Andrea Lazzari 2045247 andrea.lazzari.8@studenti.unipd.it
Tommaso Amico 2054478 tommaso.amico@studenti.unipd.it

Biathlon analysis with Bayesian inference

Abstract

In this brief analysis we study, using Bayesian Inference, shooting performances in biathlon.

Biathlon combines indeed a cross-country-skiing part of the competition and a shooting one, that can present itself on a prone or in a standing position.
We can identify 4 main race formats in the word cup,

- Sprint: athletes start at equal time intervals, whoever finishes in the least amount of time wins. They cover 3 laps, at the end of the first one they go through a prone shooting range. At the end of the second lap they shoot in the standing position.

  Every miss is a penalty lap.

- Pursuit: the athletes start with the same gap accumulated during the sprint, whoever finishes first wins. There are 5 laps with in between 4 shooting ranges, the first 2 are prone shooting ranges while the second 2 are standing.

  Every miss is a penalty lap.

- Mass start: it's similar to a pursuit race, the difference being that the athletes start all together.

- Individual: athletes start at equal time intervals, they perform 5 loops and alternate a prone shooting range followed by a standing one, another prone and the last which is a standing shooting range.
  Whoever travels the distance in the least amount of time wins.

  Every miss is an additional minute on the individual time.

The intrinsic binomial nature of a shooting: you either hit the target with probability p or you miss it[1] with probability $1 - p$, makes it perfect for an analysis using Bayesian inference. The framework we move within in this project relies on the classic inference pattern, where we choose an appropriate prior and we update our belief with the right likelihood obtaining the posterior, that is, the probability distribution function (PDF) summarizing our knowledge over the model's parameters.
A brief article with a little bit more insights on Bayesian inference for an interested reader can be found here.

Introduction

Our work is divided into two main branches, in the first part we select only races that are either pursuits or mass starts.
The idea is to analyze the performance of biathletes when they enter the last shooting range in the top 10 and when they enter it outside the top 10 separately. The goal is to seek for any difference in the standing shooting percentage in the two cases. That is, are shooters faced by being in the thick of things or do they perform independently from that?
Men and women are analyzed in the same way then, making use of a Gaussian approximation, we look for any discrepancy based on sex. This part, as described in the Methods section, will be completed via a Markov Chain Monte Carlo (MCMC) application.
The second branch is developed with in mind the idea of debugging or confirming the myth of the last shot in individual races: is it really that more difficult? With the aim of finding out if the danger of an additional minute dramatically lowers shooting percentages we divide the data set

[1] Unless you don't even shoot the bullet, in that case everything is ruined... we are thinking at you and that Anterselva individual Lisa ♡

between individual races and the other 4-range competitions, after a further filtering process a one sided Bayesian hypothesis test is performed and the results are discussed in the Results section.

The data set used is composed of every race starting from the 2014/2015 season until the last race of the 2021/2022 word cup season.

The programming language used is the R language [1], for interfacing with the *R* version of the *Jags* [2] library we make use of the *BUGS* [3] language.

The full code is available in the linked GitHub page.

Methods

Let's now see step by step the thinking process that we followed to analyze the clutchness of biathletes.
All of us have in mind a great standing shooting battle that led to glory for one contestant and to painful defeat for the other. Fillon Maillet has won many but we can clearly remember how Vetle Christiansen won bronze above him in the Beijing olympics' *Mass start*. It is also tough to forget how a near Dorothea Wierer's gold became silver thanks to a crazy performance by Marte Roiseland in Anterselva's world championship pursuit.
In this first chunk of our work we want to see who consistently prevails in this situations and who, instead, suffers.

Let's start by describing the data at our disposal, for this part 2 main data frames were exploited. The first one summarizing the result of the race, organized with the following columns,

- **Contestant identification:** 5 columns specifying the final rank of the athlete, its name, its surname, its bib in the race and the nationality.
- **Shooting:** 2 columns describing the total misses in the race and the result of each of the shooting range.
- **Time:** 3 columns specifying the total absolute time, the gap from the first position and, only for pursuits, the isolated pursuit time.
- **Race identification:** 4 columns identifying the year of the race, the venue, the format and a unique ID for each race.

The second one instead brings the necessary informations on the second to last lap, that is, the lap that leads to the final shooting range. It differs from the table described above for the following columns.

- **Loop times:** Displaying the cumulative time at the finish of the loop, the time needed to complete the lap, the time of the skiing part of the lap and finally the range, shooting and penalty time.

