# Gamma Generalized Linear Models

Samuel Liu, Trang Ngo

November 30 2022

## 1 Overview

In this paper, we will introduce the Gamma distribution, discuss the use of the link functions, as well as the estimation of the dispersion parameter $\phi$. Finally, we will apply the Gamma GLM in a case study.

### 1.1 The connection between the Gamma and exponential distribution

Consider independent random variables $Y_1, ..., Y_n$ which follows the exponential distribution with rate parameter $\lambda$ (i.e. $Y_i \sim \text{Expo}(\lambda)$. It follows that $\sum_i^n Y_i$ follows a Gamma distribution with shape parameter $n$ and rate parameter $\lambda$. (i.e. $\sum_i^n Y_i \sim \text{Gamma}(n, \lambda)$.

In other words, the exponential distribution is a special case of the Gamma distribution with shape parameter $n = 1$.

If we use the exponential distribution to model the amount of time until an event happens, then the Gamma distribution (with the same rate parameter $\lambda$ and shape parameter $n$ can be thought of as modelling the amount of time until the $n^{th}$ event occurs.

### 1.2 Modelling positive continuous data

Many applications have continuous and positive response variables, such as the amount of rainfall collected in a reservoir, the load on web servers, or the size of insurance claims. These variables have a boundary at zero, which makes them generally right-skewed. Moreover, these variables usually have an increasing mean-variance relationship.

One increasing mean-variance relationship we have seen is $V(\mu) = \mu$, which corresponds with count data. The Gamma distribution is suitable for data where the mean-variance relationship of the response variable is quadratic, i.e. $V(\mu) = \mu^2$.

# 2 The Gamma Distribution

In this section, we will derive the "basic pieces" of the Gamma distribution, including the canonical link and the deviance.

The probability density function for a Gamma distribution is usually written as:

$$f(y;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$

## 2.1 Exponential family form

We will begin by putting the Gamma's probability density function into the exponential family form:

$$\begin{aligned}
f(y;\alpha,\beta) &= \exp(\alpha \log\beta - \log\Gamma(\alpha) + (\alpha-1)(\log y) - \beta y) \\
&= \exp(\frac{-\frac{\beta y}{\alpha} - \log\frac{1}{\beta}}{\frac{1}{\alpha}} + (\alpha-1)(\log y) - \log\Gamma(\alpha)) \\
&= \exp(\frac{-\frac{\beta y}{\alpha} - \log\frac{\alpha}{\beta}}{\frac{1}{\alpha}} + (\alpha-1)(\log y) - \log\Gamma(\alpha) + \frac{\log(\alpha)}{\frac{1}{\alpha}})
\end{aligned}$$

Thus,

$$\begin{aligned}
\theta &= \frac{-\beta}{\alpha} \\
\phi &= \frac{1}{\alpha} \\
b(\theta) &= \log\frac{\alpha}{\beta} = \log\frac{-1}{\theta} \\
b'(\theta) &= -\frac{1}{\theta} = \frac{\alpha}{\beta} \\
b''(\theta) &= \frac{1}{\theta^2} = \frac{\alpha^2}{\beta^2} \\
Var(Y) &= a(\phi)b''(\theta) = (\frac{1}{\alpha})(\frac{\alpha^2}{\beta^2}) = \frac{\alpha}{\beta^2} \\
Var(\mu) &= (\frac{\alpha^2}{\beta^2}) = \mu^2
\end{aligned}$$

We can also reparameterize using the mean, $\mu$:

$$
\begin{aligned}
f(y;\alpha,\mu) &= \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} Y^{\alpha-1}\exp(\frac{-\alpha y}{\mu}) \\
&= \exp(\alpha\log(\frac{\alpha}{\mu}) - \log(\Gamma(\alpha)) + (\alpha-1)\log Y + (\frac{-\alpha y}{\mu})) \\
&= \exp(\frac{-\frac{y}{\mu} - \log(\frac{\mu}{\alpha})}{\frac{1}{\alpha}} - \log(\Gamma(\alpha)) + (\alpha-1)\log Y) \\
&= \exp(\frac{-\frac{y}{\mu} - \log(\mu)}{\frac{1}{\alpha}} - \log(\Gamma(\alpha)) + (\alpha-1)\log Y - \frac{\log\frac{1}{\alpha}}{\frac{1}{\alpha}})
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\theta &= -\frac{1}{\mu} \\
\phi &= \frac{1}{\alpha} \\
b(\theta) &= \log\mu = \log -\frac{1}{\theta} \\
b'(\theta) &= -\frac{1}{\theta} = \mu \\
b''(\theta) &= \frac{1}{\theta^2} = \mu^2 = \frac{\alpha^2}{\beta^2} \\
Var(Y) &= a(\phi)b''(\theta) = (\frac{1}{\alpha})(\frac{\alpha^2}{\beta^2}) = \frac{\alpha}{\beta^2} \\
Var(\mu) &= b''(\theta) = \mu^2
\end{aligned}
$$

## 2.2 Deviance

Using the pieces we have shown above, we can derive the deviance of the Gamma distribution. From before, we know that $\mu = \frac{-1}{\theta}$, $\theta = \frac{-1}{\mu}$, $b(\theta) = log(\frac{-1}{\theta})$, $\phi = \frac{1}{\alpha}$.

Using the same notation in class:

$$
\begin{aligned}
\hat{\theta}_i &= \theta(\hat{\mu}_i) = \frac{-1}{\mu_i} \\
\tilde{\theta}_i &= \theta(y_i) = \frac{-1}{y_i} \\
b(\hat{\theta}_i) &= log(\hat{\mu}_i) \\
b(\tilde{\theta}_i) &= log(\hat{y}_i)
\end{aligned}
$$

$$
\begin{aligned}
D(y, \hat{\mu_i}) &= 2\sum_{i=1}^{n} \left\{ y_i(\frac{-1}{y_i} + \frac{1}{\hat{\mu_i}}) - (log y_i - log\hat{\mu_i}) \right\} \\
&= 2\sum_{i=1}^{n} \left\{ y_i(\frac{-1}{y_i} + \frac{1}{\hat{\mu_i}}) - (log y_i - log\hat{\mu_i}) \right\} \\
&= 2\sum_{i=1}^{n} \left\{ \frac{y_i - \mu_i}{\mu_i} - log\frac{y_i}{\hat{\mu_i}} \right\}
\end{aligned}
$$

The unit deviance is:

$$
d(y, \mu) = 2\left\{ \frac{y - \mu}{\mu} - log\frac{y}{\mu} \right\}
$$

Note that the unit deviance approximates to the chi-squared distribution with 1 degree of freedom with the correct expected value when $\phi \leq \frac{1}{3}$ by the saddlepoint approximation. This allows us to use the same likelihood ratio tests discussed in Poisson and Binomial GLMs.

# 3 Link Functions

As shown above, the canonical link for the Gamma GLM is the inverse link. However, the log link is a much more popular link to use, because it ensures that $\mu > 0$, and also it is easier to interpret (analyzing impact of explanatory variables as multiplicative). In R, the `"log"`, `"identity"` and `"inverse"` link functions are allowed for the `"glm"` function.

Note that using the inverse link function can produce error messages, when R cannot find suitable starting points. This is often because the inverse link function does not restrict $\mu$ to be positive. You can supply starting points to `"glm()"` on the scale of the data using the input `"mustart"`. Similar problems occur with the identity link. This is why the log link is preferred: we will show this in our case study.

# 4 Estimating the Dispersion Parameter

There are multiple ways to derive approximations for the dispersion parameter: a rigorous and complete comparison between these approximations are likely beyond the scope of this paper. Instead, we will list out and derive most common approximations, and compare their uses and feasibility in the context of GLMs.

