Radomized Algorithms Problemes Fall 2019.
Solutions to (*) in HW-2

9.- (a) Notice that the probability that for $\phi(D(T(j))) = \phi(D(S))$, where there is not a match, is the probability that $D(S) \bmod p = D(T(j)) \bmod p$, with $D(S) \neq D(T(j))$. The probability of that happening once is $\leq \frac{m}{\pi(k)}$. If there is not any matching for $1 \leq j \leq n$, using Union-Bound we get $\mathbf{Pr}\left[\text{algorithm finds error}\right] \leq n\frac{m}{\pi(k)} \sim \frac{1}{c}$, (using the Prime Number Theorem and by the the choice of $k = cmn\ln(cmn)$).

(b) Time bound. Note that p is a $O(\lg n)$-bites number, so we may reasonably assume that the mod arithmetic is constant time The obvious running time of the algorithm is $O(nm)$, however we can use a clever trick to compute $\phi(D(T(j+1)))$ from $\phi(D(T(j)))$, in constant time. $T(j)$ and $T(j+1)$ differ in one the fist and last terms, so $D(T(j+1)) = 2(D(T(j)) - 2^{m-1}x_j) + x_{j+1}$ and therefore we can compute next fingerprint:
$\phi(D(T(j+1)) = 2(\phi(D(T(j)) - 2m^{m-1}x_j) + x_{j+m} \mod p$, which involves a constant number of mod p in constant time. As the loop iterates n times, the running time is $O(m+n)$.

13.- Notice for $k = 6$, $\Omega = \{(1,1),(1,2)\ldots,(1,6),\ldots(6,6)\}$ in particular if $X_1 = 3$ then we have

$$\underbrace{(3,1),(3,2),(3,3)}_{\max=X_1=3}\underbrace{(3,4),(3,5),(3,6)}_{\max=X_2}.$$

Recall for rolling twice a k-side die $|\Omega| = 36$, for $1 \leq x_1 \leq k$ and $1 \leq x_2 \leq k$, Then $Mx = \max(X_1, X_2)$ and $Mn = \min(X_1, X_2)$ are just random variables with $\mathbf{Pr}\left[(X_1, X_2) = (x_1, x_2)\right] = 1/k^2$, and $Mi(\Omega) = Mx(\Omega) = [1, k]$. Therefore, if we pick the value $i \in [0, 1]$ as the min (max) value, assuming x_1, x_2 is the outcome of the 1st, 2nd toss,

$$\mathbf{Pr}\left[Mn = i\right] = \underbrace{\sum_{x_2 \geq i} \frac{1}{k^2}}_{x_1=i} + \underbrace{\sum_{x_1 > i} \frac{1}{k^2}}_{x_2=i}$$

and $\mathbf{Pr}\left[Ma = i\right] = \frac{1}{k^2}(\sum_{x_2 \leq i} 1 + \sum_{x_1 < i} 1)$. Moreover, $\mathbf{E}\left[Ma\right], \mathbf{E}\left[Mn\right] \in [1, k]$. So $\mathbf{E}\left[Ma\right] = \sum_{i=1}^{k} i\frac{1}{k^2}(\sum_{x_2 \leq i} 1 + \sum_{x_1 < i} 1)$ and $\mathbf{E}\left[Mn\right] = \sum_{i=1}^{k} i\frac{1}{k^2}(\sum_{x_2 \leq i} 1 + \sum_{x_1 < i} 1)$

(a) $\mathbf{E}\left[Ma\right] = \sum_{i=1}^{k} i\frac{1}{k^2}(\sum_{x_2 \leq i} 1 + \sum_{x_1 < i} 1)$
$\mathbf{E}\left[Mn\right] = \sum_{i=1}^{k} i\frac{1}{k^2}(\sum_{x_2 \leq i} 1 + \sum_{x_1 < i} 1)$

Alternative notation:
($\mathbf{E}[\max(X_1, X_2)] = \sum_{x_1}\sum_{x_2} \max(x_1, x_2)\frac{1}{k^2} = \frac{1}{k^2}(\sum_{x_1}\sum_{x_2\le x_1} x_1 + \sum_{x_2>x_1} x_2)$
$\mathbf{E}[\min(X_1, X_2)] = \sum_{x_1}\sum_{x_2} \min(x_1, x_2)\frac{1}{k}\frac{1}{k} = \frac{1}{k^2}\sum_{x_1}\sum_{x_2\le x_1} x_2 + \sum_{x_2>x_1} x_1$.)

(b)

$$
\begin{aligned}
&\mathbf{E}[\max(X_1, X_2)] + \mathbf{E}[\min(X_1, X_2)] \\
&\qquad = \frac{1}{k^2}\sum_{x_1}\sum_{x_2\le x_1} x_1 + \sum_{x_2>x_1} x_2 + \frac{1}{k^2}\sum_{x_1}\sum_{x_2\le x_1} x_2 + \sum_{x_2>x_1} x_1 \\
&\qquad = \frac{1}{k^2}\sum_{x_1}(\sum_{x_2} x_2 + \sum_{x_2} x_1) = \frac{1}{k^2}\sum_{x_1}\sum_{x_2}(x_1 + x_2) = \mathbf{E}[X_1] + \mathbf{E}[X_2]
\end{aligned}
$$

(c) By using linearity of expectation twice, we get
$\mathbf{E}[\max(X_1, X_2)] + \mathbf{E}[\min(X_1, X_2)] = \mathbf{E}[\max(X_1, X_2) + \min(X_1, X_2)] = \mathbf{E}[X_1 + X_2] = \mathbf{E}[X_1] + \mathbf{E}[X_2]$

14.- (Needs problem 13)

(a) $\mathbf{Pr}[X = Y] = \sum_x (1-p)^{x-1}p(1-q)^{x-1}q$.
Recall that for geometric random variables, we have the identity
$\mathbf{Pr}[X \ge i] = \sum_{n=i}^{\infty}(1-p)^{n-1}p = (1-p)^{i-1}$ (1).
So, we obtain $\mathbf{Pr}[X = Y] = \frac{pq}{p+q-pq}$.

