工業經營學會誌 第 22卷 第 51輯 1999年 8月 53

⊠ 연구논문

베이지안 기법을 이용한 신뢰도 예측 시 관측치에 주어지는 가중치 분석에 관한 연구

- Analysis of Weights Given to Observations in the Bayesian Reliability Prediction -

양 희 중*
Yang, Hee Joong

요지

평균치에 적용되는 credibility formula를 분산에도 적용하여 응용 할 수 있는 extended credibility formula 를 개발한다. 간단한 베이지안 신뢰도 예측모형을 구축하고 이 모형에 extended credibility formula를 적용한다. 감마 사전분포 - 포아송 우도의 경우와 베타 사전분포 - 이항분포 우도 의 경우에 대해 extended credibility formula를 적용해 관측치에 주어진 가중치에 따라 사후 분산이 어떻게 변화하는지를 분석한다. 사후분산도 사후평균과 마찬가지로 사전값과 관측값의 가중평균으로 표시될 수 있다는 것을 증명한다. 가중치와 불확실성 감소율간의 관계도 연구된다. 이와 같은 가중치에 따른 사전 및 사후분포의 변화 양식에 대한 이해는 올바른 사전분포를 설정하는데 큰 도움이 될 수 있다.

1. Introduction

In the process of Bayesian prediction, the assessment of credible prior distribution is a difficult and important task since we may end up with fairly different prediction depending on the assessed prior distribution. The prior distribution is assessed based on historical data incorporating expert opinions and engineering knowledge. Lack of such information forces us to assign a large variance on the prior and consequently much weight on later observations. On the other hand, enough credible information allows us to assess a sharpened prior distribution that will not pay much attention to observations, and accordingly we can expect relatively smaller amount of change on the posterior mean by observations. We may temporarily have undesirable feature that the posterior variance increases by observations until enough data is obtained in the situation where the assessed prior variance is unsuitably small and the assessed prior expectation is far from the unobservable true mean. Such phenomena is actually due to the wrong assessment of

* 청주대학교 산업공학과

prior distribution on the part of mean as well as variances, that enables us to analyze how the weights given to observations influence the shape of posterior distributions. The knowledge about the pattern that the posterior variance behaves depending on the weight given to observations will now help us assess the proper prior distributions. We construct two simple reliability prediction models : one in which a failure rate influences the number of failures in a given time period, another in which a failure can escalate to more severe one. Based on such models, we will focus our analysis on the cases of Gamma priors-Poisson likelihood, and Beta priors-Binomial likelihood, which are frequently used in the fields of reliability prediction and forecasts of escalating failures([5],[6]).

2. Gamma Prior and Poisson Likelihood

Consider an influence diagram in figure 1, where λ denotes a failure rate and $n = n(T)$ denotes the number of failures in time period (0,T). We omit the explanation of influence diagrams which is not the scope of this work. For details of influence diagrams, readers may refer to references [3], [4]. Bayesian approach in predicting the time to next failure can be proceeded without any distributional assumptions on λ. The theory of arc reversals and node absorptions in influence diagrams can be helpful in obtaining posterior and predictive distributions. When we treat real data we firstly have to assess a prior distribution on the failure rate and likelihood on the number of failures conditional on the failure rate.

Figure 1. A Simple Reliability Prediction Model

Generally reliability systems are composed of some parallel systems or back-up systems, thus the distribution of failure rate in such systems has a peak in a low probability region and still has long tails to the high probability region. Gamma prior is a good candidate in expressing such shape. With assumption of independent failure occurrence we can assume Poisson likelihood.

We adopt the convention that if a random variable λ follows a Gamma distribution with parameters α and β, the probability density function is expressed as

$$p(\lambda) = \Gamma(\alpha, \beta) = \frac{\beta(\beta\lambda)^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}$$

The prior mean and variance of λ are obtained by

$$E[\lambda]=\frac{\alpha}{\beta} \qquad Var[\lambda]=\frac{\alpha}{\beta^2} \tag{1}$$

The posterior distribution of λ is also Gamma with parameters $\alpha'=\alpha+n$, $\beta'=\beta+T$, and the posterior mean and variance are obtained by

$$E[\lambda \mid D]=\frac{\alpha'}{\beta'}=\frac{\alpha+n}{\beta+T} \tag{2}$$

$$Var[\lambda \mid D]=\frac{\alpha'}{(\beta')^2}=\frac{\alpha+n}{(\beta+T)^2} \tag{3}$$

where D represents observed data.

