Financial Theory
+ Corporate Policy. Thomas
Copeland + F. Weston Addison Wesley
(3rd Edition
1988)

CREDIBILITY

(Rough study guides)

The study guides form background material for a lecture on Credibility given at the 1991 CAS Seminar on Ratemaking held on March 14-15 in Chicago, Illinois.

Ben Zehnwirth
Insureware Pty Ltd
15 Sidwell Avenue
E. St. Kilda, Vic 3183
Australia
Tel: +61 3 5273084
Fax: +61 3 5272349

PREFACE

These notes have grown out of a num of sets of lecture notes prepared for statistical and actuarial courses at Macquarie University and the Universit of Copenhagen. The notes are for- from perfect and far from complete, may be regarded as the first draf towards a text on credibility.

The notes are primarily intended to provide an introductory set of lectu on the subject of credibility and it intimate connections with linear regressi Bayes estimation and recursive estimation.

Sections marked with single asterisk may be omitted at first reading wi impairing the reader's understanding subsequent sections. More difficult sections are indicated by double asterisks.

References

Bailey, A.L. (1945). A generalized theory of credibility. PCAS, 32, 13-20.

Buhlmann, H. & Straub, E. (1970). Glaubwurdigk für Schadensatze. Mitteilungen der Vereinigung Schweizerischer Versicherun Mathematiker, 70, 111-133.

de Jong, P. and Zehnwirth, B. (1983). Credibility theory and the Kalman filter. Insurance: Mathematics and Economics, 2 281-286.

Jewell, W.S. (1974). Credible means are exa Bayesian for simple exponential families. ASTIN Bulletin, 8, 77-9

——— (1976). A survey of credibility. Research Report, 76-31 ORD. Uni of C Berkeley.

Whitney, A.W. (1918). The theory of experie rating. PCAS, 4, 274-29

Zehnwirth, B. (1977). The mean credibility formula is a Bayes rule. Scand. Actuar. J. 212-216.

——— (1983). Linear filtering and recursive credibility estimation. ASTIN Bulletin, 15, 19 – 35.

TABLE OF CONTENTS

§5. CREDIBILITY

§6. BUHLMANN-STRAUB WITH UNEQUAL WEIGHTS

§0. INTRODUCTION

Credibility theory (and indeed practice) forms the cornerstone of casualty actuarial mathematics. Given that credibility is intimately related to a number of fields in statistics and engineering, it should come as no surprise that the first credibility formulae were developed by Gauss in 1795.

The concept of ordinary least squares is inextricably linked with the German mathematician Karl Friedrich Gauss. Gauss derived formulae for updating regression estimates, in particular, sample means.

So, section 1 develops updating formulae for sample means using straightforward algebra. It is claimed that most credibility formulae may be derived using recursions (1·2·6).

THIS MEANS THAT GAUSS (1795) DERIVED MOST CREDIBILITY FORMULAE

Since credibility estimators are regression estimators, section 2 discusses the connections between the bivariate normal distribution, linear regression and ordinary least squares.

Gauss's formulae of section 1 are based on updating estimates using sample information. Since Bayesian estimation does not distinguish between the prior information being based on a sample or based on objectivity, Bayesian estimation can be derived using Gauss's formulae — See Section 4

So, we have the following schemeta

$$
\begin{tikzcd}
\text{Regression} \arrow[r, leftrightarrow] \arrow[d] & \text{Bayes} \arrow[r, leftrightarrow] \arrow[d] & \text{Credibility (exact)} \arrow[d] \\
\text{Linear Regression} \arrow[r, leftrightarrow] & \text{Linear Bayes} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] \arrow[dl, leftrightarrow] & \text{(Linear) Credibility} \\
\text{Conjugate priors} & \text{Sample Updating} \arrow[d, leftrightarrow] & \\
& \text{Gauss Sample Means} &
\end{tikzcd}
$$

Credibility is introduced in §5 and it becomes Immediate that

$$\begin{array}{ccc} \text{(Exact) Credibility} & \longleftrightarrow & \text{Bayes} \\ \downarrow & & \downarrow \\ \text{(Linear) Credibility} & \longleftrightarrow & \text{Linear Bayes} \end{array}$$

Bailey's (1945) generalized credibility work was well ahead of its time.

Bailey (1945) recognized the connection between credibility and linear regression and also ~~between~~ the connection between his model (which turns out to be the same as the collective risk model) and Fisher's one-way ANOVA (Analysis Of Varian

Bailey derived the credibility formula, known as the Buhlmann - Straub (1970) formula. But Bailey derived it in 194 presumably, because he used the linear regression connection.

Credibility formulae are optimal estimators derived from the postulated model.

Now, the Buhlmann - Straub model (or Bailey's model) are rarely appropriate. Why?

This is because there is abundant empirical evidence that suggests that risk parameters vary with time - hence evolutionary credibility models, and moreover may be 'explained' by various explanatory variables - hence credibility regression models à la Hachmeister.

In any case, it is always important to test the postulated model. That is subsequent to parameter estimation we need to ensure that the data does not violate the model assumptio

Evolutionary models and regression credibility models will be described in forthcoming sections of these notes.

§1. SIMPLEST STATISTICAL MODEL

(1·1) The simplest statistical model is the normal distribution. Consider a random sample $y_1, \dots, y_n$ from a normal distributio with mean μ and variance σ^2. We write:

$$y_1, \dots, y_n \overset{iid}{\sim} N(\mu, \sigma^2).$$

Equivalently,

$$y_i = \mu + \epsilon_i, \qquad (1\cdot1\cdot1)$$

where $\epsilon_1, \dots, \epsilon_n \overset{iid}{\sim} N(0, \sigma^2)$.

Equation (1·1·1) describes the simplest regression model in statistics where the independent variable X is unity.

The least squares estimator (lse) of μ is obtained by minimizing

$$J = \sum_{i=1}^{n} (y_i - \mu)^2 . \qquad (1\cdot1\cdot2)$$

Now,

$$\frac{dJ}{d\mu} = -2 \sum_{i=1}^{n} (y_i - \mu) .$$

So the 'optimal' estimator is

$$\hat{\mu}_n = \sum_{i=1}^{n} y_i / n . \qquad (1\cdot1\cdot3)$$

That is, the sample mean $\bar{Y}_n = \sum_{i=1}^{n} y_i / n$ is the lse of μ!

Note that $Var(\bar{Y}_n) = \frac{\sigma^2}{n}$.

(1.2) Recursive estimation

How is the estimator $\bar{Y}_n$ modified (or updated) on receipt of an additiona observation y_{n+1} ?

We have:

$$\hat{\mu}_n = \bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i \quad ;$$

$$\hat{\mu}_{n+1} = \bar{Y}_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} y_i .$$

So, $$(n+1)\hat{\mu}_{n+1} = \sum_{i=1}^{n+1} y_i$$

$$= \sum_{i=1}^{n} y_i + y_{n+1}$$

$$= n\hat{\mu}_n + y_{n+1} .$$

$$\therefore \quad \hat{\mu}_{n+1} = \frac{n}{n+1} \hat{\mu}_n + \frac{1}{n+1} \cdot y_{n+1}$$

i.e. $$\boxed{\hat{\mu}_{n+1} = \hat{\mu}_n + \frac{1}{n+1} (y_{n+1} - \hat{\mu}_n)} \quad (1.2.$$

Moreover,

$$\mathrm{Var}(\hat{\mu}_{n+1}) = \frac{\sigma^2}{n+1}$$
$$= \left(1-\frac{1}{n+1}\right)\cdot\frac{\sigma^2}{n}$$
$$= \left(1-\frac{1}{n+1}\right)\mathrm{Var}(\hat{\mu}_n). \quad (1.2.$$

Equation (1.2.1) tells us the the "new" or "updated" or "revised" estimate is

"the old estimate" + "weight" or "credibility" times "the prediction error"

Since $\hat{\mu}_n$ would be our prediction of Y_{n+1} (before we observe it), the term $Y_{n+1} - \hat{\mu}_n$ is called (the one step ahead) prediction error.

Put $Z_n = \frac{1}{n}$ and $C_n = \frac{\sigma^2}{n} = Z_n\sigma^2$.

We have:

$$n+1 = n + 1,$$

So, $$Z_{n+1}^{-1} = Z_n^{-1} + 1 \qquad (1.2.3)$$

So, $$Z_{n+1} = \frac{Z_n}{Z_n + 1}$$

So, we have

$$Z_{n+1} = \frac{Z_n}{1+Z_n}$$

$$= \frac{C_n \sigma^{-2}}{1 + C_n \sigma^{-2}}$$

i.e. $$\boxed{Z_{n+1} = \frac{\sigma^{-2}}{C_n^{-2} + \sigma^{-2}}} \qquad (1.2.4)$$

Therefore the "weight" or "credibility" assigned to the new observation or information Y_{n+1} is proportional to its relative precision vis a vis the estimator $\hat{\mu}_n$, for

$$Var(Y_{n+1}) = \sigma^2$$

and $$Var(\hat{\mu}_n) = C_n .$$

We can also write

$$C_{n+1} = Z_{n+1} \sigma^2$$

$$= \frac{Z_n \sigma^2}{1+Z_n}$$

$$= \frac{1}{1+Z_n} \cdot C_n$$

i.e. $$\boxed{C_{n+1} = (1 - Z_{n+1}) C_n} \qquad (1.2.5$$

Summary of results

1. $\hat{\mu}_{n+1} = \hat{\mu}_n + Z_{n+1}(X_{n+1} - \hat{\mu}_n)$

2. $Z_{n+1} = Z_n (1+Z_n)^{-1}$

$= C_n (C_n + \sigma^2)^{-1}$

3. $C_{n+1} = (1-Z_{n+1})C_n$ (1·2·6)

4. $C_{n+1}^{-1} = C_n^{-1} + (\sigma^2)^{-1}$

where,

$$Z_n = \frac{1}{n} \quad \text{and} \quad C_n = Var(\hat{\mu}_n) = Z_n \sigma^2 .$$

Since the above results are based on updating sample means and their corresponding variances, they are straightforward to derive.

INDEED, (1·2·6) are credibility formulae for model (1·1·1) and MOREOVER are very much like all credibility formulae. Many credibility formulae may be derived using (1·2·6)!!!!

§2. ~~SIMPLE~~ LINEAR REGRESSION AND RECURSIVE ESTIMATION

Preliminaries

(2.1) Bivariate normal distribution

Two random variables (Y, X) are said to have a bivariate normal distributio

$\iff$ $a_1 Y + a_2 X$ has a univariate norma distribution for all constants a_1 and a

The joint density of (Y, X) is quite involved (and does not have to be remembere and is given by

$$f_{Y,X}(y,x) = \frac{1}{2\pi\sigma_Y\sigma_X\sqrt{1-\rho^2}} \cdot \exp\left[-\tfrac{1}{2}Q(y,x)\right]$$

where the quadratic form Q is given by

$$Q(y,x) = \frac{1}{1-\rho^2}\left[\frac{(y-\mu_Y)^2}{\sigma_Y^2} + \frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(Y-\mu_Y)(X-\mu_X)}{\sigma_Y\sigma_X}\right]$$

where

$$\mu_Y = E[Y]\,;\ \mu_X = E[X]\,;\ \sigma_X^2 = \mathrm{Var}[X]\,;$$

$$\sigma_Y^2 = \mathrm{Var}[Y] \text{ and } \rho = \frac{\mathrm{cov}(Y,X)}{\sigma_Y\sigma_X} = \frac{\sigma_{YX}}{\sigma_Y\sigma_X}$$

ρ measures the linear association between Y and X.

$\rho = \pm 1 \iff \exists$ constants a and b

$: Y = aX + b.$

$\rho = 1 \iff a > 0.$

$\rho = -1 \iff a < 0.$

For a given $X = x$, there is a subpopulation of values of Y. The conditional distribution of Y given $X = x$ is also normal.

