**1. Simple statistical gradient-following algorithms for connectionist reinforcement learning**

This article presents a general class of associative reinforcement learning algorithms for connectionist networks containing stochastic units. These algorithms, called REINFORCE algorithms, are shown to make weight adjustments in a direction that lies along the gradient of expected reinforcement in both immediate-reinforcement tasks and certain limited forms of delayed-reinforcement tasks, and they do this without explicitly computing gradient estimates or even storing information from which such estimates could be computed. Specific examples of such algorithms are presented, some of which bear a close relationship to certain existing algorithms while others are novel but potentially interesting in their own right. Also given are results that show how such algorithms can be naturally integrated with backpropagation. We close with a brief discussion of a number of additional issues surrounding the use of such algorithms, including what is known about their limiting behaviors as well as further considerations that might be used to help develop similar but potentially more powerful reinforcement learning algorithms.

The analyses presented in this article, together with a variety of simulation experiments performed by the author and others, suggest that REINFORCE algorithms are useful in their own right and, perhaps more importantly, may serve as a sound basis for developing other more effective reinforcement learning algorithms. One major advantage of the REINFORCE approach is that it represents a prescription for devising statistical gradient-following algorithms for reinforcement-learning networks of units that compute their random output in essentially any arbitrary fashion. Also, because it is a gradient-based approach, it integrates well with other gradient computation techniques such as backpropagation. The main disadvantages are the lack of a general convergence theory applicable to this class of algorithms and, as with all gradient algorithms, an apparent susceptibility to convergence to false optima.

**Keywords:** Reinforcement learning, connectionist networks, gradient descent, mathematical analysis

**2. Deep Reinforcement Learning with Double Q-Learning**

The popular Q-learning algorithm is known to overestimate action values under certain conditions. It was not previously known whether, in practice, such overestimations are common, whether they harm performance, and whether they can generally be prevented. In this paper, we answer all these questions affirmatively. In particular, we first show that the recent DQN algorithm, which combines Q-learning with a deep neural network, suffers from substantial overestimations in some games in the Atari 2600 domain. We then show that the idea behind the Double Q-learning algorithm, which was introduced in a tabular setting, can be generalized to work with large-scale function approximation. We propose a specific adaptation to the DQN algorithm and show that the resulting algorithm not only reduces the observed overestimations, as hypothesized, but that this also leads to much better performance on several games.

This paper has five contributions. First, we have shown why Q-learning can be overoptimistic in large-scale problems, even if these are deterministic, due to the inherent estimation errors of learning. Second, by analyzing the value estimates on Atari games we have shown that these overestimations are more common and severe in practice than previously acknowledged. Third, we have shown that Double Q-learning can be used at scale to successfully reduce this over optimism, resulting in more stable and reliable learning. Fourth, we have proposed a specific implementation called Double DQN, that uses the existing architecture and deep neural network of the DQN algorithm without requiring additional networks or parameters. Finally, we have shown that Double DQN finds better policies, obtaining new state-of-the-art results on the Atari 2600 domain.

**Keywords:** Double Q-Learning, deep neural network

**3. Continuous control with deep reinforcement learning**

We adapt the ideas underlying the success of Deep Q-Learning to the continuous action domain. We present an actor-critic, model-free algorithm based on the deterministic policy gradient that can operate over continuous action spaces. Using the same learning algorithm, network architecture and hyper-parameters, our algorithm robustly solves more than 20 simulated physics tasks, including classic problems such as cartpole swing-up, dexterous manipulation, legged locomotion and car driving. Our algorithm is able to find policies whose performance is competitive with those found by a planning algorithm with full access to the dynamics of the domain and its derivatives. We further demonstrate that for many of the tasks the algorithm can learn policies “end-to-end”: directly from raw pixel inputs.

The work combines insights from recent advances in deep learning and reinforcement learning, resulting in an algorithm that robustly solves challenging problems across a variety of domains with continuous action spaces, even when using raw pixels for observations. As with most reinforcement learning algorithms, the use of non-linear function approximators nullifies any convergence guarantees; however, our experimental results demonstrate that stable learning without the need for any modifications between environments. Interestingly, all of our experiments used substantially fewer steps of experience than was used by DQN learning to find solutions in the Atari domain. Nearly all of the problems we looked at were solved within 2.5 million steps of experience (and usually far fewer), a factor of 20 fewer steps than DQN requires for good Atari solutions. This suggests that, given more simulation time, DDPG may solve even more difficult problems than those considered here.

A few limitations to our approach remain. Most notably, as with most model-free reinforcement approaches, DDPG requires a large number of training episodes to find solutions. However, we believe that a robust model-free approach may be an important component of larger systems which may attack these limitations (Glascher et al., 2010).

