Sampling distribution of variance

Consider experiment of rolling two fair dice

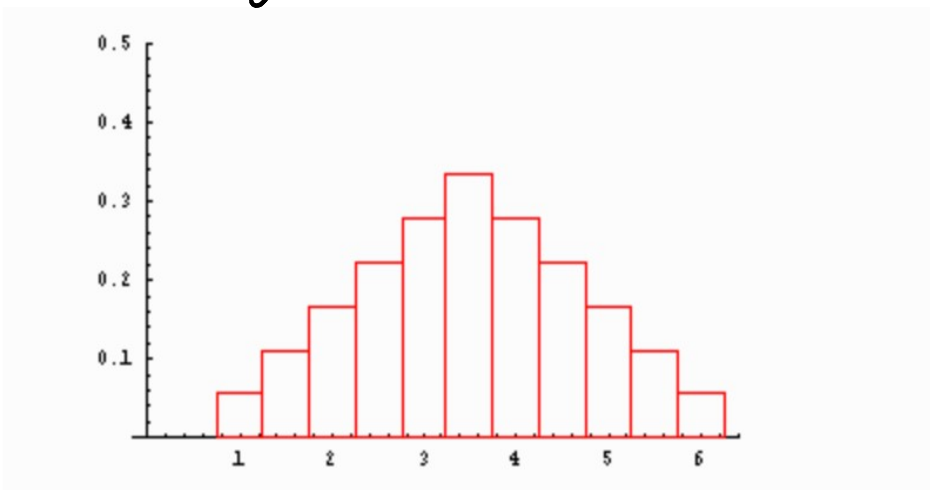
=) 36 possible outcomes (1,2,...,6 for each dice).

each pair outcome has probability 1/36.

	1	2	3	4	5	6
1	1,1 mean=1 stdev=0 var=0	1,2 mean=1.5 stdev=0.71 var=0.504	1,3 mean=2 stdev=1.41 var=1.99	1,4 mean=2.5 stdev=2.12 var=4.49	1,5 mean=3 stdev=2.83 var=8.01	1,6 mean=3.5 stdev=3.54 var=12.53
2	2,1 mean=1.5 stdev=0.71 var=0.504	2,2 mean=2 stdev=0 var=0	2,3 mean=2.5 stdev=0.71 var=0.504	2,4 mean=3 stdev=1.41 var=1.99	2,5 mean=3.5 stdev=2.12 var=4.49	2,6 mean=4 stdev=2.83 var=8.01
3	3,1 mean=2 stdev=1.41 var=1.99	3,2 mean=2.5 stdev=0.71 var=0.504	3,3 mean=3 stdev=0 var=0	3,4 mean=3.5 stdev=0.71 var=0.504	3,5 mean=4 stdev=1.41 var=1.99	3,6 mean=4.5 stdev=2.12 var=4.49
4	4,1 mean=2.5 stdev=2.12 var=4.49	4,2 mean=3 stdev=1.41 var=1.99	4,3 mean=3.5 stdev=0.71 var=0.504	4,4 mean=4 stdev=0 var=0	4,5 mean=4.5 stdev=0.71 var=0.504	4,6 mean=5 stdev=1.41 var=1.99
5	5,1 mean=3 stdev=2.83 var=8.01	5,2 mean=3.5 stdev=2.12 var=4.49	5,3 mean=5 stdev=1.41 var=1.99	5,4 mean=4.5 stdev=0.71 var=0.504	5,5 mean=5 stdev=0 var=0	5,6 mean=5.5 stdev=0.71 var=0.504
6	6,1 mean=3.5 stdev=3.54 var=12.53	6,2 mean=4 stdev=2.83 var=8.01	6,3 mean=4.5 stdev=2.12 var=4.49	6,4 mean=5 stdev=1.41 var=1.99	6,5 mean=5.5 stdev=0.71 var=0.504	6,6 mean=6 stdev=0 var=0

Sample Mean	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

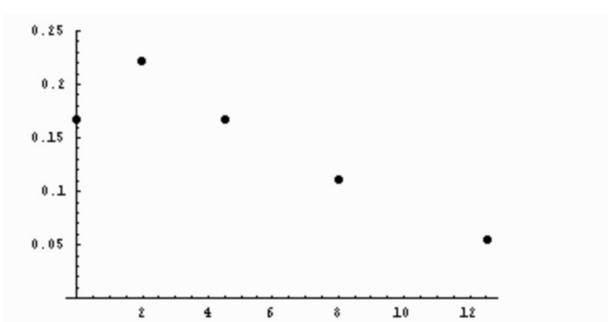
If we plot the distribution of the sample means, they will be approximately normal:

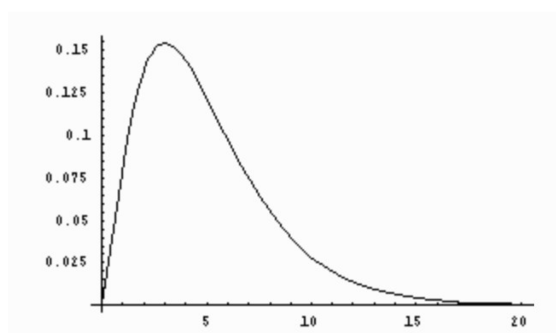


(consequence of CLT : $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$).

Sample Variance	0	0.504	1.99	4.49	8.01	12.53
Probability	6/36	10/36	8/36	6/36	4/36	2/36

On the other hand the distribution of variances is <u>not</u> approx. normal:





Instead it is close to a chi-squared (χ^2) distribution. Notice y-axis values different on the plots.

$$\frac{(n-1)}{\sigma^2} s^2 \sim \chi^2_{n-1}$$ (chi-square with n-1 degrees of freedom)

That is, sample variance times a constant $\frac{(n-1)}{\sigma^2}$ has a χ^2_{n-1} distribution.

It turns out that for independent $z_1, \ldots, z_{n-1}$:

$$Y = \sum_{i=1}^{n-1} z_i^2 = z_1^2 + z_2^2 + \ldots + z_{n-1}^2$$

where $z_i \sim N(0,1)$

$$\Rightarrow Y \sim \chi^2_{n-1}$$

$$m[Y] = m[z_1^2] + \ldots + m[z_{n-1}^2]$$

$$= 1 + \ldots + 1 = \underline{n-1}$$

$$\sigma^2[Y] = E(Y^2) - [E(Y)]^2 = \underline{2(n-1)}$$

$$P\left[\chi^2_{1-\alpha/2,\,n-1} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\alpha/2,\,n-1}\right]$$

$$= 1-\alpha$$

$$\Rightarrow P\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}}\right]$$

$$= 1-\alpha$$

If s^2 is the variance of a random sample of size n from a normal population, then:

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

is a $(1-\alpha)\times 100\%$ confidence interval for σ^2.