Refresher on statistical notations

This document introduces:

- **mathematical symbols**: $\min(x, y)$, $T \in [0, \tau]$, $\mathbb{1}_{T \leq C}$, $\sum_{i=1}^{n} N_i$, $\int_0^t \lambda(s)ds$, $\exp(-\Lambda(t))$, $\log(HR)$.
- **fundamental notions in probability** (probability, joint probability, conditional probability, expectation) and corresponding notation ($\mathbb{P}$, $\mathbb{E}$)
- **fundamental notions in statistics** (estimand, estimator, estimate, statistical inference, likelihood) and corresponding notation (ψ, $\widehat{\psi}$)

The aim of the document is to re-familiarize you with statistics in order to make the lectures easier to follow. Mathematical notations can, at first sight, seem obscur but they are very convenient shortcut, avoiding to write or say long sentences.

The recommanded reading program is the following:

- *to prepare day 1*: minimum & maximum (1.1), interval (1.2), indicator function (1.3), sum (1.4), integral (1.5), exponential (1.6), probability (2.1, 2.2, 2.3).
- *to prepare day 3*: , logarithm (1.7), estimand, estimator, estimate (3.1), statistical inference (3.2), likelihood (3.3)
- *to prepare day 9*: expectation (2.4)

⚠ This document is much more dense in mathematical notations and abstract than the lectures and exercises. Do not worry if it is not crystal clear after a first read. Maybe it will make more sense to come back to it after the lecture if something was unclear.

1 Mathematical symbols

1.1 Minimum & maximum

$\min(5, 1) = 1$ denotes the minimum between two numbers (here 5 and 1).

Considering two variables T and C, then $T^* = \min(T, C)$ is the minimum between the values taken by the variables. For instance if T is the time to death and C is the follow-up time planned in the study then T^* is the observed follow-up time since follow-up is terminated due to death of the patient or end of the study period, whichever comes first.

Note: In some books or articles $\min(5, 1)$ is denoted by $5 \wedge 1$. Similarly the maximum, e.g. $\max(5, 1)$, is denoted by $5 \vee 1$.

1.2 Interval

In epidemiology we often restrict the study to a specific time horizon, e.g., we may be interest in the COVID-19 incidence rate in 2021. Imagine we measure time (in years) since the start of the infection, where for simplicity the start is taken to be 1$^{\text{st}}$ of January 2020. Then t=1 refers to 1$^{\text{st}}$ of January 2021 and t=2 the 1$^{\text{st}}$ of January 2022.

- $t \in [1, 2]$ refers to all timepoints between t=1 and t=2, including t=1 and t=2.
- $t \in]1, 2]$ or $t \in (1, 2]$ refers to all timepoints between t=1 and t=2, excluding t=1 but including t=2.
- $t \in [1, 2[$ or $t \in [1, 2)$ refers to all timepoints between t=1 and t=2, including t=1 and excluding t=2. This corresponds to the year 2021 (i.e. from 1$^{\text{st}}$ of January 2021 to just before 1$^{\text{st}}$ of January 2022 i.e. end of 31$^{\text{st}}$ of December 2021).
- $t \in]1, 2[$ or $t \in (1, 2)$ refers to all timepoints between t=1 and t=2, excluding t=1 and t=2.

The symbol ∞ (or $+\infty$) denotes infinity, i.e. infinitely large values. So $T \in [0, +\infty[$ means a number that is 0 or above 0 and below infinity i.e. any non-negative number.

Strictly speaking intervals are only defined over a continuum of values. For discrete values such as 1, 2, 3, 4 (here 1.5 is not a possible value) one should use the set notation: $\{1, 2, 3, 4\}$ that can be abreviated into $\{1, \ldots, 4\}$.

1.3 Indicator function

$\mathbb{1}_\bullet$ is the indicator function, taking value 1 if $\bullet$ is true and 0 if $\bullet$ is false. For instance if T is the time to death and C is the planned follow-up time:

- $\mathbb{1}_{T \leq C}$ is one if death occurred during the follow-up time (no censoring) and 0 if it occurred outside of the follow-up time (censored). It is also written $\mathbb{1}\{T \leq C\}$.

