INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Class: SY MSc

Subject : Statistical & Risk Modelling - 2

Chapter: Unit 2 Chapter 1

Chapter Name: Applied Bayesian Modelling and Credibility Theory

Agenda

1. Bayes Theorem
    1. Prior and Posterior Distributions
    2. Notations
    3. Continuous Prior Distributions
    4. Priors
2. The loss function
3. Credibility
4. The Poisson/Gamma Model
5. The Normal/Normal Model
6. Empirical Bayes Credibility Theorem
    1. EBCT Model 1
    2. EBCT Model 2

0 Introduction

The Bayesian philosophy involves a completely different approach to statistics, compared to classical statistical methods. The Bayesian version of estimation is considered here for the basic situation concerning the estimation of a parameter given a random sample from a particular distribution. Classical estimation involves the method of maximum likelihood.

The fundamental difference between Bayesian and classical methods is that the parameter **θ is considered to be a random variable in Bayesian methods.**

In **classical statistics, θ is a fixed but unknown quantity**.

1 Bayes' theorem

If $B_1, B_2, \dots, B_k$ constitute a partition of a sample space S and $P(B_i) \neq 0$ for $i = 1,2, \dots, k$, then for any event A in S such that $P(A) \neq 0$:

$$\boldsymbol{P(B_r \mid A) = \frac{P(A \mid B_r)P(B_r)}{P(A)} \quad \text{where } P(A) = \sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$$

for $r = 1,2, \dots, k$.

Bayes' theorem can be adapted to deal with continuous random variables. If X and Y are continuous, then the conditional PDF of Y given X is:

$$\boldsymbol{f_{Y|X}(x,y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y}(x,y)f_Y(y)}{f_X(x)}}$$

where:

$$f_X(x) = \int_y f_{X,Y}(x,y)dy = \int_y f_{X|Y}(x,y)f_Y(y)dy$$

1.1 Prior and posterior distributions

Suppose $\underline{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from a population specified by the density or probability function $f(x;\theta)$ and it is required to estimate θ. Recall that a random sample is a set of IID random variables.

As a result of the parameter θ being a random variable, it will have a distribution. This allows the use of any knowledge available about possible values for θ before the collection of any data. This knowledge is quantified by expressing it as the prior distribution of θ.

The prior distribution summarises what we know about θ before we collect any data from the relevant population.

Then after collecting appropriate data, the posterior distribution of θ is determined, and this forms the basis of all inference concerning θ.

The Bayesian approach combines the sample data with the prior distribution. **The conditional distribution of θ given the observed data is called the posterior distribution of θ.**

1.2 Notations

As θ is a random variable, it should really be denoted by the capital Θ, and its prior density written as $f_{\Theta}(\theta)$. However, for simplicity no distinction will be made between Θ and θ, and the density will simply be denoted by $f(\theta)$.

Note that referring to a density here implies that θ is continuous.
In most applications this will be the case, as even when $\boldsymbol{X}$ is discrete (like the binomial or Poisson), the parameter (p or λ) will vary in a continuous space ($(0,1)$ or $(0,\infty)$, respectively).

Also, the population density or probability function will be denoted by $f(x \mid \theta)$ rather than the earlier $f(x;\theta)$ as it represents the conditional distribution of X given θ.

1.3 Continuous prior distributions

Suppose that $\underline{X}$ is a random sample from a population specified by $f(x \mid \theta)$ and that θ has the prior density $f(\theta)$.

In other words, $X_1, \ldots, X_n$ is a set of IID random variables whose distribution depends on the value of θ. Each of these random variables has PDF $f(x \mid \theta)$.

We now look forward to determining the posterior density

1.3 Continuous prior distributions

Determining the Posterior density

The posterior density of $\theta \mid \underline{X}$ is determined by applying the basic definition of a conditional density:

$$f(\theta \mid \underline{X}) = \frac{f(\theta, \underline{X})}{f(\underline{X})} = \frac{f(\underline{X} \mid \theta) f(\theta)}{f(\underline{X})}$$

Note that $f(\underline{X}) = \int f(\underline{X} \mid \theta) f(\theta) d\theta$. This result is like a continuous version of Bayes' theorem.

A useful way of expressing the posterior density is to use proportionality. $f(\underline{X})$ does not involve θ and is just the constant needed to make it a proper density that integrates to unity, so:

$$\boldsymbol{f(\theta \mid \underline{X}) \propto f(\underline{X} \mid \theta) f(\theta)}$$

This formula is given on page 28 of the Tables.

1.3 Continuous prior distributions

Determining the Posterior density

The formula for the posterior PDF can also be expressed as follows:

$$\boldsymbol{f_{\text{post}}}\ (\boldsymbol{\theta}) = \boldsymbol{C} \times \boldsymbol{f_{\text{prior}}}\ (\boldsymbol{\theta}) \times \boldsymbol{L}$$

where:

- $f_{\text{prior}}\ (\theta)$ is the prior PDF of θ
- $f_{\text{post}}\ (\theta)$ is the posterior PDF of θ
- L is the likelihood function obtained from the sample data
- C is a constant that makes the posterior PDF integrate to 1.

