TERG

Discussion Paper No.462

An evolutionary explanation for why the prior beliefs can be assumed to coincide with the correct ones in the simple Bayesian hypothesis testings

by Mitsunobu Miyake

February 18, 2022

TOHOKU ECONOMICS RESEARCH GROUP

Discussion Paper

GRADUATE SCHOOL OF ECONOMICS AND
MANAGEMENT TOHOKU UNIVERSITY
27-1 KAWAUCHI, AOBA-KU, SENDAI,
980-8576 JAPAN

An evolutionary explanation for why the prior beliefs can be assumed to coincide with the correct ones in the simple Bayesian hypothesis testings

by Mitsunobu MIYAKE

Graduate School of Economics and Management

Tohoku University, Sendai 980-8576, Japan

mitsunobu.miyake.b4@tohoku.ac.jp

February, 2022

Abstract: This paper considers an evolutional process, where a simple Bayesian hypothesis testing is conducted repeatedly in parallel by many testing agents with diversified prior beliefs. Assuming that each of the agents selects a (one-shot) testing strategy to maximize the fitness value believing that the (own) prior belief is true, it is shown that only the testing agents endowed with the correct prior beliefs are survived eventually. This result provides an explanation for why the prior belief of the agent can be assumed to coincide with the correct one in the Bayesian hypothesis testing, as if the agent knows the true probability assigned by the nature, without introducing the long-term learning processes.

Key words: Bayesian hypothesis testing, diagnostic testing, prior belief, Neyman-Pearson lemma, natural selection, Malthusian competition.

1. Introduction

In Bayesian hypothesis testing for simple hypothesis,[1] the testing agent is not informed of the true probability on the events "the null hypothesis is true" or "the alternative hypothesis is true" assigned by the nature, but the agent has the personal prior beliefs (subjective probabilities) on the events. Although the agent can implement the Bayesian optimal testing procedure for a given prior belief, which maximizes the (posterior) expected value of the testing outcome conditioned on the signals, there is a problem of at what level should such a prior belief be determined.

One possible answer is that the prior belief coincides with the correct one, because the correct prior belief can be derived from past information in general, applying the statistical inference theory or Bayesian learning theory.[2] For example, before a test, if sufficiently many times of repetitions of the same test have been conducted, and if the prior belief has revised based on the signal at each repetition, it is possible that the sequence of revised prior beliefs converges to the correct one.

Without introducing such a long-term learning process, this paper atempts to provide an alternative explanation for why the prior belief of the agent can be assumed to coincide with the correct one, by showing that the testing agents with correct prior beliefs are naturally selected in an evolutional process, where a Bayesian testing is conducted

---

[1] For the hypothesis testings in the philosophical literature, see Sober (2008, Ch.1). Okasha (2013) and Lo and Zhang (2021) derive the Bayesian behavioral hypotheses (up-dating of prior and utility maximization based on posterior) as the properties of long-run equilibria of evolutionary processes in which non-Bayesian acts are permitted. Zhang (2013) shows that non-Bayesian behavioral hypotheses is derived when individual's risk aversion is incorporated into the processes. In this paper, the Bayesian behavioral hypotheses are major premises, and all agents are assumed to be Bayesian agents throughout this paper.

[2] For the statistical inference theory, see DeGroot and Schervish (2012, Ch 7). For the theories of determiniation of the prior beliefs, see Berger, Bernardo and Sun (2015) and the references.

repeatedly in parallel by many testing agents.[3]

First, we consider the one-shot Bayesian testing model with just one testing agent. Because the simple Bayesian hypothesis testing model above covers a class of the diagnostic testings, as in a standard textbook (Lesaffre and Lawson, 2012), we formulates the one-shot Bayesian testing as a diagnostic testing, where testing agent (medical doctor) conducts a testing with medical treatments for the patient randomly selected in a large population. For an informational assumptuion, we assume that the agent knows neither the true prevalence rate nor the true initial disease states of the patients even after the medical treatments. But we assume that a testing agent is informed of the final health state of the patient after the medical treatment.[4]

Concretely, for a given prior belief q, we introduce the concept of the Bayesian optimal test procedure with respect to q, which is defined by the test procedure maximi-

---

[3] As a explanation of such an assertion, our approach is closely related to the pragmatistic explanation why an expert acts as if the expert knows the seemingly unknowable information which is a crucial factor determining the outcome of the act. See Milton Friedman (1953, Section III) for example. Moreover, the result also provides a reason why the common prior assumption in the game theory holds for the testing agents.

[4] Consequently, in the evolutional model, this assumption implies that the Bayesian up-dating of belief is conducted within the one-shot testing and the agents cannot use the genetically retained up-dating process. If we assume the genetically retained up-dating process or the existence of the central authority which infer the (true) posterior probabilities, it is not difficult to derive the (correct) Bayesian testing procedure as shown in Footnote 8 in Section 3 of this paper. Moreover, in a one-shot testing, if the agent can derive a large sample, the agent inferes the true hypothesis, making use of the Bayesian updating or the limiting method of the relative frequency based on Reichenbach's principle. Then, we assume that each agent derives a small sample in the one-shot testing in this paper.

zing the subjective (posterior) probability of the event that the patient is healthy.[5] For any prior belief q, it is well-known that the subjectively optimal test procedure exists and that the Bayesian optimal test procedure is a likelihood ratio test procedure (Theorem 1(i,ii)).

Moreover, for the given true probability p, we introduce the concept of the objectively optimal test procedure with respect to p, which is defined by the test procedure maximizing the objective probability of the event that the patient is healthy, based on the true probability p. Assuming that, for each probability $q \in (0, 1)$, there exists a testing agent whose subjective prior belief coincides with q, we prove that the Bayesian optimal testing procedure with respect to q is objectively optimal if and only if the prior belief q coincides with the true probability p (Theorem 2). This implies that the hypothesis testing scheme has a nice property that a Bayesian optimal testing procedure is objectively optimal only if the agent's prior belief coincides with the correct one.

Second, we introduce the evolutional testing model, where the Bayesian testing is conducted repeatedly in parallel by many testing agents, and we assume that the agents have the diversified initial prior beliefs, which are individually preserved through the generations. Moreover, we assume that the fitness value of a medical doctor is determined by the individual achievement of the agent. Specifically, we assume that the fitness value coincides with the objective probability of the event that the patient is healthy. Then, it is shown that only the testing agents with correct beliefs about the probability (the prevalence rate) are survived eventually, if the population ratio of such rational agents is not neglibible at the initial state of the model (Theorem 3). The natural selection result is

---

[5] In a standard textbook as in DeGroot and Schervish (2012, Ch 9, Section 9.8), the cost of a treatment (or decision) is introduced in order to evaluate the testing procedure numerically. In this paper, we introduce the benefit of a treatment. There is no essential defference, because minimizing the cost is equivalent to maximizing the benefit as shown in Footnote 9 in Section 3 of this paper.

derived as a consequence of the Malthusian competition among the testing agents. Namely, even if the difference of the fitness values between the testing agents with correct beliefs and the other agents is very small, the selection will appear as a consequence of a mathematical property of the exponential function, which specifies the growth rate.[6] Note that Theorem 3 holds, independent of the natural rate of population growth $\delta > 0$, which is assumed to be common for all types of agents.

Because a Bayesian hypothesis testing can be regarded as a one-shot Bayesian hypothesis testing which appears after many generations have occured, the latter result (Theorem 3) provides an explanation for why the prior belief of the agent can be assumed to coincide with the correct one in a Bayesian hypothesis testing, as if the agent knows the true probability assigned by the nature.

The next section introduces the Bayesian hypothesis testing, and Section 3 introduces the Bayesian optimal test procedure and re-states its characterization theorem (Theorem 1). Section 4 evaluates the Bayesian optimal test procedures objectively and shows the main theorem of this paper (Theorem 2). The dynamic process is introduced and Theorem 3 is shown in Section 5.

2. The (one-shot) Bayesian hypothesis testing

This section introduces the (one-shot) Bayesian hypothesis testing model with a testing agent in a human health care setting. The model can be specified by the following conditions:

R1: The patient contracts tuberculosis (TB), or the patient suffers from a normal cold (NC). The nature determines one of the two events, TB or NC exclusively for all patients in a city. The true probability of TB (true prevalence rate) is denoted by p^*. We assume that

$$p^* \in (0, 1), \tag{1}$$

[6] See Nowak (2006, Ch.2, Section 2.2.1).

R2: For a medical doctor, a patient is selected randomly from a large population of patients in the city. The medical doctor does not know the true probability p^*. But each medical doctor has a subjective probability of the patient is TB (before the test), which we call a belief. The set of all possible beliefs are denoted by $P \equiv (0, 1)$. The two simple hypotheses are given by:

the null hypothesis (H_0): The paitient suffers from a normal cold, NC.

the alternative hypothesis (H_1): The patient contracts tuberculosis, TB.

