Simultaneous Bayesian-Frequentist Sequential Testing of Nested Hypotheses

BY JAMES O. BERGER

Zhao Hongqiang

1.Introduction

1.1 background:

mainly question : The sequential testing of compound hypotheses is quite difficult.

solution:

Wald (1947) generalised the sequential probability ratio test (SPRT) by using a weighting function (some kind of prior distribution).

Subsequent references and efforts to deal with the problem can be found in Ghosh (1970) and Siegmund (1985)

The main difficulties in the calculation of the probability of error and the problems associated with choosing an stopping rule.

the problem in unconditional frequentist testing

for an $\alpha = 0.05$ level test, one reports the same error probability rejection whether the data is just at the rejection boundary or far within the rejection region.

The traditional classical way of addressing this problem is to report p-value or attained significance level

Berger, Brown & Wolpert (1994) proposed the frequentist error probabilities would equal the Bayesian posterior probabilities of the hypotheses.

1.2 The (truncated) conditional sequential t-test

```
an application of the new methodology to sequential t-testing.
```

sample :X1,X2......~$\mathcal{N}\left(\theta, \sigma^2\right)$

sequentially test

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0$$

The statistic Bn （defind by 4.2 and 4.4） forms the basis of a new (truncated) conditional sequential t-test, T*, defined as follows.

Step 1.

Choose constants O < R < 1 < A,

m:Positive integers, maximum number of observations

stopping time:

$$N \equiv N_m = \{ \text{ first } n < m \text{ such that } B_n \notin (R, A);\ \text{else, choose } n = m\}$$

R:Rejection of H0's conditions

A:Expected odds of accepting H0

Step 2.

After stopping at time N

$$\begin{cases} \text{if } B_N \leq 1, & \text{reject } H_0 \text{ and report error probability } \alpha^*\left(B_N\right) = \frac{B_N}{1+B_N} \\ \text{if } 1 < B_N < a, & \text{make no decision,} \\ \text{if } B_N \geq a, & \text{accept } H_0 \text{ and report error probability } \beta^*\left(B_N\right) = \frac{1}{1+B_N} \end{cases}$$

The constant a is computed through equation (4 · 5)

