Service of Interdeparture Time*

Urtzi Ayesta, Jonatha Anselmi and Urko Antón
Basque Center for Applied Mathematics
Bizkaia Technology Park, Derio, 48170, Spain
ayesta@bcamath.org,anselmi@bcamath.org,uanton@bcamath.org

30^{th} Augost of 2011

1 Introduction.

Fortuna's company has organized the Behobia-San Sebastián race for years. This is a 20 km race whose popularity rises every year so the number of runners and viewers rises too. So many population is a problem for the organization because that generates some bottleneck's points where people can't pass quickly and where the organization can't attend its costumers adequately. Because of this Fortuna's company asks us for advice. The two main problems to solve are these:

- The top of runners the race could have.
- To solve the bottleneck's points

For solving this we must know how the race's organization is. There are about 20.000 runners in groups of 1.500-2.000 people, and there is a color that represents each group. The color is chosen by your previously marks, so depending on the color of the group this will be faster or slower than another one. Naturally the first group who leaves the start line is the fastest one and the last one the slowest.

So in this report we have tried to solve the following things. Firstly we have studied if the arrival graphic is similar to a normal distribution graphic. Secondly we have tried to get the optimal inter departure times to do the top of the graphic smaller and to be as constant as possible. Finally we have studied how to improve the backpacks service.

2 Comparison with a normal distribution.

We knew that we had to compare the arrival graphic with another known graphic. In this way we would obtain some positive results although there weren't the exact ones. The probability theory says that the normal distribution appears in a lot of real situations so that was our first option.

*Research partially supported by grant MTM2010-17405 (Ministerio de Ciencia e Innovación, Spain) and grant PI2010-2 (Department of Education and Research, Basque Government).

2.1 The information of 2010.

Since we have thought to compare our graphic with a normal distribution we need the mean value and the standard deviation of each group.

Group	N	μ	σ
Yellow/Green (Y/G)	1438	81,5098864	9,27800215
Green (G)	1078	84,0150433	7,37830386
Red 1 (R1)	1513	91,1848865	7,30030896
Red 2 (R2)	1693	92,9345639	7,53888494
Blue 1 (B1)	1426	98,7045465	8,84973788
Blue 2 (B2)	1546	98,7404592	7,53326797
Blue 3 (B3)	1484	101,116262	8,93872895
Orange 1 (O1)	1337	106,726166	9,37150127
Orange 2 (O2)	1271	108,138028	9,57379314
White 1 (W1)	2085	110,667738	13,1528159
White 2 (W2)	2502	112,704503	13,1367428
Total	17373	100,129511	14,1245918

Table 1: Datas by group for the race of 2010 where appear the number of runners on each group (N), the mean value of each group (μ) and the standard deviation (σ).

We must know that the values have been gathered as if all the people leave the start line at the same time, but this is not true, there are some minutes between the groups. For giving importance to these values we will create the function f_y in the next subsection.

2.2 A function as the race.

For creating f_y function we will use the following notation:

$\mathbf{N}$	Number of runners
$\mathbf{K}$	Number of groups
N_r	Runners in group r
τ_r	Time between group r and $r+1$
τ	An array formed by all of the τ_r
μ_r	The mean value a person of the group r arrives to the finish line
σ_r	The standard deviation of the group r
X_r	People of group r who arrives to the finish line per minute
Y	People who arrives to the finish line per minute

Table 2: Principal notation.

So assuming that $X_i \equiv \mathcal{N}(\mu_i, \sigma_i)$, $i = 1, 2, \ldots, \mathbf{K}$ and $Y \equiv X_1 + X_2 + \ldots + X_{\mathbf{K}}$ then

$Y \equiv \mathcal{N}(\sum_{i=1}^{\mathbf{K}} \mu_i, \sqrt{\sum_{i=1}^{\mathbf{K}} \sigma_i^2})$.

Let

$$f_i(x, \tau_1, \tau_2, \ldots, \tau_{\mathbf{K}-1}) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-(\mu_i+\tau_1+\tau_2+\ldots+\tau_{i-1}}{\sigma_i})^2} \quad (1)$$

the density function of X_i for $i = 1, 2, ..., \mathbf{K}$.

Then

$$f_y(x, \tau_1, \ldots, \tau_{\mathbf{K}-1})) = \sum_{i=1}^{\mathbf{K}} \frac{N_i}{N} \frac{1}{\sigma_i \sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-(\mu_i+\tau_1+\tau_2+\ldots+\tau_{i-1})}{\sigma_i})^2} \quad (2)$$

So we want to find the bestl value of $(\tau_1, \ldots, \tau_{\mathbf{K}-1})$. Those values will be the optimal ones when the people abundance on the finish line was as minimum as possible and as constant as possible during the race. Also it must satisfy the following constraint:

$$\int_0^T f_y(x, \tau) dx < 0.99 \quad (3)$$

when T is a constant time. So the constraint means that in the minute T of the race the 99% of the runners will have crossed the finish line.

2.3 Comparison with 2010.

At this moment we have obtained all the necessary information of 2010 to substitute in the new function done, except the inter departure times, so this is we are going to do, to compare 2010 year's arrival graphic with our theoretical model graphic.

But before of that, we will show if the arrival's graphics per groups seems as the theoretical ones. As we can see on the figure 1 where are represented the pictures for the group 4 and the group 6, the approximations of the first four groups are really bad, but not the rest of them. So we can expect that the global graphic could be disarranged at the start of the picture, when the runners of the firsts groups arrive, and better at the finish of the race.

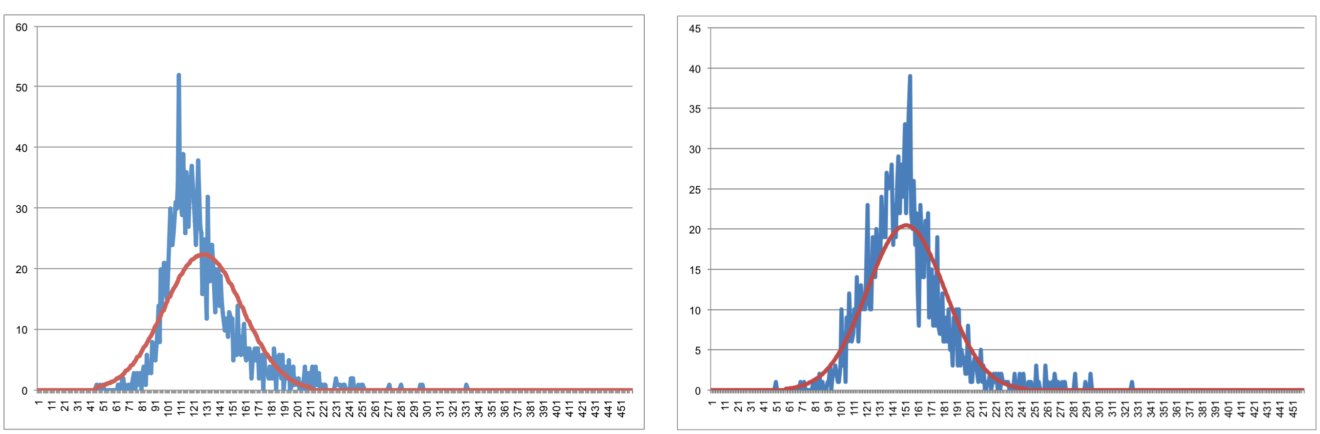

Figure 1: Left picture: Arrival graphic of 4^{th} group. Right picture: Arrival graphic of 6^{th} group

But by the moment, as Fortuna's company gave us runners times for the 0^{th}km, 5^{th}km, the 10^{th}km,and the 15^{th}km's arrival, apart from general ones at 20^{th}km, we will draw the graphics on figure 2 for all the intermediate points to see how the graphic changes and how it approximates to the theoretical graphic.