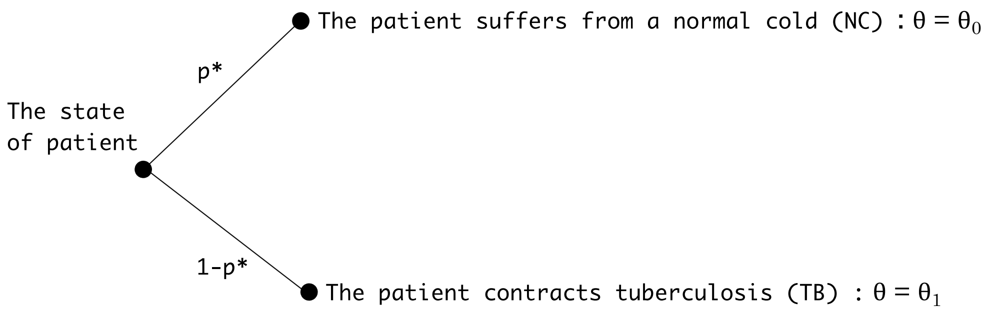

Figure 1

R3: The medical doctor examins the patient by a tuberculosis check-up kit. The check-up kit indicates a real number $x \in X \equiv \mathbb{R}$, which is assumed to be the sample mean for size m from the random device of which probability destribution is $f(x \mid \theta_0)$ in case that the paitient suffers from a normal cold ($\theta = \theta_0$). Otherwise ($\theta = \theta_1$), we assume that the check-up kit indicates a real number $x \in X$, which is assumed to be the sample mean for size m from the random device of which probability destribution is $f(x \mid \theta_1)$. We assume that:

$$f(x \mid \theta_0) \text{ and } f(x \mid \theta_1) \text{ are continuous functions of } x \in X, \tag{2}$$

and that

$$f(x \mid \theta_0) > 0 \text{ and } f(x \mid \theta_1) > 0 \text{ for all } x \in X, \tag{3}$$

R4: The medical doctor selects one of the two acts, "Not reject H_0 and prescribe streptomycin" or "Reject H_0 and prescribe antipyretics". The former act is denoted by d_0 and the latter act is denoted by d_1. The selection of act can be determined conditioned on the number in $X = \mathbb{R}$ given by R3, and the selection is represented by a function $\delta : X \to \{ d_0, d_1 \}$ such that $\delta^{-1}(d_0) \in \mathfrak{B}_X$, where $\mathfrak{B}_X$ is the σ-field of Borel subsets in X. In the following, a selection of act δ is called a testing procedure. Let Δ be the set of all testing procedures.

R5: If the event selected by the nature at R1 is consistent to the act $\delta(x)$ selected at R4, then there is no error, otherwise an error is happended. Concretely, there are two types of errors:

An error is happended if and only if $\delta(x) = d_1$, when $\theta = \theta_0$;

An error is happended if and only if $\delta(x) = d_0$, when $\theta = \theta_1$.

The former errors are called *Type I errors*, and the latter errors are called *Type II errors*. Although the doctor remenbers the act $\delta(x)$ at R4, the doctor is not informed of the true state, $(\theta = \theta_0)$ or $(\theta = \theta_1)$, and the doctor can not guess an error is happended or not.

R6: In case of the type I error is happended, the nature randomly determine the final outcome, "The paitient is healthy" or "The paitient is not healthy". The probability of "The paitient is healthy" is denoted by p_α. In case of the type II error is happended, the probability of "The paitient is healthy" is denoted by p_β. In case of no error is happended, the nature determine the final outcome "The paitient is healthy" certainly. We asume that

$$1 > p_\alpha > p_\beta > 0 . \tag{4}$$

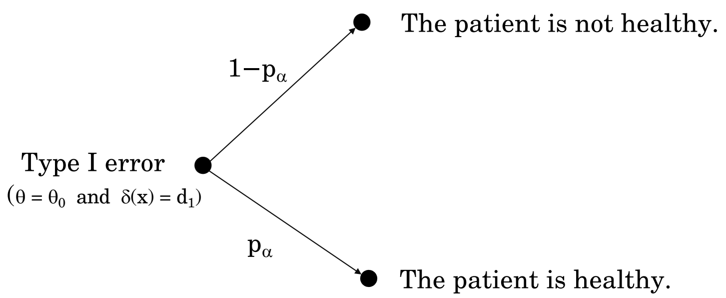

Figure 2

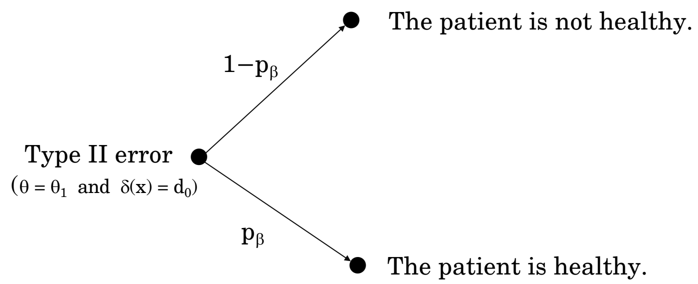

Figure 3

3. The Bayesian optimal testing procedures

In the testing model constructed in the previous section, this section introduces the Bayesian behavioral assumption for the agent to define the Bayesian optimal testing procedure. Namely, for a given prior belief, using the realization of the sample and its up-dating rule, the testing agent computes the posterior probabilities, and the agent selects the testing procudure to maximize the probability that the patient is healthy after the medical treatment.

We assume that the agent's prior belief for the occurrence of the event that a patient is TB $(\theta = \theta_1)$ is specified by a subjective probability q in P. For a given a prior belief $q \in P$, the posterior probability for $(\theta = \theta_i)$ after recieving the signal $x \in X$ is denoted by $P(\theta_i \mid x : q)$.[7] Using the Bayes theorem, we have the following well-known lemma:

[7] Setting $X_n = (x - (1/n), x + (1/n))$, we define $P(\theta_1 \mid x : q)$ by $P(\theta_1 \mid x : q) = \lim_n q \cdot P(X_n \mid \theta_1)/P(X_n)$, where $P(X_n \mid \theta_1) = \int_{X_n} f(x \mid \theta_1) d\mu$ and $P(X_n) = \int_{X_n} [q \cdot f(x \mid \theta_1) + (1 - q) \cdot f(x \mid \theta_0)] d\mu$. We can define $P(\theta_0 \mid x : q)$ by almost the same manner.

Lemma 1: $P(\theta_1 | x : q) = q\cdot f(x | \theta_1)/[q\cdot f(x | \theta_1) + (1-q)\cdot f(x | \theta_0)]$ and $P(\theta_0 | x : q) = (1-q)\cdot f(x | \theta_0)/[q\cdot f(x | \theta_1) + (1-q)\cdot f(x | \theta_0)]$.

	d_0	d_1
θ_0	1	p_α
θ_1	p_β	1

Figure 4

For a given prior belief $q \in P$, the posterior probability of "the paitient is healthy" in case of the medical treatment d_i (i=1, 2) after recieving the signal $x \in X$, denoted by $h(d_i | x: q)$, are defined by

$$h(d_1 | x: q) = P(\theta_0 | x: q)\cdot p_\alpha + P(\theta_1 | x: q)\cdot 1 \text{ and } h(d_0 | x: q) = P(\theta_0 | x: q)\cdot 1 + P(\theta_1 | x: q)\cdot p_\beta. \quad (5)$$

For a given prior belief $q \in P$, the probability of "the paitient is healthy" of a testing procedure $\delta \in \Delta$ after recieving the signal x can be denoted by $h(\delta(x) | x : q)$. As a direct consequence of Lemma 1 and (5), we have a lemma:

Lemma 2: (**i**) $h(d_1 | x: q) = [(1-q)\cdot f(x | \theta_0)\cdot p_\alpha + q\cdot f(x | \theta_1)]/[q\cdot f(x | \theta_1) + (1-q)\cdot f(x | \theta_0)]$ and $h(d_0 | x: q) = [(1-q)\cdot f(x | \theta_0) + q\cdot f(x | \theta_1)\cdot p_\beta]/[q\cdot f(x | \theta_1) + (1-q)\cdot f(x | \theta_0)]$.
(**ii**) For a given $q \in P$, $h(d_1 | x: q)$ and $h(d_0 | x: q)$ are continuous for $x \in X$.

The testing agent atempts to maximize the probabilities $h(\delta(x) | x : q)$ $(x \in X)$ by selecting the suitable testing procedure δ in Δ. Namely we can define the optimal testing procedure δ as follows: For a given prior belief $q \in P$, a test procedure $\delta \in \Delta$ is defined to be a *Bayesian optimal test procedure with respect to* q if and only if the probability of "The paitient is healthy", $h(\delta(x) | x : q)$ is maximal in the set $\{ h(d_1 | x : q), h(d_0 | x : q) \}$ for almost all possible $x \in X$.[8] Namely, it holds that

$$h(\delta(x) \mid x : q) = \mathbf{max}[\, h(d_1 \mid x : q), h(d_0 \mid x : q)] \text{ for almost all } x \in X.^{8}$$

The optimality condition above specifies a property of a test procedure only, and there is no condition for relative advantage over the other test procedures.

