CSC484 Assignment #2

Andrii Osipa

March 2017

Exercise 1.1

Given $p_1, .., p_n$ such that $\sum_{k=1}^{n} p_k = 1$.

Algorithm. We want to look at process of sampling in the following way: select uniformly bucket $b_k, k \in \overline{1, n}$ and then pick value l_k with probability p'_k or r_k with probability $1 - p'_k$, where l_k and r_k are values from $\{0, ..., n-1\}$. And such that following holds: $\sum_{l_k : l_k = i} 1/n * p'_k + \sum_{r_k : r_k = i} 1/n * (1 - p'_k) = p_i$. In other words we split original probabilities between buckets such that total probability of bucket is $1/n$ and probability of getting value i is same as it was initially.
Preprocessing: split all k into two groups: p_k less than $1/n$(L) and others(R). Then for each element $l \in L$ we pick any available element $r \in R$ and form a bucket b_k such that $l_k = l, r_k = r, p'_k = p_l$. Then we decrease probability p_r by $1/n - p_l$ and, if needed(prob p_r less that $1/n$), move r to list L. Also we delete l from list $L(\Leftrightarrow p_l := 0)$.
On k^{th} step we have $n-k$ values left in lists L and R and sum of all probabilities left is $\frac{n-k}{n}$, as on each step we decrease sum of probabilities by $1/n$. It is not possible that there is no elements left in list R, but $L \neq \emptyset$ on some step as it means that sum of all probabilities left is less then $\frac{n-k}{n}$. If we have that some elements left in R and $L = \emptyset$ this means that each of them has probability $1/n$, which is obviously the only possibility in this case.
Note: in computing case $L \neq \emptyset, R = \emptyset$ may occur due to finiteness of numbers processed, we just say that probability of each $l \in L$ is $1/n$.
Sampling: select uniformly bucket $b_k, k \in \overline{1, n}$, then pick random number in $[0, 1/n]$. If number is greater that p_k output r_k, else – l_k.
Time complexity of preprocessing: $O(n)$ as we just loop through all p_k on the beginning to create lists L and R and then we do n steps on each removing one value from list L.
Time complexity of sampling: $O(1)$ as we just do 3 operations: pick random k pick random number in $[0, 1/n]$ and "if statement".

Note: to simplify we can multiply all initial probabilities by n then total sum is n and on each step we have sum of probabilities of left values $n - k$ and each bucket has probability 1.

Exercise 2.1

Cauchy distribution density $f(x) = \left(\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)\right)^{-1}$.

$X, X_1, X_2, X_3, X_4, X_5$ from Cauchy (x_0, γ).

$$E[X] = \int\limits_{-\infty}^{\infty} \frac{x}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)} dx = \lim_{a\to\infty}\lim_{b\to-\infty}\int\limits_{b}^{a} \frac{x}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)} dx.$$

Consider $\lim\limits_{a\to\infty}\int\limits_{-a+x_0}^{a+x_0} \frac{x}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)} dx = x_0$. If the former limit exists it must be also equal x_0. Consider $\lim\limits_{a\to\infty}\int\limits_{-a}^{2a} \frac{x}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)} dx =$ $\lim\limits_{a\to\infty}\int\limits_{a}^{2a} \frac{x}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)} dx = x_0 + \lim\limits_{a\to\infty}\int\limits_{a}^{2a} \frac{x}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)} dx \geq x_0 +$ $\frac{\gamma^2}{4}$ by simple inequalities. Therefore $E[X]$ is not defined, so $V[X]$ and squared coefficient of variation.

$Y = (X_1, X_2, X_3)$. From statistics we know that given sample from distribution with density $f(x)$ and cumulative distribution function $F(x)$ we can write density for k^{th} order statistics as following: $f_k(x) = kC_n^k F(x)^{k-1} f(x)(1-F(x))^{n-k}$. For Cauchy distribution we have cumulative distribution function $F(x) = \frac{1}{\pi}\arctan\frac{x-x_0}{\gamma} + \frac{1}{2}$. Therefore for Y we have the following density function:

$$f_{median} = 6\left(\frac{1}{\pi}\arctan\frac{x-x_0}{\gamma}+\frac{1}{2}\right)\frac{1}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)}\left(\frac{1}{2}-\frac{1}{\pi}\arctan\frac{x-x_0}{\gamma}\right).$$

$$E[Y] = \frac{6}{\pi\gamma} \int\limits_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} + \frac{1}{2} \right) \frac{x}{1 + \left(\frac{x - x_0}{\gamma} \right)^2} \left(\frac{1}{2} - \frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} \right) dx =$$

$$= \frac{6}{\pi} \int\limits_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} + \frac{1}{2} \right) \frac{\frac{x - x_0}{\gamma}}{1 + \left(\frac{x - x_0}{\gamma} \right)^2} \left(\frac{1}{2} - \frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} \right) dx +$$

$$+ \frac{6}{\pi\gamma} \int\limits_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} + \frac{1}{2} \right) \frac{x_0}{1 + \left(\frac{x - x_0}{\gamma} \right)^2} \left(\frac{1}{2} - \frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} \right) dx$$

