Generalized Linear Models

Lecture 5. Models with continuous response

Continuous response

In practice, there are many situations where the response is continuous but the normal distribution does not fit:

- response can only take nonnegative values, $Y \geq 0$
- conditional probability that $Y \geq k$ increases when k increases
- variance is not constant

Examples:

- insurance claims
- time to failure (of an experiment or a machine)
- amount of rainfall

The distributions that may suit in given scenarios are **gamma, lognormal** or **inverse Gaussian** distributon

Coefficient of variation

Definition. Coefficient of variation

Coefficient of variation of a random variable Y is the ratio of its standard deviation to the mean

$$CV = \frac{\sqrt{\mathbf{D}Y}}{\mathbf{E}Y} = \frac{\sigma}{\mu}\,, \quad \hat{CV} = \frac{s_y}{\bar{y}}$$

Coefficient of variation

- is a dimensionless measure of variability (often given in %)
- allows to compare the variability of variables measured in different scales
- is also called *relative standard deviation* (RSD)
- if $\mu \to 0$ then $CV \to \infty$, i.e. it is sensitive to values of μ

CV is often used in situations where the r.v. of interest is (related to) exponential
Then $\sqrt{\mathbf{D}Y} = \mathbf{E}Y \quad \Rightarrow \quad CV = 1$
$CV < 1$ (Erlang distribution) – small relative variability
$CV > 1$ (hyperexp. dist. = mixture of exponentials) – large relative variability

Coefficient of variation for known distributions

Distribution	$\mathbf{E}Y$	$\mathbf{D}Y$	CV
$N(\mu, \sigma^2)$	μ	σ^2	$\frac{\sigma}{\mu}$
$B(n, \pi)$	$n\pi$	$n\pi(1-\pi)$	$\sqrt{\frac{1-\pi}{n\pi}}$
$Po(\lambda)$	λ	λ	$\frac{1}{\sqrt{\lambda}}$
$\Gamma(\alpha, \lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\frac{1}{\sqrt{\alpha}}$
$Exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	1

Distributions with constant CV

Consider random variables that are positive and right skewed

In that case, constant variance might not be the best assumption (think of different scales, for example), but distributions with constant CV might be quite useful

In other words, let us assume that $CV = \alpha$, which also means that **there is a linear relation between the std. deviation and the mean**:

$$\sqrt{\mathbf{D}Y} = \alpha \mathbf{E}Y$$

A known distribution with constant coefficient of variance is Gamma distribution (in what sense?)

Gamma distribution $\quad Y \sim \Gamma(\nu, \frac{\nu}{\mu})\ \ (Y \sim \Gamma(\alpha, \lambda))$

Gamma distribution is a non-negative continuous distribution with the following characteristics:

- Pdf: $f(y; \mu, \nu) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1} \exp(-\frac{\nu}{\mu}\, y), \quad y \geq 0,$

  where $\Gamma(\nu)$ is the Gamma function: $\Gamma(\nu) = \int_0^{\infty} x^{\nu-1} \exp(-x) dx$
- Mean: $\mathbf{E}Y = \frac{\nu}{\nu/\mu} = \mu$
- Variance: $\mathbf{D}Y = \frac{\mu^2}{\nu}$
- Coefficient of variation: $CV = \frac{1}{\sqrt{\nu}}$
- Skewness: $\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2}{\sqrt{\nu}} > 0$, since $\mu_3 = 2\ \frac{\mu^3}{\nu^2}$.

  In other words, the distribution is skewed to the right

Relation between parametrizations: $\alpha = \nu$, $\lambda = \frac{\nu}{\mu}$

Some known properties of the gamma function:
$\Gamma(1) = 1; \quad \Gamma(\nu) = (\nu - 1)\Gamma(\nu - 1),\ \nu > 1; \quad \Gamma(n) = (n-1)! \quad \Gamma(1/2) = \sqrt{\pi}$

