

Probability in Action: the Red Traffic Light

John A. Shanks
University of Otago

Journal of Statistics Education Volume 15, Number 1 (2007),
www.amstat.org/publications/jse/v15n1/shanks.html

Copyright © 2007 by John A. Shanks all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words:Distributions; Modelling; Optimization; Problem solving;

Abstract

Emphasis on problem solving in mathematics has gained considerable attention in recent years. While statistics teaching has always been problem driven, the same cannot be said for the teaching of probability where discrete examples involving coins and playing cards are often the norm. This article describes an application of simple probability distributions to a practical problem involving a car's approach to a red traffic light, and draws on the ideas of density functions, expected value and conditional distributions. It provides a valuable exercise in applying theory in a practical context.

1. Introduction

Problem solving resources in the teaching of probability has attracted some attention recently. Some authors have championed the cause — for example, [Leviatan \(1998, 2000, 2002\)](#), [O'Connell \(1999\)](#) and [Shaughnessy \(1992\)](#) — however resources are still scarce. This article describes an application of simple probability distributions to a practical problem involving a car's approach to a red traffic light. Although the mathematics is routine (involving little more than simple integration), the translation of the problem from words to formulae is what challenges those students unpractised at this important skill. As such it is a valuable resource for teaching and illustrating the use of probability in an applied context remote from the realm of statistical inference.

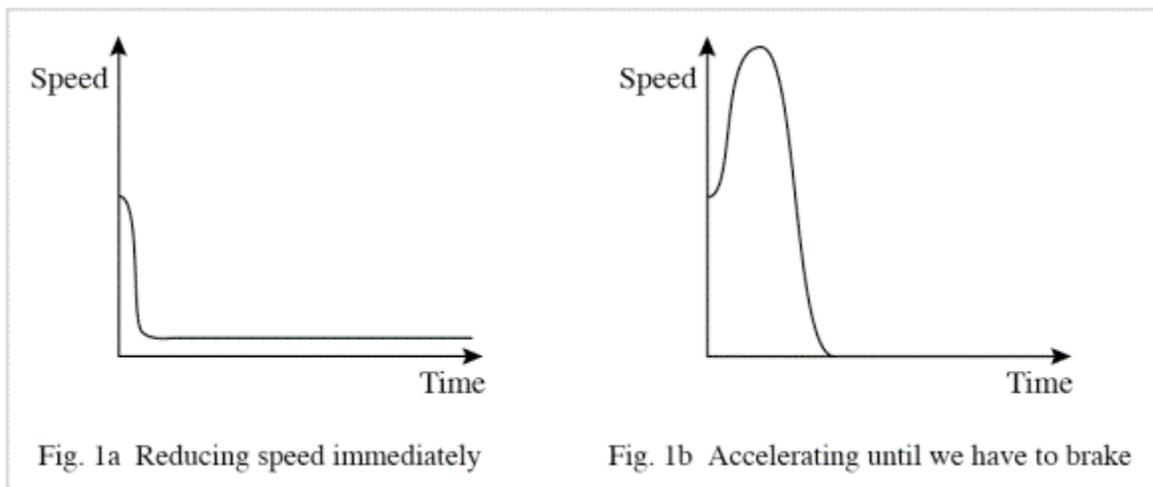
For a number of years students of a second-year paper in Probability Theory have been asked to solve this problem (broken down into bite-size pieces). Although the solution is straightforward, this is not an easy problem due partly, and perhaps mostly, to the requirement of translating from the words to the language of probability, something that the average student is not comfortable with without practice. The solution requires clear thinking about what is being asked, what the appropriate assumptions are, and what steps are needed, and an appreciation of the limiting process going from the discrete to the continuous. Only when this optimization problem has been couched in familiar terms can useful progress be made. This is in sharp contrast with the more usual assignment question that very neatly sets up a problem already phrased in probabilistic terms, such as that asking for a mean or variance of a clearly-defined distribution.

