Statistical Mediation in Early Discovery by Bayesian Analysis and Visualization

NCS 2016

Max Kuhn and Greg Steeno

Pfizer R&D

Medicinal Sciences

Pfizer's Medicinal Sciences statistical support group ($n = 3$) in Groton CT supports two main client groups:

- Chemistry (e.g. medicinal/design, synthetic, computational, ...)
- Pharmacokinetics, Dynamics, and Metabolism (PDM)

PDM is responsible for HTS as well as *in vitro* assays for primary pharmacology, early safety screens, ADME, protein binding, etc.

Chemistry (and some biology) are the *consumers* of the assay data while PDM are often the *producers*

There is about a 200:1 scientists to statistician ratio.

Statistical Mediation

As one might expect, there can be conflicts between these two groups as they work towards the goals of the business. In caricature:

> **Consumers**: the data are too variable and aren't acceptable for our needs.
>
> **Producers**: the assays are pretty good 99% of the time. You don't use them in the right way.

There is usually some recent data that is the focal point of the conflict and the statisticians are called for arbitration

Blood–Brain Penetration via Efflux Ratios

P-glycoprotein (P-gp) is a protein that can pump (i.e. efflux) drugs out of cells and is very relevant for measuring transport across the blood–brain barrier.

To measure it, an artificial membrane is created using a P-gp rich cell line.

We can measure the rate that drug *actively* crosses from side A to side B. Going from B to A reflects passive permeability.

Typically, the reported values compute averages of (unpaired) technical replicates for $A \rightarrow B$ and $A \leftarrow B$ permeability, then take their ratio.

A efflux ratio value of 2 implies that the rate of active transport into the brain is 2–fold higher than passive transport.

How Efflux Ratios are Utilized

The original goal of the assay is to determine whether the assay was significantly passing through the BBB. Historical data and known compounds have demonstrated that values above 1.5-2.5 are truly crossing the barrier.

Eventually, the assay is being used to quantify the magnitude of transport across the BBB to "drive structure–activity relationships (SAR)" as the molecule is being developed.

(In the case of in vitro clearance assays, the *in vitro assay* results have the additional usage of being propagated through *in vivo* clearance equations for clinical estimates)

What does "fit for purpose" mean ay any given time?

How Statistics Helps

Our group handles data for both sets of clients and tend to be very empirical and avoid generalizations ("that assay always produces awful/fantastic results")

There is usually no *a priori* definition of what defines a successful assay results.

Given this lack of specificity, we can do a lot to characterize what the assay can do and set expectations of what a "normal" result looks like and where the assay has excessive bias or variance issues. We try to *quantify performance and reproducibility*.

Luckily, there tends to be a lot of these data available to us. One of our main labs produces over 80% of the data published to our compound database (over 2.5B results per year).

Historical Efflux Ratio Data

A total of 39325 unique compounds were available under the same protocol. There were a total of 46505 assay values.

On average, the number of replicates was 1.2, with a minimum value of 1 replicates and a maximum of 60 assay values. The compounds were assayed between 2013-09-03 and 2016-02-25 (this version of the assay has been deprecated and is not run on new compounds).

The median time between replicates within a compound was 18 days and the largest number of days between replicates was 169 days.

The median efflux ratio was 2.6 with the majority (80%) of the values being between 1.2 and 19.8.

Some Notation

Let X represent the *reported assay result* which may be a single value or the mean of several replicates.

Also, let Θ be the underlying quantity of interest (e.g. efflux ratio).

We would like to make inferences regarding Θ based on the reported data. For example:

> *"The geometric mean of my five replicates was 5.07. What is the noise around that number? Is it getting into the brain more than our lead compound?"*

We are going to need $Pr[\Theta|X = 5.07]$ to answer this...

Bayes' Rule Review

$$Pr[\Theta|X] = \frac{Pr[\Theta] \times Pr[X|\Theta]}{Pr[X]}$$
$$= \frac{Prior \times Likelihood}{Evidence}$$

The Prior represents the distribution of the true assay parameter (for no specific compound). We dictate this or estimate it when we have a lot of data.

The Likelihood measures how probable the measured assay result is when the true value is fixed. We estimate it using data.

The Evidence is computed from the Likelihood and Prior

A different analysis is conducted for every value of X.

Priors

First, computationally, we treat Θ as a discrete parameter in a reasonably fine grid of values.

For some assays, we define a theoretical prior based on our scientific expectations. *Rare* issues for some assays:

- mixtures of normals (for percent inhibition "hit" assays)
- uniform (when we have no idea or brand new assays)
- very long tails past censoring range

For our long running ADME assay screens, we can use *nonparametrically derived priors*

Derivation of the Priors

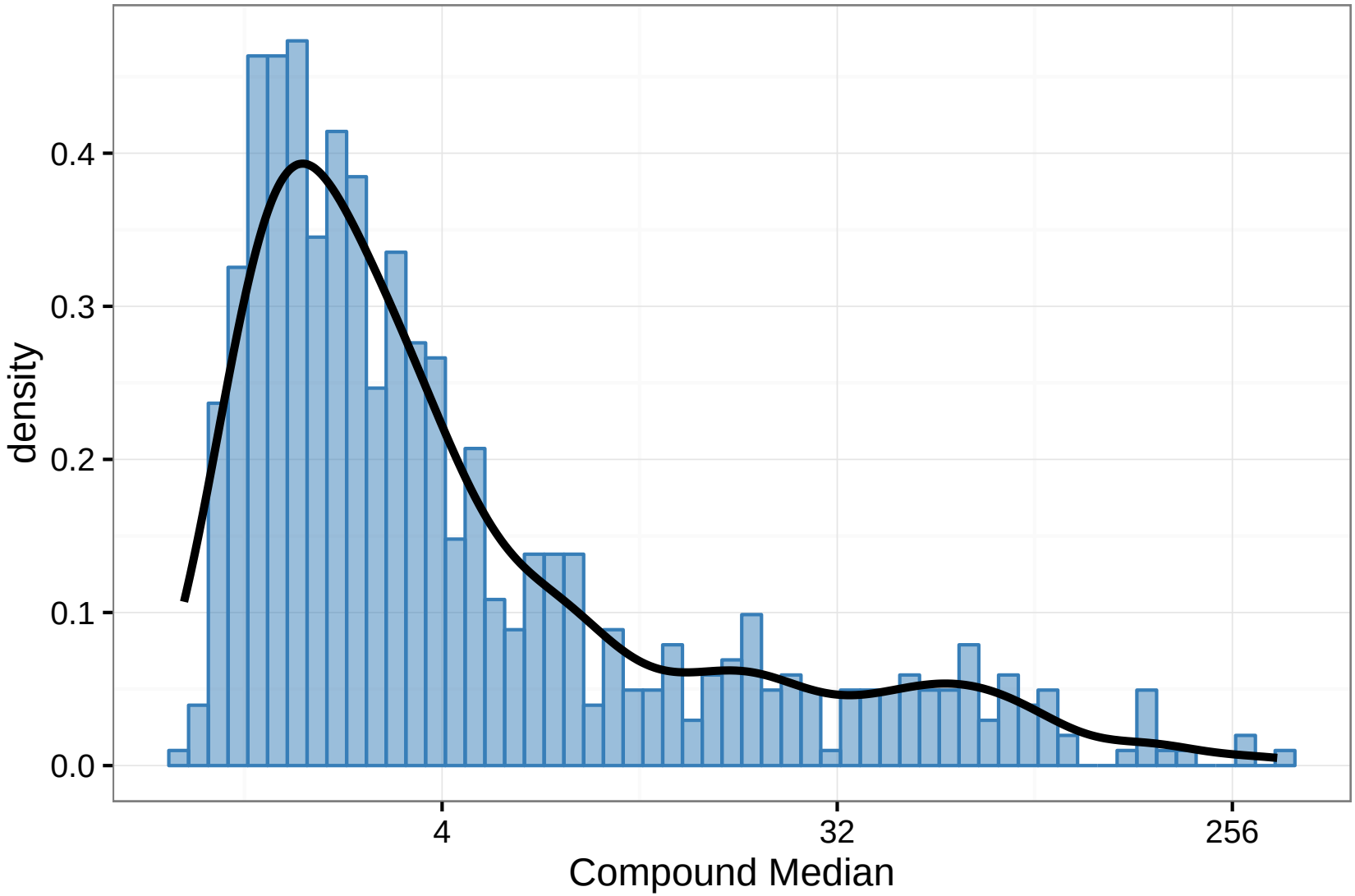
We used medians of highly replicated compounds to get a sense of the overall distribution of compounds that are exposed to the screens.

