



Bayesian Estimation of Short Term Detection Rate 2021

통계학과 석사 2022021325 이은이
통계학과 학부 2018150420 정해원

목차



1. Short Term Detection 정의
2. Dataset 소개
3. Data 전처리
4. 살인 기수 범죄의 Bayesian Approach
5. Hierarchical Bayesian Model of Each Crime and Field
6. 4-stage Hierarchical Bayesian Model and Total Joint Posterior

1

Definition of Short Term Detection

단기 검거 범죄의 정의

1. Definition of Short Term Detection

초록 ▲

본 연구는 우리나라 4개 유형 14개 범죄에 대해 범죄발생부터 검거까지의 소요시간을 추정한 결과 다음과 같은 연구결과를 도출하였다.

첫째, 검거시간을 기준으로 단기 검거 유형은 검거 소요일수가 35일 미만으로 살인기수, 살인미수 등, 방화, 상해, 폭행, 폭력행위 등, 강도, 손괴 등이 해당된다. 중기 검거유형은 35일 이상에서 80일 미만으로 절도, 체포·감금, 협박, 약취·유인이 해당된다. 끝으로 장기 검거 유형은 80일 이상으로 공갈과 마약이 상대적으로 검거에 어려움이 많은 것으로 분석되었다.

- 통상적으로 단기 검거는 **범죄 발생 이후 35일 이내 검거되는 것을 의미함**
- 범죄 유형별로 단기 검거 비율이 다름
- 단기 검거 비율을 parameter로 하는 Bayesian 분석을 진행 예정

2

Dataset 소개

Dataset 소개 – 범죄사건처리기간2021



KOSIS – 경찰의 범죄사건 처리기간 데이터 중 2021년도의 데이터

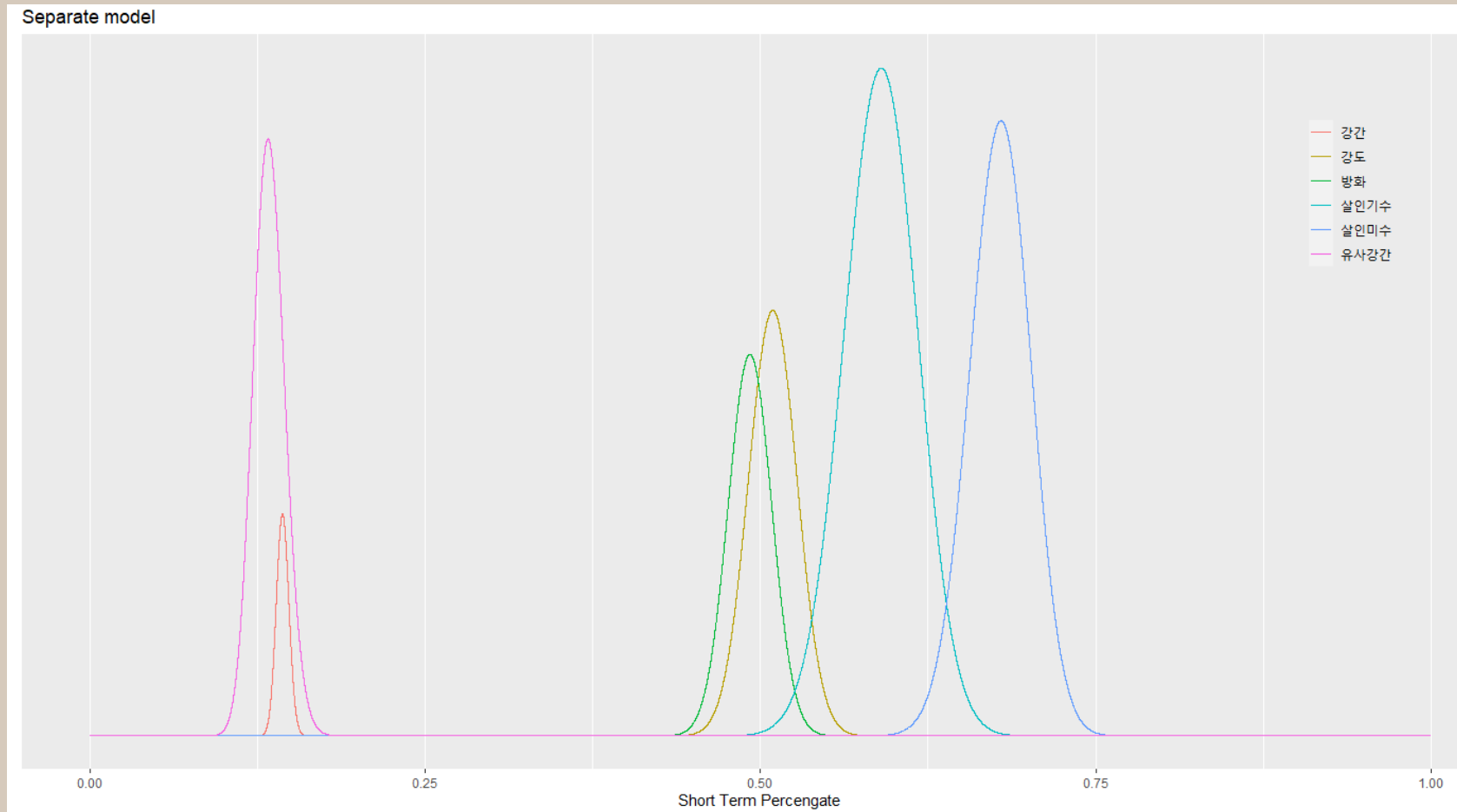
죄종별(1)	죄종별(2)	2021	2021	2021	2021	2021	2021	2021	2021
죄종별(1)	죄종별(2)	계	10일이내	20일이내	1개월이내	2개월이내	3개월이내	6개월이내	6개월초과
계	소계	1,370,722	216,067	191,813	144,908	277,259	172,031	238,432	130,21
강력범죄	소계	23,306	2,302	1,524	1,702	5,729	4,801	5,647	1,60
강력범죄	살인기수	322	161	23	6	27	19	33	5
강력범죄	살인미수등	421	239	30	17	25	35	52	2
강력범죄	강도	766	286	69	35	99	106	82	8
강력범죄	강간	5,695	459	154	205	1,035	1,206	2,002	63
강력범죄	유사강간	843	61	22	29	196	202	282	5
강력범죄	강제추행	14,030	806	1,081	1,275	4,126	3,104	2,998	64
강력범죄	기타 강간·강제추행등	276	67	14	20	38	24	72	4
강력범죄	방화	953	223	131	115	183	105	126	7
절도범죄	소계	87,567	12,951	15,107	12,249	21,957	11,234	10,431	3,63
폭력범죄	소계	269,456	35,114	45,286	36,579	69,661	36,854	34,649	11,31
폭력범죄	상해	36,006	3,993	5,367	4,798	9,906	5,607	4,998	1,33
폭력범죄	폭행	146,962	20,934	28,580	21,721	38,651	18,396	14,633	4,04
폭력범죄	체포·감금	1,573	199	140	113	357	217	382	16
폭력범죄	협박	22,483	3,282	3,433	2,982	5,980	3,122	2,774	91
폭력범죄	약취·유인	322	44	34	26	77	32	81	2
폭력범죄	폭력행위등	21,075	1,285	1,364	1,589	4,757	3,893	5,930	2,25
폭력범죄	공갈	4,445	275	156	189	590	597	1,454	1,18
폭력범죄	손괴	36,590	5,102	6,212	5,161	9,343	4,990	4,397	1,38
지능범죄	소계	284,468	22,894	11,579	12,350	37,005	39,748	92,827	68,06

각 범죄 분야별, 죄목 별로 신고 접수일에서 검거일까지 기간 별 검거 수가 수록되어있다

이 연구에서는 범죄 접수 시점부터 1개월 이내 검거된 범죄 수를 Short Term Detection으로 정의한다

Dataset 소개 – 범죄사건처리기간2021

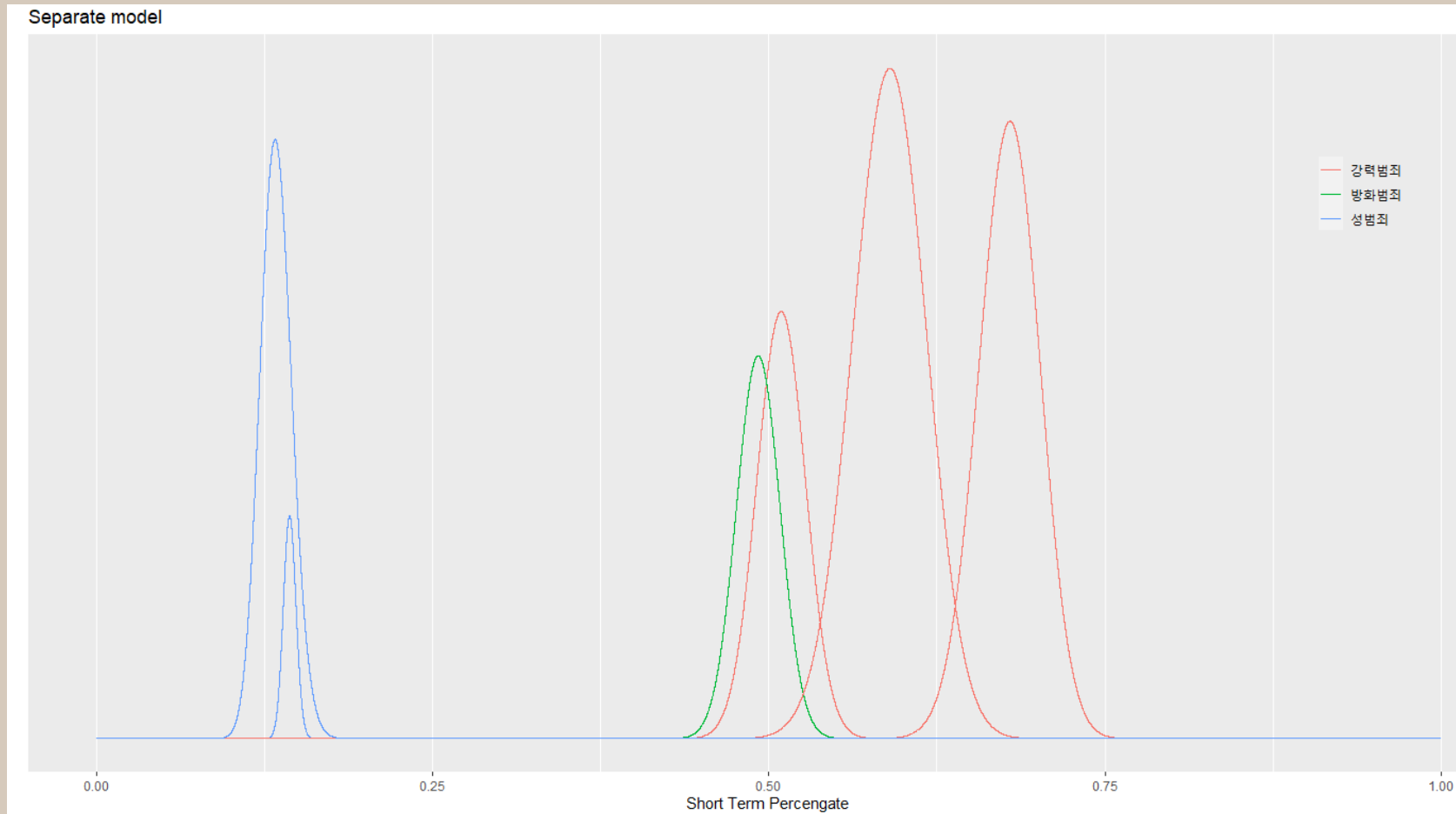
KOSIS – 경찰의 범죄사건 처리기간 데이터 중 2021년도의 데이터



범죄 죄목별로 단기 검거율에 차이가 존재한다

Dataset 소개 – 범죄사건처리기간2021

KOSIS – 경찰의 범죄사건 처리기간 데이터 중 2021년도의 데이터



범죄 분야별로 단기 검거율에 차이가 분명하게 드러난다

3

Data 전처리

전처리 1 – 분석에 사용될 범죄 선별



- 전체 범죄 중에서 의미 있는 일부 범죄만 분석에 사용함
- 범죄 분야(유사한 범죄들끼리 그룹화) 를 Field, 개별 죄목을 crime이라 정의한다

Field	Crime
강력범죄	살인기수, 살인미수, 강도
성범죄	강간, 유사강간, 강제추행
방화범죄	방화
절도범죄	절도
신체폭력	상해, 폭행, 체포/감금, 협박
재산폭력	공갈, 손괴
지능범죄	사기, 횡령
풍속범죄	도박
마약범죄	마약
교통범죄	교통범죄

