Expectation and Variance of Sample Mean, Sampling w.o Replacement

Kazu

September 6, 2017

This note will work through the probabilty theory and algebra of deriving the finite population correction for sampling without replacement.

Assume $(X_1, \cdots, X_n)$ are n samples drawn without replacement from a population of size N which has population mean μ and population variance σ^2. Let c_a denote some value X_i can assume, and let $\mathbb{P}[X_i = c_a] = n_a/N$.

Let the sample mean $\bar{X}_n$ equal $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$

Theorem 1. $\mathbb{E}[\bar{X}_n] = \mu$

Proof. $\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i\right] = \underbrace{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathbb{E}[X_i]}_{\text{by linearity of expectation}} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mu = \mu$ □

Theorem 2. $\mathrm{Var}[\bar{X}_n] = \dfrac{\sigma^2}{n}\dfrac{N-n}{N-1} = \dfrac{\sigma^2}{n}\left(1 - \dfrac{n-1}{N-1}\right)$

First, we will go through a couple lemmas, applications of the lemmas are <u>underlined</u>:
Lemma 1.

$$\mathbb{E}[X_i X_j] = \sum_{c_a}^{k}\sum_{c_b}^{k} c_a c_b \mathbb{P}(X_i = c_a, X_j = c_b) \tag{1}$$

$$= \sum_{c_a}^{k} c_a \mathbb{P}(X_i = c_a) \sum_{c_b}^{k} c_b \mathbb{P}(X_j = c_b | X_i = c_a) \tag{2}$$

$$= \sum_{c_a}^{k}\left[c_a \frac{n_a}{N}\left[\sum_{c_b}^{k}\left(c_b \frac{n_b}{N-1}\right) - c_a \frac{1}{N-1}\right]\right] \tag{3}$$

$$= \frac{1}{N(N-1)}\left[\sum_{c_a}^{k}\sum_{c_b}^{k} c_a c_b n_a n_b - \sum_{c_a}^{k} c_a^2 n_a\right] \tag{4}$$

$$= \frac{1}{N(N-1)}\left[\sum_{c_a}^{k} c_a n_a \sum_{c_b}^{k} c_b n_b - \sum_{c_a}^{k} c_a^2 n_a\right] \tag{5}$$

$$= \frac{1}{N(N-1)}\left[(N\mathbb{E}[X_i])^2 - N\mathbb{E}[X_i^2]\right] \tag{6}$$

$$= \frac{1}{N(N-1)}\left[(N\mathbb{E}[X_i])^2 - N(\mathrm{Var}[X_i] - \mathbb{E}[X_i]^2)\right] \tag{7}$$

$$= \frac{1}{(N-1)}[N\mu^2 - (\sigma^2 + \mu^2)] \tag{8}$$

$$= \frac{1}{(N-1)}[\mu^2(N-1) - \sigma^2] \tag{9}$$

$$= \mu^2 - \frac{\sigma^2}{N-1} \tag{10}$$

Lemma 2.

$$
\begin{aligned}
\mathbb{E}[(X_1 + \ldots + X_n)^2] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} X_i X_j\right] && (11)\\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i=1}^{n}\sum_{j\neq i} \mathbb{E}[X_i X_j] && (12)\\
&= \sum_{i=1}^{n} \left[\mathrm{Var}[X_i] + \mathbb{E}[X_i]^2\right] + \sum_{i=1}^{n}\sum_{j\neq i} \mathbb{E}[X_i X_j] && (13)\\
&= n[\sigma^2 + \mu^2] + \sum_{i=1}^{n}\sum_{j\neq i} \underline{\mathbb{E}[X_i X_j]} && (14)\\
&= n\sigma^2 + n\mu^2 + n(n-1)\left[\mu^2 - \frac{\sigma^2}{N-1}\right] && (15)\\
&= \frac{(N-1)n\sigma^2}{N-1} + n\mu^2 + n^2\mu^2 - n\mu^2 - \frac{n^2\sigma^2 + n\sigma^2}{N-1} && (16)\\
&= n^2\mu^2 + \frac{Nn\sigma^2 - n\sigma^2 - n^2\sigma^2 + n\sigma^2}{N-1} && (17)\\
&= n^2\mu^2 + \frac{n\sigma^2(N-n)}{N-1} && (18)
\end{aligned}
$$

Proof.

$$
\begin{aligned}
\mathrm{Var}[\bar{X}_n] &= \frac{1}{n^2}\mathrm{Var}(X_1 + \ldots + X_n) && (19)\\
&= \frac{1}{n^2}\left[\mathbb{E}[(X_1 + \ldots + X_n)^2] - \mathbb{E}[(X_1 + \ldots + X_n)]^2\right] && (20)\\
&= \frac{1}{n^2}\left[\underline{\mathbb{E}[(X_1 + \ldots + X_n)^2]} - (n\mu)^2\right] && (21)\\
&= \frac{1}{n^2}\left[n^2\mu^2 + \frac{n\sigma^2(N-n)}{N-1} - n^2\mu^2\right] && (22)\\
&= \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) && (23)
\end{aligned}
$$

□

$\frac{N-n}{N-1}$ is known as the *finite population correction.* As n approaches N, the variance of the sample mean approaches zero. Intuitively, this make sense as the sample becomes a whole population survey, making the population mean less devious ;).