

Байесовский Вывод

обзор работ

30 января 2026 г.

Содержание

1	Полезные ссылки	3
2	Байесовский анализ данных	4
2.1	Bayesian workflow	4
3	Сравнение и выбор моделей	5
3.1	Байесовский фактор и выбор модели	5
3.2	Сравнение моделей в PyMC	5
3.2.1	Информационные критерии и LOO	7
3.2.2	Усреднение по моделям	8
3.2.3	Факторы Байеса (BF)	9
3.2.4	Полезные ссылки	10
4	Гауссовы процессы	12
4.1	Калибровка измерительной системы	12
5	Анализ данных Chandra: обзор работ D. van Dyk	15
5.1	Low-Resolution Example: Hardness Ratios	16
5.1.1	Поиск рентгеновских двойных и эволюция вспышки Ross 154	19
5.2	High Resolution Deconvolution Methods	21
5.2.1	Задача 1: Spectral Analysis	22
5.2.2	Задача 2: Imaging	22
5.2.3	Задача 3: Reconstruction of a Differential Emission Measure	24
6	Астроинформатика: обзор работ Томаса Лоредо	30
6.1	Байесовский вывод: Больше чем теорема Байеса	31
6.1.1	Роль LTP в байесовском выводе	32
6.1.2	Особенности прайоров	34
6.1.3	Маргинализация по вспомогательным параметрам	36
6.1.4	Систематические ошибки и неточность модели	42
6.1.5	Увеличение дисперсии и расхождение для данных с аддитивным гауссовским шумом	45
6.1.6	Увеличение дисперсии и расхождение для пуассоновских данных	47
6.2	Пуассоновские точечные процессы	49

6.2.1	Оценка блеска TNO	49
6.2.2	Периодограмма и мультимодальность	52
6.2.3	Общая постановка байесовской задачи	55
6.2.4	Модели периодических процессов	56
6.3	Поиск экзопланет	63
6.4	Иерархические модели	65
7	Байесовское планирование эксперимента	65
A	Переводим Байесианскую Библию	66
A.1	Физика «случайных экспериментов»	66
A.1.1	10.1. Интересная корреляция	67
A.1.2	10.2. Исторические предпосылки	68
A.1.3	10.3. Как мошенничать при подбрасывании монет и кубиков	69
A.1.4	10.4. Экспериментальные доказательства	72
A.1.5	10.5. Игра в бридж	73
A.1.6	10.6. Обобщенный случайный эксперимент	75
A.1.7	10.7. И снова об индукции	78
A.1.8	10.8. А что же на счет квантовой теории?	79
A.1.9	10.9. Механика под облаками	80

1 Полезные ссылки

1. [The International Society for Bayesian Analysis](#) и [BayesComp-ISBA](#) – the Bayesian Computation Section of the International Society for Bayesian Analysis.
2. [BayesComp 2025](#): The biennial Bayes Comp meetings are organised by the Bayesian Computation Section of the ISBA. См. также [Bayes Comp 2023](#).
3. [Блог Hector McKimm](#): Н. McKimm – молодой исследователь в Imperial College London, специалист в ML и BDA. На страницах своего блога (на GitHub) он разбирает различные вопросы из области современного анализа данных. В частности, можем порекомендовать следующие посты: [Inference for a Bayesian Spectral Model with PyMC](#), [Bayesian Modelling of Photon Pile-up](#), [The Ratio Estimation Method for Simulation-Based Inference](#) и др.
4. [PHYSTAT-SBI 2024](#): рабочее совещание PHYSTAT в мае 2024 г., посвященное Simulation Based Inference in Fundamental Physics.
5. [Байесовская статистика и вероятностное программирование](#): научно-популярная статья 2017 года в блоге Олега Нагорного.
6. [The Inference Group](#): сайт группы вероятностного вывода Каведишской лаборатории в Кембридже, возглавляемой David MacKay (скончался 14 апреля 2016 года).
7. [Information Filed Theory group](#): группа Торстена Исслина (Torsten Esslin), Head of Data Science at German Centre for Astrophysics. Их основной подход – продвинутые версии Variation Inference. В частности, команда разработала программный пакет [NIFTy](#) (и его ускоренную версию под JAX - [NIFTy.re](#)), в котором реализованы...

Для проведения байесовского вывода (VI) в настоящий момент созданы специальные библиотеки, получившие в литературе название probabilistic programming languages (PPLs). Ниже мы приводим выборочный список с краткими пояснениями (более подробное описание некоторых из них находится далее в основном тексте), более полный список (со ссылками) доступен на [WIKI: Probabilistic programming](#).

- [PyMC](#): a probabilistic programming library that allows users to build Bayesian models with a simple Python API and fit them using MCMC. PyMC strives to make Bayesian modeling as simple and painless as possible, allowing users to focus on their problem rather than the methods. Uses PyTensor as its computational backend to compile through C, Numba or JAX, run your models on the GPU, and benefit from complex graph-optimizations.
- [STAN](#): a PPL for statistical inference written in C++. The Stan is used to specify a Bayesian model with an imperative program calculating the log pdf. The Stan itself can be accessed through several interfaces: 1) CmdStan – a command-line executable for the shell, 2) CmdStanR and rstan – R software libraries, 3) CmdStanPy and PyStan – libraries for the Python programming language, 4) CmdStan.rb – library for the Ruby programming language, 5) MatlabStan – integration with the MATLAB numerical computing environment, 6) Stan.jl – integration with the Julia programming language, 7) StataStan – integration with Stata, 8) Stan Playground – online at <https://stan-playground.flatironinstitute.org/>. См. также [WIKI](#), [Repository](#) и [Stan Users Guide](#), а также [Getting Started](#), [Tutorials](#) и [Case Studies](#). Про применение STAN написано несколько книг, ссылки на них можно найти здесь: [Books using Stan](#).

- [NumPyro](#): lightweight probabilistic programming library that provides a NumPy backend for Pyro. We rely on JAX for automatic differentiation and JIT compilation to GPU/CPU. Подробнее см. [Документацию](#).
- [Pyro](#): a universal PPL written in Python and supported by PyTorch on the backend. Pyro enables flexible and expressive deep probabilistic modeling, unifying the best of modern deep learning and Bayesian modeling. It was designed with these key principles: 1) Universal: Pyro can represent any computable probability distribution. 2) Scalable: Pyro scales to large data sets with little overhead. 3) Minimal: Pyro is implemented with a small core of powerful, composable abstractions. 4) Flexible: Pyro aims for automation when you want it, control when you need it. См. также [Примеры](#) и [Документацию](#).
- [EMCEE](#): stable, well tested Python implementation of the affine-invariant ensemble sampler for MCMC proposed by [\[Goodman2010\]](#). The code is open source and has already been used in several published projects in the Astrophysics literature. См. также [\[Foreman-Mackey2012\]](#) и [Документацию](#).

2 Байесовский анализ данных

По материалам статьи [\[Gelman2020\]](#). См. также ссылки в разделе [Bayesian Workflow](#).

Байесовский подход к анализу данных предоставляет мощный инструмент для работы с неопределённостью во всем – в наблюдениях, параметрах модели и ее структуре – с помощью теории вероятностей. Вероятностные языки программирования упрощают задание и подбор (fit) байесовских моделей, но всё равно оставляют нам множество вариантов построения, оценки и использования этих моделей, а также множество до сих пор нерешённых вычислительных проблем. Использование байесовского вывода для решения реальных задач требует не только статистических навыков, знаний предметной области и программирования, но и понимания решений, принимаемых в процессе анализа данных. Все эти аспекты можно рассматривать как часть сложного *рабочего процесса* (workflow) прикладной байесовской статистики. Помимо вывода, этот процесс также включает в себя итеративное построение модели, ее проверку, валидацию и устранение вычислительных проблем, понимание модели и сравнение разных моделей друг с другом.

2.1 Bayesian workflow

Байесовский вывод (Bayesian inference, BI) – это просто формулировка и вычисление условной вероятности $p(\theta|y) \propto p(\theta)p(y|\theta)$. *Байесовский (рабочий) процесс* (Bayesian workflow, BW) включает в себя несколько этапов: построение модели, вывод и проверка/улучшение модели, а также сравнение различных моделей не только для выбора или усреднения моделей, но и для лучшего понимания этих моделей.

Нам нужен BW, а не просто BI, по нескольким причинам.

1. Вычисления могут оказаться непростой задачей, и часто приходится проходить через различные этапы, включая подбор более простых или альтернативных моделей, приближенные вычисления, которые менее точны, но выполняются быстрее, а также изучение процесса подбора, чтобы прийти к выводам, которым мы сможем доверять.
2. В сложных задачах, обычно, заранее не известно, какую модель использовать, и даже в тех редких случаях, когда подходящая модель была выбрана заранее, мы, как

правило, хотим расширять её по мере сбора дополнительных данных или задавать более подробные вопросы об уже имеющихся.

3. Даже если бы наши данные были статичными, и мы знали, какую модель использовать, и у нас не возникло бы проблем с её подбором, мы всё равно хотели бы понять подобранную модель и её взаимосвязь с данными, а такое понимание часто можно получить, лишь сравнивая выводы из ряда связанных моделей.
4. Иногда разные модели приводят к разным выводам, и ни один из них не является однозначно благоприятным. В таких случаях полезно представить несколько моделей, чтобы проиллюстрировать неопределённость при выборе модели.

ПРОДОЛЖЕНИЕ СЛЕДУЕТ...

3 Сравнение и выбор моделей

3.1 Байесовский фактор и выбор модели

В этом разделе будет сделан обзор работ (и программных продуктов), посвященных вычислению *обоснованности* (evidence) – усредненной по прайору функции правдоподобия. Именно обоснованность встречается при байесовском сравнении моделей, а отношение обоснованностей двух моделей позволяет сравнивать «качество» описание моделями измеренных данных (подробнее см.).

- Статья [McEwen2021] Machine learning assisted Bayesian model comparison: learnt harmonic mean estimator, и программная реализация данного алгоритма [harmonic](#).
-

3.2 Сравнение моделей в PyMC

Данный раздел составлен по материалам главы 5 книги Osvaldo Martin, Bayesian Analysis with Python (2024), и нескольким примерам с сайта <https://www.pymc.io>. Обращаем внимание, что использованные фрагменты кода доступны в виде jupyter notebook на страничке GitHub, посвященной 3-му изданию книги: <https://github.com/alocstavodia/BAP3>. Коды предыдущих изданий доступны здесь: <https://github.com/alocstavodia/BAP>.

Любая модель некорректна! Но какие-то из них лучше, чем другие!!!

Возможно такой девиз звучит несколько странно для «классического» физика, привыкшего, что венцом научного поиска является установление закономерностей в форме той или иной «истинной формулы». Но с точки зрения практических измерений и их интерпретации никакая наша модель (*интерпретирующая модель*) не может сравниться по сложности с тем, что стоит за процессом создания данных (*генерационная модель*). В каком-то смысле, моделируя, мы и хотим выявить в данных *структуру* с помощью более простой, чем генерационная, модели. А значит, интерпретирующая модель заведомо не идентична генерационной. И все же... какие-то из них лучше, чем другие. Но что именно это означает – как правильно сравнивать модели между собой и выбирать лучшую?

Первый большой блок способов сравнения моделей, основан на *постериорных предсказаниях* или PPC (от англ. Posterior Predictive Checks). Байесовская модель позволяет получить не только сэмплы θ^s из постериорного распределения $p(\theta|\mathcal{D})$, но и генерировать новые данные \mathcal{D}_{new} из постериорного предсказательного распределения $p(\mathcal{D}_{\text{new}}|\mathcal{D}) =$

$\int p(\mathcal{D}_{\text{new}}|\theta) p(\theta|\mathcal{D}) d\theta$. Технически это происходит так: сначала выбирается какой-то из постериорных сэмплов θ^s , после чего для генерации \mathcal{D}_{new} используется то самое распределение, которое было указано в байесовской модели с аргументом `observed`. То есть при вызове `pm.sample` это распределение было функцией параметров модели θ и играло роль правдоподобия, а теперь, при вызове¹ `pm.sample_posterior_predictive`, оно является *сэмплирующим распределением данных* $p(\mathcal{D}_{\text{new}}|\theta = \theta^s)$.

Цель PPC – выяснить ограничения модели с помощью сгенерированных новых данных и сравнения их с исходными (наблюдаемыми). Никакая модель не может воспроизвести все аспекты проблемы, мы конструируем модель, подразумевая лишь какой-то определённый аспект. Разные модели могут «концентрироваться» на разных аспектах.

А что можно сравнивать?

Вариант 1: сравниваем распределения данных посредством `az.plot_ppc`.

```
idata = pm.sample(2000, idata_kwargs={'log_likelihood': True})
pm.sample_posterior_predictive(idata, model=model, extend_inferencedata=True)
az.plot_ppc(idata, num_pp_samples=100)
```

Обращаем внимание, что функции `pm.sample_posterior_predictive` надо передать как сами постериорные сэмплы параметров, так и используемую модель – иначе откуда она возьмет распределение сэмплов данных? А для чего нужен аргумент `idata_kwargs` функции `pm.sample`? Оказывается, для вычисления информационных критериев типа `waic` и `loo` (о них расскажем чуть позже) необходимо в `idata` сохранить не только постериорные сэмплы, но и значения логарифма правдоподобия.

Вариант 2: сравниваем выбранные статистики посредством `az.plot_bpv`. `Bpv` – это Bayesian p-value, и в отличие от «классического» p-value здесь весь расчет проводится не при фиксированном значении параметра, а на основе постериорных сэмплов:

```
az.plot_bpv(idata, kind="t_stat", t_stat="mean")
az.plot_bpv(idata, kind="t_stat", t_stat=iqr)
```

Первый вызов подразумевает оценку `bpv` по функции `np.mean`, а второй – по определенной пользователем функции `iqr` (например как `np.subtract(*np.percentile(x, [75, 25], axis=a))`). Вызов `az.plot_bpv` эквивалентен

```
y_pred=idata.posterior_predictive['y_pred'].stack(samples=('chain', 'draw')).values.T
T_obs = func(y_obs)
T_sim = func(y_pred, a=1)
p_value = np.mean(T_sim >= T_obs)
az.plot_kde( T_sim, plot_kwargs={'color': c}, label=f'p-value {p_value:.2f}' )
c func=np.mean или func=iqr.
```

Вариант 2:* `az.plot_bpv(idata_1, kind="u_value")`. Этот вариант `az.plot_bpv` связан с более широкой точкой зрения на статистику. Что если значение каждой точки данных \mathcal{D} будет выступать в роли статистики! Тогда мы получим целую совокупность `bpv` – столько, сколько у нас точек данных. Эту совокупность, рассматриваемую как гистограмма, снова можно `kde`-аппроксимировать и получить то, что называется распределением `u-value`. Если модель адекватно описывает данные во всем интервале измеренных значений, то распределение `u-value` будет равномерным (отсюда и название). Соответственно,

¹Интересно, что `pm.sample_posterior_predictive` допускает замену сэмплирующего распределения – лишь бы названия и размерности фигурирующих в нем параметров совпадали с теми, что передаются в нее в форме постериорного трейса. Подробнее об этом см. раздел ??.

отличия от равномерного распределения (белая линия на графике) свидетельствуют о плохом описании данных моделью.

3.2.1 Информационные критерии и LOO

Рассмотрим регрессию полиномами разного порядка q . Качество аппроксимации y_{reg} можно контролировать статистикой R^2 (в этом разделе $\mathcal{D} = \{y_i\}, i = 1, \dots, N$):

$$R^2 = \frac{\sum_i (y_{\text{reg},i} - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}, \quad \bar{y} \equiv \sum_i y_i / N$$

Чем больше q , тем R^2 ближе к 1. Плохо, если q мало (underfitting), но и если велико – тоже плохо, хотя $R^2 = 1$ (overfitting). Необходимо искать баланс между простотой и точностью! Часто про такой подход говорят как о *принципе Бритвы Оккама*.

Примечание. Различают два типа точности – within-sample accuracy (the accuracy is measured with the same data used to fit the model) и out-of-sample accuracy (the accuracy measured with data not used to fit the model). Информационные критерии как раз и придуманы так, чтобы, основываясь на within-sample accuracy (goodness-of-fit), ввести компенсацию на сложность – penalty (model complexity). Альтернативный подход получил название cross-validation: an empirical strategy based on dividing the available data into separate subsets that are alternatively used to fit and evaluate the models.

Самым простым вариантом оценки точности является минимизация $\frac{1}{N} \sum_i (y_i - \mathbb{E}[y_i|\theta])^2$, где матожидание берется по постериорному распределению. В случае не-гауссовых правдоподобий его аналогом будет goodness-of-fit = $\sum_i \ln p(y_i|\theta)$.

А чем можно штрафовать goodness-of-fit? Чем-то связанным с количеством параметров модели k . Правда, желательно использовать что-то вроде эффективного числа параметров – не все параметры равноценны с точки зрения «приобретения информации».

Совсем не-байесовским является **информационный критерий Akaike (AIC)**:

$$\text{AIC} = -2 \sum_i \ln p(y_i|\hat{\theta}_{\text{MLE}}) + 2k,$$

где $\hat{\theta}_{\text{MLE}}$ – оценка максимального правдоподобия. Обращаем внимание, что информационные критерии часто формируются именно в таком виде – с множителем «-2», т.е. лучшим моделям соответствуют меньшие критерии. В PyMC информационные критерии в функциях типа az.waic, az.loo и az.sormae можно вызывать с дополнительным аргументом scale принимающим значение "deviance" или "log" в зависимости от того, хотите ли вы использовать этот множитель или нет.

В байесовской парадигме для количественной оценки goodness-of-fit вводят понятие lpd – log pointwise predictive density:

$$\text{lpd} = \sum_i \ln p(y_i|y) = \sum_i \ln \int p(y_i|\theta) p(\theta|y) d\theta$$

и его оценку с помощью постериорных сэмплов $\{\theta^s\}, s = 1, \dots, S$:

$$\widehat{\text{lpd}} = \sum_i \ln \left(\frac{1}{S} \sum_{s=1}^S p(y_i|\theta^s) \right)$$

Widely applicable information criteria (WAIC) основан на разности $\widehat{\text{lpd}}$ и $\hat{p}_{\text{waic}} = \sum_i \mathbb{V}[\ln p(y_i|\theta)]$ – оценки эффективного числа параметров по постериорной дисперсии. На практике ее заменяют выборочной дисперсией $V_{s=1}^S[a] = \frac{1}{S-1} \sum_{s=1}^S (a_s - \bar{a})^2$:

$$\text{WAIC} = -2 \sum_i \ln \left(\frac{1}{S} \sum_{s=1}^S p(y_i|\theta^s) \right) + 2 \sum_i (V_{s=1}^S \ln p(y_i|\theta^s))$$

Кстати, есть еще DIC (Deviance Information Criterion) и BIC (Bayesian Information Criteria), но оба они хуже WAIC (а BIC, к тому же, несмотря на свое название, не совсем байесовский!).

Наконец, пожалуй, самым продвинутым и чаще всего используемым является критерий **leave-one-out cross-validation (LOO-CV)** или просто LOO:

$$\text{LOO} = \sum_i \ln \int p(y_i|\theta) p(\theta|y_{-i}) d\theta \approx \sum_i \ln \left(\frac{1}{S} \sum_{s=1}^S p(y_i|\theta_{-i}^s) \right)$$

причем постериорное распределение $p(y_i|\theta_{-i}^s)$ получают методом сэмпирования по важности со сглаживанием Парето (это отдельная тема – как-нибудь и про нее расскажу).

```
az.waic(idata)
idata_dic = {'model1':idata1, 'model2':idata2}
az.compare(idata_dic)
cmp_df = az.compare(idata_dic,method='BB-pseudo-BMA',ic="waic",scale="deviance")
az.plot_compare(cmp_df)
```

На вход функции `az.compare` передается словарь со значениями в виде трейсов, содержащих постериорные и постериорные предсказательные сэмплы, а также значения логарифма правдоподобия (именно поэтому мы изначально вызывали `pm.sample` с аргументом `idata_kwargs={'log_likelihood': True}`). По умолчанию в `az.compare` рассчитывается `loo`, а `scale="log"`.

3.2.2 Усреднение по моделям

Если уж быть совсем последовательным байесианцем, то надо не выбирать одну единственную (наилучшую) модель, а учитывать их все – усреднять по ним. Веса, по идее, должны служить величины $p(\mathcal{M}_i|\mathcal{D})$ – постериорные вероятности рассматриваемых моделей (подробнее о них в следующем разделе).

Если вес выбран, то получить взвешенное постериорное предсказательное распределение не представляет труда:

```
w = 0.5
y_12 = az.weight_predictions([idata1, idata2], weights=[w, 1-w])
az.plot_kde(y_12.posterior_predictive['y_pred'].values.T)
```

А как получить этот вес? Подсказка: посмотрите внимательнее, что возвращает `az.compare`.

Примечание. Интересно, что если у вас есть две очень хорошие и близкие друг к другу модели и еще несколько других, то функция `az.compare` одной из этих двух «хороших» моделей назначит очень низкий вес, практически нулевой даже по сравнению с остальными «плохими» моделями. Догадываетесь почему?

3.2.3 Факторы Байеса (BF)

Но байесовская парадигма дает и свой собственный критерий, не имеющий аналогов в традиционной математической статистике, который автоматически следует принципу Бритвы Оккама, т.е. учитывает как goodness-of-fit, так и сложность модели. При этом для каждой конкурирующей модели вычисляется ее постериорная вероятность

$$p(\mathcal{M}|\mathcal{D}) = \frac{p(\mathcal{M}) p(\mathcal{D}|\mathcal{M})}{p(\mathcal{D})},$$

где самым важным является множитель $p(\mathcal{D}|\mathcal{M})$, получивший название *обоснованности* (evidence) или *маргинального правдоподобия*. Последнее название связано с тем, что

$$p(\mathcal{D}|\mathcal{M}) = \int p(\mathcal{D}|\mathcal{M}, \theta) p(\theta|\mathcal{M}) d\theta$$

Отношение вероятностей двух моделей, например \mathcal{M}_1 и \mathcal{M}_2 , определяется как

$$\frac{p(\mathcal{M}_1|\mathcal{D})}{p(\mathcal{M}_2|\mathcal{D})} = \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)} \cdot \frac{p(\mathcal{D}|\mathcal{M}_1)}{p(\mathcal{D}|\mathcal{M}_2)}$$

Это отношение распадается на произведение двух сомножителей: первого, определяющего только априорными допущениями об (относительной) правдоподобности моделей, и второго – отношения обоснованностей, получившего в литературе название *байесовского фактора* BF²: $BF_{12} \equiv p(\mathcal{D}|\mathcal{M}_1)/p(\mathcal{D}|\mathcal{M}_2)$.

Только этот второй множитель зависит от данных – и именно его сложнее всего вычислить, особенно для моделей в пространстве параметров большой размерности. Собственно говоря, сэмплирование методом МСМС и придумали, чтобы получать постериорное распределение, не вычисляя этот знаменатель.

Так как же быть – как вычислять BF на практике?

Способ №1: аналитически. Очень редко, когда удается сделать оценку маргинального правдоподобия в явном виде, один из примеров – BetaBin-модель, т.е. ситуация, когда в качестве сэмплирующего данные распределения выступает биномиальное, а априорным на его параметр является бета-распределение. В этом случае $p(\mathcal{D}|\mathcal{M}) \propto B(\alpha + h, \beta + N - h)/B(\alpha, \beta)$, где $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ – бета-функция, h – количество «успехов» в данных, α, β – параметры прайора.

Другой часто используемый в образовательных целях пример – Дирихле-мультиномиальная модель или *модель Поля*. В этой модели, являющейся обобщением BetaBin-модели, $p(\boldsymbol{\theta}) = \text{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_k \theta_k^{\alpha_k - 1}$, $p(\mathcal{D}|\boldsymbol{\theta}) = \frac{N!}{\prod_k N_k!} \prod_k \theta_k^{N_k}$, $p(\boldsymbol{\theta}|\mathcal{D}) = \text{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha} + \mathbf{N})$. Отсюда можно получить выражение для обоснованности в терминах обобщенной бета-функции:

$$p(\mathcal{D}) = \frac{B(\mathbf{N} + \boldsymbol{\alpha})}{B(\boldsymbol{\alpha})}, \quad B(\boldsymbol{\alpha}) \equiv \frac{\prod_k \Gamma(\alpha_k)}{\Gamma(\sum_k \alpha_k)}.$$

Чаще данное выражение записывают следующим образом:

$$p(\mathcal{D}) = \frac{\Gamma(\alpha_0)}{\Gamma(N + \alpha_0)} \prod_k \frac{\Gamma(N_k + \alpha_k)}{\Gamma(\alpha_k)}, \quad N = \sum_k N_k, \quad \alpha_0 = \sum_k \alpha_k.$$

²Иногда название байесовский фактор закрепляется за самим отношением постериорных вероятностей (например, такое можно увидеть в книге O.Martin).

Способ №2: введением скрытой переменной, идентифицирующей модель. В приведенном ниже коде две модели отличаются лишь значениями параметров априорного бета-распределения (а в остальном идентичны). При этом вводится скрытая категориальная переменная `model_index` (со своим параметром `p`), которая в конечном итоге и будет отслеживать «частоту сэмплирования» каждой модели.

```
with pm.Model() as model_BF:
    p = np.array([0.5, 0.5])
    model_index = pm.Categorical('model_index', p=p)
    m0 = (4, 8); m1 = (8, 4)
    m = pm.math.switch(pm.math.eq(model_index, 0), m0, m1)
    # prior
    theta = pm.Beta('theta', m[0], m[1])
    # likelihood
    y = pm.Bernoulli('y', theta, observed=y_d)
    idata_BF = pm.sample(5000)
```

После сэмплирования можно не только построить постериорные трейсы (чтобы, например, убедиться как поменялось распределение параметра `p`), но и рассчитать вероятности $p(\mathcal{D}|\mathcal{M}_i)$, $i = 0, 1$, и BF_{01} :

```
az.plot_trace(idata_BF)
pM1 = idata_BF.posterior['model_index'].mean(); pM0 = 1 - pM1
BF = (pM0 / pM1) * (p[1] / p[0])
f"{BF.item():.1f}"
```

Способ №3: с помощью SMC (Sequential Monte Carlo).

```
with pm.Model() as model0:
    theta = pm.Beta('theta', 4, 8)
    y = pm.Bernoulli('y', theta, observed=y_d)
    idata0 = pm.sample_smc(2500)
```

```
with pm.Model() as model1:
    theta = pm.Beta('theta', 8, 4)
    y = pm.Bernoulli('y', theta, observed=y_d)
    idata1 = pm.sample_smc(2500)
```

```
BF_smc = np.exp(
    idata0.sample_stats["log_marginal_likelihood"].mean()
    - idata1.sample_stats["log_marginal_likelihood"].mean()
)
np.round(BF_smc).item()
```

Способ №4: используя Savage–Dickey ratio. Об этом расскажу в следующий раз!

3.2.4 Полезные ссылки

Краткое введение в эту тему, можно найти здесь: [Model comparison](#).

Очень интересный пример, который мы только начали разбирать на прошлом семинаре, но который я настоятельно рекомендую посмотреть всем (хотя бы по диагонали):

[Model building and expansion for golf putting](#). Здесь вы увидите, как на вполне реальном примере проводится итерационная процедура байесовского анализа данных: 1) создаем модель, 2) анализируем данные, 3) вносим изменения в модель и повторяем все заново. При этом модификации модели в этом примере (а речь идет о точности забивания патов в гольфе) максимально «физичны» и понятны.

Более тонкие вопросы обсуждаются в разделе 2.5 другой книги Мартина Освальда *Bayesian Modeling and Computation in Python* (в соавторстве с Ravin Kumar и Junpeng Lao): здесь можно найти подробный разбор роли параметра формы распределения Парето и интерпретации ситуаций, когда его значения велики, смысла `u_value` и аналогичного ему LOO-PIT (и сравнению графических функций Arviz `plot_bpv`, `plot_compare`, `plot_khat`, `plot_loo_pit`), и более детальный разбор усреднений по моделям. Кстати, в предыдущих разделах этой же главы разбираются в том числе и как работать при появлении `divergencies` в трейсах...

Но наиболее детально тема байесовского сравнения моделей разобрана в главах 6 и 7 книги A.Gelman, J.B.Carlin, et. al, *Bayesian Data Analysis* (третье издание этой книги мы будем обозначать просто BDA3). Несколько примеров из BDA3 разобрано в нашем документе «Уроки Машинного Обучения» (Overleaf-документ *ML Lessons*), в том числе и некоторые теоретические и практические вопросы сравнения моделей.

4 Гауссовы процессы

4.1 Калибровка измерительной системы

Раздел написан по материалам статьи D. Higdon, M. Kennedy, J.C. Cavendish, J.A. Cafo, and R.D. Ryne, Combining field data and computer simulations for calibration and prediction [Higdon2004]

Рассмотрим постановку задачи о калибровке *измерительной системы* в наиболее общем виде. Пусть целью физической системы $\zeta(x)$ является измерение наблюдаемой x . В общем случае x может быть многомерной переменной, но ее индекс $p = 1, \dots, P$ мы для простоты чаще всего будем опускать. Естественно, процесс измерения приводит к неточностям, $y = \zeta(x) + \epsilon$, здесь y – результат измерения, ϵ – ошибка измерения.

При постановке *калибровочного эксперимента* на вход измерительной системы ζ подается некоторое количество известных данных x_n (калибровочные источники, сигналы и т.п.):

$$y(x_n) = \zeta(x_n) + \epsilon_n, \quad n = 1, \dots, N$$

где $\epsilon_n \equiv \epsilon(x_n)$ – совокупность ошибок измерений. Обычно ошибки полагают независимо и одинаково распределенными (i.i.d.), а зачастую – еще и гауссовыми. Также для простоты будем считать измерения одномерными.

Проблема в том, что сама по себе измерительная система нам неизвестна, в том смысле, что у нас нет явного вида функции $\zeta(x)$. Поэтому мы хотим проанализировать результаты этого эксперимента, основываясь на некоторой его модели $\eta(x, t)$, параметризуемой параметром t (который в общем случае также многомерный – его индекс $l = 1, \dots, L$ мы тоже будем опускать). Очень часто такая модель является следствием физических законов и/или численной (компьютерной) моделью.

