Exercise 29

Sample Variance

In the following X_1, ..., X_n are the square-integrable, pairwise uncorrelated and real-valued variables with variance σ^2 and mean μ.

---

a) Test whether the formula of the arithmetic mean (1) is an $\textbf{unbiased estimator}$ of the mean μ of the population. If the estimator is not unbiased, look for an appropriate correction.

For an unbiased estimator, the following statement has to be true: Expected value = true value. If this is not the case, the estimator is biased. In the following we will therefore check if $\langle \hat{\mu} \rangle = \mu$.

$\langle \hat{\mu} \rangle = \langle \bar{X} \rangle = \langle \frac{1}{n} \sum_{i=1}^{n} X_i \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle X_i \rangle$

$= \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n} \cdot n \cdot \mu = \mu$

This means, that the arithmetic mean is an unbiased estimator of the mean μ. It'd be very well biased, if it were tested for the variance for example.

---

b) The standard error of the arithmetic mean (1) is defined as the square root of the variance of $\bar{X}$. Show that (2) is true.

$\langle (\bar{X} - \mu)^2 \rangle = \langle \bar{X}^2 - 2\bar{X}\mu + \mu^2 \rangle = \langle \bar{X}^2 \rangle - 2\mu^2 + \mu^2 = \langle \bar{X}^2 \rangle - \mu^2 = \mathrm{Var}(\bar{X})$

I wasn't completely sure how to show the first equal-symbol in (2), since the left side is literally the definition of the Variance. I will continue on with showing the right side of the equation (which is most likely more interesting) 😃

The following rules for calculating with variances will be used:

$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$

$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$

$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$

Since the random variables are pairwise uncorrelated, the Covariance term is zero.

.

$= \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}$

---

c) Test whether the formula (3) is an unbiased estimator of the variance σ^2 of the population. If the estimator is not unbiased, look for an appropriate correction.

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

$= \frac{1}{n} \langle \sum_{i=1}^n X_i^2 - 2\mu X_i + \mu^2 \rangle = \frac{1}{n} \sum_{i=1}^n \langle X_i^2 \rangle - 2\mu \langle X_i \rangle + \mu^2$

$= \frac{1}{n} \sum_{i=1}^n \langle X_i^2 \rangle - 2\mu^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n \langle X_i^2 \rangle - \mu^2$

$= \frac{1}{n} \sum_{i=1}^n \mathrm{Var}(X_i) = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2$

The given formula is an unbiased estimator for the variance and hence needs no correction.

---

d) Usually the variance σ^2 of the population is unknown and the estimator (1) is used instead of μ. Then (3) becomes (4). Test whether (4) is an unbiased estimator of the variance σ^2 of the population. If the estimator is not unbiased, look for an appropriate correction. Hint: Expand the summand with $-\mu + \mu$ and use the given relation (2).

$\langle \hat{\sigma}^2 \rangle = \langle \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \rangle$

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

$= \frac{1}{n} \langle \sum_{i=1}^n \left((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2 \right) \rangle$

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

With use of (2):

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

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

$= \frac{1}{n} \sum_{i=1}^n \langle X_i - \mu \rangle^2 - n \langle \bar{X} - \mu \rangle^2$

With the definition of the variance which was derived earlier:

$= \frac{1}{n} \left(\mathrm{Var}(X_i) - n\mathrm{Var}(\bar{X}) \right)$

And (3):

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

And with that we get the correction term of $\frac{n}{(n-1)}\langle\hat{\sigma}^2\rangle = \sigma^2$