# Bayesian adjusted inference for selected parameters

Daniel Yekutieli

Statistics and OR,
Tel Aviv University

# Multiplicity – simultaneity and Selective Inference

- In his 2001 NSF conference talks at Temple University Yosef Hochberg identified “The problem of Multiple Comparisons” with “Selective and Simultaneous Inference”.

- Benjamini and Yekutieli `05 argue that “selective inference” and “simultaneous inference” are two distinct problems encountered when trying to provide valid marginal statistical inference for multiple parameters.

# Simultaneity

The simultaneity problem is directly caused by multiplicity – it is the need to provide marginal inference that applies to all the parameters:

- Testing multiple null hypotheses – ensure that each of the null hypothesis is not falsely rejected.

- Constructing multiple confidence intervals – ensure that each of the confidence intervals covers the respective parameter.

➔ Solution – Family Wise Error rate adjusted inference:

e.g. Bonferroni CIs ( = a marginal 1- $\alpha/m$ CI for the m parameters).

# Selective Inference

Selective inference refers to the practice of providing statistical inference for parameters selected after viewing the data.

- It can occur when considering a single parameter.

- More common when initially considering multiple parameters:

  consider a 6,000 gene microarray experiment in which the researcher is only interested in the 203 genes with an observed fold-diff > 2

***Topic of this talk:***

how do we construct a marginal CI for an “interesting” gene?

# Outline

- The False Coverage-statement approach for selective inference

- Bayesian adjustment for selective inference:
  - The conditional Bayesian framework
  - Some intuition – two simulated examples
  - Relation to Storey’s Bayesian FDR (pFDR, q-values, SAM … )

- Main question:

  Is it necessary to adjust Bayesian inference for selection ?

# The selective inference framework

Benjamini and Yekutieli `05 consider the following setup:

We have a series of parameters $\theta_1, \theta_2 \cdots\ \theta_m$:

— for $i = 1 \cdots m$, $T_i$ is an estimator of $\theta_i$.

— $CI_i(T_i, \alpha)$ is a $1 - \alpha$ frequentist confidence interval for $\theta_i$

$$Pr_{T_i|\ \theta_i}\{\ \theta_i \in CI_i(T_i,\ \alpha)\ \}\ \geq\ 1 - \alpha.$$

— However we construct the CI for $\theta_i$ only if it is selected,

$$\text{i.e.}\quad i \in S(\vec{T}).$$

➔ selection biases estimation, and demonstrate that CIs for selected parameters can no longer offer any minimal coverage probability

# False Coverage-statement Rate

As solution to selective inference Benjamini and Yekutieli `05 suggest control over the False Coverage-statement Rate:

— Let $R_{CI}$ denote the number of CI constructed, $R_{CI} = |S(\vec{T})|$

— Let $V_{CI}$ denote the number of CI not covering the parameter

$$\mathbf{FCR} \;=\; E\begin{cases} V_{CI}/R_{CI} & \text{if } R_{CI} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

FCR control is a frequentist mechanism for Conditional Coverage Probability:

If $R_{CI}$ is large, then $FCR \leq 0.05$ implies that in each realization approximately 95% of the CIs cover their parameter, thus the probability that $CI_i$ randomly chosen from $S(\vec{T})$ covers $\theta_i$ is $\approx 0.95$.

# Control of the FCR

Proposition – construction of FCR adjusted CIs:

m independent test statistics and any type of selection, if R parameters are selected out of m then constructing marginal 1-(R×q)/m CIs for the selected parameters ensures FCR $\leq$ q.

Return to our example:

6,000 gene microarray, how do we construct a valid marginal 0.95 CI for the 203 selected genes?

$$\text{FCR CI:}\quad \bar{x}_i \pm \hat{\sigma} \cdot \mathbf{t}_{1-203\times 0.05/12000,df}$$

➔ Approx 95% of the 203 CIs cover the parameter

# Control of the FCR – summary

Selection can increase the probability of making a false discovery. In FCR adjusted CIs selection is quantified by R/m, and for adjusted error rate q – construct marginal CIs with error rate q × R/m:

$$\text{FCR adjusted probability} = \frac{\text{nominal probability}}{\text{Prop. of selected parameters}}$$

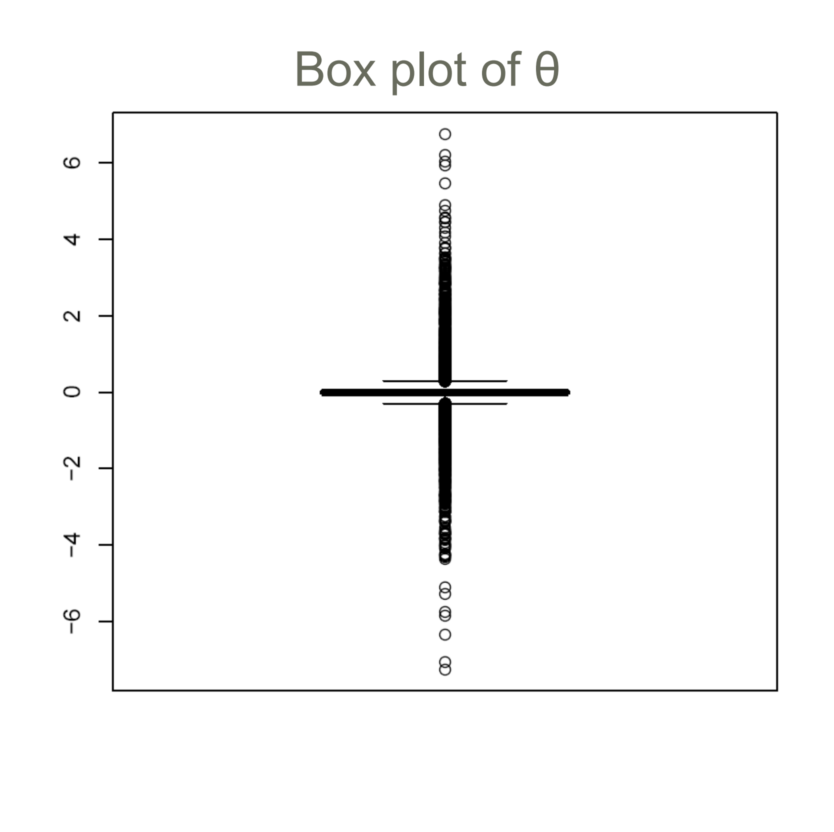
The FDR is a special case of FCR

Setting $CI = \{\theta_i : \theta_i \neq \theta_{0i}\}$ selection becomes rejection of the null hypotheses, $FCR = FDR$, and the BH procedure can be expressed in terms of FCR adjusted CIs

# simulated microarray example 1

1. Sample $50,000$ values of $\theta$:

   $\pm$ ( $24,000 \times exp(10)$ & $1,000 \times exp(1)$ )