Мы предполагаем, что при некотором (неизвестном нам пока!) значении параметра $t = \theta$ эта модель хорошо описывает данные:

$$y(x_n) = \eta(x_n, \theta) + \delta(x_n) + \epsilon_n$$

причем «невязка» δ должна быть мала: $\delta(x_n) \ll \eta(x_n, \theta)$. Для удобства использования в дальнейшем введем обозначения для вектор-столбцов $x \equiv (x_1, \dots, x_N)^\top$, $y \equiv (y_1, \dots, y_N)^\top$.

Задача калибровки модели заключается в подборе наиболее подходящего значения управляющего ею параметра на основании калибровочных данных (x, y) . Причем при байесовском подходе речь идет о вычислении *постериорного распределения* на параметр модели θ , которое позволит получить не только точечные оценки θ , но и количественную меру их неопределенности (доверительные интервалы и т.п.), а также и оценить неопределенность предсказаний самой модели (т.е. $\eta(x_{\text{new}}, \theta)$ для будущих измерений).

Реализация *байесовского вывода* (ВИ) во многом зависит от того, каким конкретным образом задана модель $\eta(x, t)$. Можно предложить три варианта:

1. Задание в явном виде – в виде формулы.
2. «Быстрый симулятор», т.е. такой, который может очень быстро генерировать отклик $\eta(x, t)$ для любых значений x и t .
3. «Медленный симулятор»: в этом случае все, что мы можем себе позволить, это генерировать отклик для некоторой ограниченной выборки значений $\{x_m^*, t_m^*\}$, $m = 1, \dots, M$. (Вопрос о том, как подбирать такую выборку, довольно сложный и мы рассмотрим его отдельно.)

Первые два варианта фактически представляют одну и ту же ситуацию, когда мы можем сгенерить сколько угодно сэмплов из модели – будь то аналитически или численно. А вот в третьем варианте модель $\eta(x, t)$ следует рассматривать статистически (*стохастически*), т.е. расширить вероятностную модель, дополнив ее как новыми параметрами, так и новыми данными (в виде $\{\eta_m^* = \eta(x_m^*, t_m^*)\}$).

ВІ для первых двух вариантов в случае $\delta = 0$ при нормальном распределении ошибок имеет традиционный вид:

$$p(\theta|y) \propto \mathcal{L}(\theta)p(\theta), \quad \mathcal{L}(\theta) \equiv p(y|\theta) \propto \exp \left\{ -\frac{1}{2}(y - \eta(\theta))^\top \Sigma_y^{-1}(y - \eta(\theta)) \right\}$$

где $\eta(\theta) \equiv (\eta(x_1, \theta), \dots, \eta(x_N, \theta))^\top$, $\Sigma_y = \sigma_y^2 \mathbb{I}_N$, σ_y^2 – дисперсия ошибки измерения, а $p(\theta)$ – прайор на интересующий нас параметр θ .

Для линейной модели, $\eta(\theta) = a\theta + b$, и при гауссовом прайоре, $p(\theta) = \mathcal{N}(\theta_0, \sigma_0^2)$, процедура ВІ может быть выполнена аналитически. В более общем случае надо воспользоваться одним из методов МСМС-сэмплирования, при котором будет производиться многократный вызов процедуры расчета $\eta(x, t)$. *SSH: Хорошо бы пояснить как именно, например, на примере РyМС*

А что же делать, если доступ к вызовам $\eta(x, t)$ (η -генератору) ограничен, как это происходит при третьем варианте задания модели измерительной системы?

Оказывается, что и в такой ситуации есть решение! Для него, правда, понадобятся дополнительные данные, получаемые за счет *ограниченного* доступа к η -генератору. Будем считать, что предварительно были получены M значений $\eta(x_m^*, t_m^*)$, $m = 1, \dots, M$. В этом случае функцию $\eta(x, t)$ можно считать заданной стохастически и прайор на нее искать в семействе *гауссовых процессов* (GP).

GP может быть определен двумя функциями – функцией среднего и ковариационной функцией. Нас интересует GP над пространством пар (x, t) , поэтому речь идет о функциях $\mu(x, t)$ и $\text{Cov}((x, t), (x', t'))$. Напоминаем, что в общем случае x и t у нас имеют размерность L и P соответственно, поэтому $\eta(x, t)$ осуществляет отображение из \mathbb{R}^{L+P} в \mathbb{R} .

Функцию среднего GP зачастую полагают равной константе, $\mu(x, t) = \mu$, а вот для выбора ковариационной функции есть много различных параметрических семейств. Рассмотрим, например следующее:

$$\text{Cov}((x, t), (x', t')) = \frac{1}{\lambda} \exp \left\{ -\sum_{l=1}^L \beta_l |x^l - x'^l|^\alpha - \sum_{p=1}^P \beta_{L+p} |t^p - t'^p|^\alpha \right\}$$

где индексы переменных x, t выписаны явным образом. Параметр λ обратно пропорционален маргинальной дисперсии η , $L + P$ компонент вектора β контролируют силу зависимости вдоль каждого из направлений аргумента функции, а α ответственен за гладкость функции.

Примечание. В [Higdon2004] по поводу α говорится следующее: «Значение $\alpha = 2$ приводит к гладким, бесконечно-дифференцируемым представлениям η , тогда как меньшие значения этого параметра дают более грубые представления. Данные выборки $\eta(x_m^*, t_m^*)$ обычно мало-информативны в отношении α , поэтому мы фиксируем α , основываясь на априорной информации об η или на численном моделировании».

Таким образом, набор параметров модели $\{\theta, \mu, \lambda, \alpha, \beta\}$ имеет размерность $P + 1 + 1 + 1 + L + P = 2P + L + 3$. Таково и общее количество параметров в нашей постановке задачи при фиксации α – ведь к этому списку надо добавить еще и ошибку измерения σ_y .

Сгруппируем данные в вектор-столбец $z = \{z_i\} = (y^\top, \eta^{*\top})^\top$ размерности $(N + M) \times 1$, соответствующий расширенный набор входных данных

$$\tilde{x} = (x_1, \dots, x_N, x_1^*, \dots, x_M^*)^\top = \begin{pmatrix} x \\ x^* \end{pmatrix}, \quad \tilde{t} = (\theta, \dots, \theta, t_1^*, \dots, t_M^*)^\top$$

Тогда

$$\mathcal{L}(z|\theta, \mu, \lambda, \beta, \sigma_y) \propto |\Sigma_z|^{-1/2} \exp \left\{ -\frac{1}{2} (z - \mu \mathbf{1}_{N+M})^\top \Sigma_z^{-1} (z - \mu \mathbf{1}_{N+M}) \right\}$$

где символом $\mathbf{1}_{N+M}$ обозначен вектор-столбец с $N + M$ единицами, а

$$\Sigma_z = \Sigma + \begin{pmatrix} \Sigma_y & \mathbf{0}_{N \times M} \\ \mathbf{0}_{M \times N} & \mathbf{0}_{M \times M} \end{pmatrix}$$

и

$$\Sigma_{ij} = \text{Cov}((\tilde{x}_i, \tilde{t}_i), (\tilde{x}_j, \tilde{t}_j)) = \frac{1}{\lambda} \exp \left\{ -\sum_{l=1}^L \beta_l |\tilde{x}_i^l - \tilde{x}_j^l|^\alpha - \sum_{p=1}^P \beta_{L+p} |\tilde{t}_i^p - \tilde{t}_j^p|^\alpha \right\}$$

Сравним эти два варианта на простом примере с нелинейной функцией $\eta(x, t) = ???$ и гауссовым прайором $p(\theta) = \mathcal{N}(0, 0.25^2)$.

Для второго варианта сгенерим $M = 20$ значений $\eta(x^*, t^*)$ и для удобства стандартизируем входные данные (чтобы они лежали в $[0, 1]^{P+L}$) и вектор η (чтобы среднее значение равнялось 0, а дисперсия 1). Для выполнения ВІ понадобятся еще прайоры на все параметры, в качестве которых автор использовал:

$$p(\mu) \propto \exp\{-\mu^2/(2v)\}, \quad p(\lambda) \propto \lambda^{a-1} \exp\{-b\lambda\}, \quad p(\beta) \propto \prod_{k=1}^{P+L} \frac{\exp\{-\beta_k\}}{\sqrt{1 - \exp\{-\beta_k\}}}$$

причем ввиду стандартизации можно зафиксировать $\mu = 0$ (т.е. положить $v = 0$), а λ целесообразно «прижать» к 1, выбрав $a = b = 5$. *SSH: Что за странное распределение у β ?*

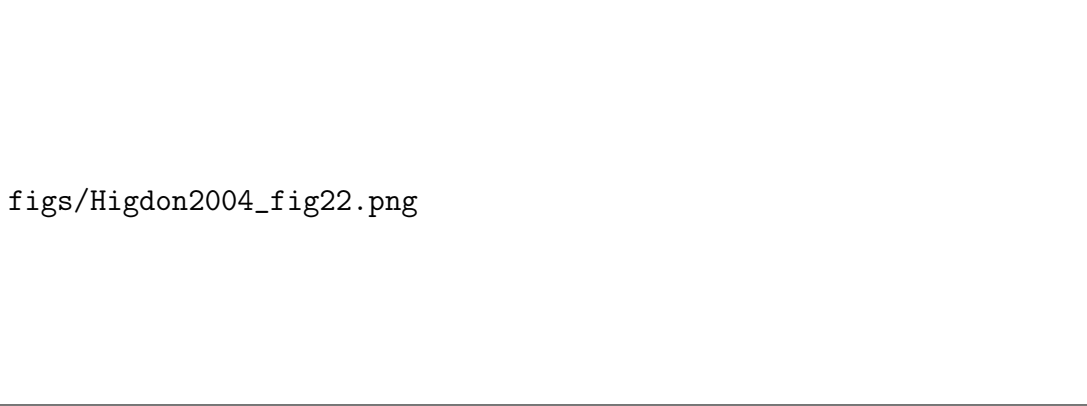
Результаты байесовского вывода для обоих вариантов при наличии всего трех точек данных изображены на рис. 1 и 2.

Теперь рассмотрим ситуацию с ненулевой невязкой δ .



figs/Higdon2004_fig21.png

Рис. 1: Априорные и постериорные неопределенности модельных предсказаний. Слева: прайор для θ и модельные функции $\eta(x, \theta)$, соответствующие его квантилям 5, 15, ..., 95%. Черными точками изображены 3 наблюдения (с 95% интервалами). Справа: постериор для θ и итоговые предсказания $\eta(x, \theta)$ для тех же квантилей.



figs/Higdon2004_fig22.png

Рис. 2: Априорные и постериорные неопределенности модельных предсказаний при медленном симуляторе. Слева: ограниченный запуск η -генератора привел к $M = 20$ значениям $\eta(x^*, t^*)$ (зеленые кружки). В центре: отдельная реализация $\eta(x, t)$, полученная по постериорному предсказательному распределению (изображены также актуальные сгенерированные точки). Справа: постериорное распределение на θ и квантили постериорных предсказательных распределений для $\eta(x, \theta)$.

5 Анализ данных Chandra: обзор работ D. van Dyk

Chandra – рентгеновский телескоп, запущенный на орбиту в 1999 году. Он получает данные в формате счета фотонов (*отсчетов*), распределенных по четырем измерениям: временному (*временной ряд*), энергетическому (*спектр*) и двумерному пространственному (*изображение*). Соответственно задачи анализа данных могут формулироваться как отдельно в каждом из этих трех блоков, так и совместно. Важно отметить, что для многих интересных источников (как точечных, так и диффузных) счет фотонов в отдельном 4D-бине может быть очень небольшим, а следовательно гауссовы аппроксимации пуассоновской статистики и следующие из нее способы анализа данных перестают работать. Именно по этой причине коллаборация Chandra в течение ряда лет развивала методы анализа данных, опирающиеся на байесовский подход. Ниже представлен обзор нескольких наиболее интересных для наших образовательных целей работ. Всех их можно найти в



Рис. 3: Прайоры и постериоры в ситуации с невязкой модели.

папке vanDyk, так как на протяжении многих лет именно David A. van Dyk из Department of Statistics, University of California курировал байесовские методы анализа Chandra.

vanDyk2004 [Highly Structured Models for Spectral Analysis in High-Energy Astrophysics](#)

vanDyk2006 [Deconvolution in High-Energy Astrophysics: Science, Instrumentation, and Methods](#)

Park2006 [Bayesian estimation of Hardness Ratios: modelling and computations](#)

Donath2024 [Joint Deconvolution of Astronomical Images in the Presence of Poisson Noise](#)

5.1 Low-Resolution Example: Hardness Ratios

Hardness Ratio (HR) – самое грубое описание спектра в виде отношения ожидаемого числа отсчетов в нижней энергетической части спектра (S, soft) к ожидаемым отсчетам в более высокой части (H, hard):

$$\mathcal{R} \equiv \mu_S / \mu_H$$

Так как при измерении источника вклад в отсчеты X вносит в том числе и фон, то, предполагая и сигнальные фотоны, и фоновые распределенными по Пуассону (\mathcal{P}), получаем (сумма пуассоновских величин является пуассоновской):

$$X_S \sim \mathcal{P}(\mu_S + \xi_S), \quad X_H \sim \mathcal{P}(\mu_H + \xi_H)$$

где ξ_S, ξ_H – ожидаемое число фоновых отсчетов в соответствующей части спектра. Для их оценки можно взять число отсчетов в кольцевой области вокруг изучаемого источника, B_S, B_H , тогда

$$B_S \sim \mathcal{P}(c\xi_S), \quad B_H \sim \mathcal{P}(c\xi_H),$$

c – известная константа, учитывающая различия в площади и времени экспозиции источника и фонового наблюдения.

Байесовский подход в данном случае заключается в вычислении постериорного распределения $p(\mathcal{R}|X_S, X_H, B_S, B_H)$ (напомним, $\mathcal{R} = \mu_S/\mu_H$).

Функция правдоподобия $p(X_S, B_S, X_H, B_H|\mu_S, \mu_H, \xi_S, \xi_H)$ естественным образом факторизуется по S и H компонентам спектра, поэтому если и прайоры выбрать факторизованными, то

$$p(\mu_S, \mu_H|X_S, B_S, X_H, B_H) = p(\mu_S|X_S, B_S) p(\mu_H|X_H, B_H)$$

Рассчитаем постериорные распределения (опуская индекс спектра)

$$p(\mu|X, B) = \int d\xi p(\mu, \xi|X, B) \propto \int d\xi p(X, B|\mu, \xi) p(\mu, \xi) = \int d\xi p(X|\mu, \xi) p(B|\xi) p(\mu, \xi)$$

Таким образом,

$$p(\mu|X, B) \propto \int d\xi \exp\{-(\mu + \xi)\} \frac{(\mu + \xi)^X}{X!} \exp\{-c\xi\} \frac{(c\xi)^B}{B!} p(\mu, \xi)$$

Расчет можно продолжить аналитически, если выбрать сопряженный прайор – в данном случае речь идет о гамма-распределении, $p(\theta) = \mathcal{G}(\alpha, \beta) \equiv 1/\Gamma(\alpha)\beta^\alpha\theta^{\alpha-1} \exp(-\beta\theta)$.

Постериорное распределение на интересующий нас параметр...

$$p(\mathcal{R}|X_S, X_H, B_S, B_H) = \int d\mu_H \mu_H p(\mathcal{R}\mu_H|X_S, B_S) p(\mu_H|X_H, B_H)$$

При низкой интенсивности источника постериорное распределение \mathcal{R} имеет тенденцию к перекосу из-за пуассоновского характера данных, к тому же \mathcal{R} принимает только положительные значения. Поэтому часто анализ данных проводят с помощью *показателя цвета*, представляющего логарифмическое преобразование \mathcal{R} , $C = \lg \mathcal{R}$, которое делает перекошенное распределение более симметричным.

В работе Park2006 сравниваются результаты оценки \mathcal{R} и C байесовским и классическими методами, см. рис. 4. При этом данные моделируются, исходя из двух сценариев: 1) $\mu_S = \mu_H = 30$, $\xi_S = \xi_H = 0.1$ – высокоинтенсивный источник, 2) $\mu_S = \mu_H = 3$, $\xi_S = \xi_H = 0.1$ – низкоинтенсивный источник. В обоих случаях $c = 100$, т.е. ожидаемое значение фоновых отсчетов всегда равно 10. Для измерений низкоинтенсивного источника постериорное и «классическое» гауссово распределения радикально отличаются друг от друга, причем последнее проявляет ряд нежелательных свойств (широкое и уходящее в нефизические области).

В работе также проверяется чувствительность результатов от выбора неинформативного прайора, рис. 5. Кроме равномерных прайоров на параметры пуассоновских распределений $(\mu_S, \mu_H, \xi_S, \xi_H)$, используются также прайор Джеффри и масштабно инвариантный прайор типа $p(\theta) \propto 1/\theta$. Все они являются частными случаями гамма-распределения (что, в частности, позволяет представить фигурирующие в теореме Байеса интегралы в виде сумм аналитических функций): $p(\theta) = \mathcal{G}(1, 0)$, $\mathcal{G}(1/2, 0)$, $\mathcal{G}(0^+, 0)$.

При больших интенсивностях источника (колонка справа) постериорные распределения стремятся быть симметричными, и влияние прайоров минимально. При малых интенсивностях (колонка слева) несимметричная форма апостериорного распределения \mathcal{R} и C становится очевидной, как и влияние выбора неинформативного прайора.

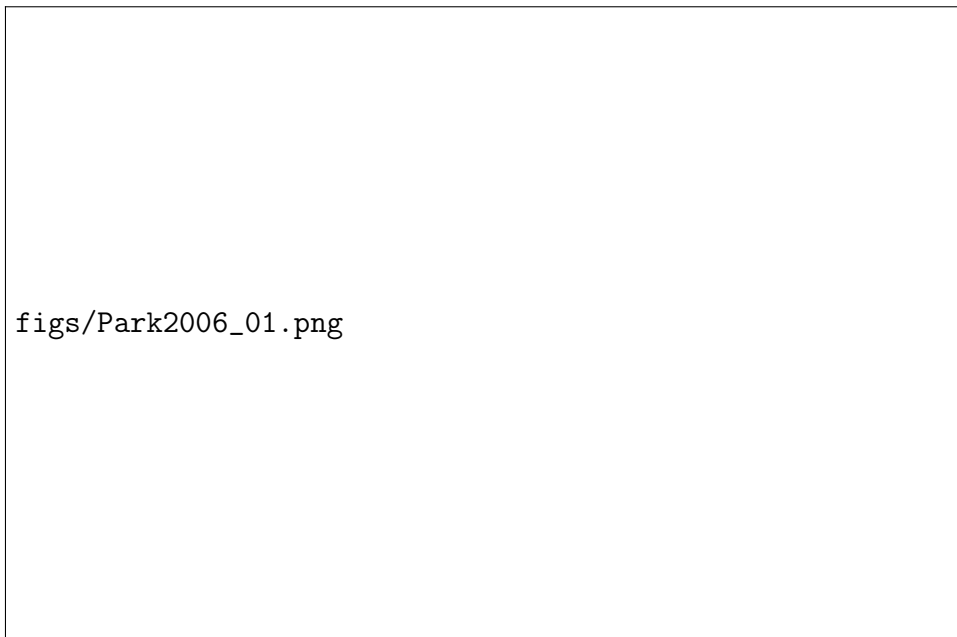


Рис. 4: Сравнение формы апостериорного распределения \mathcal{R} и \mathcal{C} с классическим гауссовым приближением для большого (слева) и для малого числа отсчетов (справа). Сплошные линии – апостериорные распределения (при равномерном прайоре), пунктирные линии – гауссово распределение со средним и стандартным отклонением, равными классическим оценкам.

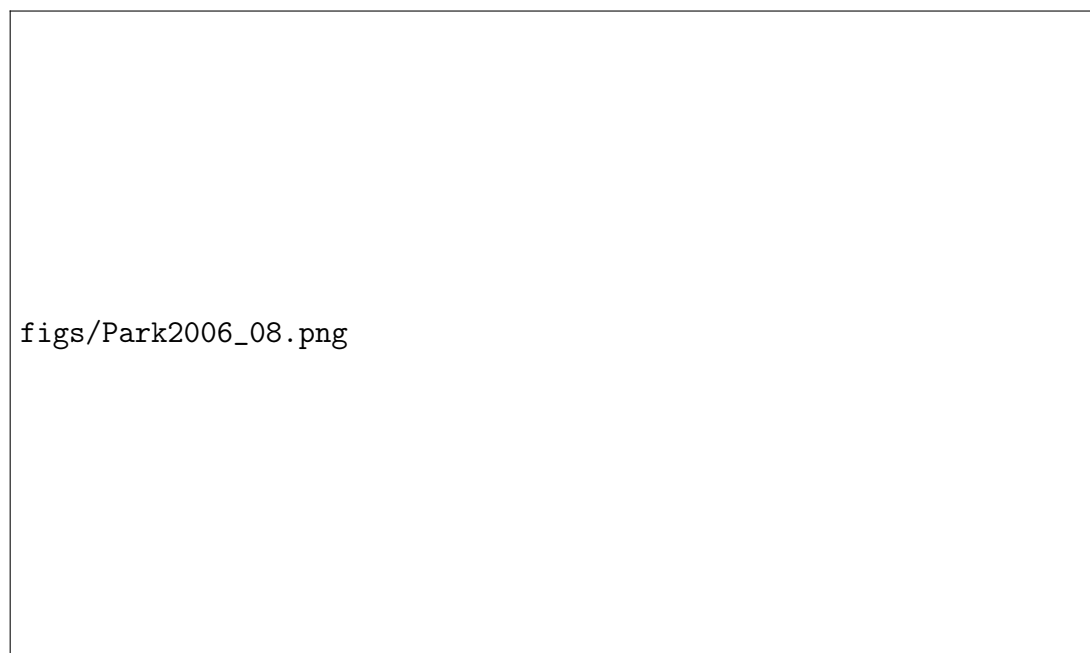


Рис. 5: Постериорные распределения \mathcal{R} и \mathcal{C} для источников разной интенсивности при различном выборе неинформативных прайоров: равномерного (красная линия), прайора Джеффри (зеленый штрих) и масштабно-инвариантного (пунктир).

5.1.1 Поиск рентгеновских двойных и эволюция вспышки Ross 154

Данный подход в работе Park2006 применен для решения двух астрофизических задач: поиску спокойных рентгеновских двойных звезд малой массы и определению эволюции вспышки на звезде Ross 154.

Маломассивные рентгеновские двойные (Low-mass X-ray binaries, LMXBs) – системы, состоящие из нейтронной звезды или черной дыры и донора с низкой массой (меньше 1-2 масс Солнца). Эти двойные системы, как правило, являются транзиентными источниками и большую часть времени проводят в состоянии покоя (quiescence, qLMXB's). Однако даже в неактивном состоянии они излучают рентгеновские лучи за счет теплового излучения с поверхности нейтронной звезды, а иногда и за счет дополнительного жесткого компонента, связанного с остаточной низкоуровневой аккрецией. Идентификация и изучение этих спокойных систем очень важны, поскольку они обеспечивают прямой обзор всего населения LMXB (и, следовательно, населения нейтронных звезд), а также ограничивают их рабочие циклы. Хотя однозначно классифицировать рентгеновский источник как qLMXB можно только по идентификации его оптических аналогов, спектры мягкого рентгеновского излучения или коэффициенты жесткости, согласующиеся с ожидаемыми для теплового излучения нейтронных звезд, очень полезны для идентификации кандидатов в qLMXB.

В работе байесовские методы применены на длительном (40 кс) наблюдении Chandra ACIS-S шарового скопления Terzan-5. Анализ данных фотометрических измерений в диапазонах 0.5-2.0 keV (S) и 2.0-6.0 keV (H) позволил расположить все 49 зарегистрированных источника на диаграмме Цвет-Блеск (использовался прайор Джеффри). Теоретически ожидается, что на этой диаграмме теплоизлучающие нейтронные звезды будут иметь рентгеновские цвета в диапазоне [0.5, 2.5]. Было показано, что байесовский метод, помимо более точных доверительных интервалов для цвета источников, позволяет оценить коэффициенты жесткости источников, которые не были обнаружены в одном из двух диапазонов. Были найдены 15 источников, чьи полосы ошибок перекрывают эту область и, таким образом, являются кандидатами в qLMXB. 8 из этих кандидатов не могли быть классифицированы классическим методом из-за слишком малой интенсивности сигнала. И наоборот, многие источники, лежащие на краю интересующей нас области, были бы ошибочно классифицированы как кандидаты в qLMXB с помощью классического метода, если бы использовались слишком большие полосы ошибок, которые он дает.

Второй физический пример был посвящен изучению временной эволюции рентгеновской вспышки (flare) на маломассивном карлике Ross 154. Звездные короны состоят из магнитно-ограниченной горячей плазмы ($T \sim 1 - 10$ МК) во внешних атмосферах звезд позднего типа и излучают оптически тонкое тепловое излучение, состоящее из континуума тормозного излучения и компонентов излучения атомных линий. Особое значение имеют вспышки, которые возникают в результате импульсного ввода энергии в корону при пересоединении сильно напряженных трубок магнитного потока и дают дополнительную информацию о динамике короны. Анализ спектров вспышек во время распада позволяет получить ограничения на физическое место и размер вспышечных петель, а также на процесс нагрева.

Полезная техника, разработанная специально для анализа гидродинамически эволюционирующих петель, заключается в отслеживании их эволюции в пространстве плотность–температура. Это делает возможным моделирование вспышек для получения физически значимых параметров, таких как длина петли, функция нагрева и т.д. К сожалению, всесторонний анализ этих параметров требует подгонки сложных модельных спектров к данным на каждом этапе эволюции вспышки, а поскольку для получения надежных результатов требуется большое число отсчетов (обычно > 500), разрешение, с которым можно


figs/Park2006_05.png

Рис. 6: Сверху: кривые блеска источника Ross 154. Наблюдаемый темп счета показан как функция времени с начала наблюдений, для счетов в мягкой (пунктирная гистограмма), средней (пунктирная гистограмма), жесткой (пунктирная гистограмма) и общей (сплошная гистограмма) частях спектра. Для удобства выбран размер бина в 150 с. Снизу: Эволюция показателя цвета \mathcal{C} для бинов размером $\delta t = 150$ с (сплошная гистограмма) и $\delta t = 40$ с (пунктирная гистограмма), вместе с соответствующими ошибками.

изучать эволюцию вспышки, сильно ограничено. Особенно это касается фазы нарастания вспышки, где наиболее интересные с физической точки зрения процессы имеют небольшую продолжительность, а количество наблюдаемых отсчетов также невелико. Однако коэффициенты жесткости могут быть построены при относительно более тонком временном разрешении и могут быть использованы в качестве косвенной подгонки параметров спектра.

Большая вспышка наблюдалась на Ross 154, близком (2.93 пк) активном М-карлике (M4.5Ve), в течение 57 ks наблюдений детектора Chandra/ACIS-S. Во время вспышки скорость счета увеличилась более чем в 100 раз. Кривая блеска источника во время вспышки показана на рис. 6 вверху, а также кривые блеска для трех полос пропускания: мягкой, 0.25-1 keV, средней, 1-2 keV и жесткой, 2-8 keV. На том же рисунке снизу показана временная динамика показателя цвета \mathcal{C} при двух разных длительностях временного бина.

Больше информации об эволюции вспышки можно получить, соотнеся между собой временную динамику цвета и интенсивности. На рис. 7 представлена траектория звезды на плоскости Цвет–Интенсивность. Для этого было произведено сэмплирование значений μ_S, μ_H из апостериорных распределений для каждого временного бина и для различных начальных временных фаз, с усреднением по размерам бинов от 40 до 400 с. На этой визуализации хорошо видно сохранение эволюционного трека вспышки Ross 154. Видна четкая картина эволюции вспыхивающей плазмы: сначала температура и интенсивность повышаются, в пике смягчаются, а затем температура и интенсивность затухают.



figs/Park2006_07.png

Рис. 7: Траектория звезды Ross 154 на плоскости Цвет–Интенсивность с усреднением по размерам бинов от 40 до 400 с. На изображение наложены точки, соответствующие каждому бину кривой блеска при размере бина 150 с (белая сплошная линия соединяет их во временной последовательности).

5.2 High Resolution Deconvolution Methods

В моделях высокого разрешения предполагается наличие большого числа параметров. При этом в качестве скрытых переменных фигурируют «идеальные бины» – значения отсчетов, которые были бы получены без учета расплывания и фона, а явные переменные (измеренные данные) – «детекторные бины» – реально полученные значения отсчетов.

Пусть J – количество идеальных бинов, L – количество детекторных бинов. Эти числа могут различаться из-за того, что спектральных каналов в реальном детекторе может быть меньше, чем то спектральное разбиение на бины, которое мы проводим в модели для скрытых переменных. (Здесь речь идет о спектральных каналах, но этот подход легко обобщается и на пространственные и временные бины.)

Переход от идеальных значений к детекторным осуществляется с помощью двух матриц. *Эффективная площадь* идеального бина j – это вероятность того, что фотон, пришедший в детектор, соответствующий идеальному бину j , будет зарегистрирован. Мы табулируем эффективную область как диагональную $J \times J$ матрицу \mathbf{A} , диагональные элементы которой равны этим вероятностям. Другими словами, $1 - \epsilon_j$ – вероятность того, что событие из идеального бина j не будет зарегистрировано детектором, где ϵ_j – диагональный элемент матрицы \mathbf{A} , эффективность регистрации события детектором.

Если регистрируется фотон, поступающий в детектор и соответствующий идеальному бину j , он может быть оказаться в одном из нескольких детекторных бинов. Это связано с приборным эффектом (размытие и т.п.). Мы табулируем приборный эффект матрицей \mathbf{P} размера $L \times J$, $\mathbf{P} = (p_{lj})$, где p_{lj} – вероятность того, что фотон из идеального бина j , зарегистрированный детектором, попадет в бин l . (Таким образом, столбцы \mathbf{P} – векторы вероятностей, сумма элементов вдоль каждого столбца матрицы \mathbf{P} должна быть равна 1, $\sum_l p_{lj} = 1$.)

