On the Dating Problem

Trevor Standley and Alan Roytman

May 7, 2007

Abstract

We introduce a game called the Dating Problem in which a male has to make decisions about proposing, rejecting, or re-dating a potential female mate under imperfect knowledge of her quality. This problem is similar to a variety of problems that arise when estimates must be used to decide the eligibility of a candidate for a position. We are interested in the relationship between the quality of the woman eventually proposed to and the asymptotic expected number of dates required to ensure a fixed quality Q. We show that every strategy must take $\Omega(\frac{1}{1-Q})$ dates in expectation, and we propose a strategy that achieves $O(\frac{1}{1-Q} \log \frac{1}{1-Q})$ dates in expectation for a given quality.

Introduction

In the Dating Problem, a man goes on a potentially infinite series of dates. Each woman has a quality q_i that lies uniformly distributed in the interval $\mathbb{R}[0, 1]$. Initially, the man only knows that the woman's quality lies between 0 and 1. At the end of every date, however, the man discovers more information about the quality of each woman. In particular, with each date the man receives another interval of size 1, the start of which is uniformly distributed between $q_i - 1$ and q_i where her true quality must also lie. Upon receiving this information, the man must decide whether to propose to the woman (and end the game), reject the woman (whereupon he will start dating a never before seen woman), or ask the woman out on another date. By using all the information he has about the current woman, the man can deduce a shrinking interval within which this woman's true quality must lie (the intersection of all the known intervals for this woman), and thus obtain a better idea of the woman's true quality. The man's goal, of course, is to minimize the number of dates it takes to ensure that he proposes to a woman whose quality is at least some given Q.

Strategy

A naïve strategy is to date each woman as many times as required to be certain that her true quality does or does not lie above the given quality Q. If the man ever sees an interval that proves the woman's true quality lies below Q, then he should reject, and if he ever sees an interval that proves the woman's true quality lies above Q he should propose. He should re-date otherwise. Although in practice, this strategy may terminate in a reasonable amount of time, the analysis will show that this strategy requires an infinite number of dates in expectation. In practice this strategy does take far longer than the following strategy.

We propose a strategy where for some threshold $Q' > Q$ (to be optimized later) the man dates a woman until he either knows her true quality is less than Q' (in which case he rejects her) or he discovers that her true quality lies above Q (in which case he proposes). If neither of these conditions hold, the man chooses to date the same woman again. Although this strategy has the potential to reject qualified women, it discovers unqualified women using fewer dates than the naïve strategy requires.

Analysis

Under this strategy, for every new woman one of three events must occur. If the woman's quality lies in the range $\mathbb{R}[Q', 1]$ the strategy will have the man eventually propose. If the woman's quality q_i lies in the range $\mathbb{R}[Q, Q')$ the strategy may have the man eventually propose, or may have the man eventually reject. And if the woman's quality lies in the range $\mathbb{R}[0, Q)$ the man will reject eventually. The following table represents these three events:

	e_1: $Q' \leq q_i$	e_2: $Q \leq q_i < Q'$	e_3: $q_i < Q$
$\Pr[e]$	$1 - Q'$	$Q' - Q$	Q
$\Pr[\text{propose this date} \mid e]$	$q_i - Q$	$q_i - Q$	0
$\Pr[\text{reject this date} \mid e]$	0	$Q' - q_i$	$Q' - q_i$

From this table, we can calculate the expected number of dates our strategy will take given the fixed quality. In particular, if X is a random variable representing the number of dates spent by the strategy, then we can express the expectation as follows:

$$E[X] = (1-Q')E[T_1]+(Q'-Q)(E[\frac{q_i - Q}{Q' - Q}T_2]+E[\frac{Q' - q_i}{Q' - Q}T_3]+E[\frac{Q' - q_i}{Q' - Q}]E[X])+Q(E[T_4]+E[X])$$

where T_1 is the random variable representing the number of dates it takes to decide $q_i \geq Q$ given that $q_i \geq Q'$, T_2 is the random variable representing the number of dates it takes to decide $q_i \geq Q$ given that $Q \leq q_i < Q'$ and the man eventually proposes, T_3 is the random variable representing the number of dates it takes to decide $q_i < Q'$ given that $Q \leq q_i < Q'$ and the man eventually rejects, and T_4 is the random variable representing the number of dates it takes to decide $q_i < Q'$ given that $q_i < Q$.