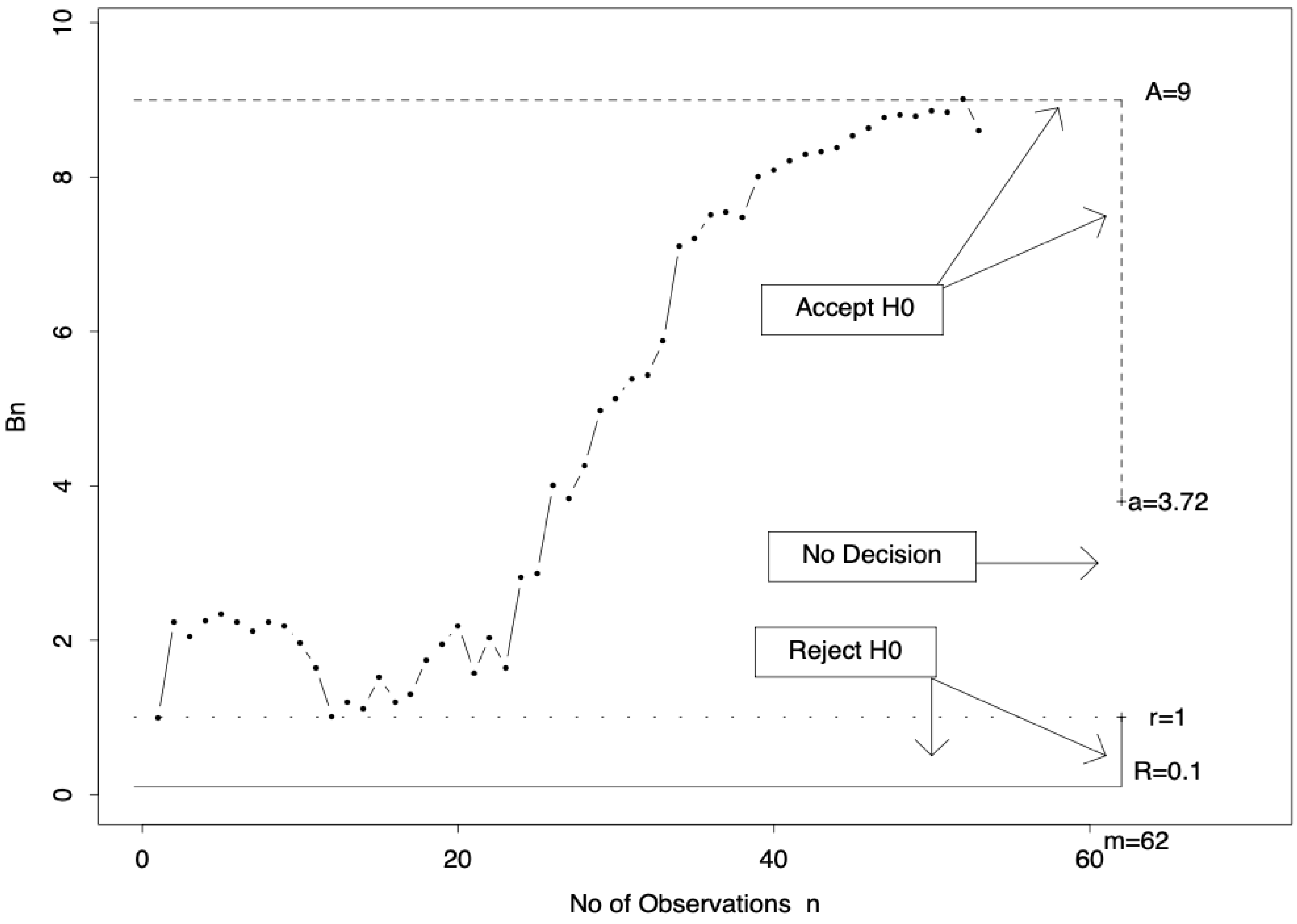
example

Table 6.3 of Armitage (1975).

Armitage used an approximate sequential t-test with a truncation at m=62. Assuming that R = 0.1 and A = 9 are chosen; these, together with m = 62, define the bounds on the stopping

and is shown in Figure 1, which also shows the data, plotted as Bn versus n.

Computation a=3.72；



The resulting decision boundary is also shown in Figure 1.

For the given data, the stopping bound A = 9 will be reached for n = 52 observations; in fact, $B_{52} = 9.017$. The test will then conclude by accepting Ho and reporting the conditional error probability

The unconditional error probabilities for Type I and Type II were calculated as a' = 0.048 and B' = 0.109.

1.3 Comments and motivation

- 1.The main motivation for the new test is that

the probability of error $\alpha^* \left(B_N\right) \beta^* \left(B_N\right)$ =H0 H1 （posterior probability）

- 2.R, A and m, have a clear intuitive interpretation.
- 3.New tests are usually computationally trivial
- 4.The new test is compatible with SRP because the reported error probability never depends on the stopping rule

2.BASIC NOTATION

$X_1, X_2, \ldots,$ be a sequence of observable random variables

$X_n = (X_1, \ldots, X_n)$and let $\mathcal{F}_n = \sigma\left\{\underset{\sim}{X_n}\right\}$ denote the corresponding sigma-algebra.

Here, n is time and Xn is the data available at time n. Let N be an appropriate stopping time (an adaptation of {Fn}) and let {FN} denote all events$D \in \mathcal{F}_\infty$The set determined before N。

2.1: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \in \Theta_1$

for some $\theta_0 \in \Theta, \theta_0 \notin \Theta_1 \subset \Theta$. usually Θ_1 given by$\Theta_1 = \{\theta \in \Theta : \theta \neq \theta_0\}$

In the default Bayesian framework （CF.Jeffreys（1961））, one specifies equal prior probabilities of 1/2 for Ho and H1 being true, and chooses a default prior density π(θ)/ 2 on Θ_1, where π(.) is a proper density on Θ_1 People accept (reject) Ho because its posterior probability is greater (less) than 1/2, reported error probability is just the posterior probability of H1 (H0).