Equation (2) can be reexpressed as (4).

$$E[\lambda \mid D]=\frac{\beta}{\beta+T}\frac{\alpha}{\beta}+\frac{T}{\beta+T}\frac{n}{T}=(1-w)E[\lambda]+w\mu \tag{4}$$

$E[\lambda]$ can be interpreted as the mean of an extreme case where the information contained in observations is completely ignored, while $\mu=n/T$ can be interpreted as the mean of another extreme case where the prior information is completely ignored, and we call μ as the observed mean. Equation (4), which is so called credibility formula, says that posterior mean can be expressed as a weighted average of prior and observed mean. Now similar analysis can be applied to cases of variances. Equation (3) can be reexpressed as equation (5).

$$Var[\lambda \mid D]=\frac{\alpha+n}{(\beta+T)^2}=\frac{\beta^2}{(\beta+T)^2}\frac{\alpha}{\beta^2}+\frac{T^2}{(\beta+T)^2}\frac{n}{T^2}=(1-w)^2\,Var[\lambda]+w^2\mu^2 \tag{5}$$

where w = T/($\beta+T$) denotes the weight given to the observations.

Var[λ] can be explained as the variance of the extreme cases where the information contained in the observation is completely ignored. On the other hand $\sigma^2=n/T^2$ can be interpreted as the variance of another extreme case where prior parameters α and β are set to zero. We call σ^2 as the observed variance or null variance. The weight assigned to the observation is determined by β and observation period T, not depending on the number of events n, Since β can be expressed as $E[\lambda]/Var[\lambda]$, small β represents a large variance on prior for the same prior mean. Thus large uncertainty on prior is reflected by a small β and gives large weight on observations. And as T becomes larger by observing data in a longer time period, the weight becomes larger and larger and eventually approaches to 1. When the weight is 1 the posterior mean is just the observed mean so that the prior information is completely ignored. Similarly, when the weight is small the posterior variance is mainly coming from the prior variance, and information contained in observations does not affect the prior uncertainty much. On the other hand, when the weight is large, most of the posterior variance is composed of the null variance, thus the amount of prior uncertainty does not affect the posterior variance much.

To study how much change on the prior mean and uncertainty reduction can be achieved by observations, let us define $d=(n/T)-E[\lambda]$ to represent a difference between observed and prior mean. Then the difference between prior and posterior mean, dE, and the amount of uncertainty reduction, dVar, can be obtained by

$$dE=E[\lambda \mid D]-E[\lambda]=wd \tag{6}$$

$$dVar=Var[\lambda]-Var[\lambda \mid D]=w(Var[\lambda]-d/(\beta+T)) \tag{7}$$

Therefore the amount of deviation of the observed mean from the assessed prior mean is reflected in the posterior mean only by the portion of the weight. If we assess a prior distribution with large variance, in other words if we tend to assign a large weight to observations, which is usual when to make a forecast without enough information, the prior prediction may be changed a lot by a single observation. On the other hand, if available information allows us to assess prior distribution with small variance or assign a small weight to observation, an observation whose mean value somewhat deviates from the prior mean will not modify the prior distribution much, thus the posterior prediction will remain similar to the prior prediction. From equation (7) it can be seen that we may have undesirable feature that the posterior variance increases when the deviation d is greater than $Var[\lambda](\beta+T)$.

As we acquire more data and thereby we assess more credible distribution, we can expect the observed mean will fall close to the prior mean. In such a steady state if the difference between observed and prior mean is negligible, the variance reduces by $wVar[\lambda]$. And the ratio of uncertainty reduction can be expressed as

$$\text{ratio of uncertainty reduction} = \frac{Var[\lambda]-Var[\lambda \mid D]}{Var[\lambda]}=w$$

The above ratio also represents the maximum ratio of uncertainty reduction that can be achieved by observations. Therefore the weight plays an important role in determining the posterior distribution

To see the increasing pattern of weights over observation period depending on the assessed variance, consider figure 2.

Figure 2 is drawn for three cases that the prior variances are 0.003, 0.03, and 0.3 with same prior mean of 0.385. It can be seen that only 20 time units increases the weight to almost 1 when the prior variance is66 large, and even 100 time units increase the weight to less than 0.5 when the prior variance is small.