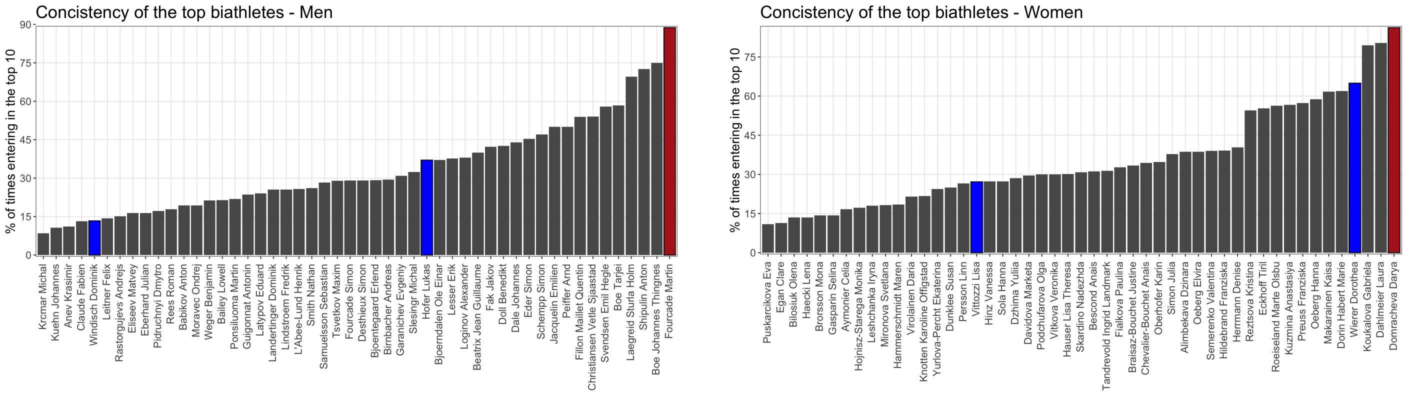
These two tables were joined through a *left join* with the left data frame being the *Results* data frame because it had records not present in the *Final loop* one, including *DNF, DSQ,* and *DNS* records.
We further managed our data by separating rows representing athletes entering the last shooting range in the top 10 from rows in which biathletes were outside of it.
To have a significant statistical sample we restricted our work to biathlethes that entered in the top 10 at least 5 times.

We show, in Figure 1, the outcome of the pre-processing part representing the most consistent biathletes, that is, the racers that percentually entered the last shooting range in top 10 the most.

We can see how Martin Fourcade unsurprisingly stands alone with more then 88% of top 10 appearances with a big lead over the 75% of Johannes Thingnes Bø.
For women instead we can see how Dahlmeier, Koukalova and Domračeva stand alone and we have another big drop after Kristina Reztsova.

Context matters though, we repeat how the data frame represented races from the $2014-2015$ season on, we have then that for some people we consider the prime of their career while for others we could follow their first steps in the word cup or their last shots before retirement, thus, this should be considered this into account in our evaluation.

After this first pre-processing step we began our actual analysis, to determine if and how much competitors were influenced by being in contention for an important position, by working in analogy with the process of determining the efficiency of a vaccine.



(a) (b)

Figure 1: Consistency of the top biathletes

When a new vaccine is being tested a set of volunteers is shot by either a placebo or an actual vaccine shot. Neither the patient nor the doctor knows the nature of the shot; based on how many volunteers develop the illness, the efficiency is computed as,

$$Efficiency = 100 \cdot \frac{placebo - vaccine}{placebo}\% \tag{1}$$

Let's see a practical example to get used with the process and with pattern of operation that we will follow throughout the length of this document.
Pfizer announced that their vaccine against COVID-19 is more then 90% effective, they studied 43538 volunteers and found 94 evaluable cases of COVID-19, the *American Food and Drug* administration set a minimum effectiveness level at 50%. The data are summarized in the following table,

	Negative patients	**Positive patients**
Placebo	21453	86
Vaccine	21991	8

As prior for the probability of developing the illness we use a *Beta* prior *Beta(3, 100)*, the analytical definition of the *Beta* distribution and a more in detail analysis of the choice of its parameters will be discussed when we'll come back to biathlon. After updating our beliefs with a *Binomial* likelihood we compute the posterior for the probability p of developing the illness making use of MCMC, again, the framework of these tools will be presented in the following. We finally compute the efficiency as in equation (1) and summarize our findings in Figure 2.
We can see how the efficiency is indeed peaked around 90%.

Let's now come back at our beloved biathlon and describe a little bit better our analogy,

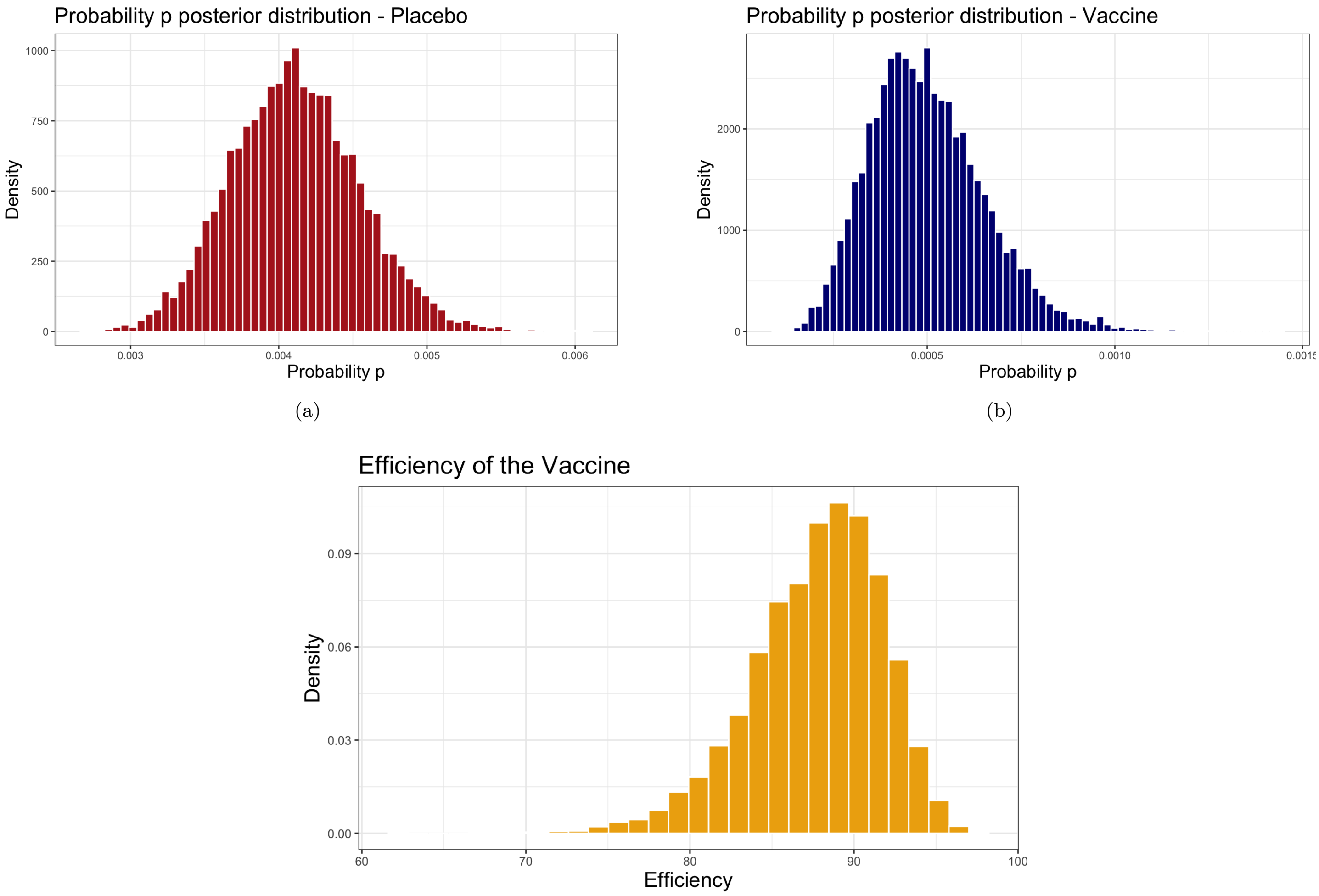
$$\begin{cases} \textbf{Placebo shots} \longmapsto \text{Shots fired when outside the top 10} \\ \textbf{Vaccine shots} \longmapsto \text{Shots fired when in the top 10} \\ \textbf{Clueless volunteer} \longmapsto \text{Clueless shooting target} \end{cases}$$