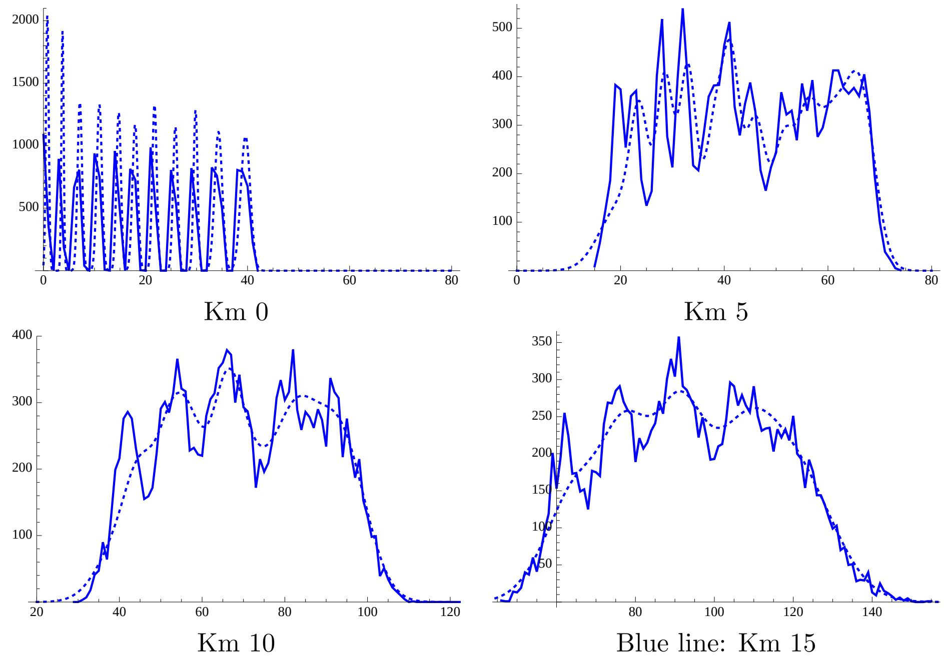

Figure 2: A comparison between the arrivals at the $0^{th}, 5^{th}, 10^{th}$ and 15^{th}km respectively (showed as a normal line) and the theoretical model for these km (showed as dotted lines).

So now we will show on figure 3 the arrival graphic where the normal line is the real graphic of 2010 as Fortuna's company gave us. The other two graphics come of the theoretical model. One of these (the dotted one) uses the inter departure times the organization thought to put, these are $(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8, \tau_9, \tau_{10}) = (3, 3, 3, 3, 3, 3, 3, 4, 4, 5)$ and the other (the dhased one) uses the values we have estimated they used,$(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8, \tau_9, \tau_{10}) = (2.916, 3.083, 3.76, 3.816, 3.083, 4.063, 4.05, 3.95, 3.983, 5.15)$

As we can see on figure 3, the dashed line is a good approximation although initially it doesn't approximate properly as we have estimated previously. So if we assume that the arrival graphic corresponds to a normal distribution, then we can see on the pictures that the organization is helping the race because they aren't using the proposed times. If they had used these, the top of the graphic would have been bigger, which is worst for the race's functioning.

So the question we will try to solve in the next section is the next one. Which had been the optimal values of $\tau = (\tau_1, \ldots, \tau_{10})$? and how to generalize those for others editions of the race?

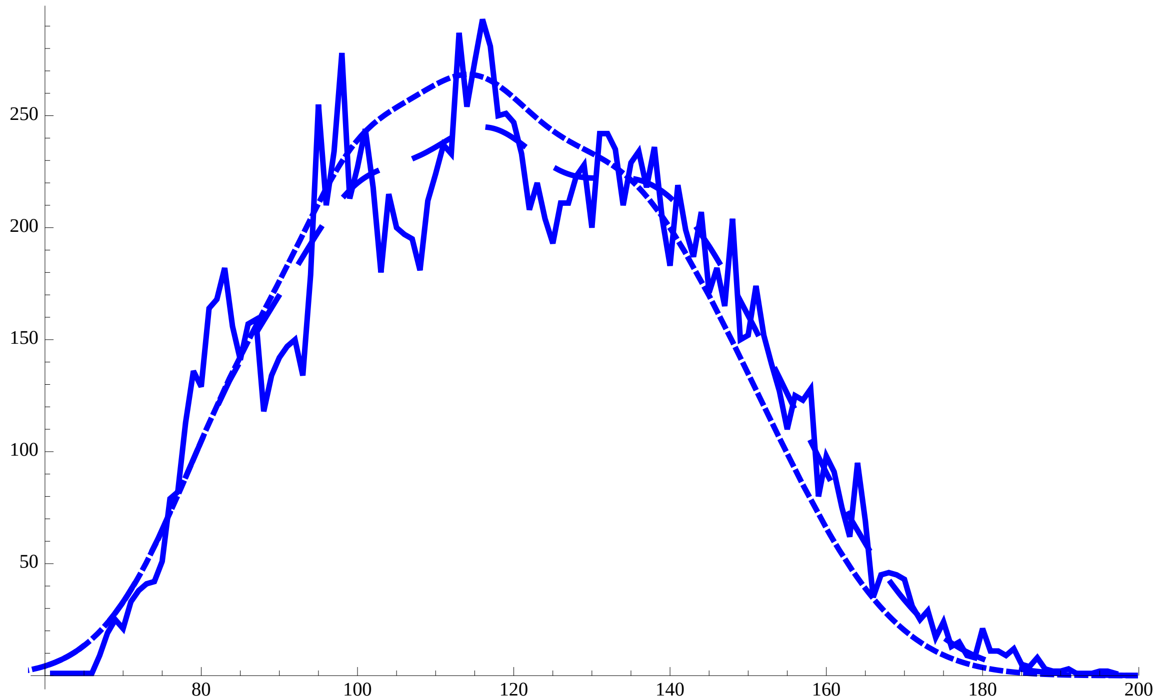

Figure 3: Arrival graphic. Normal line: real graphic. Dotted line: Theoretical model graphic with the bad times. Dashed line: Theoretical model graphic with the exactly times.

