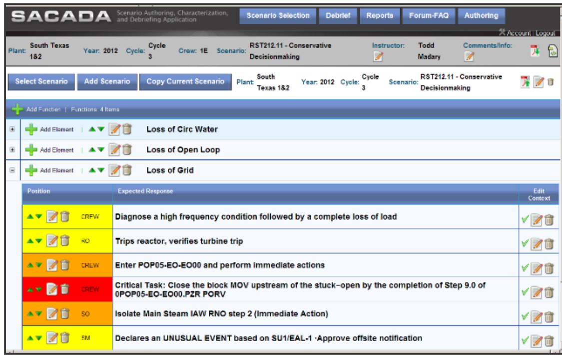
Estimating PSF Effects using Logistic Regression (LR) and Bayesian Inference of LR

Yochan Kim, Wondea Jung, Jinkyun Park

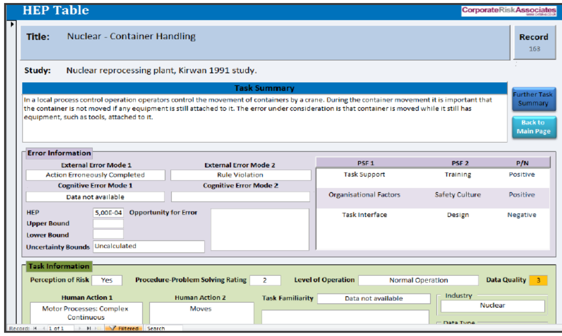
KAERI 한국원자력연구원 Korea Atomic Energy Research Institute

Introduction

- Requirement of Human Reliability Data
  - A wide range of quantitative estimates in the existing HRA methods are not supported by solid empirical bases



- Recent efforts of data collection
  - CAHR [Sträter, 1996]
  - CORE [Kirwan et al., 1997]
  - SACADA [Chang et al., 2014]
  - OPERA [Jung et al., 2016]

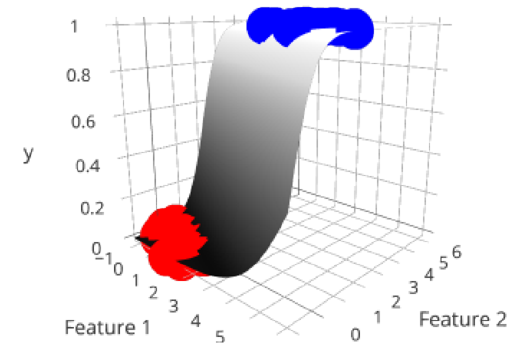


- Some estimates from the data
  - HEPs (Human error probabilities)
    - From OPERA DB [Kim et al., 2017]
    - From CORE-DATA [Basra and Kirwan, 1998]
    - From GRS event report [Preischl and Hellmich, 2013]
    - From laboratory experiments [Jang et al., 2013]
  - PSF (Performance Shaping Factor) effects
    - From laboratory experiments [Liu and Li, 2014]
    - From laboratory experiments [Kim et al., 2015]
    - From OPERA DB [Kim et al., 2018]

Logistic Regression

Logistic Regression (LR)

- Logistic regression
  - To predict a conditional probability of an event given a set of independent variables
  - Dichotomous dependent variables representing event occurrences are usually used.
  - Regression model



$$p(x) \approx \frac{p(x)}{1-p(x)} = e^{\beta_0} \cdot e^{\beta_1 x_1} \cdot \ldots \cdot e^{\beta_v x_v}$$

where $x_1,\ldots,x_v$ are the independent variables of the regression model predicting a conditional probability, $p(x)$, and $\beta_0,\ldots,\beta_v$ are the regression coefficients.