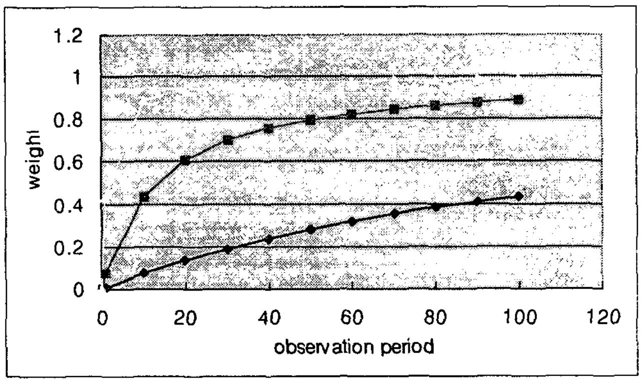

Figure 2. Weights over Observation Periods

More close look at reveals the fact that if k times smaller variance is assigned then k times longer period of observation is required to reach to the same weight. This can be proved easily. Let (α_1, β_1) and (α_2, β_2) be two different prior parameters of a Gamma distribution where

$$\frac{\alpha_1}{\beta_1} = \frac{\alpha_2}{\beta_2}, \qquad \frac{\alpha_1}{(\beta_1)^2} = k\frac{\alpha_2}{(\beta_2)^2}$$

so that the assessed prior mean is same but k times smaller variance is assessed to the second case. Then the weight to the first case when the observation period is T_1 is

$$w_1 = \frac{T_1}{\beta_1 + T_1}$$

and the weight to the second case when the observation period is T_2 is

$$w_2 = \frac{T_2}{\beta_2 + T_2} = \frac{T_2}{k\beta_1 + T_2} = \frac{\frac{1}{k}T_2}{\beta_1 + \frac{1}{k}T_2}$$

Thus if T_2/k is equal to T_1, w_2 becomes same as w_1. Therefore if one assigns k times smaller variance it can be interpreted as one is willing to wait for k times longer period until one puts same amount of trust on the observations.

Understanding such behavior of posterior mean and variance that depends on weights is very much helpful in assessing the prior distributions. It may be obscure to assign specific value for prior parameters, but we may have more confident feeling and set of ideas that is useful to assign a weight to observations.

3. Beta Priors and Binomial Likelihood

We extend a model in figure 1 to explain the failure escalation process. In figure 3, ψ denotes the escalation probability. Thus $\lambda\psi$ denotes the rate of severe failures. On the part of λ same approach in section 2 can be applied. For ψ, we assume Beta distribution with parameters a, and b, since Beta distributions are quite flexible covering almost all forms of distributions in region between 0, and 1.

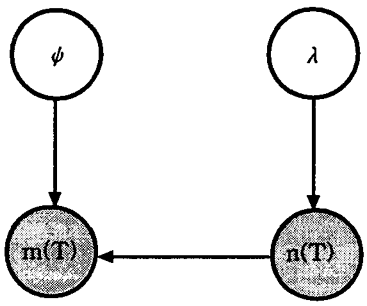

Figure 3. Failure Level Escalating Model

Since most of the failures escalate to severe ones independently, we assume Binomial likelihood for m(T) given n(T) and ψ.
In a Beta prior and Binomial likelihood model where ψ denote the failure probability following a Beta distribution with parameters a, b, assume that we have oserved m failures out of n independent trials. The prior mean and variance of ψ are obtained by

$$E[\varphi] = \frac{a}{a+b} \qquad Var[\varphi] = \frac{ab}{(a+b)^2(a+b+1)} \tag{8}$$

The posterior distribution of φ is also Beta with parameters a', b' and the posterior mean and variance are obtained by

$$E[\varphi|D] = \frac{a'}{b'} = \frac{a+m}{a+b+n} \tag{9}$$

$$Var[\varphi|D] = \frac{a'}{(a'+b')^2(a'+b'+1)} = \frac{a+m}{(a+b+n)^2(a+b+n+1)} \tag{10}$$

Similarly to the ones in the above section, equation (9) and (10) can be reexpressed as (11) and (12).

$$E[\psi|D] = \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{m}{n} = (1-w)E[\psi] + w\mu \tag{11}$$

$$Var[\psi|D]=(1-w)^2\frac{a+b+1}{a+b+n+1}\left(\frac{1+n-m}{b}+\frac{m}{a}\right)Var[\psi]+w^2\frac{n+1}{a+b+n+1}\frac{m(n-m)}{n^2(n+1)}$$

$$\approx (1-w)^3\left(1+\frac{n-m}{b}+\frac{m}{a}\right)Var[\varphi]+w^3\sigma^2 \qquad (12)$$

where w=n/(a+b+n) denotes the weight given to the observation.