It should be now clear how the higher our efficiency, the higher the effect entering the last shooting range in the top 10 is on athletes. So it's important to remember that, for our purpose, efficiency is a negative thing!
We wanna introduce now, more in depth, the analytical framework we work in.

The shooting of a target is naturally described by the *Binomial* distribution, which models the number of k successes in a series of n Bernoulli trails; a non negative random variable follows a *Binomial* if its PDF is of the form,

$$PDF(k|n,p) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \tag{2}$$

Where p is the probability of success in a single trial.
For an accurate definition of Bernoulli trials and of the *Binomial* distribution we suggest a quick peak at the Wikipedia page.



(a) (b) (c)

Figure 2: Placebo 2(a) and Vaccine 2(b) probability p posterior distribution. Efficiency 2(c) of the vaccine

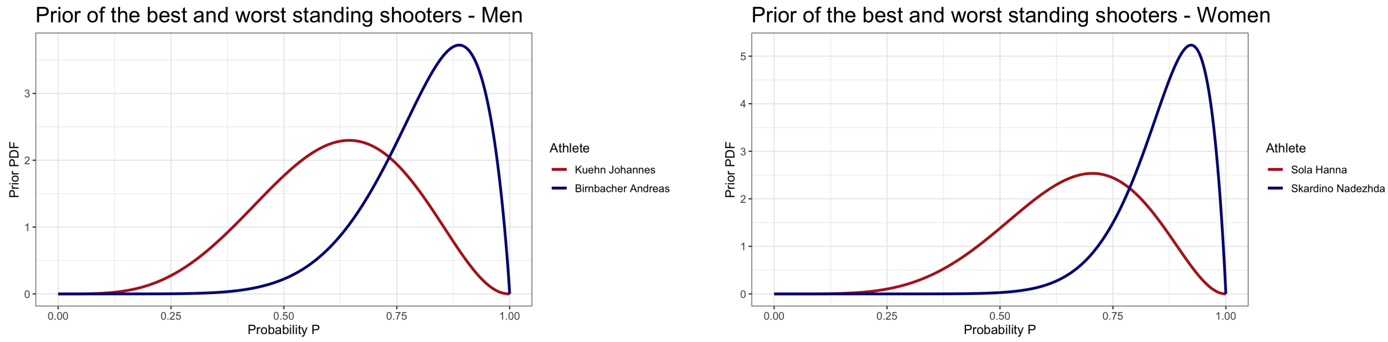
In our case n is 5 for each range, that is, the number of shots fired and k is the number of hits. p is the probability of hitting each shot.
We choose as prior, a *Beta* prior being defined between 0 and 1 and being the conjugate prior to a *Binomial* likelihood, the latter means that the posterior has the same form of the prior. A non negative random variable follows a *Beta* if its *PDF* is of the form,

$$PDF(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} \tag{3}$$

For our purpose we choose the parameters of our *Betas* in the following way. We take for each biathlete the mean and standard deviation of their standing shooting percentage. We don't consider prone shooting as the final shooting range, the one we analyze, is a standing one. Then we fix mean and std to be our mode and standard deviation of the *Beta*.
We didn't use the mean percentage as the mean of the *Beta* because wanted to keep our distribution unimodal as a prior having as mode 1 doesn't reflect our prior knowledge that tell us that there is no such thing as a perfect shooter.
In order to retrieve α from mode and variance we found the root of the correspondent 3^{rd} order equation numerically. We show in Figure 3 the priors fro the best and worst standing shooter in the data set among the one selected, with more than 5 entrances in the top 10.