2. The problem

Traffic lights: our roads would be chaotic without them, but how frustrating they can be, especially the red ones! The timing of their changing pattern and the effects they have on road users are the subjects of various mathematical models. In statistics they often feature in queuing theory (for example, [Fox and Brooks \(2004\)](#), [Baykal-Gursoy and Xiao \(2004\)](#)), and the efficient operation of a series of neighbouring lights involves network analysis and control theory. Here we consider a simpler task: how best a driver should approach a red traffic light.

Imagine that we are driving along a road, we round a corner and in front of us some distance away is a traffic light, showing red. Wanting to conserve fuel and maintain a steady ride, we would prefer not to slow down and lose momentum unduly, but to continue at our present speed and for the light to change some time before we reach the intersection. There is a possibility of course that it will not change and that we will have to slow down if not stop completely.

It is interesting to consider what strategy we should adopt to maximize our speed at the instant that the light does change, without our knowing precisely when that will be. Let us consider a couple of extreme possibilities, assuming that we are initially travelling at 50 kph. Firstly, we could brake sharply to a slow 5 kph so that we are unlikely to have to stop at the light ([Figure 1a](#)). But if we do that, it seems that we have immediately abandoned any hope of maximizing our speed when the light changes. Secondly, we could accelerate to 100 kph so that if the light changes (hopefully soon!) we will be able to speed through the intersection ([Figure 1b](#)) feeling pleased with our success. But, with this plan we are approaching the light very quickly and it is more likely that it will change after we have screeched to an annoying halt rather than while we have a high speed. Maybe some intermediate strategy is better: perhaps continue at our present speed for as long as we can, or maybe decelerate steadily to a stop at the light. Just what is the best plan?



[Figure 1](#)

In the following sections we will explore the optimum strategy that aims to maximize our speed when the light changes. Incidentally, trying to maximize the distance travelled before the light changes is of no interest: the solution in that case is to use maximum acceleration (and then maximum deceleration!) so that the light is reached in the minimum time.

First we need some assumptions. Let us be bold and assume that there are no other vehicles in our path

and that there is no speed limit, so that we can travel at whatever speed is deemed optimal. However, we will not want to reverse — that is, our speed towards the light is always non-negative. (In fact, a speed that is sometimes negative would not invalidate the following analysis, but let's be reasonable.) In most places around the world traffic lights change directly from red to green, but in a few countries, notably the U.K., there is a “get ready” intermediate stage where red and amber show together. We will assume the direct change.

3. A simple model

The crux of our dilemma is that we don't know when the light will change. If we did know, then we could adjust the speed accordingly, perhaps slowing down at first and then accelerating so that we are really motoring just as the light changes. This might be the case if we are familiar with this set of lights and we saw them change to red; we might then know exactly how long we have before the green light appears. However, we are assuming that the light is already red when we first see it, so in practice we have no definite way of predicting just when it will change. This forces us to use a stochastic model based on the distribution of the time it takes until the light changes.

The simplest model worth considering covers the situation where we don't know when the light will change, but we do know a maximum time — that is, how long the light stays red in its normal cycle. We will call this time R , which could be, say, 30 seconds, and assume that it is both known and fixed. This model is totally reasonable if we are familiar with these particular traffic lights, or if all traffic lights in our immediate neighbourhood stay red for the same time, R . Later we will consider the situation where R is unknown (and hence treated as a random variable).

We will measure time t from the moment we first see the red light. Given that the light is already red, the actual time T for it to change to green must be in the interval $[0, R]$. In fact, if our arrival is completely random then T is uniformly distributed on that interval; the density function representing the probability distribution of T is $1/R$.

