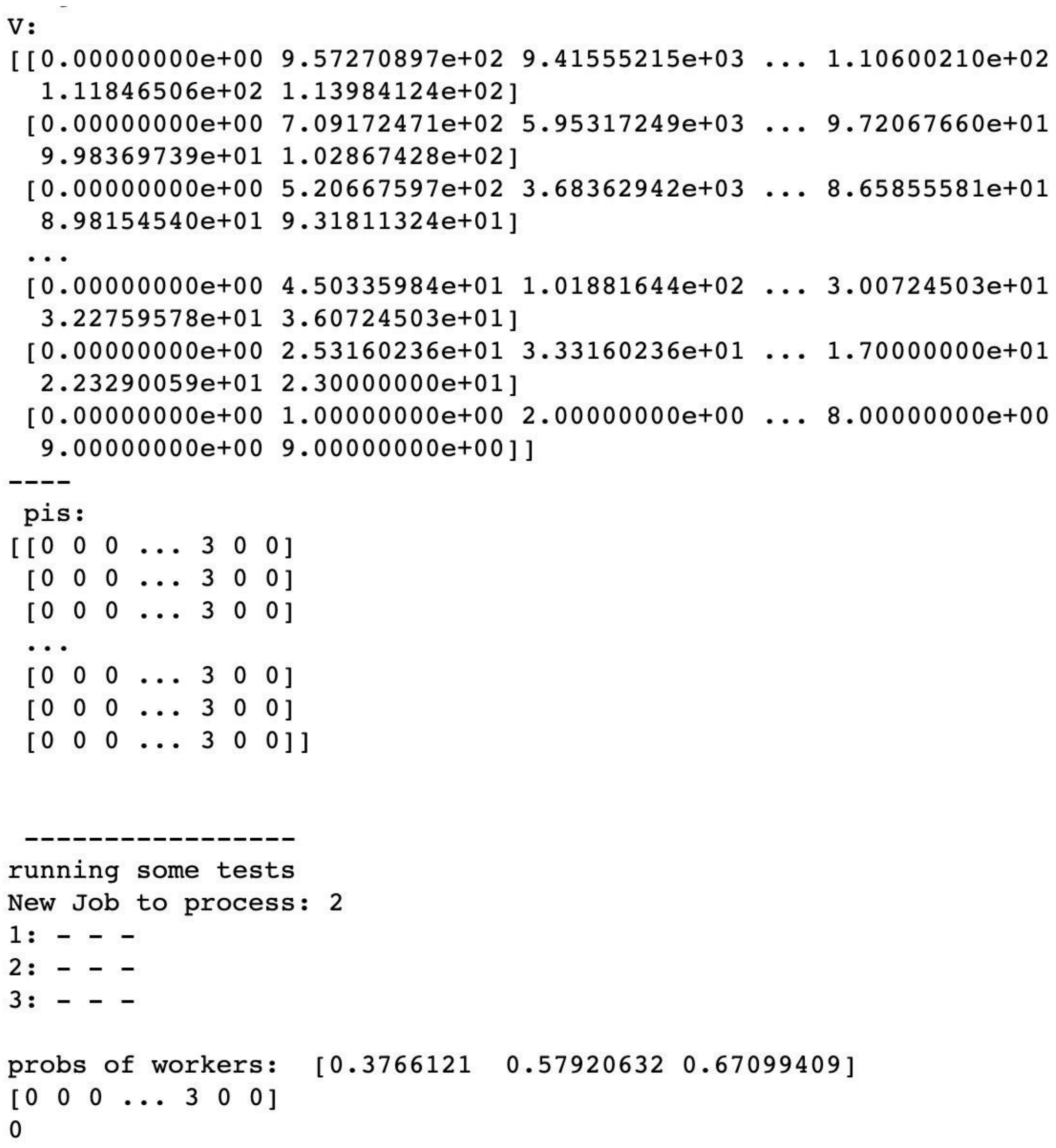
Report on Investigating the Policy:

We detected a problem that if at stage k=0 the policy seems to be idle (u=0) much more than it should be. As the queues are all empty, the policy should take an action that would assign the job to one of the queues.