The regression of Y on X, equivalently the conditional mean of Y given X is

$$\mu_{Y|X} = E[Y|X] = \mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(X - \mu_X)$$

$$= \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X). \quad (2.1.1)$$

(See Venter page)

The conditional variance

$$\sigma^2_{Y|X} = Var[Y|X]$$

$$= \sigma_Y^2 - \sigma_{YX}\,\sigma_X^{-2}\,\sigma_{YX}$$

$$= \sigma_Y^2 - \rho^2 \sigma_Y^2. \quad (2.1.2)$$

It is important to observe that the variance $Var[y|X]$ is <u>not</u> a function of the observed X. Accordingly,

$$E[Var[y|X]] = Var[y|X] . \qquad (2.1.3)$$

The conditional variance,

$$\left.\begin{aligned} \sigma^2_{y|x} &= E[(y - E[y|X])^2 | X] \\ &= E[Var(y|X)] \\ &= E[(y - E(y|x))^2] \end{aligned}\right\} (2.1.4)$$

$$= MSE \quad (\text{See } \S 3.2).$$

Equations (2.1.1), (2.1.2), (2.1.3) and (2.1.4) are <u>extremely</u> important. They have significance in the relationship between

$$\begin{array}{ccccc} \text{REGRESSION} & \longleftrightarrow & \text{BAYES} & \longleftrightarrow & \text{CREDIBILI} \\ \downarrow & & \downarrow & & \downarrow \\ \text{LINEAR REGRESSION} & \longleftrightarrow & \text{LINEAR BAYES} & \longleftrightarrow & \text{LINEAR CREDIBILI} \end{array}$$

So we see that if (Y, X) is normal then

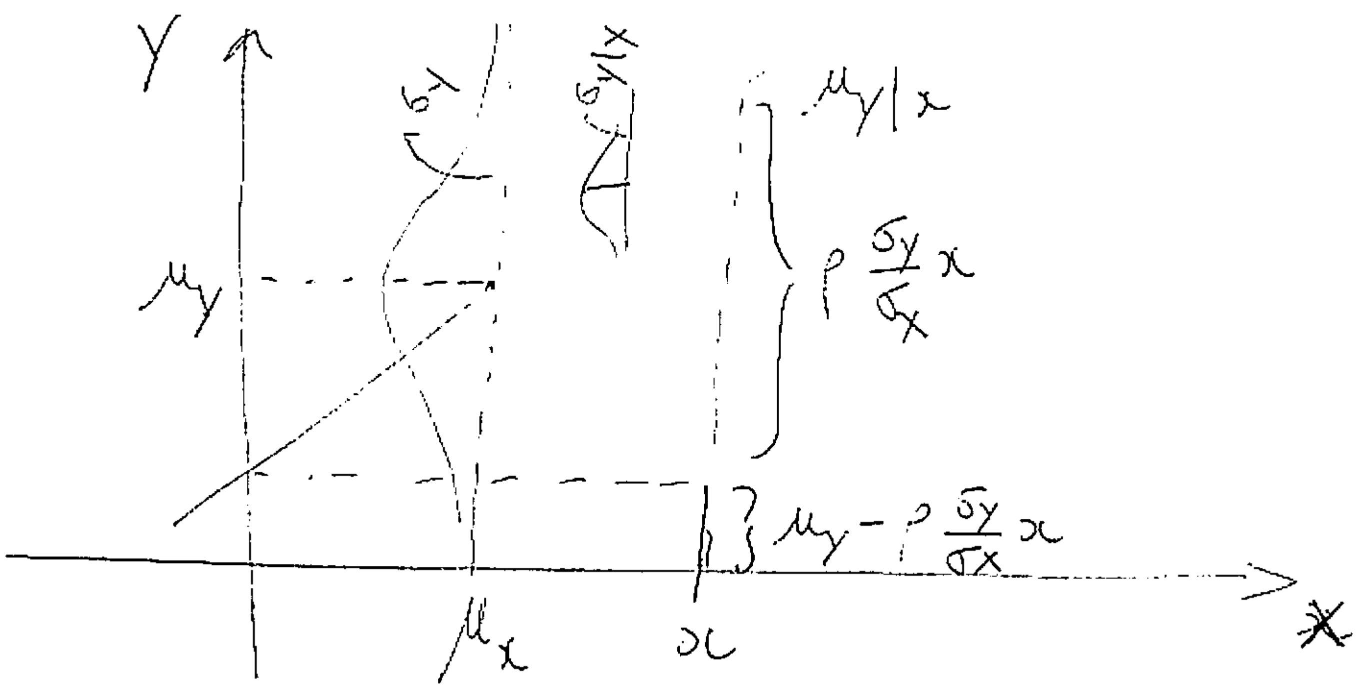
(i) the regression function $E[y|X]$ (See §   ) is <u>linear</u> in X.

(ii) $$MSE = E\,Var[y|X]$$
$$= Var[y|X],$$ where $Var[y|X]$ is <u>not</u> a function of the observed data X. (See § and Venter page   ).

Consider the conditional variance

$$\sigma^2_{Y|X} = \sigma^2_y (1-\rho^2) \qquad (2.1.5)$$

The variance of y reduces when we know the corresponding value of X.



$$\mu_{Y|x} = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_x)$$

$$\sigma^2_{Y|X} = \sigma^2_Y (1 - \rho^2)$$

$$< \sigma^2_Y$$

σ^2_Y = "TOTAL VARIATION IN Y"

$\sigma^2_{Y|X}$ = "VARIATION IN Y ABOUT X"

So, $\sigma^2_Y - \sigma^2_{Y|X}$ = "VARIATION IN Y EXPLAINED BY X"

So, $1-\rho^2 = \dfrac{\sigma^2_{y|x}}{\sigma^2_y}$

$$\Rightarrow \rho^2 = \frac{\sigma^2_y - \sigma^2_{y|x}}{\sigma^2_y} \qquad (2.1.k$$

= "PROPORTION OF VARIATION IN Y EXPLAINED BY X"

Note,

$$\sigma^2_y = \sigma^2_y - \sigma^2_{y|x} + \sigma^2_{y|x} \qquad (2.1.$$

I.e.

TOTAL VARIATION = VARIATION DUE TO X + RESIDUAL VARIATION

"TOTAL SS" = "REGRESSION SS" + "ERROR SS"

See § 2.2

2.2 The Simple linear regression model

Suppose (Y, X) are jointly normal and consider $\epsilon = Y - \mu_{Y/x}$. The random variable ϵ measures the deviation of Y from the line, alternatively its mean at $X=x$. The mean of ϵ at $X=x$ is ϕ, and the variance is $\sigma^2_{Y/x}$.

We can write:

$$Y = \mu_{Y|x} + \epsilon$$

$$= \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_x) + \epsilon$$

$$= \underbrace{\mu_Y - \rho \frac{\sigma_Y}{\sigma_X}\mu_x}_{\alpha} + \underbrace{\rho \frac{\sigma_Y}{\sigma_x}}_{\beta} x + \epsilon$$

$$\boxed{\begin{array}{l} Y = \alpha + \beta x + \epsilon, \text{ where} \\ \epsilon \sim N(0, \sigma^2_{y|x}) \end{array}} \quad (2\text{-}2)$$

Equation (2.2.1) is known as the simple linear regression model.

$$Y = \alpha + \beta x + \epsilon \; : \; \epsilon \sim N(0, \sigma^2_{Y|x}) ,$$

Note that
$$\left.\begin{aligned} \beta &= \rho \frac{\sigma_Y}{\sigma_X} = \frac{cov(Y,X)}{\sigma^2_X} \\ \text{and} \qquad & \\ \alpha &= \mu_Y - \beta \mu_X \end{aligned}\right\} \quad (2.2.2)$$

Since $\sigma^2_{Y|x}$ is <u>not</u> a function of x, it is customary to denote it by σ^2, keeping in mind that it is the conditional variance. When $\rho = 0$, $\sigma^2_{Y|x} = \sigma^2_Y$, so that none of the variation in y is explained by x.

2.3. SAMPLE ESTIMATION OF THE POPULATION PARAMETERS α, β AND σ^2.

We now discuss the estimation of the population parameters. We assume that we have a random sample $(x_1, y_1), \ldots, (x_n, y_n)$. The sample may arise in one of two ways.

(a) The values of X are arbitrarily fixed at $X = x_1$, $X = x_2, \ldots, X = x_n$; so that for $X = x_i$ we have a sub-population of Y values (that has a normal distribution). y_i is a random observation from that sub-population

(b) The values $(x_1, y_1), \ldots, (x_n, y_n)$ are random realizations from the joint distribution of (X, Y).

The regression model is a <u>conditional</u> model:

$$Y = \alpha + \beta x + \epsilon : \quad \epsilon \sim N(0, \sigma^2)$$

$$y_i = \alpha + \beta x_i + \epsilon_i : \quad \epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} N(0, \sigma^2)$$

(2.3.1)

Recall that σ^2 here denotes the variance of Y conditional on X.

From equation (2.2.2) we have,

$$\beta = \rho \frac{\sigma_y}{\sigma_x} = \frac{\sigma_{yx}}{\sigma_x^2} , \qquad (2.3.2)$$

$$\alpha = \mu_y - \beta \mu_x , \qquad (2.3.3)$$

and

$$\sigma^2 = \sigma_y^2 (1 - \rho) . \qquad (2.3.4)$$

σ^2 denotes the <u>conditional</u> variance of Y, whereas σ_y^2 is the (unconditional) variance of Y.

Let's use "moment" estimators for all parameters:

$$\hat{\sigma}_y^2 = \frac{1}{n-1} \Sigma (y_i - \bar{y})^2 \qquad (2.3.5)$$

(Since denominator is $n-1$, $\hat{\sigma}_y^2$ is <u>not</u> a moment estimator)

$$\hat{\sigma}_{yx} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{n-1}, \qquad (2\cdot3\cdot6)$$

$$\hat{\sigma}_x^2 = \frac{\Sigma(x_i-\bar{x})^2}{n-1}, \qquad (2\cdot3\cdot7)$$

$$\hat{\mu}_x = \bar{x}, \qquad (2\cdot3\cdot8)$$

$$\hat{\rho} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\,\Sigma(y_i-\bar{y})^2}},$$

$$= \frac{\hat{\sigma}_{yx}}{\hat{\sigma}_x\,\hat{\sigma}_y}. \qquad (2\cdot3\cdot9)$$

Using these estimators, we obtain:

$$\hat{\beta} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} \qquad (2\cdot3\cdot10)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x} \qquad (2\cdot3\cdot11)$$

and since

$$\epsilon_i = y_i - \alpha - \beta x_i,$$

$$\hat{\epsilon}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$$

we can estimate σ^2 by

$$\hat{\sigma}^2 = \frac{\Sigma \hat{\epsilon}_i^2}{n-2} \qquad (2\text{-}3\cdot 12)$$

Since formulae (2-3·10) and (2·3·11) are well known to readers, it is straightforward to memorize (2·3·1), (2-3-2) and (2-3-3), and (2·1·1).

Equation (2·1·1) is a credibility formula.
See § and Venter page

Now,

$$\Sigma(y_i-\bar{y})^2 = \Sigma(y_i-\bar{y})^2 - \Sigma(y_i-\hat{y}_i)^2 + \Sigma(y_i-\hat{y}_i)$$

i.e., $\Sigma(y_i-\bar{y})^2 = \Sigma(\hat{y}_i-\bar{y})^2 + \Sigma(y_i-\hat{y}_i)^2$

Analogously, (2.3.

$$\sigma_y^2 = \sigma_y^2 - \sigma^2 + \sigma^2$$

TOTAL SS = REGRESSION SS + ERROR SS

We provide a short proof of (2.3.13).

Consider $\hat{\epsilon}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$.

We have $\sum_{i=1}^{n} \hat{\epsilon}_i = n\bar{y} - n\hat{\alpha} - \hat{\beta}\bar{x}$

$= 0$.

Also, $\dfrac{\Sigma \hat{y}_i}{n} = \dfrac{\Sigma(\hat{\alpha}+\hat{\beta}x_i)}{n}$

$= \hat{\alpha} + \hat{\beta}\bar{x}$

i.e, $\bar{\hat{y}} = \bar{y}$.

Most importantly,

$$y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y}$$

So,

$$\Sigma(y_i-\bar{y})^2 = \Sigma(y_i-\hat{y}_i)^2 + CT + \Sigma(\hat{y}_i-\bar{y})^2,$$

where

$$CT = 2\,\Sigma(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

Now,

$$\Sigma(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

$$= \Sigma(y_i - \bar{y} - \hat{\beta}(x_i-\bar{x}))\,\hat{\beta}(x_i-\bar{x})$$

$$= \hat{\beta}\,\Sigma(y_i-\bar{y})(x_i-\bar{x}) - \hat{\beta}^2\,\Sigma(x_i-\bar{x})^2$$

$$= 0.$$

2.4 Ordinary least squares

We now find ordinary least squares estimators (olse) of the unknown parameters α and β based on our sample $(x_1, y_1), \ldots, (x_n, y_n)$.

The olse $\hat{\alpha}$ and $\hat{\beta}$ are obtained by minimizing the sum of squares of deviation

$$J = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \qquad (2.4.1$$

with respect to α and β.

$$\left.\begin{aligned} \frac{\partial J}{\partial \alpha} &= -2 \sum (y_i - \alpha - \beta x_i) = 0 \\ \frac{\partial J}{\partial \beta} &= -2 \sum x_i (y_i - \alpha - \beta x_i) = 0 \end{aligned}\right\} \quad (2.4.2$$

The ~~foregoing~~ equations (2.4.2) are called the <u>normal</u> (nothing to do with normal distribution) equations.

Solving the normal equations we obtain

$$\hat{\beta} = \frac{\Sigma(y_i - \bar{y})(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2}, \qquad (2\cdot4\cdot3)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad . \qquad (2\cdot4\cdot4)$$

These are the same as the 'moment' estimators (2·3·10) and (2·3·11).

2.5 Multiple linear regression

We consider the problem of predicting Y from p variables $X_1, .., X_p$. If $Y, X_1, .., X_p$ are jointly normal then the conditional mean of Y given $X_1, .., X_p$ is <u>linear</u> in $X_1 \ldots X_p$. See §2.1 and § . We are therefor motivated to consider the following model

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + \epsilon_i , \quad i = 1, .., n \qquad (2.5$$

where $\epsilon_1, .., \epsilon_n \overset{iid}{\sim} N(0, \sigma^2)$. The variance σ^2 is the <u>conditional</u> variance of Y.

Let $\underset{\sim}{y} = (y_1, .., y_n)'$, $\underset{\sim}{\beta} = (\beta_0, .., \beta_p)$, $\underset{\sim}{\epsilon} = (\epsilon_1, .., \epsilon_n)'$ and the design matrix X

$$= \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & & x_{pn} \end{bmatrix}$$

We can write (2.5.1) thus

$$\underline{y} = X \underline{\beta} + \underline{e} \qquad (2.5.2)$$

The olse of $\underline{\beta}$ is given by

$$\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{y} \qquad (2.5.3)$$

with

$$\mathrm{Var}(\hat{\underline{\beta}}) = \sigma^2 (X'X)^{-1} \qquad (2.5.4)$$

SUM OF SQUARES OF ERROR $= SSE$

$$= (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\underline{\beta})$$

TOTAL SS $= SST$

$$= \underline{y}'\underline{y} - n\bar{y}^2$$

and

REGRESSION SS $= SSR$

$$= \hat{\underline{\beta}}' X' \underline{y} - n\bar{y}^2.$$

We have: $SST = SSR + SSE$.

