ACTSC432 — Loss Models II

Classnotes for Spring 2019

by

Johnson Ng

BMath (Hons), Pure Mathematics major, Actuarial Science Minor

University of Waterloo

Table of Contents

List of Definitions

List of Theorems

List of Procedures

Preface

For this set of notes, I shall follow the format of which the course is presented, by breaking contents into modules instead of lectures. Also, I will be relying on the standard textbook for this topic, namely Klugman et al. 2012.

1 Introduction and Review of Probability

We shall first take an overview of what this course is about, and we will review on some of the relevant notions from earlier courses.

1.1 Introduction to Credibility Theory

Credibility Theory is a form of statistical inference that

- uses newly observed past events; to
- more accurately re-forecasts uncertain future events.

From Klugman et al. 2012,

> *It is a **set of quantitative tools** that allows an insurer to perform prospective **experience rating** (adjust future premiums based on past experience) on a risk or group of risks. If the experience of a policyholder is consistently better than that assumed in the underlying manual rate (also called a **pure premium**), then the policyholder may demand a rate reduction.*

That's all fancy mumbo-jumbo so let's go through an example that will hopefully enlighten us.

Example 1.1.1 (Enlightening Example to Credibility Theory)

Suppose automobile insurance policies are classified according to the following factors:

- number of drivers;

- gender of each driver;
- number of vehicles; and
- brand, model, production year, and approximate mileage driver per year.

Policies with identical characteristics are assumed to belong to the same **rating class**, which represents a group of individuals with similar risks.

Suppose there are 2 policies in the same rating class. Both policies are charged with a so-called **manual premium** of $\$1,500$ per year. This is the premium specified in the insurance manual for a policy with similar characteristics.

Let's say that after 3 years, we obtain the following data: We want

	Policy 1	Policy 2
Year 1	0	500
Year 2	200	4000
Year 3	0	2500

Table 1.1: Newly acquired past history for finding 'credibility'

to find out what's a good premium to charge to each policy for Year 4.

Remark 1.1.1

The shall leave the following as remarks.

- ***How is the policyholder's own experience account for?*** *This is a key question that will be addressed in this course.*
- *Risks in a given rating class are* ***not perfectly identical*** *(i.e., no rating system is perfect)*
- *One may refine the rating system by incorporating more factors but it is time-consuming (and no system is perfect).*

Thus, credibility theory is designed such that it

- accounts for heterogeneity within a given rating lass; and
- provides a theoretical justification to charge a premium that reflects to the policyholder's own experience.

1.2 Review of Probability

You are expected to be familiar with the following concepts:

Some examples or more detailed review will be added for each topic if I come to work through them in detail.

- Joint and Marginal Distribution
- Conditional Distribution
- Mixture Distributions (see also ACTSC431)
  - n-point Mixture
- Conditional Expectation

2 Review of Statistics

In this chapter, we will review the following notions:

- Unbiased estimation
- Maximum likelihood estimation
- Bayesian estimation ★

2.1 Unbiased Estimation

Suppose we are given a **parametric model** [1] of X, i.e. the distribution of $X \mid \Theta = \theta$ is known but θ is unknown. Furthermore, we have a **random sample** of X, i.e. we have $\{X_i\}_{i=1}^n$ is an independent and identically distributed (iid) sequence of random variables (rv) such that $X_i \sim X$.

[1] See ACTSC431.

Definition 1 (Estimate)

*An **estimate** is a specific value that is obtained when applying an estimation procedure to a set of numbers, and in our case, rvs. We usually denote an estimate by a hat $\hat{}$.*

Definition 2 (Estimator)

*An **estimator** is a rule or formula that produces an **estimate**. We usually denote an estimator by $\tilde{}$.*

Note 2.1.1

An estimate is a number or a function, while an estimator is an rv or a random function.

Remark 2.1.1

In this course, we will not make a difference between the estimator and the estimate, and will use only $\hat{}$.

