1 Introduction

In the homework for week 6, you explored the beta distribution. The beta distribution is a continuous distribution that is used to model a random variable X that ranges from 0 to 1, making it useful for modeling proportions, probabilities, or rates. The beta distribution is also known for being remarkably flexible with regards to its shape – it can be left-skewed, right-skewed, or symmetric depending on the value of the parameters that define its shape: $\alpha > 0$ and $\beta > 0$.

The beta distribution's probability density function is given by

$$f_X(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(x \in [0,1]),$$

where $I(x \in [0,1]) = 1$ when $x \in [0,1]$ and 0 otherwise. Note that this simply makes the probability density function zero where x cannot possibly take on those values, i.e., everywhere outside of [0,1].

In Lab 7, we considered four cases of the beta distribution:

- Beta($\alpha = 2$, $\beta = 5$)
- Beta($\alpha = 5$, $\beta = 5$)
- Beta($\alpha = 5$, $\beta = 2$)
- Beta($\alpha = 0.50$, $\beta = 0.50$)

You computed the population moments using numerical integration, ensuring those values matched the calculations using the formulas derived by mathematical statisticians.

$$E(X) = \frac{\alpha}{\alpha+\beta} \quad \text{(The Mean)}$$

$$var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad \text{(The Variance)}$$

$$skew(X) = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}} \quad \text{(The Skewness)}$$

$$kurt(X) = \frac{6\left[(\alpha-\beta)^2(\alpha+\beta+1) - \alpha\beta(\alpha+\beta+2)\right]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} \quad \text{(The Excess Kurtosis)}$$

Further, we interpreted these quantities in plots of the probability density functions for all four cases.

We demonstrated that our graphical and numerical summaries of random samples (data) connect to the actual population distribution. Specifically, you showed that a histogram and the estimated density were *very* close to the probability density function for each sample and, as a result, the sample mean, variance, skewness, and kurtosis are close to the population-level quantities.

Of particular import, we showed that the cumulative statistics converge to the true values that describe the actual population distribution. While the variability is high for small samples, the statistics converge to the true values and the variability decreases as the sample size increases.

Now, onto lab eight!

2 Task Six: Collect and Clean Data

Fatih (2022) suggest country death rates worldwide can be modeled with a beta distribution. Collect the data from the World Bank.[1] We will focus on 2022, the most recent year available. Note that the beta distribution has support on (0,1), so you will have to convert the data from the number of deaths per 1000 citizens to a rate (i.e., `x = number/1000`)

[1] https://data.worldbank.org/indicator/SP.DYN.CDRT.IN

3 Task Seven: What are α and β?

Compute the method of moments estimates and maximum likelihood estimates for these data. Plot a histogram of the data with the estimated method of moments and maximum likelihood distributions superimposed.
Note: Do the computations the way we have in class. Do not use the formulas below, which are just for demonstration.

MOM Scratchwork: For finding the MOM estimator, we need the first two moments:

$$\begin{aligned} E(X) &= \frac{\alpha}{\alpha+\beta} \\ E(X^2) &= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}. \end{aligned}$$

Then, we can set up the approximate system of equations:

$$\begin{aligned} \overline{X} &= \frac{\alpha}{\alpha+\beta} \\ \overline{X^2} &= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}. \end{aligned}$$

From the first equation, we get

$$\begin{aligned} \overline{X} &= \frac{\alpha}{\alpha+\beta} \\ \overline{X}(\alpha+\beta) &= \alpha \\ \overline{X}\alpha + \overline{X}\beta &= \alpha \\ \overline{X}\beta &= \alpha - \overline{X}\alpha \\ \beta &= \frac{\alpha}{\overline{X}} - \alpha. \end{aligned}$$

Plugging into the second equation, we get

$$\begin{aligned} \overline{X^2} &= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)\,(\alpha+\beta)} \\ \overline{X^2} &= \frac{(\alpha+1)\alpha}{\left(\alpha+\left(\frac{\alpha}{\overline{X}}-\alpha\right)+1\right)\left(\alpha+\left(\frac{\alpha}{\overline{X}}-\alpha\right)\right)} \\ \overline{X^2} &= \frac{(\alpha+1)\alpha}{\left(\frac{\alpha}{\overline{X}}+1\right)\left(\frac{\alpha}{\overline{X}}\right)} \\ \overline{X^2} &= \frac{(\alpha+1)}{\frac{1}{\overline{X}}\left(\frac{\alpha}{\overline{X}}+1\right)} \\ \overline{X^2} &= \frac{(\alpha+1)}{\frac{\alpha+\bar{X}}{\overline{X}^2}} \\ \overline{X^2} &= \frac{\overline{X}^2(\alpha+1)}{\alpha+\bar{X}} \\ \overline{X^2}(\alpha+\bar{X}) &= \overline{X}^2(\alpha+1) \\ \overline{X^2}\alpha + \overline{X^2}\bar{X} &= \overline{X}^2\alpha + \overline{X}^2 \\ \overline{X^2}\alpha - \overline{X}^2\alpha &= \overline{X}^2 - \overline{X^2}\bar{X} \\ \alpha &= \frac{\overline{X}^2 - \overline{X^2}\bar{X}}{\overline{X^2} - \overline{X}^2}. \end{aligned}$$

