1.3 Sum of discrete random variables

Let X and Y represent independent Bernoulli distributed random variables $B(p)$.

- Find the distribution of their sum
- What is the distribution of the sum of n independent identically distributed (i.i.d.) Bernoulli random variables?

Understanding the ideas in R:

- Use the function `sample` to generate 100 realizations of two Bernoulli variables and check the distribution of their sum.

1.4 Sum of continuous random variables

While individual values give some indication of blood manipulations, it would be interesting to also check a sequence of values through the whole season. We wish to look at the distribution of the sum of squared standardized departures from the mean value (under the null hypothesis that the athlete is not doped). Let Z denote the standardized difference from the mean value (we assume it is normally distributed), we wish to know the sum $\sum Z^2$ (we square the values, since departures in both directions are of interest). We assume that the measurements were made in intervals long enough to ensure independence between them.

- Find the formula for the distribution of the sum of two independent continuous variables ($Z = X + Y$), compare it with the formula in the discrete case
- Find the distribution of the sum $S = Z_1^2 + Z_2^2$, if Z_1 and Z_2 are standard normal variables?
  Hint: Use the equality
  $$\int_0^1 \frac{1}{\sqrt{(1-x)x}} dx = \pi$$
- Say that we get the following five standardized values for a certain athlete (Z values): $1.6, 1.5, -1.6, 1.8, 1.4$. What can we conclude about the athlete?

Understanding the ideas in R:

- Generate 10 values from the $X \sim N(148, 85)$ distribution, these values represent 10 doping test in one athlete. Standardize the values to get a $N(0,1)$ variable, square them and sum. This is the value of the first athlete, generate values for 1000 athletes in the same way. Plot a histogram of the values. Use the function `pgamma` to find the limit that is exceeded by the $\Gamma(\frac{10}{2}, \frac{1}{2})$ distribution with probability less than 0.01. Calculate the proportion of the athletes in your sample that exceed this limit.

- Say that a doped athlete has the same average, but a larger variance (the values vary more due to blood manipulation). Generate the values for 1000 athletes with a larger variance and check the proportion that exceeds the limits.

1.5 Sum of normal variables

As we have already shown, the linear transformation of a normal variable remains normally distributed. In this exercise, we shall show that the sum of two independent (standard) normal distributions in again normally distributed, but that this may not be true for two dependent normal variables.

- Show that a sum of two independent standardized normal variables is a normally distributed random variable, find its mean and standard deviation.

- Let the variable Z equal $|Y|$ if $X \geq 0$, and $Z = -|Y|$ if $X < 0$. Find the distribution of Z?

- Sketch the joint distribution of X and Z. Are the two variables independent?

- Is the sum $X + Z$ normally distributed?