1.4 Priors

- **Conjugate Prior**
  For a given likelihood, if the prior distribution leads to a posterior distribution belonging to the same family as the prior distribution, then this prior is called the conjugate prior for this likelihood.

- **Uninformative prior distributions**
  An uninformative prior distribution assumes that an unknown parameter is equally likely to take any value from a given set. In other words, the parameter is modelled using a uniform distribution.

- **Discrete prior distributions**
  When the prior distribution is discrete, the posterior distribution is also discrete. To determine the posterior distribution, we must calculate a set of conditional probabilities. This can be done using Bayes' formula.

2 The loss function

To obtain an estimator of θ, a loss function must first be specified. This is a measure of the 'loss' incurred when $g(\underline{X})$ is used as an estimator of θ.
A loss function is sought which is zero when the estimation is exactly correct, that is, $g(\underline{X}) = \theta$, and which is positive and does not decrease as $g(\underline{X})$ gets further away from θ.

There is one very commonly used loss function, called quadratic or squared error loss. Two others are also used in practice.

Then the Bayesian estimator is the $g(\underline{X})$ that minimizes the expected loss with respect to the posterior distribution.

The main loss function is quadratic loss defined by:

$$\boldsymbol{L(g(\underline{x}), \theta) = [g(\underline{x}) - \theta]^2}$$

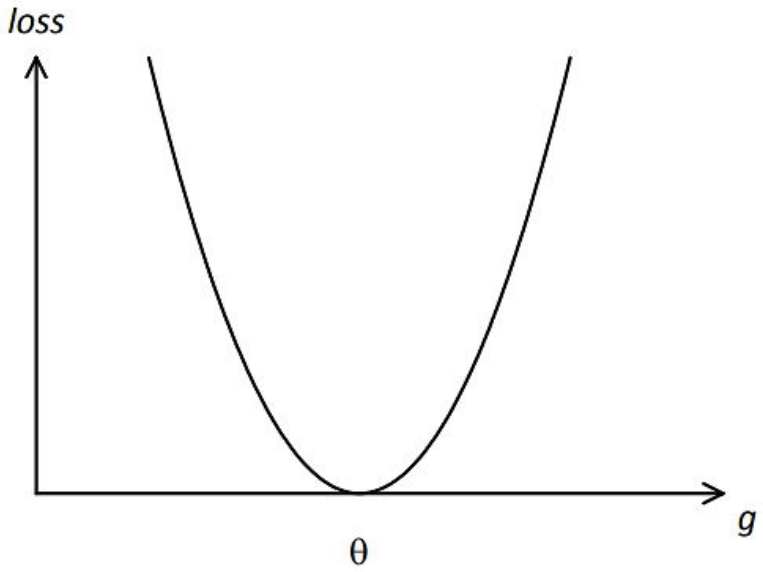
So, when using quadratic loss, the aim is to minimize:

$$E\left[(g(\underline{x}) - \theta)^2\right] = \int_{\theta} (g(\underline{x}) - \theta)^2 f_{\text{post}}\ (\theta) d\theta$$

2 The loss function

The formula for the squared error loss implies that as we move further away from the true parameter value, the loss increases at an increasing rate.

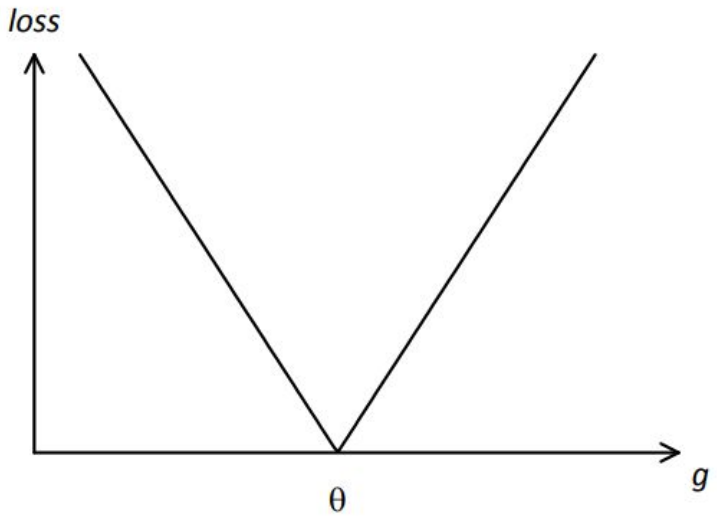
The graph of the loss function under the quadratic loss function is a parabola with a minimum of zero at the true parameter value.



