

# limdil: Limiting Dilution Analysis

## Supplementary Material: Mathematical Algorithm Implemented in limdil

Yifang Hu and Gordon Smyth

August 29, 2008

## 1 Introduction

limdil is a software application for limiting dilution analysis (<http://bioinf.wehi.edu.au/software/limdil/>).

### 1.1 Aims

The limdil function combines maximum likelihood estimation of a binomial generalized linear model to determine the response frequency for general cases, and Clopper-Pearson intervals to determine the confidence interval of response frequency for extreme special cases. The generalized linear model is fitted with cloglog link and offset  $\log(\text{dose})$ . If `observed=FALSE`, which means the total number of cells is from a Poisson distribution, the values of dose are treated as expected values, a classic Poisson single-hit model, which assumes that one cell is necessary for a response, is applied. If `observed=TRUE`, which means the total number of cells is exact, the values of dose are treated as actual cell numbers rather than expected values.

### 1.2 Variables

$M$  = number of distinct dilutions in the assay

$d_i$  = expected number of cells at dilution  $i$  when `observed=FALSE`,  $i = 1, \dots, M$

$d_i$  = actual number of cells at dilution  $i$  when `observed=TRUE`,  $i = 1, \dots, M$

$n_i$  = number of tests at dilution  $i$ ,  $i = 1, \dots, M$

$r_i$  = number of responses observed among  $n_i$ ,  $r_i \sim \text{binom}(p_i, n_i)$ ,  $i = 1, \dots, M$

$p_i$  = probability of observing a response given dose  $d_i$ ,  $i = 1, \dots, M$

$\lambda$  = proportion of cells that respond

$\frac{1}{\lambda}$  = average number of cells needed for the occurrence of a single response

## 2 Number of Cells is Random

### 2.1 General Case

When `observed = FALSE`, the number of cells in any sample at dilution  $i$  is assumed to follow a Poisson distribution, so for one test, the number of cells  $\sim \text{Pois}(d_i)$ . So the number of respondent cells  $\sim \text{Pois}(\lambda d_i)$  and  $P(\text{number of respondent cells}=0) = e^{-\lambda d_i}$ . Hence,

$$\begin{aligned}1 - p_i &= e^{-\lambda d_i} \\ \log(1 - p_i) &= -\lambda d_i \\ \log(-\log(1 - p_i)) &= \log d_i + \log \lambda \\ &= \beta_1 \log d_i + \beta_0\end{aligned}$$

where  $\beta_1 = 1$  and  $\beta_0 = \log \lambda$ .

This generalised linear model can be estimated using the R function

```
fit <- glm(y ~ offset(log(dose)),family=f,weights=tested)
```

The following items are inputs:

**y**: a vector of the response variable,  $\frac{r_i}{n_i}$ .

**offset(log(dose))**: the covariate  $\log d_i$  is included in the linear model, with known coefficient 1.

**f**: `binomial(link = "cloglog")`; “cloglog” stands for complementary log-log link.

**tested**: a vector giving number of tests at each dose,  $n_i$ .

The following items are outputs:

$\hat{\beta}_0$ : `fit$coef`, the estimated intercept.

**s.e.(\hat{\beta}\_0)**: `summary(fit)$coef[, "Std. Error"]`, which is the standard error of the estimated intercept.

Assume that  $\hat{\beta}_0$  is from a standard normal distribution and the size of the test is  $\alpha$ , the  $(1 - \alpha)100\%$  confidence interval (CI) for  $\log \lambda$  is

$$\hat{\beta}_0 \pm Z_{\frac{\alpha}{2}} \text{s.e.}(\hat{\beta}_0)$$

where  $Z_{\frac{\alpha}{2}}$  is the standard normal quantile with tail probability  $\frac{\alpha}{2}$ .

The goodness of fit of the single-hit model can be examined by testing the null hypothesis  $H_0: \beta_1 = 1$  against  $H_1: \beta_1 \neq 1$  (1). This hypothesis is tested by fitting a second model

```
fit1 <- glm(y ~ log(dose),family=f,weights=tested)
```

and conducting a likelihood ratio test, then to compare two models by using  $\chi^2$  test. If  $H_0$  is rejected, the model fit is not appropriate.

## 2.2 Extreme Special Cases

When the number of responses from all tests is zero, or when every test receives a response,  $\lambda$  and its CI are calculated differently from Section 2.1 (2).

### 2.2.1 No Responses

When no responses are observed from every test, it implies that, for each test, no cells respond. Let  $N = \sum_{i=1}^M n_i d_i$ , the total number of cells over all assays. Then

$$\begin{aligned} P(r_i = 0) &= e^{-n_i d_i \lambda} \\ \Rightarrow P(r_i = 0 \forall i) &= \exp(-\lambda N) \end{aligned}$$

A one sided  $(1 - \alpha)100\%$  CI for  $\lambda$  is  $(0, \lambda_U)$ , where  $\lambda_U$  satisfies

$$\alpha = \exp(-N\lambda_U).$$

Hence

$$\lambda_U = \frac{-\log \alpha}{N}.$$

The CI for  $1/\lambda$  is  $(1/\lambda_U, \infty)$ .