Definition 3 (Biased and Unbiased Estimator)

We say that an estimator, $\hat{\theta}$*, is* ***unbiased*** *if*

$$E[\hat{\theta} \mid \theta] = \theta$$

for all θ*. We say that an estimator is* ***biased*** *if it is not unbiased, and we define the* ***bias*** *as*

$$\mathrm{bias}_{\hat{\theta}}(\theta) = E[\hat{\theta} \mid \theta] - \theta.$$

Let's have ourselves a silly example.

Example 2.1.1

Let $(X_1, \ldots, X_n)$ be a random sample of $\mathrm{Exp}(\beta)$. The **sample mean**

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

is an unbiased estimator for the mean β; observe that by the **linearity of the expectation**, we have

$$E[\overline{X}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n}(n\beta) = \beta.$$

Example 2.1.2

Let $\{X_i\}_{i=1}^{n}$ be a random sample of $X \sim \mathrm{Unif}(0, \theta)$. Let us construct two unbiased estimators for θ using

1. the sample mean $\overline{X}$; and
2. order statistics $X_{(n)} := \max_{1 \leq i \leq n}\{X_i\}$.

Solution

1. Observe that

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\cdot n\left(\frac{\theta}{2}\right) = \frac{\theta}{2}.$$

This tells us that if we picked $\hat{\theta} = 2\overline{X}$, then we would end up with

$$E[2\overline{X}] = \theta.$$

Thus $2\overline{X}$ is an unbiased estimator of θ.

2. Using the **Darth Vader rule** [2], since the X_i's form a random sample of X, and the bounds for each X_i is 0 and θ, we have that

$$\begin{aligned} E[X_{(n)}] &= \int_0^\infty \overline{F}_{X_{(n)}}(x)\,dx \\ &= \int_0^\infty (1 - P(\max\{X_1, X_2, \ldots, X_n\}) \leq x)\,dx \\ &= \int_0^\infty (1 - P(X_1 \leq x)P(X_2 \leq x)\ldots P(X_n \leq x))\,dx \\ &= \int_0^\theta \left(1 - (\frac{x}{\theta})^n\right)dx \\ &= \theta - \frac{1}{n+1}\left(\frac{x^{n+1}}{\theta^n}\right)\Big|_{x=0}^{x=\theta} = \frac{n}{n+1}\theta, \end{aligned}$$

where we note that we can change the bounds as such since $X \sim \text{Unif}(0,\theta)$ implies that

$$P(X \leq \theta) = \begin{cases} \frac{x}{\theta} & 0 \leq x \leq \theta \\ 1 & x > \theta \end{cases}.$$

Thus, to get an unbiased estimator for θ, we simply need to consider

$$\hat{\theta} = \frac{n+1}{n}X_{(n)},$$

which then

$$E\left[\frac{n+1}{n}X_{(n)}\right] = \theta.$$

[2] The **Darth Vader rule** is given as: if X is a **non-negative** rv, then

$$E[X] = \int_0^\infty \overline{F}_X(x)\,dx,$$

where $\overline{F}_X$ is the survival function of X.

Proposition 1 (Sample Mean as the Unbiased Estimator of the Mean)

Let $\{X_i\}_{i=1}^n$ *be a random sample of* X *which has mean* μ*. Then* $\overline{X}$ *is an unbiased estimator of* μ*.*

Proof

We have that

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}(n\mu) = \mu. \qquad \square$$

Definition 4 (Sample Variance)

Let $\{X_i\}_{i=1}^n$ *be a random sample of* X *which has mean* μ *and variance* σ^2*. We define the* ***sample variance*** *as*

$$\hat{\sigma}^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Proposition 2 (Sample Variance as the Unbiased Estimator of the Variance)

Let $\{X_i\}_{i=1}^n$ *be a random sample of* X *which has mean* μ *and variance* σ^2*. Then the sample variance* $\hat{\sigma}^2$ *is an unbiased estimator of* σ^2*.*

