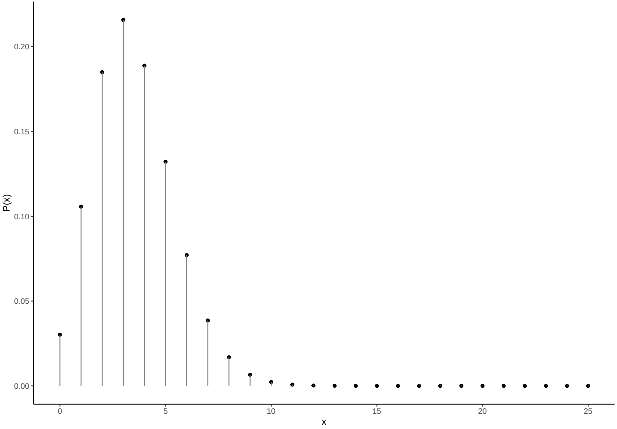
Poisson regression

Mark Andrews
Psychology Department, Nottingham Trent University

✉ mark.andrews@ntu.ac.uk

The Poisson Distribution

- The Poisson distribution is a discrete probability distribution over the non-negative integers $0, 1, 2 \ldots$.



Shown here is a Poisson distribution with $\lambda = 3.5$.

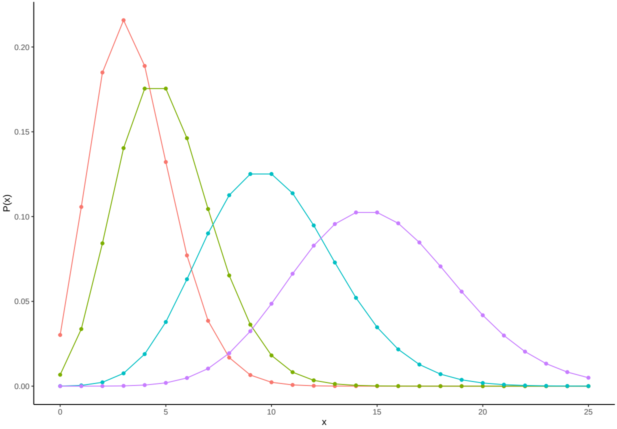
The Poisson Distribution

- The Poisson distribution is used to model the probability of a given number of events occurring in a fixed interval of time, e.g. the number of emails you get per hour, the number of shark attacks on Bondi beach every summer, etc.
- It has a single parameter λ, known as the *rate*.
- If x is a Poisson random variable whose, its probability mass function is

$$P(x = k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

The Poisson Distribution

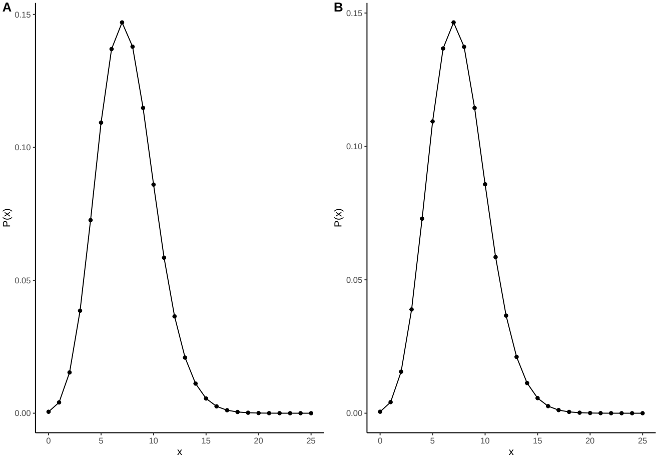
- The mean of a Poisson distribution is equal to its rate parameter λ.
- Its variance is also equal to λ.



As λ increases, so too does the variance.

The Poisson Distribution

- The Poisson distribution can be seen as the limit of a Binomial distribution as $N \to \infty$, and $\lambda = pN$.
- Shown are (left) Binomial$(N, p = \lambda/N)$ where $N \approx 10^3$ and $\lambda = 7.5$, and (right) Poisson(λ).



Poisson Regression

- In any regression problem, our data are $(y_1, x_1), (y_2, x_2) \dots (y_n, x_n)$, where each y_i is modelled as a stochastic function of x_i.
- In Poisson regression, we assume that each y_i is a Poisson random variable rate λ_i and

$$\log(\lambda_i) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki},$$

or equivalently

$$\lambda_i = e^{\beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}}.$$

Poisson Regression

- As an example of Poisson regression, we can look at the number visits to a doctor in a fixed period as a function of predictors such as gender.
- Using a data-set of over 5000 people, we estimate (using mle) that

$$\log(\lambda_i) = 1.65 + 0.43 \times x_i$$

where $x_i = 1$ for a female, and $x_i = 0$ for a male.