**Keywords:** Double Q-Learning, classic problems

**4. Policy gradient methods for reinforcement learning with function approximation**

Function approximation is essential to reinforcement learning, but the standard approach of approximating a value function and determining a policy from it has so far proven theoretically intractable. In this paper we explore an alternative approach in which the policy is explicitly represented by its own function approximator, independent of the value function, and is updated according to the gradient of expected reward with respect to the policy parameters. Williams's REINFORCE method and actor-critic methods are examples of this approach. Our main new result is to show that the gradient can be written in a form suitable for estimation from experience aided by an approximate action-value or advantage function. Using this result, we prove for the first time that a version of policy iteration with arbitrary differentiable function approximation is convergent to a locally optimal policy.

**Keywords:** function approximation

**5. A cyclic-queue model of system overhead in multiprogrammed computer systems**

A probabilistic model is presented of a multiprogrammed computer system operat- ing under demand paging. The model contains an explicit representation of system overhead, the CPU requirements and paging characteristics of the program load being described statistically. Expressions for steady-state CPU problem program time, CPU overhead time, and channel utilization are obtained. Some numerical results are given which quantify the gains in CPU utilization obtained from multiprogramming. It is also pointed out heuristically and demonstrated numerically that an actual decrease in CPU utilization results if there is too much overhead associated with multiprogramming and if the average time between page ex- ceptions decreases too rapidly with increasing number of multiprogrammed jobs.

**Keywords:** paging machines, demand paging, operating systems studies, queuing analysis

**6. The semi-markovian queue: theory and applications**

In this paper, we study a first-come-first-served single server semi-Markovian queue in which both the arrival and service mechanisms are semi-Markov processes. The interarrival time and service times may depend on one another and the marginal distribution of the service times is assumed to be phase-type. For this queue, we show that the distributions of waiting time, time in system and virtual waiting time are matrix-exponential. Further, these matrix-exponential distributions have phase-type representations. For the special case when the interarrival times are independent of the service times, we show that the queue length distribution is matrix-geometric. For this special case, we prove that the queue length distribution problem is the dual of the waiting time distribution problem, i.e., finding the solution of one problem immediately gives the solution of the other. We show that our methods are computationally feasible and report our numerical experience. We give Examples where such queues arise naturally. In particular, we discuss an application in manufacturing, a periodic queue and a queue with Markov modulated arrivals and services.

**Keywords**: Semi-Markovian queue, matrix-exponential distribution, duality

**7. A queue theory-based approach to staff software maintenance centers**

The Internet and WEB pervasivenesses are changing the landscape of several different areas, ranging from information gathering/managing and commerce to software development, maintenance and evolution. Software companies having a geographically distributed structure, or geographically distributed customers, are adopting information communication technologies to cooperate. Communication technologies and infrastructures allow the companies to create a virtual software factory. This paper proposes to adopt queue theory to deal with an economically relevant category of problems: the staffing, the process management and the service level evaluation of massive maintenance projects in a virtual software factory. Data from a massive corrective maintenance intervention were used to simulate and study different service center configurations, in particular, a monolithic configuration and a configuration corresponding to a multi-phase maintenance process where several maintenance centers cooperated. Queue theory allowed effective control of the process supporting project management decisions. The mathematical tool provided a means to assess staffing, evaluate service level and balance the workload between maintenance centers while executing the project.

**Keywords**: queue theory-based approach, staff software maintenance centers, software development, virtual software factory, corrective maintenance intervention, monolithic configuration, multi-phase maintenance process, project management decisions.

**8. Modelling and optimisation of a traffic intersection based on queue theory and markov decision control methods**

Traffic models play an important role in both today's traffic research and in many traffic applications such as traffic flow prediction, incident detection and traffic control. Modelling traffic dynamics and optimising the control signal are two interrelated problems. Modelling provides fundamental understanding of traffic dynamics and behaviour. In this paper, traffic signal is modelled as a M/M/l queueing theory. The validation of a simulation model (M/M/l queue) with different arrival rates is presented. From the result, a traffic light model was developed by applying M/M/l queue theory for single intersection. In the optimisation strategy, the Markov decision control is applied to minimize queue length and waiting time. Simulation results show the excellent potential of this approach

**Keywords**: traffic intersection, queue theory, Markov decision control, traffic flow prediction, incident detection, traffic control, traffic dynamics, M/M/l queueing theory, optimisation strategy