Some properties of gamma distribution

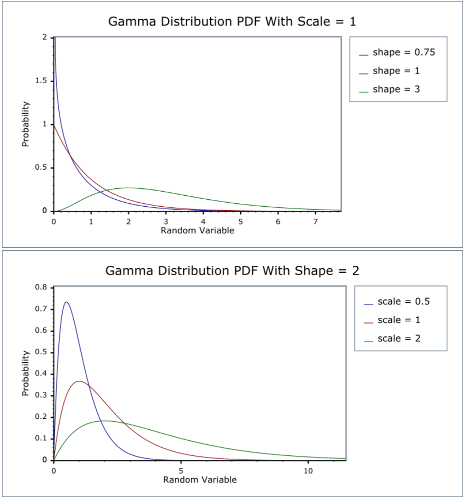
Gamma distribution has two parameters:

- shape parameter $\nu > 0$
- rate (or inverse scale) parameter $\lambda = \frac{\nu}{\mu}$ (given in this form to stress the relation to the mean)

Properties:

- if $\lambda = \frac{\nu}{\mu} = \frac{1}{2}$ then we have $\chi^2_{2\nu}$-distribution
- if $\nu = 1$ then we have exponential distribution
- if $0 < \nu \leq 1$, the density is monotone decreasing, otherwise unimodal and stretched to the right
- if $Y_i \sim \Gamma(\nu_i, \lambda)$ are independent, their sum has gamma distribution with shape $\sum_{i=1}^{n} \nu_i$ and rate λ
- if $\nu \to \infty$, gamma distribution converges to normal distribution

Density of gamma distribution



Example of gamma distribution (1)

Let us look at the duration of something which can be considered as a sum of latent period, each of which has exponential distribution with fixed intensity. Then the total duration follows gamma distribution

Lindsey (1995). Data from Liege, Belgium 1984, the duration of marriages ($n = 1699$)
Marriage is sometimes thought of as having three periods of different relationships through which a couple goes, so that gamma distribution with $\nu = 3$ might be suitable
Empirical mean is $\bar{y} = 13.85$ years and variance is $s_y^2 = 75.9$
Parameter estimates: $\hat{\nu} = \frac{\bar{y}^2}{s_y^2} = 2.53$; $\hat{\mu} = 13.85$
Histogram (next slide) shows a good fit

Example of gamma distribution (2)

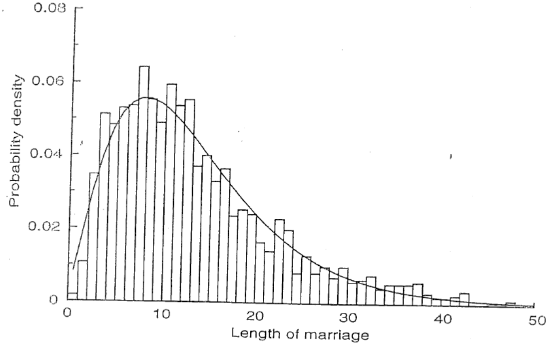


Fig. 4.20. Histogram and gamma distribution for the divorce data

Source: Lindsey (1995). Introductory Statistics. A modelling approach. Oxford

Gamma distribution as a member of exponential family

Let us now show that gamma distribution belongs to the exponential family:
Recall the pdf

$$f(y_i;\mu_i,\nu) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_i}\right)^{\nu} y_i^{\nu-1}\exp(-\frac{\nu}{\mu_i}\ y_i),$$

which can be rewritten as

$$f(y_i;\mu_i,\nu) = \exp\left\{\left(-\frac{y_i}{\mu_i} - \ln\mu_i\right)\nu - \ln\Gamma(\nu) + \nu\ln\nu + (\nu-1)\ln y_i\right\}$$

- $\theta_i = -\frac{1}{\mu_i}; \quad b(\theta_i) = -\ln(-\theta_i); \quad \varphi_i = \frac{1}{\nu}$
- Mean: $\mathbf{E}(Y|\mathbf{x}_i) = b'(\theta_i) = -\frac{1}{\theta_i} = \mu_i,\ \mu_i > 0$
- Variance function: $b''(\theta_i) = \frac{1}{\theta_i^2} = \mu_i^2$
- Variance: $\mathbf{D}(Y|\mathbf{x}_i) = \sigma_i^2 = \varphi_i b''(\theta_i) = \frac{\mu_i^2}{\nu}$ (depends on the mean)
- Coefficient of variation: $CV = \frac{\sigma_i}{\mu_i} = \frac{1}{\sqrt{\nu}}$ (is constant)

