Hw04_chunlei2

Chunlei Liu

9/24/2018

Q1:

(a). Naive method. Start from A, end at B. The length is 20.

$P_a = \frac{1}{4}(\frac{1}{3})^{19} = \frac{1}{4649045868} = 2.150979e^{-10}$

(b). One-step-look-ahead. Start from A, end at B.

$P_b = \frac{1}{4}\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{3}1\frac{1}{3}\frac{1}{2}\frac{1}{3}\frac{1}{2}\frac{1}{2}\frac{1}{3}\frac{1}{3}\frac{1}{2}\frac{1}{3}\frac{1}{3}1\frac{1}{2}\frac{1}{2}\frac{1}{3} = \frac{1}{4}(\frac{1}{3})^{11}(\frac{1}{2})^{6} = \frac{1}{45349632} = 2.20509e^{-08}$

(c). Naive method. Start from B, end at A. The length is 20.

$P_c = \frac{1}{4}(\frac{1}{3})^{19} = \frac{1}{4649045868} = 2.150979e^{-10}$

(d). One-step-look-ahead. Start from B, end at A.

$P_d = \frac{1}{4}\frac{1}{3}\frac{1}{3}\frac{1}{2}\frac{1}{3}\frac{1}{2}\frac{1}{3}\frac{1}{2}\frac{1}{3}\frac{1}{3}\frac{1}{2}\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{3}\frac{1}{2}11 = \frac{1}{4}(\frac{1}{3})^{12}(\frac{1}{2})^{5} = \frac{1}{68024448} = 1.47006e^{-08}$

Q2:

$\mu = \sum 1_{(p(T) \leq p(T_0))} P(T), P(T) \propto \frac{1}{\prod_{i=1}^{3}\prod_{j=1}^{3} n_{ij}!}, h(T) = 1_{(p(T) \leq p(T_0))}$

$\pi(T) = \frac{1}{\prod_{i=1}^{3}\prod_{j=1}^{3} n_{ij}!}, \mu = E_\pi(1_{(\pi(T) \leq \pi(T_0))})$

$g(T) = \frac{C_k^K C_{n-k}^{N-K}}{C_n^N}, W(T) = \frac{\pi(T)}{g(T)}$

$t_{11} + t_{21} + ... + t_{m1} = c_1 \Rightarrow t_{11} + r_2 + ... + r_m \geq c_1$

$t_{11} \geq c_1 - \sum_{i=2}^{m} r_i \Rightarrow max(0, c_1 - \sum_{i=2}^{m} r_i) \leq t_{11} \leq min(c_1, r_1)$. Same rule for other cells.

For each iteration:

First cell $max(0, 122 - 20 - 82) \leq t_{11} \leq min(180, 122) \Rightarrow 20 \leq t_{11} \leq 122$

Second cell $max(0, 122 - t_{11} - 82) \leq t_{21} \leq min(20, 122 - t_{11})$

Third cell $t_{31} = 122 - t_{11} - t_{21}$

Fourth cell $max(0, 26 - (20 - t_{21}) - (82 - t_{31})) \leq t_{12} \leq min(26, (80 - t_{11}))$

Fifth cell $max(0, 26 - t_{12} - (82 - t_{31})) \leq t_{22} \leq min(26 - t_{12}, 20 - t_{21})$

Sixth cell $t_{32} = 26 - t_{12} - t_{22}$

Seventh cell $t_{13} = 180 - t_{11} - t_{12}$

Eighth cell $t_{23} = 20 - t_{21} - t_{22}$

Ninth cell $t_{33} = 82 - t_{31} - t_{32}$

$\tilde{\mu} = \frac{w^{(1)}h(x^{(1)}) + ... + w^{(n)}h(x^{(n)})}{w^{(1)} + ... + w^{(n)}}$

$Var(\tilde{\mu}) \approx \frac{var_g(\frac{h(x)l(x)}{g(x)}) + \tilde{\mu}^2 var_g(\frac{l(x)}{g(x)}) - 2\tilde{\mu} cov(\frac{h(x)l(x)}{g(x)}, \frac{l(x)}{g(x)})}{nE_g^2(\frac{l(x)}{g(x)})}$