The most common estimator for $\phi$ is the Pearson estimator.

$$
\hat{\phi}_{Pearson} = \frac{1}{n - p'} \sum_{i=1}^{n} \frac{(y - \hat{\mu_i})^2}{\hat{\mu_i}^2}
$$

This is usually recommended over the mean deviance estimator,

$$\hat{\phi}_{Deviance} = \frac{D(y, \hat{\mu})}{n - p'}$$

when the accuracy of small values in the data are in doubt, since the deviance is sensitive to rounding errors as y approaches 0. We know this based on the formula for the unit deviance itself (see equation in section 2). When y is close to 0, a small change in y results in a large change in the unit deviance. In contrast, the Pearson estimator is unbiased for $\phi$ if $\mu$ is the correct mean and $V(\mu)$ is the correct variance function: this is because it is unbiased given only the first and second moment assumptions, making it more robust as it relies on fewer assumptions. In practice, the Pearson estimator tends to be more variable but less biased than the mean deviance estimator.

There is also the maximum likelihood estimator for $\phi$. This is determined by finding the value for $\phi$ that maximizes the Gamma likelihood.

We can express the likelihood for a gamma-distribution as

$$l(\mu_i, \phi) = \sum_i^n \frac{1}{\phi} \log(\frac{1}{\phi\mu_i}) + (\frac{1-\phi}{\phi}) \log(y) - \frac{y_i}{\phi\mu_i} - \log(\Gamma(\frac{1}{\phi})$$

To obtain the maximum likelihood estimator of $\phi$ we first obtain the maximum likelihood estimate of $\mu$, which is independent of $\phi$. Then the maximum likelihood equation is obtained by setting the derivative of the log-likelihood function with respect to $\phi$ equal to zero. Finding the partial derivative with respect to $\phi$, letting $\Psi(x) = \Gamma(x)'/\Gamma(x)$ and then setting it to 0 results in

$$\frac{\partial l}{\partial \phi} = \sum_i^n -\frac{1}{\phi^2} \log(\frac{1}{\phi\mu_i}) - \frac{1}{\phi^2} - \frac{\log y_i}{\phi^2} + \frac{y_i}{\phi^2\mu_i} - \Psi(\frac{1}{\phi})(\frac{-1}{\phi^2}) \stackrel{!}{=} 0$$

$$-2\sum_i^n \log\phi + \Psi(\frac{1}{\phi}) = 2\sum_i^n (\log\frac{\mu_i}{y_i} + \frac{y_i - \mu_i}{\mu_i})$$

Note that the right hand side is just the deviance for the gamma distribution. Thus, solving for $\phi$ finding the solution to this equation. Note that solving for $\phi$ this way requires iterative methods, which is why the Pearson and deviance estimates are usually used. However, it is a better unbiased estimator than both estimates when the data actually come from a Gamma GLM. In reality, when the data does not come perfectly from a Gamma distribution, the Pearson estimate is a better estimate, since the MLE estimate beecomes biased.

# 5 A Case Study: Rainfall in Australia.

We will try to apply Gamma GLM to a real-life dataset.

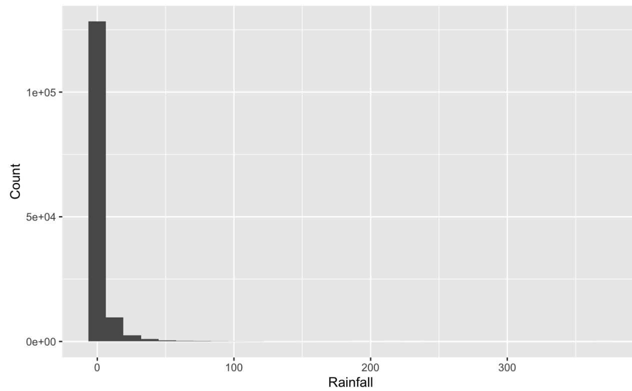


Figure 1: Histogram of rainfall.

## 5.1 Dataset

Our chosen dataset has 142199 observations and includes daily weather data from over 10 years in different locations in Australia, determined from weather stations. Variables include rainfall, minimum temperature, maximum temperature, location, maximum wind speed, humidity, and temperature for a given day.

We would like to model the amount of rain based on other weather conditions in a day using a Gamma GLM.

## 5.2 Data cleaning and exploration

### 5.2.1 Data cleaning

In our cleaning, we removed rows with missing rainfall observations. We then determined the correlation between rainfall for two consecutive days, and found that the value was 0.2. Augmenting the amount of time between days and determining correlation for rainfall, we found that at 5 days of time apart there was only -0.005. As such, we decided to take the rainfall observation for every 5th day.