And plugging back,

$$\begin{aligned}
\beta &= \frac{\left(\frac{\overline{X}^2 - \overline{X^2}\bar{X}}{\overline{X^2} - \overline{X}^2}\right)}{\overline{X}} - \left(\frac{\overline{X}^2 - \overline{X^2}\bar{X}}{\overline{X^2} - \overline{X}^2}\right) \\
&= \left(\frac{\overline{X}^2 - \overline{X^2}\bar{X}}{\overline{X^2} - \overline{X}^2}\right)\left(\frac{1}{\overline{X}} - 1\right)
\end{aligned}$$

I believe you get a different set of estimators if you solve for α first, but I leave that exercise for the interested student.

MLE Scratchwork: The likelihood function is

$$\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{x}) &= \prod_{i=1}^{n} f_X(x_i|\boldsymbol{\theta}) \\
&= \prod_{i=1}^{n} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1} \\
&= \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^n \left(\prod_{i=1}^{n} x_i^{\alpha-1}(1-x_i)^{\beta-1}\right),
\end{aligned}$$

where $\boldsymbol{\theta} = \{\alpha, \beta\}$.

For computational (and theoretical ease), we can use the loglikelihood function.

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\mathbf{x}) &= \ln\left(\prod_{i=1}^{n} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1}\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1}\right) \\
&= \sum_{i=1}^{n} \ln\left(\Gamma(\alpha+\beta)\right) - \ln\left(\Gamma(\alpha)\right) - \ln\left(\Gamma(\beta)\right) + (\alpha-1)\ln(x_i) + (\beta-1)\ln(1-x_i)
\end{aligned}$$

Interestingly, unless α or β is fixed at some convenient number (e.g., 1) this is not a trivial optimization problem with respect to $\boldsymbol{\theta} = \{\alpha, \beta\}$.

The big takeaway here is that the MOM and MLE procedures can be *very* challenging! Having a computational approach to complete these tasks is necessary for many real-world situations.

4 Task Eight: Which estimators should we use?

Let $\alpha = 8$ and $\beta = 950$. Write a `for()` loop to simulate new data ($n = 266$). Make sure to use `set.seed(7272+i)` to ensure we all work with the same samples. At each iteration (1:1000), compute and store the method of moments estimates and maximum likelihood estimates. The result is a sample of $n = 1000$ method of moments estimates for α and β, and maximum likelihood estimates for α and β. Just as we summarize data, we can summarize these estimates to see their distribution.

Plot the estimated densities for the method of moments and maximum likelihood estimates using `geom_density()` in a 2×2 grid. What do you notice about these plots? Compute the bias, precision, and mean squared error for the estimates. Report them in a table. Can you "see" these values in your plots? Use these values to choose whether we should use the MOMs or MLEs.

$$\text{bias} = E(\hat{\theta}) - \theta$$
$$\text{precision} = \frac{1}{\text{var}(\hat{\theta})}$$
$$\text{MSE} = \text{var}(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2$$

That is, for a vector of estimates `theta.hats` where the true value is `theta`, we can compute these in `R` as follows.

```
(bias <- mean(theta.hats) - theta)
(precision <- 1/var(theta.hats))
(mse <- var(theta.hats) + bias^2)
```

5 Notes about the write up

When you write up labs 7-8, think about it as an exercise of describing the beta distribution – almost a cheatsheet for the distribution. What is the beta distribution? What does it look like? What is it used for? What are its properties (you derived several of them)? What additional information do we gain from the simulations and real data analysis? The questions I've provided in labs 7-8 are aimed to guide your write up.

Do a better job than wolfram, and tell a better story than statlect. Also, this may not be helpful, but look at this cool report from NASA.

A possible sections setup could be:

1. Introduction
2. Density Functions and Parameters
3. Properties (i.e., mean, variance, etc)
4. Estimators (note these are based on samples)
5. Example (i.e., death rates data)

6 Additional Optional Coding Challenege

The `gganimate` package for `R` (Pedersen and Robinson, 2024) enables us to create animations using `ggplot2` (Wickham, 2016).

Create a plot showing how each estimator's histogram changes with sample size. That is, redo task eight for sample sizes on `seq(1,500,10)`.

References

Fatih, C. (2022). Determinants of mortality rates from COVID-19: A macro level analysis by extended-beta regression model. *Revista de Salud Pública*, 24(2):1.

Pedersen, T. L. and Robinson, D. (2024). *gganimate: A Grammar of Animated Graphics*. R package version 1.0.9.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.