2 The loss function

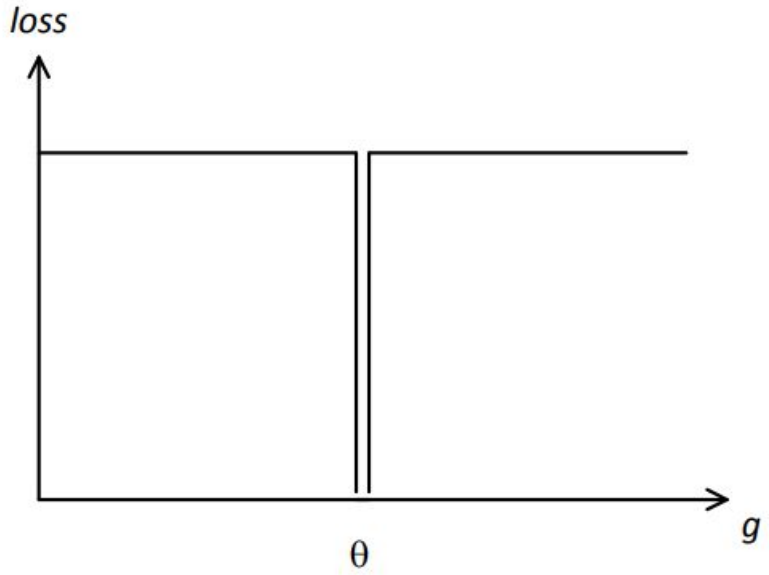
A second loss function is absolute error loss defined by:

$$L(g(\underline{x}), \theta) = |g(\underline{x}) - \theta|$$



A third loss function is ' $0/1$ ' or 'all-or-nothing' loss defined by:

$$L(g(\underline{x}), \theta) = \begin{cases} 0 & \text{if } g(\underline{x}) = \theta \\ 1 & \text{if } g(\underline{x}) \neq \theta \end{cases}$$



2 The loss function

Quadratic loss

For simplicity, g will be written instead of $g(\underline{x})$. So:

$$EPL = \int (g-\theta)^2 f(\theta \mid \underline{x}) d\theta$$

Using the formula for differentiating, we see that:

$$\frac{d}{dg} EPL = 2 \int (g-\theta) f(\theta \mid \underline{x}) d\theta$$

Equating to zero:

$$g \int f(\theta \mid \underline{x}) d\theta = \int \theta f(\theta \mid \underline{x}) d\theta$$

But $\int f(\theta \mid \underline{x}) d\theta = 1$.

2 The loss function

Quadratic loss

But $\int f(\theta \mid \underline{x})d\theta = 1$.
This is because $f(\theta \mid \underline{x})$ is the PDF of the posterior distribution. Integrating the PDF over all possible values of θ gives the value 1. So:

$$\boldsymbol{g = \int \theta f(\theta \mid \underline{x})d\theta = E(\theta \mid \underline{x})}$$

Clearly this minimizes EPL.

We can see this from the graph of the loss function or by differentiating the EPL a second time:

$$\frac{d^2}{dg^2}EPL = 2\int f(\theta \mid \underline{x})d\theta = 2 > 0 \Rightarrow \min$$

Therefore, **the Bayesian estimator under quadratic loss is the mean of the posterior distribution.**

2 The loss function

Absolute error loss

Again, g will be written instead of $g(\underline{x})$. So:

$$EPL = \int |g - \theta| f(\theta \mid \underline{x}) d\theta$$

Assuming the range for θ is $(-\infty, \infty)$, then:

$$EPL = \int_{-\infty}^{g} (g - \theta) f(\theta \mid \underline{x}) d\theta + \int_{g}^{\infty} (\theta - g) f(\theta \mid \underline{x}) d\theta$$

So:

$$\frac{d}{dg} EPL = \int_{-\infty}^{g} f(\theta \mid \underline{x}) d\theta - \int_{g}^{\infty} f(\theta \mid \underline{x}) d\theta$$

Equating to zero:

$$\int_{-\infty}^{\boldsymbol{g}} \boldsymbol{f(\theta \mid \underline{x}) d\theta} = \int_{\boldsymbol{g}}^{\infty} \boldsymbol{f(\theta \mid \underline{x}) d\theta}$$

that is, $P(\theta \leq g) = P(\theta \geq g)$, which specifies the median of the posterior distribution.

2 The loss function

All-or-nothing loss

Here the differentiation approach cannot be used. Instead, a direct approach will be used with a limiting argument.

Consider:

$$L(g(\underline{x}),\theta) = \begin{cases} 0 & \text{if } g-\varepsilon < \theta < g+\varepsilon \\ 1 & \text{otherwise} \end{cases}$$

so that, in the limit as $\varepsilon \to 0$, this tends to the required loss function.