1.4 Sum

The sum of a serie of numbers (e.g. $1+2+3+4$) or variables (e.g. $N_1+N_3+N_4+N_4+N_5$) is commonly abreviated using the capital sigma symbol: $\sum_{i=1}^{4} i$ or $\sum_{i=1}^{5} N_i$.

In the course we will use this notation to explain how to estimate an incidence rate using the ratio of the number of cases divided by the time at risk. Both the numerator and denominator are sums over the number of persons in the cohort (say n) of either the individual-specific indicator of an event ($N_i \in \{0,1\}$) or follow-up time (T_i) where i refers to the i-th individual:

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} T_i} = \frac{N_1 + N_2 + \ldots + N_n}{T_1 + T_2 + \ldots + T_n}$$

1.5 Integral

An integral can be seen as continuous sum. Imagine we are given the instantaneous risk of death[1] at time t which we will denote $\lambda(t)$:

- discrete time (say days): we cumulate the risk time using '+'
  $\sum_{s=1}^{5} \lambda(s) = \lambda(1) + \lambda(2) + \lambda(3) + \lambda(4) + \lambda(5)$.
- continuous time: we cumulate the risk using an integral: $\Lambda(5) = \int_0^5 \lambda(s)ds$ which is the area under the curve defined by the function $\lambda : t \mapsto \lambda(t)$ (see Figure 1).

If $\lambda(t)$ is piecewise constant (i.e. only change when changing day) the two are equal:

$$\begin{aligned}
\Lambda(5) = \int_0^5 \lambda(s)ds &= \int_0^1 \lambda(s)ds + \int_1^2 \lambda(s)ds + \int_2^3 \lambda(s)ds + \int_3^4 \lambda(s)ds + \int_4^5 \lambda(s)ds \\
&= \lambda(1)\int_0^1 ds + \lambda(2)\int_1^2 ds + \lambda(3)\int_2^3 ds + \lambda(4)\int_3^4 ds + \lambda(5)\int_4^5 ds \\
&= \lambda(1) + \lambda(2) + \lambda(3) + \lambda(4) + \lambda(5)
\end{aligned}$$

[1] what this exactly is will be seen during the first lecture

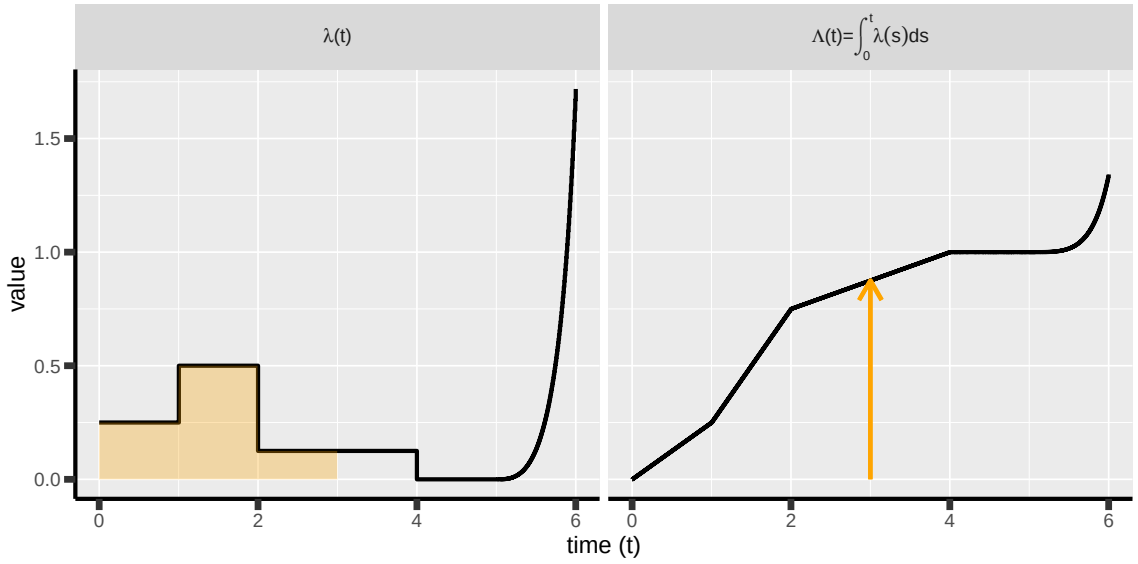

Figure 1: Instantaneous risk (left panel) and its integral over time (right panel). Each point on the right panel (say at time t) corresponds to the area under the curve from the left panel between time 0 and time t.

