STK4900/9900 - Lecture 8

Program

y continuous $Y \sim N(\mu, \sigma^2) \rightarrow$ linear regression
y binary $Y \sim Bi(1, p) \rightarrow$ logistic regression
y counts $Y \sim Po(\lambda) \rightarrow$ Poisson regression

1. Poisson distribution
2. Poisson regression
3. Generalized linear models

- Chapter 8 (except 8.2 and 8.4)
- Supplementary material on Poisson distribution

Example: Emission of alpha particles

In an experiment from 1910 Ernest Rutherford and Hans Geiger recorded the number of alpha-particles emitted from a polonium source in each of 2608 eighth-minute intervals

No.	0	1	2	3	4	5	6
Observed	57	203	383	525	532	408	273
No.	7	8	9	10	11	12	13+
Observed	139	49	27	10	4	2	0

Example: Occurrence of anencephaly in Edinburgh 1956-66

Anencephaly is a serious disorder which causes the brain of a fetus not to develop properly. The number of children born with anencephaly in Edinburgh in the 132 months from 1955 to 1966 were:

# anencephaly	0	1	2	3	4	5	6	7	8	9+
# months	18	42	34	18	11	6	0	2	1	0

We need a distribution that describes such counts

Poisson distribution

A random variable Y is Poisson distributed with parameter λ if

$$P(Y = y) = \frac{\lambda^y}{y!} e^{-\lambda}, \quad y = 0, 1, 2, \ldots$$

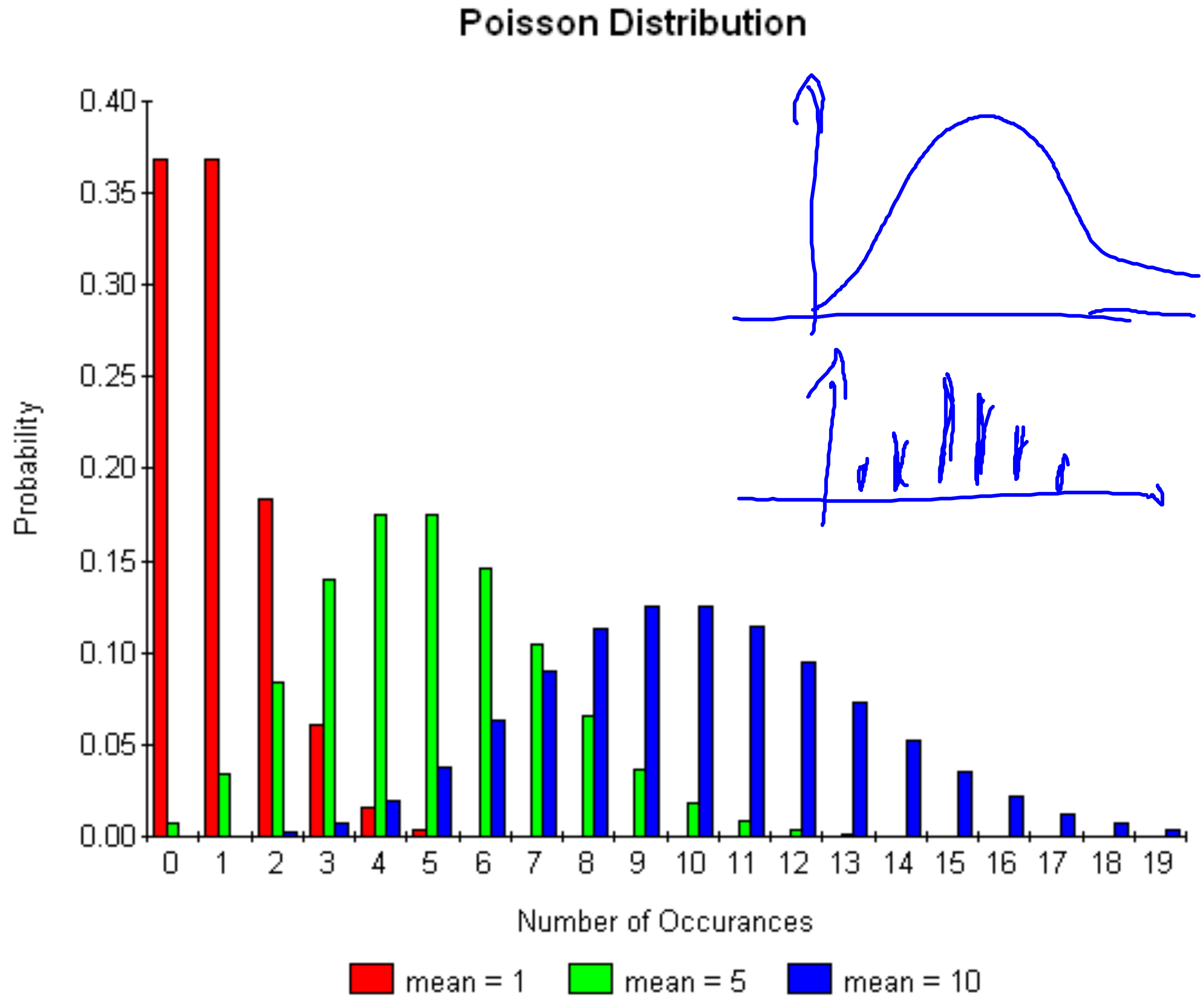
Short we may write: $Y \sim \text{Po}(\lambda)$