For the next theorem, we need some definitions and lemmas. For a given prior belief $q \in P$, if a testing agent selects a testing procedure $\delta \in \Delta$, the induced (subjective) probability of "The paitient is healthy", denoted by $G(\delta\colon q)$, is defined by

$$G(\delta\colon q) = (1-q)\cdot(1-\alpha(\delta))\cdot 1 + (1-q)\cdot\alpha(\delta)\cdot p_\alpha + q\cdot(1-\beta(\delta))\cdot 1 + q\cdot\beta(\delta)\cdot p_\beta, \qquad (6)$$

where $\alpha(\delta)$ is the probability of Type I errors of δ defined by $\alpha(\delta) = \int_{\delta^{-1}(d_1)} f(x|\theta_0)d\mu$ and $\beta(\delta)$ is the probability of Type II errors of δ defined by $\beta(\delta) = \int_{\delta^{-1}(d_0)} f(x|\theta_1)d\mu$.[9]

---

[8] When $A \in \mathfrak{B}_X$ and $\mu(A) > 0$, a statement of x holds for *almost* all $x \in A$ if and only if there exists $B \subset A$ such that: (i) $B \in \mathfrak{B}_X$, (ii) $\mu(B) = \mu(A)$, (iii) The statement of x holds for *all* $x \in B$. The function $w(x) = \mathbf{max}[h(d_1 \mid x : q), h(d_0 \mid x : q)]$ is Borel measurable, because $w(x)$ is continuous on X by Lemma 2(ii) and Royden (1988, Problems 4 and 5 in Page 34, and Problem 44 in Page 49). Suppose that there exists a statistician who can collect the information from the many agents such as (x, d, h), where x is a realization of the sample, d is the decision, and h is the final health condition of the patient. If the statistician gets sufficiently many such information, then the statistician can infer the probabilities $h(d_1 \mid x : p^*)$ and $h(d_0 \mid x : p^*)$, where p^* is the true probability, and then derives the Bayesian optimal test procedure with respect to p^* without knowing the true probabilities $[f(x \mid \theta_0), f(x \mid \theta_1), p_\alpha, p_\beta, p^*]$.

[9] The numerical value of $G(\delta\colon q)$ can be interpreted as the values of the expected utilities defined by $G(\delta\colon q)\cdot$(utility of healthy state) + $(1 - G(\delta\colon q))\cdot$(utility of not healthy state), because the values coincide with $G(\delta\colon q)$, if we set (utility of healthy state) = 1 and (utility of not healthy state) = 0. Moreover, because $G(\delta\colon q) = (1-q)\cdot(1-\alpha(\delta))\cdot 1 + (1-q)\cdot\alpha(\delta)\cdot p_\alpha + q\cdot(1-\beta(\delta))\cdot 1 + q\cdot\beta(\delta)\cdot p_\beta = 1 - [(1-q)\cdot(1-p_\alpha)\cdot\alpha(\delta) + q\cdot(1-p_\beta)\cdot\beta(\delta)]$, the maximization of $G(\delta\colon q)$ is equivalent to the minimization of the cost $C(\delta\colon q) \equiv [(1-q)\cdot(1-p_\alpha)\cdot\alpha(\delta) + q\cdot(1-p_\beta)\cdot\beta(\delta)]$.

Lemma 3: (**i**) For all $q \in P$ and all $\delta^1, \delta^2 \in \Delta_q$, it holds that $G(\delta^1: q) = G(\delta^2: q)$.

(**ii**) $G(\delta: q) = \int_X h(\delta(x) \mid x: q) d\lambda_q$, where λ_q is the measure on $(X, \mathfrak{B}_X)$ defined by $\lambda_q(B) = \int_B [q \cdot f(x \mid \theta_1) + (1 - q) \cdot f(x \mid \theta_0)] d\mu$ for all $B \in \mathfrak{B}_X$.

For a given prior belief $q \in P$, a test procedure $\delta^* \in \Delta$ is called a *subjectively optimal test procedure with respect to* q if and only if $G(\delta^*: q) \geq G(\delta: q)$ for all $\delta \in \Delta$. For a given prior belief $q \in P$, a test procedure $\delta^* \in \Delta$ is defined to be a *likelihood ratio test procedure with respect to* q if and only if

$$\delta^*(x) = d_1 \text{ for almost all } x \in C^1(q); \quad \delta^*(x) = d_0 \text{ for almost all } x \in C^0(q),$$

where $C^1(q) = \{ x \in X : f(x \mid \theta_1)/f(x \mid \theta_0) > (1 - q) \cdot (1 - p_\alpha)/q \cdot (1 - p_\beta) \}$ and $C^0(q) = \{ x \in X : f(x \mid \theta_1)/f(x \mid \theta_0) < (1 - q) \cdot (1 - p_\alpha)/q \cdot (1 - p_\beta) \}$. Let Δ_q be the set of all likelihood ratio test procedures with respect to $q \in P$.

Lemma 4: For any prior belief $q \in P$, the following assertions hold:

(**i**) $\Delta_q \neq \emptyset$.

(**ii**) If $\delta^1 \in \Delta_q$ and $\delta^2 \in \Delta$, then $h(\delta^1(x) \mid x: q) \geq h(\delta^2(x) \mid x: q)$ for almost all $x \in X$.

(**iii**) If $\delta^1 \in \Delta_q$ and $\delta^2 \in \Delta_q$, then $h(\delta^1(x) \mid x: q) = h(\delta^2(x) \mid x: q)$ for almost all $x \in X$.

(**iv**) Suppose that $\delta^1 \in \Delta_q$. If $\delta^2 \in \Delta$ satisfies $\delta^2 \notin \Delta_q$, then there exists a compact subset A in X such that $\mu(A) > 0$ and $h(\delta^1(x) \mid x: q) > h(\delta^2(x) \mid x: q)$ for all $x \in A$.

(**v**) If $\delta^1 \in \Delta_q$ and $\delta^2 \notin \Delta_q$, then $G(\delta^1: q) > G(\delta^2: q)$.

We have the following well-known theorem[10]:

[10] Theorem 1(ii)[(b)⇔(c)] is a re-statement of Neyman-Pearson lemma as in DeGroot and Schervish (2012, Ch.9, Theorem 9.2.1) and Lehmann and Romano (2005, Ch.3). For Theorem 1(ii)[(a)⇔(c)], see DeGroot and Schervish (2012, Ch.9, Section 9.8).

Theorem 1(Existence and characterization of the Bayesian optimal test procedures): Select any $q \in P$ and let q be a subjective belief of a testing agent.

(**i**) There exists a Bayesian optimal test procedure with respect to q.

(**ii**) The following three statements are mutually equivalent:

(**a**) A test procedure $\delta \in \Delta$ is Bayesian optimal with respect to q.

(**b**) A test procedure $\delta \in \Delta$ is subjectively optimal with respect to q.

(**c**) A test procedure $\delta \in \Delta$ is a likelihood ratio test procedure with respect to q.

(**iii**) Let δ^* be a Bayesian optimal test procedure with respect to q. If a test procedure δ is not a Bayesian optimal test procedure with respect to q, then $G(\delta^*: q) > G(\delta: q)$.

Although Theorem 1 is well-known, we provide a proof of the theorem for the completeness of the arguments of this paper:

Proof of Theorem 1: (**ii**) First, we prove the assertion (ii) of Theorem 1. [(**a**) ⇒ (**c**)]: Because $\Delta_q \neq \emptyset$ by Lemma 4(i), we can assume that there exists $\delta^* \in \Delta_q$. If $\delta^0 \in \Delta$ is not a likelihood ratio test procedure with respect to q, then it holds by Lemma 4(iv) that there exists a compact subset A in X such that $\mu(A) > 0$ and $h(\delta^*(x) \mid x : q) > h(\delta^0(x) \mid x : q)$ for all $x \in A$, which implies that $\delta^0 \in \Delta$ is not a Bayesian optimal test procedure with respect to q. Taking the contraposition of this, we have that a Bayesian optimal test procedure with respect to q is a likelihood ratio test procedure with respect to q. [(**c**) ⇒ (**b**)]: Let δ^* be a likelihood ratio test procedure with respect to q. Suppose that δ^* is not subjectively optimal. Then there exists $\delta^0 \in \Delta$ and there exists a compact subset A in X such that $\mu(A) > 0$ and $h(\delta^0(x) \mid x : q) > h(\delta^*(x) \mid x : q)$ for all $x \in A$. However, it holds by Lemma 4(ii) that $h(\delta^*(x) \mid x : q) \geq h(\delta(x) \mid x : q)$ for almost all $x \in A$. This is a contradiction, and we have that δ^* is subjectively optimal. [(**b**) ⇒ (**c**)]: Because $\Delta_q \neq \emptyset$ by Lemma 4(i), we can assume that there exists $\delta^* \in \Delta_q$. If $\delta \in \Delta$ is not a likelihood ratio test procedure with respect to q, then it holds by Lemma 4(v) that $G(\delta^*: q) > G(\delta: q)$, which implies that δ is not subjectively optimal with respect to q. Taking the contraposition of this, we have that a subjectively optimal test procedure with respect to q is a likelihood ratio test procedure with respect to q. [(**c**) ⇒ (**a**)]: Let δ^* be a likelihood ratio test procedure with respect to q. For any $\delta \in \Delta$, it holds by Lemma 4(ii) that $h(\delta^*(x) \mid x : q) \geq h(\delta(x) \mid x : q)$ for almost all $x \in X$, which implies that δ^* is a Bayesian optimal test procedure with respect to q. (**iii**) Theorem 1(iii) is a direct consequence of Lemma 4(v) and Theorem 1(ii)[(**a**) ⇔ (**c**)]. (**i**) Theorem 1(i) is a direct consequence of Theorem 1(ii) and Lemma 4(i). □