Наиболее общий вид высокоструктурированной деконволюционной модели

$$\boldsymbol{\lambda} = \mathbf{P}\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\xi} \quad (1)$$

где аддитивный член $\boldsymbol{\xi} = (\xi_1, \dots, \xi_L)^\top$ представляет собой ожидаемое количество фоновых фотонов в каждой ячейке детектора.

В этой модели $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_L)^\top$ – вектор ожидаемых детекторных отсчетов, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_J)^\top$ – вектор ожидаемых идеальных отсчетов. Обращаем внимание, что в обоих случаях речь идет об *ожидаемых* значениях, т.е. это параметры соответствующего распределения Пуассона: $X_l \sim \mathcal{P}(\lambda_l)$ – модель измерений. Иногда к ней стоит добавить модель фоновых измерений, $B_l \sim \mathcal{P}(c\xi_l)$. Наша цель – реконструировать (или деконволюцировать) ожидаемые идеальные значения $\boldsymbol{\mu}$ с помощью измеренных данных \mathbf{X}_{obs} и \mathbf{V}_{obs} .
SSh: А кто объяснит, почему всегда речь идет о пуассоновской статистике: какой минимальный набор предположений о характере данных (и процессе измерения) при этом подразумевается?

5.2.1 Задача 1: Spectral Analysis

В спектральной модели параметры $\boldsymbol{\mu}$ должны отражать (как минимум) три вида сигнала: непрерывную компоненту свечения (С), излучение в атомарных линиях (Е) и поглощение (А):

$$\mu_j = \left(\delta_j f(\theta^C, E_j) + \sum_{k=1}^K \theta_k^E p_{jk}^E \right) g(\theta^A, E_j)$$

где δ_j – ширина идеального энергетического бина j , $f(\theta^C, E_j)$ представляет непрерывную компоненту с параметром θ^C , θ_k^E – ожидаемое идеальное количество отсчетов за счет эмиссионной линии k (K – число линий), p_{jk}^E – доля эмиссионной линии k , попадающая в идеальный энергетический бин j . Здесь $g(\theta^A, E_j)$ – модель поглощения, параметризованная θ^A , т.е. $1 - g(\theta^A, E_j)$ – вероятность того, что фотон с энергией E_j поглощается, а следовательно, не наблюдается³. Мы также можем параметризовать p_{jk}^E , например, в терминах положения и ширины гауссового профиля линии.

Поскольку эта спектральная модель сильно структурирована, априорная информация может потребоваться лишь в незначительном объеме (особенно для относительно ярких источников). Однако для некоторых параметров, таких как расположение эмиссионных линий, априорная информация почти всегда полезна. К счастью, эта информация часто доступна с научной точки зрения. Например, наличие характерных спектральных линий, которые можно ожидать в конкретном спектре. Тем не менее, предсказуемость структуры в типичных спектрах является мощным ресурсом в этой задаче пуассоновской реконструкции.

5.2.2 Задача 2: Imaging

Реконструкция рентгеновских изображений является куда более сложной задачей, поскольку их структура гораздо менее предсказуема, чем структура спектров. Помимо гладких протяженных объектов (ЕХ, extended source), рентгеновский снимок может включать один или несколько ярких точечных источников (PS, point source).

³Обращаем внимание, что это не тоже самое, что фигурировавшая ранее при общей постановке вероятность пропуска $1 - \epsilon_j$. В данном случае речь идет о моделировании трансформации сигнала еще до попадания в детектор.

Чтобы учесть это, мы параметризуем ожидаемый идеальный счетчик как

$$\mu_j = \theta_j^{\text{ES}} + \sum_{k=1}^K \theta_k^{\text{PS}} p_{jk} \quad (2)$$

где θ_j^{ES} – ожидаемое идеальное количество отсчетов в пикселе j за счет протяженного источника, θ_k^{PS} – ожидаемое идеальное количество отсчетов от точечного источника k (K – их общее количество). Здесь p_{jk} – доля ожидаемого идеального количества отсчетов для точечного источника k , попадающих в идеальный бин j . *SSH: А где ее взять? Это ведь не PSF, верно? PSF, вроде как, - это о матрице \mathbf{P} ...*

Эта модель все еще не включает свойство гладкости, характерное для протяженного источника. Рассмотрим то, как этого можно добиться, на примере много-масштабной модели [Nowak2000]. В такой модели предлагается группировать пиксели в блоки 4×4 . Точнее, в каждом таком блоке из 16 пикселей будет 4 суб-блока – квадранта с 2×2 пикселями. Условия гладкости можно добиться, если установить определенную связь, посредством условных распределений. А именно, предлагается факторизовать априорную вероятность на идеальные отсчеты x_{ab} , $a, b = 1, 2, 3, 4$, следующим образом

$$p(\{x_{ab}\}|\theta^{\text{ES}}) = p(x|\theta^{\text{ES}}) p(x_1, \dots, x_4|x, \theta^{\text{ES}}) \prod_{a=1}^4 p(x_{a1}, x_{a2}, x_{a3}, x_{a4}|x_a, \theta^{\text{ES}})$$

где $\theta^{\text{ES}} = \{\theta_{ab}^{\text{ES}}\}$. Здесь были введены дополнительные скрытые переменные x_1, x_2, x_3, x_4 , характеризующие каждый квадрант, и x , ответственная за весь блок. Блоковая переменная предполагается пуассоновской

$$x \sim \mathcal{P}(\theta^{\text{ES}})$$

где $\theta^{\text{ES}} = \sum_{ab} \theta_{ab}^{\text{ES}}$, а условные вероятности для всех остальных скрытых переменным выбираются мультиномиальными

$$(x_1, x_2, x_3, x_4)|x \sim \text{Multinom}(\boldsymbol{\pi}_1), \quad (x_{a1}, x_{a2}, x_{a3}, x_{a4})|x_a \sim \text{Multinom}(\boldsymbol{\pi}_{2a})$$

Параметры мультиномиальных распределений – простые преобразования от θ^{ES} . Это стандартная пере-формулировка модели, позволяющая легко задать прайор, благоприятствующий гладким реконструкциям. Так, симметричный прайор Дирихле $\boldsymbol{\pi}_1 \sim \text{Dir}(\alpha_1, \alpha_1, \alpha_1, \alpha_1)$ уменьшает ожидаемые значения отсчетов в квадрантах в сторону равенства, т.е. гладкого изображения на данном уровне разрешения. Чем больше α_1 , тем более гладкой будет реконструкция при данном разрешении. Аналогично, мы задаем $\boldsymbol{\pi}_{2a} \sim \text{Dir}(\alpha_2, \alpha_2, \alpha_2, \alpha_2)$, используя один и тот же гиперпараметр для каждого мультинома на этом уровне разрешения. Однако на разных уровнях разрешения мы используем разные гиперпараметры, что позволяет добиться и разного уровня сглаживания.

Для больших изображений идея полностью аналогична: для изображения размером 256×256 существует восемь гиперпараметров (так как $2^8 = 256$). Для задания гиперпараметров $\alpha_1, \dots, \alpha_8$ можно ограничиться общим гипер-прайором типа гамма-распределения с предустановленными значениями параметра формы и масштаба.

Матрица PSF \mathbf{P} может значительно варьироваться по фокальной поверхности (ФП) фотоприемника. И даже в фиксированном месте ФП компоненты \mathbf{P} известны с погрешностью. Зачастую для генерации матрицы PSF используется моделирование, которое вносит как погрешность Монте-Карло, так и систематическую погрешность (например, поскольку моделирование требует указания неизвестного спектра изображения).

Другая проблема связана с компонентами структуры самого изображения. Астрофизики часто хотят исследовать такие структуры, как точечные источники, струи (jets) или петли горячего газа (loops of hot gas). Одной из возможных стратегий является обобщение модели (2) путем добавления дополнительных структур. Например, струя часто выглядит как цепочка скоплений, протянувшихся от источника, которую можно смоделировать с помощью набора двумерных гауссовых распределений с ограничением на их среднюю структуру.

Продемонстрируем этот подход на примере реконструкции рентгеновского изображения галактики NGC 6240. На рис. 8 слева-вверху представлено исходное X-ray изображение, полученное телескопом Chandra (в логарифмической шкале) размером 256×256 пикселей. Справа-вверху показано апостериорное среднее значение изображения μ при многомасштабном прайоре. Для многомасштабных гиперпараметров $\alpha_1, \dots, \alpha_8$ использовался гамма-прайор с параметром формы 2 и параметром масштаба 0.01. (Чтобы уменьшить чувствительность конечного результата к выбору координат в многомасштабном прайоре, на каждой итерации случайным образом выбирался пиксель, который рассматривался как начало координат в многомасштабной спецификации.) *SSh: Вот этой ремарки я не понимаю.* Показанное реконструированное изображение было сгенерировано с помощью одной МСМС-цепочки с общим количеством итераций 2000 (+1000 на warm-up, предварительно была проверена сходимость).

Полученная реконструкция явно гораздо более гладкая, чем «классические» реконструкции по методу Ричардсона-Люси⁴ (средний ряд), и не требует от пользователя принятия решения об остановке итерационного алгоритма для избежания переобучения. По сравнению с изображением, сглаженным с помощью гауссового ядра (внизу-справа), байесовская реконструкция сохранила гораздо больше структурных элементов. Обратите внимание, например, на петлю горячего газа: она также проявляется на оптическом изображении, полученном с помощью телескопа Хаббл (внизу-слева), но полностью сглажена гауссовым ядром. Авторы работы также отмечают, что при наложении реконструкции на снимок Хаббла изображения очень хорошо совпадают.

5.2.3 Задача 3: Reconstruction of a Differential Emission Measure

Покажем, как использовать (1) для исследования распределения температуры вещества в звездной короне; такое распределение получило в литературе название *дифференциальной меры излучения* (DEM, differential emission measure). Корона, внешний слой звездной атмосферы, содержит очень низкоплотный (около 10^9 частиц/см³) и очень горячий ($T > 10^6$ К) сильно ионизированный газ.

Сначала поясним о чем идет речь, на примере солнечной короны. На рисунке 9 показано изображение одного и того же участка Солнца в трех длинах волн. Хотя в видимом свете солнечные пятна выглядят как темные области на фоне яркой поверхности Солнца, в экстремальном ультрафиолете они ярко светятся. На рентгеновском снимке видны большие петли светящейся плазмы, возвышающиеся над группой солнечных пятен. Причина, по которой изображения выглядят так по-разному, заключается в том, что они фактически показывают разные слои атмосферы Солнца. Видимые фотоны происходят из фотосферы, самого нижнего и самого холодного слоя с температурой около 5000 градусов, изображение в экстремальном УФ показывает хромосферу/переходную область, которая находится над фотосферой и имеет более высокую температуру (10–100 тысяч градусов), а рентгеновские лучи происходят из солнечной короны, которая находится еще выше и

⁴То есть ML-оценка в модели пуассоновского изображения.

figs/vanDyk2006_fig08a.png

figs/vanDyk2006_fig08b.png

figs/vanDyk2006_fig05.png

Рис. 8: Реконструкция NGC 6240 по рентгеновскому снимку телескопа Chandra. Верхний ряд: слева – исходное изображение, справа – байесовская реконструкция в виде апостериорного среднего. Средний ряд: реконструкции Ричардсона-Люси, остановленные после 20 (слева) и 100 итераций (справа). Нижний ряд: слева – оптическое изображение NGC 6240, полученное телескопом Хаббл, справа – сглаженное гауссовым ядром рентгеновское изображение.

имеет еще более высокую температуру – не менее миллиона градусов.

Рентгеновский снимок иллюстрирует сложную структуру интенсивности рентгеновского излучения в солнечной короне. Структура излучения является индикатором температуры и плотности в короне. На рисунке 10 показано относительное количество вещества короны в зависимости от температуры (график солнечного DEM) в области с высокой активностью солнечных пятен (слева) и в области без активности (справа). Обратите внимание, что в спокойной области короны относительно меньше очень горячего вещества.

О температурной структуре корон других звезд известно очень мало. Звезды находятся на очень большом расстоянии, поэтому их диски невозможно разглядеть даже с помощью самых лучших существующих телескопов. Однако можно косвенно судить об этом, изучая температурную структуру. В эмиссионных линиях рентгеновских спектров звезд есть подсказки, которые можно раскрыть с помощью априорной информации, получен-



Рис. 9: Оптическое, экстремальное УФ и рентгеновское изображения Солнца. Рентгеновское изображение иллюстрирует сложный градиент интенсивности и соответствующий градиент температуры по всей солнечной короне. Звездный DEM представляет собой изображение распределения температуры вещества в звездной короне.



Рис. 10: Солнечный DEM в активной (слева) и неактивной (справа) области: относительное содержания вещества в области солнечной короны в зависимости от логарифма (по основанию 10) температуры вещества.

ной в результате детальных квантово-механических вычислений и наземных лабораторных измерений. Звездная корона состоит из различных ионов, которые можно распознать в спектре по их идентифицирующим эмиссионным линиям. Если корона относительно горячая, мы ожидаем, что эмиссионные линии более энергичных квантовых состояний будут относительно сильными. Таким образом, относительная сила эмиссионных линий несет информацию о температуре источника.

Раскрытие этой информации эквивалентно решению задачи реконструкции пуассоновского изображения. Предположим, что звездная корона состоит полностью из одного иона при определенной температуре. Квантово-механические расчеты позволяют точно предсказать, какие линии излучения возможны, и относительную вероятность появления излученного фотона в каждой из возможных линий – вектор вероятностей мультиномиального распределения. То же самое можно сделать, если известно относительное содержание различных ионов, просто создав смесь мультиномиальных распределений с весами, равными относительному содержанию ионов. Таким образом, для заданной температуры и для большого набора линий излучения мы имеем вероятность того, что излученный фотон попадает в каждую линию. Этот вектор вероятностей может быть вычислен для каждого значения температуры на сетке возможных температур короны (например, равномерно распределенных в lg-шкале). Соберем эти векторы в *матрицу излучательной способности* \mathbf{P} размера $J \times M$, где J – количество возможных линий излучения, а M – количество значений температуры на сетке.

Данные, собранные с помощью прибора HRC телескопа Chandra с установленной дифракционной решеткой LETGS, представляют собой независимые пуассоновские отсчеты по десятку тысяч линий⁵. Ожидаемое значение отсчета для каждой линии объединяем в вектор $\boldsymbol{\mu}$. Это ожидаемое идеальное значение отсчета, так как мы еще не учли инструментальные эффекты, эффективную площадь прибора, поглощение или фон. Мы хотим вычислить ожидаемое число отсчетов, соответствующее каждому из бинов температуры. Соберем ожидаемые число отсчетов в каждом бине температуры в вектор $\boldsymbol{\mu}_T$ – это и есть наше представление DEM.

ML-оценка $\boldsymbol{\mu}$ может быть вычислена путем решения уравнения $\boldsymbol{\mu} = \mathbf{P}\boldsymbol{\mu}_T$, которое является упрощенной версией модели (1). На практике вектор $\boldsymbol{\mu}$ имеет десятки тысяч компонент, в то время как $\boldsymbol{\mu}_T$ – менее 100. Можно ожидать, что это будет очень простая инверсия, поскольку точек данных гораздо больше, чем неизвестных параметров. Но информация о DEM в отсчетах линий весьма скудна: матрица \mathbf{P} может оказаться почти сингулярной, а ML-оценка вести себя очень плохо. Необходимы байесовские методы и информативная априорная информация о $\boldsymbol{\mu}_T$.

Поскольку о форме DEM известно очень мало, наложим простые сглаживающие прайоры и снова воспользуемся многомасштабным априорным распределением, но на этот раз в одном измерении (заменяв мультиномиальное распределение и распределение Дирихле биномиальным и бета соответственно).

Чтобы проиллюстрировать сложность реконструкции DEM в связи с сингулярностью \mathbf{P} , рассмотрим модельный пример, в котором наряду с хорошим приближением матрицы \mathbf{P} используются симулированные данные из известного DEM, $\boldsymbol{\mu}_T$. Для моделирования выберем 8905 линий излучения, взяв подмножество в диапазоне от 5 до 25 Å (содержатся в 1600 бинах равной ширины 0.0125 Å в единицах длины волны), количество бинов температуры – 64 ($= 2^6$).

На рисунке 11 показаны результаты четырех симуляций, соответствующих четырем выбранным DEM (значения $\boldsymbol{\mu}_T$, использованные для генерации каждого из наборов данных, представлены пунктиром). MAP-оценка DEM (вычислялась по EM-алгоритму) иллюстрируется с использованием каждого из трех многомасштабных прайоров для каждого из четырех наборов данных. Априорные распределения задаются в терминах вероятностей разделения на каждом уровне масштаба – одно и то же сопряженное бета-распределение $\mathcal{B}(\alpha, \alpha)$ на каждом уровне. Три выбора прайора соответствуют: (1) плоскому распределе-

⁵Каждый отсчет на самом деле представляет собой смесь линии и континуума, но на данный момент мы игнорируем последний.

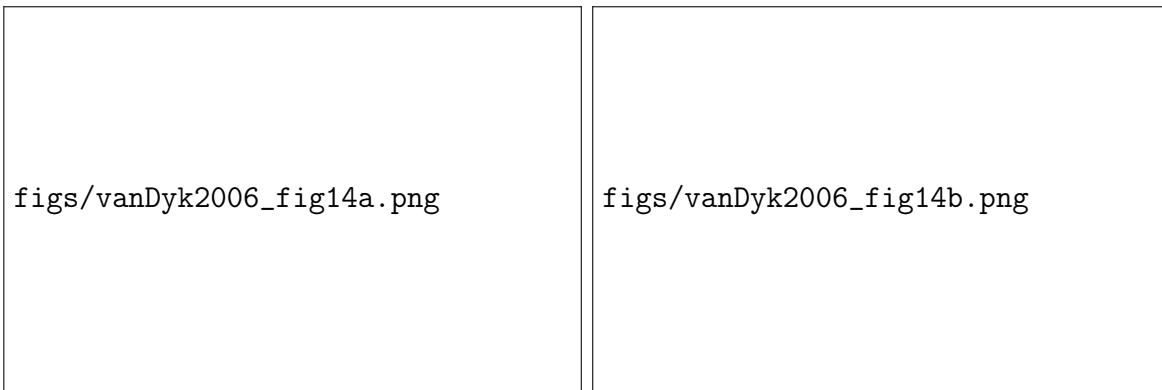


Рис. 11: MAP-оценки DEM из имитационного исследования (четыре DEM, использованные для генерации данных, изображены пунктиром). Вычисления проведены для трех разных выборов прайора. Вертикальные шкалы указаны в единицах 10^{17} .

нию с $\alpha = 1$ (No Smoothing), (2) умеренно сжимающему с $\alpha = 2$ (Medium Smoothing) и (3) сильно сжимающему с $\alpha = 3$ прайору (High Smoothing). Результаты показывают, что более сильные прайоры значительно помогают в реконструкции гладкой базовой DEM. В этих симуляциях данные таковы, что мы ожидаем надежных результатов только в диапазоне $5.5 < \lg(T) < 7.5$, поэтому при оценке реконструкций сосредоточимся именно на этом диапазоне. Еще раз подчеркнем: даже с таким большим количеством энергетических бинов ML-оценки не удовлетворительны и лишь априорное предположение о гладкости DEM помогает достичь приемлемых результатов.

С помощью MCMC-сэмплирования можно вычислить апостериорное среднее для μ_T и оценки его апостериорной вариативности. На этот раз использовался более широкий диапазон линий излучения – мы выбрали 26 005 линий (взяв подмножество в диапазоне от 3 до 180 Å, что соответствует 14 160 бинам). Для учета более крупного набора данных применялся и более сильный прайор – сопряженное бета-распределение с $\alpha = 10$. На рисунке 12 показаны результаты для имитационных моделей DEM 3 и 4. Для каждого набора данных была запущена цепь Маркова на 300 000 итераций, сохраняя при этом каждый 100-й сэмпл (общий размер выборки 3000 сэмплов, из которых последние 2000 использовались для MC-интегрирования). На рисунке сравниваются μ_T с его апостериорным средним значением и 95% апостериорными интервалами. Опять же, в диапазоне $5.5 < \lg(T) < 7.5$ реконструкции очень хорошие.

Наконец, попытаемся оценить DEM яркой рентгеновской звезды Капелла (α Aur). Исходный спектр Капеллы, собранный с помощью HRT с использованием дифракционной решетки LETGS, представлен на рисунке 13 и демонстрирует множество эмиссионных линий. Выберем 7741 линию, взяв подмножество в диапазоне от 6 до 20 Å (1120 бин шириной 0.0125 Å).

В эту реконструкцию мы также включим непрерывную компоненту (континуум), которую будем считать известной с точностью до нормирующего множителя, и учтем процессы, описанные ранее при формулировке (1) (включая фон, который также считаем известным до нормализующей константы). MAP-оценку DEM вычисляем с помощью EM-алгоритма (снова начиная с плоского DEM и используя тот же критерий сходимости, что и в имитационном исследовании), в бета-гиперпрайоре $\alpha = 3$.

Модель снова была профитирована с помощью MCMC. На рисунке 14 постериорное распределение μ_T охарактеризовано его средним и 95% интервалами. Важным преимуществом данного байесовского подхода является принципиальный способ присвоения ошибок



Рис. 12: Полосы ошибок реконструированных DEM из имитационного исследования: пунктир – генерационная DEM, сплошная линия – апостериорное среднее при бета-прайоре с $\alpha = 10$, заштрихованная область представляет собой 95% апостериорный маргинальный интервал.

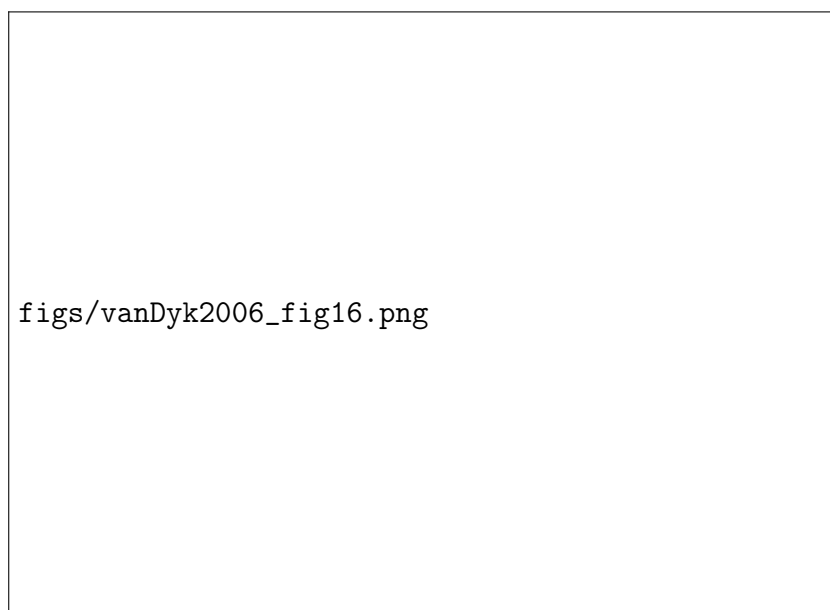




Рис. 13: Спектр Капеллы (α Aur), полученный с помощью HRC Chandra с дифракционной решеткой LETGS. На первом панели увеличен низкочастотный конец спектра. Обратите внимание на многочисленные линии эмиссии, составляющие спектр; несколько важных линий эмиссии помечены.

аппроксимирующей DEM. Ни одна из предыдущих попыток реконструкции звездной DEM не учитывала неопределенность в фитирующей DEM и атомных данных – фактически, это первый метод глобальной аппроксимации спектра, который использует всю доступную информацию в виде индивидуальных излучательных способностей линий. Это позволяет концептуально легко вносить улучшения, учитывая ошибки в атомных базах данных и включая полезную априорную информацию.



figs/vanDyk2006_fig17a.png



figs/vanDyk2006_fig17b.png

Рис. 14: Реконструкция DEM звезды Капелла, слева – в линейном масштабе, справа – в логарифмическом. Сплошная линия – апостериорное среднее DEM при бета-априорном распределении с $\alpha = 3$, заштрихованная область – 95% апостериорный интервал для DEM. (Пунктирная линия справа – реконструкция DEM Капеллы, выполненная [Dupree1993] в рамках другой модели и по другим данным.).

6 Астроинформатика: обзор работ Томаса Лоредо

Томас Лоредо (Thomas Loredo) – один из основоположников байесовского подхода в астрофизике и астрономии, он фактически является одним из лидеров направления, получившего название *астростатистика* и *астроинформатика*. В частности, именно Т.Лоредо инициировал...

Работы Loredo по астростатистике условно можно каталогизировать по четырем блокам: 1) Баейсовский вывод: общие вопросы (BI). 2) Пуассоновские точечные процессы (PPP). 3) Поиск экзопланет (EXO). 4) Иерархические байесовские модели (HBM). В соответствии с этим ниже мы приводим список наиболее интересных для нас работ Т.Loredo (в хронологическом порядке, по первому автору и году публикации).

Gregory1992 A new method for the detection of a periodic signal of unknown shape and period (PPP).

Loredo1993 Bayesian Inference With Log-Fourier Arrival Time Models and Event Location Data (PPP).

Loredo2004 Accounting for Source Uncertainties in Analyses of Astronomical Survey Data (PPP).

Loredo2011a Rotating Stars and Revolving Planets: Bayesian Exploration of the Pulsating Sky (PPP+EXO).

Loredo2011b Bayesian Methods for Analysis and Adaptive Scheduling of Exoplanet Observations (EXO).

Loredo2012a Sines, steps and droplets: Semiparametric Bayesian modeling of arrival time series (PPP).

Loredo2012b Bayesian astrostatistics: a backward look to the future (BI).

Loredo2017 Improving Exoplanet Detection Power: Multivariate Gaussian Process Models for Stellar Activity (EXO).

Loredo2019 Multilevel and hierarchical Bayesian modeling of cosmic populations (HBM).

Loredo2021 GPU-Accelerated Hierarchical Bayesian Inference with Application to Modeling Cosmic Populations: CUDAHM (HBM).

Loredo2023 Splines 'n Lines: Rest-frame galaxy spectral energy distributions via Bayesian functional data analysis (HBM).

Некоторые ранние работы Томаса Лоредо и слайды его презентаций доступны по ссылке: <https://hosting.astro.cornell.edu/~loredo/bayes/tjl.html>.

6.1 Байесовский вывод: Больше чем теорема Байеса

Раздел написан по материалам статьи Thomas J. Loredo and Robert L. Wolpert, Bayesian inference: More than Bayes's theorem [Loredo2024].

Байесовский вывод получил свое название от *теоремы Байеса* (ВТ, Bayes's theorem), выражающей постериорные вероятности гипотез о процессе генерации данных как (нормированное) произведение априорных вероятностей (прайоров) и функции правдоподобия. Однако в байесовском выводе используется вся теория вероятностей, а не только ВТ. Многие гипотезы, представляющие научный интерес, являются *составными* (composite hypotheses), причем доказательная сила гипотезы зависит от информации о дополнительных факторах, таких как значения *вспомогательных параметров* (nuisance parameters, например, неопределенные показатели фона или калибровочные коэффициенты). Многие важные возможности байесовских методов обусловлены использованием *закона полной вероятности* (LTP, law of total probability), который предписывает вычислять вероятности для составных гипотез путем *маргинализации* по вспомогательным факторам.

Здесь мы разберем роль маргинализации в байесовском выводе в контексте следующих тем:

- *Правдоподобие и вероятность*: рассмотрим введенное Фишером понятия правдоподобия как дополнения к вероятности, а также использование ВТ для «переворачивания» условного распределения и создания из функции правдоподобия постериорного распределения.
- *Прайоры – не (просто) штрафы*: априорное распределение не просто смещает пик функции правдоподобия, оно преобразует правдоподобие в величину, которую можно осмысленно интегрировать. В пространствах даже скромных размерностей так называемое «проклятие размерности» (curse of dimensionality) подразумевает, что интегрированные вероятности могут неожиданным образом накапливаться вдали от пика правдоподобия.
- *Маргинализация и оптимизации по вспомогательным параметрам*: большинство задач анализа данных в астрономии опирается на модели, включающие как параметры, представляющие непосредственный научный интерес, так и вспомогательные параметры (например, описывающие фон), которые необходимы для связи интересующих нас параметров с данными. Байесовские методы маргинализируют (усредняют) постериорные распределения по вспомогательным параметрам, чтобы учесть их неопределенность, в то время как в популярном альтернативном подходе по ним проводится оптимизация. Мы также обсуждаем так называемые *проблемы ошибки измерения* (термин из статистики), где различия между маргинализацией и оптимизацией могут усиливаться.
- *Маргинализация и систематическая ошибка*: кратко описываем использование маргинализации для определения и распространения систематической ошибки в условиях, когда стандартное «распространение ошибки» (статистический «дельта-метод») не работает.

6.1.1 Роль ЛТР в байесовском выводе

Параметрическая модель данных определяет, как можно предсказать или смоделировать гипотетические данные \mathcal{D} как функцию вектора параметров фиксированной размерности θ . То есть параметрическая модель – это *набор выборочных распределений* (collection of sampling distributions), функция одновременно гипотетических значений параметра и гипотетических значений данных: $p(\mathcal{D}|\theta) = f(\mathcal{D}; \theta)$. Каждое выборочное распределение – это распределение вероятностей *для данных*. Функция правдоподобия *фиксирует* $\mathcal{D} = \mathcal{D}_{\text{obs}}$ и тем самым является функцией только θ : $\mathcal{L}(\theta) \equiv p(\mathcal{D}_{\text{obs}}|\theta) = f(\mathcal{D}_{\text{obs}}; \theta)$.

Функция правдоподобия определяет, насколько хорошо каждое из выборочных распределений-кандидатов предсказывает наблюдаемые значения данных. Обратите внимание, что статистическая модель задает выборочные распределения для данных и функцию правдоподобия для параметров. Термин «правдоподобие данных» является некорректным – он полностью упускает смысл введения правдоподобия.

Постериорная плотность функции распределения⁶ (PDF, probability density function) тесно связана с функцией правдоподобия и получается простым умножением ее на априорную PDF (и нормировкой): $p(\theta|\mathcal{D}_{\text{obs}}) \propto p(\theta) \times \mathcal{L}(\theta)$. Постериорная PDF – это распределение вероятности по θ , имеющее осмысленные интегралы!