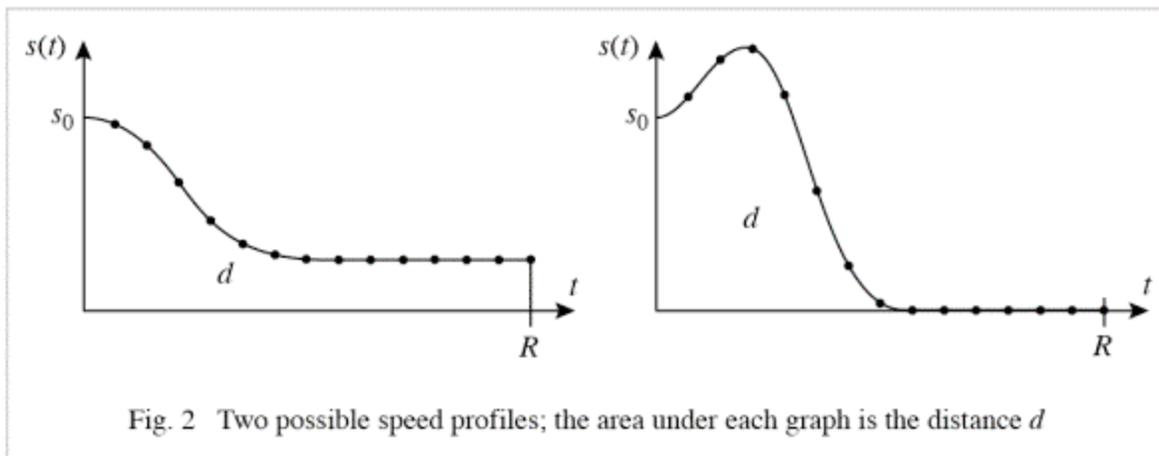
We need two more constants: the speed at which we are initially travelling, s_0 , and the distance to the light, d . Let the subsequent speed be given by the function $s(t)$ of time t . Then we have the following restrictions on s :

$$s(t) \geq 0, \quad s(0) = s_0, \quad \text{and} \quad \int_0^R s(t) dt = d \quad (1)$$

Recall that the area under the graph of any speed function gives the distance travelled. This distance is given by the definite integral in (1): in time R we must not travel more than d , and if we are looking for an optimum solution there is no point in travelling less than d — to do so means that we could have used a higher speed throughout. The choice of the upper limit R deserves some comment. It is tempting to use T here, as T is the actual time when the light changes — but T is unknown and we cannot possibly choose a speed function $s(t)$ in this case. Instead we have to plan our speed $s(t)$ over the *whole* interval $[0, R]$ in such a way as to cover the distance d .

Another important point to note is that $s(t)$ represents our *planned speed* to solve our maximization problem; once the light changes to green, our *actual* speed could well change. To help with understanding both this point and our maximization problem, [Figure 2](#) shows two possible speed profiles, one with a speed that is initially reduced and then held constant, the other with an acceleration before a sharp deceleration to zero speed. In both cases the area under the graph is equal to d . The value of T will fall somewhere between 0 and R . It may help to look at the discretized version of the problem: if T can occur only at a finite number of times (for example, every second), the marked points on each

speed curve represent the corresponding (equally-likely) speeds. The mean of the marked speeds is the expected speed when the light changes. The continuous analogue of this discrete mean is the integral of the speed divided by the time R as we will see next.



[Figure 2](#)

Returning to the continuous problem, if the light changes at time T then at that instant we will be travelling at speed $s(T)$. If T has a probability density function $f(t)$, then using the standard formula for expected value, the expected speed E at the time the light changes to green will be given by

$$E = \int_0^R s(t) f(t) dt \quad (2)$$

Since, for our simple model, $f(t) = 1/R$, we have

$$E = \int_0^R s(t) \frac{1}{R} dt = \frac{1}{R} \int_0^R s(t) dt = \frac{1}{R} d$$

Both d and R are fixed, so this result tells us that the expected speed when the light changes is a constant, not dependent on the speed function at all! We can therefore choose any speed profile, consistent with the conditions (1), and the result will be the same. So the two extreme cases we looked at earlier (and all others that fit (1), for example those in [Figure 2](#)) are in fact equivalent: our expected speed when the light changes will be d/R , which coincidentally is the average speed to travel the distance d in time R . This would certainly be the expected speed if we were to drop our speed instantaneously to the constant d/R at the start and maintain it for the time R — whenever the light changes, our speed would be d/R .