For any prior belief $q \in P$, the existence of a Bayesian optimal test procedure with respect to q is ensured by Theorem 1(i). The optimality concept of (**a**) is defined by the best responces against the possible signals, whereas the optimality concept of (**b**) is defined by the comparisons of the values directly on the set of all test procedures. In terms of the game theory, the former corresponds to the optimality in the behavioral strategies in the extenstive form games, the latter corresponds to the optimality in the pure strategies in the strategic form games. Theorem 1(ii)[(**a**) ⇔ (**b**)] shows that such two optimality concepts are equivalent in the Bayesian testing scheme. Theorem 1(iii) strengthens the optimality condition by showing that the value of G at the Bayesian optimal test procedure is *strictly* greater than the values of G at the non-optimal test procedures.

Example 1: Set $p^* = 0.4$. We assume that the real number $x \in X$ is the sample mean for size m from $N(0, 1)$ in case that the paitient suffers from a normal cold $(\theta = \theta_0)$. Otherwise $(\theta = \theta_1)$, we assume that $x \in \mathbb{R}$ is the sample mean for size m from $N(1, 1)$. For simplicity, if $\theta = \theta_0$, we assume that the emergence of x is subject to the nomal distribution $N(0, 1/m)$ (Note that $1/m$ is the variance), otherwise $(\theta = \theta_1)$, we assume that the emergence of x is subject to the nomal distribution $N(1, 1/m)$. Cosequently, the two conditional probability destributions, $f(x \mid \theta_0)$ and $f(x \mid \theta_1)$ can be defined by

$f(x\,|\,\theta_0) = A\cdot exp[-(½)\cdot m\cdot x^2]$ and $f(x\,|\,\theta_1) = A\cdot exp[-(½)\cdot m\cdot(x-1)^2]$ for all $x \in X$,

where $A = m^{½}\cdot(2\pi)^{-½}$ and $exp(y) = e^y$ for all $y \in X$. When $m = 2$, setting $A^* = \pi^{-½}$, the graph of $f(x\,|\,\theta_0) = A^*\cdot exp[-x^2]$ and $f(x\,|\,\theta_1) = A^*\cdot exp[-(x-1)^2]$ can be drawn as follows:

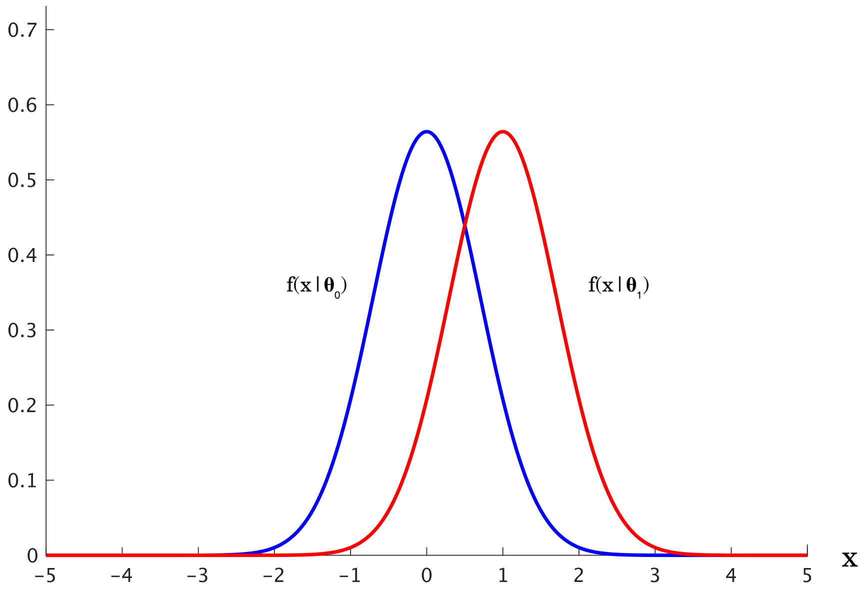

Figure 5

Lemma 5: (**i**) $L(x) \equiv f(x\,|\,\theta_1)/f(x\,|\,\theta_0) = exp[m\cdot(x-(½))]$ is an increasing function of x.

(**ii**) $c(q) \equiv L^{-1}((1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta))$ is a decreasing function of q.

(**iii**) $\lim_{q\to 0} c(q) = +\infty$ and $\lim_{q\to 1} c(q) = -\infty$.

(**iv**) The likelihood ratio test procedure with respect to q can be re-defined by

$\delta^*(x) = d_1$ for *almost* all $x \in \{ x \in \mathbb{R} : x > c(p) \}$;

$\delta^*(x) = d_0$ for *almost* all $x \in \{ x \in \mathbb{R} : x < c(p) \}$.

For all $q \in P$, $c(q)$ is called the *critical value* of the likelihood ratio test procedure with respect to q. When $m = 2$, $p_\alpha = 0.3$, $p_\beta = 0.1$, $q = q(\theta_1) = 0.2$, $q(\theta_0) = 0.8$, it holds that:

$c(q) = c(0.2) = (1/2)\cdot log(4) + (1/2)\cdot log(7/9) + (1/2) = 1.0675,$

$L(x) \equiv f(x\,|\,\theta_1)/f(x\,|\,\theta_0) = A^*\cdot exp[-(x-1)^2]/A^*\cdot exp[-x^2] = exp(2x-1)$

$P(\theta_0 \mid x : q) = (1 - q)\cdot f(x \mid \theta_0)/[\, q\cdot f(x \mid \theta_1) + (1 - q)\cdot f(x \mid \theta_0)] = 4/\,[L(x) + 4]$,

$P(\theta_1 \mid x : q) = q\cdot f(x \mid \theta_1)/[\, q\cdot f(x \mid \theta_1) + (1 - q)\cdot f(x \mid \theta_0)] = 1/[\,1 + 4(1/L(x)] = L(x)/[L(x)+4]$,

$h(d_1 \mid x: q) = P(\theta_0 \mid x: q)\cdot p_\alpha + P(\theta_1 \mid x: q)\cdot 1 = [L(x)+(1.2)]\,/[L(x) + 4]$,

$h(d_0 \mid x: q) = P(\theta_0 \mid x: q)\cdot 1 + P(\theta_1 \mid x: q)\cdot p_\beta = [(0.1)L(x)+4]\,/[L(x) + 4]$.

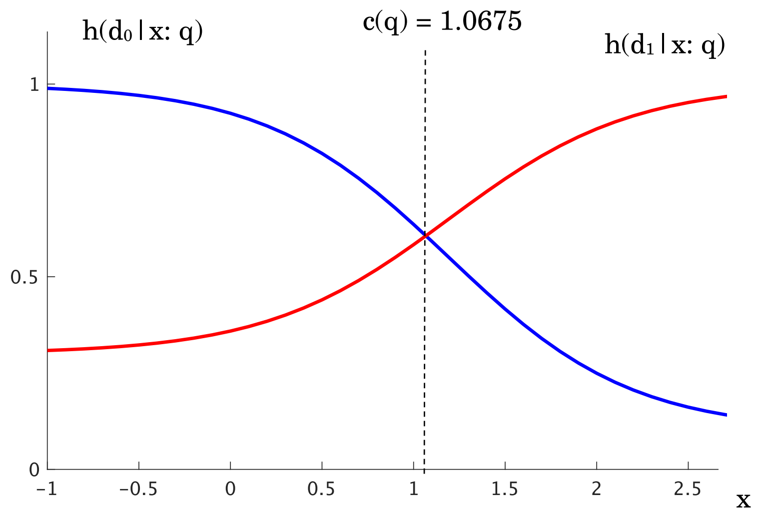

Figure 6

4. The objective evaluation of the Bayesian optimal testing procedures

It holds by Theorem 1 that the testing agent selects a likelihood ratio test procedure in Δ_q, where q is the prior belief of the agent. This section evaluates the selected test procedures objectively and proves that the selected test procedure is objectively optimal if and only if the testing agent's prior belief coincides with the true probability.

Suppose that $p^* \in (0, 1)$ is the true probability (true prevalence rate). We assume that:

A1: For any $y \in \mathbb{R}_{++}$, there exists some $x \in \mathbb{R}$ such that $y = f(x \mid \theta_1)/f(x \mid \theta_0)$.