전처리 2 - 데이터 정수화 / 변수 생성



- 1개월 이내 검거된 범죄 : Short_Term, 나머지를 Long_Term로 변수 변환하여 Binomial로 취급한다

$$\text{Short Term} \sim \text{Binom}(\text{Total}_i, \theta_i)$$

```
> Crime <- Crime %>% mutate(Short_Term = D10+D20+M1, Long_Term = Total - Short_Term) %>% select(Field, crime, Total, Short_Term, Long_Term)
> Crime
# A tibble: 19 × 5
```

	Field <chr>	crime <chr>	Total <int>	Short_Term <int>	Long_Term <int>
1	강력범죄	살인기수	322	190	132
2	강력범죄	살인미수등	421	286	135
3	강력범죄	강도	766	390	376
4	강력범죄	강간	5695	818	4877
5	강력범죄	유사강간	843	112	731
6	강력범죄	강제추행	14030	3162	10868
7	강력범죄	방화	953	469	484
8	절도범죄	소계	87567	40307	47260
9	폭력범죄	상해	36006	14158	21848
10	폭력범죄	폭행	146962	71235	75727
11	폭력범죄	체포 · 감금	1573	452	1121
12	폭력범죄	협박	22483	9697	12786
13	폭력범죄	공갈	4445	620	3825
14	폭력범죄	손괴	36590	16475	20115
15	지능범죄	사기	231529	34559	196970
16	지능범죄	횡령	29852	8135	21717
17	풍속범죄	도박범죄	15007	6682	8325
18	마약범죄	소계	10623	3196	7427
19	교통범죄	소계	316934	210694	106240

4

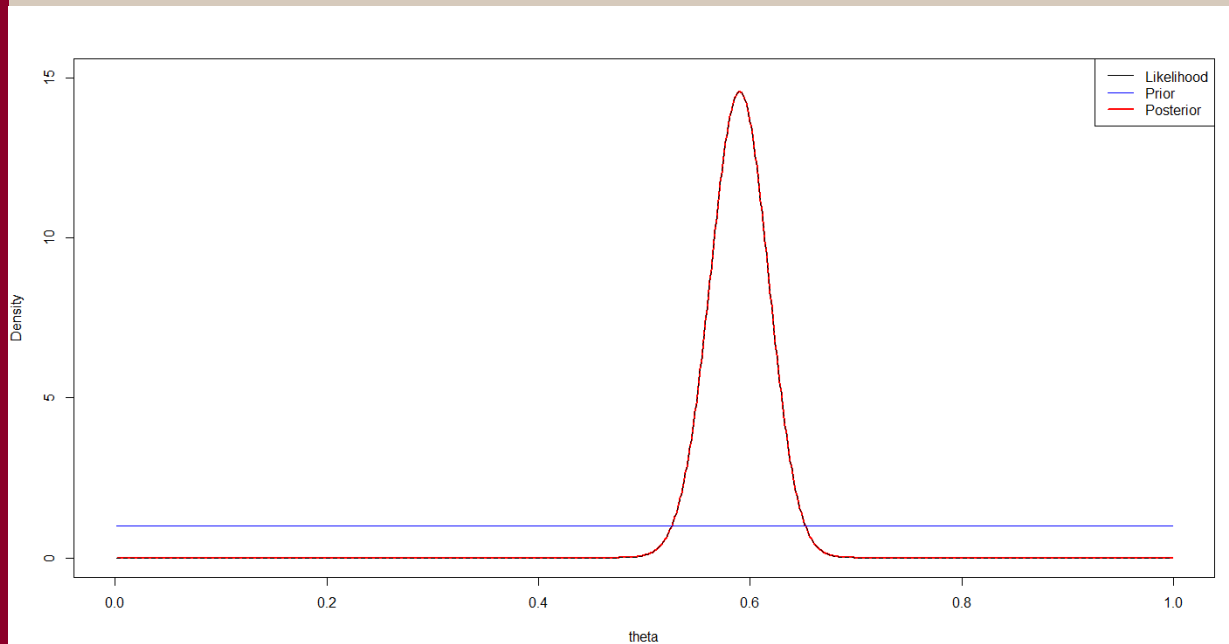
살인기수 범죄의 Bayesian Approach

살인기수 범주의 Bayesian Approach

- 총 322회의 살인기수 범죄 발생 건수 중 190회 1개월 이내 검거

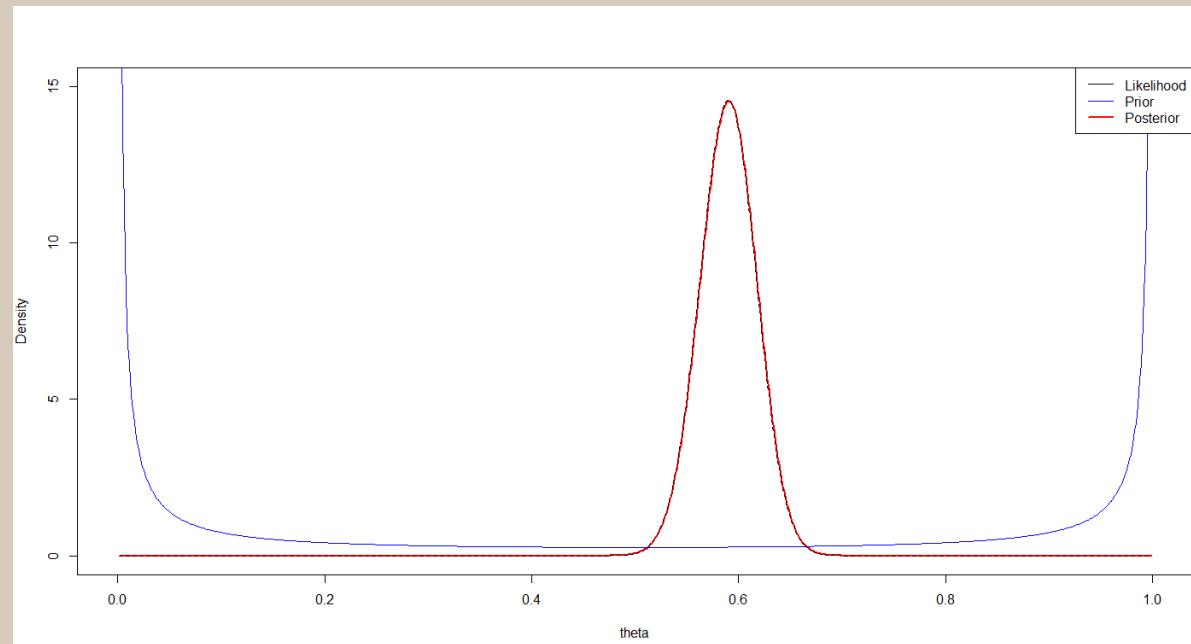
단기 검거율 $X|\theta \sim \text{Binom}(322, 0.5900)$

$\theta \sim \text{Beta}(1, 1)$, Uniform Prior



Mean : 0.58951
95% Bayesian CI : [0.5355, 0.6424]

$\theta \sim \text{Beta}(0, 0)$, Hadlane's Prior



Mean : 0.59006
95% Bayesian CI : [0.5359, 0.6431]

살인기수 범죄의 Bayesian Approach

- 총 322회의 살인기수 범죄 발생 건수 중 190회 1개월 이내 검거

살인 기수 범죄는 보통
단기간에 검거된다(긍정)

VS

살인 기수 범죄는
단기간에 검거되지 않는다(부정)

각자의 생각을 반영한 Prior을 설정해준다

보통 25건 중 18건 정도는 단기에 검거

VS

50건의 사건이 있다면 겨우 5건만 단기 검거

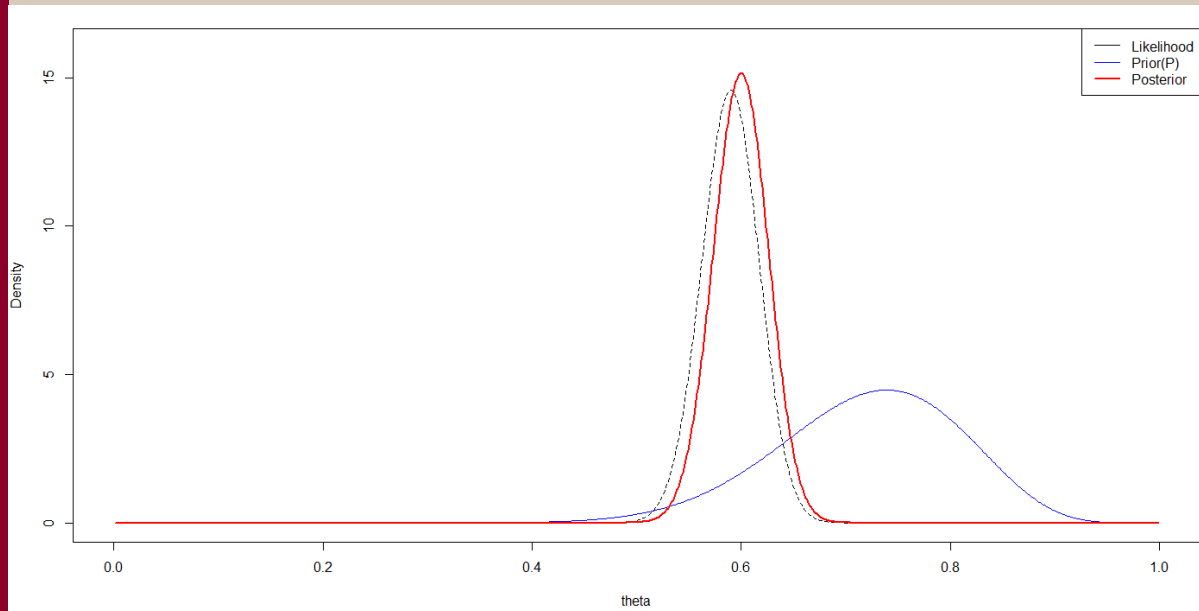
$$\theta_+ \sim \text{Beta}(19, 8)$$

$$\theta_- \sim \text{Beta}(6, 46)$$

살인기수 범죄의 Bayesian Approach

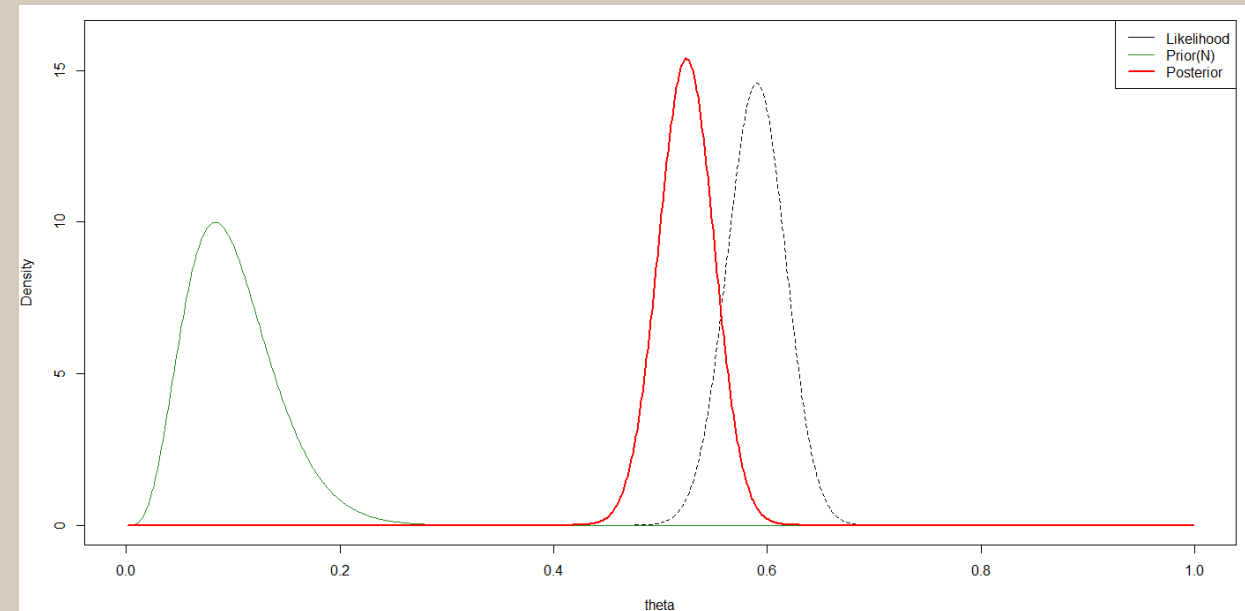
- 총 322회의 살인기수 범죄 발생 건수 중 190회 1개월 이내 검거

$$\theta_+ \sim \text{Beta}(19, 8)$$



Mean : 0.59942
95% Bayesian CI : [0.5474, 0.6503]

$$\theta_- \sim \text{Beta}(6, 46)$$



Mean : 0.52419
95% Bayesian CI : [0.4734, 0.5747]