In particular:

$T_1 = \frac{1}{q_i - Q}$, $T_2 = \frac{1}{q_i - Q}$, $T_3 = \frac{1}{Q' - q_i}$, $T_4 = \frac{1}{Q' - q_i}$. Thus, the expectations needed to calculate $E[X]$ are:

$$\begin{aligned}
E[T_1] &= \frac{1}{1-Q'} \int_{Q'}^{1} \frac{1}{q_i - Q}\, dq_i = \frac{1}{1-Q'} \ln(\frac{1-Q}{Q'-Q}) \\
E[\frac{q_i - Q}{Q' - Q}T_2] &= \frac{1}{Q'-Q} \int_{Q}^{Q'} \frac{1}{Q'-Q}\, dq_i = \frac{1}{Q'-Q} \\
E[\frac{Q' - q_i}{Q' - Q}T_3] &= \frac{1}{Q'-Q} \int_{Q}^{Q'} \frac{1}{Q'-Q}\, dq_i = \frac{1}{Q'-Q} \\
E[\frac{Q' - q_i}{Q' - Q}] &= \frac{1}{Q'-Q} \int_{Q}^{Q'} \frac{Q - q_i}{Q'-Q}\, dq_i = \frac{1}{2} \\
E[T_4] &= \frac{1}{Q} \int_{0}^{Q} \frac{1}{Q' - q_i}\, dq_i = \frac{1}{Q} \ln(\frac{Q'}{Q'-Q})
\end{aligned}$$

Putting everything together, we find:

$$E[X] = \ln(\frac{1-Q}{Q'-Q}) + (1 + 1 + \frac{1}{2}(Q'-Q)E[X]) + \ln(\frac{Q'}{Q'-Q}) + QE[X]$$

Solving for E[X]:

$$E[X] = \frac{\ln(\frac{1-Q}{Q'-Q}) + 2 + \ln(\frac{Q'}{Q'-Q})}{1 - \frac{Q'+Q}{2}}$$

We now have an expression involving only constants and the variable Q' which we are trying to optimize. Note that this equation explains why the naïve strategy discussed earlier requires an infinite number of dates in expectation. In particular,

$$\lim_{Q'\to Q} E[X] = \frac{\lim_{Q'\to Q}(\ln(\frac{1-Q}{Q'-Q}) + 2 + \ln(\frac{Q'}{Q'-Q}))}{1-Q} = \infty$$

Returning to our expression for $E[X]$, we can take the derivative $\frac{dE[X]}{dQ'}$ and arrive at the following expression:

$$\frac{(1-Q-\frac{Q'-Q}{2})(\frac{-Q'-Q}{Q'(Q'-Q)}) + \frac{1}{2}(\ln(\frac{(1-Q)Q'}{(Q'-Q)^2}) + 2)}{(1-Q-\frac{Q'-Q}{2})^2}$$

Unfortunately, the roots of this expression cannot be found analytically. However, using Newton's method, we have arrived at the following table:

Q	Q'
0.2	0.703401
0.4	0.789132
0.6	0.853577
0.7	0.8837
0.8	0.9154
0.85	0.9328

Q	Q'
0.9	0.9518
0.925	0.9622
0.95	0.9733
0.975	0.9856
0.985	0.9910
0.995	0.996755

The expected number of women the man sees under our strategy can be calculated as follows:

$$E[W] = Pr[\text{man rejects woman } i]E[W] + 1$$

$$E[W] = (\frac{1}{Q'-Q}\int_{Q}^{Q'} (Q'-q_i)\,dq_i + Q)E[W] + 1 = (\frac{Q'-Q}{2} + Q)E[W] + 1$$

$$E[W] = \frac{2}{2-Q-Q'}$$

Also, because the derivative is close to 0 for values near the optimal, Q' can be approximated without too much loss with the following expression: $\frac{Q+1}{2}$. Notice that if we use this value for Q' for the expected number of women our strategy requires $\frac{4}{3}\frac{1}{1-Q}$ women in expectation.

Lemma: Every strategy requires $\Omega(\frac{1}{1-Q})$ women.

Proof: The probability of a given woman having quality less than Q is Q. Thus, we have the following recursive equation:

$$E[W] = QE[W] + 1 \implies E[W] = \frac{1}{1-Q}$$

Because it takes this many women in expectation before any woman of the desired quality comes up, it must be the case that every strategy requires $\Omega(\frac{1}{1-Q})$ women in expectation.

We therefore see that our strategy is within a constant factor of the optimal number of women seen by the man. In other words, even if the man was clairvoyant and could instantly tell the exact quality of each woman, he would still require within a constant factor the same amount of women as our strategy requires.

