Theory supporting the net benefit and Peron's scoring rules

Brice Ozenne

April 1, 2019

This document describe the relationship between the net benefit and traditional parameter of interest (e.g. hazard ratio). It also present how Peron's scoring rules for the survival and competing setting were derived.

In the examples we will use a sample size of:

```
n <- 1e4
```

and use the following R packages

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

Contents

1 Parameter of interest

Let consider two independent real valued (univariate) random variables X and Y. Informally X refer to the outcome in the experimental group while Y refer to the outcome in the control group. For a given threshold $\tau \in \mathbb{R}^{+*}$, the net benefit can be expressed as:

$$\Delta_\tau = \mathbb{P}\left[X \geq Y + \tau\right] - \mathbb{P}\left[Y \geq Y + \tau\right]$$

To relate the net benefit to known quantities we will also consider the case of an infinitesimal threshold τ:

$$\Delta_+ = \mathbb{P}\left[X > Y\right] - \mathbb{P}\left[Y > X\right]$$

In any case, X and Y play a symetric role in the sense that given a formula for $\mathbb{P}\left[X \geq Y + \tau\right]$ (or $\mathbb{P}\left[X > Y\right]$), we can substitute X to Y and Y to X to obtain the formula for $\mathbb{P}\left[Y \geq X + \tau\right]$ (or $\mathbb{P}\left[Y > X\right]$).

2   Binary variable

2.1   Relationship between Δ_+ and the prevalence

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

Using the independence between X and Y:

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

By symmetry:

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

So

$$\Delta_+ = \mathbb{P}\left[X = 1\right] - \mathbb{P}\left[Y = 1\right]$$

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
```

4

3 Continuous variable

3.1 Relationship between Δ and Cohen's d

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

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

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

Considering $Z \sim \mathcal{N}(d, 2)$ with $d = \frac{\mu_X - \mu_Y}{\sigma}$, we express:

$$\mathbb{P}[X > Y] = \mathbb{P}[\sigma(Y - X) > 0] = \mathbb{P}[Z > 0] = \Phi(\frac{d}{\sqrt{2}})$$

By symmetry

$$\mathbb{P}[Y > X] = \mathbb{P}[Z < 0] = 1 - \Phi(\frac{d}{\sqrt{2}})$$

So

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

3.2 In R

Settings:

```
meanX <- 0
meanY <- 2
sdXY <- 1
```

Simulate data:

```
set.seed(10)
df <- rbind(data.frame(tox = rnorm(n, mean = meanX, sd = sdXY), group = "C"),
            data.frame(tox = rnorm(n, mean = meanY, sd = sdXY), 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 <- (meanY-meanX)/sdXY
2*pnorm(d/sqrt(2))-1
```

```
[1] 0.8427008
```

4 Survival

4.1 Relationship between Δ and the hazard ratio

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

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

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

- $X \sim Exp(\alpha_X)$. The corresponding hazard function is $\lambda(t) = \alpha_X$.

- $Y \sim Exp(\alpha_Y)$. The corresponding hazard function is $\lambda(t) = \alpha_Y$.

So the hazad ratio is $HR = \frac{\alpha_X}{\alpha_Y}$. 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.

$$
\begin{aligned}
\mathbb{P}\left[X > Y\right] &= \int_0^\infty \mathbb{P}\left[x > Y\right] d\mathbb{P}\left[x > X\right] \\
&= \int_0^\infty \left(\int_0^x \alpha_Y \exp(-\alpha_Y y) dy\right) (\alpha_X \exp(-\alpha_X x) dx) \\
&= \int_0^\infty \left[-\exp(-\alpha_Y y)\right]_0^x (\alpha_X \exp(-\alpha_X x) dx) \\
&= \int_0^\infty (1 - \exp(-\alpha_Y x)) (\alpha_X \exp(-\alpha_X x) dx) \\
&= \int_0^\infty \alpha_X \left(\exp(-\alpha_X x) - \exp(-(\alpha_X + \alpha_Y)x)\right) dx \\
&= \left[\exp(-\alpha_X x) - \frac{\alpha_X}{\alpha_X + \alpha_Y} \exp(-(\alpha_X + \alpha_Y)x)\right]_0^\infty \\
&= 1 - \frac{\alpha_X}{\alpha_X + \alpha_Y} = \frac{\alpha_Y}{\alpha_X + \alpha_Y} \\
&= \frac{1}{1 + HR}
\end{aligned}
$$

So $\mathbb{P}\left[Y > X\right] = \frac{\alpha_X}{\alpha_Y + \alpha_X} = 1 - \frac{1}{1+HR}$ and:

$$
\Delta_+ = 2\frac{1}{1 + HR} - 1 = \frac{1 - HR}{1 + HR}
$$

4.2   Scoring rule in presence of censoring

Let's consider the following random variables:

- X the time to the occurrence of the event in the experimental group.

- $\tilde{X}$ the censored event time in the experimental group, i.e. $\tilde{X} = X \wedge C_X$ where C_X denotes the censoring time in the experimental group.

- $\varepsilon_X = \mathbb{1}_{X \leq C_X}$ the event time indicator in the experimental group.

- Y the time to the occurrence of the event in the control group.

- $\tilde{Y}$ the censored event time in the control group, i.e. $\tilde{Y} = X \wedge C_Y$ where C_Y denotes the censoring time in the control group.

- $\varepsilon_Y = \mathbb{1}_{Y \leq C_Y}$ the event time indicator in the control group.

We observe one realization $(\tilde{x}, \tilde{y}, e_X, e_Y)$ of the random variables $\left(\tilde{X}, \tilde{Y}, e_X, e_Y \right)$. We use the short notation $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$. We assume to know the expected survival in each group (respectively S_C and S_T) at each timepoint.

4.2.1   Case: $S_T = 0, S_C = 1$

Probability in favor of the treatment:

$$
\begin{aligned}
\mathbb{P}\left[X \geq Y + \tau \middle| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C \right] &= \mathbb{P}\left[X \geq \tilde{y} + \tau \middle| X > \tilde{x} \right] \\
&= \frac{\mathbb{P}\left[X \geq \tilde{y} + \tau, X > \tilde{x} \right]}{\mathbb{P}\left[X > \tilde{x} \right]} \\
&= \begin{cases} 1 \text{ if } \tilde{x} \geq \tilde{y} + \tau \\ \frac{S_T((\tilde{y}+\tau)_-)}{S_T(\tilde{x})} \text{ if } \tilde{x} < \tilde{y} + \tau \end{cases}
\end{aligned}
$$

Probability in favor of the control:

$$
\begin{aligned}
\mathbb{P}\left[Y \geq X + \tau \middle| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C \right] &= \mathbb{P}\left[\tilde{y} \geq X + \tau \middle| X > \tilde{x} \right] \\
&= 1 - \mathbb{P}\left[\tilde{y} < X + \tau \middle| X > \tilde{x} \right] \\
&= 1 - \frac{\mathbb{P}\left[X > \max\left(\tilde{x}, \tilde{y} - \tau \right) \right]}{\mathbb{P}\left[X > \tilde{x} \right]} \\
&= \begin{cases} 0 \text{ if } \tilde{x} \geq \tilde{y} - \tau \\ 1 - \frac{S_T(\tilde{y}-\tau)}{S_T(\tilde{x})} \text{ if } \tilde{x} < \tilde{y} - \tau \end{cases}
\end{aligned}
$$

4.2.2   Case: $S_T = 1, S_C = 0$

By symmetry we have:
Probability in favor of the treatment:

$$
\mathbb{P}\left[X \geq Y + \tau \middle| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C \right] = \begin{cases} 0 \text{ if } \tilde{y} \geq \tilde{x} - \tau \\ 1 - \frac{S_C(\tilde{x}-\tau)}{S_C(\tilde{y})} \text{ if } \tilde{y} < \tilde{x} - \tau \end{cases}
$$

Probability in favor of the control:

$$\mathbb{P}\left[Y \geq X + \tau \big| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C\right] = \mathbb{P}\left[\tilde{y} \geq X + \tau \big| X > \tilde{x}\right]$$

$$= \begin{cases} 1 \text{ if } \tilde{y} \geq \tilde{x} + \tau \\ \frac{S_C((\tilde{x}+\tau)_-)}{S_C(\tilde{y})} \text{ if } \tilde{y} < \tilde{x} - \tau \end{cases}$$

4.2.3   Case: $S_T = 0, S_C = 0$

Probability in favor of the treatment:

$$\mathbb{P}\left[X \geq Y + \tau \big| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C\right]$$
$$= \mathbb{P}\left[X \geq Y + \tau \big| X > \tilde{x}, Y > \tilde{y}\right]$$
$$= \mathbb{P}\left[(X \geq Y + \tau) \cap (\tilde{x} \geq Y + \tau) \big| X > \tilde{x}, Y > \tilde{y}\right] + \mathbb{P}\left[(X \geq Y + \tau) \cap (\tilde{x} < Y + \tau) \big| X > \tilde{x}, Y > \tilde{y}\right]$$
$$= \mathbb{P}\left[\tilde{x} \geq Y + \tau \big| Y > \tilde{y}\right] + \frac{\mathbb{P}\left[(X \geq Y + \tau) \cap (\tilde{x} < Y + \tau) \cap (X > \tilde{x}) \cap (Y > \tilde{y})\right]}{\mathbb{P}\left[(X > \tilde{x}) \cap (Y > \tilde{y})\right]}$$
$$= \mathbb{P}\left[\tilde{x} \geq Y + \tau \big| Y > \tilde{y}\right] + \frac{\mathbb{P}\left[(X \geq Y + \tau) \cap (Y > \max(\tilde{y}, \tilde{x} - \tau))\right]}{\mathbb{P}\left[(X > \tilde{x}) \cap (Y > \tilde{y})\right]}$$

where we have used that:

$$(X \geq Y + \tau) \cap (\tilde{x} < Y + \tau) \implies X > \tilde{x}$$

Since:

$$\mathbb{P}\left[A > B\right] = \int_{-\infty}^{+\infty} \mathbb{P}\left[A > t\right] d\mathbb{P}\left[B \leq t\right]$$

$$\mathbb{P}\left[(A > B + \tau) \cap (B > b)\right] = \int_{b+}^{+\infty} \mathbb{P}\left[A > t + \tau\right] d\mathbb{P}\left[B \leq t\right]$$

$$= -\int_{b+}^{+\infty} \mathbb{P}\left[A > t + \tau\right] d\mathbb{P}\left[B > t\right]$$

we obtain for $A = X$, $B = Y$, $b = \max(\tilde{y}, \tilde{x} - \tau)$:

$$\mathbb{P}\left[X \geq Y + \tau \big| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C\right]$$
$$= \mathbb{P}\left[\tilde{x} \geq Y + \tau \big| Y > \tilde{y}\right] - \frac{\int_{\max(\tilde{y}, \tilde{x}-\tau)^+}^{\infty} \mathbb{P}\left[(X \geq t + \tau)\right] d\mathbb{P}\left[Y > t\right]}{S_T(\tilde{x}) S_C(\tilde{y})}$$
$$= \mathbb{P}\left[\tilde{x} \geq Y + \tau \big| Y > \tilde{y}\right] - \frac{\int_{\max(\tilde{y}, \tilde{x}-\tau)^+}^{\infty} S_T((t+\tau)_-) dS_C(t)}{S_T(\tilde{x}) S_C(\tilde{y})}$$

So using the results of the case $S_T = 1, S_C = 0$ we obtain:

$$\mathbb{P}\left[X \geq Y + \tau \middle| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C\right]$$

$$= \begin{cases} -\frac{\int_{\tilde{y}+}^{\infty} S_T((t+\tau)_-)dS_C(t)}{S_T(\tilde{x})S_C(\tilde{y})} & \text{if } \tilde{y} \geq \tilde{x} - \tau \\ 1 - \frac{S_C(\tilde{x}-\tau)}{S_C(\tilde{y})} - \frac{\int_{(\tilde{x}-\tau)+}^{\infty} S_T((t+\tau)_-)dS_C(t)}{S_T(\tilde{x})S_C(\tilde{y})} & \text{if } \tilde{y} < \tilde{x} - \tau \end{cases}$$

Probability in favor of the control: By symmetry we have:

$$\mathbb{P}\left[Y \geq X + \tau \middle| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C\right]$$

$$= \begin{cases} -\frac{\int_{\tilde{x}+}^{\infty} S_C((t+\tau)_-)dS_T(t)}{S_T(\tilde{x})S_C(\tilde{y})} & \text{if } \tilde{x} \geq \tilde{y} - \tau \\ 1 - \frac{S_T(\tilde{y}-\tau)}{S_T(\tilde{x})} - \frac{\int_{(\tilde{y}-\tau)+}^{\infty} S_C((t+\tau)_-)dS_T(t)}{S_T(\tilde{x})S_C(\tilde{y})} & \text{if } \tilde{x} < \tilde{y} - \tau \end{cases}$$

9

4.2.4 Synthesis

Probability in favor of the treatment: $\mathbb{P}\left[X \geq Y + \tau \big| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C\right]$

(e_X, e_Y)	$\tilde{x} \leq \tilde{y} - \tau$	$|\tilde{x} - \tilde{y}| < \tau$	$\tilde{x} \geq \tilde{y} + \tau$
(1, 1)	0	0	1
(1, 0)	0	0	$1 - \frac{S_C(\tilde{x}-\tau)}{S_C(\tilde{y})}$
(0, 1)	$\frac{S_T((\tilde{y}+\tau)_-)}{S_T(\tilde{x})}$	$\frac{S_T((\tilde{y}+\tau)_-)}{S_T(\tilde{x})}$	1
(0, 0)	$-\frac{\int_{t>\tilde{y}}^{\infty} S_T((t+\tau)_-)dS_C(t)}{S_T(\tilde{x})S_C(\tilde{y})}$	$-\frac{\int_{t>\tilde{y}}^{\infty} S_T((t+\tau)_-)dS_C(t)}{S_T(\tilde{x})S_C(\tilde{y})}$	$1 - \frac{S_C(\tilde{x}-\tau)}{S_C(\tilde{y})} - \frac{\int_{t>\tilde{x}-\tau}^{\infty} S_T((t+\tau)_-)dS_C(t)}{S_T(\tilde{x})S_C(\tilde{y})}$

Probability in favor of the control: $\mathbb{P}\left[Y \geq X + \tau \big| \tilde{x}, \tilde{y}, e_X, e_Y, S_T, S_C\right]$

(e_X, e_Y)	$\tilde{x} \leq \tilde{y} - \tau$	$|\tilde{x} - \tilde{y}| < \tau$	$\tilde{x} \geq \tilde{y} + \tau$
(1, 1)	1	0	0
(1, 0)	1	$\frac{S_C((\tilde{x}+\tau)_-)}{S_C(\tilde{y})}$	$\frac{S_C((\tilde{x}+\tau)_-)}{S_C(\tilde{y})}$
(0, 1)	$1 - \frac{S_T(\tilde{y}-\tau)}{S_T(\tilde{x})}$	0	0
(0, 0)	$1 - \frac{S_T(\tilde{y}-\tau)}{S_T(\tilde{x})} - \frac{\int_{t>\tilde{y}-\tau}^{\infty} S_C((t+\tau)_-)dS_T(t)}{S_T(\tilde{x})S_C(\tilde{y})}$	$-\frac{\int_{t>\tilde{x}}^{\infty} S_C((t+\tau)_-)dS_T(t)}{S_T(\tilde{x})S_C(\tilde{y})}$	$-\frac{\int_{t>\tilde{x}}^{\infty} S_C((t+\tau)_-)dS_T(t)}{S_T(\tilde{x})S_C(\tilde{y})}$

Probability neutral to the treatment: $\mathbb{P}\left[|X - Y| < \tau \big| \tilde{x}, \tilde{y}, \theta, \eta, S_T, S_C\right]$

$$= 1 - \mathbb{P}\left[X \geq Y + \tau \big| \tilde{x}, \tilde{y}, \theta, \eta, S_T, S_C\right] - \mathbb{P}\left[Y \geq X + \tau \big| \tilde{x}, \tilde{y}, \theta, \eta, S_T, S_C\right]$$

10

4.3 Partially known survival curve

In the case where $x^* < y^* - \tau$, we need an estimate of $S_X(y^* - \tau)$ to compute the probability in favor of the control. If we can only have an estimate of S_X up to $x_{max} < y^* - \tau$ then we can use the following inequality:

$$S_X(x_{max}) \geq S_X(y^* - \tau)$$

$$\mathbb{P}\left[x \geq y - \tau | x \geq x^*, y = y^*\right] \geq 1 - \frac{S_X(x_{max})}{S_X(x^*)}$$

Probability of being neutral:

$$\mathbb{P}\left[|x - y| \leq \tau | x \geq x^*, y = y^*\right] = 1 - \mathbb{P}\left[x \geq y + \tau | x \geq x^*, y = y^*\right] - \mathbb{P}\left[y \geq x + \tau | x \geq x^*, y = y^*\right]$$
$$= \frac{S_X(y^* - \tau \vee x^*) - S_X(y^* + \tau \vee x^*)}{S_X(x^*)}$$

Consider the case x^* If $x_{max} > y^* - \tau$ then

$$\mathbb{P}\left[|x - y| \leq \tau | x \geq x^*, y = y^*\right] \geq \frac{S_X(y^* - \tau) - S_X(x_{max})}{S_X(x^*)}$$

otherwise

$$\mathbb{P}\left[|x - y| \leq \tau | x \geq x^*, y = y^*\right] \geq 0$$

Probability of being uninformative: It is computed as the complement to 1 of the sum of the probability of being in favor of the treatment, in favor of the control, and neutral.

EXAMPLE:

- when $x^* > y^* + \tau$, the probability of being favorable is 1 so the probability of being uninformative is 0.

- when $|x^* - y^*| < \tau$, the probability of being in favor of the control is 0. If we know the survival in the experimental group up to time y^*, then we can only say that the probability of being favorable is bounded below by 0. The probability of being neutral bounded below by $1 - S_T(y^*)/S_T(x^*)$. The probability of being uninformative is then $S_T(y^*)/S_T(x^*)$. Clearly this probability becomes small when $S_T(y^*)$ is small. The approximation by the lower bound becomes exact when $S_T(y^*)$ tends to 0.

4.4 In R

Settings:

