

1. MGB VI.5 Let X_1, \dots, X_n be a random sample from a distribution which has a finite fourth moment. Define $\mu = E(X_1)$, $\sigma^2 = Var(X_1)$, $\mu_3 = E[(X_1 - \mu)^3]$, $\mu_4 = E[(X_1 - \mu)^4]$, $\bar{X} = (1/n) \sum X_i$, $S^2 = [1/(n-1)] \sum (X_i - \bar{X})^2$.

(a) Does $S^2 = [1/2n(n-1)] \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2$?

Solution:

Well,

$$\begin{aligned}
 \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2 &= \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X} + \bar{X} - X_j)^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^n \left[(X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - X_j) + (\bar{X} - X_j)^2 \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})^2 + 2 \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})(\bar{X} - X_j) + \sum_{i=1}^n \sum_{j=1}^n (\bar{X} - X_j)^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})^2 + 2 \sum_{i=1}^n (X_i - \bar{X}) \cdot 0 + \sum_{i=1}^n \sum_{j=1}^n (\bar{X} - X_j)^2 \\
 &= \sum_{i=1}^n n \cdot (X_i - \bar{X})^2 + n \cdot \sum_{j=1}^n (\bar{X} - X_j)^2 \\
 &= 2n \cdot \sum_{i=1}^n (X_i - \bar{X})^2.
 \end{aligned}$$

Dividing both sides of the equality above by $1/2n(n-1)$ establishes the equality.

□

(b) Show $Var(S^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right)$.

Solution:

Note that

$$\begin{aligned}
 S^2 &= \frac{1}{n-1} \sum_{i=1}^n \left[(X_i - \mu) - (\bar{X} - \mu) \right]^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2 - 2 \sum_{i=1}^n \sum_{j=1}^n (X_i - \mu)(\bar{X} - \mu) + \sum_{i=1}^n (\bar{X} - \mu)^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2 + 0 - n(\bar{X} - \mu)^2.
 \end{aligned}$$

So,

$$\begin{aligned}
(S^2)^2 &= \frac{1}{(n-1)^2} \left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \right]^2 \\
&= \frac{1}{(n-1)^2} \left[\sum_{i=1}^n (X_i - \mu)^2 - \frac{1}{n} (n\bar{X} - n\mu)(n\bar{X} - n\mu) \right]^2 \\
&= \frac{1}{(n-1)^2} \left[\sum_{i=1}^n (X_i - \mu)^2 - \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \sum_{j=1}^n (X_j - \mu) \right]^2 \quad (1) \\
&\stackrel{\dagger}{=} \frac{1}{(n-1)^2} \left[\sum_{i=1}^n \sum_{j=1}^n A_{i,j} - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n B_{i,j,k} + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n C_{i,j,k,l} \right],
\end{aligned}$$

where

$$\begin{aligned}
A_{i,j} &= (X_i - \mu)^2 (X_j - \mu)^2, \\
B_{i,j,k} &= (X_i - \mu)^2 (X_j - \mu) (X_k - \mu), \\
C_{i,j,k,l} &= (X_i - \mu) (X_j - \mu) (X_k - \mu) (X_l - \mu).
\end{aligned}$$

By linearity, let us calculate the expectation of $(S^2)^2$ by calculating the expectation of the right hand side of \dagger by each term separately. The main strategy is to partition each integer lattice into disjoint sets and add over them.^{1 2} That is,

$$\begin{aligned}
E \left[\sum_{i=1}^n \sum_{j=1}^n A_{i,j} \right] &= E \left[\sum_{i=j} (X_i - \mu)^4 + \sum_{i \neq j} (X_i - \mu)^2 (X_j - \mu)^2 \right] \\
&\stackrel{\dagger}{=} \sum_{i=j} E(X_i - \mu)^4 + \sum_{i \neq j} E(X_i - \mu)^2 E(X_j - \mu)^2 \\
&\stackrel{*}{=} n\mu_4 + n(n-1)\sigma^2 \cdot \sigma^2,
\end{aligned}$$