Then the expected posterior loss is:

$$\boldsymbol{EPL = 1 - \int_{g-\varepsilon}^{g+\varepsilon} f(\theta \mid \underline{x})d\theta = 1 - 2\varepsilon . f(g \mid \underline{x})} \quad \textbf{for small } \boldsymbol{\varepsilon}$$

The EPL is minimised by taking g to be the mode of $f(\theta \mid \underline{x})$.

2 The loss function

A loss function, such as quadratic (or squared) error loss, absolute error loss or all-or-nothing $(0/1)$ loss gives a measure of the loss incurred when $\hat{\theta}$ is used as an estimator of the true value of θ.

In other words, it measures the seriousness of an incorrect estimator.

Under squared error loss, the mean of the posterior distribution minimizes the expected loss function.

Under absolute error loss, the median of the posterior distribution minimizes the expected loss function.

Under all-or-nothing loss, the mode of the posterior distribution minimizes the expected loss function.

Question

CS1A September 2023 Q7

Total losses in a particular company are modelled by a random variable Y with density function:

$$f(y) = \begin{cases} \dfrac{c}{y^{c+1}}, & y > 1, \quad c > 0 \\ 0, & \text{otherwise.} \end{cases}$$

An analyst wishes to estimate the unknown parameter c.

(i) Derive the maximum likelihood estimate for parameter c: [3]

The analyst assumes a gamma prior distribution for c with parameters (a, b).

(ii) Determine the posterior distribution of c with all its parameters. [6]

(iii) Comment on the relationship between the prior distribution and the posterior distribution of c. [1]

(iv) Determine the Bayesian estimate of parameter c under quadratic loss. [2] [Total 12]

Solution

(i)

Correct answer: C [3]

The likelihood for the parameter c given n independent randomly sample is

$L(c) = \prod_{i=1}^{n} \frac{c}{y_i^{c+1}} = c^n \prod_{i=1}^{n} \frac{1}{y_i^{c+1}}$

$l(c) = nlog(c) - (c+1)\ \sum_1^n \log(y_i)$.

The corresponding partial derivative is $\frac{\partial l}{\partial c} = \frac{n}{c} - \sum_1^n \log(y_i)$

$\frac{\partial l}{\partial c} = 0$ then $c = \frac{n}{\sum_1^n \log(y_i)}$.

The MLE of c is $\hat{c} = \frac{n}{\sum_1^n \log(y_i)}$.

Solution

(ii)

The prior distribution is: $f(c) \propto c^{a-1}e^{-bc}$. [1]

The likelihood is: $L(c) = c^n e^{-(c+1)\sum_i^n \log(y_i)} \propto c^n e^{-c\sum_i^n \log(y_i)}$ [1½]

The posterior distribution is given as:

$P(c) \propto c^{a-1}e^{-bc}c^n e^{-c\sum_i^n \log(y_i)} = c^{n+a-1}e^{-c(b+\sum_i^n \log(y_i))}$ [1½]

The posterior distribution of the parameter c is a $Gamma(n+a,\ b+\sum_i^n \log(y_i))$. [2]

(iii)

The posterior and the prior distributions are from the same family, [½]

therefore the prior is a conjugate prior. [½]

(iv)

Under a quadratic loss, the Bayesian estimate is the posterior mean [1]

i.e. $\frac{n+a}{b+\sum_i^n \log(y_i)}$ [1]

[Total 12]

3 Revisiting Conditional expectations

If X and Y are discrete random variables, then:

$$E(X \mid Y = y) = \sum_{x} xP(X = x \mid Y = y)$$

Similarly, if X and Y are continuous random variables, then:

$$E(X \mid Y = y) = \int_{x} x f_{X|Y}(x, y)dx$$

3 Revisiting Conditional expectations

Manipulation of conditional expectations is an important technique in credibility theory, as it is in many other areas of actuarial science. Some results are:

For any random variables X and Y (for which the relevant moments exist):

$$E[X] = E[E(X \mid Y)]$$

Another important concept is that of conditional independence. If two random variables X_1 and X_2 are conditionally independent given a third random variable Y, then:

$$E[X_1 X_2 \mid Y] = E[X_1 \mid Y]E[X_2 \mid Y]$$

3 Credibility

Credibility Premium Formula

The basic idea underlying the credibility premium formula is intuitively very simple and very appealing. The following information is available:

- $\bar{X}$ is an estimate of the expected aggregate claims / number of claims for the coming year based solely on data from the risk itself.
- μ is an estimate of the expected aggregate claims / number of claims for the coming year based on collateral data, ie data for risks similar to, but not necessarily identical to, the particular risk under consideration.

The credibility premium formula (or credibility estimate of the aggregate claims / number of claims) **for this risk is:**

$$\boldsymbol{Z\bar{X} + (1 - Z)\mu}$$

where Z is a number between zero and one and is known as the credibility factor.

3 Credibility

Credibility Factor

The credibility factor Z is just a weighting factor. Its value reflects how much 'trust' is placed in the data from the risk itself, $\bar{X}$, compared with the data from the larger group, μ, as an estimate of next year's expected aggregate claims or number of claims - the higher the value of Z, the more trust is placed in $\bar{X}$ compared with μ, and vice versa.

