Economics Bulletin

Volume 30, Issue 4

A stochastic analysis of goods allocation by queuing and the prevention of violence

Amitrajeet A Batabyal
Department of Economics, Rochester Institute of Technology

Gamini Herath
Monash University, Sunway Campus, Malaysia

Abstract

When demand for a good exceeds its supply, queuing mechanisms are commonly used to allocate the good in question to citizens. However, very long queues result in excessive wait times and this can lead to violence. As such, the purpose of this paper is to analyze two stochastic models of goods allocation with queuing and the possibility of violence. In the first model, there is no capacity constraint. Using this model, we compute the long run delay per citizen in allocating the pertinent good. Next, we discuss the computation of the equilibrium probabilities for our discrete-time Markov chain theoretic model. In the second model, we capture the violence aspect of the underlying story explicitly with a capacity constraint. Then, we compute the long run fraction of citizens who are not provided the relevant good and the long run fraction of time the good allocating public official is busy.

We thank the Editor John P. Conley and an anonymous referee for their helpful comments on a previous version of this paper. In addition, Batabyal acknowledges financial support from the Gosnell endowment at RIT. The usual disclaimer applies.
Citation: Amitrajeet A Batabyal and Gamini Herath, (2010) "A stochastic analysis of goods allocation by queuing and the prevention of violence", *Economics Bulletin*, Vol. 30 no.4 pp. 3143-3151.
Submitted: Sep 22 2010. **Published:** November 27, 2010.

1. Introduction

Public officials in many countries and particularly those in developing countries and in transition economies often use queuing mechanisms to allocate scarce goods to citizens. Examples of goods that have been allocated using queuing include rice in Sri Lanka (Gunawardana 2000), banking services in Nigeria (Woldie 2003), formal sector jobs in Nicaragua (Pisani and Pagan 2003), and jobs in general in urban Ethiopia (Serneels 2007). A key feature of these goods allocation mechanisms is that they involve waiting in line by citizens. Put differently, citizens have to actually wait in queue to obtain the good that is being allocated by the relevant public official.

The reader will understand that these queuing mechanisms take on particular salience during times of severe scarcity brought on by natural factors and/or by governmental policies. Batabyal (2005a) tells us that in such hard times, the available supplies of the germane scarce goods are allocated by public officials in many developing countries in particular using queuing mechanisms of one sort or another. From the standpoint of this paper, what is significant is that in times of acute scarcity, demand for a good greatly exceeds its supply. Therefore, if a good characterized by excess demand is to be allocated to citizens by means of a queuing mechanism then, in the time period of interest, it is important for a public provider to regulate the *length* of the queue. Concrete examples from two different countries explain why.

In Venezuela, President Hugo Chavez's desire to bring the powerful state owned oil company *Petroleos de Venezuela* (PDVSA) under his control has caused considerable political upheaval. As reported by the Economist (Anonymous, 2003, 2009), there was a severe shortage of both cooking gas and petrol in the country. As a result, in some provider facilities, there were queues of citizens that could be measured not in hours but in days. This situation gave rise not only to a secondary market for places in the fuel queues but it also resulted in considerable economic unrest and *violence*.

In Egypt, the policies of President Hosni Mubarak's government have made life for many of this nation's 75 million citizens rather difficult. In particular, the prices of many food items have been rising and it is now common for public officials to allocate scarce goods such as bread by means of queuing mechanisms. The Economist (Anonymous, 2008) describes instances where, *inter alia*, rising food prices have led to spontaneous protests and to urban riots that have led, as in Venezuela, to *violence*.

Very recently, violence broke out among the 50,000 people who queued to grab the last batch of Olympic tickets that went on sale in Beijing, China.[1] This example and the examples in the preceding two paragraphs help explain why it is necessary for a public provider—who uses a queuing mechanism to allocate a particular good—to regulate the *length* of the pertinent queue. Such an action effectively caps the length of the queue. In turn, this prevents the phenomenon of overly long citizen wait times and thereby reduces the likelihood of violence.