- Useful to derive quantitative effects of PSFs
  - HEP quantification model in HRA method

$$HEP = NHEP \cdot PSFmultiplier_1 \cdot x_1 \ldots PSFmultiplier_v \cdot x_v$$

PSF Effect Estimation by LR(1)

Data Collection[Kim et al., 2018]

- Data of training records from full-scope simulators
  - Conventional MCR
  - 223 records

Reference plant type	Scenario	Number of collected records
Westinghouse-type plant	Interfacing System Loss of Coolant Accident (ISLOCA)	10
	Steam Generator Tube Rupture (SGTR) following Main Steam Line Break (MSLB)	8
Combustion engineering-type plant (OPR1000)	Control element assembly Deviation	14
	Charging system volume control tank outlet valve failure	18
	Pressurizer level controller failure	22
	Reactor coolant pump cyclone filter blockage	8
	Condensate polishing system valve close	8
	Reactor containment pan cooler high vibration	18
	Deaerator level controller failure and inlet valve blockage	13
	Condensate tube loss	40
	Condenser vacuum lowering	13
	Compressed instrument air loss	19
	Emergency seal oil pump spurious start	22
	04SN bus power loss	10

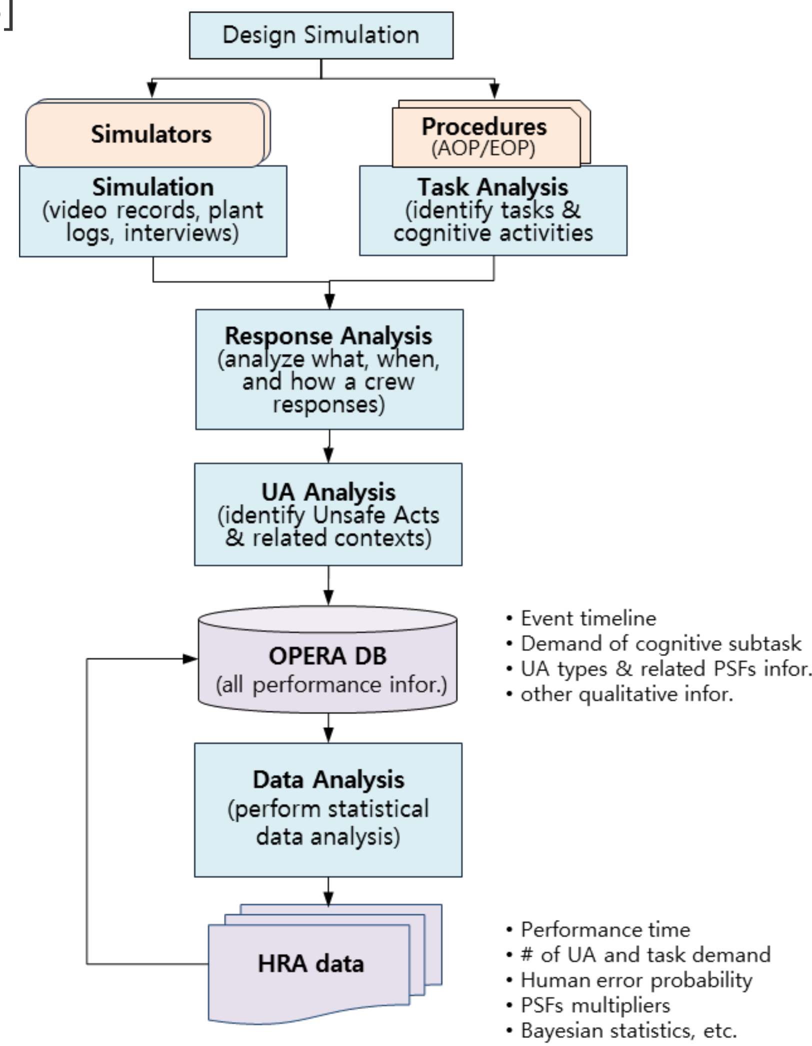
Kim et al., "Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data", RESS 173 (2018) 12–22

PSF Effect Estimation by LR(2)

Statistical Analysis[Kim et al., 2018]

- Dependent variable (DV) [Kim et al., 2018]
  - Unsafe act occurrences (1: occurred, 0: not)
  - For 6 types of unsafe act types
    - Information gathering (~~EOO,~~ EOC)
    - Situation interpreting (~~EOO,~~ EOC)
    - Response planning (EOO, EOC)
    - Execution (EOO, EOC)