2.6. RECURSIVE REGRESSION ESTIMATION

Consider the linear model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \epsilon_i .$$

$$i = 1, \dots, n$$

Let the vector $\underset{\sim}{x}_i' = (1, x_{1i}, \dots, x_{pi})^{\neq}$.

That is $\underset{\sim}{x}_i$ is the 'design' for the ith observation. The sum of squares of deviation

$$J = \sum_{i=1}^{n} (y_i - \underset{\sim}{x}_i' \underline{\beta})^2 . \qquad (2\text{-}6\text{-}1$$

The olse of $\underline{\beta}$ is obtained by minimizing (2-6-1) with respect to $\underline{\beta}$.

$$\frac{\partial J}{\partial \underline{\beta}} = -2 \sum_{i=1}^{n} \underline{x}_i (y_i - \underline{x}_i' \underline{\beta}) = \underline{0}$$

$$\therefore \quad \sum_{i=1}^{n} \underset{\sim}{x}_i \underset{\sim}{x}_i' \hat{\underline{\beta}}_n = \sum_{i=1}^{n} \underset{\sim}{x}_i y_i \qquad (2.6.$$

Note that $X = \begin{pmatrix} \underset{\sim}{x}_1' \\ \vdots \\ \underset{\sim}{x}_n' \end{pmatrix}$

$\hat{\beta}_n$ is the olse of β based on the observations $(y_1, x_{11}, \ldots, x_{p1}), \ldots, (y_n, x_{1n}, \ldots, x_{pn})$

How is the estimator $\hat{\beta}_n$ updated on receipt of additional information $(y_{n+1}, x_{1\,n+1}, \ldots, x_{p\,n+1})$?

Consider (2-6-2), viz.,

$$\sum_{i=1}^{n} \underset{\sim}{x}_i \underset{\sim}{x}_i' \hat{\underline{\beta}}_n = \sum_{i=1}^{n} \underline{x}_i y_i \qquad (2\text{-}6\text{-}3)$$

Note, $X'X = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i'$, $X'\underline{y} = \sum_{i=1}^{n} \underline{x}_i y_i$.

Put $Var(\hat{\beta}_n) = \sigma^2 (X'X)^{-1} = C_n = \sigma^2 P_n$, say

Also, let $\underline{b}_n = \sum_{i=1}^{n} \underline{x}_i y_i$.

$$P_{n+1}^{-1} = \sum_{i=1}^{n+1} \underline{x}_i \underline{x}_i'$$

$$= P_n^{-1} + \underline{x}_{n+1} \underline{x}_{n+1}' . \quad (2.6.4$$

and similarly

$$\underline{b}_{n+1} = \underline{b}_n + \underline{x}_{n+1} y_{n+1} \quad (2.6.5$$

But,

$$P_{n+1}^{-1} \hat{\beta}_{n+1} = \underline{b}_{n+1}$$

$$= \underline{b}_n + \underline{x}_{n+1} y_{n+1} \quad (2.6.6$$

from 2.6.3

So,

$$P_{n+1}^{-1} \hat{\underline{\beta}}_{n+1} = P_n^{-1} \hat{\underline{\beta}}_n + \underline{x}_{n+1} y_{n+1} \quad (2.6.7)$$

Alternatively,

$$\hat{\underline{\beta}}_{n+1} = P_{n+1} P_n^{-1} \hat{\beta}_n + P_{n+1} \underline{x}_{n+1} y_{n+1} \quad (2.6.8$$

Let $\underline{K}_{n+1} = P_{n+1} \underline{x}_{n+1}$, $\quad$ (2.6.

Since, $\quad P_{n+1}^{-1} = P_n^{-1} + \underline{x}_{n+1} \underline{x}'_{n+1}$, it follow

that

$$I = P_{n+1} P_n^{-1} + P_{n+1} \underline{x}_{n+1} \underline{x}'_{n+1} ,$$

$$\therefore \quad P_n = P_{n+1} + P_{n+1} \underline{x}_{n+1} \underline{x}'_{n+1} P_n ,$$

$$\text{so,} \quad P_n = P_{n+1} + \underline{K}_{n+1} \underline{x}'_{n+1} P_n$$

$$\therefore \quad P_{n+1} = (I - \underline{K}_{n+1} \underline{x}'_{n+1}) P_n \qquad (2\text{-}6.1$$

If we let

$$Z_{n+1} = \underline{K}_{n+1} \underline{x}'_{n+1} \quad \text{then}$$

we have

$$C_{n+1} = (I - Z_{n+1}) C_n \qquad (2.6.11)$$

which is identical to (1.2.5)!

Let's push on. Since

$$P_{n+1} = P_n - K_{n+1} \underline{x}'_{n+1} P_n$$

$$\Rightarrow \quad P_{n+1} \underline{x}_{n+1} = P_n \underline{x}_{n+1} - K_{n+1} \underline{x}'_{n+1} P_n \underline{x}_{n+1}$$

i.e.

$$K_{n+1} = P_n \underline{x}_{n+1} - K_{n+1} \underline{x}'_{n+1} P_n \underline{x}_{n+1}$$

$$\therefore \; K_{n+1} \left(I + x'_{n+1} P_n x_{n+1} \right) = P_n x_{n+1}$$

So, $$K_{n+1} = P_n x_{n+1} \left(I + x'_{n+1} P_n x_{n+1} \right)^{-1} \qquad (2.6.12$$

Equivalently,

$$K_{n+1} = C_n x_{n+1} \left(\sigma^2 I + x'_{n+1} C_n x_{n+1} \right)^{-1} \qquad (2.6.1$$

and

$$Z_{n+1} = C_n x_{n+1} \left(\sigma^2 I + x'_{n+1} C_n x_{n+1} \right)^{-1} x'_{n+1} \qquad (2.6.14)$$

It can be shown using matrix manipulations that

$$\cancel{Z_{n+1} = 2(\sigma^{-2} I +}$$

$$Z_{n+1} = \left(x_{n+1} \sigma^{-2} x'_{n+1} + C_n^{-1} \right)^{-1} x_{n+1} \sigma^{-2} x'_{n+1} . \qquad (2.6.$$

This formula bears strong resemblance to 2

of (1.2.6).

Since

$$K_{n+1} = P_{n+1} \, x_{n+1}$$

$$= \underset{\sim}{P_n} \, \underline{x}_{n+1} \left(I + \underset{\sim}{x}'_{n+1} P_n \, \underset{\sim}{x}_{n+1} \right)^{-1}$$

multiplying by $\underline{x}'_{n+1} \, P_{n+1}$

$$\underset{\sim}{x}'_{n+1} \, P^2_{n+1} \, x_{n+1} = \underset{\sim}{x}'_{n+1} \, P_{n+1} \, P_n \, x_{n+1} \left(\qquad \right)^{-1}$$

Use Bx $\vec{x}$, $\vec{i}$

§3. REGRESSION, LINEAR REGRESSION AND NORMALITY

Preliminaries 2.3

(3.1) Introduction

Regression is concerned with the prediction of one (or more) variable(s) y on the basis of information provided by ~~either~~ other measurements or variables $(X_1, \ldots, X_p) = \underline{X}'$.

y is "response" or "dependent variable"

X_i is "predictor variable"

or "independent variable"

or "carrier"

3.2 Minimum mean square error predictor

We wish to determine a function of the data $\underline{X}$ which best predicts y.

Let $g(\underset{\sim}{X})$ be any function of the data (or sample) $\underline{X}$.

Define

$$g_0(\underset{\sim}{X}) = E[Y \mid \underset{\sim}{X}].$$

That is, g_0 represents the mean of Y given or conditional on the data $\underset{\sim}{X}$.

(R1): The mean square error

$E[(Y - g(X))^2]$ is minimized by the function g_0. That is, <u>the minimum mean square error predictor is the conditional mean</u>.

<u>Proof</u>: First we note that since

$$E[Y] = E[E[Y \mid \underset{\sim}{X}]]$$
$$= E[g_0(\underset{\sim}{X})], \text{ we have}$$

$$E[(Y - g_0)(g_0 - g)]$$
$$= E[E[(Y - g_0)(g_0 - g) \mid \underset{\sim}{X}]]$$
$$= 0.$$

So,

$$E[(Y-g)^2]$$
$$= E[(Y-g_0)^2] + E[(g-g_0)^2] \quad (3.4)$$
$$\geq E[(Y-g_0)^2] \quad \forall g \ (\neq g_0).$$

Therefore g_0 is the minimum mean square error predictor (or estimator) of Y based on $\underline{X}$.

<u>Definition</u>: The function (of $\underline{X}$)

$$g_0 = E[Y|\underline{X}]$$ is called the <u>regression</u> (or <u>regression function</u>) of Y on $\underline{X}$.

(R2): The mean square error (mse) of g_0 is $E[Var[Y|X]]$.

<u>Proof</u>: $$mse = E[(Y-g_0)^2]$$
$$= E[(Y-E(Y|\underline{X}))^2]$$
$$= E[E[(Y-E(Y|\underline{X}))^2|\underline{X}]]$$
$$= E[Var[Y|\underline{X}]].$$

The mse of g_0 is "<u>the average variance of Y conditional on $\underset{\sim}{X}$</u>".

(R3). g_0 maximizes the correlation between Y and g, viz. $\rho(Y, g)$.

(L1): $\quad \mathrm{cov}(Y, g_0) = \mathrm{Var}[g_0] \qquad (3\text{-}2.2)$

<u>Proof</u>: $\mathrm{cov}(Y, g_0)$

$$= E[Y g_0] - E[Y]E[g_0]$$
$$= E[E[Y g_0 | \underset{\sim}{X}]] - E[Y]E[g_0]$$
$$= E[g_0^2] - E^2[g_0]$$
$$= \mathrm{Var}[g_0]$$

(L2): $\quad \mathrm{cov}(Y, g) = \mathrm{cov}(g_0, g) \qquad (3\text{-}2.3)$

<u>Proof</u>: $\mathrm{cov}(Y, g)$

$$= E[Yg] - E[Y]E[g]$$

$$= E[E[Yg \mid \underset{\sim}{X}]] - E[g_0]E[g]$$

$$= E[g g_0] - E[g_0]E[g]$$

$$= \operatorname{cov}(g, g_0)$$

Proof of (R3).

$$\rho^2(Y, g_0) = \frac{\operatorname{cov}^2(Y, g_0)}{\operatorname{Var}[Y] \operatorname{Var}[g_0]}$$

$$= \frac{\operatorname{Var}^2[g_0]}{\operatorname{Var}[Y]\operatorname{Var}[g_0]}, \quad \text{using (3.2.2).}$$

$$= \frac{\operatorname{Var}[g_0]}{\operatorname{Var}[Y]}. \qquad (3.2.4)$$

Accordingly,

$$\rho^2(g, Y) = \frac{\operatorname{cov}^2(g, Y)}{\operatorname{Var}[g]\operatorname{Var}[Y]}$$

$$= \frac{\operatorname{cov}^2(g, g_0)}{\operatorname{Var}[g]\operatorname{Var}[Y]}, \quad \text{using (3.2.3).}$$

$$= \frac{\mathrm{cov}^2(g, g_0)}{\mathrm{Var}[g]\,\mathrm{Var}[g_0]} \cdot \frac{\mathrm{Var}[g_0]}{\mathrm{Var}[Y]}$$

$$= \rho^2(g, g_0) \cdot \rho^2(Y, g_0), \text{ using } (3.4)$$

So $\rho^2(g, Y) \leq \rho^2(Y, g_0) \ \forall g$.

So g_0 is also the 'best' predictor of Y in terms of maximizing the correlation between Y and any function of the data $\underline{X}$.

Consider decomposition (3.2.1) (which is almost ubiquitous) for the case $g = E[Y]$. We have

$$E[(Y - E(Y))^2] = E[(Y - g_0)^2] + E[(g_0 - E[g_0])^2]$$

That is,

$$\boxed{Var[Y] = E[Var[Y|\underline{X}]] + Var[E[Y|\underline{X}]]} \qquad (3.2.5)$$

See §2.1

Alternatively,

$$\sigma_Y^2 = E[\sigma_{Y|\underline{X}}^2] + \sigma_{g_0}^2 \qquad (3.2.6)$$

This is analogous to

"Total SS" = "ERROR SS" + "REGRESSION SS"

"ERROR SS" = AVERAGE VARIATION ABOUT $\underline{X}$

Alternatively,

Total variation in Y = Average variation of Y about $\underline{X}$

+ Variation in regression function.

The square of the maximum correlation attained is

$$\rho^2(Y, g_0) = \frac{Var[g_0]}{Var[gY]}, \text{ using } (3.2.4)$$

$$= \frac{\sigma^2_{g_0}}{\sigma^2_y}$$

$$= 1 - \frac{E[\sigma^2_{y|\underline{x}}]}{\sigma^2_y} \qquad (3.2.7$$

i.e. $\rho^2(Y, g_0) = \dfrac{\text{Variation in regression function}}{\text{Total variation in } Y}$

$= R^2$!

See formulae (2.1.6) and (2.1.7).

3.3 LINEAR REGRESSION

Recall from §2.1 that if (Y, X) are jointly normal then the regression $E[Y|X]$ is <u>linear</u> in X.