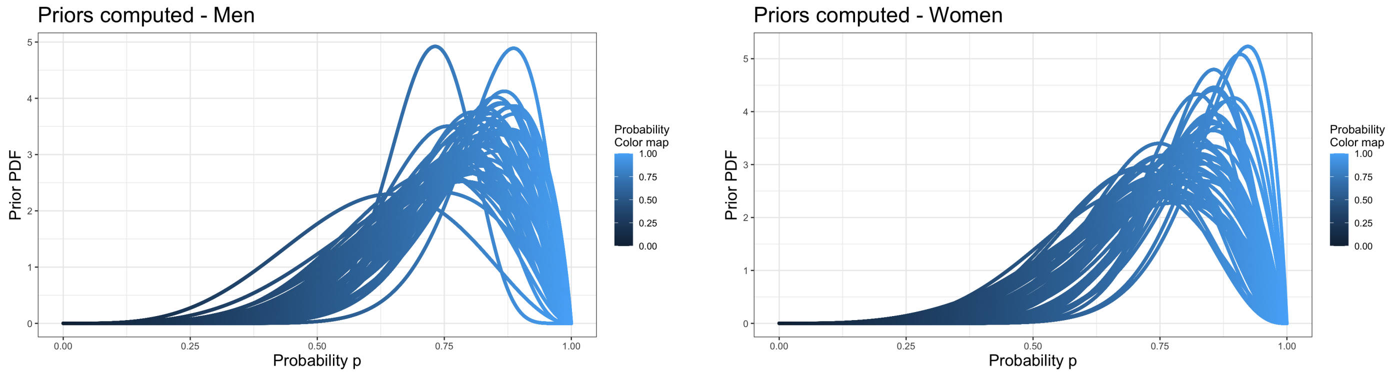
We can see how something we know very well is confirmed in Figure 3, Johannes Khun has never been a standing shooting specialist and Hanna Sola isn't likely to win a shoot off against her teammate Skardino.
All the priors of all the biathletes are instead shown in Figure 4.

It's interesting to notice how, while women have a pretty smooth transition from low to high probabilities, men show a couple of outliers: a low precision but high consistency curve and a low precision low consistency curve being wider then the rest.



(a) (b)

Figure 3: Priors of the best and worst standing shooters in the data set for men 3(a) and women 3(b).



(a) (b)

Figure 4: All the priors for men 4(a) and women 4(b).

Making the assumption of independent shots we can factorize the likelihood and obtain a *Beta* posterior $B(x|\alpha', \beta')$ having the parameters written below.

$$\begin{aligned} \alpha' &= \alpha + y \\ \beta' &= \beta + n - y \end{aligned} \tag{4}$$

Where, for each athlete, α and β are the prior's parameter, y, is the number of hits and n is the number of total attempts.

We proceeded by computing the posterior in this way for each biathlete both for the table of top 10 entrances and for the remaining one, we then sampled 10000 records from the two posteriors and computed the efficiency thanks to equation 1, the results are shown and discussed in the Results section.
To inspect a potential difference in clutchness between men and women we used a Gaussian approximation for the difference of the two efficiencies. For the purpose of this point we analyzed all the athletes together divided by sex in order to retrieve a sort of global efficiency.
To look at the difference between the two efficiency we made use of a Gaussian approximation. We took the 10000 samples as described above and then used their difference as data to put in input of a Markov Chain Monte Carlo.
For the reader not used to work with MCMC we suggest looking at this brief introduction.

The Markov chain algorithm we used was the Gibbs sampling algorithm, exploited by JAGS [2]. The parameters to be inferred were the mean μ of the Gaussian and the precision, which is defined as $\tau = \dfrac{1}{\sigma^2}$.
We used a normal prior for the mean and a lognormal one for the precision τ.

Let's now instead see what was the process used to inspect the last shot in an individual race.

How many medals, podiums or even careers are made or broke in the last shot of an individual race?

Our minds immediately immediately go at the last 2 olympics in which Tarjei Bø in the first and Maxim Tsvetkov in the second had the shot to win gold and they missed it.

We consider here only 4-polygons races and separate individual races from the rest. We then select all the records where athletes hit their first 19 shots and were therefore perfect up to the very end.
We use this two sets by performing a one-sided hypothesis test, we set up the Null hypothesis H_0 as the probability of hitting the last shot is lesser or equal in pursuits and mass starts then in individual. The alternative hypothesis H_1 instead is what we want to prove, that is, the pressure and weight of hitting the last shot in individuals makes percentages much lower.

Now, contrary to what we did in the clutchness part, we don't keep each biathlete separate but we consider all the men together and all the women together.
The prior, that should be next to irrelevant due to the number of records, is again chose to be a *Beta* distribution for all the same reasons we explained before. To retrieve its parameters we follow the same thought process, also here by considering only standing shooting and compute mean and standard deviation of the percentages for all the men and for all the women in individual races.

We proceeded by setting up a threshold that represents the average shooting percentage in pursuits and mass starts computed as total hits over total misses. Our Null hypothesis H_0 is represented by the probability of hitting the last shot in individuals, after a perfect 19 out of 19, being less or equal than this threshold while what we want to prove is that it is greater; our aim is then to reject the Null hypothesis.

Our hypothesis test is a Bayesian one so, after finding the posterior using the update laws in 4, we integrated our posterior for values greater than our threshold. We proceeded by fixing a significance level α and if our integral value is lower than α, we can reject the Null hypothesis with probability $(1-\alpha)\%$ and successfully accept our alternative hypothesis H_1.

Results

Let's now show a couple of results of the efficiency part of the analysis.
We begin by showing the plot the efficiency of the most and least consistent men biathlethes.