5

Hierarchical Bayesian Model of Each Crime and Field

각 범죄 Field의 계층적 베이지안 모델

Hierarchical Bayesian Model – About Prior

3 Different Priors

❶ 해당 범죄에 대한 사전 지식이 없음

- Uniform Prior ($\theta \sim \text{Beta}(1,1)$)

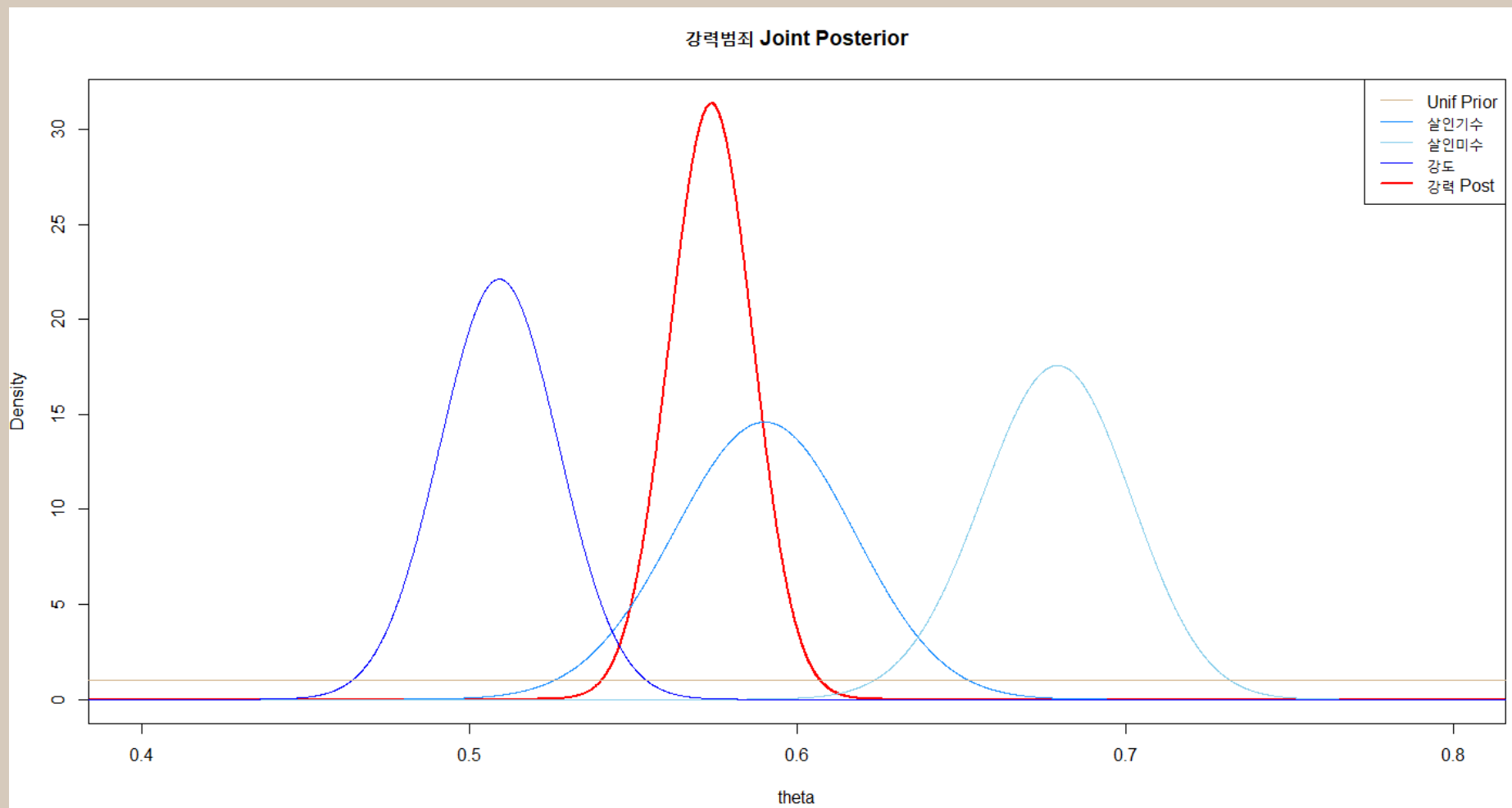
❷ 해당 범죄의 검거율이 높을 것이라고 믿음

- Positive Beta Prior ($\theta \sim \text{Beta}(71,31)$)

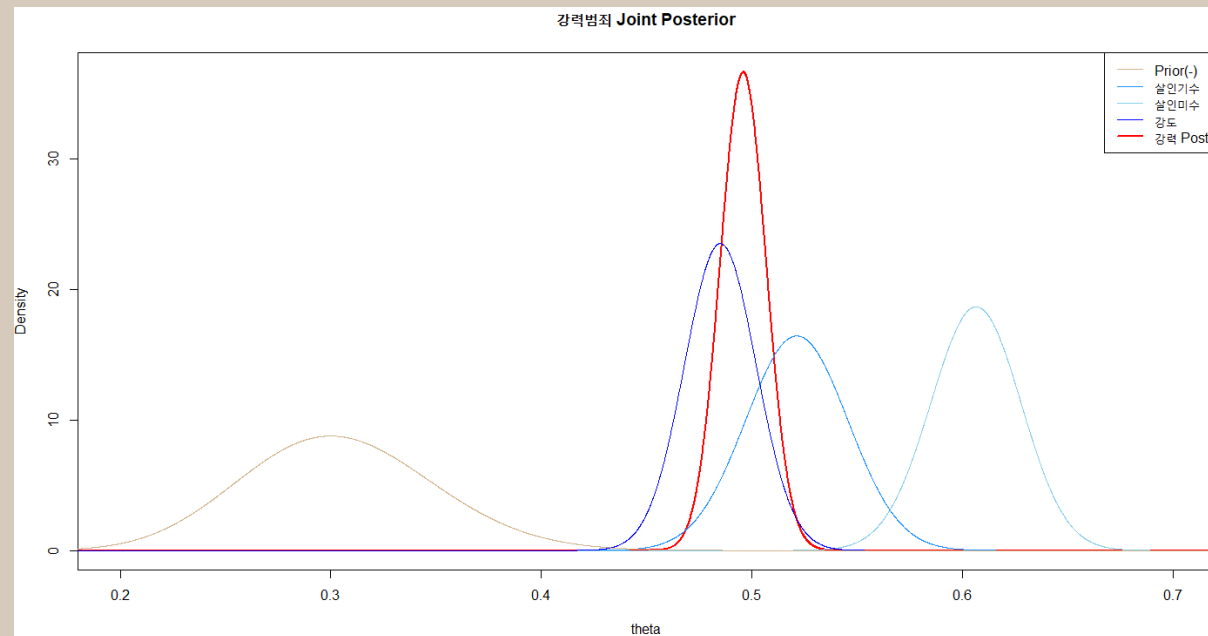
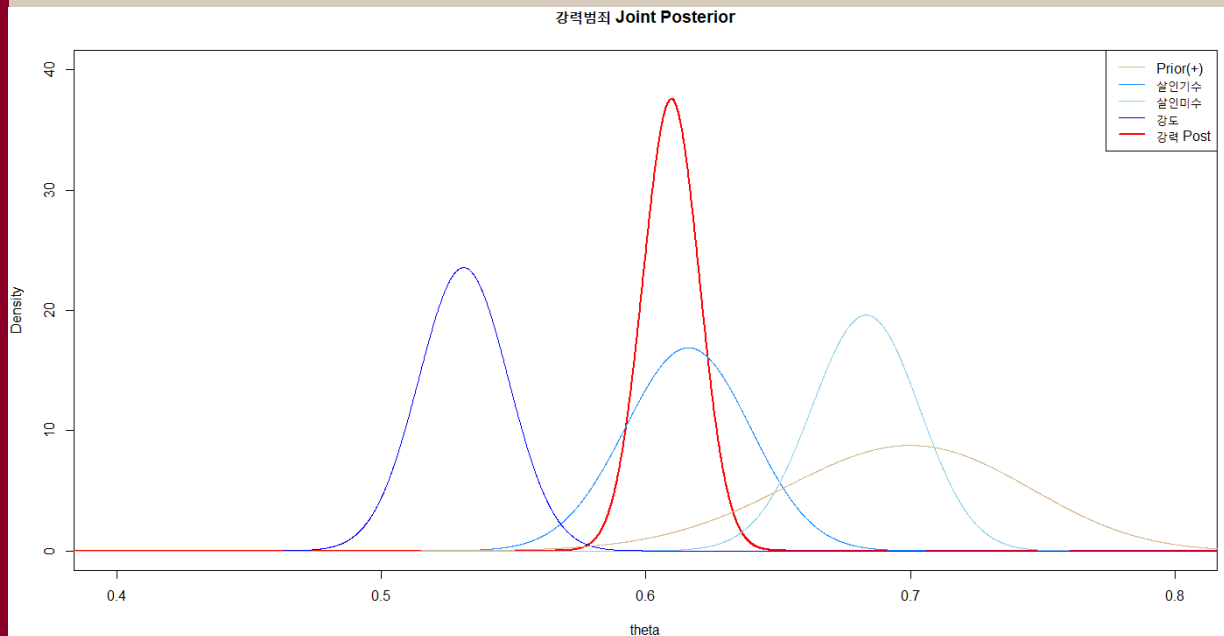
❸ 해당 범죄의 검거율이 낮을 것이라고 믿음

- Negative Beta Prior ($\theta \sim \text{Beta}(31,71)$)

Hierarchical Bayesian Model – Field 1 : 강력범죄



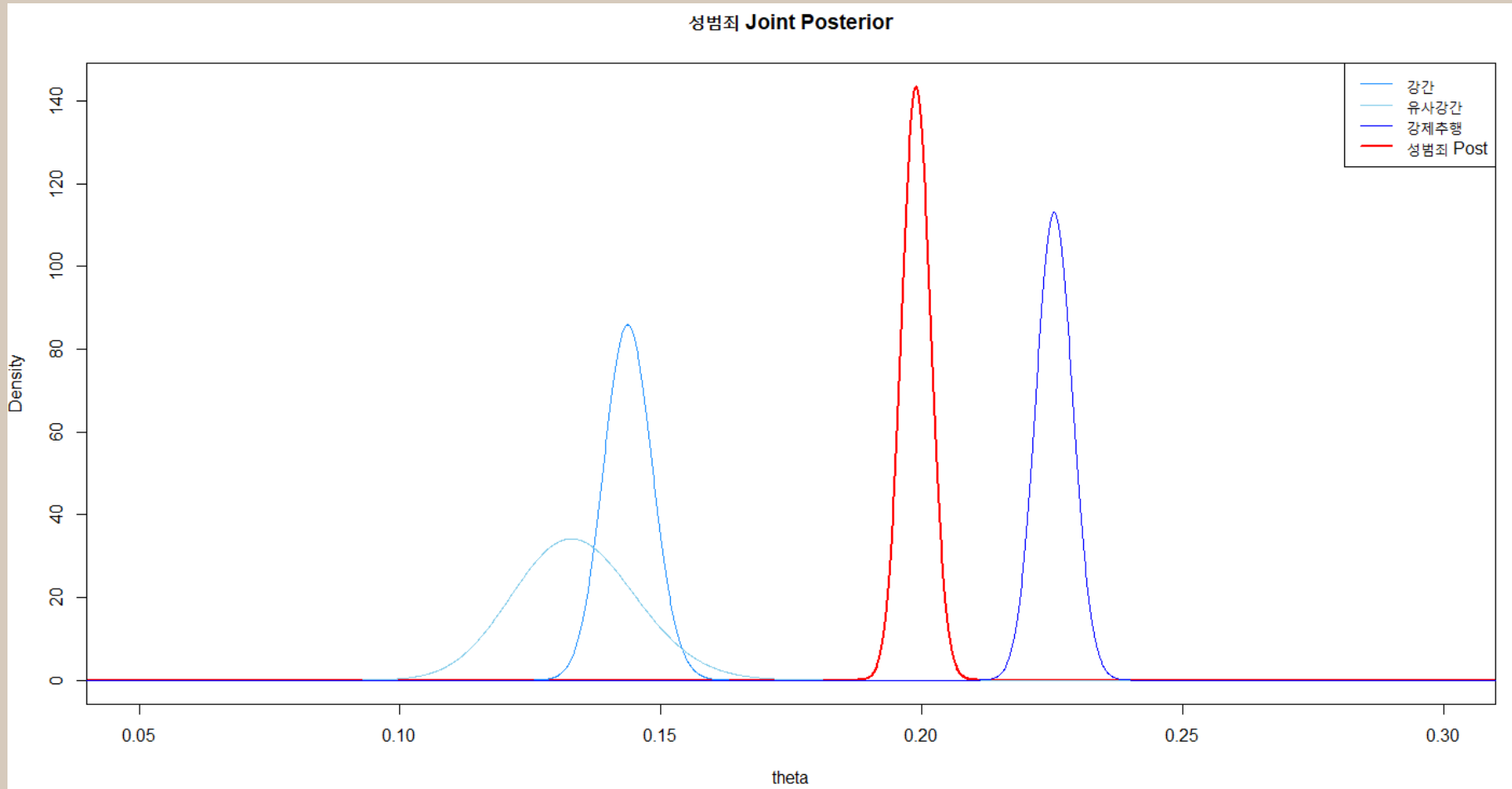
Hierarchical Bayesian Model – Field 1 : 강력범죄