From these, we use smoothed histograms to get a sense of $Pr[\Theta]$. In some cases, we still have to use significant prior belief to specify the distributions outside of the censoring range.

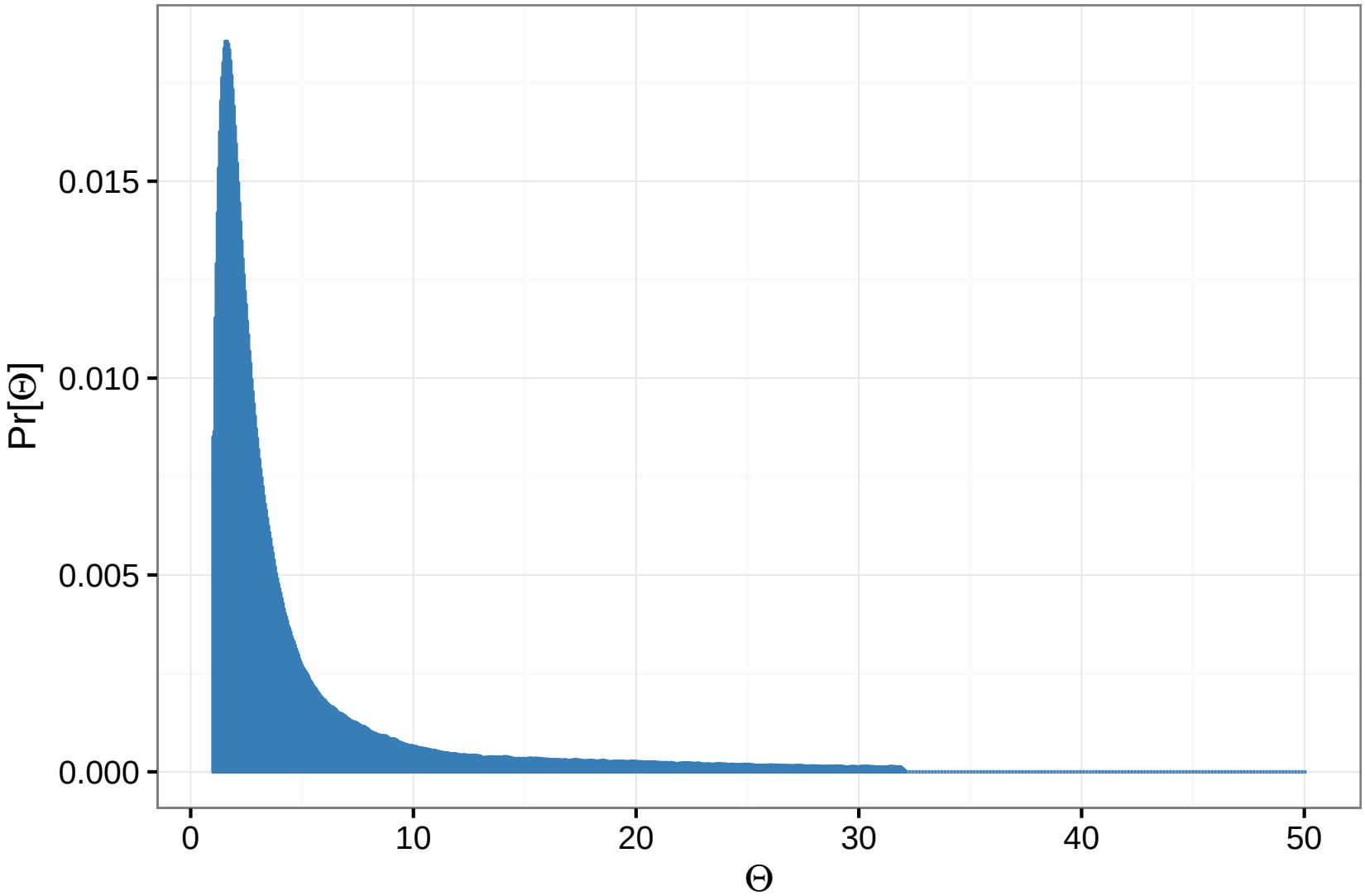
Given a kernel smoothed prior, we can simulate very large sample sizes to get smoothed histogram to use as priors

We sample the medians and simulate normal data using the bandwidth estimate as the standard deviation.

Efflux Ratio: Medians for $m_i \geq 3$



Efflux Ratio Prior Distribution



Likelihood Calculations

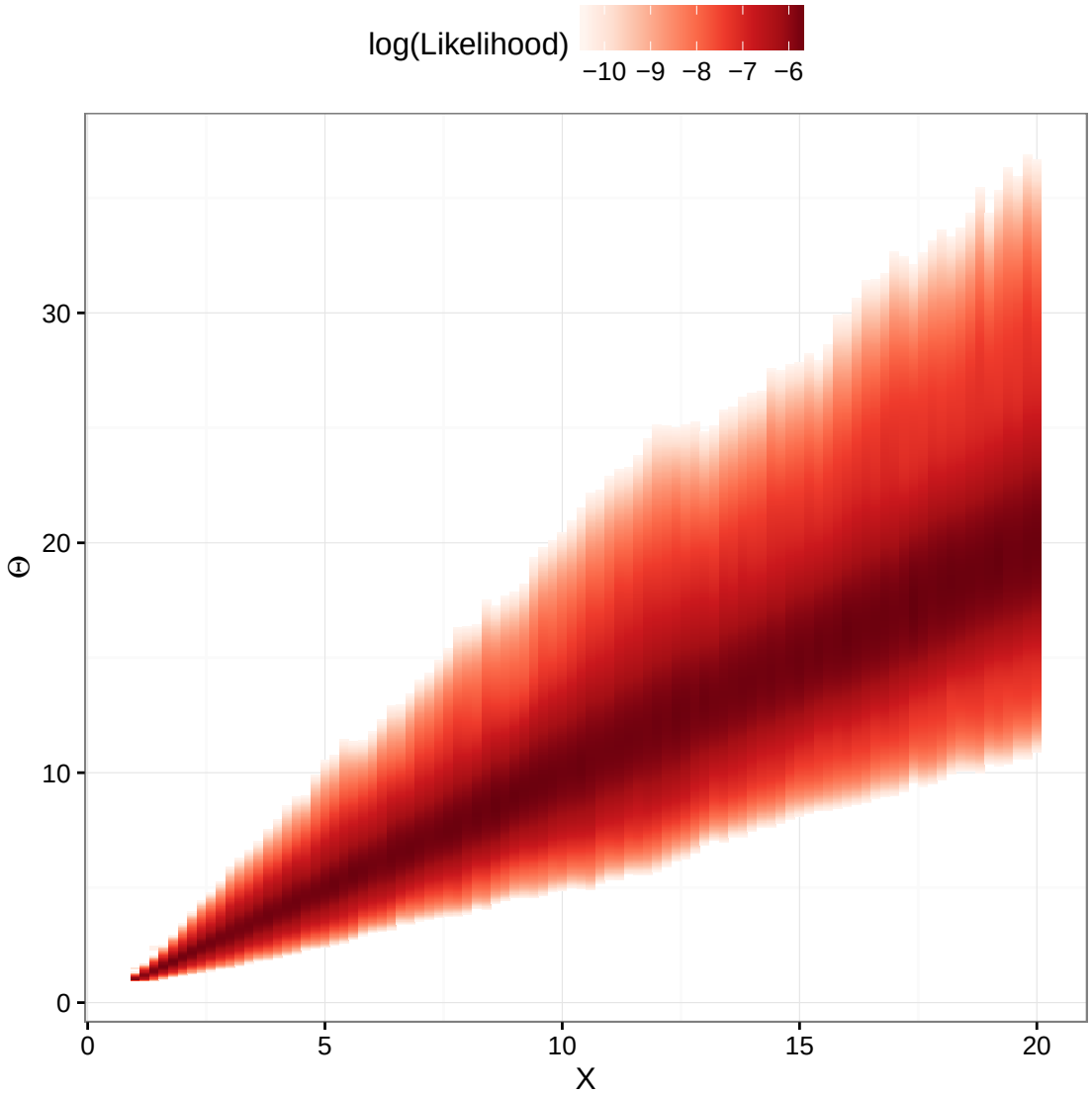
As with the priors, some assay likelihoods are estimated via simple parametric approaches (but may require a variance function over X). We can also compute a *local* nonparametric likelihood. Recall that