---

1 Go to http://www.abc.net.au/news/stories/2008/07/25/2315138.htm for additional details on this incident.

Despite the demonstrated importance of having a capacity constraint for queue based goods allocation mechanisms, to the best of our knowledge, this question has received virtually no theoretical attention in the literature. Lui (1985), Batabyal and Nijkamp (2004), and Batabyal and Yoo (2007) have used queuing models to study bribery and corruption. Athreya and Majumdar (2005) have analyzed a queuing model in which a reorganization of two government departments may increase efficiency. Although all four of these papers refer either to the optimal queue capacity determination issue or to the expected queue size in a stochastic equilibrium, they do *not* explicitly analyze the question of allocating a good with a queuing mechanism when violence is a potential problem. Recently, Batabyal (2005a) has shed light on the question of violence prevention in the context of goods allocation with queuing mechanisms. Specifically, Batabyal (2005a) uses the so called $M/M/1$ queuing model—see Ross (2003, pp. 480-490) or Tijms (2003, pp. 188-190) for textbook expositions—to determine the smallest capacity that will keep the likelihood of violence below an exogenously specified value.

In this paper, we use a new discrete-time Markov chain theoretic framework and extend the analysis in Batabyal (2005a) in particular by analyzing hitherto unstudied questions concerning the allocation of goods using queuing mechanisms. Specifically, we analyze two stochastic models of goods allocation with queuing and the possibility of violence. In the first model, there is no capacity constraint. Using this model, we first compute the long run delay per citizen in allocating the pertinent good and then we discuss the computation of our Markov chain model's equilibrium probabilities. In the second model, we capture the violence aspect of the underlying story with a capacity constraint. Then, we ascertain the long run fraction of citizens who are not provided the relevant good and the long run fraction of time the good allocating public official is busy.

The rest of this paper is organized as follows. Section 2.1 delineates our discrete-time Markov chain theoretic model of a good allocation process with no capacity constraint. Section 2.2 computes the long run delay per citizen in allocating the germane good. Section 2.3 discusses how the equilibrium probabilities for our discrete-time Markov chain theoretic model might actually be computed by demonstrating that the equilibrium distribution of this model has a geometric tail. Next, section 3.1 describes a discrete-time Markov chain theoretic model of a good allocation process with a capacity constraint. Section 3.2 determines the long run fraction of citizens who are not provided the good in question. Section 3.3 computes the long run fraction of time the good allocating public official is busy. Finally, section 4 concludes and then discusses one way in which the research in this paper might be extended.

2. A Good Allocation Model With No Capacity Constraint

2.1. Preliminaries

Consider a particular region in a developing country or in a transition economy in a time of acute economic scarcity. Within this region, we focus on a single public official who is in charge of allocating an essential but scarce good such as bread or rice to citizens. It takes this official a probabilistic amount of time to allocate the good in question to citizens. We suppose that these probabilistic good allocation times are independent random variables with a common Erlang (r,μ)

distribution.[2] For reasons of mathematical tractability, we suppose that $r/\mu<1$.

In each time period, a citizen arrives at our public official's distribution facility to obtain a single unit of the good under study. By "single unit of the good" we mean something like a single loaf of bread, a single bag of rice, or a single cannister of water. Our public official interacts with a single citizen at a point in time. This official's distribution facility has enough space so that arriving citizens who find the official busy wait in a queue until it is their turn to be allocated the good by this official.

The reader will note that in the model of this section, we do *not* have a capacity constraint on the number of queuing citizens that our public official's distribution facility can hold. Therefore, it is certainly possible that violence will break out among the queuing citizens if the long run average delay per citizen in allocating the pertinent good gets too long. Given this state of affairs, let us now derive an expression for this long run average delay per citizen.

2.2. Long run average delay

In order to use a discrete-time Markov chain theoretic framework, let us recall the contents of footnote 2 and use the fact that the Erlang (r,μ) distributed good allocation time can be thought of as the sum of r independent subtasks each of which has an exponential distribution with mean $(1/\mu)$. Now, from the standpoint of our good allocating public official, let X_n denote the number of uncompleted subtasks present just before the arrival epoch of the nth citizen. Then, the stochastic process $\{X_n\}$ is a discrete-time Markov chain with state space $I=\{0,1,...\}$.[3]

The one-step transition probabilities for this Markov chain are given by

$$p_{ij}=e^{-\mu}\frac{\mu^{i+r-j}}{(i+r-j)!},\ for\ \ i=0,1,...,\ j=1,...,i+r, \qquad (1)$$

where $p_{i0}=1-\Sigma_{j=1}^{i+r}\ p_{ij}$. The reader should note that in writing equation (1) and the expression for p_{i0},