3 Optimal inter departure times.

In this section we will try to find the optimal values of the inter departure times to do the arrival graphic as constant as possible in the top and to do the top smaller. Also we will need a time constraint because if we don't put anyone the race could extend too much time and the organization doesn't want that the race lasts more than 180 minutes.

3.1 The first contact with the problem

To have the first contact with the problem and to understand how the graphic changes depending on the inter departure times we will draw some graphics for some different values of τ. These values will be painted with the theoretical curve of 2010 to a faster and more comfortable comparison.

If we study the graphics of the figure 4 we can not see a graphic that minimizes considerably the top except probably the picture with $\tau_i = 5 \forall i$, but it seems that it extends too much because the race can't last more than 180 minutes.

To help ourselves to solve the problem we will build the following table. For each graphic of figure 4 we will find the minute when each graphic arrives to the top and the value of the top to help the organization to prevent the bottlenecks. Also we will find the expected minute when the 95%, the 97.5%, the 99%, the

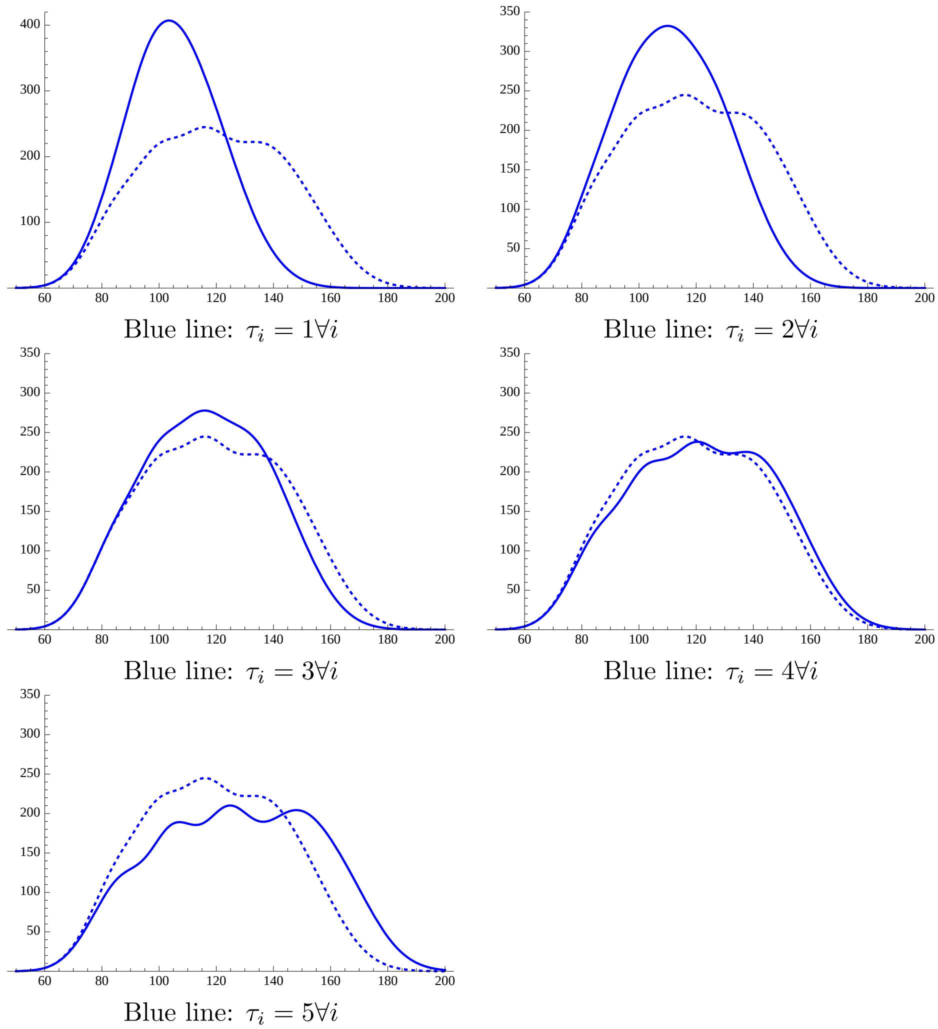

Figure 4: A comparison between f_y with the times of 2010 (dotted line) and and f_y with the times given.

99.5%, and the 99.9% of the runners arrive.

Since the others experiments and studies we have done didn't help us to understand how we can optimize the inter departure times, we can only conclude that the organization was working very well and advice it to fix in four minutes the times if the following editions of the race if it contains more or less the same number of groups.

One thing we must think is that the race's fame rises every year so in the next section we will add runners and groups to 2010's race and we will determine the new top of the curve and how long the race would be.

3.2 Adding population.

f_y	$\tau_i = 1$	$\tau_i = 2$	$\tau_i = 3$	$\tau_i = 4$	$\tau_i = 5$	τ proposed for 2010	τ given in 2010
Minute	103.455	110.039	115.919	120.456	124.863	114.127	116.066
Top	407.43	332.419	277.804	238.223	210.188	268.403	244.931
Constraint 95%	133.7	143.040	152.488	162.015	171.599	155.642	159.452
Constraint 97,5%	138.963	148.476	158.064	167.704	177.385	161.397	165.217
Constraint 99%	144.898	154.518	164.189	173.900	183.645	167.648	171.476
Constraint 99,5%	148.821	158.482	168.187	177.929	187.705	171.706	175.538
Constraint 99,9%	156.651	166.360	176.112	185.902	195.727	179.727	183.567

Table 3: Expected number of people as maximum on the finish line and expected minutes for each constraint.

3.2.1 First case.

Let's suppose that we have in the race two more white groups with the following data.

2010	N	μ	σ
Group W3	2100	113	13
Group W4	2300	112	14

Table 4: Adding two more white groups to 2010's race.

The datas obtained with the new distribution are showed on the table 5:

f_y two more groups	$\tau_i = 1$	$\tau_i = 2$	$\tau_i = 3$	$\tau_i = 4$	$\tau_i = 5$
Minute	108.015	122.006	136.011	147.641	159.771
Top	458.078	370.235	323.447	291.630	266.475
Constraint 95%	138.487	149.124	159.909	170.813	181.822
Consttraint 97,5%	143.534	154.302	165.217	176.263	187.425
Constraint 99%	149.227	160.112	171.158	182.350	193.669
Constraint 99,5%	153.020	163.977	175.107	186.391	197.807
Constraint 99,9%	160.693	171.796	183.094	194.554	206.143

Table 5: Expected number of people as maximum on the finish line and expected minutes for each constraint for the new distribution of the race.

In this case the main problems are on one hand the time the race need and on the other hand the big population in the lasts groups. We can see how the curve rises in the lasts minutes in the figure 5.

So we think that the organization could have been got a good idea in its proposed times adding a minute between the lasts groups or that too many population needs more groups. Maybe this is not a good idea because the race might be too long but we will study in the next section.

If we study the solutions for each possibility: $\tau = (3, \ldots, 3, 4), \tau = (3, \ldots, 3, 4, 4), \ldots, \tau = (3, 4, \ldots, 4)$ we will know that the best solution for the interests of the organization is to fix in four minutes the inter departure time between the orange's groups until the last two groups, i.e., $\tau = (3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4)$.

