The Variance of R_n:
We follow similar steps to find the variance of R_n.

$$Var[R_n] = Var\left[1 + \sum_{i=2}^{n-1} I_i\right] = Var\left[\sum_{i=2}^{n-1} I_i\right] \tag{1.3}$$

Note that we can not just run the variance operator through the sum as we did with the expectation operator above. This would only be valid if the indicator variables are independent. Just because we are assuming that, for $i \neq j$, X_i and X_j are independent of one another does not necessarily imply that I_i and I_j are independent of one another. In fact, for certain relationships between i and j, they are most certainly not! The problem is that I_i is a function of X_{i-1}, X_i, and X_{i+1}. Similarly, I_{i+1}, for example, is a function of X_i, X_{i+1}, and X_{i+2}. There is a relationship between I_i and I_{i+1} caused by the overlapping indices in the X's.
Using the covariance operator

$$Cov(X,Y) := \mathsf{E}\left[(X - \mathsf{E}[X])(Y - \mathsf{E}[Y])\right] = \mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]$$

and the facts (which may be verified from the definition of covariance) that

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = Var(X)$
- $$\begin{array}{rcl} Cov(X_1 + X_2, Y_1 + Y_2) & = & Cov(X_1,Y_1) + Cov(X_1,Y_2) \\ & & +Cov(X_2,Y_1) + Cov(X_2,Y_2) \end{array},$$

we may write (1.3) as

$$Var[R_n] = \sum_{i=2}^{n-1} \sum_{j=2}^{n-1} Cov(I_i, I_j) \tag{1.4}$$

Now,

$$\mathsf{E}[I_i^2] = (1)^2 \cdot P(I_i = 1) + (0)^2 \cdot P(I_i = 1) = 1 \cdot P(X_i \text{ starts a run }) = \tfrac{2}{3}.$$

So,

$$Cov(I_i, I_i) = Var[I_i] = \mathsf{E}[I_i^2] - (\mathsf{E}[I_i])^2 = \frac{2}{3} - \left(\frac{2}{3}\right)^2 = \frac{2}{9}.$$

For $i \neq j$, we write

$$\begin{aligned}
\mathsf{E}[I_i I_j] &= P(X_i \text{ starts a run and } X_j \text{ starts a run}) \\
&= P(X_{i-1} < X_i > X_{i+1} \text{ and } X_{j-1} < X_j > X_{j+1}) \\
&\quad + P(X_{i-1} < X_i > X_{i+1} \text{ and } X_{j-1} > X_j < X_{j+1}) \\
&\quad + P(X_{i-1} > X_i < X_{i+1} \text{ and } X_{j-1} < X_j > X_{j+1}) \\
&\quad + P(X_{i-1} > X_i < X_{i+1} \text{ and } X_{j-1} > X_j < X_{j+1}) \\
&\overset{Why?}{=} 2 \cdot P(X_{i-1} < X_i > X_{i+1} \text{ and } X_{j-1} < X_j > X_{j+1}) \\
&\quad + 2 \cdot P(X_{i-1} < X_i > X_{i+1} \text{ and } X_{j-1} > X_j < X_{j+1})
\end{aligned}$$

We will illustrate the computation of these two probabilities in the case where $|i-j| = 1$ (or, more simply, we will consider $j = i+1$ since $Cov(I_i, I_j) = Cov(I_j, I_i)$), and leave the other similar computations as an exercise.

$$P(X_{i-1} < X_i > X_{i+1} \text{ and } X_{j-1} < X_j > X_{j+1}) \overset{j=i}{=} P(X_{i-1} < X_i > X_{i+1} \text{ and } X_i < X_{i+1} > X_{i+2}$$

since X_i can not be simultaneously less than and greater than X_{i+1}.

$$\begin{aligned}
P(X_{i-1} < X_i > X_{i+1} \text{ and } X_{j-1} > X_j < X_{j+1}) &\overset{j=i}{=} P(X_{i-1} < X_i > X_{i+1} < X_{i+2}) \\
&= \int_{-\infty}^{\infty} P(X_{i-1} < X_i > X_{i+1} < X_{i+2} | X_i = x) \cdot f(x)\, dx \\
&= \int_{-\infty}^{\infty} P(X_{i-1} < x > X_{i+1} < X_{i+2} | X_i = x) \cdot f(x)\, dx \\
&\overset{indep.}{=} \int_{-\infty}^{\infty} P(X_{i-1} < x > X_{i+1} < X_{i+2}) \cdot f(x)\, dx \\
&\overset{indep.}{=} \int_{-\infty}^{\infty} P(X_{i-1} < x) \cdot P(x > X_{i+1} < X_{i+2}) \cdot f(x)\, dx
\end{aligned}$$

