Norman Rzepka, 754644
Markus Dücker, 779867

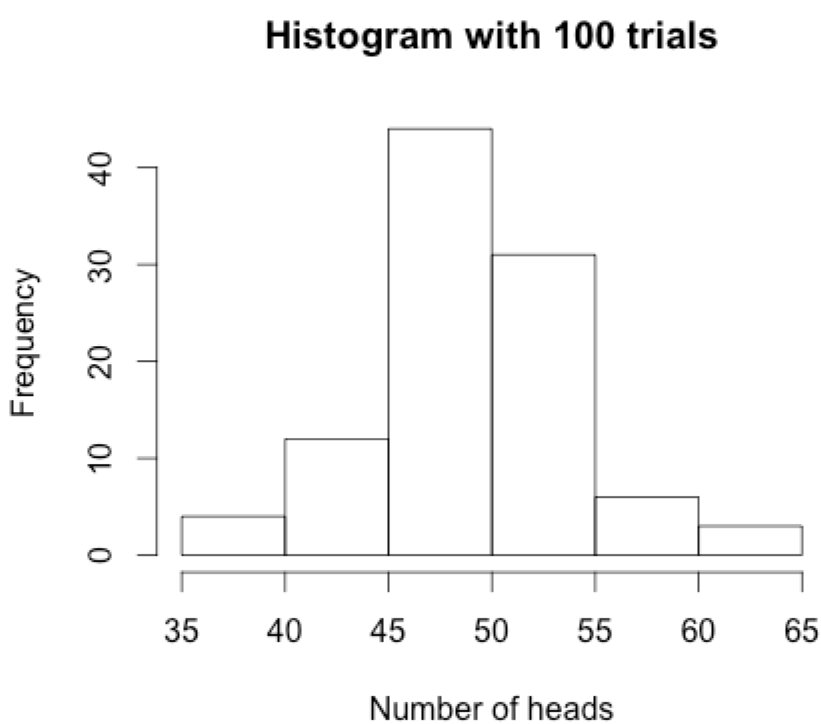
Week 2

Ass. 1

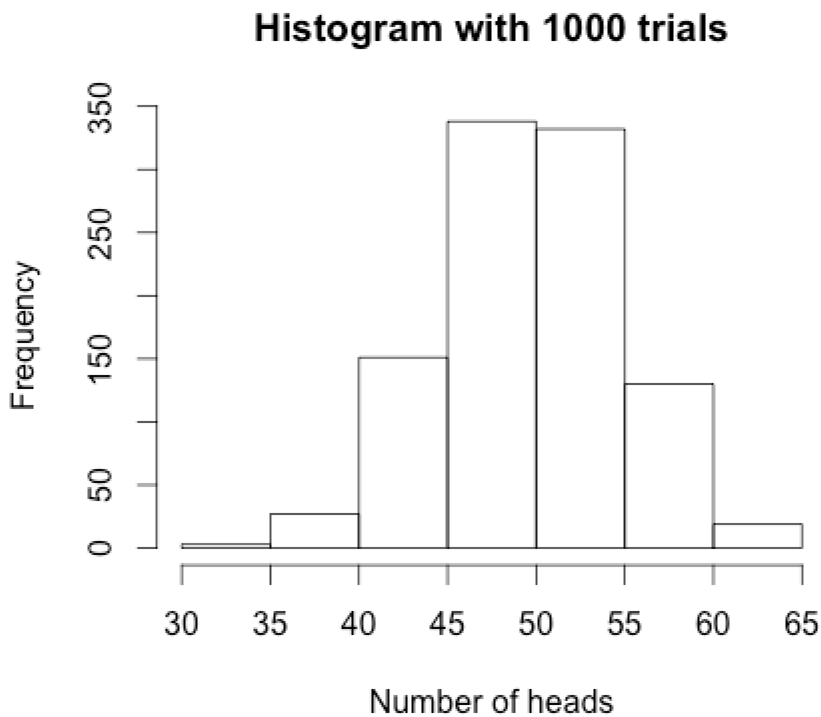
empVar(X)	43440685.140496
sampVar(X)	47784753.654545
var(X)	47784753.654545

