Report: Analysis of Independence between Sample Mean and Coefficient of Variation

Introduction

we analyze the independence between the sample mean and the coefficient of variation (CV). If the sample mean and CV are independent, the underlying data distribution must be a Gamma distribution. This independence is unique to the Gamma distribution and does not apply to other common distributions.

In the context of identifying heterogeneity in SAR imaging data, two main models are used: the Gamma SAR distribution and the G_I^0 distribution. The Gamma SAR distribution is easier to handle and allows homogeneous areas to be identified without having to resort to the G_I^0 distribution.

Methodology

To evaluate this property, we simulated data from both the Gamma SAR and G_I^0 distributions. The goal is to determine if the mean and CV are independent for these distributions.

Simulations

The following R code was used to generate 10000 samples for each distribution and analyze the independence between the mean and CV.

Gamma SAR Distribution

For the Gamma SAR distribution, we simulated the data and plotted the CV versus the mean.

G_I^0 Distribution

For the G_I^0 distribution, we also simulated the data and plotted the CV versus the mean.

The scatter plot for the Gamma SAR distribution shows a random distribution of points, indicating that the mean and CV are independent. This confirms the property of the Gamma distribution where the sample mean and CV do not reveal any correlation.

In contrast, the scatter plot for the G_I^0 distribution shows a log-like pattern, indicating dependence between the mean and CV. This dependency is due to the parameter α in the G_I^0 distribution, which measures the roughness of the data. When α approaches $-\infty$, the G_I^0 distribution approximates the Gamma SAR distribution, and the mean and CV become independent. However, for values such as $\alpha = -1.5$, the mean and CV are dependent.

Model Fitting

To model the dependency observed in the G_I^0 distribution, we used the following model:

$$\text{model}(\bar{x}) = \sqrt{n}\left(1 - \exp\left(-(\beta_0 + \beta_1 \bar{x})\right)\right)$$

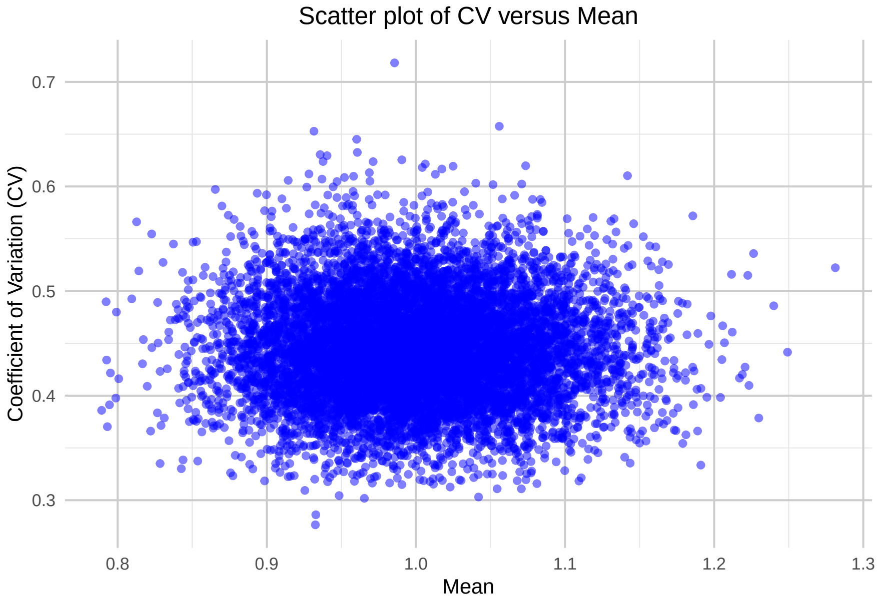

Figure 1: Scatter plot of CV versus Mean.

where β_0 and β_1 are parameters, the p-value for β_1 plays a crucial role in understanding the fit of the model and the nature of the data distribution.

```
##
## Formula: CV ~ sqrt(sample.size) * (1 - exp(-1 * (beta0 + beta1 * Mean)))
##
## Parameters:
##       Estimate Std. Error t value Pr(>|t|)
## beta0 -0.04401    0.00266  -16.55   <2e-16 ***
## beta1  0.37712    0.00300  125.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4687 on 9998 degrees of freedom
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 3.4e-06
```

Interpretation of p-Values for β_1

High p-Value for β_1

- **Implication**: A high p-value (typically greater than 0.05) for β_1 suggests that the coefficient β_1 is not significantly different from zero. This implies that the relationship between the mean ($\bar{x}$) and the coefficient of variation in the data can be explained by the Gamma SAR distribution.
- **Interpretation**: When β_1 is not statistically significant, it supports the hypothesis that the data fits the Gamma SAR distribution. In this case, the sample mean and CV are independent, which is a characteristic feature of the Gamma distribution.