- We have discretized Θ and X to reasonably fine grids
- For our application, we precompute the posterior across different values of X
- $Pr[\Theta|X = x] \propto Pr[\Theta] \times Pr[X = x|\Theta]$
- For a **fixed value of** X, the posterior is computed across many values of Θ

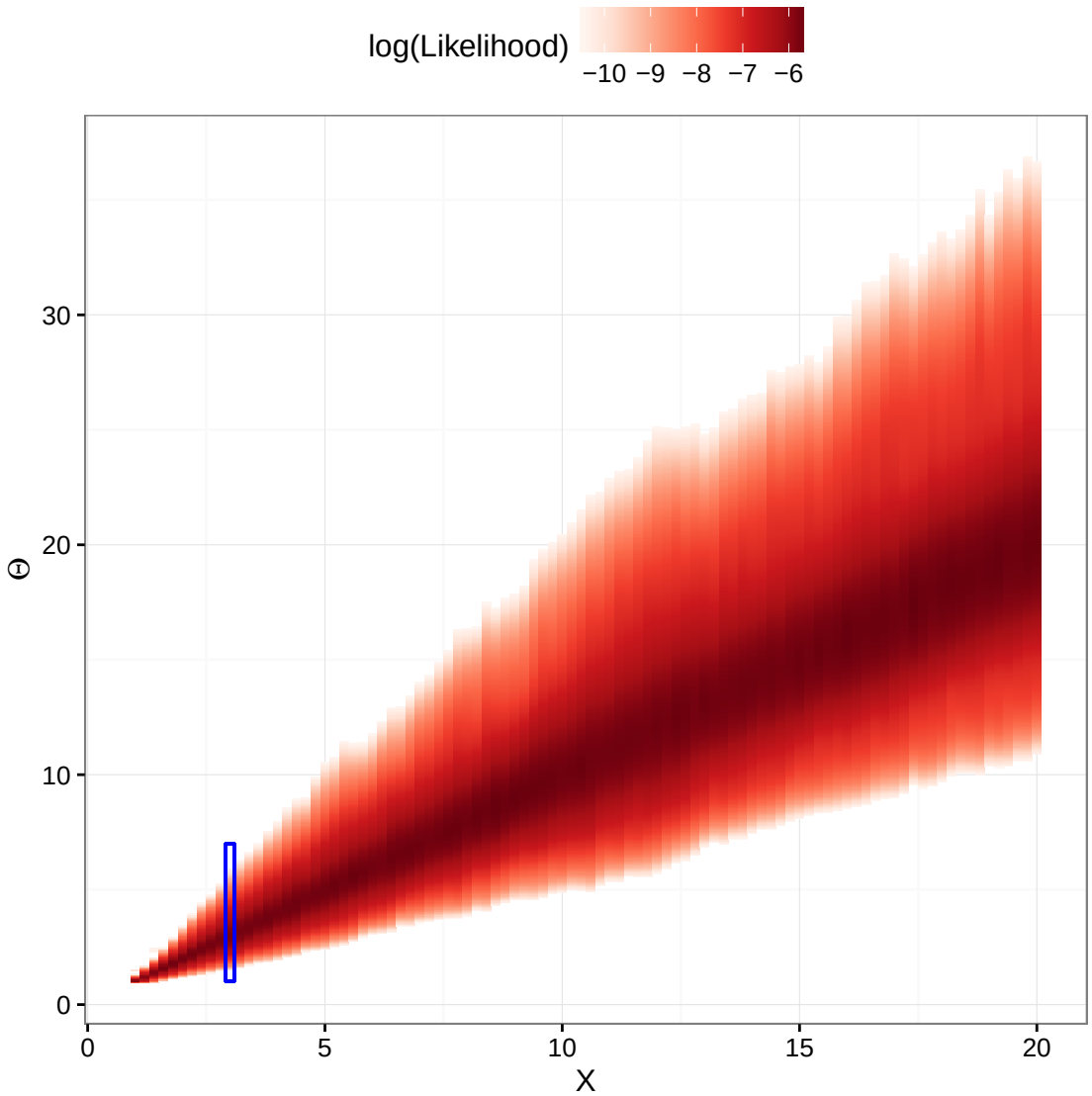
For computational efficiency, our approach is to first fix Θ then compute $Pr[X|\Theta = \theta]$ across X.

We do this for all values of X and use the results to get $Pr[X = x|\Theta]$ for a single fixed value of X.

$Pr[X|\Theta]$ Heatmap for $n = 1$



$Pr[X = 3.0|\Theta]$ Heatmap for $n = 1$



Likelihood aka $Pr[X = 3.0|\Theta]$ aka $\ell(\Theta; X)$

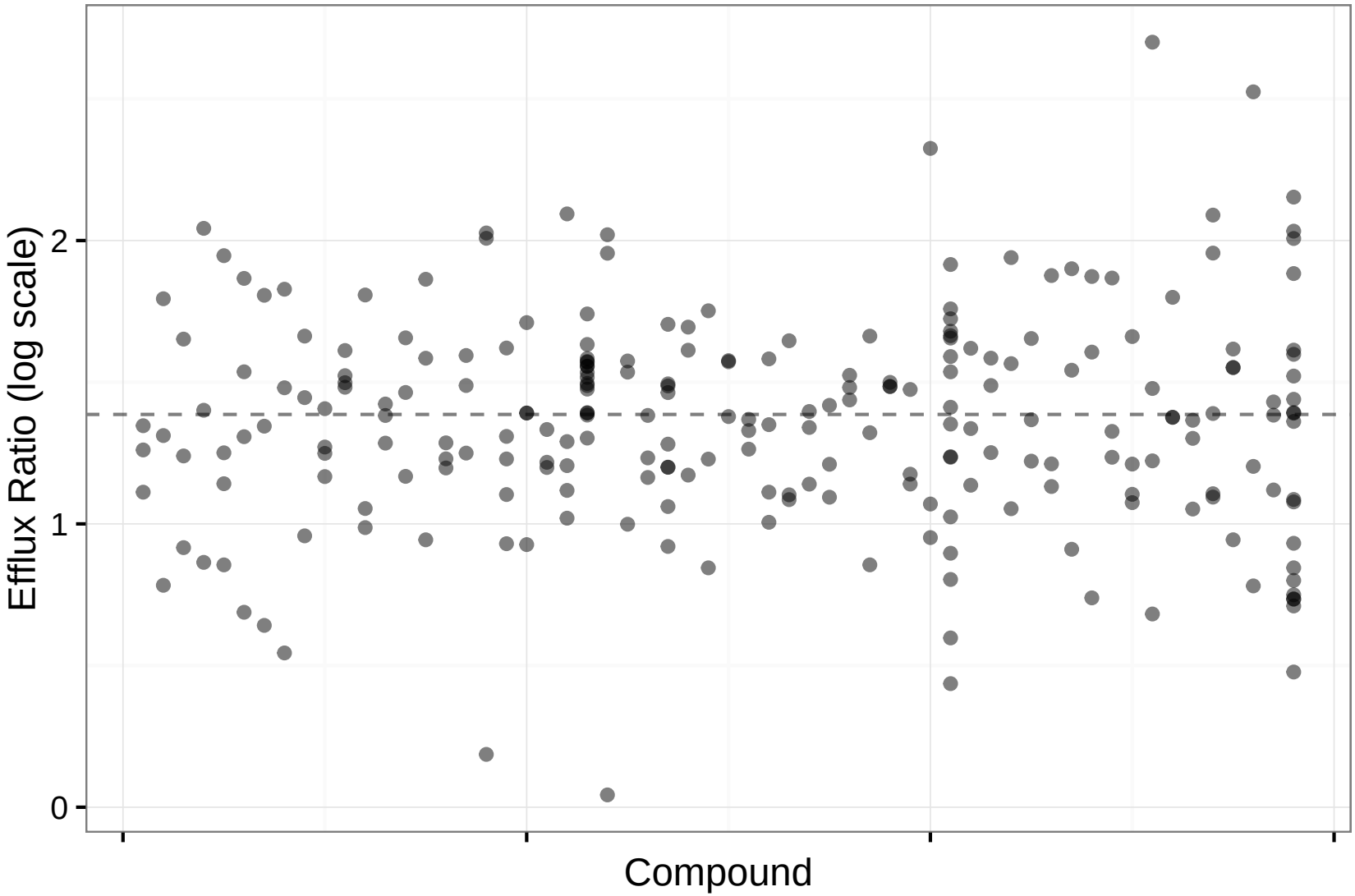
Computing $Pr[X|\Theta = \theta]$

Given the large amount of data on–hand, we can first estimate $Pr[X|\Theta = \theta]$ non–parametrically:

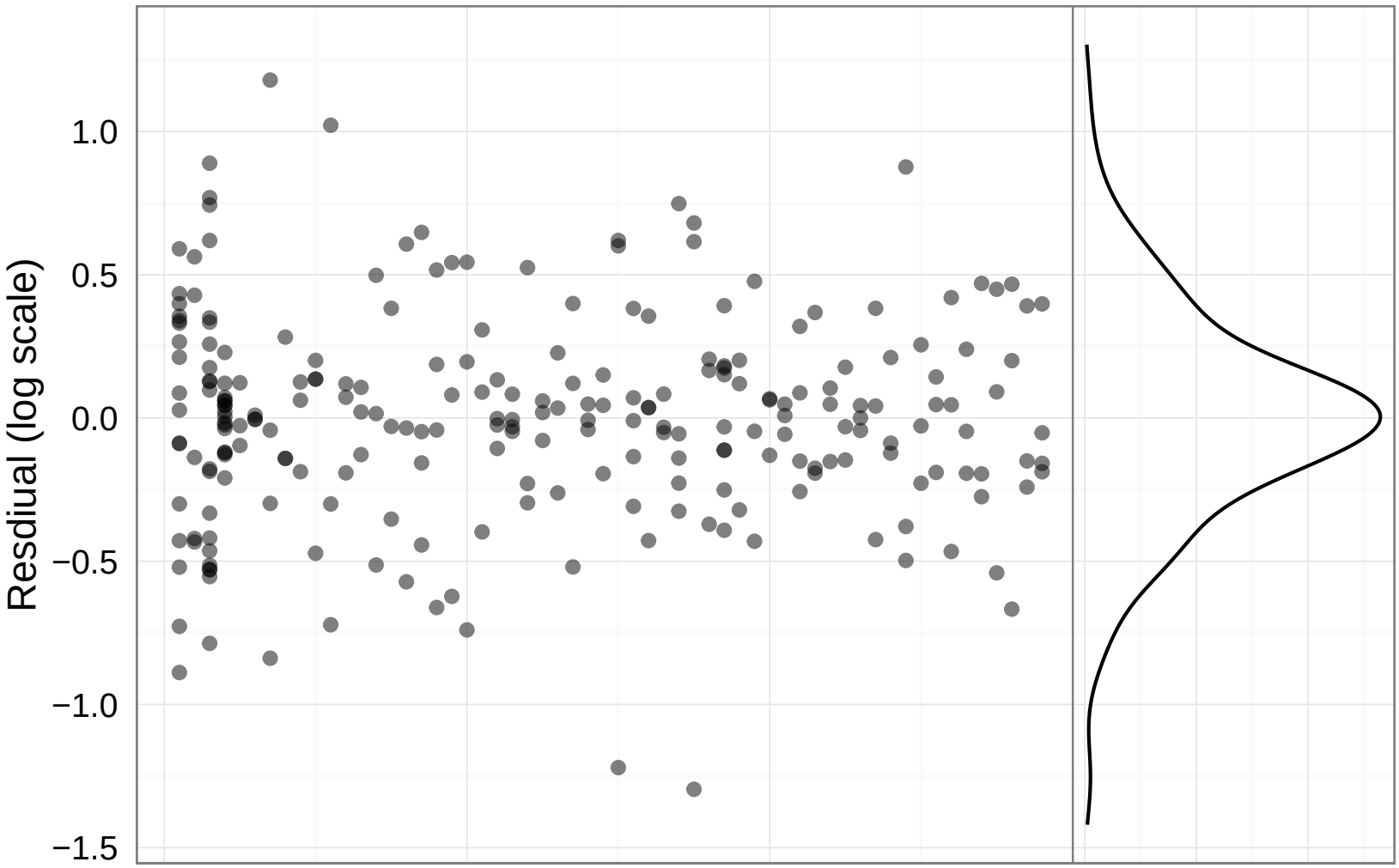
1. For a given value of θ, find the *well–replicated* compounds whose average X_i is within $\theta \pm \epsilon$ ($i = 1 \ldots m$ compounds)
2. Calculate the within-compound residuals $e_{ij} = X_{ij} - X_i$ ($j = 1 \ldots m_i$ replicates)
3. Use the bootstrap to estimate the distribution $Pr[X|\Theta = \theta]$ using a histogram or kernel density estimate

Compounds Near $\Theta = 4.0$

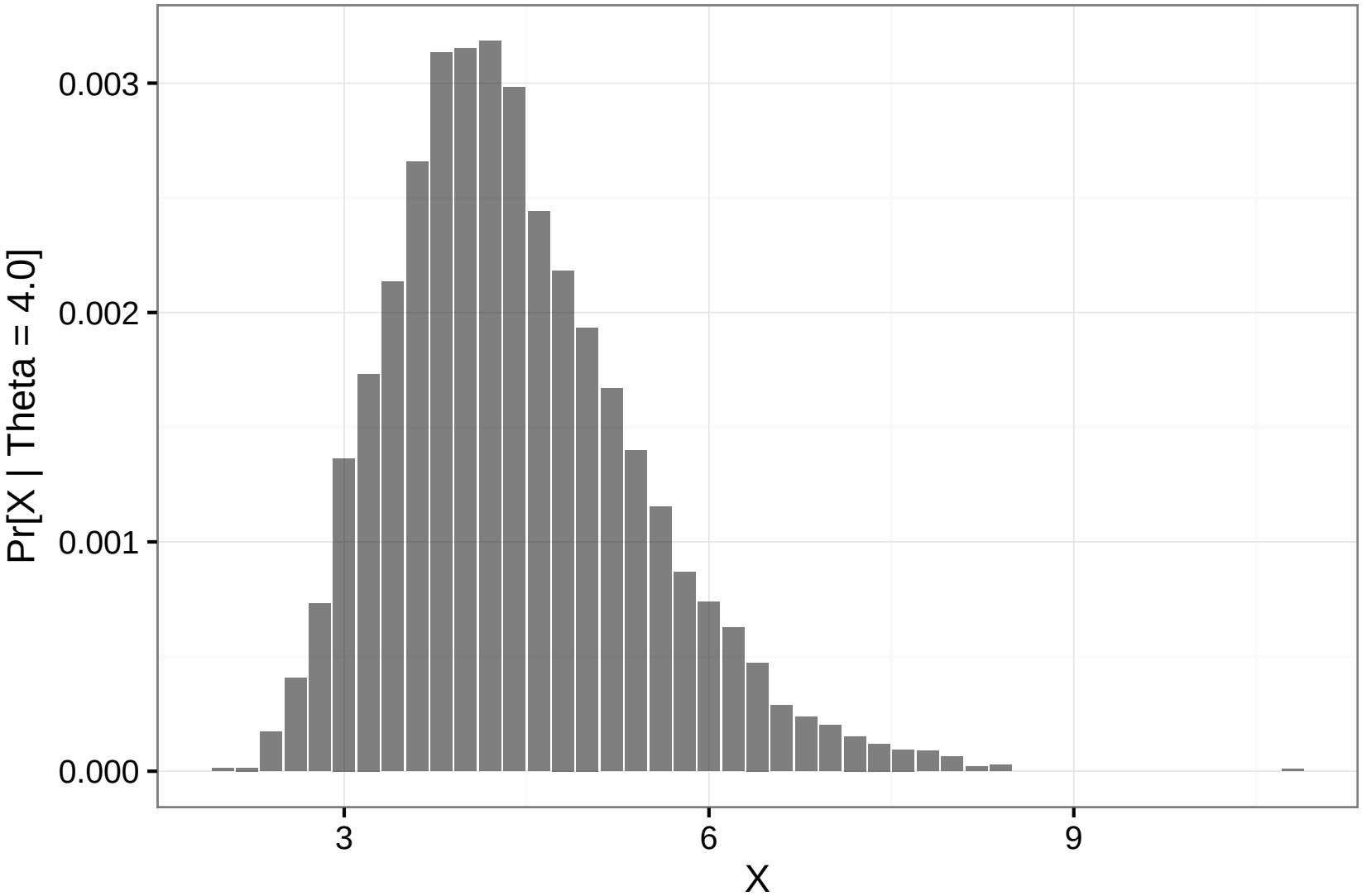
$\epsilon = 0.15$ log units, $m_i \geq 3$ ($m = 58$ compounds and $n = 240$ values)