We can rewrite integrals as following, skipping constants:

$$\int\limits_{-\infty}^{\infty} (\frac{1}{4} - \frac{1}{\pi^2} \arctan^2 z) \frac{z}{1 + z^2} dz,$$

$$\lim_{z \to +/-\infty} z(\frac{1}{4} - \frac{1}{\pi^2} \arctan^2 z) = +/- \frac{1}{\pi} \Rightarrow \exists c > 0 : \forall |z| > c : \ |z(\frac{1}{4} - \frac{1}{\pi^2} \arctan^2 z)| < C$$

$$\Rightarrow 0 \leq \int\limits_{0}^{\infty} (\frac{1}{4} - \frac{1}{\pi^2} \arctan^2 z) \frac{z}{1 + z^2} dz < \int\limits_{0}^{\infty} \frac{C}{1 + z^2} < \infty;$$

$$\Rightarrow 0 \geq \int\limits_{-\infty}^{0} (\frac{1}{4} - \frac{1}{\pi^2} \arctan^2 z) \frac{z}{1 + z^2} dz > \int\limits_{-\infty}^{0} \frac{-C}{1 + z^2} > -\infty.$$

Therefore this integral converges and easy to see that it is equal to 0(function symmetric).
Second integral in the sum above: $\int\limits_{-\infty}^{\infty} (\frac{1}{4} - \frac{1}{\pi^2} \arctan^2 z) \frac{1}{1 + z^2} dz = \frac{\pi}{6}$ is known result.
We obtain the following: $E[Y] = x_0$.
Variance of Y.

$$V[Y] = \frac{6}{\pi\gamma} \int\limits_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} + \frac{1}{2} \right) \frac{(x - x_0)^2}{1 + \left(\frac{x - x_0}{\gamma} \right)^2} \left(\frac{1}{2} - \frac{1}{\pi} \arctan \frac{x - x_0}{\gamma} \right) dx$$

so we are interested in the following integral:

$$\int\limits_{-\infty}^{\infty} \left(\frac{1}{\pi} \arctan z + \frac{1}{2} \right) \frac{z^2}{1 + z^2} \left(\frac{1}{2} - \frac{1}{\pi} \arctan z \right) dz$$

Lets use the following inequality $\arctan x \geq \frac{\pi}{2} - \frac{1}{x}$. Then $\left(\frac{1}{\pi}\arctan z + \frac{1}{2}\right)\frac{z^2}{1+z^2}\left(\frac{1}{2} - \frac{1}{\pi}\arctan z\right) >$ $\left(1/4 - 1/4 + \frac{2}{\pi z} - \frac{1}{\pi^2 z^2}\right)\frac{z^2}{1+z^2} = \frac{2z}{\pi(1+z^2)} - \frac{1}{\pi^2(1+z^2)}$ and this holds for all $z \neq 0$. And $\int\limits_{(-\infty,-a)\cup(a,\infty)} \frac{2z}{\pi(1+z^2)}$ does not converge, but $\int\limits_{(-\infty,-a)\cup(a,\infty)} \frac{z^2}{1+z^2}$ converge. Therefore our integral is also does not converge:

$$\int\limits_{-\infty}^{\infty} \frac{z^2}{1+z^2}\left(\frac{1}{4} - \frac{1}{\pi^2}\arctan^2 z\right) dz =$$

$$= \int\limits_{-a}^{a} \frac{z^2}{1+z^2}\left(\frac{1}{4} - \frac{1}{\pi^2}\arctan^2 z\right) dz + \int\limits_{(-\infty,-a)\cup(a,\infty)} \frac{z^2}{1+z^2}\left(\frac{1}{4} - \frac{1}{\pi^2}\arctan^2 z\right) dz >$$

$$> \int\limits_{-a}^{a} \frac{z^2}{1+z^2}\left(\frac{1}{4} - \frac{1}{\pi^2}\arctan^2 z\right) dz + \int\limits_{(-\infty,-a)\cup(a,\infty)} \frac{2z}{\pi(1+z^2)} + \int\limits_{(-\infty,-a)\cup(a,\infty)} \frac{z^2}{1+z^2}$$

Therefore $V[Y]$ is not defined in this case and so the squared coefficient of variation.

Case $Z = median(X_1, X_2, X_3, X_4, X_5)$**.** By very same logic as for Y we can show that $E[Z] = x_0$. Then by using inequality $\arctan x \leq \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3}$ we can easily show that variance of Z converges. But it‘s exact value still remains unknown, so the squared coefficient of variation...
But in this case it is at least defined.