```
alphaX <- 2
alphaY <- 1
```

Simulate data:

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

Buyse test:

```
BuyseTest(group ~ tte(time, censoring = event), data = df,
          method.inference = "none", trace = 0, scoring.rule = "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" = alphaY/alphaX, "Delta" = 2*alphaX/(alphaY+alphaX)-1)
c("HR.cox" = HR, "Delta" = (1-HR)/(1+HR))
```

```
       HR     Delta
0.5000000 0.3333333
   HR.cox     Delta
0.4918256 0.3406392
```

5 Competing risks

5.1 Theory

5.1.1 General case (no censoring)

Let consider:

- X_E^* the time to the occurrence of the event of interest in the experimental group.

- Y_E^* the time to the occurrence of the event of interest in the control group.

- X_{CR}^* the time to the occurrence of the competing event of interest in the experimental group.

- Y_{CR}^* the time to the occurrence of the competing event of interest in the control group.

Let denote $\varepsilon_X = 1 + \mathbb{1}_{X_E^* > X_{CR}^*}$ the event type indicator in the experimental group and $\varepsilon_Y = 1 + \mathbb{1}_{Y_E^* > Y_{CR}^*}$ the event type indicator in control group ($= 1$ when the cause of interest is realised first and 2 when the competing risk is realised first).

For each subject either the event of interest or the competing event is realized. We now define:

$$X = \begin{cases} X_E^* \text{ if } \varepsilon_X = 1 \\ +\infty \text{ if } \varepsilon_X = 2 \end{cases} \text{ and } Y = \begin{cases} Y_E^* \text{ if } \varepsilon_Y = 1 \\ +\infty \text{ if } \varepsilon_Y = 2 \end{cases}$$

i.e. when the event of interest is not realized we say that the time to event is infinite.

We thus have:

$$\begin{aligned}
\mathbb{P}\left[X > Y\right] =& \mathbb{P}\left[X > Y | \varepsilon_X = 1, \varepsilon_Y = 1\right] \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 1\right] \\
& + \mathbb{P}\left[X > Y | \varepsilon_X = 1, \varepsilon_Y = 2\right] \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 2\right] \\
& + \mathbb{P}\left[X > Y | \varepsilon_X = 2, \varepsilon_Y = 1\right] \mathbb{P}\left[\varepsilon_X = 2, \varepsilon_Y = 1\right] \\
& + \mathbb{P}\left[X > Y | \varepsilon_X = 2, \varepsilon_Y = 2\right] \mathbb{P}\left[\varepsilon_X = 2, \varepsilon_Y = 2\right] \\
=& \mathbb{P}\left[X > Y | \varepsilon_X = 1, \varepsilon_Y = 1\right] \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 1\right] \\
& + 0 * \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 2\right] \\
& + 1 * \mathbb{P}\left[\varepsilon_X = 2, \varepsilon_Y = 1\right] \\
& + 0 * \mathbb{P}\left[\varepsilon_X = 2, \varepsilon_Y = 2\right]
\end{aligned}$$

So $\mathbb{P}\left[X > Y\right] = \mathbb{P}\left[X > Y | \varepsilon_X = 1, \varepsilon_Y = 1\right] \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 1\right] + \mathbb{P}\left[\varepsilon_X = 2, \varepsilon_Y = 1\right]$ and:

$$\begin{aligned}
\Delta =& \left(\mathbb{P}\left[X > Y | \varepsilon_X = 1, \varepsilon_Y = 1\right] - \mathbb{P}\left[X < Y | \varepsilon_X = 1, \varepsilon_Y = 1\right]\right) \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 1\right] \\
& + \mathbb{P}\left[\varepsilon_X = 2, \varepsilon_Y = 1\right] - \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 2\right]
\end{aligned}$$

5.1.2 Exponential distribution (no censoring)

Now let's assume that:

- $X_E \sim Exp(\alpha_{E,X})$.

- $Y_E \sim Exp(\alpha_{E,Y})$.

- $X_{CR} \sim Exp(\alpha_{CR,X})$.

- $Y_{CR} \sim Exp(\alpha_{CR,Y})$.

Then:

$$\mathbb{P}\left[X_E > Y_E\right] = \mathbb{P}\left[X_E > Y_E | \varepsilon_X = 1, \varepsilon_Y = 1\right] \mathbb{P}\left[\varepsilon_X = 1, \varepsilon_Y = 1\right] + \mathbb{P}\left[\varepsilon_X = 2, \varepsilon_Y = 1\right]$$

$$= \frac{1}{(\alpha_{E,X} + \alpha_{CR,X})(\alpha_{E,Y} + \alpha_{CR,Y})} \left(\alpha_{E,X}\alpha_{E,Y}\frac{\alpha_{E,X}}{\alpha_{E,X} + \alpha_{E,Y}} + \alpha_{CR,X}\alpha_{E,Y}\right)$$

Just for comparison let's compare to the cumulative incidence. First we only consider one group and two competing events whose times to event follow an exponential distribution:

- $T_E \sim Exp(\alpha_E)$. The corresponding hazard function is $\lambda(t) = \alpha_E$.

- $T_{CR} \sim Exp(\alpha_{CR})$. The corresponding hazard function is $\lambda(t) = \alpha_{CR}$.

The cumulative incidence function can be written:

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

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

Then applying this formula in the case of two groups gives:

$$CIF_1(t|group = X) = \frac{\alpha_{E,X}}{\alpha_{E,X} + \alpha_{CR,X}} \left(1 - \exp(-(\alpha_{E,X} + \alpha_{CR,X}) * t_-)\right)$$

$$CIF_1(t|group = Y) = \frac{\alpha_{E,Y}}{\alpha_{E,Y} + \alpha_{CR,Y}} \left(1 - \exp(-(\alpha_{E,Y} + \alpha_{CR,Y}) * t_-)\right)$$

5.2 In R

5.2.1 BuyseTest (no censoring)

Setting:

```
alphaE.X <- 2
alphaCR.X <- 1
alphaE.Y <- 3
alphaCR.Y <- 2
```

Simulate data:

```
set.seed(10)
df <- rbind(data.frame(time1 = rexp(n, rate = alphaE.X), time2 = rexp(n, rate =
    alphaCR.X), group = "1"),
            data.frame(time1 = rexp(n, rate = alphaE.Y), time2 = rexp(n, rate =
    alphaCR.Y), group = "2"))
df$time <- pmin(df$time1,df$time2) ## first event
df$event <- (df$time2<df$time1)+1 ## type of event
```

BuyseTest:

```
e.BT <- BuyseTest(group ~ tte(time, censoring = event), data = df,
                  method.inference = "none", scoring.rule = "Gehan",
                  trace = 0)
summary(e.BT, percentage = TRUE)
```

```
    Generalized pairwise comparisons with 1 endpoint

> statistic       : net benefit (delta: endpoint specific, Delta: global)
> null hypothesis : Delta == 0
> treatment groups: 1 (control) vs. 2 (treatment)
> censored pairs  : uninformative pairs
> results
endpoint threshold total favorable unfavorable neutral uninf   delta   Delta
    time     1e-12   100      41.6       45.12   13.28     0 -0.0352 -0.0352
```

Note that without censoring one can get the same results by treating time as a continuous variable that take value ∞ when the competing risk is observed:

```
df$timeXX <- df$time
df$timeXX[df$event==2] <- max(df$time)+1
e.BT.bis <- BuyseTest(group ~ cont(timeXX), data = df,
                  method.inference = "none", trace = 0)
summary(e.BT.bis, percentage = TRUE)
```

```
    Generalized pairwise comparisons with 1 endpoint

> statistic       : net benefit (delta: endpoint specific, Delta: global)
> null hypothesis : Delta == 0
```

```
> treatment groups: 1 (control) vs. 2 (treatment)
> results
endpoint threshold total favorable unfavorable neutral uninf    delta    Delta
  timeXX     1e-12   100      41.6       45.12   13.28       0 -0.0352 -0.0352
```

Expected:

```
weight <- (alphaE.X+alphaCR.X)*(alphaE.Y+alphaCR.Y)
exp <- list()
exp$favorable <- 1/weight*(alphaE.X*alphaE.Y*alphaE.X/(alphaE.X+alphaE.Y)+(alphaE.X*
    alphaCR.Y))
exp$unfavorable <- 1/weight*(alphaE.X*alphaE.Y*alphaE.Y/(alphaE.X+alphaE.Y)+(alphaE.Y*
    alphaCR.X))
exp$neutral <- alphaCR.X*alphaCR.Y/weight

100*unlist(exp)
```

```
favorable unfavorable      neutral
 42.66667    44.00000     13.33333
```

5.2.2   BuyseTest (with censoring)

Simulate data:

```
df$eventC <- df$event
df$eventC[rbinom(n, size = 1, prob = 0.2)==1] <- 0
```

BuyseTest (biased):

```
e.BTC <- BuyseTest(group ~ tte(time, censoring = eventC), data = df,
                   method.inference = "none", scoring.rule = "Gehan",
                   trace = 0)
summary(e.BTC, percentage = TRUE)
```

```
    Generalized pairwise comparisons with 1 endpoint

> statistic        : net benefit (delta: endpoint specific, Delta: global)
> null hypothesis : Delta == 0
> treatment groups: 1 (control) vs. 2 (treatment)
> censored pairs  : uninformative pairs
> uninformative pairs: no contribution at the current endpoint, analyzed at later endpoints (if a
> results
endpoint threshold total favorable unfavorable neutral uninf    delta    Delta
    time     1e-12   100      31.1       35.15    8.65  25.1 -0.0406 -0.0406
```

BuyseTest (unbiased):

```
e.BTCC <- BuyseTest(group ~ tte(time, censoring = eventC), data = df,
                    method.inference = "none", scoring.rule = "Gehan",
                    correction.uninf = 2,
                    trace = 0)
summary(e.BTCC, percentage = TRUE)
```

    Generalized pairwise comparisons with 1 endpoint

> statistic   : net benefit (delta: endpoint specific, Delta: global)
> null hypothesis : Delta == 0
> treatment groups: 1 (control) vs. 2 (treatment)
> censored pairs : uninformative pairs
> uninformative pairs: no contribution, their weight is passed to the informative pairs using IPC
> results
endpoint threshold total favorable unfavorable neutral uninf   delta   Delta
    time   1e-12   100   41.52   46.94   11.54   0 -0.0542 -0.0542

5.2.3 Cumulative incidence

Settings:
```
alphaE <- 2
alphaCR <- 1
```

Simulate data:
```
set.seed(10)
df <- data.frame(time1 = rexp(n, rate = alphaE), time2 = rexp(n, rate = alphaCR),
    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 = alphaE/(alphaE+alphaCR)*(1-exp(-(alphaE+alphaCR)*(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
```

6 Inverse probability weighting

In case of censoring we can use an inverse probability weighting approach. Let denote $\delta_{c,X}$ (resp. $\delta_{c,Y}$) the indicator of no censoring relative to $\tilde{X}$ (resp $\tilde{Y}$), $\tilde{X}_E$ and $\tilde{Y}_E$ the censored event time. We can use inverse probability weighting to compute the net benefit:

$$\Delta^{IPW} = \frac{\delta_{c,\tilde{X}}\delta_{c,\tilde{Y}}}{\mathbb{P}\left[\delta_{c,\tilde{X}}\right]\mathbb{P}\left[\delta_{c,\tilde{Y}}\right]}(\mathbb{1}_{\tilde{Y}>\tilde{X}} - \mathbb{1}_{\tilde{Y}<\tilde{X}})$$

$$= \begin{cases} \frac{1}{\mathbb{P}\left[\delta_{c,\tilde{X}}\right]\mathbb{P}\left[\delta_{c,\tilde{Y}}\right]}(\mathbb{1}_{Y>X} - \mathbb{1}_{Y<X}), \text{ if no censoring} \\ 0, \text{ if censoring} \end{cases}$$

This is equivalent to weight the informative pairs (i.e. favorable, unfavorable and neutral) by the inverse of the complement of the probability of being uninformative. This is what is done by the argument `correction.tte` of `BuyseTest`. This works whenever the censoring mechanism is independent of the event times and we have a consistent estimate of $\mathbb{P}\left[\delta_c\right]$ since:

$$\mathbb{E}\left[\Delta^{IPW}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\delta_{c,\tilde{X}}\delta_{c,\tilde{Y}}}{\mathbb{P}\left[\delta_{c,\tilde{X}}\right]\mathbb{P}\left[\delta_{c,\tilde{Y}}\right]}(\mathbb{1}_{\tilde{Y}>\tilde{X}} - \mathbb{1}_{\tilde{Y}<\tilde{X}})\bigg|\tilde{X},\tilde{Y}\right]\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\frac{\delta_{c,\tilde{X}}\delta_{c,\tilde{Y}}}{\mathbb{P}\left[\delta_{c,\tilde{X}}\right]\mathbb{P}\left[\delta_{c,\tilde{Y}}\right]}\bigg|\tilde{X},\tilde{Y}\right]\right]\mathbb{E}\left[\mathbb{1}_{Y>X} - \mathbb{1}_{Y<X}\right]$$

$$= \frac{\mathbb{E}\left[\delta_{c,\tilde{X}}\delta_{c,\tilde{Y}}\right]}{\mathbb{P}\left[\delta_{c,\tilde{X}}\right]\mathbb{P}\left[\delta_{c,\tilde{Y}}\right]}\Delta = \frac{\mathbb{E}\left[\delta_{c,\tilde{X}}\right]\mathbb{E}\left[\delta_{c,\tilde{Y}}\right]}{\mathbb{P}\left[\delta_{c,\tilde{X}}\right]\mathbb{P}\left[\delta_{c,\tilde{Y}}\right]}\Delta$$

$$= \Delta$$

where we used the law of total expectation (first line) and the independence between the censoring mecanisms.

7 Asymptotic distribution

In this section we restrict ourself to the GPC as defined in (Buyse, 2010), i.e. we do not consider Peron scoring rule nor any correction (like inverse probability weighting).

7.1 Theory

We consider two independent samples $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$ where the first one contains iid realisations of a random variable X and the second one contains iid realisations of a second variable Y. For each realisation we observe p endpoints.

The estimator of the net benefit can be written as the difference between two estimators:

$$\hat{\Delta}_\tau = \widehat{\mathbb{P}}[X \geq Y + \tau] - \widehat{\mathbb{P}}[Y \geq Y + \tau]$$

We denote by ϕ_k the scoring rule relative to $\mathbb{P}[X \geq Y + \tau]$ for the endpoint k, e.g. $\phi_k(x_1, y_1) = \mathbb{1}_{x_1 > y_1}$ for a binary endpoint. The scoring rule may depend of additional arguments, e.g. a threshold τ but this will be ignored for the moment. Finally, we denote by k_{ij} the endpoint at which the pair (i, j) is classified as favorable or unfavorable. If this does not happen then $k_{ij} = p$. With this notations, the estimator $\widehat{\mathbb{P}}[X \geq Y + \tau]$ can be written as a U-statistic:

$$\widehat{\mathbb{P}}[X \geq Y + \tau] = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \phi_{k_{ij}}(x_1, y_1)$$

This is a two sample U-statistic of order $(1,1)$ with kernel $\phi_{k_{ij}}(x_1, y_1)$ (trivially symmetric in x and y separately). From the U-statistic theory (e.g. see appendix A), it follows that $\widehat{\mathbb{P}}[X \geq Y + \tau]$ is unbiased, normally distributed, and its iid decomposition is the Hájek projection:

$$\begin{aligned}
H^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau]) &= \frac{1}{m} \sum_{i=1}^{m} \left(\mathbb{E}\left[\phi_{k_{ij}}(x_i, y_j) \Big| x_i \right] - \mathbb{P}\left[X \geq Y + \tau \right] \right) \\
&+ \frac{1}{n} \sum_{j=1}^{n} \left(\mathbb{E}\left[\phi_{k_{ij}}(x_i, y_j) \Big| y_j \right] - \mathbb{P}\left[X \geq Y + \tau \right] \right) \\
&= \sum_{l=1}^{m+n} H_l^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau])
\end{aligned}$$

where $H_l^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau])$ are the individual terms of the iid decomposition. For instance in the binary case, the term relative to the $i-th$ observation of the experimental group is:

$$H_i^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau]) = \frac{\mathbb{E}\left[\phi_{k_{ij}}(x_i, y) \Big| x_i \right] - \mathbb{P}\left[X \geq Y + \tau \right]}{m} = \begin{cases} \frac{1 - p_y - \mathbb{P}[X \geq Y + \tau]}{m} & \text{if } x = 1 \\ \frac{-\mathbb{P}[X \geq Y + \tau]}{m} & \text{if } x = 0 \end{cases}$$

where p_y is the proportion of 1 in the control group.

7.2 Variance estimator based on the first order H-decomposition

The Hájek projection can be used to estimate the variance of $\widehat{\mathbb{P}}[X \geq Y + \tau]$. Indeed, since:

$$\widehat{\mathbb{P}}[X \geq Y + \tau] = H^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau]) + o_p\left(n^{-\frac{1}{2}}\right)$$

We obtain that asymptotically:

$$\mathbb{V}ar\left[\widehat{\mathbb{P}}[X \geq Y + \tau]\right] = \frac{\sigma_{1,0}^2}{n} + \frac{\sigma_{0,1}^2}{m} = \sum_{l=1}^{m+n}\left(H_l^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau])\right)^2$$

$$\text{where } \sigma_{1,0}^2 = \frac{1}{m}\sum_{i=1}^{m}\left(\mathbb{E}\left[\phi_{k_{ij}}(x_i, y)\Big| x_i\right] - \mathbb{P}[X \geq Y + \tau]\right)^2 = m\sum_{i=1}^{m}\left(H_i^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau])\right)^2$$

$$\text{and } \sigma_{0,1}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\mathbb{E}\left[\phi_{k_{ij}}(x_i, y)\Big| y_j\right] - \mathbb{P}[X \geq Y + \tau]\right)^2 = n\sum_{j=1}^{n}\left(H_j^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau])\right)^2$$

Similarly we obtain:

$$\mathbb{V}ar\left[\widehat{\mathbb{P}}[Y \geq X + \tau]\right] = \sum_{l=1}^{m+n}\left(H_l^{(1)}(\widehat{\mathbb{P}}[Y \geq X + \tau])\right)^2$$

$$\mathbb{C}ov\left[\widehat{\mathbb{P}}[X \geq Y + \tau], \widehat{\mathbb{P}}[Y \geq X + \tau]\right] = \sum_{l=1}^{m+n}\left(H_l^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau])H_l^{(1)}(\widehat{\mathbb{P}}[Y \geq X + \tau])\right)$$

7.3 Variance estimator based on the second order H-decomposition

An better estimator (i.e. unbiased) of the variance of $\widehat{\mathbb{P}}[X \geq Y + \tau]$ can be obtained using a second order H-decomposition. As explained in the appendix A, the formula for the variance becomes:

$$\mathbb{V}ar\left[\widehat{\mathbb{P}}[X \geq Y + \tau]\right] = \frac{1}{nm}\left((m-1)\sigma_{1,0}^2 + (n-1)\sigma_{0,1}^2 + \sigma_{1,1}^2\right)$$

$$\text{where } \sigma_{1,1}^2 = \mathbb{E}\left[\phi_{k_{ij}}(x_i, y_j)^2\right] - \mathbb{P}[X \geq Y + \tau]^2$$

Note that since we consider binary scores, $\phi_{k_{ij}}(x_i, y_j)^2 = \phi_{k_{ij}}(x_i, y_j)$ so $\sigma_{1,1}^2 = \mathbb{P}[X \geq Y + \tau](1 - \mathbb{P}[X \geq Y + \tau])$. When computing the covariance $\sigma_{1,1}^2 = -\mathbb{P}[X \geq Y + \tau]\mathbb{P}[Y \geq X + \tau]$ because $\mathbb{1}_{x_i \geq y_j + \tau}\mathbb{1}_{y_j \geq x_i + \tau} = 0$.

7.4 Example

Let's consider a case with 2 observations per group:

```
d <- data.table(id = 1:4, group = c("C","C","T","T"), toxicity = c(1,0,1,0))
d
```

```
   id group toxicity
1:  1     C        1
2:  2     C        0
3:  3     T        1
4:  4     T        0
```

We can form 4 pairs:

```
d2 <- data.table(pair = c("3-1","4-1","3-2","4-2"),
                 type = c("1-1","0-1","1-0","0-0"),
                 favorable = c(0,0,1,0),
                 unfavorable = c(0,1,0,0))
d2
```

```
   pair type favorable unfavorable
1:  3-1  1-1         0           0
2:  4-1  0-1         0           1
3:  3-2  1-0         1           0
4:  4-2  0-0         0           0
```

So $U = \mathbb{P}[X > Y]$ equals:

```
U <- 1/4
```

and the iid terms are:

$$H_1^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau]) = \frac{1}{n}\left(\mathbb{E}\left[\mathbb{1}_{x>y_1}\middle|y_1\right] - U\right) = \frac{\frac{\mathbb{1}_{x_1>y_1}+\mathbb{1}_{x_2>y_1}}{2} - 1/4}{2} = \frac{0 - 1/4}{2} = -1/8$$

$$H_2^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau]) = \frac{1}{n}\left(\mathbb{E}\left[\mathbb{1}_{x>y_2}\middle|y_2\right] - U\right) = \frac{\frac{\mathbb{1}_{x_1>y_2}+\mathbb{1}_{x_2>y_2}}{2} - 1/4}{2} = \frac{1/2 - 1/4}{2} = 1/8$$

$$H_3^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau]) = \frac{1}{m}\left(\mathbb{E}\left[\mathbb{1}_{x_1>y}\middle|x_1\right] - U\right) = \frac{\frac{\mathbb{1}_{x_1>y_1}+\mathbb{1}_{x_1>y_2}}{2} - 1/4}{2} = \frac{1/2 - 1/4}{2} = 1/8$$

$$H_4^{(1)}(\widehat{\mathbb{P}}[X \geq Y + \tau]) = \frac{1}{m}\left(\mathbb{E}\left[\mathbb{1}_{x_2>y}\middle|x_2\right] - U\right) = \frac{\frac{\mathbb{1}_{x_2>y_1}+\mathbb{1}_{x_2>y_2}}{2} - 1/4}{2} = \frac{0 - 1/4}{2} = -1/8$$

We can use the method `iid` to extract the iid decomposition in the BuyseTest package:

```
BuyseTest.options(order.Hprojection = 1)
e.BT <- BuyseTest(group ~ bin(toxicity), data = d,
                  keep.pairScore = TRUE,
                  method.inference = "u-statistic", trace = 0)
iid(e.BT)
```

```
       favorable unfavorable
[1,]     -0.125        0.125
[2,]      0.125       -0.125
[3,]      0.125       -0.125
[4,]     -0.125        0.125
```

This leads to the following estimates for the variance covariance:

```
crossprod(iid(e.BT))
```

```
            favorable unfavorable
favorable      0.0625     -0.0625
unfavorable   -0.0625      0.0625
```

Which is precisely what is stored in `e.BT`:

```
e.BT@covariance
```

```
             favorable unfavorable covariance netBenefit winRatio
toxicity_0.5    0.0625      0.0625    -0.0625       0.25        4
```

Note that we could also estimate the variance via the formula given in (Bebu and Lachin, 2015), e.g.:

$$\sigma^2_{favorable} = \mathbb{P}\left[X \geq Y_1, X \geq Y_2\right] - \mathbb{P}\left[X \geq Y\right]^2$$
$$= 1/8 - 1/16 = 0.0625$$

Indeed to compute $\mathbb{P}\left[X \geq Y_1, X \geq Y_2\right]$ we distinguish 2*2*2=8 cases ($X \in \{x_1, x_2\}$, $Y_1 \in \{y_1, y_2\}$, and $Y_2 \in \{y_1, y_2\}$) and only one satisfyies $X \geq Y_1, X \geq Y_2$ (when $X = x_1$ and $Y_1 = Y_2 = y_2$). This is what is performed when calling:

```
e2.BT <- BuyseTest(group ~ bin(toxicity), data = d,
                keep.pairScore = TRUE,
                method.inference = "u-statistic-bebu", trace = 0)
e2.BT@covariance
```

```
             favorable unfavorable covariance netBenefit winRatio
toxicity_0.5    0.0625      0.0625    -0.0625       0.25        4
```

Let's now consider the second order decomposition. For the variance of $\widehat{\mathbb{P}}[Y \geq X + \tau]$:

- $\sigma^2_{1,1} = 1/4(1 - 1/4) = 3/16$

- $\sigma^2_{1,0} = \frac{(-1/4)^2 + (1/4)^2}{2} = 1/16$

- $\sigma^2_{0,1} = \frac{(1/4)^2 + (-1/4)^2}{2} = 1/16$

So $\sigma^2_{favorable}$ becomes $\frac{1}{2*2}\left((2-1)\frac{1}{16}+(2-1)\frac{1}{16}+\frac{3}{16}\right)=\frac{5}{64}$

For the covariance between $\widehat{\mathbb{P}}[Y \geq X + \tau]$ and $\widehat{\mathbb{P}}[X \geq Y + \tau]$:

- $\sigma^2_{1,1} = -(1/4)(1/4) = -1/16$

- $\sigma^2_{1,0} = -1/16$

- $\sigma^2_{0,1} = -1/16$

So the covariance betwen the estimators equals $\frac{1}{2*2}\left((2-1)\frac{-1}{16}+(2-1)\frac{-1}{16}-\frac{1}{16}\right)=\frac{3}{64}$. This is exactly what BuyseTest outputs:

```
BuyseTest.options(order.Hprojection = 2)
e.BT <- BuyseTest(group ~ bin(toxicity), data = d,
                  keep.pairScore = TRUE,
                  method.inference = "u-statistic", trace = 0)
e.BT@covariance
```

	favorable	unfavorable	covariance	netBenefit	winRatio
toxicity_0.5	0.078125	0.078125	-0.046875	0.25	4

7.5 Type 1 error in finite sample

7.5.1 Binary endpoint

```
tpsBin <- system.time(
    eBin.power <- powerBuyseTest(sim = simBuyseTest, n.rep = 1e3, cpus = 4,
                                 formula = Treatment ~ bin(toxicity),
                                 sample.size = c(10,25,50,100,250),
                                 method.inference = "u-statistic", trace = 0,
                                 transform = c(TRUE,FALSE), order.Hprojection = 1:2)
)
```

```
tpsBin
```

```
user   system elapsed
1.47     0.14  211.06
```

```
summary(eBin.power, statistic = c("netBenefit","winRatio"),
        legend = FALSE, col.rep = FALSE)
```

```
        Simulation study with Generalized pairwise comparison
```

```
> statistic   : net benefit
```

n.T	n.C	mean.estimate	sd.estimate	order	mean.se	rejection (FALSE)	rejection (TRUE)
10	10	0.0023	0.2235	1	0.2116	0.085	0.113
				2	0.2116	0.085	0.113
25	25	-6e-04	0.1482	1	0.1385	0.084	0.089
				2	0.1385	0.084	0.089
50	50	-0.0015	0.1003	1	0.0990	0.059	0.059
				2	0.0990	0.059	0.059
100	100	-0.0018	0.0694	1	0.0704	0.044	0.044
				2	0.0704	0.044	0.044
250	250	-0.0011	0.0423	1	0.0446	0.045	0.045
				2	0.0446	0.045	0.045

```
> statistic   : win ratio
```

n.T	n.C	mean.estimate	sd.estimate	order	mean.se	rejection (FALSE)	rejection (TRUE)
10	10	1.6606	2.2207	1	1.6772	0.1301	0.0381
				2	1.6540	0.1301	0.0381
25	25	1.2083	0.8376	1	0.7044	0.1120	0.0620
				2	0.7035	0.1120	0.0620
50	50	1.0795	0.4534	1	0.4366	0.0730	0.0590
				2	0.4365	0.0730	0.0590
100	100	1.0327	0.2948	1	0.2937	0.0520	0.0440
				2	0.2936	0.0520	0.0440
250	250	1.0099	0.1715	1	0.1810	0.0490	0.0450
				2	0.1810	0.0490	0.0450

7.5.2 Continuous endpoint