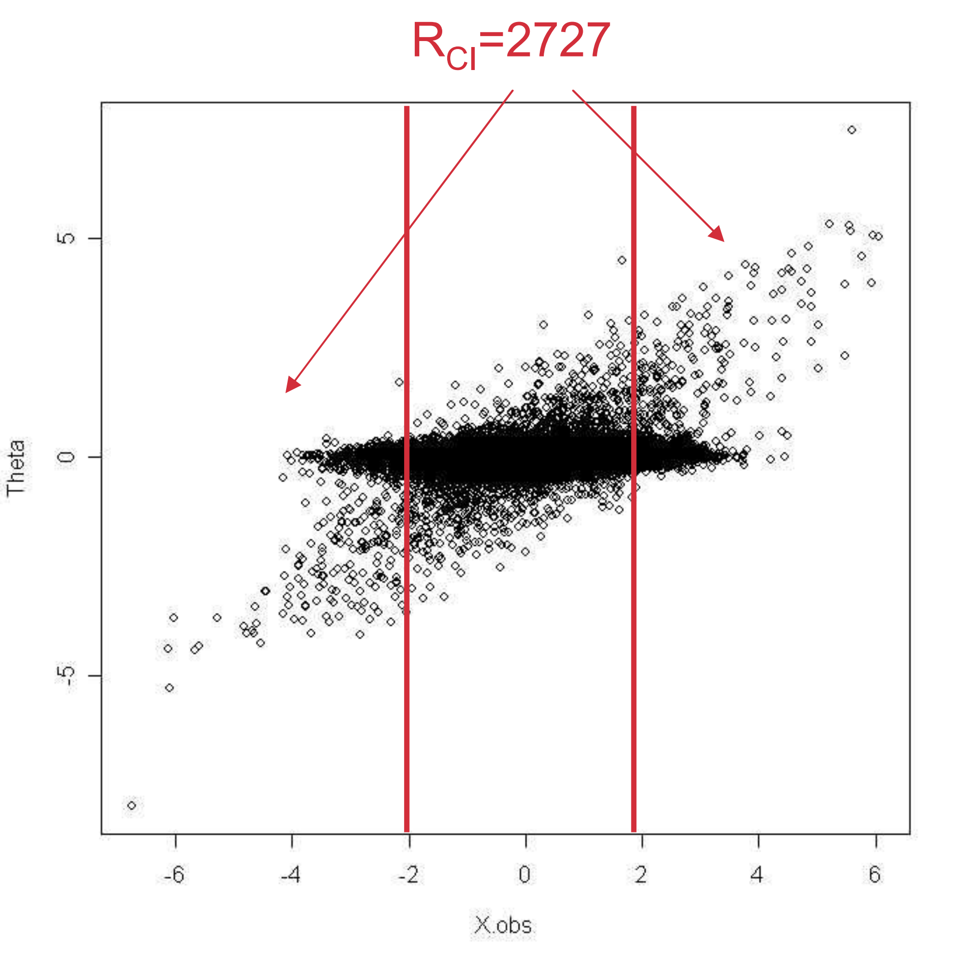
# simulated microarray example 1

1. Sample $50,000$ values of $\theta$:

   $$\pm \; ( \; 24,000 \times exp(10) \; \& \; 1,000 \times exp(1) \; )$$

2. For each observed $\theta$ sample $x$ from $N(\theta, \; 1)$

# simulated microarray experiment



$R_{CI}$=2727

# simulated microarray example 1

1. Sample $50,000$ values of $\theta$:

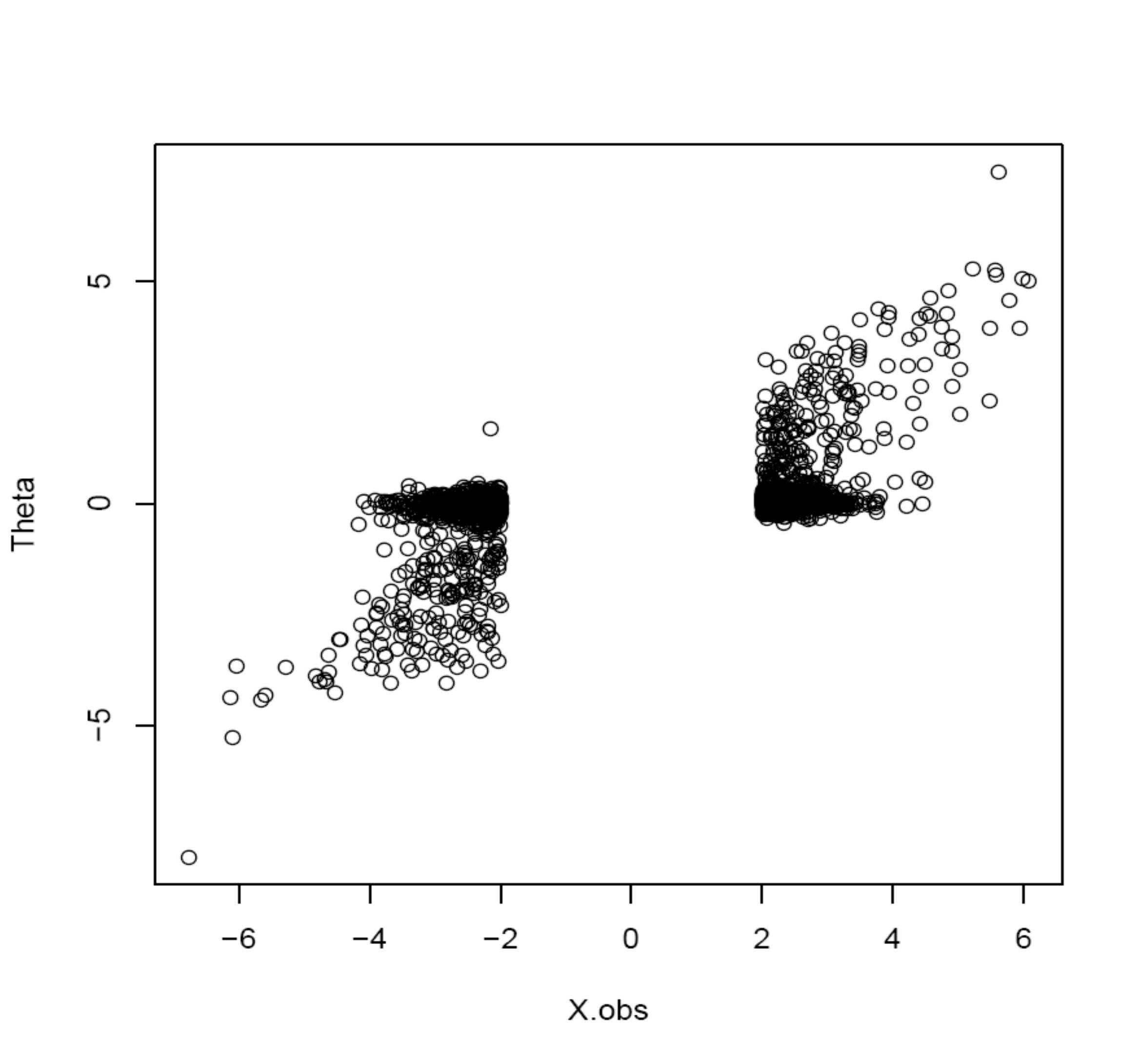
   plus / minus – $24,000 \times exp(10)$ & $1,000 \times exp(1)$

2. For each observed $\theta$ sample $x$ from $N(\theta,\ 1)$

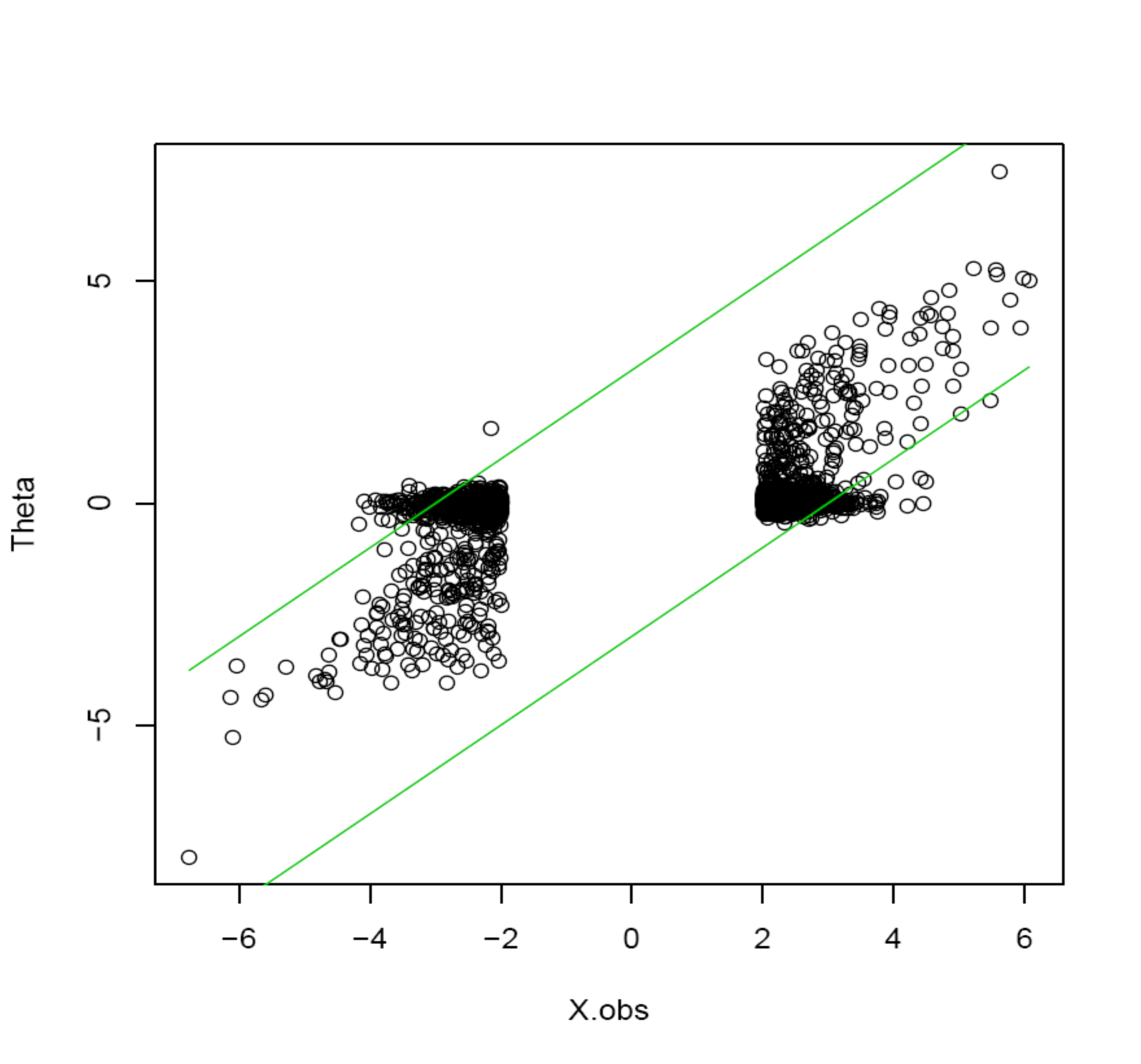
3. For each $|x| \geq 2$ construct:

   — 0.95 FCR CI: $x \pm \mathbf{Z}_{1-R\cdot 0.05/100{,}000}$, $R = \#\{|x| \geq 2\}$

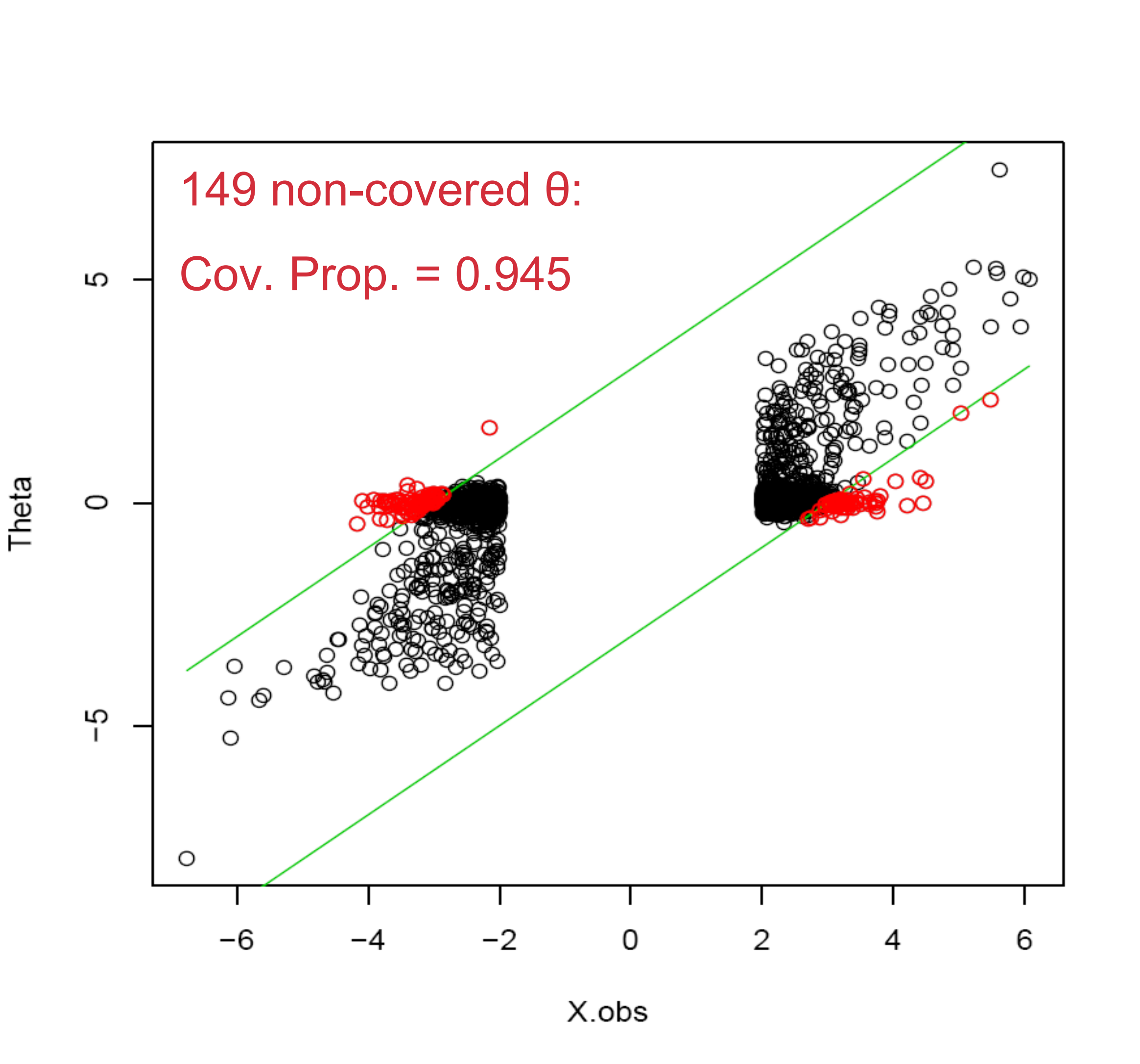
# simulated microarray example 1



# simulated microarray example 1



# simulated microarray example 1



## simulated microarray example 1

In large, non-sparse, problems FCR adjustment corrects for selection:

i.e. Control FCR = q ➔ mean( conditional-coverage ) ≅ 1-q.

However the FCR approach has several limitations …

# Limitation 1: inability to incorporate external information

The validity of FCR adjusted inference is based on the notion of exchangeability between the selected parameters, which only holds if we have no prior knowledge on the parameters.

➔ Prior knowledge breaks down the mechanism for inferring Conditional Coverage Probability for the individual parameters:

Overlooking this point creates an opportunity for “cheating” with the FDR – a *typical* example:

suppose we control the FDR at level $q \leq 0.05$ and get 100 discoveries, and we have prior knowledge that 98 of the discoveries are true discoveries, what is our confidence that the 2 remaining discoveries, for which we have no prior knowledge, are true discoveries?

# Limitation 2: applies same selection correction to all $\theta_i$

Corrects all parameters for selection by dividing the error rate by R/m.

We would like individual selection adjustment for each $\theta_i$ , determined by the prior information on the parameter and its selection probability, but also by the observed $X_i$:

Large $|X_i|$ ➔ $\theta_i$ likely to be selected ➔ Smaller adjustment for selection

# Limitation 3: adjustment unrelated to mode of selection

The FCR adjusted CI – same CI with smaller error rate.

We would also like the mode of selection to determine the selection adjustment:

e.g. if the selection is independent of $X_i$

→ the CI should not be adjusted for selection

# Limitation 4: adjustment for “simple” selection

FCR adjustment applies for providing valid marginal inference for one out of R parameters when initially considering m parameters.

➔ How do you adjust for selection a function of several selected parameters?

➔ How do you adjust for selection the results of hierarchical FDR procedures?

➔ After model selection how do you adjust the CIs for the model coefficients?

# Bayesian selective inference?

Selection irrelevant to Bayesian analysis:

Bayesian analysis conditions on the observed data – the notion that if the experiment could be repeated an “interesting” parameter would no longer be “interesting” seems irrelevant, thus should not, in any way, affect the inference.

Technically: Given a multivariate prior on $\vec{\theta} = \{\theta_1, \cdots \theta_m\}$ and likelihood $f_{X|\,\vec{\theta}}$, then conditioning on $\vec{X} = \vec{x}$ yields a posterior distribution on $\vec{\theta}$, and integrating the posterior over the other $\theta_j$ yields a valid marginal posterior for *any* $\theta_i$.

# Bayesian selective inference!

## Actually, selection may be relevant

When drawing information from the event $\vec{X} = \vec{x}$
the other potential outcomes of $\vec{X}$ are important:
— should we consider all possible potential values of $\vec{X}$?
— or should we only consider selected values of $\vec{X}$?

## ➔ Selection affects the likelihood

Let $S = S(\vec{X})$ denote the event of selection, then for each $\vec{\theta}$, instead of considering the likelihood $f(\vec{x};\ \vec{\theta})$, consider the conditional likelihood:

$$f_S(\vec{x};\ \vec{\theta}) \ = \ f(\vec{x};\ \vec{\theta})/Pr(S;\vec{\theta}).$$

# The Bayesian selective inference framework

## Bayesian adjusted inference for a selected parameter θ$_i$

— $\pi(\vec{\theta})$ is the prior distribution of $\vec{\theta}$, $f(\vec{X};\ \vec{\theta})$ is the likelihood.

