The Samaritani Formula – more details

Adam Atkinson (ghira@mistral.co.uk)
Presented at G4G12 30th March to 4th April 2016
(Updated after real world events in late 2016 to early 2017)

"Fads and Fallacies" is one of my favourite Gardner books and I think the Samaritani Formula, at least as presented in the modern day, is the sort of thing Gardner would have found amusing and/or maddening.

This article contains more information than my 5 minute talk could possibly hope to. At the time of writing this, if I google "Samaritani Formula" the only hits in English I find relate to me talking about it elsewhere in 2015. If you look for "formula di Samaritani" you should find many hits in Italian, not all of which have anything to do with me. Googling "ritardi" and "Samaritani" and "lotto" in some combination should find quite a lot. And not all of it will be mathematicians arguing with Samaritani boosters.

If terms like "idiot" and "conperson" seem a bit strong, please bear in mind that bankruptcies and suicides are involved in this whole affair. I would regard the victims in this as, say, merely ill-informed or gullible. Alternatively, we could argue that "person who says untrue things without realising they are untrue, having spent some time thinking about them" and "person who says untrue things knowing them to be untrue" are much longer and would make the text unwieldy.

Italy has a lot of people offering or selling advice on how to pick lottery numbers, and while a lot of the stuff one sees is usual abuse of (some form of) the law of large numbers or the central limit theorem, one thing I've only come across in Italy is *quantitative* nonsense in the form of the Samaritani Formula. It's possible that the formula is also used nonsensically elsewhere under the same or a different name and if so I'd love to hear about it.

The amusing part is that actually the S.F. is not as nonsensical as it might seem at first glance: it merely doesn't mean what its modern boosters say it means. It is, of course, of no help whatsoever in playing the lottery.

My encounters with the Samaritani Formula have almost all involved people turning up in Italian newsgroups (typically it.scienza.matematica but also it.scienza.fisica, it.hobby.enigmi and others including the now defunct it.fan.dewdney) asking how many draws in a row a number in the Italian lottery can fail to turn up. If you are very brave, look at it.hobby.lotto. Well, actually they ask what the "theoretical maximum delay" in Lotto is. ("ritardo massimo teorico"). We have always taken this to mean "largest possible value" but at least a few ritardisti say it means something else.

Of course we all say "there is no limit" or "infinity" or similar but some of these people want to be told the answer is 220, or tell us that it is. (It seems likely the number they give must have been updated for reasons which will become clear below.)

If you ask them why 220 is the limit, they say the Samaritani Formula says so.

$$\log_{(17/18)} 1/300000 = 220$$

The first time I and others came across this, when we said it wasn't clear why this should make any sense at all, it was explained to us that 300000 was 50x6000. Which indeed it is. So the clarified version of the formula was:

$$\log_{(17/18)} 1/(50*6000)=220$$

In case this isn't immediately convincing either, now may be the time to explain where all the numbers in this come from.

The lottery we're talking about here is the main one, “Lotto”. It has 11 lottery machines, called “ruote” (wheels), each of which has 90 numbers. At each draw, 5 numbers are picked from each of the 11 wheels. So “Lotto” is 11 parallel “90 pick 5” lotteries. You can bet on various things, right down to the level of betting on a single number from one of the wheels turning up. This is apparently a popular bet, and also one of the least worst deals. It pays about 10 to 1, and of course has a 1 in 18 chance of working. Note that all the bets are fixed odds: there is no jackpot type mechanism. (So “Avoid common combinations if you have to play at all” does not apply here.)

In the formula above, 17/18 is the probability of a given number failing to appear, 6000 is the number of draws that had taken place in the history of Lotto at some point in the past, and 50 is the number of numbers drawn in each draw. Of course, you will have noticed that 50 is not 55. The reason for this is that in the late 1800s there were 6 wheels, then 7 later, 8 when Samaritani was alive, 10 some time later (including the late 1990s), and 11 now. You may see the Samaritani formula quoted as $R_{max}=\log_{(17/18)} 1/d$ where R_{max} is the maximum run of absences being calculated and d is the number of numbers that have ever been drawn. In simulations I use a fixed number of wheels and at least for now will live with fixed numbers of wheels in attempts to model Lotto as well.

Of course when we understood they were asking about the maximum possible run of absences in 6000 draws not in general, we changed our answer from infinity to 6000. In case 6000 draws seems like a large value to use in examples, bear in mind that Lotto has been around for well over a century. The frequency of draws has increased over the years, and is currently at three a week.

Some Italian lottery players (and the media in Italy) refer to the number of draws a number has been absent as its “delay” (“ritardo” in Italian) and people who claim to obtain insight from studying these delays are referred to (perhaps not by themselves) as “ritardisti” (singular: ritardista). I will use the term “delay” in this article because it's shorter than “block of absences”. I have occasionally used the term “delay theorist” when talking about ritardisti in English but I'll use “ritardisti” in this article just in case there actually are people called “delay theorists” who are not idiots. Or conpeople.

Some Samaritani fans claim that Samaritani claimed that the maximum possible block of absences of a number in the Italian lottery was the logarithm (to base 17/18) of one over the number of numbers that had been drawn. Clearly this is a number that increases slowly with time. Equally clearly, if Samaritani said any such thing he was some kind of idiot. Or con artist. (Some other Samaritani fans are less clear about the meaning of the number that comes out of the formula, and may say that it is rarely breached, and only by a little, or that it is “extra information”, without saying what use it might be.)

They then say that clearly you can guarantee winning a bet on a single number on the Italian lottery by

waiting for one to reach a delay of (say) 160, then making progressively larger bets on it every draw until it appears. (Over almost 20 years of this, the 160 and 60 version has been very consistent. They are all getting this from somewhere, I have to assume.) Since single number bets pay about 10 to 1, you only need to increase your bet by about 11% (actually 10.46 but I'm rounding up) each time you fail. And if 220 is a real limit you only need to be able to do this 60 times in a row because, at least some ritardisti say, a delay of 221 is impossible. (Maybe they have amended this to a larger value in the last 20 years). Of course, you need to have several thousand euros available to be able to "follow" a number for 60 draws like this and only stand to win 10 euros if you start with bets of 1 euro. If you have more money, you can start earlier. Of course, in real life you risk losing everything. At least one ritardista who claims "theoretical maximum" doesn't mean what we think it does has still used the 160/220 example, which suggests to me that he thinks or wants victims to think 220 has magic powers.

There is some fun to be had with ritardisti in a "mockery through participation" sort of way by taking a hard limit of 220 at face value. e.g. if multiple numbers reach a delay of 220 at the same time, they all have to appear in the next draw. But what if more than 5 numbers on the same wheel do? Clearly they can't. This would force some of them to come out the draw before. One could clearly then imagine a situation where the same 5 numbers came out 204 times in a row and among the others we have 5 with a delay of 220, 5 with a delay of 219, and so on. A hard limit Samaritani fan would have to say that this would force the next 17 draws on that wheel completely to avoid violating the hard limit of 220. You could then bet on combinations of 5 with guaranteed success 17 times in a row. This of course sounds absurd, yet we're just taking what (at least some) ritardisti say and running with it. For some unaccountable reason this sometimes causes offence.

Actually, at least some ritardisti might respond to the above mockery by saying that a number can't appear 204 times in a row. Applying the Samaritani formula to runs of successes instead of failures will "of course" tell us some sort of magic limit on the number of successive appearances of a number.

When particular numbers reach very high "delays" one sees stories in the papers about people going bankrupt and/or killing themselves. It's not clear that I can blame ritardisti or the Samaritani Formula directly for this but they are part of a whole environment which gives the impression that ritardi are of some significance. If you google "53 Venezia" you should find some articles about events in 2005 when number 53 in Venice had not been drawn for a very long time.