- Independent variable (IV)
  - 26 variables in OPERA database
    - Crew information and training information
    - Environmental issues
    - Overall crew characteristics including communication and leadership
    - Task type
    - Component/system type to be controlled
    - Time pressure
    - Human-machine interface attributes
    - Communication quality
    - Task complexity
    - Task familiarity
    - Procedure quality
    - Recovery information
    - …



<HuREX Process>

Kim et al., "Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data", RESS 173 (2018) 12–22

PSF Effect Estimation by LR(3)

Result[Kim et al., 2018]

Cognitive activity	Error mode	Multiplier Estimator (coefficient, exponentiated coefficient)	P-value (likelihood ratio test)
Information gathering and reporting	EOC	•(Intercept) (-6.43, 1.61e-03) •Confusing statement = TRUE (2.52, 1.24e+01)	1.13E-05 ***
Situation interpreting	EOC	•(Intercept) (-2.08, 1.25e-01) •Simulation mode = EMERGENCY and Time pressure = INSIGNIFICANT (-16.5, 6.92e-08) •Simulation mode = ABNORMAL and Time pressure = URGENT (3.18, 2.40e+01)	1.33E-03 **
Response planning and instruction	EOO	•(Intercept) (-2.40, 9.09e-02) •Instruction contents = DISCRETE CONTROL (2.18, 8.83) •Instruction contents = INFORMATION(-1.61, 2.00e-01) •Instruction contents = EX-CONTROL (2.15, 8.58) •Instruction contents = PROCEDURE (0.0691, 1.07) •Continuous action step = TRUE (1.53, 4.62) •Training experience = TRUE (-4.00, 1.84e-02) •Simulation mode = EMERGENCY (-2.83, 5.87e-02) •Multiple constraint = TRUE (1.98, 7.26)	4.44E-32 ***
Response planning and instruction	EOC	•(Intercept) (-5.46, 4.26e-03) •Simulation mode = EMERGENCY and Contingency action part = FALSE (-0.181, 8.34e-01) •Simulation mode = EMERGENCY and Contingency action part = TRUE (3.42, 3.06e+01) •Description of object = TRUE (-1.83, 1.60e-01)	3.16E-13 ***
Execution	EOO	•(Intercept) (-6.49, 1.52e-03) •Number of manipulation (0.159, 1.17)	7.44E-09 ***
Execution	EOC	•(Intercept) (-5.77, 3.12e-03) •Confusing statement = TRUE (1.88, 6.54)	6.13e-02

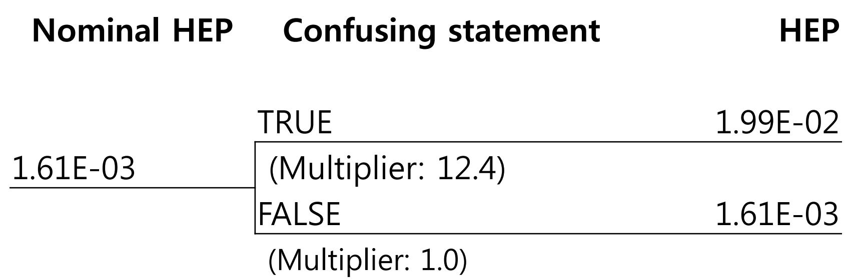
Kim et al., "Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data", RESS 173 (2018) 12–22

PSF Effect Estimation by LR(4)