For each fixed n, the marginal density functions of Xn under Ho and H1 in (2.1) are given, respectively, by

2.2: $m_{0,n}\left(x_n\right)=f_n\left(x_n \mid \theta_0\right)$ and $m_{1,n}\left(x_n\right)=\int_{\Theta_1} f_n\left(x_n \mid \theta\right) \pi(\theta) d\theta$

2.3: $H_0^{\prime}: X_n \sim m_{0,n}\left(x_n\right), \quad \forall n \geq 1,$ versus $H_1^{\prime}: X_n \sim m_{1,n}\left(x_n\right), \quad \forall n \geq 1,$

which is based, for each fixed n>=1, on the corresponding likelihood ratio for Xn,

2.4 : $B_n=\frac{m_{0,n}(\underset{\sim}{x_n})}{m_{1,n}\left((\underset{\sim}{x_n})\right.}$

H0 being true is

2.5: $\alpha^* (B_n) \equiv \mathrm{pr}\left(H_0 \mid \underset{\sim}{x_n}\right) = B_n / (1 + B_n)$

H1 being true is

2.6: $\beta^* (B_n) \equiv \mathrm{pr}(H_1 \mid x_n) = 1 / (1 + B_n)$

The classical sequentially test hypothesis (2.3) is to construct rejection and acceptance regions that report the probability of error for type I and type II given by a′ and B′

When considering compound choices, such as H1 in (2.1), the probability of a type II error is a function of θ

$\beta(\theta)=\mathrm{pr}(\text{ Accepting } H_0 \mid \theta) \equiv P_\theta(\text{ Accepting } H_0)$, for $\theta \in \Theta_1$。 Note, that (α', β') and $\beta(\theta)$will depend on the stop rule used.

3. THE MODIFIED BAYES-SEQUENTIAL TEST

In the test case of Section 2, let N be the appropriate stopping time for the sequential experiment (i.e., $\mathrm{pr}(N < \infty \mid \theta) = 1 \quad \forall \theta \in \Theta$).

Let Pi denote (2.3) in H_i', $\quad i = 0, 1$, under the probability. That is, for any$\mathcal{D} \in \mathcal{F}_N$ 。

$P_0(\mathcal{D}) = \mathrm{pr}\big(\mathcal{D} \mid H_0'\big) \equiv P_{\theta_0}(\mathcal{D})$, while

$3.1 : P_1(\mathcal{D}) = \mathrm{pr}\big(\mathcal{D} \mid H_1'\big) \equiv \int_{\Theta_1} P_\theta(\mathcal{D})\pi(\theta)d\theta$

For $i = 0, 1$, let $F_i(\cdot)$ denote the distribution function of $B_N : F_i(b) \equiv P_i\left(B_N \leq b\right)$, $b \in \mathbb{R}$. Wherever they exist, we write F_i^{-1} for the inverse function of $F_i, i = 0, 1$, and denote

3.2: $\psi(s) = F_0^{-1}\left(1 - F_1(s)\right)$ and $\psi^{-1}(s) = F_1^{-1}\left(1 - F_0(s)\right)$

Condition I. Assume that the range of B_N is of the form $\mathcal{B} = (R_L, R_U] \cup [A_L, A_U)$, where $R_U \leq 1 \leq A_L, R_L$ could be zero, and A_U could be infinity.

assume that ψ exists on $(R_L, R_U]$ and ψ^{-1} exists on $[A_L, A_U)$. Since we are dealing with continuous densities, this condition will be satisfied by all but very strange stopping rules. Here are few examples.

Example 1. A standard "open-ended" stopping rule, familiar from the SPRT, is

$$3.3: N=\min \{n \geq 1: \quad B_n \notin (R, A)\}$$

where $R < 1 < A$. If the B_n can range from zero to infinity, it is easy to see that $(R_L, R_U] = (0, R]$ and $[A_L, A_U) = [A, \infty)$, and the remaining part of Condition I can be easily verified.

Example 2. With N as in $(3 \cdot 3)$, consider the truncated at m stopping time $N_m = \min(N, m)$ (see also Section $(1 \cdot 2)$). Clearly, this is a proper stopping time. Since the range of B_m must include 1, so must that of B_N; hence $R_U = 1 = A_L$.

For the examples in this paper, the range of B_n is of the form $(0, C_n)$, where the C_n are increasing in n. Thus $R_L = 0$ and $A_U = C_m$ and the range of B_N is therefore $\mathcal{B} = (0, C_m)$. The remaining part of Condition I can be easily verified.

