math612.5 612.5 Topics in statistics and probability

Gunnar Stefansson

19. desember 2016

Acknowledgements
MareFrame is a EC-funded RTD project which seeks to remove the barriers preventing more widespread use of the ecosystem-based approach to fisheries management.
`http://mareframe-fp7.org`
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no.613571.
`http://mareframe-fp7.org`
Háskóli Íslands
`http://www.hi.is/`

Efnisyfirlit

1 Independence, expectations and the moment generating function 3
1.1 Independent random variables . . . 3
1.1.1 Details . . . 3
1.2 Independence and expected values . . . 4
1.2.1 Details . . . 4
1.2.2 Examples . . . 4
1.3 Independence and the covariance . . . 5
1.4 The moment generating function . . . 5
1.4.1 Examples . . . 5
1.5 Moments and the moment generating function . . . 5
1.5.1 Details . . . 5
1.6 The moment generating function of a sum of random variables . . . 6
1.6.1 Details . . . 6
1.7 Uniqueness of the moment generating function . . . 6

2 The gamma distribution 6
2.1 The gamma distribution . . . 6
2.1.1 Details . . . 6
2.2 The mean, variance and mgf of the gamma distribution . . . 6
2.2.1 Details . . . 7
2.3 Special cases of the gamma distribution: The exponential and chi-squared distributions 8
2.3.1 Details . . . 8
2.4 The sum of gamma variables . . . 9
2.4.1 Details . . . 9

3 Notes and examples: The linear model 10
3.1 Simple linear regression in R . . . 10
3.1.1 Details . . . 10
3.2 Multiple linear regression . . . 11
3.3 The one-way model . . . 11
3.3.1 Details . . . 11
3.3.2 Examples . . . 12
3.4 Random effects in the one-way layout . . . 12
3.4.1 Details . . . 12
3.5 Linear mixed effects models (lmm) . . . 13
3.5.1 Details . . . 13
3.5.2 Examples . . . 13
3.6 Maximum likelihood estimation in lmm . . . 14
3.6.1 Details . . . 14

4 Some regression topics 14
4.1 Poisson regression . . . 14
4.2 The generalized linear model (GLM) . . . 15
4.2.1 Details . . . 15

1 Independence, expectations and the moment generating function

1.1 Independent random variables

Recall that two events, A and B, are independent if,

$$P[A \cap B] = P[A]P[B]$$

Since the conditional probability of A given B is defined by:

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

We see that A and B are independent if and only if

$$P[A|B] = P[A](when P[B] > 0)$$

Two continuous random variables, X and Y, are similarly independent if,

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$$

1.1.1 Details

Two continuous random variables, X and Y, are similarly independent if,

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$$

Now suppose X has p.d.f. f_X and Y has p.d.f. f_Y. Then,

$$P[X \in A] = \int_A f_X(x)dx$$

$$P[Y \in B] = \int_B f_Y(y)dy$$

So X and Y are independent if:

$$P[X \in, Y \in B] = \int_A f_X(x)dx \int_B f_Y(y)dy$$

$$= \int_A f_X(x)(\int_B f_Y(y)dy)dx$$

$$= \int_A \int_B f_X(x) f_Y(y)dydx$$

But, if f is the joint density of X and Y then we know that

$$P[X \in A, Y \in B]$$

$$\int_A \int_B f(x,y)dydx$$

Hence X and Y are independent if and only if we can write the joint density in the form of,

$$f(x,y) = f_X(x) f_Y(y)$$

1.2 Independence and expected values

If X and Y are independent random variables then $E[XY] = E[X]E[Y]$.

Further, if X and Y are independent random variables then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ is true if g and h are functions in which expectations exist.

1.2.1 Details

If X and Y are random variables with a joint distribution function $f(x,y)$, then it is true that for $h : \mathbb{R}^2 \to \mathbb{R}$ we have

$$E[h(X,Y)] = \int\int h(x,y)f(x,y)dxdy$$

for those h such that the integral on the right exists.

