Non-normal models

Other discrete and continuous distributions can serve as the basis for "AOV", regression, etc. The key statistical concepts are likelihood, maximum likelihood estimation, and likelihood ratio tests.

Sums of squares, "least squares", etc. are not used (they arise from the normal distribution).

Likelihood (R. A. Fisher): the **likelihood** (or likelihood function) is the probability model for the data, evaluated at the data. One can think of it as the probability, if the random process could be repeated, that the particular outcome represented by the data would reoccur.

ex. $Y \sim$ binomial(n, π)

$$P(Y = y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}$$

data: $n = 10$, $y = 6$

$$L = \frac{10!}{6!4!}\pi^6(1-\pi)^4$$

(L depends on the value of the unknown parameter π)

ex. Random sample from a Poisson distribution (sample plots in a field; count dandelions in each plot) $Y_1, Y_2, ...,$ Y_n , where $Y_i \sim$ Poisson(μ)

$$P(Y_i = y) = \frac{e^{-\mu}\mu^y}{y!} \qquad (y = 0, 1, 2, ...)$$

data: 0, 3, 0, 1, 4, 6, 2, 1

$$L = P(Y_1 = 0 \textit{ and } Y_2 = 3 \textit{ and } \cdots \textit{ and } Y_8 = 1)$$

$$= P(Y_1 = 0)P(Y_2 = 3)\cdots P(Y_8 = 1)$$

$$= \frac{e^{-\mu}\mu^0}{0!}\frac{e^{-\mu}\mu^3}{3!}\cdots\frac{e^{-\mu}\mu^1}{1!}$$

ex. Random sample from a normal distribution
$Y_1, Y_2, ..., Y_n$, where $Y_i \sim \text{normal}(\mu, \sigma^2)$

data: 18, 21, 23, 22, 17, 20, 21, 19

(probability model for a random sample from a continuous distribution is the product of probability density functions)

$$f(y) = \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad \text{(pdf for normal)}$$

$$L = f(18) f(21) \cdots f(19) =$$

$$\frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(18-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(21-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(19-\mu)^2}{2\sigma^2}}$$

$$= \left(\sigma^2 2\pi\right)^{-\frac{8}{2}} e^{-\frac{1}{2\sigma^2}\left[(18-\mu)^2 + (21-\mu)^2 + \cdots + (19-\mu)^2\right]}$$

Maximum likelihood parameter estimates (R. A. Fisher): the values of the unknown parameters that maximize the likelihood are the **maximum likelihood** (ML) estimates.

ML estimates have good statistical properties (small variances, approx. unbiased, etc.)

ex. Binomial $L = \frac{10!}{6!4!}\pi^6(1-\pi)^4$

calculus: $\widehat{\pi} = \frac{6}{10}$ ML

ex. Poisson random sample $y_1, y_2, ..., y_n$

$$\widehat{\mu} = \overline{y}$$

ex. Normal random sample $y_1, y_2, ..., y_n$

$\widehat{\mu}$ minimizes $(y_1 - \mu)^2 + (y_2 - \mu)^2 + \cdots + (y_n - \mu)^2$

$$\widehat{\mu} = \overline{y}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\left[(y_1 - \widehat{\mu})^2 + (y_2 - \widehat{\mu})^2 + \cdots + (y_n - \widehat{\mu})^2\right]$$

(ML uses n, not $n-1$, in denominator)

Maximized likelihood: $\widehat{L}$ is the likelihood value calculated using the ML estimates of the parameters.

ex. Binomial $\widehat{L} = \frac{10!}{6!4!}\left(\frac{6}{10}\right)^6\left(1-\frac{6}{10}\right)^4$

ex. Normal

$$\widehat{L} = \left(\widehat{\sigma}^2 2\pi\right)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

Likelihood ratio hypothesis test (J. Neyman, E. S. Pearson, S. S. Wilks)

$\theta = (\theta_1, \theta_2, ..., \theta_r)$ parameters of interest in a model (ex. $\mu_1, \mu_2, ..., \mu_t$ in an AOV)

$\psi = (\psi_1, \psi_2, ..., \psi_s)$ other parameters in the model ("nuisance parameters", ex. σ^2 in an AOV)

H_0: parameters in θ are constrained (ex. $\mu_1 = \mu_2 = \cdots = \mu_t = \mu$)

H_a: parameters are not constrained

$\widehat{L}_0$: likelihood maximized under model H_0

$\widehat{L}_a$: likelihood maximized under model H_a

$\widehat{L}_0/\widehat{L}_a$: likelihood ratio comparing the two models (if large, H_0 favored; if small, H_a favored)

Theorem (Wilks): if H_0 is true, then

$$G^2 = -2\log_e\left(\frac{\widehat{L}_0}{\widehat{L}_\text{a}}\right) \underset{\text{approx}}{\sim} \text{chi-square}(v)$$

where v = (# parameters estimated under H_a) − (# parameters estimated under H_0)

ex. 3 fields, model $Y_{ij} \sim \text{Poisson}(\mu_i)$ $i = 1, 2, 3$

field	n_i	observations	$\hat{\mu}_i$
1	6	2, 7, 3, 2, 4, 6	4.0
2	8	1, 1, 1, 2, 2, 0, 1, 1	1.125
3	7	2, 4, 3, 0, 2, 4, 1	2.2857143

$\hat{\mu} = 2.3333333$

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu$

H_a: $\mu_1 \neq \mu_2 \neq \mu_3$

$\log_e\left(\widehat{L}_0\right) = -39.86337$ $\log_e\left(\widehat{L}_\text{a}\right) = -33.823003$

$$G^2 = -2\left[\log_e\left(\widehat{L}_0\right) - \log_e\left(\widehat{L}_\text{a}\right)\right] = 12.080748$$

df = 3 − 1 = 2 $\chi^2_{0.05} = 5.992$

$G^2 > \chi^2_{0.05}$ $\therefore$ reject H_0