We have that:

$$E(Y) = \text{Var}(Y) = \lambda$$

The Poisson distribution arises as:

- an approximation to the distribution of $Y \sim \text{bin}(n, p)$ when p is small and n is large $(\lambda = np)$
- from a Poisson process



$Po(\lambda=1)$ $Po(\lambda=5)$ $Po(\lambda=10)$

Poisson approximation to the binomial distribution

When n is large and p is small, we have with $\lambda = np$

$$\binom{n}{y} p^y (1-p)^{n-y} \approx \frac{\lambda^y}{y!} e^{-\lambda}$$

$np = 500 \cdot \frac{1}{1000} = 0.5 = \lambda$

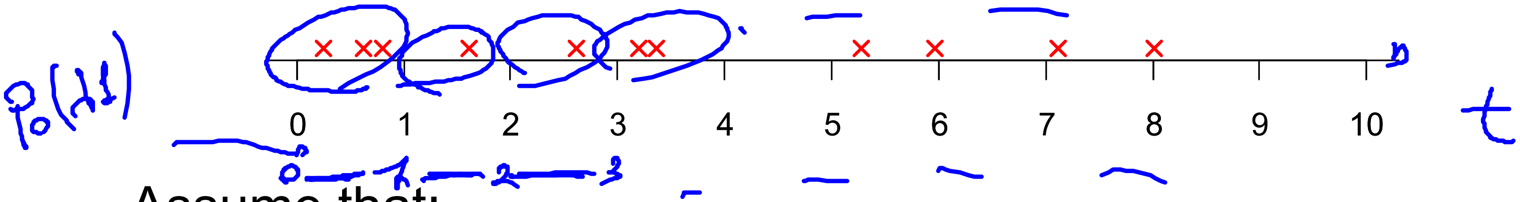
Illustration:

	Poisson	Binomial	Binomial	Binomial
		$n = 500$	$n = 50$	$n = 5$
x	$\lambda = 0.5$	$p = 0.001$	$p = 0.01$	$p = 0.1$
0	0.6065	0.6064	0.6050	0.5905
1	0.3033	0.3035	0.3056	0.3280
2	0.0758	0.0758	0.0756	0.0729
3	0.0126	0.0126	0.0122	0.0081
4	0.0016	0.0016	0.0015	0.0005

The Poisson distribution is often an appropriate model for "rare events"

Poisson process

We are observing events (marked by x) happening over time:



Assume that:

- the rate of events λ is constant over time
  (rate = expected number of events per unit of time)
- the number of events in disjoint time-intervals are independent
- events do not occur together