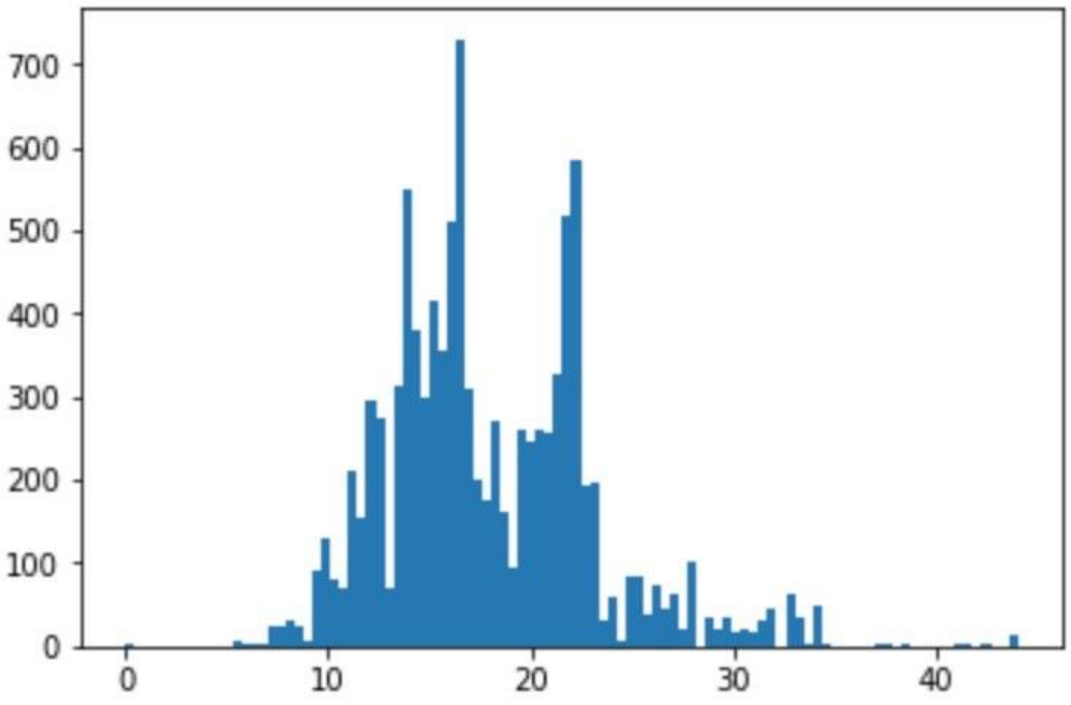
```
V:
[[0.00000000e+00 9.57270897e+02 9.41555215e+03 ... 1.10600210e+02
  1.11846506e+02 1.13984124e+02]
 [0.00000000e+00 7.09172471e+02 5.95317249e+03 ... 9.72067660e+01
  9.98369739e+01 1.02867428e+02]
 [0.00000000e+00 5.20667597e+02 3.68362942e+03 ... 8.65855581e+01
  8.98154540e+01 9.31811324e+01]
 ...
 [0.00000000e+00 4.50335984e+01 1.01881644e+02 ... 3.00724503e+01
  3.22759578e+01 3.60724503e+01]
 [0.00000000e+00 2.53160236e+01 3.33160236e+01 ... 1.70000000e+01
  2.23290059e+01 2.30000000e+01]
 [0.00000000e+00 1.00000000e+00 2.00000000e+00 ... 8.00000000e+00
  9.00000000e+00 9.00000000e+00]]
----
 pis:
[[0 0 0 ... 3 0 0]
 [0 0 0 ... 3 0 0]
 [0 0 0 ... 3 0 0]
 ...
 [0 0 0 ... 3 0 0]
 [0 0 0 ... 3 0 0]
 [0 0 0 ... 3 0 0]]


 -----------------
running some tests
New Job to process: 2
1: - - -
2: - - -
3: - - -

probs of workers:  [0.3766121  0.57920632 0.67099409]
[0 0 0 ... 3 0 0]
0
```