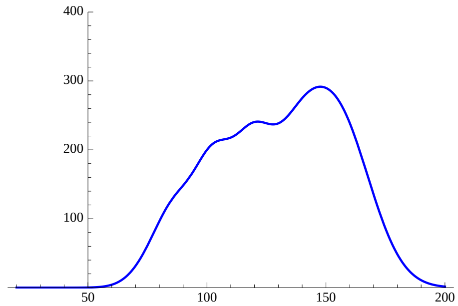

Figure 5: Graphic with the new data for $\tau_i = 4, i \in \{1, \ldots, 12\}$

Let's see the graphic for the new value of τ.

Minute:	140.114
Top:	292.82
Const 95%:	163.813
Const 97.5%:	169.26
Const 99%:	175.35
Const 99.5%:	179.39
Const 99.9%:	187.55

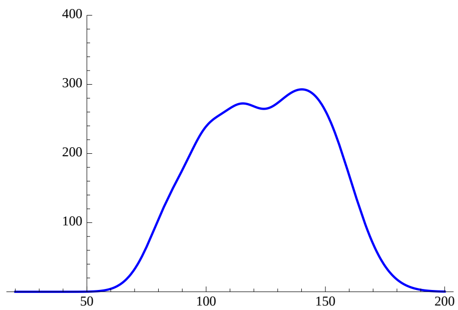

Figure 6: Graphic for 13 groups and $\tau = (3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4)$.

Logically we have found a lot of better solutions but as these don't follow a pattern theses are been rejected.

3.2.2 Second case.

As we have seen in the previous case a big population in the lasts groups could generate problems. To solve or to minimize these problems we will add another group. Since the last four groups, the white ones, have a lot of runners, we will create another white group so now we have 5 white groups with the following datas. By simplicity the groups have the same number of runners, the same mean value and the same standard deviation.

Group	N	μ	σ
W1	1797	112.12	13.33
W2	1797	112.12	13.33
W3	1797	112.12	13.33
W4	1798	112.12	13.33
W5	1798	112.12	13.33

Table 6: New distribution of the white groups.

And again the best solution is when $\tau = (3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4)$ with the following datas.

Minute:	142.702
Top:	274.061
Const 95%:	166.094
Const 97.5%:	171.581
Const 99%:	177.655
Const 99.5%:	181.656
Const 99.9%:	189.661

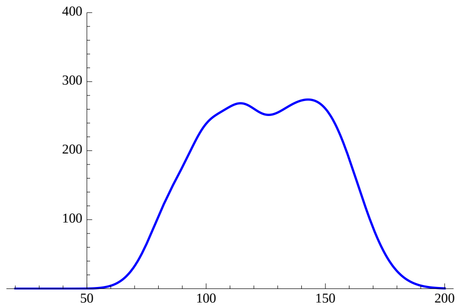

Figure 7: Graphic for 14 groups and $\tau = (3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4)$.

If we compare the figure 6 and the figure 7 we can see that if we add one more group the race will take two more minutes more or less but the top will increase considerably so it might be a good idea to add more groups instead of distributing the groups with too much people.

4 Backpacks.

The backpacks service is one of the most important services the organization gives because there is a big population of runners that don't live in San Sebastian and they need to change their clothes when they finish the race without long queues and bottlenecks.

The organization estimates that the number of people that use this service is about a 50% of the people but it knows that when the service works properly the number of people will rise, but at the moment we only want to guarantee a good service for the half of the population.

Let's describe how the organization gave this service in 2010 for understanding why it doesn't work correctly. This year the backpacks were distributed in 20 arcs. In each arc there were six volunteers distributing the backpacks, so there was 120 people for the job. Each runner depending on his/her dorsal had his/her arc to receive the backpack. Then the arc number one was for the runners with the dorsal between 1-1000, the number two for 1001-2000, ..., and the las one for 19001-20000. As the first group who left the start point was the group with the dorsals between 1-2000, the second was the group with the dorsals 2001-4000, ... then there was an overflow on 2 or 3 arcs and the others completely empty during all the race because ρ, the coefficient of server utilization (we will define later), was bigger than one, that means that the number of people arrived to the backpacks service was bigger than the number of people left. So is indispensable to change the model.

4.1 A new model.

The model that is being used has some advantages. Is too easy for the volunteers to order and to find the backpacks. But there are some disadvantages too, the main one is the saturation on the arcs.

So we proposed some models, all of them based on randomizing the arc in which each runner have to go, and the model that Fortuna preferred because this model can randomize properly the arc and because with it the volunteers can order and find the backpacks more quickly than with the other proposed models is the following one: To randomize depending on the last two numbers of the dorsal. For example, in 2010 could have done: The runners whose dorsal finished in 01-02-03-04-05 had to go to the first arc, if their dorsal finished in 06-07-08-09-10 they had to go to the second arc, . . . , and if their dorsal finished in 96-97-98-99-00 they had to go to the last arc, the 20^{th} one.

4.2 Applying the model.

In this section we will see why is too important to randomize the arc to avoid bottlenecks and to disregard of some volunteers.

But before, we must learn the M/M/c theoretical model of the queueing theory.

In the mathematical theory of random processes, the M/M/c model is a multi-server queue model.

Following Kendall's notation it indicates a system where:

- Arrivals are a Poisson process.
- Service time is exponentially distributed.
- There are c servers
- The length of queue in which arriving users wait before being served is infinite
- The population of users, or requests, available to join the system is infinite

Such a system can be shaped by a birth-death process, where each state represents the number of users in the system. As the system has an infinite queue and the population is unlimited, the number of states the system can occupy is infinite: state 0 (no users in the system), state 1 (1 user), state 2 (two users), etc. As the queue will never be full and the population size being infinite, the birth rate (arrival rate), λ, is constant for every state. The death rate (service rate), μ, is not a constant because it depends on the relation between the state k and the number of servers c, i.e, if $0 \leq k \leq c \Rightarrow \mu_k = k\mu$ else $\mu_k = c\mu$. There is a relation between this two variables called coefficient of server utilization or traffic intensity and it is noted by ρ. It's relation is $\rho = \frac{\lambda}{c\mu}$. If the traffic intensity is greater than one then the queue will grow without bound and the system is unstable but if not we can obtain some interesting datas.

So we will define these new variables and the told ones on the table 7 and later we will write the formulas for all of them.

$$p_0(\lambda, \mu, c) = \{(\sum_{n=o}^{c-1} \frac{(\frac{\lambda}{\mu})^n}{n!}) + \frac{(\frac{\lambda}{\mu})^c}{c!(1 - \frac{\lambda}{c\mu})}\}^{-1} \tag{4}$$

Λ	Number of people that crosses the finish line per minute.
λ	Number of people that arrives to the backpacks service per minute.
μ	Number of people each volunteer attend per minute.
c	Number of servers / volunteers in each arc.
ρ	Coefficient of server utilization.
p_o	Probability of being 0 runners in the system.
L	Expected number of runners in the system.
L_q	Expected number of runners in the queue.
W	Expected time each runner have to wait in the system.
W_q	Expected time each runner have to wait in the queue.