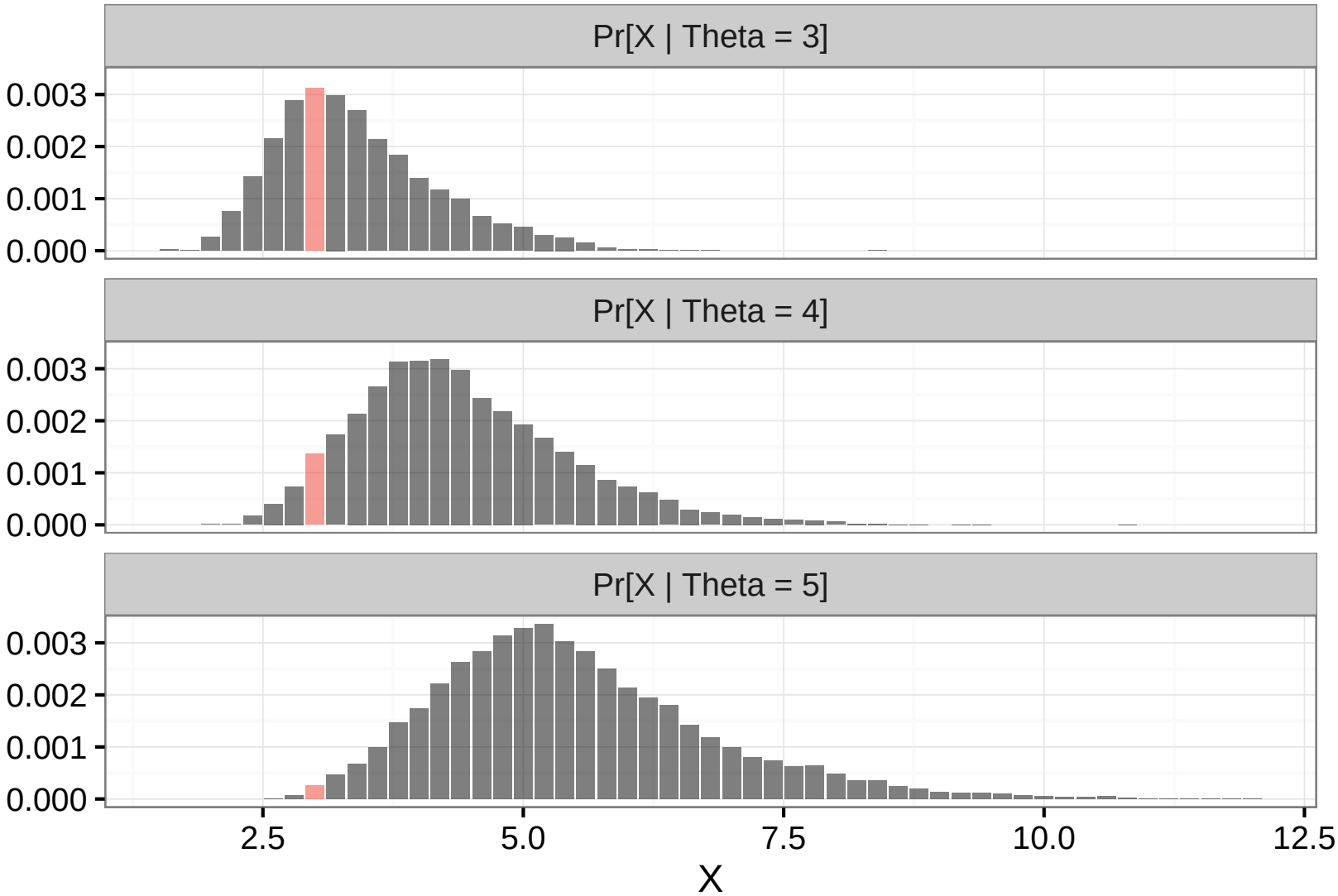
Residuals and Density Estimate



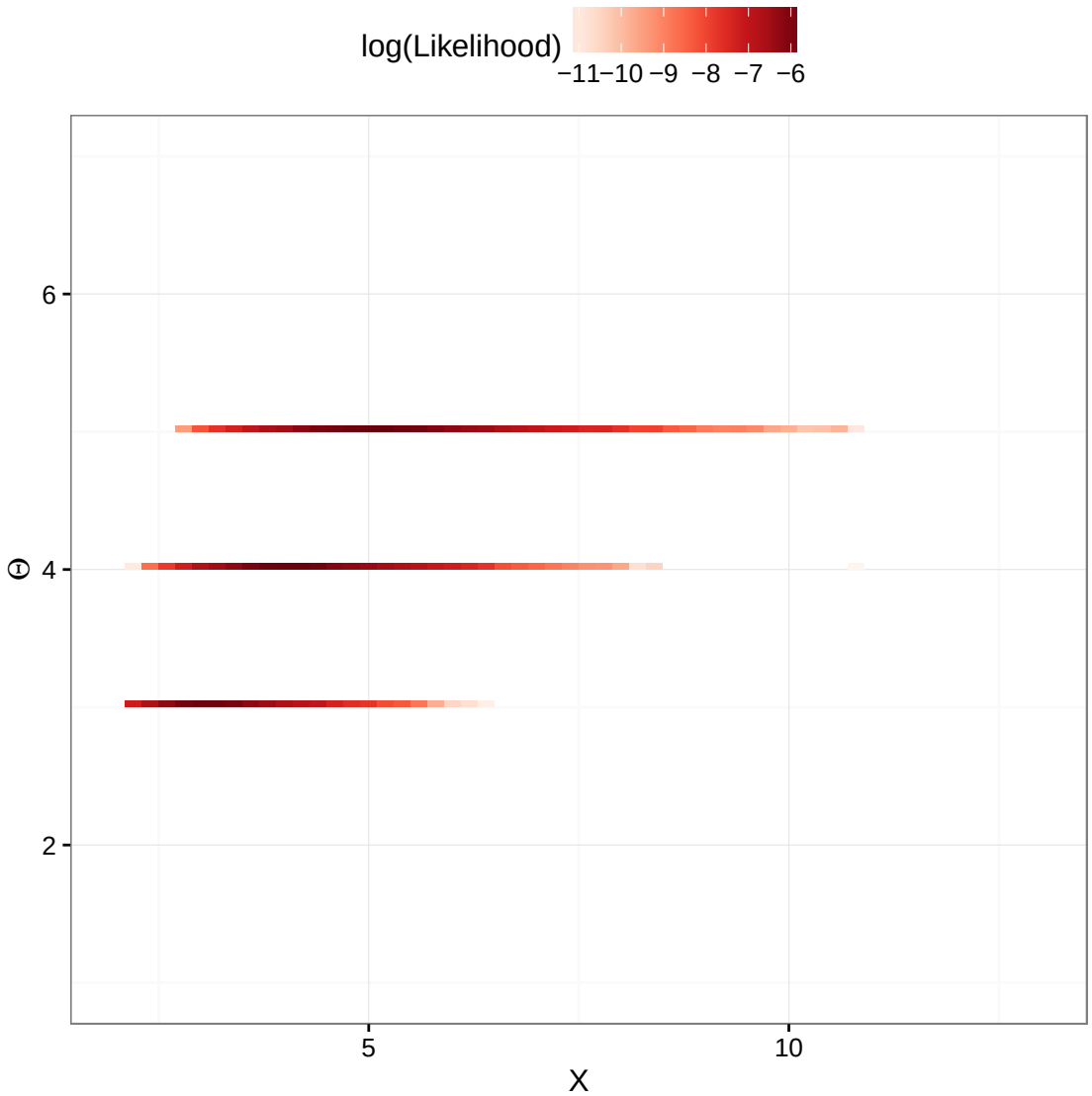
$Pr[X|\Theta = 4.0] \neq$ Likelihood