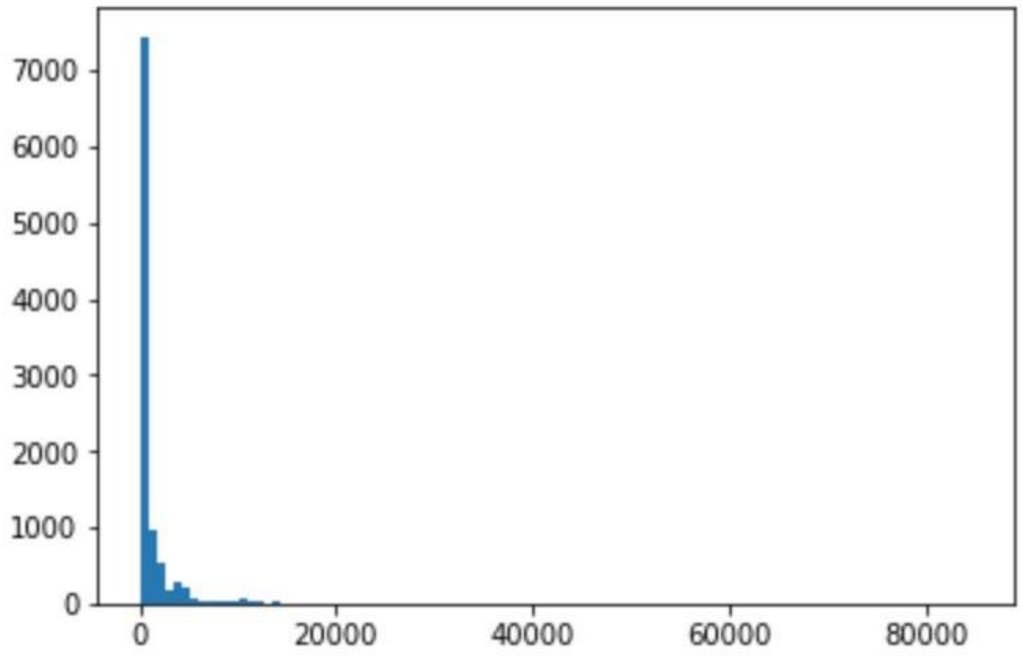
Moreover, we can compare stage k=0 and stage k=9:

Stage **k=9**:



We have a well distributed Value function.

