Converting Raw Moments to Central Moments

STA 325 - Particle Clustering Project

2025-10-15

Goal

Convert raw moments $E[X]$, $E[X^2]$, $E[X^3]$, $E[X^4]$ to interpretable summary statistics:

- **Mean**: $\mu = E[X]$
- **Variance**: $\sigma^2 = E[X^2] - \mu^2$
- **Skewness**: $\gamma = E[(X - \mu)^3]/\sigma^3$
- **Kurtosis**: $\kappa = E[(X - \mu)^4]/\sigma^4$

These will be the response variables for our machine learning models.

Load Data

```
library(tidyverse)

# Load training data
train_data <- read_csv("data-train.csv")

# Preview
head(train_data)
```

```
## # A tibble: 6 x 7
##      St    Re      Fr R_moment_1 R_moment_2 R_moment_3 R_moment_4
##   <dbl> <dbl>   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1  0.1    224   0.052   0.00216     0.130       14.4      1586.
## 2  3      224   0.052   0.00379     0.470       69.9     10404
## 3  0.7    224 Inf       0.00291     0.0435       0.822      15.6
## 4  0.05    90 Inf       0.0635      0.0907       0.467       3.27
## 5  0.7    398 Inf       0.000369    0.00622      0.126       2.57
## 6  2       90   0.3     0.148       2.01        36.2       672.
```

Calculate Summary Statistics

Formulas

For a random variable X with raw moments $m_1 = E[X]$, $m_2 = E[X^2]$, $m_3 = E[X^3]$, $m_4 = E[X^4]$:

Variance:

$$\sigma^2 = E[X^2] - \mu^2$$

Third Central Moment:

$$E[(X-\mu)^3] = E[X^3] - 3\mu E[X^2] + 2\mu^3$$

Skewness:

$$\gamma = \frac{E[(X-\mu)^3]}{\sigma^3}$$

Fourth Central Moment:

$$E[(X-\mu)^4] = E[X^4] - 4\mu E[X^3] + 6\mu^2 E[X^2] - 3\mu^4$$

Kurtosis:

$$\kappa = \frac{E[(X-\mu)^4]}{\sigma^4}$$

Implementation

```
train_with_stats <- train_data %>%
  mutate(
    # Mean (first raw moment)
    mean = R_moment_1,

    # Variance (second central moment)
    variance = R_moment_2 - R_moment_1^2,
    sd = sqrt(variance),

    # Third central moment
    mu3 = R_moment_3 - 3*R_moment_1*R_moment_2 + 2*R_moment_1^3,

    # Skewness (standardized third central moment)
    skewness = mu3 / (sd^3),

    # Fourth central moment
    mu4 = R_moment_4 - 4*R_moment_1*R_moment_3 +
          6*R_moment_1^2*R_moment_2 - 3*R_moment_1^4,

    # Kurtosis (standardized fourth central moment)
    kurtosis = mu4 / (sd^4)
  ) %>%
  # Remove intermediate calculations
  select(-mu3, -mu4)
```

Example Calculation

First row of training data:

```
# Raw moments
cat("Raw moments:\n")
```

```
## Raw moments:
```

```
cat("  E[X]    =", train_data$R_moment_1[1], "\n")
```

```
##   E[X]    = 0.002157
```

```
cat("  E[X^2]  =", train_data$R_moment_2[1], "\n")
```

```
##   E[X^2]  = 0.13035
```

```
cat("  E[X^3]  =", train_data$R_moment_3[1], "\n")
```

```
##   E[X^3]  = 14.374
```

```
cat("  E[X^4]  =", train_data$R_moment_4[1], "\n\n")
```

```
##   E[X^4]  = 1586.5
```

```
# Calculated statistics
cat("Calculated statistics:\n")
```

```
## Calculated statistics:
```

```
cat("  Mean:      ", train_with_stats$mean[1], "\n")
```

```
##   Mean:       0.002157
```

```
cat("  Variance:  ", train_with_stats$variance[1], "\n")
```

```
##   Variance:   0.1303453
```

```
cat("  Std Dev:   ", train_with_stats$sd[1], "\n")
```

```
##   Std Dev:    0.3610337
```

```
cat("  Skewness:  ", train_with_stats$skewness[1], "\n")
```

```
##   Skewness:   305.428
```

```
cat("  Kurtosis:  ", train_with_stats$kurtosis[1], "\n")
```

```
##   Kurtosis:   93371.66
```

Summary of Calculated Measures

```
train_with_stats %>%
  select(mean, variance, skewness, kurtosis) %>%
  summary()
```

```
##       mean             variance            skewness         kurtosis
##  Min.   :0.000222   Min.   :   0.0001   Min.   : 11.97   Min.   :   150.5
##  1st Qu.:0.002157   1st Qu.:   0.0245   1st Qu.: 72.55   1st Qu.:  5622.3
##  Median :0.002958   Median :   0.0808   Median :110.12   Median : 12158.7
##  Mean   :0.040394   Mean   :  92.4855   Mean   :162.81   Mean   : 39749.6
##  3rd Qu.:0.087868   3rd Qu.:   0.5268   3rd Qu.:269.54   3rd Qu.: 72732.4
##  Max.   :0.172340   Max.   :1044.2759   Max.   :344.91   Max.   :132136.7
```

Save Processed Data

```
write_csv(train_with_stats, "data-train-processed.csv")
cat("Saved: data-train-processed.csv\n")
```

```
## Saved: data-train-processed.csv
```

The processed file contains all original columns plus the calculated statistics: `mean`, `variance`, `sd`, `skewness`, `kurtosis`.