Low p-Value for β_1

- **Implication**: A low p-value (typically less than 0.05) for β_1 indicates that β_1 is significantly different from zero. This suggests that the model does not fit the data well if the assumption of independence

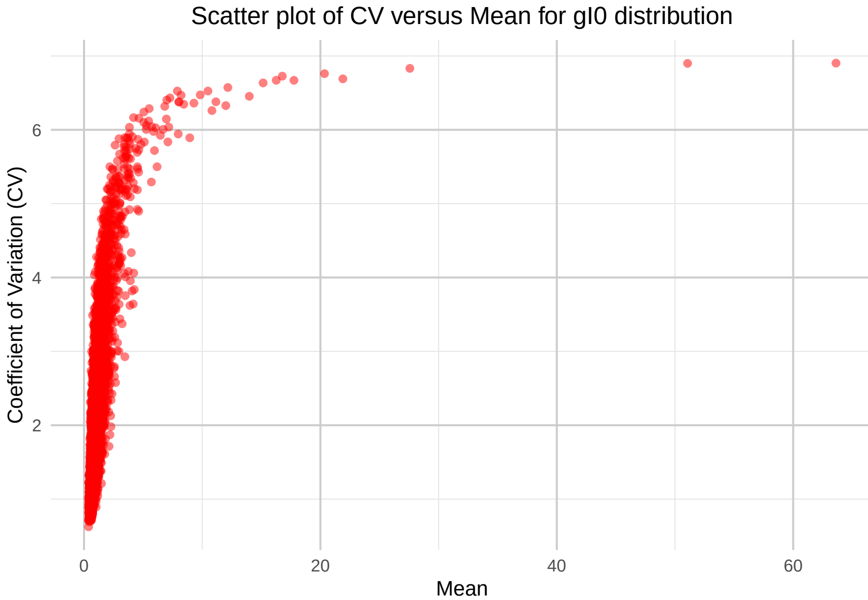

Figure 2: Scatter plot of CV versus Mean, $\alpha = -1.5$.

between the mean and CV is made.

- **Interpretation**: When β_1 is statistically significant, it implies that the data is better described by the G_I^0 distribution . In this scenario, the sample mean and CV are dependent, reflecting a more complex relationship between these two metrics. Specifically, this dependency indicates that the data shows a logarithmic behavior in the relationship between the mean and CV, as described by the G_I^0 distribution. The parameter α of the G_{I^0} distribution, which approaches -1 in extreme cases, captures this dependency.

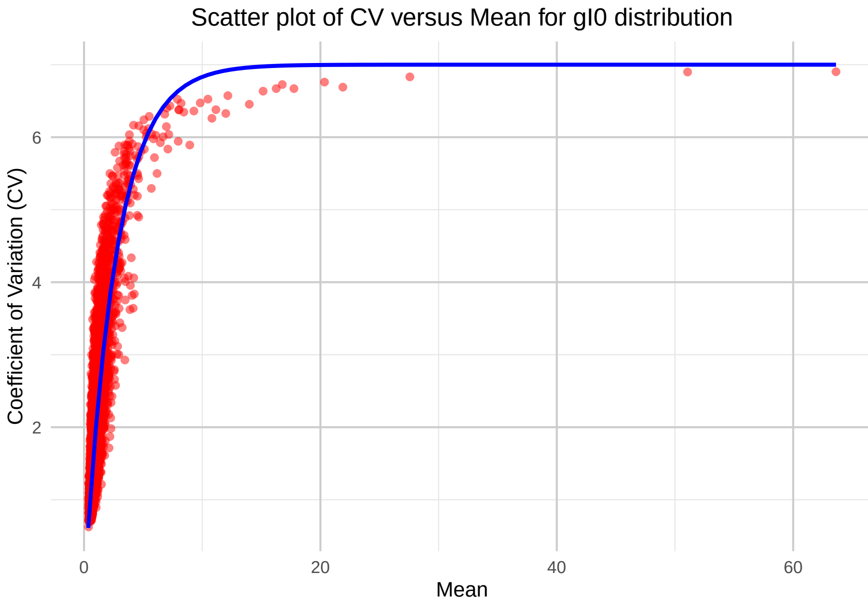

Figure 3: Scatter plot of CV versus Mean, $\alpha = -1.5$.

Table 1: Regression Results for Different Alpha Values.

α	**Parameter**	**Estimate**	**Std. Error**	t **Value**	p **Value**
-1.50000	β_0	-0.04401	0.00266	-16.55000	<2e-16
	β_1	0.37712	0.00300	125.71000	<2e-16
-3.00000	β_0	-0.06769	0.00260	-26.06000	<2e-16
	β_1	0.21955	0.00260	84.62000	<2e-16
-6.00000	β_0	0.01723	0.00194	8.90400	<2e-16
	β_1	0.08462	0.00193	43.82500	<2e-16
-20.00000	β_0	0.05538	0.00137	40.34000	<2e-16
	β_1	0.02043	0.00137	14.91000	<2e-16
-100.00000	β_0	0.06253	0.00121	51.45000	<2e-16
	β_1	0.00491	0.00121	4.04500	5.28e-05
-200.00000	β_0	0.06371	0.00118	53.93300	<2e-16
	β_1	0.00268	0.00118	2.27200	0.0231
-1000.00000	β_0	0.06636	0.00119	56.01200	<2e-16
	β_1	-0.00056	0.00118	-0.47200	0.637