Mathematical Statistics II

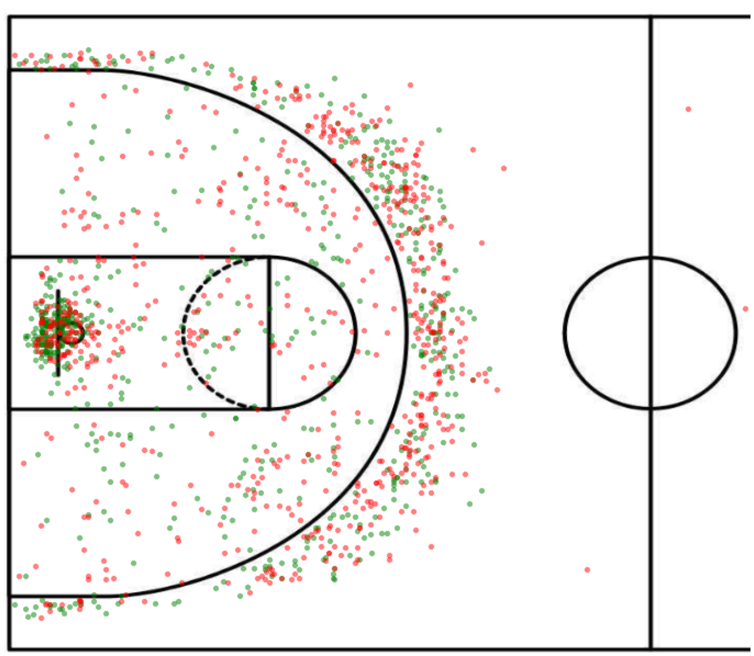
Statistical Computing Activity: Module 2

One purpose of these Statistical Computing Activities (SCAs) is to give you a chance to explore statistics when the random variables do not follow a Normal distribution. Another purpose is to give you more skills in thinking about the randomness that is life.

Usually, like here, these SCAs will have a theme and several problems dealing with that theme or purpose. The reason for that extra layer of complexity is to tie what we do in the class with what we can use these techniques for in our lives as statisticians and/or consultants and/or full members of a democratic society.

⋆ ⋆ ⋆

What to submit. On the due date, please submit a LaTeX-typed. Each problem starts on a new page. Please include the entire code at the end of the submission in an appendix.



The ability of a basketball player to make 3-point shots is an important skill; it helps the team win games. As an advisor for the Portland Trail Blazers, I find myself in need to estimate the proportion of times that Patrick Williams will make a 3-point shot. I need to know this so that I can estimate his value to the team should we be able to lure him from the Chicago Bulls.

Designs of Experiment. To estimate this, I have a few options for designing my experiment. First, I could have him shoot until he first misses, then estimate π from that information. Second, I could have him shoot 10 times and use the number of made shots to estimate π. Third, I could just guess that $\pi = 0.25$.

Note that these three methods (data-collection schemes) use different distributions. The first uses a Geometric distribution; the second, a Binomial distribution; the third, a Dirac distribution. Also note that I will be performing this experiment only once ($n = 1$).

⋆⋆⋆

Problem 1: Estimators

As each of the three methods utilizes a different distribution, the estimators between them may differ. By hand, determine the maximum likelihood estimator for the first scheme, the method of moments estimator for the second scheme, and the obvious estimator for the third scheme. There is no `R` for this problem.

For reference, the two probability mass functions are

$$p_1(x) = \pi(1-\pi)^x \qquad\qquad x \in \left\{0, 1, 2, \ldots\right\}$$

$$p_2(x) = \binom{10}{x}\pi^x(1-\pi)^{10-x} \qquad\qquad x \in \left\{0, 1, \ldots, 10\right\}$$

So, for this problem, you will need to calculate $\hat{\pi}_1$, $\hat{\pi}_2$, and $\hat{\pi}_3$.

Problem 2: Bias

By hand, estimate the bias of each of the three estimators as a function of π. That is, determine $\mathbb{E}\left[\hat{\pi}_1 - \pi\right]$, $\mathbb{E}\left[\hat{\pi}_2 - \pi\right]$, and $\mathbb{E}\left[\hat{\pi}_3 - \pi\right]$. Note that you can calculate the exact bias for estimators 2 and 3, but that you can only approximate it for estimator 1. There is no `R` for this problem.

Problem 3: Mean Square Error

Again, by hand, determine the bias of each of the three estimators as a function of π. That is, determine $\mathbb{E}\left[(\hat{\pi}_1 - \pi)^2\right]$, $\mathbb{E}\left[(\hat{\pi}_2 - \pi)^2\right]$, and $\mathbb{E}\left[(\hat{\pi}_3 - \pi)^2\right]$. Again, while you can calculate the MSE exactly for estimators 2 and 3, you may have to estimate it for estimator 1. There is no `R` for this problem.

Problem 4: Interpretation

Which of the three estimators is best according to the usual MSE criteria? You should provide a graphic showing the MSE values for all possible values of π. This will make comparisons much easier, allowing you to state something like

> When the population proportion is large, one should use $\hat{\pi}_r$. When the population proportion is between approximately 40% and 80%, one should use $\hat{\pi}_f$. At no time should one use $\hat{\pi}_q$, because its MSE is never the lowest.

This problem *will* require `R` for the graphic and for calculating the MSE for "all possible values" of π.

Now that you have given a mathematical/statistical reason for your estimator choices, give a non-mathematical/non-statistical reason why one would not want to use the estimator.