1.6 Exponential

The exponential function is a mathematical function defined on the real line. It is strictly increasing, taking values between 0 and ∞. It is denoted $\exp(x)$ or equivalently e^x. One key property is that:

- $\exp(a+b) = \exp(a)\exp(b)$, thus $\exp(0) = 1$ and $\exp(-a) = 1/\exp(a)$

During the course we will use the exponential function to model the covariate effect on the hazard: $\lambda(t|X) = \lambda_0(t)\exp(\beta X)$:

- when $X = 0$ then $\lambda(t|X=0) = \lambda_0(t)\exp(0) = \lambda_0(t)$.
- the hazard ratio $\frac{\lambda(t|X=x+1)}{\lambda(t|X=x)} = \frac{\exp(\beta(x+1))}{\exp(\beta x)} = \exp(\beta(x+1-x)) = \exp(\beta)$.

It will also be used to model the survival[2] based on the modeled cumulative hazard: $S(t|X) = \exp(-\Lambda(t|X)) = \exp\left(-\int_0^t \lambda(s|X)ds\right)$:

- the survival decreases as the cumulative hazard increases.
- the survival equals 1 when the cumulative hazard is 0 and goes to 0 as the cumulative hazard increases to infinity.

[2] what this exactly is will be seen during the first lecture. It is also introduced at the end of the next page

1.7 Natural logarithm

The natural logarithm is a mathematical function defined on positive numbers. It is strictly increasing, taking values between $-\infty$ and ∞. It is denoted $\log(x)$. One key property is that:

- $\log(a) + \log(b) = \log(ab)$, thus $\log(1) = 0$ and $\log(1/a) = -\log(a)$

The logarithm is the inverse of the exponential function (i.e. $\log(\exp(x)) = x$) which can be, for instance, used to re-express regression coefficient as a ratio of log survival. Indeed under the model $S(t|X) = \exp(-\Lambda(t|X)) = \exp(-\Lambda_0(t)\exp(\beta X))$:

$$\frac{\log(S(t|X = x + 1))}{\log(S(t|X = x))} = \frac{-\Lambda(t|X = x + 1)}{-\Lambda(t|X = x)} = \exp(\beta)$$

2 Probabilistic background

2.1 Probability of an event

The probability of an event is a number between 0 and 1 indicating how likely is an event to occur. Such an event could be A: "landing on tail" or B: "landing on head" when tossing a fair coin.

$$\mathbb{P}[A] = 0.5 \tag{1}$$

is then a concise way to express "A fair coin has probability 0.5 to land on tail".

Medical studies typically involve more complex outcome than binary, e.g. disease diagnosis is the result of an examination that revealed a number of symptoms associated with the disease. **Random variables** are functions that maps the set of possible outcomes to numerical values. In the coin example the possible outcomes are "landing on tail" and "landing on head". Example of random variables are:

- A_{t}: mapping "landing on tail" to 1 and "landing on head" to 0.
- A_{h}: mapping "landing on tail" to 0 and "landing on head" to 1.
- A_0: mapping "landing on tail" to -0.5 and "landing on head" to 0.5.

Probabilities can then be expressed relative to random variables, e.g. $\mathbb{P}[A_{\mathrm{t}} = 1] = 0.5$, $\mathbb{P}[A_{\mathrm{h}} = 0] = 0.5$, $\mathbb{P}[A_0 = -0.5] = 0.5$ are all equivalent to Equation 1. What to use depends on the aim and on convenience.

During the course we will consider more complex probabilities. For instance consider investing how long patients survive after being diagnosed of lung cancer. We will typically introduce T the time to death, implicitely defining a random variable mapping the history of the patient (date of birth, date of long cancer diagnosis, date of death) to a number.

⚠ here the definition is ambiguous. One would typically understand that the number is the difference between the date of death and the date of long cancer diagnosis, expressed in years.