Prior의 영향을 받아 mean이 변하고 variance 는 감소한다

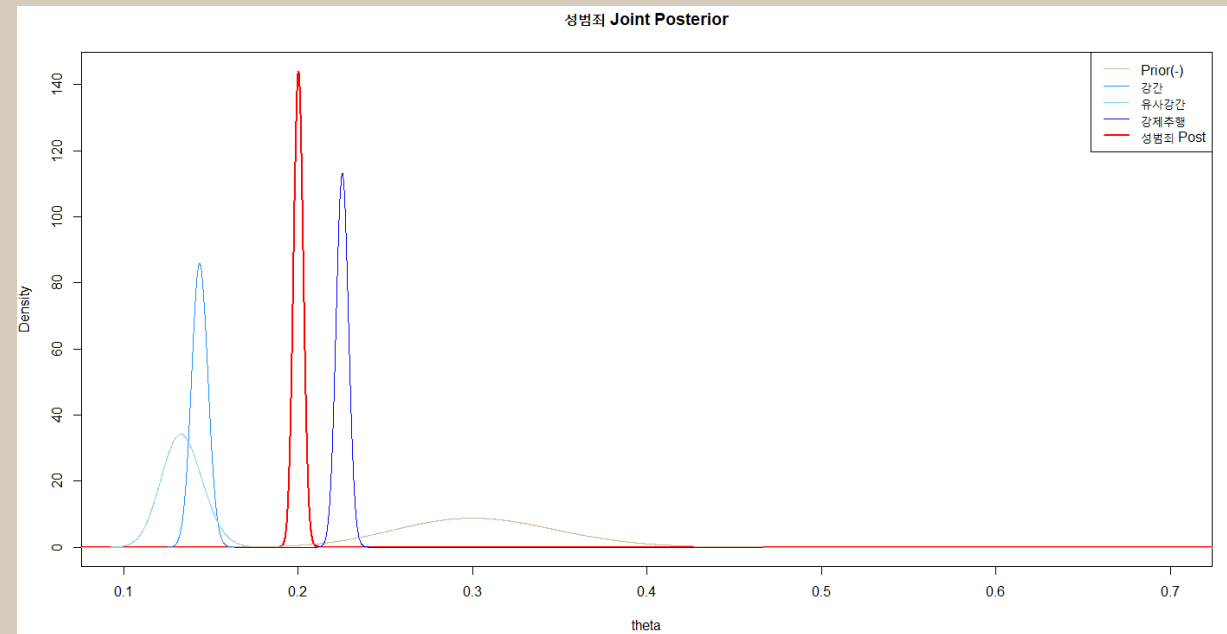
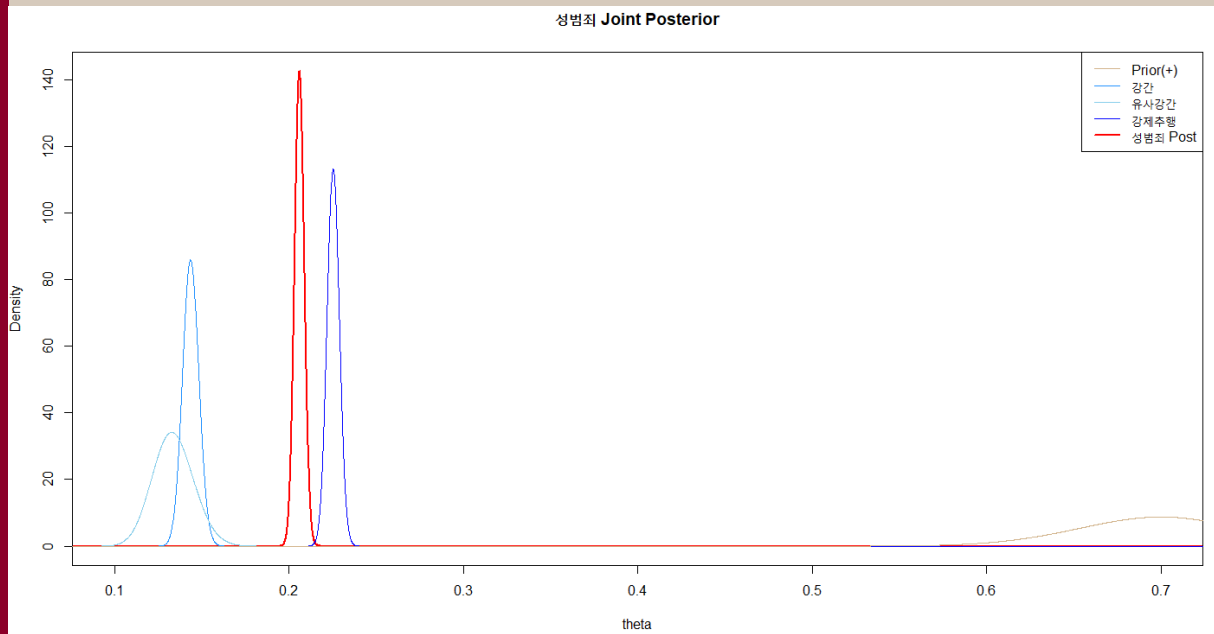
그러나 Prior의 영향력이 적다

Hierarchical Bayesian Model – Field 2 : 성범죄



성범죄 Field의 Joint Posterior mean이 강력범죄 Field의 posterior mean보다 작다

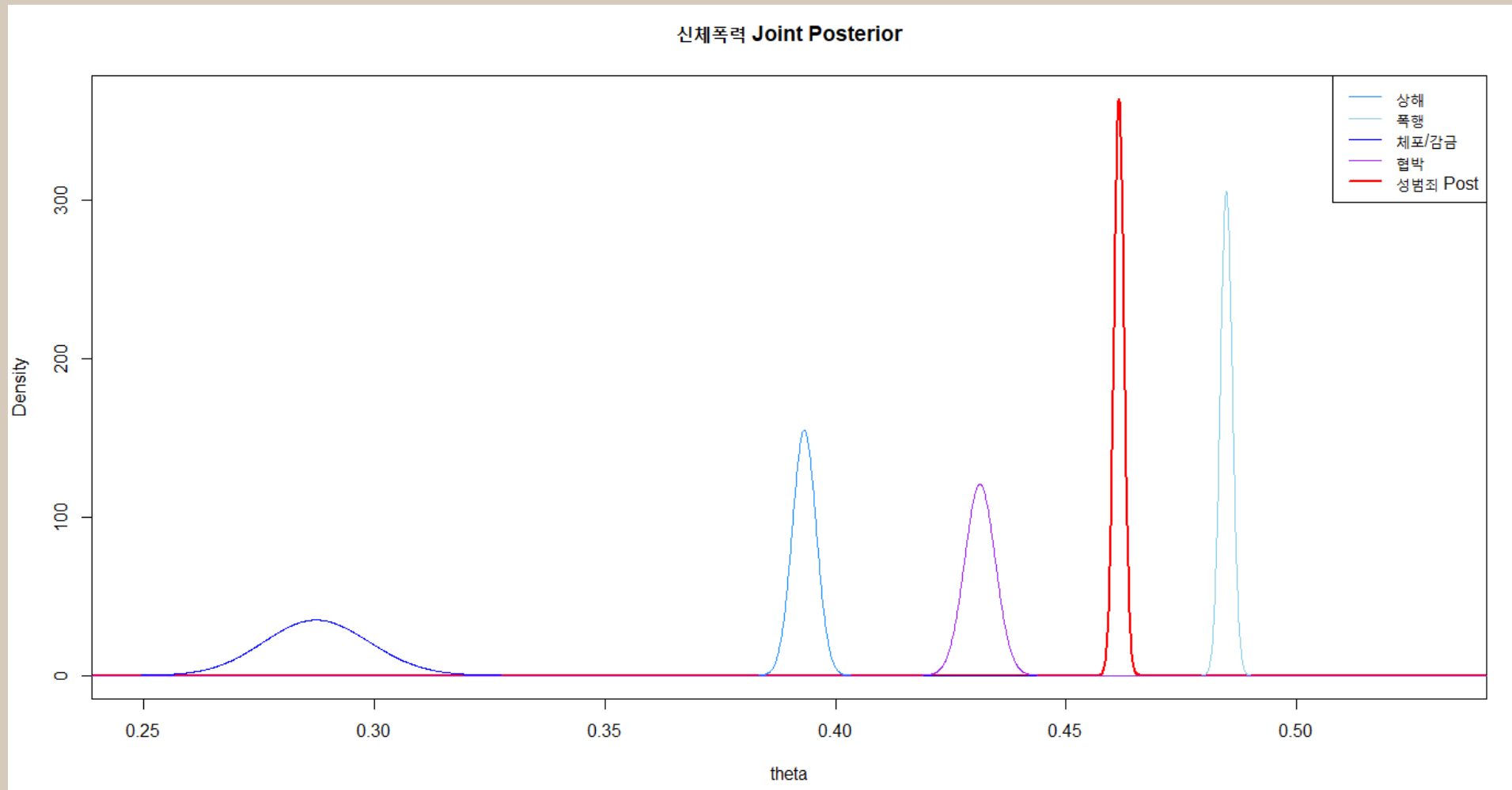
Hierarchical Bayesian Model – Field 2 : 성범죄