Кратко перечислим тот спектр задач, который позволяет решать интегральная мера.

- *Нормализация апостериорных PDF*: Апостериорная предсказательная вероятность (стоит в знаменателе ВТ) играет роль нормализующей константы, поскольку не зависит от θ .

$$p(\mathcal{D}_{\text{obs}}) = \int d\theta p(\theta) p(\mathcal{D}_{\text{obs}}|\theta)$$

Ее также часто называют *маргинальным правдоподобием* (marginal likelihood) и *обоснованностью* (evidence).

- *Доверительные области* (credible regions): Аналогичным образом можно вычислить вероятность принадлежности θ доверительной области R :

$$p(\theta \in R|\mathcal{D}_{\text{obs}}) = \int_R d\theta p(\theta \in R|\theta) p(\theta|\mathcal{D}_{\text{obs}})$$

Может существовать множество областей, имеющих желаемое значение вероятности. Наименьшая такая область обычно уникальна, ее плотность в R выше, чем за ее пределами, она получила название доверительной области наибольшей апостериорной плотности (HPD, highest posterior density).

- *Маргинализация по вспомогательным параметрам*: В большинстве реальных задач анализа данных есть параметры, необходимые для моделирования данных, но не представляющие непосредственного интереса. Пусть ψ обозначает интересные (сигнальные) параметры, а η – вспомогательные (nuisance) параметры. Неопределенность в ψ оценивается *маргинальным апостериорным PDF*,

$$p(\psi|\mathcal{D}_{\text{obs}}) = \int d\eta p(\psi, \eta|\mathcal{D}_{\text{obs}})$$

Если постериор исследуется с помощью сэмплирования (например, посредством MCMC), этот интеграл можно аппроксимировать, просто построив гистограмму ψ -компонент апостериорных сэмплов (ψ, η) .

⁶В случае дискретной случайной величины (например, пуассоновской) мы будем говорить о *массовой функции распределения* - PMF (probability mass function).

- *Распространение неопределенности* (propagation of uncertainty): Предположим, мы вычислили апостериорную PDF для параметров модели θ , но нас интересует не непосредственно θ , а величина $f = F(\theta)$ для известной функции $F(\theta)$. Фреквентистские методы «распространения ошибок» опираются на приближения, предполагающие, что постериор является гауссовым. При байесовском подходе мы можем распространять неопределенность более аккуратно и точно, используя ЛТР для вычисления маргинального постериора для f :

$$p(f|\mathcal{D}_{\text{obs}}) = \int d\theta p(f, \theta|\mathcal{D}_{\text{obs}}) = \int d\theta p(\theta|\mathcal{D}_{\text{obs}})p(f|\theta) = \int d\theta p(\theta|\mathcal{D}_{\text{obs}}) \delta(f - F(\theta))$$

При сэмплировании этот интеграл можно аппроксимировать, просто построив гистограмму значений $F(\theta)$, вычисленных по постериорным сэмплам θ .

- *Прогнозирование* (prediction): В контексте модели с параметрами θ для наблюдаемых данных \mathcal{D}_{obs} , когда мы хотим сделать предсказания относительно будущих данных \mathcal{D}' , которые мы можем получить, можно воспользоваться ЛТР для вычисления апостериорного предсказательного распределения:

$$p(\mathcal{D}'|\mathcal{D}_{\text{obs}}) = \int d\theta p(\mathcal{D}'|\theta) p(\theta|\mathcal{D}_{\text{obs}})$$

- *Сравнение моделей* (model comparison): Чтобы сравнить конкурирующие параметрические модели, задаваемые контекстами M_i (каждая со своими параметрами θ_i), мы определяем апостериорные шансы или факторы Байеса. Для этого необходимо вычислить маргинальное правдоподобие каждой модели,

$$p(\mathcal{D}_{\text{obs}}|M_i) = \int d\theta_i p(\theta_i|M_i) p(\mathcal{D}_{\text{obs}}|\theta_i, M_i)$$

Это говорит о том, что вероятность модели (в целом, то есть с учетом неопределенности ее параметров) – это среднее значение функции правдоподобия для параметров этой модели.

- *Усреднение по моделям* (model averaging): Снова рассмотрим ситуацию с конкурирующими параметрическими моделями с контекстами M_i , но теперь пусть все модели имеют некоторые общие интересующие нас параметры ϕ , дополненные наборами собственных вспомогательных параметров η_i . В условиях, когда ни одна конкретная модель не является более предпочтительной, чем ее конкуренты, при оценке ϕ мы бы хотели учесть неопределенность модели. Пусть \mathcal{C} обозначает общий контекст, в котором собраны все конкурирующие модели. Усреднение по моделям использует ЛТР для распространения неопределенности как модели, так и вспомогательных параметров:

$$p(\phi|\mathcal{D}_{\text{obs}}, \mathcal{C}) = \sum_i p(M_i|\mathcal{D}_{\text{obs}}, \mathcal{C}) p(\phi|\mathcal{D}_{\text{obs}}, M_i) \propto \sum_i p(\mathcal{D}_{\text{obs}}|M_i) \int d\eta_i p(\phi, \eta_i|\mathcal{D}_{\text{obs}}, M_i)$$

где в последнем переходе предположено, что модели априори одинаково вероятны, так что вероятность каждой пропорциональна ее маргинальному правдоподобию $p(\mathcal{D}_{\text{obs}}|M_i)$.

6.1.2 Особенности прайоров

Различие между вероятностью и правдоподобием указывает на двойственную роль априорных вероятностей в байесовском выводе. В байесовских учебниках прайоры обычно рассматриваются как регуляторы вероятности для учета базовых ставок, ограничений на параметры (положительная определенность и т.п.) или результатов предыдущих экспериментов. И это действительно важно. Но более фундаментальная роль прайора заключается в том, чтобы осуществить переход от правдоподобия к вероятности. Имея в руках распределение вероятностей, мы можем использовать ЛТР для решения множества вопросов, связанных с рассмотрением составных гипотез.

Прайоры в байесовском выводе не просто модулируют функцию правдоподобия, но, что более важно, преобразуют *интенсивное* правдоподобие в *экстенсивную* вероятность (аналогия с физическими понятиями «температуры» и «теплоты»). В байесовском выводе область пространства параметров с высоким правдоподобием может иметь малую вероятность, если ее объем мал или если априорная PDF приписывает этой области низкую вероятность. Подходы, основанные исключительно на правдоподобии, фокусируются на конкретных гипотезах, которые являются «горячими» (например, с максимальным правдоподобием), в то время как байесовский вывод фокусируется на наборах гипотез с наибольшей «теплотой» (т.е. с большой вероятностью).

Фреквентистская статистика включает в себя метод *штрафования максимального правдоподобия*, при котором функцию правдоподобия сначала умножают на штрафную функцию $r(\theta)$ (регуляризатор), а затем находят максимум этого произведения: $\hat{\theta} = \arg \max_{\theta} r(\theta)\mathcal{L}(\theta)$. Штрафная функция смещает положение максимума. Она похожа на прайор, так как умножается на функцию правдоподобия, но поскольку байесовские вычисления интегрируют, а не максимизируют по пространству параметров, прайор может сделать гораздо больше, чем просто сдвинуть положение моды (пика постериорного распределения) относительно максимума правдоподобия.

Ряд концепций из различных областей математики может помочь понять следствия наличия (суммируемой/интегрируемой) *меры*, а не просто упорядочивания предпочтений в пространстве параметров. Перечислим лишь некоторые из них: проклятие размерностей (curses of dimensionality) в многомерной геометрии, концентрации меры (concentration of measure) в теории меры, типичные множества (typical sets) в теории информации.

Для высокоразмерных пространств мода распределения, скорее всего, будет весьма нетипичной, в том смысле, что случайные выборки из распределения вряд ли будут похожи на его моду (не важно речь идет о распределении параметров модели или данных).

Априорные вероятности позволяют получить из функций правдоподобия нечто интегрируемое, что дает возможность проводить различные полезные вычисления, когда мы интересуемся составными гипотезами. Однако «экстенсивная» (суммируемая) природа априорных и постериорных распределений означает, что эффекты объема пространства параметров могут оказать неожиданное влияние на вывод. Если рассматривать прайоры только как модулирующие или штрафующие функцию правдоподобия, то кажется, что равномерный прайор не отвлечет фокус внимания от пика правдоподобия. Но понятие типичных множеств и поведение равномерных прайоров в многомерном выводе предупреждают нас о том, что *даже плоские прайоры могут сместить внимание от области максимального правдоподобия, если пространство параметров является высокоразмерным*. С одной стороны, это предупреждение о том, что оптимизация может ввести в заблуждение: она может сакцентировать внимание на тех частях пространства параметров, которые являются очень нетипичными. С другой стороны, это также предостережение против того, чтобы полагаться на интуицию, полученную в пространствах малого чис-

figs/Loredo2024_fig05.png

Рис. 15: 10 сэмплов векторов f_i из равномерного на симплексе прайора, построенных как гистограммы на единичном интервале. Слева сэмплы из распределения с $K = 5$ бинами, справа – для $K = 30$.

ла измерений (например, отождествляя равномерность прайора с неинформативностью). Практический урок заключается в том, чтобы *всегда рассматривать выборки из прайора*, особенно при работе в высокоразмерных пространствах параметров.

Рассмотрим оценку PDF для некоторой непрерывной наблюдаемой величины по данным гистограммы (скажем, потоков галактик, сгруппированных в несколько десятков бинов по потоку). Мы можем смоделировать гистограмму отсчетов с помощью *мультиномиального распределения* с вероятностями f_k для объектов, попадающих в каждый из K бинов ($\sum_k f_k = 1$). Тогда функция правдоподобия, связанная с набором наблюдаемых отсчетов n_k , имеет вид

$$\mathcal{L}(f_1, \dots, f_K) = \frac{N!}{\prod_k n_k!} \prod_k f_k^{n_k}$$

где $N = \sum_k n_k$ – общее количество отсчетов. Легко показать, что оценки максимального правдоподобия (MLE) равны $\hat{f}_k = n_k/N$, что интуитивно кажется привлекательным результатом.

Для байесовского анализа заманчивым выбором прайора, неинформативного по f_i , является плоский или равномерный прайор на симплексе $\sum_k f_k = 1$. Но что такое распределение говорит нам о наших априорных ожиданиях для $\{f_i\}$?

На рис. 15 мы пытаемся разобраться в этом вопросе, рассматривая случайные выборки из прайора. На левой панели показана выборка из 10 случайных векторов f_i из равномерного прайора для $K = 5$. Сэмплы демонстрируют достаточное разнообразие: некоторые выглядят почти равномерными, а некоторые имеют один или несколько бинов с большим f_i . На правой же панели показана другая выборка из 10 случайных векторов – теперь для $K = 30$. Возможно это покажется удивительным, но теперь разнообразие невелико – все сэмплы практически равномерные.

Простой анализ подтверждает визуальное впечатление. Можно вычислить маргинальную PDF для конкретного значения f_k (маргинализируя по всем остальным параметрам). Для $K = 2$ эта маргинальная PDF будет равномерной, но с ростом K она все больше концентрируется вблизи $f_k = 0$, образуя скошенное распределение с матожиданием $1/K$ и все уменьшающейся дисперсией. Случайные выборки и анализ говорят нам о том, что с ростом размерности большая часть объема симплекса находится в области, где все f_k почти равны друг другу и расположены вблизи центра симплекса. (Примечательно, что это отличается от поведения гиперкубов в «проклятии размерности», где объем накапливается на границе.)

Как мы можем это исправить? Рассмотрим семейство *симметричных прайоров Дирихле*

$$p(f_1, \dots, f_K) \propto \delta \left(1 - \sum_k f_k \right) \prod_k f_k^{\alpha-1}$$

(δ -функция накладывает нормирующее ограничение на параметры f_k). Обратите внимание, что зависимость от f_k представлена произведением степеней каждого f_k – та же форма зависимости, что и в функции правдоподобия. У этого прайора есть единственный (скалярный) «гиперпараметр» $\alpha > 0$, с которым мы можем экспериментировать, чтобы контролировать его поведение (он называется параметром концентрации). При $\alpha = 1$ симметричный Дирихле становится плоским (равномерным) прайором. При других значениях чистый эффект умножения прайора на функцию правдоподобия заключается в добавлении $\alpha - 1$ к каждому отсчету n_k ; грубо говоря, прайор добавляет $\alpha - 1$ «априорных отсчетов» к функции правдоподобия. (Прайор Дирихле является *сопряженным* для мультиномиального правдоподобия.)

Ну и как задать α ? Одна из привлекательных идей заключается в том, что бинирование может быть выполнено различными способами, и при выборе прайора следует стремиться к некоторой согласованности между различными вариантами бинирования. Рассмотрим, в частности, *согласованность агрегации*: разработаем правило выбора α таким образом, чтобы применение этого правила, скажем, для $K = 10$, а затем агрегирование соседних пар бинов для получения модели из 5 бинов давало тот же самый прайор, что и при применении правила для α непосредственно для $K = 5$. Нетрудно показать, что условию согласованности агрегирования удовлетворяет выбор $\alpha = C/K$ для произвольной константы C , в то время как любое постоянное значение, такое как $\alpha = 1$ (соответствующее равномерному прайору), использованное для рис. 15, не удовлетворяет.

На рис. 16 показаны выборки из прайора согласованной агрегации с $C = 2$: $\alpha = 2/K$. Для $K = 2$ бинов он становится равномерным, что, как мы видели, имеет приемлемое поведение для малых K . На рисунке показаны сэмплы из прайора при $K = 30$. В сравнении с правой панелью рис. 16 теперь сэмплы уже не все плоские: в большинстве случаев один или несколько бинов имеют значительно большие значения f_k , чем в других бинах.

Мы сосредоточились на прайорах, а как это отразится на постериорах? Для совместной оценки значений f_k всех K бинов влияние незначительно, если в каждом бине есть хотя бы несколько отсчетов; новый прайор эффективно вычитает *SSH: свернуть с оригиналом* значение $C/K - 1$ из каждого числа отсчетов n_k , что не сильно меняет ситуацию по сравнению с использованием равномерного прайора. Однако если нас интересует только один или несколько бинов, и поэтому мы будем делать маргинализацию по остальным, разница может оказаться значительной. Кроме того, если мы хотим провести сравнение моделей, сравнивая модели с разным количеством бинов посредством их маргинальных правдоподобий (т.е. маргинализируя по всем параметрам f_k), то эти два прайора могут привести к очень разным результатам.

6.1.3 Маргинализация по вспомогательным параметрам

Очень часто (практически повсеместно) для решения задачи моделирования данных требуется введение параметров, которые не представляют непосредственного научного интереса, но необходимы для описания того, как данные соотносятся с интересующими нас величинами. Оценка интересующих параметров модели должна каким-то образом учитывать неопределенность в «неинтересных» вспомогательных величинах.

figs/Loredo2024_fig06.png

Рис. 16: Как на предыдущем рисунке, но с сэмплами из симметричного распределения Дирихле с согласованностью агрегации для $K = 30$ бинов.

Рассмотрим типичную ситуацию «сигнал+фон» с интересующим нас параметром сигнала s и фоновым вспомогательным параметром b . Мы предполагаем, что имеющиеся данные предоставляют полезную информацию о b , поэтому совместная постериорная PDF $p(s, b | \mathcal{D}_{\text{obs}})$ демонстрирует некоторую зависимость между s и b . Например, изображение \mathcal{D}_{obs} может включать независимые данные о b из измерений вдали от предполагаемого источника. В такой ситуации данные вблизи источника обеспечивают зашумленное измерение $s + b$, а данные вне источника дают отдельное зашумленное измерение b . Тогда совместная функция правдоподобия для s и b обычно напоминает изображенную слева на рис. 17, демонстрируя отрицательную корреляцию – чем больше b , тем меньше s .

Спецификация значения одного параметра в многопараметрической задаче является *составной гипотезой*: указание только s соответствует утверждению, что для данного s выполняется некоторый выбор параметров из множества $\{(s, b) : b \in [b_l, b_u]\}$, хотя мы не знаем какой именно⁷.

Следствие для s с учетом неопределенности b формулируется посредством маргинальной постериорной PDF

$$p(s | \mathcal{D}_{\text{obs}}) = \int db p(s, b | \mathcal{D}_{\text{obs}}) \propto \int db p(s, b) \mathcal{L}(s, b)$$

Совместный прайор может быть представлен как произведение прайора на s , $p(s)$, и условного прайора на b , $p(b | s)$. Во многих случаях может не быть контекстной информации, связывающей b и s априори, что делает последний фактор независимым от s ; для упрощения мы также предположим такую независимость.

Поскольку маргинализации проводится по b , прайор на s можно исключить из интеграла:

$$p(s | \mathcal{D}_{\text{obs}}) \propto p(s) \int db p(b | s) \mathcal{L}(s, b) = p(s) \mathcal{L}_m(s), \quad \mathcal{L}_m \equiv \int db p(b | s) \mathcal{L}(s, b),$$

⁷ $[b_l, b_u]$ обозначает диапазон допустимых значений b , например, $[0, \infty]$.

figs/Loredo2024_fig07.png

Рис. 17: Слева: пример совместной функции правдоподобия в условиях «сигнал+фон». Серый горизонтальный пунктир показывает \hat{b} – глобальную MLE для b . Синяя линия демонстрирует \hat{b}_s – MLE для b при заданном s ; она пересекает контуры функции совместного правдоподобия там, где они становятся вертикальными. Зеленый вертикальный пунктир указывает на срез совместной функции правдоподобия при $s = s_1$. Справа: ингредиенты аппроксимации маргинальной функции правдоподобия для s , зеленая кривая – срез совместной функции правдоподобия при $s = s_1$ как функция b , синий пунктир – (условный) прайор для b .

где была введена *маргинальная функция правдоподобия* \mathcal{L}_m (marginal likelihood function). Поскольку интегрирование, лежащее в основе маргинализации, изолировано в \mathcal{L}_m , в дальнейшем мы будем ориентироваться на нее. Чтобы понять, что достигается маргинализацией, аппроксимируем интеграл по b в два этапа (компоненты этой аппроксимации показаны на рис. 17 справа):

1. Предположим, что данные \mathcal{D}_{obs} информативны, поэтому прайор для b почти постоянен в области, где интеграл в определении \mathcal{L}_m имеет большую часть своей площади. Это позволяет извлечь фактор $p(b)$ из интеграла, установив его значение равным его значению в месте расположения пика среза функции правдоподобия.
2. Приблизим оставшийся интеграл $\mathcal{L}(s, b)$ по b (площадь под функцией правдоподобия в измерении вспомогательного параметра) произведением высоты правдоподобия и некоторой удобной меры ее ширины.

Для реализации этих шагов введем некоторые обозначения. Пусть \hat{b}_s – значение b , максимизирующее функцию правдоподобия при фиксированном s , т.е. вдоль вертикального среза слева на рис. 17 (изображен срез при $s = s_1$): $\hat{b}_s = \arg \max_b \mathcal{L}(s, b)$. Графически \hat{b}_s как функция от s отслеживает точки, в которых касательные к контурам $\mathcal{L}(s, b)$ становятся вертикальными. Относительная неинформативность прайора на шаге 1 подразумевает

$$\int db p(b|s) \mathcal{L}(s, b) \approx p(\hat{b}_s|s) \int db \mathcal{L}(s, b)$$

Интеграл по b дает площадь под кривой функции правдоподобия. Пусть δb_s обозначает некоторую меру ширины среза функции правдоподобия. Приблизим интеграл произведением высоты и ширины среза правдоподобия, $\int db \mathcal{L}(s, b) \approx \mathcal{L}(s, \hat{b}_s) \delta b_s$. Первый фактор – совместная функция правдоподобия, максимизированная по b для каждого s . Это функция исключительно s ; она называется *профильной функцией правдоподобия* (profile likelihood function) для s , $\mathcal{L}_p \equiv \mathcal{L}(s, \hat{b}_s)$.

Собрав эти результаты, мы получаем следующее полезное приближение для маргинальной функции правдоподобия (для простоты предположим априорную независимость параметров):

$$\mathcal{L}_m(s) \approx p(\hat{b}_s) \mathcal{L}_p(s) \delta b_s$$


Два последних фактора зависят от данных. Большая часть работы, проделанной маргинализацией, отражена в факторе профильного правдоподобия. Наивный подход к работе со вспомогательным параметром заключается в том, чтобы зафиксировать его на наилучшем значении, – это соответствует горизонтальному срезу $\mathcal{L}(s, b)$ вдоль серой пунктирной линии (через точку глобальной MLE (\hat{s}, \hat{b})). Вместо этого функция профильного правдоподобия учитывает тот факт, что выбор s подразумевает другой «лучший» выбор для b , чем глобальное MLE. Но маргинальная функция правдоподобия включает дополнительный фактор, δb_s , который учитывает, как *неопределенность* во вспомогательном параметре может меняться в зависимости от интересующего параметра. По сути, маргинализация говорит о том, что нужно начать с профилирования (а не просто использовать наилучший подходящий вспомогательный параметр), а затем скорректировать вероятность профиля, взвешивая области, где неопределенность вспомогательного параметра велика, гораздо выше, чем области, где она мала.

На рис. 18 показаны два сценария, в которых маргинальная и профильная функции правдоподобия могут заметно отличаться друг от друга (здесь s и b обозначают обобщенные интересующий и вспомогательный параметры соответственно). На верхней левой панели показана «вспыхивающая» («flaring») функция правдоподобия для двух переменных, когда неопределенность b значительно меняется в зависимости от s . На нижней левой панели показано, как вспыхивание смещает маргинальную функцию правдоподобия в сторону от совместного пика, отражая, что большая неопределенность b справа делает более вероятным, что истинные значения параметров находятся правее точки MLE. На верхней правой панели показана «бананообразная» («banana-shaped») функция правдоподобия двух переменных. Они часто возникают в ситуациях, когда предсказания зависят от параметров по степенному закону (таким образом, предсказания изменяются слабее вдоль кривой, чем ортогонально ей). Здесь увеличенный вес δb_s профильной функции правдоподобия возникает потому, что по мере того, как контуры правдоподобия наклоняются вверх с увеличением s , неопределенность в b возрастает.

Часто бывает, что поправка на объем вспомогательного параметра при одном единственном измерении заметна, но не очень существенна, она может сдвинуть профильную функцию правдоподобия лишь на малую долю ее ширины. Но для аналогичных измерений эффект, как правило, имеет сходное направление. При объединении информации по многим измерениям систематический эффект от фактора объема вспомогательного параметра может накапливаться, и его игнорирование может серьезно испортить агрегированные выводы, такие как демографические.

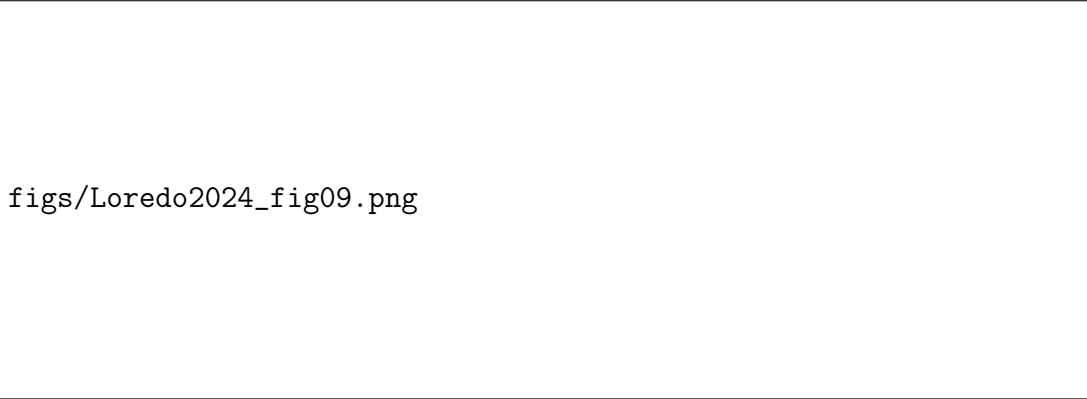
Рис. 19 иллюстрирует это на несколько надуманном, но наглядном примере. Для каждого из множества N объектов мы проводим пару измерений некоторого свойства объекта, μ_i для объекта i , с помощью прибора, который обеспечивает измерения с аддитивным гауссовским шумом с нулевым средним и фиксированным, но неизвестным стандартным отклонением σ . Обозначим пару измерений x_i и y_i . Для объекта i существует функция правдоподобия для его μ_i и общего параметра σ , которая является произведением двух нормальных распределений,

$$l_i(\mu_i, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right)$$



figs/Loredo2024_fig08.png

Рис. 18: Геометрии двухпараметрических функций правдоподобия, когда маргинальная и профильное правдоподобие могут приводить к существенно различным выводам. На верхних панелях изображены совместные функции правдоподобия для интересующего нас параметра s и вспомогательного параметра b . Синие кривые показывают \hat{b}_s , MLE для b при заданном s , путь, определяющий профильную функцию правдоподобия. На нижних панелях приведены профильная (зеленая) и маргинальная (синяя) функции правдоподобия для s . Слева: функция правдоподобия со «вспыхивающими» контурами, когда неопределенность вспомогательного параметра значительно меняется как функция интересующего параметра. Справа: функция правдоподобия с «бананообразными» контурами, когда неопределенность вспомогательного параметра меняется из-за кривизны пути профиля.



figs/Loredo2024_fig09.png

Рис. 19: Совместная, маргинальная и профильная функции правдоподобия для задачи о нормальных парах (Неймана-Скотта), пояснения в тексте.

Из одной пары измерений можно получить значительную неопределенность как в μ_i , так и в σ . На левой панели рисунка показаны контуры $l_i(\mu_i, \sigma)$ для пары измерений (изображенных точками вблизи оси ординат), контуры ограничивают приблизительные доверительные области 25%, 50%, 75%, 90%, 95%, 99% и 99.9%. Примечательно, что это пример вспыхивающей функции правдоподобия, которая, как мы видим, не является редким явлением: она актуальна даже для привычного нормального распределения, когда существует неопределенность не только в его среднем, но и в стандартном отклонении.

Можно улучшить вывод, объединив информацию по многим парам измерений. Каждая

пара привносит новую информацию о σ , позволяя нам откалибровать уровень шума прибора. Вся информации о σ в совокупности должна позволить нам более точно измерить и каждый μ_i , как если бы параметр σ стал известен. На правой панели рисунка показаны компоненты для вывода σ из объединенных данных. Сплошные кривые показывают профильные (оранжевые) и маргинальные (синие) функции правдоподобия по одной паре измерений (для маргинализации использовались равномерные прайоры на μ_i). Аналогично показанному на рис. 18 пик маргинальной кривой смещен вправо от пика профильной. Более того, маргинальная кривая значительно шире. При этом оба правдоподобия существенно перекрываются. При измерении одной пары выводы, сделанные на основе этих двух кривых, будут отличаться, но, вероятно, не очень сильно.

Пунктирные кривые на правой панели рис. 19 показывают, что происходит, когда мы накапливаем информацию по нескольким парным измерениям (в данном случае взятым из широкого равномерного распределения, хотя описанные здесь результаты не зависят от этого выбора). Формально, для этого мы вычисляем функцию правдоподобия для всего набора данных, которая является произведением функций правдоподобия для пар, $\mathcal{L}(\vec{\mu}, \sigma) = \prod_{i=1}^N l_i(\mu_i, \sigma)$, где $\vec{\mu}$ обозначает набор параметров μ_i . Мы снова можем обобщить следствия для σ , профилируя или маргинализируя по μ_i – но теперь по всем N параметрам μ_i .

Пунктирные кривые показывают итоговые профильную и маргинальную функции правдоподобия. При $N = 50$ они становятся значительно более узкими. Но профильная функция правдоподобия сходится в стороне от истинного значения ($\sigma_{\text{truth}} = 1$). Можно показать, что MLE является несостоятельной оценкой для σ в этом случае – она сходится к неверному значению при $N \rightarrow \infty$ (это остается верным, даже если с объектом производится более двух измерений, при условии, что число измерений на объект ограничено).

Примечание. Несложно показать, что MLE решение асимптотически сходится к значению $\sigma/\sqrt{2}$. Конечно, заманчиво надеяться, что по мере сбора все большего количества данных асимптотика придет на помощь и гарантирует сходимость к истине, а все неопределенности эффективно «усреднятся». Но приведенный выше пример показывает, что это не так. Причина в том, что наличие неопределенностей источников (здесь μ_i) означает, что каждый источник приносит с собой новый параметр, который необходимо оценить, явно или неявно. Нейман и Скотт назвали эти параметры «случайными» (incidental), в отличие от «структурного» параметра σ , общего для всех измерений. Общее число параметров растет с увеличением количества данных, поэтому допустимый объем пространства параметров не уменьшается асимптотически до нуля, и именно эта особенность не гарантирует выполнение обычных асимптотик.

Чтобы показать, что такие проблемы не редкость в астрономии, на рис. 20 показаны функции правдоподобия из смоделированных данных для изолированных тусклых точечных источников, наблюдаемых с помощью CCD-камеры с критически дискретизированной гауссовой PSF (critically-sampled Gaussian PSF). Слева – для источника $\sim 6\sigma$, истинное положение которого находится в середине *SSH: или все-таки в центре?* пикселя. Контурные представлены в пространстве потока F и одной из направляющих координат x (положение на CCD-матрице; другая координата фиксирована в своем наилучшем значении). Контурные демонстрируют вспышку: они почти симметричны по x , но очень асимметричны по F . Это показывает значительную зависимость потока от направления: позиции, удаленные от наилучшего положения, в любом направлении, подразумевают более низкие оценки потока. Рост площади контуров с уменьшением потока означает, что *маргинальное правдоподобие для потока смещается вниз от совместного наилучшего потока*. Сдвиг происходит на уровне процентов, что гораздо меньше, чем в приведенной выше задаче парных измере-

figs/Loredo2024_fig10.png

Рис. 20: Правдоподобие как функция потока F и одной из координат положения x для тусклых точечных источников на астрономическом изображении. Перекрестие показывает истинные значения (x, F) ; ромбы – оценки максимального правдоподобия. Слева – 6σ -источник, справа – 3σ -источник.