The weight assigned to the observation is determined by a+b and total number of trials n, not depending on the number of failures m. Small a+b represents a large variance on prior for the same prior mean. Thus large uncertainty on prior is reflected by a small a+b and gives large weight on observations. Similar explanation on equations (4) and (5) can be applied to equations (11) and (12).

Equation (6) and (7) can be rewritten as followings for the case of Beta prior and Binomial likelihood.

$$dE = E[\psi|D] - E[\psi] = wd \qquad (13)$$

$$dVar=Var[\psi]-Var[\psi|D]\approx w\left(Var[\psi]-\frac{b+n-m}{(a+b+n)^2}d\right) \qquad (14)$$

As the case of Gamma prior and Poisson likelihood the deviation of the observed mean is taken into account by the portion of the weight when to determine the posterior mean. And when the difference between observed and prior mean is negligible, the variance reduces by wVar[ψ]. Thus the ratio of uncertainty reduction is

$$\text{ratio of uncertainty reduction}=\frac{Var[\psi]-Var[\psi|D]}{Var[\psi]}=w$$

The last argument in previous section is still valid for Beta prior and Binomial likelihood. Let(a_1,b_1) and ((a_2,b_2) be two different prior parameters of a Beta distribution where

$$\frac{a_1}{a_1+b_1}=\frac{a_1}{a_2+b_2},\ \frac{a_1b_1}{(a_1+b_1)^2(a_1+b_1+1)}=k\frac{a_2b_2}{(a_2+b_2)^2(a_2+b_2+1)}$$

Then

$$w_1=\frac{n_1}{a_1+b_1+n_1}$$

$$w_2=\frac{n_2}{a_2+b_2+n_2}\approx\frac{n_2}{ka_1+kb_1+n_2}=\frac{\frac{1}{k}n_2}{a_1+b_1+\frac{1}{k}n_2}$$

thus if n_2/k is equal to n_1, w_2 is approximately same as w_1. Therefore if one assigns k times smaller variance, it can be interpreted as one is willing to perform k times more sampling until one puts same amount of trust on the observations.

4. Summary and Conclusions

The assessment of proper distribution is crucial in obtaining a credible prediction. The shape(wide vs. narrow) and location(deviation from the true mean) of prior distribution have combined effects when to respond to observations and determine the posterior distribution. The relationship between prior and posterior distribution can be explained by so called "credibility formula", which says that the posterior mean is a weighted average of prior mean and observed mean. Credible prior information enables us to asses a sharpened prior distribution (small weight on observations) that will not pay much attention to observation, and accordingly we can expect relatively smaller amount of change on the posterior mean by observation. Such a role of weights in determining the posterior mean can be expressed as a function of weights.

The ratio of uncertainty reduction is also expressed in terms of weights. We also find that if one assigns k times smaller prior variance it reflects that one is willing to wait for k times longer period until one puts same amount of trust on the observations. Such a full knowledge about how weights work in the process of getting posterior distributions make it possible for us to asses more adequate prior distributions.

References

[1] Aitchison, J., and Dunsmore, I.R. [1975], Statistical Prediction Analysis, Cambridge University Press.

[2] DeGroot, M.H. [1970], Optimal Statistical Decision, McGtaw-Hill Book Company.

[3] Oliver, R.m., and Yang, Heejoong [1990], "Updating Event Tree Parameters to Predict High Risk Incidents", Influence Diagrams, Belief Nets and Decision Analysis, edited by Oliber and Smith, John wiley & Sons, p277-296.

[4] Shachter, Ross D., [1987], "Evaluating Influence Diagrams" Operations Research v. 34, p871-882.

[5] Yang, Heejong[1990], "An Approximation Method in Bayesian Prediction of Nuclear Power Plant Accidents", Journal of the Korean Institute of Industrial Engineers, Vol. 16, No.2, p135-147.

[6] Yang, Heejoong[1993],"predicting Nuclear Power Plant Accidents in Korea", IE Interfaces, Vol. 6, No. 2, p79-90.