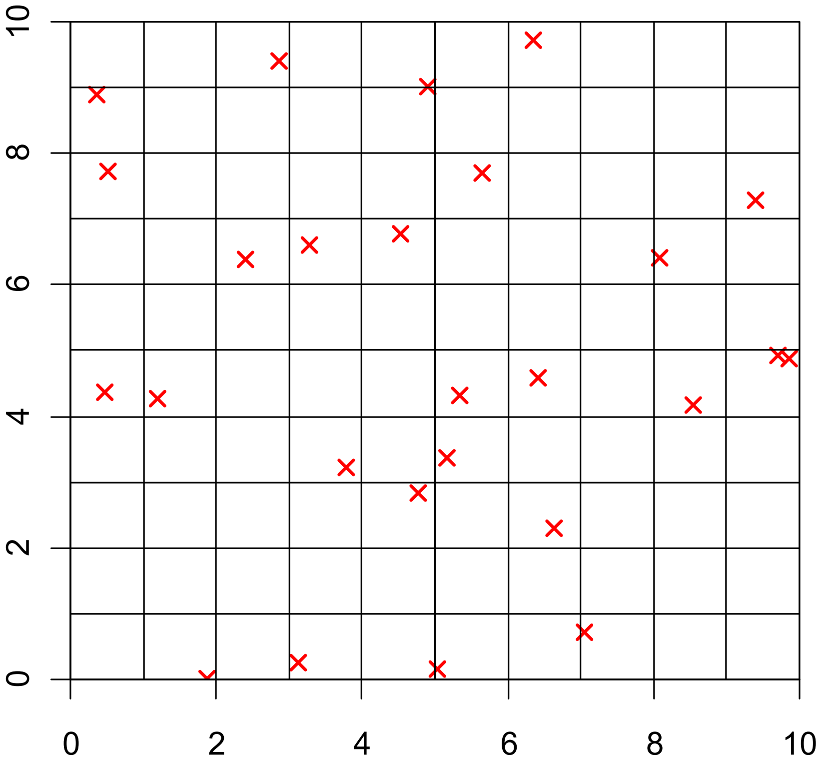
Then we have a Poisson process

The Poisson process is an appropriate model for events that are happening "randomly over time"

Let Y be the number of events in an interval of length t

Then: $Y \sim \mathrm{Po}(\lambda t)$

In a similar manner we may have a Poisson process in the plane:



Assume that:

- the rate of points λ is constant over the region (rate = expected number of points in an area of size one)
- the number of points in disjoint areas are independent
- points do not coincide

Then we have a Poisson process in the plane (spatial process)

This is a model for "randomly occurring" points

Let Y be the number of events in an area of size a

Then: $Y \sim \text{Po}(\lambda a)$

Overdispersion

For a Poisson distribution, the expected value and the variance are equal

$E(Y) = \lambda = Var(Y)$

$\bar{y}$ $\quad$ s^2

One way of checking whether the Poisson distribution is appropriate for a sample $y_1, y_2, ..., y_n$ is to compare

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{with} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

For a Poisson distribution both $\bar{y}$ and s^2 are estimates of λ so they should not differ too much

We may compute the coefficient of dispersion: $CD = \frac{s^2}{\bar{y}}$

If CD is (substantially) larger than 1, it is a sign of overdispersion

For the alpha particles we have

$$\bar{y} = 3.88 \quad \text{and} \quad s^2 = 3.70$$

which gives

$$CD = \frac{3.70}{3.88} = 0.95$$

For the anencephaly data we have

$$\bar{y} = 1.97 \quad \text{and} \quad s^2 = 2.41$$

which gives

$$CD = \frac{2.41}{1.97} = 1.22$$

The two examples do not show signs of overdispersion

Test of Poisson distribution

$Y \overset{?}{\sim} Po(\lambda)$

Data: $y_1, y_2, \ldots, y_n$

Null hypothesis H_0: data are Poisson distributed

	0	1	2	3	4	5	6	7	--
E_j	17	23	24	16	8	3	2	1	
O_j	16	22	27	21	9	4	3	2	

5-7

$\bar{E}$ 6

O 9

Procedure:

- Estimate (MLE): $\hat{\lambda} = \bar{y}$
- Compute expected frequencies under H_0 : $\boxed{E_j} = n \cdot \overbrace{\left(\hat{\lambda}^j / j!\right) e^{-\hat{\lambda}}}$
- Compute observed frequencies: $O_j =$ number of y_i equal to j
- Aggregate groups with small expected numbers, so that all E_j's are at least five. Let K be the number of groups thus obtained

$K=6$

- Compute Pearson chi-squared statistic: $\chi^2 = \sum \frac{\left(O_j - E_j\right)^2}{E_j}$
- Under H_0 the Pearson statistic is approximately chi-squared distributed with $K-2$ degrees of freedom

Example: Emission of alpha particles

There is a good agreement between observed and expected frequencies:

No.	0	1	2	3	4	5	6	7	8	9	10	11	12	13+
Observed	57	203	383	525	532	408	273	139	49	27	10	4	2	0
Expected	54	210	407	525	509	395	255	141	68	30	11	4	1	1