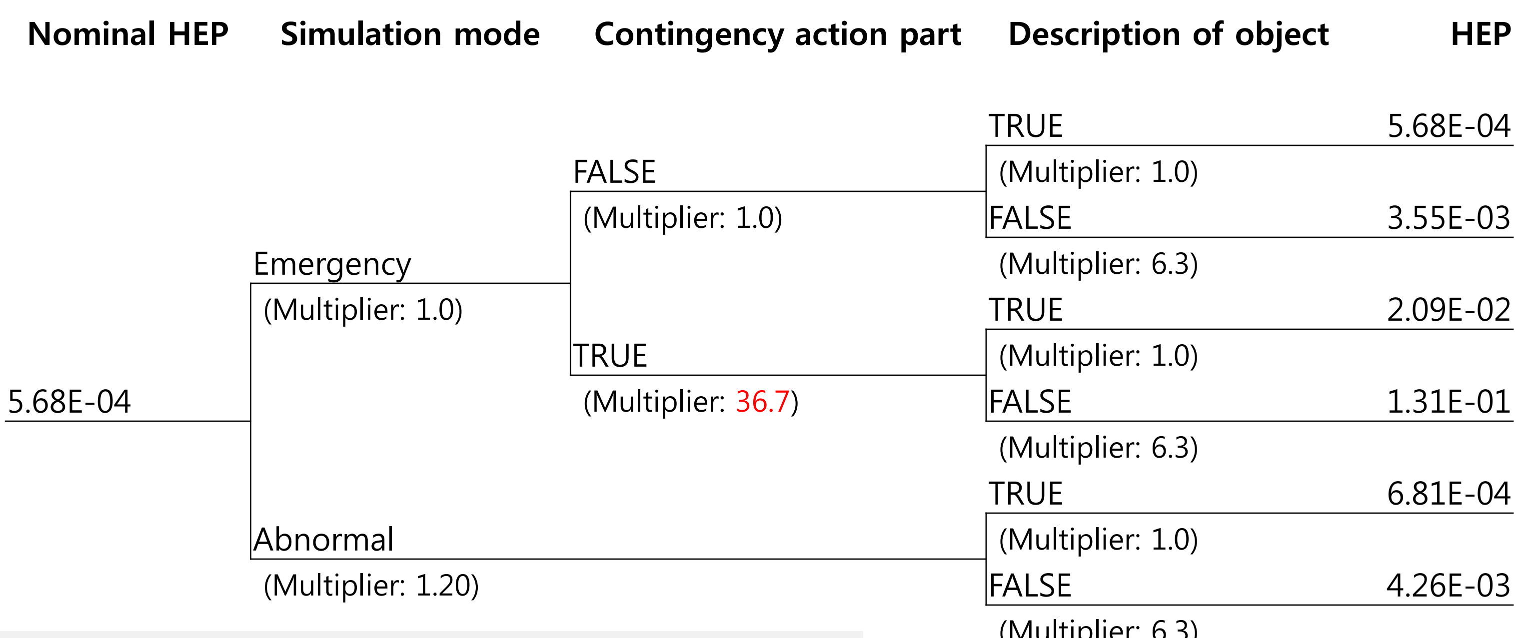
Decision Tree based on the Result[Kim et al., 2018]