### 5.2.2 Exploratory Data Analysis

From Figure 1, we can see that the rainfall data is heavily right-skewed. Furthermore, the data has an overabundance of zero's.

Figure 2 plots the log-transformed rainfall. Note that there are still a lot of repeated values: in this case the original data has 8761 observations where rainfall equals 0.2.

Next, we examine the bivariate relationship between the covariates and the response variable. To aid with visual interpretation, we decided to plot each covariate against the log of rainfall in Figure 3. It is not obvious whether any

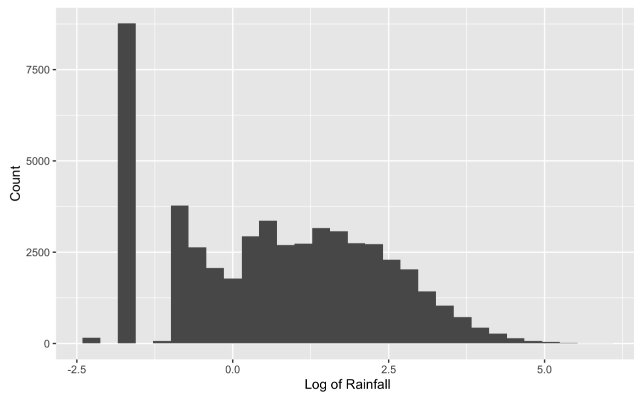


Figure 2: Histogram of log of rainfall.

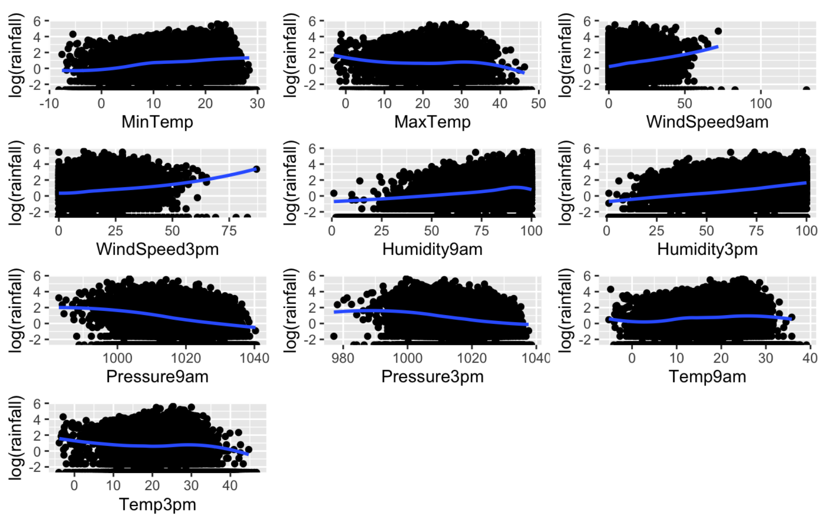


Figure 3: Log of Rainfall against quantitative variables.

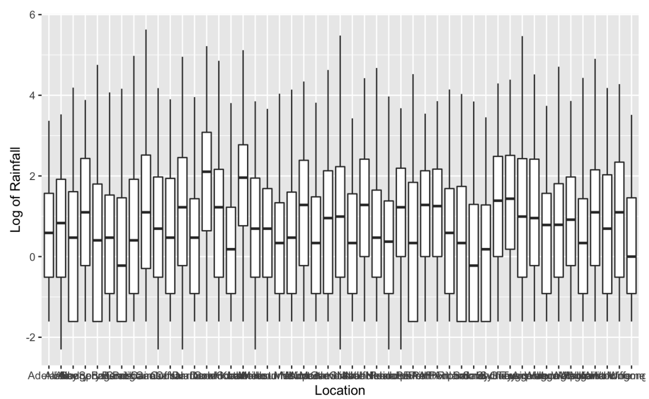


Figure 4: Log of Rainfall for different locations.

of these quantitative variables might be insignificant, so we decided to keep all of them in our model.

