Exercise F.3

We are given an urn initially filled with n black balls and n white balls, and we draw two balls from the urn, without replacement. Let U be the number of black balls on the first draw, and V be the number of black balls on the second draw. Since the draws are without replacement, U and V are dependent; for instance, if $n = 1$ then the value of V can be determined if the value of U is known.

Properties of U and V can be determined by examining the possible outcomes of the first two draws and their probabilities:

$$\begin{aligned}
\Pr\{BB\} &= \frac{n}{2n} \cdot \frac{n-1}{2n-1} = \frac{n-1}{4n-2} \\
\Pr\{BW\} &= \frac{n}{2n} \cdot \frac{n}{2n-1} = \frac{n}{4n-2} \\
\Pr\{WB\} &= \frac{n}{2n} \cdot \frac{n}{2n-1} = \frac{n}{4n-2} \\
\Pr\{WW\} &= \frac{n}{2n} \cdot \frac{n-1}{2n-1} = \frac{n-1}{4n-2}
\end{aligned}$$

As a consistency check, note that these probabilities sum to one.

(a) What is ρ_{UV}?

From definition F.3,

$$\rho_{UV} = \frac{\gamma_{UV}}{\sigma_U \sigma_V}$$

and using the alternate definition of γ_{UV} we have

$$\gamma_{UV} = \mathrm{E}[UV] - \mu_U \mu_V.$$

Thus we need to determine the mean and standard deviation of U and V, and the mean of UV.

Using definition 6.1.4, we can compute mean and variance of U and V using the probabilities listed above:

$$\begin{aligned}
\mu_U &= 1 \cdot (\Pr\{BB\} + \Pr\{BW\}) + 0 \cdot (\Pr\{WB\} + \Pr\{WW\}) &= \cdots = \tfrac{1}{2} \\
\mu_V &= 1 \cdot (\Pr\{BB\} + \Pr\{WB\}) + 0 \cdot (\Pr\{BW\} + \Pr\{WW\}) &= \cdots = \tfrac{1}{2} \\
\sigma_U^2 &= 1^2 \cdot (\Pr\{BB\} + \Pr\{BW\}) + 0^2 \cdot (\Pr\{WB\} + \Pr\{WW\}) - \mu_U^2 &= \cdots = \tfrac{1}{4} \\
\sigma_V^2 &= 1^2 \cdot (\Pr\{BB\} + \Pr\{WB\}) + 0^2 \cdot (\Pr\{BW\} + \Pr\{WW\}) - \mu_V^2 &= \cdots = \tfrac{1}{4}
\end{aligned}$$

Similarly, we can compute $\mathrm{E}[UV]$:

$$\mathrm{E}[UV] = 1 \cdot 1 \cdot \Pr\{BB\} + 1 \cdot 0 \cdot \Pr\{BW\} + 0 \cdot 1 \cdot \Pr\{WB\} + 0 \cdot 0 \cdot \Pr\{WW\} = \frac{n-1}{4n-2}$$

We then obtain

$$\begin{aligned}
\gamma_{UV} &= \mathrm{E}[UV] - \mu_U \mu_V \\
&= \frac{n-1}{4n-2} - \frac{1}{2} \cdot \frac{1}{2} \\
&\vdots \\
&= -\frac{1}{8n-4}
\end{aligned}$$

Finally, we obtain

$$\begin{aligned}\rho_{UV} &= \frac{\gamma_{UV}}{\sigma_U \sigma_V} \\ &= \frac{\frac{-1}{8n-4}}{\frac{1}{2}\cdot\frac{1}{2}} \\ &\vdots \\ &= -\frac{1}{2n-1}\end{aligned}$$

(b) Why does $\rho_{UV} \to 0$ as $n \to \infty$?

As the urn becomes more full, the proportion of black balls is less affected by the first draw. More formally, the probability that the second draw is black becomes less dependent on the outcome of the first draw. Indeed, the difference of these probabilities $\Pr\{WB\} - \Pr\{BB\} = \cdots = \frac{1}{2n-1}$ tends to zero as n approaches infinity. Since V becomes less dependent on U as n increases, and $\rho_{UV} = 0$ when U and V are independent, we should have $\rho_{UV} \to 0$.

(c) Verify using Monte Carlo simulation for the case $n = 5$

We can simulate draws from the urn by keeping track of the number of black and white balls in the urn, and generating the appropriate Bernoulli random variates:

```
black = n;
white = n;
U = Bernoulli( black / (black+white) );
if (U) black--; else white--;
V = Bernoulli( black / (black+white) );
```

The above experiment is replicated as many times as needed. We can measure μ_U, μ_V, σ_U, σ_V, and $\mathrm{E}[UV]$ using Welford's one-pass algorithm (algorithm 4.1.1). ρ_{UV} can be computed from these values using the same equations as above.

The theoretical values and the simulation results for $n = 5$ are shown in the table below, for $1,000,000$ replications and an intial seed of 123456789.

	Theory	Simulation
μ_U	0.500000	0.500407
μ_V	0.500000	0.500246
σ_U	0.500000	0.500000
σ_V	0.500000	0.500000
$\mathrm{E}[UV]$	0.222222	0.222772
γ_{UV}	-0.027777	-0.027555
ρ_{UV}	-0.111111	-0.110218