```
tpsCont <- system.time(
    eCont.power <- powerBuyseTest(sim = simBuyseTest, n.rep = 1e3, cpus = 4,
                               formula = Treatment ~ cont(score),
                               sample.size = c(10,25,50,100,250),
                               method.inference = "u-statistic", trace = 0,
                               transform = c(TRUE,FALSE), order.Hprojection = 1:2)
)
```

```
tpsCont
```

```
user   system elapsed
1.86    0.16  195.00
```

```
summary(eCont.power, statistic = c("netBenefit","winRatio"),
        legend = FALSE, col.rep = FALSE)
```

```
        Simulation study with Generalized pairwise comparison
```

> statistic   : net benefit

n.T	n.C	mean.estimate	sd.estimate	order	mean.se	rejection (FALSE)	rejection (TRUE)
10	10	0.0056	0.2642	1	0.2562	0.076	0.130
				2	0.2615	0.073	0.130
25	25	0.0048	0.1593	1	0.1632	0.061	0.088
				2	0.1647	0.054	0.087
50	50	0.0063	0.1156	1	0.1154	0.064	0.080
				2	0.1160	0.060	0.080
100	100	0.0015	0.0825	1	0.0816	0.054	0.062
				2	0.0818	0.054	0.062
250	250	6e-04	0.052	1	0.0516	0.050	0.053
				2	0.0517	0.050	0.053

> statistic   : win ratio

n.T	n.C	mean.estimate	sd.estimate	order	mean.se	rejection (FALSE)	rejection (TRUE)
10	10	1.1973	0.79	1	0.6710	0.090	0.041
				2	0.6856	0.082	0.033
25	25	1.0652	0.3568	1	0.3574	0.063	0.036
				2	0.3607	0.062	0.033
50	50	1.041	0.2466	1	0.2438	0.053	0.054
				2	0.2449	0.052	0.053
100	100	1.0167	0.1676	1	0.1672	0.054	0.049
				2	0.1676	0.053	0.048
250	250	1.0067	0.1047	1	0.1042	0.050	0.049
				2	0.1044	0.050	0.049

7.5.3 Time to event endpoint (Gehan method)

```
tpsGehan <- system.time(
    eGehan.power <- powerBuyseTest(sim = simBuyseTest, n.rep = 1e3, cpus = 4,
                                   formula = Treatment ~ tte(eventtime,
                                                             censoring = status),
                                   scoring.rule = "Gehan",
                                   sample.size = c(10,25,50,100,250),
                                   method.inference = "u-statistic", trace = 0
                                   transform = c(TRUE,FALSE), order.Hprojection = 1:2)
)
```

```
tpsGehan
```

```
user   system elapsed
1.38     0.25  177.58
```

```
summary(eGehan.power, statistic = c("netBenefit","winRatio"),
        legend = FALSE, col.rep = FALSE)
```

```
        Simulation study with Generalized pairwise comparison
```

```
 > statistic   : net benefit
   n.T n.C rep.estimate rep.se mean.estimate sd.estimate mean.se rejection.rate
1:  10  10          1000   1000      -0.003120     0.14812 0.14413          0.087
2:  50  50          1000   1000       0.001308     0.06445 0.06620          0.052
3: 100 100          1000   1000      -0.000690     0.04785 0.04690          0.049
4: 250 250          1000   1000      -0.000647     0.02929 0.02978          0.050
```

```
 > statistic   : win ratio
   n.T n.C rep.estimate rep.se mean.estimate sd.estimate mean.se rejection.rate
1:  10  10           974    974         1.873      3.2268  2.0924        0.04339
2:  50  50          1000   1000         1.092      0.4696  0.4510        0.04500
3: 100 100          1000   1000         1.038      0.3045  0.2983        0.04500
4: 250 250          1000   1000         1.012      0.1818  0.1819        0.04900
```

7.5.4 Time to event endpoint (Peron method)

```
tpsPeron <- system.time(
    ePeron.power <- powerBuyseTest(sim = simBuyseTest, n.rep = 1e3, cpus = 4,
                                   formula = Treatment ~ tte(eventtime,
                                                             censoring = status),
                                   scoring.rule = "Peron",
                                   sample.size = c(10,25,50,100,250),
                                   method.inference = "u-statistic", trace = 0
                                   transform = c(TRUE,FALSE), order.Hprojection = 1:2)
)
```

```
tpsPeron
```

```
user   system elapsed
1.16     0.13  198.24
```

```
summary(ePeron.power, statistic = c("netBenefit","winRatio"),
        legend = FALSE, col.rep = FALSE)
```

```
        Simulation study with Generalized pairwise comparison

 > statistic   : net benefit
   n.T n.C rep.estimate rep.se mean.estimate sd.estimate mean.se rejection.rate
1:  10  10         1000   1000     0.0048942     0.22201 0.19280          0.172
2:  50  50         1000   1000     0.0028917     0.11738 0.08918          0.167
3: 100 100         1000   1000     0.0004554     0.10206 0.06325          0.250
4: 250 250         1000   1000     0.0036080     0.08643 0.04016          0.358

 > statistic   : win ratio
   n.T n.C rep.estimate rep.se mean.estimate sd.estimate mean.se rejection.rate
1:  10  10         1000   1000         1.137      0.6089 0.47252          0.086
2:  50  50         1000   1000         1.037      0.2597 0.19203          0.144
3: 100 100         1000   1000         1.023      0.2186 0.13397          0.231
4: 250 250         1000   1000         1.023      0.1838 0.08482          0.354
```

8 References

Bebu, I. and Lachin, J. M. (2015). Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics*, 17(1):178–187.

Buyse, M. (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in medicine*, 29(30):3245–3257.

Lee, A. J. (1990). U-statistics: Theory and practice. statistics: Textbooks and monographs 110. *Dekker, New York. MR*, 10754:17.

A   Recall on the U-statistic theory

This recall is based on chapter 1 of Lee (1990).

A.1   Motivating example

We will illustrate basic results on U-statistics with the following motivating question: "what is the asymptotic distribution of the empirical variance estimator?". For a more concrete example, imagine that we want to provide an estimate with its 95% confidence interval of the variability in cholesterol measurements. We assume that we are able to collect a sample of n independent and identically distributed (iid) realisations $(x_1, \ldots, x_n)$ of the random variable cholesterol, denoted X. We ignore any measurement error.

A.2   Estimate, estimator, and functionnal

We can compute an **estimate** of the variance using the following **estimators** $\hat{\mu}$ and $\hat{\sigma}^2$:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \tag{2}$$

Given a dataset the estimator $\hat{\sigma}^2$ outputs a deterministic (i.e. not random) quantity, called the estimate of the variance. For instance if we observe:

```
x <- c(1,3,5,2,1,3)
```

then s equals:

```
mu <- mean(x)
sigma2 <- var(x)
sigma2
```

[1] 2.3

   In general the value of the estimate depends on the dataset. The estimator acts like a function f_n that takes as argument some data and output a quantity of interest. This is often refer to as a **functionnal**, e.g. $\hat{\sigma}^2 = f_n(x_1, \ldots, x_n)$. Here we use the hat notation to emphasise that $\hat{\sigma}^2$ is a random quantity: for each new realisation $(x_1, \ldots, x_n)$ of X corresponds a realisation for $\hat{\sigma}^2$ i.e. a possibly different value for the variance. If mechanism generating the data has cumulative distribution function F then we can also define the true value as $\sigma^2 = f_{\sigma^2}(F)$ (which is a deterministic value) where:

$$\mu(F) = f_\mu(F) = \int_{-\infty}^{+\infty} x \, dF(x) \tag{3}$$

$$\sigma^2(F) = f_{\sigma^2}(F) = \int_{-\infty}^{+\infty} (x - f_\mu(F))^2 \, dF(x) \tag{4}$$

This can be understood as the limit $f(F) = \lim_{n \to \infty} f_n(x_1, \ldots, x_n)$. Because σ^2 and f_{σ^2} are very close quantities we will not distinguish them in the notation, i.e. write $\sigma^2 = \sigma^2(F)$. This corresponds to formula (1) in Lee (1990).

When we observe a sample, we use it to plug-in formula (3) and (4) an approximation $\hat{F}$ of F. Usually our best guess for F is $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x \leq x_i}$ where $\mathbb{1}_.$ is the indicator function taking value 1 if . is true and 0 otherwise. One can check that when plug-in $\hat{F}$ formula (3) and (4) becomes formula (1) and (2).

To summarize:

- an estimator is a random variable whose realisation depends on the data. Its realization is called estimate.

- an estimate is a deterministic value that we obtain using the observed data (e.g. observed variability is 2.3)

- a functionnal (of an estimator) is the rule by which an estimator transforms the data into an estimate.

A.3 Aim

Using formula (1) and (2) we can easily estimate the variance based on the observed realisations of X (i.e. the data). However how can we get an confidence interval? What we want is to quantify the incertainty associated with the estimator, i.e. how the value output by the functionnal is sensitive to a change in the dataset. To do so, since the estimator $\hat{\sigma}^2$ is a random variable, we can try to characterize its distribution. This is in general difficult. It is much easier to look at the distribution of the estimator $\hat{\sigma}^2$ if we would have an infinite sample size. This is what we will do, and rely on similations to see how things go in finite sample size. As we will see, the asymptotic distribution of the variance is a Gaussian distribution with a variance that we can estimate:

```
n <- length(x)
k <- mean((x-mu)^4)
var_sigma2 <- (k-sigma2^2)/n
var_sigma2
```

[1] 0.4898611

So we obtain a 95% confidence intervals for the variance doing:

```
c(estimate = sigma2,
  lower = sigma2 + qnorm(0.025) * sqrt(var_sigma2),
  upper = sigma2 + qnorm(0.975) * sqrt(var_sigma2))
```

```
 estimate      lower      upper
2.3000000 0.9282197 3.6717803
```

We can see that it is not a very good confidence interval since it symmetric - we know that the variance is positive so it should extend more on the right side. But this only problematic in small sample sizes. In large enough sample sizes the confidence interval will be correct and we focus on this case.

In summary, we would like:

- to show that our estimator $\hat{\sigma}^2$ is asymptotically normally distributed.

- to have a formula for computing the asymptotic variance.

To do so we will use results from the theory on U-statistics.

NOTE: we can already guess that the estimator $\hat{\sigma}^2$ (as most estimators) will be asymptotically distributed because it can be expressed as a average (see formula (2)). If we would know the mean of X, then the terms $x_i - \mu$ are iid so the asymptotically normality of $\hat{\sigma}^2$ follows from the central limit theorem. It does not give us a formula for the asymptotic variance though.

A.4   Definition of a U-statistic and examples

A U-statistic with kernel h of order k is an estimator of the form:

$$\hat{U} = \frac{1}{\binom{n}{k}} \sum_{(\beta_1, \ldots, \beta_k) \in \beta} h\left(x_{\beta_1}, \ldots, x_{\beta_k}\right)$$

where β is the set of all possible permutations between k integers choosen from $\{1, \ldots, n\}$. We will also assume that the kernel is symmetric, i.e. the order of the arguments in h has no importance. Note that because the observations are iid, $\hat{U}$ is an unbiased estimator of U.

EXAMPLE 1: the simplest example of a U-statistic is the estimator of mean for which $k = 1$ and h is the identity function:

$$\hat{\mu} = \frac{1}{\binom{n}{1}} \sum_{(\beta_1) \in \{1, \ldots, n\}} x_{\beta_1} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

EXAMPLE 2: our estimator of the variance is also a U-statistic, but this requires a little bit more work to see that:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i^2 - 2x_i\hat{\mu} + \hat{\mu}^2\right)$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i^2 - 2x_i\frac{1}{n}\sum_{j=1}^{n} x_j + \hat{\mu}^2\right) = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} \left(x_i^2 - 2x_ix_j + \hat{\mu}^2\right)$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} \left((x_i - x_j)^2 - x_j^2 + \hat{\mu}^2\right) = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)^2 - \frac{1}{n-1}\sum_{j=1}^{n}\left(x_j^2 - \hat{\mu}^2\right)$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)^2 - \hat{\sigma}^2$$

$$\hat{\sigma}^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{(x_i - x_j)^2}{2} = \frac{2}{n(n-1)} \sum_{i=1}^{n}\sum_{i<j} \frac{(x_i - x_j)^2}{2} = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n}\sum_{i<j} \frac{(x_i - x_j)^2}{2}$$

where we have used that $\sum_{i=1}^{n} \left(-2x_i\hat{\mu} + \hat{\mu}^2\right) = \left(-2n\hat{\mu}^2 + n\hat{\mu}^2\right) = \sum_{i=1}^{n} -\hat{\mu}^2$. So the variance estimator is a U-statistic of order 2 with kernel $h(x_1, x_2) = \frac{(x_1 - x_2)^2}{2}$.

We can verify that numerically:

```
M.combn <- combn(length(x),2) ## create all pairs
xi_minus_xj <- apply(M.combn,2, function(iPair){(x[iPair[1]]-x[iPair[2]])})
mean(xi_minus_xj^2/2) - var(x)
```

```
[1] 2.3
```

EXAMPLE 3: another classical example of U-statistic is the signed rank statistic which enable to test non-parametrically whether the center of a distribution is 0. This corresponds to:

```
wilcox.test(x)
```

```
        Wilcoxon signed rank test with continuity correction

data:  x
V = 21, p-value = 0.03501
alternative hypothesis: true location is not equal to 0

Warning message:
In wilcox.test.default(x) : cannot compute exact p-value with ties
```

Let's take two random realisation of X and denote thoses X_1 and X_2 (they are random variables). The parameter of interest (or true value) is $U = \mathbb{P}\left[X_1 + X_2 > 0\right]$ and the corresponding estimator is:

$$\hat{U} = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{i<j} \mathbb{1}_{x_i+x_j>0}$$

A.5   A major result from the U-statistic theory

So far we have seen that our estimator for the variance was a U-statistic. We will now use the U-statistic theory to obtain its asymptotic distribution.

Theorem (adapted from Lee (1990), theorem 1 page 76)
Let $\hat{U}$ be a U-statistic of order k with non-zero first component in its H-decomposition. Then $n^{\frac{1}{2}}(\hat{U} - U)$ is asymptotically normal with mean zero and asymptotic variance σ_1^2 where σ_1^2 is the variance of the first component in the H-decomposition of $\hat{U}$.

So under the assumption that the first term of the H-decomposition of the variance is non 0 then we know that the asymptotic distribution of our variance estimator is normal and if we are able to compute the variance of the first term of the H-decomposition then we would also know the variance parameter of the asymptotic distribution. So it remains to see what is this H-decomposition and how can we characterize it.

A.6 The first term of the H-decomposition

The H-decomposition (short for Hoeffling decomposition) enables us to decompose the estimator of a U-statistic of rank k into a sum of k uncorrelated U-statistics of increasing order (from 1 to k) with variances of decreasing order in n. As a consequence the variance of the U-statistic will be asymptotically equal to the variance of the first non-0 term in the decomposition.

Before going further we introduce:

- $X_1, \ldots, X_n$ the random variables associated with each sample.

- $\mathcal{L}_2$ the space of all random variables with zero mean and finite variance.
  It is equiped with the inner product $\mathbb{C}ov\,[X, Y]$.

- the subspaces $\left(\mathcal{L}_2^{(j)}\right)_{j \in \{1, \ldots, k\}}$ where for a given $j \in \{1, \ldots, k\}$, $\mathcal{L}_2^{(j)}$ is the subspace of $\mathcal{L}_2$ containing all random variables of the form $\sum_{(\beta_1, \ldots, \beta_j) \in \beta} \psi(X_{\beta_1}, \ldots, X_{\beta_j})$ where β is the set of all possible permutations between j integers choosen from $\{1, \ldots, n\}$. For instance $\mathcal{L}_2^{(1)}$ contains the mean, $\mathcal{L}_2^{(2)}$ contains the variance, and $\mathcal{L}_2^{(j)}$ contains all U-statistics of order j with square integrable kernels.

We can now define the H-decomposition as the projection of $\hat{U} - U$ on the subspaces $\mathcal{L}_2^{(1)}$, $\mathcal{L}_2^{(2)} \cap \left(\mathcal{L}_2^{(1)}\right)^{\perp}$, $\ldots$, $\mathcal{L}_2^{(k)} \cap \left(\mathcal{L}_2^{(k-1)}\right)^{\perp}$. Here $A^{\perp}$ indicates the space orthogonal to A. So the first term of the H-decomposition, denoted $H^{(1)}$, is the projection of $\hat{U} - U$ on $\mathcal{L}_2^{(1)}$; this is also called the Hájek projection. Clearly all terms of the projection are mutually orthogonal (or uncorrelated), they are unique (it is a projection) and they correspond to U-statistics of increasing degree (from 1 to k). It remains to get a more explicit expression for these term and show that their variance are of decreasing order in n.

We now focus on the first term and show that $H^{(1)} = \sum_{i=1}^n \mathbb{E}\left[\hat{U} - U | X_i\right]$. Clearly this term belongs to $\mathcal{L}_2^{(1)}$. It remains to show that $\hat{U} - U - H^{(1)}$ is orthogonal to $\mathcal{L}_2^{(1)}$. Let consider an element $V \in \mathcal{L}_2^{(1)}$:

$$
\begin{aligned}
\mathbb{C}ov\left[\hat{U} - U - H^{(1)}, V\right] &= \mathbb{E}\left[(\hat{U} - U - H^{(1)})V\right] \\
&= \sum_{i'=1}^n \mathbb{E}\left[(\hat{U} - U - H^{(1)})\psi(X_{i'})\right] \\
&= \sum_{i'=1}^n \mathbb{E}\left[\mathbb{E}\left[\hat{U} - U - H^{(1)}\big| X_{i'}\right]\psi(X_{i'})\right]
\end{aligned}
$$

So it remains to show that $\mathbb{E}\left[\hat{U} - U\big| X_{i'}\right] = \mathbb{E}\left[H^{(1)}\big| X_{i'}\right]$. This follows from:

$$
\begin{aligned}
\mathbb{E}\left[H^{(1)}\big| X_{i'}\right] &= \mathbb{E}\left[\sum_{i=1}^n \mathbb{E}\left[\hat{U} - U | X_i\right]\bigg| X_{i'}\right] = \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\hat{U} - U | X_i\right]\big| X_{i'}\right] \\
&= \mathbb{E}\left[\hat{U} - U | X_i\right] + \sum_{i \neq i'}^n \mathbb{E}\left[\mathbb{E}\left[\hat{U} - U | X_i\right]\big| X_{i'}\right] \\
&= \mathbb{E}\left[\hat{U} - U | X_i\right] + \sum_{i \neq i'}^n \mathbb{E}\left[\mathbb{E}\left[\hat{U} - U | X_i\right]\right]^{\nearrow 0}
\end{aligned}
$$

where we have used that X_i and $X_{i'}$ are independent and $\mathbb{E}\left[\mathbb{E}\left[\hat{U} - U | X_i\right]\right] = \mathbb{E}\left[\hat{U} - U\right] = 0$.

We can now re-express the first term of the H-decomposition more explicitly:

$$
\begin{aligned}
H^{(1)} &= \sum_{i=1}^{n} \mathbb{E}\left[\hat{U} - U \Big| X_i\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\frac{1}{\binom{n}{k}} \sum_{(\beta_1,\ldots,\beta_k)\in\beta} h\left(x_{\beta_1},\ldots,x_{\beta_k}\right) - U \Big| X_i\right] \\
&= \frac{1}{\binom{n}{k}} \sum_{(\beta_1,\ldots,\beta_k)\in\beta} \sum_{i=1}^{n} \mathbb{E}\left[h\left(x_{\beta_1},\ldots,x_{\beta_k}\right)\Big| X_i\right] - U \\
&= \frac{1}{\binom{n}{k}} \sum_{(\beta_1,\ldots,\beta_k)\in\beta} \sum_{i=1}^{n} \mathbb{1}_{i\in\beta}\mathbb{E}\left[h\left(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_k\right)\Big| x_i\right] + \mathbb{1}_{i\notin\beta}*0 - U \\
&= \frac{1}{\binom{n}{k}} \sum_{i=1}^{n} \mathbb{P}\left[i\in\beta\right]\mathbb{E}\left[h\left(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_k\right)\Big| x_i\right] - U \\
&= \frac{\binom{n-1}{k-1}}{\binom{n}{k}} \sum_{i=1}^{n} \mathbb{E}\left[h\left(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_k\right)\Big| x_i\right] - U \\
H^{(1)} &= \frac{k}{n} \sum_{i=1}^{n} \mathbb{E}\left[h\left(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_k\right)\Big| x_i\right] - U
\end{aligned}
$$

Let's now compute the variance of $\hat{U}$:

$$
\begin{aligned}
\mathbb{V}ar\left[\hat{U}\right] &= \binom{n}{k}^{-2} \mathbb{V}ar\left[\sum_{(\beta_1,\ldots,\beta_k)\in\beta} h\left(x_{\beta_1},\ldots,x_{\beta_k}\right)\right] \\
&= \binom{n}{k}^{-2} \mathbb{C}ov\left[\sum_{(\beta_1,\ldots,\beta_k)\in\beta} h\left(x_{\beta_1},\ldots,x_{\beta_k}\right), \sum_{(\beta'_1,\ldots,\beta'_k)\in\beta'} h\left(x_{\beta'_1},\ldots,x_{\beta'_k}\right)\right] \\
&= \binom{n}{k}^{-2} \sum_{(\beta_1,\ldots,\beta_k)\in\beta} \sum_{(\beta'_1,\ldots,\beta'_k)\in\beta'} \mathbb{C}ov\left[h\left(x_{\beta_1},\ldots,x_{\beta_k}\right), h\left(x_{\beta'_1},\ldots,x_{\beta'_k}\right)\right]
\end{aligned}
$$

Using the symmetry of the kernel we see that the terms in the double sum only depends on the number of common observations. To determine a term with j common observations, a choose:

- k observations among the n for the first kernel: $\binom{n}{k}$ possibilities

- c common index for the two kernels among the k: $\binom{k}{c}$ possibilities

- $k - c$ observations among the remaining $n - k$ observations for the second kernel: $\binom{n-k}{k-c}$ possibilities

So denoting $\sigma_c^2 = \mathbb{C}ov\left[h\left(x_1, \ldots, x_k\right), h\left(x_1, \ldots, x_c, x'_{c+1}, \ldots, x'_k\right)\right]$ this gives:

$$
\begin{aligned}
\mathbb{V}ar\left[\hat{U}\right] &= \binom{n}{k}^{-2} \sum_{c=0}^{n} \binom{n}{k}\binom{k}{c}\binom{n-k}{k-c}\sigma_c^2 \\
&= \sum_{c=0}^{k} \frac{k!(n-k)!}{n!} \frac{k!}{c!(k-c)!} \frac{(n-k)!}{(k-2k+c)!(n-c)!}\sigma_c^2 \\
&= \sum_{c=0}^{k} \frac{k!^2}{c!(k-c)!^2} \frac{(n-k)!^2}{(n-2k+c)!n!}\sigma_c^2 \\
&= \sum_{c=0}^{k} \mathcal{O}\left(\frac{(n-k)!^2}{(n-2k+c)!n!}\right)\sigma_c^2 \\
&= \sum_{c=0}^{k} \mathcal{O}\left(\frac{(n-k)\ldots(n-2k+c+1)}{n\ldots(n-k+1)}\right)\sigma_c^2 \\
&= \sum_{c=0}^{k} \mathcal{O}\left(\frac{n^{-k+2k-c}}{n^k}\right) = \sum_{c=0}^{k} \mathcal{O}\left(n^{-c}\right)\sigma_c^2
\end{aligned}
$$

So if $\sigma_1^2 \neq 0$ then the asymptotic variance only depends on the variance of the first term, i.e.:

$$
\begin{aligned}
\mathbb{V}ar\left[\hat{U}\right] = \mathbb{V}ar\left[H^{(1)}\right] &= \frac{k^2}{n^2}\mathbb{V}ar\left[\sum_{i=1}^{n}\mathbb{E}\left[h\left(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k\right)\middle|x_i\right]\right] \\
&= \frac{k^2}{n^2}\sum_{i=1}^{n}\mathbb{V}ar\left[\mathbb{E}\left[h\left(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_k\right)\middle|x_i\right]\right] \\
&= \frac{k^2}{n^2}n\mathbb{V}ar\left[\mathbb{E}\left[h\left(x, x_2, \ldots, x_k\right)\middle|x\right]\right] \\
\mathbb{V}ar\left[\hat{U}\right] &= \frac{k^2}{n}\mathbb{V}ar\left[\mathbb{E}\left[h\left(x, x_2, \ldots, x_k\right)\middle|x\right]\right]
\end{aligned}
$$

In summary we have obtained a formula for the asymptotic variance of the U-statistic.

EXAMPLE 1: Sample mean
We first compute the Hájek projection of the mean:

$$
H_{\hat{\mu}}^{(1)} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[x_i|x_i\right] - \mu = \frac{1}{n}\sum_{i=1}^{n}x_i - \mu
$$

And then compute the asymptotic variance as:

$$
\mathbb{V}ar\left[\hat{\mu}\right] = \mathbb{V}ar\left[H_{\hat{\mu}}^{(1)}\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}ar\left[x_i - \mu\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}
$$

EXAMPLE 2: Sample variance

We first compute the Hájek projection of the variance:

$$
\begin{aligned}
H_{\hat{\sigma}^2}^{(1)} &= \frac{2}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{(x_i - X_2)^2}{2}\bigg| x_i\right] - \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[x_i^2 - 2x_i X_2 + X_2^2 \big| x_i\right] - \sigma^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(x_i^2 - 2x_i \mu + \sigma^2 + \mu^2\right) - \sigma^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left((x_i - \mu)^2 - \sigma^2\right)
\end{aligned}
$$

And then compute the asymptotic variance as:

$$
\begin{aligned}
\mathbb{V}ar\left[\hat{\sigma}^2\right] &= \mathbb{V}ar\left[H_{\hat{\sigma}^2}^{(1)}\right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}ar\left[(x_i - \mu)^2 - \sigma^2\right] \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[(x - \mu)^4\right] - \mathbb{E}\left[(x - \mu)^2\right]^2 \\
&= \frac{\mu_4 - (\sigma^2)^2}{n}
\end{aligned}
$$

where $\mu_4 = \mathbb{E}\left[(x - \mu)^4\right]$ is the fourth moment of the distribution. For a better approximation in small sample size we could account for the variance of the second term of the H-decomposition. We would obtain (Lee (1990), page 13):

$$
\mathbb{V}ar\left[\hat{\sigma}^2\right] = \frac{\mu_4}{n} - \frac{(n-3)(\sigma^2)^2}{n(n-1)}
$$

When $\frac{n-3}{n-1}$ is close to 1 then the first order approximation is sufficient.

EXAMPLE 3: Signed rank statistic

We first compute the Hájek projection of the signed rank statistic:

$$
\begin{aligned}
H_{\hat{U}}^{(1)} &= \frac{2}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{1}_{x_i + X_2 > 0}\big| x_i\right] - U = \frac{2}{n} \sum_{i=1}^{n} \mathbb{P}\left[X_2 > -x_i \big| x_i\right] - \mathbb{P}\left[X_2 > -X_1\right] \\
&= \frac{2}{n} \sum_{i=1}^{n} (1 - F(-x_i)) - \mathbb{E}_x\left[(1 - F(-x))\right]
\end{aligned}
$$

Since under the null, the distribution is symmetric $F(-x) = 1 - F(x)$:

$$
H_{\hat{U}}^{(1)} = \frac{2}{n} \sum_{i=1}^{n} F(x_i) - \mathbb{E}_x\left[F(x)\right]
$$

We will use that for continuous distribution $F(x)$ is uniformly distribution and therefore has variance $\frac{1}{12}$. So we can compute the asymptotic variance as:

$$