Table 7: Principal notation.

$$L_q(\lambda, \mu, c) = \frac{p_o(\lambda, \mu, c) * (\frac{\lambda}{\mu})^c * (\frac{\lambda}{c\mu})}{c! * (1 - \frac{\lambda}{c\mu})^2} \tag{5}$$

$$L(\lambda, \mu, c) = L_q(\lambda, \mu, c) + \frac{\lambda}{\mu} \tag{6}$$

$$W_q(\lambda, \mu, c) = \frac{L_q(\lambda, \mu, c)}{\lambda} \tag{7}$$

$$W(\lambda, \mu, c) = \frac{L(\lambda, \mu, c)}{\lambda} \tag{8}$$

So now we are ready to do a simulation of how could have done the backpacks service if Fortuna had used the new model. But we need the exact values of some dates for obtaining the rest.

As the organization estimated the half of the population uses this service, so $\lambda = \frac{1}{2}\Lambda$, also they estimated that $\mu = 3$ runners/minute so we will use that data although it seems a very quick process.

As the organization used in 2010 20 arcs to distribute the backpacks, we will use the same number of arcs, so the number of runners that arrived to each arc per minute will be $\frac{\lambda}{20}$.

In 2010 $max\{\Lambda\} = 293$ so we will work for $\Lambda_1 = 300$ and $\Lambda_2 = 250$.

Let's be $\Lambda_1 = 300$, then $\frac{\lambda}{20} = 7.5$ runners / minute, and we obtain the following table:

So if we need to attend the arrival of 300 runners (150 to the backpacks service) without complications we would need only 3 volunteers per arc, the system wouldn't be too much occupied and each runner would wait 0.80 minutes=48 seconds.

c	ρ	L	W
3	0.8333	6.011	0.8015
4	0.625	3.033	0.4044
5	0.5	2.6303	0.3507

Table 8: Datas depending on the number of volunteers in each arc if $\Lambda = 300$

But this would happen only in the maximum of the race. It would be more convenient to see what it happens if $\Lambda = 250$ because it was an affluence more typical.

Let's be $\Lambda_2 = 250$, then $\frac{\lambda}{20} = 6.25$ runners / minute, and we obtain the following table:

c	ρ	L	W
2	1.04	##	##
3	0.6944	3.1839	0.5094
4	0.5208	2.2953	0.3673

Table 9: Datas depending on the number of volunteers in each arc if $\Lambda = 250$

That is a good simulation to see how the coefficient of server utilization and the number of volunteers can decrease without bottlenecks, but we will have better information in the next section with the datas of 2011.

Also we can obtain the maximum value of Λ depending on the number of volunteers in each arc. It will be the maximum when $\rho = 1$. But we must know the values we are going to obtain won't be very realistic because with $\rho = 1$ the system is unstable and also $\mu = 3$ seems a very optimistic data so let's do a table that showed us these datas depending on the number of servers and differing between $\rho = 1$ or $\rho = 0.85$ and $\mu = 3$ or $\mu = 2$ respectively.

	$\rho = 1$ and $\mu = 3$	$\rho = 1$ and $\mu = 2$	$\rho = 0.85$ and $\mu = 3$	$\rho = 0.85$ and $\mu = 2$
$c = 2$	240	160	204	136
$c = 3$	360	240	306	204
$c = 4$	480	320	408	272
$c = 5$	600	400	510	340
$c = 6$	720	480	612	408

Table 10: Maximum number of runners could arrive to the finish line point depending on the servers if $\rho = 0.85$ and $\mu = 3$

So as we can see this service was working very badly because it has too much capacity if we compared with the maximum arrival in 2010.

5 2011 estimations.

For the race of this year the organization gave us the following datas showed on the table 11

Dorsal color	Runners	Dorsal
Yellow	276	1-500
Green	2586	501-4000
Red	3993	4001-9000
Blue	6300	9001-16000
Orange	3640	16001-20000
White	6276	20001-26999
TOTAL	23071	

Table 11: Datas of 2011

We know that the group of yellow's runners go out with the green group, so if we join the two groups we obtain a group with 2861 runners.

When we compared the datas of 2009 and 2010 we saw that the mean value and the standard deviation is more or less equal for the groups with the same color, so for doing the estimation we will use the weighted mean value for both of them. Also for simplicity we will divide the groups of the next way as we show on the table 12.

Group	Color	Runners	μ	σ
1	Yellow/Green (Y/G)	1431	82.71	8.21
2	Green (G)	1431	82.71	8.21
3	Red 1 (R1)	1997	92.53	7.80
4	Red 2 (R2)	1996	92.53	7.80
5	Blue 1 (B1)	2100	100.30	8.59
6	Blue 2 (B2)	2100	100.30	8.59
7	Blue 3 (B3)	2100	100.30	8.59
8	Orange 1 (O1)	1820	107.89	9.65
9	Orange 1 (O2)	1820	107.89	9.65
10	White 1 (W1)	2092	112.09	13.22
11	White 2 (W2)	2092	112.09	13.22
12	White 3 (W3)	2092	112.09	13.22

Table 12: Datas for 2011

5.1 Solutions depending on τ

Depending on the value of τ we obtain the following graphics as the figure 8 show us with their respectively datas as we can see on the table 13.

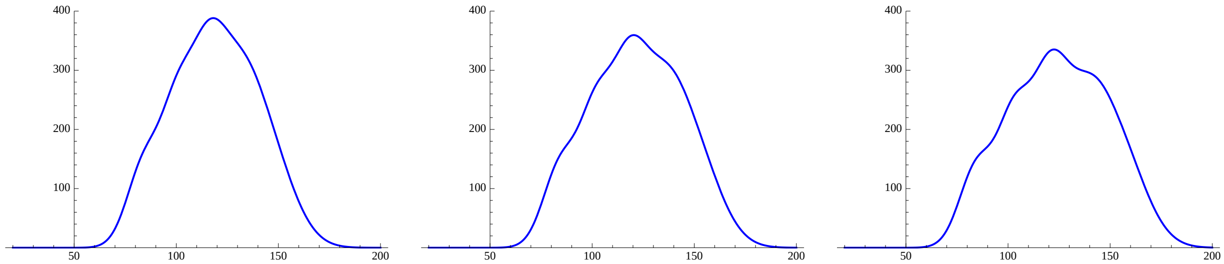

Figure 8: Estimated graphics for 2011 if $\tau = 3$, $\tau = 3.5$ or $\tau = 4$ respectively.

	$\tau = 3$	$\tau = 3.5$	$\tau = 4$
Minute:	118.111	120.388	122.537
Top:	388.146	359.428	335.121
Const 95%:	154.414	159.429	164.471
Const 97.5%:	160.021	165.11	170.221
Const 99%:	166.165	171.303	176.465
Const 99.5%:	170.172	175.336	180.523
Const 99.9%:	178.12	183.327	188.561

Table 13: Datas for 2011 depending on the value of τ.