This is both a startlingly simple result and perhaps also a disappointing one — we may have hoped for some very special optimum strategy to adopt, some cleverly conceived formula that we alone could follow. But then we did choose a simple model and maybe there is more interest to be had when we extend that model.

4. A realistic model

The general problem we wish to study is how to choose $s(t)$ to maximize

$$E = \int_0^R s(t) f(t) dt$$

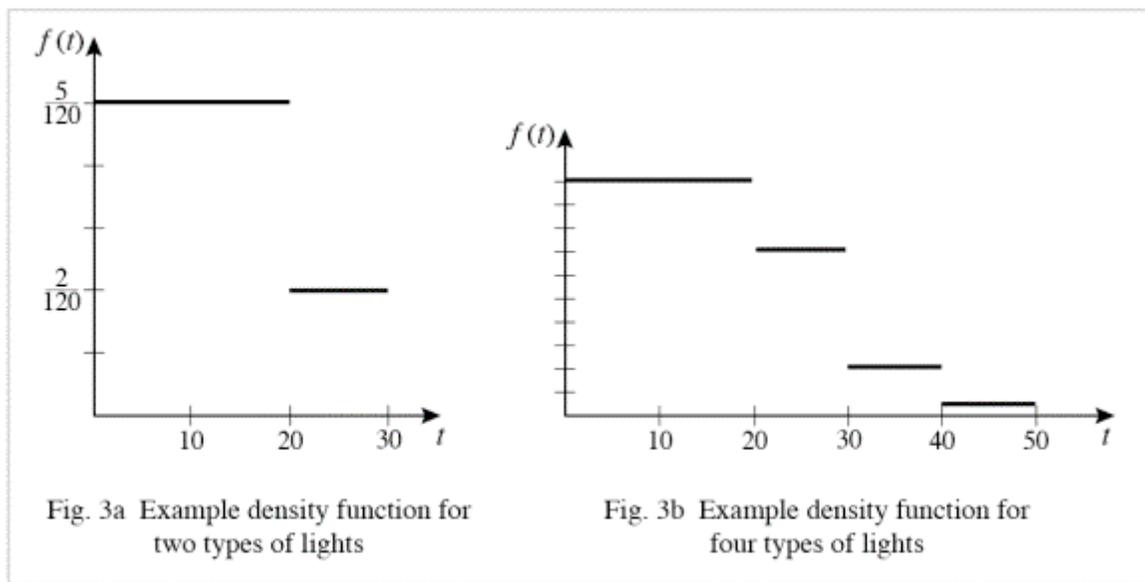
subject to (1), where $f(t)$ represents the probability distribution of the time T until the light changes. While we may find solutions for particular density functions $f(t)$, it is unlikely that a solution can be found for a general $f(t)$.

However, let us look a little closer at the types of distribution relevant to our particular situation. As we have seen and used earlier, for a *given* traffic light already showing red, the time T until it changes to green is uniformly distributed on the interval $[0, R]$, where R is the total time the light stays on red. The situation is made more complicated by the fact that R is unknown when we are approaching an unfamiliar set of lights. In fact R has its own probability distribution. If we consider all traffic lights in the world there will be a minimum value r_{\min} and a maximum value r_{\max} for R . We can only guess what these may be. Perhaps $r_{\min} = 10$ secs and $r_{\max} = 60$ secs. In a particular city, there will be less extreme values, maybe $r_{\min} = 15$ secs and $r_{\max} = 40$ secs.

Initially let us take a very simple distribution for R . (We will later consider when R can take on a continuum of values.) Suppose in our city half the traffic lights stay red for 20 secs and half for 30 secs. So $R = 20$ with a probability of $1/2$, and $R = 30$ also with a probability of $1/2$. The uniform density function for the time T corresponding to the first type of lights is $f_1(t) = 1/20$ for t in $[0, 20]$, and that for the second type is $f_2(t) = 1/30$ for t in $[0, 30]$. If we don't know to which type our traffic light belongs, the overall density function for T is $f(t) = 1/2 f_1(t) + 1/2 f_2(t)$, or

$$f(t) = \begin{cases} \frac{1}{40} + \frac{1}{60} = \frac{5}{120}, & \text{on } [0, 20] \\ \frac{1}{60} = \frac{2}{120}, & \text{on } [20, 30] \end{cases}$$

[Figure 3a](#) shows the density function $f(t)$, a simple example of a *mixture* distribution.