Some examples from mainstream sources, in English, would be:
http://www.theguardian.com/world/2004/dec/04/italy.johnhooper
"Italy's unlucky number"

http://www.theguardian.com/world/2005/feb/11/italy.sophiearie
"No 53 puts Italy out of its lottery agony"

http://news.bbc.co.uk/1/hi/world/europe/4256595.stm
"Number 53 brings relief to Italy"

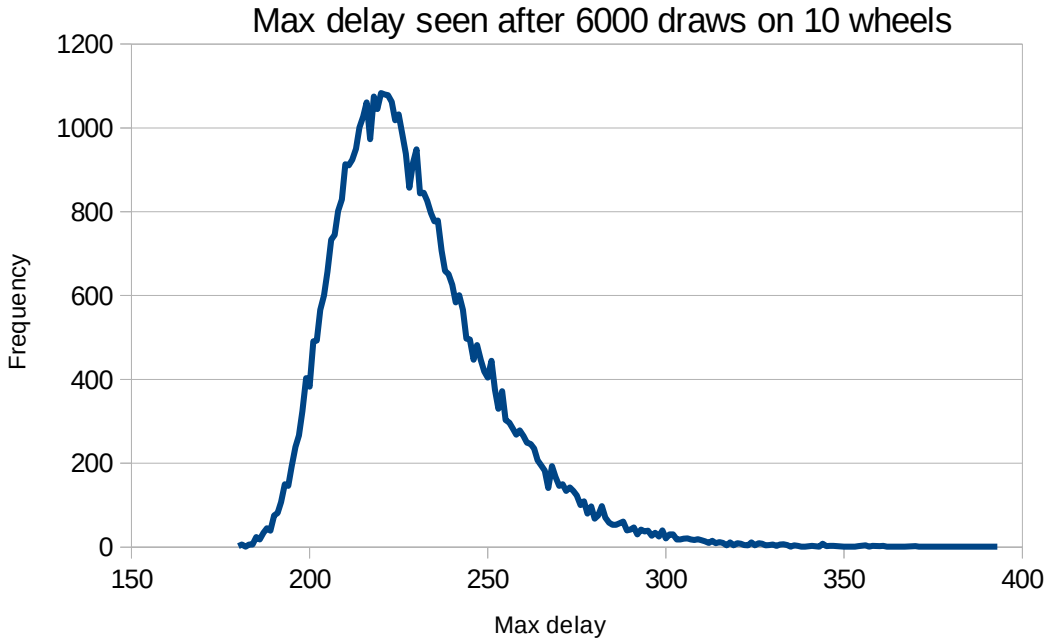
Note that number 53 in Venice in 2005 did not reach even a delay of 220, let alone whatever the magic Samaritani number would have been by then. One does have to wonder what will happen if 220, or the Samaritani number at the time, is ever exceeded. (See 2017 extras at the end: this has now happened.) My suspicion is that the Samaritani formula is known only within a fairly small group of lottery players

and the Venice 53 victims were just following the usual “It's bound to appear soon” type reasoning.

Also look on youtube for videos of “Il lotto alle otto”, a TV program about Lotto from many years ago now. I’ve seen one episode where the presenter said “This number hasn’t appeared for (number) draws. It’s ripe.” or similar, and people wore tabards with the most delayed numbers on them. No doubt if asked, the TV people would have said there was no significance to this.

One very strange event during the 2005 Venice incident was Codacons, an Italian consumer rights group, asking that number 53 be removed from the Venice draw to solve the problem of people ruining themselves. If there’s a misunderstood genius anywhere in this whole story, clearly it’s Codacons. Maybe in the long run there would be fewer and fewer numbers in Lotto? Or just replace Venice 53 with Venice 91, etc.? Also amusing.

Before we try anything else, let's simulate 6000 draws of a 10-wheel Italian lottery and see what maximum delay we get. Actually, let's do this a few thousand times so we see what the common values are and what kind of spread we see. Clearly anything from 17 to 6000 is possible though 17 and 6000 themselves will be very very unlikely.



As it happens, in the above the most common value is 220, though there may be an element of luck here. As the simulation ran the mode changed quite often between various values from 220 to 226.
The mean is a little over 229, the median is 226 and the standard deviation is about 22.2.

220 really doesn't seem to be a particularly high value for the maximum delay so if ritardisti were trying to invent some sort of magic threshold you'd think it would need to be much higher than this. Have they really never checked how commonly it is exceeded in simulations? But it is the most common value, or at any rate one of the most common.

In case it is not clear, at the beginning of a simulation, the delays of all numbers are set to 0. Then after each draw, the delays of the 5 numbers per wheel which are drawn are set to 0 and the delays of the other 85 numbers in each wheel increase by 1. (If new wheels are added other than at the start, their numbers also all have initial delays of 0 so in the first draw of a new wheel, we will see 5 “0”s and 85 “1”s.)

I contributed to some articles about this in the Italian computing magazine “MC-Microcomputer” in 1999. The other people involved were Elio Fabri and Francesco Romani of Pisa University, Dani Ferrari (a retired engineer in Rome) and Corrado Giustozzi, the author of Intelligiochi, the recreational maths column in MC-Microcomputer for 15 years (and thus Italy's answer to Martin Gardner). Giustozzi’s column hosted articles but he was not otherwise involved as far as I recall. Francesco Romani wrote the Mathematica column in the same magazine and dedicated at least one episode of it to the Samaritani formula and lottery discussions. Back issues of the magazine are now online. See e.g. https://issuu.com/adpware/docs/mc191 for the January 1999 edition. Some of our discussions took place on a mailing list whose contents are most likely now lost, and on the “matenigmici” forum on the BBC/ISP MC-Link in Rome run by the same company as the magazine. The contents of this and other forums are also presumably now lost. Italian Wikipedia has a list of topics covered in Intelligiochi over the years at https://it.wikipedia.org/wiki/Intelligiochi

At least one web site about delays switched from saying “ritardo massimo teorico” to “ritardo massimo modale” at some point after our articles appeared. That is pretty funny.

When pressed, the ritardisti I have encountered mostly appear to say they believe that lottery draws are independent and that all combinations of numbers are equally likely. Clearly one could imagine lotteries where all combinations were not equally likely or where the draws were not independent: when I first encountered ritardisti and the Samaritani Formula, Italian lottery draws were done by hand by blindfolded orphans, or so I am told. The balls were actually inserted into the draw apparatus in a fixed order, and mixed a little before the first number was drawn and a little more before each later number. Some people claimed that there was some evidence the distribution of the first number drawn was non-uniform. It's hard to believe this was enough to make bets on single numbers a good deal, but an exaggerated version of this could serve as an example of a significantly non-uniform lottery. Of course, with independent draws but different probabilities for the numbers, ritardisti are even wronger than usual since the rarer numbers will tend to reach higher “delays” so playing “delayed” numbers would be harmful to one's winning chances.

We could also turn this into a plausible example of a lottery where there was dependence between draws. Let's imagine that the order in which balls are put into the machine is not fixed, but depends on their delays. Perhaps we could make the most delayed numbers more likely to be drawn (at least as the first ball in a draw) or more entertainingly make the most delayed numbers less likely to be drawn. Even in this case, it wouldn't matter what the exact value of e.g. the highest delay was. It could even be that some dependence was present in the orphan-based version of Lotto: possibly if the orphans knew they were drawing from a wheel with a very “delayed” number they might behave somewhat differently. It seems unlikely this effect would be very large. Perhaps someone reading this has access to a large number of blindfolded orphans and would like to find out.

One could imagine a lottery where single numbers all had the same probabilities of appearing, but all combinations of five did not. For example, maybe 1-5, 6-10, etc. are the only combinations that can

appear. Of course I/we are explicitly assuming that draws are independent and that all combinations of 5 numbers are equally likely. We say that if ritardisti are also saying this then they need to accept the consequences. If they explicitly say that lottery draws can NOT be modelled this way that is a different problem. (Examination of actual draws shows no obvious sign that higher delay numbers are more likely to come out, for example. This will not surprise you.) I think some ritardisti are trying to say that Samaritani proved that independence has limits, or that independent events aren't. It's quite hard to work out what they are actually claiming, a lot of the time. "We don't know what independence is, or are hoping that you don't" is mostly what I take them to be saying.