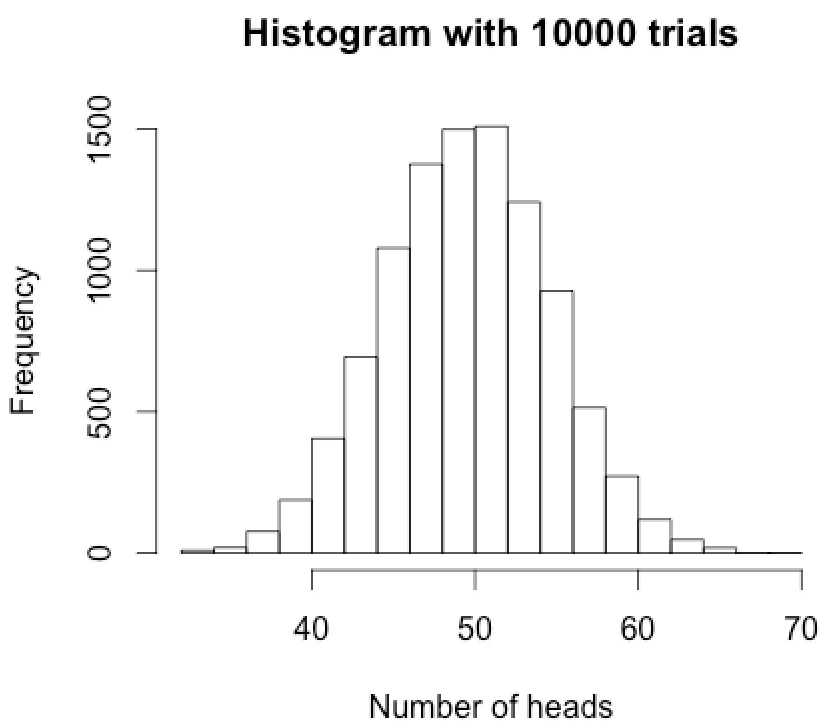
Based on our observations, var(X) calculates the sample variance. The empirical variance is lower than the sample variance.

Ass. 2a



Norman Rzepka, 754644
Markus Dücker, 779867





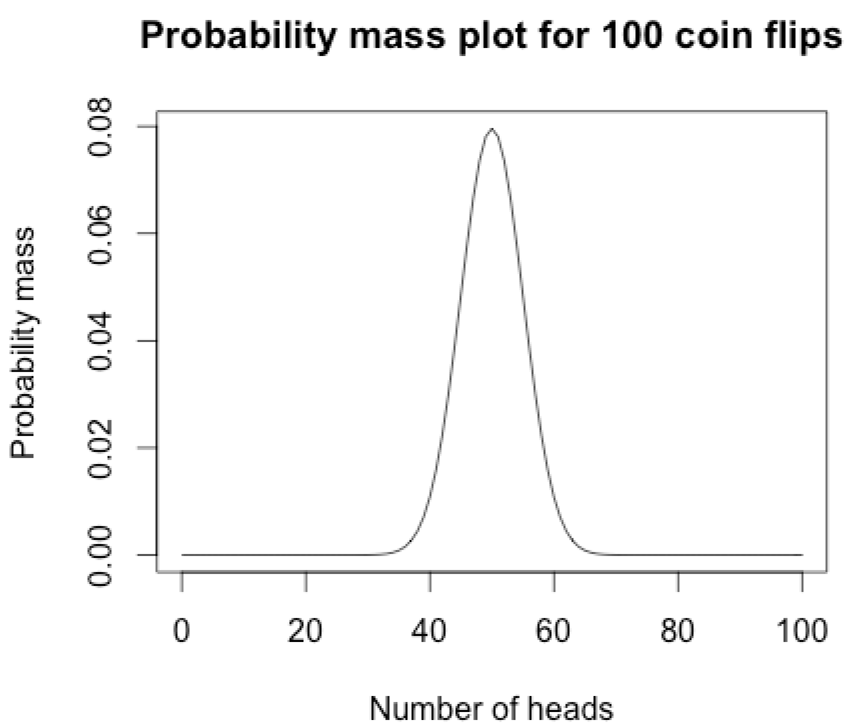
Norman Rzepka, 754644
Markus Dücker, 779867