Deviance

If the shape parameter ν is known, the log-likelihood can be written as:

$$l(\mu_i; y_i, \nu) = \left(-\frac{y_i}{\mu_i} - \ln \mu_i\right)\nu - \ln\Gamma(\nu) + \nu \ln \nu + (\nu - 1)\ln y_i$$

Thus, the sample log-likelihood of a current model is

$$l(\hat{\boldsymbol{\mu}}; \boldsymbol{y}, \nu) = \nu \sum\left(-\frac{y_i}{\hat{\mu}_i} - \ln \hat{\mu}_i\right) + C$$

and for a saturated model (i.e. $\hat{\mu}_i = y_i$)

$$l(\boldsymbol{y}; \boldsymbol{y}, \nu) = -\nu \sum(1 + \ln y_i) + C$$

Deviance for gamma model

$$D(\boldsymbol{y}, \hat{\boldsymbol{\mu}}) = -2\sum\left(\ln \frac{y_i}{\hat{\mu}_i} - \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)$$

and $D \overset{a}{\sim} \varphi\chi^2_{n-p}$, where $\varphi = \frac{1}{\nu}$

Link functions

Canonical link of the gamma model is $g(\mu_i) = -\frac{1}{\mu_i}$
Nevertheless, the default link function offered in most statistical packages is the inverse function:

$$g(\mu_i) = \frac{1}{\mu_i}$$

Why?

Note that by construction we have a restriction $\mu_i > 0$, which also implies $\eta_i > 0$

Other possible choices:

- log-link $g(\mu_i) = \ln \mu_i$ (very often used)
- identity link $g(\mu_i) = \mu_i$ (NB! restriction $\mu_i > 0$)
- power function link $g(\mu_i) = \mu_i^{\alpha}$ (also, e.g. $\sqrt{\mu_i}$)

Example. Gamma model with inverse responses (1)

One of the most widely used models with gamma distributed response (physics, biochemistry) has the following form:

$$\frac{1}{y} = a + b\frac{1}{x}$$

and is called Briggs/Haldane equation or Michaelis/Menten/Hollings equation

We recognize this is a model with inverse linear response:

$$\eta = \beta_0 + \frac{\beta_1}{x} \quad \Rightarrow \; \mu_i = \frac{x}{\beta_0 x + \beta_1}$$

which gives us hyperbolic form for μ against x with slope $\frac{1}{\beta_1}$ and asymptote $\frac{1}{\beta_0}$

See also next slide, graph (a), where $\beta_0 = 1, \; \beta_1 = 1$, i.e. the asymptote is 1

Nelder, 1966

Example. Gamma model with inverse responses (2)

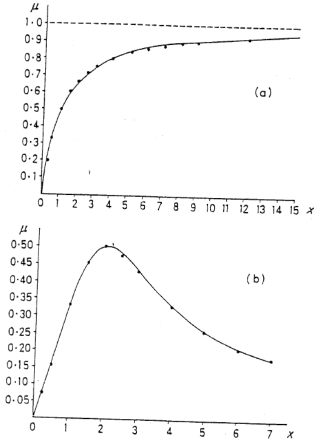


Fig. 7.3. *Inverse polynomials: (a) the inverse linear* $\mu^{-1} = 1 + x^{-1}$; *(b) the inverse quadratic* $\mu^{-1} = x - 2 + 4x^{-1}$.

(a) Inverse linear response, (b) Inverse quadratic response

Example. Gamma model with inverse responses (3)

If we include term x in the model, we obtain

$$\eta = \beta_0 + \frac{\beta_1}{x} + \gamma x,$$

which gives inverse quadratic link (see previous slaid, graph (b))

Slope at the origin is $\frac{1}{\beta_1}$, maximum is reached at $x = \sqrt{\frac{\beta_1}{\gamma}}$

On graph $\beta_0 = -2,\ \beta_1 = 4,\ \gamma = 1$, maximum is reached at $x = 2$