EOO: error of omission;
EOC: error of commission

<Information gathering and reporting (EOC)>



<Response planning and instruction (EOC)>

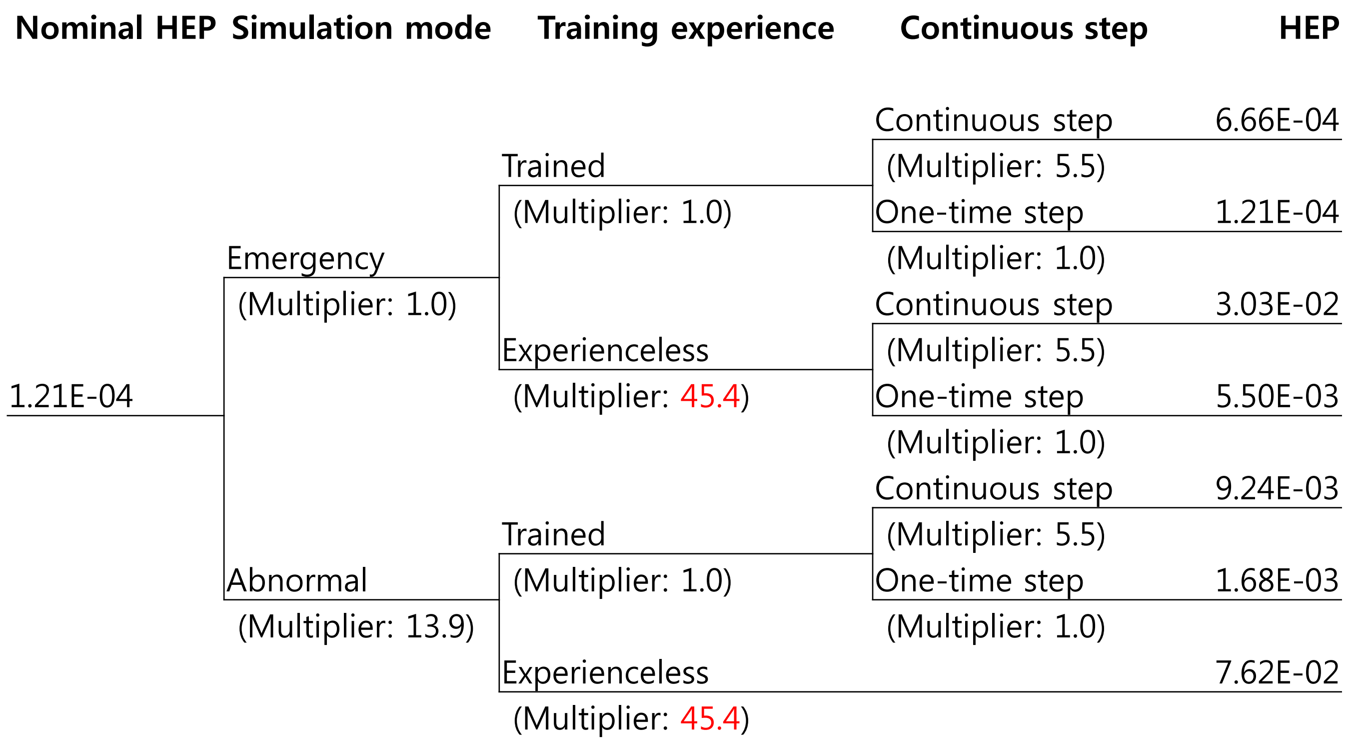


Kim et al., "Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data", RESS 173 (2018) 12–22

PSF Effect Estimation by LR(5)

Decision Tree based on the Result[Kim et al., 2018]

<Response planning and instruction (EOO), Task: procedure progression>



Kim et al., "Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data", RESS 173 (2018) 12–22

PSF Effect Estimation by LR(6)

Decision Tree based on the Result[Kim et al., 2018]

<Execution (EOO)>

Nominal HEP	Manipulation #	HEP
1.52E-03	1 (Multiplier: 1.17^ 1)	1.78E-03
	2 (Multiplier: 1.17^ 2)	2.08E-03
	...	...
	27 (Multiplier: 1.17^ 27)	1.10E-01

<Execution (EOC)>

Nominal HEP	Confusing statement	HEP
3.12E-03	TRUE (Multiplier: 6.5)	2.04E-02
	FALSE (Multiplier: 1.0)	3.12E-03

Kim et al., "Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data", RESS 173 (2018) 12–22

Limitation of Statistical Analysis

Findings from the Results[Kim et al., 2018]

- Difference between estimates in expert judgment and statistical analysis
  - Response planning (EOO)

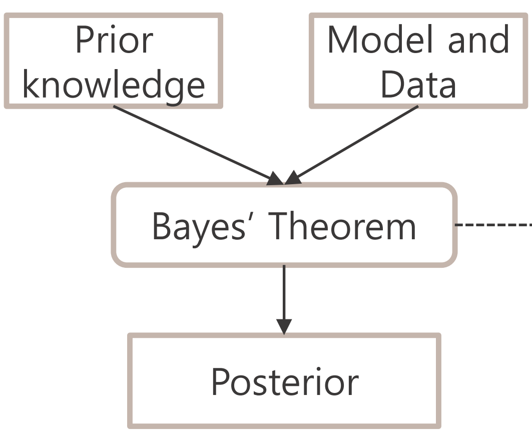
Statistical Analysis with OPERA DB		SPAR-H		HEART	
Variable	PSF Effect estimates	PSF	PSF multiplier	PSF	PSF multiplier
Continuous action step	5.5 (True) 1 (False)	Complexity	5 (Highly complex) 2 (Moderately complex) 1 (Nominal) 0.1 (Obvious diagnosis)	A channel capacity overload	6
Training experience	45.4 (False) 1 (True)	Experience/ Training	10 or 3 (Low) 1 (Nominal) 0.5 (High)	Unfamiliarity	17