Of course, ritardisti don't just say things which aren't true. They say:
(i) Things which are false.

e.g There is some advantage to playing a "delayed" number.

(ii) Things which are true but irrelevant.

An example of this is if they say that a number with delay 160 is very likely to appear within 60 draws. It is, but so is any other number.

(iii) Things to which no real meaning can be assigned.

Mathematical formulae which make no sense at all might fall into this category.

Clearly, someone who claims that draws are independent and combinations are all equally likely BUT that their study of delays allows them to predict results better than chance is either a conperson or an idiot. In any particular case it's hard to say. It's also possible that some ritardisti, like characters in Raymond Smullyan puzzle books, *believe* that they believe that draws are independent, but actually believe that draws are not independent.

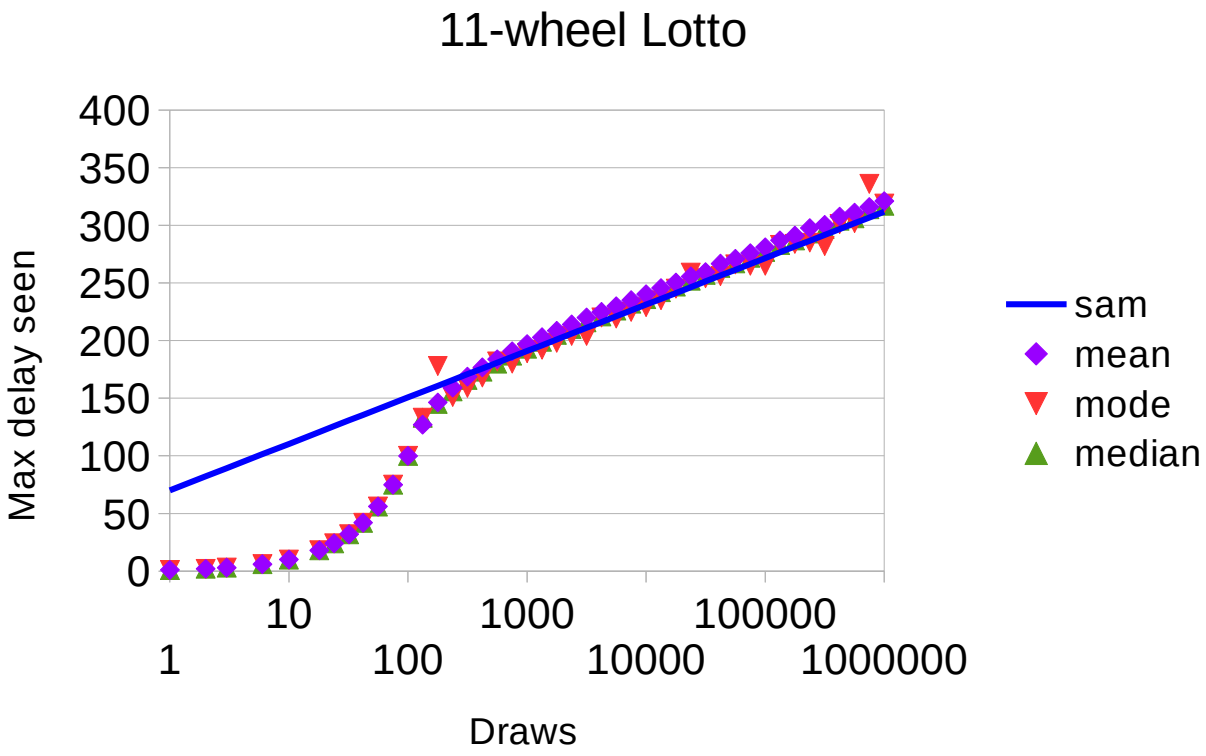
It's obvious enough that the 220 that comes out of the Samaritani Formula is not any kind of a limit on how large delays can become. At the time of the discussions during my first encounter with the Formula in the late 90s, one of the ritardisti seemed to feel that it was significant that the maximum delay ever recorded was 202, which was less than 220. Of course this did not strike most of us as very significant. Lots of numbers are greater than 202 and just because the S.F. produced one that did not seem very interesting. The 202 record was beaten in 2006, becoming 203. See the end of the article for events from late 2016 and early 2017.

As we have seen, if you simulate 6000 draws of the 10-wheel version of the Italian lottery you find that the maximum delay falls both sides of 220, but that 220 is in the fairly narrow range of values it takes fairly frequently. This of course could be a coincidence, but if you simulate other numbers of draws you find that the number produced by the Samaritani Formula is always one of the relatively plausible values, at least for numbers of draws over 1000 or so. This continues to work if you change the number of "wheels" in the lottery to 11 as it is today. This is more interesting. The SF clearly doesn't produce a maximum value delays can assume, as that would be n after n draws.

Here is a chart of Samaritani's number for an 11-wheel Italian lottery after various numbers of draws from 1 to a million, along with the mean, mode and median of the maximum delays seen in many

simulations made of each number of draws. Fewer runs were made of the longer simulations so the mode in particular is less reliable.

As we can see, the mean, mode and median are very close to each other, and the Samaritani number



agrees with all of them for numbers of draws over a few hundred. For such numbers of draws the distribution of maximum delays has the usual Brontosarus type shape one often sees: thin at one end, much much thicker in the middle, and thin again at the far end. The red triangle above the Samaritani graph for n=178 is not an artefact: see much later for note about second peak for small *n*.

More interesting (and not shown on this graph) is that the standard deviation of the maximum delay seems to vary very little with the number of draws (once the simulated and Samaritani values have started agreeing). This was a surprise at least to me.

Dani Ferrari, one of the group in the original discussions, went to the Biblioteca Nazionale (National Library) in Rome and found a copy of Samaritani's 1937 book “La teoria e il calcolo matematico dei ritardi. Studio teorico e pratico sul giuoco del Lotto”. (Theory and practice of delays. Theoretical and practical study of the game of Lotto.) “Giuoco” instead of “gioco” is rather old-fashioned, but this *is* a book from the 1930s. He looked through it and said that as far as he could see, Samaritani was neither a conman nor an idiot, and that his formula was intended to estimate how long the maximum delay seen in *n* draws of the lottery would be, not to give an upper bound on it. One of these decades I might try to buy a copy of this book but it seems to be hard to find.

How can we derive the Samaritani Formula?

At this point it seems like rather than being just nonsense, the S.F. could actually be the answer to a different question. We are now playing mathematical Jeopardy. What question is the S.F. the answer to? Preferably a non-insane question about the Italian Lotto. It doesn't need to be anything that helps choose Lotto numbers. Of course it's obvious it can't do that.

One of Francesco Romani's articles is here: http://www.digitanto.it/mc-online/PDF/Articoli/191_166_169_0.pdf

Considering the destiny of just one number in the Italian lottery, we can treat it as n throws of a biased coin, where a Head, probability 1/18, is that number being drawn, and a Tail, probability 17/18, is that number not being drawn. We then ask ourselves what can be said about the distribution of the length of the longest block of tails. And we regard two successive heads as having a 0-length block of tails between them, which is perhaps not entirely usual.

(I only went looking to see how "real people" approached this problem some considerable time after we first ran into the S.F.: It is apparently known that the longest run of tails in n tosses of a biased coin with p(Head)=*p*, p(Tail)=*q* is likely to be about $\log_q(1/np)$ if *np*>>1. See, for example, http://www.johndcook.com/blog/2012/11/14/probability-of-long-runs/ and http://www.csun.edu/~hcmth031/tspolr.pdf (referred to in the previous article). I have rewritten the formula in the second article to look as much like the Samaritani formula as possible.)

np is the number of times this number is expected to be drawn. Each draw is preceded by a run of failures (possibly of length 0). For the sake of coming up with an estimate, let's say the number comes up exactly *np* times and see where that gets us.