ний. Но, как и в той задаче, сдвиг *систематический*, в одном направлении и одинакового масштаба для всех 6σ -источников; он влияет на точность (смещение), больше, чем на погрешности, и будет доминировать над статистическими ошибками при анализе, усредненном по нескольким сотням источников или более. На правой панели рисунка показаны контуры вероятности для подпорогового измерения 3σ -источника с четвертьпиксельным смещением. Функция правдоподобия быстро усложняется для источников тусклее $\sim 5\sigma$, что усиливает необходимость детального учета неопределенности вспомогательных параметров при изучении распределения потока тусклых источников.

Для тех, кому требуется фреквентистское решение задач, связанных с ошибками измерений, существует обширная и постоянно пополняющаяся литература, в которой представлены многочисленные методы, специализированные на решение конкретных задач. С байесовской точки зрения, специализированные методы вовсе не нужны: маргинализация по скрытым параметрам решает такие проблемы в общем и гибко (хотя возникающие вычислительные проблемы могут потребовать нетривиальных численных алгоритмов, специализированных под разные условия). Примечательно, что даже статистики, которые обычно отдают предпочтение фреквентистским методам, рекомендуют байесовские подходы с маргинализацией по скрытым параметрам в сложных ситуациях, особенно для нелинейного моделирования данных с гетероскедастической ошибкой измерения (т.е. с измерениями, имеющими различные стандартные отклонения). Астрономические данные обычно имеют гетероскедастическую ошибку измерения, и способность маргинализации гибко учитывать это осложнение стала сильным стимулом для распространения байесовских методов для решения задач ошибки измерения в астрономии.

6.1.4 Систематические ошибки и неточность модели

Известный афоризм, приписываемый Джорджу Боксу (George Box), мотивирует нашу последнюю область применения маргинализации: «Все модели неверны, но некоторые из них полезны». В некоторых случаях анализа данных мы можем быть в значительной степени уверены в нашей модели, но все же хотим изучить возможность того, что выводы могут быть испорчены потенциальными влияниями, упущенными в модели. В других случаях мы можем сознательно принять приблизительную модель и захотеть учесть недостаточ-

ную точность данных, когда будем делать выводы о существенных особенностях модели. Здесь мы кратко опишем роль составных гипотез и маргинализации для улучшения количественной оценки неопределенности в таких условиях. Мы рассматриваем два случая: во-первых, когда систематическая ошибка возникает из-за неопределенности выбора модели (например, включать ли условия для ожидаемых повреждающих эффектов), и во-вторых, когда модель нацелена на учет только основных характеристик явления, и мы хотим учесть неопределенность в заведомо немоделируемых деталях.

Довольно распространенный способ, с помощью которого астрономы пытаются учесть систематическую ошибку от выявленных потенциально повреждающих эффектов, заключается в использовании классической проверки на значимость нулевой гипотезы (NHST, null hypothesis significance test), чтобы проверить, есть ли в данных значимые доказательства наличия эффектов. Для проверки наличия потенциального эффекта разрабатывается статистика; если нулевая гипотеза об отсутствии эффекта не отвергается, анализ продолжается в предположении, что эффект отсутствует. Существенная проблема этого подхода заключается в том, что неспособность отвергнуть нулевую гипотезу не обязательно должна соответствовать сильному предпочтению нулевой гипотезы перед альтернативной. В идеале мы хотели бы учитывать потенциальные систематические эффекты, взвешивая их таким образом, чтобы оценить доказательную силу в пользу или против их наличия.

Маргинальные правдоподобия могут обеспечить такое взвешивание, однако они обычно более чувствительны к выбору прайора, чем оценки параметров, в частности, к диапазонам пространств параметров, рассматриваемых для конкурирующих моделей, и использовать их для учета систематической ошибки следует с осторожностью.

Примером количественной оценки систематических ошибок с использованием маргинального правдоподобия и факторов Байеса является работа [Drell2000], в которой изучалось потенциальное влияние систематики, связанной с эволюцией источников, на ранние исследования доказательств существования темной энергии по результатам измерений кривых блеска сверхновых типа Ia (SN Ia) («SN-космология»). В работах, посвященных открытию темной энергии, не было найдено никаких существенных свидетельств эволюции источников SN Ia (зависимость свойств кривой блеска от красного смещения), поэтому при анализе предполагалось *нулевая эволюция*. В [Drell2000] были рассчитаны факторы Байеса, указывающие на неоднозначность данных в отношении наличия эволюции. Важным аспектом вычислений было рассмотрение различных прайоров для параметров моделей эволюции, чтобы убедиться в надежности полученных результатов. Допущение малой вероятности эволюции и маргинализация ее влияния значительно ослабляли доказательную силу присутствия темной энергии, *если только не считать космологию плоской*, что в то время было неоправданно. К счастью, в течение нескольких лет накапливающиеся доказательства других космологических явлений позволили установить плоскую космологию. С точки зрения анализа систематических ошибок [Drell2000], только в контексте этих более поздних доказательств открытие темной энергии стало надежным.

В астрономии мы часто сознательно используем модели, которые отражают только основные черты явления. Это особенно верно при первом исследовании пограничной области, где изучение таких особенностей может дать значительное понимание, даже если модель, отражающая основные черты, упускает те или иные детали. В качестве примера можно привести подсчет импульсов в кривых свечения гамма-всплесков (где нас может волновать только количество импульсов и их временные масштабы, а не всякие мелкие колебания) или моделирование функций светимости с нарушением степенного закона (где мы хотели бы оценить степенной наклон и светимость, независимо от мелких неровностей в распределении). Простое принятие модели основных признаков и игнорирование неточ-


ной спецификации модели может испортить выводы, в частности, создать искусственно жесткие ограничения на параметры базовой модели.

Один из подходов к решению данной проблемы заключается в том, чтобы просто «раздуть» выборочные распределения. Например, в статистической литературе при анализе данных счетчиков принято с самого начала использовать отрицательное биномиальное (NB), а не пуассоновское распределение, поскольку у NB-распределения есть дополнительный параметр, который может быть использован для расширения (overdispersion) предсказательного распределения отсчетов, и оно включает распределение Пуассона в качестве частного случая. [Vonamante2023] адаптировал эту идею для учета систематической ошибки при анализе пуассоновских данных в астрономических наблюдениях в режиме, когда χ^2 -фитирование номинально точно, предложив новое *расширенное распределение* χ^2 .

Мы разрабатываем дополнительные методы моделирования характерных признаков пуассоновских данных и данных точечных процессов, которые тесно связаны с использованием NB-распределений. Однако мы не принимаем расширенные распределения (overdispersed distributions) как есть. В отличие от большинства применений распределения Пуассона в статистической литературе, во многих астрофизических приложениях есть серьезные физические основания для принятия распределения Пуассона в качестве описания вариативности повторной выборки, ожидаемой в данных – так называемой *алеаторной неопределенности* (от латинского aleator – «игрок в кости»). Мы используем увеличение дисперсии для отражения комбинации алеаторных и эпистемических источников неопределенности. (Эпистемический компонент отражает систематические ошибки, которые могут не изменяться случайным образом в разных повторениях.)

Мы строим расширенные распределения с помощью представлений, которые связывают их с базовым алеаторным распределением выборки через правдоподобный механизм расхождения (discrepancy mechanism) (адаптируя понятие аддитивных функций расхождения из эмуляции гауссовых процессов в компьютерных моделях; подробнее см. ниже). При этом фактическая функция интенсивности (частота событий) для управляющего данными пуассоновского процесса является произведением параметрического показателя базовой модели, который мы хотим оценить, и неотрицательного коэффициента расхождения. На рис. 21 показано построение модели для подсчета числа фотонов от меняющегося во времени источника (с разбивкой на временные бины), используемой для оценки параметров базовой модели кривой блеска. На синей гистограмме показаны предсказания пуассоновской модели для количества фотонов в бинах в виде ожидаемых значений λ_i . Они получены с помощью базовой модели, идеализированного описания истинной кривой блеска, которая отражает ключевые характеристики, представляющие научный интерес (такие как местоположение, продолжительность и амплитуда). Чтобы предсказать наблюдаемые числа n_i , мы умножаем каждую интенсивность λ_i на случайный *коэффициент расхождения* α_i , взятый из распределения с единичным средним, как показано на вставке. Точнее, для α_i мы используем гамма-распределение, дисперсия которого, при фиксированном среднем, контролируется единственным параметром формы β . В процессе фитирования мы проводим маргинализацию по всем коэффициентам α_i , а β рассматриваем как неопределенный параметр, оцениваемый по данным.

Выбор гамма-распределения позволяет провести маргинализацию коэффициентов расхождения аналитически и получить предсказательную РМФ, которая оказывается NB с дисперсией, увеличенной по сравнению с пуассоновской. Такая схема явно признает, что фактическое распределение вариативности отсчетов должно быть пуассоновским. Она также указывает на обобщения, которые могут быть применены к точечным данным и которые могут учитывать корреляции в расхождениях между бинами.



figs/Loredo2024_fig11.png

Рис. 21: Схематическое изображение мультипликативного подхода. Синяя гистограмма показывает РМФ с распределением Пуассона, полученную в рамках базовой модели. Зеленые точки изображают РМФ, полученную в результате умножения пуассоновских матожиданий на результаты процесса расхождения с гамма-распределением. На вставке: гамма-распределение с единичным средним, использованное для коэффициентов расхождения.

6.1.5 Увеличение дисперсии и расхождение для данных с аддитивным гауссовским шумом

Рассмотрим аппроксимацию кривой (регрессию) зашумленных измерений, y_i , которые являются функциями известных значений x_i . Например, фитирование данных спектра, где y_i – спектральная плотность потока, а x_i – длина волны ($i = 1, \dots, N$). Пусть $f(x; \psi)$ – модель, предсказывающая истинное значение y как функцию x , с параметрами ψ . Функция $f(x; \psi)$ может быть простой формулой (например, спектр черного тела или степенной закон) или может быть вычислена только с помощью компьютерного моделирования (например, из кода радиационного переноса, используемого для фитирования спектральных данных атмосферы экзопланет). Наша цель – оценить параметры ψ .

Предположим, что измерения имеют аддитивные гауссовские ошибки с нулевым средним, каждая из которых имеет известное стандартное отклонение σ_i . Тогда мы можем записать модель в виде

$$y_i = f(x_i; \psi) + \epsilon_i; \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$$

где $\mathcal{N}(0, \sigma^2)$ обозначает нормальное распределение с нулевым средним и дисперсией σ^2 . Функция правдоподобия может быть вычислена из произведения N нормальных PDF для измерений y_i ,

$$\mathcal{L}(\psi) \propto \frac{1}{\prod_i \sigma_i} \exp \left[-\frac{\chi^2(\psi)}{2} \right]$$

где $\chi^2(\psi)$ – обычная взвешенная сумма квадратов остатков

$$\chi^2(\psi) \equiv \sum_i \left(\frac{y_i - f(x_i; \psi)}{\sigma_i} \right)^2$$

Байсовский вывод предполагает умножение функции правдоподобия на прайор и исследование полученного апостериорного распределения. Но важным компонентом байсовского метода (да и фреквентистского тоже) является проверка модели – предоставление некоторого теста на адекватность предполагаемой модели, обычно путем измерения того, насколько типичны или нетипичны наблюдаемые данные по сравнению с тем, что предсказывает модель. Фреквентистская проверка модели обычно исследует минимальное значение χ^2 . При байсовской проверке, чтобы учесть неопределенность параметров модели, оценивается постериорно-усредненное значение χ^2 . В любом случае, если это значение слишком сильно превышает N , то это повод для беспокойства по поводу выбранной модели.

Что же делать дальше? Один из путей – сформулировать конкурирующие функции $f(x; \psi)$ и сравнить их производительность с исходной моделью (например, с помощью байсовского сравнения моделей). Но еще один способ, которым исходная модель могла потерпеть неудачу, заключается в том, что масштаб шума мог быть неправильно определен, т.е. значения σ_i неверны. Неформальный способ, которым астрономы иногда справляются с этой проблемой, заключается в увеличении значений σ_i до тех пор, пока минимальное χ^2 не станет равным N (что имеет желательный эффект расширения функции правдоподобия). С байсовской точки зрения мы можем реализовать это, введя параметр расширенной дисперсии α , и изменив модель распределения шума на

$$\epsilon_i \sim \mathcal{N}(0, \alpha^2 \sigma_i^2)$$

что поменяет функцию правдоподобия на

$$\mathcal{L}(\psi, \alpha) \propto \frac{1}{\alpha^N \prod_i \sigma_i} \exp \left[-\frac{\chi^2(\psi)}{2\alpha^2} \right] \quad (3)$$

Затем мы можем ввести прайор на α (например, обеспечивая $\alpha > 0$ и, возможно, отражая ожидания, что α будет близка к единице и, скорее всего, будет больше единицы) и провести маргинализацию по нему для оценки ψ , учитывая неопределенность в истинных масштабах шума σ_i . В итоге маргинализация расширяет функцию правдоподобия несколько больше, чем подход $\min \chi^2 = N$, поскольку учитывает неопределенность в α (т.е. в истинном масштабе шума).

Примечание. Эта маргинализация может быть выполнена аналитически, если выбрать равномерный прайор на α или $\log \alpha$, или, что более информативно, в виде обратного гамма-распределения – двухпараметрического семейства распределений, которое, например, может выражать наше предпочтение значений α , близких к единице.

Этот подход с увеличением дисперсии отражает гипотезу о том, что реальный уровень шума больше, чем первоначально предполагаемый. Потенциально мы можем проверить ее с помощью повторных наблюдений за одной и той же системой, при этом ожидается, что остатки будут варьироваться с увеличенным масштабом шума.

Альтернативный способ интерпретации слишком большого χ^2 заключается в том, что модель, $f(x; \psi)$, отражает существенное или важное поведение системы, но существуют небольшие отклонения от f , не отраженные в базовой модели, достаточно большие, чтобы мы могли рассмотреть, как они могут повлиять на оценку ψ . Это можно выразить, записав

$$y_i = f(x_i; \psi) + \delta_i + \epsilon_i$$

где мы добавили в модель новый член – расхождение δ_i , предназначенный для описания разницы между реальной функцией, предсказывающей y_i , и нашей моделью f (теперь уже

явно считающейся приближенной). Нам нужно назначить совместное распределение вероятностей для всех значений δ_i . Простейший выбор – описать неопределенность расхождения с помощью независимых нормально распределенных величин с нулевыми средними и со стандартными отклонениями s_i , пропорциональными масштабам шума, $s_i = \beta\sigma_i$. Сумма двух независимых нормальных сама является нормальной, а ее дисперсия равна сумме дисперсий отдельных величин. Таким образом, чистый эффект этой модели с расхождениями заключается в масштабировании σ_i в $1 + \beta$ раз. Если мы ре-параметризуем модель, положив $\alpha = 1 + \beta$, то в итоге получим ту же функцию правдоподобия (3). Байесовский вывод будет происходить, как и раньше, но интерпретация отличается. Теперь δ_i представляет собой эпистемическую неопределенность относительно того, как истинная функция, предсказывающая y_i , соотносится с базовой, но не совсем точной модельной функцией f . В этой ситуации, если бы мы многократно наблюдали одну и ту же систему, мы бы не ожидали, что остатки будут варьироваться в масштабах, превышающих σ_i . Скорее, мы ожидали бы, что δ_i будут одинаковыми при повторных наблюдениях и мы смогли бы оценить их, выявив тем самым различия базовой и истинной модели.

Этот вид моделирования с расхождениями довольно часто возникает в условиях, когда f требует больших вычислительных затрат и должен быть интерполирован по выборке прогонов компьютерной модели. Модели гауссовых процессов – гибкие интерполяторы, часто используемые для таких задач, в том числе в астрономии (см., например, [Higdon2012], [Cromer2024]⁸). В таких условиях используются более сложные и лучше мотивированные модели с δ_i , чем та, которую мы описали выше. Обычно δ_i моделируются корреляционным образом, обеспечивая определенный уровень гладкости расхождения как функции x (для этого используется дополнительный гауссовский процесс). Это позволяет проводить аналитическую маргинализацию возможных структурированных функций расхождения [Higdon2004].

6.1.6 Увеличение дисперсии и расхождение для пуассоновских данных

Многие астрономические измерения представляют собой подсчеты дискретных объектов или событий, которые часто хорошо моделируются с помощью распределения Пуассона. Это особенно характерно для данных изображений, спектральных и временных рядов, которые подсчитывают фотоны, где фундаментальная физика процессов излучения и обнаружения часто убедительно оправдывает использование распределений Пуассона. Интересным вопросом является то, как идеи перераспределения и расхождения могут быть реализованы в пуассоновских условиях.

Для конкретности рассмотрим временной ряд, в котором счет фотонов разбит на временные бины, n_i фотонов в бине i , а длительность бинов одноква, Δ . Предположим, что у нас есть модель, которая предсказывает скорость счета фотонов (функция интенсивности точечного процесса Пуассона), $\lambda(t; \psi)$, как функцию времени t и параметров модели ψ . Функция правдоподобия для ψ задается произведением пуассоновских распределений для отсчетов в каждом бине:

$$\mathcal{L}(\psi) = \prod_i \frac{[\lambda(t_i; \psi)\Delta]^{n_i}}{n_i!} \exp[-\lambda(t_i; \psi)\Delta]$$

(для простоты мы предположили, что скорость не сильно меняется с ростом Δ , поэтому

⁸[Cromer2024] представляет собой ссылку на GitHub-репозиторий с библиотекой Ostrich, осуществляющей суррогатное моделирование при помощи PCA и GP-интерполяции. Некоторое описание данного продукта приведено в статье [Cromer2021].

ожидаемое количество фотонов определяется произведением скорости в середине времени бина t_i и длительности бина).

Теперь предположим, что для этой модели есть признаки несоответствия, то есть разница между наблюдаемыми количествами n_i и ожидаемыми количествами $\lambda(t_i; \psi)\Delta$ больше, чем пуассоновское стандартное отклонение, определяемое квадратным корнем из ожидаемых количеств. В некоторых случаях допущение Пуассона может быть просто неправдой точкой, не имеющей убедительного физического обоснования, и можно искать альтернативную РМФ, которая обеспечит больший контроль над дисперсией. Для распределения Пуассона среднее и стандартное отклонение задаются величиной λ . Отрицательное биномиальное (NB) распределение – это распределение отсчетов, которое предлагает отдельный контроль среднего и стандартного отклонения, что позволяет добиться увеличения дисперсии.

В задачах с бинарными исходами (скажем, успех и неудача) с вероятностью успеха θ вероятность увидеть n успехов в N испытаниях задается биномиальным распределением, в котором N фиксировано априори, а n случайно. Распределение NB дает вероятность для смежной ситуации, когда подсчитывается количество неудач n (переопределили) до достижения заданного (целевого) числа успехов r . Число неудач является случайным и может варьироваться от 0 (когда все r успехов происходят в начале последовательности) до ∞ . Таким образом, общее число испытаний является случайным и может составлять от n до ∞ . РМФ распределения NB имеет вид

$$p(n|\theta, r) = \frac{(n+1-r)!}{n!(r-1)!} (1-\theta)^n \theta^r \equiv \text{NB}(n; r, \theta)$$

(Это просто биномиальная вероятность увидеть $r-1$ успехов в $n+(r-1)$ испытаниях, а затем увидеть успех в следующем испытании.) Математически эта РМФ для n хорошо определена даже для нецелых (положительных) значений r (в этом случае факториалы должны быть вычислены с помощью гамма-функций). Параметры θ и r позволяют отдельно контролировать среднее $\mu = r(1-\theta)/\theta$ и стандартное отклонение величины n . NB становится пуассоновским со средним числом отсчетов μ в пределе $\theta \rightarrow 1$ при фиксированном μ (что соответствует заданию $r = \mu\theta/(1-\theta)$). Байесовский вывод ψ в условиях с увеличенной дисперсией отсчетов осуществляется с помощью NB, обычно перепараметризованного в терминах μ (например, заданного $\lambda(t; \psi)\Delta$) и параметра коэффициента увеличения.

Как можно внедрить идею расхождения в моделирование пуассоновских данных? Покажем, что для этого полезно использовать распределение NB. Предположим, что $\lambda(t; \psi)$ – базовая модель, не четко определяющая реальное пуассоновское среднее. Запишем фактическое пуассоновское среднее в бине i как произведение базового среднего и мультипликативного коэффициента расхождения α_i (играющего роль, аналогичную роли δ_i в гауссовском случае), так что функция правдоподобия теперь имеет вид

$$\mathcal{L}(\psi; \{\alpha_i\}) = \prod_i \frac{[\alpha_i \lambda(t_i; \psi)\Delta]^{n_i}}{n_i!} \exp[-\alpha_i \lambda(t_i; \psi)\Delta]$$

Мы описываем неопределенность в коэффициентах расхождения с помощью PDF для значений α_i , матожидание которого равно единице, так что ожидается, что базовая модель будет точной в среднем (по всем бинам). Удобным выбором является гамма-распределение (степенной закон, умноженный на экспоненту α) с единичным средним и единственным параметром, управляющим его шириной (разбросом коэффициентов расхождения). При

таким выборе мы можем аналитически маргинализовать параметры α_i – это даст NB-распределение для числа отсчетов в каждом бине. Маргинальная функция правдоподобия для параметров функции скорости ψ может быть записана как

$$\mathcal{L}_m(\psi) = \prod_i \text{NB}(n_i; \beta \lambda(t_i; \psi) \Delta, \beta / (1 + \beta))$$

где β – параметр, определяющий ширину гамма-распределения коэффициентов расхождений. При такой параметризации стандартное отклонение Пуассона увеличивается в $\sqrt{1 + 1/\beta}$ раз. Здесь распределение NB возникает не для описания расширенной дисперсии, которую мы ожидаем увидеть в виде избыточной изменчивости отсчетов при повторных наблюдениях одной и той же системы, а для учета эпистемической неопределенности в отношении расхождения между базовой моделью $\lambda(t_i; \psi)$ и истинной частотой событий. Эта схема особенно уместна в условиях, когда физика сильно мотивирует модель Пуассона как описание алеаторной неопределенности (ожидаемой изменчивости), поскольку она предполагает наличие скрытой модели, которая описывает данные с помощью распределения Пуассона.

Гамма-распределение позволяет аналитически маргинализовать коэффициенты расхождения. Но выбор гамма-распределения имеет и более фундаментальную подоплеку. Можно представить себе ре-биннинг данных (либо деление, либо агрегирование бинов). Мы хотели бы, чтобы модель расхождений удовлетворяла *принципу согласованности агрегации*: применение подхода к исходным данным, а затем ре-биннинг должно давать те же выводы, что и применение подхода к ре-биннинговым данным. Гамма-распределение обеспечивает это требование согласованности путем масштабирования β размером бина.

6.2 Пуассоновские точечные процессы

Данное направление разработано в ряде работ Лоредо в 1990-е. Наиболее детально метод был сформулирован в совместной работе Gregory and Loredo, A new method for the detection of a periodic signal of unknown shape and period [Gregory1992]. Некоторые детали чуть подробнее были рассмотрены в работе Loredo, Bayesian Inference With Log-Fourier Arrival Time Models and Event Location Data [Loredo1993]. Однако этой темой Томас занимался вплоть до начала 2010-х, а в дальнейшем очень часто приводил в качестве примеров для пояснения и популяризации байесовского подхода.

6.2.1 Оценка блеска TNO

Большинство работ Loredo по PPP, в том числе и [Gregory1992], посвящено выявлению периодических процессов в данных, представленных временами прихода одиночных фотонов. Однако в [Loredo2004] PPP используются для данных измерений блеска так называемых *транснептуновых объектов* (TNO). В этой работе, в частности, говорится о том, как неопределенности в «точечных» данных могут приводить к деформации распределений – гистограмм, построенных путем разбиения данных по бинам. Впервые на эту проблему обратил внимание Jeffreys еще в 1938 году. С распределенными по бинам данными происходит что-то подобное тому, что и при размытии изображения в результате конволюции с PSF. Поэтому Jeffreys критиковал предложенные тогда Эддингтоном решения, которые производили де-конволюцию как задачу обращения, а не как *логический вывод* (inference).

В [Loredo2004] в качестве данных выступают измеренные значения *относительной светимости* (блеска) m TNO в некотором обзоре неба (с телесным углом обзора ω), $\mathcal{D} = \{d_i\}$, $i = 1, \dots, N$ – номер измеренного TNO.

figs/Loredo2004_fig04.png

Рис. 22: К построению функции правдоподобия для идеализированного обзора как точечного процесса по m с пустыми интервалами размера Δ_α и TNO, обнаруженными в «точках» m_i в небольших интервалах размера δm .

Сначала рассматривается идеализированная ситуация, когда измерение является точным, $d_i = m_i$ (m_i – истинное значение блеска), а эффективность детектирования – пороговой, т.е. как TNO идентифицируются все *надпороговые измерения*, т.е. с $m_i > m_{\text{thr}}$ с некоторым (известным) порогом m_{thr} . В этом случае для построения правдоподобия $p(\mathcal{D}|\theta)$, где θ – параметр, контролирующий распределение TNO по блеску, достаточно разбить (одномерную) шкалу m на интервалы Δ_α , где TNO не обнаружен, и узкие интервалы δm , где обнаружен всего один TNO (фактически ширина δm определяется разрешением прибора обзора), рис. 22, и ввести ожидаемое (среднее) число объектов на каждом из таких интервалов:

$$\mu_\alpha = \omega \int_{\Delta_\alpha} dm \sigma(m), \quad \mu_i = \omega \delta m \sigma(m)$$

Здесь $\sigma(m)$ – плотность распределения по блеску, т.е. $\sigma(m) dm d\omega$ – вероятность оказаться TNO со светимостью в интервале $[m, m + dm]$ внутри телесного угла обзора $d\omega$. Она как раз и контролируется интересующим нас параметром(и) θ .

Вероятность отсутствия TNO в интервале Δ_α – это пуассоновская вероятность отсутствия событий при ожидаемых μ_α , равная $\exp(-\mu_\alpha)$. Вероятность увидеть TNO с блеском m_i в интервале δm – это пуассоновская вероятность одного события при ожидаемых μ_i , т.е. $\mu_i \exp(-\mu_i)$. Перемножение этих вероятностей дает правдоподобие (ожидаемые значения в экспонентах суммируются и дают интеграл от $\sigma(m)$ по всем возможным значениям m)

$$\mathcal{L}(\theta) \equiv p(\mathcal{D}|\theta) = (\omega \delta m)^N \exp \left\{ -\omega \int dm \Theta(m_{\text{thr}} - m) \sigma(m) \right\} \prod_{i=1}^N \sigma(m_i) \quad (4)$$

где функция Хевисайда Θ в подынтегральном выражении специфицирует принятую нами идеальную пороговую эффективность регистрации.

При переходе от идеального обзора к реальному придется сделать два существенных изменения. Во-первых, вместо пороговой эффективности ввести некоторую функцию $\eta(m)$, определяемую критериями отбора событий в обзоре и их идентификации как TNO (будем считать ее заданной). А во-вторых, надо будет ввести функцию правдоподобия самого измерения, $l_i(m) \equiv p(d_i|m)$, как вероятность получения измерения d_i источника i в предположении, что источник имеет блеск m (часто она может быть адекватно аппроксимирована гауссовой функцией, заданной наилучшим – максимумом правдоподобия – значением m для TNO и его неопределенностью).

В байесовском подходе истинные значения блеска m_i выступают в качестве вспомогательных (nuisance) параметров и по ним необходимо провести интегрирование.

$$p(\mathcal{D}|\theta) = \int p(\{d_i\}, \mathcal{N}, \{m_j\}|\theta) \{dm_j\} = \int \{dm_j\} p(\{m_j\}, \mathcal{N}|\theta) p(\{d_i\}|\{m_j\}, \mathcal{N}, \theta)$$

Обращаем внимание, что здесь кроме измерений в данные вошло еще предположение \mathcal{N} , утверждающее, что других объектов не детектировано, $\mathcal{D} = \{d_i\}, \mathcal{N}$, а $\{m_j\}$ также понимается как утверждение, что в соответствующих интервалах ширины δm существуют TNO со светимостью m_j .

Первый фактор вполне аналогичен рассмотренной выше идеализированной ситуации с той лишь разницей, что, ввиду наличия предположения \mathcal{N} , под μ_α теперь понимается не общее ожидаемое число TNO в интервале Δ_α , а ожидаемое число детектированных TNO, определяемое в том числе и эффективностью, $\mu_\alpha = \omega \int_{\Delta_\alpha} dm \eta(m) \sigma(m)$. Выражение для μ_j при этом не поменяется, $\mu_j = \omega \delta m \sigma(m_j)$ (они ведь уже были детектированы!). Поэтому

$$p(\{m_j\}, \mathcal{N} | \theta) = \prod_{\alpha} e^{-\mu_\alpha} \prod_j \mu_j e^{-\mu_j} = (\omega \delta m)^N \prod_{j=1}^N \sigma(m_j) \exp \left\{ -\omega \int dm \eta(m) \sigma(m) \right\} \quad (5)$$

где в последнем переходе объединение интервалов Δ_α было заменено на весь интервал изменения светимостей, а вкладом в экспоненциальный член от $\sigma(m_j)$ пренебрежено (он пропорционален δm).

Не совсем понятно: а если окажется, что δm сравнимы с Δ_α – как быть тогда? Есть ощущение, что тут закралась ошибка и надо отдельно учитывать интеграл по m по интервалам без событий (они входят с множителем η) и с событием (без η). Тогда приведенное выше выражение придется домножить на $\exp\{-\omega \delta m \sum_i \sigma(m_i)\}$, но при этом интеграл распространить лишь на объединение интервалов Δ_α . Фактически же это будет эквивалентно исходному выражению (5) (с интегрированием по всему интервалу изменения блеска) и умножением на фактор $\exp\{\omega \delta m \sum_i \sigma(m_i)(\eta(m_i) - 1)\}$.

