In the memory of Professor John Nelder.

From Dutch book, posterior, likelihood, fiducial probability, relevant subsets to epistemic confidence, extended likelihood and new package for statistical science.

Youngjo Lee

Dankook University

May 14th, 2021

Co-workers: Hangbin Lee, Jeongseop Han, Maengseok Noh, Ildo Ha, Rezzy Eko Caraka, Smart Sarpong, Yudi Pawitan.

1. I. The Confidence.
   - Introduction
   - Relevant subsets
   - Confidence distribution
   - Ancillary statistics
   - Examples
   - Discussion
2. II. H-likelihood.
   - Extended likelihood and H-likelihood
3. III. Various Models.
4. IV. A New Package for Statistical Science.

Introduction

- Given data $Y = y$ from a model $p_\theta(y)$ with scalar parameter θ, a confidence interval CI(y) is computed with coverage probability

$$P_\theta(\theta \in \text{CI(Y)}) = \gamma.$$

- We are interested in the *epistemic confidence*, defined as the sense of confidence in the observed CI(y).
- However, the frequentist view is emphatic that the probability γ does not apply to the observed CI(y) but to the procedure CI(Y).
- Traditionally, only the Bayesians have no problem in stating that their subjective probability is epistemic.
- Is there a way to make a non-Bayesian confidence epistemic?

Introduction

- For interpretation, we introduce a logical device, called the Dutch Book, classically proposed by Ramsey(1926) and de Finetti(1931). One's subjective probability of event E is defined as the personal betting price one puts on E.
- Though subjective, the price is not arbitrary, it is a price such that no external agent can construct a Dutch Book against them, i.e., make a risk-free profit.
- We define confidence to be epistemic if it is protected from the Dutch Book.
- We assume a betting market of a crowd of independent intelligent players, where bets are like a commodity with supply and demand.

Introduction

- Assuming a perfect market condition – full competition, perfect information and no transaction cost –, there is an equilibrium price at which there is balance between supply and demand. 'Perfect information' here means all players have access to the data y and the sampling model $p_\theta(y)$.
- For the betting market, assuming an objective probability model, there is no arbitrage if the price is determined by the objective probability. (Ross, 1976)
- The Bayesian setup does not presume the betting market, hence bets are only between two people. To avoid Dutch Book, you should make your bets internally consistent by following probability laws.
- However, even if your bets are internally consistent , if your prices do not match the market prices, under our setting a risk-free profit can be made by playing between you and the market.

Introduction

- How the confidence can be applied to the observed interval?
- Intuitively, this is when you are sure that you have used all the available information in the data, so nobody can construct a Dutch Book against you. Theoretically, to construct the Dutch Book, an external agent must exploit unused information in the form of a relevant subset, conditional on whether he can get a different coverage probability.
- Pawitan and Lee (2021) showed that the confidence is an extended likelihood (Lee et al., 2017). The likelihood principle (Birnbaum, 1962) and its extension (Bjørnstad, 1996) state that the likelihood contains all the information in data.
- Intuitively, this implies that the full likelihood leaves no relevant subset, and is thus protected from the Dutch Book. Our aim is to establish the theoretical justification for this intuitive notion.

Relevant subsets

- A statistic $R(y)$ is defined to be relevant if the conditional coverage is non-trivially and consistently biased in one direction. This means that there is $\epsilon > 0$ free of θ, such that

$$\begin{aligned} \text{positively biased}: &\quad P_\theta(\theta \in \mathrm{CI}(Y)|R(y)) \geq \gamma + \epsilon \quad \text{for all } \theta, \\ \text{or negatively biased}: &\quad P_\theta(\theta \in \mathrm{CI}(Y)|R(y)) \leq \gamma - \epsilon \quad \text{for all } \theta. \end{aligned}$$

- If it exists, $R(y)$ can be used to construct Dutch Book against betting price γ, so the confidence level γ is not epistemic.
- Conversely, If there is no relevant subset, the betting price determined by the confidence level is protected from the Dutch Book. Thus, we establish that confidence is epistemic by showing that there is no relevant subset.

Example 1: relevant subsets

- Let $y = (y_1, y_2)$ be an iid sample from a uniform distribution on $\{\theta - 1, \theta, \theta + 1\}$, where the parameter θ is an integer.
- Let $y_{(1)}$ and $y_{(2)}$ be the minimum and maximum values of y_1 and y_2.
- A confidence interval $\text{CI}(y) \equiv [y_{(1)}, y_{(2)}]$ has a coverage probability

$$P_\theta(\theta \in \text{CI}) = 7/9.$$

- Suppose that we observe $y_{(1)} = 3$ and $y_{(2)} = 5$, then we have $\text{CI}(y) = [3, 5]$.
- This interval is formally 78% CI for θ, but we can actually be sure that $\theta = 4$.

Example 1: relevant subsets

- Here the range $R(y) \equiv y_{(2)} - y_{(1)}$ is relevant. For all θ, we have

$$P_\theta(\theta \in \text{CI} | R = 2) = 1 > 7/9$$
$$P_\theta(\theta \in \text{CI} | R = 1) = 1 > 7/9$$
$$P_\theta(\theta \in \text{CI} | R = 0) = 1/3 < 7/9.$$

- In the betting market, $R(y)$ will be used by the intelligent players to settle prices at these conditional probabilities. If $y_1 = 3$ and $y_2 = 5$ are observed, the intelligent players will not use 7/9 as the price and will instead use 1.