The special case when the regression function $E[Y|X]$ is <u>linear</u> has been studied extensively.

It can be shown that g_0 is linear in $X_1, \ldots, X_p$

$\iff$

1. $E[Y|\underset{\sim}{X}] = \mu_Y + \mathrm{cov}(Y, \underline{X}')\,\mathrm{Var}^{-1}(\underset{\sim}{X})(\underset{\sim}{X} - \mu_{\underline{X}})$ (3.3

2. $$\begin{aligned} MSE &= E[\mathrm{Var}[Y|\underset{\sim}{X}]] \\ &= \mathrm{Var}[Y] - \mathrm{cov}(Y, \underline{X}')\,\mathrm{Var}^{-1}[\underline{X}]\,\mathrm{cov}(\underset{\sim}{X}, Y) \end{aligned}$$ (3.3

For the case $p=1$, that is, only one predictor variable.

$$E[Y|X] = \mu_Y + \frac{cov(Y,X)}{Var[X]}(X - \mu_X) \qquad (3\cdot3\cdot3)$$

This is identical to (2·1·1).

and

$$MSE = E[Var[Y|X]]$$

$$= Var[Y] - \rho^2 Var[Y] \qquad (3\cdot3\cdot4)$$

which is identical to (2·1·2).

Put $Z = \frac{cov(Y,X)}{V(X)}$ then

(3·3·3) may be re-cast

$$\boxed{\begin{aligned} \mu_{Y|X} &= \mu_Y + Z(X - \mu_X) \\ E[Var[Y|X]] &= MSE \\ &= (1 - Z)\, Var[Y] \end{aligned}} \qquad (3\cdot3\cdot5)$$

See formulae 1. and 3. of (1·2·6).

The matrix version of (3.3.5) ($p \neq 1$) is given by

$$
\boxed{
\begin{aligned}
&\mu_{Y|\underline{X}} = \mu_Y + Z(\underline{X} - \mu_{\underline{X}}) \\
&MSE = (I - Z)\,\mathrm{Var}[\underline{Y}] \\
&\text{where} \\
&Z = \mathrm{cov}(Y, \underline{X}')\,\mathrm{Var}^{-1}(\underline{X})
\end{aligned}
}
\qquad (3.3.6)
$$

The above formulae are easy to memorize as they bear strong resemblance to (2.3.1) to (2.3.13) of the simple linear regression.

We conclude this section by remarking that linearity is synonymous with normality. [illegible]

§4. UPDATING, BAYES AND LINEAR BAYES

§4.1 SIMPLE FUNDAMENTAL UPDATING FORMULAE.

Suppose you have available two pieces of information, equivalently two estimators of the same parameter θ.

Denote the first estimator by $\hat{\theta}_1$ and the second estimator by $\hat{\theta}_2$. Both estimators are ubiased, so that

$E[\hat{\theta}_1] = E[\hat{\theta}_2] = \theta$, and the

respective variances are given by

$$v_1 = \mathrm{Var}[\hat{\theta}_1] \text{ and } v_2 = \mathrm{Var}[\hat{\theta}_2].$$

How do we combine the two pieces of information. It is reasonable to use the weighted least squares estimator that minimizes

$$J = v_1^{-1}(\theta - \hat{\theta}_1)^2 + v_2^{-1}(\theta - \hat{\theta}_2)^2. \quad (4.1)$$

The weight v_i^{-1} is proportional to the precision v_i^{-1} of the estimator $\hat{\theta}_i$.

Now,

$$\frac{dJ}{d\theta} = 2 v_1^{-1}(\theta - \hat{\theta}_1) + 2 v_2^{-1}(\theta - \hat{\theta}_2).$$

Setting $\frac{dJ}{d\theta} = 0$, we obtain

$$\begin{aligned} \hat{\theta} &= (1-z)\,\hat{\theta}_1 + z\,\hat{\theta}_2 \\ &= \hat{\theta}_1 + z(\hat{\theta}_2 - \hat{\theta}_1) \end{aligned} \qquad (4.1.2)$$

where,

$$z = \frac{v_2^{-1}}{v_1^{-1} + v_2^{-1}} = \frac{1}{1 + \frac{v_2}{v_1}} \qquad (4.1.3)$$

Moreover,

$$\mathrm{Var}[\hat{\theta}] = (1-z)\,\mathrm{Var}[\hat{\theta}_1]. \qquad (4.1.4)$$

We shall call formulae (4.1.2) to (4.1.4) the <u>fundamental updating formulae</u>.

Note that recursions (1·2·6) can be easily derived using the foregoing formulae (4·1·2) to (4·1·4).

§4.2. Bayes formula

Suppose we observe one observatio y that comes from a normal distribut with mean μ and variance σ^2.

Before we observe y, our knowledge (or uncertainty) about the known parameter μ is expressed by a prior distribution that is normal with mean μ_0 and variance σ_0^2.

We write:

$$y \mid \mu \sim N(\mu, \sigma^2)$$

and

$$\mu \sim N(\mu_0, \sigma_0^2).$$

Based on §3-2, the minimum mean square error predictor of μ is

$$E[\mu \mid y] \qquad (4.2.1)$$

with mean square error (or average ~~cov~~ variance)

$$E[\mathrm{Var}[\mu \mid y]]. \qquad (4.2.2)$$

The predictor $E[\mu|y]$ is also the Bayes rule (with respect to square error loss function).

Indeed,

$$E[\mu|y]$$

represents the mean value of μ, conditional on the sample information y and the 'prior' information μ_0 (and σ_0^2).

Since the random variables (μ, y) are jointly normal, we have

$$E[\mu/y] = E[\mu] + cov(\mu, y)\, Var^{-1}[y] \cdot (y - E[y]), \quad (4.2.3$$

using result (2.1.1) and

$$Var[\mu|y] = Var[\mu] - \sigma_{y\mu}\, Var^{-1}[y]\, \sigma_{\mu y} . \quad (4.2.$$

We now have the following results:

(R1): $E[Y] = E[E[Y|\mu]]$

$= E[\mu]$

So, $\boxed{E[Y] = \mu_0}$

(R2): $Var[Y] = Var[E[Y|\mu]] + E[Var[Y|\mu]]$

$= Var[\mu] \quad + E[\sigma^2]$

So, $\boxed{Var[Y] = \sigma_0^2 \quad + \sigma^2}$

(R3): $\sigma_{Y\mu} = \sigma_{\mu Y}$

$= cov(Y, \mu)$

$= E[Y\mu] - E[Y]E[\mu]$

$= E[E[Y\mu|\mu]] - E[\mu]E[\mu]$

$= E[\mu^2] - E^2[\mu]$

$= Var[\mu]$

So, $\boxed{\sigma_{Y\mu} = \sigma_0^2}$

i.e. $\boxed{cov(Y, \mu) = Var[\mu]}$

Indeed, we can show in general that for any random variable y for which $E[Y|\mu]$

$$\boxed{Cov(Y,\mu) = Var[\mu]} \qquad (4.2.5)$$

See also equation (3.2.2)

Substituting above results into (4.2.3) and (4.2.4), we obtain

$$E[\mu|Y] = \mu_0 + Z(Y-\mu_0), \qquad (4.2.6)$$

$$Var[\mu|Y] = (1-Z)\sigma_0^2 \qquad (4.2.7)$$

where

$$\boxed{\begin{aligned} Z &= \frac{Cov(\mu,Y)}{Var[Y]} \\ &= \frac{Var[\mu]}{Var[Y]} \\ &= \frac{1}{1+\dfrac{E[Var[Y|\mu]]}{Var[E[Y|\mu]]}} \\ &= \frac{1}{1+\dfrac{\sigma^2}{\sigma_0^2}} \end{aligned}} \qquad (4.2.8)$$

We can write

$$Z = \frac{1}{1 + \frac{EV}{VE}},$$

where, $EV = E[Var[y|\mu]]$

= "Average process variance",

$VE = Var[E[y|\mu]]$

= "Variance of Hypotheses"

≡ "Variance in parameter risk μ".

See Venter page 379, page 393, page 395

Based on the results of section 3.3 the Bayes rule (4.2.6) is the linear Bayes rule (even when we omit normality assumption) and (4.2.7) is the MSE, viz., $E[Var[\mu/y]]$.

4.3 Derivation of Bayes rule using fundamental updating formulae.

The (linear) Bayes estimator (4.2.6) and its MSE given by (4.2.7) may be derived using the fundamental updating formulae of section 4.1.

Before we observe y we have a "prior" estimate of μ given by μ_0 with variance σ_0^2. So, set $\theta = \mu$, and

$$\hat{\theta}_1 = \mu_0$$

with $\quad v_1 = \sigma_0^2.$

Based on the sample y we have

$$\hat{\theta}_2 = y, \text{ with}$$

$$v_2 = \sigma^2.$$

So, using (4.1.2) to (4.1.4) we ~~have~~ obtain (4.2.6) and (4.2.7).

Note that for v_1 we substitute

$$\sigma_0^2 = \mathrm{Var}[\mu]$$

$$= \text{"Variance in the parameter"}$$

and for v_2 we substitute

$$E[\mathrm{Var}[y|\mu]]$$
$$= \text{"Mean process variance"}$$
$$= \sigma^2, \text{ in this case.}$$

<u>By way of summary</u>, if we have data y for which

(i) $E[y|\mu] = \mu$

(ii) $E[\mu] = \mu_0$

(iii) $E[\mathrm{Var}[y|\mu]] = \sigma^2$

(iv) $\mathrm{Var}[\mu] = \sigma_0^2$

then the linear Bayes rule (4.2.6) and its MSE (4.2.7) can be derived assuming

$$y|\mu \sim N(\mu, \sigma^2)$$

and

$$\mu \sim N(\mu_0, \sigma_0^2)$$

<u>even</u> if $y|\mu$ is <u>not</u> normal!

ALSO, (4.2.6) and (4.2.7) can be derived assuming that the prior estimate μ_0 is based on a sample

Bayesian formulation

μ random with $E[\mu]=\mu_0$, $Var[\mu]=\sigma_0^2$

$E[y|\mu]=\mu$ and $E[Var[y|\mu]]=\sigma^2$

Prior sample formulation

μ fixed with $E[\mu_0]=\mu$, $Var[\mu_0]=\sigma_0^2$

$E[y]=\mu$ and $Var[y]=\sigma^2$ (obtained from the Bayesian formulation)

$\boxed{\text{BOTH FORMULATIONS PRODUCE SAME ANSWERS}}$

Exercise: Derive (4.2.6) and (4.2.7) using (1.2.6).

4.3 Bayes and Regression.

The (linear) Bayes rule $E[\mu/Y]$ is the regression of μ on Y. See sections 2 and 3. So what is the difference between the Bayesian model of §4. and regression model of section 2.3, say

In the Bayesian model the parameter μ is unobservable whereas in a typical regression model where the regression function is $E[Y|X]$, say, both Y and X are observable.

4.4 Conjugate priors

We observed that with a normal sampling distribution and a normal prior the Bayes rule is linear. Are there any other sampling-prior pairs that yield Bayes rules that are linear?

For a given model $f_Y(y;\theta)$, we can look for a family of prior distributions such that the posterior distributions are in the same family as the sampling distribution, only the indexing quantities being changed. Such a family is called <u>conjugate</u>.

Suppose we have data from

$$f_Y(y;\theta) = \exp[a(\theta)b(y) + c(\theta) + d(y)],$$

then the conjugate prior is

$$f_{\Theta}(\theta; k_1, k_2) = \exp(k_1 a(\theta) + k_2 c(\theta)).$$

That is, the prior (save a normalizing constant) is obtained from the sampling distribution by making the parameter in the sampling distribution the argument of the prior density.

Example 1: $f_Y(y;\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}$,

then the sampling density as a function of θ is also normal. So, the conjugate prior for θ is the normal distribution.

Example 2: $f_Y(y;\theta) = \frac{e^{-\theta}\theta^y}{y!}$; $y=0,1,\ldots$ $\theta>0$

The prior density for θ is

$$\propto e^{-\theta}\theta^y,$$ and so

the Gamma distribution is the prior density.

RESULT

If the prior distribution is conjugate, then the Bayes rule is linear.
Moreover, the Bayes rule is given by (2.1.1) and the MSE ($E[Var[y|x]]$) by equation (2.1.2)

(4.5) Linear Bayes rules

Recall from §3.2 that the Bayes rule, equivalently, the regression function is the optimal mean square error predictor.

Additionally, the Bayes rule is linear in the data for sampling-conjugate paires (§4.3 & §4.4) and moreover can be derived assuming normality or the simple statistical fixed parameter model of §1.2.

If in §3.2 we confine our estimators (or functions) to (inhomogeneous) linear or affine functions of the data $\underset{\sim}{X}$, i.e functions $g(\underset{\sim}{X})$:

$$g(\underset{\sim}{X}) = a_0 + \underset{\sim}{a}'\underset{\sim}{X}, \quad \text{then}$$

the minimum mean square error estimator amongst the linear estimators is given by equation (3.3.1), which is identical to the Bayes estimator assuming normality. The MSE is given by (3.3.2), which is the MSE assuming normality.

See Venter pages 416-419.

Example 1.

