Variability within multi-component systems

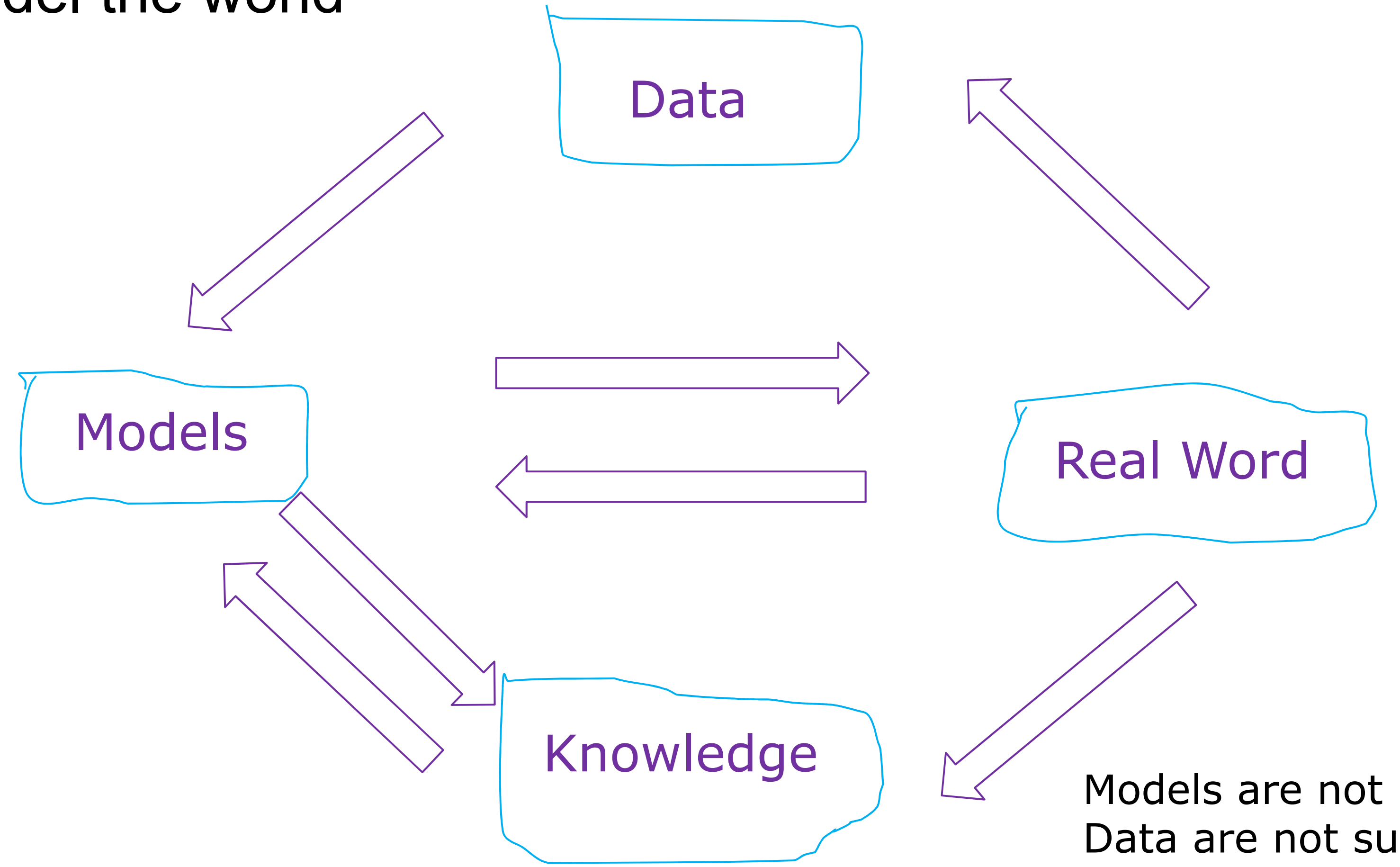
Bayesian inference in probabilistic risk assessment
The current state of the art

Shuoyun Zhang

ETH Zurich

Introduction

Model the world



Models are not precise
Data are not sufficient
Natural variability
Knowledge is subjective

Introduction

- Uncertain and parametric model

- Bayesian theorem as a inference tool

Data

Prior Knowledge

New Knowledge

$$\pi_1(\theta \mid x) = \frac{f(x \mid \theta)\,\pi_0(\theta)}{\int_\Theta f(x \mid \theta)\,\pi_0(\theta)\,d\theta}$$

Introduction

- MCMC and the software package

  MCMC: Markov Chain Monte Carlo

  Software: Winbugs Openbugs

- Bayesian inference is applied broadly

  e.g. behavioral science, finance, human health, process control and ecological

Content

- Hierarchical modeling of variability (3)
- Modeling of time-dependent reliability (4)
- Modeling of random duration (5)
- Treatment of uncertain and missing data (6)
- Bayesian Regression models (7)
- Model selection and validation (8)
- Summary and conclusion (9,10)

Hierarchical modeling of variability

- Hierarchical Bayesian modeling

$$\pi(\theta) = \int_{\Phi} \pi_1(\theta \mid \varphi)\pi_2(\varphi)\,d\varphi$$

First-stage prior

hyperprior

$\pi_1(\theta \mid \varphi)$ is the population variability in θ for a given value φ