Confidence distribution

- Let $t \equiv T(y)$ be a statistic of θ, and define the right-side P-value function

$$C_m(\theta; t) \equiv P_\theta(T \geq t).$$

  Here the subscript m is used to indicate that it is a 'marginal' confidence.
- It generally behaves like a proper cumulative distribution function of θ for each t, so that $C_m(\theta; t)$ is called the confidence distribution of θ.
- For continuous θ, the confidence density and likelihood are defined by

$$c_m(\theta) \equiv c_m(\theta; t) \equiv \partial C_m(\theta; t)/\partial\theta, \qquad L(\theta) \equiv L(\theta; t) \equiv -\partial C_m(\theta; t)/\partial t,$$

  and

$$C_m(\theta \in \text{CI}) \equiv \int_{CI} c_m(\theta) d\theta \tag{1}$$

  to convey the confidence of θ belonging in the CI.

Confidence distribution

- The frequentist CI is defined by

$$\mathrm{CI}_\gamma(T) = \left(q_{\gamma_2}^{-1}(T), q_{\gamma_1}^{-1}(T)\right)$$

where $q_\gamma^{-1}(T)$ is the inverse function of γ-qunatile function of T and $\gamma_2 - \gamma_1 = \gamma$, to have a coverage probability $P_\theta(\theta \in \mathrm{CI}_\gamma(T)) = \gamma$.

Lemma

Under the regularity condition R1,

$$P_\theta(\theta \in CI_\gamma(T)) = C_m(\theta \in CI_\gamma(t); t).$$

- Fisher (1930, 1933) called $C_m(\theta; t)$ the fiducial probability of θ. Lemma 1 states when Fisher's fiducial probability becomes the coverage probability for continuous t. For discrete t, it holds only asymptotically.

Confidence distribution

- Define the implied prior as

$$c_0(\theta) \equiv c_0(\theta; t) \equiv m(t)\frac{c_m(\theta; t)}{L(\theta; t)}, \tag{2}$$

where $m(t)$ cancels out all the terms not involving θ in $c_m(\theta; t)/L(\theta; t)$. Then the full confidence density is defined by

$$c_f(\theta) \equiv c_f(\theta; y) \propto c_0(\theta)L(\theta; y). \tag{3}$$

- The subscript f indicates that it is associated with the full likelihood based on the whole data y. When T is sufficient, $c_m(\theta) = c_f(\theta)$. But in general they are not equal. $c_f(\theta; y)$ looks like a Bayesian posterior. However, the implied prior is not subjectively selected, and can be improper and even data-dependent.

Confidence distribution

Regularity Conditions.

R1. T is a continuous scalar statistic whose quantile function $q_\alpha(\theta)$, defined by

$$P_\theta(T \le q_\alpha(\theta)) = \alpha,$$

is strictly increasing function of θ for any $\alpha \in (0, 1)$.

R2. $c_0(\theta)$ is positive and locally integrable on the parameter space Θ such that

$$\int_J c_0(\theta) d\theta < \infty, \quad \text{for any compact subsets } J \subseteq \Theta.$$

R3. $\log c_0(\theta)$ is uniformly continuous in y.

R4. The confidence interval $\mathrm{CI}(y) = (b_L(y), b_U(y))$ is locally bounded, i.e., for any compact set K in the sample space of y, there exist M_1 and M_2 such that

$$|b_L(y)| \le M_1 \quad \text{and} \quad |b_U(y)| \le M_2 \quad \text{for any } y \in K.$$

Confidence distribution

Theorem

- *Consider the full confidence density $c_f(\theta) \propto c_0(\theta)L(\theta; y)$, with $c_0(\theta)$ being the implied prior defined by (2) satisfying R2 and R3, and $T(Y)$ satisfies R1.*
- *Let γ be the degree of confidence for the observed confidence interval $CI(y)$ that satisfies R4, such that*

$$\gamma = \int_{CI(y)} c_f(\theta)d\theta, \quad \text{for all } y.$$

- *Then $c_f(\theta)$ does has no relevant subsets.*

- $c_0(\theta) \equiv c_0(\theta; y)$ can be an arbitrary function that satisfies R2 & R3 and leads to proper $c(\theta; y)$. In particular, it does not have to be an implied prior $c_0(\theta; t)$. For example, if $c_0(\theta)$ is proper Bayesian prior, then $c_f(\theta)$ is a Bayesian posterior, shown by Robinson's (1979) not to have relevant subsets.

Confidence distribution

- If T is sufficient, the frequentist CI satisfies

$$P_\theta(\theta \in \mathrm{CI}_\gamma(Y)) = C_m(\theta \in \mathrm{CI}_\gamma(y)) = C_f(\theta \in \mathrm{CI}_\gamma(y)) = \gamma,$$

  for all θ and for all y. Note that $P_\theta(\theta \in \mathrm{CI}(Y)) = C_f(\theta \in \mathrm{CI}(y))$ holds asymptotically, regardless whether y is continuous or discrete.
- If you use an arbitrary $c_0(\theta; y)$ that is not the implied prior, your price γ will differ from the market price. Then, I can construct a Dutch Book against you. It means that, in this case, the theorem is meaningful only for two people betting repeatedly against each other, with gains or loses expressed in terms of expected value or long-term average. It is exactly the setting described by Buehler (1959) and Robinson (1979).