$Y_1, \dots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where σ^2 is known, $E[Y_i] = E[\mu] = \mu_0$ and $Var[\mu] = \sigma_0^2$. Reduction by sufficiency yields

$$\bar{Y}|\mu \sim N(\mu, \frac{\sigma^2}{n}).$$

Applying §4.2, the (linear) Bayes rule for μ is

$$\hat{\mu} = \mu_0 + Z(\bar{Y} - \mu_0), \text{ with}$$

$$E[Var[\mu|\bar{Y}]] = (1-Z)\sigma_0^2, \text{ where}$$

$$\begin{aligned} Z &= cov(\mu, \bar{Y}) Var^{-1}[\bar{Y}] \\ &= \frac{Var[\mu]}{Var[\bar{Y}]} \\ &= \frac{1}{1 + \frac{E[Var[\bar{Y}|\mu]]}{Var[E[\bar{Y}|\mu]]}} \\ &= \frac{n}{n + \frac{E[Var[Y_i|\mu]]}{Var[E[Y_i|\mu]]}} \\ &= \frac{n}{n + \sigma^2/\sigma_0^2} \end{aligned}$$

We can also derive the above formulae using (1·2·6) as follows:

Assume "first" estimate of μ is μ_0 with variance σ_0^2.

So, put $\hat{\mu}_1 = \mu_0$ and $C_1 = \sigma_0^2$.

We wish to update the first estimate based on additional information $\bar{Y}$ with variance $\frac{\sigma^2}{n}$.

So,

$$\hat{\mu}_2 = \hat{\mu}_1 + Z_2 (\bar{Y} - \hat{\mu}_1)$$

$$= \mu_0 + Z_2 (\bar{Y} - \mu_0) .$$

Now, Z_2 in (1·2·6) is given by

$$C_1 (C_1 + \sigma^2)^{-1} \qquad (*)$$

where σ^2 is the variance of the data at 'time' 2. So here replace σ^2 in (*) by $\frac{\sigma^2}{n}$. So

$$Z_2 = \frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}} = \frac{n}{n + \sigma^2/\sigma_0^2} \quad !$$

The weight or credibility Z_2 is proportional to the relative precision of $\bar{Y}$ v/s & v/s μ_0.

<u>Example 2</u>. (See Venter page).

Suppose $Y_1, \ldots, Y_n \overset{iid}{\sim} Po(\lambda)$, then

$$f_{Y_i}(y_i; \lambda) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} \; ; \; y = 0, 1, \ldots$$

The conjugate prior is Gamma(α, β). α is shape parameter and β is scale parameter.

$$f_\Lambda(\lambda) = \frac{\lambda^{\alpha-1} e^{-\lambda/\beta}}{\Gamma(\alpha)\,\beta^\alpha} \; ; \quad \lambda > 0.$$

$$E[\Lambda] = \alpha\beta = \lambda_0, \text{ say}$$

$$Var[\Lambda] = \alpha\beta^2 = \sigma_0^2, \text{ say}.$$

Posterior density of $\Lambda \mid Y_1 \ldots Y_n$ is

$$\propto \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} \cdot \frac{\lambda^{\alpha-1}e^{-\lambda/\beta}}{\Gamma(\alpha)\beta^{\alpha}}$$

$$= \text{Gamma}\left(\Sigma y_i + \alpha,\ \frac{\beta}{n\beta+1}\right).$$

So, $$E[\Lambda \mid Y_1 \ldots Y_n] = (\Sigma Y_i + \alpha)\frac{\beta}{\beta+1}$$

$$= \alpha\beta \cdot \frac{1}{1+n\beta} + \frac{n\beta}{n\beta+1}\cdot \bar{Y}$$

$$= \alpha\beta + Z(\bar{Y} - \alpha\beta)$$

$$= \lambda_0 + Z(\bar{Y} - \lambda_0)$$

where $$Z = \frac{n\beta}{n\beta+1}$$

$$= \frac{n}{n + 1/\beta}$$

$$= \frac{n}{n + \frac{\lambda_0}{\sigma_0^2}}$$

$$= \frac{n}{n + EV/VE}$$

Had we assumed

$$\bar{Y} | \lambda \sim N(\lambda, \lambda_0)$$

and

$$\lambda \sim N(\lambda_0, \sigma_0^2)$$

then we would have obtained the same Bayes rule and the same linear Bayes rule (due to conjugacy).

§5. CREDIBILITY

§5.1 Introduction

Credibility theory is the name given by American actuaries to heuristic linear estimation formulae developed in the years before World War I for insurance ratemaking problems. Although it has traditionally been a cornerstone of casualty actuarial mathematics it turns out to be related to other fields including Bayes estimation (Section 4), linear Bayes estimation (Section 4), linear regression (Sections and 3), linear filtering theory (de Jong and Zehnwirth, 1981; Zehnwirth, 1983 and Zehnwirth 1988).

§5.2 Early Work

The general credibility formula

$$C = (1-Z)B + ZA \qquad (5.2.1)$$

originated in the U.S. in the context of workers' compensation insurance.

The quantity B is the industry-wide premium rate currently charged for a particular occupational class.

An employer having had a favorable safety record ~~with~~ this particular occupational class presses for recognition in his insurance premium and prefers to be charged a quantity A, the rate based on ~~his~~ its own experience. In view of the fact that the employers' own experience is usually scanty and accordingly A is not a reliable (credible) ~~fore~~cast (estimate of cost, Whitney (1918) suggested a balance C between two extremes B and A. The quantity Z is called the credibility fac~~tor~~; it provides for the mixing of the fair industry-wide premium B and the employer based premium A, with increasing credibility attached to the latter as the volume of experience of the employer increases

Whitney (1918) derived (5.2.1) for the following model:

He assumed the number of observed claims Y for employer is distributed

Binomial (N, p). N is # of

employees and p is probability of claim

Whitney (1918) also assumed that a priori

$$p \sim N(p_0, \sigma_0^2).$$

So we have

$$E\left[\frac{Y}{N} \mid p\right] = p$$

$$Var\left[\frac{Y}{N} \mid p\right] = \frac{p(1-p)}{N}$$

So,

$$E\left[Var\left[\frac{Y}{N} \mid p\right]\right] = \frac{p_0 - \sigma_0^2 - p_0^2}{N}$$

$$= \frac{p_0(1-p_0) - \sigma_0^2}{N}$$

So, based on Section 4, the credibility estimator (linear Bayes estimator) is –

$$\hat{p} = p_0 + z\left(\frac{Y}{N} - p_0\right) \qquad (5.2.2)$$

where

$$z = \frac{1}{1 + \frac{EV}{VE}} = \frac{1}{1 + \frac{\frac{p_0(1-p_0) - \sigma_0^2}{N}}{\sigma_0^2}}$$

$$= N\sigma_0^2 / \left[(N-1)\sigma_0^2 + p_0(1-p_0)\right]$$

Whitney's credibility factor

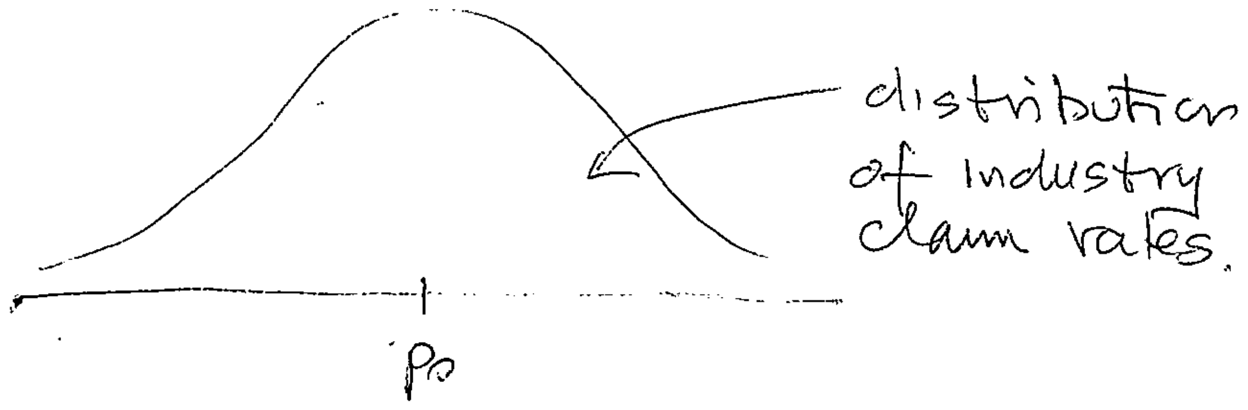
$$Z' = N\sigma_0^2/[N\sigma_0^2 + p_0(1-p_0)],$$

is <u>incorrect</u>. He replaced

$E[p(1-p)]$ by $p_0(1-p_0)$ instead of

$p_0(1-p_0)-\sigma_0^2$!

The quantity p_0 represents the industry-wide claim rate whereas σ_0^2 represents the variance of claim rates within the industry. The estimate of σ_0^2 was guided partly by general reasoning and partly by an appeal to underwriting judgments.



Based on the results of §4, the credibility estimator 5.2.2 is the exact Bayes rule for a prior that has a Beta distribution.

Definition of credibility estimator

A linear estimator $a_0 + b_0 \cdot \frac{Y}{N}$ is a (linear) credibility estimator for p

$\Leftrightarrow$

$$E\left[\left(a_0 + b_0 \frac{Y}{N} - p\right)^2\right] \leq E\left[\left(a + b \cdot \frac{Y}{N} - p\right)^2\right]$$

$$\forall a, b \in \mathbb{R}^1$$

The credibility estimator is the <u>linear Bayes rule</u>. It is said to be "exact" if it is the (actual) Bayes rule.

The connection between (5.2.1) and Bayes theorem falls under the heading of <u>exact credibility</u>. It was formalized by Jewell (1974) and Diaconis and Ylvisaker (1979) for the exponential family of sampling distributions. Zehnwirth (1977) formalized the connection in the nonparametric case

Consider another example similar to the one studied by Whitney.

Suppose we have an insurance policy for which the 'manual' premium is μ_0

That is, μ_0 is the premium for similar policies (risks). Let Y_i be the loss size for the ith year the individual policy is in force. After n years we have at our disposal observations $Y_1, \ldots, Y_n$ that are assumed to be independent and identically distributed

The average annual loss size is:

$$\overline{Y_n} = \frac{1}{n} \sum_{i=1}^{n} Y_i \ .$$

How different are μ_0 and $\overline{Y_n}$?

For the first year the policy is in force we presumably charge μ_0 ?

What do we charge in subsequent years as we obtain more information regarding the individual policy (risk)?

After n years, the two extreme

choices for the pure premium are μ_0 and $\overline{Y}_n$.

A "compromise" would be

$$(1-Z_n)\mu_0 + Z_n \overline{Y}_n$$
$$= \mu_0 + Z_n(\overline{Y}_n - \mu_0),$$

where Z_n is a 'weight' or credibili[ty] lying between 0 and 1.

The credibility Z_n should depend on

1. n; as the number of years increases it is reasonable to attach more weight to $\overline{Y}_n$;

2. how different is $\overline{Y}_n$ to μ_0 (or μ to μ_0). That is, how different is the individual risk (policy) to the collection of risks in the portfolio?

We need some assumptions. We accordingly describe the <u>collective risk model</u>.

§5.3 Collective risk model

The concept of the collective of risks plays a central role in credibility theory.

Definition: A collective is a set or collection of individual risks. Each risk is labelled by a parameter ϑ. The collective is denoted by $\Theta = \{\vartheta\}$.

Example: Consider automobile hull insurance policy. The annual loss sizes $Y_{1j}, \ldots, Y_{nj}$ can only take on value ϕ or 1. The probability that $Y_{ij} = 1$, depends on type of motor car ϑ_j, say. Assume type of motor car is unknown. However, we can assert that ϑ has a distribution, i.e., when I draw a car at random, what is the probability that it is a Honda Accord, say.

Definitions: The parameter ϑ is called the risk parameter and its distribution

$$U(\vartheta) = \Pr[\Theta \le \vartheta]$$ is called the structural distribution.

The distribution U describes the risk structure of the portfolio.

Suppose we have data on p risks from the collective of risks.

The p risks are labelled $\vartheta_1, .., \vartheta_p$.

For risk ϑ_j we have data $Y_{1j}, .., Y_{nj}$ assumed to be identically and independent distributed given ϑ_j.

The p risks (policies), $\vartheta_1, .., \vartheta_p$ may be regarded as a random selection from a (conceptual) population of risks, called the collective. That is, the unobservable quantities $\theta_1, .., \theta_p$ may be regarded as independent realizations from the structureural distribution.

Jewell (   ) describes the foregoing collective risk model as a double urn problem. We first select a risk, alternative the quantity θ from an ~~the~~ urn containing the conceptual population of risks. The value θ selected (but unobservable) determines a second urn from which the claims experience is selected.

$$
\begin{array}{cccc}
\theta_1, & \theta_2, & \cdots, & \theta_p \\
Y_{11} & Y_{12} & & Y_{pP} \\
Y_{21} & Y_{22} & & Y_{2p} \\
\vdots & \vdots & & \vdots \\
Y_{n1} & Y_{n2} & & Y_{np}
\end{array}
$$

Suppose (i) $E[Y_{ij} | \theta_j] = \mu(\theta_j) = \mu_j$

(ii) $Var[Y_{ij} | \theta_j] = \sigma^2(\theta_j) = \sigma_j^2$

(iii) $E[\mu_j] = \mu_0$

(iv) $Var[\mu_j] = \sigma_0^2$

(v) $E[\sigma_j^2] = \sigma^2$

Note that μ_0 represents the pure premium for the collective whereas μ_j is the pure premium for risk θ_j.

The (linear) credibility estimator for μ_j minimizes the least squares

$$E[(a_0 + \sum_{i,j} a_{ij} Y_{ij} - \mu_j)^2]$$ over all

constants a_0; $\{a_{ij}\}$.