Proof

First, note that

$$\begin{aligned}
\operatorname{Var}(\overline{X}) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} n \operatorname{Var}(X_i) \\
&= \frac{1}{n}\sigma^2.
\end{aligned}$$

Thus

$$\begin{aligned} E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] &= E\left[\sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2\right] \\ &= \sum_{i=1}^{n} E\left[(X_i - \mu)^2\right] + \sum_{i=1}^{n} E\left[(\mu - \overline{X})^2\right] \\ &\quad + 2E\left[\sum_{i=1}^{n}(X_i - \mu)(\mu - \overline{X})\right] \\ &= n\sigma^2 + n\,\mathrm{Var}(\overline{X}) \;^{3} + 2nE[(\overline{X} - \mu)(\mu - \overline{X})] \;^{4} \\ &= n\sigma^2 - n\,\mathrm{Var}(\overline{X}) \\ &= n\sigma^2 - n\left(\frac{1}{n}\sigma^2\right) \\ &= (n-1)\sigma^2. \end{aligned}$$

[4] This relies on the fact that $\overline{X}$ is the unbiased estimator for μ (cf. Proposition 1). We then use the definition of the variance to achieve this.

[4] We used the fact that

$$\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n} X_i - n\mu = n\overline{X} - n\mu.$$

Also, note that

$$\mathrm{Var}(\overline{X}) = E[(\overline{X} - \mu)^2].$$

It follows that

$$E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \sigma^2. \qquad \square$$

Remark 2.1.2

In general, unbiasedness is ***not preserved*** *under parameter transformations. E.g.,* $\frac{1}{\overline{X}}$ *is generally not unbiased for* μ*, where* μ *is the mean of* $\overline{X}$*.*

Some unbiased estimators can also be unreasonable.

Example 2.1.3

Consider $X \sim \mathrm{Poi}(\lambda)$, where $\lambda > 0$. Note that

$$E[(-1)^X] = e^{\lambda(-1-1)} = e^{-2\lambda}$$

by the **probability generating function** method, and we see that $(-1)^X$ is an unbiased estimator of $e^{-2\lambda}$. However, we see that $(-1)^x$ only takes on values ± 1, which is nowhere close to $e^{-2\lambda}$.

Intuitively, $e^{-2\overline{X}}$ would be a "better" estimator despite the fact that it is biased.

Despite shortcomings like the above, unbiasedness is generally a good property for an estimator to have.

2.2 *Mean Squared Error*

Definition 5 (Mean Squared Error)

Suppose $\hat{\theta}$ is an estimator for the parameter θ. The ***mean squared error (MSE)*** *of $\hat{\theta}$ is defined as*

$$\mathrm{MSE}_{\hat{\theta}}(\theta) := E\left[(\hat{\theta} - \theta)^2\right] = \mathrm{Var}(\hat{\theta}) + \mathrm{bias}_{\hat{\theta}}(\theta)^2.$$

Proof

It is not immediately clear how the two expressions are the same. We shall prove it here. First, note that $\mathrm{bias}_{\hat{\theta}}(\theta) = E[\hat{\theta}] - \theta$ is a real value. Using a similar idea as in Proposition 2, we see that

$$\begin{aligned} E\left[\left(\hat{\theta} - \theta\right)^2\right] &= E\left[\left(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta\right)^2\right] \\ &= E\left[(\hat{\theta} - E[\hat{\theta}])^2\right] + E\left[\left(E[\hat{\theta}] - \theta\right)^2\right] \\ &\quad + 2E[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)] \\ &= \mathrm{Var}(\hat{\theta}) + \mathrm{bias}_{\hat{\theta}}(\theta)^2 \\ &\quad + 2\,\mathrm{bias}_{\hat{\theta}}(\theta)\, \cancelto{0}{E[\hat{\theta} - E[\hat{\theta}]]} \\ &= \mathrm{Var}(\hat{\theta}) + \mathrm{bias}_{\hat{\theta}}(\theta)^2. \end{aligned}$$