As for categorical variables, note that `Location` must be included in the model since the data is collected across different locations in Australia. We can see the variance in rainfall among different locations in Figure 4.

| Temperature | Humidity | Pressure | Wind Speed |
|---|---|---|---|
| 0.86 | 0.67 | 0.96 | 0.52 |

Correlation between 9am and 3pm measurements.

Note that some measurements (including temperature, humidity, wind speed, and atmospheric pressure) were recorded twice per day, at 9am and 3pm. We found that these pairs of measurements are highly correlated, as shown in Table 5.2.2. Thus, we decided to only include the 9am measurements in our model.

Since the Gamma `glm` function in R does not work with data points that are zeros, we tried adding 0.01 to the response variables for all observations. However, the GLM did not converge. We ended up removing all zero's from our dataset. Our research question then became: Can we model the amount of rain based on weather data for a given day *given that it does rain*?

## 5.3 Analysis

Using sequential F tests, we found that the significant variables are `Humidity9am`, `Pressure9am`, `Location`, `WindSpeed9am`, `MaxTemp`, `MinTemp`. The output is shown in Figure 5

After refitting the model after dropping the non-significant variables, we decided to do diagnostics on our model by looking at the residuals. We looked at the working residuals, binned Pearson residuals, and binned Deviance residuals. We find that the working residuals do not look great, but the binned Pearson residuals do not show signs of telescoping, which is good. We also don't see

```
Analysis of Deviance Table

Model: Gamma, link: log

Response: Rainfall

Terms added sequentially (first to last)


            Df Deviance Resid. Df Resid. Dev        F    Pr(>F)
NULL                         8939      21147
Humidity9am   1  1176.55      8938      19971 484.4332 < 2.2e-16 ***
Pressure9am   1  1644.70      8937      18326 677.1904 < 2.2e-16 ***
Location     44  1921.49      8893      16404  17.9807 < 2.2e-16 ***
WindSpeed9am  1   421.43      8892      15983 173.5214 < 2.2e-16 ***
MaxTemp       1    70.34      8891      15913  28.9633 7.564e-08 ***
MinTemp       1   408.97      8890      15504 168.3885 < 2.2e-16 ***
Temp9am       1     2.84      8889      15501   1.1705    0.2793
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5: ANOVA output in R

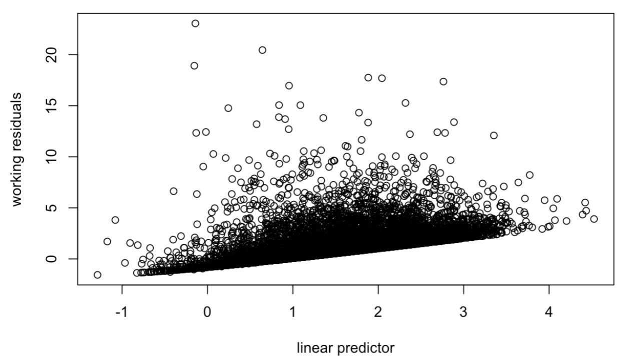


Figure 6: Working residuals

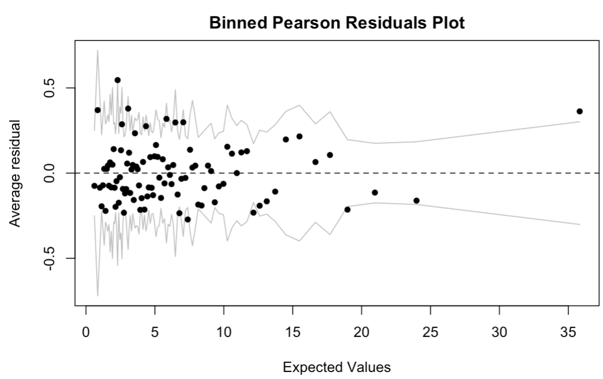


Figure 7: Binned residuals plot with Pearson residuals.

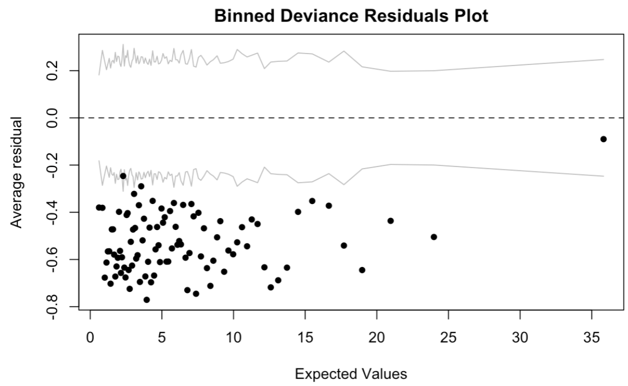


Figure 8: Binned residuals plot with deviance residuals.

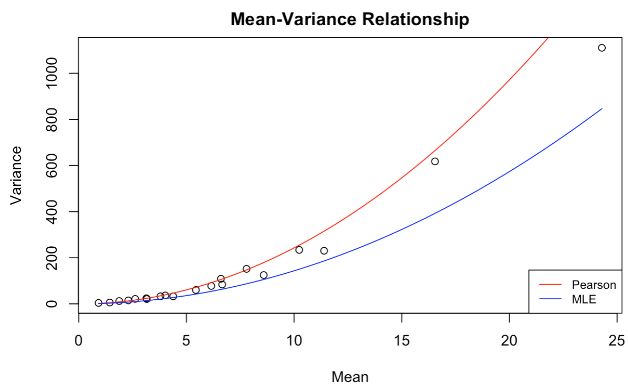


Figure 9: ANOVA output in R

obvious pattern in the deviance residuals, although strangely all the residuals are less than 0, which does not really make sense, because it means that every single predicted value is less than the actual: we did not see this in the Pearson residual case. We think there may be something going on with the calculation for the deviance residual itself, where the log term in the deviance makes the calculation very susceptible to error when y is small, as we mentioned above. We believe that because we have a lot of data close to 0, and that the data is accurate only up to one decimal place, that this may have an effect on the calculation for individual deviances. Still, that does not explain why all the binned deviance residuals are negative. In any case, we believe that the lack of obvious pattern in the residuals are a sign that the model is not violating obvious assumptions.

We then decided to look at the mean variance relationship for our data to ensure that the Gamma GLM was appropriate here. We plotted the mean