— The researcher must specify $S(\vec{X})$: “objectively“ determine the values of $\vec{X}$ for which he would be intersted in $\theta_i$.

— Compute the *conditional posterior*:

$$
\begin{aligned}
p_S(\vec{\theta};\ \vec{x}) &= \frac{f(\vec{x};\ \vec{\theta})/Pr(S;\ \vec{\theta}) \cdot \pi(\vec{\theta})}{\int f_S(\vec{x};\ \vec{\theta})/Pr(S;\vec{\theta}) \cdot \pi(\vec{\theta})\ d\vec{\theta}} \\
&= \frac{f_S(\vec{x};\ \vec{\theta})\pi(\vec{\theta})}{\int f_S(\vec{x};\ \vec{\theta})\pi(\vec{\theta})\ d\vec{\theta}}
\end{aligned}
$$

$\Rightarrow$ For the marginal posterior of $\theta_i$ integrate $p_S(\vec{\theta};\ \vec{x})$ over the other $\theta_j$.

# The Bayesian adjustment for selective inference follows the same principle as the FCR adjustment

In FCR adjusted CIs the error rate, which is the probability that the CI fails to cover the parameter, is divided by the selection probability

$$\text{FCR adjusted probability} = \frac{\text{nominal probability}}{\text{selection probability}}$$

Similarly, for the conditional posterior the likelihood is weighted proportionally to 1/ (the selection probability)

$$p_S(\vec{\theta};\ \vec{x}) \quad = \quad \frac{f(\vec{x};\ \vec{\theta})/Pr(S;\ \vec{\theta}) \cdot \pi(\vec{\theta})}{\int f_S(\vec{x};\ \vec{\theta})/Pr(S;\vec{\theta}) \cdot \pi(\vec{\theta})\ d\vec{\theta}}$$

# simulated microarray example 2

1000 gene microarray experiment: inference for $\theta_1$ given that the rank of its observed expression is $\geq R$.

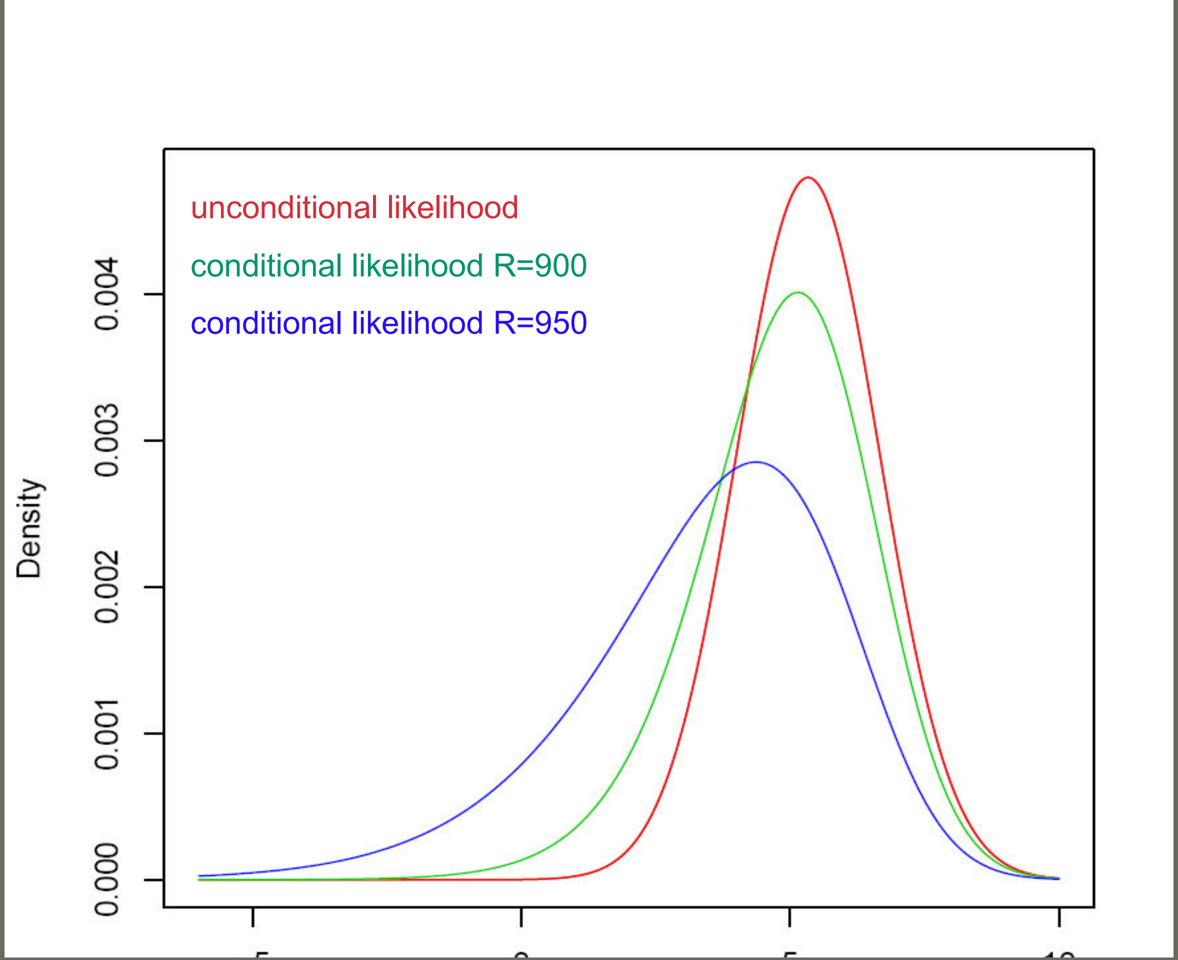
Default simulation parameters:

$X_1=4.5$, $R=900$, Likelihood of all genes $N(\theta_i, 1)$, Prior of all genes $N(0,2)$

Simulation

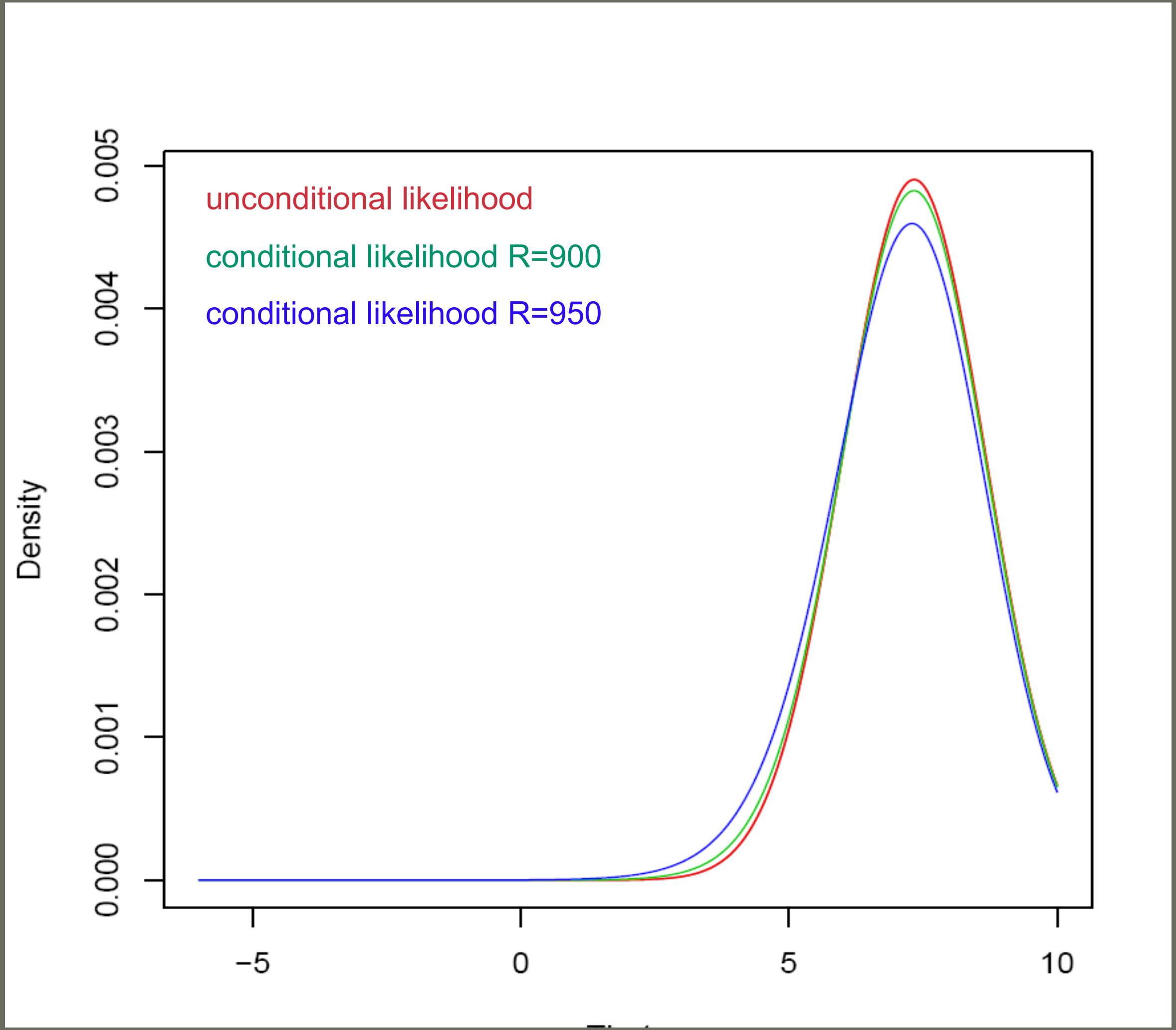
1. Study the effect of R on the conditional likelihood.
2. Study the effect of $x_1$ and R on the conditional likelihood
3. Study the effect of the prior of the other $\theta_i$ on the conditional likelihood.
4. Study the effect of the prior of $\theta_1$ on the conditional posterior.

