Explicit formula for the net benefit

Brice Ozenne, Julien Péron

July 10, 2018

Contents

1 Parameter of interest

Let consider two independent random variables X and Y. We are interested in:

$$\Delta = \mathbb{P}[Y > X] - \mathbb{P}[X > Y]$$

In the examples we will use a sample size of:

```
n <- 1e4
```

and use the following R packages

```
library(BuyseTest)
library(riskRegression)
```

2 Binary variable

2.1 Theory

$$\mathbb{P}[Y > X] = \mathbb{P}[Y = 1, X = 0]$$

Using the independence between Y and X:

$$\mathbb{P}[Y > X] = \mathbb{P}[Y = 1]\,\mathbb{P}[X = 0] = \mathbb{P}[Y = 1]\,(1 - \mathbb{P}[X = 0]) = \mathbb{P}[Y = 1] - \mathbb{P}[Y = 1]\,\mathbb{P}[X = 1]$$

By symmetry:

$$\mathbb{P}[X > Y] = \mathbb{P}[X = 1] - \mathbb{P}[Y = 1]\,\mathbb{P}[X = 1]$$

So

$$\Delta = \mathbb{P}[Y = 1] - \mathbb{P}[X = 0]$$

2.2 In R

Settings:

```
prob1 <- 0.4
prob2 <- 0.2
```

Simulate data:

```
set.seed(10)
df <- rbind(data.frame(tox = rbinom(n, prob = prob1, size = 1), group = "C"),
        data.frame(tox = rbinom(n, prob = prob2, size = 1), group = "T"))
```

Buyse test:

```
BuyseTest(group ~ bin(tox), data = df, method.inference = "none", trace = 0)
```

```
endpoint threshold   delta   Delta
     tox       0.5 -0.1981 -0.1981
```

Expected:

```
prob2 - prob1
```

```
[1] -0.2
```

3 Continuous variable

3.1 Theory

Let's consider two normally distributed variables with common variance:

- $X \sim \mathcal{N}\left(\mu_X, \sigma^2\right)$
- $Y \sim \mathcal{N}\left(\mu_Y, \sigma^2\right)$

Denoting $d = \frac{\mu_Y - \mu_X}{\sigma}$:

- $X^* \sim \mathcal{N}\left(0, 1\right)$
- $Y^* \sim \mathcal{N}\left(d, 1\right)$

$$\mathbb{P}\left[Y > X\right] = \mathbb{E}\left[\mathbb{1}_{Y>X}\right] = \mathbb{E}\left[\mathbb{1}_{Y*>X*}\right] = \mathbb{E}\left[\mathbb{1}_{Z>0}\right]$$

where $Z \sim \mathcal{N}\left(d, 2\right)$ so $\mathbb{P}\left[Y > X\right] = \Phi(\frac{d}{\sqrt{2}})$

By symmetry

$$\Delta = 2 * \Phi(\frac{d}{\sqrt{2}}) - 1$$

3.2 In R

Settings:

```
mean1 <- 0
mean2 <- 2
sd12 <- 1
```

Simulate data:

```
set.seed(10)
df <- rbind(data.frame(tox = rnorm(n, mean = mean1, sd = sd12), group = "C"),
          data.frame(tox = rnorm(n, mean = mean2, sd = sd12), group = "T"))
```

Buyse test:

```
BuyseTest(group ~ cont(tox), data = df, method.inference = "none", trace = 0)
```

```
endpoint threshold  delta  Delta
     tox     1e-12 0.8359 0.8359
```

Expected:

```
d <- (mean2-mean1)/sd12
2*pnorm(d/sqrt(2))-1
```

```
[1] 0.8427008
```

4 Survival

4.1 Theory

For a given cumulative density function $F(x)$ and a corresponding probability density function $f(x)$ we define the hazard by:

$$\begin{aligned}\lambda(t) &= \left.\frac{\mathbb{P}\left[t \leq T \leq t+h | T \geq t\right]}{h}\right|_{h\to 0^+} \\ &= \left.\frac{\mathbb{P}\left[t \leq T \leq t+h\right]}{\mathbb{P}\left[T \geq t\right] h}\right|_{h\to 0^+} \\ &= \frac{f(t)}{1-F(t)}\end{aligned}$$

Let now consider two times to events following an exponential distribution:

- $T1 \sim Exp(\alpha_1)$. The corresponding hazard function is $\lambda(t) = \alpha_1$.
- $T2 \sim Exp(\alpha_2)$. The corresponding hazard function is $\lambda(t) = \alpha_2$.

So the hazad ratio is $HR = \frac{\lambda_1}{\lambda_2}$. Note that if we use a cox model we will have:

$$\lambda(t) = \lambda_0(t)\exp(\beta \mathbb{1}_{group})$$

where $\exp(\beta)$ is the hazard ratio.

$$\begin{aligned}\mathbb{P}\left[T_1 > T_2\right] &= \int_0^\infty \alpha_1 \exp(-\alpha_1 t) \int_0^{t_1} \alpha_2 \exp(-\alpha_2 t) dt_2 dt_1 \\ &= \int_0^\infty \alpha_1 \exp(-\alpha_1 t)[\exp(-\alpha_2 t)]_{t_1}^0 dt_1 \\ &= \int_0^\infty \alpha_1 \exp(-\alpha_1 t)(\exp(-\alpha_2 t_1) - 1) dt_1 \\ &= \frac{\alpha_1}{\alpha_1+\alpha_2}[\exp(-(\alpha_1+\alpha_2)t)]_\infty^0 - [\exp(-\alpha_1 t)]_\infty^0 \\ &= 1 - \frac{\alpha_1}{\alpha_1+\alpha_2} \\ &= 1 - \frac{HR}{1+HR}\end{aligned}$$

So $\mathbb{P}\left[T_2 > T_1\right] = \frac{HR}{1+HR}$.