Two ways to get the known result seem to be (left as exercises for the interested reader):

1) Write down the probability that an observation from our 0-based geometric distribution will be <=*k*. Raise this to the power *np* to get the probability that *np* independent observations from a geometric distribution will all be <=*k*. (They're not really independent since e.g. there's a constraint on what their sum can be. And there won't necessarily be exactly *np* of them of course.) Take the derivative of this with respect to *k* to get the density (pretending that *k* is continuous rather than discrete). Do it again and set to 0 to find the mode. (This at least feels like we're trying to calculate the mode of something.) I get the S.F. with a -1 on the end if I do this though I may have made a mistake in my algebra. Given that I round to the nearest integer and I imagine Samaritanists round down, this seems close enough. When using a computer I of course work the discrete values out.

or

2) Consider a list of the lengths of the block of failures before each success. 1/18 of the numbers will be 0, 17/18^2 will be 1, and so on. Ask yourself what length of block of failures is such that on average you expect one number in your list to be at least that big. Of course, the run of failures at the end of our sequences of tosses is invisible to this method. This gets the Samaritani formula exactly. (Some ritardisti seem to think the S number is such that an average of one number will come out with *exactly* that delay. That's not correct.) It's not obvious to me that

what are are doing here is trying to calculate the mode of anything. But if we're after an order of magnitude of any/all measures of central location, that's also fine of course.

Should I/we have recognised the S.F. as being this known result about the maximum run length when tossing a biased coin? Possibly. I didn't recall having seen it before and the (moderate numbers of) people I've told about the Samaritani formula over the last 20 years have so far never said "Oh, that's someone failing to understand (result)!". Perhaps I've just never told it to probabilists. In retrospect I should have realised that it was, or was related to, a known trick where you ask students to either toss a coin 100 times or pretend to have done so, and then tell them whether they cheated or not.

How do we extend this to an 11-wheel Italian lottery? Clearly the destinies of numbers in the same wheel are not really independent of one another. To take an extreme case, if you know that five of the numbers in a wheel have appeared, you know that none of the other 85 in that wheel can appear in this draw. However, as an approximation, we could treat the current Italian lottery as 990 independent coins being thrown in parallel. This would limit the maximum possible delay after n throws to n in all cases. Or to make it easier to apply the formula in these articles, treat the Italian lottery as a single coin being thrown $990n$ times. The problem here is that this would allow delays longer than n, and would also allow delays spanning multiple blocks thus counting a longer delay than would appear in any single block. However, for n large enough neither of these problems really makes much difference.

At this point treating the Italian lottery as $990n$ tosses of a biased coin using method (1), the Samaritani formula drops out immediately!

$$\log_{(17/18)}(18/990n) = \log_{(17/18)}(1/55n)$$

Was the result about longest runs known before 1937? Indeed it was.
https://math.stackexchange.com/questions/59738/probability-for-the-length-of-the-longest-run-in-n-bernoulli-trials refers to "*A History of Probability and Statistics and Their Applications before 1750*" by Hald, with solutions by de Moivre (1738), Simpson (1740), Laplace (1812), and Todhunter (1865)." Of course this is for a single coin. So Samaritani seems not to have been the first,
though quite possibly he did not know this at the time.

With method (2) above we know we have exactly $55n$ successes in our $990n$ trials so the Samaritani formula drops out again. Of course the argument about 1/18 of the numbers in our list being 0 etc. is unconvincing for small n as the first 55 numbers are all 0, the next 55 are all 0 or 1, etc. And as mentioned delays after a number's last appearance (or of a number which never appears) are invisible in this method.

However, let's call the distribution of the largest of $55n$ (or whatever) observations from our 0-based geometric distribution "The Samaritani Distribution" (with parameter $55n$ or whatever).

Improving on Samaritani

The approach Francesco Romani writes about in his January 1999 article is to use a recursive calculation to get the exact distribution of the maximum delay for a single number, then take the 990^{th} power of its cumulative distribution function to get the approximate c.d.f. for the maximum delay of the entire Italian lottery. Then de-cumulativize that to see what the most probable maximum delay is. Of course as

mentioned before the 990 numbers are not really independent. And the calculations need to be done using arbitrary precision integers (using Maple or Mathematica or similar) as otherwise horrible things can happen. However, this approach is probably better than Samaritani for small number of draws, albeit impractical in 1937 or without a computer today.

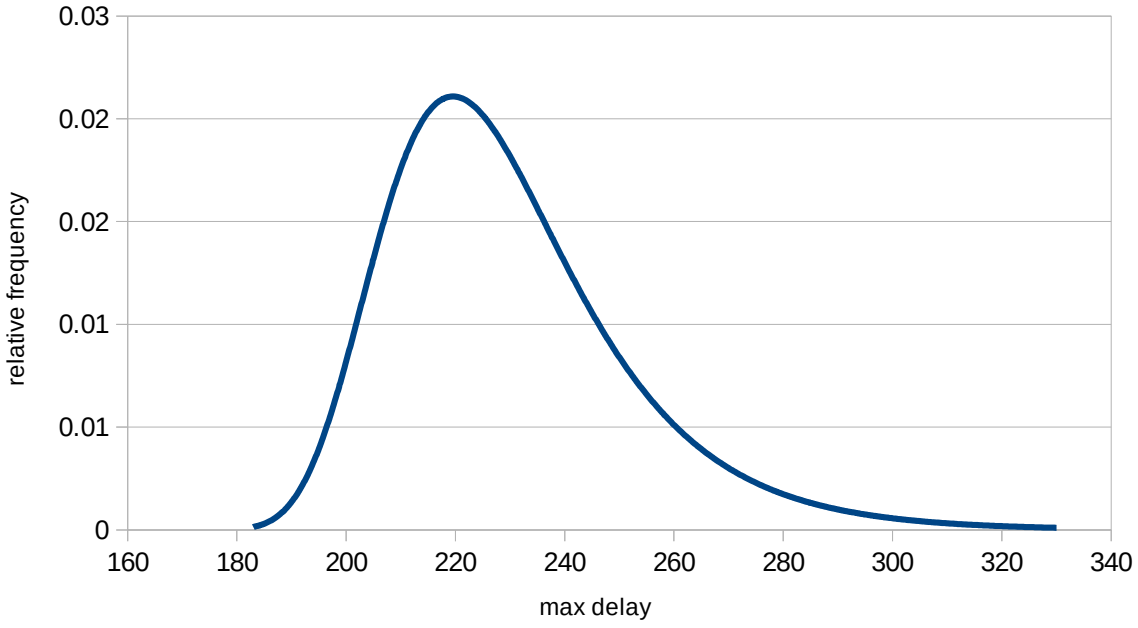
Let us consider a single number in the Italian lottery, (equivalently, a biased coin). Let p=1/18, q=17/18.

Let us define $g(n,r)$ = probability that after n draws a delay of length r has been seen.