Lemma: The optimal strategy for the dating problem takes $O(\frac{1}{1-Q}\ln\frac{1}{1-Q})$ dates.

Proof: By using a suboptimal value for Q', we cannot possibly decrease the expected number of dates. So, plugging $Q' = \frac{Q+1}{2}$ into E[X], we have:

$$E[X] = \frac{\ln(\frac{1-Q}{Q'-Q}) + 2 + \ln(\frac{Q'}{Q'-Q})}{1 - \frac{Q'+Q}{2}}$$

$$E[X] = \frac{\ln(\frac{1-Q}{\frac{Q+1}{2}-Q}) + 2 + \ln(\frac{\frac{Q+1}{2}}{\frac{Q+1}{2}-Q})}{1 - \frac{\frac{Q+1}{2}+Q}{2}}$$

$$E[X] = \frac{1}{\frac{3}{4}(1-Q)}(2 + \ln(2) + \ln(\frac{1+Q}{1-Q}))$$

Since Q is bounded above by 1, we have $E[X] = O(\frac{1}{1-Q}\ln\frac{1}{1-Q})$.
In particular, since such a strategy exists, it must be the case that the optimal strategy cannot be worse.

Data

The data from the following table assumes $Q' = \frac{Q+1}{2}$. The empirical portion of the table was generated using a simulation that ran for ten million iterations:

	Analytic Values		Empirical Averages:		
Q	$E[X]$	$E[W]$	X	W	Final q_i
0.7	19.6789	4.444444	19.6746	4.44343	0.883356
0.75	24.7416	5.333333	24.735	5.33264	0.902779
0.8	32.6025	6.666667	32.6026	6.66678	0.922208
0.85	46.2707	8.888889	46.2739	8.89005	0.941677
0.9	75.1678	13.33333	75.1757	13.3337	0.961113
0.925	105.57	17.77778	105.594	17.7821	0.970839
0.95	169.512	26.66667	169.51	26.6655	0.98056
0.975	376.672	53.33333	376.802	53.3557	0.990282
0.985	673.642	88.88889	673.449	88.8743	0.994168

The following table was calculated using optimal values for Q'. We can see that the difference between the optimal expectations/averages are very close to those obtained using the $Q' = \frac{Q+1}{2}$ approximation. The empirical data was also obtained using a simulation that ran for ten million iterations:

Q	$E[X]$	Empirical X
0.6	13.43	13.4292
0.7	19.5114	19.5141
0.75	24.586	24.5863
0.8	32.4727	32.4751
0.85	46.1876	46.1912
0.9	75.1517	75.182
0.925	105.569	105.568
0.95	169.382	169.422
0.975	375.148	375.133
0.985	668.601	667.959
0.995	2277.5	x

Merits, Limitations, and Open Problems

Relating this problem to reality assumes an infinite number of women. We believe that this is justified because there are more women in the world than any man could date, and therefore it would always be possible for a man to find another date. This problem also assumes the quality of every woman comes uniformly distributed between $\mathbb{R}[0,1]$. This too is a realistic assumption when one considers the concept of quality as it relates to percentiles. In particular, for any ranked set, the percentile of an element chosen at random from that set is uniformly distributed. In addition, the interval model of receiving information about a woman's true quality is realistic. If one considers a model in which every woman has several traits where each trait contributes to her quality, one can imagine learning of the presence or absence of specific traits during a date, yielding maximum and minimum possible qualities.

Because in the real world applicants for positions (women in our case) are pre-screened and applicants of higher quality are more likely to be interviewed, pre-screening has a large effect on the analysis of real world situations. This is not reflected by our model. Perhaps the largest limitation of our model is that women are assumed to always agree to marriage when the man proposes to the woman. A better model might assign probabilities based on quality to each woman representing the chance that she will reject or accept if she is proposed to, and so a more qualified woman might be less likely to marry the man when compared to a woman of lesser quality. One open problem we propose might be to analyze a model that reflects this subtlety, as it should make for a more challenging analysis. Another interesting problem might be to analyze the same Dating Problem where each woman's true quality q_i and the estimates come from some Gaussian distribution. However, in this case, it would be impossible to guarantee that the woman eventually proposed to has a quality which is at least any fixed Q.

Although we were unable to prove that our strategy's requirements were a constant factor away from the optimal strategy's, we suspect that because every strategy requires no less than $\Omega(\frac{1}{1-Q})$ women, and because a woman's quality becomes significantly more difficult to estimate as it gets closer to one (so the number of dates per woman probably also must increases as Q approaches one), our strategy may be optimal to within a constant factor.