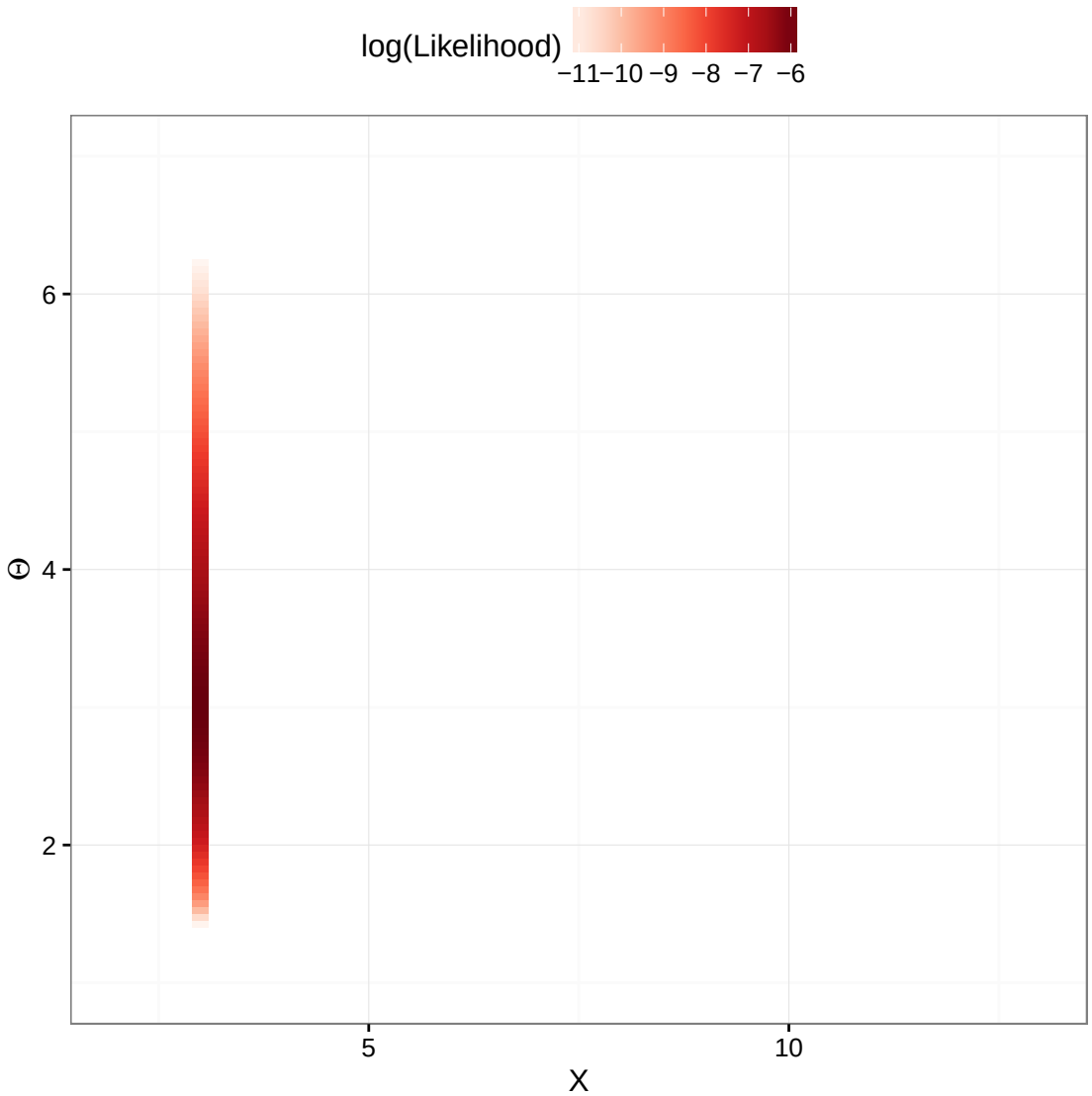
$Pr[X|\Theta \in \{3, 4, 5\}]$



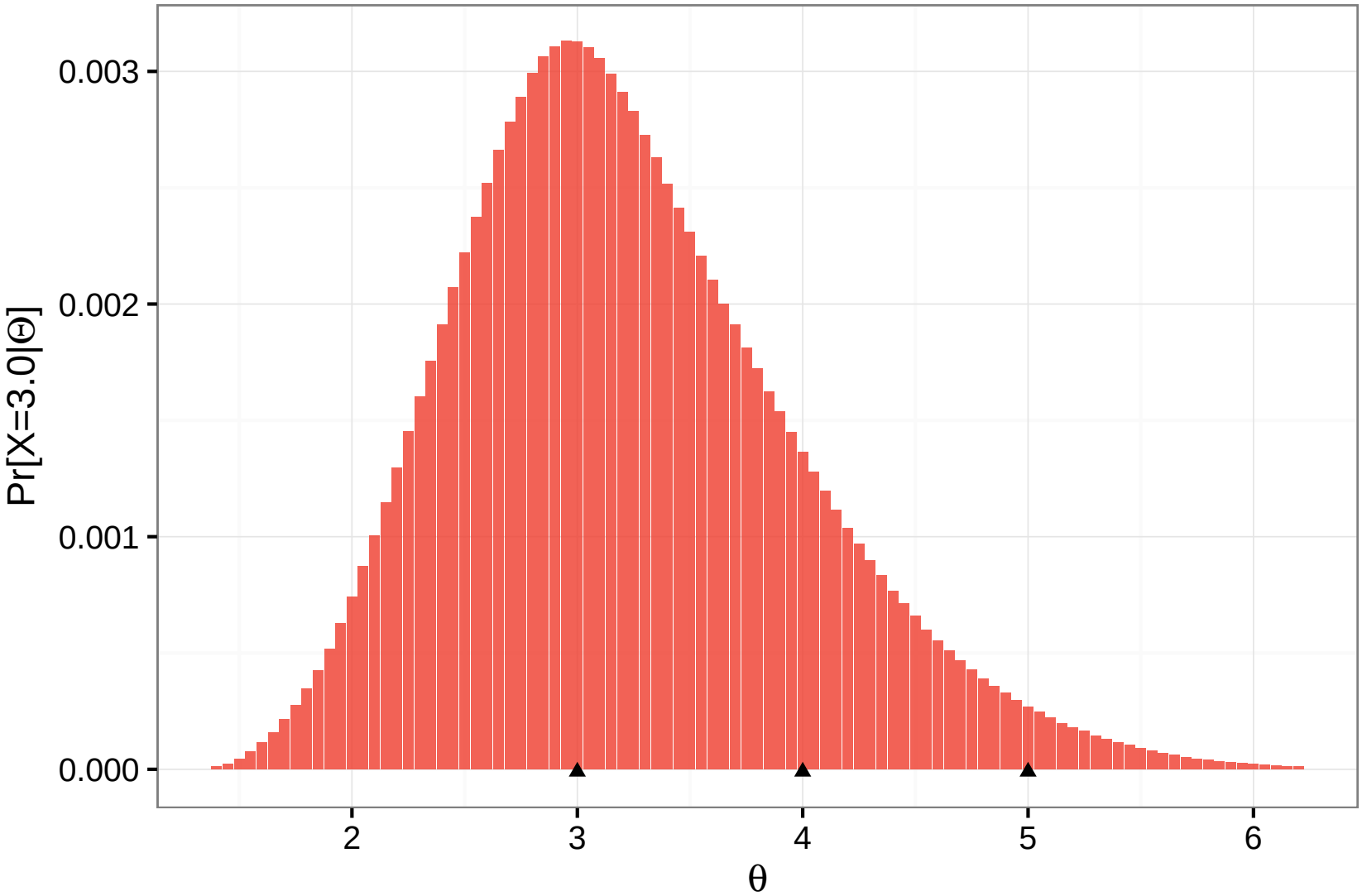
$$Pr[X|\Theta \in \{3, 4, 5\}]$$



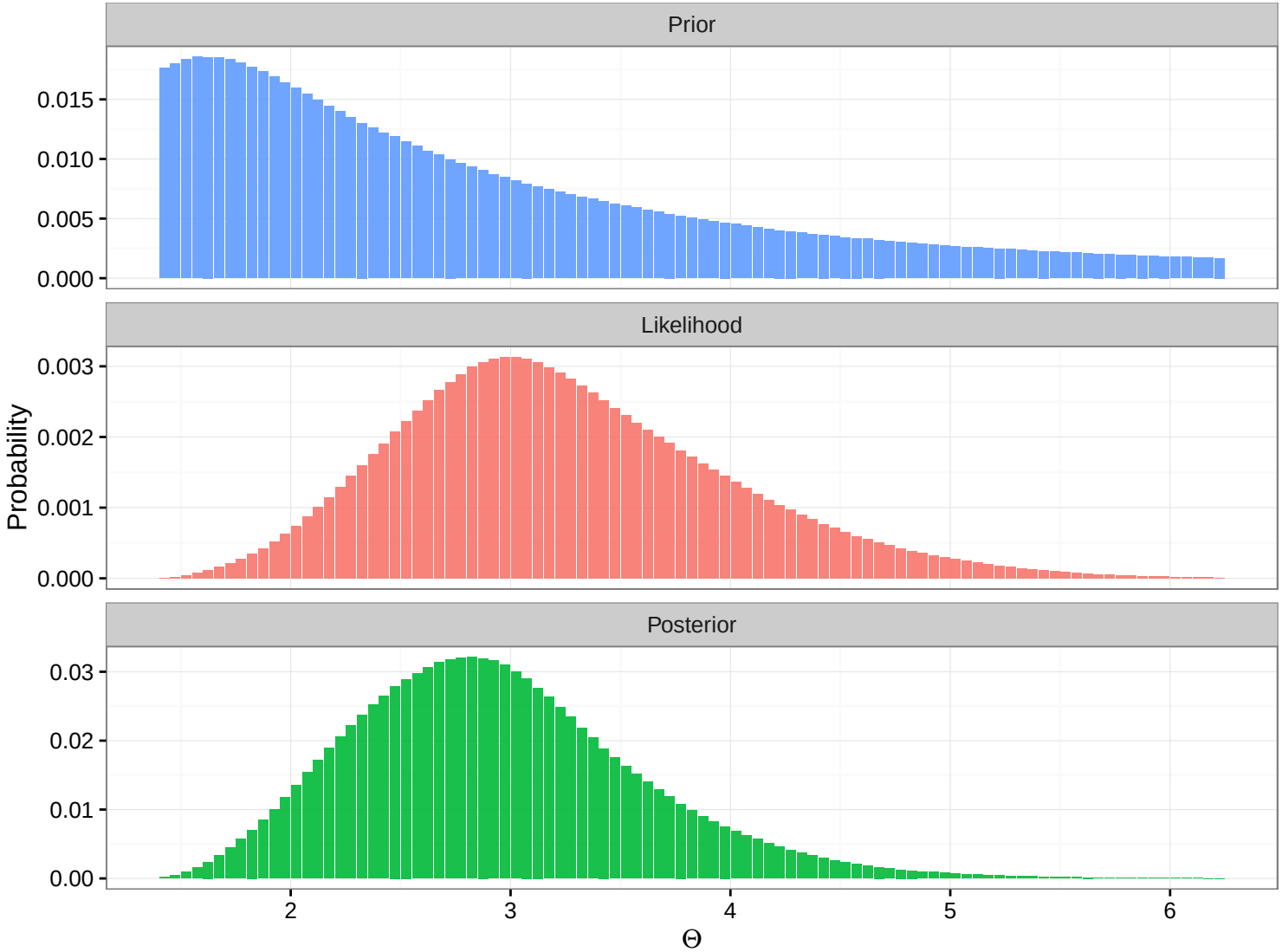