11+

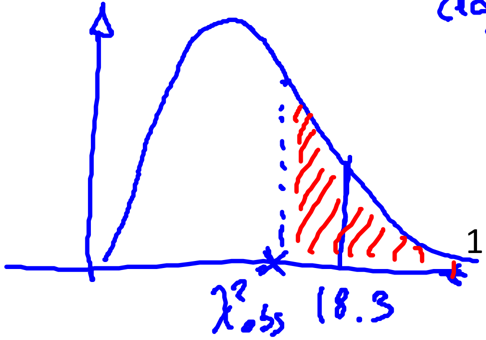
We aggregate the three last groups, leaving us with K = 12 groups

Pearson chi-squared statistic: $\chi^2 = 10.42$ $(df = 10)$

P-value: 40.4%

The Poisson distribution fits nicely to the data

df = 10



Example: Occurrence of anencephaly in Edinburgh 1956-66

Here as well there is a good agreement between observed and expected frequencies:

$\bar{y} = \hat{\lambda}$

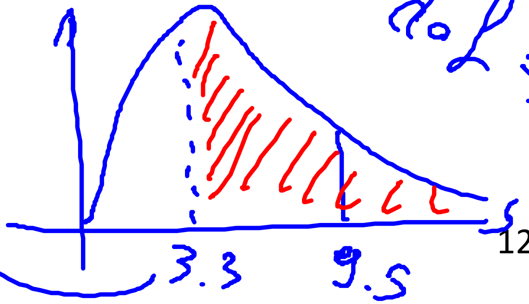
# anencephaly	0	1	2	3	4	5	6	7	8	9+
# observed	18	42	34	18	11	6	0	2	1	0
# expected	18.4	36.3	35.7	23.5	11.5	4.5	1.5	0.4	0.1	0.03

We aggregate the five last groups, leaving us with *K* = 6 groups

Pearson chi-squared statistic: $\chi^2 = 3.3$ $(df = 4)$

P-value: 50.9%

The Poisson distribution fits nicely to the data



d.o.f = 3

Example: Mite infestations on orange trees

A mite is capable of damaging the bark of orange trees

An inspection of a sample of 100 orange trees gave the following numbers of mite infestations found on the trunk of each tree:

# infestations	0	1	2	3	4	5	6	7	8+
# observed	55	20	21	1	1	1	0	1	0
# expected	44.5	36.0	14.6	3.9	0.8	0.13	0.02	0.00	0.00

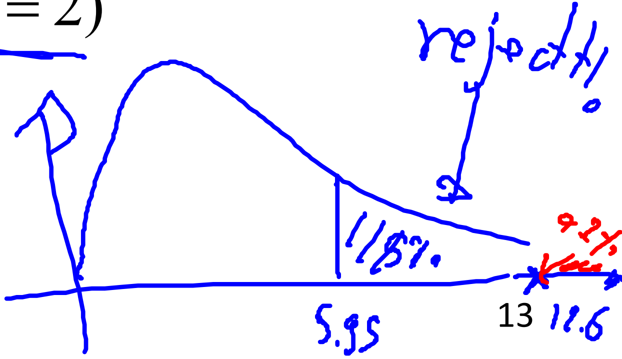
$$\sum \frac{y_i}{n} = \frac{20+42+3+4+5+7}{100} = \frac{81}{100} \quad 0.81 = \lambda$$

We aggregate the six last groups, leaving us with $K = 4$ groups

Pearson chi-squared statistic: $\chi^2 = 12.6 \qquad (df = 2)$

P-value: 0.2%

The Poisson distribution does not fit the data



O	55	20	21	4
E	44.5	36	14.6	4.85

Poisson regression

$N(\mu,\sigma^2)$ $\quad \mu = \beta_0 + \beta_1 x$

$Bi(n,p)$ $\quad p = \frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)}$

So far we have considered the situation where the observations are a sample from a Poisson distribution with parameter λ (which is the same for all observations)

We will now consider the situation where the Poisson parameter may depend on covariates, and hence is not the same for all observations

We assume that we have independent data for each of n subjects:

$$y_i, x_{1i}, x_{2i}, \ldots, x_{pi} \qquad i = 1, \ldots, n$$

$$y_i = \text{a count for subject no. } i$$

$$x_{ji} = \text{predictor (covariate) no. } j \text{ for subject no. } i$$

In general we assume that the responses y_i are realizations of independent Poisson distributed random variables $Y_i \sim \text{Po}(\lambda_i)$ where $\lambda_i = \lambda(x_{1i}, x_{2i}, \ldots., x_{pi})$ is a function of the covariates