Prior를 반영해도 큰 차이가 나지 않는다

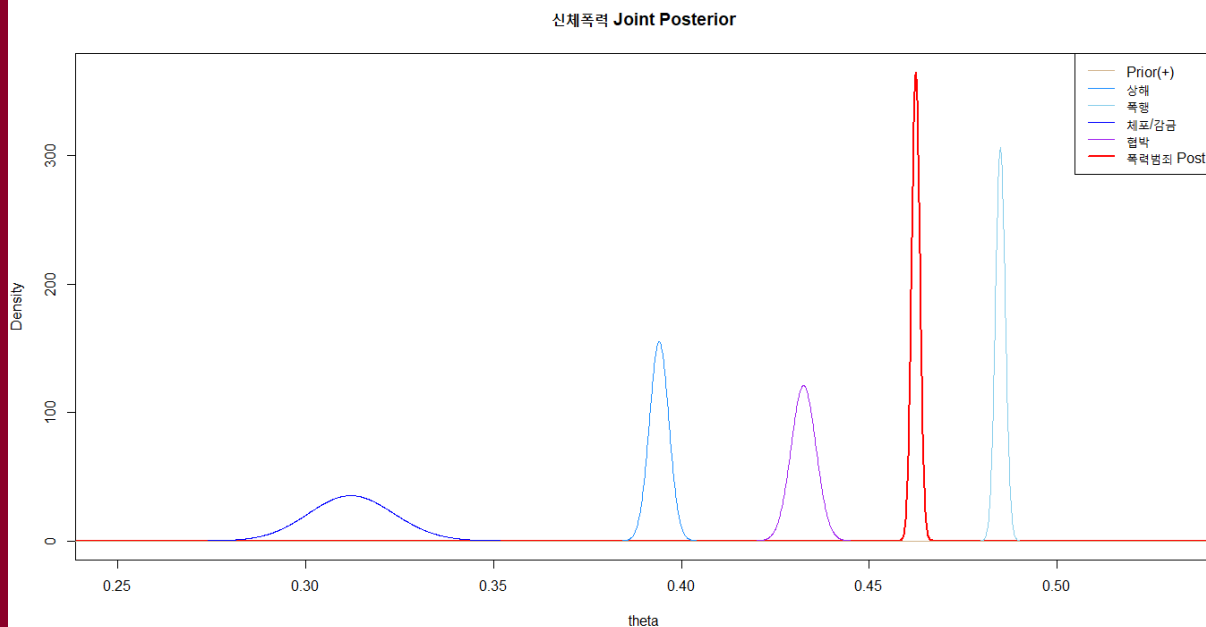
성범죄 Field의 범죄들이 강력범죄에 비해 단기 검거되는 비율이 낮다

Hierarchical Bayesian Model – Field 3 : 신체폭력범죄

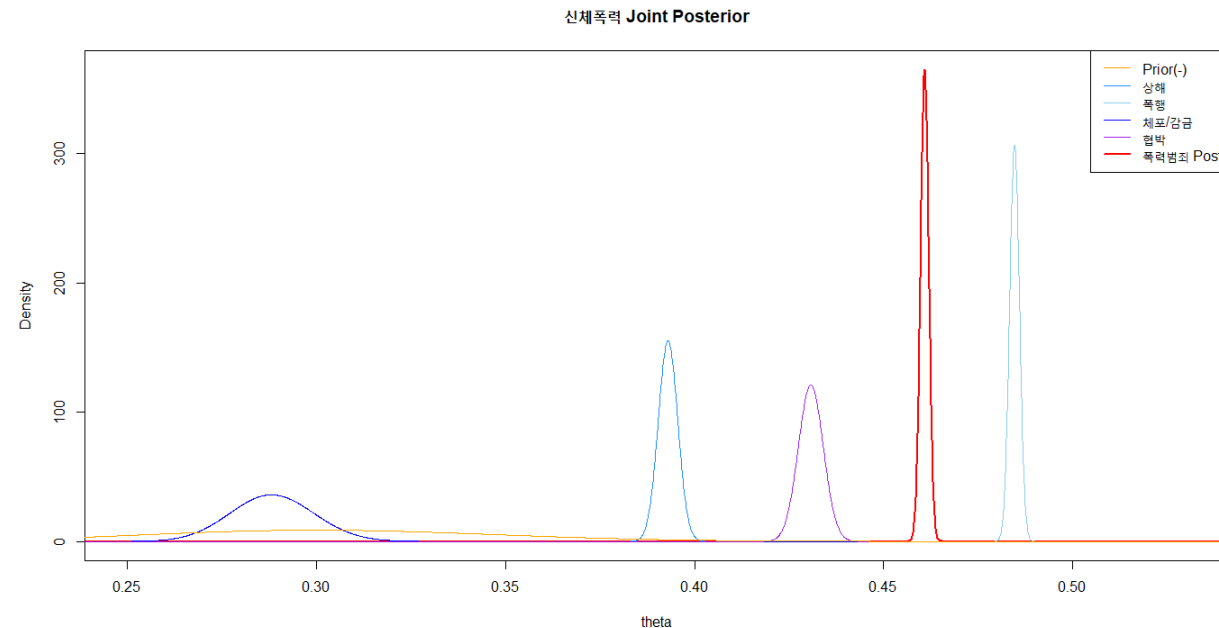


신체폭력범죄 Field의 Joint Posterior mean이 강력범죄의 mean보다는 작고, 성범죄 mean보다는 크다

Hierarchical Bayesian Model – Field 3 : 신체폭력범죄



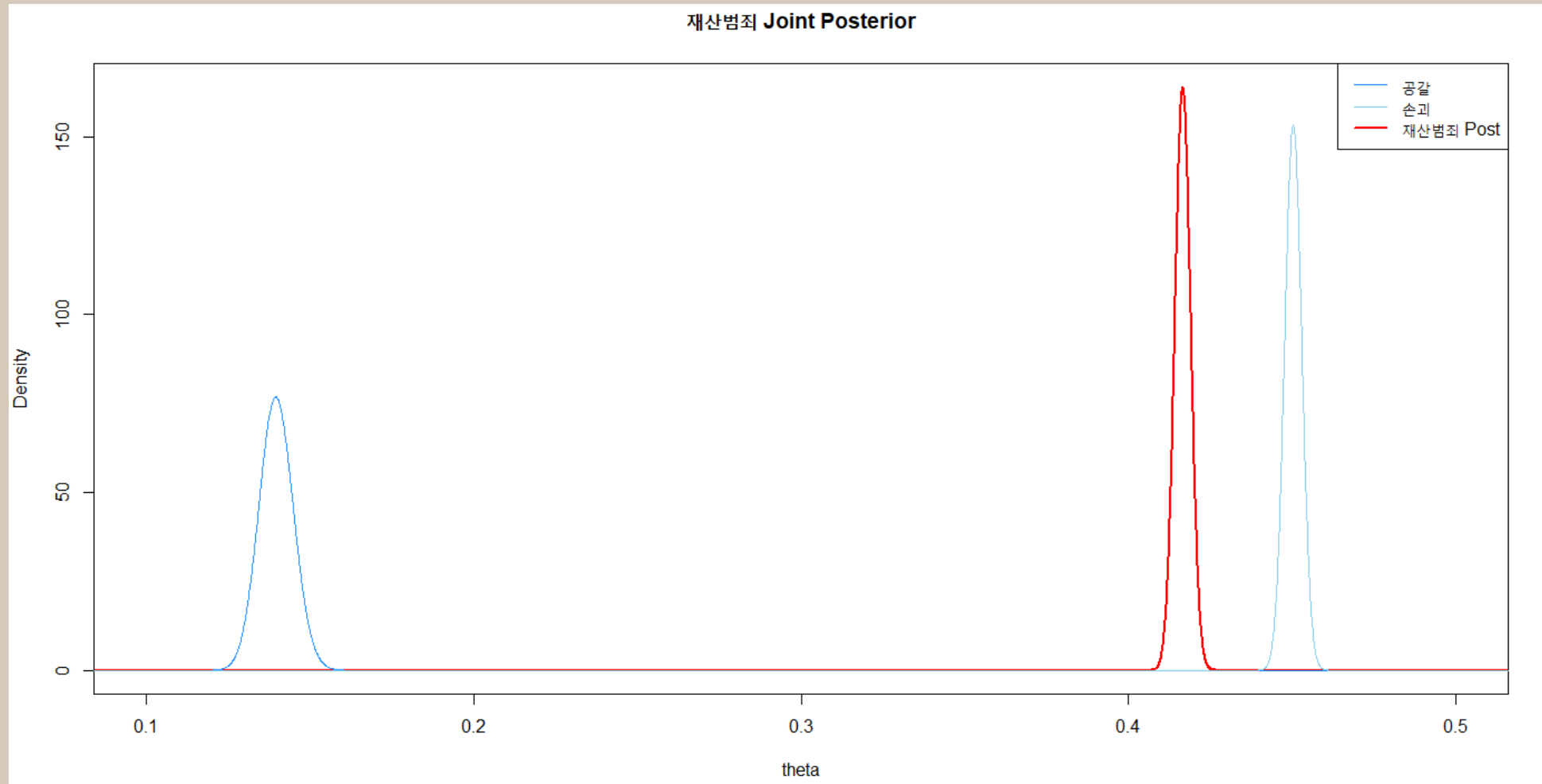
Mean : 0.4614244 Mode : 0.4614241
Variance : 1.1998×10^{-6}



Mean : 0.4616176 Mode : 0.4616172
Variance : 1.1998×10^{-6}

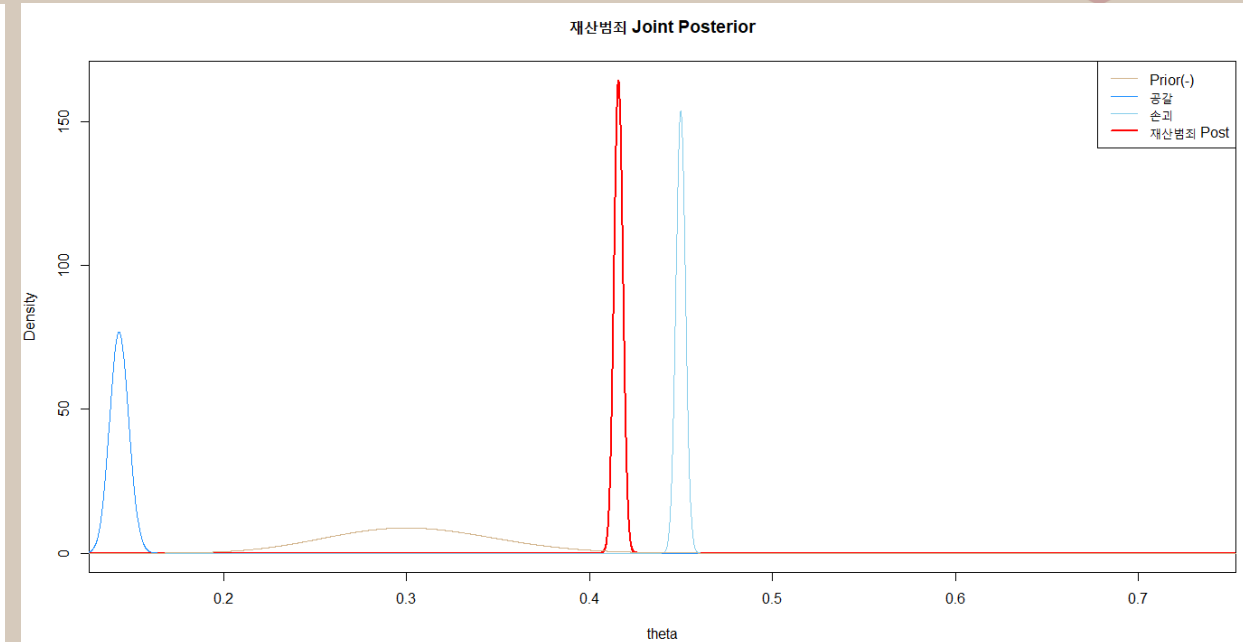
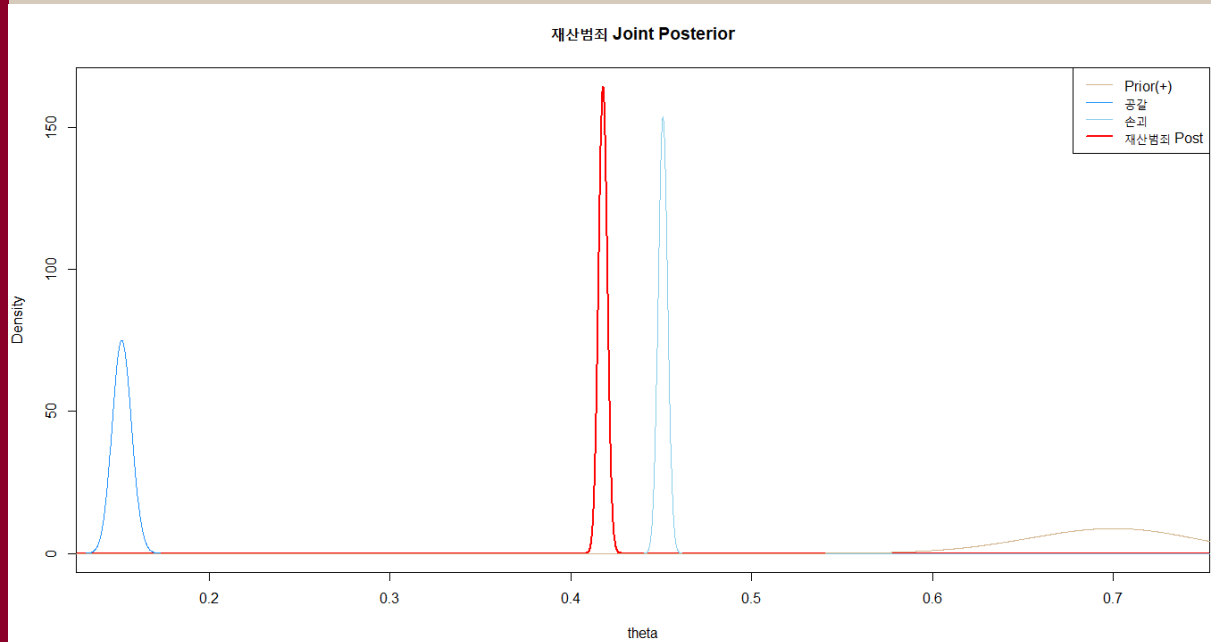
Prior를 반영해도 차이가 미미하며, Mean과 mode의 차이가 매우 작다