We can then define the survival probability as:

$$S(t) = \mathbb{P}[T > t]$$

which is the probability that time to death is (strictly) greater than t. When t is expressed in years, $S(1)$ denotes the probability that time to death is (strictly) greater than 1 year, i.e. the 1 year survival.

2.2 Joint probability

The joint probability is the probability of two events co-occurring. If we flip two fair coins and denote by A_1 the event "the first coin lands on tail" and by A_2 the event "the second coin lands on tail", then:

$$\mathbb{P}[A_1, A_2] = 0.25$$

is a concise way to express "Two fair coins have probability 0.25 to both land on tail".

In epidemiology it is frequent that T is a time to first event. One event will be the event of interset, say long cancer diagnosis ($\Delta = 1$), and the other a competing event, say death ($\Delta = 2$). Then:

$$\mathbb{P}[T \leq t, \Delta = 1] = 0.03$$

is a concise way to express "the probability that the first event happen before time t and it is a long cancer diagnosis is 0.03". When t is expressed in years, a more 'common language' formulation would be "the risk for the patient to be diagnosed with long cancer within t-years is 0.03" (since to be diagnosed one has to be alive it is unecessary to precise which event came first). Similarly:

$$\mathbb{P}[T \leq t, \Delta = 2] = 0.2$$

would be a concise way to express "the risk for the patient to die within t-years free of long cancer is 0.2".

2.3 Conditional probability

The conditional probability is the probability of one event occurring given that another has occurred. Consider a study following people after long cancer diagnosis that may ($X = 1$) or may not ($X = 0$) have previously suffered from a stroke attack. If we denote by T the time to death then:

$$\mathbb{P}[T \leq t | X = 1]$$

expresses the probability of death within t years after a diagnosis of cancer for a patient with history of stroke. This would typically bigger than:

$$\mathbb{P}[T \leq t] = \mathbb{P}[T \leq t | X = 0]\mathbb{P}[X = 0] + \mathbb{P}[T \leq t | X = 1]\mathbb{P}[X = 1]$$

the probability of death within t years after a diagnosis of cancer for a person for which we do not know his history stroke at diagnostic. As shown on the right hand

side[3], this probability would be a weighted average of the probability of death before time t for someone without history of stroke and the probability of death before time t for someone with history of stroke, the weights being relative to the probability to not having experience a stroke and the probability to have experienced a stroke.

- An interactive visualization of the conditional probability can be found at https://seeing-theory.brown.edu/compound-probability/index.html#section3.

⚠ Conditioning on future events is typically a bad idea as it can lead to substantial bias. For instance conditioning on whether a cancer patient experience a stroke **during** the study time may severely bias downward the probability of death since to be able to experience a stroke one needs to be alive. This is commonly refered to as immortal time bias.

2.4 Expectation

One common example of **expectation** is the gain one would obtain in average when playing a gambling game. More precisely suppose one can bet on 5 numbers returning respectively p_1, p_2, p_3, p_4, p_5 DKK. If numbers are drawn with equal probability 0.2 the expected gain is:

$$\mathbb{E}[X] = 0.2 * p_1 + 0.2 * p_2 + 0.2 * p_3 + 0.2 * p_4 + 0.2 * p_5$$

where X denotes the random variable mapping what has been drawn to the gain. For instance if only number 5 provides a gain and this gain is 100DKK then:

$$\mathbb{E}[X] = 0.2 * 0 + 0.2 * 0 + 0.2 * 0 + 0.2 * 0 + 0.2 * 100 = 20DKK$$

This concept will mostly be useful at the end of the course when we will consider models where the risk of death of the patient may depend on the treatment E and covariates X, say sex, in a complicated way. Denote by $\mathbb{P}[T \leq 1|E, X]$ the 1 year risk of death for a person under treatment E with gender X. One could then compute the expected 1-year risk of death in the population as:

$$\begin{aligned}\mathbb{E}[\mathbb{P}[T \leq 1|E, X]] =& \mathbb{P}[T \leq 1|E = \text{placebo}, X = \text{male}]\,\mathbb{P}[E = \text{placebo}, X = \text{male}] \\ &+ \mathbb{P}[T \leq 1|E = \text{placebo}, X = \text{female}]\,\mathbb{P}[E = \text{placebo}, X = \text{female}] \\ &+ \mathbb{P}[T \leq 1|E = \text{active}, X = \text{male}]\,\mathbb{P}[E = \text{active}, X = \text{male}] \\ &+ \mathbb{P}[T \leq 1|E = \text{active}, X = \text{female}]\,\mathbb{P}[E = \text{active}, X = \text{female}]\end{aligned}$$

i.e. the sum over all patient profiles of the risk under patient profile times the frequency that this patient profile is met in the population.