2

Economists such as Lui (1985) and Batabyal (2005a, 2005b) who have studied the allocation of goods with queuing mechanisms have frequently assumed that these goods allocation—or more generally service—times are exponentially distributed. We are using the Erlang distribution in this paper because this distribution is a more general distribution than the exponential distribution. Specifically, an Erlang (r,μ) distributed random variable can be decomposed as the sum of r independent random variables each of which is exponentially distributed with the same mean $(1/\mu)$. See Tijms (2003, pp. 442-443) for additional details on the Erlang distribution.

3

To keep the subsequent mathematical analysis straightforward, we suppose that this Markov chain satisfies assumption 3.3.1 in Tijms (2003, p. 98). In other words, we are assuming that our discrete-time Markov chain has a regeneration state that is ultimately reached from every initial state with probability one and that the expected number of steps needed to return from this regeneration state to itself is finite.

we have used the fact that when our public official is not idle, his service completions of subtasks, as it were, occur in accordance with a Poisson process—see Ross (2003, pp. 288-329) or Tijms (2003, pp. 1-31) for textbook accounts—with parameter μ.

Let $\{\pi_j\}$ denote the equilibrium distribution for our discrete-time Markov chain. This means that

$$\pi_j = \Sigma_{k \in I}\ \pi_k\ p_{kj},\ j \in I. \tag{2}$$

With this definition in place, some thought tells us that the long run average delay per citizen in allocating the germane good is given by

$$\textit{Long run average delay} = \{\Sigma_{j=1}^{\infty}\ j\pi_j\}/\mu. \tag{3}$$

From a practical perspective, equation (3) is of great interest because the likelihood of violence in our public official's distribution facility depends fundamentally on the ratio on the right-hand-side (RHS) of this equation. In particular, if the probability of violence is to be kept low then this RHS ratio must be made small. Note that this ratio can be made small by making μ, a parameter of the Erlang distributed good allocation times, as large as possible.

A second way in which the likelihood of violence can be kept low is for our public official to put in place an explicit capacity constraint for his distribution facility. Practically, what this means is that when the number of queuing citizens reaches a certain finite number, no further citizens are allowed into this facility during the relevant time period. We shall study the impact of an explicit capacity constraint in section 3 below but before we do that, we would like to briefly discuss whether, in a given practical setting, the equilibrium probabilities for our Markov chain model, i.e., the $\pi_j's$, can actually be computed.

2.3. The geometric tail

The computation of the equilibrium probabilities for our Markov chain model is made difficult by the fact that the state space $I = \{0, 1, ...\}$ for this Markov chain $\{X_n\}$ is *infinite*. In such cases, it is standard to use the so called "geometric tail approach"—see Tijms (2003, pp. 111-114)—to compute the equilibrium probabilities of the relevant Markov chain. However, before this approach can be used, one must first show that the Markov chain under consideration has a geometric tail. We now show that the Markov chain $\{X_n\}$ that we have been studying thus far in this section does indeed have a geometric tail.

We shall use the discussion in Tijms (2003, pp. 111-113) to make our argument. Now, to show that our Markov chain $\{X_n\}$ has a geometric tail, we have to show that condition B in Tijms (2003, p. 113) is satisfied. To this end, let our Erlang distribution's parameter r equal s in condition

B. In addition, let $e^{-\mu}\mu^{r-k}/(r-k)!$ equal α^k in condition B. Then, using condition B we see that as $j\to\infty$, $\pi_j \sim \gamma\eta^j$, where γ is a constant and η is the unique root of the equation $\omega^r - e^{-\mu(1-\omega)} = 0$—also see equation 3.4.9 in Tijms (2003, p. 113)—on the unit interval $(0,1)$.

We have now shown that condition B in Tijms (2003, p. 113) is satisfied and hence the Markov chain of interest $\{X_n\}$ does have a geometric tail. This means that the infinite system of linear equations for the π_j can be reduced to a *finite* system of linear equations using the geometric tail method described in Tijms (2003, pp. 113-116). From a practical perspective, this means that the equilibrium probabilities for our Markov chain $\{X_n\}$ can actually be computed. We now proceed to analyze a discrete-time Markov chain theoretic model of a good allocation process with a capacity constraint.