□

Note 2.2.1

The MSE is a measure to evaluate the ***quality of estimators****. The smaller the MSE, the better the estimator.*

2.3 Maximum Likelihood Estimation

Definition 6 (Likelihood Function)

*Let $\{X_i\}_{i=1}^n$ be a random sample of X with density $f(x;\underline{\theta})$, where $\underline{\theta}$ is possibly a vector of parameters. The **likelihood function** for $\underline{\theta}$ is defined as*

$$L(\underline{\theta}) = \prod_{i=1}^{n} f(X_i;\underline{\theta}).$$

Definition 7 (Maximum Likelihood Estimation)

*The **maximum likelihood estimation (MLE)** of $\hat{\underline{\theta}}$ of $\underline{\theta}$ is an approach that maximizes $L(\hat{\underline{\theta}})$.*

Note 2.3.1

*Heuristically, under the MLE, $\hat{\underline{\theta}}$ is the **most likely parameter** for the sample $(X_1, \ldots, X_n)$ to be realized.*

Sometimes, the likelihood function is difficult to work with. Fortunately, since $\ln x$ is a increasing bijective function that preserves monotonicity, we can make us of this property to ensure maximality.

Definition 8 (Log-likelihood Function)

*The **log-likelihood function** is defined as*

$$l(\underline{\theta}) = \sum_{i=1}^{n} \ln(f(X_i;\underline{\theta})).$$

Example 2.3.1

Let $\{X_i\}_{i=1}^n$ be a random sample for $\mathrm{N}(\mu, v)$. Find the MLE for μ, v.

Solution

First, we shall work on getting an MLE for μ. The likelihood function here is

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} f(X_i; \mu) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\ &\propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \right\}. \end{aligned}$$

Evaluating the derivative and equating it to 0 would be fruitless, since this is an exponentiation. Thus we appeal to the log-likelihood, which is

$$l(\mu) \propto \sum_{i=1}^{n} (X_i - \mu)^2.$$

The derivative log-likelihood is thus

$$\frac{dl}{d\mu} \propto -2 \sum_{i=1}^{n} (X_i - \mu).$$

Equating the above to 0, we get

$$\hat{\mu} = \overline{X}.$$

Now for an MLE of σ^2. For sanity, let us denote $\tau = \sigma^2$. Then the likelihood function, focusing on τ, is

$$\begin{aligned} L(\tau) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(X_i-\mu)^2}{2\tau}} \\ &\propto \tau^{-\frac{n}{2}} e^{-\frac{1}{2\tau} \sum_{i=1}^{n} (X_i-\mu)^2}. \end{aligned}$$

Again, the likelihood involves an exponentiation, so we appeal to the log-likelihood, which is

$$l(\tau) \propto -\frac{n}{2} \ln \tau - \frac{1}{2\tau} \sum_{i=1}^{n} (X_i - \mu)^2.$$

The derivative of the log-likelihood is

$$\frac{dl}{d\tau} = -\frac{n}{2\tau} + \frac{1}{2\tau^2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Equating the above to 0, we get

$$n = \frac{1}{\hat{\tau}}\sum_{i=1}^{n}(X_i - \hat{\mu})^2,$$

and so

$$\hat{\sigma}^2 = \hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

2.4 Bayesian Estimation

From Klugman et al. 2012,

> *The Bayesian approach assumes that only the data actually observed are relevant and it is the population distribution that is variable.*

Definition 9 (Prior Distribution)

The ***prior distribution*** *is a probability distribution over the space of possible parameter values. It is denoted $\pi(\theta)$ and represents our opinion concerning the relative chances that various values of θ are the true value of the parameter.*

Note 2.4.1

- *The parameter θ may be scalar or vector valued.*
- *Determining the prior distribution has always been one of the barriers to the widespread acceptance of the Bayesian methods, since it is almost certainly the case that your experience has provided you with some insight about possible parameter values before the first data point has been observed.*

We shall use the following concepts from multivariate statistics to obtain the following definitions.