Poisson Regression

- Using this example, we see that for a female

$$\lambda_{\text{Female}} = e^{1.65+0.43} = 8.004$$

  and for males

$$\lambda_{\text{Male}} = e^{1.65} = 5.2$$

- In other words, the expected value for females is 8.2 and for males it is 5.2.

Coefficients

- In Poisson regression, coefficients can be understood as follows.
- In the previous example,

$$\begin{aligned}\lambda_{\text{Female}} &= e^{1.65+0.43}, \\ &= e^{1.65}e^{0.43}, \\ \lambda_{\text{Male}} &= e^{1.65}.\end{aligned}$$

- This means that the exponent of the gender coefficient, i.e. $e^{0.43}$, signifies the multiplicative increase to the average rate of doctor visits for women relative men.
- In other words, women visit doctors on average $e^{0.43} = 1.53$ times more than men.

Coefficients

- In an arbitrary example with a single continuous predictor variable,
$$\begin{aligned}\lambda &= e^{\alpha+\beta x_i},\\ &= e^{\alpha}e^{\beta x_i},\end{aligned}$$
If we increase x_i by 1, we have
$$\begin{aligned}\lambda^+ &= e^{\alpha+\beta(x_i+1)},\\ &= e^{\alpha+\beta x_i+\beta},\\ &= e^{\alpha}e^{\beta x_i}e^{\beta},\end{aligned}$$
- As $\lambda^+ = \lambda e^{\beta}$, we see that e^{β} is the multiplicative effect of an increase in one unit to the predictor variable.

Exposure and offset

- In some problems, the length of time during which events are measured varies across individuals.
- In the doctor visits example, we might have recordings of number of visits per year for some people and number of visits per 9 months, etc, for others.
- These situations are dealt with using an *exposure* term for each individual.

Exposure and offset

- When using an exposure term, we use the original count data as before, but treat
$$y_i \sim \text{Poisson}(\lambda_i/u_i),$$
where u_i is a term signifying the relative exposure time for person i.
- According to this,
$$\log(\lambda_i/u_i) = \alpha + \beta x_i,$$
$$\log(\lambda_i) = \alpha + \beta x_i + \log(u_i)$$
- In other words, $y_i \sim \text{Poisson}(\lambda_i/u_i)$ is equivalent to $y_i \sim \text{Poisson}(\lambda_i)$, where $\log(\lambda_i) = \alpha + \beta x_i + \log(u_i)$.

Exposure and offset

- For example, suppose we monitor people's drinking at social occasions. We find that three people drink 12, 7 and 3 drinks over the course of 7, 5 and 2 hours, respectively.
- If we are trying to predict drinking as a function of predictor variables, we ought to calibrate by the different time frames.
- Treating e.g. 12 as a draw from $\text{Poisson}(\lambda_i/7)$ where $\log(\lambda_i/7) = \alpha + \beta x_i$ is identical to treating 12 as a draw from $\text{Poisson}(\lambda_i)$ where $\log(\lambda_i) = \alpha + \beta x_i + \log(7)$.

Exposure and offset

- In general, exposure terms are treated as fixed offsets.
- If our data is $(y_1, x_1), (y_2, x_2) \ldots (y_n, x_n)$ with exposures $u_1, u_2 \ldots u_n$, then we treat

$$y_i \sim \text{Poisson}(\lambda_i),$$

where

$$\log(\lambda_i) = \log(u_i) + \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}.$$

Model Fit with Deviance

- As in the case of logistic regression, we estimate the parameters, e.g. α and β, using maximum likelihood estimation.
- Once we have the maximum likelihood estimate for the parameters, we can calculate *goodness of fit* using deviance.
- As before, the *deviance* of a model is defined

$$-2 \log L(\hat{\alpha}, \hat{\beta} | \mathcal{D}),$$

  where $\hat{\alpha}$, $\hat{\beta}$ are the mle estimates.

Model Fit with Deviance: Model testing

- In a model with one predictor, a null model would be that $P(y_i)$ is not a function of x_i.
- The difference in the deviance of the null model minus the deviance of the full model is

$$\Delta_D = D_0 - D_1.$$

- Under the null hypothesis, Δ_D is distributed as χ^2 with 1df.

Deviance based model testing

- In general, we can compare any two *nested* models using χ^2 test applied to differences in deviance.
- The deviance of the subset model minus that of the full model will always be (approximately) distributed a χ^2 with df equalling the difference in the number of parameters between the two models.
- In other words, under the null hypothesis that subset and full models are identical, the difference in the deviances will be distributed as a χ^2 with df equal to the difference in the number of parameters between the two models.
- This is identical to the case of logistic regression.