**9. Steady‐state diffusion approximations for discrete‐time queue in hospital inpatient flow management**

In this article, we analyze a discrete‐time queue that is motivated from studying hospital inpatient flow management, where the customer count process captures the midnight inpatient census. The stationary distribution of the customer count has no explicit form and is difficult to compute in certain parameter regimes. Using the Stein's method framework, we identify a continuous random variable to approximate the steady‐state customer count. The continuous random variable corresponds to the stationary distribution of a diffusion process with state‐dependent diffusion coefficients. We characterize the error bounds of this approximation under a variety of system load conditions—from lightly loaded to heavily loaded. We also identify the critical role that the service rate plays in the convergence rate of the error bounds. We perform extensive numerical experiments to support the theoretical findings and to demonstrate the approximation quality. In particular, we show that our approximation performs better than those based on constant diffusion coefficients when the number of servers is small, which is relevant to decision making in a single hospital ward.

**Keywords**: discrete queue, steady‐state analysis, Stein's method, state‐dependent diffusion

**10. Reducing Queues in a Nigerian Hospital Pharmacy**

Queues are characterized structures formed to maintain order and create a hold on time, money and human contribution towards development and efficient performance of any system. The aim of this work was to characterize the queue, describe the queue discipline of the outpatient pharmacy, to institute a cross-sectional intervention by streamlining queue behaviour and to measure the impact of streamlining queue characteristics and queue discipline on waiting time of patients. Results showed that queue characteristics existing at the pharmacy during the situation analysis was a single servermultiple queue model. However, after the intervention was done involving staff re-orientation, the streamlined process reduced waiting time from 167.0 to 55.1 min. Queue discipline was strictly instituted by designed tally cards that were serially numbered. The characterization and discipline that was instituted handled and/or eliminated the challenge of shunting, balking or jockeying and reduced reneging. The workflow chart was sketched and drawn to scale, aiding the collation of baseline data and for proposed structural modifications. Other results obtained include the waiting area to pharmacy space ratio, which gave a good result of 1:9. Effort should therefore be intensified by hospital pharmacists to reduce patient queues and improve efficiency of services, following the results of snapshots from this work.

**Keywords**: Queue characteristics, queue discipline, outpatient pharmacy

**11. Estimating Business Loss to a Hospital Emergency Department from Patient Reneging by Queuing-Based Regression**

Viewing a hospital emergency department (ED) as a production network, patients arrive to the processing queue and may renege (leave) before entering service. Reneging is not just a function of the delay in the queue or the number in the queue, but also the attitude of the patients toward perceived delay or fullness which vary between EDs. In this paper, we present evidence that reneging in an ED queue can be well estimated by nonlinear regression on pK, the ‘full queue’ formula of the M/M/1/K, over a wide range of ED situations. The regression curve is used for estimating business loss to production in the ED. We argue that the curve is useful outside the range of data used for fitting because the curve’s functional form is not ad hoc as is most regression; instead it represents a universal queuing behavior.

**Keywords:** Emergency department, queuing theory, regression, business loss, reneging.

**12. An MRP-Based architecture of plan resources and to manage waiting queue in hospital systems**

In the last decades, the interest in the development of healthcare planning and control systems is quickly spreading. Hospitals quality improvement, in fact, is actually a continuous process mainly aiming at improving the professional services rendered to patients, not only in terms of effectiveness of cures but also in terms of efficiency of supplying services system. For a long time, researchers have discussed about how the benefits in terms of reduction of waiting times and waiting queue, obtained by the IT application in manufacturing, can be achieved also in hospital systems without forgetting that healthcare supplies the basic good: health. Resources coordination according to patients who are in hospital, allows the reduction of “slacks” at resources due to discharge delays, late-start surgeries and slow laboratories turnaround. Therefore a resources planning system has been implemented in order to reduce the high waiting times and increase resources utilization in hospitals. The paper describes the methodology to create a dependent demand starting from patients needs and proposes the implementation of an MRP (Material Requirements Planning) procedure for hospitals. The PDTs (Diagnostic Terapeutic Path) for each patient in hospital are generated. All the PDTs are used to calculate resources, materials and facilities requirement in short-mid term, after having linked resources to the services that hospitals are equipped to provide (BORM – Bill of Resources and Materials). In this way the MRP procedure is able to plan resources, facilities, materials and HR in accordance with the real “demand of patients” and highlight potential overloads and problems.

The output of the described system is the activities Gantt chart which represents a calendar of the activities that have to be realized inside hospitals according to patients demand.