$\pi_2(\varphi)$ is the distribution representing the uncertainty in φ

Hierarchical modeling of variability

- Illustrative example: Poisson process

Possion process: $$f\left(x\,|\,\lambda\right)=\frac{\left(\lambda t\right)^{x}e^{-\lambda t}}{x!},x=0,1$$

Source	Failures	Exposure time (h)
1	0	87,600
2	7	525,600
3	1	394,200
4	0	87,600
5	8	4,555,200
6	0	306,600
7	0	394,200
8	0	569,400
9	5	1,664,400
10	1	3,766,800
11	4	3,241,200
12	2	1,051,200

We have a prior belief that there is a source-to-source variability in λ

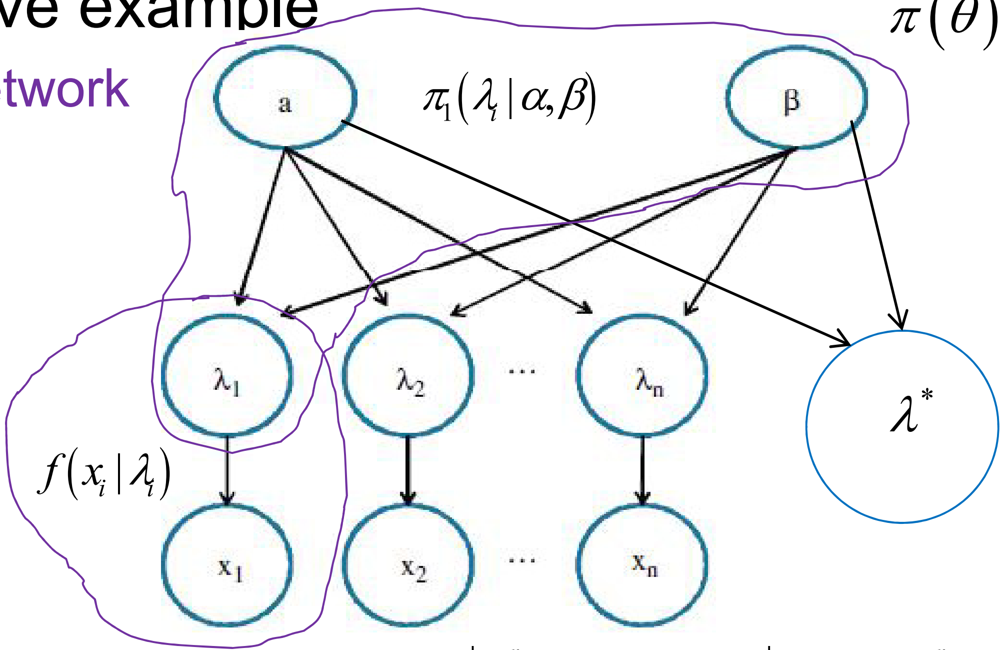
Assuming $$\pi_1\left(\lambda\,|\,\varphi\right)=\pi_1\left(\lambda\,|\,\alpha,\beta\right)\sim Gamma\left(\alpha,\beta\right)$$

Hierarchical modeling of variability

- Illustrative example

$$\pi(\theta)=\int_{\Phi}\pi_1(\theta\,|\,\varphi)\pi_2(\varphi)\,d\varphi$$

Bayesian Network



$$\pi_2\left(\alpha,\beta\,|\,\tilde{x},\tilde{t}\right)\propto f\left(\alpha,\beta,\tilde{x},\tilde{t}\right)=\int_0^\infty\int_0^\infty\cdots\int_0^\infty\left[\prod_{i=1}^{n}\pi_1\left(\lambda_i\,|\,\alpha,\beta\right)\right]\pi\left(\alpha,\beta\right)\prod_{i=1}^{n}f\left(\tilde{x}_i\,|\,\lambda_i\right)d\lambda_1\cdots d\lambda_n$$

$$\pi\left(\lambda_i\,|\,\tilde{x},\tilde{t}\right)=\int_0^\infty\int_0^\infty\cdots\int_0^\infty\left\{\iint\left[\prod_{i=1}^{n}\pi_1\left(\lambda_i\,|\,\alpha,\beta\right)\right]\pi_1\left(\lambda^*\,|\,\alpha,\beta\right)\pi_2\left(\alpha,\beta\,|\,\tilde{x},\tilde{t}\right)d\alpha\,d\beta\right\}$$

$$\times d\lambda_1 d\lambda_2\cdots d\lambda_{i-1}d\lambda_{i+1}\cdots d\lambda_n d\lambda^*$$

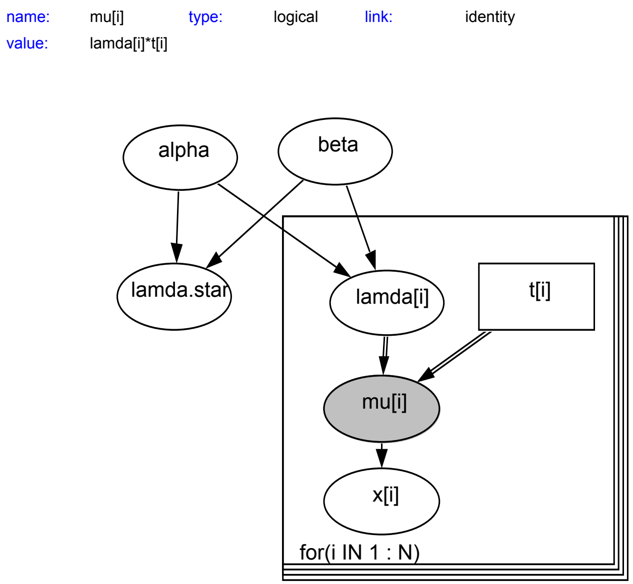
$$=\iint\pi_1\left(\lambda_i\,|\,\tilde{x},\tilde{t},\alpha,\beta\right)\pi_2\left(\alpha,\beta\,|\,\tilde{x},\tilde{t}\right)d\alpha\,d\beta$$