```
 (Intercept)  Humidity9am  Pressure9am WindSpeed9am      MaxTemp      MinTemp 
4.330470e+21 1.023828e+00 9.520851e-01 1.021225e+00 9.218932e-01 1.093173e+00 
```

Figure 10: Coefficient output in R on exponential scale

and variance of data grouped by the value of their linear predictor based on the model in Figure 9. We found that the $\mu^2$ relationship was present, which provided evidence for the Gamma GLM. We fit the relationship using both the MLE and Pearson estimate for $\phi$, and found that the Pearson estimate was better at capturing the dispersion. However, it did look like it was a little larger then what the true dispersion was, but this is not entirely implausible, considering that the squared term in the Pearson estimate can tend to result in overestimates of $\phi$.

The `R` output for our coefficients are included in Figure 10. In brief, we expect a 2.3 percent increase in rainfall with an increase in humidity by one percent, given all other variables are constant. We expect a 4.8 percent decrease in rainfall with an increase in pressure by one Hg, given all other variables are constant. We expect a 2.2 percent increase in rainfall with an increase in wind speed by 1 kilometer per hour, given all other variables are constant. We expect a decrease in rainfall by 8 percent with an increase in one Celsius for the maximum temperature, given all other variables are constant. Finally, we expect an increase in rainfall by 9 percent with an increase in one Celsius for minimum temperature. We believe this relationship with temperature may be because very hot days are drier with little rainfall, and thus there is that negative relationship with maximum temperature.

## 5.4 Conclusion

Gamma GLM can be a useful tool for modelling positive continuous data. However, from our example above, we saw that Gamma GLM might not be appropriate for all types of positive continuous data. In our case, even with the zero's removed, our Gamma GLM was still not great. We suspect this might be due to the fact that we have a lot of observations with the same amount. For zero-inflated, positive continuous data, a truncated model or a Tobit model might be more appropriate.