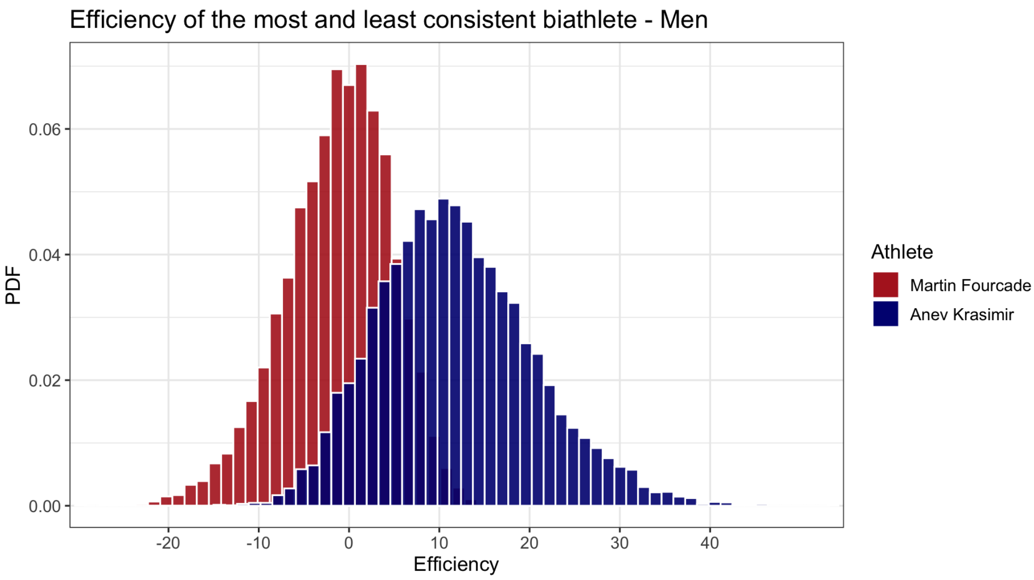
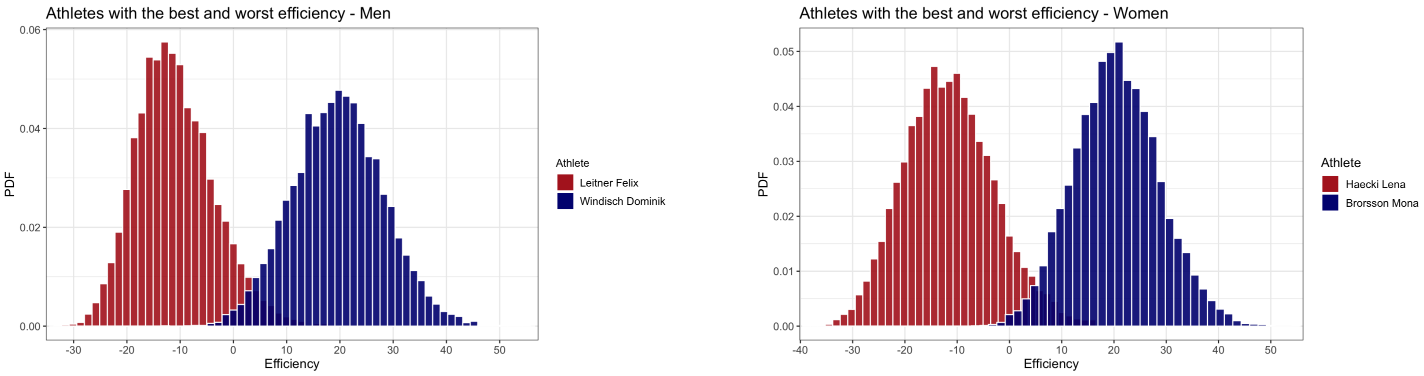


Figure 5: Efficiency of the most and least consistent biathlete

The thing we can immediately notice is how centered around 0 Martin Fourcade's efficiency is. Nonetheless it is something we would expect, there's no one used to the top positions as Monsieur Le Biathlon, as Figure 1(a) certifies.
As far as Anev instead we can see how his percentages worsens when being in the top 10.

Let's transition to more extreme curves that show a more significant improvement or a sign of



(a) (b)

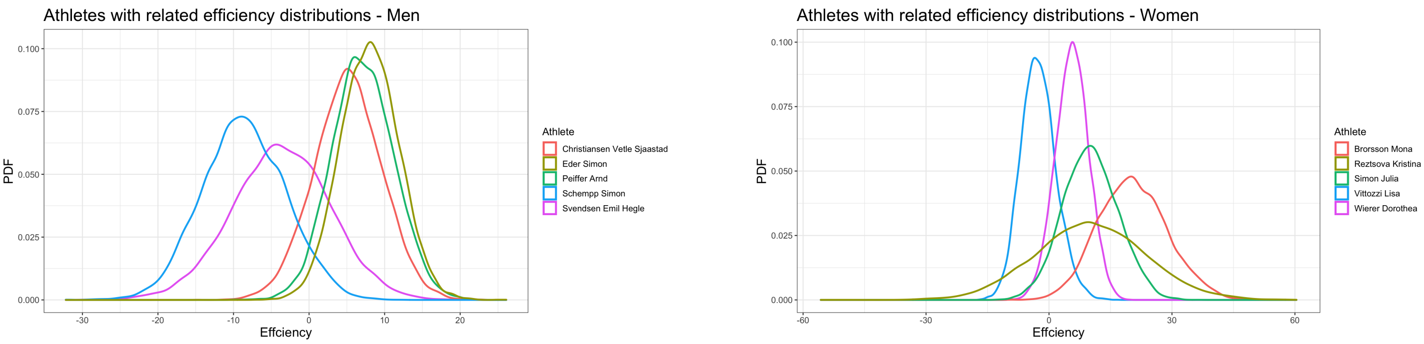
Figure 6: Best and worst efficiency for men 6(a) and women 6(b)

crumbling under pressure in Figure 6.