We will consider regression models for the rates of the form:

$$\lambda_i = \lambda(x_{1i}, x_{2i}, \ldots, x_{pi})$$

$$= \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}\right)$$

$\lambda = \exp(\beta_0 + \beta_1 x)$

$\log(\lambda) = \beta_0 + \beta_1 x_1 + \ldots \beta_p x_p$

This ensures that the rates are positive, as they should

λ

If we consider two subjects with values $x_1 + \Delta$ and x_1, for the first covariate and the same values for all the others, their rate ratio (RR) becomes

$\Delta = 1$

$$\frac{\lambda(x_1+\Delta, x_2, \ldots, x_p)}{\lambda(x_1, x_2, \ldots, x_p)} = \frac{\exp(\beta_0 + \beta_1 (x_1+\Delta) + \beta_2 x_2 + \ldots + \beta_p x_p)}{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)} = e^{\beta_1 \Delta}$$

In particular e^{β_1} is the rate ratio corresponding to one unit's increase in the value of the first covariate *holding all other covariates constant*

In many applications we have data on an aggregated form

We then record counts for groups of individuals who share the same values of the covariates

Example: Insurance claims

We consider data on accidents in a portfolio of private cars in an English insurance company during a three months period

The variables in the data set are as follows:

- Age of the driver (1=less than 30 year, 2= 30 years or more)
- Motor volume of the car (1=less than 1 litre, 2=1-2 litres, 3=more than 2 litres)
- Number of insured persons in the group (defined by age and motor volume)
- Number of accidents in the group

	age	vol	num	acc
<30	1	1	846	137
	1	2	2421	444
	1	3	207	52
>30	2	1	4101	402
	2	2	14412	1869
	2	3	1372	247

When our observations are aggregated counts, an observation y_i is a realization of

$$Y_i \sim \text{Po}(w_i \lambda_i) \qquad (*)$$

$$e^{\log(x)} = x$$

$$e^{\log(w_i)} = w_i$$

where the weight w_i is the number of subjects in group i

When we combine (*) with the regression model on slide 14, we may write:

$$E[Y|X] = \lambda_i$$

$$E[Y|X] = w_i \lambda_i$$

$$E(Y_i) = w_i \boxed{\lambda_i}$$

$$E[Y_i|X_i] = w_i \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots. + \beta_p x_{pi}\right)$$

$$= \exp\left(\log(w_i) + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots. + \beta_p x_{pi}\right)$$

$$1 = \beta_{-1}$$

Formally $\log(w_i)$ is a "covariate" where the regression coefficient is known to equal 1. Such a "covariate" is called an offset

Example: Insurance claims

R commands:

```
car.claims=read.table("http://www.uio.no/studier/emner/matnat/math/STK4900/data/car-claims.txt", header=T)
fit.claims=glm(acc~offset(log(num))+factor(age)+factor(vol), data=car.claims,family=poisson)
summary(fit.claims)
```

R output (edited):

```
                  Estimate        Std. Error        z value         Pr(>|z|)
(Intercept)        -1.916          0.055           -34.83          < 2e-16
factor(age)2       -0.376          0.044            -8.45          < 2e-16
factor(vol)2        0.244          0.048             5.09          3.57e-07
factor(vol)3        0.570          0.072             7.90          2.85e-15
```

λ

$\lambda = e^{\beta_0}$

$e^{\beta_1} > 1$

Note e.g. that

- $e^{-1.916} = 0.147$ is expected number of claims for a driver younger than 30 years with a small car
- $e^{-0.376} = 0.687$ is the rate ratio for a driver 30 years or older compared with a driver younger than 30 years (with same type of car)

Maximum likelihood estimation

We have : $P(Y_i = y_i) = \frac{(w_i\lambda_i)^{y_i}}{y_i!}\exp(-w_i\lambda_i)$

$e^{\beta_0} = \exp\{\beta_0\}$

The likelihood is the simultaneous distribution

$$L = \prod_{i=1}^{n} \frac{(w_i\lambda_i)^{y_i}}{y_i!}\exp(-w_i\lambda_i)$$

$\exp(\beta_0 + \beta_x x, x_i \ldots)$

considered as a *function of the parameters* $\beta_0, \beta_1, \ldots, \beta_p$ for the observed values of the y_i

$L(\beta_0, \beta_1, \ldots \beta_p)$

The maximum likelihood estimates (MLE) $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ maximize the likelihood, or equivalently the log-likelihood $l = \log L$