Definition 10 (Joint Distribution)

Let $\{X_i\}_{i=1}^n$ be a random sample of the rv X, and Θ another rv that is independent of the X_i's [5]*, with pdf π. Let $\vec{X} = (X_1, X_2, \ldots, X_n)$. Then the **joint distribution** of $\vec{X}$ and Θ is defined as*

$$f_{\vec{X},\Theta}(\vec{x}, \theta) = f_{\vec{X}\mid\Theta}(\vec{x} \mid \theta)\pi(\theta).$$

[5] Note that Θ does not necessarily have a similar distribution to X.

Definition 11 (Marginal Distribution)

Let $\{X_i\}_{i=1}^n$ be a random sample of the rv X, and Θ another rv that is independent of the X_i's [6]*, with pdf π. Let $\vec{X} = (X_1, X_2, \ldots, X_n)$. Then the **marginal distribution** of $\vec{X}$ is defined as*

$$f_{\vec{X}}(\vec{x}) = \int f_{\vec{X}\mid\Theta}(\vec{x} \mid \theta)\pi(\theta)\, d\theta.$$

[6] Note that Θ does not necessarily have a similar distribution to X.

Once we have obtained data, we can look back at our prior distribution and "update" it to...

Definition 12 (Posterior Distribution)

Let $\{X_i\}_{i=1}^n$ be a random sample of the rv X, and Θ another rv that is independent of the X_i's [7]*, with pdf π. The **posterior distribution**, denoted by $\pi_{\Theta\mid\vec{X}}(\theta \mid \vec{x})$, is the conditional probability distribution of the parameters given the observed data.*

[7] Note that Θ does not necessarily have a similar distribution to X.

It is easy to find out what the general formula of the posterior distribution is. One simply needs to make use of Definition 10 and Definition 11. The proof of the following proposition is left as an easy brain exercise for the reader.

Exercise 2.4.1
Prove Proposition 3.

Proposition 3 (Formula for the Posterior Distribution)

With the assumptions in Definition 12, we have that the posterior distribution can be computed as

$$\begin{aligned}\pi_{\Theta|\vec{X}}(\theta \mid \vec{x}) &= \frac{f_{\vec{X},\Theta}(\vec{x},\theta)}{f_{\vec{X}}(\vec{x})} \\ &= \frac{\left(\prod_{i=1}^{n} f_{X_i|\Theta}(x_i \mid \theta)\right)\pi(\theta)}{\int_{\forall\theta}\left(\prod_{i=1}^{n} f_{X_i|\Theta}(x_i \mid \theta)\right)\pi(\theta)\,d\theta}.\end{aligned}$$

Definition 13 (Posterior Mean)

The ***posterior mean*** *is defined as the expected value of the posterior distribution.*

Definition 14 (Bayes Estimator)

The ***Bayes estimator*** *of Θ is the posterior mean of Θ, defined as*

$$\hat{\theta}_B := E[\Theta \mid \vec{X} = \vec{x}] = \int_{\forall\theta} \theta \cdot \pi_{\Theta|\vec{X}}(\theta \mid \vec{x}).$$

Note 2.4.2

It can be shown that $\hat{\theta}_B$ minimizes the mean square error

$$\min_{\hat{\theta}} E\left[\left(\hat{\theta} - \Theta\right)^2 \mid \vec{X} = \vec{x}\right].$$

2.4.1 Conjugate Prior Distributions and the Linear Exponential Family

Definition 15 (Conjugate Prior Distribution)

A prior distribution is said to be a ***conjugate prior distribution*** *for a given model if the resulting posterior distribution is from the same family as the prior, although possibly with different parameters.*

[8]

[8] More examples should be added here.