[3] which follows from the law of total probability

3 Statistical background

3.1 Terminology

An **estimand** is a quantity of interest to be estimated in a statistical analysis. It is **deterministic** (i.e. a number) in frequentist statistics.

- in the coin example, the estimand could be the probability of landing on tail (which equals 0.5 for a fair coin).
- other examples include "the 1-year risk of being diagnosed with long cancer" or "the difference in probability of surviving at least a year between people exposed to pollution ($E = 1$) and people non-exposed to pollution ($E = 0$)". If we denote by ψ the estimand the previous sentences could be abreviated by $\psi = \mathbb{P}[T \leq 1, \Delta = 1]$ or $\psi = S(t|E = 1) - S(t|E = 0)$. In the latter case, if the survival time follows an exponential distribution with rate 2 under pollution and 1 otherwise, then $\psi = \exp(-2) - \exp(-1) \approx -0.2325$:

An **estimator** is a method to deduce from the data at hand the value of the estimand (typically under minimal distributional assumptions).

- the empirical average is an example of estimator: it sums the observations and divide by the number of observations. Consider again the fair coin example, and denote by X the random variable which takes value 1 if the coin lands on tail and value 0 if the coin lands on head (X could be the earning of a gambling game). Say we flip the coin 5 times and observe the following five realizations of X: x_1, x_2, x_3, x_4, x_5. The empirical average of X in this sample is

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

- another commonly used estimator is maximum likelihood estimator (MLE) which find values for the model parameters minimizing the discrepancy between modeled values and observed values. This is what is used when doing linear regression (`lm` in ), logistic regression (`glm` in ), or Poisson regression (`glm` in ).

An **estimate** is the value provided by the estimator for a given dataset. It is often represented by adding an hat to the estimand, e.g. $\widehat{\psi}$ and is stochastic as it will fluctuate from dataset to dataset.

- following the fair coin example, if we observed head ($x_1 = 0$), tail ($x_2 = 1$), tail ($x_3 = 1$), head ($x_4 = 0$), head ($x_5 = 0$), the empirical average of X is:

$$\overline{x} = \frac{0 + 1 + 1 + 0 + 0}{5} = 0.4$$

- similarly, in the pollution example, had we been able to collect data we could estimate the survival under pollution and free of pollution to estimate the pollution effect. For instance with 10 individuals in each group, we may measure:

```
set.seed(12)
sample1.T0 <- round(rexp(10,1),3) ## no pollution
sample1.T1 <- round(rexp(10,2),3) ## pollution
rbind(sample1.T0,sample1.T1)
```

```
            [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10]
sample1.T0 2.189 0.636 0.117 2.857 1.817 0.283 3.930 4.224 1.264 0.628
sample1.T1 0.599 0.608 0.376 0.505 0.166 1.115 1.323 0.210 0.786 0.947
```

In absence of censoring, the survival in each group can be estimated by averaging whether the survival time are greater than t:

```
mean(sample1.T1>1) - mean(sample1.T0>1)
```

```
[1] -0.4
```

These values are **stochastic**, i.e., dependent on the sample. Had we had different observations we would have obtained different values:

```
set.seed(12)
sample2.T0 <- round(rexp(10000,1),3) ## pollution
sample2.T1 <- round(rexp(10000,2),3) ## no pollution
mean(sample2.T1>1) - mean(sample2.T0>1)
```

```
[1] -0.2363
```

- the estimate will get closer to the estimand as the number of observations increases (provided certain assumptions, e.g. independent and identically distributed observations, no confounding, ...). See https://seeing-theory.brown.edu/basic-probability/index.html#section2 for an illustration.

