Exercise sheet - Week 17 (Data stream algorithms)

EXERCISE 1 -Variance of Morris' Algorithm

Recall Morris' algorithm :

```
C=0
for x in [x_1,x_2,...,x_n]:
   C=C+1 with probability 2^(-C)
return 2^C - 1
```

Denote by C_i the value of `C` after having seen x_i, recall $\mathbb{E}[2^{C_i}] = i + 1$.

1. Check that for every $i \geq 0$

$$\mathbb{E}[(2^{C_{i+1}})^2 \mid C_i] = (2^{C_i})^2 + 3 \times 2^{C_i}.$$

2. Compute Variance$(2^{C_n} - 1)$.
3. Let $A > 0$. Use Bienaymé-Chebychev's inequality to find an upper bound for the error $\mathbb{P}\left(\left|2^{C_n} - 1 - n\right| \geq An\right)$.

EXERCISE 2 -Variance of MinCount

For a parameter $k \geq 2$, the output of Giroire's MinCount algorithm has the same distribution as $\frac{k-1}{M_r^{(k)}}$, where

$$M_r^{(k)} = \min\left\{U_1, \ldots, U_r\right\},$$

where U_i's are independent uniform random variables in $(0, 1)$.

1. Compute the variance of $\left(\dfrac{k-1}{M_r^{(k)}}\right)$.

(Recall that $M_r^{(k)}$ has density $r\binom{r-1}{k-1}x^{k-1}(1-x)^{r-k}$.)

EXERCISE 3 -Empty bins in HitCount

In the analysis of HitCount one considers b balls drawn uniformly at random in m bins, let V be the number of empty bins.

1. For $0 \leq k \leq m$ justify that

$$\mathbb{P}(V = k) = \frac{m!}{k!m^n}\left\{ b \atop m-k \right\},$$

where $\{\}$ is the notation for *Stirling numbers of the second kind*, *i.e.* $\left\{ r \atop q \right\}$ enumerates partitions of a set of size r into q non-empty classes.

2. Prove that

$$\sum_{k,b} \mathbb{P}(V = k)\frac{1}{b!}u^k z^b = \left(e^{z/m} - 1 - u\right)^m.$$

(This is more difficult, for a solution see p.109 in : Marianne Durand. *Combinatoire analytique et algorithmique des ensembles de données*. PhD Thesis, École Polytechnique (2004), `https://pastel.hal.science/pastel-00000810/`.)