С другой стороны,

$$p(\{d_i\} | \{m_j\}, \mathcal{N}, \theta) = p(\{d_i\} | \{m_j\}) = \prod_{i=1}^N l_i(m_j), \quad l_i(m) \equiv p(d_i | m)$$

Поэтому

$$p(\mathcal{D} | \theta) = (\omega \delta m)^N \exp \left\{ -\omega \int dm \eta(m) \sigma(m) \right\} \prod_{i=1}^N \int \sigma(m) l_i(m) dm \quad (6)$$

Примечание. Отсутствие фактора эффективности η в последнем подинтегральном выражении требует пояснения. Пусть \mathcal{D}_i обозначает, что TNO действительно детектирован по данным полученным от TNO-кандидата номер i . Нас интересует совместная вероятность $p(\mathcal{D}_i, d_i | m)$, так как именно она умножается на $\sigma(m)$. По правилу произведения вероятностей

$$p(\mathcal{D}_i, d_i | m) = p(d_i | m) p(\mathcal{D}_i | d_i, m) = p(\mathcal{D}_i | m) p(d_i | \mathcal{D}_i, m)$$

Первое выражение непосредственно даст необходимый нам результат, так как $p(d_i | m) = l_i(m)$, а $p(\mathcal{D}_i | d_i, m) \equiv 1$ ввиду того, что d_i – данные от детектированного TNO (детектирование – это критерий того, что наблюдаемые данные находятся в некотором допустимом множестве, и по определению данные детектированного TNO должны находиться в этом допустимом множестве). Аналогично во втором выражении $p(\mathcal{D}_i | m) = \eta(m)$ по определению, а вот вероятность $p(d_i | \mathcal{D}_i, m)$ надо рассчитывать аккуратно – она отнюдь не равна $l_i(m)$ (правильное значение получается по теореме Байеса и равно $l_i(m)/\eta(m)$ и поэтому факторы η при перемножении сокращаются).

В [Loredo2004] данная вероятностная модель была проверена на модельных данных с $\sigma(m) = \sigma_{23} 10^{\alpha(m-23) + \alpha'(m-23)^2}$ (т.е. $\theta = \{\alpha, \alpha'\}$) В соответствие с этим распределением была сгенерирована выборка светимостей TNO, а затем была сформирована выборка числа пуассоновских отсчетов с ожидаемым значением, пропорциональным потоку от этого источника при простом однопиксельном измерении. Если количество отсчетов превышало пороговое значение n_{thr} , TNO детектировалось, и его пуассоновское правдоподобие использовалось для $l(m)$ (использовались значения $\alpha = 0.75$, $\alpha' = -0.05$, порог и ожидаемое число были подобраны так, чтобы самые тусклые детектированные TNO имели неопределенность 33%).

Реконструкция производилась следующим образом. Зафиксировав N (для сравнения использовалось $N = 100$ и $N = 1000$), генерировались N выборок надпороговых данных $\{d_i\}$. Выбрав равномерный прайоры на α и α' вычислялись моды маргинальных постериорных распределений $p(\alpha|\mathcal{D})$, $p(\alpha'|\mathcal{D})$, полученных по (6), которые затем наносились как точки на диаграмму $\alpha - \alpha'$. Распределение так полученных реконструированных точек θ сравнивалось с точкой истинных значений.

Кроме того, демонстрировалось как такое распределение деформируется, если не использовать неопределенность данных, т.е. предполагая $d_i = m_i$, и правдоподобие задается (4). Оказалось, что оценки, не учитывающие погрешность измерений, систематически завышают α и занижают α' .

Основной вывод работы: погрешности измерений должны тщательно учитываться при анализе данных обзора. В частности, влияние неопределенности не «усредняется» по мере увеличения размера наборов данных, а, наоборот, может усиливаться. Байесовский вывод оказывается идеальным инструментом для решения этой проблемы. Учет объемов в пространстве параметров – особенно объемов, связанных со случайными параметрами, возникающими из-за погрешности измерений, – позволяет байесовскому анализу точно учитывать искажения в данных.

Более детально PPP рассмотрены в работах [Gregory1992], [Loredo1993] и [Loredo2011a]: в первых двух⁹ подробно изложен предлагаемый авторами метод реконструкции периодических процессов, в третьей проведено сравнение с имеющимися в литературе ad hoc (фреквентистскими) методами. Ниже мы рассмотрим эти работы параллельно.

6.2.2 Периодограмма и мультимодальность

В астрономии наиболее известные и влиятельные статистические методы обнаружения и определения характеристик периодических сигналов используют *периодограммы*.

Пусть даны N эквидистантных измерений d_i , в которых мы пытаемся обнаружить следы некоторой периодической функции, т.е. $d_i = g(t) + \epsilon_i$, причем $g(t + 1/f) = g(t)$ с неизвестной нам частотой f и независимыми ошибками измерения ϵ_i . Общее время измерений обозначим $T = (N - 1)\delta t$.

Периодограммой Шустера называют функцию

$$\mathcal{P}(\omega) = \frac{1}{N} \left| \sum_{i=1}^N d_i e^{i\omega t_i} \right|^2$$

Она возникает сразу в нескольких статистических контекстах обнаружения периодического сигнала. Базовый рассматривает в качестве нулевой гипотезу отсутствия периодической

⁹Обращаем внимание, что работа [Loredo1993] является по сути техническим отчетом и не публиковалась в реферируемых журналах. К сожалению, в ней замечен ряд опечаток, особенно в нормировочных множителях и коэффициентах – так что при сравнении с другими статьями следует быть аккуратным.

компоненты и независимо нормально распределенные (i.i.d.) ошибки измерений,

$$\mathcal{H}_0 : g(t) \equiv 0, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

В этом случае можно показать, что при $\omega_j = 2\pi f_j = 2\pi j/T$ величины $\mathcal{P}(\omega_j)$ также являются i.i.d., причем

$$\mathcal{P}(\omega_j) \sim \chi_2^2 \equiv \text{Exp}(\lambda = 1/2)$$

Частоты ω_j получили название *частот Фурье*, всего их $N_F = N/2$. *SSH: Хм, а почему не N? Видимо, для вещественного сигнала вторая половина зависима от первой.* Независимость $\mathcal{P}(\omega_j)$ подразумевает, что непрерывная функция $\mathcal{P}(\omega)$ может иметь значимую структуру на масштабах угловой частоты $\sim 2\pi/T$. *SSH: Не совсем понимаю почему?*

Теперь легко сформулировать первый способ обнаружения периодической компоненты в измерениях – в стиле проверки статистических гипотез. Для этого надо найти самое большое значение периодограммы на фурье-спектре, вычислить по нему p-value p_1 для χ_2^2 -распределения и модифицировать ее к общему p-value как $p = p_1 N_F$: если p мало (скажем, $p < 0.01$), можно утверждать, что есть значимое доказательство наличия периодической компоненты.

На практике, когда в измерениях действительно есть периодический сигнал, он скорее всего не приходится на фурье-частоту. Чтобы сделать обнаружение более чувствительным, приходится совершать оверсэмплинг с фактором M (обычно не очень большим числом), т.е. рассматривать значения периодограммы на MN_F частотах, отстоящих друг от друга на $1/(MT)$. Проблема в том, что теперь значения периодограммы не являются независимыми даже в случае гипотезы \mathcal{H}_0 и приходится использовать Монте-Карловское моделирование, чтобы вычислять поправочный коэффициент к p-value.

Другой способ выхода на периодограмму соответствует рассмотрению параметрически заданной задачи. Пусть мы пытаемся обнаружить в измерениях периодический сигнал вида $g(t) = A \cos(\omega t + \phi)$ и будем это делать стандартным методом наименьших квадратов (LS), т.е. минимизируя

$$Q(\omega, A, \phi) = \sum_{i=1}^N [d_i - A \cos(\omega t_i + \phi)]^2$$

по параметрам A и ϕ (фиксируя ω). Подставляя в Q LS-оценки $\hat{A}(\omega)$, $\hat{\phi}(\omega)$, получаем профильную статистику $Q_p(\omega) = Q(\omega, \hat{A}(\omega), \hat{\phi}(\omega))$, причем в предположении \mathcal{H}_0

$$Q_p = \text{const} - \mathcal{P}(\omega)$$

В рамках вероятностной интерпретации LS-метода, речь идет о максимизации функции правдоподобия $\mathcal{L}(\omega, A, \phi) = -\frac{1}{2}Q(\omega, A, \phi)$, получаемое при этом правдоподобие $\mathcal{L}_p = \mathcal{L}(\omega, \hat{A}(\omega), \hat{\phi}(\omega))$ также называют *профильным*.

Следствием этой тесной связи между параметрическим гармоническим анализом и периодограммами является то, что сильная изменчивость (непараметрической) периодограммы подразумевает сильную мультимодальность функции правдоподобия гармонической модели (а следовательно, и апостериорного распределения в байесовском гармоническом анализе) на частотных масштабах $\sim 1/T$.

Обобщение периодограммы на случай не-эквидистантных измерений было дано в работах [Lomb1976] и [Scargle1982] и получило название *периодограммы Ломба-Скаргла* (LSP). Jaynes (1987) и Bretthorst (1988) вместо того, чтобы *максимизировать* функцию правдоподобия по амплитуде и фазе, «поступили по-байесовски» – провели *маргинализацию* для

этих параметров. При этом логарифм маргинальной плотности для частоты пропорционален периодограмме; аналогичная связь с LSP для не-эквидистантных выборок получена в [Bretthorst2001]. Но это было нечто большее, чем повторное открытие прежних результатов. В рамках байесовского подхода расчеты по преобразованию значений периодограммы в вероятностные утверждения о сигнале разительно отличаются от их аналогов в фреквентистском спектральном анализе.

Самое главное отличие проявляется не при оценке параметров, а при обнаружении сигнала путем сравнения моделей. Условная вероятность того, что периодический сигнал присутствует на априорно известной частоте, приблизительно равна экспоненте от периодограммы. Но частота никогда не известна точно априори. Для обнаружения новых периодических источников приходится проводить «слепой поиск» в большом диапазоне частот. Даже для восстановления известного сигнала в новых данных неопределенность (предсказания) частоты, основанная на предыдущих измерениях, обычно значительна. В байесовских расчетах неопределенность частоты учитывается путем вычисления не максимального, а маргинального правдоподобия. При этом у частот Фурье нет никакой особой роли – ни в их расположении, ни в количестве; на самом деле, нужно оценивать периодограмму на стольких частотах, сколько необходимо для точного вычисления интеграла по *непрерывной* экспоненте от периодограммы.

Подводя итог, можно отметить следующие ключевые моменты: (1) Мы ожидаем, что правдоподобие (а следовательно, и постериорная вероятность) будет очень мультимодальным в частотном измерении. (2) Масштаб изменчивости правдоподобия в частотном измерении $\sim 1/T$, где T – общее время измерений. Для задач с длительными наборами данных и значительной априорной неопределенностью частоты анализ частотного измерения будет сложным. (3) Ключевое различие между байесовским и фреквентистским подходами возникает в том, как обрабатывается неопределенность частоты (и других параметров), например, максимизируется ли она и затем корректируется для множественных тестов, или маргинализируется, позволяя путем усреднения вероятности неявно учитывать размер пространства параметров.

Для решения большинства задач обнаружения периодических сигналов по данным о времени прихода астрономы используют фреквентистские методы, вдохновленные подходом периодограмм в условиях аддитивного шума. Применяются различные статистики, но на практике доминируют три. Все они начинаются с формирования данных по модулю пробного периода, т.е. для каждой *пробной частоты* ω по времени t_i рассчитывается фаза $\varphi_i \in [0, 2\pi)$. Статистика должна выявлять отклонения от равномерности в распределении по фазе.

Статистика Рэлея $R(\omega)$ определяется как

$$R^2 = \frac{1}{N} \left[\left(\sum_{i=1}^N \sin \varphi_i \right)^2 + \left(\sum_{i=1}^N \cos \varphi_i \right)^2 \right]$$

Величина $2R^2(\omega)$ называется *рэлеевской мощностью*. Это аналог периодограммы Шустера для точечных процессов, и при нулевой гипотезе асимптотически $2R^2 \sim \chi_2^2$. *SSh: В каком смысле асимптотически?* На практике статистика Рэлея хорошо подходит для обнаружения сигналов, имеющих гладкие кривые блеска с одним пиком на период.

Вторая статистика определяется суммой мощностей для M гармоник

$$Z_M^2 = 2 \sum_{m=1}^M R^2(m\omega)$$

и при нулевой гипотезе асимптотически $Z_M^2 \sim \chi_{2M}^2$. Число гармоник M обычно задается априори (популярный выбор – $M = 2$), хотя можно также адаптировать значение M к данным.

Третий часто используемый метод – χ^2 -epoch folding (усреднение по эпохам, χ^2 -EF). Для каждой пробной частоты полученные фазы разбиваются на K равных по ширине фазовых бинов, и для проверки нулевой гипотезы о равномерности распределения фаз используется пирсоновский χ^2 . Количество бинов выбирается априори. Количество отсчетов в каждом бине (для выбранного ω) зависит от начала отсчета времени: перемещение начала отсчета изменит фазы и перераспределит события между бинами. Чтобы учесть это, статистику χ^2 можно усреднить по фазе. Это изменяет ее распределение при нулевой гипотезе, поэтому ее исследование проводят путем Монте-Карло моделирования.

Статистики Z_M и χ^2 -EF могут быть более чувствительны к структурированным крикам блеска, чем статистика Рэлея, но при этом возникают дополнительные сложности в виде Монте-Карло распределений или необходимости априорно фиксировать параметры структуры (количество гармоник или бинов).

6.2.3 Общая постановка байесовской задачи

Рассмотрим байесовские альтернативы традиционным тестам, построенные с использованием моделей *неоднородного пуассоновского точечного процесса* с зависящей от времени интенсивностью $r(t)$ – ожидаемой частотой событий в единицу времени. Для периодических моделей параметры $r(t)$ будут включать амплитуду A , угловую частоту ω , начальную фазу ϕ (соответствующую определению начала времени) и один или несколько параметров формы \mathcal{S} , которые параметризуют форму кривой свечения.

Данные представлены N значениями моментов прихода (событий) $\mathcal{D} = \{t_i\}, i = 1, \dots, N$, зафиксированных регистрирующим прибором на интервале времени длительностью T .

Функция правдоподобия имеет вид,

$$\mathcal{L}(A, \omega, \phi, \mathcal{S}) = \exp \left\{ - \int_T dt r(t) \right\} \prod_{i=1}^N r(t_i) \quad (7)$$

Обращаем внимание, что время T – это общая длительность интервалов, для которых известно, что они содержат одно или ни одного события. Оно не обязательно должно быть непрерывным – допускается наличие разрывных промежутков (*гэпов*), для которых нет данных.

Примечание. Как получена формула (7)? В полной аналогии с PPP задачи о распределении TNO по блеску, мы можем разбить временную ось на интервалы длительностью δt такие, что в каждый из них может быть обнаружено одно или ни одного события. Из распределения Пуассона для вероятности наблюдения n событий $p_n(t) = (r(t)\delta t)^n \exp\{-r(t)\delta t\}/n!$, следует выражение для функции правдоподобия

$$p(\mathcal{D}|r) = \prod_{i=1}^N p_1(t_i) \prod_{\alpha=1}^Q p_0(t_\alpha)$$

где Q – количество временных интервалов, в которые не детектировано ни одного события, $p_0(t) = \exp\{-r(t)\delta t\}$, $p_1(t) = r(t)\delta t \exp\{-r(t)\delta t\}$. Подставляя выражения для p_0, p_1 , получаем

$$p(\mathcal{D}|r) = (\delta t)^N \prod_{i=1}^N r(t_i) \exp \left\{ - \sum_{j=1}^{N+Q} r(t_j)\delta t \right\}$$

эквивалентное (7) с областью интегрирования $T = (N + Q)\delta t$. Длина интервала δt определяется временным разрешением регистрирующего прибора. Но когда используется теорема Байеса, фактор $(\delta t)^N$ входит и в числитель и знаменатель и в итоге сокращается.

Так как наша задача заключается в поиске в данных периодических сигналов, то представим модельный сигнал $r(t)$ в виде

$$r(t) = A\rho(\omega t + \phi), \quad A \equiv \frac{1}{P} \int_P dt r(t), \quad P \equiv 2\pi/\omega$$

с периодической функцией $\rho(\varphi)$ с периодом 2π , нормированной $\int_0^{2\pi} \rho(\varphi) = 2\pi$. Тем самым в качестве одного из параметров модели мы выделили амплитуду сигнала A . При этом

$$\mathcal{L}(A, \omega, \phi, \mathcal{S}) = A^N e^{-AT} \prod_{i=1}^N \rho(\omega t_i + \phi)$$

Здесь мы предположили, что T охватывает много периодов, так что интеграл по времени в экспоненте хорошо аппроксимируется значением AT . Если к тому же предположить априорную независимость A от других параметров модели, то выражение для маргинализованного правдоподобия на частоту, фазу и параметры формы упрощается до

$$\mathcal{L}(\omega, \phi, \mathcal{S}) \propto \prod_{k=1}^N \rho(\omega t_i + \phi)$$

Само постериорное распределение на амплитуду слабо зависит от прайора: $p(A|\mathcal{D}) \propto p(A)A^N \exp(-AT)$. В случае равномерного на некотором интервале $[0, A_{\max}]$ прайора постериорное распределение амплитуды является гамма-распределением $\mathcal{G}(N + 1, 1/T)$ со средним $A_{\text{mean}} = (N + 1)/T$ и дисперсией $\sigma_A^2 = (N + 1)/T^2$ (мода равна $A_m = N/T$). Если же в качестве прайора выбрать сопряженное правдоподобию экспоненциальное распределение¹⁰ $p(A) = \frac{1}{A_0} \exp(-A/A_0)$, то постериор также окажется гамма-распределением со средним, модой и дисперсией, равными

$$A_{\text{mean}} = (N + 1)/(T + 1/A_0), \quad A_m = N/(T + 1/A_0), \quad \sigma_A^2 = (N + 1)/T(T + 1/A_0)$$

Таким образом, в наиболее распространенной с физической точки зрения ситуации с $A_0 T \gg 1$ постериорное распределение на A не зависит от масштаба A_0 в прайоре (и совпадает с тем, что получается при равномерном прайоре).

Можно также показать, что для экспоненциального прайора на A маргинальное правдоподобие (evidence) может быть представлено в виде:

$$p(\mathcal{D}) = \frac{N!}{A_0(T + 1/A_0)^{N+1}} \int d\omega \int d\phi \int d\mathcal{S} p(\omega, \phi, \mathcal{S}) \prod_{i=1}^N \rho(\omega t_i + \phi)$$

6.2.4 Модели периодических процессов

Модель Log-Sin

Чтобы двигаться дальше, надо определить модели для $\rho(\varphi)$, помня при этом о вычислительных трудностях. В частности, поскольку нам нужно будет интегрировать функцию

¹⁰Сопряженным к правдоподобию, конечно же, является гамма-распределение $\mathcal{G}(k, \theta)$, а экспоненциальное – лишь его частный случай, $\mathcal{G}(1, 1/\lambda) = \text{Exp}(\lambda)$. Напомним также, что для гамма-распределения матожидание и дисперсия равны соответственно $k\theta$ и $k\theta^2$. Легко убедиться, что если $A \sim \mathcal{G}(k, A_0)$, то $A|\mathcal{D} \sim \mathcal{G}(k + N, A_0/(1 + A_0 T))$.

правдоподобия по пространству параметров (для вычисления маргиналов при оценивании параметров и маргинального правдоподобия при сравнении моделей), мы в первую очередь ищем модели, позволяющие выполнить как можно большую часть этого интегрирования аналитически. Здесь мы сосредоточимся на двух взаимодополняющих вариантах, один из которых позволяет провести аналитически маргинализацию по фазе, а другой – полупараметрическую модель, позволяющую аналитически маргинализовать параметры формы.

Первый вариант соответствует *Log-Sin модели*, которая с учетом нормировки ρ приводит к распределению фон Мизеса:

$$\rho(\varphi) = \frac{1}{I_0(\kappa)} e^{\kappa \cos(\varphi)}$$

где $I_0(\kappa)$ – модифицированная функция Бесселя 0-го порядка. Эта модель имеет единственный параметр формы, параметр концентрации κ , который одновременно управляет шириной пика на кривой блеска и отношением пик-долина (или долей импульса).

При равномерном прайоре на фазу маргинализация по ней проводится аналитически и приводит к итоговому результату (вывод приведен в [Loredo1993]):

$$\mathcal{L}(\omega, \kappa) = \frac{I_0(\kappa N R(\omega))}{(I_0(\kappa))^N}$$

Статистика Рэлея возникает здесь как *достаточная статистика* для оценки частоты и концентрации для Log-Sin модели. Интересно, что зависимость от κ определяется только значением R и размером выборки. Используя асимптотические свойства функции Бесселя, можно показать, что при наличии потенциального свидетельства наличия сигнала на определенной частоте (т.е. при $R > N$), правдоподобие является приблизительно гамма-распределением по κ . Кроме того, функция правдоподобия сильно коррелирует ω и κ , так что правдоподобие будет наибольшим на частотах, для которых концентрация оценена как большая (что интуитивно понятно). Прайор для κ в виде гамма-распределения является асимптотически сопряженным.

Примечание. В более общем случае правдоподобие на все параметры модели можно представить в виде *SSH: Во всех формулах этого раздела, содержащих R нужна перепроверка: в работах Лоредо есть опечатки в ее нормировках*

$$\mathcal{L}(\omega, \phi, \kappa) \propto [I_0(\kappa)]^{-N} \exp \left\{ \kappa \sqrt{N} R(\omega) \cos(\phi - \phi') \right\}$$

где выражение для $\phi' = \phi'(\omega)$ определено как

$$\tan \phi' = \frac{\sum_i \sin(\omega t_i)}{\sum_i \cos(\omega t_i)}$$

Фактически ϕ' задает оценку максимального правдоподобия для ϕ при заданных ω и κ . Более того, так как для $\phi \approx \phi'$ имеем $\cos(\phi - \phi') \approx 1 - (\phi - \phi')^2/2$, то гауссова аппроксимация (логарифма) правдоподобия позволяет оценить стандартное отклонение как $\sigma_\phi = 1/\sqrt{\kappa R(\omega) \sqrt{N}}$.

Зависимость правдоподобия от κ можно рассмотреть в пределах больших и малых значений κ , воспользовавшись асимптотиками модифицированной функции Бесселя:

$$I_n(x) \approx \left(\frac{x}{2}\right)^2 \left[\frac{1}{n!} + \frac{x^2}{4(n+1)!} \right], \quad x \rightarrow 0$$

и

$$I_n(x) \approx \frac{e^x}{\sqrt{2\pi x}} \left[1 - \frac{4n^2 - 1}{8x} \right], \quad x \rightarrow \infty$$

При этом для малых значений κ (точнее, при $\kappa \ll 1/N$)

$$\mathcal{L}(\omega, \kappa) \approx C_1(\omega) \left(1 + \frac{N\kappa^2}{4} (R^2(\omega) - 1) \right)$$

В пределе больших κ , когда $\kappa R(\omega) \gg 1/\sqrt{N}$

$$\ln \mathcal{L}(\omega, \kappa) \approx C_2(\omega) \kappa^{(N-1)/2} \exp \left\{ -\kappa(N - R(\omega)\sqrt{N}) \right\}$$

Так как $R < \sqrt{N}$, то $\mathcal{L}(\omega, \pi)$ экспоненциально падает при больших κ . Асимптотическое поведение подразумевает, что при $R < 1$ максимум правдоподобия расположен в $\hat{\kappa} = 0$, что соответствует отсутствию периодической компоненты. Ненулевая мода возможна только при $R(\omega) > 1$. Можно показать, что при $\hat{\kappa}N \ll 1$ выражение для моды $\hat{\kappa}^2 \approx 8 \frac{R^2 - 1}{NR^4 - 1}$, а при $\hat{\kappa}R \gg 1/\sqrt{N}$ $\hat{\kappa} \approx \frac{N-1}{2(N-R\sqrt{N})}$.

Расчет не асимптотических значений моды $\hat{\kappa}$ аналитически невозможен, о том как сделать это численно, а также оценить моду $\hat{\omega}$ и маргинальное правдоподобие $p(\mathcal{D}|\mathcal{M}_{\log-\sin})$ см. в работе [Loredo1993].

Модель Log-Fourier

Основным ограничением модели Log-Sin является наличие всего одного параметра, задающего профиль сигнала. В частности, параметр κ определяет как длительность пика, так и отношение пик-долина, что сложно обосновать физически. Поэтому в [Loredo1993] был предложен другой класс профилей, обобщающий Log-Sin на случай, когда под экспонентой стоит не одна гармоника, а целый их ряд – в стиле ряда Фурье:

$$\rho(t) = \frac{1}{I(\boldsymbol{\kappa}, \boldsymbol{\phi})} \exp \left\{ \sum_{m=1}^M \kappa_m \cos(m\omega t + \phi_m) \right\} \quad (8)$$

где $\boldsymbol{\kappa} = \{\kappa_m\}$, $\boldsymbol{\phi} = \{\phi_m\}$, $m = 1, \dots, M$, а $I(\boldsymbol{\kappa}, \boldsymbol{\phi})$ определяется (к сожалению не аналитически) из условия нормировки $\int_0^{2\pi/\omega} dt \rho(t) = 1/\omega$. При $M = 1$ эта модель, конечно же, совпадает с моделью Log-Sin.

Правдоподобие по частоте и параметрам формы в этом случае представимо в виде

$$\mathcal{L}(\omega, \boldsymbol{\phi}, \boldsymbol{\kappa}) \propto [I(\boldsymbol{\kappa}, \boldsymbol{\phi})]^{-N} \exp \left\{ \sum_{m=1}^M \kappa_m \sqrt{N} R(m\omega) \cos(\phi_m - \phi'_m) \right\}$$

с теми же, что и раньше, определениями R и ϕ' , с той лишь разницей, что в них подставляется частота $\omega_m = m\omega$.

В работе [Loredo1993] показана связь этого правдоподобия с широко используемой при фреквентистском подходе тестовой статистикой $Z_M^2 = 2 \sum_{m=1}^M R^2(m\omega)$. Так, при $\phi = \phi'_m$ и $\kappa_m = 1$

$$p(\omega, \boldsymbol{\phi} = \boldsymbol{\phi}', \boldsymbol{\kappa} = 1|\mathcal{D}) \propto \exp \left\{ \sum_m \sqrt{N} R(m\omega) \right\}$$

причем

$$\left(\sum_m R(m\omega) \right)^2 = \sum_m R^2(m\omega) + \sum_m \sum_{m' \neq m} R(m\omega) R(m'\omega)$$

Таким образом,...

Модель PCS

Сложность реальных кривых свечения пульсаров указывает на то, что модель, допускающая больше структуры, чем один гладкий пик за период, будет лучше обнаруживать пульсары, чем простая $\log\text{-sin}$ модель. В [Gregory1992] была предложена модель с *кусочно-постоянным профилем* (PCS, piecewise constant shape) для $\rho(\varphi)$, при которой ρ постоянна на каждом из K фазовых бинов равной ширины. Если K определяется данными, то эта модель становится полупараметрической по духу. Обозначим это семейство моделей \mathcal{M}_K , при этом \mathcal{M}_1 соответствует модели с постоянной интенсивностью, $r(t) = A$, т.е. отсутствию в данных периодической компоненты. В дальнейшем мы часто будем опускать символ \mathcal{M}_K после знака обуславливания, если из контекста понятно о какой модели идет речь (т.е. когда не производится сравнение моделей с разным K).

Функция PCS-профиля может быть записана

$$\rho(\varphi) = AK f_{k(\varphi)}, \quad k(\varphi) = [1 + K(\varphi \bmod 2\pi)/2\pi]$$

где параметры высоты $\mathbf{f} = \{f_k\}$ задают для каждой ступеньки длительностью $1/K$ периода; при такой параметризации параметры \mathbf{f} лежат на единичном симплексе, $f_k \geq 0$, $\sum_k f_k = 1$. Маргинальная функция правдоподобия для угловой частоты, фазы и формы имеет вид мультиномиального распределения:

$$\mathcal{L}(\omega, \phi, \mathbf{f}) \propto K^N \prod_{k=1}^K f_k^{n_k}$$

где $n_k = n_k(\omega, \phi)$ – число событий, попавших на ступеньку кривой свечения номер k при выбранных значениях ω и ϕ .

Привлекательность модели PCS заключается в простой зависимости от \mathbf{f} . Это позволяет аналитически маргинализировать по параметрам профиля, если использовать сопряженный прайор. В [Gregory92] было выбран равномерный прайор,

$$p(\mathbf{f}|\mathcal{M}_K) = \prod_k \pi(f_k), \quad \pi(f_k) = 1/K!$$

при котором

$$\mathcal{L}(\omega, \phi) \propto \frac{K^N (K-1)!}{(N+K-1)!} \left[\frac{n_1! n_2! \dots n_K!}{N!} \right] \quad (9)$$

В байесовских условиях *обратная кратность* (reciprocal multiplicity, член в квадратных скобках) дает больше, чем простая тестовая статистика: она позволяет рассчитать апостериорные вероятности для частоты, фазы и количества бинов. Кроме того, усредняя модель (по K , фазе и частоте), можно оценить форму кривой блеска без привязки к определенному разбиению на бины: набор точечных оценок $\rho(\varphi)$ является гладким (пусть и несколько «коробчатым»), хотя рассматриваемая как функция эта оценка находится за пределами исходной PCS-модели.

Стоит обратить внимание на связь модели PCS с χ^2 -EF. Используя приближение Стирлинга для факториалов при $n \gg 1$, $\ln n! \approx n \ln n - n + \frac{1}{2} \ln(2\pi n)$, можно показать, что для больших чисел отсчетов в бинах,

$$\ln \frac{N!}{n_1! \dots n_K!} \approx N\mathcal{H} - \frac{1}{2} \sum_k \ln n_k + \frac{1}{2} \ln N - \frac{K-1}{2} \ln(2\pi)$$

где была введена величина $\mathcal{H} \equiv -\sum_k \frac{n_k}{N} \ln \frac{n_k}{N}$, имеющая смысл *энтропии* распределения $\{n_k\}$. В этом пределе $p(\omega, \phi) \propto \exp\{-N\mathcal{H}\}$, т.е. наиболее вероятным значениям частоты и фазы (при выбранном количестве бинов K) соответствует кривая свечения с минимальной энтропией.