In general terms, the credibility factor would be expected to behave as follows:

- The more data there are from the risk itself, the higher should be the value of the credibility factor.
- The more relevant the collateral data, the lower should be the value of the credibility factor.

3 Credibility

Credibility Factor

One final point to be made about the credibility factor is that, while its value should reflect the amount of data available from the risk itself, its value should not depend on the actual data from the risk itself, ie on the value of $\bar{X}$.

If $\boldsymbol{Z}$ were allowed to depend on $\bar{X}$ then any estimate of the aggregate claims/number of claims, say ϕ, taking a value between $\bar{X}$ and μ could be written in the form of Z by choosing Z to be equal to $(\phi - \mu)/(\bar{X} - \mu)$.

The problems remain of how to measure the relevance of collateral data and how to calculate the credibility factor Z .

There are two approaches to these problems: **Bayesian credibility and empirical Bayes credibility theory.**

3.1 Bayesian Credibility

The Bayesian approach to credibility involves the same steps as Bayesian estimation, described in the last chapter:

- We start with a prior distribution for the unknown parameter under consideration (eg the claim frequency), which summarizes any knowledge we have about its possible values. The form of the prior distribution should be derived from information provided by the collateral data.
- We then collect relevant data and use these values to obtain the likelihood function.
- The prior distribution and likelihood function are combined to produce the posterior distribution.
- A loss function is specified to quantify how serious misjudging the parameter value would be. The loss function should be based on commercial considerations of the financial effect on the insurer's business of incorrectly estimating the parameter (and hence the premium rates).
- The Bayesian estimate of the parameter value is then calculated.

4 The Poisson/gamma model

Suppose the claim frequency for a risk, ie the expected number of claims in the coming year, needs to be estimated.
The problem can be summarized as follows.

- The number of claims each year is assumed to have a Poisson distribution with parameter λ.
- The value of λ is not known, but estimates of its value are possible along the lines of, for example, 'there is a $\mathbf{50}\%$ chance that the value of λ is between 50 and 150'.
- More precisely, before having available any data from this risk, the feeling about the value of λ is that it has a $\mathrm{Gamma}(\alpha, \beta)$ distribution.
- The gamma distribution is the conjugate prior for Poisson data. So, if the number of claims each year has a Poisson distribution with parameter λ and we use a gamma distribution as the prior distribution for λ, the posterior distribution for λ will also be a gamma distribution.
- Data from this risk are now available showing the number of claims arising in each of the past n years.

4 The Poisson/gamma model

This problem fits exactly into the framework of Bayesian statistics and can be summarized more formally as follows.

- The random variable $\boldsymbol{X}$ represents the number of claims in the coming year from a risk.
- The distribution of $\boldsymbol{X}$ depends on the fixed, but unknown, value of a parameter, λ.
- The conditional distribution of $\boldsymbol{X}$ given λ is Poisson(λ).
- The prior distribution of λ is $\mathrm{Gamma}(\alpha, \beta)$.

The problem is to estimate λ given the data $\underline{x}$, and the estimate wanted is the Bayes estimate with respect to quadratic loss, ie $E(\lambda \mid \underline{x})$.

Combining the prior distribution and the sample data, we see that:

$$f_{\mathbf{post}}(\boldsymbol{\lambda}) \propto \boldsymbol{\lambda}^{\alpha-1} e^{-\beta\lambda} \times e^{-n\lambda} \boldsymbol{\lambda}^{\sum x_i} = \boldsymbol{\lambda}^{\sum x_i + \alpha - 1} e^{-(n+\beta)\lambda}, \ \boldsymbol{\lambda > 0}$$

The posterior distribution of λ given $\underline{x}$ is $\mathrm{Gamma}(\alpha + \sum_{i=1}^{n} x_i, \beta + n)$.

5 The Normal/normal model

The problem is to estimate the pure premium, ie the expected aggregate claims, for a risk. Let $\boldsymbol{X}$ be a random variable representing the aggregate claims in the coming year for this risk. The following assumptions are made.

- The distribution of $\boldsymbol{X}$ depends on the fixed, but unknown, value of a parameter, θ.
- The conditional distribution of X given θ is $N(\theta, \sigma_1^2)$.
- The uncertainty about the value of θ is modelled in the usual Bayesian way by regarding it as a random variable.