Since Y_{ij} is independent of $Y_{kj'}$, it follows the credibility estimator for μ_j does not involve data on the other risks. Moreover, since $Y_{1j}, \dots, Y_{nj}$ are independent and identically distributed conditional on θ_j, it follows that

$$a_{1j} = a_{2j} = \dots = a_{nj}.$$

So, credibility estimator for μ_j is linear in $\overline{Y}_{\cdot j} = \frac{1}{n} \sum_{i=1}^{n} Y_{ij}$.

Now, $E[\overline{Y}_{\cdot j} \mid \theta_j] = \mu_j$ and

$$E[\mathrm{Var}[\overline{Y}_{\cdot j} \mid \theta_j]] = \frac{\sigma^2}{n}.$$

So, in order to derive the linear Bayes rule for μ_j (equivalently, the credibility formula) for μ_j, we may assume

$$\overline{Y}_{\cdot j} \mid \mu_j \sim N\left(\mu_j, \frac{\sigma^2}{n}\right)$$

and

$$\mu_j \sim N(\mu_0, \sigma_0^2)$$

— SEE SECTION 4

So,

$$\hat{\mu}_j = \mu_0 + Z(\bar{Y}_{\cdot j} - \mu_0) \qquad (5.3.1)$$

with,

$$E[(\mu_j - \hat{\mu}_j)^2] = E[Var[\mu_j \mid \bar{Y}_{\cdot j}]]$$

$$= (1-Z)\sigma_0^2 \qquad (5.3.2)$$

where,

$$Z = \frac{(\sigma^2/n)^{-1}}{\sigma_0^{-2} + \left(\frac{\sigma^2}{n}\right)^{-1}}$$

$$= \frac{n}{n + \sigma^2/\sigma_0^2}$$

$$= \frac{1}{1 + \dfrac{E[Var[\bar{Y}_{\cdot j} \mid \theta_j]]}{Var[E[\bar{Y}_{\cdot j} \mid \theta_j]]}} \qquad (5.3.3)$$

See Venter pages 416-423.

Example: Suppose μ_j represents the claim rate of policyholder j. The observable Y_{ij} denotes the number of claims made by policyholder j in year i. So $Y_{1j}, \ldots, Y_{nj}$ are independently and identically distributed Poisson variables with mean μ_j (conditional on μ_j).
Here, $E[\mathrm{Var}[\bar{Y}_{\cdot j} | \mu_j]] = E[\frac{\mu_j}{n}] = \frac{\mu_0}{n}$.

So credibility estimator

$$\hat{\mu}_j = \mu_0 + Z(\bar{Y}_{\cdot j} - \mu_0) \qquad (5.3.4)$$

with

$$Z = \frac{n}{n + \frac{\mu_0}{\sigma_c^2}}. \qquad (5.3.5)$$

If the structural function is a Gamma distribution then (5.3.4) is the exact credibility (Bayes) estimator and (5.3.2) is the Bayes risk when Z is given by (5.3.5).

It is interesting to note the followin[g]. Suppose μ_j were also observable so that one could display a scatter plot of the number pairs $(\mu_j, \bar{Y}_{\cdot j})$. If we fit a straight line regression thru these number pairs in order to predict μ from $\bar{Y}_{\cdot}$, the regression line would be

$$\tilde{\mu}_j = \bar{Y}_{\cdot\cdot} + b(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}) \qquad (5\text{-}3\text{-}6)$$

with $b = \operatorname{cov}(\mu_j, \bar{Y}_{\cdot j}) / \operatorname{Var}[\bar{Y}_{\cdot j}]$

$$= Z \qquad (5\text{-}3\text{-}7)$$

So (5-3-6) is same as (5-3-1) and (5-3-4) except that we substitute for μ_0 its estimate $\bar{Y}_{\cdot\cdot}$ based on the data!!

5.4 Buhlmann - Straub model

The Buhlmann Straub (1970) model has the same prescription as the collective risk model of the previous section save that

$$\mathrm{Var}[Y_{ij} \mid \theta_j] = \sigma_j^2 / P_{ij}$$

where P_{ij}'s are known constants reflecting volume measures.

The (linear) credibility estimator here is

$$\hat{\mu}_j = \mu_0 + Z_j (\bar{Y}_{\cdot j} - \mu_0), \qquad (5.4.1)$$

with

$$E[(\mu_j - \hat{\mu}_j)^2] = (1 - Z_j)\,\sigma_0^2 \qquad (5.4.2)$$

where

$$Z_j = \frac{P_{\cdot j}}{P_{\cdot j} + \sigma^2/\sigma_0^2} \qquad (5.4.3)$$

and

$$P_{\cdot j} = \sum_{i=1}^{n} P_{ij} \, .$$

5.5 BAILEY'S MODEL, ONE-WAY ANOVA AND THE COLLECTIVE RISK MODEL.

Bailey's (1945) paper is remarkably modern. Not only did he derive (5.3.1) to (5.3.3) and the corresponding empirical versions which predates Buhlmann's model but also predates the Lindley estimator (the James-Stein estimator that shrinks towards the grand mean, Stein (1962)).

The principal reason that Bailey's work is so remarkable in that it was conducted before (i) Robbins (1955) introduced the empirical Bayes field of statistics, (ii) Wald (1950) formalized statistical decision theory, and ~~before~~ (iii) Savage (19. formalized Bayesian statistics.

It appears to me that Bailey's important and early contribution has been overlooked by many actuaries perhaps as a result of his complicated notation.

Since Bailey used some results from Fishers one-way ANOVA set-up, we first study one-way ANOVA.

Preliminaries

§5.6 ONE-WAY ANOVA

The data consist of a double array $\{Y_{ij}\}$ (just as in the collective risk model) where Y_{ij} is the ith observation in the sample from the jth population.

$$\begin{array}{cccc} Y_{11} & Y_{12} & \cdots & Y_{1p} \\ Y_{21} & & & Y_{2p} \\ \vdots & & & \vdots \\ Y_{n1} & & & Y_{np} \end{array}$$

The model customarily chosen for the data in a one way ANOVA is

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij} \qquad (5.6.1)$$

$$i = 1, \ldots, n$$
$$j = 1, \ldots, p$$

where μ denotes the overall mean

$\mu_j = \mu + \alpha_j$ denotes the mean of the jth population and ϵ_{ij} is random (unexplained) variation with mean ϕ and variance τ^2.

An important distinction in the model assumptions over whether the conclusion are to apply strictly to the p populati under observation or whether they are to apply to a wider class of populations of which the p populations are a representative. In the first instant the p populations are viewed as fixed, whereas in the second they are viewed as random (just like in the collective risk model).

Fixed Effects

The complete model is

$$Y_{ij} = \mu_j + \epsilon_{ij} \qquad (5\text{-}6\text{-}2$$

or

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij}$$

for $i=1,..,n$; $j=1,..,p$, where $\epsilon_{ij} \sim N(0,\sigma^2)$. To avoid identifiabili problems $\sum_{j=1}^{p} \alpha_j = 0$.

The ordinary least squares estimator of μ_j is $\bar{Y}_{\cdot j}$, the mean of the jth sample

Applying straightforward algebra, the total variation

$$\Sigma\Sigma(Y_{ij}-\bar{Y}_{\cdot\cdot})^2 \quad \text{may be}$$

decomposed:

$$\Sigma\Sigma(Y_{ij}-\bar{Y}_{\cdot\cdot})^2 = \Sigma\Sigma(Y_{ij}-\bar{Y}_{\cdot j})^2 + \Sigma\Sigma(\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot})^2 \qquad (5.6.$$

Notationally

$$SST = SSW + SSB \qquad (5.6.$$

The total sum of squares SST measures the variability in the population comprising the p populations.

The sum of squares within populations SSW measures the random variation in the data. Indeed,

$$\hat{\sigma}^2 = MSE$$

$$= \frac{SSW}{N-p} \qquad (5.6.5)$$

where, $N = np$.

The variation between the p population is measured by SSB. So,

$$\boxed{\begin{array}{ccccc} \text{TOTAL VARIABILITY} & = & \text{VARIABILITY WITHIN} & + & \text{VARIABILITY BETWEEN} \end{array}} \quad (5.6.1)$$

Compare this decomposition with the following that applies to the collective ns model

$$\boxed{Var[Y_{ij}] = E[Var[Y_{ij}|\mu_j]] + Var[E[Y_{ij}|\mu_j]]}$$

Consider the null hypothesis

$$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_p$$

equivalently,

$$H_0: \quad \alpha_j = 0, \quad \forall j$$

versus

$$H_1: \quad \alpha_j \neq 0 \quad \forall j$$

If H_0 is true then

$$\bar{Y}_{.j} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and

$$\bar{Y}_{..} \sim N\left(\mu, \frac{\sigma^2}{N}\right).$$

So under H_0

$$E\left[\frac{SSB}{p-1}\right] = \sigma^2.$$

Let $MSW = MSE$ and

$$MSB = \frac{SSB}{p-1}.$$

Then, under H_0 we exept the F-ratio

$$F = \frac{MSB}{MSW}$$ to be close to 1.

That is, under H_0 we expect the variation between groups

$\simeq$ the variation within groups

If H_0 is not true, then the μ_j's are different, so that we would expect

$$MSB > MSW.$$

We therefore are inclined to reject the null hypothesis of equal means when the F-ratio > 1.

§5.7 Random effects and collective risk model.

In the random effects model of the one-way ANOVA, the p populations that are considered are not the only ones of interest. They are merely representative of a wider class of populations from which they have been selected.

In the random effects model the populations means $\mu_1, \mu_2, .., \mu_p$ are regarded as independent realizations from a 'prior' distribution*. So

$$\boxed{\text{ONE-WAY RANDOM EFFECTS} \equiv \text{COLLECTIVE RISK MODEL}}$$

See Venter page 420

Indeed, Bailey (1945)

considered the one-way random effects model (collective risk model) and derived the credibility estimator (5.3.1), (5.3.2) and (5.3.3) as the linear regression of μ_j on $\bar{Y}_{.j}$.

* 'prior' distribution $\equiv$ structural distribution

Let's rewrite (5.3.1).

$$\hat{\mu}_j = \mu_0 + Z(\bar{Y}_j - \mu_0) \qquad (5.7.1)$$

with

$$Z = \frac{n}{n + \sigma^2/\sigma_0^2} \qquad (5.7.2)$$

Bailey (1945) went a step further by estimating the hyperparameters $\mu_0, \sigma^2, \sigma_0$ and consequently producing an empirical version of (5.7.1).

<u>Bailey's (1945)</u> analysis uses Fisher's one-way ANOVA results of § 5.6.

Venter page 432-433 also discusses the collective risk model but does not mention that the ~~first~~ credibility formulae and empirical versions were first developed by <u>Bailey in 1945</u>.

An unbiased estimator for μ_0, the pure premium for the collective is

$$\overline{Y}_{\cdot\cdot} = \frac{1}{p}\sum_{j=1}^{n} \overline{Y}_{\cdot j} = \frac{1}{N} \sum\sum Y_{ij}$$

Consider the quantity

$$s_j^2 = \sum_{i=1}^{n} (Y_{ij} - \overline{Y}_{\cdot j})^2/(n-1).$$

s_j^2 is the sample variance based on the sample from the jth population. So, s_j^2 is an unbiased estimator of

$$Var[\, Y_{ij} \mid \theta_j \,]$$

So, $$E\left[\sum_{j=1}^{p} s_j^2/p\right] = E[\, Var[Y_{ij} \mid \theta_j]\,] = \sigma^2 \qquad (5.7.3$$

Accordingly, <u>MSW is an unbiased estimator of σ^2</u>.

Now, consider

$$MSB = \sum_{j=1}^{p} \sum_{i=1}^{n} (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 / (p-1)$$

$$= n \sum_{j=1}^{p} (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 / (p-1)$$

So, $\frac{MSB}{n}$ is an unbiased estimator o

$$Var[\bar{Y}_{\cdot j}] = Var[E[\bar{Y}_{\cdot j} | \theta_j]] + E[Var[\bar{Y}_{\cdot j} | \theta_j]]$$

$$= Var[\mu_j] + \frac{\sigma^2}{n}$$

$$\therefore \boxed{Var[\bar{Y}_{\cdot j}] = \sigma_0^2 + \frac{\sigma^2}{n}} \qquad (5.7.4)$$

Application of (5.7.3) and (5.7.4) yields unbiased estimators of σ^2 and σ_0^2 thus

$$\hat{\sigma}^2 = MSW$$

$$\hat{\sigma}_0^2 = \frac{MSB}{n} - \frac{MSW}{n}$$

$$= \frac{1}{n}(MSB - MSW)$$

So, an estimate of Z is given by

$$\hat{Z} = \frac{n}{n + \dfrac{MSW}{\frac{1}{n}(MSB - MSW)}}$$

$$= \left(1 - \frac{1}{F}\right), \text{ where } F \text{ is the}$$

F-ratio statistic of §5.6.

So, the empirical linear credibility estimator derived by Bailey (1945) is given by

$$\hat{\mu}_j = \bar{Y}_{..} + \left(1 - \frac{1}{F}\right)\left(\bar{Y}_{.j} - \bar{Y}_{..}\right) \quad (5.7.5$$

The last formula was also derived by Buhlmann & Straub in (1970).

The estimator (5·7·5) is called a 'smooth' test-estimator.

If the null hypothesis $H_0: \mu_1 = \mu_2 = \dots = \mu_p$ is true then $F \doteqdot 1$ so that

$$\hat{\mu}_j = \bar{Y}_{\cdot\cdot} \, .$$

That is, if the risks are the same (homogen) (no variation within the collective), the estimate of the pure premium for risk j is equal to the estimate of the pure premium for the collective.

