

MATH3823 Generalized Linear Models

Robert G Aykroyd

2024-01-18

Table of contents

Appendix A: Basic material	3
A.1 Notation	3
A.2 Rules of probability & statistics	4
A.3 Standard distributions	5

Appendix A: Basic material

This Appendix contains a brief summary of ideas which are considered **pre-requisite**. Although you should be familiar with the ideas and definitions, the notation and layout may be slightly different to that you are used to. If this is the case then please take a little time to get accustomed to the style below, as it will be used through this module. At the University of Leeds, most ideas are covered by the pair of first-year modules, MATH1710 and MATH1712, with a few ideas from MATH2715 in second year. If, however, you have not seen this material previously then please consult a standard introductory text, such as

- Rice, JA (1995). Mathematical Statistics and Data Analysis, 2nd Edition. Duxbury Press. (Other Editions are, of course, fine.)

Please note that this is the first time these notes have been used. If you see errors or you think of some important points which are not covered, then please let me know.

A.1 Notation

Where ever possible, upper case letters from near the end of the alphabet will be used for random variables, for example X and Y , or X_1, \dots, X_n etc and corresponding lower case letters for particular values of a random variable, for example, x and y , or x_1, \dots, x_n . Constants will usually be denoted using lower case letters from the start of the alphabet, for example, a, b, c etc. Further, lower case letters such as i, j, n, m etc are often reserved for integer constants.

In general, parameters of distributions, or of models, will be denoted using lower case Greek letters, such as α, β , with a generic parameter often being denote θ (as below). A common exception to this convention is the probability in the binomial, geometric etc, where the letter p is often used.

The probability mass function of a discrete random variable will be denoted $p(\cdot)$, for example $p(x)$, or $p_X(x)$ when there are multiple distributions being considered at the same time. In these cases, the probability of the event occurring is being given, for example $p_X(x) = Pr(X = x)$. For a continuous random variable, the probability density function will be denoted $f(\cdot)$, for example $f(x)$ or $f_X(x)$. For all types of distribution, the cumulative distribution function will be written $F(\cdot)$, for example $F(x)$ or $F_X(x)$, referring to the probability that the random variable is less than or equal to the specified value, for example $F(x) = Pr(X \leq x)$.

Further, where useful, parameters of distributions may be added to this notation, for example as $f(x; \theta)$. Note that here a semi-colon is used to separate the variable from the parameter. In some areas of statistics, in particular Bayesian methods, similar notation will be used for conditional distributions, such as $f(x|\theta)$, but in that context θ is also being considered a random variable and hence it is a true conditional distribution.

Note that p is one of the most overused symbols as it might represent the number of variables or parameters in a model, $p \in (1, 2, \dots)$, the probability in, say, a geometric distribution, $\text{Geom}(p)$ with $p \in (0, 1)$, a discrete distribution, $p(x)$, or a simple event probability, $p(A)$. Which of these is being considered should be clear by context, but it is important not to get confused.

A.2 Rules of probability & statistics

- **Rules of expectation.** Expectation is linear so for random variables X_1 and X_2 , and constants a , b and c , then

$$\mathbb{E}[aX_1 + bX_2 + c] = a\mathbb{E}[X_1] + b\mathbb{E}[X_2] + c.$$

There is no change to this rule when X_1 and X_2 are independent.

- **Rules of variance.** By definition, the $\text{Var}[X]$ is given by

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

with the variance of the sum of two random variables

$$\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] - 2\text{Cov}[X_1, X_2]$$

where $\text{Cov}[X_1, X_2]$ is the covariance which is defined in the following section. Further,

$$\text{Var}[aX_1 + bX_2 + c] = a^2\text{Var}[X_1] + b^2\text{Var}[X_2] - 2ab\text{Cov}[X_1, X_2].$$

If, however, X_1 and X_2 are independent, then the above reduces to

$$\text{Var}[aX_1 + bX_2] = a^2\text{Var}[X_1] + b^2\text{Var}[X_2].$$

- **Definition of covariance and correlation.** By definition, the covariance, $\text{Cov}[X_1, X_2]$, is given by

$$\text{Cov}[X_1, X_2] = \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])]$$

which has the alternate form

$$\text{Cov}[X_1, X_2] = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$$

and also the correlation

$$\text{Cor}[X_1, X_2] = \frac{\text{Cov}[X_1, X_2]}{\sqrt{\text{Var}[X_1]\text{Var}[X_2]}}.$$

In the case when X_1 and X_2 are independent random variables then $E[X_1 X_2] = E[X_1]E[X_2]$ hence $\text{Cov}[X_1, X_2] = 0$ and so $\text{Cor}[X_1, X_2] = 0$ also. This gives justification for the variance formula for the sum of independent random variables. Note that

$$\text{Cov}[aX_1 + c, bX_2 + d] = ab\text{Cov}[X_1, X_2]$$

but that

$$\text{Cor}[aX_1 + c, bX_2 + d] = \text{Cor}[X_1, X_2].$$

That is correlation is unaffected by linear transformations of the variables.

- **Joint distribution for independent events.** Suppose that X_1 and X_2 are a pair of independent continuous random variables then their joint probability density function is given by

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2).$$

Further, if $\mathbf{X} = \{X_1, \dots, X_n\}$ are a collection of independent continuous random variables then their joint probability function is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i).$$

When dealing with discrete random variables, the density functions are simply replaced by the appropriate probability mass functions.