The **prior distribution of $\boldsymbol{\theta}$ is $\boldsymbol{N(\mu, \sigma_2^2)}$.**

So, **the posterior distribution of $\boldsymbol{\theta}$ given $\underline{x}$ is:**

$$\boldsymbol{N\left(\frac{\mu\sigma_1^2 + n\sigma_2^2\bar{x}}{\sigma_1^2 + n\sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + n\sigma_2^2}\right)}$$

where:

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

5 The Normal/normal model

The Bayesian estimate of θ under quadratic loss is the mean of this posterior distribution:

$$E(\theta \mid \underline{x}) = \frac{\mu\sigma_1^2 + n\sigma_2^2\bar{x}}{\sigma_1^2 + n\sigma_2^2}$$

$$= \frac{\sigma_1^2}{\sigma_1^2 + n\sigma_2^2}\mu + \frac{n\sigma_2^2}{\sigma_1^2 + n\sigma_2^2}\bar{x}$$

or:

$$\boldsymbol{E(\theta \mid \underline{x}) = Z\bar{x} + (1 - Z)\mu}$$

where:

$$Z = \frac{n}{n + \sigma_1^2/\sigma_2^2}$$

5 The Normal/normal model

There are some further points to be made about the credibility factor, Z, given by:

- It is always between zero and one.
- It is an increasing function of $\boldsymbol{n}$, the amount of data available.
- It is an increasing function of σ_2, the standard deviation of the prior distribution.

These features are all exactly what would be expected for a credibility factor.

Question

CS1A September 2023 Q7

Let $X_1, X_2, \ldots, X_n$ be independent observations from a Bernoulli distribution with P(Xi = 1) = p, i = 1, ..., n. The parameter p has a beta prior distribution with parameters (a, b).

(i) Determine the posterior distribution of parameter p. [6]

(ii) Determine the Bayesian estimate of parameter p under quadratic loss. [1]

(iii) Determine the Bayesian estimate of parameter p under quadratic loss as a credibility estimate, stating the credibility factor. [2]

[Total 9]

Solution

(i)
Likelihood function:
L(p) = product(i in1:n) {p^x_i * (1-p)^(1-x_i)}
= p^{sum(i in1:n) x_i} * (1-p)^{n- sum(i in1:n) x_i } [1]

Prior density:
f(p) is proportional to: p^(a-1) * (1-p)^(b-1) [1]

Posterior density:
f(p|x) proportional to L(p) * f(p) [½]
= p^{sum(i in1:n) x_i} * (1-p)^{n- sum(i in1:n) x_i } * p^(a-1) * (1-p)^(b-1)
= p^{sum(i in1:n) x_i+ a-1}} * (1-p)^{n+b - sum(i in1:n) x_i -1} [1½]

So, the posterior is a beta distribution [1]
with parameters sum(i in1:n) x_i+ a and n+b - sum(i in1:n) x_i [1]

(ii)
The Bayes estimate under quadratic loss is the posterior mean, i.e. [½]
{sum(i in1:n) x_i+ a} / (a+b+n) [½]

Solution

(ii)
The Bayes estimate under quadratic loss is the posterior mean, i.e. [½]
{sum(i in1:n) x_i+ a} / (a+b+n) [½]

(iii)
We can write the estimate as
{sum(i in1:n) x_i+ a} / (a+b+n) = {n/(n+a+b) * {sum(i in1:n) x_i / n} + (a+b)/(n+a+b) * {a/(a+b)}
= Z * xbar + (1-Z) * prior mean, [1½]

where Z = n/(n+a+b) is the credibility factor [½]
[Total 9]

Question

CS1A April 2024 Q7

(i) Explain what is meant by a conjugate prior distribution. (1)

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables from the Poisson distribution with parameter m, and m follows, a priori, a gamma distribution with probability density function given by

$$f(m) = \frac{s^a}{\Gamma(a)} m^{a-1} e^{-sm}, \text{ with } a, s > 0.$$

(ii) Show that this prior distribution is conjugate for m. [3]

(iii) Determine the mean and variance of the posterior distribution of m. [2]

(iv) Comment on how the prior distribution affects the posterior for large sample sizes, n. [2]

Consider now the parameter 1/m.

(v) (a) Find the prior mean for parameter 1/m:

(b) Find the posterior mean for parameter 1/m:

Solution

(i)

The prior distribution of a parameter is called conjugate if the posterior distribution is of the same family as the prior.

(ii)

The posterior distribution can be written as

$$f(m|\boldsymbol{x}) \propto f(\boldsymbol{x}|m)f(m)$$

$$= \prod_{i=1}^{n}\left(\frac{m^{x_i}e^{-m}}{x_i!}\right)\frac{s^a}{\Gamma(a)}m^{a-1}\,e^{-sm}$$

$$\propto m^{a+n\bar{x}-1}e^{-(s+n)m},$$

which implies that $m|\boldsymbol{x}$ follows a Gamma $(a+n\bar{x}, s+n)$.

Solution

(iii)

The posterior mean is given by

$$E[m|\boldsymbol{x}] = \frac{a + n\bar{x}}{s + n},$$

and the posterior variance

$$V[m|\boldsymbol{x}] = \frac{a + n\bar{x}}{(s + n)^2}.$$

(iv)

As n tends to ∞, $E[m|\boldsymbol{x}]$ tends to $\bar{x}$ and $V[m|\boldsymbol{x}]$ tends to $\frac{\bar{x}}{n}$. For large values of n, the posterior mean and variance are not affected by the prior specified for m through a and s.