and

$$\begin{aligned}
E \left[\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n B_{i,j,k} \right] &= E \left[\sum_{i=j=k} B_{i,j,k} + \sum_{\substack{i \neq j \\ j=k}} B_{i,j,k} + 2 \sum_{\substack{i=j \\ j \neq k}} B_{i,j,k} + \sum_{i \neq k \neq j} B_{i,j,k} \right] \\
&\stackrel{\dagger}{=} \sum_{i=j=k} E(X_i - \mu)^4 + \sum_{\substack{i \neq j \\ j=k}} E(X_i - \mu)^2 E(X_j - \mu)^2 + 2 \sum_{\substack{i=j \\ j \neq k}} 0 + \sum_{i \neq k \neq j} 0 \\
&\stackrel{*}{=} n\mu_4 + n(n-1)\sigma^2 \cdot \sigma^2,
\end{aligned}$$

¹In many of the equalities above, we use the independence of distinct X_i and the fact that $g(X_i, \dots)$ are also independent for arbitrary g . These are labeled with \dagger .

²We use the fact that the number of terms in the double sum with two non equal indices is ${}_nP_2 = n(n-1)$. These are labeled with $*$.

and

$$\begin{aligned}
E \left[\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n C_{i,j,k,l} \right] &= E \left[\sum_{i=j=k=l} C_{i,j,k,l} + 3 \sum_{\substack{i=j \\ j \neq k \\ k=l}} C_{i,j,k,l} + \dots \right. \\
&\quad \left. 4 \sum_{\substack{i \neq j \\ j=k=l}} C_{i,j,k,l} + 4 \sum_{\substack{i \neq j \neq k \\ k=l}} C_{i,j,k,l} + \sum_{i \neq j \neq k \neq l} C_{i,j,k,l} \right] \\
&\stackrel{\ddagger}{=} \left[\sum_{i=j=k=l} E(X_i - \mu)^4 + 3 \sum_{\substack{i=j \\ j \neq k \\ k=l}} E(X_i - \mu)^2 E(X_k - \mu)^2 + \dots \right. \\
&\quad \left. 4 \sum_{\substack{i \neq j \\ j=k=l}} 0 + 4 \sum_{\substack{i \neq j \neq k \\ k=l}} 0 + \sum_{i \neq j \neq k \neq l} 0 \right] \\
&\stackrel{*}{=} n\mu_4 + 3n(n-1)\sigma^4.
\end{aligned}$$

So the expected value of \ddagger ,

$$\begin{aligned}
E[(S^2)^2] &\stackrel{\ddagger}{=} \frac{1}{(n-1)^2} \left[(n\mu_4 + n(n-1)\sigma^4) - \frac{1}{n} (n\mu_4 + n(n-1)\sigma^4) + \frac{1}{n^2} (n\mu_4 + n(n-1)3\sigma^4) \right] \\
&= \frac{1}{(n-1)^2} \left[n\mu_4 \left(1 - \frac{2}{n} + \frac{1}{n^2} \right) + n(n-1)\sigma^4 \left(1 - \frac{2}{n} + \frac{3}{n^2} \right) \right] \\
&= \frac{1}{(n-1)^2} \left[n\mu_4 \left(\frac{n^2 - 2n + 1}{n^2} \right) + n(n-1)\sigma^4 \left(\frac{n^2 - 2n + 3}{n^2} \right) \right] \\
&= \frac{1}{(n-1)^2} \left[\mu_4 \left(\frac{(n-1)^2}{n} \right) + (n-1)\sigma^4 \left(\frac{n^2 - 2n + 3}{n} \right) \right] \\
&= \frac{1}{n} \left[\mu_4 + \sigma^4 \left(\frac{n^2 - 2n + 3}{n-1} \right) \right].
\end{aligned}$$

So,

$$\begin{aligned}
Var[S^2] &= E[(S^2)^2] - E(S^2)^2 \\
&= \frac{1}{n} \left[\mu_4 + \sigma^4 \left(\frac{n^2 - 2n + 3}{n-1} \right) \right] - \sigma^4 \\
&= \frac{1}{n} \left[\mu_4 + \sigma^4 \left(\frac{n^2 - 2n + 3 - n(n-1)}{n-1} \right) \right] \\
&= \frac{1}{n} \left[\mu_4 - \sigma^4 \left(\frac{n-3}{n-1} \right) \right].
\end{aligned}$$

□

(c) Find $Cov(\bar{X}, S^2)$ and note that $Cov(\bar{X}, S^2) = 0$ if $\mu_3 = 0$.