The condition **A1** means the range of $L(x) \equiv f(x \mid \theta_1)/f(x \mid \theta_0)$ coincides with the full set $\mathbb{R}_{++}$.

If a testing agent selects a testing procedure $\delta \in \Delta$, the induced probability of "The paitient is healthy", denoted by $G(\delta: p^*)$, can be defined by

$$G(\delta: p^*) = (1 - p^*)\cdot(1 - \alpha(\delta))\cdot 1 + (1 - p^*)\cdot\alpha(\delta)\cdot p_\alpha + p^*\cdot(1 - \beta(\delta))\cdot 1 + p^*\cdot\beta(\delta)\cdot p_\beta . \qquad (7)$$

Then we have a lemma:

Lemma 6: (**i**) For all $q \in P$ and all $\delta^1, \delta^2 \in \Delta_q$, it holds that $G(\delta^1: p^*) = G(\delta^2: p^*)$.

(**ii**) $G(\delta: p^*) = \int_X h(\delta(x) \mid x : p^*)d\lambda$, where λ is the measure on $(X, \mathfrak{B}_X)$ defined by $\lambda(B) = \int_B [p^*\cdot f(x \mid \theta_1) + (1 - p^*)\cdot f(x \mid \theta_0)]d\mu$ for all $B \in \mathfrak{B}_X$.

(**iii**) $\Delta_q \cap \Delta_r = \emptyset$ for all $q, r \in P$ such that $q \neq r$.

We assume that:

> **A2**: For each $q \in P$, there exists a continuum of agents N_q ($\mu(N_q) > 0$) such that if $i \in N_q$ then i's prior belief coincides with q.

Because the value $G(\delta: p^*)$ is independent of the selection of δ in Δ_q as shown by Lemma 6(i), we can define a function $F(\cdot, p^*)$ on P by

$$F(q: p^*) = G(\delta: p^*) \text{ for some } \delta \in \Delta_q. \qquad (8)$$

The function $F(\cdot, p^*)$ need not to be an injection on P. See Example 1 (Figure 7) at the end of this section. The value of $F(q: p^*)$ can be interpreted as the the population ratio of "The paitient is healthy", when all agents in N_q take the test procedures in $\Delta(q)$. Hence the resulting measure $\mu(M_q)$ is given by

$$\mu(M_q) = \mu(N_q) \cdot F(q: p^*). \qquad (9)$$

In the following of this paper, Δ_q is called a *testing strategy corresponding to* q. Note that $\Delta_q \cap \Delta_r = \emptyset$ for all $q, r \in P$ such that $q \neq r$. (Lemma 6(iii)). A testing strategy Δ_q is *objectively optimal with respect to* p^* if and only if $F(q: p^*) \geq F(r: p^*)$ for all $r \in P$. As a main result of this paper, we have the following theorem:

Theorem 2(Existence and characterization of the objectively optimal testing strategy):

(**i**) There exists uniquely an objectively optimal testing strategy with respect to p^*.

(**ii**) The following three statements for a prior belief $q \in P$ are mutually equivalent:

(**a**) A testing strategy Δ_q corresponding to q is objectively optimal with respect to p^*.

(**b**) If $\delta^* \in \Delta_q$ and $\delta \notin \Delta_q$, then $G(\delta^*: p^*) > G(\delta: p^*)$.

(**c**) The prior belief q coincides with p^*.

Proof of Theorem 2: (**ii**) First, we prove the assertion (ii) of Theorem 2. [(**a**) ⇒ (**c**)]: Suppose that (**a**) and $q \neq p^*$ hold. If $\delta \in \Delta_q$, it holds by Lemma 6(iii) that $\delta \notin \Delta_{p^*}$. Fix any $\delta^* \in \Delta_{p^*}$. It holds by Theorem 1(iii) that $G(\delta^*: p^*) > G(\delta: p^*)$, which implies that $F(p^*: p^*) > F(q: p^*)$. This contradicts with (**a**). Hence we have that [(**a**) ⇒ (**c**)]. [(**c**) ⇒ (**b**)]: Suppose that $\delta^* \in \Delta_{p^*}$ and $\delta \notin \Delta_{p^*}$. It holds by Theorem 1(iii) and Lemma 6(ii) that $G(\delta^*: p^*) > G(\delta: p^*)$. [(**b**) ⇒ (**a**)]: Let $r \in P$ be a prior belief such that $r \neq q$. If $\delta \in \Delta_r$, then it holds by Lemma 6(iii) that $\delta \notin \Delta_q$. Fix any $\delta^* \in \Delta_q$. We have by (**b**) that $G(\delta^*: p^*) > G(\delta: p^*)$, which implies that $F(q: p^*) > F(r: p^*)$. (**i**) Theorem 2(i) holds by Theorem 2(ii)[(**b**) ⇒ (**a**)] and Lemma 4(i) . □

For a given true probability $p^* \in P$, Theorem 2(i) ensures that there exists uniquely an objectively optimal testing strategy with respect to p^*. Theorem 2(ii)[(**a**) ⇔ (**c**)] implies that a subjectively optimal testing procedure with respect to a prior belief q is objectively optimal only if the prior belief q coincides with the true probability p^*. Namely it holds that $F(p^*: p^*) > F(q: p^*)$ for all $q \neq p^*$. The objective optimality condition selects a specific strategy Δ_q among the class of all testing strategies $\{ \Delta_q : q \in P \}$, whereas the condition (**b**) in Theorem 2(ii) selects a specific strategy Δ_q such that each test procedure in Δ_q dominates all of test procedures in $(\Delta_q)^c = \Delta/\Delta_q$, which is an optimality concept in Δ. Theorem 2(ii) [(**a**) ⇔ (**b**)] shows that such two optimality concepts are equivalent in the testing scheme.

Example 1(Continue): Using Lemma 5, we can draw the graph of the function F(q: p*), when $m = 2$, $p^* = 0.4$, $p_\alpha = 0.3$, $p_\beta = 0.1$.

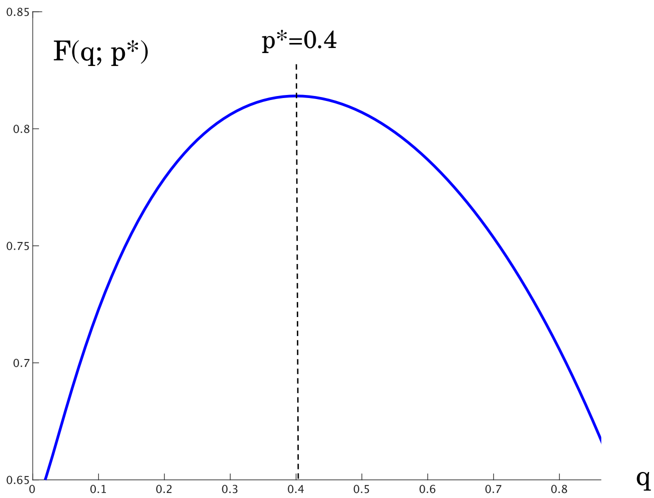

Figure 7

5. An dynamic process for selecting the testing agents with correct prior beliefs

This section introduces an evolutional testing model, where the Bayesian testing is conducted repeatedly in parallel by many testing agents, and it is shown that only the testing agents with correct prior belief are survived eventually.

(**Time structure**): The time is discrete and infinite, and it is denoted by $t = 1, 2, \cdots$.

(**Initial condition**): Let $\mathfrak{B}_P$ be the σ-field of Borel subsets in $P = (0, 1)$. Without specifying the set of testing agents, we assume that the population distribution on the set of testing strategies at $t = 1$ is represented by a discrete measure β on $(P, \mathfrak{B}_P)$, where a *discrete* measure on $(P, \mathfrak{B}_P)$ is a measure π on $(P, \mathfrak{B}_P)$ such that there exists a finite subset Q of P such that $\pi(Q) = \pi(P)$. In the following, a discrete measure π on $(P, \mathfrak{B}_P)$ is called a *population distribution on P*.

(**Transition rule**): At the biginning of period t, for each medical doctor who are active at t, a patient is selected randomly from a large population of patients. The patient is tested, and the medical doctor provides the medical treatment depending on the test result (signal) as specified by R3. If a patient's final health condition is good, the medical doctor can leave an offspring, otherwise the doctor can not leave an offspring. This assumption means that the fitness value of the strategy $q \in P$ is assumed to coincide with the value $F(q; p^*)$. The offspring's prior belief coincides with the pearent's prior belief. At the end of period *t*, all the medical doctors enter retirement, and the offsprings appear. We assume that the measure of each type of offsprings is increased by the natural rate of population growth $\gamma_t > 0$. All the offsprings will become active medical doctors in the next period *t+1*. A sequence of strategy distributions $\{ \pi_t \}$ is called a *dynamic process*. Then we have the following theorem:

Theorem 3: Let p^* be the true probability (true prevalence rate), and let β be the initial population distribution on P. Suppose that all of the testing agents satisfies the Bayesian behavioral assumption. The above rules determine uniquely a dynamic process and the process is given by a history $\{ \pi_t \}_{t=1}^{\infty}$ such that:
(**i**) $\pi_1 = \beta$, and (**ii**) $\pi_{t+1}(\{q\}) = (1+\gamma_t)\cdot F(q; p^*)\cdot\pi_t(\{q\})$ for all $q \in \text{supp}(\beta)$ and all $t = 1, 2, 3, \cdots$. We suppose additionally that $p^* \in \text{supp}(\beta)$. Then $\lim_{t\to+\infty} \pi_t(\{ q \})/\pi_t(P) = 1$ holds if and only if $q = p^*$.

