ST346 Chapter 3

Contents

Preface

These slides are a slight adaptation from the original slides developed by Prof Martyn Plummer for the module.

If you find any typos, please inform the module leader.

These materials are solely for your own use and you must not distribute these in any format. **Do not upload these materials to the internet or any filesharing sites nor provide them to any third party or forum.**

Chapter 3 Poisson regression models

3.1 Moments and cumulants

3.1.1 Moments

The **moments** of a random variable Y are the expected values of positive integer powers:

$$\begin{aligned} m_1 &= \mathbb{E}(Y), \\ m_2 &= \mathbb{E}(Y^2), \\ m_3 &= \mathbb{E}(Y^3), \end{aligned}$$

and so on.

The moments of a distribution do not always exist. For example, for the t-distribution on d degrees of freedom, m_k does not exist for $k \geq d$.

When all moments exist, they uniquely characterize the distribution, if Carleman's condition is satisfied:

$$\sum_{k=1}^{\infty} (m_{2k})^{-\frac{1}{2k}} = +\infty.$$

Informally this condition ensures that the absolute value of moment m_k does not grow too quickly with k.

For example, the lognormal distribution has moments

$$m_k = \exp(k^2/2)$$

which grow rapidly with k. Thus the lognormal distribution does not satisfy Carleman's condition and is not uniquely determined by its moments.

If the **moment generating function** $M(t)$ can be defined for t in some neighbourhood of zero, then it uniquely characterizes the distribution. Recall the definition of the moment generating function:

$$\begin{aligned} M(t) &= \mathbb{E}\left(\exp(tY)\right) \\ &= \\ &= \\ &= \sum_{r=0}^{\infty} \frac{t^r m_r}{r!} \end{aligned}$$

By definition

$$M(0) \quad = \quad \mathbb{E}\Big(\exp(0)\Big) \quad = \quad 1.$$

The moments can be recovered from the derivatives of $M(t)$ at zero:

$$m_r \quad = \quad \frac{d^r M(0)}{dt^r}.$$

3.1.2 Cumulant generating function

The **cumulant generating function** is the log of the moment generating function:

$$K(t) \quad = \quad \log\Big[M(t)\Big] \quad = \quad \log\Big[\mathbb{E}\Big(\exp(tY)\Big)\Big]$$

The rth cumulant κ_r is defined as the rth derivative of the cumulant generating function evaluated at $t = 0$:

$$\kappa_r \quad = \quad \frac{d^r K(0)}{dt^r}.$$

The cumulants can be expressed in terms of moments up to the same order (you don't need to memorise these!):

$$\begin{aligned} \kappa_1 &= m_1 \\ \kappa_2 &= m_2 - m_1^2 \\ \kappa_2 &= m_3 - 2m_1 m_2^2 + 2m_1^3 \\ \ldots \end{aligned}$$

Note:

- The first 2 cumulants are well-known quantities: the mean and the variance.
- The 3rd cumulant is a measure of asymmetry, related to the skewness:

$$\text{Skewness} \quad = \quad \frac{\kappa_3}{\kappa_2^{3/2}}.$$

- Higher order cumulants are harder to interpret.

Unlike moments, cumulants are additive for independent random variables.

If $S = \sum_{i=1}^{N} Y_i$ for independent $Y_1, \ldots, Y_n$, then

$$K_S(t) \quad = \quad \sum_{i=1}^{n} K_{Y_i}(t).$$

Proof:

$$\begin{aligned}
K_S(t) &= \log\Big(\mathbb{E}\Big[\exp(tS)\Big]\Big) \\
&= \\
&= \\
&= \\
&= \\
&= \sum_{i=1}^{n} K_{Y_i}(t).
\end{aligned}$$

It follows that the rth cumulant of S is the sum of the rth cumulants of $Y_1, \ldots, Y_n$:

$$\frac{d^r K_S(0)}{dt^r} = \sum_{i=1}^{n} \frac{d^r K_{Y_i}(0)}{dt^r}.$$

We will be dealing with a family of distributions called **exponential dispersion models (EDMs)**.

- All moments and cumulants are finite for EDMs. $M(t)$ and $K(t)$ are always defined.
- $K(t)$ takes a particularly simple form for EDMs.
- We will use $K(t)$ to derive the mean and variance of EDMs.
- In some proofs we will also use the fact that $M(t)$ or $K(t)$ uniquely defines a distribution.