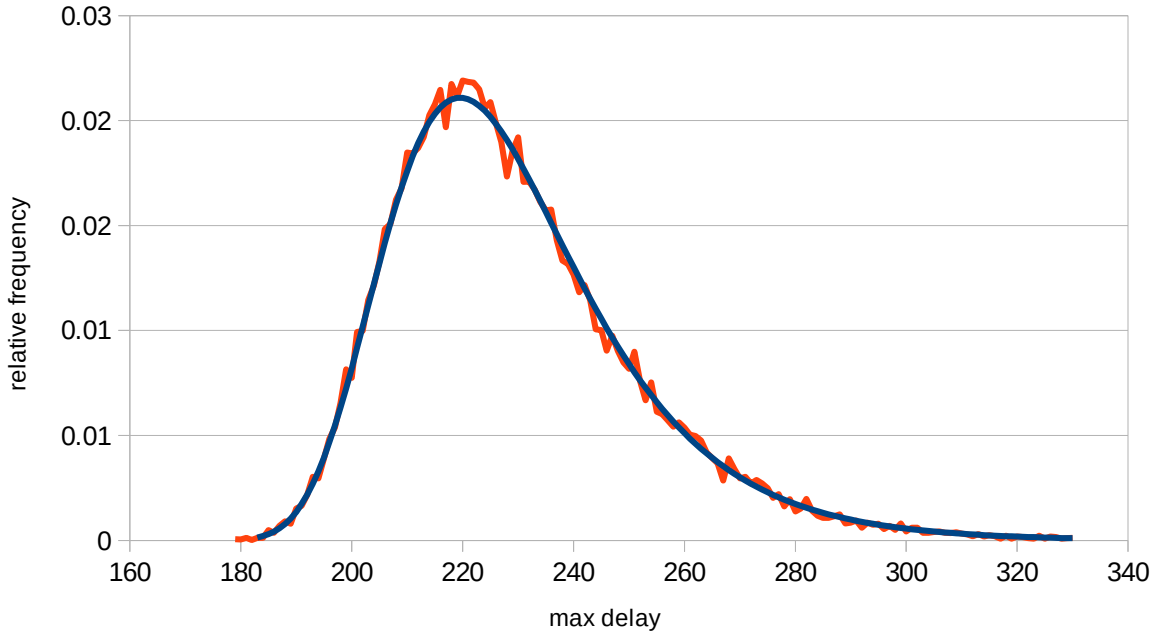
Then $g(n,r)=0$ if $n<r$.
$g(r,r)=q^r$
$g(n,r)=g(n-1,r)+(1-g(n-r-1))pq^r$ for larger n, since for there to have been a delay of r by draw n, it had either been seen at draw n-1 or was new at draw n, in which case there had been no delay of r in the first n-1-r draws, then our number appeared, then it failed to for r draws in a row.

Then the probability that after n draws the maximum delay seen is r is $g(n,r)$-$g(n,r+1)$. As seen earlier we then take the cumulative version of this (1-$g(n,r+1)$: if no delay of r+1 has been seen then the maximum is less than or equal to r) to get the probability the max delay is <=r, and raise this to the 990th power to get the probability that the maximum of 990 independent numbers is <=r,. (We know they aren't really independent but let's try it as an approximation). Then the difference between these values for r-1 and r tells us the probability that the maximum delay is exactly r.(The maximum delay is <=r but it's not <=r-1)

Here is a graph of what Romani's technique gives for the distribution of the longest delay in 6000 draws of a 10-wheel lottery: (the mode is 220). Since this is a 10 wheel lottery I have only raised things to the 900th power, not the 990th, on this occasion.

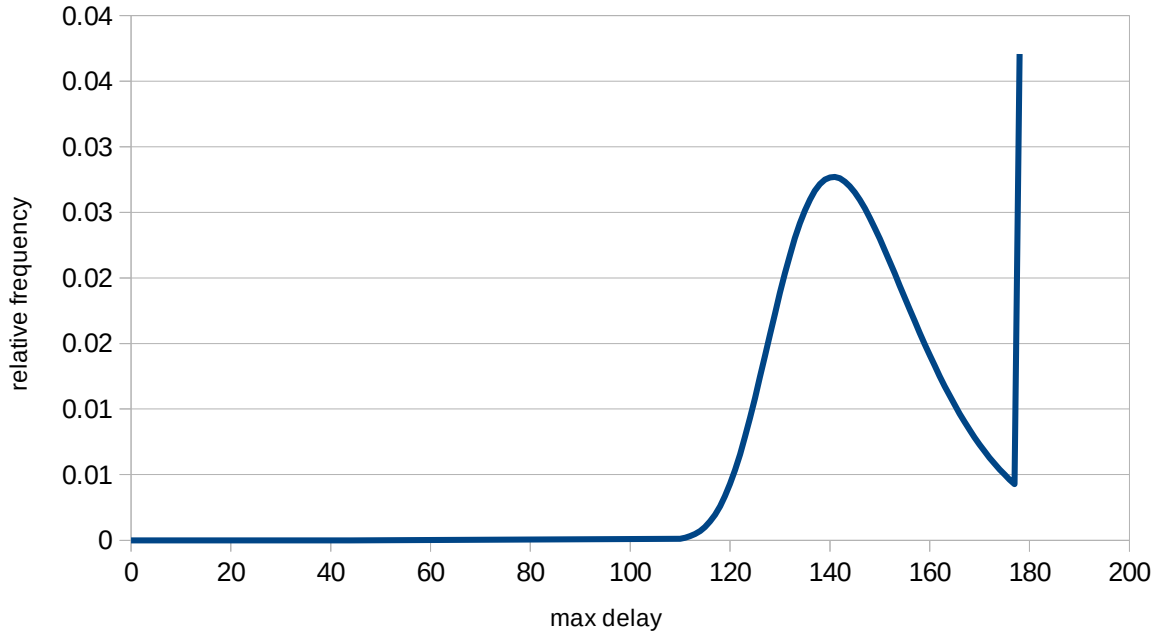


And here it is compared with the results of the simulations from earlier:

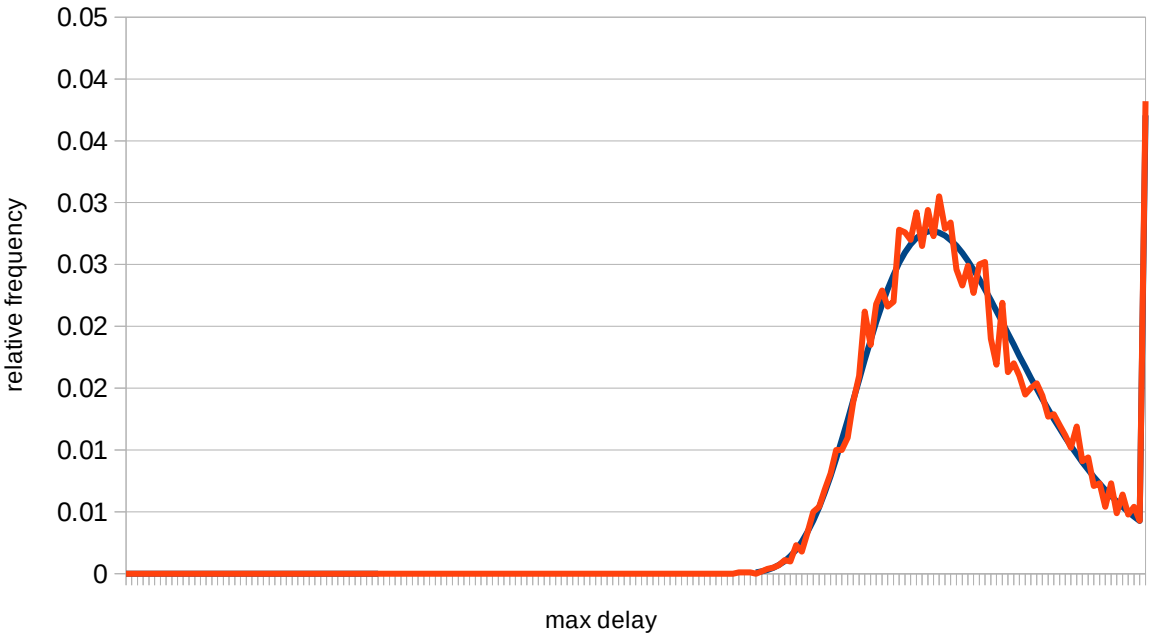


I'd say for the purposes we're using it for, this is fine.