So the final p value is 0.0001291898, the standard error is 9.264994e-05.

```
set.seed(40)
n = 1000
W_m = rep(0, n)
P_m = rep(0, n)
```

```r
P_0 = 1*10^45*(1/factorial(90))*(1/factorial(13))*(1/factorial(19))*(
  1/factorial(12))*(1/factorial(1))*(1/factorial(13))*(1/factorial(
    78))*(1/factorial(6))*(1/factorial(50))
for (i in 1:n) {
  #initial value
  W = 1*10^45
  P = 1*10^45
  #First cell t_11
  t_11_min = max(0, 122 - 20 - 82)
  t_11_max = min(122, 180)
  t_11 = rhyper(1, t_11_max, t_11_max, t_11_max + t_11_min)
  P = P*1/factorial(t_11)
  W = W*(1/factorial(t_11))/dhyper(t_11, t_11_max, t_11_max, t_11_max + t_11_min)
  W = W*1*10^45
  #Second cell t_21
  t_21_min = max(0, 122 - t_11 - 82)
  t_21_max = min(20, 122 - t_11)
  t_21 = rhyper(1, t_21_max, t_21_max, t_21_max + t_21_min)
  P = P*1/factorial(t_21)
  W = W*(1/factorial(t_21))/dhyper(t_21, t_21_max, t_21_max, t_21_max + t_21_min)
  W = W*1*10^45
  #Third cell t_31
  t_31 = 122 - t_11 - t_21
  P = P*1/factorial(t_31)
  W = W*(1/factorial(t_31))/1
  W = W*1*10^45
  #Fourth cell t_12
  t_12_min = max(0, 26 - (20 - t_21) - (82 - t_31))
  t_12_max = min(26, 180 - t_11)
  t_12 = rhyper(1, t_12_max, t_12_max, t_12_max + t_12_min)
  P = P*1/factorial(t_12)
  W = W*(1/factorial(t_12))/dhyper(t_12, t_12_max, t_12_max, t_12_max + t_12_min)
  W = W*1*10^45
  #Fifth cell t_22
  t_22_min = max(0, 26 - t_12 - (82 - t_31))
  t_22_max = min(26 - t_12, 20 - t_21)
  t_22 = rhyper(1, t_22_max, t_22_max, t_22_max + t_22_min)
  P = P*1/factorial(t_22)
  W = W*(1/factorial(t_22))/dhyper(t_22, t_22_max, t_22_max, t_22_max + t_22_min)
  W = W*1*10^45
  #Sixth cell t_32
  t_32 = 26 - t_12 - t_22
  P = P*1/factorial(t_32)
  W = W*(1/factorial(t_32))/1
  W = W*1*10^45
  #Seventh cell t_13
  t_13 = 180 - t_11 - t_12
  P = P*1/factorial(t_13)
  W = W*(1/factorial(t_13))/1
  W = W*1*10^45
  #Eighth cell t_23
  t_23 = 20 - t_21 - t_22
  P = P*1/factorial(t_23)
```

```
  W = W*(1/factorial(t_23))/1
  W = W*1*10^45
  #Ninth cell t_33
  t_33 = 82 - t_31 - t_32
  P = P*1/factorial(t_33)
  W = W*(1/factorial(t_33))/1
  W = W*1*10^45
  W_m[i] = W
  P_m[i] = P

}

for (j in 1:n) {
  if (P_m[j] <= P_0) {
    P_m[j] = 1
  } else {
    P_m[j] = 0
  }

}
mu = (sum(W_m*P_m))/sum(W_m)
mu
```

```
## [1] 0.0001291898
```

```
sqrt((var(W_m*P_m) + mu^2*var(W_m) - 2*mu*cov(W_m*P_m, W_m))/(n*(mean(W_m))^2))
```

```
## [1] 9.264994e-05
```

Q3:

$|\Omega| = \sum_{T\in\Omega} \frac{1}{q(T)} q(T), q(T) = \frac{1}{u-l+1} \sim U(l, u)$

For each iteration:

Similar to problem 1 except the proposal distribution changes from hypergeometric distribution to the uniform distribution.

Draw independent samples $T^{(1)}...T^{(n)}$ from $q(T)$.

Estimate the number of tables by $|\tilde{\Omega}| = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{q(T)}$

Estimate the standard error by $\tilde{\sigma}$

The final estimate for the number of the tables is 351226.4, the standard error is 5505.392.

```
set.seed(20)
N = 3000
Q_m = rep(0, N)
for (k in 1:N) {
  #initial value
  Q = 1
  #First cell t_11
  t_11_min = max(0, 122 - 20 - 82)
  t_11_max = min(122, 180)
  t_11 =  sample(t_11_min:t_11_max, 1)
  Q = Q*(1/(t_11_max - t_11_min + 1))
  #Second cell t_21
  t_21_min = max(0, 122 - t_11 - 82)
```

```
  t_21_max = min(20, 122 - t_11)
  t_21 =  sample(t_21_min:t_21_max, 1)
  Q = Q*(1/(t_21_max - t_21_min + 1))
  #Third cell t_31
  t_31 = 122 - t_11 - t_21
  Q = Q*1
  #Fourth cell t_12
  t_12_min = max(0, 26 - (20 - t_21) - (82 - t_31))
  t_12_max = min(26, 180 - t_11)
  t_12 =  sample(t_12_min:t_12_max, 1)
  Q = Q*(1/(t_12_max - t_12_min + 1))
  #Fifth cell t_22
  t_22_min = max(0, 26 - t_12 - (82 - t_31))
  t_22_max = min(26 - t_12, 20 - t_21)
  t_22 =  sample(t_22_min:t_22_max, 1)
  Q = Q*(1/(t_22_max - t_22_min + 1))
  #Sixth cell t_32
  t_32 = 26 - t_12 - t_22
  Q = Q*1
  #Seventh cell t_13
  t_13 = 180 - t_11 - t_12
  Q = Q*1
  #Eighth cell t_23
  t_23 = 20 - t_21 - t_22
  Q = Q*1
  #Ninth cell t_33
  t_33 = 82 - t_31 - t_32
  Q = Q*1
  Q_m[k] = Q
}
mean(1/Q_m)
```

```
## [1] 351226.4
```

```
sqrt(1/N*var(1/Q_m))
```

```
## [1] 5505.392
```