Solution:

As in (b), we seek to decompose $S^2\bar{X}$ into sums of $X_i - \mu$. That is, first note

$$S^2\bar{X} = \frac{1}{n} S^2 \sum_{i=1}^n (X_i - \mu) + S^2\mu. \quad (2)$$

So we consider

$$\begin{aligned} S^2 \left(\sum_{i=1}^n X_i - n\mu \right) &= \frac{1}{n-1} \left(\sum_{i=1}^n (X_i - \mu)^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (X_i - \mu)(X_j - \mu) \right) \sum_{i=1}^n (X_i - \mu) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sum_{j=1}^n (X_i - \mu)^2 (X_j - \mu) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (X_i - \mu)(X_j - \mu)(X_k - \mu) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n \sum_{j=1}^n A_{i,j} - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n B_{i,j,k} \right), \end{aligned}$$

where

$$\begin{aligned} A_{i,j} &= (X_i - \mu)^2 (X_j - \mu), \\ B_{i,j,k} &= (X_i - \mu)(X_j - \mu)(X_k - \mu). \end{aligned}$$

We proceed as in (b), by partitioning each sum into disjoint sets of tuples of integers such that we can calculate expected values. I.e.

$$\begin{aligned} E \left[\sum_{i=1}^n \sum_{j=1}^n A_{i,j} \right] &= \sum_{i=j} E(X_j - \mu)^3 + \sum_{i \neq j} E(X_j - \mu)^2 (E(X_j) - \mu) \\ &= n\mu_3, \end{aligned}$$

and

$$\begin{aligned} E \left[\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n B_{i,j,k} \right] &= \sum_{i=j=k} E(X_i - \mu)^3 + 3 \sum_{\substack{i=j \\ i \neq k}} 0 + \sum_{i \neq j \neq k} 0 \\ &= n\mu_3. \end{aligned}$$

Taking the expected value of (2),

$$E[S^2\bar{X}] = \frac{1}{n(n-1)} (n\mu_3 - \mu_3) + \frac{\sigma^2\mu}{n}$$

so

$$\begin{aligned} Cov(S^2, \bar{X}) &= E[S^2\bar{X}] - E[S^2] E[\bar{X}] \\ &= \frac{1}{n(n-1)} (n\mu_3 - \mu_3) + \sigma^2\mu - \sigma^2\mu = \frac{1}{n(n-1)} (n\mu_3 - \mu_3) \end{aligned}$$

□

2. MGB VI.6

(a) From a random sample of size 2 from a population with finite $(2r)$ th moment, find $E(M_r)$ and $Var(M_r)$, where $M_r = (1/n) \sum_{i=1}^n (X_i - \bar{X}_n)$.

Solution:

Note that $\bar{X} = (X_1 + X_2)/2$, and thus we can first calculate

$$\begin{aligned} M_r &= \frac{1}{2} \left(X_1 - \frac{X_1 + X_2}{2} \right)^r + \frac{1}{2} \left(X_2 - \frac{X_1 + X_2}{2} \right)^r \\ &\stackrel{\dagger}{=} \frac{1}{2} \left(\frac{X_1 - X_2}{2} \right)^r + \frac{1}{2} \left(\frac{X_2 - X_1}{2} \right)^r. \end{aligned}$$

When r is odd, we have that $M_r \equiv 0$, hence $E(M_r) = 0$. Otherwise

$$\begin{aligned} E[M_r] &= E \left[\left(\frac{X_1 - X_2}{2} \right)^r \right] \\ &= E \left[2^{-r} \sum_{i=0}^r (-1)^i X_1^{r-i} X_2^i \right] \\ &= 2^{-r} \sum_{i=0}^r (-1)^i E[X_1^{r-i} X_2^i] \\ &= 2^{-r} \mu'_r \sum_{i=0}^r (-1)^i \\ &= 2^{-r} \mu'_r, \end{aligned}$$

Where the last equality follows from r being even. For the variance, note that if r is odd $E[M_r^2] = 0$ by \dagger . If r is even,

$$\begin{aligned} E[M_r^2] &= E \left[\left(\frac{X_1 - X_2}{2} \right)^{2r} \right] \\ &= E[M_{2r}] \\ &= 2^{-2r} \mu'_{2r}. \end{aligned}$$

Thus, $Var[M_r] = 2^{-2r}(\mu'_{2r} - \mu'^2_r)$.