(b) We split the event $\mathbf{Pr}[\min(X, Y) = k]$ into two disjoint events.

$$
\begin{aligned}
\mathbf{Pr}[\min(X, Y) = k] &= \mathbf{Pr}[(X = k) \cap (Y > k)] + \mathbf{Pr}[(X > k) \cap (Y = k)] \\
&= \mathbf{Pr}[(X = k)]\,\mathbf{Pr}[(Y \ge k)] + \mathbf{Pr}[(X > k)]\,\mathbf{Pr}[(Y = k)].
\end{aligned}
$$

Using again $\mathbf{Pr}[X \ge i] = (1-p)^{i-1}$, we get
$\mathbf{Pr}[(X > k)] = \mathbf{Pr}[(X \ge k)] - \mathbf{Pr}[(X - k)] = (1-p)^{k-1}(1-p)$.
Finally,

$$
\begin{aligned}
\mathbf{Pr}[\min(X, Y) = k] &= (1-p)^{k-1}p(1-q)^{k-1} + (1-p)^{k-1}(1-p)(1-q)^{k-1}q \\
&= ((1-p)(1-q))^{k-1}(p + (1-p)q) = ((1-p)(1-q))^{k-1}(p + q - pq).
\end{aligned}
$$

(c) We know from problem 11 that $\mathbf{E}[\max(X_1, X_2)] = \mathbf{E}[X_1] + \mathbf{E}[X_2] - \mathbf{E}[\min(X_1, X_2)]$. From part (b), we know that $\min(X, Y)$ is a geometric random variable mean $p + q - pq$. Therefore, $\mathbf{E}[\min(X, Y)] = 1/(p + q - pq)$, and we get
$\mathbf{E}[\max(X, Y)] = \frac{1}{p} + \frac{1}{q} + \frac{1}{p+q+pq}$.

(d) $\mathbf{E}[X|X \le Y] = \sum_{x\ge 1} x\mathbf{Pr}[X = x|x \le Y] = \sum_x x\frac{\mathbf{Pr}[(X=x)\cap(x\le Y)]}{\mathbf{Pr}[X\le Y]}$.

Let's consider the denominator.

$$\begin{aligned}\mathbf{Pr}\left[X \le Y\right] &= \sum_{z\ge 1} \mathbf{Pr}\left[(X=z)\cap(z\le Y)\right] = \sum_z \mathbf{Pr}\left[(X=z)\right]\mathbf{Pr}\left[(z\le Y)\right] \\ &= \sum_z (1-p)^{z-1}p(1-q)^{z-1} = \sum_z ((1-p)(1-q))^{z-1}p = p\sum_z (1-p-q+pq)^{z-1} \\ &= \frac{p}{p+q-pq} \text{using again eq. (1).}\end{aligned}$$

Now we can compute the whole equation.

$$\begin{aligned}\mathbf{Pr}\left[X \le Y\right] &= \frac{p+q-pq}{p}\sum_x x\mathbf{Pr}\left[X=x\right]\mathbf{Pr}\left[x\le Y\right] = \frac{p+q-pq}{p}\sum_x x(1-p)^{x-1}p(1-q)^{x-1} \\ &= (p+q-pq)\sum_x x(1-p-q+pq)^{x-1}.\end{aligned}$$

This is equal to the expectation of a geometric random variable with mean $p+q-pq$. So $\mathbf{E}\left[X \le Y\right] = \frac{1}{p+q-pq}$.

15.- Let $b_1, b_2, \ldots, b_t$ the first t streamed objects. Let X_t a rv which takes the value of the new object in memory after b_t. We have to prove that $\mathbf{Pr}\left[X_t = b_i\right] = 1/t$, for all $1 \le i \le t$. Use induction on t. For $t=1$, it is true as $X_t = b_1$ with probability $=1$. Assume after k observations we have $\mathbf{Pr}\left[X_t = b_i\right] = 1/t$ for all $1 \le i \le k$. In the next observation for $t+1$, we have that $X_{t+1} = b_{t+1}$ with probability $1/(t+1) \Rightarrow \mathbf{Pr}\left[X_{t+1} = b_{n+1}\right] = 1/(n+1)$.
For $1 \le i \le t$

$$\begin{aligned}\mathbf{Pr}\left[X_{t+1} = b_t\right] &= \mathbf{Pr}\left[\text{no change after observation} t \text{ i } X_t = b_i\right] \\ &= \mathbf{Pr}\left[\text{no change after} t\right]\mathbf{Pr}\left[X_t = b_i\right] \\ &= \left(1 - \frac{1}{t+1}\right)\frac{1}{t} = \frac{1}{t+1}.\end{aligned}$$

Any of the 2 cases ($X_{t+1} = b_i$ o $X_{t+1} = b_{i+1}$) happen with probability $1/(n+1)$.
If from t to m there were not changes in the observations we would have that observacions tindriem:
$\mathbf{Pr}\left[X_m = b_i\right] = \frac{1}{t}\left(1 - \frac{1}{t+1}\right)\left(1 - \frac{1}{t+2}\right)\cdots\left(1 - \frac{1}{m}\right) = \frac{1}{m}$.