Example. Hurn , 1945

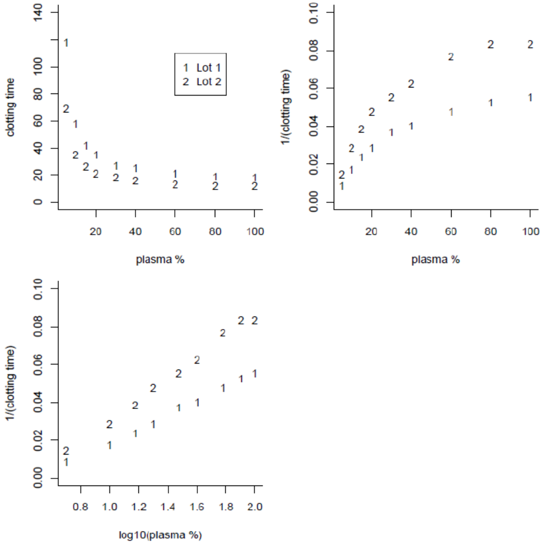
Data about blood clotting
y – blood clotting time (sec); $L = \{LOT1, LOT2\}$ – 2 different scenarios
u – plasma concentration (%)

u	*Lot1*	*Lot2*
5	118	69
10	58	35
15	42	26
20	35	21
30	27	18
40	25	16
60	21	13
80	19	12
100	18	12

Bliss (1970) estimated a hyperbolic model using gamma distribution and log-link
Results ($x = \log_{10} u$, see next slide for the explanation of this transform):

$$LOT1: \ \hat{\mu}^{-1} = -0.0166 + 0.0153x; \quad LOT2: \ \hat{\mu}^{-1} = -0.0239 + 0.0236x$$

Example. Hurn , 1945



Measures of explained variation (R^2) for Gamma model

Two estimates for R^2 are proposed:

- estimate based on residual sums of squares:

$$R^2_{SS} = 1 - \frac{RSS(\hat{\boldsymbol{\mu}})}{RSS(\bar{\boldsymbol{\mu}})}$$

- estimate based on deviances:

$$R^2_D = 1 - \frac{D(\boldsymbol{y}, \hat{\boldsymbol{\mu}})}{D(\boldsymbol{y}, \bar{\boldsymbol{\mu}})}$$

Adjustments by degrees of freedom and certain shrinkage factor to given estimates is also proposed

M. Mittlböck, H. Heinzl (2002). Measures of explained variation in Gamma regression models. *Commun. Statist. – Simulat. and Comput.*, 31(1), 67-73.

Estimation of the shape parameter (and model scale)

If the shape parameter ν is not known, we can find ν and φ using the maximum likelihood method (derivative of log-likelihood by ν, ...),
which is approximately

$$\hat{\varphi} = \hat{\nu}^{-1} = \frac{D(\boldsymbol{y}, \hat{\boldsymbol{\mu}})}{n}$$

or, a bias-corrected version:

$$\hat{\varphi} = \hat{\nu}^{-1} = \frac{D(\boldsymbol{y}, \hat{\boldsymbol{\mu}})}{n - p}$$

These estimates are sensitive to small values of y_i and are undefined if $y_i = 0$

The moment-based Pearson χ^2 is thus often preferred

$$\hat{\varphi} = \hat{\nu}^{-1} = \frac{\chi^2}{n - p}$$

R: `summary(model)$disp` gives Pearson estimate for dispersion,
`gamma.shape(model)` uses iterations, starts from deviance based estimate

Remarks to gamma models

- If the response has small CV (even 0.6), it is difficult to distinguish normal model for $\log y$ and log-linear gamma model (i.e. gamma with *log*-link) (Atkinson, 1982)
- In situations where censoring is possible (reliability and survival data), gamma models may have issues in practical applications

Lognormal model

Definition. Lognormal distribution

Random variable Y has lognormal distribution if its logarithm $Z = \ln(Y)$ has normal distribution