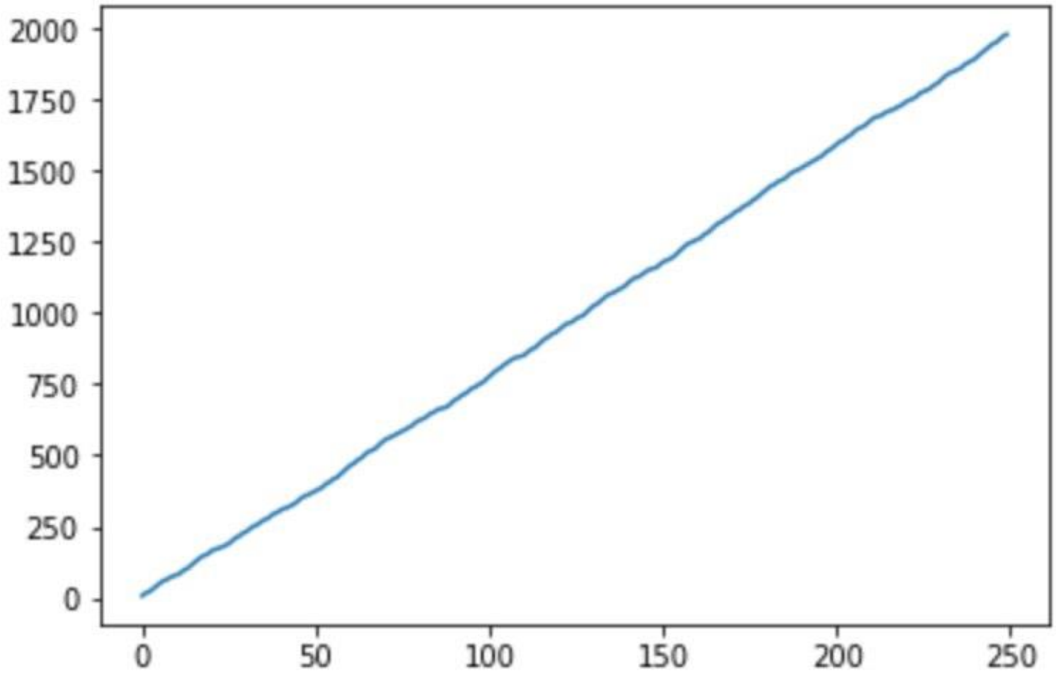
For stage **k=0** we get:



Which shows that our value function not properly converges. The probabilities are probably not normalized properly.

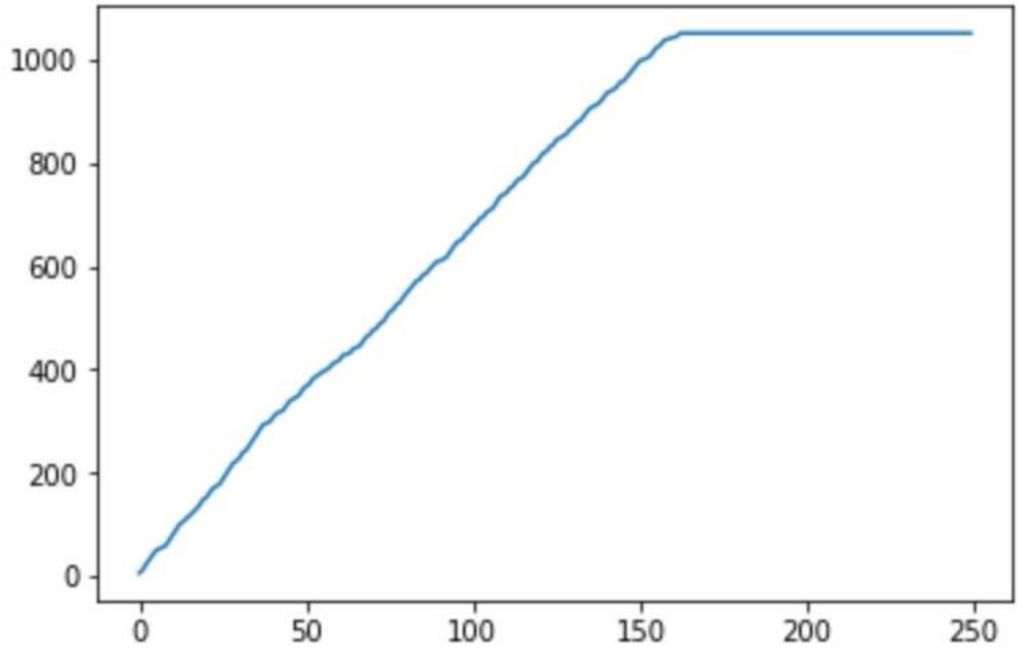
Investigating the Policy for 250 iterations

The local costs at the stage **k=9** are:



It can be seen the policy is not that good evolved since it only had a lookout of one stage in advance.

If we compare this to the first stage **k=0**:



The policy here converges for more iterations since it also takes the future stages into account.