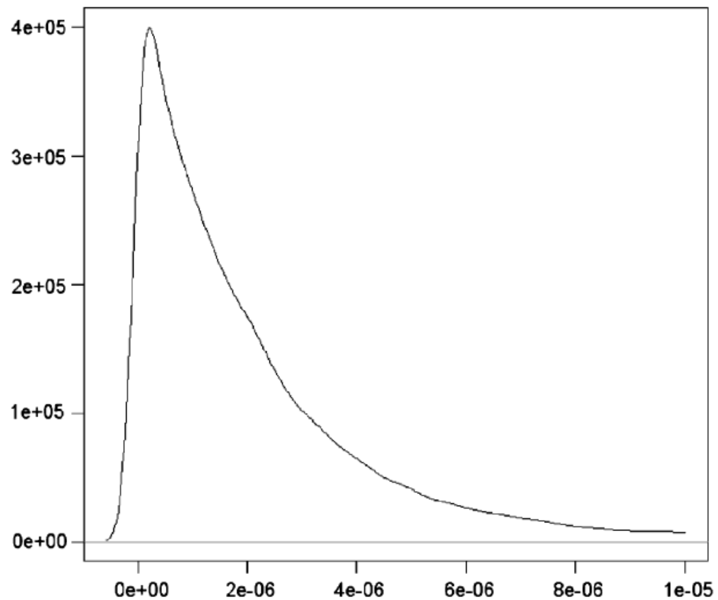
$$\pi\left(\lambda^*\,|\,\tilde{x},\tilde{t}\right)=\iint\pi_1\left(\lambda^*\,|\,\alpha,\beta,\tilde{x},\tilde{t}\right)\pi_2\left(\alpha,\beta\,|\,\tilde{x},\tilde{t}\right)d\alpha\,d\beta$$

Hierarchical modeling of variability

➢ Illustrative example

Winbugs Doodle model





$\pi\left(\lambda^* \mid \tilde{x}, \tilde{t}\right)$

mean 2.5×10^{-6}

90% credible interval $\left(3.9\times10^{-8}, 8.0\times10^{-6}\right)$

Hierarchical modeling of variability

- Illustrative example: Emergency Diesel Generators (EDG)

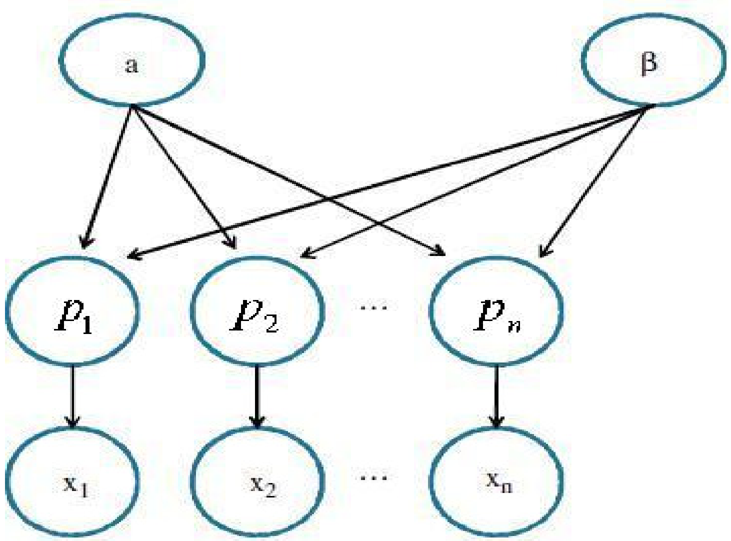
EDG	Failures	Successes	Demands
1	0	140	140
2	0	130	130
3	0	130	130
4	1	129	130
5	2	98	100
6	3	182	185
7	3	172	175
8	4	163	167
9	5	146	151
10	10	140	150
Total	28	1430	1458

The failure probability p_1 of No.1 disel is of interest. (binomial distribution)

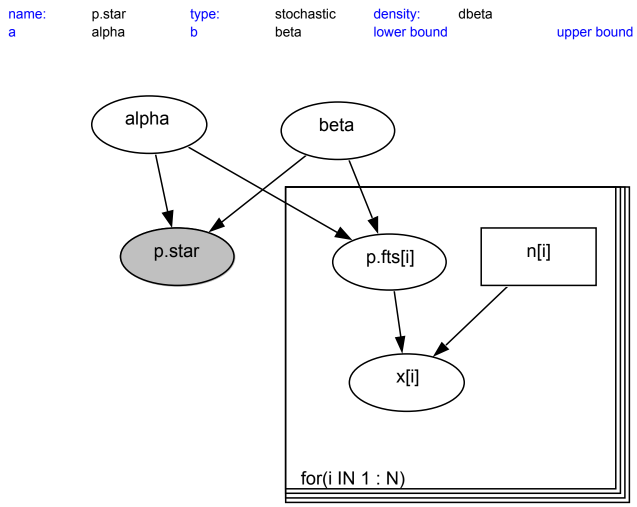
$x \sim Bin(p,n) \quad \pi_1(p \mid \alpha,\beta) = Beta(\alpha,\beta)$