□

(b) For a random sample of size n from a population with mean μ and r th central moment μ_r , show that

$$E \left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^r \right] = \mu_r$$

Solution:

Well,

$$\begin{aligned} E \left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^r \right] &= \frac{1}{n} \sum_{i=1}^n E [(X_i - \mu)^r] \\ &= \frac{1}{n} \sum_{i=1}^n \mu_r \\ &= \frac{1}{n} n \mu_r = \mu_r. \end{aligned}$$

□

3. MGB VI.9 Suppose that \bar{X}_1 and \bar{X}_2 are means of two samples of size n from a population with variance σ^2 . Determine n so that the probability will be about .01 that the two sample means will differ by more than σ . (Consider $Y = \bar{X}_1 - \bar{X}_2$.)

Solution:

We assume that \bar{X}_1 and \bar{X}_2 are independent. Thus

$$Var(Y) = Var(\bar{X}_1) + Var(-\bar{X}_2) = \frac{2\sigma^2}{n}.$$

By Chebychev's inequality,

$$\begin{aligned} P(|Y| \geq \sigma) &= P \left(|Y| \geq \sqrt{\frac{n}{2}} \sqrt{\frac{2}{n}} \sigma \right) \\ &\leq \frac{2}{n}. \end{aligned}$$

Thus if $\frac{2}{n} < .01$, or rather $n > 200$, we can guarantee $P(|Y| \geq \sigma) < .01$.

Alternatively, let us assume[†] that the population is distributed such that n is large enough so that $F_n(x) \approx \Phi(x)$ and $G_n(x) \approx \Phi(x)$ where Φ is the c.d.f. for the standard normal distribution and F_n and G_n are the c.d.f.'s for the random variables $(\bar{X}_{1,2} - \mu)/(\sigma/\sqrt{n})$ respectively (guaranteed by the central limit theorem). Thus,

$$\begin{aligned} Y &= \bar{X}_1 - \bar{X}_2 \\ &= \left(\frac{\bar{X}_1 - \mu}{\frac{\sigma}{\sqrt{n}}} - \frac{\bar{X}_1 - \mu}{\frac{\sigma}{\sqrt{n}}} \right) \frac{\sigma}{\sqrt{n}} \\ &\approx (Z_1 - Z_2) \frac{\sigma}{\sqrt{n}}, \end{aligned}$$

Where Z_1 and Z_2 are independent and identically distributed $N(0, 1)$. It is known that linear combinations of normal random variables are distributed normally. In particular, $Y \sim N \left(\mu - \mu, \frac{\sigma^2}{n} + \frac{\sigma^2}{n} \right) \sim N \left(0, \frac{2\sigma^2}{n} \right)$. Thus,

$$\begin{aligned} P(|Y| \geq \sigma) &\approx 2\Phi \left(-\frac{\sigma}{2\sigma/\sqrt{n}} \right) \\ &= 2\Phi \left(-\frac{\sqrt{n}}{2} \right). \end{aligned}$$

If we require that this probability be approximately .01, we can approximate the quantile .01/2 of the standard normal in R, and solve for n to arrive at an estimation. That is,

```
> (2*qnorm(.01/2))**2
[1] 26.53959
```

Thus, if $n \geq 27$ and the population meets assumption \dagger for $n = 27$, we can guarantee two samples will differ by more than σ with probability .01.

□

4. MGB VI.10 Suppose that light bulbs made by a standard process have an average life of 2000 hours with a standard deviation of 250 hours, and suppose that it is considered worthwhile to replace the process if the mean life can be increased by at least 10 percent. An engineer wishes to test a proposed new process, and he is willing to assume that the standard deviation of the distribution of lives is about the same as for the standard process. How large a sample should he examine if he wishes the probability to be about .01 that he will fail to adopt the new process if in fact it produces bulbs with a mean life of 2250 hours?