Solution

(v) (a)

Correct answer is D.

$$E\left(\frac{1}{m}\right) = \frac{s}{a-1}$$

We can write:

$$E[1/m] = \int_0^\infty \frac{f(m)}{m} dm = \int_0^\infty \frac{s^a m^{a-1} e^{-sm}}{m\Gamma(a)} dm = \int_0^\infty \frac{s^a m^{a-2} e^{-sm}}{\Gamma(a)} dm$$

$$= \frac{s}{a-1} \int_0^\infty \frac{s^{a-1} m^{a-2} e^{-sm}}{\Gamma(a-1)} dm = \frac{s}{a-1} \times 1$$

Since the last integral is the pdf of a Gamma $(\alpha - 1, s)$.

Solution

(b)

Correct answer is A.

$$E\left(\frac{1}{m} \mid x\right) = \frac{s+n}{a+n\bar{x}-1}$$

We have:

$$E\left[\frac{1}{m} \mid x\right] = \int_0^\infty \frac{f(m \mid x)}{m} dm = \int_0^\infty \frac{(s+n)^{(a+n\bar{x})} m^{a+n\bar{x}-2} e^{-(s+n)m}}{\Gamma(a+n\bar{x})} dm$$

$$= \frac{s+n}{a+n\bar{x}-1} \int_0^\infty \frac{(s+n)^{a+n\bar{x}-1}}{\Gamma(a+n\bar{x}-1)} m^{a+n\bar{x}-2} e^{-(s+n)m} dm$$

$$= \frac{s+n}{a+n\bar{x}-1} \times 1$$

Since the last integral is the pdf of a Gamma $(\alpha + n\bar{x} - 1, s+n)$.

6 Empirical Bayes Credibility Theory

In this section we will discuss two Empirical Bayes Credibility Theory (EBCT) models. These models can be used to estimate the annual claim frequency or risk premium based on the values recorded over the last n years.

Model 1 gives equal weight to each risk in each year.

Model 2 is more sophisticated and takes into account the volume of business written under each risk in each year.

6.1 EBCT: Model 1

Model 1: Specification

The problem of interest is the estimation of the pure premium, or possibly the claim frequency, for a risk.

Let $X_1, X_2, \ldots$ denote the aggregate claims, or the number of claims, in successive periods for this risk. A more precise statement of the problem is that having observed the values of $X_1, X_2, \ldots, X_n$, the expected value of X_{n+1} needs to be estimated.

From now on $X_1, X_2, \ldots, X_n$ will be denoted by $\underline{X}$.

6.1 EBCT: Model 1

Model 1: Assumptions

The following assumptions will be made for EBCT Model 1.

Assumption 1: The distribution of each $\boldsymbol{X_j}$ depends on a parameter, denoted θ, whose value is fixed (and the same for all the $X_j s$) but is unknown.

Assumption 2: Given θ, the X_j 's are independent and identically distributed.
The parameter θ is known as the risk parameter. It could, as in Section 3 of the previous chapter, be a real number or it could be a more general quantity such as a set of real numbers.

A consequence of these two assumptions is that the random variables $\{\boldsymbol{X_j}\}$ are identically distributed.

An important point to note is that the $\boldsymbol{X_j}$'s are not (necessarily) unconditionally independent.

6.1 EBCT: Model 1

Model 1: The credibility premium

Next some notation is introduced. Define $m(\theta)$ and $s^2(\theta)$ as follows

$$m(\theta) = E\left(X_j \mid \theta\right)$$
$$s^2(\theta) = \text{var}\left(X_j \mid \theta\right)$$

The credibility premium for Risk i is

$$\boldsymbol{Z\bar{X} + (1 - Z)E[m(\theta)].}$$

where:

$$\bar{X} = \sum_{j=1}^{n} x_j/n \text{ and } Z = \frac{n}{n + E[s^2(\theta)]/\text{var}[m(\theta)]}$$

6.1 EBCT: Model 1

Model 1: Parameter estimation

Unbiased estimators for $E[m(\theta)], E[s^2(\theta)]$ and $\text{var}[m(\theta)]$ are given as follows:

Parameter	Estimator

These formulae are given on page 29 of the Formulae and Tables for Examinations of the Faculty of Actuaries and the Institute of Actuaries.

6.2 EBCT: Model 2

Model 2: Specification

EBCT Model 2 is a generalization of Model 1. Although some of the formulae for Model 2 look similar to those for Model 1, it is important to appreciate the differences between the models (in terms of their assumptions and results).

Definitions: Y_{ij} represents the number of claims (or aggregate claim amount) for risk $i(i = 1, \ldots, N)$ in year $j(j = 1, \ldots, n)$.

