Calculation of processing and overprocessing

Ilya Verenich

$$\widetilde{W_\sigma} = \sum_{i=1}^{N}(P_i^r \cdot \prod_{k=1}^{i-1}(1 - P_k^r) \cdot (\sum_{j=1}^{i} E_j)) \tag{1}$$

1 For Bondora log

Minimal possible number of checks, averaged over 11 test sets :

```
round(sum(dat$minimum_check_number)/length(filenames))
```

```
## [1] 21570
```

Average number of checks per test set that one would do if they follow **our ordering**:

```
round(sum(dat$nr_checks_our)/length(filenames))
```

```
## [1] 21833
```

Average number of checks that one would do if they apply **Wil's method** (constant reject probabilities):

```
round(sum(dat$nr_checks_Wil)/length(filenames))
```

```
## [1] 22398
```

Average number of checks that one would do if for every case they do checks in **random order**

```
round(sum(dat$nr_checks_rand)/length(filenames))
```

```
## [1] 22800
```

Average overprocessing (in %) with our ordering:

```
round(100*(sum(dat$nr_checks_our) - sum(dat$minimum_check_number))/(length(filenames)*sum(dat$minimum_check
```

```
## [1] 1.22
```

Average overprocessing with Wil ordering:

```
round(100*(sum(dat$nr_checks_Wil) - sum(dat$minimum_check_number))/(length(filenames)*sum(dat$minimum_check
```

```
## [1] 3.84
```

Average overprocessing with random ordering:

```
round(100*(sum(dat$nr_checks_rand) - sum(dat$minimum_check_number))/(length(filenames)*sum(dat$minimum_chec
```

```
## [1] 5.7
```

Distribution of the number of checks performed

```
print(tt)
```

```
##                       1   2    3
## count_checks_our    972 259 6781
## count_checks_Wil    640 358 7014
## count_checks_rand   412 412 7188
## minimal            1233   0 6779
```

2 For Environmental permit log

Minimal possible number of checks, averaged over 12 test sets :

```
round(sum(dat$minimum_check_number)/length(filenames))
```

```
## [1] 412
```

Average number of checks per test set that one would do if they follow **our ordering**:

```
round(sum(dat$nr_checks_our)/length(filenames))
```

```
## [1] 575
```

Average number of checks that one would do if they apply **Wil's method** (constant reject probabilities):

```
round(sum(dat$nr_checks_Wil)/length(filenames))
```

```
## [1] 574
```

Average number of checks that one would do if for every case they do checks in **random order**

```
round(sum(dat$nr_checks_rand)/length(filenames))
```

```
## [1] 656
```

Average overprocessing (in %) with our ordering:

```
round(100*(sum(dat$nr_checks_our) - sum(dat$minimum_check_number))/(length(filenames)*sum(dat$minimum_check
```

```
## [1] 39.66
```

Average overprocessing with Wil ordering:

```
round(100*(sum(dat$nr_checks_Wil) - sum(dat$minimum_check_number))/(length(filenames)*sum(dat$minimum_check
```

```
## [1] 39.26
```

Average overprocessing with random ordering:

```
round(100*(sum(dat$nr_checks_rand) - sum(dat$minimum_check_number))/(length(filenames)*sum(dat$minimum_chec
```

```
## [1] 59.14
```

Distribution of the number of checks performed

```
print(tt)
```

```
##                         1     2     3
## count_checks_our      2.5 157.6  85.9
## count_checks_Wil      2.4 159.4  84.2
## count_checks_rand     1.4  79.5 165.1
## minimal             163.0   0.0  83.0
```