[Figure 3](#)

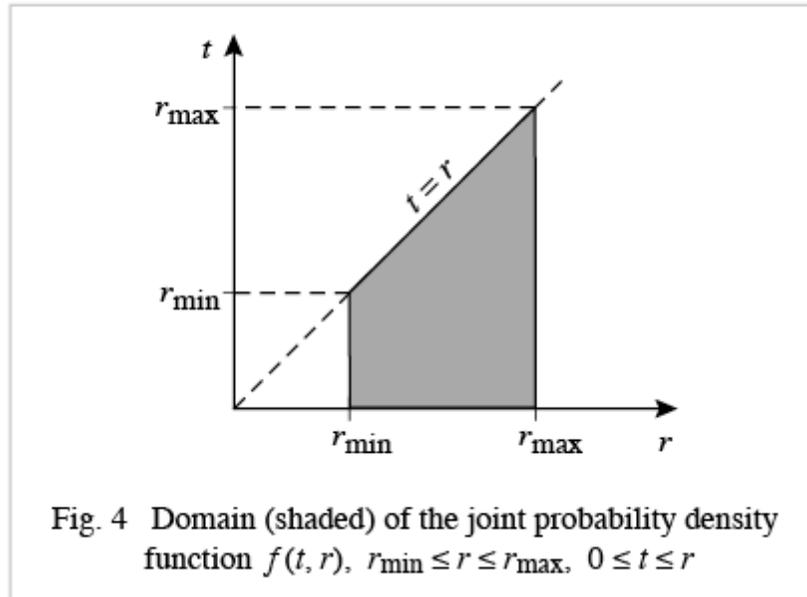
For a more complicated example, consider when 20% of lights stay red for 20 secs, 50% stay red for 30 secs, 20% for 40 secs and 10% for 50 secs. Then the overall density function for an unfamiliar traffic light is (see [Figure 3b](#)) $f(t) = 0.2 f_1(t) + 0.5 f_2(t) + 0.2 f_3(t) + 0.1 f_4(t)$, or

$$f(t) = \begin{cases} \frac{2}{10} \frac{1}{20} + \frac{5}{10} \frac{1}{30} + \frac{2}{10} \frac{1}{40} + \frac{1}{10} \frac{1}{50} = \frac{101}{3000}, & \text{on } [0, 20] \\ \frac{5}{10} \frac{1}{30} + \frac{2}{10} \frac{1}{40} + \frac{1}{10} \frac{1}{50} = \frac{71}{3000}, & \text{on } [20, 30] \\ \frac{2}{10} \frac{1}{40} + \frac{1}{10} \frac{1}{50} = \frac{21}{3000}, & \text{on } [30, 40] \\ \frac{1}{10} \frac{1}{50} = \frac{6}{3000}, & \text{on } [40, 50] \end{cases}$$

Note that in the interval $[0, r_{\min}] = [0, 20]$, $f(t)$ is made up of contributions from all the $f_i(t)$. For t larger than r_{\min} , only relevant terms are included, that is, terms for which f_i is non-zero at that particular time.

The analysis here is based on joint and conditional distributions. Since both T and R are random variables, there is an underlying joint probability distribution given by some density function $f(t, r)$, where $0 \leq t \leq r$. The individual density functions $f_i(t)$ are each conditional on a discrete value r_i of the parameter R . These could be written $f_i(t) = f(t | R = r_i)$ to show their conditional nature, where the r_i are the possible values of R . The resulting distribution for T , given by the density $f(t)$, is a marginal distribution. (For example see [Lindgren \(1994\)](#), [Wackerly, et al. \(2002\)](#).) The process can be extended to the more general case where R has a probability distribution given by the continuous density function

$g(r)$ for $r_{\min} \leq r \leq r_{\max}$. In this case the joint density function is given by $f(t, r) = f(t | R = r) g(r)$; [Figure 4](#) shows its domain.