Hierarchical modeling of variability

- Illustrative example: Emergency Diesel Generators



Bayesian Network



Winbugs Doodle model

Result:

	5th	50th	95th	Mean
Empirical Bayes	4.7E−04	4.4E−03	1.7E−02	5.9E−03
Two-stage Bayes	1.2E−04	3.3E−03	1.8E−02	5.2E−03
Hierarchical Bayes	5.9E−05	4.5E−03	1.9E−02	6.3E−03

Content

- Hierarchical modeling of variability
- Modeling of time-dependent reliability (4)
- Modeling of random duration
- Treatment of uncertain and missing data
- Bayesian Regression models
- Model selection and validation
- Summary and conclusion

Modeling of time-dependent reliability

- Modeling time trends

- Modeling failure with repair

Modeling time trends

- **Motivation**

  It is sometimes the case that the usual Poisson and binomial models are rendered invalid because the parameter of interest (λ or p, respectively) is not constant over time.

- **Principal idea**

  Generalized Linear Model (GLM)

  for λ, use $\log(\lambda) = a + bt$

  for p, use Logit function $\log\left(\dfrac{p}{1-p}\right) = a + bt$

  if $b = 0$, there is no time trend

  if $b > 0$, p or λ is increasing over time.

Modeling time trends

- Example: valve leakage

  if there appears to be any systematic time trend in p

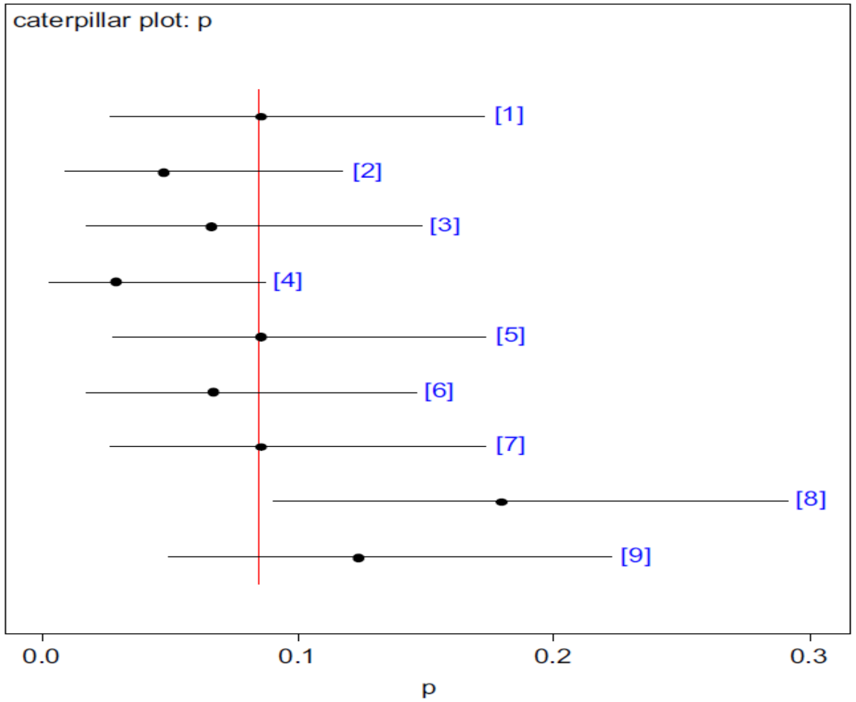
Valve leakage data

Year	Number of failures	Demands
1	4	52
2	2	52
3	3	52
4	1	52
5	4	52
6	3	52
7	4	52
8	9	52
9	6	52

Modeling time trends

➢ Example: valve leakage

To check the clouds in our mind, we don't use any GLM model firstly, instead we update the Jeffreys prior with the data for each year, the result is:



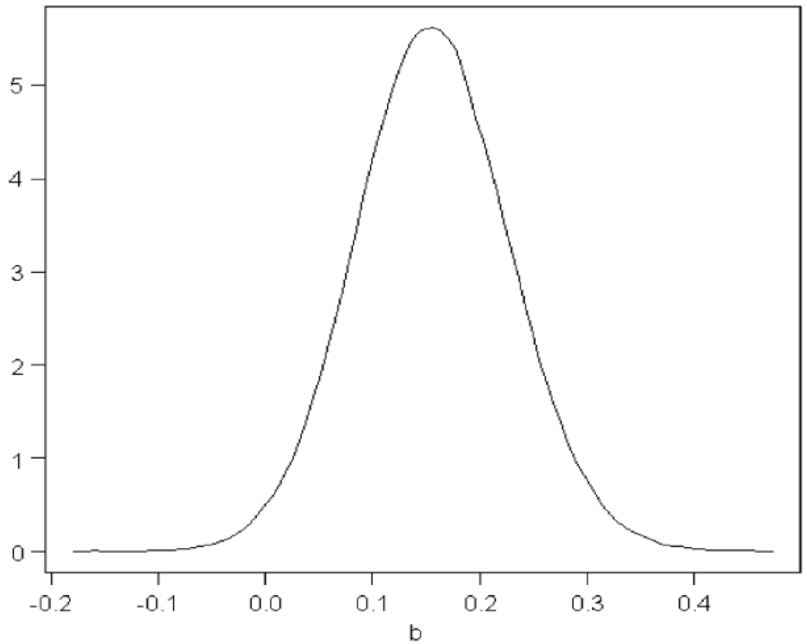
It looks likes that there is an increasing trend with time, but significant uncertainty in the individual estimates clouds this conclusion.