Hierarchical Bayesian Model – Field 4 : 재산폭력범죄



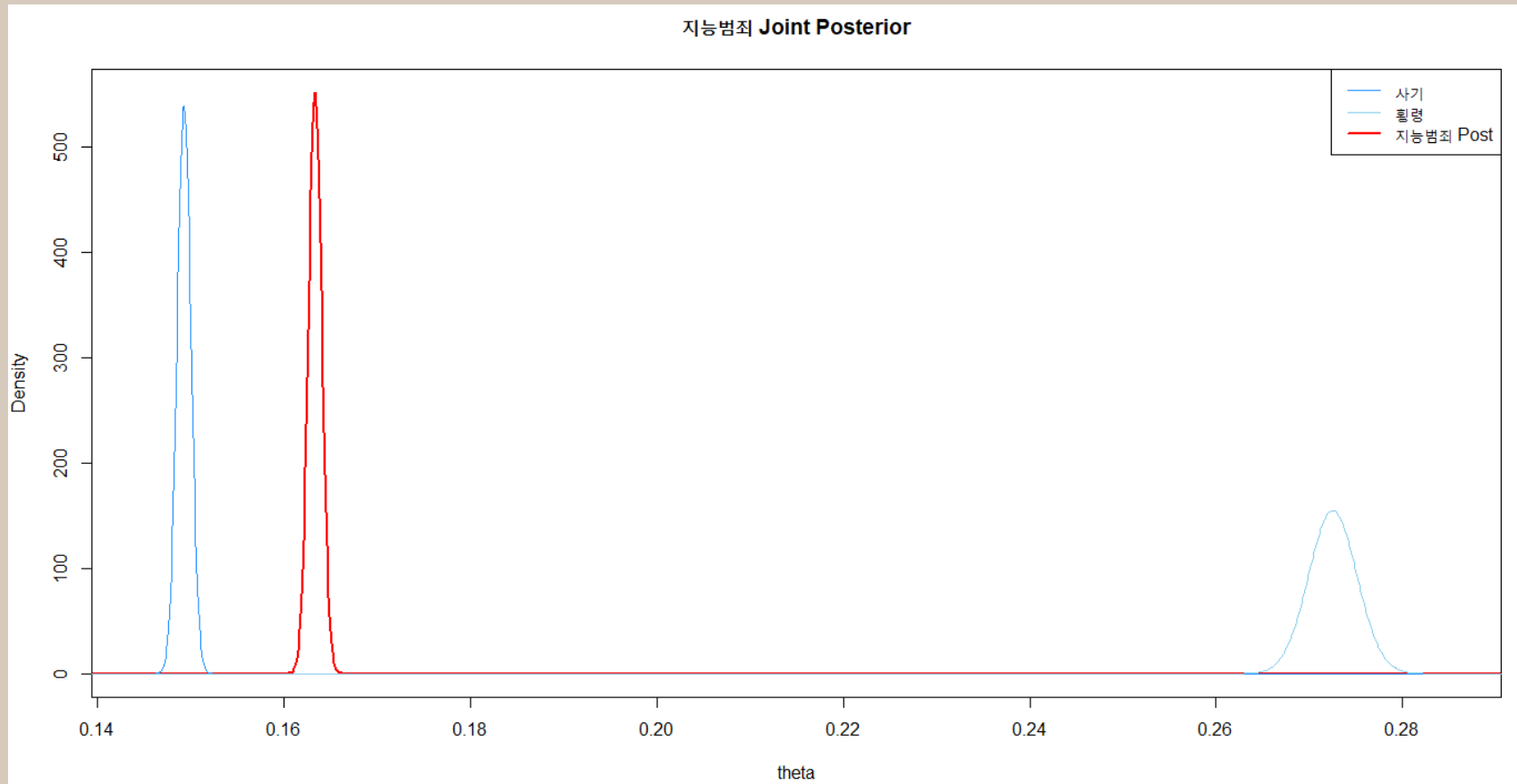
빈도수가 공갈죄(연 4445회 발생)에 비해 더 많은 손괴죄(연 36590회 발생)의 영향으로 Joint Posterior가 손괴죄의 likelihood에 가깝게 편향되어 있다.

Hierarchical Bayesian Model – Field 4 : 재산폭력범죄



Prior를 반영해도 차이가 미미하며, Mean과 mode의 차이가 매우 작다

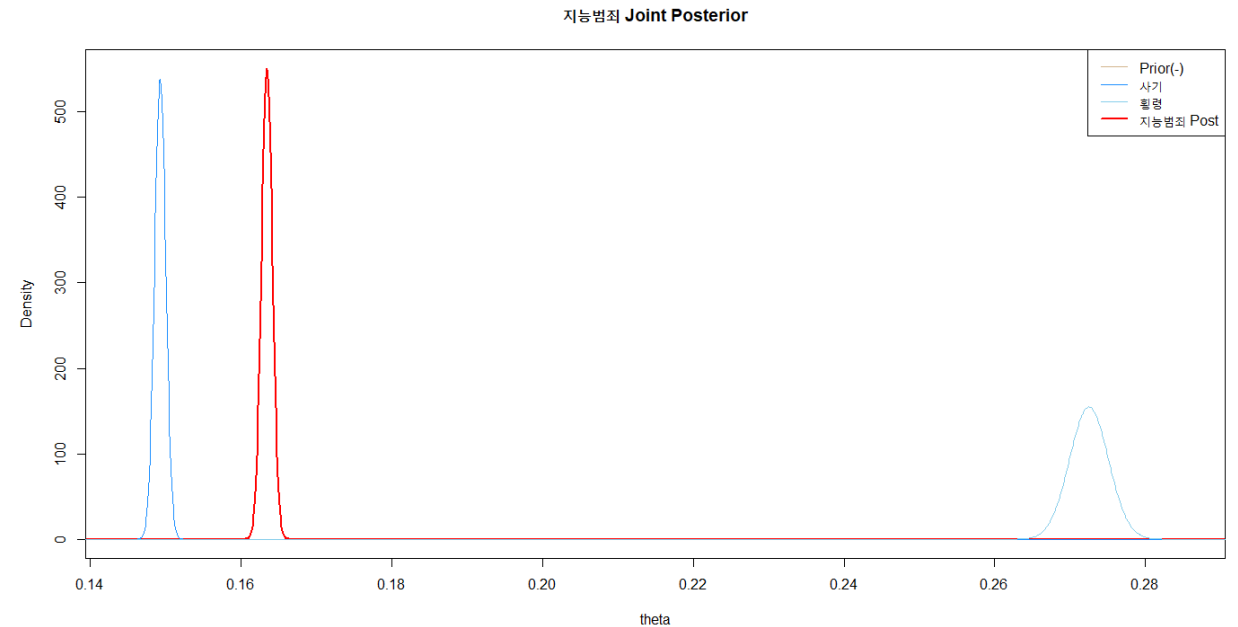
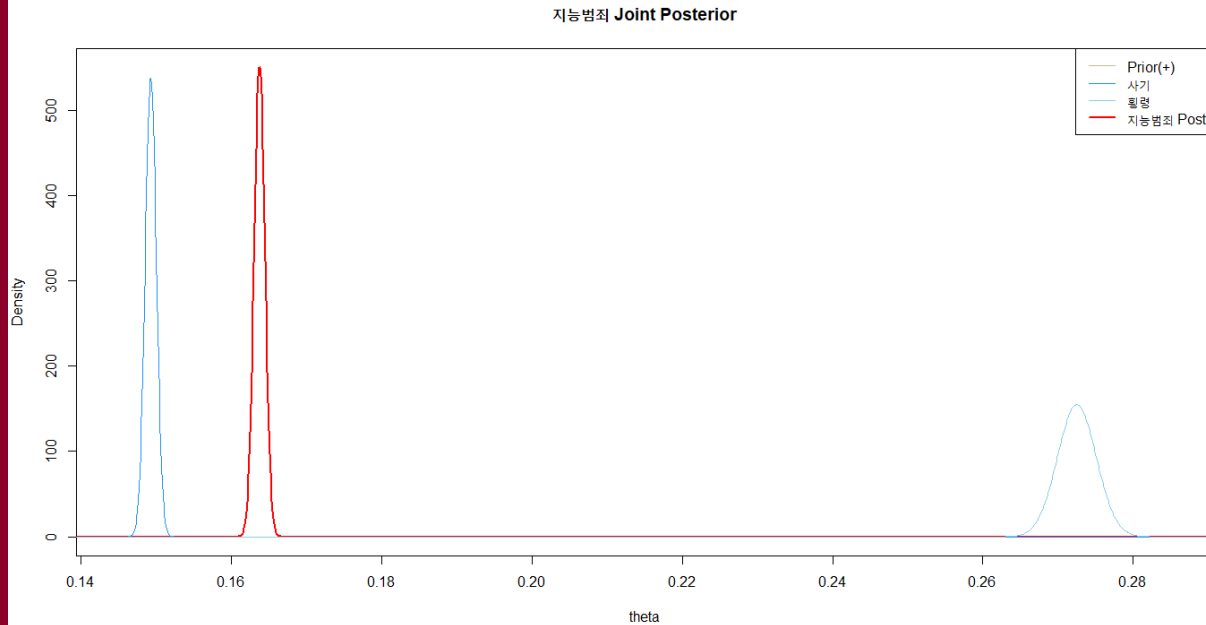
Hierarchical Bayesian Model – Field 5 : 지능범죄



빈도수가 횡령죄(연 29852회 발생)에 비해 더 많은 사기죄(연 231529회 발생)의 영향으로 Joint Posterior가 사기죄의 likelihood에 가깝게 편향되어 있다.

사기죄는 likelihood 단계에서부터 Variance가 매우 작다

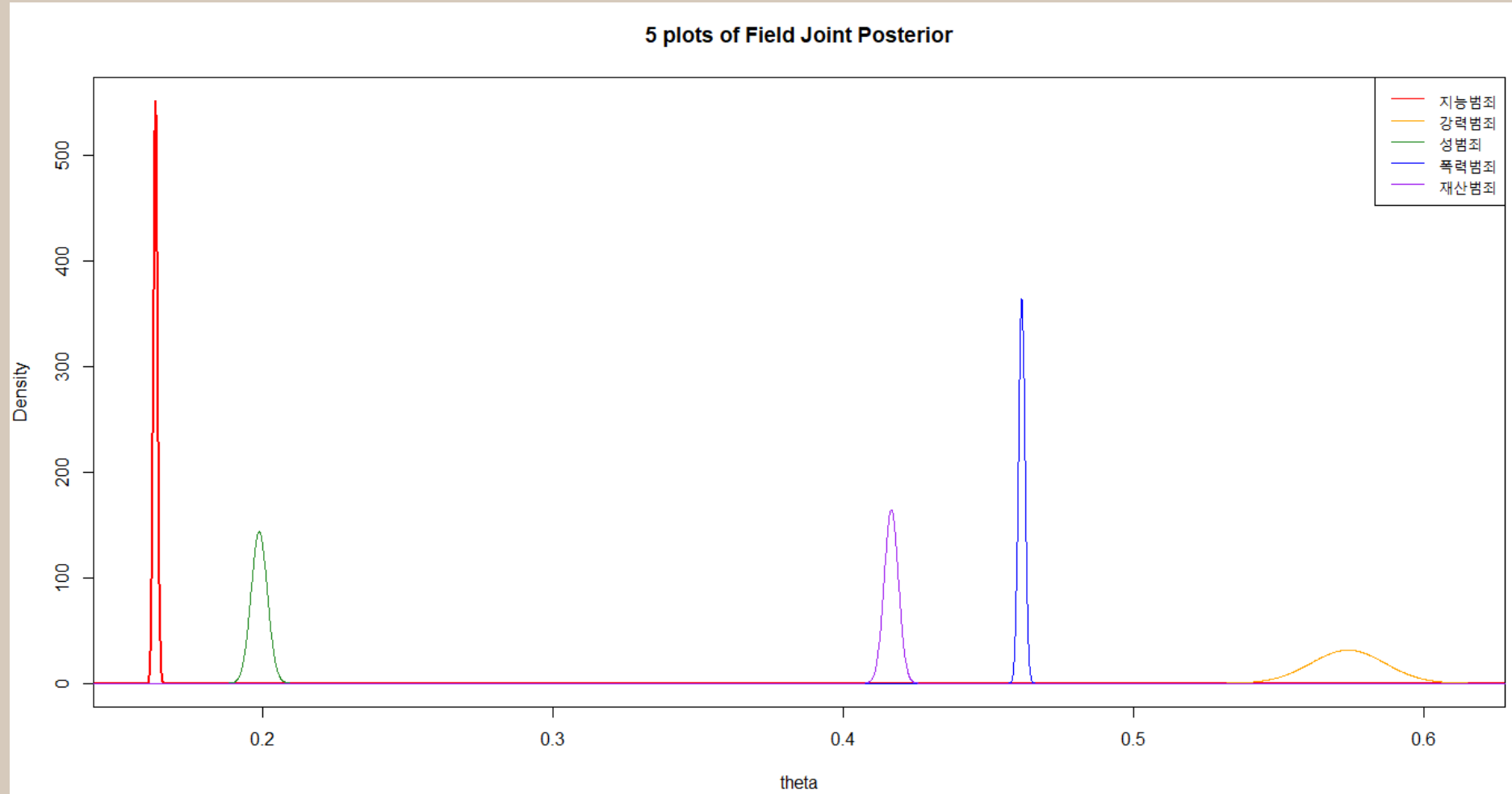
Hierarchical Bayesian Model – Field 5 : 지능범죄



Prior의 영향력이 전무하다 = Data의 영향력이 매우 강하다

지능 범죄의 경우, 특히 사기죄의 단기 검거율이 매우 낮고, 이로 인해 지능 범죄 분야 전체의 Posterior mean도 낮다.