- Limitation of statistical analysis: Sensitive to data
  - Multi-collinearity
  - Missing values in some data area
  - PSF level definition
  - Insufficient samples
  - Effects of latent variable

Kim et al., "Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data", RESS 173 (2018) 12–22

Bayesian Inference

- Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available. [Wikipedia]
- The Bayesian approach provides a formal mechanism for combing all available information [Smith, 2017]
  - Including engineering and qualification test data, field experience, expert judgment, and data from similar systems



From our knowledge of the problem beyond what we call "data"

From the model representing the process providing the data

$$\underbrace{\pi_1(\lambda|data)}_{\text{Posterior}} = \frac{\overbrace{f(data|\lambda)}^{\text{Likelihood}}\overbrace{\pi_0(\lambda)}^{\text{Prior}}}{\underbrace{\int f(data|\lambda)\pi_0(\lambda)d\lambda}_{\text{Normalizing constant}}}$$

From application of Bayes' theorem to obtain the probability of a hypothesis conditional upon our knowledge and applicable data

"Total probability" over all specified hypotheses H

[Smith, 2017]

Bayesian Inference in Logistic Regression

Bayesian update of regression coefficients

- Likelihood: logistic model from OPERA DB
- Prior: independent prior distributions for regression coefficients
  - Normal distribution, $\beta_i \sim N(\mu_i, \sigma_i^2)$
  - **Case 1: noncommittal broad prior, $N(0, 10^2)$**
  - Case 2: highly informative prior to 0, $N(0, 1^2)$
  - Case 3: highly informative prior from SPAR-H method, $N(PSF_{SPAR\text{-}H}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$
  - Case 4: highly informative prior by CREAM method, $N(PSF_{CREAM}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$

Prior X Likelihood

$N(0, 10^2)$

$$ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots$$

OPERA from HuREX framework

Bayesian Inference in Logistic Regression

Bayesian update of regression coefficients

- Likelihood: logistic model from OPERA DB
- Prior: independent prior distributions for regression coefficients
  - Normal distribution, $\beta_i \sim N(\mu_i, \sigma_i^2)$
  - Case 1: noncommittal broad prior, $N(0, 10^2)$
  - **Case 2: highly informative prior to 0, $N(0, 1^2)$**
  - Case 3: highly informative prior from SPAR-H method, $N(PSF_{SPAR\text{-}H}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$
  - Case 4: highly informative prior by CREAM method, $N(PSF_{CREAM}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$

Prior **X** **Likelihood**

$N(0, 1^2)$

$$ln\left(\frac{p(x)}{1-p(x)}\right) = \boxed{\beta_0} + \boxed{\beta_1} x_1 + \boxed{\beta_2} x_2 + \boxed{\beta_3} x_3 + \ldots$$

OPERA from HuREX framework

Bayesian Inference in Logistic Regression

Bayesian update of regression coefficients

- Likelihood: logistic model from OPERA DB
- Prior: independent prior distributions for regression coefficients
  - Normal distribution, $\beta_i \sim N(\mu_i, \sigma_i^2)$
  - Case 1: noncommittal broad prior, $N(0, 10^2)$
  - Case 2: highly informative prior to 0, $N(0, 1^2)$
  - **Case 3: highly informative prior from SPAR-H method, $N(\mathrm{PSF}_{\text{SPAR-H}}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$**
  - Case 4: highly informative prior by CREAM method, $N(\mathrm{PSF}_{\text{CREAM}}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$