Если же продолжить огрублять метод и допустить, что во всех бинах число отсчетов не сильно отклоняется от среднего значения, $n_k \approx N/K$, то можно «выйти» и на χ^2 -статистику. Действительно, так как при $x \rightarrow 1$ $\ln x \approx (x-1) - \frac{1}{2}(x-1)^2$, то для $x = \frac{N/K}{n_k}$ имеем

$$\mathcal{H} = \frac{1}{N} \sum_k \left(n_k \ln K + n_k \ln \frac{N/K}{n_k} \right) \approx \ln K - \frac{1}{2N} \sum_k \frac{(n_k - N/K)^2}{n_k}$$

Другими словами,

$$p(\omega, \phi | \mathcal{D}) \propto \exp\{\chi^2/2\}, \quad \chi^2 = \sum_k \frac{(n_k - N/K)^2}{n_k}$$

Резюмируя: в пределе $n_k \approx N/K$

$$2 \ln \mathcal{L}(\omega, \phi) \approx \chi^2 + \sum_{k=1}^K \ln n_k + C(K)$$

где $C(K)$ – константа, зависящая только от K , а χ^2 – та же статистика, что и в методе χ^2 -EF. На самом деле, $\exp(\chi^2/2)$ может быть хорошим приближением к маргинальному правдоподобию для ω и ϕ . Несмотря на это, при имитационном моделировании модель PCS оказывается лучше способной обнаруживать слабые периодические сигналы, чем усредненная по фазе статистика χ^2 . Причина, вероятно, связана не столько с плохой аппроксимацией, сколько с тем, что с байесовской точки зрения правильной величиной для усреднения по фазе является не χ^2 , а именно $\exp(\chi^2/2)$. Ad hoc усреднение χ^2 для устранения вспомогательного параметра фазы по сути «сглаживает», а не маргинализирует зависимость модели от фазы.

Все дальнейшие вычисления будут опираться на следующее выражение для функции правдоподобия, маргинализованной по амплитуде и параметрам профиля (в обоих случаях – с равномерным прайором):

$$\mathcal{L}(\omega, \phi) \equiv p(\mathcal{D} | \omega, \phi, \mathcal{M}_K) = \frac{(\delta t/T)^N \gamma(N+1, A_{\max} T)}{A_{\max} T} \frac{(K-1)!}{(N+K-1)!} \frac{K^N}{W_K(\omega, \phi)} \quad (10)$$

с $W_K(\omega, \phi) = \frac{N!}{n_1! \dots n_K!}$ и $\gamma(N+1, A_{\max} T) = \int_0^{A_{\max} T} y^N \exp\{-y\} dy$.

Рассмотрим сначала ситуацию, когда по тем или иным причинам частота и фаза сигнала известны (т.е. ω и ϕ расположены справа от знака обуславливания). В этом случае нашей задачей является обнаружение периодического сигнала, т.е. сравнение гипотезы \mathcal{M}_1 со сложной гипотезой, представляющей собой совокупность \mathcal{M}_K с K меняющимся от 2 до некоторого максимального значения K_{\max} (ограничение связано с тем, что данные не могут свидетельствовать в пользу бесконечно тонко структурированного сигнала, на практике будут браться значения $K_{\max} = 12 - 15$).

Если ввести *шансы* (odds ratio) гипотезы \mathcal{M}_K против гипотезы \mathcal{M}_1 как

$$\mathcal{O}_{K,1} \equiv \frac{p(\mathcal{M}_K | \mathcal{D})}{p(\mathcal{M}_1 | \mathcal{D})} = \frac{p(\mathcal{M}_K) p(\mathcal{D} | \mathcal{M}_K)}{p(\mathcal{M}_1) p(\mathcal{D} | \mathcal{M}_1)} = \mathcal{O}_{K,1}^{\text{prior}} \cdot \frac{p(\mathcal{D} | \mathcal{M}_K)}{p(\mathcal{D} | \mathcal{M}_1)}$$

то

$$\mathcal{O}_{\text{per.}} = \sum_{K=2}^{K_{\max}} \mathcal{O}_{K,1} = \frac{\sum_{K=2}^{K_{\max}} p(\mathcal{D}|\mathcal{M}_K) \cdot \mathcal{O}_{K,1}^{\text{prior}}}{p(\mathcal{D}|\mathcal{M}_1)},$$

и вероятность обнаружения периодической компоненты можно будет представить как

$$p(\text{per.}|\mathcal{D}) = \frac{\mathcal{O}_{\text{per.}}}{1 + \mathcal{O}_{\text{per.}}}, \quad p(\text{non per.}|\mathcal{D}) = \frac{1}{1 + \mathcal{O}_{\text{per.}}}$$

Подстановка $p(\mathcal{D}|\mathcal{M}_K) \equiv \mathcal{L}$ дает

$$\mathcal{O}_{K,1}(\omega, \phi) = \mathcal{O}_{K,1}^{\text{prior}} \cdot \frac{K^N}{C_{N+K-1}^N W_K(\omega, \phi)}$$

Теперь перейдем к более реалистичной ситуации, когда известен только период (частота) пульсаций, а фаза нет. Например, такое может быть, когда мы изучаем рентгеновские пульсации, период которых был заранее надежно установлен по радио-сигналу. При этом для равномерного по фазе прайора

$$\mathcal{O}_{K,1}(\omega) = \mathcal{O}_{K,1}^{\text{prior}} \cdot \frac{K^N}{C_{N+K-1}^N} \int_0^{2\pi} \frac{d\phi/2\pi}{W_K(\omega, \phi)}$$

причем численную оценку интеграла достаточно будет провести на интервале $[0, 2\pi/K]$ и умножить результат на K : сдвиги фазы более чем на один бин просто соответствуют циклической перестановке f_k , которая не меняет значения кратности W_K .

Наконец, если неизвестны ни частота, ни фаза, то при масштабнo-инвариантном прайоре $p(\omega) \propto 1/\omega$

$$\mathcal{O}_{K,1} = \mathcal{O}_{K,1}^{\text{prior}} \cdot \frac{K^N}{\ln(\omega_{\max}/\omega_{\min}) C_{N+K-1}^N} \int_{\omega_{\min}}^{\omega_{\max}} \frac{d\omega}{\omega} \int_0^{2\pi} \frac{d\phi/2\pi}{W_K(\omega, \phi)}$$

Пределы по частоте в дальнейшем выберем равными $\omega_{\min} = 10 \cdot 2\pi/T$, $\omega_{\max} = N \cdot 2\pi/T$, что заведомо накрывает интервал возможных для анализа частот, исходя из длительности интервала измерений и их количества.

Если на этапе сравнения предсказаний различных моделей была выбрана модель \mathcal{M}_K как наиболее вероятная, то после этого может быть поставлена задача оценки частоты. Ее можно сделать по одномерному маргиналу

$$p(\omega|\mathcal{D}, \mathcal{M}_K) \propto \frac{1}{\omega} \int \frac{d\phi}{W_K(\omega, \phi)}$$

или по совместному с фазой распределению

$$p(\omega, \phi|\mathcal{D}, \mathcal{M}_K) \propto \frac{1}{\omega W_K(\omega, \phi)}$$

Эти маргинальные распределения показывают, что кратность $W_K(\omega, \phi)$ играет роль достаточной статистики при оценках частоты и фазы периодического сигнала в модели PCS. Обращаем внимание, что константы прайоров, A_{\max} , 2π , $\ln(\omega_{\max}/\omega_{\min})$, не вошли в постериорные распределения на параметры модели: результат байесова вывода весьма нечувствителен к интервалу прайоров. Это сильно контрастирует с задачей сравнения моделей: штрафующий фактор Оккама во многом определяется этими интервалами.

На самом деле наиболее последовательным с байесовской точки зрения вариантом оценки частоты, будет не подстановка одной какой-то модели, а усреднение (с постериорными вероятностями) по всем моделям, описывающим периодический сигнал (т.е. при $K \geq 2$): $p(\omega|K > 1, \mathcal{D}) = \sum_{K=2}^{K_{\max}} p(\mathcal{M}_K|\mathcal{D}) p(\omega|\mathcal{D}, \mathcal{M}_K)$.

А как в рамках модели PCS оценивать кривую свечения?

Так как постериорное распределение по амплитуде факторизуется, то фактически речь идет об оценке профиля $\rho(t)$. При этом следует помнить, что оценка отдельного параметра f_k еще не дает временную зависимость, так как индекс $k(t)$ зависит также от частоты и фазы (которые, обычно, неизвестны). Фактически распределением формы сигнала является полное постериорное распределение, маргинализованное по амплитуде

$$p(\mathcal{M}_K, \omega, \phi, \mathbf{f}|\mathcal{D}) = p(\mathcal{M}_K|\mathcal{D}) p(\omega|\mathcal{M}_K, \mathcal{D}) p(\phi|\omega, \mathcal{M}_K, \mathcal{D}) p(\mathbf{f}|\phi, \omega, \mathcal{M}_K, \mathcal{D}) \quad (11)$$

Здесь мы разложили совместное распределение так, чтобы было удобно провести грубый поиск моды – выполняя последовательно максимизацию каждого фактора: сначала ищем наиболее вероятное разбиение на бины, потом, зафиксировав модель \mathcal{M}_K , подбираем наиболее вероятную для нее частоту, затем, основываясь на уже определенных величинах K и ω , подбираем фазу, и наконец – набор параметров формы. Полученный в результате профиль будет *почти* наиболее вероятным. Наверное, не удивительно, что при этом получится $\hat{f}_k = n_k(\hat{\omega}, \hat{\phi})!$

К сожалению, такой подход практически ничего не говорит нам о неопределенности распределения профиля, т.е. о разнообразии профилей, согласованных с измерениями. Гораздо лучше для этих целей подходит МСМС-сэмплирование...

Рассмотрим подробнее последний фактор в (11):

$$p(\mathbf{f}|\omega, \phi, \mathcal{M}_K, \mathcal{D}) = \frac{(N + K - 1)!}{N!} W_K(\omega, \phi) \prod_{k=1}^K f_k^{n_k}$$

Моду этого условного распределения надо искать с учетом связи $\sum_k f_k = 1$, т.е. при помощи множителя Лагранжа. В итоге получим, как и анонсировалось выше, $\hat{f}_k = n_k/N$. Если же нас интересует не мода, а среднее этого распределения, то

$$\langle f_k|\omega, \phi, K \rangle \equiv \int d\mathbf{f} f_k p(\mathbf{f}|\omega, \phi, \mathcal{M}_K, \mathcal{D}) = \frac{n_k + 1}{N + K}$$

Кроме того,

$$\sigma_k = \sqrt{\langle f_k^2 \rangle - \langle f_k \rangle^2} = \sqrt{\frac{\langle f_k \rangle (1 - \langle f_k \rangle)}{N + K + 1}}$$

Если же частота и фаза нам априори неизвестны, то по ним надо провести маргинализацию

$$\langle f(t)|K \rangle = \int d\omega \int d\phi \int d\mathbf{f} f_{k(t)} p(\omega, \phi, \mathbf{f}|\mathcal{M}_K, \mathcal{D}) = \int d\omega \int d\phi \langle f_{k(t)}|\omega, \phi, K \rangle p(\omega, \phi|\mathcal{M}_K, \mathcal{D})$$

Примечание. С точки зрения байесовского подхода простая модификация может улучшить возможности модели PCS. Более эффективным, чем равномерный, будет *прайор Дирихле*, он также сохранит аналитичность вычислений. В наиболее общем виде распределение Дирихле описывается набором из K положительных параметров $\boldsymbol{\alpha} \equiv \{\alpha_k\}$, $k = 1, \dots, K$:

$$f \sim \text{Dir}(\boldsymbol{\alpha}): p(\mathbf{f}|\mathcal{M}_K) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K f_k^{\alpha_k - 1}$$

где $\alpha_0 = \sum_k \alpha_k$. Несложно показать, что маргинализация по параметрам амплитуды (при равномерном прайоре) и формы сигнала (при прайоре Дирихле) приводит к следующему результату для правдоподобия относительно частоты и фазы

$$\mathcal{L}(\omega, \phi) = \frac{(\delta t)^N K^N}{A_{\max} T^{N+1}} \gamma(N+1, A_{\max} T) \frac{\Gamma(\alpha_0) \Gamma(N + \alpha_0 - K + 1)}{\prod_k \Gamma(\alpha_k) \Gamma(N + \alpha_0)} \cdot \frac{1}{W(\omega, \phi)} \quad (12)$$

Здесь

$$W(\omega, \phi) = \frac{(N + \alpha_0 - K)!}{(n_1 + \alpha_1)! \dots (n_K + \alpha_K)!}$$

с факториалами от нецелых аргументов, понимаемыми в смысле гамма-функции, $x! = \Gamma(x+1)$, а символом γ обозначена *неполная гамма-функция*: $\gamma(n, x) = \int_0^x y^{n-1} \exp\{-y\} dy$, такая что $\Gamma(n) = \gamma(n, \infty)$.

Для равномерного прайора на параметры формы, $\alpha_k = 1$, $\alpha_0 = K$, выражение для W упрощается

$$W(\omega, \phi) = \frac{N!}{n_1! \dots n_K!}$$

и становится равным *кратности* (multiplicity), т.е. числу способов, которыми можно получить набор значений $\{n_k\}$ при распределении N объектов по K бинам (при этом (12) соответствует (9)).

При использовании прайора Дирихле остается незакрытым вопрос выбора его параметров α_k . По нескольким причинам (см., например, раздел 6.1.2) их значения имеет смысл сделать зависящими от K . Рассмотрим, например, семейство *симметричных* прайоров Дирихле, т.е. таких, что $\alpha_1 = \dots = \alpha_K = \alpha$, $\alpha_0 = K\alpha$ и


$$\pi(\mathbf{f}) \propto \delta\left(1 - \sum_k f_k\right) \prod_{k=1}^K f_k^{\alpha-1}$$

Одним из предпочтительных выборов параметра α является $\alpha \propto 1/K$ (коэффициент пропорциональности можно специфицировать на основании исследования известных кривых свечения изучаемых объектов), при котором семейство прайоров получается *делимым* (divisible):...

6.3 Поиск экзопланет

Метод радиальных скоростей – один из наиболее успешных методов обнаружения экзопланет. Он основан на определении вариаций в скорости звезды-хозяина, вызванных гравитационным воздействием движущейся по орбите вокруг нее планеты, а именно скорости вдоль луча зрения. Эта компонента скорости получила название *радиальной скорости* звезды.

Группа Лоредо разработала набор БАД-инструментов для поиска экзопланет, оценки параметров их орбит и адаптивного планирования будущих наблюдений. Экзопланета обнаруживается путем наблюдения за «колебаниями» звезды-хозяина, реагирующей на гравитационное притяжение орбитальной планеты. Временной ряд значений радиальной скорости звезды определяется в рамках ньютоновской механики. Однако эта аналитическая модель сильно нелинейна, что приводит к сложным, мультимодальным функциям правдоподобия, она становится еще более сложной в присутствии нескольких планет. Размерность пространства параметров варьируются от нескольких единиц до нескольких десятков. Поскольку орбиты периодичны, анализ облегчают байесовские обобщения методов



figs/Loredo2011a_fig04.png

Рис. 23: Процесс оценки параметров по RV-данным звезды HD 3651 (из работы [Fischer_2003]: 1) определение RV в отдельные(неэквидистантные) моменты времени, 2) оценка периода по периодограмме Ломба-Скаргла, 3) результат фитирования остальных параметров - минимизация хи-квадрат), 4) полученные оценки для периода, эксцентриситета, произведения массы планеты на синус угла наклона эклиптики и длины большой полуоси орбиты.

периодограмм. При этом модель должна быть линейно сепарабельной, что позволяет провести частичную аналитическую маргинализацию и уменьшить размерность пространства параметров. Последующий анализ использует адаптивные методы МСМС и адаптивное же сэмплирование по важности для вычисления интегралов, необходимых как для логического вывода (обнаружение планет и измерение их орбит), так и для информационно-максимизирующего проектирования последующих измерений (для адаптивного планирования наблюдений).

Основные результаты работ группы Лоредо опубликованы в работах [Loredo_2011a] и [Loredo_2011b]. (Обращаем внимание, что окончательная версия второй работы датирована аж 2018 годом!) Ниже мы проанализируем эти работы совместно.

Наиболее распространенным методом обнаружения экзопланет является доплеровский метод определения радиальной скорости (RV) звезды-хозяйки. Вместо того, чтобы измерять положение звезды на небе в зависимости от времени (что потребовало бы необычайной угловой точности), этот метод измеряет скорость звезды в прямой видимости как функцию времени – колебание от нас к нам, а не из стороны в сторону. Это возможно с помощью высокоточных спектроскопических наблюдений за линиями в спектре звезды: из-за эффекта Доплера длины волн этих линий очень незначительно смещаются во времени. Таким образом можно измерить RV с точностью до метра в секунду!

Работа [Loredo_2017] посвящена обоснованию возможности поиска экзопланет относительно малой массы, сравнимой с массой Земли. При этом амплитуды колебаний RV очень малы – сантиметры в секунду, что...

6.4 Иерархические модели

7 Байесовское планирование эксперимента

Первое знакомство с тем как байесовский вывод (BI) может помочь при планировании эксперимента, несомненно надо начинать с главы 7 учебника D.Sivia, носящее само за себя говорящее название Experimental design. (Адаптированный русский перевод этой главы вошел как раздел 1.7 спецкурса MDA.) Именно здесь приведены примеры влияния на результаты эксперимента аппаратной функции детектора,...

Дальнейшее погружение в тему, получившую в англоязычной литературе название Bayesian experimental design, предлагается совершить на основе следующих работ:

- [Chaloner1995](#) Kathryn Chaloner and Isabella Verdinelli, Bayesian Experimental Design: A Review. Statistical Science, 1995, Vol. 10(3), 273-304.
- [Sebastiani2000](#) Paolo Sebastiani and Henry P. Wynn, Maximum entropy sampling and optimal Bayesian experimental design. J. R. Statist. Soc. B (2000) 62, Part 1, 145-157
- [Loredo2011a](#) Thomas J. Loredo, Rotating Stars and Revolving Planets: Bayesian Exploration of the Pulsating Sky.

А Переводим Байесовскую Библию

Байесовскому подходу посвящено большое количество монографий и несчетное число публикаций. Но одна книга стоит особняком. Речь идет о библии байесовского подхода *Probability Theory: The Logic of Science*, написанной Эдвином Джейнсом ([J.T. Jaynes](#)) и выпущенной под редакцией его ученика G. Larry Bretthorst в 2003 году в издательстве Cambridge University Press.

Часть глав этой фундаментальной книги Джейнс не успел завершить – он ушел из жизни в 1998 году. В последние годы жизни он поручил своему ученику дописать недостающий материал. Но Бретхорст решил оставить книгу такой, какой она вышла из под пера самого учителя (лишь внеся необходимую редакторскую правку), оставив не заполненной часть книги.

На сайте родного для обоих Университета в Вашингтоне Бретхорст создал небольшую страничку, посвященную [Probability Theory As Extended Logic](#), которую снабдил вот такими словами:

Е.Т. Джейнс был одним из первых, кто осознал, что созданная Лапласом теория вероятностей, является обобщением аристотелевской логики, которое сводится к дедуктивной логике в том частном случае, когда наши гипотезы либо истинны, либо ложны.

Книга Джейнса (в дальнейшем мы будем ее называть по первым буквам слов из названия, как [PT-LS]) относится к разряду MUST READ. Во многом, и свой курс мы строили, опираясь на этот фундаментальный труд. Однако некоторые выдержки из книги (и даже целые разделы) хотелось бы представить перед читателем в оригинальном виде. Ниже приведен перевод наиболее интересных, с нашей точки зрения, «кусочков» книги. Их изучение ни в коей мере не заменит чтение [PT-LS], однако позволит хотя бы на немного проникнуться тем очарованием, которым наполняет вдумчивого читателя этот шедевр.

Собственные комментарии по поводу прочитанного мы будем выделять курсивом.

А.1 Физика «случайных экспериментов»

Глава №10 книги носит название «Физика случайных экспериментов», в которой Джейнс разворачивает увлекательную дискуссию вокруг объективности понятия вероятности в Физике, и в частности в практике физического эксперимента.

Главу предваряет эпиграф в виде цитаты G.E.P. Vox (1962):

«Полагаю, что было бы очень трудно убедить толкового физика в разумности существующей статистической практики, но было бы гораздо меньше проще с подходом, основанным на правдоподобии и теореме Байеса».

Сразу после этого Джейнс комментирует:

...идея о том, что вероятности являются физически реальными вещами, основанными на наблюдаемых частотах случайных величин, казалось бы, превращают ее [теорию вероятности] в раздел экспериментальной науки. Мы уже видели некоторые трудности, к которым приводит такой взгляд; в отдельных реальных физических экспериментах различие между случайными и неслучайными величинами настолько неясна и искусственна, что приходится прибегать к черной магии, чтобы вообще втиснуть это различие в проблему. В настоящей главе мы подробно остановимся на физических соображениях, показывающих принципиальную сложность понятия «случайных» экспериментов.

Ниже мы приводим полный перевод всех разделов этой главы книги.

Начинает Джейнс рассказа о том, как теория вероятностей после работ могучего

Лапласа погрузилась в такую «ортодоксальную яму». Оказывается, не последнюю роль в этом заняли биологи и философы, которые в начале 20-го столетия отодвинули от статистического анализа физиков – а те, почему-то, плохо споротивлялись...

А.1.1 10.1. Интересная корреляция

Всегда находились несогласные с «частотной» точкой зрения, которые вместе с Лапласом утверждали, что теорию вероятностей правильно рассматривать как «исчисление индуктивных рассуждений» и она вообще не имеет принципиального отношения к случайным экспериментам. Основная цель настоящей работы – продемонстрировать, что теория вероятностей может последовательно и с пользой работать не только с частотами в случайных экспериментах (если только ей будет позволено это делать). Согласно этой точке зрения, рассмотрение случайных экспериментов – лишь одно из специализированных приложений теории вероятностей, и далеко не самое важное. На самом деле теория вероятностей как логика решает гораздо более общие проблемы рассуждений, которые не имеют ничего общего с шансами или случайностью, но в значительной степени связаны с реальным миром. В настоящей главе мы пойдем дальше и покажем, что «частотная» теория вероятностей сталкивается с серьезными логическими трудностями при работе с очень случайными экспериментами, для которых она и была изобретена.

Изучая литературу по этим вопросам, можно обнаружить удивительную корреляцию: те, кто защищает нечастотную точку зрения, как правило, являются физиками, тогда как до самого недавнего времени математики, статистики и философы почти всегда отдавали предпочтение частотной интерпретации. Таким образом, оказывается, что вопрос этот не только философский или математический – каким-то образом, пока еще не совсем понятным, это связано и с физикой.

Математик склонен думать о случайном эксперименте как об абстракции – на самом деле, просто как о последовательности чисел. Чтобы определить «природу» случайного эксперимента, он вводит утверждения – по-разному называемые допущениями, постулатами или аксиомами, – которые определяют выборочное пространство и утверждают существование и некоторые другие свойства предельных частот. Но в реальном мире случайный эксперимент – отнюдь не абстракция, свойства которой можно определять по своему желанию. Он, безусловно, подчиняется законам физики. Тем не менее, признание этого явно отсутствует в частотных интерпретациях теории вероятностей. Даже словосочетания «законы физики» в них не найти. Но определение вероятности как частоты – это не просто предлог для игнорирования законов физики; все гораздо серьезнее. Мы хотим показать, что частотная интерпретация, исключая все другие, *требует* игнорировать практически все профессиональные знания, которыми обладают ученые о реальных явлениях. Вряд ли на это стоит опираться, если цель состоит в том, чтобы делать выводы о реальных явлениях.

Как только описывается конкретный случайный эксперимент, физику свойственно начинать думать не об абстрактном примере, а о физическом механизме наблюдаемого явления. Вопрос о том, совместимы ли обычные постулаты теории вероятностей с известными законами физики, поддается логическому анализу, результаты которого имеют прямое отношение к вопросу не о математической непротиворечивости частотной и нечастотной теорий вероятности, а их применимости к реальным ситуациям. В нашей вступительной цитате статистик Г.Э.П.Бокс отметил это; давайте проанализируем его заявление в свете как истории, так и физики.

А.1.2 10.2. Исторические предпосылки

Как мы знаем, теория вероятностей началась с рассмотрения игровых ситуаций Джероламо Кардано в 16-м веке и Паскаля и Ферма в 17-м. Однако ее дальнейшее развитие, в 18-м и 19-м веках, было стимулировано приложениями в астрономии и физике и было работой специалистов – Джеймса и Даниила Бернулли, Лапласа, Пуассона, Лежандра, Гаусса, Больцмана, Максвелла, Гиббса – большинство из которых мы сегодня назвали бы математическими физиками.

Но реакция против Лапласа началась уже в середине девятнадцатого века, когда Cournot, Ellis, Boole и Venn (никто из которых не имел физического образования) не смогли понять доводы Лапласа и подвергли их критике, фактически просто игнорируя все его успешные результаты. В частности, Джон Венн, философ, не обладавший ни малейшими познаниями Лапласа ни в физике, ни в математике, тем не менее считал себя способным писать язвительные и саркастические нападки на работы Лапласа. (В главе 16 мы отмечаем его возможное позднее влияние на молодого Р.А.Фишера.) Буль неоднократно показывает, что не понимает роли априорных вероятностей Лапласа (как представления *состояния знания*, а не физического факта). Другими словами, он тоже подвержен заблуждению проекции разума. Так, на стр. 380 он отвергает назначение в качестве априорного равномерного распределения из-за его «произвольности» и явно отказывается исследовать его следствия – тем самым он сам лишает себя возможности понять, что на самом деле делал Лаплас и почему.

Лапласа решительно защищали математик Август де Морган и физик У. Стэнли Джевонс, которые понимали мотивы Лапласа и для которых его прекрасная математика была скорее радостью, чем болью. Тем не менее нападки Буля и Венна нашли сочувственный отклик среди нефизиков в Англии. Возможно, это произошло потому, что биологи, чья физико-математическая подготовка была по большей части не намного лучше, чем у Венна, пытались найти эмпирические подтверждения теории Дарвина и понимали, что необходимо собрать и проанализировать большие массивы данных, чтобы обнаружить небольшие, медленные тренды (которые они воспринимали как инструменты, с помощью которых происходит эволюция). Считая, что математические труды Лапласа слишком обширны, чтобы их переварить (ведь профессии статистика еще не существовало), они, естественно, приветствовали предложение о том, что не нужно читать Лапласа вообще.

В любом случае радикальная перемена произошла примерно в начале этого [двадцатого] столетия, когда на арену вышла новая группа специалистов, но не физиков. Их интересовали в основном биологические проблемы, и при поддержке Венна они отвергали практически все, что делал Лаплас. Чтобы заполнить вакуум, они стремились заново развить эту область, основываясь на совершенно иных принципах, в которых вероятности приписывались только данным и ничему другому. По-началу это действительно упростило математику, так как многие проблемы, решаемые методами Лапласа, теперь оказались за пределами их методов. И пока они рассматривали только относительно простые задачи (технически это были задачи с достаточной статистикой и без вспомогательных параметров), этот недостаток не вызывал беспокойства. Эта чрезвычайно агрессивная школа вскоре стала доминировать в области настолько, что ее методы стали известны как «ортодоксальная» статистика. Современная профессия статистика развилась в основном из этого направления.

Одновременно с этим физики – за почти единственным исключением сэра Harold Jeffreys – постепенно ушли из этой науки, а статистический анализ исчез из учебных программ по физике. Это исчезновение было настолько полным, что если бы сегодня кто-нибудь провел опрос среди физиков, думается, ни один из ста опрошенных не смог бы

назвать такие имена, как Фишер, Нейман, Вальд, или такие термины, как максимальное правдоподобие, доверительный интервал, дисперсионный анализ.

Такой ход событий – ведущая роль физиков в разработке первоначальных байесовских методов и последующее их отлучение от ортодоксальной статистики – не был случайным. В качестве еще одного доказательства того, что между ортодоксальной статистической доктриной и физикой существует своего рода фундаментальный конфликт, мы можем отметить, что два самых красноречивых сторонника нечастотных определений в начале 20-го века – Пуанкаре и Джеффрис – были математическими физиками высочайшей компетентности, как и Лаплас. Таким образом, утверждение профессора Бокса имеет под собой четкую историческую подоплеку.

Но какова природа этого конфликта? Что есть такого в знаниях физика, что заставляет его отвергать то самое, что другие считают придающим «объективность» теории вероятностей? Чтобы увидеть, в чем заключается сложность, мы рассмотрим несколько простых случайных экспериментов с точки зрения физика. Факты, на которые мы хотим указать, настолько элементарны, что трудно поверить, что они действительно неизвестны современным авторам теории вероятностей. Постоянное появление новых учебников, игнорирующих их, просто иллюстрирует то, что мы, преподаватели физики, и так всегда знали: вы можете научить ученика законам физики, но вы не можете научить его искусству осознания актуальности этих знаний, не говоря уже о привычке их фактического применения в своих повседневных задачах.

А.1.3 10.3. Как смошенничать при подбрасывании монет и кубиков

Крамер (1946) принимает за аксиому, что «любая случайная величина имеет уникальное распределение вероятностей». Из более позднего контекста становится ясно, что на самом деле он имеет в виду, что она имеет уникальное частотное распределение. Если предположить, что число, полученное при бросании игральной кости, является случайной величиной, это приводит к выводу, что частота, с которой выпадает определенная грань кости, является ее физическим свойством – точно так же, как ее масса, момент инерции или химический состав. Так, Крамер (стр. 154) утверждает:

«Числа p_r должны фактически рассматриваться как физические константы конкретной игральной кости, и вопрос об их числовых значениях не может быть решен аксиомами теории вероятностей, точно также как размер и вес игральной кости не определяется геометрическими и механическими аксиомами. Однако опыт показывает, что в хорошо сделанной игральной кости частота любого события r в длинной серии бросков обычно приближается к $1/6$, и соответственно мы будем часто полагать, что все p_r равны $1/6$...»

Для физика это утверждение кажется полнейшим пренебрежением к известным законам механики. Результаты многократных подбрасываний игральной кости не сообщают нам какого-либо определенного числа, характерного только для этой кости. Напротив, они рассказывают кое-что о том, как бросали кости. Если вы подбрасываете «заряженные» кости по-разному, вы можете легко изменить относительную частоту выпадения граней. Даже если ваши кости «честны» вы все равно сможете добиться этого с чуть большей сложностью.