3.2 Statistical inference

In previous section, we consider an (artificial) example where we wanted to assess the effect of pollution on the 1-year risk of being diagnosed with long cancer. With one sample (n=10 per group) we obtained effect of -0.4 while with another, much larger, sample we obtained an effect of -0.2363. This variability, i.e. dependence of the estimated treatment effect on the sample, makes us unsure whether the observed negative effect is just a coincidence or would happen with most samples.

- to limit the number of incorrect claims, conclusions should be drawn not only with respect to the estimate but also to the statistical uncertainty. For instance one could use a test comparing proportion of events:

```
library(exact2x2)
## first sample
dp1 <- binomMeld.test(x1 = sum(sample1.T0>1), x2 = sum(sample1.T1>1),
                      n1 = length(sample1.T0), n2 = length(sample1.T1))
## second sample
dp2 <- binomMeld.test(x1 = sum(sample2.T0>1), x2 = sum(sample2.T1>1),
                      n1 = length(sample2.T0), n2 = length(sample2.T1))

M.res <- rbind(sample1 = c(dp1$estimate, dp1$conf.int,dp1$p.value),
               sample2 = c(dp2$estimate, dp2$conf.int,dp2$p.value))
colnames(M.res) <- c("estimate","lower","upper","p.value")
M.res
```

```
        estimate      lower      upper    p.value
sample1  -0.4000 -0.7670161  0.1216111 0.16980235
sample2  -0.2363 -0.2479586 -0.2245869 0.00000002
```

Confidence intervals (CIs) can be understood as displaying the range of effects that are compatible with the data - see (Greenland et al., 2016) for a formal definition.

- with sample 1, we see that an effect (in absolute value) as large as -0.767 and as small as -0.12 are compatible with the data so no clear conclusion can be drawn. Nevertheless we can exclude large positive (i.e. protective) effects.
  With sample 2, the CI is narrower with values between 0.22 and 0.24 not only showing evidence for a decreased survival but also quantifying it.

- Formally, if we were making many studies on pollution and compute each time a 95% CI, then 95% of them will contain the estimand (i.e. we would only be mistaken in 5% of the cases). See https://seeing-theory.brown.edu/frequentist-inference/index.html#section2 for an interactive visualization of the frequentist properties of confidence intervals. In practice this often require certain assumptions to be met (e.g. independent observations, large enough sample).

- In any case, the width of the CIs reflects the statistical uncertainty: broad CIs point toward insufficient or too noisy data[4].

Typically one wishes to show the presence of an effect. To do so a study is set up where the absence of this effect (null hypothesis) should be incompatible wiht the measurements. **P-values** are then used to quantify the evidence against the null hypothesis. A low p-value (typically below 0.05) indicates evidence against the null hypothesis, i.e., typically evidence against no treatment effect i.e. evidence in favor of a treatment effect. As mentionned in Greenland et al. (2016), this interpretation implicitely require that modeling assumptions are correct but to keep things simple we will assume this is the case during the course.

⚠ a large p-value does NOT indicate evidence for the null hypothesis, i.e. does NOT indicate no effect. The above example illustrate this: in the first (small) sample the p-value was 0.170 but that does NOT support no pollution effect. Looking at the confidence interval we see rather large effects values compatible with the data. See Greenland et al. (2016) for other common misconceptions.

3.3 Deriving estimators using the likelihood

The likelihood quantifies how likely it is to observe the data under the current model. It is commonly used to estimate model parameters (say θ) and quantify the estimation uncertainty. The main ideas are the following:

- define a statistical model, e.g. treatment modifies (multiplicatively) the survival probability.
- express the likelihood of the model (denoted $\mathcal{L}(\theta)$) i.e. relate the model parameters to the observed data
- estimate the model parameters by finding the value they should take to maximize the likelihood. This value is called the maximum likelihood estimate (MLE) and denoted $\widehat{\theta}$. This is often equivalent to finding the value where the first derivative of the log-likelihood (denoted $\mathcal{S}(\theta) = \mathcal{L}'(\theta)$) is 0.
- estimate the estimation uncertainty by evaluating the second derivative of the log-likelihood (denoted $\mathcal{H}(\theta) = \mathcal{S}'(\theta)$). Intuitively, the sharper the second derivative the easier it is to identify where the first derivative is 0 (compare the blue and red curve in Figure 2). More precisely the standard error of the estimator (denoted $\sigma_{\widehat{\theta}}$) equals the square root of the inverse of the opposite of the second derivative of the log-likelihood $\sqrt{\frac{1}{-\mathcal{H}(\widehat{\theta})}}$