Alternatively, if the risks are heterogeneous, so that $F >> 1$, then

$$\hat{\mu}_j \simeq \bar{Y}_{\cdot j} \, .$$

That is we can only use the individual risk's experience to estimate its pure premium.

§6 Bühlmann Straub with unequal weights

Interchange role of i and j

in the collective risk model §5-3+

Let ϑ_i represent risk i and

X_{ij} is the jth observation.

$i = 1, .., n$ (p) and $j = 1, .., m$ (n)

§5.3+ §5.3+

The risk θ_i is characterized by an individual risk profile ϑ_i which is itself the realization of a r.v. Θ_i

(A1) Conditional on $\Theta_i = \theta_i$

$\{X_{ij} : j = 1, .., m\}$ are independent

with $E[X_{ij} | \theta_i] = \mu(\theta_i) = \mu_i$

$Var[X_{ij} | \theta_i] = \dfrac{\sigma^2(\theta_i)}{w_{ij}} = \dfrac{\sigma_i^2}{w_{ij}}$

(A2) The pairs $(\underset{\sim}{\theta}_1, \underset{\sim}{X}_1), \dots,$ are independent and $\theta_1, \theta_2,$ are iid.

$\Rightarrow \underset{\sim}{X}_1, .., \underset{\sim}{X}_n$ are also iid

Let $\mu_0 = E[\mu(\theta_i)]$

$\sigma^2 = E[\sigma^2(\theta_i)]$

$\sigma_0^2 = Var[\mu(\theta_i)] = v^2$

$X_i = \sum_j w_{ij} X_{ij} / w_{i\cdot}$

where $w_i = \sum_j w_{ij}$

$\bar{X} = \sum w_i X_i / w_{\cdot\cdot}$

where $w_{\cdot\cdot} = \sum w_i = \sum\sum w_{ij}$

(1) $Var[X_{ij}] = E[Var[X_{ij}|\theta_i]] + Var[E(X_{ij}|\theta_i)]$

$= \dfrac{\sigma^2}{w_{ij}} + \sigma_0^2$

(2) $Var[X_i] = EV + VE$

$E[X_i | \theta_i] = \mu(\theta_i)$

$$Var(X_i|\theta_i) = \frac{\Sigma_j w_{ij}^2 \, Var(X_{ij}|\theta_i)}{w_{i\cdot}^2}$$

$$= \frac{\sigma^2(\theta_i)}{w_{i\cdot}}$$

$\therefore EV = \frac{\sigma^2}{w_{i\cdot}}$

$$VE = Var(\mu|\theta_i)$$

$$= \sigma_0^2 \quad (= v^2)$$

$\therefore Var[X_i] = \frac{\sigma^2}{w_{i\cdot}} + \sigma_0^2$

(3) $Var(\bar{X}_{\cdot\cdot}) = \frac{\Sigma w_{i\cdot}^2 \, Var(X_i)}{w_{\cdot\cdot}^2}$

as $X_1, X_2, \ldots$ are independent

$$\therefore \mathrm{Var}(\bar{X}) = \frac{1}{w_{\cdot\cdot}^2} \sum \left(w_{i\cdot}^2 \left(\frac{\sigma^2}{w_{i\cdot}} + \sigma_0^2 \right) \right)$$

$$= \frac{1}{w_{\cdot\cdot}^2} \left(w_{\cdot\cdot}\, \sigma^2 + \sum w_{i\cdot}^2\, \sigma_0^2 \right)$$

$$= \frac{\sigma^2}{w_{\cdot\cdot}} + v^2 \sum \frac{w_{i\cdot}^2}{w_{\cdot\cdot}^2}$$

(4) Decomposition of variance

$$SST = SSW + SSB.$$

$$\sum\sum w_{ij}(x_{ij} - \bar{x})^2 = \sum\sum w_{ij}(x_{ij} - x_i)^2 + \sum\sum w_{ij}(x_i - \bar{x})^2$$

Just write

$$x_{ij} - \bar{x} = x_{ij} - x_i + x_i - \bar{x}$$

and the cross product term is zero.

Note $SS\,B = \sum\sum w_{ij}(x_i - \bar{x})^2$

$$= \sum w_{i\cdot}\,(x_i - \bar{x})^2$$

(5) Consider $SSW = \sum_i \sum_j w_{ij} (X_{ij} - X_i)^2$

and $S_i^2 = \frac{1}{m-1} \sum_j w_{ij} (X_{ij} - X_i)^2$

S_i^2 is a "sample variance" for ith risk

$$\sum_j w_{ij} (X_{ij} - X_i)^2$$
$$= \sum_j w_{ij} (X_{ij} - \mu_i + \mu_i - X_i)^2$$
$$= \sum_j w_{ij} (X_{ij} - \mu_i)^2 - w_{i\cdot} (X_i - \mu_i)^2$$

$$\therefore E[S_i^2 | \theta_i] = \frac{1}{m-1} \left\{ \sum_j w_{ij} \operatorname{Var}[X_{ij} | \theta_i] - w_{i\cdot} \operatorname{Var}[X_i | \theta_i] \right\}$$

$$= \frac{1}{m-1} \left\{ m\, \sigma^2(\theta_i) - \sigma^2(\theta_i) \right\}$$

$$= \sigma^2(\theta_i)$$

$$\therefore E[S_i^2] = \sigma^2$$

S_i^2 is ube for σ^2

$\hat{\sigma}^2 = \frac{1}{n} \sum_i S_i^2$ is ube for σ^2

$$\boxed{\begin{array}{l}(6)\ \frac{1}{n(m-1)} \sum\sum w_{ij} (X_{ij}-X_i)^2 \text{ is ube for } \sigma^2 \\ \hat{\sigma}^2 = \frac{1}{n(m-1)} \cdot SSW \text{ is ube for } \sigma^2\end{array}}$$

(7) Consider $SSB = \sum_i w_i (X_i - \bar{X})^2$

$$= \sum_i w_i (X_i - \mu_0 + \mu_0 - \bar{X})^2$$

$$= \sum_i w_i (X_i - \mu_0)^2 - w_{..} (\bar{X} - \mu)^2$$

(we have done this kind of calculation with S_i^2)

$$\therefore E[SSB] = \sum w_i \, Var[X_i] - w_{..} \, Var[\bar{X}]$$

$$= \sum w_i \left(\frac{\sigma^2}{w_i} + \sigma_0^2 \right) - w_{..} \left(\frac{\sigma^2}{w_{..}} + \sigma_0^2 \frac{\sum w_i^2}{w_{..}^2} \right)$$

 See p 106.

$$E[SSB] = n\sigma^2 + w_{..}\sigma_0^2 - \sigma^2 - \sigma_0^2 \frac{\sum w_i^2}{w_{..}}$$

$$= (n-1)\sigma^2 + \sigma_0^2\left(w_{..} - \frac{\sum w_i^2}{w_{..}}\right)$$

$$= (n-1)\sigma^2 + \sigma_0^2 w_{..}\left(1 - \frac{\sum w_i^2}{w_{..}^2}\right)$$

$$\therefore E\left[\frac{SSB}{w_{..}}\right] = \frac{(n-1)\sigma^2}{w_{..}} + \sigma_0^2\left(1 - \frac{\sum w_i^2}{w_{..}^2}\right)$$

Consider $1 - \dfrac{\sum w_i^2}{w_{..}^2}$

$$= \sum \frac{w_i}{w_{..}}\left(1 - \frac{w_i}{w_{..}}\right)$$

and let $d = \left[\sum \dfrac{w_i}{w_{..}}\left(1 - \dfrac{w_i}{w_{..}}\right)\right]^{-1}$

(8) $$\therefore E\left[\frac{d \cdot SSB}{w_{..}} - \frac{d(n-1)\hat{\sigma}^2}{w_{..}}\right] = \sigma_0^2$$

(9) $d\left\{\sum_{i=1}^{n} \frac{w_{i\cdot}}{w_{\cdot\cdot}}(X_i-\bar{X})^2 - \left(\frac{n-1}{w_{\cdot\cdot}}\right)\hat{\sigma}^2\right\}$

is ube for $\sigma_0^2 (= v^2)$.

This is the same as Buhlmann Gisler page 95

Can also write as

$$\frac{d}{w_{\cdot\cdot}}\left(\sum_{i=1}^{n} w_i (X_i-\bar{X})^2 - \frac{(n-1)}{w_{\cdot\cdot}}\hat{\sigma}^2\right)$$

Now, based on (6) $\hat{\sigma}^2 = \frac{1}{n(m-1)} SSW$

So now we have

$\hat{\sigma}_0^2 = d\left\{\frac{SSB}{w_{\cdot\cdot}} - \frac{(n-1)}{w_{\cdot\cdot}}\frac{1}{n(m-1)} SSW\right\}$ is

ube for σ_0^2

using the decomposition $\hat{\sigma}_0^2$ can be written

$$d\left\{\frac{SST-SSW}{w_{\cdot\cdot}} - \frac{(n-1)}{w_{\cdot\cdot}}\frac{SSW}{n(m-1)}\right\}$$

$$= \frac{d}{w_{\cdot\cdot}} SST - \frac{d\,SSW}{w_{\cdot\cdot}} - \frac{d(n-1)}{w_{\cdot\cdot}\, n(m-1)} SSW$$

$$= \frac{d}{w_{\cdot\cdot}} SST - SSW \left\{ \frac{d}{w_{\cdot\cdot}} + \frac{d(n-1)}{w_{\cdot\cdot}\, n(m-1)} \right)$$

$$= \frac{d}{w_{\cdot\cdot}} SST - \frac{SSW}{n(m-1)} \left\{ \frac{d\, n(m-1)}{w_{\cdot\cdot}} + \frac{d(n-1)}{w_{\cdot\cdot}} \right\}$$

$$= d\left\{ \frac{SST}{w_{\cdot\cdot}} - \frac{SSW}{n(m-1)} \left(\frac{1}{w_{\cdot\cdot}}\right)\left(nm - \cancel{n} + \cancel{n} - 1\right) \right\}$$

$$= d\left\{ \frac{SST}{w_{\cdot\cdot}} - \frac{\hat{\sigma}^2 (nm-1)}{w_{\cdot\cdot}} \right\}$$

$$= \frac{d(nm-1)}{w_{\cdot\cdot}} \left\{ \frac{SST}{(nm-1)} - \hat{\sigma}^2 \right\} \quad — (10)$$

Now $\quad P^{*}$ (Schmidh) $\displaystyle = \frac{1}{nm-1} \sum w_i \left(1 - \frac{w_i}{w_{\cdot\cdot}}\right)$

$$= \frac{w_{\cdot\cdot}}{nm-1} \sum \frac{w_i}{w_{\cdot\cdot}} \left(1 - \frac{w_i}{w_{\cdot\cdot}}\right)$$

$$= \frac{w_{\cdot\cdot}}{(nm-1)\, d}$$

$\therefore$ (10) can be rewritten

$$\hat{\sigma}_0^2 (= \hat{v}^2)$$

$$= \frac{1}{p^*} \left\{ \frac{SST}{nm-1} - \hat{\sigma}^2 \right\}$$

QED!!

$$= d \left\{ \sum_{i=1}^{n} \frac{w_i}{w_{\cdot\cdot}} (x_i - \bar{x})^2 - \left(\frac{n-1}{w_{\cdot\cdot}}\right) \hat{\sigma}^2 \right\}$$

$$= \frac{d}{w_{\cdot\cdot}} \left\{ SSB - \frac{(n-1)}{w_{\cdot\cdot}} \frac{1}{n(m-1)} SSW \right\}.$$

The Mack Method
+ credibility – using Paul's formulation.

$$\text{ldf}_{i,j} = \frac{y_{ij}}{x_{ij}} = X_{ij}$$

$$w_{ij} = x_{ij}$$

$$E[X_{ij} \mid \theta_i] = E\left[\frac{y_{ij}}{x_{ij}} \mid \theta_i\right] = r(\theta_i)$$

mean ratio for division i.

$$\text{Var}[X_{ij} \mid \theta_i] = \text{Var}\left[\frac{y_{ij}}{x_{ij}} \mid \theta_i\right] = \frac{\sigma^2(\theta_i)}{x_{ij}}$$

for the Mack Method.

$$X_i = \frac{\sum_j x_{ij} \frac{y_{ij}}{x_{ij}}}{\sum_j x_{ij}} = \frac{\sum y_{ij}}{x_i}$$

$$= \text{ldf}_i$$

$$x = \sum x_i = \sum\sum x_{ij}$$

$$\bar{X} = \frac{\sum x_i X_i}{\sum x_i}$$

$$= \frac{\sum\sum y_{ij}}{x}$$

$$= ldf(s)$$

where $x = \sum x_i$

$$v^2 = Var[\, r(\vartheta_i)\,]$$

$$\sigma^2 = E[\, \sigma^2(\vartheta_i)\,]$$

$$\hat{\sigma}^2 = \frac{1}{n(m-1)} \sum\sum x_{ij}(ldf_{i,j} - ldf(s)_i)^2$$

$\hat{v}^2$ (as in Paul's note)

$$= \frac{1}{x^*}\left\{ \left(\frac{1}{nm-1}\right) \sum\sum x_{ij} (ldf_{i,j} - ldf(s))^2 - \hat{\sigma}^2 \right\}$$

$$x^* = \frac{1}{nm-1} \sum x_i (1 - x_i/x)$$

A better way for expressing the estimator for v^2 is

$$\hat{v}^2 = \frac{d}{x}\left\{ \sum_{i=1}^{n} x_i \left(ldf(d)_i - ldf(s)\right)^2 - (n-1)\sigma^2 \right\}.$$

The credibility adjusted Mack (equivalently, volume weighted link ratio) is

$$ldf(s) + \hat{Z}_i\left(ldf(d)_i - ldf(s)\right)$$

where $$\hat{Z}_i = \frac{1}{1+\frac{\hat{\sigma}^2}{x_i \hat{v}^2}}$$

According to our model

$$Z_i^{+} = \frac{1}{1+\frac{\hat{\sigma}_i^2}{x_i \hat{v}^2}}$$

where $\hat{\sigma}_i^2 = \frac{1}{M-1} \sum_j x_{ij}\left(ldf_{i,j} - ldf(d)_i\right)^2$

which is the u.b.e. for $\sigma^2(\theta_i) = \sigma_i^2$

$\hat{\sigma}_i^2$ is the sample variance for the ith risk (division i)

and $\dfrac{\hat{s}_i^2}{x_i}$ is the estimator

of the variance of the Mack link ratio for risk i.