3. A Good Allocation Model With a Capacity Constraint

3.1. Preliminaries

The stochastic environment in which our public official now allocates the good in question to arriving citizens is the same as in section 2.1. The only difference is that in order to reduce the likelihood of violence arising from an inordinately lengthy long run delay per citizen in allocating the germane good (see equation (3)), our public official puts in place an explicit constraint on how many citizens may queue in his distribution facility.

In particular, we suppose that once K citizens are in the distribution facility, no more citizens are allowed into this facility. Citizens who arrive at this facility when it is full, i.e., when K citizens are already present, are turned away and they leave without obtaining the scarce good under study. In this setting, our first task now is to determine the long run fraction of citizens who are turned away from the distribution facility and hence not provided with the good in question.

3.2. Long run fraction of citizens

Our discrete-time Markov chain $\{X_n\}$ is the same as in section 2 except that the state space is now different. Specifically, the state space is $I = \{0, 1, ..., (K+1)r\}$, where K is the finite capacity put in place by our public official to reduce the likelihood of violence in the distribution facility. We now have to specify the one-step transition probabilities for our Markov chain $\{X_n\}$ with the constricted state space. To this end, note that for any state i with $0 \le i \le Kr$, we have

$$p_{ij} = e^{-\mu}\frac{\mu^{i+r-j}}{(i+r-j)!} \quad for \;\; 1 \le j \le i+r, \tag{4}$$

and, for any state i with $Kr < i \le (K+1)r$, we have

$$p_{ij} = e^{-\mu} \frac{\mu^{i-j}}{(i-j)!} \; for \; 1 \le j \le i. \tag{5}$$

Following the methodology employed in section 2.2, let us denote the equilibrium distribution for the Markov chain of this section by $\{\pi_i, \; 0 \le i \le (K+1)r\}$. Now, using the definition of the equilibrium distribution and some thought we can deduce that the long run fraction of citizens who are turned away from the distribution facility and hence not provided the good in question is given by

$$Long \; run \; fraction \; of \; citizens \; without \; good = \Sigma_{i=Kr+1}^{(K+1)r} \; \pi_i. \tag{6}$$

Equation (6) tells us that when our public official's focus is on reducing the likelihood of violence in his distribution facility, the long run fraction of citizens who will come away from this distribution facility without the pertinent good is equal to the *sum* of the constricted state space Markov chain's equilibrium probabilities. Inspecting equation (6) it is straightforward to confirm that the larger the individual state equilibrium probabilities for the states involved in the summation on the RHS of equation (6), the larger is the long run fraction of citizens who do not obtain the good in question. We now proceed to our final task in this paper and that is to ascertain the long run fraction of time our good allocating public official is busy.

3.3. Long run fraction of time

To compute this fraction, first note that the arrival rate of citizens who are admitted into our public official's distribution facility and not turned away is given by $\Sigma_{i=0}^{Kr} \; \pi_i$. Now using Little's formula[4] with the preceding result, we reason that the long run fraction of time our public official is busy is given by

$$Long \; run \; fraction \; of \; time \; busy = (r/\mu)\Sigma_{i=0}^{Kr} \; \pi_i. \tag{7}$$

Equation (7) tells us that the long run fraction of time our public official is busy allocating the good in question to citizens is given by the product of the Erlang distribution's two parameters (r/μ) and the sum of the equilibrium probabilities of being in a specific number of states. Inspecting equation (7), it is straightforward to confirm two results. First, as in the case of equation (3) in section 2.2, the long run fraction of time our public official is busy can be *reduced* by increasing the value of a particular parameter of the Erlang distribution, namely, μ. Second, as either the Erlang distribution parameter r or the finite capacity K increases, the long run fraction of time our public official is busy also *increases*.

4

Little's formula is a well known result that applies to virtually all queuing systems. See Ross (2003, pp. 476-478) and Tijms (2003, pp. 50-53) for textbook expositions of Little's formula.