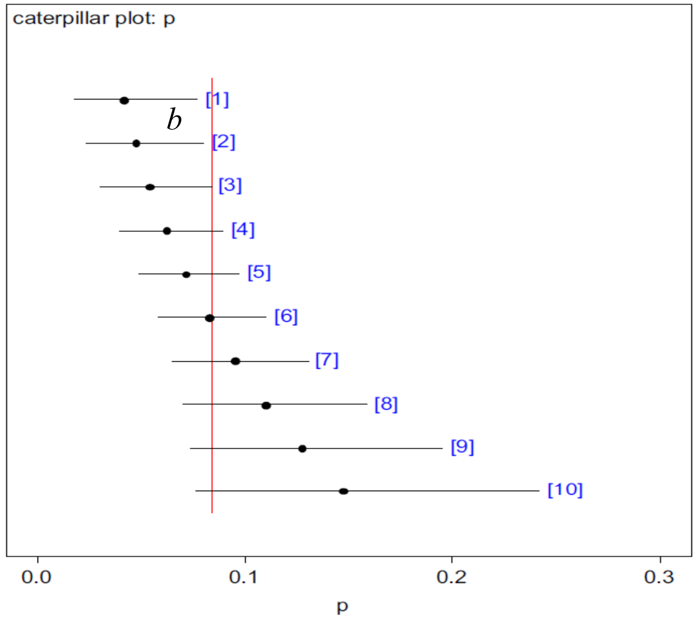
Modeling time trends

Example: valve leakage

Take the GLM to check, logit function

Restult:





Density is highest for $b > 0$, suggesting an increasing trend in p

Modeling of time-dependent reliability

- Modeling time trends

- Modeling failure with repair

Modeling failure with repair

- Motivation

- Most of repair model is focus on "same as new" assumption, under the assumption of that the stochastic point process being observed is a renewal process. Less work has addressed the more reasonable assumption that repairs make the component "same as old".

- Under the "same as new" assumption, the times between failures are independently, and identically distributed (iid). Under the "same as old" assumption for repair, the inter-arrival times are not iid; the distribution for the ith time is dependent upon the i-1th time.

Modeling failure with repair

- Principal idea

for Poisson process:

$$f(x) = \frac{\left(\int_0^t \lambda(t)\,dt\right)^x}{x!} \exp\left(-\int_0^t \lambda(t)\,dt\right)$$

If $\lambda(t)$ is constant, it is the homogenous Poisson process (HPP) ;
if $\lambda(t)$ is dependent on time, it is the nonhomogenous Poisson process (NHPP) ;

if $\lambda(t)$ is increasing with time, the times between failures are decreasing with time; the component is aging or wearing out.
If $\lambda(t)$ is decreasing with time, the times between failures are increasing with time, the component is experiencing reliability growth.

Modeling failure with repair

- Principal idea

Common form to model $\lambda(t)$ include power-law process:

$$\lambda(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}$$

the loglinear model

$$\lambda(t) = \exp(a + bt)$$

and the linear model

$$\lambda(t) = a + bt$$

Content

- Hierarchical modeling of variability
- Modeling of time-dependent reliability
- Modeling of random duration (5)
- Treatment of uncertain and missing data
- Bayesian Regression models
- Model selection and validation
- Summary and conclusion

Modeling of random duration

➢ Motivation

- The simplest stochastic model for applications where time is the random variable of interest is the exponential distribution, which means that time-independent rate.
- There are numerous applications for time-dependent case, e.g. the rate of recovering offsite ac power at a commercial nuclear plant is often a decreasing function of time after power is lost, analyst is led to models that allow for time-dependent recovery rates, e.g. Weibull or lognormal distribution.
- Bayesian inference is more complicated when the likelihood function is other than exponential. The difficulty of Bayesian approach has led analyst to use frequentist methods, such as MLE. The advent of Winbugs allows a fully Bayesian approach to the problem to be implemented.