This discussion with details will continue with next set of notes.

Note variance of credibility estimator is

$$(1-\hat{Z}_i)\, Var[ldf(s)]$$

$$= (1-\hat{Z}_i)\frac{1}{(mn-1)\,x} \cdot \sum\sum x_{ij}(ldf(i,j)-ldf(s))^2$$

This is an important statistic

Schmidli's analyses
- further comments

Why does Schmidli express the estimator of v^2 ($= \sigma_0^2$) in terms of SST rather than SSB?

After all, v^2 represents the variation in risk means across different risk groups.

Bottom of page 54, his starting point for estimating v^2 is SST, yet bottom of page 48 the starting point is SSB for the same model with weights equal to one! SST is <u>not</u> a natural starting point for developing an estimator for v^2.

SSB is the "natural" starting point as he himself says at bottom of

Link ratios formulated as regression estimators.

- Generalization of B-S model to other weighted averages

Consider again the credibility factor Z_i given by

$$Z_i = \frac{1}{1 + \frac{\sigma^2}{x_i \cdot \upsilon^2}}$$

On page 33° of the notes we find that $E[S_i^2] = \sigma^2$

S_i^2 from page 40° in the Mack formulation of volume weighted averages is given by

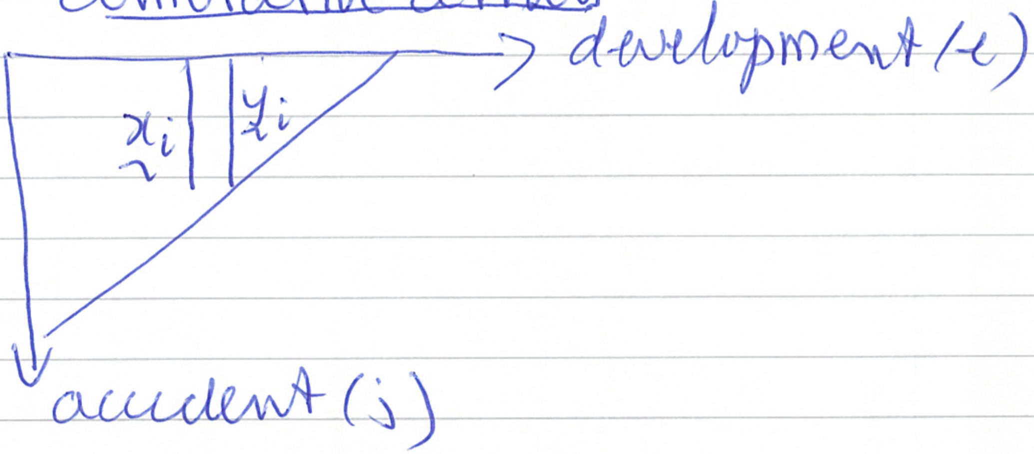
$$\frac{1}{m-1} \sum_j x_{ij} (ldf_{i,j} - ldf(a)_i)^2$$

Importantly, $\frac{S_i^2}{x_i}$ (estimate of $\frac{\sigma^2}{x_i}$)

is the estimator of the variance of the volume weighted average link ratio (that is the Mack estimator).

We now prove this and also generalize to other weighted average link ratios.

cumulative array



$$y_i = (y_{i1}, \ldots, y_{im})', \quad x_i = (x_{i1}, \ldots x_{im})'$$

$$ldf(i,j) = \frac{y_{ij}}{x_{ij}}$$

Consider the regression model

$$y_{ij} = r_i x_{ij} + \epsilon_{ij} \; ; \; j = 1, \ldots, m$$

where $Var(\epsilon_{ij}) = \sigma_i^2 x_{ij}^{\delta}$

The weighted least squares estimator of $r(i)$ minimizes

$$\sum_j x_{ij}^{-\delta} (y_{ij} - r(i) x_{ij})^2$$

$\hat{r}(i)$ the weighted least squares estimator is given by

$$\hat{r}(i) = \frac{\sum_j \frac{x_{ij}^2}{x_{ij}^{\delta}} \cdot \frac{y_{ij}}{x_{ij}}}{\sum_j \frac{x_{ij}^2}{x_{ij}^{\delta}}} \qquad (6.1)$$

and

$$\mathrm{Var}(\hat{r}(i)) = \frac{\sum_j \frac{x_{ij}^4}{x_{ij}^{2\delta}} \cdot \frac{\sigma_i^2 x_{ij}^{\delta}}{x_{ij}^2}}{\left(\sum_j \frac{x_{ij}^2}{x_{ij}^{\delta}}\right)^2} \qquad (6.2)$$

$$\hat{\sigma}_i^2 = \frac{1}{m-1} \sum_j x_{ij}^{-\delta} (y_{ij} - \hat{r}(i) x_{ij})^2 \qquad (6.3)$$

Case 1. $\delta = 1$, $w_{ij} = x_{ij}$ in B-S model

$$\hat{r}(i) = \sum_j y_{ij} / \sum_j x_{ij} = \text{volume weighted average} \qquad (6.4)$$

$$Var(\hat{r}(i)) = \sigma_i^2 \frac{\sum_j x_{ij}}{(\sum_j x_{ij})^2}$$

$$= \frac{\sigma_i^2}{\sum_j x_{ij}} = \frac{\sigma_i^2}{x_i} \qquad (6.5)$$

$$\hat{\sigma}_i^2 = \frac{1}{m-1} \sum_j x_{ij}^{-1} (y_{ij} - \hat{r}(i) x_{ij})^2$$

$$= \frac{1}{m-1} \sum_j x_{ij} \left(\frac{y_{ij}}{x_{ij}} - \hat{r}(i) \right)^2$$

$$= \frac{1}{m-1} \sum_j x_{ij} (ldf(i,j) - ldf(i))^2$$

$$= s_i^2 \qquad (6.6)$$

$$\therefore \hat{Var}(\hat{r}(i)) = \frac{s_i^2}{x_i} = \frac{\hat{\sigma}_i^2}{x_i} \qquad (6.7)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i s_i^2 \qquad (6.8)$$

and $$\hat{z}_i = \frac{1}{1 + \frac{\hat{\sigma}^2}{x_i \hat{v}^2}} \qquad (6.9)$$

Case 2. $\delta = 2$, $w_{ij} = 1$ in B-S model

Using (6.1), (6.2) + (6.3) we obtain

$$\hat{r}(i) = \frac{1}{m} \sum_j y_{ij}/x_{ij} = \text{arithmetic average} \qquad (6.10)$$

$$\mathrm{Var}(\hat{r}(i)) = \frac{\sum_j \sigma_i^2}{\left(\sum_j 1\right)^2}$$

$$= \frac{\sigma_i^2}{m} \qquad (6.11)$$

$$\hat{\sigma}_i^2 = \frac{1}{m-1} \sum_j x_{ij}^{-2} \left(y_{ij} - \hat{r}(i)\, x_{ij}\right)^2$$

$$= \frac{1}{m-1} \sum_j \left(ldf(i,j) - \hat{r}(i)\right)^2 \qquad (6.12)$$

$$\hat{\mathrm{Var}}(\hat{r}(i)) = \frac{\hat{\sigma}_i^2}{m} \qquad (6.13)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i \hat{\sigma}_i^2 \qquad (6.14)$$

$$\hat{z}_i = \frac{1}{1 + \frac{\hat{\sigma}^2}{m v^2}}$$

Case 3 : $\delta = 0$, $w_{ij} = x_{ij}^2$ in B-S model

Again using (6.1), (6.2) + (6.3) we obtain

$$\hat{r}(i) = \frac{\sum_j x_{ij}^2 \; y_{ij}/x_{ij}}{\sum_j x_{ij}^2}$$

= volume^2 weighted average (6.15)

$$Var(\hat{r}(i)) = \sigma_i^2 \frac{\sum_j x_{ij}^2}{(\sum_j x_{ij}^2)^2}$$

$$= \frac{\sigma_i^2}{\sum x_{ij}^2} \qquad (6.16)$$

$$\hat{\sigma}_i^2 = \frac{1}{m-1} \sum_j (y_{ij} - \hat{r}(i) x_{ij})^2$$

$$= \frac{1}{m-1} \sum_j x_{ij}^2 (ldf(i,j) - \hat{r}(i))^2$$

(6.17)

$$\therefore Var(\hat{r}(i)) = \frac{\hat{\sigma}_i^2}{\sum x_{ij}^2} \qquad (6.18)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum \hat{\sigma}_i^2 \qquad (6.19)$$

$$\hat{z}_i = \frac{1}{1 + \frac{\hat{\sigma}^2}{\sum x_{ij}^2 \, V^2}} \qquad (6.20)$$

Buhlmann-Straub for Mack method (model ($\delta=1$)) revisited

The linear Bayes rule (credibility estimator) that minimizes the Bayes risk

$$E[(a+b\,ldf(d)_i - r(\theta_i))^2] \quad (6.21)$$

is

$$ldf(s) + Z_i\,(ldf(d)_i - ldf(s))$$

$$= (1-Z_i)\,ldf(s) + Z_i\,ldf(d)_i$$

where

$$Z_i = \frac{1}{1+\frac{\sigma^2}{x_i v^2}}$$

The E operator in equation (6.21) is over the distributions of $y_{ir}|\theta_i$ and θ_i.

Note:

1. The Bayes risk averages over all possible values of the hyperparameter (structure parameter) θ_i.

2. In the credibility factor

$$\hat{Z}_i = \frac{1}{1 + \frac{\hat{\sigma}^2}{w^* \, V^2}} \, ,$$

the value $\frac{\hat{\sigma}^2}{w^*}$ is equal to

$$\frac{1}{w^*} \frac{1}{n} \sum_i \hat{\sigma}_i^2$$

where $\frac{\hat{\sigma}_i^2}{w^*}$ is the variance of the individual risk link ratio estimator $\hat{r}(i)$.

Would it not be better to use

the variance $\frac{\hat{\sigma}_i^2}{w_*}$ of $\hat{r}(i)$, instead of the average of all variances $\frac{\hat{\sigma}^2}{w_*}$?

Both $\hat{\sigma}^2$ and $\hat{\sigma}_i^2$ are unbiased for σ^2 in the B-S model, but $\hat{\sigma}_i^2$, alternatively, the precision $\frac{1}{\hat{\sigma}_i^2}$ is reflective of the uncertainty in $\hat{r}(i)$, not $\frac{1}{\hat{\sigma}^2}$.

Moreover, if $\hat{r}$, the estimator of the average link ratio across all divisions is significantly different to $\hat{r}(i)$, (the link ratio estimator for individual division), should we adjust $\hat{r}(i)$? The answer is definitely in the negative!

Alternative formulation of the problem

Based on the preceding discussion it does not make sense to adjust (pull) the estimator $\hat{r}(i)$ towards $\hat{r}$ if the two estimators are statistically different. $\hat{r}(i)$ should only be adjusted if the null hypothesis

$$H_0 : r(i) = r$$

cannot be rejected.

Using the regression approach this hypothesis can be tested.

If indeed, $H_0: r(c) = r$

is not rejected than the best linear weighted least squares estimator ~~#~~ $\hat{r}_c(c)$ of $r(c)$ is

$$(1 - z_i^{*})\,\hat{r} + z_i^{*}\,\hat{r}(c)$$

where
$$z_i^{*} = \frac{1/\mathrm{Var}(\hat{r}(c))}{\dfrac{1}{\mathrm{Var}(\hat{r})} + \dfrac{1}{\mathrm{Var}(\hat{r}(c))}}$$

That is, $\hat{r}_c(c)$ is a weighted average based on the relative precisions of both estimators.

$$\mathrm{Var}[\hat{r}_c(c)] = (1 - z_i^{*})\,\mathrm{Var}(\hat{r}),$$

Alternatively,
$$\frac{1}{\mathrm{Var}[\hat{r}_c(c)]} = \frac{1}{\mathrm{Var}[\hat{r}]} + \frac{1}{\mathrm{Var}[\hat{r}(c)]}$$

Notes:

1. The regression model for $\hat{r}$ and $\hat{r}(i)$ do <u>not</u> have to be based on the same value of δ.

2. The number of accident years used in estimating $r(i)$ and r do <u>not</u> have to be the same

3. $\hat{r}$ is based on all divisions <u>except</u> division i.

4. $\hat{r}$ does <u>not</u> have to be based on <u>all</u> divisions other than i. It could be based on any number of divisions other than i.

5. By computing $\hat{r}(i)$ and $Var(\hat{r}(i))$ for all divisions one may be able to select the divisions that statistically have the same ratio as i, before any further analyses.