Wald tests and confidence intervals

- $\hat{\beta}_j$ = MLE for β_j
- $se(\hat{\beta}_j)$ = standard error for $\hat{\beta}_j$

To test the null hypothesis $H_{0j}: \beta_j = 0$ we use the Wald test statistic:

$$z = \frac{\hat{\beta}_j - 0}{se(\hat{\beta}_j)}$$

which is approximately N(0,1)-distributed under H_{0j}

95% confidence interval for β_j: $\hat{\beta}_j \pm 1.96 \cdot se(\hat{\beta}_j)$

$RR_j = \exp(\beta_j)$ is the rate ratio for one unit's increase in the value of the *j*-th covariate *holding all other covariates constant*

We obtain a 95% confidence interval for RR_j by transforming the lower and upper limits of the confidence interval for β_j

Rate ratios with confidence intervals for the insurance example

R command (using the function from slide 10 of Lecture 7):

e

```
expcoef(fit.claims)
```

R output (edited):

```
                  expcoef        lower        upper
(Intercept)       0.15           0.13         0.16
factor(age)2      0.69           0.63         0.75
factor(vol)2      1.28           1.16         1.40
factor(vol)3      1.77           1.53         2.04
```

$$\exp\left(\hat{\beta}_0 - 1.56\ se(\hat{\beta}_0)\right)$$

$$\exp\left(\hat{\beta}_0 + 1.56\ se(\hat{\beta}_0)\right)$$

$$\hat{\beta}_0 \pm 1.56 \cdot se(\hat{\beta}_0)$$

Deviance and likelihood ratio tests

We want to test the null hypothesis H$_0$ that q *of* the β_j's are equal to zero, or equivalently that there are q linear restrictions among the β_j's

Procedure:

- $\tilde{l}$ is the maximum possible value of the log-likelihood, obtained for the saturated model with no restrictions on the λ_i
- $\hat{l} = \log \hat{L}$ is the log-likelihood for the full Poisson regression model
- $\hat{l}_0 = \log \hat{L}_0$ is the log-likelihood under H$_0$
- Deviances $D = 2(\tilde{l} - \hat{l})$ and $D_0 = 2(\tilde{l} - \hat{l}_0)$
- Test statistic $G = D_0 - D = -2\log\left(\hat{L}_0 / \hat{L}\right)$ is chi-squared distributed with q df under H$_0$

$G \sim \chi^2_q$

Example: Insurance claims

R commands:

```
fit.null=glm(acc~offset(log(num)), data=car.claims,family=poisson)
fit.age=glm(acc~offset(log(num))+factor(age), data=car.claims,family=poisson)
fit.age.vol=glm(acc~offset(log(num))+factor(age)+factor(vol), data=car.claims,family=poisson)
fit.interaction=glm(acc~offset(log(num))+factor(age)+factor(vol) +factor(age):factor(vol),
                    data=car.claims,family=poisson)
anova(fit.null,fit.age,fit.age.vol,fit.interaction,test="Chisq")
```

vol { 1 2 3 [illegible]

R output (edited):

```
Analysis of Deviance Table

Model 1: acc ~ offset(log(num))
Model 2: acc ~ offset(log(num)) + factor(age)
Model 3: acc ~ offset(log(num)) + factor(age) + factor(vol)
Model 4: acc ~ offset(log(num)) + factor(age) + factor(vol) + factor(age):factor(vol)

       Resid. Df      Resid. Dev      Df      Deviance      P(>|Chi|)
1          5            126.11
2          4             63.93        1        62.18        3.132e-15
3          2              1.98        2        61.95        3.534e-14
4          0              0.00        2         1.98        0.371
```

2

We end up with model 3 with no interaction (cf slides 17 and 19)

Generalized linear models

The models for

- Multiple linear regression
- Logistic regression
- Poisson regression

are the most common generalized linear models (GLMs)

A GLM consists of three parts

- A family of distributions
- A linear predictor
- A link function

Gaussian → linear regr.
Binomial → logistic regr
Poisson → Poisson regr.
Gamma → Gamma regr

Example GLM: (standard) Poisson-regression

The three parts are for Poisson-regression

- Family: The observations Y_i are independent and Poisson distributed with means $\mu_i = E(Y_i) \simeq \lambda_i$
- The linear predictor: A linear expression in regression parameters and covariates