Modeling of random duration

- Illustrative Example: Recovery of offsite ac power

Times to recover offsite ac power for grid-related disturbances

Site	Date	Potential recovery time (min)
Davis-Besse	14/8/2003	657
Fermi	14/8/2003	384
Fitzpatrick/nine mile point 1	14/8/2003	142
Ginna	14/8/2003	54
Indian point	16/6/1997	42
Indian point	14/8/2003	102
Nine mile point 2	14/8/2003	110
Palo verde	14/6/2004	37
Peach bottom	15/9/2003	16
Perry	14/8/2003	87
Summer	11/7/1989	100
Vermont yankee	17/8/1987	17

Use Weibull and lognormal distribution to model the rocovery time

Modeling of random duration

- Illustrative Example: Recovery of offsite ac power

Times to recover offsite ac power for grid-related disturbances

Site	Date	Potential recovery time (min)
Davis-Besse	14/8/2003	657
Fermi	14/8/2003	384
Fitzpatrick/nine mile point 1	14/8/2003	142
Ginna	14/8/2003	54
Indian point	16/6/1997	42
Indian point	14/8/2003	102
Nine mile point 2	14/8/2003	110
Palo verde	14/6/2004	37
Peach bottom	15/9/2003	16
Perry	14/8/2003	87
Summer	11/7/1989	100
Vermont yankee	17/8/1987	17

Use Weibull and lognormal distribution to model the model the rocovery time.

We are interested in parameters:

Weibull distribution: α (shape) , β (scale)

Lognormal distribution: μ, σ

Modeling of random duration

- Illustrative Example: Recovery of offsite ac power

MLE result:

Weibull: α (shape) = 0.929, β (scale) = 2.332 h
Lognormal: $\mu = 0.300$, $\sigma = 1.064$

Bayesian result:

for Weibull

Parameter	Posterior mean	90% credible interval
α	0.914	(0.812, 1.254)
β	2.58	(1.343, 4.299)

for Lognormal

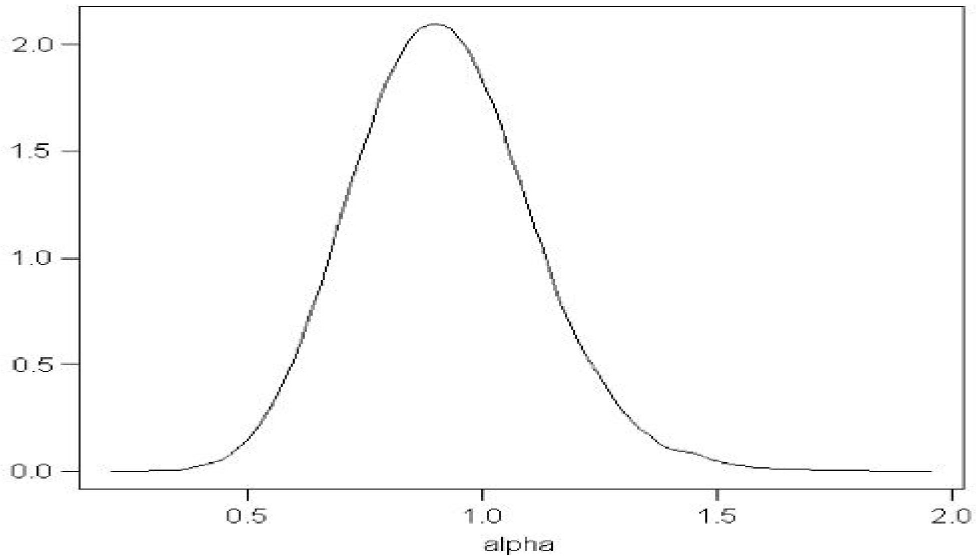
Parameter	Posterior mean	90% interval
μ	0.299	(−0.269, 0.886)
σ	1.194	(0.833, 1.725)

Modeling of random duration

- Illustrative Example: Recovery of offsite ac power

  Bayesian result:

  for Weibull distribution



It is shown that the posterior distribution of α has significant probability mass centered about 1.0, indicating that an exponential distribution might be a reasonable model.

Content

- Hierarchical modeling of variability
- Modeling of time-dependent reliability
- Modeling of random duration
- Treatment of uncertain and missing data (6)
- Bayesian Regression models
- Model selection and validation
- Summary and conclusion

Treatment of uncertain and missing data

➢ Motivation

- It is common in risk applications to encounter situations in which the observed data, which would normally enter into Bayes' theorem via the likelihood function, are either missing or the exact values are not known with certainty.

- Bayesian framework supplies a good tool for this case. We can simply assign the parameter of interest with a distribution that quantifies our available information, again reinforcing the idea that the Bayesian methodology encodes information via probability distribution.

Treatment of uncertain and missing data

- Illustrative example: motor-operated valves failing to open on demand

The number of demands is nominally 381, but could have been as high as about 440, and as low as 275. Assume that we have seen four failures to open.

How to estimate the probability of failure p ?

Treatment of uncertain and missing data

- Illustrative example: motor-operated valves failing to open on demand

Assign the demand number $n \sim U(275, 440)$

Give the failure probability a Jeffrey prior $p \sim Beta(0.5, 0.5)$

Result:

mean: 0.01284

95% credible interval(0.003738, 0.02804)

Comparison result: (n use 381)

mean: 0.01179

95% credible interval(0.003604, 0.02483)

$$\frac{4}{381} = 0.01050$$

Treatment of uncertain and missing data

- Illustrative example: motor-operated valves failing to open on demand

  If we don't know the failure number exactly, we take it as a uncertain value, and assume it could be 3, 4, 5, or 6, with $\Pr(3)=0.1, \Pr(4)=0.7, \Pr(5)=0.15$ and $\Pr(6)=0.05$.