Something amusing that happens for fairly small numbers of draws is that the distribution is significantly bimodal. Here is 178 draws of an 11-wheel lottery using our approximation (the red triangle above the graph earlier):

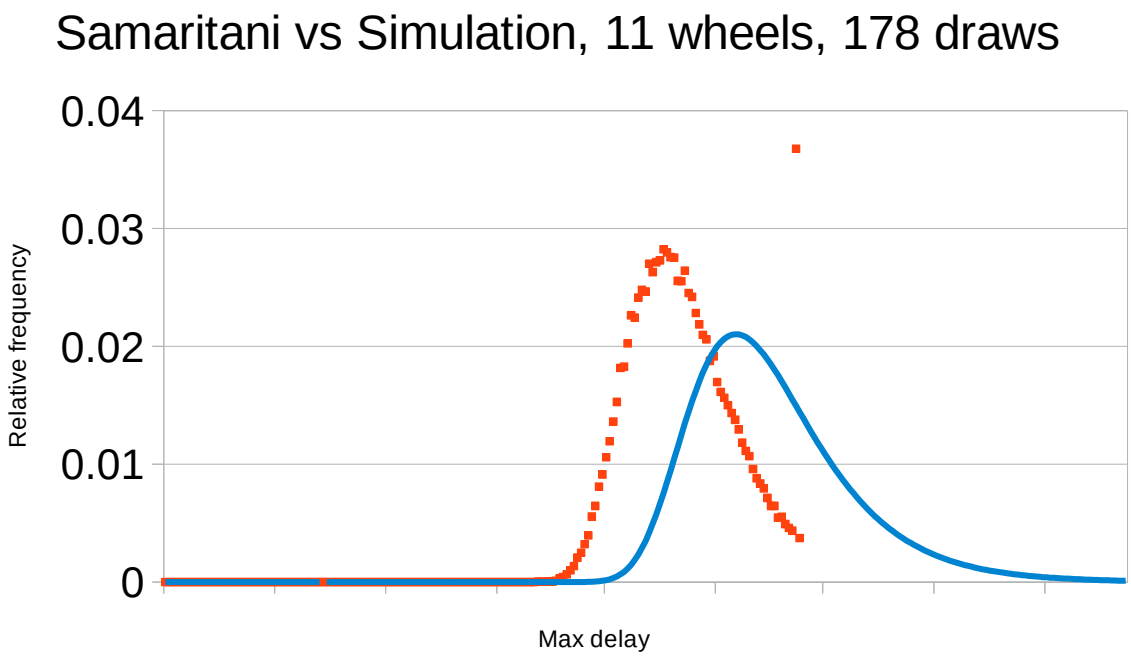


And here it is with the results of 10000 simulated runs superimposed:



Very good, again. Of course, like the original, our improvement is no use for playing the lottery.

For comparison, here is what the distribution of the maximum of 178 observations from a 0-based geometric distribution looks like:



As *n* increases, the mean and median presumably change fairly smoothly but the mode will suddenly jump down as *n* goes beyond some threshold value.

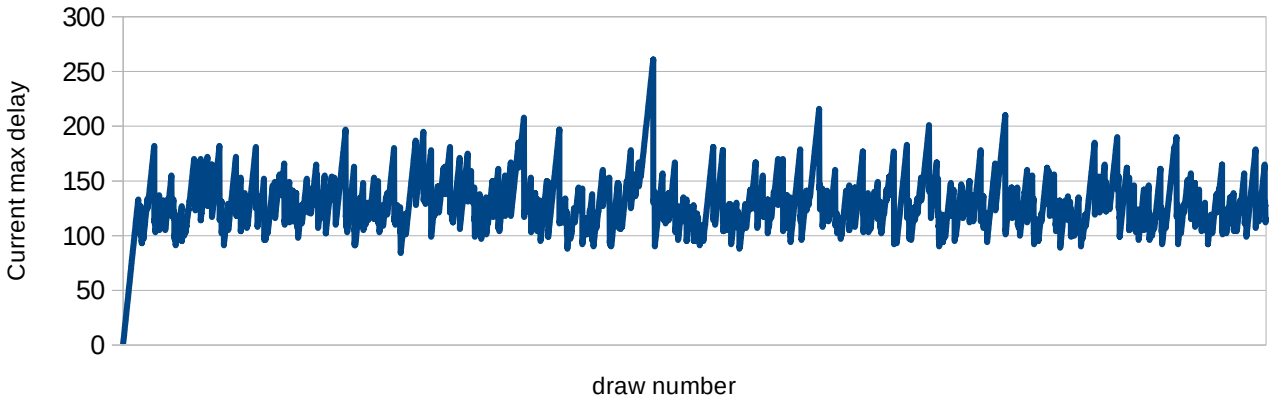
For *n*=100, the peak at 100 is pretty well all you can see.

This effect is present, less dramatically, for higher numbers of draws too. This feature is one that simpler methods would not pick up: if you're treating the whole thing as 990*n* coin tosses or as 55*n* observations of a geometric distribution, you will allow delays over *n*.

Centenari

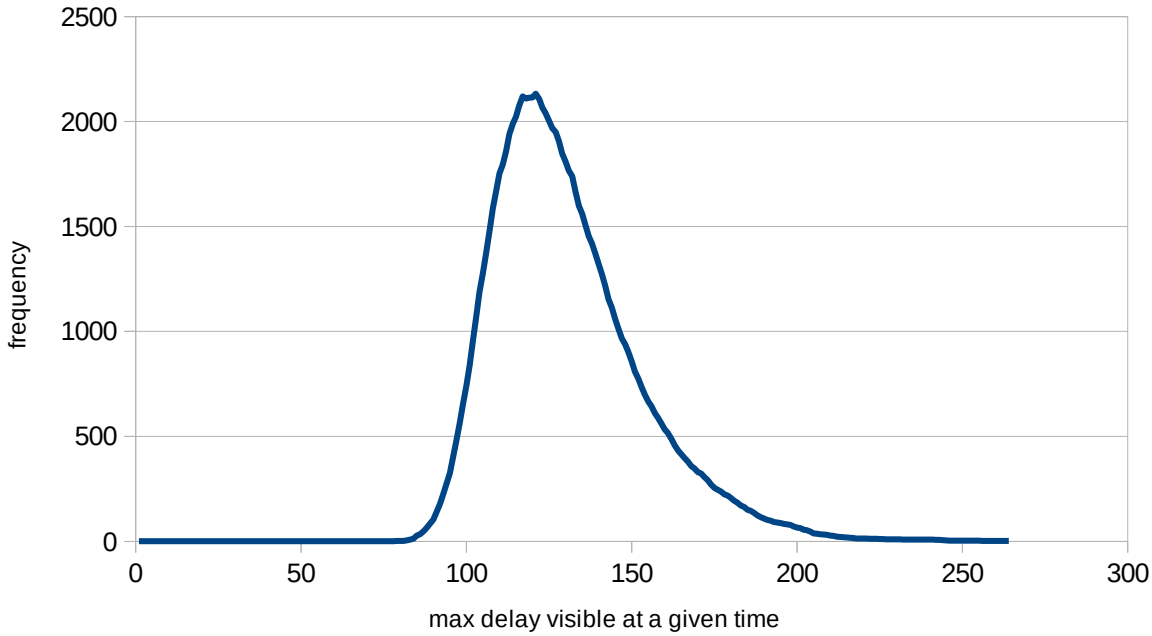
Some people act as though there's something special about numbers with delays over 100 and call them "centenari" (centenarians). Are they even rare enough to be notable?

Let's graph "max current delay" for 10000 draws and see what we get.



Conclusion: "centenari" aren't rare at all. Indeed, they're pretty usual. And beyond the initial ramp up it seems pretty stable.

Running 1 million draws and graphing the frequencies of the max current delays seen I get this:



which also suggests that having at least one number with a delay over 100 is pretty normal. No doubt they were less so back when there were only 8 (or fewer) wheels, but still equally meaningless.

One reasonable question is how the S.F. can be a con if it is given away for free. Perhaps this can be taken to suggest that ritardisti are more likely to be idiots than conpeople, but it's also possible the "basic" S.F. is a freebie intended to help sell software to analyse delays, magazines about lottery advice, or similar. It also seems that there may be more "advanced" kinds of delays than just those of single numbers. I have for example seen people talk about delays of pairs of numbers and Samaritani type results for those. Of course, if ritardisti talk about "level delays", "compensated delays" or "delays with garlic, oil and chili" (some or all of these names are made up), it's all obvious nonsense if draws are independent.