\mathbb{V}ar\left[\hat{U}\right] = \mathbb{V}ar\left[H_U^{(1)}\right] = \frac{4}{n^2} \sum_{i=1}^{n} \mathbb{V}ar\left[F(x_i) - \mathbb{E}_x\left[F(x)\right]\right] = \frac{4}{n^2} n \frac{1}{12} = \frac{1}{3}
$$

A.7 Two sample U-statistics

So far we have assumed that all our observations were iid. But in the case of GPC, we study two populations (experimental arm and control arm) so we can only assume to have two independent samples $x_1, x_2, \ldots, x_m$ and $y_1, y_2, \ldots, y_n$ where the first one contains iid realisations of a random variable X and the second one contains iid realisations of a second variable Y. We can now define a two-sample U-statistic of order k_x and k_y as:

$$\hat{U} = \frac{1}{\binom{m}{k_x}\binom{n}{k_y}} \sum_{(\alpha_1, \ldots, \alpha_{k_x}) \in \alpha} \sum_{(\beta_1, \ldots, \beta_{k_y}) \in \beta} h(x_{\alpha_1}, \ldots, x_{\alpha_{k_x}}, y_{\beta_1}, \ldots, y_{\beta_{k_y}})$$

where α (resp. β) is the set of all possible permutations between k_x (resp. k_y) intergers chosen from $\{1, \ldots, m\}$ (resp. $\{1, \ldots, n\}$) and the kernel $h = h(x_1, \ldots, x_{k_x}, y_1, \ldots, y_{k_y})$ is permutation symmetric in its first k_x arguments and its last k_y arguments separately. Once more it follows from the independence and iid assumptions that $\hat{U}$ is an unbiased estimator of $U = \mathbb{E}\left[h(X_1, \ldots, X_{k_x}, Y_1, \ldots, Y_{k_y})\right]$ where $X_1, \ldots, X_{k_x}$ (resp. $Y_1, \ldots, Y_{k_y}$) are the random variables associated to distinct random samples from X (resp. Y). The two-sample case is a specific case of the Generalized U-statistics introduced in section 2.2 in Lee (1990).

Many results for U-statistics extends to two sample U-statistics. For instance the Hájek projection of $\hat{U} - U$ becomes:

$$H^{(1)} = \frac{k_x}{m} \sum_{i=1}^{m} \left(\mathbb{E}\left[h(x_1, x_2, \ldots, x_{k_x}, y_1, \ldots, y_{k_y}) \big| x_i\right] - U\right) + \frac{k_y}{n} \sum_{j=1}^{n} \left(\mathbb{E}\left[h(x_1, \ldots, x_{k_x}, y_1, y_2, \ldots, y_{k_y}) \big| y_j\right] - U\right)$$

Before stating any asymptotic results, we need to define what we now mean by asymptotic (since we have two sample sizes m and n). We now mean by asymptotic that we create an increasing sequence of m and n indexed by v such that:

- $m_v \xrightarrow[v \to \infty]{} \infty$

- $n_v \xrightarrow[v \to \infty]{} \infty$

- there exist a $p \in]0; 1[$ satisfying $\frac{m}{n+m} \xrightarrow[v \to \infty]{} p$ and $\frac{n}{n+m} \xrightarrow[v \to \infty]{} 1 - p$.

Informally speaking, this means that m and n goes to infinity at the same speed. Let's denotes:

$$\mathbb{V}ar\left[\mathbb{E}\left[h(x, x_2, \ldots, x_{k_x}, y_1, \ldots, y_{k_y}) \big| x\right]\right] = \sigma_{1,0}^2$$
$$\mathbb{V}ar\left[\mathbb{E}\left[h(x_1, \ldots, x_{k_x}, y, y_2, \ldots, y_{k_y}) \big| y\right]\right] = \sigma_{0,1}^2$$

We then have the following result:

Theorem (adapted from Lee (1990), theorem 1 page 141)
Let $\hat{U}$ be a U-statistic of order k_x and k_y with non-zero first component (i.e. $\sigma_{1,0}^2 > 0$ and $\sigma_{0,1}^2 > 0$) in its H-decomposition. Then $(m+n)^{\frac{1}{2}}(\hat{U} - U)$ is asymptotically normal with mean zero and asymptotic variance $p^{-1}k_x^2\sigma_{1,0}^2 + (1-p)^{-1}k_y^2\sigma_{0,1}^2$ which is the variance of the first component in the H-decomposition of $\hat{U}$.

EXAMPLE 4: Mann-Whitney statistic

If our parameter of interest is $\mathbb{P}[X \leq Y]$ then the estimator:

$$\hat{U} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{1}_{x_i \leq y_j}$$

is a U-statistic of order $k_x = 1$ and $k_y = 1$ with kernel $h(x,y) = \mathbb{1}_{x \leq y}$ We first compute the Hájek projection of the signed rank statistic:

$$
\begin{aligned}
H_{\hat{U}}^{(1)} &= \frac{1}{m} \sum_{i=1}^{m} \left(\mathbb{E}\left[\mathbb{1}_{x_i \leq y} \middle| x_i \right] - U \right) + \frac{1}{n} \sum_{j=1}^{n} \left(\mathbb{E}\left[\mathbb{1}_{x \leq y_j} \middle| y_j \right] - U \right) \\
&= \frac{1}{m} \sum_{i=1}^{m} \left(\mathbb{P}\left[Y \geq x_i\right] - U \right) + \frac{1}{n} \sum_{j=1}^{n} \left(\mathbb{P}\left[X \leq y_j\right] - U \right) \\
&= \frac{1}{m} \sum_{i=1}^{m} \left(1 - F_{-,y}(x_i) - U \right) + \frac{1}{n} \sum_{j=1}^{n} \left(F_x(y_j) - U \right) \\
&= -\frac{1}{m} \sum_{i=1}^{m} \left(F_{-,y}(x_i) - \mathbb{E}_x[F_{-,x}(x)] \right) + \frac{1}{n} \sum_{j=1}^{n} \left(F_x(y_j) - \mathbb{E}_y[F_y(y)] \right)
\end{aligned}
$$

where F_- is the left limit of F, F_x(resp. F_y) denoting the cumulative distribution function of X (resp. Y). For continuous distributions $F_- = F$ and under the null hypothesis that $F_x = F_y$, we get that:

$$\mathbb{V}ar\left[\hat{U}\right] = \mathbb{V}ar\left[H_{\hat{U}}^{(1)}\right] = \frac{1}{m}\frac{1}{12} + \frac{1}{n}\frac{1}{12} = \frac{nm}{12(m+n)}$$

If we are not under the null we end up with the formula:

$$\mathbb{V}ar\left[\hat{U}\right] = \frac{1}{m^2} \sum_{i=1}^{m} \mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x_i \leq y} \middle| x_i\right] - U\right] + \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x \leq y_j} \middle| y_j\right] - U\right]$$

Noticing that:

$$\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{x_i \leq y} \middle| x_i\right] - U\right] = \mathbb{E}\left[\mathbb{1}_{x_i \leq y}\right] - U = 0$$

We can compute the variance as:

$$
\begin{aligned}
\mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x_i \leq y} \middle| x_i\right] - U\right] &= \mathbb{E}\left[\left(\mathbb{E}\left[\mathbb{1}_{x_i \leq y} \middle| x_i\right] - U\right)^2\right] \\
&= \int_x \left(\int_y (\mathbb{1}_{x \leq y} - U)\, dF_Y(y)\right) \left(\int_y (\mathbb{1}_{x \leq y} - U)\, dF_Y(y)\right) dF_X(x) \\
&= \int_x \left(\int_{y_1} (\mathbb{1}_{x \leq y_1} - U)\, dF_Y(y_1)\right) \left(\int_{y_2} (\mathbb{1}_{x \leq y_2} - U)\, dF_Y(y_2)\right) dF_X(x) \\
&= \int_x \int_{y_1} \int_{y_2} (\mathbb{1}_{x \leq y_1} - U)(\mathbb{1}_{x \leq y_2} - U)\, dF_Y(y_1) dF_Y(y_2) dF_X(x) \\
&= \mathbb{E}\left[(\mathbb{1}_{x \leq y_1} - U)(\mathbb{1}_{x \leq y_2} - U)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{x \leq x_1} \mathbb{1}_{x \leq y_2}\right] - \mathbb{E}\left[\mathbb{1}_{x \leq y_1}\right] U - \mathbb{E}\left[\mathbb{1}_{x \leq y_2}\right] U + U^2 \\
&= \mathbb{P}\left[x \leq y_1, x \leq y_2\right] - \mathbb{P}\left[x \leq y\right]^2
\end{aligned}
$$

So the variance is:

$$\mathbb{V}ar\left[\hat{U}\right] = \frac{1}{m}\left(\mathbb{P}\left[x \leq y_1, x \leq y_2\right] - \mathbb{P}\left[x \leq y\right]^2\right) + \frac{1}{n}\left(\mathbb{P}\left[x_1 \leq y, x_2 \leq y\right] - \mathbb{P}\left[x \leq y\right]^2\right)$$

$$= \frac{\sigma_{1,0}^2}{m} + \frac{\sigma_{0,1}^2}{n}$$

In fact we could have a more precise formula by accounting for the second term in the H-decomposition. Lee (1990) (Theorem 2 page 38, formula 2) give the general formal for the variance that becomes in the case of a two sample U statistic of degree 1:

$$\mathbb{V}ar\left[\hat{U}\right] = \frac{\sigma_{1,0}^2}{m} + \frac{\sigma_{0,1}^2}{n} + \frac{\sigma_{1,1}^2 - \sigma_{0,1}^2 - \sigma_{1,0}^2}{nm}$$

$$= \frac{1}{nm}\left((n-1)\sigma_{1,0}^2 + (m-1)\sigma_{0,1}^2 + \sigma_{1,1}^2\right)$$

where $\sigma_{1,1}^2 = \mathbb{P}\left[x \leq y\right]\left(1 - \mathbb{P}\left[x \leq y\right]\right)$. Indeed the second term of the H-decomposition would be the projection of $\mathbb{1}_{X \leq Y}$ on X, Y where we substract components of the Hájek projection to get the orthogonality between $H_{\hat{U}}^{(1)}$ and $H_{\hat{U}}^{(2)}$ (see formula 11 page 33 of Lee (1990) for a generic formula):

$$H_{\hat{U}}^{(2)} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\mathbb{E}\left[\mathbb{1}_{x_i \leq y_j}\middle|x_i, y_j\right] - U\right) - \left(\mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right] - U\right) - \left(\mathbb{E}\left[\mathbb{1}_{x \leq y_j}\middle|y_j\right] - U\right)$$

$$= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}_{x_i \leq y_j} - \mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right] - \mathbb{E}\left[\mathbb{1}_{x \leq y_j}\middle|y_j\right] + U$$

and we retrieve the formula given page 39 of Lee (1990) for the variance of the Mann-Whitney U-statistic:

$$\mathbb{V}ar\left[H_{\hat{U}}^{(2)}\right] = \frac{1}{(mn)^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{V}ar\left[\mathbb{1}_{x_i \leq y_j}\right] + \mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right]\right] + \mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x \leq y_j}\middle|y_j\right]\right]$$

$$- 2\mathbb{C}ov\left[\mathbb{1}_{x_i \leq y_j}, \mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right]\right] - 2\mathbb{C}ov\left[\mathbb{1}_{x_i \leq y_j}, \mathbb{E}\left[\mathbb{1}_{x \leq y_j}\middle|y_j\right]\right]$$

$$= \frac{1}{(mn)^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{V}ar\left[\mathbb{1}_{x_i \leq y_j}\right] - \mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right]\right] - \mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x \leq y_j}\middle|y_j\right]\right]$$

$$= \frac{\sigma_{1,1}^2 - \sigma_{0,1}^2 - \sigma_{1,0}^2}{nm}$$

where the second line follows from:

$$\mathbb{C}ov\left[\mathbb{1}_{x_i \leq y_j}, \mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right]\right] = \mathbb{E}\left[\left(\mathbb{1}_{x_i \leq y_j} - U\right)\left(\mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right] - U\right)\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\left(\mathbb{1}_{x_i \leq y_j} - U\right)\mathbb{E}\left[\mathbb{1}_{x_i \leq y} - U\middle|x_i\right]\middle|x_i\right]\right] \quad = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{x_i \leq y} - U\middle|x_i\right]^2\right]$$

$$= \mathbb{V}ar\left[\mathbb{E}\left[\mathbb{1}_{x_i \leq y}\middle|x_i\right]\right]$$

and the last line follows from:

$$\mathbb{V}ar\left[\mathbb{1}_{x_i \leq y_j}\right] = \mathbb{E}\left[\mathbb{1}_{x_i \leq y_j}^2\right] - \mathbb{E}\left[\mathbb{1}_{x_i \leq y_j}\right]^2 = \mathbb{E}\left[\mathbb{1}_{x_i \leq y_j}\right] - \mathbb{E}\left[\mathbb{1}_{x_i \leq y_j}\right]^2$$

$$= \mathbb{E}\left[\mathbb{1}_{x_i \leq y_j}\right]\left(1 - \mathbb{E}\left[\mathbb{1}_{x_i \leq y_j}\right]\right) = \mathbb{P}\left[x \leq y\right]\left(1 - \mathbb{P}\left[x \leq y\right]\right)$$

A final useful result is about the covariance between two 2-samples U-statistics.

$$
\mathbb{C}ov\left[H^{(2)}_{\hat{U}_1}, H^{(2)}_{\hat{U}_2}\right] = \frac{1}{(mn)^2} \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{C}ov\left[U_{1,ij} - \mathbb{E}\left[U_{1,ij}\big|i\right] - \mathbb{E}\left[U_{1,ij}\big|j\right], U_{2,ij} - \mathbb{E}\left[U_{2,ij}\big|i\right] - \mathbb{E}\left[U_{2,ij}\big|j\right]\right]
$$

$$
= \frac{1}{(mn)^2}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{C}ov\left[U_{1,ij}, U_{2,ij}\right] + \mathbb{C}ov\left[\mathbb{E}\left[U_{1,ij}\big|i\right], \mathbb{E}\left[U_{2,ij}\big|i\right]\right] + \mathbb{C}ov\left[\mathbb{E}\left[U_{1,ij}\big|j\right], \mathbb{E}\left[U_{2,ij}\big|j\right]\right]
$$

$$
- \mathbb{C}ov\left[U_{1,ij}, \mathbb{E}\left[U_{1,ij}\big|i\right]\right] - \mathbb{C}ov\left[U_{2,ij}, \mathbb{E}\left[U_{1,ij}\big|i\right]\right]
$$

$$
- \mathbb{C}ov\left[U_{1,ij}, \mathbb{E}\left[U_{1,ij}\big|j\right]\right] - \mathbb{C}ov\left[U_{2,ij}, \mathbb{E}\left[U_{1,ij}\big|j\right]\right]
$$

$$
= \frac{1}{(mn)^2}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{C}ov\left[U_{1,ij}, U_{2,ij}\right] - \mathbb{C}ov\left[\mathbb{E}\left[U_{1,ij}\big|i\right], \mathbb{E}\left[U_{2,ij}\big|i\right]\right] - \mathbb{C}ov\left[\mathbb{E}\left[U_{1,ij}\big|j\right], \mathbb{E}\left[U_{2,ij}\big|j\right]\right]
$$

So:

$$
\mathbb{C}ov\left[U_1, U_2\right]
$$
$$
= \mathbb{C}ov\left[H^{(1)}_{\hat{U}_1}, H^{(1)}_{\hat{U}_2}\right] + \mathbb{C}ov\left[H^{(2)}_{\hat{U}_1}, H^{(2)}_{\hat{U}_2}\right]
$$
$$
= \frac{1}{(mn)^2}\sum_{i=1}^{n}\sum_{j=1}^{m}(m-1)\mathbb{C}ov\left[\mathbb{E}\left[U_{1,ij}\big|i\right], \mathbb{E}\left[U_{2,ij}\big|i\right]\right] + (n-1)\mathbb{C}ov\left[\mathbb{E}\left[U_{1,ij}\big|j\right], \mathbb{E}\left[U_{2,ij}\big|j\right]\right] + \mathbb{C}ov\left[U_{1,ij}, U_{2,ij}\right]
$$

where the last term can be computed using that:

$$
\mathbb{C}ov\left[U_{1,ij}, U_{2,ij}\right] = \mathbb{E}\left[(U_{1,ij}U_{2,ij})^2\right] - \mathbb{E}\left[U_{1,ij}U_{2,ij}\right]^2
$$

B Information about the R session used for this document