Suppose X and Y are independent continuous r.v., then

$$f(x,y) = f_X(x)f_Y(y)$$

Thus,

$$E[XY] = \int\int xyf(x,y)dxdy$$

$$= \int\int xyf_X(x)f_Y(y)dxdy$$

$$= \int xf_X(x)dx \int yf_Y(y)dy$$

$$= E[X]E[Y]$$

Note 1.1. Note that if X and Y are independent then $E[h(X)g(Y)] = E[h(X)]E[g(Y)]$ is true whenever the functions h and y have expected values.

1.2.2 Examples

Example 1.1. Suppose $X,Y \in U(0,2)$ are i.i.d then,

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

and similarly for f_Y.

Next, note that,

$$f(x,y) = f_X(x)f_Y(y) = \begin{cases} \frac{1}{4} & \text{if } 0 \leq x,y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Also note that $f(x,y) \geq 0$ for all $(x,y) \in \mathbb{R}^2$ and

$$\int\int f(x,y)dxdy = \int_0^2 \int_0^2 \frac{1}{4}dxdy = \frac{1}{4}.4 = 1$$

It follows that,

$$E[XY] = \int_{-\alpha}^{\alpha}\int_{-\alpha}^{\alpha} xyf(x,y)dxdy$$

$$= \int_{y=0}^{2} \int_{x=0}^{2} xy.\frac{1}{4}dxdy$$

$$= \int_{y=0}^{2} (\int_{x=0}^{2} xy\frac{1}{4}dx)dy$$

$$= \int_{y=0}^{2} [\frac{1}{4}y.\frac{1}{2}x^2]_{x=0}^{2} dy$$

$$= \int_{y=0}^{2} \frac{1}{4}y(\frac{1}{2}.2^2 - \frac{1}{2}.0)dy$$

$$\int_0^2 \frac{1}{4}y2dy = \int_0^2 \frac{1}{2}ydy = \frac{1}{2}.\frac{1}{2}y^2|_y^2 = \frac{1}{4}.2^2 = 1$$

But

$$E[X] = E[Y] = \int_{y=0}^{2} x.\frac{1}{2}dx = 1$$

So

$$E[XY] = E[X]E[Y]$$

1.3 Independence and the covariance

If X and Y are independent then $Cov(X,Y) = 0$.

In fact, if X and Y are independent then $Cov(h(X), g(Y)) = 0$ for any functions in which expected values exist.

1.4 The moment generating function

If X is a random variable we define the moment generating function when t exists as: $M(t) := E(e^{tX})$.

1.4.1 Examples

Example 1.2. If $X \sim b(n, p)$ then $M(t) = \sum_{x=0}^{n} e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p \cdot (1-p)^{n-x}$

1.5 Moments and the moment generating function

If $M_X(t)$ is the moment generating function (mgf) of X, then $M_X^{(n)}(0) = E[X^n]$.

1.5.1 Details

Observe that $M(t) = E[e^{tX}] = E[1 + X + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \ldots]$ since $e^a = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots$. If the random variable $e^{|tX|}$ has a finite expected value then we can switch the sum and the expected valued to obtain:

$$M(t) = E[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}] = \sum_{n=0}^{\infty} \frac{E[(tX)^n]}{n!} = \sum_{n=0}^{\infty} t^n \frac{E[X^n]}{n!}$$

This implies that the n^{th} derivative of $M(t)$ in $t = 0$ is exactly $E[X^n]$

1.6 The moment generating function of a sum of random variables

$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ if X and Y are independent.

1.6.1 Details

Let X and Y be independent random vaiables, then

$$M_{X+Y}(t) = E[e^{Xt+Yt}] = E[e^{Xt}e^{Xt}] = E[e^{Xt}]E[e^{Xt}] = M_X(t)M_Y(t)$$

1.7 Uniqueness of the moment generating function

Moment generating functions (m.g.f.) uniquely determine the probability distribution function for random variables. Thus, if two random variables have the same m.g.f, then they must also have the same distribution.