3.1.3 Summary on moments and cumulants

- For the distributions considered in this module, the moment generation function $M(t)$ and the cumulant function $K(t)$ are always defined and **uniquely characterise the distribution.**
- The **moment generating function** $M(t)$ of Y is given by

$$M(t) \quad = \quad \mathbb{E}\left(\exp(tY)\right) \quad = \quad \sum_{r=0}^{\infty} \frac{t^r m_r}{r!}.$$

- $m_r \ = \ \mathbb{E}(Y^r)$ is the rth **moment** and satisfies

$$m_r \quad = \quad \frac{d^r M(0)}{dt^r}.$$

- The **cumulant generating function** $K(t)$ of Y is given by

$$K(t) \quad = \quad \log\Big[M(t)\Big]$$

  where the rth **cumulant** κ_r is defined as

$$\kappa_r \quad = \quad \frac{d^r K(0)}{dt^r}.$$

- $\kappa_1 = \mathbb{E}(Y)$ and $\kappa_2 = \mathbb{V}ar(Y)$.
- If $S = \sum_{i=1}^{N} Y_i$ for independent $Y_1, \ldots, Y_n$, then

$$K_S(t) \quad = \quad \sum_{i=1}^{n} K_{Y_i}(t).$$

- The rth cumulant of S is the sum of the rth cumulants of $Y_1, \ldots, Y_n$:

$$\frac{d^r K_S(0)}{dt^r} \quad = \quad \sum_{i=1}^{n} \frac{d^r K_{Y_i}(0)}{dt^r}.$$

Exercise 6

Show that

$$\begin{aligned} m_r &= \frac{d^r M(0)}{dt^r} \\ \kappa_1 &= m_1 \\ \kappa_2 &= m_2 - m_1^2 \end{aligned}$$

Exercise 7

Work through Sections 1 - 3 of Computer Practical 2.

3.2 The Poisson distribution

The **Poisson distribution** is a discrete distribution with support on $\mathbb{N}_0$:

$$\mathbb{P}(Y = r) \quad = \quad \frac{\mu^r}{r!} \exp(-\mu)$$

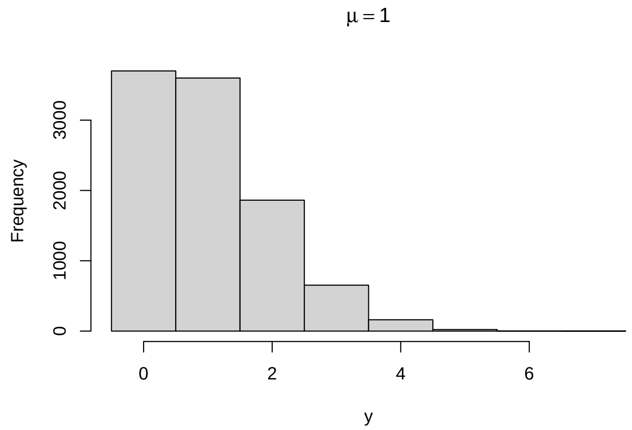
for $r = 0, 1, 2, \ldots$ where

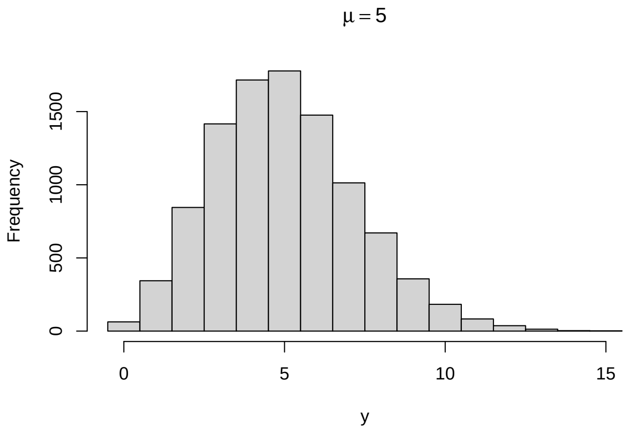
$$\mu \quad = \quad \mathbb{E}(Y).$$

The Poisson distribution is commonly used to model count data.