[4]The width of the confidence intervals is often an interesting metric to discuss when asked whether a study was sufficiently powered

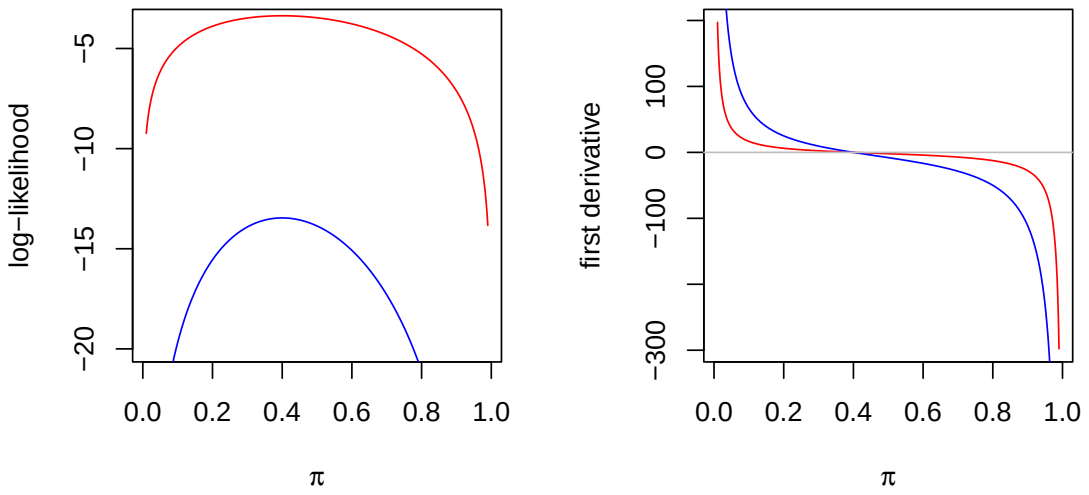

Figure 2: Log-likelihood and its first derivative (score) for $m = 2$ and $n = 3$ or $m = 8$ and $n = 12$

Consider a very simple example where we want to estimate the survival probability: $\pi = \mathbb{P}[Y = 1]$ (here we omit the reference to time to further simplify)

- the statistical model is $\mathbb{P}[Y = 1] = \pi$ and $\mathbb{P}[Y = 0] = 1 - \pi$
- given that m patients died and n survived the likelihood is $\mathcal{L}(\pi) = (1-\pi)^m \pi^n$. The log-likelihood is $\log(\mathcal{L}(\pi)) = -m\log(1-\pi) + n\log(\pi)$ whose values are displayed in the left panel of Figure 2 for various values of π.
- its derivative is $\mathcal{S}(\pi) = -\frac{m}{1-\pi} + \frac{n}{\pi}$ whose values are displayed in the right panel of Figure 2 for various values of π. It equals 0 when $\widehat{\pi} = \frac{n}{m+n}$.
- taking one more derivative we obtain $\mathcal{H}(\pi) = -\frac{m}{(1-\pi)^2} - \frac{n}{\pi^2} = -\frac{m+n}{\pi(1-\pi)}\left(\frac{\pi}{1-\pi} + \frac{\frac{n}{m+n}(1-2\pi)}{\pi(1-\pi)}\right)$. At the MLE this is equal to $\mathcal{H}(\widehat{\pi}) = -\frac{m+n}{\widehat{\pi}(1-\widehat{\pi})}\left(\frac{\widehat{\pi}}{1-\widehat{\pi}} + \frac{\widehat{\pi}(1-2\widehat{\pi})}{\widehat{\pi}(1-\widehat{\pi})}\right) = -\frac{m+n}{\widehat{\pi}(1-\widehat{\pi})}$ so $\widehat{\sigma}_{\widehat{\theta}} = \sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{m+n}}$

4 Reference

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., and Altman, D. G. (2016). Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, 31(4):337–350.