4.2 In R

Settings:

```
alpha1 <- 2
alpha2 <- 1
```

Simulate data:

```
set.seed(10)
df <- rbind(data.frame(time = rexp(n, rate = alpha1), group = "C", event = 1),
        data.frame(time = rexp(n, rate = alpha2), group = "T", event = 1))
```

Buyse test:

```
BuyseTest(group ~ tte(time, censoring = event), data = df,
      method.inference = "none", trace = 0, method.tte = "Gehan")
```

```
endpoint threshold  delta  Delta
    time     1e-12 0.3403 0.3403
```

Expected:

```
e.coxph <- coxph(Surv(time,event)~group,data = df)
HR <- as.double(exp(coef(e.coxph)))
c("HR" = alpha2/alpha1, "Delta" = (alpha2/alpha1)/(1+alpha2/alpha1))
c("HR.cox" = HR, "Delta" = (HR)/(1+HR))
```

```
       HR     Delta
0.5000000 0.3333333
   HR.cox     Delta
0.4918256 0.3296804
```

5 Competing risks

5.1 Theory

Let now consider two competing events whose times to event follow an exponential distribution:

- $T1 \sim Exp(\alpha_1)$. The corresponding hazard function is $\lambda(t) = \alpha_1$.
- $T2 \sim Exp(\alpha_2)$. The corresponding hazard function is $\lambda(t) = \alpha_2$.

The cumulative incidence function can be written:

$$
\begin{aligned}
CIF_1(t) &= \int_0^t \lambda_1(s) S(s_-) ds \\
&= \int_0^t \alpha_1 \exp(-(\alpha_1 + \alpha_2) * s_-) ds \\
&= \frac{\alpha_1}{\alpha_1 + \alpha_2} \left[\exp(-(\alpha_1 + \alpha_2) * s_-)\right]_t^0 \\
&= \frac{\alpha_1}{\alpha_1 + \alpha_2} \left(1 - \exp(-(\alpha_1 + \alpha_2) * t_-)\right)
\end{aligned}
$$

where $S(t)$ denote the event free survival and s_- denotes the right sided limit.

Now if we consider two groups such that:

- $T1 \sim Exp(\alpha_{1,T})$ in group T and $T1 \sim Exp(\alpha_{1,C})$ in group C
- $T2 \sim Exp(\alpha_{2,T})$ in group T and $T2 \sim Exp(\alpha_{2,C})$ in group C

Then:

$$
\begin{aligned}
CIF_1(t|group = T) &= \frac{\alpha_{1,T}}{\alpha_{1,T} + \alpha_{2,T}} \left(1 - \exp(-(\alpha_{1,T} + \alpha_{2,T}) * t_-)\right) \\
CIF_1(t|group = C) &= \frac{\alpha_{1,C}}{\alpha_{1,C} + \alpha_{2,C}} \left(1 - \exp(-(\alpha_{1,C} + \alpha_{2,C}) * t_-)\right)
\end{aligned}
$$

Let denote ε_T the event type indicator (1 cause of interest and 2 competing risk) in group T and ε_C the event type indicator in group C:

$$
\begin{aligned}
\Delta &= \frac{1}{\mathbb{P}\left[\varepsilon_T = 1, \varepsilon_C = 1\right]} \frac{\alpha_{1,T}}{\alpha_{1,T} + \alpha_{1,C}} - \frac{1}{\mathbb{P}\left[\varepsilon_T = 1, \varepsilon_C = 2\right]} + \frac{1}{\mathbb{P}\left[\varepsilon_T = 2, \varepsilon_C = 1\right]} \\
&=
\end{aligned}
$$

5.2 In R

Settings:

```
alpha1 <- 2
alpha2 <- 1
```

Simulate data:

```
set.seed(10)
df <- data.frame(time1 = rexp(n, rate = alpha1), time2 = rexp(n, rate = alpha2), group
    = "1", event = 1)
df$time <- pmin(df$time1,df$time2)
df$event <- (df$time2<df$time1)+1
```

Cumulative incidence (via risk regression):

```
e.CSC <- CSC(Hist(time, event) ~ 1, data = df)
vec.times <- unique(round(exp(seq(log(min(df$time)),log(max(df$time)),length.out = 12))
    ,2))
e.CSCpred <- predict(e.CSC, newdata = data.frame(X = 1), time = vec.times , cause = 1)
```

Expected vs. calculated:

```
cbind(time = vec.times,
      CSC = e.CSCpred$absRisk[1,],
      manual = alpha1/(alpha1+alpha2)*(1-exp(-(alpha1+alpha2)*(vec.times)))
      )
```

```
     time    CSC     manual
[1,] 0.00 0.0000 0.00000000
[2,] 0.01 0.0186 0.01970298
[3,] 0.02 0.0377 0.03882364
[4,] 0.05 0.0924 0.09286135
[5,] 0.14 0.2248 0.22863545
[6,] 0.42 0.4690 0.47756398
[7,] 1.24 0.6534 0.65051069
[8,] 3.70 0.6703 0.66665659
```

Could also be obtained treating the outcome as binary:

```
mean((df$time<=1)*(df$event==1))
```

```
[1] 0.6375
```

Now with Buyse test:

```
df11 <- df[]
```