# effect of R on the conditional likelihood



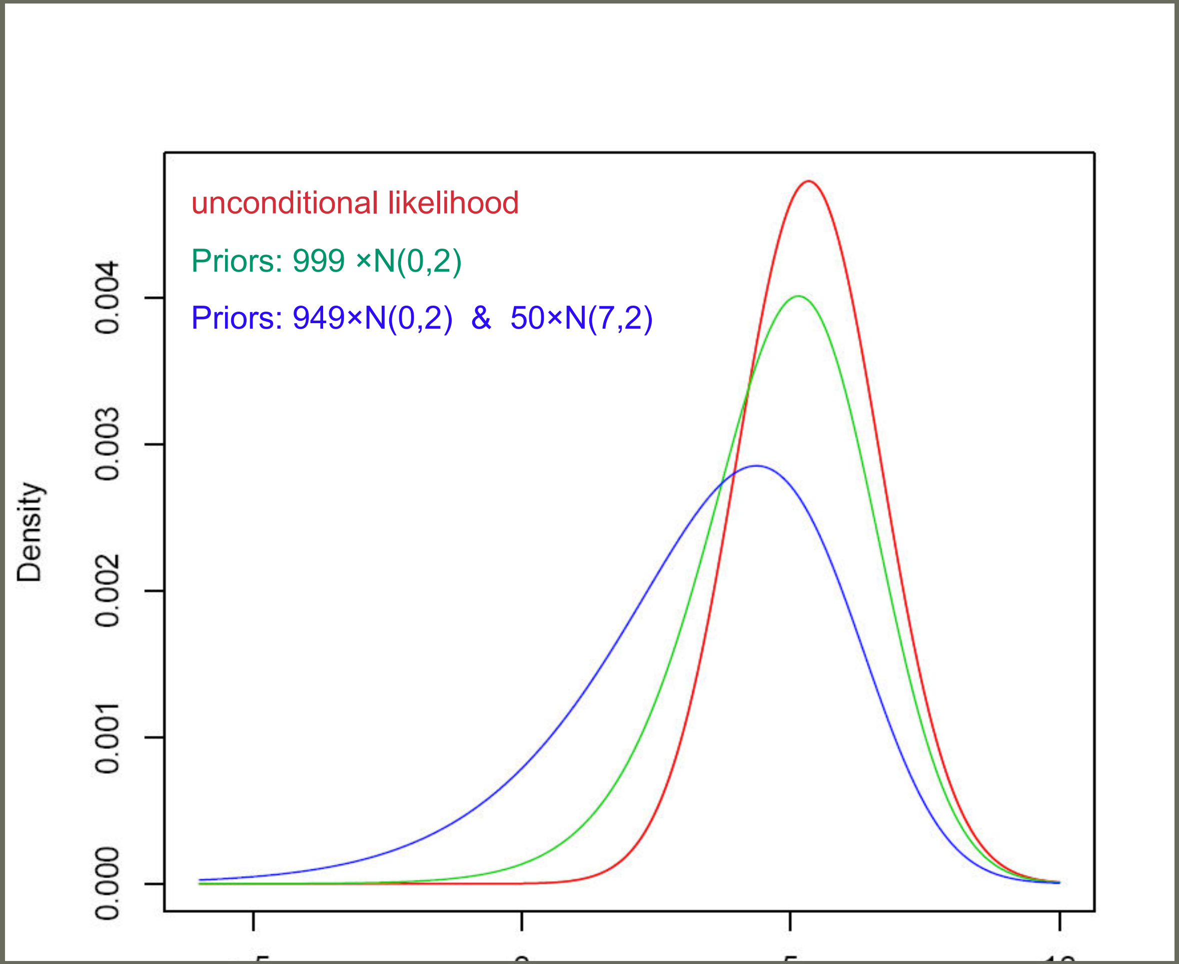
As selection increases ➔ smaller & flatter Conditional likelihood

## effect of $x_1$ and R on the conditional likelihood: $x_1 = 6$ instead of 4.5



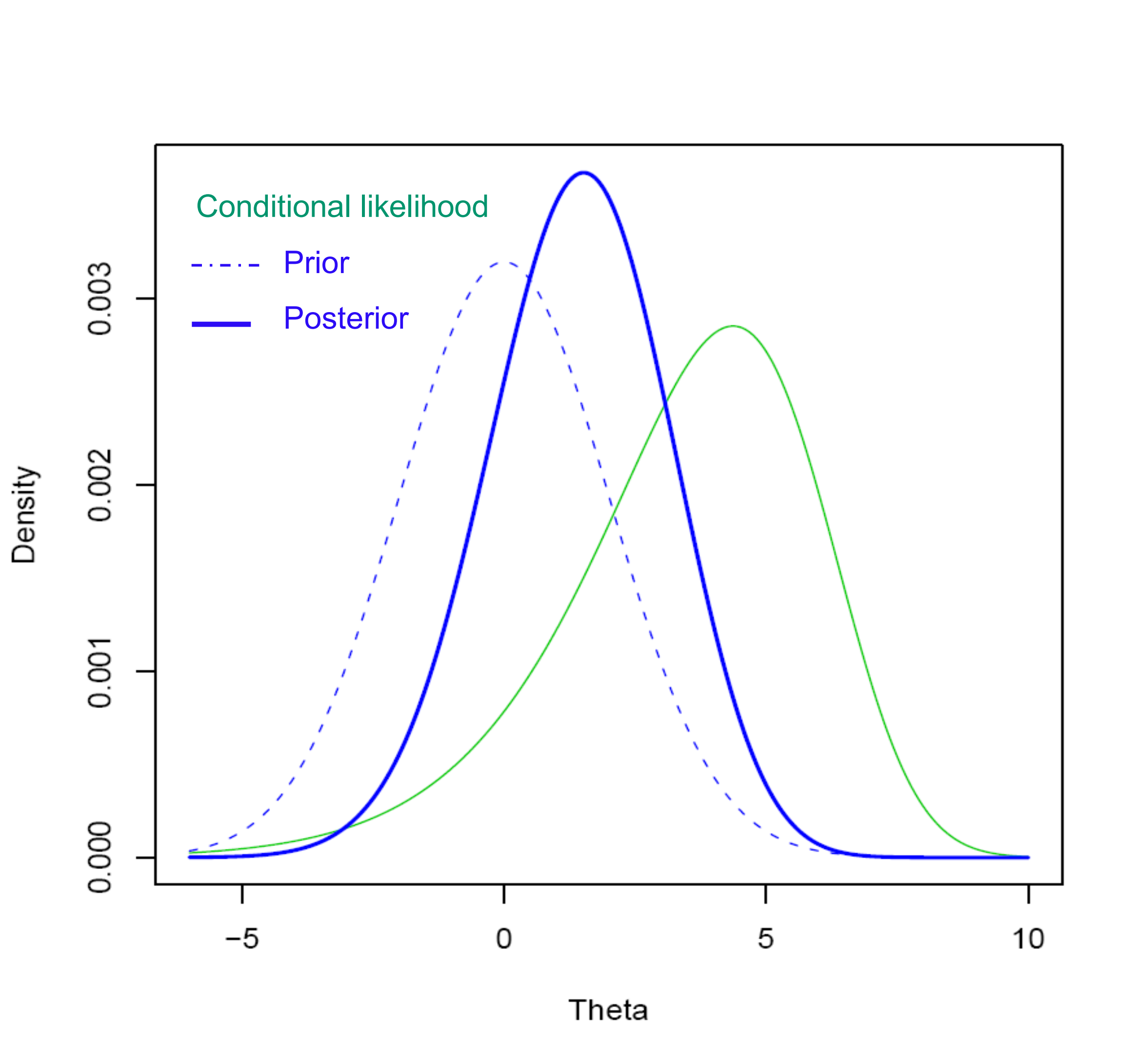
➔ Selection effect almost disappears for large $x_1$

effect of prior of other $\theta_j$ on the conditional likelihood: $x_1 = 4.5$, R=900