Treatment of uncertain and missing data

- Illustrative example: motor-operated valves failing to open on demand

.

$$\pi_{avg}\left(p\,|\,x,n\right)=\sum_{i=1}^{N}\left[\int_{n_{lower}}^{n_{upper}}\frac{f\left(x_i\,|\,p,n\right)\pi\left(p\right)}{\int_0^{\infty}\int_{n_{lower}}^{n_{upper}} f\left(x_i\,|\,p,n\right)\pi\left(p\right)\pi\left(n\right)dndp}\pi\left(n\right)dn\right]\Pr\left(x_i\right)$$

Result:

If $n \sim U\left(275,440\right)$

mean: 0.01371

95% credible interval: (0.003759, 0.0302)

If n=381

mean: 0.0122

95% credible interval: (0.003495, 0.02609)

Content

- Hierarchical modeling of variability
- Modeling of time-dependent reliability
- Modeling of random duration
- Treatment of uncertain and missing data
- Bayesian Regression models (7)
- Model selection and validation
- Summary and conclusion

Bayesian Regression models

- Regression model

The unknown parameters denoted as β; this may be a scalar or a vector of length k

The independent variables, X

The dependent variable, Y.

$$Y = f\left(\beta, X\right)$$

Regression model could be used to predict the value of interest which will provide information to decision-maker prior.

- Bayesian Regression Model

The analysis was frequentist in nature, probability distributions representing epistemic uncertainty in the input parameters were not available. The Bayesian perspective is needed to operate the regression process.

Bayesian Regression models

- Illustrative example: predict model for ill-fated launch of Challenger in January 1986

There are six O-rings on the shuttle, during each launch, the number of distress events defined as erosion or blow-by of a primary field O-ring, is modeled as binomial with parameters p and 6: $X \sim Binomial(p,6)$

Use regression model:

logit(p) = a+b *temp + c*press

To predict the distress events at $31^\circ F$ which is the approximate temperature for the disastrous launch of the Challenger.

Flight	Distress events	Temp (°F.)	Press (p sig
1	0	66	50
2	1	70	50
3	0	69	50
5	0	68	50
6	0	67	50
7	0	72	50
8	0	73	100
9	0	70	100
41-B	1	57	200
41-C	1	63	200
41-D	1	70	200
41-G	0	78	200
51-A	0	67	200
51-C	2	53	200
51-D	0	67	200
51-B	0	75	200
51-G	0	70	200
51-F	0	81	200
51-I	0	76	200
51-J	0	79	200
61-A	2	75	200
61-B	0	76	200
61-C	1	58	200

Bayesian Regression models

- Illustrative example: predict model for ill-fated launch of Challenger in January 1986

logit(p) = a+b *temp + c*press

Result:

Parameter	Mean	Standard dev.	95% credible interval
a (intercept)	2.24	3.74	(−4.71, 9.92)
b (temp. coeff.)	−0.105	0.05	(−0.20, −0.02)
c (press. coeff.)	0.01	0.009	(−0.004, 0.03)

the model predicts about 4 distress events at 31$^{\circ}F$

logit(p) = a+b *temp

Content

- Hierarchical modeling of variability
- Modeling of time-dependent reliability
- Modeling of random duration
- Treatment of uncertain and missing data
- Bayesian Regression models
- Model selection and validation (8)
- Summary and conclusion

Model selection and validation

- Motivation
- Bayesian posterior predictive statistics and Bayesian p-value
- Deviance information criterion (DIC)

Model selection and validation

➢ Motivation

- If the selected prior distribution is appropriate? How to check it?

- If there are some random models for candidate to model the variability, which one is more appropriate? How to check it?

Model selection and validation

- To solve these problem, summary statistics derived from the posterior predictive distribution provides a very good tool, such as Bayesian χ^2 and Cramer-von Mises statistic, which lead to a Bayesian analog of p-value.

- The deviance information criterion (DIC), a Bayesian analog of a penalized likelihood measure also provides a good framework for the model selection.

Bayesian posterior predictive statistics and Bayesian p-value

- Bayesian χ^2

use the observed values of χ to form the observed statistic:

$$x_{obs}^2 = \sum_i \frac{\left(x_{obs,i} - \mu_i\right)^2}{\sigma_i^2}$$

generate replicate values of χ from its posterior predictive distribution to construct an analogous statistic:

$$x_{rep}^2 = \sum_i \frac{\left(x_{rep,i} - \mu_i\right)^2}{\sigma_i^2}$$

μ_i is the ith mean value, σ_i^2 is the ith variance.

If the model is very good, p-value $\Pr\left(\chi_{rep}^2 \geq \chi_{obs}^2\right)$ should be around 0.5.

If we want to judge the model validation, we can set the cutoff, we will reject a model with a p-value to below the cutoff. We also can select a model from many candidate models whose p-value is closest to 0.5

Bayesian posterior predictive statistics and Bayesian p-value

- Cramer-von Mises statistics

Rank the observed data x_{obs} as $x_{obs}^{(1)} \leq x_{obs}^{(2)} \leq \cdots \leq x_{obs}^{(N)}$

construct statistic

$$D_{obs} = \sum_i \left(F_{obs,i} - \frac{2i-1}{2N} \right)^2 \quad \text{where} \quad F_{obs,i} = \Pr\left(X < x_{obs}^{(i)} \right)$$

generate replicate values of x_{rep} from its posterior predictive distribution , then rank the replicate value as $x_{rep}^{(1)} \leq x_{rep}^{(2)} \leq \cdots \leq x_{rep}^{(N)}$ to construct an analogous statistic:

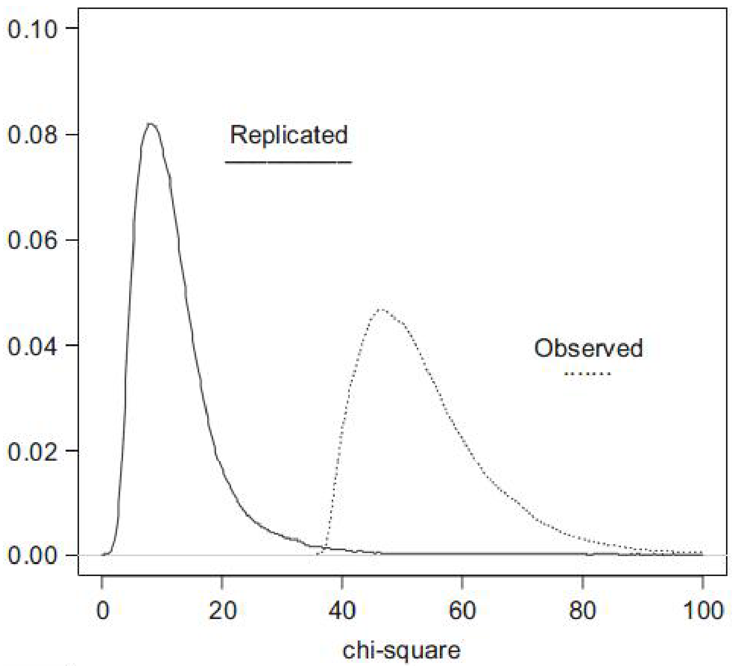
$$D_{rep} = \sum_i \left(F_{rep,i} - \frac{2i-1}{2N} \right)^2 \quad \text{where} \quad F_{rep,i} = \Pr\left(X < x_{rep}^{(i)} \right)$$

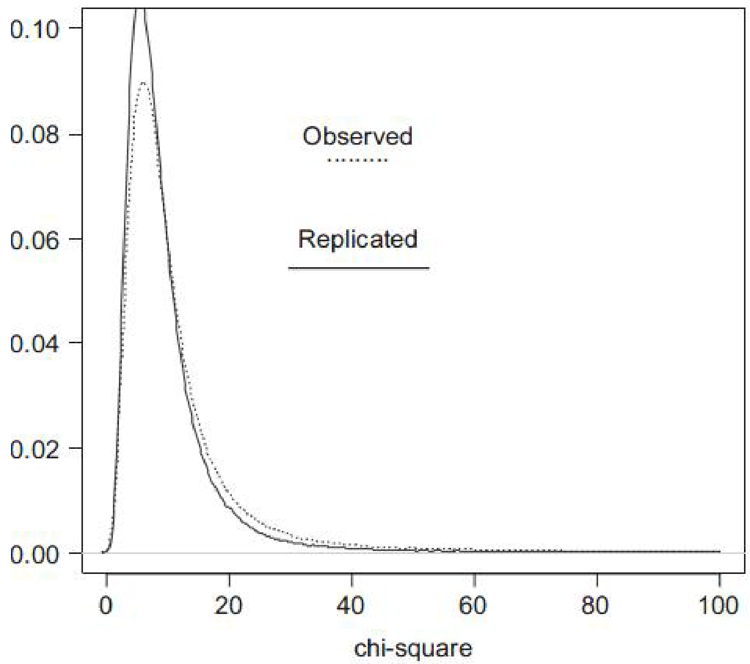
p-value is $\Pr\left(D_{rep} \geq D_{obs} \right)$

Bayesian posterior predictive statistics and Bayesian p-value

- Illustrative example: source to source variability

assume the constant λ | assume variable λ





P-value

Model	Bayesian p-value
Variable λ	0.46
Constant λ	0.002

Model selection and validation

➢ Deviance information criterion (DIC)

- The DIC is a hierarchical modeling generalization of the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by (MCMC) simulation.
- Define the deviance as $D(\theta) = -2\log\left(p\left(y|\theta\right)\right)$, $p\left(y|\theta\right)$ is the likelihood function . The expectation $\overline{D} = E_{\theta}\left[D(\theta)\right]$ is a measure of how well the model fits the data; the larger this is, the worse the fit. The effective number of parameters of the model is computed as $p_D = \overline{D} - D\left(\overline{\theta}\right)$. $\overline{\theta}$ is the expectation of θ . The deviance information criterion is calculated as:

$$DIC = p_D + \overline{D} = 2\overline{D} - D\left(\overline{\theta}\right)$$

Rule: The small the DIC is, the better the model fit.

DIC can be get from Winbugs.

Summary and conclusion

- A variety of other “model of the world” are amenable to inference via Bayes’ theorem , including Bayesian belief networks(BBN), influence diagram, and fault tree.
- Uncertain parameters (in the epistemic sense) are inputs to the models used to infer the values of future observations, leading to an increase in scientific knowledge.
- The advent of MCMC-based sampling methods, coupled with easy-to-use software and powerful computers, allows us to encode information via Bayes’ Theorem for a large variety of problems, domains, model types, data sets, and complications.
- It is best done via Bayesian inference with modern computational tools, which eliminate the need for the approximations and ad hoc approaches of the past.

Thank you for your attentions!