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}$$

- The link function: Linking μ_i and η_i

$$\eta_i = g(\mu_i) = \log(\mu_i)$$

$f(x) = x$

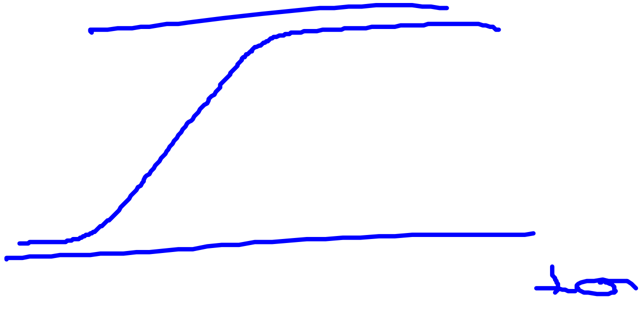
$\eta = \mu$

For the multiple linear regression model the family is normal and the link function is an identity function $\eta_i = g(\mu_i) = \mu_i$

For logistic regression: binary / binomial family and link function is the logit function

$$\eta_i = g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$$

Other link functions may also be specified:



For binary responses:

- Complementary log-log link:
  $\eta_i = g(\mu_i) = \log(-\log(1-\mu_i))$
- Probit link: $\eta_i = g(\mu_i) = \Phi^{-1}(\mu_i)$ where $\Phi(z)$ is the cumulative N(0,1)-distribution

For Poisson responses:

- Identity link: $\eta_i = g(\mu_i) = \mu_i$
- Square root link: $\eta_i = g(\mu_i) = \sqrt{\mu_i}$

Statistical inference in GLMs is performed as illustrated for logistic regression and Poisson regression

$$\mathcal{L} = \prod_{i=1}^{n} \binom{1}{y_i} p_i^{y_i} (1-p_i)^{1-y_i}$$

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

Estimation:

- Maximum likelihood (MLE)

$$H_0: \beta_1 = 0$$

Testing:

- Wald tests
- Deviance/likelihood ratio tests

$$Z = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim N(0;1)$$

$$\chi^2_q$$

A particular feature of the GLMs is the variance function $V(\mu)$ which is specific for each family of distributions. The variance functions describe how the variance depends on the mean μ.

- For the Poisson distribution: $V(\mu) = \mu$ $= \lambda$
- For binary data: $V(\mu) = \mu(1-\mu)$
- For normal data we define $V(\mu) = 1$ since the variance does not depend on the mean thus $\mathrm{Var}(Y_i) = \sigma^2 1 = \sigma^2 V(\mu_i)$

1

μ

$\parallel$

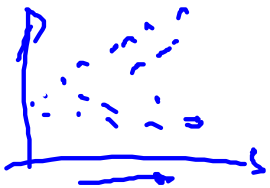
$E[Y] = p$

$\mathrm{Var}(Y) = p(1-p)$

$\sigma^2 \cancel{1} \mu$

As discussed previously in these slides there may be overdispersion relative to a Poisson model. This could be allowed for by specifying a model $\mathrm{Var}(Y_i) = \phi V(\mu_i)$

ϕ



Example: Number of sexual partners

Study of sexual habits, Norwegian Institute of Public Health, 1988

Y_i = no. sex-partners, $i = 1, \ldots, n = 8553$

A Poisson-regression found that the expected value increased with

- Age, being single, having had HIV-test and was higher for men

However, the data was overdispersed. A “Pearson X2” statistic is

$$X^2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} = 51927$$

which is large compared with residual degrees of freedom 8544. An overdispersion term is estimated as

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} = \frac{51927}{8544} = 6.08$$

and should have been close to 1 if the Poisson model was correct.

Standard errors and inference needs correction for overdispersion!

Correction for overdispersion

A overdispersed Poisson model is given by

- $\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots. + \beta_p x_{pi}$
- $\mathrm{Var}(Y_i) = \phi\mu_i$

This model can be fitted as a standard Poisson-regression, but the standard errors must be corrected to

$$se^* = se\sqrt{\hat{\phi}}$$

where se is the standard error from the Poisson-regression and the overdispersion $\hat{\phi}$ is estimated as on the previous slide. Similarly the z-values become

$$z^* = z / \sqrt{\hat{\phi}}$$

and p-values must be corrected correspondingly

$$z = \frac{\hat{\beta}}{se(\hat{\beta})}$$

$$z^* = \frac{\overset{z}{\overbrace{\hat{\beta}}}}{se(\hat{\beta})\sqrt{\hat{\phi}}} = \frac{z}{\sqrt{\hat{\phi}}}$$

