1 Variance

Let X_i for $i = 1, 2, \ldots, n$ be independent and identically-distributed random variables with $\sigma^2 = \mathrm{Var}(X_i)$ and $\mu = E(X_i)$. We denote

$$\hat{\sigma}^2 = \frac{1}{K_1(n)} \sum_{i=1}^{n} (X_i - \mu)^2$$

and

$$S^2 = \frac{1}{K_2(n)} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

where $\bar{X} = \sum_{i=1}^{n} X_i / n$.

1. (a) Find $K_1(n)$ satisfying $E(\hat{\sigma}^2) = \sigma^2$. You can find $K_1(n)$ mathematically (without normality assumption).

   Solution:

   Since

   $$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} X_i^2 - 2\mu \sum_{i=1}^{n} X_i + n\mu^2$$

   and

   $$E(X_i^2) = \mathrm{Var}(X_i) + \mu^2 = \sigma^2 + \mu^2,$$

   we have

   $$E\left[\sum_{i=1}^{n} (X_i - \mu)^2\right] = n(\sigma^2 + \mu^2) - 2\mu(n\mu) + n\mu^2 = n\sigma^2.$$

   Thus, $K_1(n) = n$. NOTE: we did not use any distribution assumption.

   (b) Find $K_2(n)$ satisfying $E(S^2) = \sigma^2$. You can find $K_2(n)$ mathematically (without normality assumption).

Solution:

Since

$$
\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - n\bar{X}^2,
\end{aligned}
$$

$$E(X_i^2) = \mathrm{Var}(X_i) + \mu^2 = \sigma^2 + \mu^2,$$

and

$$E(\bar{X}^2) = \mathrm{Var}(\bar{X}) + \left\{E(\bar{X})\right\}^2 = \sigma^2/n + \mu^2$$

we have

$$E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = n(\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2) = (n-1)\sigma^2.$$

Thus, $K_2(n) = n - 1$. NOTE: we did not use any distribution assumption.

2. (a) Using the Monte Carlo simulation (R program), estimate $K_1(n)$ satisfying $E(\hat{\sigma}^2) = \sigma^2$ under the assumption that X_i are from a normal distribution. (Set $n = 5$ for example).

   (b) Using the Monte Carlo simulation (R program), estimate $K_2(n)$ satisfying $E(S^2) = \sigma^2$ under the assumption that X_i are from a normal distribution. (Set $n = 5$ for example).

3. (a) Using the Monte Carlo simulation (R program), estimate $K_1(n)$ satisfying $E(\hat{\sigma}^2) = \sigma^2$ under the assumption that X_i are from a **exponential** distribution ($\mu = 1$). (Set $n = 5$ for example).

   (b) Using the Monte Carlo simulation (R program), estimate $K_2(n)$ satisfying $E(S^2) = \sigma^2$ under the assumption that X_i are from a **exponential** distribution ($\mu = 1$). (Set $n = 5$ for example).

2 Standard deviation

Let X_i for $i = 1, 2, \ldots, n$ be independent and identically-distributed random variables with $\sigma = \sqrt{\mathrm{Var}(X_i)}$. We denote

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$.

1. Find $c_4(n)$ satisfying $E(S/c_4(n)) = \sigma$ under the assumption that X_i are from a normal distribution. You can find $c_4(n)$ mathematically.

Solution: Earlier we proved that the sample variance is unbiased under the normal distribution assumption. That is

$$E[S^2] = E[\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)] = \sigma^2.$$

Given that the variance estimate is unbiased, a natural question arises. *Is the sample standard deviation also unbiased?* We already know that, *if* the sample standard deviation is unbiased, then the relation below holds

$$E[S] = \sigma,$$

where $S = \left[\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)\right]^{1/2}$. As it turns out, the sample standard deviation is *not* unbiased under the normal distribution assumption but, by calculating the expectation of S explicitly, we can derive a bias-corrected expression which is unbiased. The derivation follows below.

Under the normal distribution assumption, we can prove that $(n-1)S^2/\sigma^2$ has the chi-square distribution with $n-1$ degrees of freedom which is equivalent to the gamma distribution with $\alpha = (n-1)/2$ and $\theta = 2$. Now, it is well known that

$$E\left[Y^c\right] = \frac{\Gamma(\alpha + c)\theta^c}{\Gamma(\alpha)}$$

when Y has the gamma distribution with parameters α and θ. Clearly, for $c = 1/2$, we have

$$E\left[\sqrt{Y}\right] = \frac{\Gamma(\alpha + 1/2)\sqrt{\theta}}{\Gamma(\alpha)}.$$

Now let $Y = (n-1)S^2/\sigma^2$ so that Y has the gamma distribution with parameters $\alpha = (n-1)/2$ and $\theta = 2$. Then using the relation above, we obtain

$$E\left[\sqrt{(n-1)S^2/\sigma^2}\right] = \frac{\Gamma(n/2)\sqrt{2}}{\Gamma(n/2 - 1/2)}.$$

We then take the previous expectation and simplify under the square-root on the left hand side of the above equation and obtain

$$\frac{\sqrt{n-1}}{\sigma} \cdot E[S] = \frac{\Gamma(n/2)\sqrt{2}}{\Gamma(n/2 - 1/2)}.$$

But this implies that

$$E[S] = c_4\,\sigma$$

where

$$\boxed{c_4 = \sqrt{\frac{2}{n-1}} \cdot \frac{\Gamma(n/2)}{\Gamma(n/2 - 1/2)}.}$$

Thus, the estimator S/c_4 is unbiased for σ.

2. Can you find $c_4(n)$ satisfying $E(S/c_4(n)) = \sigma$ under any other distribution?

3. Using the Monte Carlo simulation (R program), estimate $c_4(n)$ satisfying $E(S/c_4(n)) = \sigma$ under the assumption that X_i are from a normal distribution.

4. Using the Monte Carlo simulation (R program), estimate $c_4(n)$ satisfying $E(S/c_4(n)) = \sigma$ under the assumption that X_i are from a **exponential** distribution ($\mu = 1$).