Then we recommend to put the inter departure time in 4 minutes if we only contemplate those three options and if the organization consider properly that about 127 people (estimated) won't finish the race in the 180 minutes if we only contemplate those three options.

But doing more simulations we have discovered that using an inter departure times that simulate a normal picture, i.e., a short periods of time at the beginning and the end of the start and longer at the medium we can improve the maximum of the race and to do more constant the arrivals in this point. For illustrate it we will print the best solution we have found on the graphic 9, so those have been our times recommended for 2011 before we were working in another different model and more precise than the other one. But as we wanted to add a new red group to try to allow the first pick of the graphic and as the Organization could have some problems to distribute the groups with the times proposed we decide to fix the times in $\tau = (2, 2, 3, 3, 3, 5, 6, 4, 4, 4, 4, 4)$ which are a good times in agreement to the race needs. In the following picture, graphic 10, we will can see that. Obviously the second graphic seems too much better than the first one, its problem is that the last runners will arrive too late, but as the organization this year could have a lot of problems in case more than 300 runners arrive at the same time it prefers do the race longer in time.

Minute:	117.749
Top:	293.391
Const 95%:	163.343
Const 97.5%:	169
Const 99%:	175.16
Const 99.5%:	179.17
Const 99.9%:	187.12

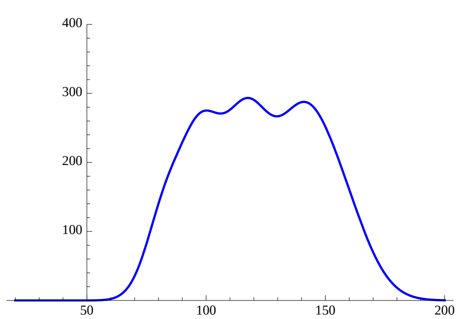

Figure 9: Graphic and datas for $tau = (2, 2, 3, 4, 5, 6, 5, 5, 4, 3, 3)$.

Minute:	128.683
Top:	277.384
Const 95%:	172.604
Const 97.5%:	178.883
Const 99%:	185.818
Const 99.5%:	190.364
Const 99.9%:	199.407

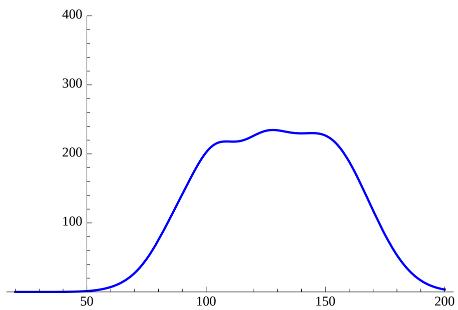

Figure 10: Graphic and datas for $tau = (2, 2, 3, 3, 3, 5, 6, 4, 4, 4, 4, 4)$.

5.2 Backpacks

For the backpacks study we know that not all the people will participate in the race only about the 85%-90% of the inscriptions, but if we work with all the people we will assure a perfect work.

The organization told us that this year they will can use 30 arcs and 100 volunteers, there will be arcs with only one volunteer attending to one ending, arcs with two attending to two, with three and with four volunteers so the number of runners arrive to each arc will be $\lambda\frac{c}{100}$. This happens because there are some arcs with space problems. Then we must see if depending on the number of volunteers there could be a bottleneck or not.

Then we will see how the system works for $\Lambda = 350$, $\Lambda = 300$ and $\Lambda = 250$ variating the values of μ because it doesn't seem a good approximation the organization gave us ($\mu = 3$ runners/minute), it seem bigger than the real value.

On the table 14 we can see how the system works if $\Lambda = 350 \Rightarrow \lambda = 175$

c	$\lambda\frac{c}{100}$	μ	ρ	L	W
1	1.75	1.5	1.16667	##	##
1	1.75	2	0.875	7	4
1	1.75	2.5	0.7	2.33333	1.33333
2	3.5	1.5	1.16667	##	##
2	3.5	2	0.875	7.46667	2.13333
2	3.5	2.5	0.7	2.7451	0.784314
3	5.25	1.5	1.16667	##	##
3	5.25	2	0.875	8.03808	1.53106
3	5.25	2.5	0.7	3.2488	0.61882
4	7	1.5	1.16667	##	##
4	7	2	0.875	8.66503	1.23786
4	7	2.5	0.7	3.80019	0.542885

Table 14: Datas if $\Lambda = 350$ depending on c and μ values

Then the organization must be careful because if $\mu = 1.5$ there will be bottlenecks, so they have to assure that $\mu \geq 2$ for attending to 350 runners.

On the table 15 we can see how the system works if $\Lambda = 300 \Rightarrow \lambda = 150$

c	$\lambda \frac{c}{100}$	μ	ρ	L	W
1	1.5	1.5	1	##	##
1	1.5	2	0.75	3	2
1	1.5	2.5	0.6	1.5	1
2	3	1.5	1	##	##
2	3	2	0.75	3.42857	1.14286
2	3	2.5	0.6	1.875	0.625
3	4.5	1.5	1	##	##
3	4.5	2	0.75	3.95327	0.878505
3	4.5	2.5	0.6	2.33212	0.518248
4	6	1.5	1	##	##
4	6	2	0.75	4.5283	0.754717
4	6	2.5	0.6	2.65997	0.471761

Table 15: Datas if $\Lambda = 300$ depending on c and μ values

As in the previous case the system doesn't work if $\mu = 1.5$.

On the table 16 we can see how the system works if $\Lambda = 250 \Rightarrow \lambda = 125$

c	$\lambda \frac{c}{100}$	μ	ρ	L	W
1	1.25	1.5	0.833333	5	4
1	1.25	2	0.625	1.66667	1.33333
1	1.25	2.5	0.5	1	0.8
2	2.5	1.5	0.833333	5.45455	2.18182
2	2.5	2	0.625	2.05128	0.820513
2	2.5	2.5	0.5	1.33333	0.533333
3	3.75	1.5	0.833333	6.01124	1.603
3	3.75	2	0.625	2.52066	0.672176
3	3.75	2.5	0.5	1.73684	0.463158
4	5	1.5	0.833333	6.62194	1.32439
4	5	2	0.625	3.03309	0.606619
4	5	2.5	0.5	2.17391	0.434783

Table 16: Datas if $\Lambda = 250$ depending on c and μ values

Then we can conclude that if the organization can assure that $\mu \geq 2$ then the backpacks service will work properly although in the arc with only one server the time a runner have to wait might be very big.

6 Aims for the next race

Although we can expect a better working of the race than the previous year, we know there is more things we can do to improve 2012's race. These are some of the ideas for the next year.

