2}\sigma^4, \\ \mathrm{MSE}[S_a^4] &= \frac{(n-1)^2}{a^2} \cdot \frac{1}{n}\left(\gamma_2 + \frac{2n}{n-1}\right)\sigma^4 + \frac{(n-1)^2}{a^2}\sigma^4 + \left(1 - \frac{2(n-1)}{a}\right)\sigma^4 \\ &= \left[x^2 \cdot \frac{1}{n}\left(\gamma_2 + \frac{2n}{n-1}\right) + x^2 - 2x + 1\right]\sigma^4 \qquad \left(\text{let } x = \frac{n-1}{a}\right) \\ &= \left[\left(\frac{\gamma_2}{n} + \frac{2}{n-1} + 1\right)x^2 - 2x + 1\right],\end{aligned}$$

which is maximized when

$$x = \frac{1}{\dfrac{\gamma_2}{n} + \dfrac{2}{n-1} + 1} = \frac{n-1}{a} \quad \Rightarrow \quad a = n + 1 + \frac{n-1}{n}\gamma_2.$$

In case of normal distribution with mean μ and variance σ^2, since

$$\gamma_2 = \mathrm{E}\left[\frac{(X-\mu)^4}{\sigma^4}\right] - 3 = \mathrm{E}[Z^4] - 3,$$

where Z follows a standard normal distribution and $\mathrm{E}[Z^4]$ is the 4th moment of it, given by

$$\mathrm{E}[Z^4] = \left.\frac{\mathrm{d}^4 m_Z(t)}{\mathrm{d}t^4}\right|_{t=0} = \left.(t^4 + 6t^2 + 3)e^{\frac{1}{2}t^2}\right|_{t=0} = 3,$$

and thus

$$a = n + 1.$$

Mean and Variance for Estimators

We know that the mean square error is given by

$$\begin{aligned}
\mathrm{MSE}[\widehat{\theta}] &= \mathrm{E}[(\widehat{\theta} - \theta)^2] \\
&= \mathrm{E}[(\widehat{\theta} - \mathrm{E}[\widehat{\theta}] + \mathrm{E}[\widehat{\theta}] - \theta)^2] \\
&= \mathrm{E}[(\widehat{\theta} - \mathrm{E}[\widehat{\theta}])^2] + \mathrm{E}[2(\widehat{\theta} - \mathrm{E}[\widehat{\theta}])(\mathrm{E}[\widehat{\theta}] - \theta)] + \mathrm{E}[(\mathrm{E}[\widehat{\theta}] - \theta)^2] \\
&= \mathrm{Var}\ \widehat{\theta} + \mathrm{bias}^2.
\end{aligned}$$

MSE is an overall measurement of the quality of the estimator. This means that under the same MSE, estimators with lower bias have a larger variance, and vice versa. This is called the bias-variance trade-off. Therefore, depending on the context, we might prefer a smaller bias or smaller variance.

Here we analyze the mean and variance for unbiased estimators that we have seen in lectures and assignments. For now on, suppose we have sample $X_1, \ldots, X_n$ from a population X with mean μ and variance σ^2.

Mean

The unbiased estimator for mean is given by

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Then the mean and variance for this estimator are given by

$$\begin{aligned}
\mathrm{E}[\widehat{\mu}] &= \mathrm{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[X_i] = \mu, \\
\mathrm{Var}\ \widehat{\mu} &= \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \frac{1}{n^2} \mathrm{Var}\ X_i = \frac{\sigma^2}{n}.
\end{aligned}$$

Variance

The unbiased estimator for variance is given by

$$\widehat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Then

$$\begin{aligned}
\mathrm{E}[\widehat{\sigma^2}] &= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^n X_i^2 - 2X_i\overline{X} + \overline{X}^2\right] \\
&= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^n X_i^2 - n\overline{X}^2\right] \\
&= \frac{1}{n-1}\left(\sum_{i=1}^n \mathrm{E}[X_i^2] - n\mathrm{E}[\overline{X}^2]\right) \\
&= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) \\
&= \sigma^2.
\end{aligned}$$

For variance of this estimator, we have

$$\begin{aligned}
\mathrm{Var}\ \widehat{\sigma^2} &= \mathrm{Var}\left(\frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2\right) \\
&= \mathrm{E}\left[\left(\frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2 - \sigma^2\right)^2\right].
\end{aligned}$$