Ancillary statistics

- Suppose that $A(y) = a$ is ancillary, $T(y) = t$ is not sufficient but (t, a) is sufficient. In this case, $A(y)$ is called a maximal ancillary and

$$L(\theta; y) = L(\theta; t, a) \propto p_\theta(t|a)p(a) \propto p_\theta(t|a) = L(\theta; t|a). \tag{4}$$

- Thus, conditioning a non-sufficient statistic by a maximal ancillary has recovered the lost information and restored the full-data likelihood.
- Define the conditional confidence distribution given $A(y) = a$,

$$C_c(\theta; t|a) \equiv P_\theta(T \geq t|a), \qquad c_c(\theta; t|a) \equiv \partial C_c(\theta; t|a)/\partial\theta.$$

Corollary

Under the regularity condition R1,

$$P_\theta(\theta \in CI|a) = C_c(\theta \in CI; t|a).$$

where CI is the confidence interval based on the conditional distribution of $T|a$.

Ancillary statistics

- As before, we define the implied prior

$$c_0(\theta) \equiv c_0(\theta; t|a) \equiv m(t, a)\frac{c_c(\theta; t|a)}{L(\theta; t|a)},$$

  and the full confidence $c_f(\theta) \propto c_0(\theta)L(\theta; y)$.
- In particular, the conditional confidence becomes the full confidence:

$$c_c(\theta; t|a) = c_f(\theta).$$

Corollary

If $A(y) = a$ is maximal ancillary for $T(y)$ and CI is constructed from the conditional confidence density based on $T|a$, then under R1-R4, the conditional confidence $C_c(\theta \in CI; t|a)$ has no further relevant subsets.

Updating formula

In practical, $c_f(\theta; y)$ is hard to compute!!!

- Suppose that, for sample size $n = 1$, there is a statistic $t_1 \equiv T(y_1)$ that allows us to construct a valid confidence density $c_m(\theta, t_1)$. Then we can compute $c_0(\theta)$ based on $c_m(\theta; t_1)/L(\theta; t_1)$.
  When $c_0(\theta)$ is free of data, the updating formula gives

$$\begin{aligned} c_f(\theta; y) &\propto c_m(\theta; t_1) L(\theta; y_1|t_1) L(\theta; y_2 \cdots y_n) \\ &\propto c_0(\theta) L(\theta; t_1) L(\theta; y_1|t_1) L(\theta; y_2 \cdots y_n) \\ &= c_0(\theta) L(\theta; y_1) L(\theta; y_2 \cdots y_n) \\ &= c_0(\theta) L(\theta; y). \end{aligned} \tag{5}$$

- The statistic t_1 trivially exists if y_1 itself leads to a valid confidence density.

Location family

- Suppose that $y_1, \ldots, y_n$ are an iid sample from the location family with density

$$p_\theta(y_i) = f(y_i - \theta),$$

where $f(\cdot)$ is an arbitrary but known density function. Immediately, based on y_1 alone, the confidence density is

$$c(\theta; y_1) = f(y_1 - \theta) = L(\theta; y_1),$$

so the implied prior $c_0(\theta) = 1$. Using formula (5), the full confidence density is

$$c_f(\theta) \propto L(\theta) = \prod_{i=1}^{n} f(y_i - \theta). \tag{6}$$

This is a remarkably simple way to arrive at the confidence density of θ.

Exponential family

- Suppose that y is an iid sample from the exponential family with log-density

$$\log p_\theta(y_i) = \sum_{j=1}^{J} h_j(\theta) t_j(y_i) - A(\theta) + c(y_i).$$

The standard evaluation of confidence requires the tail probability of the distribution of the MLE, which in general has no closed form formula.

Small Sample Asymptotics for MLE !!!

- Barndorff-Nielsen's (1983) magical formula gives an approximation,

$$p_\theta(\hat{\theta}|a) = k|I(\hat{\theta})|^{1/2} \frac{L(\theta)}{L(\hat{\theta})} + O(n^{-1}),$$

where k is a normalizing constant that is free of θ. His formula is magical but often impossible to use in practice.

Example 2: curved exponential family

- Let $y_1, \ldots, y_n$ be iid sample from $N(\theta, \theta^2)$ for $\theta > 0$. First consider the confidence distribution based on y_1,

$$C_m(\theta; y_1) = P_\theta(Y_1 \geq y_1) = 1 - \Phi\left(\frac{y_1 - \theta}{\theta}\right).$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$. Then $C_m(\theta = \infty; y_1) = 1 - \Phi(-1) = 0.84 < 1$, so that y_1 does not give a valid confidence distribution. With whole data, the MLE is given by

$$\hat{\theta} = \hat{\theta}(y) = \frac{1}{2n}\left\{-\sum y_i + \sqrt{\left(\sum y_i\right)^2 + 4\sum y_i^2}\right\}$$

with a maximal ancillary $A(y) = \frac{\sum y_i}{\sqrt{\sum y_i^2}}$.

Example 2: curved exponential family

We compared several methods for the confidence density of θ.