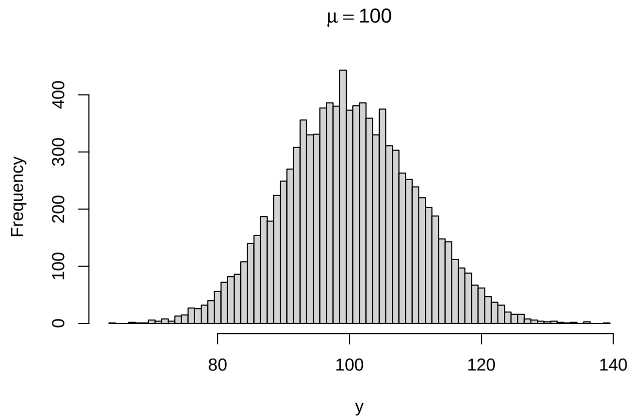
Below are some example histograms for an iid sample of size $n = 10,000$ from a Poisson(μ) distribution. For small values of μ, the Poisson distribution is skewed with a high probability of $Y = 0$ and a right tail that decays quickly.

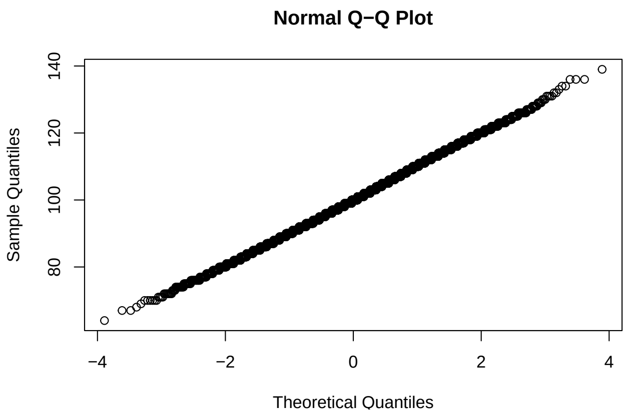
For moderate values of μ the Poisson distribution is more bell-shaped, but still positively skewed.





For large values of μ the Poisson distribution converges to a normal distribution.





3.2.1 Characteristics of the Poisson distribution

The moment generating function of a Poisson(μ) distribution is

$$M(t) \quad = \quad \mathbb{E}\Big(\exp(tY)\Big) \quad = \quad \exp\Big(\mu\big[\exp(t)-1\big]\Big)$$

with the corresponding cumulant generating function

$$K(t) \quad = \quad \log\Big(M(t)\Big) \quad = \quad \mu\Big[\exp(t)-1\Big].$$

Proof:

$$\begin{aligned}
M(t) \quad &= \quad \mathbb{E}\Big(\exp(tY)\Big) \\
&= \\
&= \\
&= \\
&= \\
&= \\
&= \\
&= \quad \exp\Big(\mu\big[\exp(t)-1\big]\Big).
\end{aligned}$$

It follows that $K(t) = \mu\,[\exp(t) - 1]$ and thus

$$\begin{aligned}
K'(t) \quad &= \quad \mu \exp(t), \\
K''(t) \quad &= \quad \mu \exp(t).
\end{aligned}$$

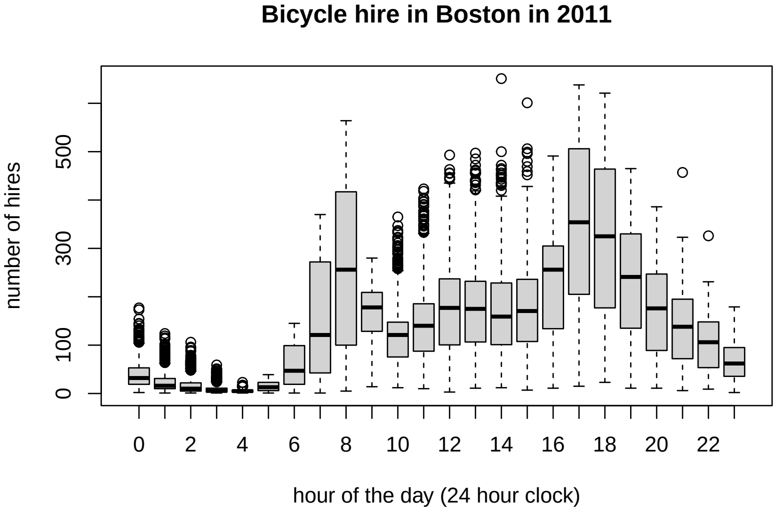
Therefore

$$\begin{aligned}
\mathbb{E}(Y) \quad &= \quad K'(0) \quad = \quad \mu(\exp(0)) \quad = \quad \mu, \\
\mathbb{V}ar(Y) \quad &= \quad K''(0) \quad = \quad \mu(\exp(0)) \quad = \quad \mu.
\end{aligned}$$

Thus the variance of a Poisson distribution is equal to its mean.