A.3 Standard distributions

- **Bernoulli**, with parameter θ ($0 < \theta < 1$). A discrete random variable X with probability mass function

$$p(x; \theta) = \theta^x(1 - \theta)^{1-x} \quad x = 0, 1$$

$$E[X] = \theta \text{ and } \text{Var}[X] = \theta(1 - \theta).$$

- **Geometric**, with parameter θ ($0 < \theta < 1$). A discrete random variable X with probability mass function

$$p(x; \theta) = \theta(1 - \theta)^{x-1} \quad x = 1, 2, \dots$$

$$E[X] = 1/\theta \text{ and } \text{Var}[X] = (1 - \theta)/\theta^2.$$

- **Binomial**, with parameters n and θ (where n is a known positive integer and $0 < \theta < 1$). A discrete random variable X with probability mass function

$$p(x; n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad x = 0, 1, \dots, n$$

$$E[X] = n\theta \text{ and } \text{Var}[X] = n\theta(1 - \theta).$$

- **Poisson**, with parameter λ ($\lambda > 0$). A discrete random variable X with probability mass function

$$p(x; \theta) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, \dots$$

$$E[X] = \lambda \text{ and } \text{Var}[X] = \lambda.$$

- **Negative Binomial**, with parameters r and θ ($r > 0$ and $0 < \theta < 1$). A discrete random variable X with probability mass function

$$p(x; r, \theta) = \binom{x-1}{r-1} \theta^r (1 - \theta)^{x-r} \quad x = r, r+1, \dots$$

$$E[X] = \frac{r}{\theta} \text{ and } \text{Var}[X] = \frac{r(1 - \theta)}{\theta^2}.$$

- **Uniform**, with parameter θ ($\theta > 0$). A continuous random variable X , with probability density function

$$f(x; \theta) = \frac{1}{\theta} \quad 0 < x < \theta$$

$$E[X] = \frac{\theta}{2} \text{ and } \text{Var}[X] = \frac{\theta^2}{12}.$$

- **Exponential**, with parameter λ ($\lambda > 0$). A continuous random variable X , with probability density function

$$f(x; \lambda) = \lambda e^{-\lambda x} \quad x > 0$$

$$E[X] = \frac{1}{\lambda} \text{ and } \text{Var}[X] = \frac{1}{\lambda^2}.$$

- **Gamma** with parameters α and β ($\alpha, \beta > 0$). A continuous random variable X , with probability density function

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad x > 0$$

$E[X] = \alpha/\beta$ and $\text{Var}[X] = \alpha/\beta^2$. Note: $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for all α and $\Gamma(\alpha + 1) = \alpha!$ for integers $\alpha > 1$. BEWARE some authors replace β by $1/\beta$ but still calling it Gamma(α, β). You always need to be clear which parameterization is being used.

- **Beta** with parameters α and β ($\alpha, \beta > 0$). A continuous random variable X , with probability density function

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad 0 < x < 1$$

where

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$E[X] = \frac{\alpha}{\alpha+\beta} \text{ and } \text{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Also referred to as the Beta(α, β) distribution.

- **Normal** with parameters μ and σ^2 ($-\infty < \mu < \infty$ and $\sigma^2 > 0$). A continuous random variable X , with probability density function

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad -\infty < x < \infty$$

$$E[X] = \mu \text{ and } \text{Var}[X] = \sigma^2.$$

Also referred to as the N(μ, σ^2) distribution.

- **Pareto** with parameters θ and α ($\theta, \alpha > 0$). A continuous random variable X , with probability density function

$$f(x; \theta, \alpha) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} \quad x > \theta$$

$$E[X] = \frac{\alpha\theta}{(\alpha-1)} \text{ and } \text{Var}[X] = \frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)} \quad (\alpha > 2).$$

- **Chi-Square** with parameter n (n is a positive integer). A continuous random variable X , with probability density function

$$f(x; n) = \left(\frac{1}{2}\right)^{\frac{n}{2}} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{\Gamma(\frac{n}{2})} \quad x > 0$$

$$E[X] = n \text{ and } \text{Var}[X] = 2n.$$

Also referred to as the χ_n^2 distribution, with n degrees of freedom.

Note: A χ_n^2 distribution is also a Gamma($\frac{n}{2}, \frac{1}{2}$) distribution.

- **Student's t** with parameter n (n is a positive integer). A continuous random variable X , with probability density function

$$f(x; n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2}) \left[1 + \frac{x^2}{n}\right]^{\frac{n+1}{2}}} \quad -\infty < x < \infty$$

$$E[X] = 0 \quad (n > 1) \quad \text{and} \quad \text{Var}[X] = \frac{n}{n-2} \quad (n > 2).$$

Also referred to as the t distribution, with n degrees of freedom.

- **F** with parameters m and n (m, n are positive integers). A continuous random variable X , with probability density function

$$f(x; m, n) = \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{\Gamma(\frac{m+n}{2}) x^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2}) \left[1 + \frac{mx}{n}\right]^{\frac{m+n}{2}}} \quad x > 0$$

$$E[X] = \frac{n}{n-2} \quad (n > 2) \quad \text{and} \quad \text{Var}[X] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (n > 4).$$

Also referred to as the F distribution, with m and n degrees of freedom. Note: The order of m and n is important.