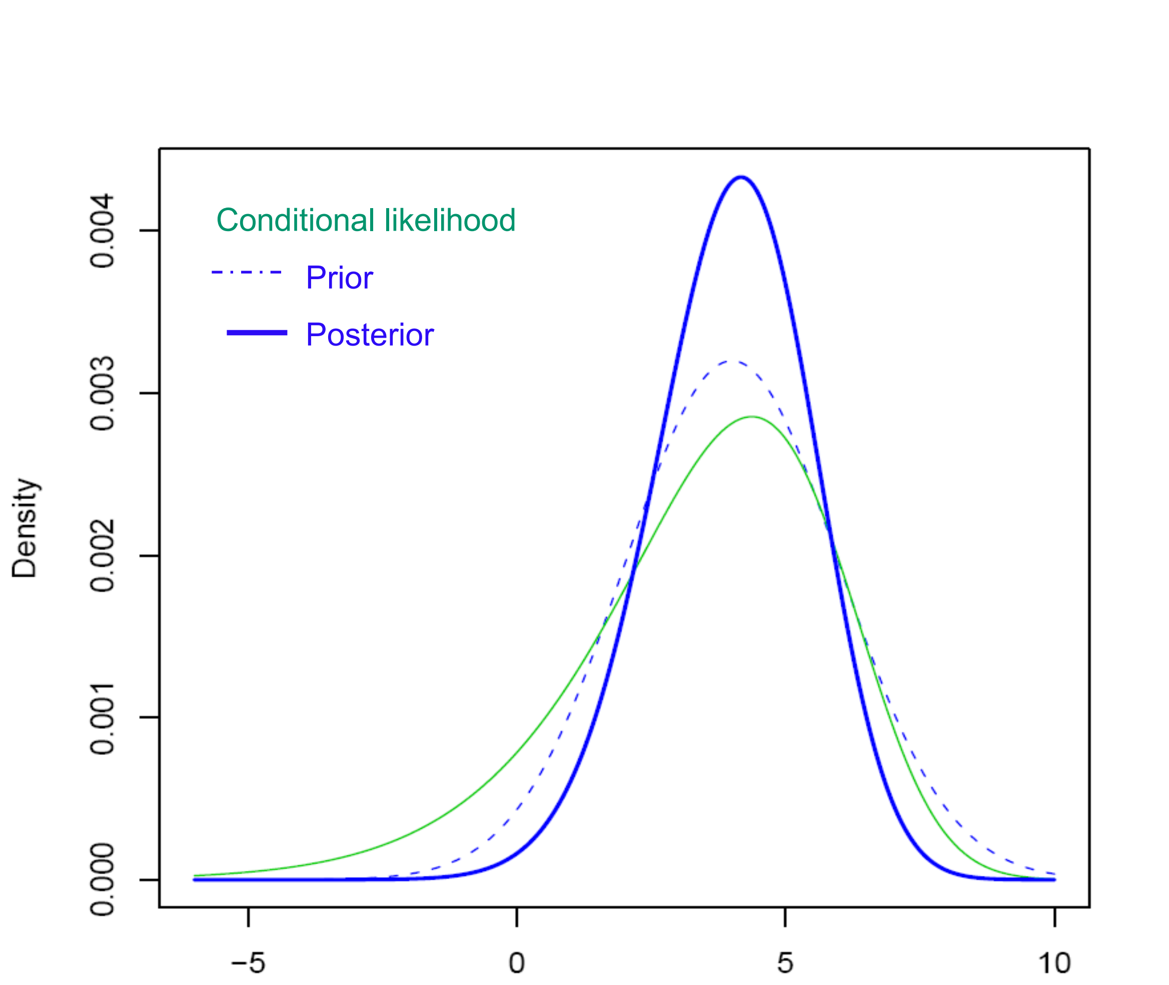


Larger priors for other θ ➔ stronger selection and smaller & flatter Conditional likelihood

# effect of the prior of $\theta_1$ on the conditional posterior: $x_1 = 4.5$, R=950, N(0,2) prior



## effect of the prior of $\theta_1$ on the conditional posterior: $x_1 = 4.5$, R=950, N(4,2) prior



Similarly to non-conditional Bayesian:

Prior fits likelihood ➔ less diffusive posterior

# Bayesian selective inference ⇔ Bayesian FDR

Bayesian FDR ideas were introduced in a series of papers:

Storey `02; Storey `03; Efron, Tibshirani, Storey, and Tusher `01 …

In Storey's Bayesian FDR approach the parameter of interest is pFDR – the conditional probability of making a false discovery given that a discovery has been made for a fixed discovery region Γ

$$pFDR(\Gamma) \; = \; Pr(\; H = 0 |\; T \in \Gamma \;).$$

# Bayesian FDR framework

Storey's Bayesian FDR only applies in the following mixture model:

For $i = 1 \cdots m$:

— $H_i \sim Bernoulli(1 - \pi_0) \quad ( \ H_i = 0$ – a true null hypotheses )
— $T_i | \ H_i = 0 \sim F_0, \quad T_i | \ H_i = 1 \sim F_1$

The pFDR is then be defined

$$pFDR(\Gamma) \ = \ \frac{\pi_0 Pr_{H_i=0}( \ T \in \Gamma \ )}{Pr( \ T \in \Gamma \ )}$$

Estimating pFDR (e.g. SAM)

$$pFDR(\Gamma) \hat{=} \frac{\hat{\pi}_0 \cdot \text{p-value}(\Gamma)}{R(\Gamma)/m}$$

( = adaptive BH adjusted p-value)

# Limitations of the Bayesian FDR

The Bayesian FDR is the Bayesian analogue of the FDR, it provides the same type of inference – and has the same limitations:

the pFDR is computed for the selection rule, it only applies to the specific hypotheses through the exchangeability in the mixture model.

Unlike the FCR,

the only inference it provides is the rejection of the null hypothesis.

# Bayesian selective inference in the mixture model

In Bayesian selective inference we also consider the conditional probability making a false discovery given a discovery rule, however …

— For each $\theta_i$ we can have different $\pi_0$, $F_0$, and $F_1$,
($\pi_0$ is no longer the proportion of true nulls)
— and we can also use the information in $T_i = t_i$, for $t_i \in \Gamma$.

We explicitly compute the conditional posterior probability of making a false discovery for a specific discovered parameter

$$p_\Gamma(H_i = 0; t_i)$$
$$= \frac{\pi_{0i} \cdot Pr_{H_i=0}(\, T_i = t_i | T_i \in \Gamma \,)}{\pi_{0i} \cdot Pr_{H_i=0}(\, T_i = t_i | T_i \in \Gamma \,) + \pi_{1i} \cdot Pr_{H_i=1}(\, T_i = t_i | T_i \in \Gamma \,)}.$$

# Many Bayesians may find Bayesian selective inference unacceptable

- From the Bayesian perspective there is only one way to compute the likelihood – the unconditional probability of observing the data.

- Furthermore, Bayesian selective inference violates an important principle in Bayesian analysis – the stopping rule principle, which implies that the researcher’s intentions should not affect the statistical inference drawn from the data.