3.2.2 Example - Boston bicyle hires

The data shows the count of bicycles hired in each hour.



There is more variation at times of high demand than at times of low demand.

3.2.3 Sums of independent Poisson random variables.

Consider independent random variables $Y_1, \ldots, Y_n$ where

$$Y_i \quad \sim \quad \text{Poisson}(\mu_i).$$

Let $S = \sum_{i=1}^n Y_i$, then

$$S \quad \sim \quad \text{Poisson}(\mu_S)$$

where

$$\mu_S \quad = \quad \sum_{i=1}^n \mu_i.$$

Proof:

The cumulant generating function of S is given by

$$\begin{aligned} K_S(t) &= \sum_{i=1}^n K_{Y_i}(t) \\ &= \\ &= \\ &= \Big[\exp(t) - 1\Big]\mu_S \end{aligned}$$

which is the cumulant generating function of a Poisson distribution with mean μ_S.

3.3 Poisson regression for counts

In a regression model, we explain the expectation μ_i in terms of a p-vector of predictor variables $\boldsymbol{x}_i$ and parameters $\boldsymbol{\beta}$.

For Poisson outcome data, we normally use a log link to construct a regression model:

$$\log(\mu_i) \quad = \quad \boldsymbol{x}_i^T \boldsymbol{\beta}.$$

The log link ensures that the expectation μ_i is always positive as

$$\mu_i \quad = \quad \exp\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right) \quad > \quad 0$$

for all $\boldsymbol{\beta} \in \mathbb{R}^p$.

Other link functions are possible, but they constrain the possible values of $\boldsymbol{\beta}$.

3.4 Poisson regression for rates

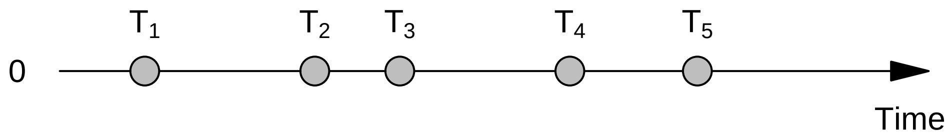
3.4.1 Poisson process

A **Poisson process** is a continuous-time stochastic process, often used to represent queues. Events occur in time with constant rate λ, that is

$$\lim_{\delta \to 0} \frac{\mathbb{P}(\text{event} \in (t, t+\delta])}{\delta} \quad = \quad \lambda,$$

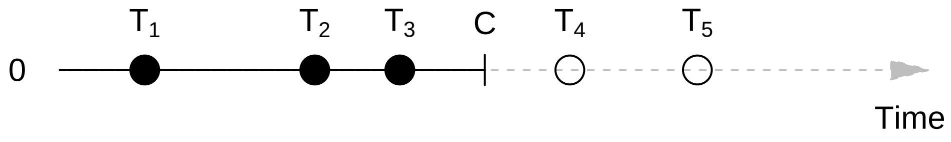
independent of the history of the process and of t.

Event times $T_1, T_2, \ldots$ can be represented as points on the time line.



The inter-event times have independent exponential distributions with common mean $\frac{1}{\lambda}$.

Suppose we observe the Poisson process up to a censoring time C.



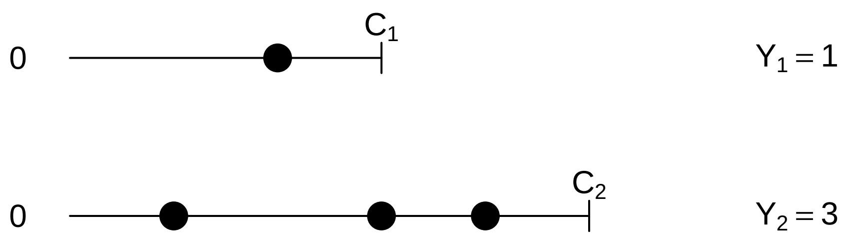
Events occurring after the censoring time are not counted. Let Y be the number of events occurring before time C then

$$Y \quad \sim \quad \text{Poisson}(\mu) \qquad \text{with} \qquad \mu = \lambda C.$$

Observations may have different censoring times. With different observation windows, we have

$$Y_i \quad \sim \quad \text{Poisson}(\mu_i) \qquad \text{with} \quad \mu = \lambda_i C_i.$$

The event count will vary between observations, even if they have the same rate $\lambda_i = \lambda$.