```
sessionInfo()
```

R version 3.5.1 (2018-07-02)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 7 x64 (build 7601) Service Pack 1

Matrix products: default

locale:
[1] LC_COLLATE=Danish_Denmark.1252 LC_CTYPE=Danish_Denmark.1252   LC_MONETARY=Danish_Denmark.12
[5] LC_TIME=Danish_Denmark.1252

attached base packages:
 [1] stats4   parallel tools   stats   graphics grDevices utils   datasets methods   ba

other attached packages:
 [1] lava_1.6.4   doParallel_1.0.14   iterators_1.0.10   foreach_1.4.4   BuyseTest_1.
 [8] spelling_1.2   roxygen2_6.1.0.9000 butils.base_1.2   Rcpp_1.0.0   data.table_1

loaded via a namespace (and not attached):
 [1] compiler_3.5.1   prettyunits_1.0.2   base64enc_0.1-3   remotes_2.0.2
 [7] pkgload_1.0.2   lattice_0.20-35   memoise_1.1.0   rlang_0.3.0.1
[13] commonmark_1.6   RcppArmadillo_0.9.200.4.0 withr_2.1.2   stringr_1.3.1
[19] fs_1.2.6   grid_3.5.1   rprojroot_1.3-2   glue_1.3.0
[25] survival_2.42-6   sessioninfo_1.1.1   callr_3.0.0   purrr_0.2.5
[31] backports_1.1.2   ps_1.1.0   splines_3.5.1   assertthat_0.2
```