The proposed logic allows a more suitable management of the resource’s capacity and availability and, consequently, the improvement of waiting queues management by basing the activities planning and the resources scheduling on requirements. In fact, after building a Gantt’s chart of the monitored resources, their availability (facilities, personnel, laboratories, etc.) can be known in real time or even in advance. When resources are overloaded and it is not possible to realize one or more of the planned activities, patients can be rescheduled in order to meet the available capacity. At the same way, when an emergency occurs it is possible to reschedule not-emergency patients exactly when the resources they need are available. In this manner, the eventuality that a patient is admitted in advance and uselessly occupies resources (beds or food service) in waiting for services that cannot be supplied in that moment is avoided. The update of the system, moreover, allows the real time check of resources overloads or delays in supplying services for patient who are already in the hospital.

The described system can be seen as the first step toward the development of a network able to manage more hospitals and able to sort out patients in healthcare structures according to the required services and the real availability of resources.

**Keywords:** Healthcare System Management, MRP logic, resources planning systems.

**13. Using Queuing Theory and Simulation Modelling to Reduce Waiting Times in An Iranian Emergency Department**

Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4709818/>

This is a cross-sectional study in which simulation software (Arena, version 14) was used. The input information was extracted from the hospital database as well as through sampling. The objective was to evaluate the response variables of waiting time, number waiting and utilization of each server and test the three scenarios to improve them.

Running the models for 30 days revealed that a total of 4088 patients left the ED after being served and 1238 patients waited in the queue for admission in the ED bed area at end of the run (actually these patients received services out of their defined capacity). The first scenario result in the number of beds had to be increased from 81 to179 in order that the number waiting of the “bed area” server become almost zero. The second scenario which attempted to limit hospitalization time in the ED bed area to the third quartile of the serving time distribution could decrease the number waiting to 586 patients.

**Keywords:** Computer simulation, Emergency department, Hospital bed capacity, Length of stay, Queuing theory

**14. Review of predicting number of patients in the queue in the hospital using monte carlo simulation**

Healthcare is essential to the general welfare of society. It provides for the prevention, treatment, and management of illness through the services offered by medical and allied health professions. Emergency Department crowding causes a series of negative effects, e.g. medical errors, poor patient treatment and general patient dissatisfaction. In light of these challenges, a need for review and reform of our healthcare practices has become apparent. One road to improve the typical clinical system is to describe the patient flow in a model of the system and how the system is constrained by available equipment. Various predictive control models have been developed to try and ease overcrowding in hospitals. Such models are the Model Predictive Control to control the queuing systems. In this study the research will compare the existing prediction models and come up with Monte Carlo Simulation model to predict the number of patients in the queue.

**Keywords:** Emergency Department, Prediction Model, Queuing System, Monte Carlo Simulation

**15. Technical Note—An Equivalence Between Continuous and Discrete Time Markov Decision Processes**

A continuous time Markov decision process with uniformly bounded transition rates is shown to be equivalent to a simpler discrete time Markov decision process for both the discounted and average reward criteria on an infinite horizon. This result clarifies some earlier work in this area.

**Keyword:** Markov Decision Process

**16. An Improved Approximation for the Gaussian Q-Function**

We present a novel, simple and tight approximation for the Gaussian Q-function and its integer powers. Compared to other known closed-form approximations, an accuracy improvement is achieved over the whole range of positive arguments. The results can be efficiently applied in the evaluation of the symbol error probability (SEP) of digital modulations in the presence of additive white Gaussian noise (AWGN) and the average SEP (ASEP) over fading channels. As an example we evaluate in closed-form the ASEP of differentially encoded QPSK in Nakagami-m fading.

**Keywords:** approximation theory, AWGN channels, error statistics, Nakagami channels, quadrature phase shift keying

**Off-Policy vs. On-Policy**

Reinforcement Learning algorithms which are characterized as off-policy generally employ a separate behavior policy that is independent of the policy being improved upon; the behavior policy is used to simulate trajectories. A key benefit of this separation is that the behavior policy can operate by sampling all actions, whereas the estimation policy can be deterministic (e.g., greedy) [[1]](https://pemami4911.github.io/blog/2016/08/21/ddpg-rl.html#References). Q-learning is an off-policy algorithm, since it updates the Q values without making any assumptions about the actual policy being followed. Rather, the Q-learning algorithm simply states that the Q-value corresponding to state s(t)s(t) and action a(t)a(t) is updated using the Q-value of the next state s(t+1)s(t+1) and the action a(t+1)a(t+1) that maximizes the Q-value at state s(t+1)s(t+1). On-policy algorithms directly use the policy that is being estimated to sample trajectories during training (Emami [blog], n.d.).

On-policy methods attempt to evaluate or improve the policy that is used to make decisions, whereas off-policy methods evaluate or improve a policy different from that used to generate the data (Sutton, 1998).