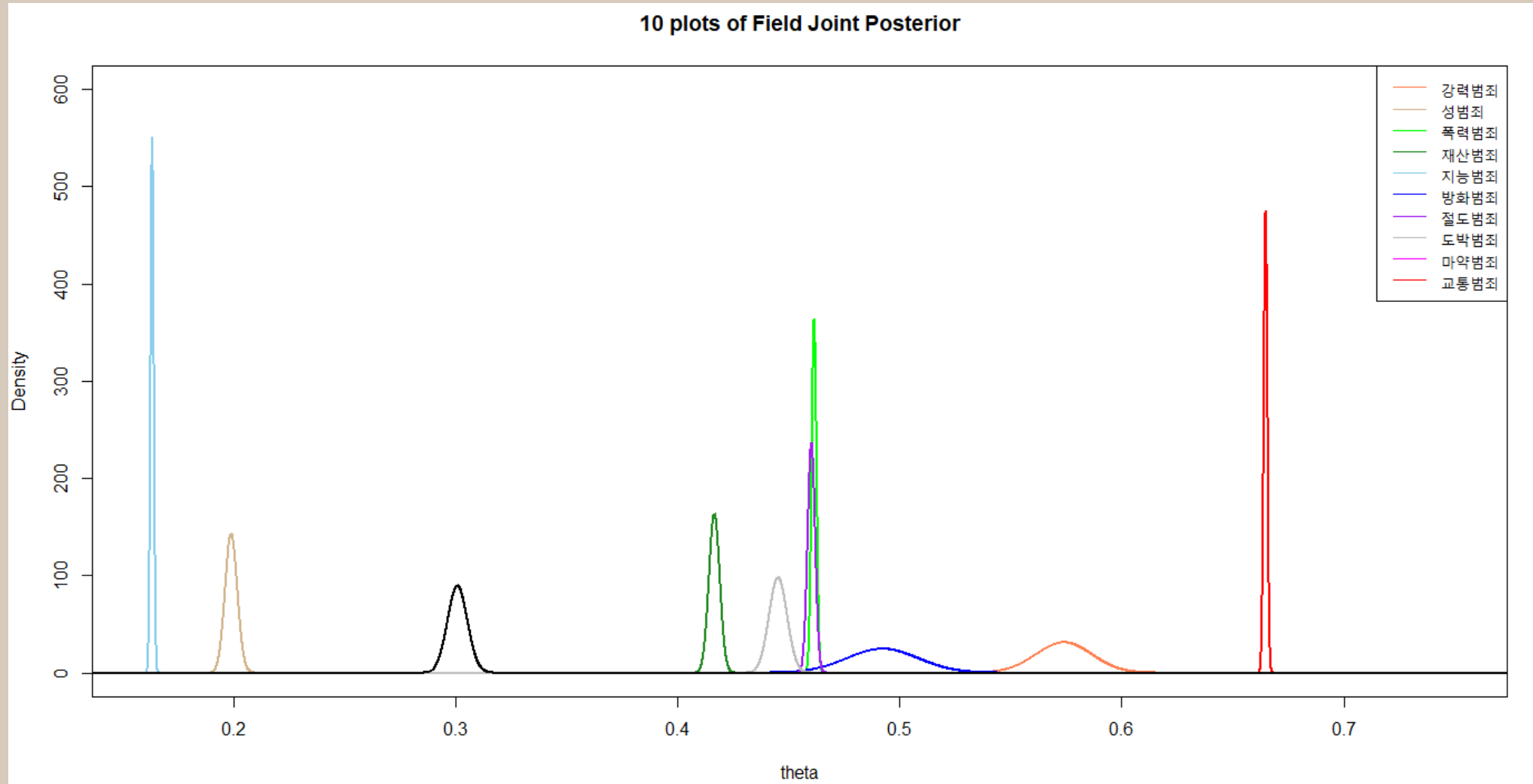
5개 Field들의 Posterior



Field별로 단기 범죄 검거율에 큰 차이를 보인다

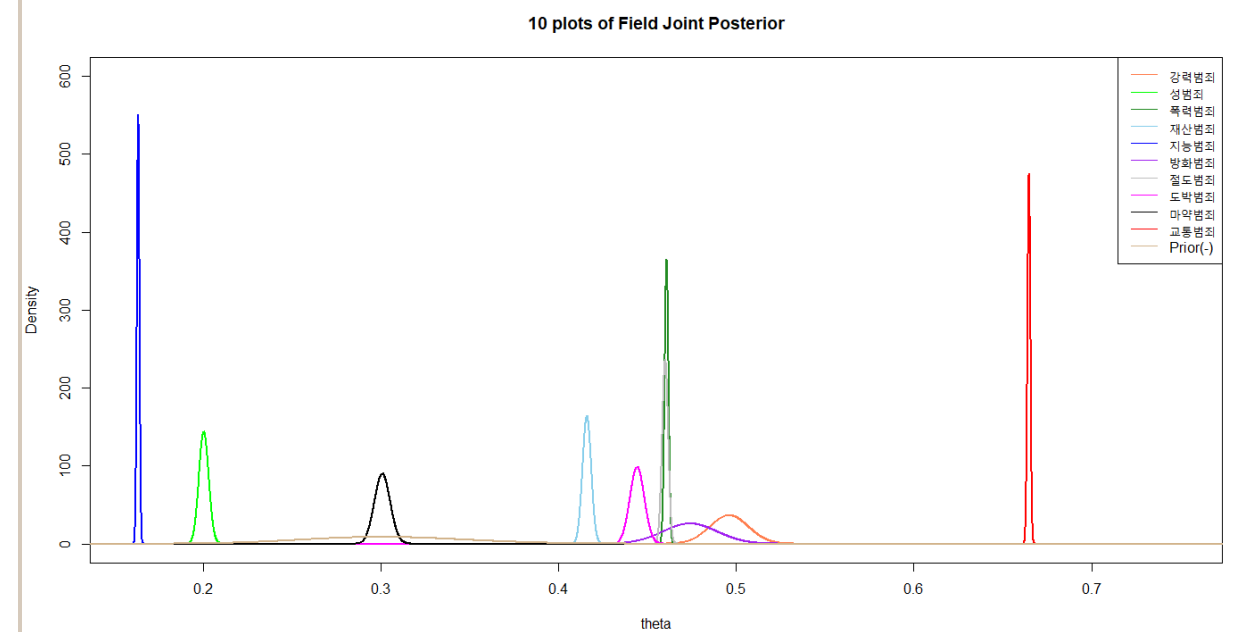
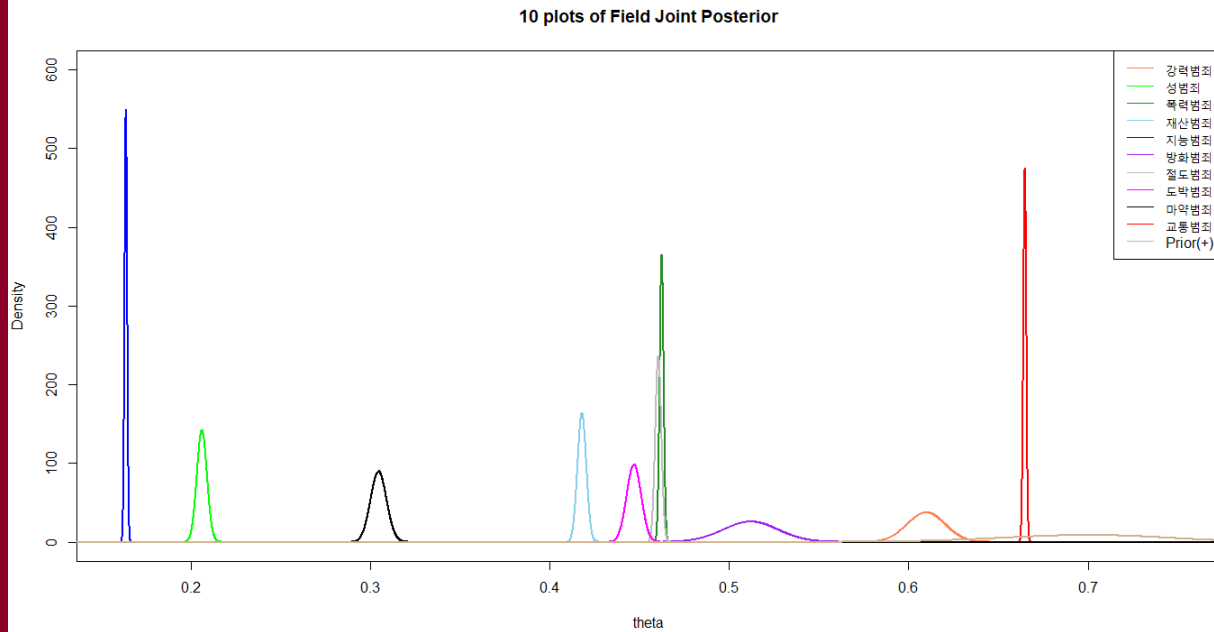
범죄 수가 많을수록 variance는 감소하고, kurtosis는 증가한다

10개 Field들의 Posterior



일부 Field에는 하나의 범지만 소속되어있다

10개 Field들의 Posterior



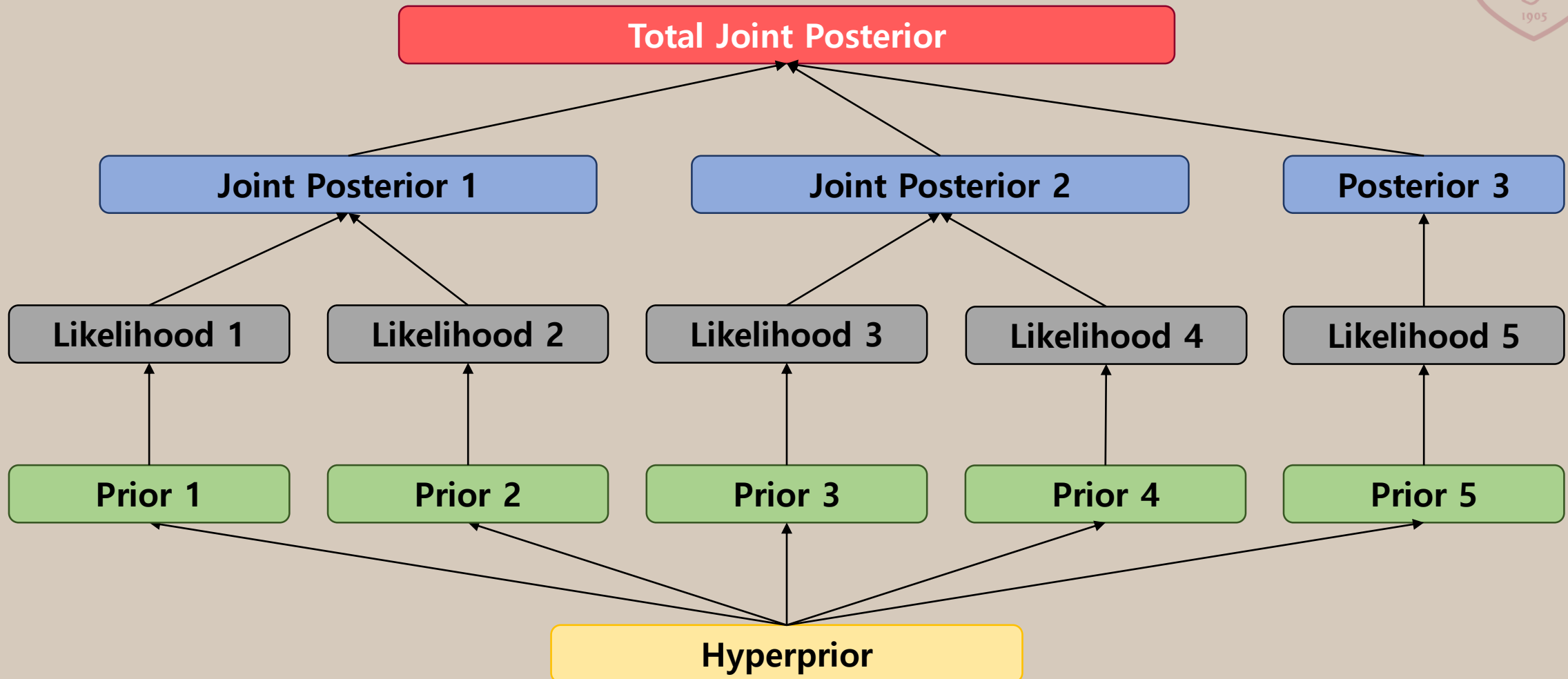
강력 범죄 Field는 N이 상대적으로 작은 편이어서 prior의 영향을 많이 받음