[Figure 4](#)

The marginal distribution for T is obtained by integrating $f(t, r)$ with respect to r — that is, along horizontal lines through the domain shown in [Figure 4](#). For $0 \leq t \leq r_{\min}$ this involves integrating over the whole interval $[r_{\min}, r_{\max}]$. However, for $r_{\min} \leq t \leq r_{\max}$, since the domain is restricted by $t \leq r$, the integration is over the interval $[t, r_{\max}]$, corresponding to the triangular part of the domain. Since, for any given value r of R , the time T is uniformly distributed on $[0, r]$, we have that $f(t | R = r) = 1/r$, and so, putting this altogether, we find that the marginal density function for T is

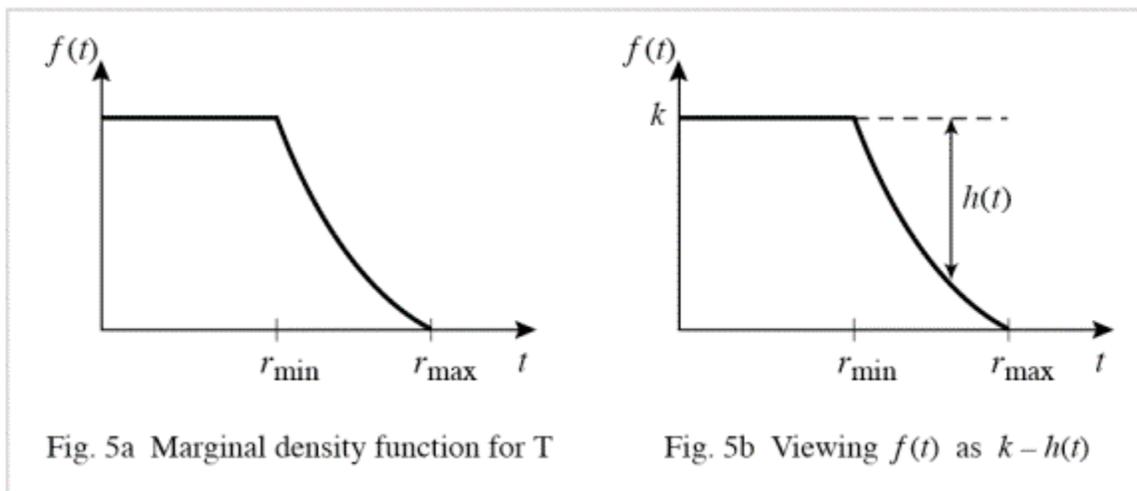
$$f(t) = \begin{cases} \int_{r_{\min}}^{r_{\max}} \frac{g(r)}{r} dr, & \text{if } 0 \leq t \leq r_{\min} \\ \int_t^{r_{\max}} \frac{g(r)}{r} dr, & \text{if } r_{\min} \leq t \leq r_{\max} \end{cases} \quad (3)$$

The first integral does not depend on t , and in the second integral t appears only in the lower limit, so it is clear that $f(t)$ will always be constant on $[0, r_{\min}]$ and decrease to 0 as t increases to r_{\max} , for any distribution $g(r)$.