- A better approximation: As we have seen, our actual model is good for having a first approximation but of course is not perfect. In fact it is not able to predict the big picks the arrival's picture has. Then, we are building a new one which is more optimistic than this one.
- Means of transport: As the race starts in a different city it finishes and as it is very famous between runners , people has to use a vehicle for going to Behobia or for returning of Donosti or for both of them. Then is needed an study of the percentage of people that uses each mean of transport to avoid futures bottlenecks.
- A future problem in the start point (Behobia): Is known that the Behobia-Donostia called race's popularity rises every year, and that the finish line point hasn't got an infinitive capacity. Moreover, the organization doesn't want that the race extends up to one hour because it must gather all the things to restore the traffic. So Fortuna's workers think a solution could be to anticipate the starting time. We must study if this possible solution could do a bottleneck in Behobia.
- Gipuzkoa Square: This is next to the finish line. It has a lot of services and the runners can leave the square in some points, it depends on their needs. Then this year the organization is going to use some movement sensors that can help us to see the percentages of people that leave the square in each exit point to know the people that use each service. After the study of this datas we can conclude how the better way of distributing the services is to can avoid futures bottlenecks.

7 Behobia-Donosti 2011^{st} edition.

7.1 Introduction.

In this section we will show on one hand how the race was such a good race for the organization: no bottlenecks, no high picks, good backpacks service... and how the race was the worst race for the runners. The problem of this was the sun and the high temperatures.

To the inexperts in long race's topic, we must inform that running with too many degrees is dangerous and also very hard. Being well prepared is necessary. Considering that the race is in November in the Basque Country it was unbelievable knowing that there were about 21-25C, although the Organization noticed the runners of this while the weather forecast was changing. Moreover there was a very hard sun and south wind which it do the race harder.

The main question we have done ourselves is if it was possible to support the level of the previous years. Obviously not. As we can see on the picture 11 the time one runner needed too complete the race in 2011 is higher than the previous one, whereas the pictures of the years 2009 and 2010 are too similar. The values of the runners are given in percentages, so if we want to know the number of runners at some point we have to multiply by the total number of runners the race had.

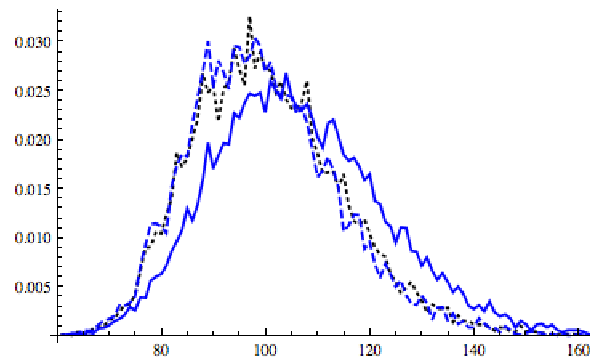

Figure 11: Minutes runners needed for have the race finished in 2009, black dotted line, in 2010, blue dashed line, and in 2011, blue normal line.

So considering that, you will can be doing the following questions:

- If this year have had unexpected values then your predictions were bad.
- The time a runner have to do to finish the race has risen for all the runners or only for the worse ones?
- People lost the time constantly or the more kilometers the worse time?

These questions and another ones we will explain in the following subsections.

7.2 Datas and first comparisons.

In this subsection we will write the real datas of the 2011's race and later we will show some comparison's table between the lasts editions of the race.

The first thing the table 17 where the real datas are.

As we can see, the means values and the standard deviations of each group and consequently of each color, seem higher than other editions. On the table 18, we can observe how the values have risen this year respect the lasts editions depending on the color of the dorsal.

We can see on the table that the mean values of 2011 year's race are too high. There are such a high values that its are comparable with the mean values a previous level (color) of another year had whereas if we compare the year 2009, too much cold, rain and hail, and the year 2010, really a good weather for running, the mean values only decrease a few. So we think the hot weather influences too much more than

Group	N	μ	σ
Y/G	1596	86.5664056	13.536966
G	1017	87.1246477	9.0701237
R1	1133	96.3972051	9.4198328
R2	1112	96,2304556	8.3985098
R3	1194	97.8884841	9.449979
A1	1823	105.751893	10.303723
A2	1785	105.803716	10.178712
A3	1694	107.806759	10.766862
O1	1545	114.868457	12.54092
O2	1518	116.361012	12.258426
W1	1854	118.066046	15.715823
W2	1962	118.102752	14.552625
W3	1471	119.650963	14.810813
Total	19704	106.824365	16.345945

Table 17: Datas by group for the race of 2011 where appear the number of runners on each group (N), the mean value of each group (μ), the standard deviation (σ) and the color of each group.

	2009			2010				2011			
Color	N	μ	σ	N	μ	diff(%)	σ	N	diff(%)	μ	σ
Y/G	2134	82,9	7,9	2516	82,6	(-0.3%)	8.6	2613	86,8	(+5.1%)	12,0
R	2963	93,0	8,2	3206	92,1	(-0.9%)	7.5	3439	96,9	(+5.2%)	9.1
B	3547	101,3	8,8	4456	99,5	(-1.7%)	8.5	5302	106,4	(+6.9%)	10,5
O	1927	108,5	9,9	2608	107,4	(-1.0%)	9.5	3063	115,6	(+7.6%)	12,4
W	4290	111,8	13,1	4587	111,8	(-0.1%)	12.6	5287	118,5	(+6.0%)	15,1

Table 18: Comparison of datas between the last years.

the cold one. It makes it harder.

Also another question is solved which is the following one. Everybody independently on his / her level loses more or less the same time (on percentage). So the level is not important for studying the losses of time, although the best groups lost fewer than the others, possibly because they are better prepared and trained. We can appreciate this better on the figure 12.

7.3 Intermediate points.

As we have obtained that the generalized loss of time is not blame of the worst runners in this subsection we will study if the more kilometers each runner run the more time each runner spent on it. How can we do this?

There are three intermediate points along the race: one in the km5, another in the km10 and the other in the km15, i.e., one intermediate point for each five kilometers. So lets show a table with the times spent on each 5 kilometers in 2010 and afterwards with 2011.

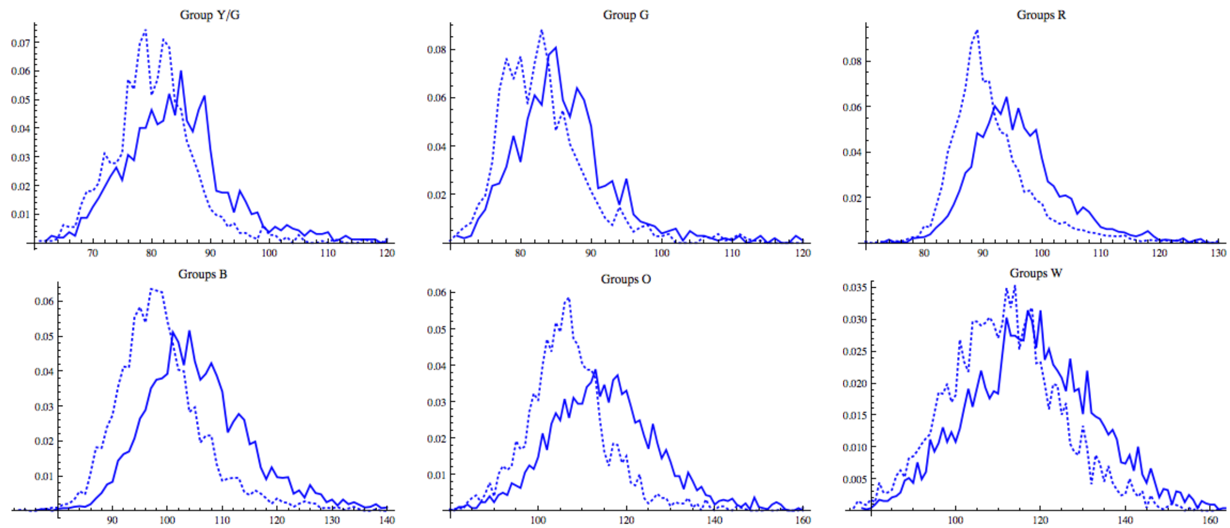

Figure 12: Comparison between the times for 2010 and 2011 collected by groups.