In Figure 6 we see next to no overlap between the two curves, Felix Leitner definitely performed better when being in contention for a good spot while Dominik lowered his percentages. We all know that Dominik isn't the most consistent but also how he can be devastating in big events, in Oestersund you can probably still hear the echoes of that mass start win [2].

Shifting to the women's side, kudos to Lena Haecki to a very low efficiency. More interestingly though we see confirmation in something we already suspected: if we in fact think at someone that has often lost a podium position in the last shooting Mona Brorsson easily comes to mind. She has gained her first top 3 finish this year in Antholz but many times pressure played a tricky role in her career, the fact that almost the whole efficiency curve is outside 0 is a strong suggestion that the worsening of her percentages when in top 10 has nothing to do with chances.

In Figure 7 we show the density estimate of the efficiency curves for athletes for which we have a very significant sample, considering only racers with tons of word cup starts.



(a) (b)

Figure 7: Efficiency density estimate for big-name biathletes

In men what immediately comes to mind is how high the efficiency is for Simon Eder, we are talking of someone who is going to retire as the athlete with the most 0s in history. Furthermore what is certain is that we have a significant statistic on Simon Eder as he has raced most of the competitions from 2014. Even by non integrating the curve to rely on an analytical support we can conclude that Simon Eder's percentage worsen when he enters the last shooting range in the top 10.
A similar thought process can be applied to Arnd Peiffer while Vetle Christiansen's curve that

[2]Note here how the win in the Oestersund word championship is featured in the data frame of records that represents entrances in the last shooting range outside the top 10. So basically one of the clutchest shooting of all times works against Dominik. In the process of establishing reasonable constraints we had to make some hard cuts and it's possible that contradictory things like that happen. Nonetheless what happened in 2019 can be considered more then a rarity.

continues to suggest the precence of the effect, is more centered around to 0.
Svendsen and Schempp are more shifted toward negative efficiency values, in Svendsen case though the curve includes 0 too much to say anything decisive.
We will return on Christiansen and Svendsen's cases a little bit more later on.

In the women's plot, where Mona Brorsson is represented to provide a comparison, we can see how both Lisa Vittozzi and Dorothea Wierer's bells are sharpely peaked, suggesting standing shooting percentages that, in their slight improvement or worsening, are very consistent.
Kristina Reztsova's instance is pretty interesting due to its very high variability, this is certainly highly favoured by her low number of tryes, but does also suggest a inconsistent performance shooting in and shooting out.

Other then the analysis described in the Methods section, we repeated it making the threshold at the $5-th$ position. So we separated last-range shootings where the biathlete entered in the top 5 from the ones he entered outside of it.
As expected, the general trend shows curves that are similar to the ones obtained in the top 10 analyisis, at times a small shift can be observed, usually to the right.
This doesn't prevent us from noticing a couple of outliers that can be found among all records. In Figure 8 a side-by-side comparison of the top-10 and top-5 case for Emil Svendsen and Vetle Christiansen is shown.

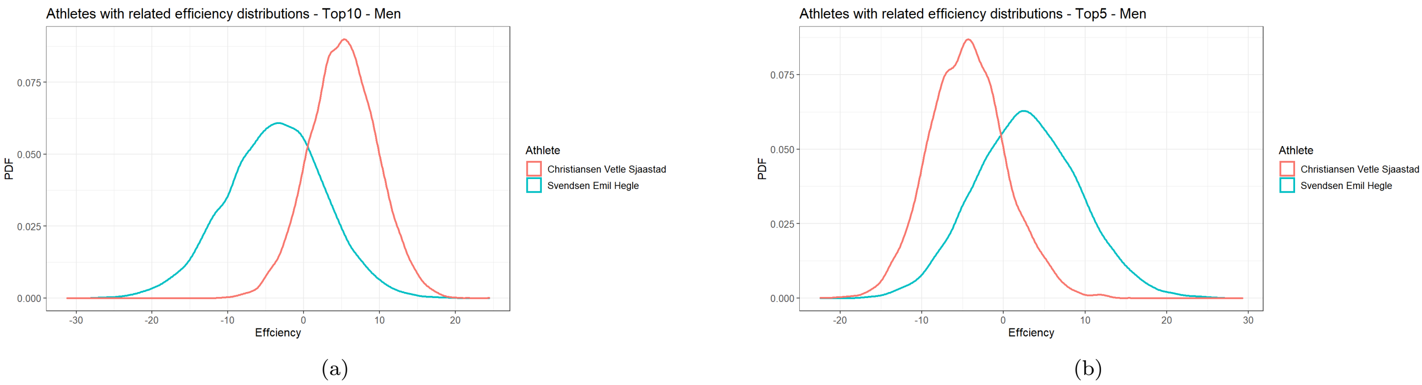


Figure 8: Side by side comparison of Christiansen and Svendsen's top 5 and top 10 performances

In the Figure above a big change from the common behaviour can be seen. While the top 10 comparison suggests an Emil Svendsen that is, so to speak, clutcher than Christiansen, the top 5 study tells the opposite. This tells us mainly two things: the first is that Svendsen performs better when he is between the $6-th$ and the $10-th$ position while Christiansen as an opposite behaviour.
The second one is that no definitive conclusion should be made without clear evidence: only if the efficiency curve doesn't embeds 0 we should make any claim an effect of the "vaccine".
A more complete and in depth analysis should provide a proper hypothesis test for each of the tested competitors to determine, at some fixed confidence level, if the value of the efficiency can be attributed to chance alone.