P_{ij} represents the corresponding risk volume (eg number of policies or premium income).

The P_{ij} 's are assumed to be known.

$$\boldsymbol{x_{ij} = Y_{ij}/P_{ij}}$$

6.2 EBCT: Model 2

Model 2: Assumptions

The assumptions that specify EBCT Model 2 are as follows.

Assumption 1: The distribution of each $\boldsymbol{X_j}$ depends on the value of a parameter, $\boldsymbol{\theta}$, whose value is the same for each j but is unknown.

Assumption 2: Given θ, the $\boldsymbol{X_j}$'s are independent (but not necessarily identically distributed).

Assumption 3: $E(X_j \mid \theta)$ does not depend on $\boldsymbol{j}$.

Assumption 4: $P_j \mathrm{var}(X_j \mid \theta)$ does not depend on j.

6.2 EBCT: Model 2

Model 2: The credibility premium

Having made Assumptions 9 and 10, $m(\theta)$ and $s^2(\theta)$ can be defined as follows:

$$m(\theta) = E\left(X_j \mid \theta\right)$$
$$s^2(\theta) = P_j \mathrm{var}\left(X_j \mid \theta\right)$$

The credibility premium for Risk i is:

$$\boldsymbol{Z\bar{X} + (1 - Z)E[m(\theta)]}$$

where:

$$\overline{\boldsymbol{X}} = \frac{\sum_{j=1}^{n} \boldsymbol{P}_j \boldsymbol{X}_j}{\sum_{j=1}^{n} \boldsymbol{P}_j} = \frac{\sum_{j=1}^{n} Y_j}{\sum_{j=1}^{n} P_j}$$

and:

$$Z = \frac{\sum_{j=1}^{n} P_j}{\sum_{j=1}^{n} P_j + \frac{E[s^2(\theta)]}{\mathrm{var}[m(\theta)]}}$$

6.2 EBCT: Model 2

Model 2: Parameter estimation

Unbiased estimators for $E[m(\theta)], E[s^2(\theta)]$ and $\text{var}[m(\theta)]$ are given as follows:

Parameter	Estimator

These formulae are given on page 30 of the Formulae and Tables for Examinations of the Faculty of Actuaries and the Institute of Actuaries.

Question

CT6 April 2015 Q5

An insurance company has for five years insured three different types of risk. The number of policies in the j^{th} year for the i^{th} type of risk is denoted by P_{ij} for $i = 1,2,3$ and $j = 1,2,3,4,5$. The average claim size per policy over all five years for the i^{th} type of risk is denoted by $\bar{X}_i$. The values of P_{ij} and $\bar{X}_i$ are tabulated below.

Number of policies						Mean claim size
Risk type i	Year 1	Year 2	Year 3	Year 4	Year 5	
1	17	23	21	29	35	850
2	42	51	60	55	37	720
3	43	31	62	98	107	900

The insurance company will be insuring 30 policies of type 1 next year and has calculated the aggregate expected claims to be 25,200 using the assumptions of Empirical Bayes Credibility Theory Model 2.

Calculate the expected annual claims next year for risks 2 and 3 assuming the number of policies will be 40 and 110 respectively.
[9]

Solution

$$\bar{P}_1 = 17 + 23 + 21 + 29 + 35 = 125$$

$$\bar{P}_2 = 42 + 51 + 60 + 55 + 37 = 245$$

$$\bar{P}_3 = 43 + 31 + 62 + 98 + 107 = 341$$

$$\bar{P} = 125 + 245 + 341 = 711$$

$$\bar{X} = \frac{(850 \times 125 + 720 \times 245 + 900 \times 341)}{711} = 829.18$$

expected claims per policy for risk 1 next year $= \frac{25,200}{30} = 840$

so $\quad 840 = Z_1 \times 850 + (1 - Z_1) \times 829.18$

$$Z_1 = \frac{840 - 829.18}{850 - 829.18} = 0.519594$$

so $\quad \frac{125}{125 + \frac{E(s^2(\theta))}{\text{Var}[m(\theta)]}} = 0.51969$

so $\quad \frac{E(s^2(\theta))}{\text{Var}[m(\theta)]} = \frac{125 - 0.51969 \times 125}{0.51969} = 115.57217$

so $\quad Z_2 = \frac{245}{245 + 115.528} = 0.6794756$

$$Z_3 = \frac{341}{341 + 115.528} = 0.74686$$

Solution

So credibility premium per policy are

$$\text{Type 2: } 0.67956 \times 720 + (1 - 0.67956) \times 829.18 = 755.0$$
$$\text{Type 3: } 0.74694 \times 900 + (1 - 0.74694) \times 829.18 = 882.1$$

so overall expected claims

$$\text{Type 2: } 754.98 \times 40 = 30{,}200$$
$$\text{Type 3: } 882.08 \times 110 = 97{,}028$$