# §24. Sampling Distribution of the Sampling Variance

Let $X$ be a random variable and let $X_1, \ldots, X_n$ be a random sample from $X$.

**Definition 1**

A quantity computed from a sample is called a **statistic**.

**Example 1**

The sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is a statistic.

So is the sample variance

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

**Remark**
The statistics are r.v. themselves $\because$ they depend on the random sample and their distributions are called **sampling distributions**.

**Example 2**

$p(\overline{x})$ sampling distribution of the mean.

$p(s^2)$ sampling distribution of the sampling variance.

To keep the discussion reasonably simple we will only consider the case of a sample $X_1, \ldots, X_n$ from a Gaussian population $X \sim N(\mu, \sigma)$

Let's find out what is the distribution of $s^2$.

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \overline{x})^2 &= \sum_{i=1}^{n}\left[(x_i - \mu) - (\overline{x} - \mu)\right]^2 \\ &= \sum_{i=1}^{n}(x_i - \mu)^2 - 2\sum_{i=1}^{n}(x_i - \mu)(\overline{x} - \mu) + \sum_{i=1}^{n}(\overline{x} - \mu)^2\end{aligned}$$

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

$$\Rightarrow \quad \frac{(n-1)s^2}{\sigma^2} + \frac{(\bar{x} - \mu)^2}{\sigma^2/n} = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2}$$

It can be shown that in a sampling distribution from a Gaussian $\overline{x}$ and $s^2$ are independent.

$$M\frac{(n-1)s^2}{\sigma^2}(t) \cdot \underbrace{M\frac{(\bar{x}-\mu)^2}{\sigma^2/n}(t)}_{\substack{\text{Square of} N(0,1) \\ \chi^2(1)}} = \underbrace{M\sum_{i=1}^{u}\frac{(x_i-\mu)^2}{\sigma^2}(t)}_{\substack{\text{Sum of n squares} \\ N(0,1) \sim \chi^2(n)}}$$

$$M\frac{(n-1)s^2}{\sigma^2}(t) \cdot (1-2t)^{-1/2} = (1-2t)^{-n/2}$$
$$M\frac{(n-1)s^2(t)}{\sigma^2} = (1-2t)^{-\frac{n-1}{2}}$$

$$M\frac{(n-1)s^2(t)}{\sigma^2} = (1-2t)^{-\frac{n-1}{2}} \quad \Rightarrow \quad \boxed{\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)}$$

Since the expected value of $\chi^2(k)$ random variable is $k$, we have

$$E\left(s^2\right) = \frac{n-1}{n-1}\sigma^2 \Rightarrow E\left(s^2\right) = \sigma^2$$

Thus with

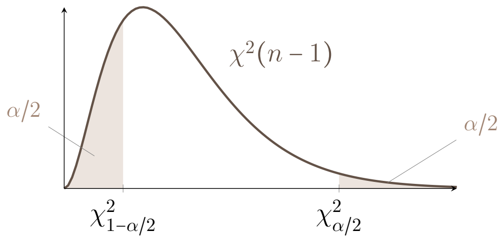
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$s^2$ is an unbiased estimation for $\sigma^2$

The variance of $\chi^2(k)$ is $2k$, so we have

$$\text{Var}\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{(n-1)^2}{\sigma^4}\text{Var}\left(s^2\right) = 2(n-1) \quad \Rightarrow \quad \text{Var}\left(s^2\right) = \frac{2\sigma^4}{n-1}$$

## Confidence Interval of the Variance and Standard Deviation of a Gaussian Random Variable




$$P\left(\chi^2_{1-\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1-\alpha$$

↗ $1-\alpha$ of all samples have sample variance not in the $\alpha/2$ tails

$$\Rightarrow P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$$

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

$100(1-\alpha)\%$ CI for the variance of normal

**Example 3**

Data on pH of rain in Michigan.

5.47 5.37 5.38 4.63 5.37 3.74 3.71 4.96 4.64 5.11
5.65

Assuming that the pH is Gaussian compute a 95% CI for the rain pH variance and for the standard deviation.

Using Open Office: $s^2 = 0.4508, n-1 = 10$ d.f.