Looking for a difference in clutchness based on sex we do see a slight difference, with men performing a little bit better when in the top 10, so to speak. The discrepancy is though very narrow and, as it can be seen in Figure 9(c) every possible hypothesis test would state that the difference can be attributed to chance alone.
As described in the Methods section, we computed the parameters of the Gaussian approximation for the difference between the two efficiencies thanks to MCMC. The posteriors we got are shown in Figure 9(a) and 9(b). We can see their pretty nice convergence and unimodal behaviour. The 4 chains shown in Figure 10 testimony instead how the burn in chosen was proper and that they do not show any auto-correlation behaviour.

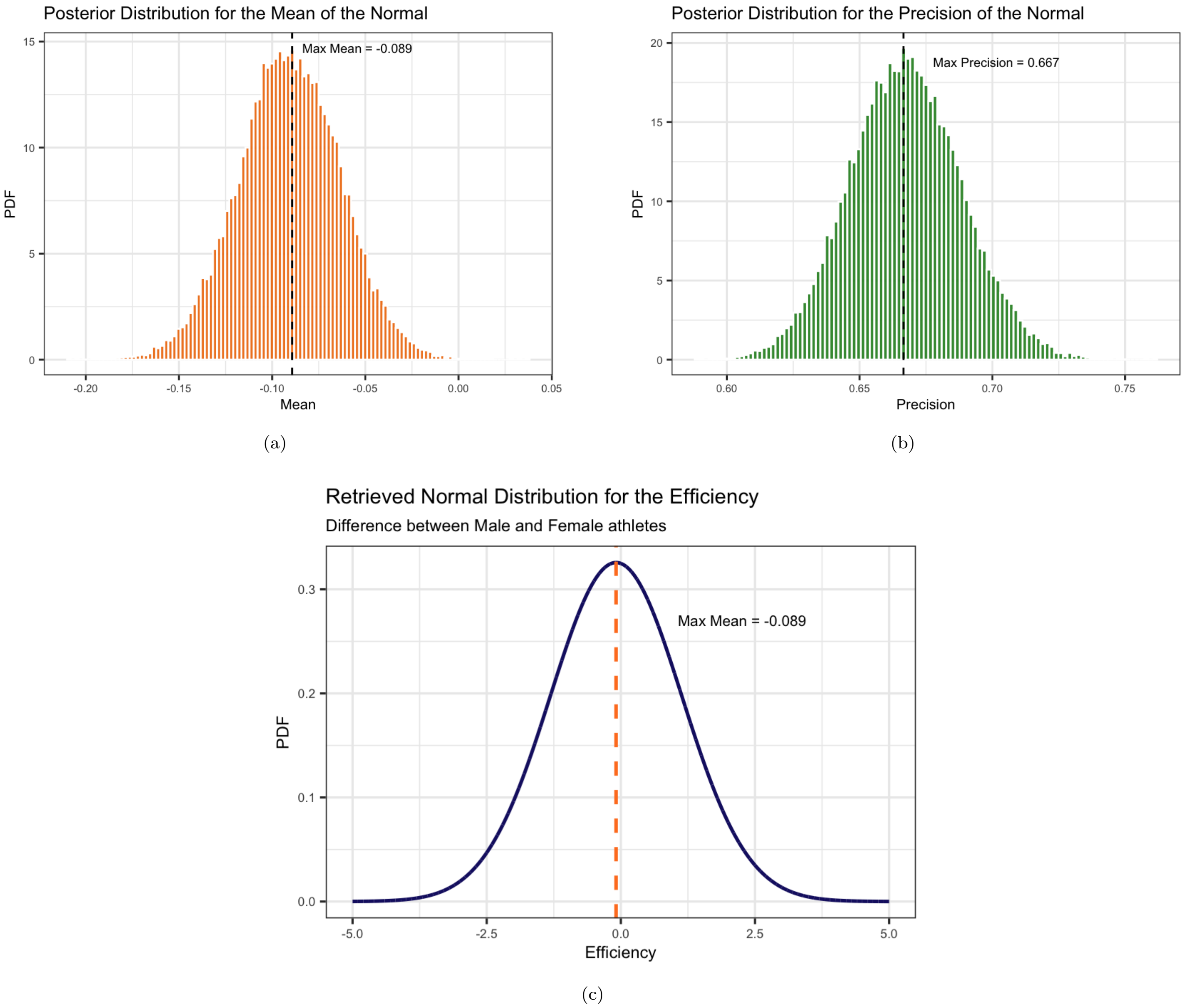


Figure 9: Posteriors and results of our MCMC analysis.

We can now shift to the infamous last shot in an individual race.
In Figure 11 the result of our one-sided hypothesis test is shown and a clear-cut result stands out.

This is indeed the most clear and significant result of the whole analysis. We briefly restate the goal: our Null hypotesis N_0 we wanna disproof is the fact that the last shot in an individual is not more difficult than the last one in mass starts and pursuits (when hitting the first 19). The alternative hypothesis N_1, what we wanna show, is that the last shot in an individual is the most difficult.

Integrating the part of the posterior curve that is beyond the threshold described in the Methods section we obtain a value of 0.007 for men and 0.002 for women.
The final result is, that we can reject the Null hypothesis at a 99% significance level for both men and women so, we can say that with 99% probability the last shot in an individual is the toughest.