For example, if R is uniformly distributed between r_{\min} and r_{\max} , then $g(r) = 1/(r_{\max} - r_{\min})$ on that interval and

$$f(t) = \begin{cases} \frac{\ln(r_{\max}) - \ln(r_{\min})}{r_{\max} - r_{\min}}, & \text{if } 0 \leq t \leq r_{\min} \\ \frac{\ln(r_{\max}) - \ln(t)}{r_{\max} - r_{\min}}, & \text{if } r_{\min} \leq t \leq r_{\max} \end{cases}$$

(See [Figure 5a.](#))



[Figure 5](#)

So in practice, whatever the distribution of R , the probability density $f(t)$ will be constant, say k , from $t = 0$ to time $t = r_{\min}$ and then will decrease to 0, either continuously or in discrete jumps, as t increases to r_{\max} . In fact it will be helpful to write $f(t)$ in the form

$$f(t) = k - \begin{cases} 0, & \text{if } 0 \leq t \leq r_{\min} \\ h(t), & \text{if } r_{\min} \leq t \leq r_{\max} \end{cases}$$

where $h(t)$ is a function having the property $0 = h(r_{\min}) \leq h(t) \leq h(r_{\max}) = k$ on the interval $[r_{\min}, r_{\max}]$ (see [Figure 5b](#)).

We can now return to our optimization problem: we want to maximize the expected speed E at the time the light changes to green, where

$$E = \int_0^{r_{\max}} s(t) f(t) dt = \int_0^{r_{\max}} s(t) k dt - \int_{r_{\min}}^{r_{\max}} s(t) h(t) dt = kd - \int_{r_{\min}}^{r_{\max}} s(t) h(t) dt$$

The integral subtracted from the constant kd is positive and so we maximize E by minimizing this integral. We easily do this (without needing to know the details of $h(t)$) by taking $s(t) = 0$ on the interval $[r_{\min}, r_{\max}]$. Note that this choice imposes no restrictions on $s(t)$ over the interval $[0, r_{\min}]$, other than the conditions (1); so just like in the simple model we can choose the speed how we wish.

The conclusion we draw depends on our knowing r_1 for our locality: we should aim to travel the distance d in this time by any speed profile we wish. The fact that $s(t) = 0$ for $t \geq r_{\min}$ fits in with the fact that the red light might not change to green at r_{\min} so we need to have stopped moving. Therefore:

Choose any speed function so that we are stopped at the traffic light at time r_{\min} .

In my own home town R appears to vary between 15 seconds and 35 seconds, so I should ensure that I have stopped precisely at the red light after 15 seconds. I don't want to be short of the light at that time nor should I be travelling at speed. I could however choose to reach the light and stop there before the 15 seconds.

We may wish to consider how a speed profile $s(t)$ can be chosen to satisfy the requirements:

$$s(t) \geq 0 \text{ on } [0, r_{\min}], \quad s(0) = s_0, \quad s(r_{\min}) = 0 \quad \text{and} \quad \int_0^{r_{\min}} s(t) dt = d$$

There are many types of profile that we could adopt, and there are some interesting mathematical curve-fitting concepts that can be tackled here. But we will not pursue these: the driver has no time to perform curve fitting, nor the computational means, presumably. An experienced driver should be able to achieve the desired result — being stopped at the light at time r_{\min} — with reasonable accuracy. If a sensible fairly smooth profile is chosen then there will be savings of both fuel and wear on the engine and brakes.

5. Related problems

What speed should we adopt if we *do* see the light change to red? If we are familiar with this traffic light then the problem is no longer stochastic and becomes one of kinematics; in this case, to avoid unrealistic solutions, we have to impose limits on acceleration and build in a speed limit. If the light is unknown to us, then we know that it will change at a time in the interval $[r_{\min}, r_{\max}]$, on which we will need to consider some distribution $f(t)$. The problem is certainly closely related to that studied above, but the solution is more complex in that again we need to impose reality conditions to achieve any useful result. We leave such considerations to the reader, who may wonder why it is that adding one extra detail produces a more difficult problem.

6. Using this exercise in the classroom

This has proved to be a stimulating problem and a worthwhile exercise in probability especially when posed as a group exercise. Listening to some of the discussions that arise in such group sessions exposes many difficulties in couching the problem in probabilistic terms, coping with the joint distribution, and then finding an optimum solution. Certainly students need to be comfortable with joint and marginal probabilities if they are to make good headway with little help.