Example 3. Variants on the stopping time $(3 \cdot 3)$

$(i) \quad N_m^1 = \min \{n \geq m : B_n \notin (R, A)\}$, of a "two-stage" study (with an initial sample of $m > 1$ observations taken in a single batch and then followed by a sequential sampling),

$(ii) \quad N_m^2 = m \cdot \inf \{n \geq 1 : \quad B_{nm} \notin (R, A)\}$, of a "group-sequential" study which samples (sequentially) m units at-a-time.

Corresponding to the constants R_U and A_L defined above, let

$$3.4 : r = \min \left(R_U, \psi^{-1} (A_L)\right), \quad a = \max \left(A_L, \psi \left(R_U\right)\right)$$

These constants, r and a, define the "decision boundaries" for the modified Bayessequential test, T, as follows: $\begin{cases}\text { if } B_{N} \leq r, & \text { reject } H_{0}, \text { and report the } \operatorname{CEP} \alpha^{*}\left(B_{N}\right)=B_{N} /\left(1+B_{N}\right) \\ \text { if } r<B_{N}<a, & \text { make no decision (though the experiment is stopped) } \\ \text { if } B_{N} \geq a, & \text { accept } H_{0}, \text { and report the } \operatorname{CEP} \beta^{*}\left(B_{N}\right)=1 /\left(1+B_{N}\right)\end{cases}$

3.6 $S\left(B_{N}\right)=\min \left(B_{N}, \psi^{-1}\left(B_{N}\right)\right)$

$$
\begin{gathered}
\alpha(s)=\operatorname{pr}\left(B_N \leq r \mid \theta_0, S\left(B_N\right)=s\right) \\
\beta(s \mid \theta)=\operatorname{pr}\left(B_N \geq a \mid \theta, S\left(B_N\right)=s\right), \theta \in \Theta_1
\end{gathered}
$$

THEOREM 1 . Consider the test T^*, of the simple versus composite hypotheses $(2 \cdot 1)$, under Condition I and with the conditioning statistic $S\left(B_N\right)$ given in $(3 \cdot 6)$. Then, in the rejection and acceptance regions, respectively,

$$
3.8: \alpha^*\left(B_N\right)=\alpha(s) \quad \text { and } \quad \beta^*\left(B_N\right)=\int_{\Theta_1} \beta(s \mid \theta) \pi(\theta \mid s) d \theta
$$

where $\pi(\theta \mid s)$ denotes the posterior density of θ conditional on H_1 being true and on the observed value s of $S\left(B_N\right)$.

The first equality in $(3 \cdot 8)$ provides the key result that the conditional Type I error probability in $(3 \cdot 7)$ and the posterior probability of H_0 in $(2 \cdot 5)$ are equal.

he second equality states that $\beta^* \left(B_N\right)$ in $(2 \cdot 6)$ (the posterior probability of $H_1)$ is the average of the conditional Frequentist Type II error probability

分层贝叶斯（Hierarchical bayesian）

让 y_j 成为一个观察和 θ_j 控制数据生成过程的参数 y_j. 进一步假设参数 $\theta_1, \theta_2, \ldots, \theta_j$ 由一个共同的群体交换生成, 分布由一个超参数控制 ϕ. 贝叶斯分层模型包含以下阶段:

Stage I: $y_j \mid \theta_j, \phi \sim P\left(y_j \mid \theta_j, \phi\right)$

Stage II: $\theta_j \mid \phi \sim P\left(\theta_j \mid \phi\right)$

Stage III: $\phi \sim P(\phi)$

在第一阶段看到的可能性是 $P\left(y_j \mid \theta_j, \phi\right)$, 和 $P\left(\theta_j, \phi\right)$ 作为其先验分布。请注意, 可能性取决于 ϕ 只有通过 θ_j. 第一阶段的先验分布可以分解为: $P\left(\theta_j, \phi\right)=P\left(\theta_j \mid \phi\right) P(\phi)$ 【来自条件概率的定义】 和 ϕ 作为具有超先验分布的超参数, $P(\phi)$. 因此, 后验分布与:
$P\left(\phi, \theta_j \mid y\right) \propto P\left(y_j \mid \theta_j, \phi\right) P\left(\theta_j, \phi\right)$ [使用贝叶斯定理]

$P\left(\phi, \theta_j \mid y\right) \propto P\left(y_j \mid \theta_j\right) P\left(\theta_j \mid \phi\right) P(\phi)$[13]

4. APPLICATION

In this section we illustrate the application of the proposed conditional sequential testing procedure to the "two-sided" normal testing problem.