Similarly,

$$
\begin{aligned}
P(x > X_{i+1} < X_{i+2}) &= \textstyle\int_{-\infty}^{\infty} P(x > X_{i+1} < X_{i+2} | X_{i+1} = y) \cdot f(y)\, dy \\
&= \textstyle\int_{-\infty}^{\infty} P(x > y < X_{i+2} | X_{i+1} = y) \cdot f(y)\, dy \\
&\overset{indep.}{=} \textstyle\int_{-\infty}^{\infty} P(x > y < X_{i+2}) \cdot f(y)\, dy \\
&= \textstyle\int_{-\infty}^{x} P(y < X_{i+2}) \cdot f(y)\, dy \\
&= \textstyle\int_{-\infty}^{x} [1 - F(y)] \cdot f(y)\, dy \\
&\overset{u=F(y)}{=} \textstyle\int_{0}^{F(x)} [1 - u]\, du = F(x)\left[1 - \frac{1}{2}F(x)\right]
\end{aligned}
$$

So, we have

$$
\begin{aligned}
P(X_{i-1} < X_i > X_{i+1} &\quad \text{and } X_{j-1} > X_j < X_{j+1}) \\
&= \textstyle\int_{-\infty}^{\infty} P(X_{i-1} < x) \cdot P(x > X_{i+1} < X_{i+2}) \cdot f(x)\, dx \\
&= \textstyle\int_{-\infty}^{\infty} F(x) \cdot F(x) \left[1 - \frac{1}{2}F(x)\right] \cdot f(x)\, dx \\
&\overset{u=F(x)}{=} \textstyle\int_{0}^{1} u^2 \cdot \left[1 - \frac{1}{2}u\right]\, du = \frac{5}{24}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
Cov(I_i, I_{i+1}) &= \mathsf{E}[I_i I_{i+1}] - \mathsf{E}[I_i]\mathsf{E}[I_{i+1}] \\
&= (2)(0) + (2)\left(\tfrac{5}{24}\right) - \left(\tfrac{2}{3}\right)\left(\tfrac{2}{3}\right) \\
&= -\tfrac{1}{36}
\end{aligned}
$$

Similar computations show that

$$Cov(I_i, I_j) = \frac{1}{180} \qquad \text{for } |i - j| = 2$$

and

$$Cov(I_i, I_j) = 0 \qquad \text{for } |i - j| > 2.$$

Hence, (**??**) becomes

$$\begin{aligned} Var[R_n] &= \textstyle\sum_{i=2}^{n-1}\sum_{j=2}^{n-1} Cov(I_i, I_j) \\ &= (n-2)\cdot Var(I_2) + 2\cdot(n-3)\cdot Cov(I_2, I_3) + 2\cdot(n-4)\cdot Var(I_2, I_4) \\ &= (n-2)\cdot\left(\tfrac{2}{9}\right) + 2\cdot(n-3)\cdot\left(-\tfrac{1}{36}\right) + 2\cdot(n-4)\cdot\left(\tfrac{1}{180}\right) \\ &= \tfrac{16n-29}{90}. \end{aligned}$$

For large n, we now know that

$$R_n \overset{approx.}{\sim} N\left(\frac{2n-3}{3}, \frac{16n-29}{90}\right)$$

and that

$$Z_n := \frac{R_n - \mu_{R_n}}{\sigma_{R_n}}$$

where $\mu_{R_n} := \frac{2n-3}{3}$ and $\sigma^2_{R_n} := \frac{16n-29}{90}$.

Exercises for Section 1.3.2

1.3.1 In this section, we used the fact that if $X_1, X_2, \ldots, X_n$ is a random sample from a distribution, then

$$P(X_j > X_{j-1}, X_j > X_{j+1}) = P(X_j < X_{j-1}, X_j < X_{j+1}).$$

(a) Explain why this is true "in words".

(b) Now prove it!

1.3.2 For the "Runs Up and Runs Down Test", provide the details for the computation of $Var[R_n]$ by showing the following.

(a) $Cov(I_i, I_j) = \frac{1}{180}$ for $|i-j| = 2$

(b) $Cov(I_i, I_j) = 0$ for $|i-j| > 2$

(c)

$$\sum_{i=2}^{n-1}\sum_{j=2}^{n-1} Cov(I_i, I_j) = (n-2)\cdot Var(I_2) + 2\cdot(n-3)\cdot Cov(I_2, I_3) + 2\cdot(n-4)\cdot Var(I_2, I_4)$$