Solution:

Let X be the random variable representing the new bulb life. We assume that it has $\mu = 2250$ and $\sigma^2 = 250$. Let X_1, X_2, \dots, X_n be a random sample from this sample with population mean \bar{X}_n . We know that the process will not be adopted if our estimate for μ , \bar{X}_n , is less than 10 percent more than the original process – i.e. $\bar{X}_n < 2200$. Well,

$$\begin{aligned} P(\bar{X}_n \leq 2200) &= P(\bar{X}_n - \mu \leq -50) && \text{(recall } \mu = 2250\text{)} \\ &\leq P(\bar{X}_n - \mu \leq -50) + P(\bar{X}_n - \mu \geq 50) \\ &= P(|\bar{X}_n - \mu| \geq 50) && \text{since the events are disjoint,} \\ &= P\left(|\bar{X}_n - \mu| \geq \frac{\sqrt{n}}{5} \cdot \frac{250}{\sqrt{n}}\right) \end{aligned}$$

Recall that $250/\sqrt{n} = \sigma/\sqrt{n}$ is the standard deviation for \bar{X}_n , hence Cheyechev's inequality says

$$P(\bar{X}_n \leq 2200) < \frac{25}{n}.$$

In which case, if $n > 2500$ we are guaranteed with probability .99 that the process will not be adopted.

Alternatively, if we assume that n is large enough so that $\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right)$ is distributed approximately normal, say $Z \sim N(0, 1)$ (guaranteed by central limit theorem).

Then,

$$\begin{aligned} P(\bar{X}_n \leq 2200) &= P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{-50}{250/\sqrt{n}}\right) \\ &\approx P\left(Z \leq -\frac{\sqrt{n}}{5}\right) \\ &= \Phi\left(-\frac{\sqrt{n}}{5}\right) \end{aligned}$$

If we require that this probability be less than .01, we can find the quantile associated with .01 (via R) and solve for n . That is:

```
> (5*qnorm(.01))**2
[1] 135.2974
```

So, if $n > 136$ and n is sufficient to make the normal approximation valid, then there is a probability of .01 that the process will not be adopted.

□

5. MGB VI.14 Find the third moment about the mean of the sample mean for sample sizes of size n from a Bernoulli population. Show that it approaches 0 as n becomes large (as it must if the normal approximation is to be valid).

Solution:

We first show generically for any random variable X with finite third moment that

$$\begin{aligned} E[(\bar{X} - \mu)^3] &= E\left[\left(\sum_{i=1}^n X_i/n - n\mu/n\right)^3\right] \\ &= \frac{1}{n^3} E\left[\left(\sum_{i=1}^n X_i - n\mu\right)^3\right] \\ &= \frac{1}{n^3} E\left[\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (X_i - \mu)(X_j - \mu)(X_k - \mu)\right] \\ &= \frac{1}{n^3} E\left[\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (X_i - \mu)(X_j - \mu)(X_k - \mu)\right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^3} \left[\sum_{i=j=k} E(X_i - \mu)^3 + 3 \sum_{\substack{i=j \\ i \neq k}} E(X_i - \mu)^2 (E(X_k) - \mu) + \dots \right. \\
&\quad \left. \sum_{i \neq j \neq k} (E(X_i) - \mu)(E(X_j) - \mu)(E(X_k) - \mu) \right] \\
&= \frac{1}{n^3} \left[\sum_{i=j=k} E(X_i - \mu)^3 + 3 \sum_{\substack{i=j \\ i \neq k}} 0 + \sum_{i \neq j \neq k} 0 \right] \\
&= \frac{1}{n^2} \mu_3.
\end{aligned}$$

This shows that, in general, as $n \rightarrow \infty$ the third moment about μ of \bar{X}_n (when it exists) approaches 0 as predicted by the central limit theorem. (The normal distribution has zero odd moments.) In particular for the Bernoulli distribution, we note that its moment generating function is $M_X(t) = q + pe^t$ of which $M_X^{(n)}(0) = p$, so

$$\begin{aligned}
\mu_3 &= E(X - \mu)^3 \\
&= E(X^3) - 3E(X^2)\mu + 3E(X)\mu^2 - \mu^3 \\
&= p - 3p^2 + 2p^3.
\end{aligned}$$

Remark: Note that we could use the fact that $Y = \sum X_i$ is distributed Binomial to find $E(\bar{X} - p)^3 = 1/n^3 E(\sum X_i - np)^3 = \mu_Y^3$ for a less general solution.

□