X1,X2...~i.i.d $\mathcal{N}\left(\theta, \sigma^2\right)$

case1 In Table 1 below we compute a for the common default choice of $\xi = 2$ and various values of R and A. including the unconditional probabilities of Type I and Type II errors, $\alpha' = \mathrm{pr}\left(B_N < r \mid H_0'\right)$ and $\beta' = \mathrm{pr}\left(B_N > a \mid H_1'\right)$

the expected stopping times $(E_0(N), E_1(N))$ and variances $(\mathrm{var}_0(N), \mathrm{var}_1(N))$ under H_0' and H_1'

the corresponding probabilities of no-decision, $p_0 = \mathrm{pr}\left(r < B_N < a \mid H_0'\right)$,and $p_1 = \mathrm{pr}\left(r < B_N < a \mid H_1'\right)$

Table 1. Two-sided normal sequential testing with known σ^2

A	a	α'	β'	p_0	p_1	$E_0(N)$	$E_1(N)$	$var_0(N)$	$var_1(N)$
						$R = 0.1$			
3.0	3.176	0.034	0.207	0.254	0.082	8.6	7.6	66	84
4.0	4.084	0.039	0.177	0.185	0.046	14.5	10.9	212	289
5.0	5.075	0.042	0.153	0.137	0.027	22.5	14.2	611	616
6.0	6.043	0.044	0.134	0.108	0.018	31.8	18.0	1168	1154
7.0	7.040	0.046	0.118	0.084	0.012	43.0	21.6	2083	2120
8.0	8.025	0.047	0.106	0.067	0.008	55.8	25.2	3817	2947
9.0	9.023	0.048	0.096	0.054	0.006	70.4	29.6	6072	6000
10.0	10.015	0.049	0.088	0.043	0.004	86.8	33.7	10550	8548
						$R = 0.05$			
3.0	3.187	0.017	0.204	0.276	0.089	9.0	8.9	106	163
4.0	4.091	0.019	0.176	0.209	0.051	15.3	12.9	344	461
5.0	5.087	0.021	0.152	0.163	0.032	23.7	17.1	983	1327
6.0	6.052	0.022	0.133	0.133	0.022	33.5	21.3	1632	2182
7.0	7.051	0.023	0.118	0.110	0.016	45.4	26.3	2962	3605
8.0	8.034	0.023	0.105	0.095	0.012	58.8	29.9	5272	4440
9.0	9.033	0.024	0.096	0.081	0.009	74.0	35.3	8791	8182
10.0	10.023	0.025	0.088	0.070	0.007	91.2	39.9	13904	12738

We continue with the sequential test of (4· 1), but now assume that both (θ, σ^2) are unknown. using a hierarchical prior structure defined as follows.

For the first-stage prior distribution of θ, take $\pi_1\left(\theta \mid \sigma^2, \xi\right)=\mathcal{N}\left(\theta_0, \xi \sigma^2\right)$.

For the second-stage prior of $\left(\sigma^2, \xi\right)$, take $\pi_2\left(\sigma^2, \xi\right)=\sigma^{-2} g(\xi) d \sigma^2 d \xi$,

$$4.2: B_n=\left(n-1+y_n\right)^{-n / 2} \times\left[\int_0^{\infty} \frac{(1+n \xi)^{(n-1) / 2}}{\left[(n-1)(1+n \xi)+y_n\right]^{n / 2}} g(\xi) d \xi\right]^{-1}$$

where $y_n=n\left(\bar{x}-\theta_0\right)^2 / S_n^2$ and S_n^2 is the usual sample variance.