Note that the convergence is independent of the value of δ, because it is neutral. Hence γ_t can be variable over time. For example, $\gamma_t = sin(t)$.

Proof of Theorem 3: It holds by Theorem 1 that the initial distribution π coincides with the distribution β. For all $q \in \text{supp}(\beta)$, it holds by (ii) that $\pi_t(\{q\}) = [\prod_{n=1}^{t}(1+\gamma_n)]\cdot[F(q; p^*)]^t \cdot\pi_1(\{q\})$, which implies

$$\pi_t(\{q\})/\pi_t(P) = [F(q; p^*)]^t\cdot\pi_1(\{q\})/ [\textstyle\sum_{q \in \text{supp}(\beta)} [F(q; p^*)]^t\cdot\pi_1(\{q\})$$

$$= [F(q; p^*)/F(p^*; p^*)]^t/[\textstyle\sum_{q \in \text{supp}(\beta)} [F(q; p^*)/F(p^*; p^*)]^t .$$

Case 1($q = p^*$): It holds by Theorem 2 that

$$\lim_{t\to+\infty} [F(p^*; p^*)/F(p^*; p^*)]^t = \lim_{t\to+\infty} 1^t = 1 \text{ and}$$

$$\lim_{t\to+\infty} [F(r; p^*)/F(p^*; p^*)]^t = 0 \text{ for all } r \neq p^*.$$

Hence we have that

$$\lim_{t\to+\infty}[\textstyle\sum_{r \in \mathrm{supp}(\beta)} [F(r; p^*)/F(p^*; p^*)]^t = \lim_{t\to+\infty} [F(p^*; p^*)/F(p^*; p^*)]^t = 1 \text{ and}$$

$$\lim_{t\to+\infty} \pi_t(\{q\})/\pi_t(P) = \lim_{t\to+\infty} \pi_t(\{q^*\})/\pi_t(P) = 1.$$

Case 2($q \neq p^*$): Because $\lim_{t\to+\infty} [F(q; p^*)/F(p^*; p^*)]^t = 0$ and $\lim_{t\to+\infty} [\sum_{r\in \mathrm{supp}(\beta)} [F(r; p^*)/F(p^*; p^*)]^t = 1$, we have that $\lim_{t\to+\infty}\pi_t(\{q\})/\pi_t(P) = 0/1 = 0$. □

Example 1 (Continue) : When $q_k = 0.1\cdot k$ for $k = 1, 2, \cdots, 8$, and $m = 2$, $p^* = 0.4$, $p_\alpha = 0.3$, $p_\beta = 0.1$, $\delta = 0.3$, we assume that $\mu(N_k(1)) = 1/8$ for all $k = 1, 2, \cdots, 8$.

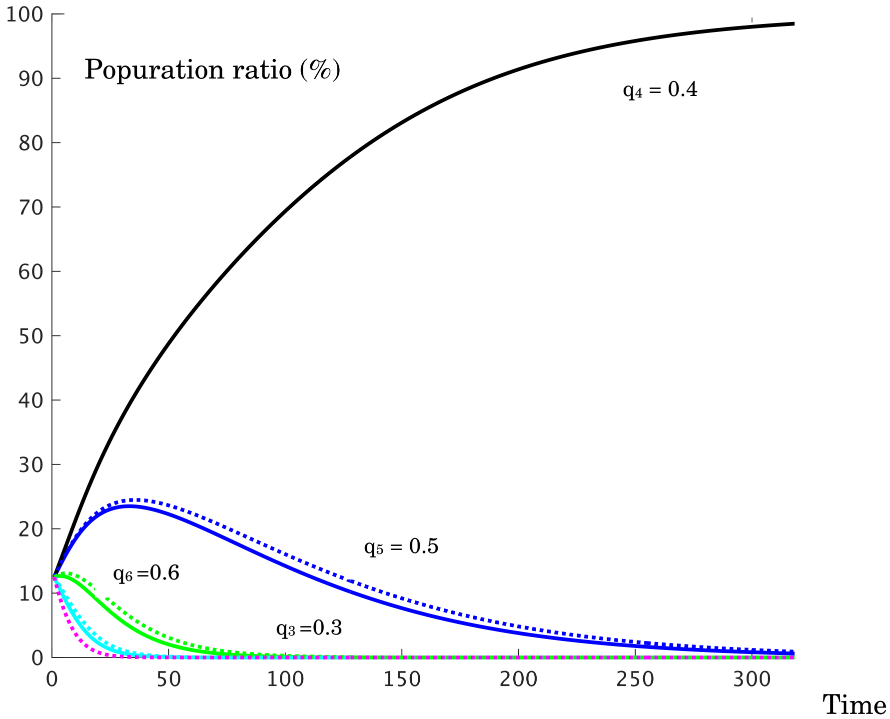

Figure 8

6. Proof of lemmas

Proof of Lemma 1: Setting $X_n = (x - (1/n), x + (1/n))$, it holds by the definition of $P(\theta_1 | x : p)$ that $P(\theta_1 | x : q) = \lim_n q \cdot P(X_n | \theta_1)/P(X_n)$, where $P(X_n | \theta_1) = \int_{X_n} f(x | \theta_1) d\mu$ and $P(X_n) = \int_{X_n} [q \cdot f(x | \theta_1) + (1 - q) \cdot f(x | \theta_0)] d\mu = q \cdot P(X_n | \theta_1) + (1 - q) \cdot P(X_n | \theta_0)$. It follows from the mean-value theorem that there exists a sequence of $\{ x_n \}$ such that: (**i**) $\lim x_n = x$, (**ii**) $x_n \in X_n$ for all n, (**iii**) $p(X_n | \theta_i) = f(x_n | \theta_i)(2/n)$ for $i = 0, 1$ and all n. Because f is continuous, we have that

$$P(\theta_1 | x : q) = \lim_n q \cdot P(X_n | \theta_1)/P(X_n)$$

$$= \lim_n q \cdot f(x_n | \theta_1)(2/n)/[q \cdot f(x_n | \theta_1)(2/n) + (1 - q) \cdot f(x_n | \theta_0)(2/n)]$$

$$= \lim_n q \cdot f(x_n | \theta_1)/[q \cdot f(x_n | \theta_1) + (1 - q) \cdot f(x_n | \theta_0)] = q \cdot f(x | \theta_1)/[q \cdot f(x | \theta_1) + (1 - q) \cdot f(x | \theta_0)].$$

Using almost the same manner, we can prove that $P(\theta_0 | x : q) = (1 - q) \cdot f(x | \theta_0)/[q \cdot f(x | \theta_1) + (1 - q) \cdot f(x | \theta_0)]$. □

Proof of Lemma 3: (**i**) We can prove Lemma 3(i) as a direct consequence of (6) and Lemma A below:

Lemma A: Let g be a bounded continuous function on S in $\mathfrak{B}_X$ such that $g(x) > 0$ for all $x \in S$. For a given null set A in $\mathfrak{B}_X$, the value of the Lebesgue integral of g on S coincides with the value of the Lebesgue integral of g on S/A.

Lemma A is a direct consequence of Royden (1988, Ch.4, Proposition 5(iii), Page 82).

(**ii**) Setting $k(x) = q \cdot f(x | \theta_1) + (1 - q) \cdot f(x | \theta_0)$, $X^1 = \delta^{-1}(d_1)$ and $X^0 = \delta^{-1}(d_0)$, we have that

$$G(\delta: q) = (1 - q) \cdot p_\alpha \cdot \alpha(\delta) + q \cdot (1 - \beta(\delta)) + (1 - q) \cdot (1 - \alpha(\delta)) + q \cdot p_b \cdot \beta(\delta)$$

$$= (1 - q) \cdot p_\alpha \cdot \int_{X^1} f(x | \theta_0)\, d\mu + q \cdot \int_{X^1} f(x | \theta_1) d\mu$$

$$+ (1 - q) \cdot \int_{X^0} f(x | \theta_0)\, d\mu + q \cdot p_b \cdot \int_{X^0} f(x | \theta_1) d\mu$$

$$= \int_{X^1} [(1 - q) \cdot f(x | \theta_0) \cdot p_\alpha + q \cdot f(x | \theta_1)] d\mu + \int_{X^0} [(1 - q) \cdot f(x | \theta_0) + q \cdot f(x | \theta_1) \cdot p_\beta] d\mu$$

$$= \int_{X^1} h(d_1 | x : q) \cdot k(x) d\mu + \int_{X^0} h(d_0 | x : q) \cdot k(x) d\mu$$

$$= \int_{X^1} h(d_1 | x : q) d\lambda_q + \int_{X^0} h(d_0 | x : q) d\lambda_q = \int_X h(\delta(x) | x : q) d\lambda_q.$$ □

Proof of Lemma 4: Select any $q \in P$ and let q be a subjective belief of a testing agent.