지능 범죄, 교통 범죄 등 N이 매우 큰 Field는 prior의 영향을 거의 받지 않음

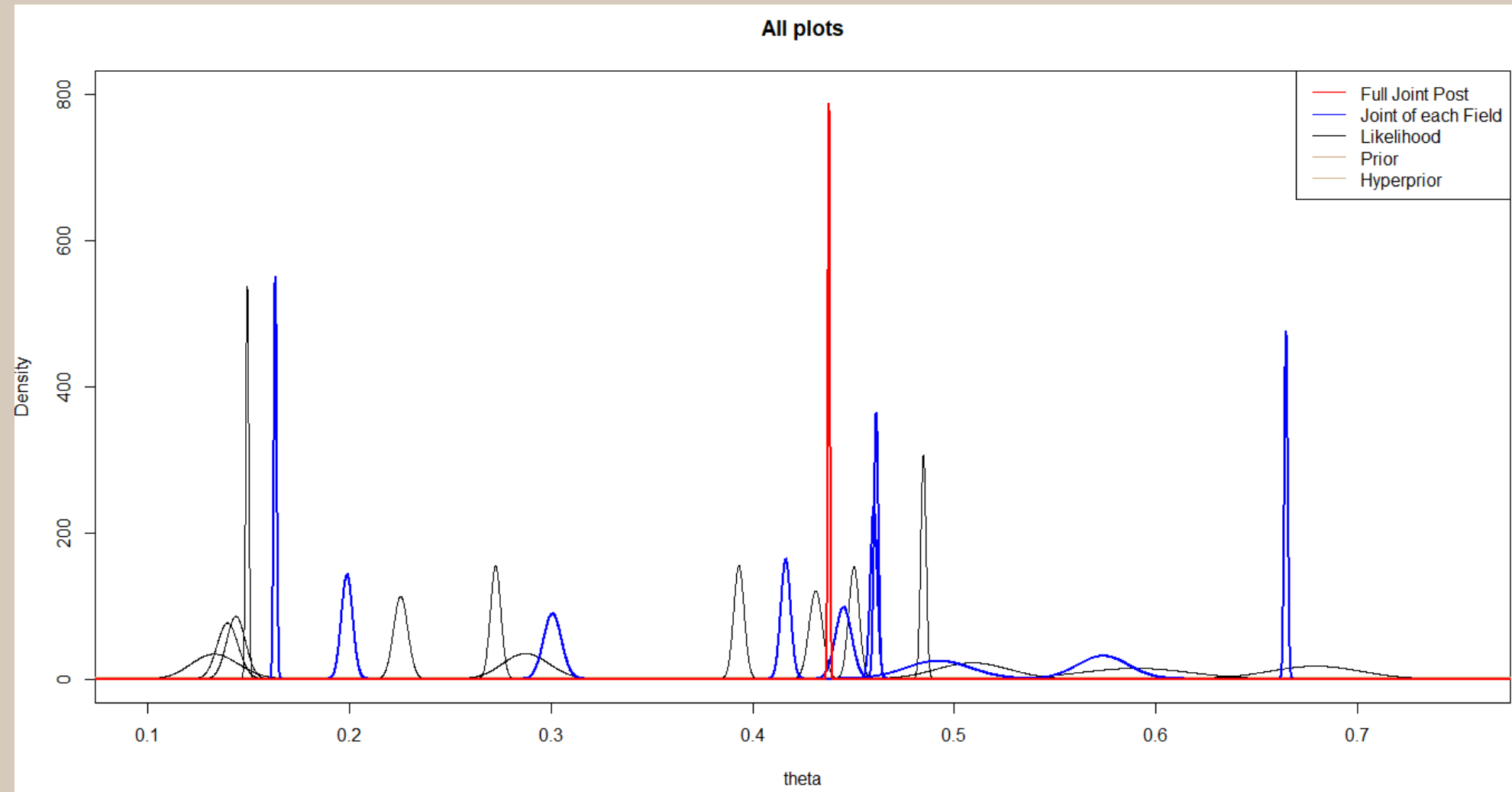
6

4-stage Hierarchical Bayesian Model and Total Joint Posterior

Structure of 4-stage Hierarchical Bayesian Model

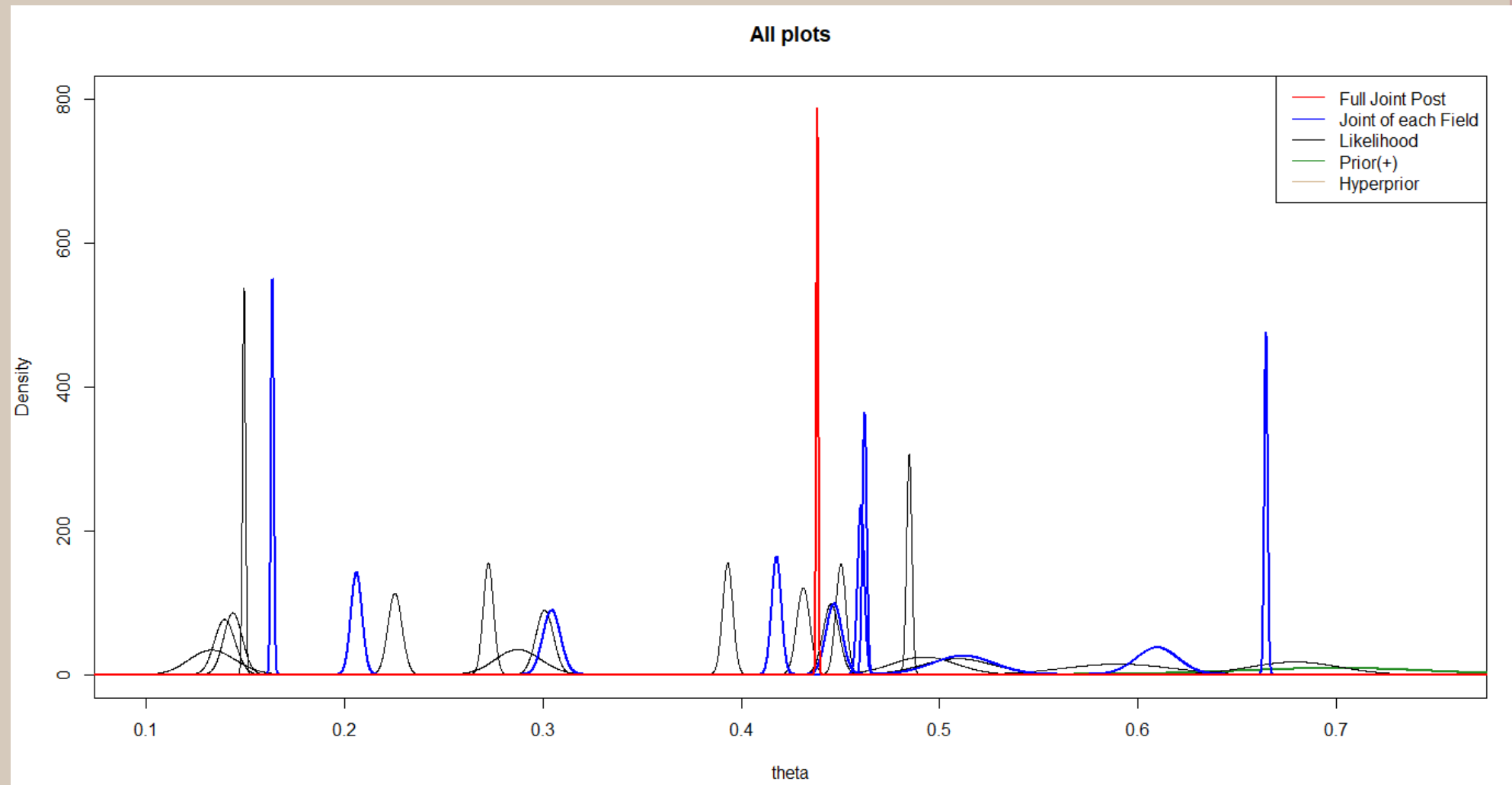


4-stage Hierarchical Bayesian Model (Flat Hyperprior, Uniform Priors)



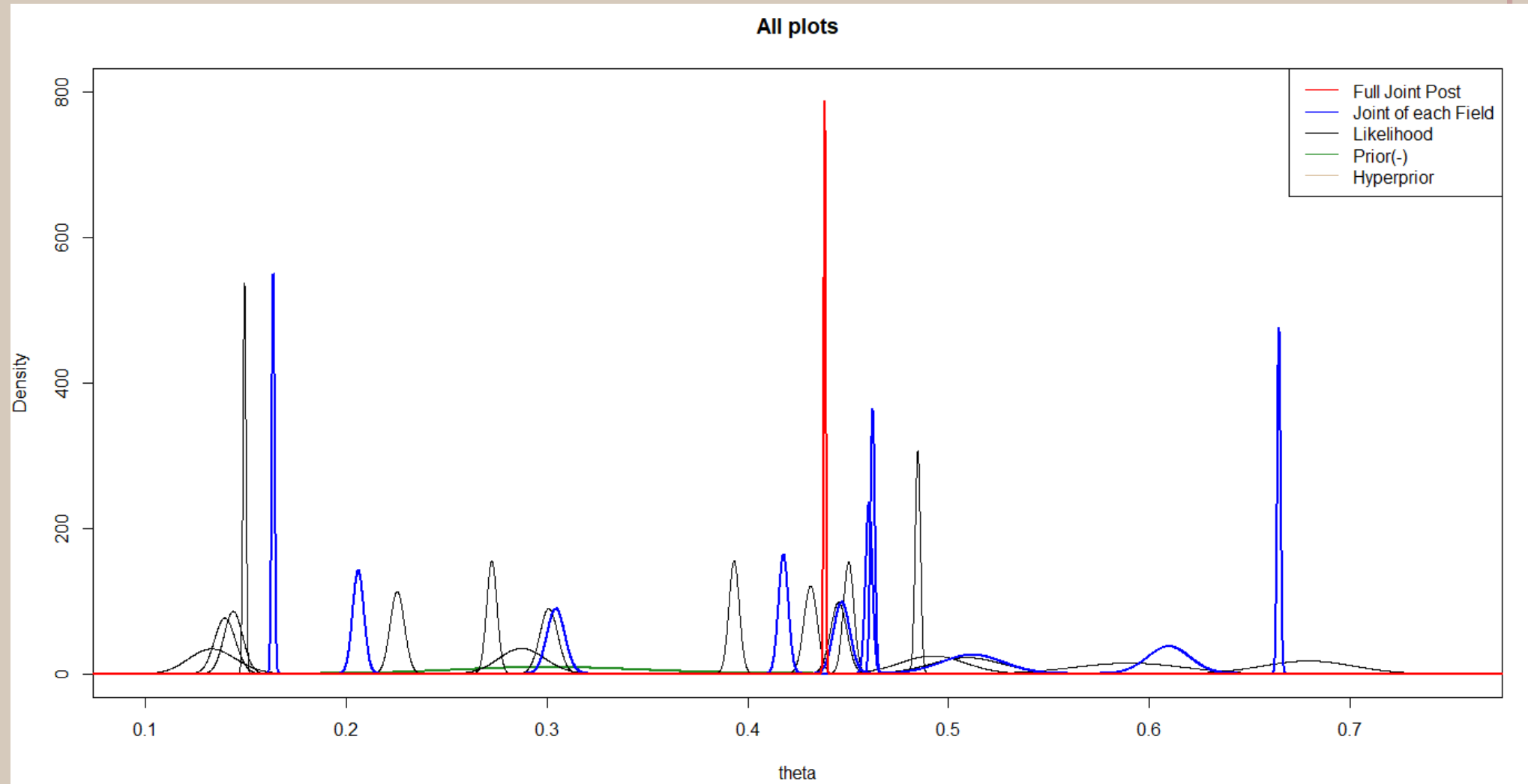
Mean = 0.4380, Variance = 2.5817×10^{-7}

4-stage Hierarchical Bayesian Model (Flat Hyperprior, Beta Priors)



Mean = 0.4385

4-stage Hierarchical Bayesian Model (Flat Hyperprior, Beta Priors)



Mean = 0.4377



한글

결론



- (1) 범죄 Field별로 단기 검거율에 분명한 차이가 존재하고, 이를 시각화를 통해 확인하였음
- (2) 살인죄의 경우, 상당수의 범죄가 1달 이내 검거됨
- (3) 사기죄의 경우, 소액 사기를 포함하여 범죄 발생 건수가 많고, 단기에 검거되는 비율은 극소수임
- (4) 대부분의 범죄 field에서 같은 field내의 개별 범죄들은 단기 검거율에 큰 차이를 보이지 않음

한계점

- (1) Prior의 다양성 부족
- (2) 모든 범주의 종류를 고려하지 않음
- (3) 각 범주의 상대적 경중을 고려하지 않음
- (4) 각 Field별 mean이 다르다는 것을 가설검정을 통해 확인하지 못함