Likelihood aka $Pr[X = 3.0|\Theta]$ aka $\ell(\Theta; X)$



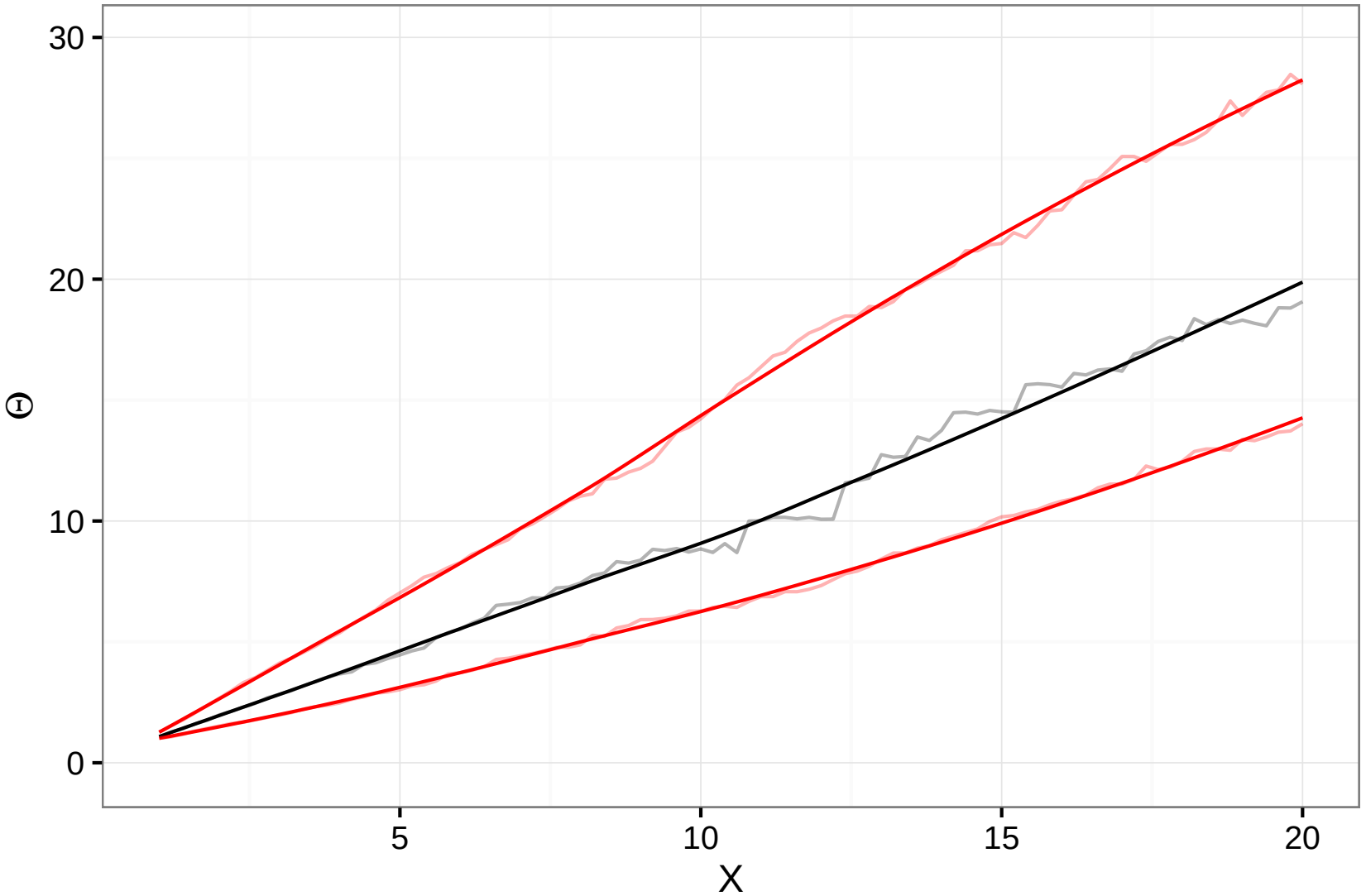
Likelihood aka $Pr[X = 3.0|\Theta]$ aka $\ell(\Theta; X)$