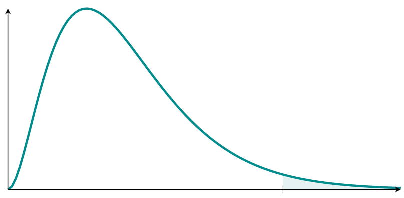
$\chi^2_{\alpha/2} = 20.48,$ $\chi^2_{1-\alpha/2} = 3.25$

$$\frac{(10)\cdot(0.4508)}{20.48} \le \sigma^2 \le \frac{10(0.4508)}{3.25}$$

$$0.2201 \le \sigma^2 \le 1.3871 \quad \text{with 95\% confidence}$$
$$0.4691 \le \sigma^2 \le 1.1777 \quad \text{with 95\% confidence}$$

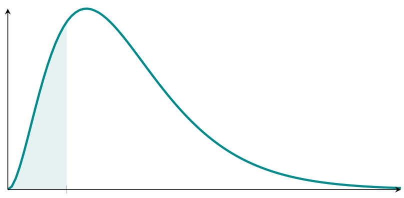
## Lower and Upper Confidence Bounds

The $100(1-\alpha)\%$ lower and upper confidence bounds on $\sigma^2$ are




$$\frac{(n-1)s^2}{\chi^2_\alpha} \leq \sigma^2$$

**Remark** Lower Confidence Bound




$$\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha}}$$

**Remark** Upper Confidence Bound

**Example 4**

An automatic filling machine fills bottles with beer. A random sample of 20 bottles has a sample variance of the fill volume of $s^2 = 2.957\, ml^2$. If the variance of the fill volume is too high an acceptable portion of the bottles will be under-or over filled. Assuming that the fill volume is normal construct a 95% upper confidence bound for $\sigma$.

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.95}}$$

$$\sigma^2 \leq \frac{19(2.957)}{10\cdot 117} \quad \Rightarrow \quad \sigma^2 \leq 5.5533 \quad \text{with 95\% confidence}$$

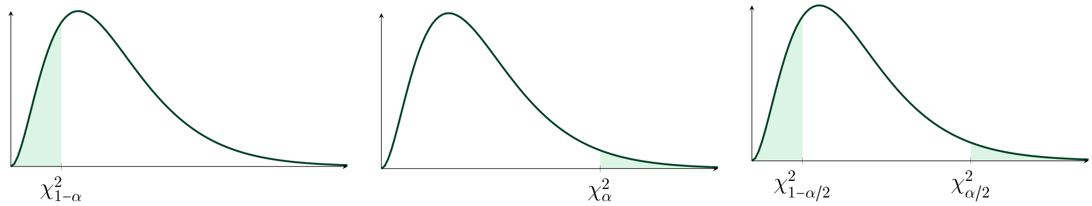
$$\Rightarrow \quad \sigma \leq 2.3565 \text{ ml} \quad \text{with 95\% confidence}$$

## Hypothesis Testing for the Variance and Standard Deviation of Gaussian

Let $X$ be a normal random variable and let $X_1, \ldots, X_n$ be a random sample from $X$. Let $s^2$ be the sample variance.

We would like to test the hypothesis that $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$ or $\sigma^2 < \sigma_0^2$ or $\sigma^2 \neq \sigma_0^2$ at significance level $\alpha$ (probability for Type I error).

The critical (rejection) regions are:




$H_1: \ \sigma^2 < \sigma_0^2$ $\qquad$ $H_1: \ \sigma^2 > \sigma_0^2$ $\qquad$ $H_1: \ \sigma^2 \neq \sigma_0^2$

with test statistic

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

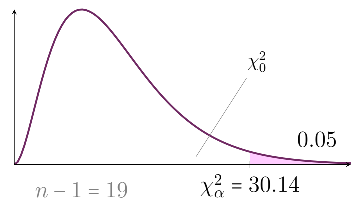
**Example 5**

For the beer bottle example test $H_0 : \sigma^2 = 2$ versus $H_1 : \sigma^2 > 2$ at the 5% level of significance.

$$\chi_0^2 = \frac{(19)(2.957)}{2} = 28.09 < 30.14 = x_\alpha^2$$

Fail to reject $H_0$. This sample does not provide enough evidence that the fill volume variance exceeds 2 mL$^2$.

The $p$−value for the test is $P\left(\chi^2 > \chi_0^2\right) = 0.0817$. The sample is not unlikely enough to contradict H$_0$.




**Example 6**

The sugar content of the syrup in canned peaches is normal and the variance is thought to be $\sigma^2 = 18$ mg$^2$. A sample of a dozen cans gives $s^2 = 6.7$ mg$^2$. Test $H_0 : \sigma^2 = 18$ vs. $H_1 : \sigma^2 < 18$ at $\alpha = 0.05$ level of significance. What is the $p$−value of this test.

$$\chi_0^2 = \frac{(4-1)s^2}{\sigma_0^2} = \frac{11(6.7)}{18} = 4.0944 < \chi_{1-\alpha}$$

Reject $H_0$. Accept $H_1$. The variance of the sugar content of canned peaches is less that 18 mg$^2$.

$$p-\text{value} \ = 0.0329 < 0.05 = \alpha$$

The sample contradicts $H_0$ strongly enough.