Ass. 2b

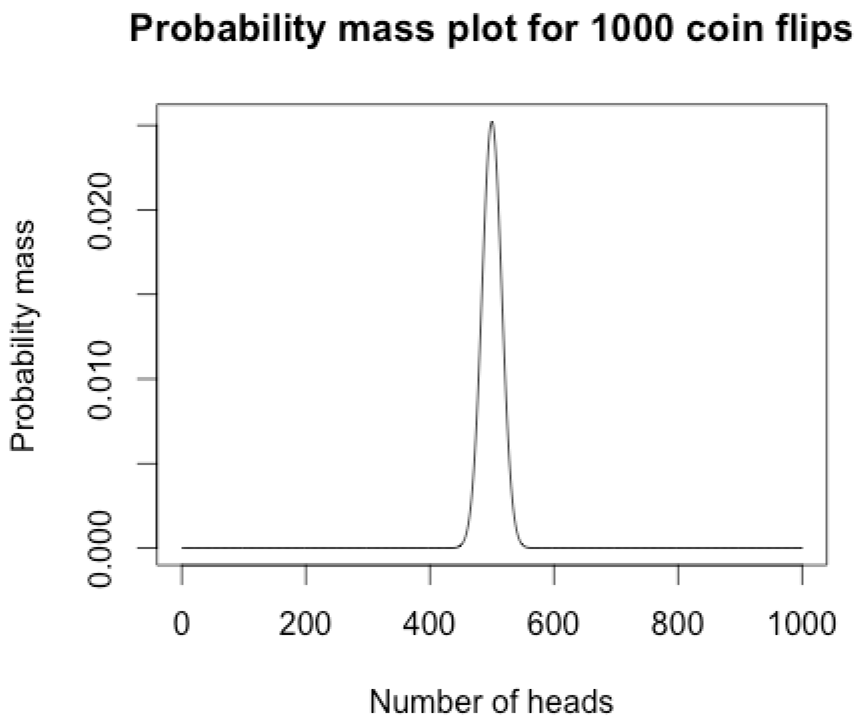
trials	mean	median	empVar	sampVar
100	49.590000	50	28.869600	29.161212
1000	50.055000	50	24.216839	24.241080
10000	49.978400	50	25.156792	25.159308

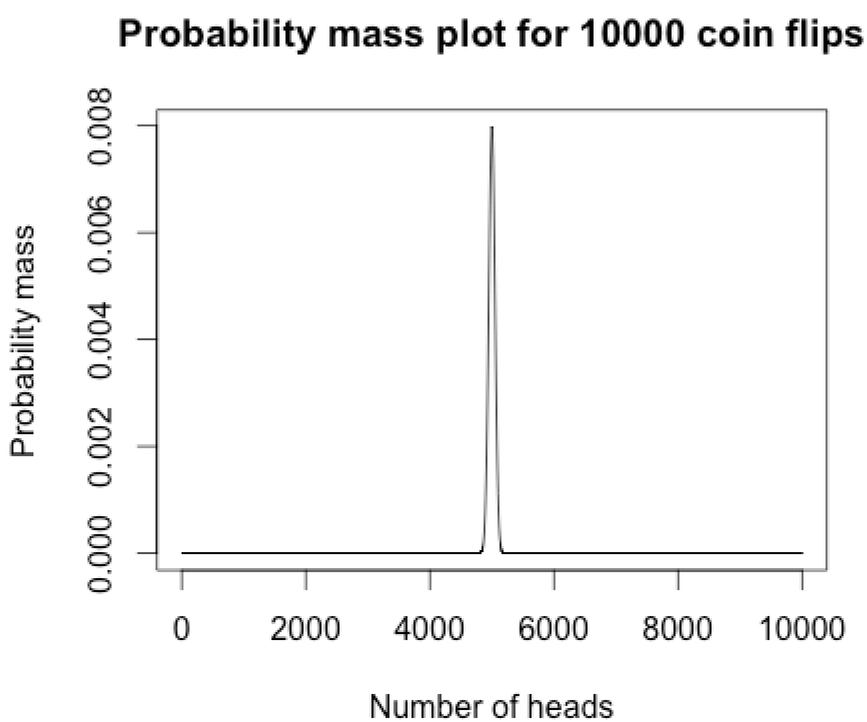
We can observe that all experiments lead to a median of 50 and a mean very close to 50. The mean gets closer to 50 as we increase the number of trials. With increasing number of trials the variances decrease. We can also observe that the difference between empVar and sampVar decreases with increased number of trials.