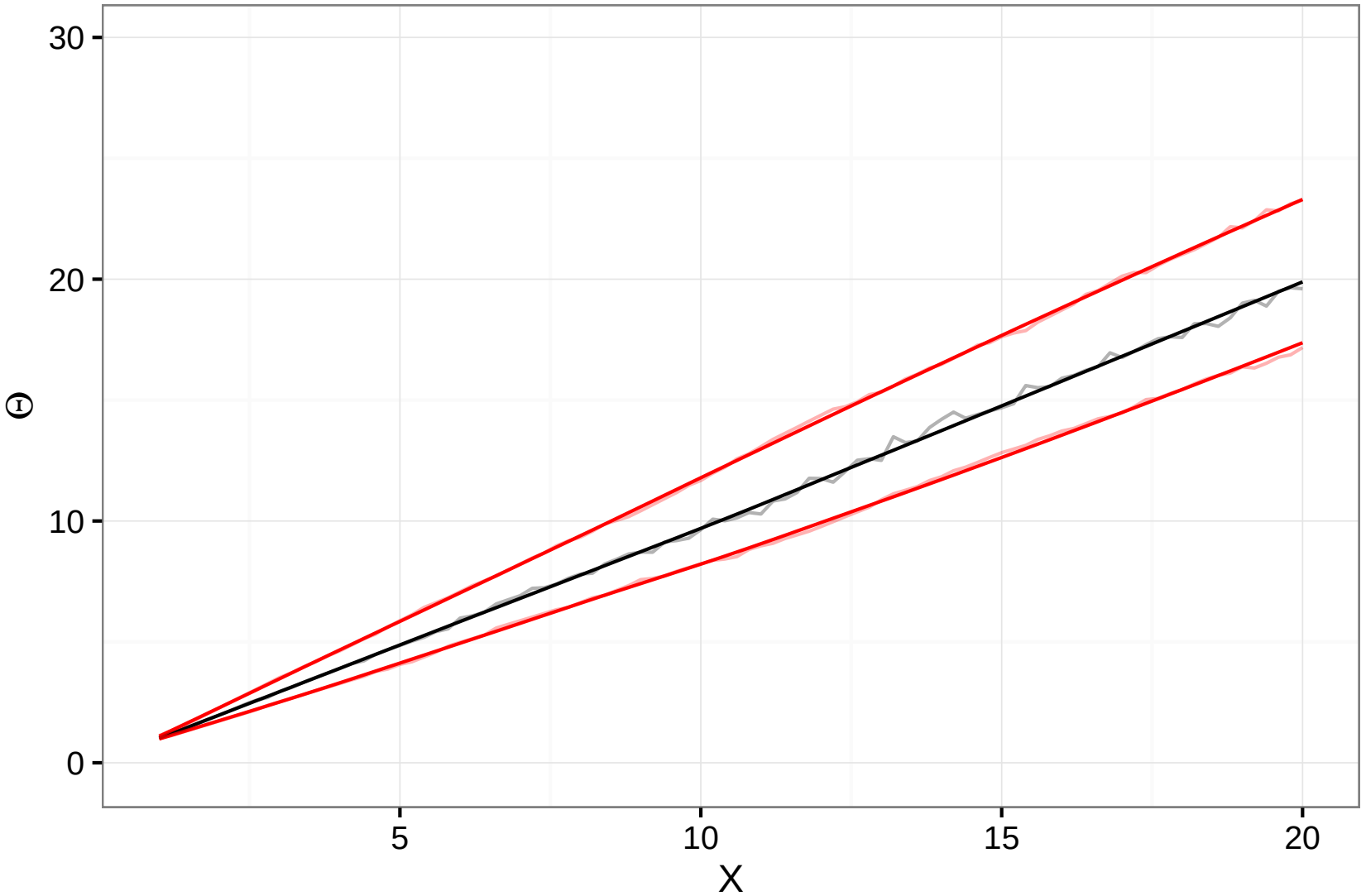
Posterior $Pr[\Theta|X = 3.0]$ for $n = 1$



90% Credible Intervals Across X for $n = 1$



90% Credible Intervals Across X for $n = 5$



Derived Properties of Interest

From the posterior(s), we can estimate a number of important quantities:

- Binning probabilities: $Pr[L \leq \Theta \leq U|X]$
- Fold–difference distribution: $Pr[\Theta_1/\Theta_2|X_1, X_2]$
- Rank–order probability: $Pr[\Theta_1 > \Theta_2|X_1, X_2]$
- Selectivity ratio distribution: $Pr[\Theta/\Lambda|X, Y]$

Making It Accessible

Our primary concern in exposing the scientists to these analyses is to make them approachable and concise. We tried to:

- eliminate jargon (e.g. "posterior", "prior")
- display/isolate small snippets of results
- visualize most results with text to accentuate
- let users *interact* with the analyses
- focus on the *scientific* questions and how we think that they should answer them
- provide a simple and well organized URL to the analyses

(demo!)

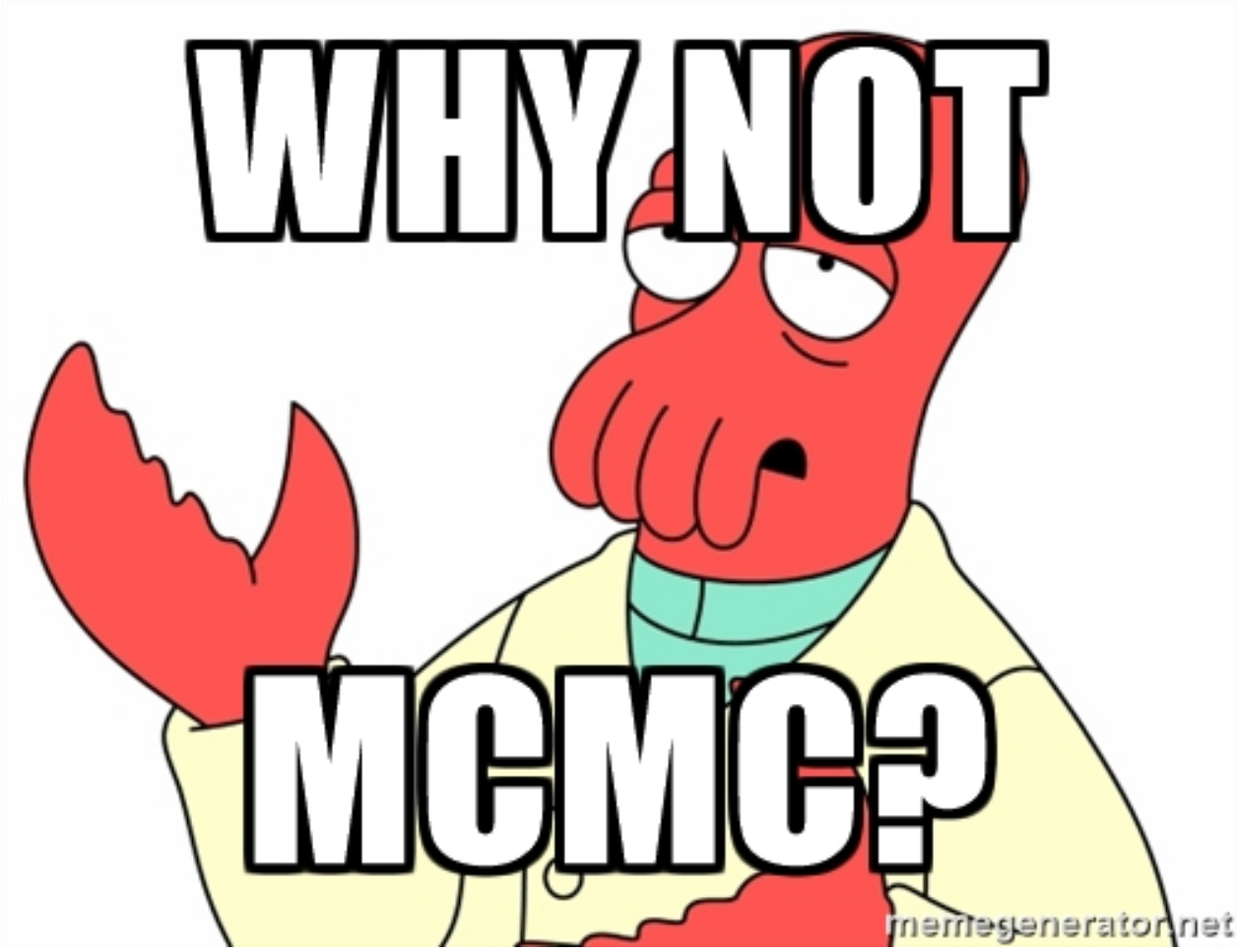
Next Steps

- keep adding more assays
- do a proper Bayesian analysis using MCMC and hyper–priors (i.e. avoid running separate analyses over X)
- add in effects for important molecular properties (e.g. surface area for permeability and efflux assays, IC_{50} curve parameters)
- different priors for different contexts (e.g. lead compounds vs early screening)

Ackowelgements

- Kat Gore for the invitation to speak
- Phillip Yates, Woody Burchett, Kjell Johnson, and Phil Woodward for statistical contributions/technical developments
- Matt Troutman, Chris Keefer, Anthony Carlo and our PDM clients for their support and engagement
- Mike Miller for helping to get shiny servers into production

Backup Slides


WHY NOT
MCMC?
memegenerator.net

Why Not MCMC?

At first, we thought

- we would not be able to use nonparametric priors or properly handle variance heterogeneity over the measurement scale
- on-the-fly computation time was excessive for a user interface (i.e. no pre-computations)
- a hierarchical model would need individual replicate values to estimate posteriors on new compounds

Why Not MCMC?

Now that I'm better informed (i.e. I hired a Bayesian), a hierarchical model would negate these issues.

Hyperparameters at the compound–level would allow for different variances over the measurment range

Computationally, we can precompute and evauate the MCMC fitting then use the posteriors for the hyperparameters as *priors* when we get new data. This tends to converge quickly and we can get predictions of the compound mean.