How do we account for this in our Poisson regression model?

Using the log link we have

$$\begin{aligned} \log(\mu_i) &= \log(\lambda_i C_i) \\ &= \log(\lambda_i) + \log(C_i) \end{aligned}$$

If we assume

$$\lambda_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

then

$$\log(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} + \log(C_i).$$

The term $\log(C_i)$ is called an **offset.**

An offset is not the same as a predictor variable. If we put $\log(C_i)$ in our model as a predictor, then we get

$$\log(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} + \gamma \log(C_i)$$

with a parameter γ to be estimated from the data. With an offset, we fix $\gamma = 1$.

The Poisson model with log link and an offset is used whenever we have a denominator D_i such that

$$\mathbb{E}(Y_i) = D_i \, \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

The denominator does not need to be time. It could be, for example

- spatial area,
- population size,
- volume.

3.4.2 Aggregation of Poisson data

Suppose we have Poisson observations $Y_1, Y_2, \ldots, Y_m$ that have the same observed predictor values – hence the same rate λ – but different observation times $C_1, C_2, \ldots, C_m$.

Then the log-likelihood function is given by

$$\begin{aligned} l(\lambda) &= \sum_{i=1}^{m}(Y_i \log(\lambda) - \lambda C_i) + \ldots \\ &= \log(\lambda)\left(\sum_{i=1}^{m} Y_i\right) + \lambda\left(\sum_{i=1}^{m} C_i\right) + \ldots \end{aligned}$$

where terms not depending on λ have been discarded. So the total event count $\sum_{i=1}^{m} Y_i$ and the sum of censoring times $\sum_{i=1}^{m} C_i$ are sufficient for λ as

$$\sum_{i=1}^{m} Y_i \sim \text{Poisson}\Big(\lambda \sum_{i=1}^{m} C_i\Big).$$

(See also the result on sums of independent Poisson random variables in Section 3.2.3!)

If all our predictor variables are categorical, we can aggregate the data. For each combination of predictor variables, all we need is

- the total number of events $\sum_{i=1}^{m} Y_i$ and
- the sum of observation times $\sum_{i=1}^{m} C_i$.

We apply Poisson regression with a log link and use $\log(\sum_{i=1}^{m} C_i)$ as an offset.

3.4.3 Example - The ships data

The dataset `ships` in the `MASS` package gives the number of damage incidents and aggregate months of service by ship type, year of construction, and period of operation.

```
library(MASS)
data(ships)
head(ships,7)
  type year period service incidents
1    A   60     60     127         0
2    A   60     75      63         0
3    A   65     60    1095         3
4    A   65     75    1095         4
5    A   70     60    1512         6
6    A   70     75    3353        18
7    A   75     60       0         0
```

This is aggregated data. The variable `service` gives the aggregate months of service for each group in the data.

As the groups have different total service times and thus different exposure to the risk of a damage incident, we are interested in the rate of damage incidents per month of service time.

Some groups have a zero service time as the period of operation considered is before the year in which the ships in that group were constructed. These are excluded from further analysis.

```
ships <- subset(ships, service > 0)

rate <- sum(ships$incidents)/sum(ships$service)
round(rate, 5)
[1] 0.00218
```

We use `log(service)` as an offset and fit a Poisson GLM with log-link. Here we are fitting a null model, that is a model with no predictor variables.

```
glm.out <- glm(incidents ~ offset(log(service)),
               family=poisson(link="log"), data=ships)
round(coef(glm.out), 2)
(Intercept)
      -6.13
round(exp(coef(glm.out)), 5)
(Intercept)
    0.00218
```

Suppose we would like to examine the incident risks for different types of ships, then we may fit the following Poisson GLM:

```
glm.out <- glm(incidents ~ 0 + type + offset(log(service)),
               family=poisson(link="log"), data=ships)
round(exp(coef(glm.out)), 4)
 typeA  typeB  typeC  typeD  typeE
0.0044 0.0018 0.0019 0.0038 0.0062
```

Exercise 8

Work through Sections 3 and 4 of Computer Practical 2.