(**i**) Define a function $\delta : X \to \{ d_0, d_1 \}$ by

$\delta(x) = d_1$ for all $x \in \{ x \in X : f(x \mid \theta_1)/f(x \mid \theta_0) \geq (1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta) \}$;

$\delta(x) = d_0$ for all $x \in \{ x \in X : f(x \mid \theta_1)/f(x \mid \theta_0) < (1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta) \}$.

It suffices to prove that $\{ x \in X : f(x \mid \theta_1)/f(x \mid \theta_0) < (1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta) \} \in \mathfrak{B}_X$, which is a direct consequence of the fact that $L(x) \equiv f(x \mid \theta_1)/f(x \mid \theta_0)$ is continuous on X.

(**ii**) Fix any $\delta^1 \in \Delta_q$ and $\delta^2 \in \Delta$. It follows from Lemma 2(i) that δ^1 has the property:

$\delta^1(x) = d_0$ for almost all x in

$$h(d_1 \mid x{:}\ q) = (1-q)\cdot f(x \mid \theta_0)\cdot p_\alpha + q\cdot f(x \mid \theta_1) < (1-q)\cdot f(x \mid \theta_0) + q\cdot f(x \mid \theta_1)\cdot p_\beta = h(d_0 \mid x{:}\ q);$$

$\delta^1(x) = d_1$ for almost all x in $h(d_1 \mid x{:}\ q) > h(d_0 \mid x{:}\ q)$;

Do either if $h(d_1 \mid x{:}\ q) = h(d_0 \mid x{:}\ q)$.

This means that $h(\delta^1(x) \mid x : q) = \mathbf{max}[h(d_1 \mid x : q), h(d_0 \mid x : q)]$ for almost all $x \in X$. Because $\mathbf{max}[h(d_1 \mid x : q), h(d_0 \mid x : q)] \geq h(\delta^2(x) \mid x : q)$ for all $x \in X$, we have that

$$h(\delta^1(x) \mid x : q) \geq h(\delta^2(x) \mid x : q) \text{ for almost all } x \in X.$$

(**iii**) Lemm 4(iii) is a direct consequence of Lemma 4(ii).

(**iv**) Fix any $\delta^1 \in \Delta_q$, and suppose that $\delta^2 \in \Delta$ satisfies $\delta^2 \notin \Delta_q$.

Case I($\mu(C^1(q)) > 0$ and $\mu(C^0(q)) > 0$): We need a claim:

Claim 1: At least one of the following two statements holds:

(**a**) There exists $A^* \in \mathfrak{B}_X$ such that $A^* \subset C^1(q)$, $\mu(A^*) > 0$ and $\delta^2(x) = d_0$ for all $x \in A^*$.

(**b**) There exists $B^* \in \mathfrak{B}_X$ such that $B^* \subset C^0(q)$, $\mu(B^*) > 0$ and $\delta^2(x) = d_1$ for all $x \in B^*$.

Proof of Claim 1: Set $A^* = C^1(q) \cap \{ x \in X : \delta^2(x) = d_0 \}$ and $B^* = C^0(q) \cap \{ x \in X : \delta^2(x) = d_1 \}$. If $\mu(A^*) = 0$ and $\mu(B^*) = 0$, then $\mu(C^1(q) \cap \{ x \in X : \delta^2(x) = d_1 \}) = \mu(C^1(q))$ and $\mu(C^0(q) \cap \{ x \in X : \delta^2(x) = d_0 \}) = \mu(C^0(q))$. It holds by $\mu(C^1(q)) > 0$ and $\mu(C^0(q)) > 0$ that

$$\delta^2(x) = d_1 \text{ for almost all } x \in C^1(q) \text{ and } \delta^2(x) = d_0 \text{ for almost all } x \in C^0(q).$$

Hence it holds that $\delta^2 \in \Delta_q$, which is a contradiction. Thus we have that $\mu(A^*) > 0$ or $\mu(B^*) > 0$. □

First we prove Lemma 4(iv) in case (**a**). It holds by Claim 1 and Royden (1988, Proposition 15, Ch.3, Page 63) that there exists a *compact* subset A in $C^1(q)$ such that $\mu(A) > 0$ and $\delta^2(x) = d_0$ for all $x \in A$. Because $f(x \mid \theta_1)/f(x \mid \theta_0) > (1 - q)\cdot(1 - p_\alpha)/q\cdot(1 - p_\beta)$ implies $(1 - q)\cdot f(x \mid \theta_0)\cdot p_\alpha + q\cdot f(x \mid \theta_1) > (1 - q)\cdot f(x \mid \theta_0) + q\cdot f(x \mid \theta_1)\cdot p_\beta$ for all $x \in A$, it holds by Lemma 2(i) that

$$h(d_1 \mid x: q) > h(d_0 \mid x: q) \text{ for all } x \in A. \tag{10}$$

It holds by the definition of δ^1 that

$$\delta^2(x) = d_0 \text{ and } \delta^1(x) = d_1 \text{ for almost all } x \in A. \tag{11}$$

We have by (10) and (11) that $h(\delta^1(x) \mid x : q) > h(\delta^2(x) \mid x : q)$ for almost all $x \in A$. It holds by Royden (1988, Proposition 15, Ch.3, Page 63) that there exists a compact subset A^+ in A such that $\mu(A^+) > 0$ and $h(\delta^1(x) \mid x : q) > h(\delta^2(x) \mid x : q)$ for all $x \in A^+$. Second, we can prove Lemma 4(iv) in case (**b**) by almost the same manner.

Case II ($\mu(C^1(q)) > 0$ and $\mu(C^0(q)) = 0$): We need a claim:

Claim 2: There exists $A^* \in \mathfrak{B}_X$ such that $A^* \subset C^1(q)$, $\mu(A^*) > 0$ and $\delta^2(x) = d_0$ for all $x \in A^*$.

Proof of Claim 2: Set $A^* = C^1(q) \cap \{ x \in X : \delta^2(x) = d_0 \}$. If $\mu(A^*) = 0$, then $\mu(C^1(q) \cap \{ x \in X : \delta^2(x) = d_1 \}) = \mu(C^1(q))$. It holds by $\mu(C^1(q)) > 0$ that $\delta^2(x) = d_1$ for almost all $x \in C^1(q)$. Hence it holds that $\delta^2 \in \Delta_q$, which is a contradiction. Thus we have that $\mu(A^*) > 0$. □

It holds by Claim 2 and Royden (1988, Proposition 15, Ch.3, Page 63) that there exists a *compact* subset A such that $A^* \subset C^1(q)$, $\mu(A) > 0$ and $\delta^2(x) = d_0$ for all $x \in A$. Because $f(x \mid \theta_1)/f(x \mid \theta_0) > (1 - q)\cdot(1 - p_\alpha)/q\cdot(1-p_\beta)$ implies $(1 - q)\cdot f(x \mid \theta_0)\cdot p_\alpha + q\cdot f(x \mid \theta_1) > (1 - q)\cdot f(x \mid \theta_0) + q\cdot f(x \mid \theta_1)\cdot p_\beta$ for al l $x \in A$, it holds by Lemma 2(i) that

$$h(d_1 \mid x: q) > h(d_0 \mid x: q) \text{ for all } x \in A. \tag{12}$$

It holds by the definition of δ^1 that

$$\delta^2(x) = d_0 \text{ and } \delta^1(x) = d_1 \text{ for almost all } x \in A. \tag{13}$$

We have by (12) and (13) that $h(\delta^1(x) | x : q) > h(\delta^2(x) | x : q)$ for almost all $x \in A$. It holds by Royden (1988, Proposition 15, Ch.3, Page 63) that there exists a compact subset A^+ in A such that $\mu(A^+) > 0$ and $h(\delta^1(x) | x : q) > h(\delta^2(x) | x : q)$ for all $x \in A^+$.

Case III ($\mu(C^1(q)) = 0$ and $\mu(C^0(q)) > 0$): We can prove Lemma 4(iv) in this case by almost the same manner used in the proof of the case II above.

Case IV($\mu(C^1(q)) = 0$ and $\mu(C^0(q)) = 0$): In this case, $f(x | \theta_1)/f(x | \theta_0) = (1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta)$ for all x. Hence we have that $\Delta_q = \Delta$, which implies that there is no $\delta^2 \in \Delta$ which satisfies $\delta^2 \notin \Delta_q$. Thus Lemma 4(iv) holds in this case.

(**v**) Lemm 4(v) is a direct consequence of Lemma 3(ii) and Lemma 4(ii, iv). □

Proof of Lemma 5: (**i**) It holds that $L'(x) = exp[m\cdot(x-(½))] > 0$, which implies that $L(x)$ is an increasing function of x.