Prior X **Likelihood**

$N(0, 10^2)$ $N(\mathrm{PSF}_{\text{SPAR-H}}, 1^2)$

$$ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots$$

OPERA from HuREX framework

Bayesian Inference in Logistic Regression

Bayesian update of regression coefficients

- Likelihood: logistic model from OPERA DB
- Prior: independent prior distributions for regression coefficients
  - Normal distribution, $\beta_i \sim N(\mu_i, \sigma_i^2)$
  - Case 1: noncommittal broad prior, $N(0, 10^2)$
  - Case 2: highly informative prior to 0, $N(0, 1^2)$
  - Case 3: highly informative prior from SPAR-H method, $N(PSF_{SPAR\text{-}H}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$
  - **Case 4: highly informative prior by CREAM method, $N(PSF_{CREAM}, 1^2)$, noncommittal broad prior on intercept, $N(0, 10^2)$**

Prior X Likelihood

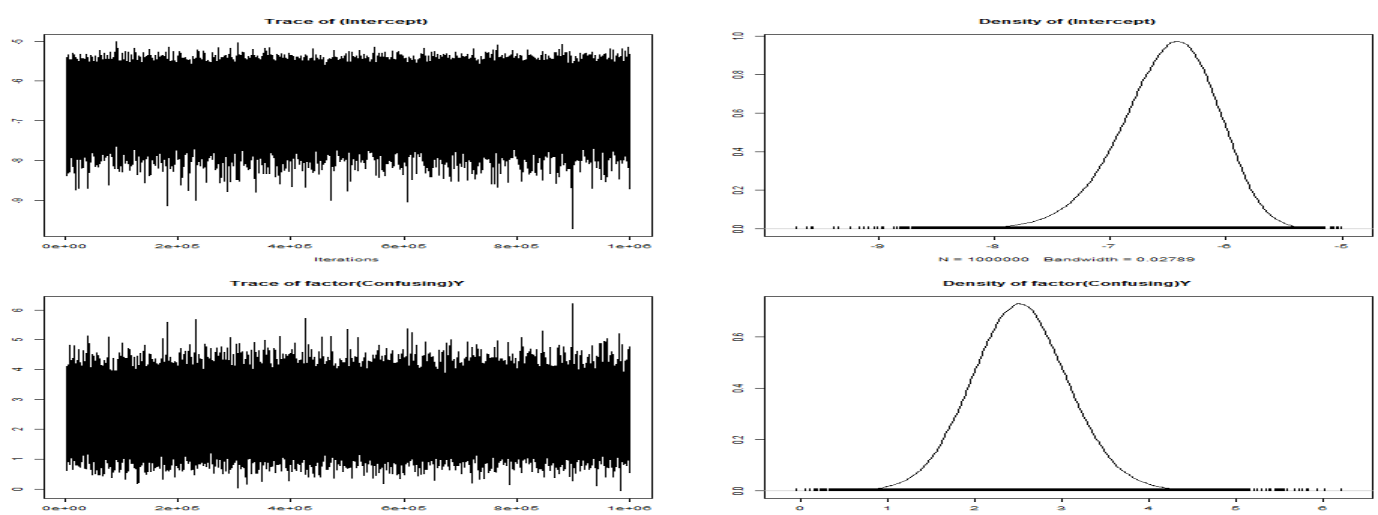
$N(0, 10^2)$ $N(PSF_{CREAM}, 1^2)$

$$ln\left(\frac{p(x)}{1-p(x)}\right) = \boxed{\beta_0} + \boxed{\beta_1}x_1 + \boxed{\beta_2}x_2 + \boxed{\beta_3}x_3 + \ldots$$

OPERA from HuREX framework

Bayesian Inference Method

- Calculation algorithm for posterior probabilities
  - MCMC (Markov Chain Monte Carlo) algorithm: sampling from a probability distribution based on constructing a Markov chain that has the desired distribution as its equilibrium distribution
- R statistical tool with “MCMCpack” package was employed
- The number of Metropolis iterations for the sampler: 1,000,000
- The number of burn-in iterations for the sampler: 1,000



Result

Errors of commission in information gathering

* ML: maximum likelihood, BU: Bayesian update

PSF	ML	BU with $N(0, 10^2)$ (Case 1)	BU with $N(0, 1^2)$ (Case 2)	BU with SPAR-H (Case 3)	BU with CREAM (Case 4)	SPAR-H	CREAM
Nominal HEP	1.61.E-03	1.50.E-03	3.41.E-03	1.69.E-03	1.88.E-03	-	-
Confusing statement	12.38	12.45	4.43	10.08	8.14	5 (poor procedure)	2 (inappropriate proc.)