2 The gamma distribution

2.1 The gamma distribution

If a random variable X has the density

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

where $x > 0$ for some constants α, $\beta > 0$, then X is said to have a gamma distribution.

2.1.1 Details

The function Γ is basically chosen so that f integrates to one, i.e.

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt$$

It is not too hard to see that $\Gamma(n) = (n-1)!$ if $n \in \mathbb{N}$. Also, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for all $\alpha > 0$.

2.2 The mean, variance and mgf of the gamma distribution

Suppose $X \sim G(\alpha, \beta)$ i.e. X has density

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

Then,

$$E[X] = \alpha\beta$$

$$M(t) = (1-\beta t)^{-\alpha}$$

$$V[X] = \alpha\beta^2$$

2.2.1 Details

The expected value of X can be computed as follows:

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{0}^{\infty} x \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} dx \\
&= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^{\alpha}} \int_{0}^{\infty} \frac{x^{(\alpha+1)-1} e^{-x/\beta}}{\Gamma(\alpha+1)\beta^{\alpha+1}} dx \\
&= \frac{\alpha\Gamma(\alpha)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^{\alpha}}
\end{aligned}$$

so $E[X] = \alpha\beta$.

Next, the m.g.f.is given by

$$\begin{aligned}
E[e^{tX}] &= \int_{0}^{\infty} e^{tx} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_{0}^{\infty} x^{\alpha-1} e^{tx-x/\beta} dx \\
&= \frac{\Gamma(\alpha)\phi^{\alpha}}{\Gamma(\alpha)\beta^{\alpha}} \int_{0}^{\infty} \frac{x^{(\alpha-1)} e^{-x/\phi}}{\Gamma(\alpha)\phi^{\alpha}} dx
\end{aligned}$$

if we choose ϕ so that $\frac{-x}{\phi} = tx - x/\beta$ i.e. $\frac{-1}{\phi} = t - \frac{1}{\beta}$ i.e. $\phi = -\frac{1}{t-1/\beta} = \frac{\beta}{1-\beta t}$ then we have

$$\begin{aligned}
M(t) &= \left(\frac{\phi}{\beta}\right)^{\alpha} \\
&= \left(\frac{\beta/(1-\beta t)}{\beta}\right)^{\alpha} \\
&= \frac{1}{(1-\beta t)^{\alpha}}
\end{aligned}$$

or $M(t) = (1-\beta t)^{-\alpha}$. It follows that

$$M'(t) = (-\alpha)(1-\beta t)^{-\alpha-1}(-\beta) = \alpha\beta(1-\beta t)^{-\alpha-1}$$

so $M'(0) = \alpha\beta$. Further,

$$\begin{aligned}
M''(t) &= \alpha\beta(-\alpha-1)(1-\beta t)^{-\alpha-2}(-\beta) \\
&= \alpha\beta^{2}(\alpha+1)(1-\beta t)^{-\alpha-2}
\end{aligned}$$

$$\begin{aligned}
E[X^{2}] &= M''(0) \\
&= \alpha\beta^{2}(\alpha+1) \\
&= \alpha^{2}\beta^{2} + \alpha\beta^{2}
\end{aligned}$$

Hence,

$$
\begin{aligned}
V[X] &= E[X]^2 - E[X]^2 \\
&= \alpha^2\beta^2 + \alpha\beta^2 - (\alpha\beta)^2 \\
&= \alpha\beta^2
\end{aligned}
$$

2.3 Special cases of the gamma distribution: The exponential and chi-squared distributions

Consider the gamma density,

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

For parameters $\alpha, \beta > 0$.

If $\alpha = 1$ then

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

and this is the density of exponential distribution.