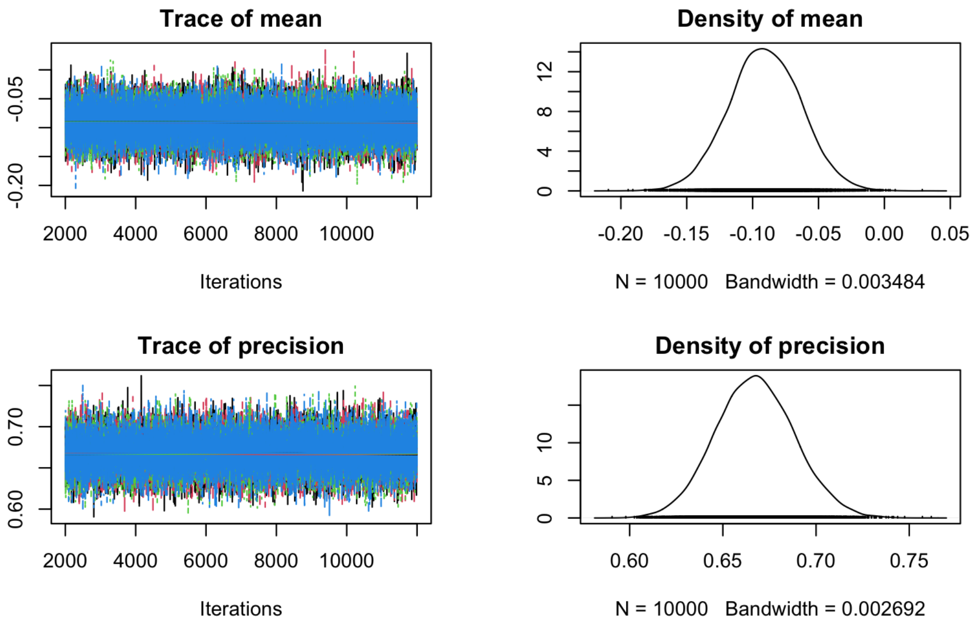
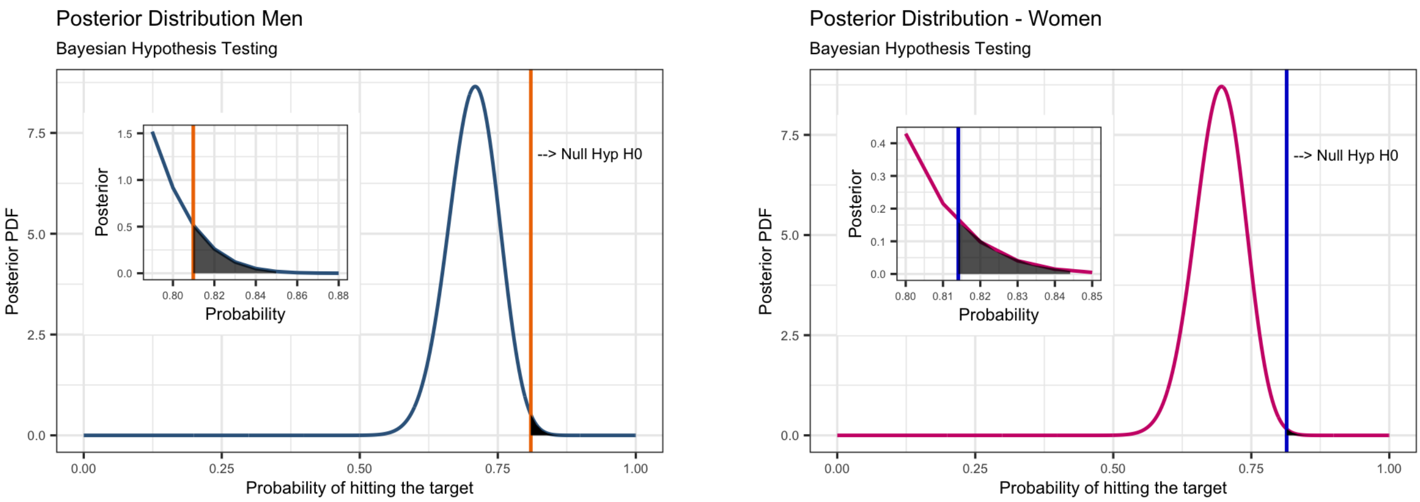


Figure 10: MCMC result's outline plotted thanks to the *Coda*[4] library



(a) (b)

Figure 11: Hypothesis test result for women (right) and men (left)

1 Conclusions

In this work we have analyzed, thanks to Bayesian inference, biathlon shooting.
In the first part we tried to seek for any change in performance among each athlete depending on the importance of the last shooting range.
Interestingly we found out how some racers improve their percentages while other lower them. We discovered also great shooters can suffer and obtained confirmation on the reputation of word cup veterans.
This first branch was developed differently for every competitor, merging a little bit the results seeking for discrepancy between men's and women's behaviour, which we did not found.

In the later part we made use of a one-sided hypothesis test to investigate individual's shooting, the result obtained was that with 99% probability the last shot in an individual race is the most difficult between all formats.

References

[1] Datacamp. R documentation;. [Version 2.0]. Available from: "https://www.rdocumentation.

org".

[2] JAGS. JAGS source code;. Available from: "https://mcmc-jags.sourceforge.io".

[3] Cambridge University. BUGS project;. Available from: "https://www.mrc-bsu.cam.ac.uk/software/bugs/".

[4] Martyn Plummer. Coda library's documentation;. Available from: "https://www.rdocumentation.org/packages/coda/versions/0.19-4".