- $c_c(\theta; \hat{\theta}|a)$ from the exact conditional density $f(\hat{\theta}|a)$:

$$c_c(\theta; \hat{\theta}|a) \propto \theta^{-1} L(\theta; y)$$

- Updating formula from $c_c(\theta; \hat{\theta}_i|a_i)$:

$$c_f(\theta; y) \propto c_c(\theta; \hat{\theta}_i|a_i) L(\theta; y_{(-i)}) \propto \theta^{-1} L(\theta; y)$$

– Updating formula from $c_m(\theta; \hat{\theta}_i)$:

$$c_{fi}(\theta; \hat{\theta}_i, y_{(-i)}) \propto c_m(\theta; \hat{\theta}_i) L(\theta; y_{(-i)}) \propto \theta^{-1} L(\theta; y) \frac{L(\theta; \hat{\theta}_i)}{L(\theta; y_i)}$$

+ Marginal confidence from the marginal density $f(\hat{\theta}_1, \cdots, \hat{\theta}_n)$:

$$c_m(\theta; \hat{\theta}_1, \cdots, \hat{\theta}_n) \propto \theta^{-1} L(\theta; y) \prod_{i=1}^{n} \frac{L(\theta; \hat{\theta}_i)}{L(\theta; y_i)}$$

Example 2: curved exponential family

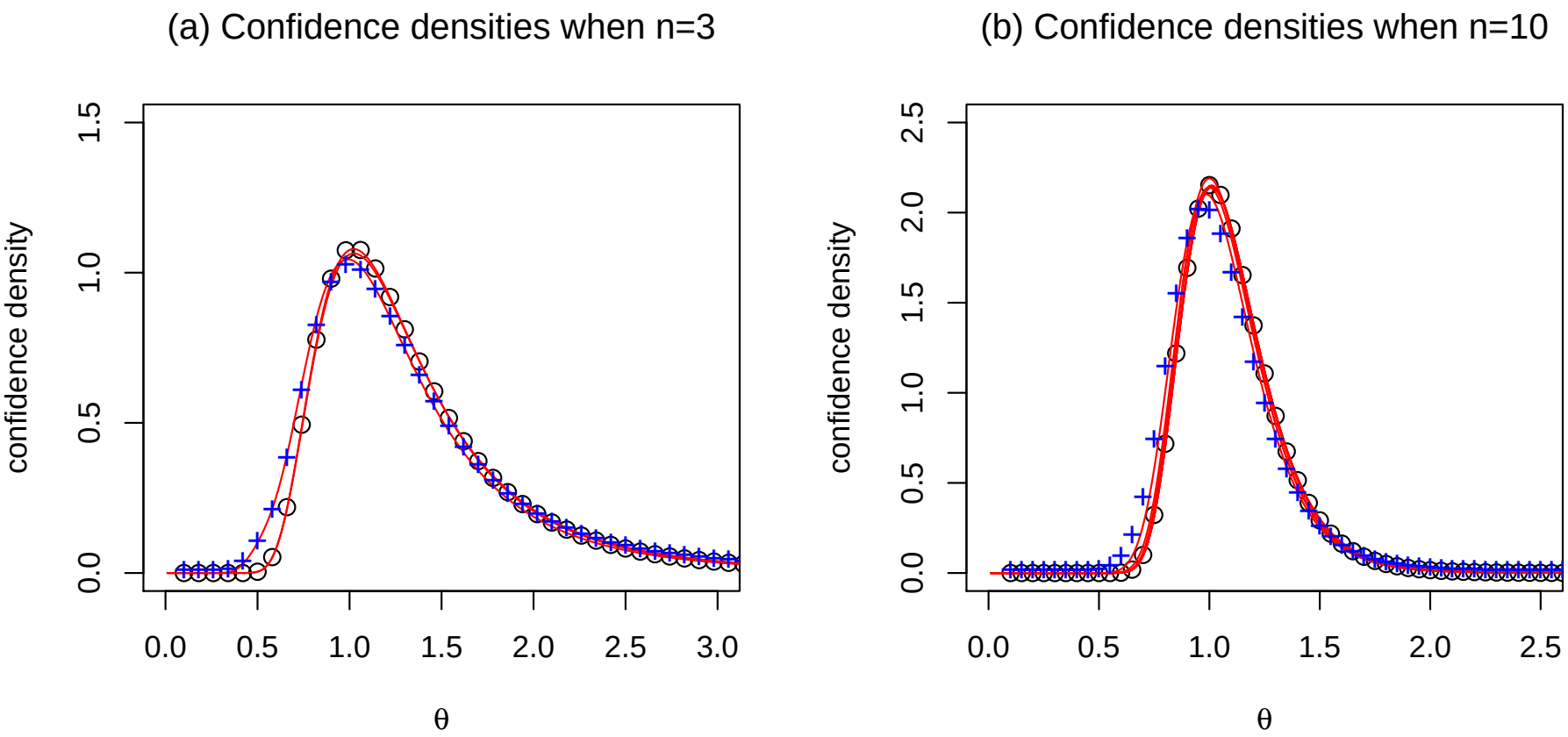


Figure: Confidence densities $c_f(\theta; y)$ (circle), $c_{fi}(\theta; y)$ (solid), $c_m(\theta; \hat{\theta}_1, \cdots, \hat{\theta}_n)$ (corss) for (a) $n = 3$, (b) $n = 10$.

Example 2: curved exponential family

Exact Finite Sample Inference !!!