Consider next the case $\alpha = \frac{v}{2}$ and $\beta = 2$ where v is an integer, so the density becomes,

$$f(x) = \frac{x^{\frac{v}{2}-1}e^{\frac{-x}{2}}}{\Gamma(\frac{v}{2})Z^{\frac{v}{2}}}, x > 0$$

This is the density of a chi-squared random variable with v degrees of freedom.

2.3.1 Details

Consider, $\alpha = \frac{v}{2}$ and $\beta = 2$ where v is an integer, so the density becomes,

$$f(x) = \frac{x^{\frac{v}{2}-1}e^{\frac{-x}{2}}}{\Gamma(\frac{v}{2})Z^{\frac{v}{2}}}, x > 0$$

This is the density of a chi - squared random variable with v degrees of freedom.

This is easy to see by starting with $Z \sim n(0,1)$ and defining $W = Z^2$ so that the c.d.f. is:

$$H_(w) = P[W \leq w] = P[Z^2 \leq w]$$

$$= P[-\sqrt{w} \leq Z \leq \sqrt{w}]$$

$$= 1 - P[|Z| > \sqrt{w}]$$

$$= 1 - 2p[Z < -\sqrt{w}]$$

$$= 1 - 2\int_{-\alpha}^{\sqrt{w}} \frac{e^{\frac{-t^2}{2}}}{\sqrt{2w}} dt = 1 - 2\phi(\sqrt{w})$$

The p.d.f. of w is therefore,

$$h(w) = H'(w)$$

$$= 0 - 2\phi'(\sqrt{w})\frac{1}{2}w^{\frac{1}{2}-1}$$

but

$$\phi(x) = \int_{-\alpha}^{x} \frac{e^{\frac{-t^2}{2}}}{2\Pi} dt; \phi'(x) = \frac{d}{dx}\int_{\alpha}^{x} \frac{e^{\frac{-t^2}{2}}}{2\Pi} dt = \frac{e^{\frac{-x^2}{2}}}{2\Pi}$$

So

$$h[w] = -2\frac{e^{\frac{-w}{2}}}{2\Pi}.\frac{1}{2}.w^{\frac{1}{2}-1}$$

$$h[w] = \frac{w^{\frac{-1}{2}-1}e^{\frac{-w}{2}}}{2\Pi}, w > 0$$

We see that we must have $h = f$ with $v = 1$. We have also shown $\Gamma(\frac{1}{2})2^{\frac{1}{2}} = \sqrt{2\Pi}$, i.e $\Gamma(\frac{1}{2}) = \sqrt{\Pi}$. Hence we have shown the χ^2 distribution on 1 df to be $G(\alpha = \frac{v}{2}, \beta = 2)$ when v = 1.

2.4 The sum of gamma variables

In the general case if $X_1 \ldots X_n \sim G(\alpha, \beta)$ are i.i.d. then $X_1 + X_2 + \ldots X_n \sim G(n\alpha, \beta)$.

In particular, if $X_1, X_2, \ldots, X_v \sim \chi^2$ i.i.d. then $\sum_{i=1}^{v} X_i \sim \chi_v^2$.

2.4.1 Details

If X and Y are i.i.d. $G(\alpha, \beta)$, then

$$M_X(t) = M_Y(t) = \frac{1}{(1-\beta t)^{\alpha}}$$

and

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \frac{1}{(1-\beta t)^{2\alpha}}$$

So

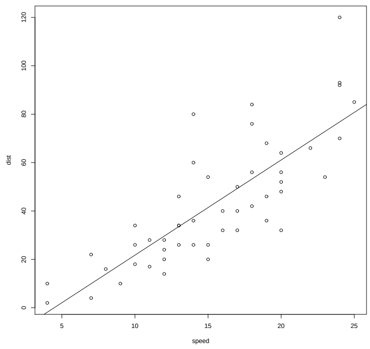
$$X + Y \sim G(2\alpha, \beta)$$

In the general case if $X_1 \ldots X_n \sim G(\alpha, \beta)$ are i.i.d. then $X_1 + X_2 + \ldots X_n \sim G(n\alpha, \beta)$. In particular, if $X_1, X_2, \ldots, X_v \sim \chi^2$ i.i.d., then $\sum_{i=1}^{v} X_i \sim \chi_v^2$