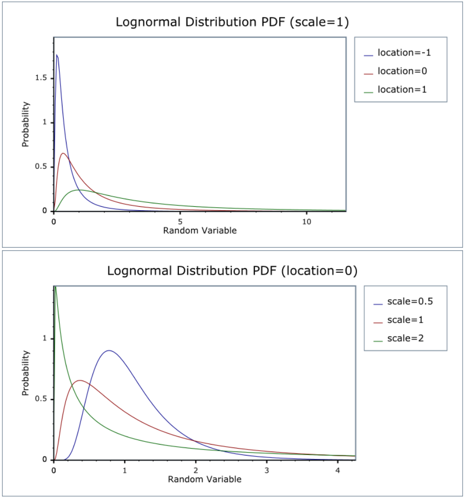
If $Z \sim N(\mu, \sigma^2)$ and $Z = \ln(Y)$ then $Y \sim LN(\mu, \sigma^2)$

- Pdf: $f_Y(y; \mu, \sigma) = \frac{1}{y} f_Z(\ln y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\{-\frac{(\ln y - \mu)^2}{2\sigma^2}\}$, $y > 0$
- Expectation: $\mathbf{E}Y = \exp(\mu + \frac{1}{2}\sigma^2)$
- Variance: $\mathbf{D}Y = \exp(2\mu + \sigma^2)(\exp\sigma^2 - 1)$
- Coefficient of variation: $CV = \sqrt{\exp\sigma^2 - 1}$

Variance is proportional to the square of the mean (like gamma distribution)

In case of small values of σ^2 ($CV < 0.7$), it is quite close to gamma distribution

Pdf of lognormal distribution



Usage of lognormal distribution

Lognormal is widely used in different areas, e.g.

- measurements in biology and medicine (size of living tissue (length, skin area, weight), length of inert appendages (hair, claws, nails, teeth) of biological specimens, blood pressure of adult humans)
- extreme events in hydrology (daily rainfall and river discharge volumes)
- income of most of the population (the distribution of higher-income individuals follows a Pareto distribution)
- financial and insurance models (claim amounts, Black-Scholes model, changes in the logarithm of exchange rates, price indices, and stock market indices are assumed normal)
- size of cities
- different processes in technology

Lognormal distribution and exponential family

Question

Does lognormal distribution belong to the exponential family?

Answer:

- Based on the definition of natural 1-parameter exponential family, log-normal distribution does not belong to this family
- On the other hand, lognormal distribution belongs to 2-parameter exponential family with natural parameters and natural statistics, respectively, given by

$$\left(-\frac{1}{2\,\sigma^2}, \frac{\mu}{\sigma^2}\right) \quad \text{and} \quad (\ln^2(Y), \ln(Y))$$

Inverse Gaussian model

Definition. Inverse Gaussian distribution

Random variable Y has inverse Gaussian distribution, $Y \sim IG(\mu, \lambda)$ if its pdf has the following form:

$$f(y; \mu, \lambda) = \left[\frac{\lambda}{2\pi y^3}\right]^{1/2} \exp\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\}\ ,\ y > 0$$

- Mean: $\mathbf{E}Y = \mu$
- Variance: $\mathbf{D}Y = \frac{\mu^3}{\lambda}$ – variance is proportional to mean cubed
- Coefficient of variation: $CV = \sqrt{\frac{\mu}{\lambda}}$

NB! The name can be misleading: "inverse" means that while the Gaussian describes a Brownian motion's level at a fixed time, the inverse Gaussian describes the distribution of the time a Brownian motion with positive drift takes to reach a fixed positive level.

Usage of inverse Gaussian distribution

IG has sharper peak and heavier tails (as compared to lognormal), thus it is used in areas related to more extreme events

- insurance
- financial mathematics
- meteorology (wind energy applications)

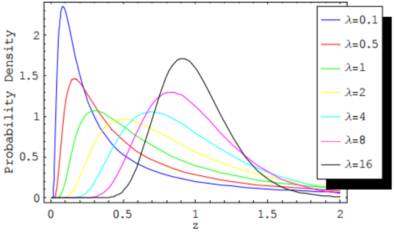
Its hazard function is $\cap$-shaped as for lognormal and Weibull distributions, thus IG is also used in

- survival analysis
- risk analysis (e.g. analysis of noise effects)