In section 2 with no capacity constraint, we derived a single metric for the likelihood of violence, namely, equation (3). However, in this section, we have two metrics for the likelihood of violence and these two metrics capture two different aspects of this likelihood. To see this, observe that violence can occur either inside or outside our public official's distribution facility. Equation (6) pertains to citizens who have been turned away from the distribution facility and hence this equation is a measure of the likelihood of violence *outside* the distribution facility. In contrast, equation (7) concerns citizens who are inside the distribution facility and hence this equation is a measure of the likelihood of violence *inside* our public official's distribution facility. The upshot of this discussion is that our public official is aware that violence can occur both outside and inside the distribution facility and he is *not* only paying attention to the likelihood of violence inside the distribution facility.

If we think of the capacity constraint K as a control variable for our public official then inspection of equations (6) and (7) tells us that both these long run fractions are increasing functions of K. Therefore, one possible course of action for our public official would be to choose K and optimize a weighted sum of the two metrics in equations (6) and (7) where the weights sum to unity and magnitude of the weights reflects the emphasis placed by this public official on the likelihood of violence either outside or inside the distribution facility. The optimal value of K would be the solution to this weighted optimization problem and this value may or may not be "small." Having said this, note that if the concern is that the optimal value of K might be "extremely small" then one could solve this weighted optimization problem with an explicit constraint requiring K to be bigger than some desired positive integer and this positive integer would then become the effective lower bound for K. This completes our discussion of the good allocation model with a capacity constraint.

4. Conclusions

In this paper, we analyzed two discrete-time Markov chain theoretic models of goods allocation by a public official with queuing and with the possibility of violence. In the first model, there was no capacity constraint. Using this model, we first computed the long run delay per citizen in allocating the pertinent good. Next, we discussed a way in which the equilibrium probabilities for our Markov chain model might actually be computed. In the second model, we captured the violence aspect of the underlying story with a capacity constraint. Then, we first determined the long run fraction of citizens who are not provided the relevant good and next we ascertained the long run fraction of time the good allocating public official is busy.

The analysis in this paper can be extended in a number of different directions. In what follows, we suggest one potential extension. Consistent with the discussion in the last paragraph of section 3, it would be useful to set up and solve an optimization problem for our public official in which this official selects the finite capacity K to, for instance, optimize the net social benefit to the region under study from the allocation of the pertinent scarce good. Studies that analyze this aspect of the problem will enhance our understanding of the nexuses between the allocation of goods by means of queuing mechanisms and the question of preventing one or more kinds of violence.

References

Anonymous. (2003) "Pyrrhus of Caracas" *The Economist*, January 2.

Anonymous. (2008) "Will the dam burst?" *The Economist*, September 11.

Anonymous. (2009) "Oblivious to the coming storm" *The Economist*, February 9.

Athreya, K.B. and M. Majumdar (2005) "A note on closing the windows: Some lessons from queuing theory" *International Journal of Economic Theory* **1**, 73-81.

Batabyal, A.A. (2005a) "On the allocation of commodities by queuing and the prevention of violence" *Economics Bulletin* **15**, 1-7.

Batabyal, A.A. (2005b) "On bribing and balking in a simple queuing model of resource allocation" *Economics Bulletin* **16**, 1-10.

Batabyal, A.A. and P. Nijkamp (2004) "Favoritism in the public provision of goods in developing countries" *Economics Bulletin* **15**, 1-10.

Batabyal, A.A. and S.J. Yoo (2007) "Corruption, bribery, and wait times in the public allocation of goods in developing countries" *Review of Development Economics* **11**, 507-17.

Gunawardana, P.J. (2000) "Concessional sales, open market demand, and the consumption of rice in Sri Lanka" *International Journal of Social Economics* **27**, 847-61.

Lui, F.T. (1985) "An equilibrium queuing model of bribery" *Journal of Political Economy* **93**, 760-81.

Pisani, M.J. and J.A. Pagan (2003) "Sectoral queuing in a transitional economy: The case of Nicaragua in the 1990s" *Labour* **17**, 571-97.

Ross, S.M. (2003) *Introduction to Probability Models*, 8^{th} edition, Academic Press: San Diego, California.

Serneels, P. (2007) "The nature of unemployment among young males in urban Ethiopia" *Review of Development Economics* **11**, 170-86.

Tijms, H.C. (2003) *A First Course in Stochastic Models*. Wiley: Chichester, UK.

Woldie, A. (2003) "Nigerian banks—quality of services" *Journal of African Business* **4**, 69-87.