3 Notes and examples: The linear model

3.1 Simple linear regression in R

To test the effect of one variable on another, simple linear regression may be applied. The fitted model may be expressed as $y = \alpha + \hat{\beta}x$, where α is a constant, $\hat{\beta}$ is the estimated coefficient, and x is the explanatory variable.



Example taken from R of a fitted model using linear regression.

3.1.1 Details

Below is the linear regression output using the R's data set "car". Notice that the output from the model may be divided into two main categories:

1. output that assesses the model as a whole, and
2. output that relates to the estimated coefficients for the model

```
Call:
lm(formula = dist ~ speed, data = cars)

Residuals:
   Min 1Q Median 3Q Max
-29.069 -9.525 -2.272 9.215 43.201

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.5791 6.7584 -2.601 0.0123 *
speed 3.9324 0.4155 9.464 1.49e-12 ***
---

Residual standard error: 15.38 on 48 degrees of freedom
Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438
F-statistic: 89.57 on 1 and 48 DF, p-value: 1.490e-12
```

Notice that there are four different sets of output (`Call`, Residuals, Coefficients, and `Results`) for both the constant α and the estimated coefficient $\hat{\beta}$ speed variable.

The estimated coefficients describe the change in the dependent variable when there is a single unit increase in the explanatory variable given that everything else is held constant.

The standard error is a measure of accuracy and is used to construct the confidence interval. Confidence intervals provide a range of values for which there is a set level of confidence that the true population mean will be within the given range. For example, if the CI is set at 95% percent then the probability of observing a value outside the given CI range is less than 0.05.

The p-value is represented as a percentage. Specifically, the p-value indicates the percentage of time, given that your null hypothesis is true, that you would find an outcome at least as extreme as the observed value. If your calculated p-value is 0.02 then 2
In the overall model assessment the R-squared is the explained variance over the total variance. Generally, a higher R^2 is better but data with very little variance makes it easy to achieve a higher R^2, which is why the adjusted R^2 is presented.

Lastly, the F-statistic is given. Since the t-Statistic is not appropriate to compare two or more coefficients, the F-statistic must be applied. The basic methodology is that it compares a restricted model where the coefficients have been set to a certain fixed level to a model which is unrestricted. The most common is the sum of squared residuals F-test.

3.2 Multiple linear regression

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Formally, the model for multiple linear regression, given n observations, is $yi = 0 + 1xi1 + 2xi2 + \ldots pxip + i fori = 1, 2, \ldots n.$

The definition above was taken from: http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm

3.3 The one-way model

The one-way ANOVA model is of the form:

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

or

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

3.3.1 Details

The one-way ANOVA model is of the form:

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

where Y_{ij} is observation j in treatment group i and μ_i are the parameters of the model and are means of treatment group i. The ε_{ij} are independent and follow a normal distribution with mean zero and constant variance σ^2 often written as $\varepsilon \sim N(0, \sigma^2)$.

The ANOVA model can also be written in the form:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where μ is the overall mean of all treatment groups and α_i is the deviation of mean of treatment group i from the overall mean. The ε_{ij} follow a normal distribution as before.

The expected value of Y_{ij} is μ_i as the expected value of the errors is zero, often written as $E[Y_{ij}] = \mu_i$.

3.3.2 Examples

Example 3.1. In the rat diet experiment the model would be of the form:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

where y_{ij} is the weight gain for rat j in diet group i, μ_i would be the mean weight gain in diet group i and εij would be the deviation of rat j from the mean of its diet group.