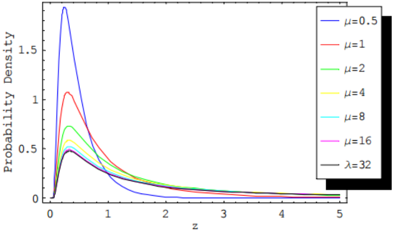
IG is first mentioned by Schrödinger (1915) and Smoluchowski (1915), name "inverse Gaussian" is proposed by Tweedie (1941)

Same class of distributions is discussed by Wald (1947). If $\mu = 1$, inverse Gaussian is called Wald's distribution

Pdf of inverse Gaussian distribution $IG(\mu, \lambda)$



A) $\mu = 1$ and varying λ



B) Varying μ and $\lambda = 1$

Source: Matsuda, K (2005). Inverse Gaussian Distribution.

Inverse Gaussian as a member of exponential family

Let us rewrite the pdf

$$f(y_i;\mu_i,\lambda)=\left[\frac{\lambda}{2\pi y_i^3}\right]^{1/2}\exp\{-\frac{\lambda(y_i-\mu_i)^2}{2\mu_i^2 y_i}\}\ ,\ y>0$$

in the following form:

$$\begin{aligned} f(y_i;\mu_i,\lambda) &= \exp\{-\frac{\lambda(y_i-\mu_i)^2}{2\mu_i^2 y_i}+\frac{1}{2}\ln(\frac{\lambda}{2\pi y_i^3})\} \\ &= \exp\left\{\frac{-y_i/2\mu_i^2+1/\mu_i}{1/\lambda}-\frac{\lambda}{2y_i}+\frac{1}{2}\ln(\frac{\lambda}{2\pi y_i^3})\right\} \end{aligned}$$

Now

- canonical parameter: $\theta_i=-\frac{1}{2\mu_i^2}$
- $b(\theta_i)=-\frac{1}{\mu_i}=-(-2\theta_i)^{\frac{1}{2}}$
- $\varphi_i=\frac{1}{\lambda}$
- $b'(\theta_i)=(-2\theta_i)^{-\frac{1}{2}}=\mu_i$
- $b''(\theta_i)=\mu_i^3$

Prove it!

Link functions used for inverse Gaussian model

Canonical link: $g(\mu_i) = -\frac{1}{2\mu_i^2}$ – not used very often in this exact form

Default link in most statistical packages: squared inverse, $g(\mu_i) = \frac{1}{\mu_i^2}$, which implies $\mu_i = \frac{1}{\sqrt{\eta_i}}$

Other possibilities:

- log-link $g(\mu_i) = \ln \mu_i$ – often used, especially when squared inverse has convergence or negativity issues
- identity $g(\mu_i) = \mu_i$ – always a simple choice, problems if $\eta_i < 0$

Deviance

Sample log-likelihood for inverse Gaussian:

$$l(\boldsymbol{\mu}; \boldsymbol{y}, \lambda) = \sum_i \left\{ \frac{-y_i/2\mu_i^2 + 1/\mu_i}{1/\lambda} - \frac{\lambda}{2y_i} + \frac{1}{2}\ln(\frac{\lambda}{2\pi y_i^3}) \right\}$$

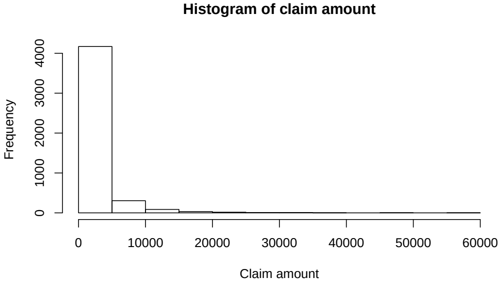
Thus the deviance is (as $\varphi = \frac{1}{\lambda}$)

$$\begin{aligned} D &= -\frac{2}{\lambda}\{l(\hat{\boldsymbol{\mu}}; \boldsymbol{y}, \lambda) - l(\boldsymbol{y}; \boldsymbol{y}, \lambda)\} \\ &= 2\sum_i \left\{ \left(\frac{y_i}{2\hat{\mu}_i{}^2} - \frac{1}{\hat{\mu}_i} \right) - \left(\frac{y_i}{2y_i^2} - \frac{1}{y_i} \right) \right\} \\ &= \sum_i \left\{ \frac{y_i}{\hat{\mu}_i^2} - \frac{2}{\hat{\mu}_i} + \frac{1}{y_i} \right\} = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2} \end{aligned}$$