Ass. 2c



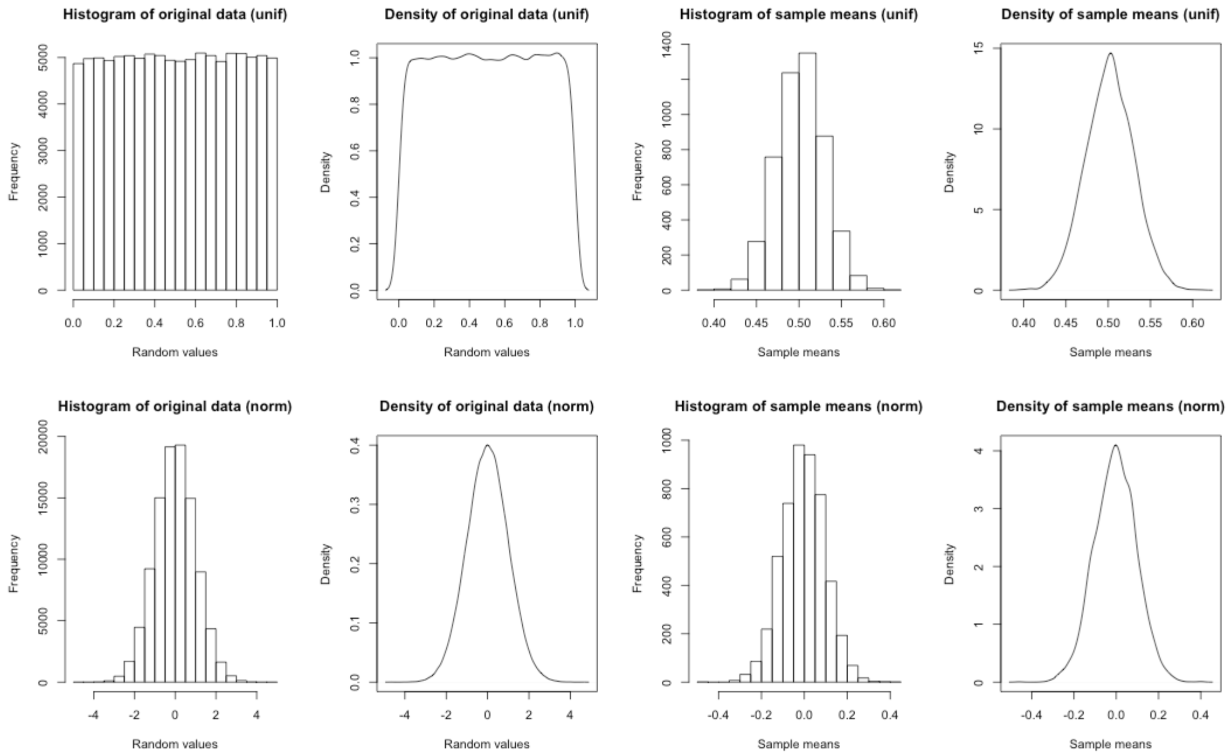
Norman Rzepka, 754644
Markus Dücker, 779867





Norman Rzepka, 754644
Markus Dücker, 779867

Ass. 3



Ass. 4

a)

$H_0 : \mu_A = \mu_B$

b)

Two-sided test is more appropriate because we're dealing with equality in our hypotheses and therefore need to care about both ends of the sampling distribution.

c)

Using a confidence level of 0.95 means that we'll reject hypotheses if the drawn sample is not within the bounds of $[\mu - 1.96\sigma, \mu + 1.96\sigma]$ of the sampling distribution.

d)

$\overline{A} = 0.8954545,\ \overline{B} = 0.8909091,\ df = 10$

e)

$\overline{d} = \overline{A} - \overline{B} = 0.004545455$

Norman Rzepka, 754644
Markus Dücker, 779867

$s_d = \sqrt{\frac{s_A^2}{11} + \frac{s_B^2}{11}} = 0.01444854$
$Y = \frac{\bar{d}\sqrt{k}}{s_d} = \frac{0.004545455 \cdot 3.316625}{0.01444854} = 1.043397$

f)
$t_{10,0.025} = 2.228$
$1.043397 > 2.228$ is not correct.
Therefore the hypothesis H_0 is not rejected and both algorithms are not considered significantly different.

Ass. 5

$H_0 : \mu = 2$
$\sigma = 0.1$
$\overline{x} = 2.012$
$n = 200$

$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = 0.00707,\ \ z = \frac{\mu - \overline{x}}{\sigma_{\overline{x}}} = 1.697$
$P(z < 1.697) = 0.9545 \approx 95\%$

Assuming a significance level of 0.05 the presented information supports the hypothesis that $\mu = 2$.

Ass. 6

- Ease of implementation: If an algorithm is easier to implement the developer is less likely to introduce bugs. This is true for initial implementation and further maintenance. Thus the easier algorithm is favorable.
- Runtime metrics: The algorithm that requires less time to compute a result is favorable because it enables faster feedback loops.
- Community support: An algorithm that creates more interest in the community is more likely to get further improved. Our own implementation may benefit from the improvements other community members propose.

Ass. 7

a. Regular dice: 2.584963
b. Manipulated dice: 2.089861