2010	Km0 - Km5	Km5 Km10	Km10 Km15	Km15 Km20
Grupo Y/G	20.29	20.87	20.03	20.32
Grupo G	20.80	21.60	20.82	20.80
Grupo R1	22.58	23.36	22.57	22.67
Grupo R2	23.21	23.80	23.00	22.92
Grupo B1	24.62	25.24	24.33	24.51
Grupo B2	24.77	25.24	24.30	24.43
Grupo B3	25.27	25.88	24.92	25.05
Grupo O1	26.65	27.22	26.27	26.59
Grupo O2	27.11	27.53	26.61	26.89
Grupo W1	27.81	28.07	27.28	27.51
Grupo W2	28.38	28.53	27.92	27.87

Table 19: Minutes per each section in 2010.

As we can see on the table 19 in 2010 the times were more or less equal except in the section between the kilometer 5 and the kilometer 10 where it is the pass called Gaintxurizketa and where the runners lose a few minutes due to the hardness of it. Moreover, although the times are more or less equal in the rest of sections the last two sections are faster than the first one.

Lets see what happened in 2011

At first, the first five kilometers of 2011 seems as the 2010 ones, but then the times were too much wore than the 2010 ones. This means that the runners didn't know how to run with the atypical weather there was, and they didn't decrease their times constantly as they should have done. We will see better on the figure 13.

2011	Km0 - Km5	Km5 Km10	Km10 Km15	Km15 Km20
Grupo Y/G	20.93	21.90	21.62	22.12
Grupo G	20.59	22.47	21.86	22.20
Grupo R1	22.52	24.62	24.36	24.90
Grupo R2	22.46	24.57	24.35	24.85
Grupo R3	23.11	25.08	24.69	25.00
Grupo B1	24.87	26.95	26.68	27.25
Grupo B2	24.89	26.99	26.63	27.30
Grupo B3	25.34	27.48	27.22	27.76
Grupo O1	26.92	29.31	28.99	29.65
Grupo O2	27.42	29.78	29.35	29.80
Grupo W1	27.90	30.25	29.70	30.21
Grupo W2	28.06	30.31	29.62	30.12
Grupo W3	28.41	30.79	29.99	30.46

Table 20: Minutes per each section in 2011.

7.4 Normal model.

So the following question you will be doing may be this one: "Then the predictions you did were bad". Obviously yes, the graphic we estimated there would be is not the most satisfactory one because the mean value and the standard deviation of each group change. But there is something positive with the results which is that the approximation of the normal model to the real graphic using the real datas is the better approximation we have found at the moment so apparently the hard weather do the distribution of the arrivals more similar than a normal distribution.

On the following picture, the figure **??**, we will can see a comparison between the normal model real graphic, and with the normal model graphic that we had done with the real number of runners on each group (the approximation we did before don't have the real number of runners if not the inscriptions).

FOOOOOOOOOOOOOOOOOOOOOOOOOOOOOTOOOOOOOOOOOOOOOOOOOOOOOOO

FOOOOOOOOOOOOOOOOOOOOOOOOOOOTOOOOOOOOOOOOOOOOOOOOOOOOO

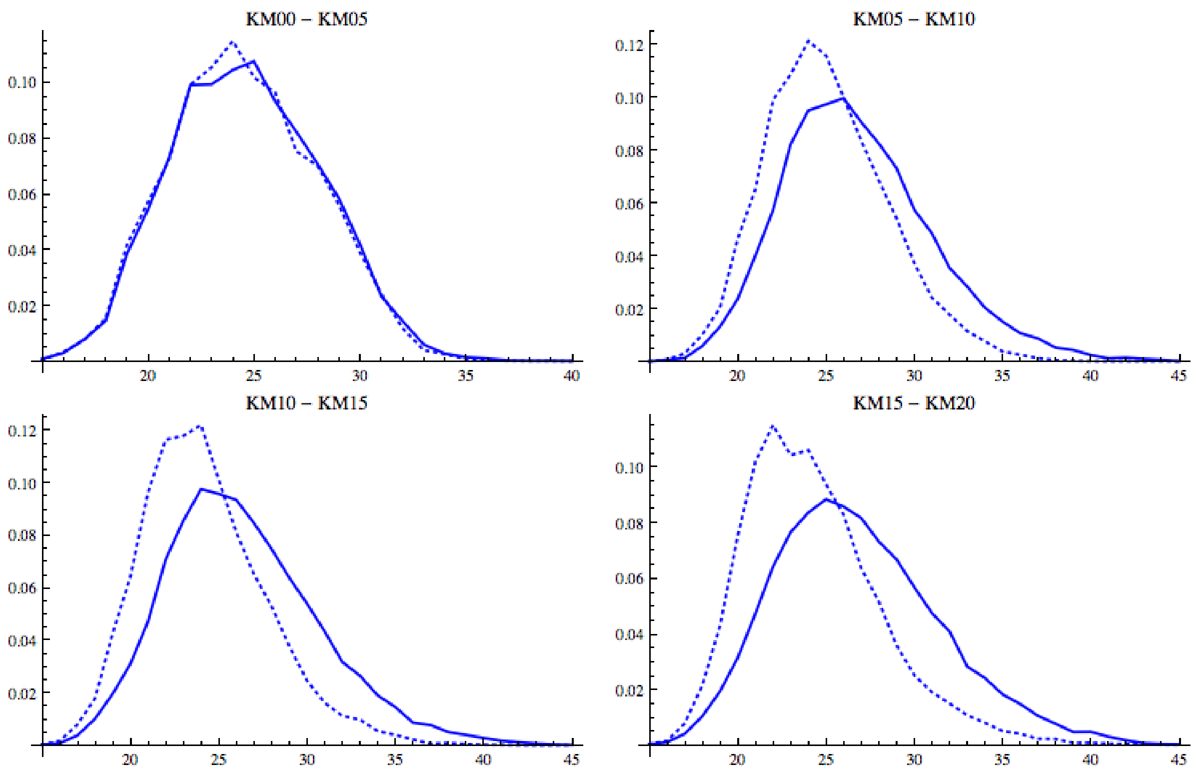

Figure 13: Runners per minute between each intermediate point.