# Bayesian selective inference = change of prior

The unconditional likelihood can be expressed

$$f_X(x;\theta) \;=\; f_{X_S}(x;\theta)\cdot Pr(S;\;\theta),$$

$$\text{where } X_S \text{ is } X \text{ given selection.}$$

Let $\pi(\theta)$ denote the prior, then we can define an alternative prior

$$\pi_S(\theta) \;=\; \pi(\theta)\cdot Pr(S;\;\theta)$$

And with this prior the conditional posterior equals the unconditional posterior

$$\begin{aligned} p(\theta|\;x) &\propto f_X(x;\;\theta)\cdot\pi(\theta) \\ &= f_{X_S}(x;\;\theta)\cdot Pr(S;\;\theta)\cdot\pi(\theta) \\ &= f_{X_S}(\theta;\;x)\cdot\pi_S(\theta) \\ &\propto P_S(\theta|x) \end{aligned}$$

# Bayesian Selective inference:
# a Bayesian calibration problem

“Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician”,  Donald Rubin, *Annals of Stat ’84:*

“… The applied statistician should be Bayesian in principle and calibrated to the real world – appropriate frequency calculations help to define such a tie“

“ … The applied statistician should avoid models that are contradicted by observed data in relevant ways – frequency calculations for hypothetical replications can monitor a model’s adequacy and help to suggest more appropriate models”

➔ Fortunately, in microarrays we have numerous replications …

## Return to simulated microarray example 1

1. Sample $50,000$ values of $\theta$:

   Both Positive and Negative — $24,000$ $exp(10)$ and $1,000$ $exp(1)$

2. For each observed $\theta$ sample $x$ from $N(\theta,\ 1)$

3. For each $|x| \geq s$ construct:

   — 0.95 FCR CI: $x \pm qnorm(1 - R \cdot 0.05/10,000)$, $R = \{|x| \geq s\}$

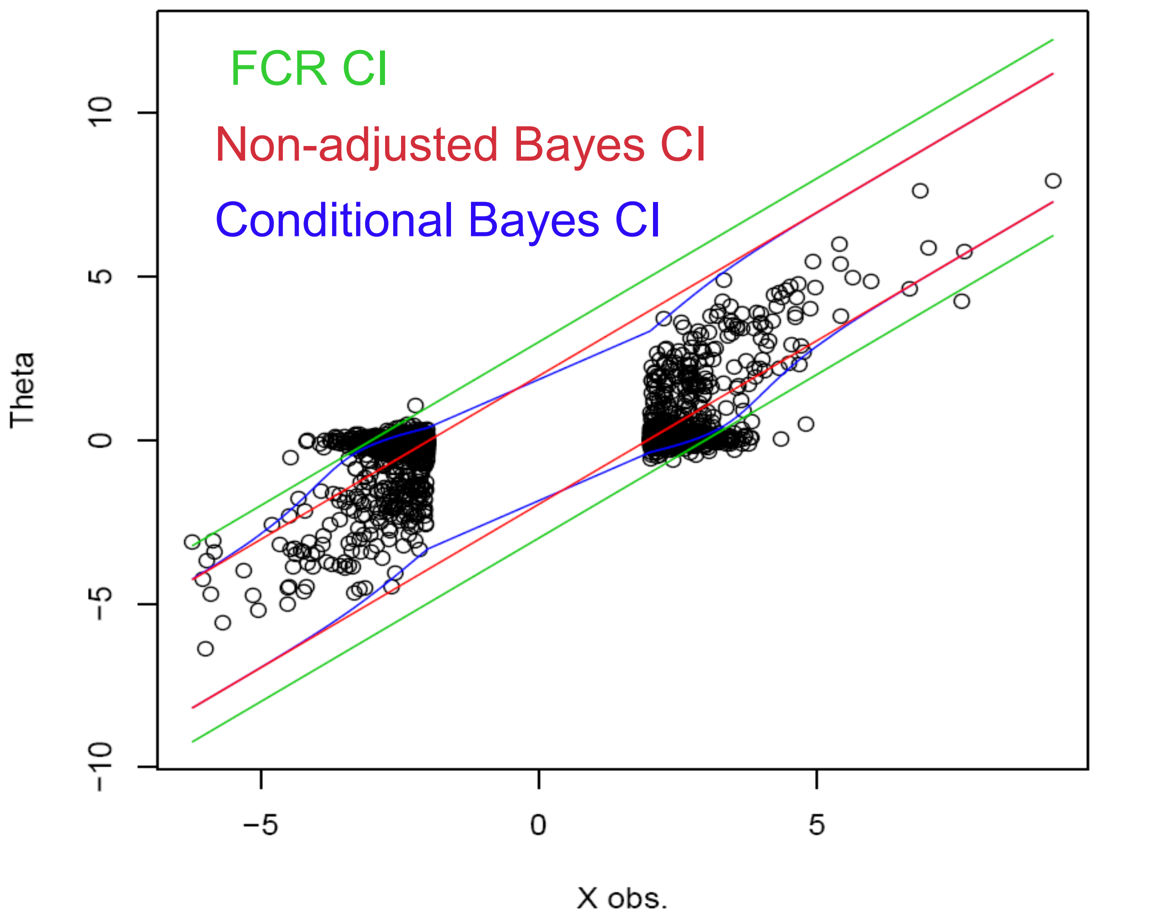
   — non-adjusted Bayesian 0.95 CI

   — adjusted Conditional Bayesian 0.95 CI

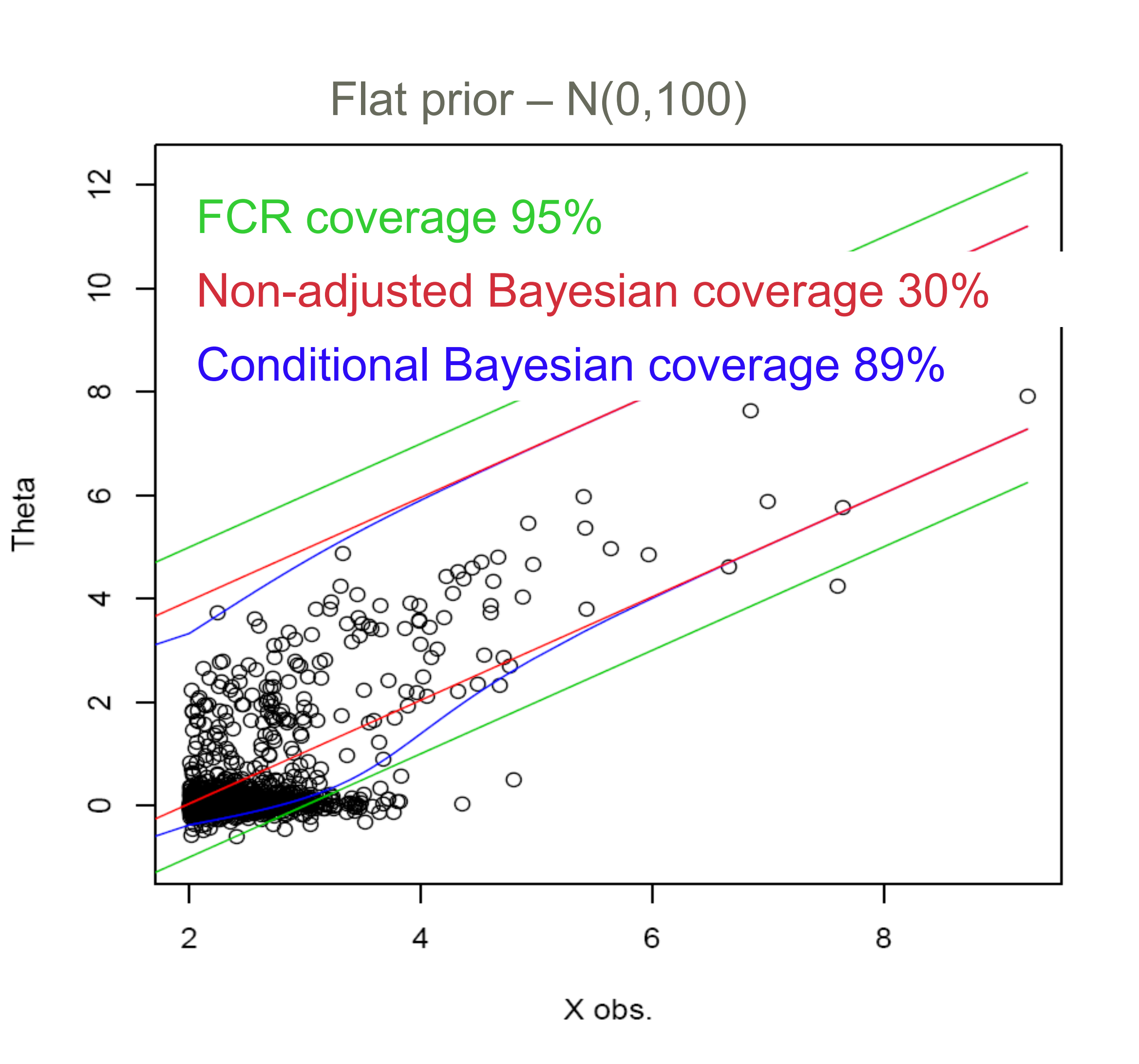
Use the same prior for all parameters and examine if the CI fit the data