(**ii**) Because $c(q) = (1/2)+(1/m)\cdot log\,[(1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta)] = (1/m)\cdot log(1/q-1)+(1/m)\cdot log(1-p_\alpha)/(1-p_\beta) + (1/2)$, it holds that $c'(q) = (1/m)\cdot 1/[1/q-1]\cdot(-1/q^2) = -(1/m)\cdot(1/q)\cdot[1/(1-q)] < 0$, which implies that $c(q)$ is a decreasing function of q.

(**iii**) Because $\lim_{q\to 0} log(1/q-1) = +\infty$, we have that

$$\lim_{q\to 0} c(q) = \lim_{q\to 0} [(1/m)\cdot log(1/q-1)+(1/m)\cdot log(1-p_\alpha)/(1-p_\beta) + (1/2)] = +\infty.$$

Because $\lim_{p\to 1} log(1/q-1) = -\infty$, we have that

$$\lim_{p\to 1} c(q) = \lim_{p\to 1} [(1/m)\cdot log(1/q-1) + (1/m)\cdot log(1-p_\alpha)/(1-p_\beta) + (1/2)] = -\infty.$$

(**iv**) It holds that $\{ x \in X : f(x | \theta_1)/f(x | \theta_0) > (1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta) \} = \{ x \in X : x > (L^{-1}((1-q)\cdot(1-p_\alpha)/p\cdot(1-p_\beta)) \} = \{ x \in X : x > c(q) \}$, and that $\{ x \in X : f(x | \theta_1)/f(x | \theta_0) < (1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta) \} = \{ x \in X : x < (L^{-1}((1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta)) \} = \{ x \in X : x < c(q) \}$. □

Proof of Lemma 6: (**i**) We can prove Lemma 6(i) by almost the same manner used in the proof of Lemma 3(i).

(**ii**) We can prove Lemma 6(ii) by almost the same manner used in the proof of Lemma 3(ii).

(**iii**) Because $p \neq q$ implies $(1-p)\cdot(1-p_\alpha)/p\cdot(1-p_\beta) \neq (1-q)\cdot(1-p_\alpha)/q\cdot(1-p_\beta)$, Lemma 6(iii) holds by the full range assumption of $L(x)$ and the continuity of $L(x)$. □

7. Appendix

The full range assumption: $\{\, y \in \mathbb{R}_{++} : \ y = f(x \mid \theta_1)/f(x \mid \theta_0)$ for some $x \in X. \,\} = \mathbb{R}_{++}$ is indispensable for Theorems 2 and 3. We construct a counter example, on which the full range assumption, Theorems 2 and 3 do not hold :

Example 2: Set $p^* = 0.2$, and set $m = 2$, $p_\alpha = 0.3$, $p_\beta = 0.1$. The two conditional probability destributions, $f(x \mid \theta_0)$ and $f(x \mid \theta_1)$ are defined by

$$f(x \mid \theta_0) = A\cdot exp[-x^2] \quad \text{and} \quad f(x \mid \theta_1) = A\cdot exp[-x^2]\cdot g(x) \quad \text{for all } x \in X,$$

where $A = 2^{½}\cdot(2\pi)^{-½}$, $exp(y) = e^y$ for all $y \in X$ and $g(x) = 2\cdot e^x/(e^x+1)$ for all $x \in X$.

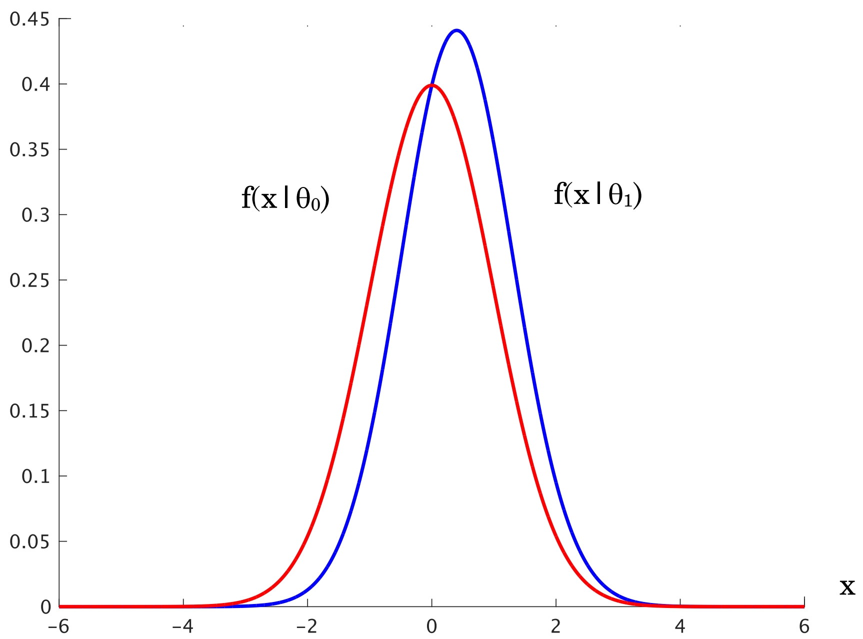

Figure 9

$L(x) \equiv f(x \mid \theta_1)/f(x \mid \theta_0) = g(x) = 2 \cdot e^x/(e^x+1)$, which implies that the monotonicity condition holds, but $\lim_{x \to +\infty} L(x) = 2$ and $\lim_{x \to -\infty} L(x) = 0$. Because $(1-p) \cdot (1-p_\alpha)/p \cdot (1-p_\beta) = (1-p) \cdot 7/p \cdot 9 = (7/9) \cdot (1/p - 1) > 2$ iff $p < (7/25) = 0.28$, we have that

$$\Delta_q = \Delta_r \text{ for all } q, r \in P \text{ such that } q, r < 0.28,$$

and that

$$F(q: 0.2) = 0.82 \qquad \text{for all } q < 0.28$$

$$= 1 - [0.56 \cdot \int_{X^1} f(x \mid \theta_0) dx + 0.18 \cdot \int_{X^0} f(x \mid \theta_1) dx] \qquad \text{for all } q \geq 0.28,$$

where $X^1 = \delta^{-1}(d_1)$ and $X^0 = \delta^{-1}(d_0)$.

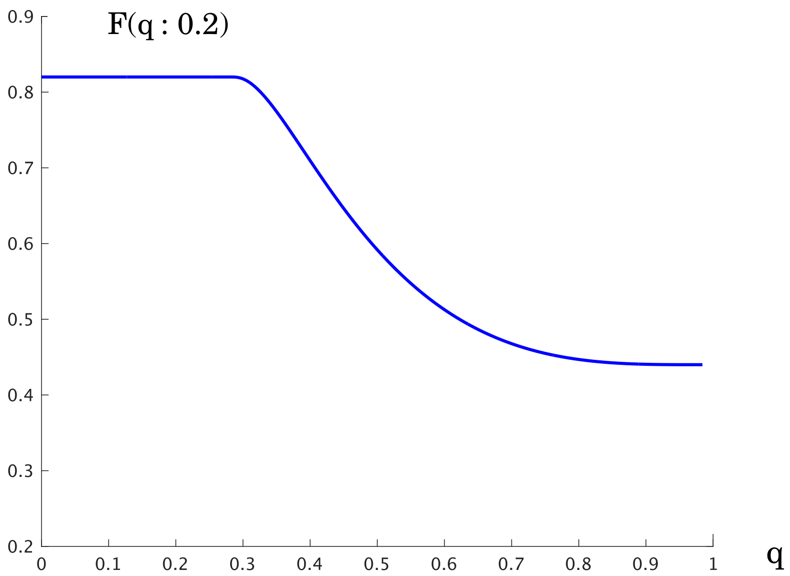

Figure 10

8. References

Berger, J. O., Bernardo, J. M., and Sun, D. (2015). Overall objective priors. *Bayesian Analysis*, **10** (1), 189-221.

DeGroot, M. H., and Schervish, M. J. (2012). "*Probability and Statistics, 4rth Ed.*" Pearson Education, Harlow, UK.

Friedman, M. (1953). "*Essays in positive economics*." University of Chicago Press.

Lehmann, E.L. and J. P. Romano, (2005), "*Testing statistical hypotheses, 3rd Ed*." Springer, New York.

Lesaffre, E., and Lawson, A. B. (2012). "*Bayesian biostatistic*s." John Wiley & Sons.

Lo, A. W., and R. Zhang, (2021). The evolutionary origin of Bayesian heuristics and finite memory. *iScience* **102853**.

Nowak, M. A. (2006). "*Evolutionary dynamics: exploring the equations of life*". Harvard University Press.

Okasha, S. (2013). The evolution of Bayesian updating. *Philosophy of Science* **80** (5), 745-757.

Royden, H.L. (1988). "*Real analysis 3rd Ed*". Macmillan, London.

Sober, E. (2008). "*Evidence and evolution: The logic behind the science*". Cambridge University Press.

Zhang, H. (2013). Evolutionary justifications for non-Bayesian beliefs. *Economics Letters*, **121** (2), 198-201.