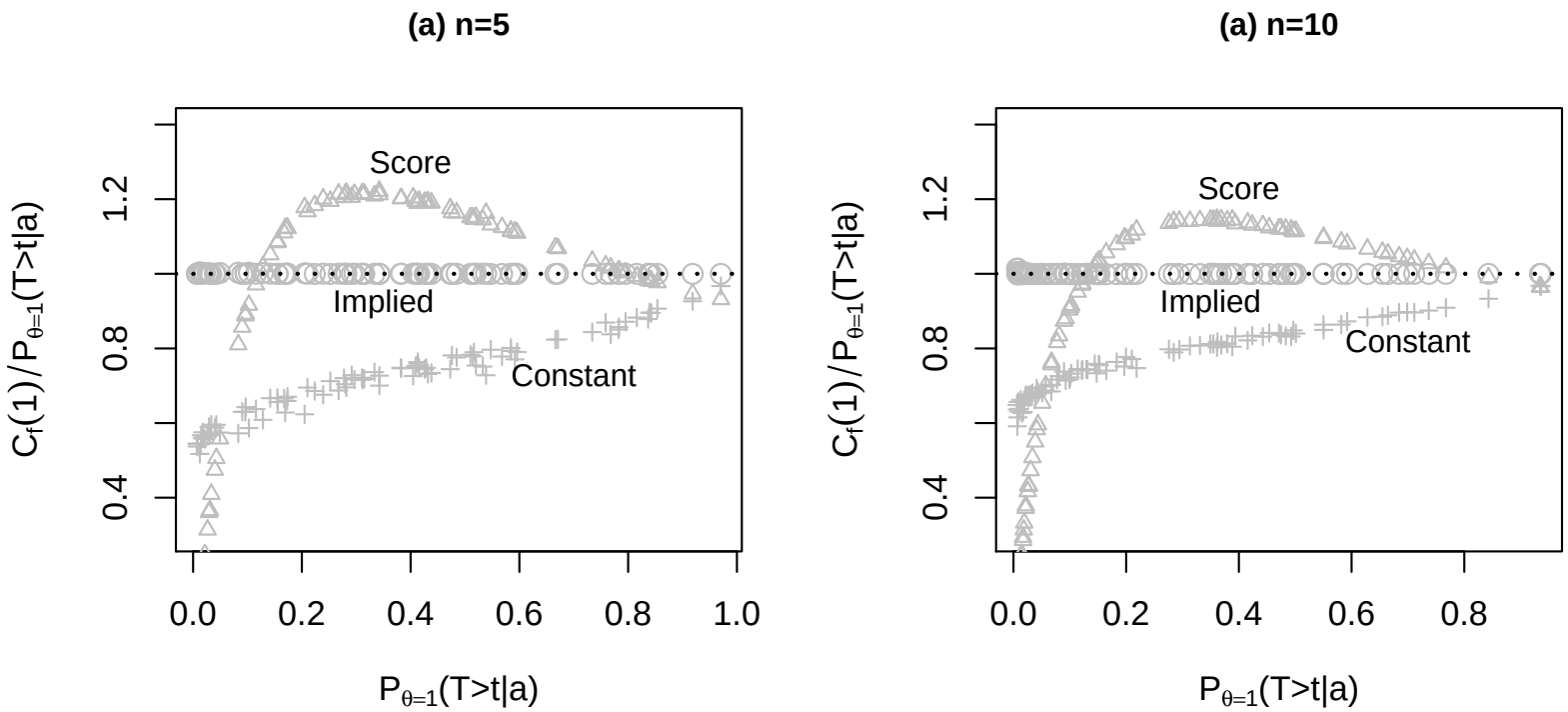


Figure: Conditional p-value of $H_0 : \theta = 1$ vs. confidence $C_f(1)$ using the implied prior (circle), constant prior (cross), corresponding p-value from the score test (triangle). To show the quality of approximation for small p-values, y-axis is expressed as a ratio.

Discussion

- We have described the market-linked Dutch Book argument to establish the epistemic confidence that is meaningful for the observed confidence interval.
- Fisher tried to achieve the same purpose with the fiducial probability, but the use of the word 'probability' generated much confusion and controversies, so the concept of fiducial probability has been practically abandoned. It is actually the extended likelihood of Lee and Nelder (1996).
- Our results show that we can turn a classical likelihood into a confidence density by multiplying it with an implied prior. Furthermore, we get epistemic confidence by establishing the absence of relevance subsets.

Extended Likelihood and H-likelihood

Considering follow three types of object.

Types of object	P-value, $\mathrm{P}_\theta\left(T_n \geq t\right)$, for confidence
$\theta \in \mathbb{R}^p$: unknown parameter	θ
$v \in \mathbb{R}^q$: unobservable random quantity	$v = T_n$
y : observed data	$y = t$

The extended likelihood is defined by

$$\begin{aligned} L_e\left(\theta, v\right) &= f_\theta\left(y|v\right) f_\theta(v) \\ &= f_\theta\left(y\right) f_\theta(v|y). \end{aligned}$$

Extended Likelihood and H-likelihood

Based on the **classical likelihood principle** of Birnbaum (1962), the marginal likelihood

$$L(\theta) = f_\theta(y)$$

carries all the information in the data about θ.

This means that the predictive distribution

$$L(\theta, v; v|y) = f_\theta(v|y)$$

should not carry any information about θ, because

$$L_e(\theta, v) = f_\theta(y) f_\theta(v|y).$$

Extended Likelihood and H-likelihood

Suppose there exists a function $v(\theta, y)$ such that

$$\begin{aligned} & L\left(\theta, v(\theta, y); v|y\right) = c. \\ \Leftrightarrow \quad & L_e\left(\theta, v(\theta, y)\right) \propto L(\theta), \end{aligned}$$

where c is free of θ, and $\propto$ means that both sides are equivalent in terms of θ.

