Stein's Estimator on NBA players' data

Giovanni Mantovani, Santiago Lunque, Marco Lomele

Background

Maximum Likelihood Estimators. When estimating the parameters of a determined distribution, Maximum Likelihood Estimators (MLEs) are often the choice of preference. Under regularity conditions, they are asymptotically unbiased, consistent, and efficient. The idea under which MLEs operate is greatly intuitive: seek the parameter value that maximises the likelihood function, which measures how well the model explains the observed data. Given both functions reach their maximum at the same point, in practice, the MLEs seek to maximise the log-likelihood function:

$$\ell(\theta|X) = logL(\theta|X)$$

In other words, MLEs answer the following question: given the data, what is the most probable value of the parameter? It is therefore intuitive to think that MLEs are the best estimators independently of the dimensional space being evaluated.

James-Stein Estimators. However, in 1961 the latest version of Stein's estimator was proposed, often referred to as the James-Stein estimator (JSE). This estimator achieves a lower mean square error (MSE) than MLEs when observing data points coming from three or more distributions. To explain the idea behind it, assume the parameters of interest are the population means. Instead of proposing the MLE for each of the three means, it is better to "shrink" them, as the JSE proposes.

$$\widehat{\theta}_{Stein} = (1 - \frac{(p-2)\boldsymbol{\sigma}^2}{\|X\|^2}) \times X \quad for\, p \geq 3$$

Geometrically, JSE "shrinks" the data points toward the origin. For example, with three normal distributions of equal variance, the three-dimensional data points form a spherical shape centred on the means, indicated by the red dot in Figure 1. Shrinkage moves points near the origin (the "near end") away from the true means, while points farther out (the "far end") move closer to the true means. But how can one confirm that the "far end" outweighs the "near end"?

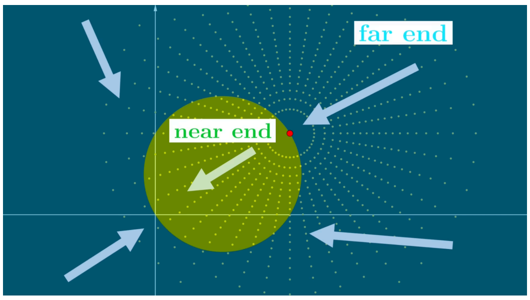

Figure 1: Near end and far end areas. Source: https://www.youtube.com/watch?v=cUqoHQDinCM&t=978s

In one and two dimensions, it is not always true that the "near end" is less than the "far end". From three dimensions onward, the "far end" becomes larger than the "near end". Thus, the amount of points moved towards the means is greater than the amount of points moved away from the mean. As a result, the James-Stein estimator becomes evidently better than MLEs.

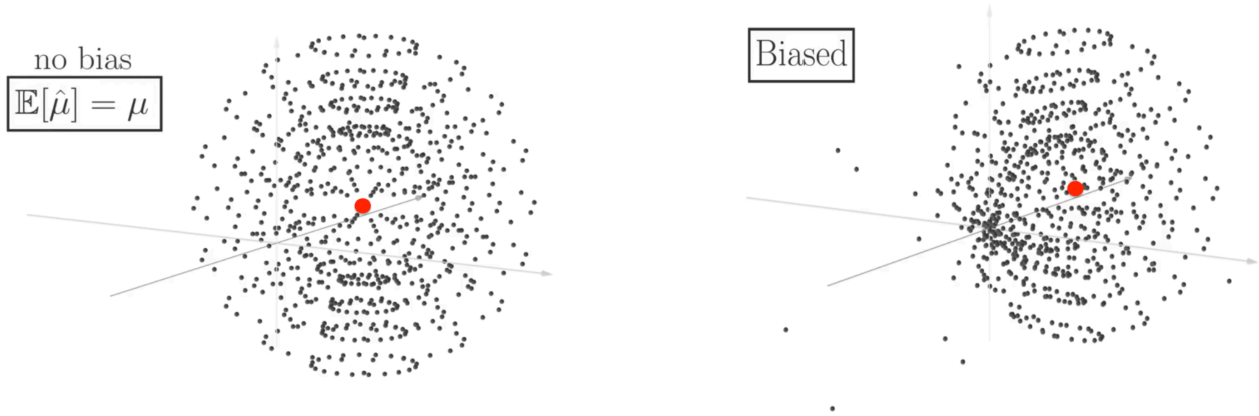

Figure 2: Shrinkage. Source: https://youtu.be/cUqoHQDinCM?si=HsWUkomepv6cI6cQ

Application. A basketball team traditionally consists of five distinct positions, each having different physical characteristics:

- the point guard (PG), typically the shortest (175-188 cm) and lightest (75-85 kg) player, who orchestrates the offence;
- the shooting guard (SG), slightly taller (188-198 cm) but similarly lean (80-90 kg), specialising in perimeter scoring;
- the small forward (SF), whose moderate height (198-205 cm) and athletic build (95-105 kg) enable versatility in both offence and defence;
- the power forward (PF), who combines significant height (205-210 cm) with substantial mass (105-115 kg) to control the paint and rebound;
- the centre (C), the tallest (208-215 cm) and heaviest (110-125 kg) player, who protects the rim and dominates the area near the basket.

Research

Goal. The aim is to illustrate the admissibility of Stein's estimator using player data from the National Basketball Association (NBA). Imagine a situation where the players' mean height and mean weight per position are not known. This information interests upcoming players who need to decide which position to aim for, coaches who need to allocate players effectively, and fans who might be looking for insights. The following example shows how Stein's estimator provides more accurate results compared to standard Maximum Likelihood Estimation.

Procedure. The first step is to perform exploratory data analysis. The data, taken from Kaggle, describes NBA players' height and weight. Data is then grouped by position in order to plot multiple histograms.

Figure 3 shows the distributions of the heights and weights per position, which appear to be normally distributed. However, heteroskedasticity is present in the data, meaning that the variance differs across the distributions of the positions. For practicality, it is assumed that this difference is negligible.

Secondly, samples of 50 observations from each basketball position are taken. This resembles what a researcher would do if the true mean height and mean weight were not available.

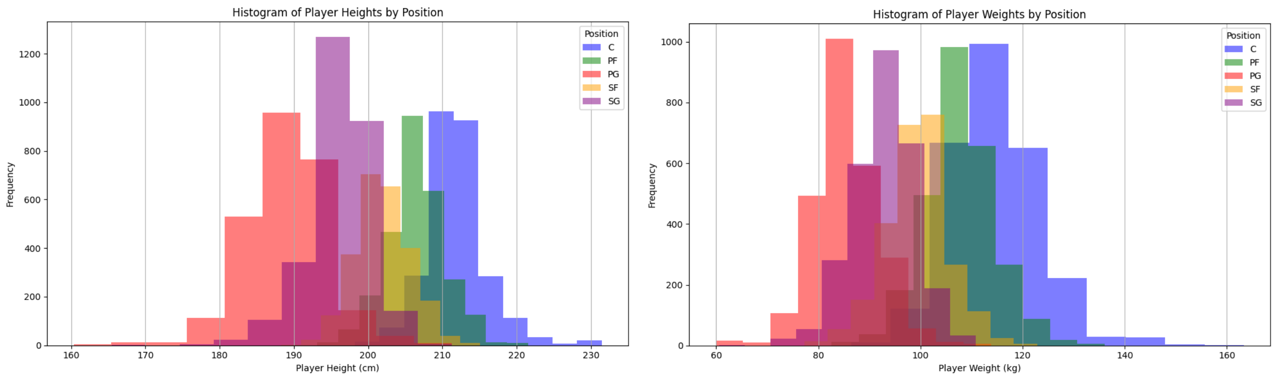

Figure 3: Population distribution of heights and weights by basketball position.

Furthermore, the true mean of each variable, per position, is estimated using a sample mean, the equivalent of taking the MLE. Then, Stein's shrinkage factor is computed and applied to the sample means estimated from the MLE. This way it's possible to obtain the JS estimates.

Finally, given that the true mean height and true mean weight are known, it's possible to compute the MSE for both estimates to compare the performance of the two estimation methods.

Results

	Height		**Weight**	
	MLE	*JS*	*MLE*	*JS*
Mean Squared Error	260.786759	258.565291	519.218052	514.046988
Difference	0.85%		0.99%	

Table 1. MSE comparison between ML and JS

As shown in Table 1, the JS estimator achieves a lower MSE than that of MLEs. In particular, the MSE was reduced by 0.85% and 0.99% for height and weight, respectively.

Conclusion

James-Stein proved that in large dimensions his estimator dominantes MLEs. In this example, however, it is important to note that this fact does not imply that for each position the JS estimator dominates MLEs. It could occur that, for a specific position, the MLE is closer to the true mean than the JS estimator.

Nevertheless, JS estimation is still highly practical. In contexts that deal with high dimensionality, like finance, genomics, or image processing, Stein's shrinkage-based approach improves parameter estimation compared to the MLE at the expense of introducing some bias.