Exercise 1

For the following random experiments, decide what the distribution of X should be. In nearly every case, there are additional assumptions that should be made for the distribution to apply. Please identify those assumptions yourself (which may or may not strictly hold in practice).

(a) We throw a dart at a dart board. Let X denote the squared linear distance from the bullseye to where the dart landed.

(b) We randomly choose a textbook from the shelf at the bookstore and let P denote the proportion of the total pages of the book devoted to exercises.

(c) We measure the time it takes for the water to completely drain out of the kitchen sink.

(d) We randomly sample strangers at the grocery store and ask them how long it will take them to drive home.

Exercise 2

The response variable Y from a placebo controlled randomized clinical trial (RCT I) is $N(5, 1)$ and $N(5+\beta_{\text{shift}}, 1)$ for the placebo and treatment group, respectively. The expected treatment effect is $\beta_{\text{shift}} = 1.5$.

```
set.seed(2018)
n <- 25
m.plc <- 5
placebo <- rnorm(n, mean = m.plc, sd = 1)
```

(a) Simulate n response variables for the treatment group (`treatment`). For this, assume the transformation function $h(y)$ is defined as $h(y) = \text{id}$.

(b) Plot the empirical cumulative distribution functions (ecdf) in the same graph and add a legend.

(c) What is $\mathsf{E}(Y \mid \text{treatment}) - \mathsf{E}(Y \mid \text{placebo})$? How can you interpret it?

For the analysis of an additional randomized clinical trial (RCT II), we assume the transformation function $h(y)$ is redefined as $h(y) = \log(y) + 1$, and $\beta_{\text{shift}} = 2.02$.

(d) Based on this, simulate $n = 100$ observations $(Y_1, Y_2, ...Y_{100})$ for patients allocated to placebo and treatment group, respectively.

- $F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y \mid \text{placebo}) = \Phi(h(y)) = \Phi(\log(y) + 1)$
- $F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y \mid \text{treatment}) = \Phi(h(y) - \beta) = \Phi(\log(y) + 1 - \beta_{shift})$

Careful, $\Phi = N(0, 1)$ stands for the standard normal distribution with $\mu = 0$ and $\sigma^2 = 1$.

(e) Plot the cumulative distribution function (cdf) and the empirical cumulative distribution function (ecdf) = separately for the placebo and treatment group.
What's your conclusion?

Exercise 3

In this exercise you will explore how different choices of F_Z and h influence the shape of the cumulative distribution function of the response Y.

(a) Plot the four distributions F_Z that were introduced in the lecture, i.e., Standard-Normal, -Logistic, -Gompertz, -Gumbel, on an appropriate range. State the support of a random variable following the above-mentioned distributions.

luisa.barbanti@uzh.ch
lucasheinrich.kook@uzh.ch

(b) For the following transformation functions of the response Y plot the resulting cdf of Y, i.e., $F_Y(y) = F_Z(h(y))$ for all error distributions from (a) and arbitrary $b > 0$, $a \in \mathbb{R}$ (which you may choose yourself).

- $h(y) = a + by$
- $h(y) = a + b\log(y)$
- $h(y) = a + b\exp(y)$
- $h(y) = \tan(y)$

Also state the support of the random variable $h(Y)$.

luisa.barbanti@uzh.ch
lucasheinrich.kook@uzh.ch