Result

Errors of omission in response planning

* ML: maximum likelihood, BU: Bayesian update

PSF	ML	BU with $N(0, 10^2)$ (Case 1)	BU with $N(0, 1^2)$ (Case 2)	BU with SPAR-H (Case 3)	BU with CREAM (Case 4)	SPAR-H	CREAM
Nominal HEP	4.68.E-05~ 1.03.E~03	7.83.E-06~ 1.75.E-04	4.15.E-05~ 6.74.E-04	1.25.E-05~ 2.61.E-04	1.54.E-05~ 3.15.E-04	-	-
Continuous action step	5.50	5.57	4.44	5.48	5.47	5 (high complexity)	5 (more than capacity)
Training experience	45.40	45.20	15.65	30.45	25.28	10 (low training)	5 (inadequate training)
Simulation mode	13.86	13.49	6.32	10.53	9.30	-	-

Result

Errors of commission in response planning

* ML: maximum likelihood, BU: Bayesian update

PSF	ML	(Case 1) BU with $N(0, 10^2)$	(Case 2) BU with $N(0, 1^2)$	(Case 3) BU with SPAR-H	(Case 4) BU with CREAM	SPAR-H	CREAM
Nominal HEP	5.68.E-04	4.64.E-04	6.84.E-04	4.93.E-04	4.95.E-04	-	-
Simulation mode	1.20	1.19	2.88	1.83	2.13	-	-
Contingency action	36.73	39.96	24.17	34.92	32.17	5 (high complexity)	2 (low experience)
Description of object	6.26	6.65	5.70	6.13	6.10	5 (poor procedure)	5 (inappropri ate proc.)

Result

Errors of omission in execution

* ML: maximum likelihood, BU: Bayesian update

PSF	ML	BU with $N(0, 10^2)$ (Case 1)	BU with $N(0, 1^2)$ (Case 2)	BU with SPAR-H (Case 3)	BU with CREAM (Case 4)	SPAR-H	CREAM
Nominal HEP	1.80.E-03	1.59.E-03	5.46.E-03	1.58.E-03	-	-	-
# of manipulation	1.17	1.17	1.11	1.18	-	2 (moderately complexity)	-

Result

Errors of commission in execution

* ML: maximum likelihood, BU: Bayesian update

PSF	ML	BU with $N(0, 10^2)$ (Case 1)	BU with $N(0, 1^2)$ (Case 2)	BU with SPAR-H (Case 3)	BU with CREAM (Case 4)	SPAR-H	CREAM
Nominal HEP	3.12.E-03	2.80.E-03	7.10.E-03	2.92.E-03	3.27.E-03	-	-
Confusing statement	6.54	5.57	1.71	5.55	3.56	5 (poor procedure)	2 (inappropriate proc.)

Summary and Discussion

- The human reliability data was analyzed with the Bayesian logistic regression.
  - 4 kinds of prior knowledge were applied.
  - BU incorporates the empirical data with prior knowledge.
  - The PSF effects by BU were less sensitive to data characteristics.
  - BU allows measuring the uncertainties. (not addressed today)
- Most coefficients by BU approached to the expected values by incorporating priors.
  - The most conservative prior, 'BU with $N(0, 1^2)$', suppressed the effects of PSFs from ML estimation.
- Some PSF effects were still large.
  - Training may have more effects on reliability than our expectation.
  - Effects of contingency action part can involve interactive influences of two or more factors.
- Which prior is suitable to PSF modeling is important to appropriate estimation.
- Quality and quantity of empirical data is still valuable.

Research Plan

- New training records from a digital MCR is being collected.
  - Similar statistical analysis can be performed on new records

THANK
YOU

감사합니다.

yochankim@
kaeri.re.kr