# Simulated microarray example 1 – N(0, 100) prior

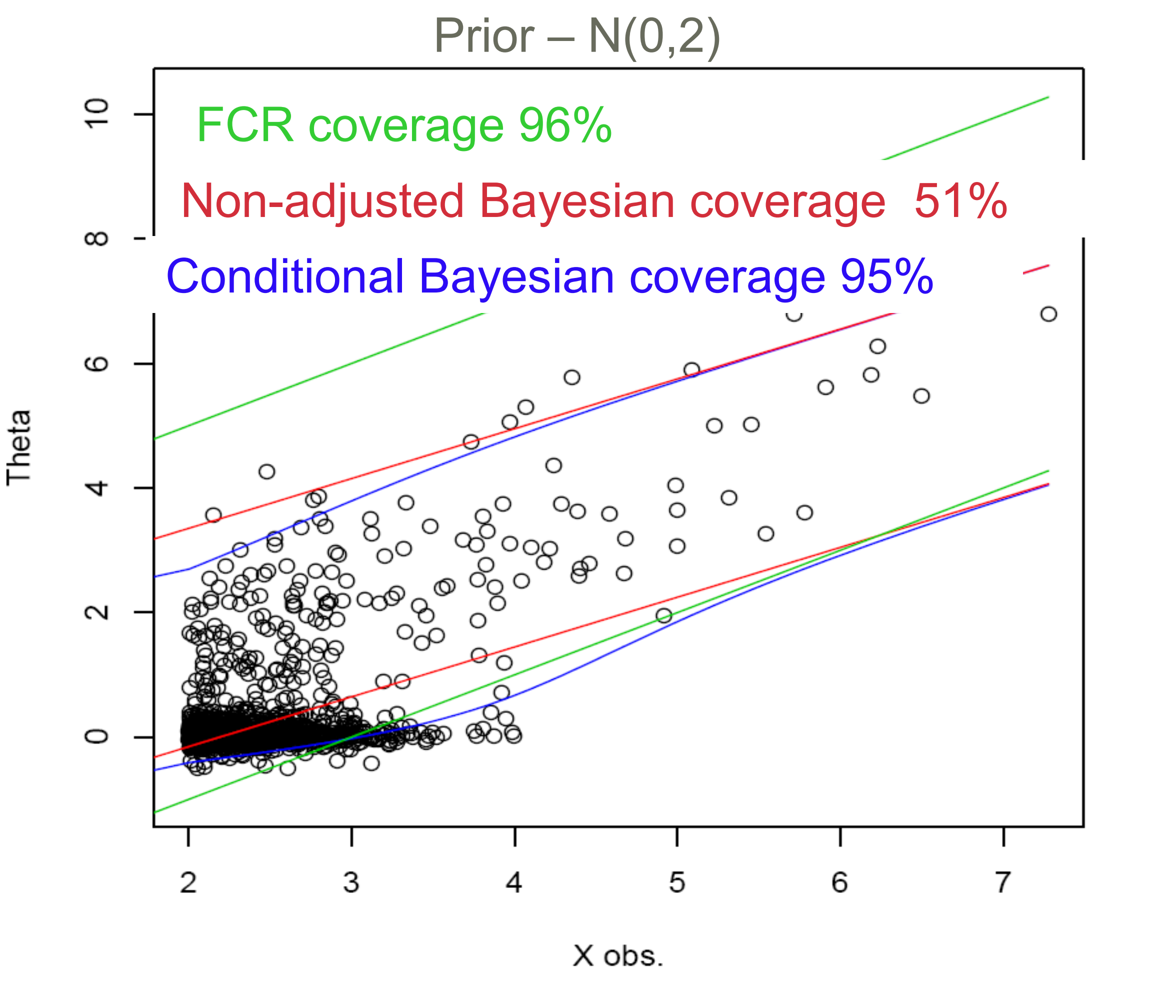
R=2,734 selected genes



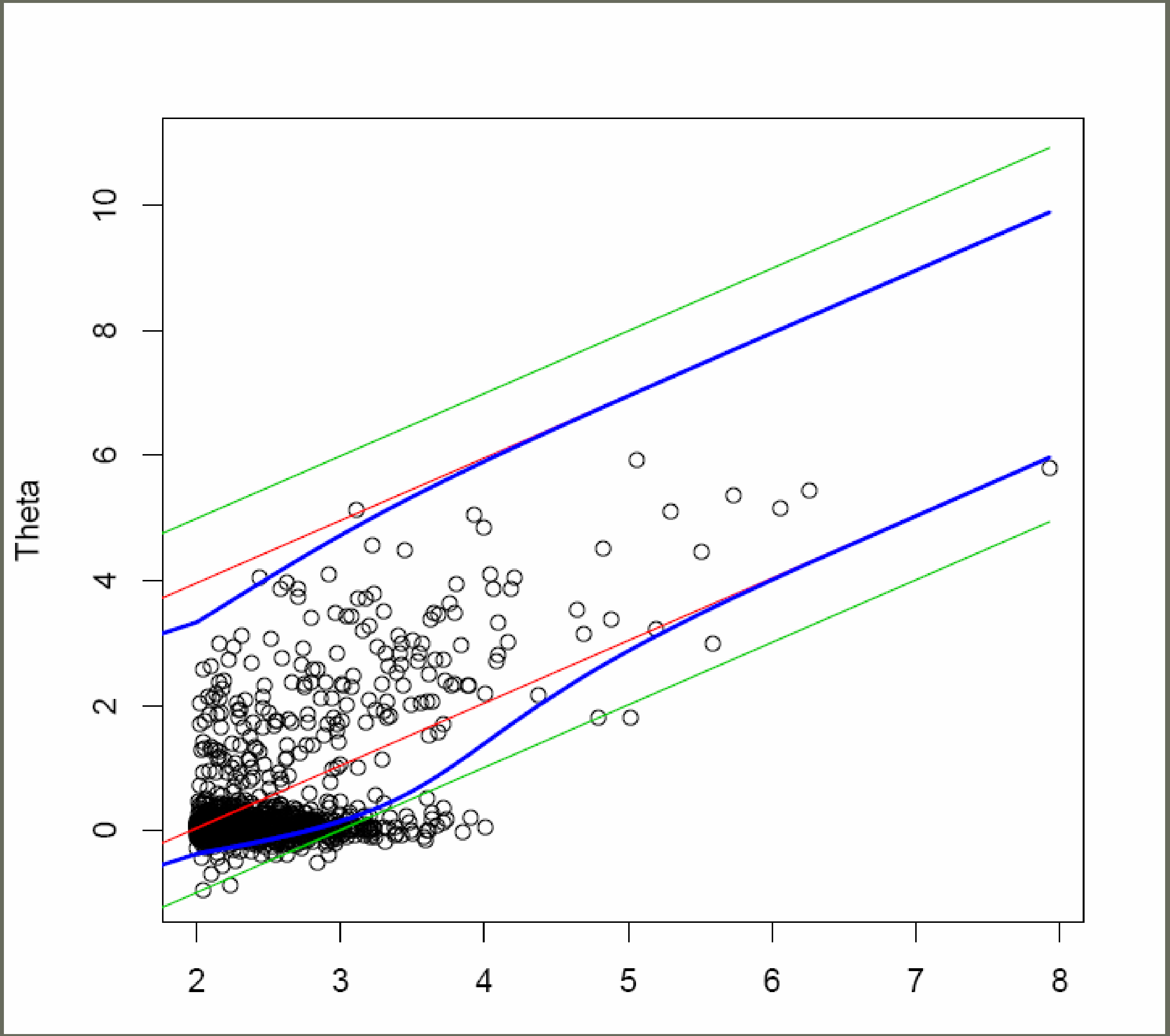
# Simulated microarray example 1 – detail



# Simulated microarray example 1 – N(0, 2) prior



# Return to study non-informative prior CIs



## Summary:

- Frequentists: the conditional Bayesian approach provides better selective inference than the FCR approach

- Bayesians: Bayesian inference may need to be adjusted for selection

Acknowledgement:

Felix Abramovich
Yisrael Parmet
Yoav Benjamini

Thank you !