Here, $v(\theta, y)$ is called **canonical function** of v if $v(\theta, y)$ satisfies

$$L_e(\theta, v(\theta, y)) \propto L(\theta).$$

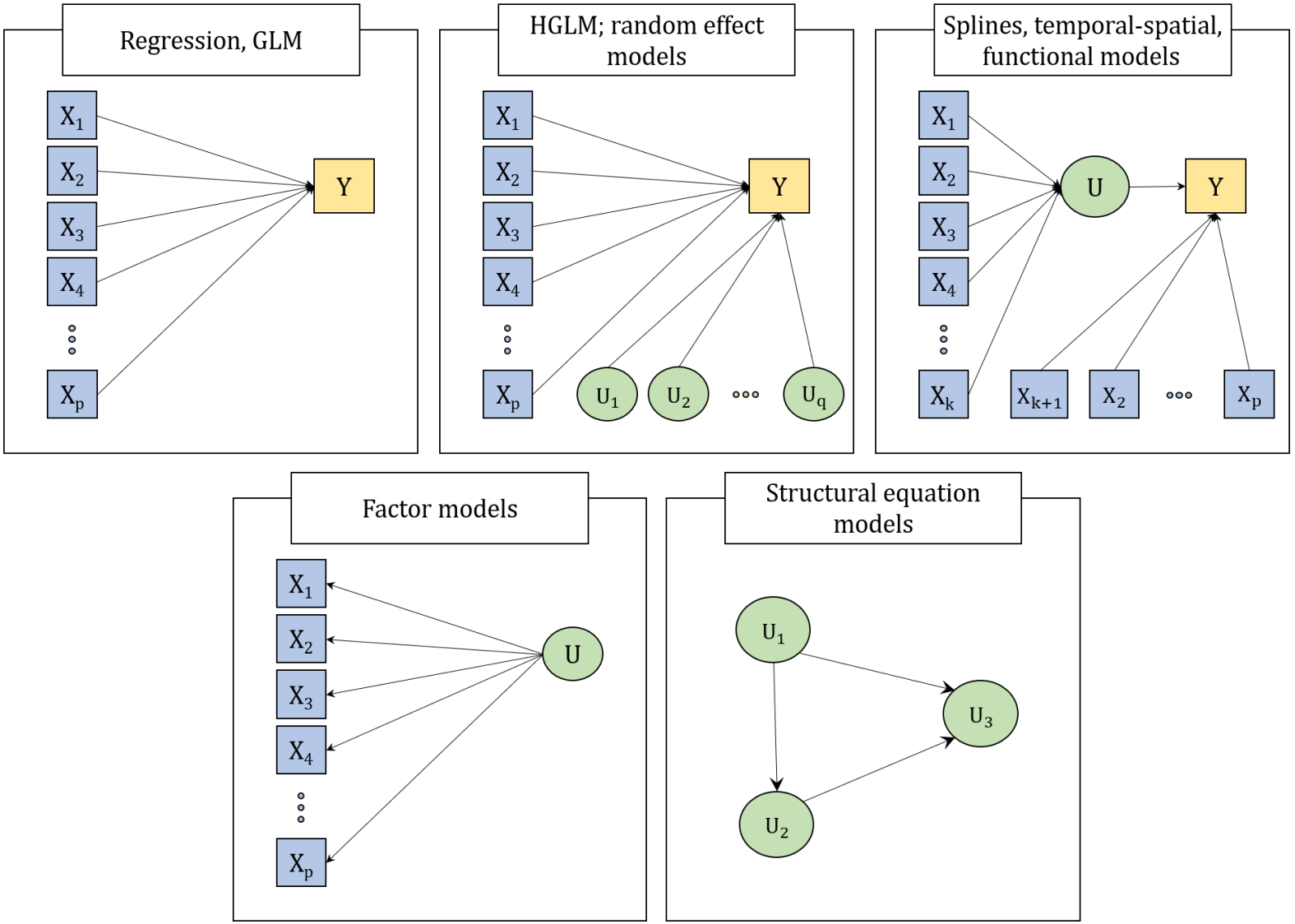
Extended Likelihood and H-likelihood

If we find an extended likelihood whose joint maximization of θ and v gives

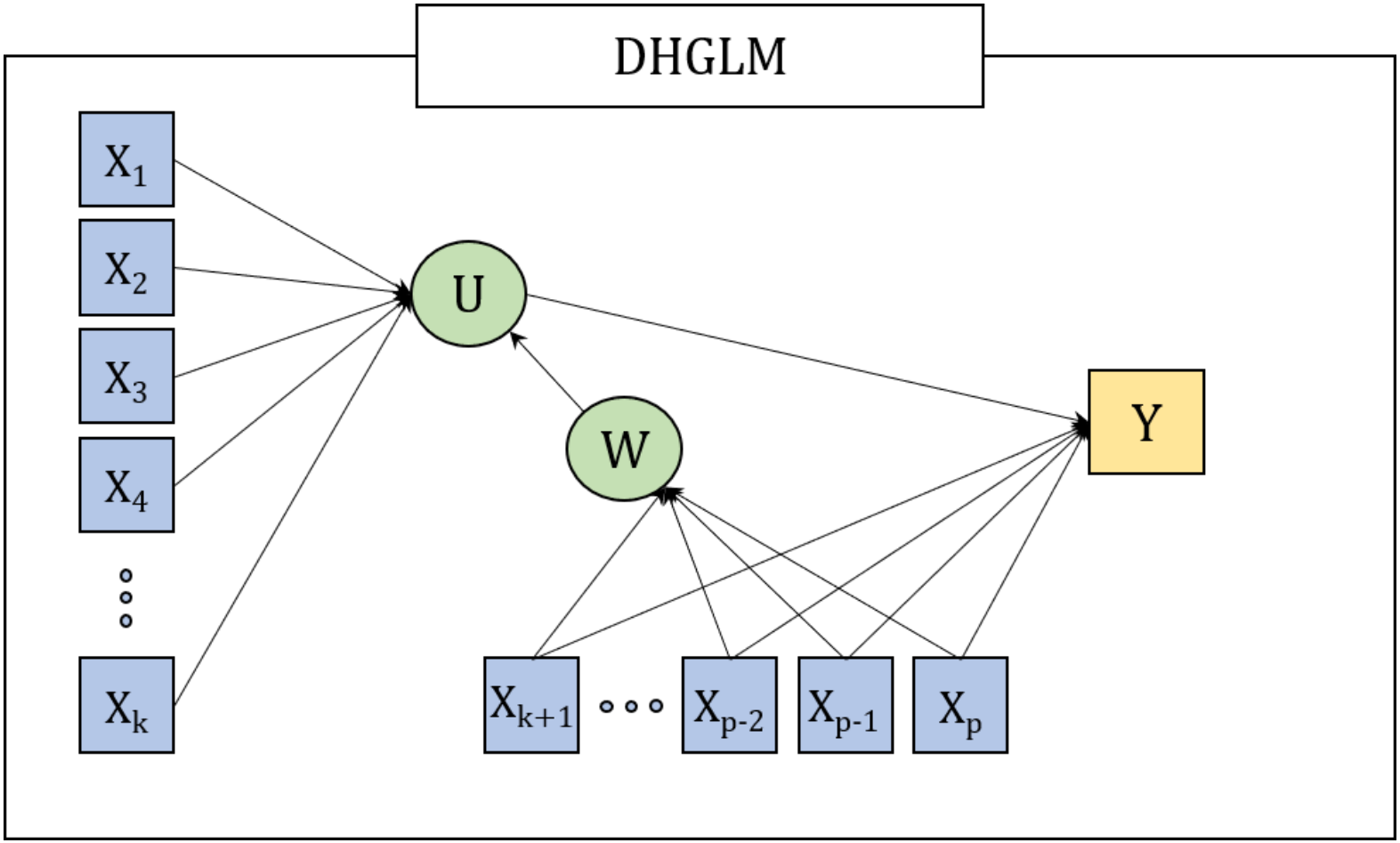
$$\begin{aligned}\hat{v} &= \arg\max_v h(\theta, v) = v(\hat{\theta}, y),\\ \hat{\theta} &= \arg\max_\theta h(\theta, v),\end{aligned}$$

Lee, Nelder and Pawitan (2017) defined such a v-scale as **canonical scale** where v is continuous and $H(\theta, v) = L_e(\theta, v)$ as **H-likelihood**. Joint optimization of h-likelihood provides proper likelihood inferences.