3.4 Random effects in the one-way layout

The random effects model is written as: $y_{ij} = \mu + \alpha_1 + \varepsilon_{ij}$
where

$$j = 1, \ldots, J$$
$$i = 1, \ldots, I$$

and assumes $\varepsilon_{ij} \sim n(0, \sigma_A^2)$, $\alpha_i \sim n(0, \sigma_A^2)$, and that they are all independent.

3.4.1 Details

Note that this is considerably different from the fixed effect model

$$Ey_{ij} = \mu$$
$$Vy_{ij} = \sigma_A^2 + \sigma^2$$

we have

$$\begin{aligned} cov(y_{ij}, y_{i'j'}) &= cov(\alpha_i + \varepsilon_{ij}, \alpha_{i'} + \varepsilon_{i'j'}) \\ &= E[(\alpha_i + \varepsilon_{ij})(\alpha_{i'} + \varepsilon_{i'j'})] \\ &= E[\alpha_i \alpha_{i'}] + E[\varepsilon ij \alpha_{i'}] + E[\alpha_i \varepsilon_{i'j'}] + E[\varepsilon_{ij}\varepsilon_{i'j'}] \end{aligned}$$

Note 3.1. Note that $E[UW] = E[U]E[W]$ if U, W are independent

So,

$$E[\varepsilon_{ij}\alpha_{i'}] = E[\alpha_i \varepsilon_{i'j'}] = E\alpha_i E\varepsilon_{i'j'} = 0$$

Further,

$$E[\varepsilon_{ij}\varepsilon i'j'] = \begin{cases} \sigma^2 & \text{if } i = i', j = j' \\ 0 & \text{otherwise} \end{cases}$$

and

$$E[\alpha_i \alpha i'] = \begin{cases} \sigma_A^2 & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases}$$

so

$$Cov(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_A^2 + \sigma^2 & \text{if } i = i', j = j' \\ \sigma_A' & \text{if } i = i', j \neq j' \\ 0 & \text{otherwise} \end{cases}$$

It follows that the correlation between measurements y_{ij} and $y_{ij'}$ (within the same group) are

$$\begin{aligned} cor(y_{ij}, y_{ij'} &= \frac{Cov(y_{ij}, y_{ij'})}{\sqrt{v[y_{ij}]v[y_{ij'}]}} \\ &= \frac{\sigma_A^2}{\sqrt{(\sigma_A^2+\sigma^2)^2}} \end{aligned}$$

$\Rightarrow Cor(y_{ij}, y_{ij'}) = \frac{\sigma_A^2}{\sigma_A^2+\sigma^2}$

This is the intra-class correlation.

3.5 Linear mixed effects models (lmm)

The simplest mixed effects model is

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

where $\mu, \alpha_1, \alpha_2, \ldots, \alpha_i$ are unknown constants,
$\beta_j \sim n(0, \sigma_\beta^2)$
$\varepsilon_i j \sim n(0, \sigma^2)$
(β_j and $\varepsilon_i j$ independent).

3.5.1 Details

The μ and α_i are the fixed effects and β_j is the random effects.

Recall that in the simple one-way layout with $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, we can write the model in matrix form $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ where $\underline{\beta} = (\mu, \alpha_1, \ldots, \alpha_I)'$ and X is appropriately chosen.

The same applies to the simplest random effects model $y_{ij} = \mu + \beta_j + \varepsilon_{ij}$ where we can write $\underline{y} = \mu \cdot \underline{1} + Z\underline{U} + \underline{\varepsilon}$ where $\underline{1} = (1, 1, \ldots, 1)'$, $\underline{U} = (\beta_1, \ldots, \beta_J)'$.

In general, we write the mixed effects models in matrix form with $\underline{y} = X\underline{\beta} + Z\underline{U} + \underline{\varepsilon}$, where $\underline{\beta}$ contains the fixed effects and $\underline{U}$ contains the random effects.

3.5.2 Examples

Example 3.2.