Rewriting sample variance as

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - \frac{1}{n}\left(\sum_{i=1}^n X_i\right)^2\right) \\
&= \frac{1}{n(n-1)}\left((n-1)\sum_{i=1}^n X_i^2 - 2\sum_{i<j} X_iX_j\right) \\
&= \frac{1}{n(n-1)}\sum_{i<j}(X_i-X_j)^2,
\end{aligned}$$

we have

$$\begin{aligned}
\mathrm{E}[(\widehat{\sigma^2}-\sigma^2)^2] &= \mathrm{E}[S^4 - 2\sigma^2S^2 + \sigma^4] \\
&= \mathrm{E}\left[\left(\frac{1}{n(n-1)}\sum_{i<j}(X_i-X_j)^2\right)^2\right] - \sigma^4 \\
&= \mathrm{E}\left[\left(\frac{1}{n(n-1)}\sum_{i<j}(X_i-X_j)^2\right)\left(\frac{1}{n(n-1)}\sum_{p<q}(X_p-X_q)^2\right)\right] - \sigma^4.
\end{aligned}$$

Expanding the product above, we can have the following cases regarding the values of i, j and p, q. Denote the kth moment of X as $\mathrm{E}[X^k] = m_k$.

- $\underline{i = j, p = q}$. Since X_i and X_j are independent,

$$\begin{aligned}\mathrm{E}[(X_i - X_j)^4] &= \mathrm{E}[X_i^4 - 4X_i^3X_j + 6X_i^2X_j^2 - 4X_iX_j^3 + X_j^4] \\ &= 2m_4 - 8m_1m_3 + 6m_2^2.\end{aligned}$$

  There are $\dfrac{n(n-1)}{2}$ ways of choosing i, j, and thus there are the same number of such terms.

- $\underline{|\{i,j\} \cap \{p,q\}| = 1}$. $|\cdot|$ is the cardinality of set. Suppose for now that $j = q$.

$$\begin{aligned}\mathrm{E}[(X_i - X_j)^2(X_p - X_q)^2] &= \mathrm{E}[(X_iX_p - X_iX_q - X_pX_q + X_q^2)^2] \\ &= \mathrm{E}[X_i^2X_p^2 + X_i^2X_q^2 + X_p^2X_q^2 + X_q^4 - \\ &\qquad - 2X_i^2X_pX_q - 2X_iX_p^2X_q + 2X_iX_pX_q^2 + \\ &\qquad + 2X_iX_pX_q^2 - 2X_iX_q^3 - 2X_pX_q^3] \\ &= m_4 + 3m_2^2 - 4m_1m_3.\end{aligned}$$

  There are $n(n-1)(n-2)$ ways of such terms.

- $\underline{i \neq j \neq p \neq q}$. Since the four variables are independent,

$$\begin{aligned}\mathrm{E}[(X_i - X_j)^2(X_p - X_q)^2] &= \big(\mathrm{E}[(X_i - X_j)^2]\big)^2 \\ &= \big(\mathrm{E}[X_i^2 - 2X_iX_j + X_j^2]\big)^2 \\ &= \big(2m_2 - 2m_1^2\big)^2 \\ &= 4m_2^2 - 8m_1^2m_2 + 4m_1^4.\end{aligned}$$

  There are $\dfrac{n(n-1)(n-2)(n-3)}{4}$ ways of choosing i, j, p, q, and thus there are the same number of such terms.

Summing up, we have

$$\begin{aligned}\mathrm{E}[\widehat{\sigma^2}] &= \frac{1}{n^2(n-1)^2}\mathrm{E}\left[\left(\sum_{i<j}(X_i - X_j)^2\right)^2\right] \\ &= \frac{1}{n}\left(m_4 - 4m_1m_3 - \frac{n-3}{n-1}m_2^2 + \frac{4(2n-3)}{n-1}m_1^2m_2 - \frac{2(2n-3)}{n-1}m_1^4\right),\end{aligned}$$

which coincides with what is given in Assignment 4.2, where

$$\begin{aligned}
\frac{1}{n}\left(\gamma_2 + \frac{2n}{n-1}\right)\sigma^4 &= \frac{1}{n}\left(\mathrm{E}[(X-\mu)^4] - \frac{n-3}{n-1}(m_2 - m_1^2)^2\right) \\
&= \frac{1}{n}\left(\mathrm{E}[X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4] - \frac{n-3}{n-1}(m_2 - m_1^2)^2\right) \\
&= \frac{1}{n}\left(m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4 - \frac{n-3}{n-1}(m_1^4 - 2m_1^2m_2 + m_2^2)\right) \\
&= \frac{1}{n}\left(m_4 - 4m_1m_3 + \frac{8n-12}{n-1}m_1^2m_2 - \frac{n-3}{n-1}m_2^2 - \frac{4n-6}{n-1}m_1^4\right).
\end{aligned}$$