$n \geq 2$, we write the density of y_n as $f\left(y_n \mid \mu\right)$, where $\mu=\left(\theta-\theta_0\right) / \sigma$. Then the test can be rewritten as a test of $H_0: \mu=0$, which is a simple hypothesis. under H_1, **the hierarchical prior defined earlier becomes:**$\pi_1(\mu \mid \xi)$ is $\mathcal{N}(0, \xi)$, while $\xi>0$ still has proper prior $g(\xi)$. The implied prior, $\pi(\mu)=\int \pi_1(\mu \mid \xi) g(\xi) d \xi$

in this case, the marginal density of y_n under H_0 and H_1, respectively, becomes $m_{0, n}\left(y_n\right)=m\left(y_n \mid 0\right)$ and $m_{1, n}\left(y_n\right)=\int m\left(y_n \mid \xi\right) g(\xi) d \xi$, where $K_n=\Gamma\left(\frac{n}{2}\right)(n-1)^{(n-1) / 2} / \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n}{2}-\frac{1}{2}\right)$ and

marginal density

$$
4.3: m\left(y_n \mid \xi\right)=\int f\left(y_n \mid \mu\right) \pi_1(\mu \mid \xi) d \mu=K_n \frac{{y_n}^{-1 / 2}(1+n \xi)^{(n-1) / 2}}{\left[(n-1)(1+n \xi)+y_n\right]^{n / 2}}
$$

prior：

$$
4.4: g(\xi)=(2 \pi)^{-\frac{1}{2}} \xi^{-\frac{3}{2}} \exp \left\{-\frac{1}{2 \xi}\right\}, \quad \xi>0
$$

For some predetermined stopping boundaries R and $A (R < 1 < A)$, we consider, as in Section $1 \cdot 2$, the truncated at m stopping time N_m, discussed in Example 2 . Along the same lines, it can be easily verified that the test T^* in $(3 \cdot 5)$ applies here with $R_U = A_L = 1$, so that, by $(3 \cdot 4)$, a satisfies the equation

$$F_0(a) = 1 - F_1(1)$$

A	m	a	α'	β'	p_0	p_1	$E_0(N_m)$	$E_1(N_m)$	$var_0(N_m)$	$var_1(N_m)$
						$R = 0.1$				
8	50	3.670	0.048	0.116	0.129	0.059	44.4	16.6	92	393
	100	6.050	0.042	0.103	0.094	0.035	56.2	21.9	551	936
	200	8.014	0.038	0.104	0.084	0.017	62.7	25.9	1581	1845
	300	8.016	0.035	0.104	0.083	0.013	64.9	27.0	2330	2428
9	50	3.640	0.049	0.117	0.128	0.059	48.0	17.3	83	420
	100	4.540	0.045	0.097	0.091	0.036	67.3	23.6	517	1079
	200	9.008	0.040	0.094	0.073	0.019	77.7	28.5	1901	2208
	300	9.015	0.037	0.094	0.071	0.012	80.8	30.4	3005	3103
10	50	3.630	0.048	0.117	0.129	0.060	48.1	17.2	78	421
	100	4.010	0.046	0.095	0.089	0.039	77.6	24.8	453	1195
	200	7.850	0.042	0.086	0.065	0.021	92.7	31.3	2070	2644
	300	10.008	0.040	0.086	0.061	0.013	97.9	34.8	3651	3960
						$R = 0.05$				
8	50	3.775	0.039	0.115	0.139	0.063	44.8	17.9	81	416
	100	6.075	0.031	0.103	0.108	0.038	57.1	123.9	545	1025
	200	8.016	0.023	0.104	0.101	0.020	64.1	29.0	1648	2173
	300	8.018	0.020	0.104	0.100	0.016	66.8	30.4	2577	2940
9	50	3.795	0.040	0.115	0.138	0.062	48.0	18.3	69	437
	100	4.875	0.033	0.096	0.106	0.042	68.3	26.1	498	1176
	200	9.013	0.025	0.094	0.092	0.021	55.1	31.8	1933	2585
	300	9.019	0.021	0.094	0.089	0.015	55.1	35.3	3216	3913
10	50	3.775	0.039	0.115	0.141	0.064	48.5	18.4	63	435
	100	4.345	0.034	0.093	0.102	0.044	78.6	27.5	411	1263
	200	10.003	0.026	0.086	0.084	0.024	94.3	35.2	2069	3027
	300	10.011	0.023	0.086	0.078	0.016	100.6	39.5	3847	4759

QUESTION

1. How to explain this table
2. at Hierarchical Bayesian