The conclusion then is that the Samaritani formula is not nonsense, but has been misinterpreted, possibly wilfully, by modern ritardisti who use an approximate formula obtained by considering independent events to claim that those same events are not in fact independent. Until I have read the book for myself I cannot be sure that Samaritani was not a ritardista, but I am happy to imagine that he was not: Dani Ferrari is pretty solid.

It continues to be the case that if someone talks about "the Samaritani Formula" there is a pretty good chance they are a ritardista (except me in this article, clearly) so if you do want to use it for anything, it might be best to call it something else to avoid being mistaken for one. The same "independent coins being tossed" approach works fairly well for, e.g., dice or random decimal numbers, where you treat these as single-wheel lotteries where 1 number is drawn from 6 or 10, respectively.

Also, while the Samaritani formula is no use for playing the lottery normally, if someone says they are about to launch an Italian-style lottery and offers you the chance to bet on what the longest run of absences will have been 100 years from now, maybe it could be useful for that.

2017 Update

In late 2016 the nice people at fateilnostrogioco.it wrote to ask if I'd heard that a number (Number 53 on the National wheel) had reached a delay of 233. This was not only more than the previous record, and more than the magic number of 220 from the late 90s, it was even more than the updated magic number for late 2016. So I kept an eye on Italian news and lottery web sites for a while. This number eventually reached a delay of 257 before appearing. At least some ritardisti are going to need a new story now. It might be possible to get some idea of the idiot/conperson balance among ritardisti by seeing how many of them bankrupted or killed themselves in late 2016 and 2017 but I have found no useful information about this.

Something that happened in 2009 provides us with an opportunity to apply the Samaritani Formula and our improvement to another scenario. Apparently in 2009 the Samaritani limit for a pair of numbers was violated. It's possible that this is only true if you fail to update the 6000-draw, 10-wheel version for things as they stood in 2009 but it was mentioned on Lotto forums so there are people who track this kind of thing. But don't notice that they're using an outdated nonsense magic number for comparison.

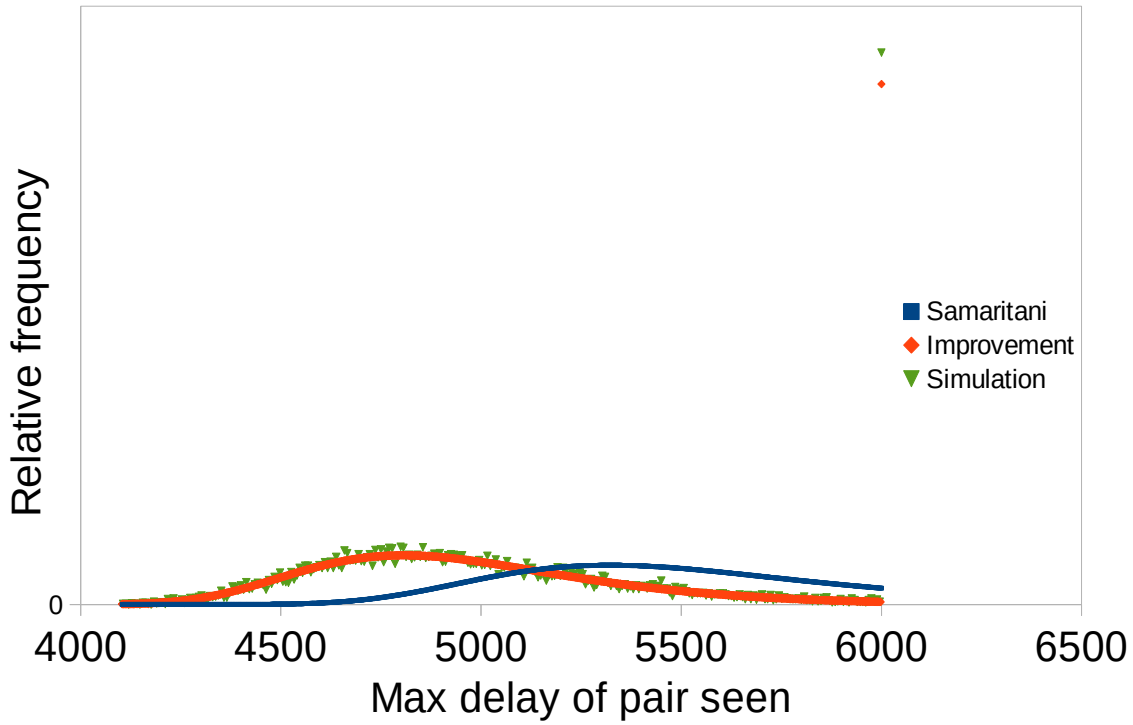
How does the Samaritani model apply to pairs of numbers? In actual Lotto the 5 numbers drawn on each

wheel form 10 pairs of numbers. And there are 4005 possible pairs of numbers that can be formed from the full 90. Our p is now 10/4005 instead of 1/18 and our q is now 3995/4005 instead of 17/18. If we have 10 wheels and 6000 draws as in the late 90s, then we have drawn 10*10*6000 pairs of numbers. And we pretend they are all independent though clearly they are not. So in this case we need to calculate

$$\log_{(3995/4005)} 1/600000 = 5321.871202$$

I guess ritardisti would round this down to 5321? Look for both 5321 and 5322 when googling. See for example http://forum.lottoced.com/forum/lottoced/statistica/58937-il-memorial-day-del-gioco-del-lotto

In any event, let us compare simulations, Samaritani and our improvement:



We see a spike at 6000 in the simulation and in our improvement. This seems a lot like the n=178 example for the base case: n is too small for the approximation to work. Our improvement clearly beats Samaritani. While the 178th draw was long enough ago for the n=178 graph to be perhaps mean, that is clearly not the case here. One has to wonder if as well as being idiots or conmen, ritardisti are also lazy. Perhaps they just imagine that their potential victims are and won't check something like this.

I have not worked out what the late 2016 replacement for 5321 or 5322 should be. Clearly it will be quite a lot bigger. It is curious that at least some lottery people who want to use the Samaritani Formula for some purpose continue to tell each other about values it produced with parameters from 20 years ago. The pair of numbers which reached a delay of 5322 in 2009 finally came out in 2016. It was on the Bari wheel, added in 1874. Since then two wheels were added in 1939 and one more in 2005. When Samaritani's book was published, then, there were only 8 wheels.

If someone wants to run a clone of Lotto and offers you a chance to bet on what the maximum delay for a pair of numbers will have been in 100 years, our improvement or a simulation is definitely a better idea than the Samaritani Formula.

I've even seen people applying the S.F. to groups of 3 or more numbers. Why is not clear. There are 117480 groups of 3 numbers you can form from 90. Since the 5 numbers drawn from a single wheel cover 10 of these, even with incredible luck we can't see all triples in a single wheel until 11748 draws have taken place. And that's if it's possible to avoid duplicates: haven't checked.

There is an archive of draws from 1939 until the present at
Archive of results https://www.lottomaticaitalia.it/STORICO_ESTRAZIONI_LOTTO/storico.zip
Of course, since Lotto started before 1939, the initial delays are unknown so it might be prudent to only consider each number after it has appeared at least once. This does not apply to wheels added after 1939: one assumes the initial delays of all numbers on a new wheel are 0.

Succcess runs

An exercise for the reader: use Samaritani-type reasoning to estimate the length of the longest run of appearances of a single number in, say, 6000 draws of a 10-wheel Lotto. (This is of course not a magic limit on what it can be.)

Improving the improvement

Clearly original Samaritani copes with varying numbers of wheels perfectly fine since it only uses how many numbers have been drawn. It would claim that a huge number of wheels and a single draw was the same situation as one wheel and a large number of draws.