Example 2.4.1

The following are some important/prominent examples of conjugate prior distributions:

Table 2.1: Important/Prominent Conjugate Prior Distributions

$\pi(\theta)$	$f(x \mid \theta)$	$\pi(\theta \mid \vec{x})$
Gamma	Poisson	Gamma
Normal	Normal	Normal
Beta	Binomial	Beta
Beta	Geometric	Beta

Definition 16 (Linear Exponential Family)

An rv X is said to belong to the ***linear exponential family*** *if its pdf is of the form*

$$f(x, \theta) = \frac{p(x) e^{x r(\theta)}}{q(\theta)},$$

where $p(x)$ is some function of x, and $r(\theta), q(\theta)$ are some functions of θ, and the support of f does not depend on θ.

Example 2.4.2

Some members of the linear exponential family include

- $\mathrm{Exp}(\theta) : f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, where $p(x) = 1$, $r(\theta) = -\frac{1}{\theta}$ and $q(\theta) = \theta$.
- $\mathrm{Gam}(\alpha, \theta) : f(x, \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha - 1} e^{-\frac{x}{\theta}}$.
- $\mathrm{Poi}(\theta) : f(x, \theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{\frac{1}{x!} e^{x \ln \theta}}{e^\theta}$
- $\mathrm{N}(\theta, v) : f(x, \theta, v) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-\theta)^2}{2v}} = \frac{(2\pi v)^{-\frac{1}{2}} e^{-\frac{x^2}{2v}} e^{x \frac{\theta}{v}}}{e^{\theta^2 / 2v}}$

Note 2.4.3

Basically, functions the belong to a linear exponential family is a linear-like function with an exponent.

Theorem 4 (Conjugate Prior Distributions of Linear Exponential Distributions)

Suppose that given $\Theta = \theta$ the rvs $\vec{X}$ are iid with pf

$$f_{X_j \mid \Theta}(x_j \mid \theta) = \frac{p(x_j)e^{r(\theta)x_j}}{q(\theta)},$$

where Θ has the pdf

$$\pi(\theta) = \frac{[q(\theta)]^{-k}e^{\mu k r(\theta)}r'(\theta)}{c(\mu, k)},$$

where μ and k are parameters of the distribution and $c(\mu, k)$ is the ***normalizing constant*** [9]*. Then the posterior pf $\pi_{\Theta \mid \vec{X}}(\theta \mid \vec{x})$ is of the same form as $\pi(\theta)$, i.e. $\pi(\theta)$ is a conjugate prior distribution function.*

[9] The normalizing constant is used to reduce any probability function to a probability density function with a total probability of 1. (Source: Wikipedia)

Proof

Notice that the posterior distribution is

$$\begin{aligned}
\pi(\theta \mid \vec{x}) &= \frac{\left(\prod_{i=1}^n f_{X_i \mid \Theta}(x_i \mid \theta)\right)\pi(\theta)}{\int_{\forall \theta}\left(\prod_{i=1}^n f_{X_i \mid \Theta}(x_i \mid \theta)\right)\pi(\theta)\,d\theta} \\
&\propto \left(\prod_{i=1}^n f_{X_i \mid \Theta}(x_i \mid \theta)\pi(\theta)\right) \\
&= \left(\prod_{i=1}^n \frac{p(x_j)e^{r(\theta)x_j}}{q(\theta)}\right)\left(\frac{[q(\theta)]^{-k}e^{\mu k r(\theta)}r'(\theta)}{c(\mu, k)}\right) \\
&\propto q(\theta)^{-(n+k)}e^{\mu k + n\bar{x} r(\theta)}r'(\theta) \\
&= q(\theta)^{-k^*}e^{\mu^* k^* r(\theta)}r'(\theta),
\end{aligned}$$

where

$$k^* = k + n, \text{ and } \mu^* = \frac{\mu k + \sum x_j}{k + n} = \frac{k}{k+n}\mu + \frac{n}{k+n}\bar{x},$$

and we see that the posterior distribution has the same form as $\pi(\theta)$. $\square$

Example 2.4.3

One non-example is mentioned in Example 2.4.1: the distribution of X_i is not from the linear exponential family, but we still obtain that the posterior distribution has a similar distribution to the posterior distribution.

Bibliography

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons Inc., Hoboken, New Jersey, 4th edition.

Index