### 2.2.2 All Tests Respond

When every test receives a response, it implies that, for every test, at least one cell from the dilution responds. So

$$\begin{aligned} P(r_i = n_i) &= (1 - e^{-d_i \lambda})^{n_i} \\ \Rightarrow P(r_i = n_i \forall i) &= \prod_{i=1}^M (1 - e^{-d_i \lambda})^{n_i} \end{aligned}$$

A one sided  $(1 - \alpha)100\%$  CI for  $\lambda$  is  $(\lambda_L, 1)$ , where  $\lambda_L$  satisfies

$$\alpha = \prod_{i=1}^M (1 - e^{-d_i \lambda_L})^{n_i}$$

Since the equation above is a little hard to solve, we give a conservative one sided  $(1 - \alpha)100\%$  CI which is available in closed form. Let  $d$  be the smallest dose among all and  $N$  be the sum of all  $n_i$  at dose  $d$ . The CI is  $(\lambda_L, 1)$  where  $\lambda_L$  satisfies

$$\alpha = (1 - e^{-d\lambda_L})^N.$$

Hence

$$\lambda_L = \frac{\log(1 - \alpha^{1/N})}{-d}.$$

The CI for  $1/\lambda$  is  $(1, 1/\lambda_L)$ .

## 3 Number of Cells is Observed

### 3.1 General Case

When observed = TRUE, the number of cells in the dose is exact, so for each dilution,  $P(r_i = 0) = (1 - \lambda)^{d_i}$ . Hence,

$$\begin{aligned}1 - p_i &= (1 - \lambda)^{d_i} \\ \log(1 - p_i) &= d_i \log(1 - \lambda) \\ \log(-\log(1 - p_i)) &= \log d_i + \log(-\log(1 - \lambda)) \\ &= \beta_1 \log(d_i) + \beta_0\end{aligned}$$

where  $\beta_1 = 1$  and  $\beta_0 = \log(-\log(1 - \lambda))$

This generalised linear model can be estimated using the R function

```
fit <- glm(y ~ offset(log(dose)),family=f,weights=tested)
```

The inputs are the same as those in Section 2.1.

The following items are outputs:

$\hat{\beta}_0$ : `fit$coef`, the estimated intercept.

s.e.( $\hat{\beta}_0$ ): `summary(fit)$coef[, "Std. Error"]`, which is the standard error of the estimated intercept.

Assume that  $\hat{\beta}_0$  is from a standard normal distribution and the size of the test is  $\alpha$ . The  $(1 - \alpha)100\%$  CI for  $\log(-\log(1 - \lambda))$  is

$$\hat{\beta}_0 \pm Z_{\frac{\alpha}{2}} \text{s.e.}(\hat{\beta}_0)$$

where  $Z_{\frac{\alpha}{2}}$  is the standard normal quantile with tail probability  $\frac{\alpha}{2}$ .

As in Section 2.1, the goodness of fit of the single-hit model can be examined by testing the null hypothesis  $H_0: \beta_1 = 1$  against  $H_1: \beta_1 \neq 1$ . If  $H_0$  is rejected, the model fit is not appropriate.

### 3.2 Extreme Special Cases

When the number of responses from all tests is zero, or when every test receives a response,  $\lambda$  and its CI are calculated differently from Section 3.1.

#### 3.2.1 No Responses

When no responses are observed from every test, it implies that, for each test, no cells respond. Let  $N = \sum_{i=1}^M n_i d_i$ , the total number of cells over all assays. Then

$$\begin{aligned}P(r_i = 0) &= (1 - \lambda)^{n_i d_i} \\ \Rightarrow P(r_i = 0 \forall i) &= (1 - \lambda)^N\end{aligned}$$

A one sided  $(1 - \alpha)100\%$  CI for  $\lambda$  is  $(0, \lambda_U)$ , where  $\lambda_U$  satisfies

$$\alpha = (1 - \lambda_U)^N.$$

Hence

$$\lambda_U = 1 - \alpha^{1/N}.$$

The CI for  $1/\lambda$  is  $(1/\lambda_U, \infty)$ .

### 3.2.2 All Tests Respond

When every test receives a response, it implies that, for every dilution, at least one cell from the dilution responds. So

$$\begin{aligned} P(r_i = n_i) &= (1 - (1 - \lambda)^{d_i})^{n_i} \\ \Rightarrow P(r_i = n_i \forall i) &= \prod_{i=1}^M (1 - (1 - \lambda)^{d_i})^{n_i} \end{aligned}$$

A one sided  $(1 - \alpha)100\%$  CI for  $\lambda$  is  $(\lambda_L, 1)$ , where  $\lambda_L$  satisfies

$$\alpha = \prod_{i=1}^M (1 - (1 - \lambda_L)^{d_i})^{n_i}$$

Since the equation above is a little hard to solve, we give a conservative one sided  $(1 - \alpha)100\%$  CI which is available in closed form. Let  $d$  be the smallest dose among all and  $N$  be the sum of all  $n_i$  at dose  $d$ . The CI is  $(\lambda_L, 1)$  where  $\lambda_L$  satisfies

$$\alpha = (1 - (1 - \lambda_L)^d)^N.$$

Hence,

$$\lambda_L = 1 - \exp\left(\frac{\log(1 - \alpha^{1/N})}{d}\right).$$

The CI for  $1/\lambda$  is  $(1, 1/\lambda_L)$ .

## References

- [1] Thierry Bonnefoix, Philippe Bonnefoix, Paul Verdiela, and Jean-Jacques Sotto. Fitting limiting dilution experiments with generalized linear models results in a test of the single-hit poisson assumption. *J Immunol Methods*, 194:113–119, 1996.
- [2] C.J. Clopper and E.S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26:404–413, 1934.