Example. Australian insurance claims 2004-2005

Data: 67856 insurance policies, 4624 claims
How to model the individual claim amount?

claimcst0 – claim amount (0 if no claim) (min 200, max 55922)
gender – gender of driver: M, F
area – driver's area of residence: A, B, C, D, E, F
agecat – driver's age category: 1 (youngest), 2, 3, 4, 5, 6



Example. Results (1)

```
> m1 = glm(claimcst0~factor(agecat)+gender+area,
                      data=claims, family="inverse.gaussian"(link="log"))
> summary(m1)
...
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      7.70839    0.09853  78.230  < 2e-16 ***
factor(agecat)2 -0.15845    0.10349  -1.531 0.125836
...
factor(agecat)6 -0.31865    0.12076  -2.639 0.008349 **
genderM          0.15283    0.05119   2.986 0.002846 **
areaB           -0.02976    0.07287  -0.408 0.682977
...
areaF            0.35539    0.13049   2.723 0.006485 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
...
AIC: 77162
```

How to interpret the results?

Example. Results (2)

```
> m2 = glm(claimcst0~factor(agecat)+gender+area,
                     data=claims, family="Gamma")

> m3 = glm(log(claimcst0)~factor(agecat)+gender+area,
                          data=claims, family="gaussian")
> m1$aic
[1] 77162.32

> m2$aic
[1] 79331.75

> m3$aic
[1] 14707.79
```

Which model is best?

Lognormal model revisited. Log-likelihood

Sample likelihood for lognormal:

$$L(\boldsymbol{y}; \boldsymbol{\mu}, \sigma) = \prod_i \frac{1}{y_i} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(\ln y_i - \mu_i)^2}{2\sigma^2})$$

Sample log-likelihood for lognormal:

$$l(\boldsymbol{y}; \boldsymbol{\mu}, \sigma) = -\sum_i \ln(y_i) - n \ln(\sigma\sqrt{2\pi}) - \sum_i \frac{(\ln y_i - \mu_i)^2}{2\sigma^2}$$

NB! Compare with the log-likelihood of normal applied to log-sample:

$$l(\ln \boldsymbol{y}; \boldsymbol{\mu}, \sigma) = -n \ln(\sigma\sqrt{2\pi}) - \sum_i \frac{(\ln y_i - \mu_i)^2}{2\sigma^2}$$

Lognormal model. AIC

$$AIC = -2\text{log-likelihood} + 2p$$

AIC for lognormal:

$$AIC_{LN} = -2\left(-\sum_i \ln(y_i) - n\ln(\sigma\sqrt{2\pi}) - \sum_i \frac{(\ln y_i - \mu_i)^2}{2\sigma^2}\right) + 2p$$

AIC for normal applied to log-sample:

$$AIC_N = -2\left(-n\ln(\sigma\sqrt{2\pi}) - \sum_i \frac{(\ln y_i - \mu_i)^2}{2\sigma^2}\right) + 2p$$

$\Rightarrow AIC_{LN} = AIC_N + 2\sum_i \ln(y_i)$

Lognormal model. How to proceed?

We just need to calculate $2\sum_i \ln(y_i)$ and add it to the AIC found for normal model

```
> sum(log(claims$claimcst0))
[1] 31489.81

> m3$aic + 2*sum(log(claims$claimcst0))
[1] 77687.41
```

Thus the correct value for lognormal model is 77687

Recall that the other models had $AIC_{IG} = 77162$ and $AIC_{Gamma} = 79331$

These numbers are now comparable and we conclude that inverse Gaussian fits the data best

Conclusive remarks

Some remarks about residual analysis

- Normal distribution: constant variance (for standardized residuals)
- If variance varies
  – in case of gamma (and lognormal) variance is proportional to squared mean,
  – in case of inverse Gaussian variance is proportional to cubed mean
- Graphs:
  a) in case of log-link, argument vs $\log y_i$ should be linear
  b) in case of inverse link, argument vs $1/y_i$ should be linear