The following teaching plan describes how the problem has been posed to students taking a second-year

paper in Probability that begins a more theoretical look at distributions, including discrete, continuous, joint and marginal distributions. (These students have previously passed an introductory paper in Statistics which covers only the very basics of probability, and a Calculus paper.) As indicated, the problem is tackled first in class and then in group sessions with assistance from tutors if required. Students are then expected to write up their own individual answers for grading. The essential steps of the plan follow this paper quite closely.

- Class: Initial discussion by the instructor describing the problem and inviting participation from the class at all times. Covers [Section 2](#), and introduces notation and basic ideas.
- Class: [Section 3](#). Discussion of the simple case when R is known, of the conditions imposed in [\(1\)](#), and derivation of the quantity we wish to minimize [\(2\)](#), leading to the result $E = d/R$. In particular, answer the questions:
 1. If our speed is $s(t)$ and the light changes at a time T with density function $f(t)$, then how can we express the expected value E of the speed when the light changes?
 2. What is E if T is uniformly distributed over $[0, R]$ (with R known)?
- Class: Brief outline of the tasks ahead: considering R as unknown and hence having its own probability distribution; the discrete case; the continuous case.
- Groups: Questions relating to [Section 4](#).
 3. If R is unknown but takes the value 20 secs with a probability of 1/2 and the value of 30 secs with a probability of 1/2, find and sketch the density function $f(t)$. (Students need to appreciate that there are two density functions here, each conditional on a value of R . They also need to be able to combine them to form $f(t)$.)
 4. Find the density function $f(t)$ for a more complicated discrete example such as that with four types of lights leading to [Figure 3b](#).
 5. If R has a continuous distribution $g(t)$ over some known interval $[r_{\min}, r_{\max}]$, draw the domain ([Figure 4](#)) of the joint distribution $f(t, r)$.
 6. For a value of t in the interval $[0, r_{\max}]$, how do we find the marginal density function $f(t)$? What complication is there in this process? Hence find an expression for $f(t)$ and make a typical sketch ([Figure 5](#)).
 7. What is E and how do we maximize it in this case? What then is the answer to the original problem?

Experience shows that most groups will require some assistance. Even the more able students benefit from knowing that they are on the right track.

The exercise could be capped off with some data collection: for the sake of mathematics, if not fuel economy, students could set out about town armed with a stopwatch to find their local value of r_{\min} .

Acknowledgement

The author wishes to thank the three referees and an Associate Editor for helpful comments which resulted in the clarification of a number of issues.

References

Baykal-Gursoy, M., Xiao, W., (2004), "Stochastic Decomposition in M/M/? Queues with Markov Modulated Service Rates," *Queueing Systems*, 48, 1-2.

Fox, J. and Brooks, G. (2004), "Optimisation of Traffic Signals Using an Expert System," Joint Thesis, Curtin University of Technology, Perth.

Leviatan, T. (1998), "Misconceptions, fallacies, and pitfalls in the teaching of probability," *Proceedings of ICM*, Berlin: IMU.

Leviatan, T. (2000), "Strategies and principles of problem solving in probability," *Proceedings of ICME-9*, Springer.

Leviatan, T. (2002), "On the use of paradoxes in the teaching of probability," *Proceedings of ICOTS-6*, South Africa: ISI, IASE.

Lindgren, B. W. (1994), *Statistical Theory*, London: Chapman and Hall.

O'Connell, A. (1999), "Understanding the Nature of Errors in Probability Problem-Solving," *Educational Research and Evaluation*, 5, 1.

Shaughnessy, J. M. (1992), "Research in probability and statistics: Reflections and directions," In Grouws, D. A. (Ed.), *Handbook of Research on Mathematics Teaching and Learning*, New York: Macmillan

Wackerly, D. D., Mendenhall, W. and Scheaffer, R. L. (2002), *Mathematical Statistics with Applications*, 6th Edition, Pacific Grove, California: Duxbury.

John A Shanks
Department of Mathematics and Statistics
University of Otago
Dunedin, New Zealand
jshanks@maths.otago.ac.nz

[Volume 15 \(2007\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Information Service](#) | [Editorial Board](#) | [Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)