

Sampling distributions

In the three forms of random sampling that we have discussed, $X_1, X_2, \dots, X_n$ are independent random variables with the same distribution. Suppose that the distribution has expectation μ and variance σ^2 .

We might be interested in:

- the sample total $T = X_1 + X_2 + \dots + X_n$;
- the sample mean $\bar{X} = \frac{T}{n}$;
- the sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right]$.

These are all random variables, so we write them with capital letters, and they each have an expectation and a variance.

Theorem 18 (a) $\mathbb{E}(T) = n\mu$.

(b) $\text{Var}(T) = n\sigma^2$.

(c) $\mathbb{E}(\bar{X}) = \mu$.

(d) $\text{Var}(\bar{X}) = \sigma^2/n$.

(e) $\mathbb{E}(S^2) = \sigma^2$.

Proof (a) $\mathbb{E}(T) = \mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n)$
 $= \mu + \mu + \dots + \mu = n\mu$.

$$\begin{aligned}
\text{(b)} \quad \text{Var}(T) &= \text{Var}(X_1 + X_2 + \cdots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n) \\
&= \sigma^2 + \sigma^2 + \cdots + \sigma^2 = n\sigma^2.
\end{aligned}$$

The middle step uses independence.

$$\text{(c)} \quad \mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{T}{n}\right) = \frac{1}{n} \mathbb{E}(T) = \frac{n\mu}{n} = \mu, \text{ using (a).}$$

$$\text{(d)} \quad \text{Var}(\bar{X}) = \text{Var}\left(\frac{T}{n}\right) = \frac{1}{n^2} \text{Var}(T) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}, \text{ using (b).}$$

$$\text{(e)} \quad \mathbb{E}(S^2) = \mathbb{E}\left[\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)\right] = \frac{1}{n-1} \left[\sum_{i=1}^n \mathbb{E}(X_i^2) - n\mathbb{E}(\bar{X}^2)\right]. \text{ But}$$

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

and

$$\mathbb{E}(\bar{X}^2) = \text{Var}(\bar{X}) + (\mathbb{E}(\bar{X}))^2 = \frac{\sigma^2}{n} + \mu^2,$$

from (c) and (d), so

$$\mathbb{E}(S^2) = \frac{1}{n-1} \left[n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \right] = \frac{1}{n-1} [(n-1)\sigma^2] = \sigma^2. \quad \blacksquare$$

This explains why we use the divisor $n-1$ when we calculate the sample variance.

In spite of this theorem, we do not know the whole of the distribution of T except in some special cases.

If X_i is Bernoulli then T is binomial.

If X_i is binomial then T is binomial.

If X_i is Poisson then T is Poisson.

If X_i is normal then T is normal.

If n is large, the Central Limit Theorem shows that the distributions of the sample total T and the sample mean $\bar{X}$ are both approximately normal. If we know the mean μ and the variance σ^2 , we can use standard normal tables to find out about the distribution of T and $\bar{X}$.