Count data with over-dispersion – Quasi-likelihood

Although the corrections for overdispersion shown on the previous slide should be simple to carry out it is convenient that it is already implemented in R through a so-called

- Quasi-likelihood

The family-specification in the glm-command is given as “quasi” with arguments

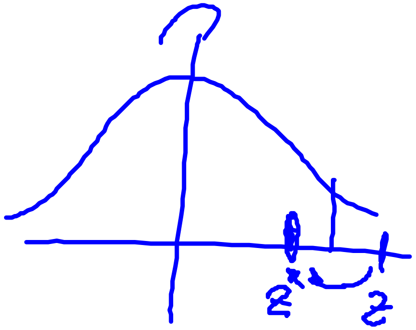
- var=“mu”
- link=log

```
glm(partners~Gender+Married+factor(HIVtest)+factor(agegr),
     family=quasi(link=log,var="mu"),data=part)
```

Results from over-dispersed Poisson model on no. of sexual partner data.

```
                    Estimate  Std. Error  z value   Pr(>|z|)

(Intercept)         1.82862   0.07665   23.857   < 2e-16 ***
Gender             -0.49038   0.02145  -22.859   < 2e-16 ***
Married            -0.43997   0.02521  -17.449   < 2e-16 ***
factor(HIVtest)2 0.35017   0.03254   10.763   < 2e-16 ***
factor(HIVtest)3 0.14901   0.05657    2.634   0.00845 **
factor(agegr)2    0.57142   0.06721    8.502   < 2e-16 ***
factor(agegr)3    0.90489   0.06767   13.372   < 2e-16 ***
factor(agegr)4    1.04673   0.06550   15.981   < 2e-16 ***
factor(agegr)5    0.84322   0.06806   12.389   < 2e-16 ***
---
(Dispersion parameter for quasi family taken to be 6.07765)
Null deviance:       53136 on 8553 degrees of freedom
Residual deviance: 40002 on 8544 degrees of freedom
```



Although the associations are still all strongly significant they have been scaled down a factor $2.45 = \sqrt{6.08} = \sqrt{\hat{\phi}}$ σ

Heteroscedastic linear model

1
2
3
4

Assume that the linear structure

$$E(Y_i) = \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots. + \beta_p x_{pi}$$

was found acceptable, but that the variance depended on μ_i as

$$\mathrm{Var}(Y_i) \approx \phi \mu_i$$

One way to handle the non-constant variance could then be to specify a quasi-likelihood model with identity link and variance function “mu”

R can also handle variance structures $\phi\mu^2$ and $\phi\mu^3$

Generalized additive models (GAM)

We have encountered GAMs for

$$E[Y|X] = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots$$

- Multiple linear regression
- Logistic regression

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots$$

Any generalized linear model (GLM) can be extended to a GAM including Poisson regression models

A GAM consists of three parts

- A family of distributions
- A link function
- An additive predictor

$N(\mu, \sigma^2)$

$Bi(1, p)$

$Po(\lambda)$

I

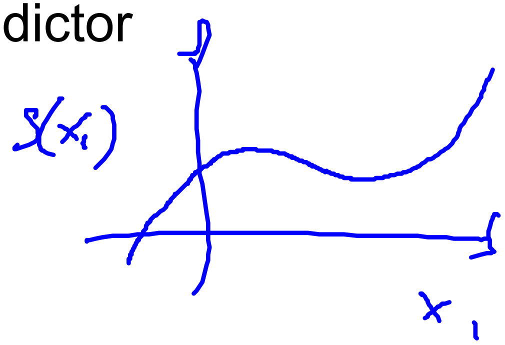
logit

[illegible]

~~Linear~~

GAM, continued

Thus the first two components of a GAM are the same as for a GLM, *but* for the last component we replace the linear predictor



linear predictor →

$$\eta_i = \beta_0 + \underline{\beta_1}\,\underline{x_{1i}} + \beta_2\, x_{2i} + \ldots + \beta_p\, x_{pi}$$

with an additive predictor

additive predictor →

$$\eta_i = \beta_0 + \underline{f_1(x_{1i})} + \underline{f_2(x_{2i})} + \ldots + f_p(x_{pi})$$

where the linear terms $\beta_j\, x_{ji}$ are replaced by smooth functions $f_j(x_{ji})$

Before fitting and plotting a GAM-model the library gam must be invoked (and installed).

Examples of use of GAM is found in Lecture 5, slide 19 and Lecture 7, slide 35.

library(gam)