To adapt our improvement to cope with varying numbers of wheels we just multiply the results for single wheels, each with its own number of draws. We could have n throws of 6*90 coins, n-a of another 90, n-b of another 90 and n-c of yet another 90 etc. to account for wheels 7 to 11 being added partway through the history of the game.

I have made no attempt to actually do this to see if what difference it makes modelling the exact history of Lotto itself. Mostly because I don't know the exact numbers of draws that have taken place on each wheel. I could approximate it by using whole numbers of years, potentially.

Neither Samaritani nor our improvement would cope with what the UK lottery did when it changed from 49 numbers to 59.

Was Samaritani an idiot or conperson?

I am still not sure about this. It seems that he must have understood better than his modern followers do how his formula was derived and what its limitations were. However, if he knew it was entirely useless for any normal Lotto player (Betting on max delay in a new lottery 100 years from now is not a normal option) what was the point of his book? If I ever get hold of a copy, I may write a review of it. Dani Ferrari is the only member of our gang of four to have seen the book and he didn't think Samaritani seemed like an idiot or a conperson during his fairly brief time with a copy.

Magic Coins

Elio Fabri suggested way back when that one way to make money out of ritardisti would be to offer to sell them coins that had been tossed until they'd come up the same way up, say, 10 times in a row. A ritardista (or victim, possibly) informed me in it.scienza.matematica in 2017 that the S.F. puts the lie to the idea that a coin tossed 12 times can come up heads 12 times, and seems to say that the coin can't come up heads twelve times in a row unless it has been tossed about 10 thousand times. Now possibly we shouldn't take the value 10 thousand as an exact value but if by "about 10 thousand" he means at least some specific number bigger than 12 (which if 12 is impossible he must?) maybe we have a possible customer. We would need to video ourselves tossing the coin. Would we need to film ourselves getting coins directly from the mint? Does it matter if the coins have flipped whilst falling down chutes in the mint? Does a coin's memory of how many times it has been tossed and with what results fade if it is not tossed for a while? Does a change of owners have any effect? Do ritardisti imagine coins have nanotechnological gyroscopes and/or motors in them as well as memory? Or is magic involved? Back when numbers were drawn by blindfolded orphans, presumably they thought the orphans' hands were pushed towards delayed numbers by mystical forces.

I'm not sure in this particular case, but I believe I've seen a ritardista or victim using the S.F. to treat a coin as a lottery wheel with two numbers. Using Samaritani for runs of either face of a coin suggests we are treating the two faces of a coin as independent. This seems much bolder than treating lottery numbers on the same wheel as independent. I'm not sure if this particular victim or ritardista is misunderstanding results about expected times until k heads in a row, or is perhaps using an inverse Samaritani calculation to find the lowest number of tosses such that 12 heads becomes the most likely longest run, then failing to understand what that would mean.

Also-rans

Two examples of vaguely or supposely mathematical Lotto-related myths which are less interesting than the Samaritani Formula:

The first is the "Law of the Third" ("La Legge del Terzo"). People who claim this is useful observe that about one third of numbers should have delays in a particular range, and that somehow you can pick better lottery numbers based on this. This is of course nonsense but it's obvious boring nonsense based on an obvious and boring misunderstanding.

The second is called "Ciclometria" in Italian. You draw the numbers from 1 to 90 in a circle, draw regular polygons on it, and claim that this helps pick better lottery numbers. This isn't even a misunderstanding of something real. It's just absurd. Indeed it's so absurd that even on lottery forums at least some people seem to feel it sounds like nonsense.

If you want to see a non-mathematical lottery myth, google "La Smorfia" or explore the site www.lottomatica.it – at least when I last looked they described various methods people use to pick numbers. They don't say they're reasonable, but they also don't say they're not. And they do tell you what the most "delayed" numbers on each wheel are. While they don't say this is significant, many people visiting the site might imagine that it must be otherwise they wouldn't bother telling you.

History of Lotto

Information I have from https://it.wikipedia.org/wiki/Lotto#Storia

Before 1863: don't know number of wheels or draw frequency, or if beginning is well-defined. Wikipedia claims draw frequency increased to once every 2 weeks in 1807 and was previously 2 to 3 a year. http://www.ilcomplottoforum.com/t15454-la-storia-del-lotto claims there was only one wheel from 1682 to 1871. I'll go with Wikipedia and ignore everything before 1863.

(Italian unification wasn't even complete in 1863, apart from anything else.)
1863 6 wheels, 2 draws a month
1871 7 wheels, 1 draw a week
1874 8 wheels, 1 draw a week
1939 10 wheels, 1 draw a week
1997 10 wheels, 2 draws a week
2005 11 wheels, 3 draws a week

This would suggest something like:

years	years in range	wheels	Draws / week	draws	numbers drawn	cum. Draws	cum. Nos
1863-1870	8	6	0.5	208	6240	208	6240
1871-1873	3	7	1	156	5460	364	11700
1874-1938	65	8	1	3391	135640	3755	147340
1939-1996	58	10	1	3026	151300	6781	298640
1997-2004	8	10	2	834	41700	7615	340340
2005-2017	13	11	3	2034	111870	9649	452210

The above makes some attempt to account for leap years.

So 300000 numbers drawn is about right for the late 90s when we first ran into this. Assuming we're counting from 1863. Elsewhere I have said I got 480k numbers for late 2016 so clearly I've made a mistake either there or here. In any event the updated magic Samaritani number for late 2017 seems to be 227 (if we round down) or 228 (rounding to nearest integer, or allowing for a bit more time to pass). Of course I have no idea how many draws from 1862 and earlier I should be allowing for. Googling does find some people updating each other on what the magic Samaritani number has now become at various times over the years.

Blindfolded orphans stopped being used as a source of random numbers in 2009 (and started to be phased out in 2005).

So what's wrong with ritardisti?

One could suppose they are merely ignorant, but this suggestion seems to offend them. I've had one claim he's been studying probability longer than I have been alive, for example.

If someone has been studying probability for over 50 years and don't understand independence, then "idiot" seems like a fair assessment (that's about as bad as me with group theory), and it does seem to be a common opinion in Italy that ritardisti are idiots.

I suspect, however, that a significant fraction of them have to be conpeople, but of course it's hard to know how to prove this except as mentioned by tracking suicides and bankruptcies of ritardisti in late 2016 to early 2017.

It's amusing to imagine that some of the conpeople could be aiming to reduce their own tax burden by encouraging other Italians to pay lots of voluntary taxes.

It's amusing, though cruel and implausible, to imagine that some of the conpeople could be eugenicists, trying to improve the Italian gene pool by driving the more gullible elements of the Italian population to bankruptcy and/or death.

Whatever else is wrong with them, the results re pairs of numbers suggest that I should add a "laziness" axis as well.

Is it even possible that by telling people to wait until a delay of 160 appears they are actually trying to get people to play Lotto less? Perhaps in that case they should suggest waiting for a delay of 500, or just say not to play at all. But suggesting exponentially increasing bets as part of this doesn't feel like something you'd do if trying to encourage people not to play.

How well-known is it?

I think it's pretty obscure. The "Fate il Nostro Gioco" people had never heard of it and they've been running an anti-gambling campaign and roadshow for years. It is mentioned by name by some people on lottery forums, and occasional visitors to maths and science groups every few years. When exploring lottery forums during the "Nazionale 53" event in late 2016 to 2017 I found some posts where people told each other there was a formula, or it had been shown that, etc., without naming Samaritani. I have to suspect that in the general population, basically no-one has heard of it. Italian mathematicians generally don't seem to have heard of it and usually when a new ritardista or victim turns up on it.scienza.matematica only the people who were around for the previous visit know about Samaritani.

Please contact me if you've seen things like this

If you have seen something similar to this outside Italy, I'd love to hear about it. Or if you've come across it in Italy between 1937 and the late 90s I'd also like to hear about that. Has it been circulating ever since Samaritani wrote his book, or did someone uncover it decades later and start spreading it?