1. $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ (SLR)
2. $y_{ij} = \mu + \alpha_i + \beta_i x_{ij} + \varepsilon_{ij}$ only fixed effects (ANCOVA)
3. $y_{ijk} = \mu + \alpha_i + b_j + \varepsilon ijk$ where α_i are fixed but b_j are random.
4. $y_{ijk} = \mu + \alpha_i + b_j x_{ij} + \varepsilon_{ijk}$ where α_i are fixed but b_j are random slopes.

3.6 Maximum likelihood estimation in lmm

The likelihood function for the unknown parameters $L(\beta, \sigma_A^2, \sigma^2)$ is

$$\frac{1}{(2\pi)^{n/2}\left|\Sigma_y\right|^{n/2}} e^{-1/2(\mathbf{y}-X\beta)'\Sigma_y^{-1}(\mathbf{y}-X\beta)}$$

where $\Sigma_y = \sigma_A^2 ZZ' + \sigma^2 I$.

Maximising L over $\beta, \sigma_A^2, \sigma^2$ gives the variance components and the fixed effects. May also need $\hat{\mathbf{u}}$, this is normally done using BLUP.

3.6.1 Details

Recall that if W is a random variable vector with $EW = \mu$ and $VW = \Sigma$ then

$$E[AW] = A\mu$$
$$V[AW] = A\Sigma A'$$

In particular, if $W \sim n(\mu, \Sigma($ then $AW \sim n(A\mu, A\Sigma A')$.

Now consider the lmm with

$$y = X\beta + Zu + \varepsilon$$

where

$$u = (u_1, \ldots, u_m)'$$
$$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)'$$

and the random variables $U_i \sim n(0, \sigma_A^2)$, $\varepsilon_i \sim n(0, \sigma^2)$ are all independent so that $u \sim n(0, \sigma_A^2 I)$ and $\varepsilon \sim n(\mathbf{0}, \sigma^2 I)$.
Then $Ey = X\beta$ and

$$\begin{aligned} Vy &= \Sigma_y \\ &= V[Zu + V[\varepsilon] \\ &= Z(\sigma_A^2 I)Z' + \sigma^2 I \\ &= \sigma_A^2 ZZ' + \sigma^2 I \end{aligned}$$

and hence $y \sim n(X\beta, \sigma_A^2 ZZ' + \sigma^2 I)$.

Therefore the likelihood function for the unknown parameters $L(\beta, \sigma_A^2, \sigma^2)$ is

$$= \frac{1}{(2\pi)^{n/2}\left|\Sigma_y\right|^{n/2}} e^{-1/2(\mathbf{y}-X\beta)'\Sigma_y^{-1}(y} - X\beta)$$

where $\Sigma_y = \sigma_A^2 ZZ' + \sigma^2 I$. Maximizing L over $\beta, \sigma_A^2, \sigma^2$ gives the variance components and the fixed effects. May also need $\hat{u}$, which is normally done using BLUP.

4 Some regression topics

4.1 Poisson regression

Data y_i are from a Poisson distribution with mean μ_i and $\ln \mu_i = \beta_1 + \beta_2 x_i$. A likelihood function can be written and the parameters can be estimated using maximum likelihood.

4.2 The generalized linear model (GLM)

Data y_i are from a distribution within the exponential family, with mean μ_i and $g(\mu_i) = x_i\beta'$ for some link function, g. A likelihood function can now be written and the parameters can be estimated using maximum likelihood.

4.2.1 Details

Data y_i are from a distribution within the exponential family, with mean μ_i and $g(\mu_i) = x_i\beta'$ for some link function, g.

The exponential family includes distributions such as the Gaussian, binomial, Poisson, and gamma (and thus exponential and chi-squared).

The link functions are typically

- identity(with the Gaussian)
- log (with the Poisson and the gamma)
- logistic (with the binomial)

A likelihood function can be written and the parameters can be estimated using maximum likelihood.