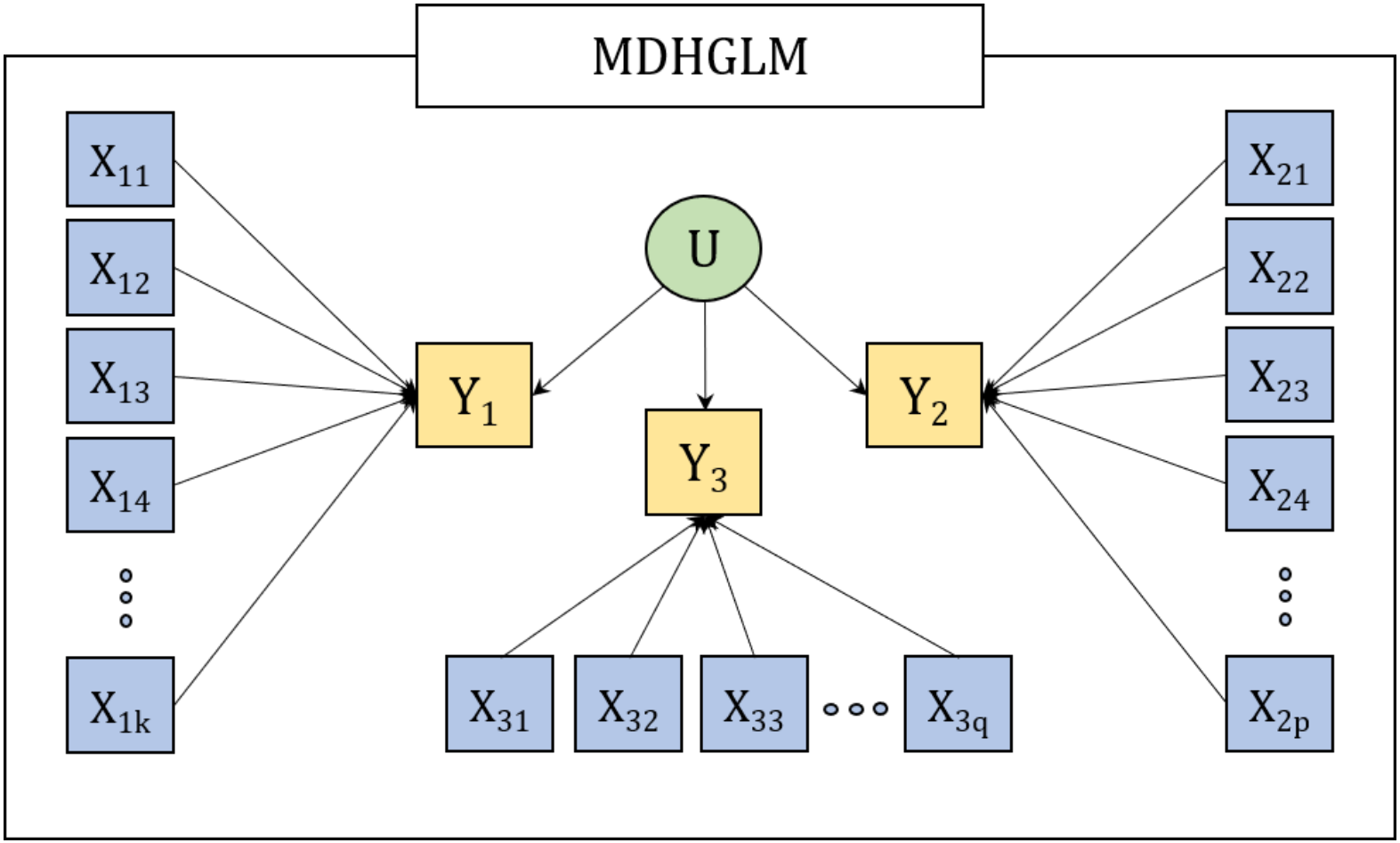
Various Models



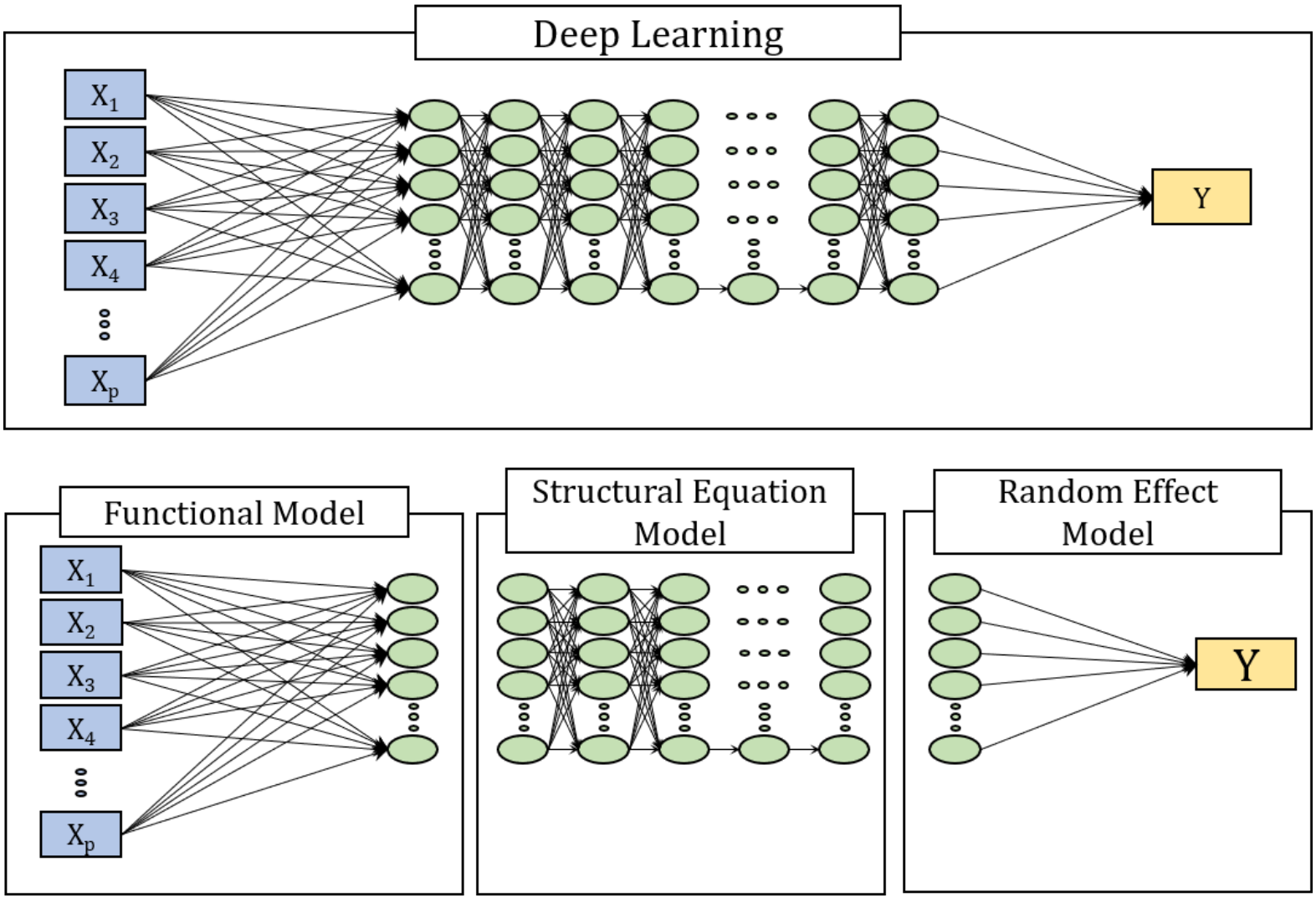
Various Models



Various Models



Various Models



Albatross Analytics

Web-based program for data analysis, including various random effect models: HGLM, DHGLM, frailty models, SEM, etc.

- `https://kicloud.ksc.re.kr/`