Хотя принципы останутся теми же, будет проще обсудить случайный эксперимент только с двумя возможными исходами в каждом испытании. Поэтому рассмотрим «предвзятую» (смешенную, *biased*) монету, о которой И.Дж.Гуд (1962) заметил:

«Большинство из нас, вероятно, думают о предвзятой монете, как если бы она имела физическую вероятность. Независимо от того, определено ли это понятие в терминах

частоты или просто возникает из другого типа теории, думаю, мы действительно рассуждаем именно так. Подозреваю, что даже самый крайний субъективист, такой как де Финетти, должен был бы согласиться с тем, что он действительно иногда думал таким образом, хотя, возможно, и не стал бы делать это публично.»

Мы не знаем личного мнения на этот счет самого де Финетти, но заметим, что как раз его знаменитая теорема об обмене (exchangeability theorem) показывает нам, как проводить вероятностный анализ смещенной монеты, не рассуждая предложенным выше способом.

В любом случае легко продемонстрировать, как будет анализировать эту проблему физик. Предположим, что центр тяжести монеты лежит на ее оси, но смещен на расстояние x от ее геометрического центра. Если мы согласны с тем, что результат подбрасывания этой монеты является «случайной величиной», то согласно аксиоме, сформулированной Крамером и на которую намекает Гуд, должна существовать определенная функциональная связь между частотой выпадения орла и x : $p_H = f(x)$.

Заметим, что данное утверждение выходит далеко за рамки традиционной свободы математика изобретать произвольные аксиомы и вторгается в область физики – ибо законы механики вполне компетентны сказать нам, существует ли такое функциональное отношение или нет.

Еще проще проанализировать игру, которую используют, чтобы решить такие практические вопросы, как выбор начальной стороны на футбольном матче. Сначала ваш оппонент выбирает «орла» или «решку» по своему желанию. Затем уже вы подбрасываете монету в воздух, ловите ее в руку и, не глядя на нее, показываете своему оппоненту – он выигрывает, если указал правильно. Кроме того, «справедливым» считается подбрасывание, при котором монета поднимается в воздух не менее чем на девять футов и, таким образом, находится в свободном полете не менее 1.5 секунд.

Законы механики говорят здесь нам следующее. Эллипсоид инерции тонкого диска представляет собой сплюснутый сфероид с эксцентриситетом $1/\sqrt{2}$. Смещение x не влияет на симметрию этого эллипсоида, поэтому согласно построению Poinsot, найденному в учебниках по динамике твердого тела, polhodes остаются кругами, концентрическими с осью монеты. Вследствие этого характер кувыркания у смещенной монеты в полете точно такой же, как и у несмещенной, за исключением того, что у смещенной монеты параболическую траекторию «свободной частицы» описывает центр тяжести, а не геометрический центр.

Важной особенностью этого кувыркающегося движения является сохранение углового момента: во время полета угловой момент монеты сохраняет фиксированное направление в пространстве (а угловая скорость – нет, и поэтому кувырканье может показаться глазу хаотичным). Обозначим это фиксированное направление единичным вектором n ; это может быть любое направление, которое вы выберете, и оно определяется особым поворотом, который вы придаете монете в момент запуска. Независимо от того, смещена монета или нет, она будет показывать одну и ту же грань на протяжении всего движения, если смотреть с этого направления (если, конечно, n не точно перпендикулярно оси монеты).

Поэтому, чтобы узнать, какая грань окажется верхней в вашей руке, нужно всего лишь выполнить следующую процедуру. Обозначим через k единичный вектор, проходящий через монету вдоль ее оси с точкой на стороне «орла». Теперь подбросьте монету, повернув ее так, чтобы k и n образовали острый угол, затем поймите ее ладонью в плоскости, перпендикулярной n . При последовательных бросках вы можете позволить направлению n , величине углового момента и углу между n и k широко варьироваться; тогда кувыркающееся движение будет казаться глазу совершенно разным при различных бросках, и потребуются почти сверхчеловеческая наблюдательность, чтобы обнаружить вашу страте-

гию.

Таким образом, любой, кто знаком с законом сохранения углового момента, может после некоторой практики сжульничать в обычной игре с подбрасыванием монеты и делать ставки со 100-процентной точностью. Вы можете получить любую частоту «орла», которую захотите. И *смещение монеты вообще никак не влияет на результаты!*

Конечно, как только этот секрет будет раскрыт, кто-нибудь возразит, что анализируемый эксперимент слишком «прост». Другими словами, те, кто постулировал физическую вероятность смещенной монеты, не говоря об этом, на самом деле имели в виду более сложный эксперимент, в котором некая «случайность» имеет больше возможностей дать о себе знать.

Принимая эту критику во внимание, не можем отказаться от очевидного комментария: просматривая литературу по теории вероятностей, не странно ли, что так много математиков, обычно куда более аккуратных, чем физики, при перечисления всех необходимых для истинности утверждения свойств, должно быть не узрели этой необходимости здесь? Однако для большей конструктивности мы можем с тем же успехом проанализировать и более сложный эксперимент.

Предположим, что теперь вместо того, чтобы ловить монету в руку, мы бросаем ее на стол и позволяем ей вращаться и подпрыгивать в различных направлениях, пока не остановится. Является ли этот эксперимент достаточно «случайным», чтобы проявилась истинная «физическая вероятность»? Пожалуй, ответом будет то, что случайность недостаточна, если монету подбрасывать вверх всего на шесть дюймов, но она станет «честным» экспериментом, если мы подбросим ее несколько выше.

Насколько высоко мы должны подбросить монету, чтобы можно было измерить ее истинную физическую вероятность? Это не простой вопрос, и мы не пытаемся ответить на него здесь. Однако может показаться, что всякий, кто утверждает существование физической вероятности монеты, должен быть готов ответить на него; в противном случае трудно понять, какое содержание имеет это утверждение (то есть нет возможности его подтвердить или опровергнуть).

Мы не отрицаем, что смещение монеты теперь будет иметь некоторое влияние на частоту выпадения орла; мы утверждаем только, что величина этого влияния очень сильно зависит от того, как именно вы подбрасываете монету. Так что, опять же, в этом эксперименте не существует определенного числа $p_H = f(x)$, описывающего физическое свойство монеты. В самом деле, даже направление этого влияния может быть изменено различными способами подбрасывания!

Как бы высоко мы ни подбрасывали монету, закон сохранения углового момента остается в силе; и поэтому мы можем подбросить ее *методом А*: чтобы гарантировать, что орел окажется сверху, когда монета впервые ударится о стол, нам нужно только держать ее орлом вверх и подбрасывать так, чтобы общий угловой момент был направлен вертикально. Опять же, мы можем варьировать величину углового момента и угол между n и k , так что на глаз движение будет выглядеть совершенно по-разному при разных бросках, и потребуются очень внимательное наблюдение, чтобы заметить, что орел остается сверху на протяжении всего полета. Несмотря на то, что все усложняется после удара монеты о стол, тот факт, что орел сначала находился сверху, оказывает сильное влияние на результат (это больше всего заметно для больших угловых моментов).

Многие люди научились подбрасывать монету *методом В*, при котором она проходит стадию стояния на ребре и быстрого вращения вокруг вертикальной оси, прежде чем, наконец, падает на одну или другую сторону. Если вы подбросите монету таким образом, эксцентричное положение центра тяжести будет иметь преобладающее влияние и сделает

практически достоверным, ее падение на одну и ту же сторону. Можно предположить, что монета предпочитает падать так, чтобы центр тяжести был ниже, т.е. если центр тяжести смещен в сторону решки, то монета должна будет иметь тенденцию выпадать орлом. Однако по интересной механической причине, которую мы оставляем вам для выяснения из принципов динамики твердого тела, метод Б на самом деле приводит к противоположному результату: монета предпочитает падать так, чтобы ее центр тяжести находился выше.

С другой стороны, смещение центра тяжести монеты имеет довольно небольшое влияние в противоположном направлении, если мы подбрасываем ее *методом С*: монета вращается вокруг горизонтальной оси, перпендикулярной ее оси, и так отскакивает, пока не сможет больше не перевернуться. *SSH: Я не понял этого метода*

Следовательно, и в этом эксперименте человек, знакомый с законами механики, может подбросить смещенную монету так, что она по его желанию выпадет преимущественно либо орлом, либо решкой. Кроме того, эффект метода А сохраняется независимо от того, является ли монета смещенной или нет; т.е. это можно сделать даже с совершенно «честной» монетой. Наконец, хотя мы рассматривали только монеты, по существу те же механические соображения (с более сложными деталями) применимы и к подбрасыванию любого другого предмета, например игральной кости.

Автор никогда не думал о смещенной монете, «как если бы она имела физическую вероятность», просто потому что, будучи профессиональным физиком, знал, что у нее нет физической вероятности. Из того факта, что мы видели сильный перевес выпадении, например, орла, мы не можем на законных основаниях заключить, что центр тяжести монеты смещен: он может быть смещенным, а может монету просто подбросили таким образом, что систематически выпадает именно орел. Точно так же из того факта, что мы видели одинаковое количество оров и решек, мы не можем заключить, что монета «честная». Она может быть честным, а может быть, ее подбрасывали так, чтобы свести на нет эффект смещения.

А.1.4 10.4. Экспериментальные доказательства

Поскольку только что приведенные выводы находятся в прямом противоречии с тем, что почти повсеместно постулируется при изложении теории вероятностей, стоит отметить, что можно легко проверить их, поэкспериментировав несколько минут на кухне. Прекрасную «смещенную монету» обеспечивает металлическая крышка от небольшой банки для джема, которая не имеет «резьбы» снаружи и край которой завернутый внутрь, а не наружу, так что внешняя поверхность круглая и гладкая, и настолько симметричная, что при взгляде на крышку сбоку нельзя сказать, какая ее сторона является верхней.

Подозревая, что многие не очень сведующие в физике люди просто не поверят только что заявленным вещам без экспериментального подтверждения, мы провели опыты с крышкой от банки диаметром $d = 2\frac{5}{8}$ дюйма, высотой $h = \frac{3}{8}$ дюйма. При одинаковой толщине металла центр тяжести должен быть смещен от геометрического центра на расстояние $x = dh/(2d + 8h) = 0.120$ дюйма, что было подтверждено путем подвешивания крышки за край и измерением угла, под которым она останавливается в покое. Обычно ожидается, что это смещение заставит крышку падать нижней (т.е. внутренней) стороной вверх – поэтому назовем ее «орлом». Крышку подбрасывали примерно на 6 футов и она падала на гладкий линолеум. Я сделал по десять тренировочных бросков каждым из трех описанных методов, после чего выполнил основную часть эксперимента (результаты приведены в таблице 10.1): методом А с намеренным предпочтением орла, методом А с намеренным предпочтением решки, методом В и методом С.

В методе А способ подбрасывания полностью доминировал над результатом (эффект смещения, по-видимому, был бы еще больше, если бы «монету» бросали на поверхность с большим коэффициентом трения). В методе Б смещение полностью определяло результат (интересно, что примерно в тридцати таких бросках какое-то время казалось, что выпадет решка, как можно было бы наивно ожидать; но каждый раз «монета» в конце концов выправлялась и переворачивалась – как и предсказывают законы динамики твердого тела). В методе С не было никаких значимых доказательств влияния смещения. Результаты весьма красноречивы.

Несогласный всегда может заявить, что подбрасывание монеты любым из четырех конкретных описанных способов является «мошеничеством» и что существует «честный» способ ее подбрасывания, при котором «истинные» физические вероятности монеты выводятся из эксперимента. Но опять-таки человек, утверждающий это, должен быть готов точно определить, в чем состоит этот справедливый метод, иначе утверждение лишено содержания. По-видимому, честный метод подбрасывания должен представлять собой какую-то случайную смесь методов А(Н), А(Т), В, С и каких-то других – но тогда какой «справедливый» относительный вес им придать? Трудно понять, как можно определить «справедливый» метод подбрасывания, кроме как при условии, что он должен приводить к определенной частоте выпадения орла – и мы попадаем в замкнутый круг.

Этот анализ можно продолжить дальше, что мы и сделаем ниже; но, возможно, уже сейчас становится понятно, что анализ подбрасывания монеты и кости не является проблемой абстрактной статистики, в которой можно свободно вводить постулаты о «физических вероятностях», игнорирующие законы физики. Это проблема механики, очень сложная и не относящаяся к теории вероятностей, за исключением того, что она заставляет нас более тщательно думать о том, как должна быть сформулирована теория вероятностей, чтобы ее можно было применять к реальным ситуациям. Выполнение случайного эксперимента с монетой не говорит нам, какова физическая вероятность выпадения орла; это может сказать нам кое-что о смещении, но также и кое-что о том, как подбрасывается монета. В самом деле, если мы не знаем, как она подбрасывается, мы не сможем сделать каких-либо надежных выводов о его смещении из эксперимента.

Однако из вышеизложенного может быть неясно, что выводы такого типа вполне справедливы для случайных экспериментов и никоим образом не зависят от конкретных механических свойств монет и игральных костей. Чтобы проиллюстрировать это, рассмотрим совершенно другой вид случайного эксперимента с точки зрения физика.

А.1.5 10.5. Игра в бридж

В другом месте мы приводим цитаты заявлений профессора У.Феллера об использовании теоремы Байеса в тестировании контроля качества (quality control testing, глава 17), о правиле наследования Лапласа (rule of succession, глава 18) и о концепции Даниэля Бернулли о функции полезности для теории принятия решений (utility function, глава 13). Не подведет он нас и здесь. В своем интересном учебнике (Феллер, 1951) он пишет: «Число возможных раздач карт в бридже составляет почти 10^{30} . Обычно мы соглашаемся считать их равновероятными. Для проверки этого необходимо более 10^{30} опытов – потребовалось бы миллиард миллиардов лет, если бы каждый живущий человек играл одну игру каждую секунду, днем и ночью». И снова у нас точка зрения, что раздача карт в бридже обладает «физической вероятностью», что приписывание одинаковых вероятностей является «соглашением» и что окончательным критерием его правильности должны быть наблюдаемые частоты в случайном эксперименте.

Ошибка здесь в том, что никто из нас – даже сам Феллер – не захотел бы воспользоваться этим критерием с реальной колодой карт. Потому что, если мы знаем, что колода честная, здравый смысл подсказывает нам нечто более весомое, чем 10^{30} случайных экспериментов. На самом деле мы согласились бы принять результат случайного эксперимента только в том случае, *если бы он согласовывался с нашим предвзятым представлением о том, что все распределения равновероятны.*

Многим это последнее утверждение покажется чистым богохульством: оно резко противоречит тому, чему нас всех учили – правильному отношению к теории вероятностей. Однако для того, чтобы понять, почему это так, нам нужно только представить себе, что эти 10^{30} опытов были проведены, а равномерного распределения не получилось. Если все раздачи карт имеют одинаковую частоту, то любая комбинация из двух указанных карт выпадет вместе в данной руке в среднем один раз за $(52 \times 51)/(13 \times 12) = 17$ раздач. Но предположим, что комбинация (Валет червей – Семерка треф) вместе выпадала в три раза чаще. Примем ли мы тогда как установленный факт, что в конкретной комбинации (Валет червей – Семерка треф) есть что-то такое, что делает ее более вероятной, чем другие?

Полагаем, что нет. Мы скорее отказались бы от результатов эксперимента и сказали бы, что карты не были должным образом перетасованы. И тем самым снова попали бы в замкнутый круг, потому что нет никакого способа определить «правильный» метод перетасовки, кроме как при условии, что он должен производить все распределения с одинаковой частотой!

Но любая попытка найти такое определение сопряжена с еще более глубокими логическими трудностями; никто не осмеливается подробно описать процедуру перетасовки, потому что это разрушило бы «случайность» и сделало бы точный результат предсказуемым и всегда одним и тем же. Для того чтобы эксперимент оставался «случайным», необходимо описать процедуру не полностью, чтобы результат в разных попытках был разным. Но как можно доказать, что не полностью определенная процедура будет производить все распределения с одинаковой частотой? Нам кажется, что попытка поддержать постулат Феллера о физических вероятностях раздач в игре в бридж прямо приводит к логическому противоречию.

Традиционное образование считает, что определение вероятностей должно основываться в основном на частотах; и что любое другое обоснование в лучшем случае подозрительно, в худшем – иррационально с катастрофическими последствиями. Напротив, приведенный выше пример очень ясно показывает, что *существует принцип определения вероятностных отнесений, не имеющий ничего общего с частотами, но настолько убедительный, что имеет приоритет над любым объемом частотных данных.* Если нынешнее обучение не допускает существования этого принципа, то только потому, что наша интуиция зашла так далеко впереди логического анализа – как это происходит, например, в элементарной геометрии – что мы никогда не утруждали себя изложением этого логического анализа в строгой математической форме. Но если мы научимся это делать, можно ожидать, что математическая формулировка может быть применена к гораздо более широкому классу задач, где одной нашей интуиции уже вряд ли будет достаточно.

Проводя вероятностный анализ раздач карт в бридже, действительно ли нас интересуют физические вероятности? Или индуктивное рассуждение? Чтобы ответить на это, рассмотрим следующий сценарий. Дата – 1956 год, когда писатель встретил Вилли Феллера и поговорил с ним на эту тему. Предположим, я сказал бы ему, что сдавал карты 1000 раз, каждый раз тасуя «честно»; и что в каждом случае в моей руке оказывалась семерка треф. Какой бы была его реакция? Думаю, он мысленно представил бы число $(1/4)^{1000} = 10^{-602}$ и тут же заключил, что я не сказал правды – и никакие уговоры с

моей стороны не поколебали бы этого суждения. Но чем объясняется сила его веры? Понятно, это не может быть оправдано, если наше присвоение равных вероятностей всем раздачам карт (и следовательно, вероятность $1/4$ того, что семерка трюф окажется в руке раздающего) является просто «соглашением», которое может быть изменено в свете экспериментальных доказательств. Он отверг бы представленные мной экспериментальные данные, точно также как и мы выше. Еще более очевидно, что он не воспользуется никакими знаниями и о результатах эксперимента с 10^{30} раздачами в бридж.

Тогда *какие у него есть дополнительные доказательства*, которые, как подсказывает ему здравый смысл, имеют больший вес, чем любое количество случайных экспериментов, но чью помощь он отказывается признать при написании своих учебников? Чтобы утверждать, что теория вероятностей является экспериментальной наукой, опирающейся в основном не на логический вывод, а на частоты исходов случайного эксперимента, необходимо скрыть часть имеющейся информации. Однако именно эта скрытая информация позволяет нашим выводам приблизиться к достоверности дедуктивного рассуждения в этом и многих других примерах.

Скрытое свидетельство – это, конечно, просто наше признание *симметрии* ситуации. Единственная разница между семеркой и восьмеркой заключается в том, что на лицевой стороне карты напечатано другое число. Наш здравый смысл подсказывает нам, что то, куда пойдет карта при тасовании, зависит только от приложенных к ней механических сил, а не от того, какой номер напечатан на ее лицевой стороне. Если мы наблюдаем какую-либо систематическую тенденцию к попаданию одной карты в руке крупье, которая сохраняется при бесконечном повторении эксперимента, мы можем заключить из этого только то, что в процедуре тасования существует некоторая систематическая тенденция, которая одна и определяет исход эксперимента.

Таким образом, снова эксперимент ничего не говорит нам о «физических вероятностях» различных раздач в бридж. Он говорит лишь о том, как тасуются карты. Но пока еще не раскрыта вся мощь симметрии как убедительного доказательства, так что пришло время рассмотреть этот вопрос подробнее.

В следующих двух разделах Джейнс показывает, что отсутствие физических вероятностей может быть доказано в любом реальном физическом эксперименте. И снова обращается к проблеме использования индуктивного вывода в науке. И особой роли физиков в этом!

А.1.6 10.6. Обобщенный случайный эксперимент

После вышеизложенных доводов все еще можно занять следующую позицию (как это и сделал один из слушателей моей лекции): «Вы показали только то, что монеты, кости и карты представляют собой исключительные случаи, когда физические соображения устраняют обычные постулаты вероятности, т.е. они на самом деле не являются “случайными экспериментами”. Но это не имеет большого значения, так как все они использовались лишь в иллюстративных целях; а вот в более достойных случайных экспериментах, которые заслуживают серьезного внимания ученого, физическая вероятность существует».

Чтобы ответить на это, отметим два момента. Во-первых, повторим, что когда кто-либо утверждает существование физической вероятности в каком-либо эксперименте, то на нем лежит обязанность определить точные обстоятельства, при которых эта физическая вероятность может быть измерена; в противном случае утверждение лишено содержания.

Следует подчеркнуть: те, кто утверждает существование физических вероятностей, делают это, полагая, что устанавливают тем самым «объективность» своей позиции, ко-

торой не хватает тем, кто говорит только о «состоянии знания». Однако утверждать как факт то, что не может быть ни доказано, ни опровергнуто наблюдением фактов, есть противоположность объективности; это утверждение того, истинность чего никто не мог бы знать. Такое утверждение даже нельзя назвать описанием «состояния знания».

Во-вторых, обратите внимание, что любой конкретный эксперимент, для которого утверждается существование физической вероятности, может быть подвергнут физическому анализу, подобному только что приведенному, что в конечном итоге приведет к пониманию его механизма. Но как только это понимание будет достигнуто, тогда и этот новый эксперимент окажется исключительным случаем, как и предыдущие, когда физические соображения исключили обычные постулаты физических вероятностей.

Ибо, как только мы поняли механизм любого эксперимента, логически не остается места для какого-либо постулата о том, что различные исходы обладают физической вероятностью. И вопрос «Каковы вероятности различных исходов (O_1, O_2, \dots)?» сразу сводится к вопросу «Каковы вероятности соответствующих начальных условий (I_1, I_2, \dots), которые привели к этим исходам?»

Можно предположить, что сами возможные начальные условия $\{I_k\}$ эксперимента E обладают физическими вероятностями. Но тогда мы рассмотрим предшествующий случайный эксперимент E' , который создает условия I_k в качестве возможных результатов: $I_k = O'_k$. Мы можем проанализировать физический механизм E' , и как только он будет понят, вопрос снова вернется к следующему: «Каковы вероятности различных начальных условий I'_k для эксперимента E' ?»

Очевидно, мы вовлечены в бесконечную регрессию $\{E, E', E''\}$; попытка ввести физическую вероятность потерпит неудачу на каждом уровне, где наше знание физических законов позволяет нам анализировать задействованный механизм. Понятие «физической вероятности» должно постоянно переходить с одного уровня на другой по мере развития знания.

Таким образом, мы находимся в ситуации, очень похожей на «войну между наукой и теологией» прежних времен. В течение нескольких столетий теологи, не имевшие фактических знаний в астрономии, физике, биологии и геологии, тем не менее считали себя компетентными делать догматические фактические утверждения, которые вторгались в эти области – и от которых они впоследствии были вынуждены отказаться одно за другим перед лицом развития знаний.

Ясно, что теория вероятностей должна быть сформулирована таким образом, чтобы избежать утверждений о фактах, которые определенно принадлежат другим областям и от которых позже потребуются отказаться (как это сейчас имеет место со многими утверждениями касающимися монет, игральных костей и карт). Нам кажется, что единственная формулировка, которая достигает этого и в то же время обладает аналитической силой для решения актуальных проблем науки, – это та, которая была обнаружена и изложена на интуитивной основе Лапласом и Джеффрисом. Ее достоверность является вопросом логики и не зависит ни от каких физических предположений.

Как мы видели в главе 2, основной вклад в эту логику внес Р.Т.Кокс (1946), (1961), показавший, что эти интуитивные основания можно заменить теоремами. Думаем, что не случайно Ричард Кокс был также физиком (профессором физики и деканом аспирантуры в Университете Джонса Хопкинса), которому с самого начала были очевидны отмеченные нами выше моменты.

Формулировка теории вероятностей Лапласа-Джеффриса-Кокса не требует от нас делать эти вынужденные шаги вниз по этому бесконечному регрессии. Она полагает, что все, что – подобно детскому привидению – постоянно ускользает от света детального рас-

смотрения, может существовать только в нашем воображении. Те, кто более всего верит в физические вероятности, как и те, кто верит в астрологию, кажется, никогда не задаются вопросом, что представляет собой контролируемый эксперимент, способный подтвердить или опровергнуть их веру.

Действительно, примеры с монетами и картами должны убедить нас в том, что такие контролируемые эксперименты в принципе невозможны. Выполнение любого из так называемых случайных экспериментов не скажет нам, что такое «физические вероятности», потому что *не существует такой вещи, как «физическая вероятность»*. Эксперимент сообщает нам, пусть довольно грубым и неточным образом, кое-что о том, как начальные условия меняются от одного повторения к другому.

Гораздо более эффективным способом получения этой информации было бы непосредственное наблюдение за начальными условиями. Однако во многих случаях это выходит за рамки наших нынешних возможностей – как при определении безопасности и эффективности нового лекарства. Здесь единственным полностью удовлетворительным подходом был бы анализ подробной последовательности химических реакций, которые следуют за приемом этого лекарства у людей с любым мыслимым состоянием здоровья. Проведя такой анализ, можно было бы предсказать для каждого отдельного пациента, каким именно будет эффект от приема лекарства.

Поскольку в настоящее время о таком анализе вообще не может быть и речи, единственным возможным способом получения информации об эффективности лекарства является проведение «случайного» эксперимента. Нет двух пациентов с одинаковым состоянием здоровья, и неизвестные вариации этого фактора составляют переменные начальные условия эксперимента, а выборочное пространство содержит множество различных реакций на лекарство. Наше использование теории вероятностей в этом случае является стандартным примером индуктивного рассуждения, которое сводится к следующему.

Если исходные условия эксперимента (т.е. физиологические состояния приходящих к нам больных) в будущем будут продолжать изменяться в том же неизвестном диапазоне, что и в прошлом, то относительная частота излечений в будущем будет приближаться к той, которую мы наблюдали в прошлом. В отсутствие положительных свидетельств, объясняющих причину, по которой должны произойти некоторые изменения в будущем, и указывающих, в каком направлении эти изменения должны идти, у нас нет оснований предсказывать какие-либо изменения в любом направлении, и поэтому мы можем только предполагать, что все будет продолжаться более-менее одинаково. По мере того, как мы наблюдаем, что относительная частота излечений и побочных эффектов остается стабильной в течение все более и более продолжительного времени, мы все больше и больше убеждаемся в этом выводе. Но это только индуктивное рассуждение – нет никакого дедуктивного доказательства того, что частоты в будущем не будут полностью отличаться от тех, что были в прошлом.

Предположим теперь, что привычки в еде или какой-либо другой аспект образа жизни населения начинают меняться. Тогда состояние здоровья поступающих пациентов будет варьироваться в другом диапазоне, нежели раньше, и частота выздоровления при одном и том же лечении может начать дрейфовать вверх или вниз. Вероятно, мониторинг этой частоты может оказаться полезным индикатором того, что привычки населения меняются, а это, в свою очередь, может привести к новой политике в отношении медицинских процедур и просвещения в вопросах здоровья.

Мы видим, что применяемая здесь логика практически идентична логике промышленного контроля качества, обсуждавшейся в главе 4. Но рассмотрение ее в более общем виде заставляет нас увидеть роль индукции в науке совершенно иначе, чем представляли себе

это раньше некоторые философы.

А.1.7 10.7. И снова об индукции

Как мы отмечали в главе 9, некоторые философы отвергли индукцию на том основании, что нет способа доказать ее «правильность» (теории никогда не могут достичь высокой вероятности). Но функция индукции состоит в том, чтобы сообщить нам не о том, какие предсказания верны, а о том, на какие предсказания указывает наше нынешнее знание. Если предсказания сбываются, то мы довольны и становимся более уверенными в наших нынешних знаниях – но при этом мы мало чему научились.

На реальную роль индукции в науке ясно указал более шестидесяти лет назад Гарольд Джеффрис (1931, глава 1), однако, насколько нам известно, ни один математик или философ никогда не обращал ни малейшего внимания на сказанное им:

«Общим аргументом в пользу индукции является то, что индукция всегда срабатывала в прошлом и, следовательно, можно ожидать, что она будет действовать и в будущем. Возражали, что это сам по себе индуктивный аргумент и не может использоваться в поддержку индукции. Что почти никогда не упоминается, так это то, что индукция часто терпела неудачу в прошлом и что прогресс в науке в значительной степени является следствием прямого внимания к случаям, когда индуктивный метод приводил к неверным предсказаниям.»

Говоря более строго, мы узнаем что-то новое о реальном мире только тогда, когда наши индуктивные выводы ошибочны. Таким образом, для ученого самый быстрый путь к открытию состоит в том, чтобы исследовать те ситуации, в которых наиболее вероятно, что индуктивный вывод из наших текущих знаний не верен. Но эти выводы должны быть нашими лучшими выводами, которые полностью используют все наши знания. Всегда можно сделать индуктивные выводы, которые бесполезно ошибочны, просто игнорируя убедительную информацию.

Именно это и сделал Поппер. Его попытка интерпретировать саму вероятность как выражение физической причинности не только наносит вред приложениям теории вероятностей, как мы видели в главе 3 (это помешало бы нам правильно определить примерно половину всех условных вероятностей, потому что они выражают логические связи, а не причинно-физические) – это приводит к тому, что человек наколдовывает воображаемые причины, игнорируя при этом то, что уже известно о реальных физических причинах. Это может свести наши выводы к уровню донаучных, необразованных суеверий, даже если у нас есть хорошие данные.

Почему физики видят это с большей легкостью, чем другие? Потому что, создав это знание физических законов, мы кровно заинтересованы в нем и хотим, чтобы оно сохранялось и использовалось. Интерпретации частоты или склонности (propensity) начинаются с того, что отбрасываются практически все профессиональные знания, над получением которых мы трудились веками. Те, кто этого не понял, не в состоянии рассуждать с нами о философии науки или о надлежащих методах вывода.

В заключительных разделах главы Джейнс замахивается на, казалось бы, «святое»: на вероятностную интерпретацию квантовой физики. Причем, Jaunes сначала делает это несколько философски, а потом мастерски иллюстрирует наглядным примером развита альтернативной науки.

А.1.8 10.8. А что же на счет квантовой теории?

Те, кто цепляется за веру в существование «физических вероятностей», могут отреагировать на приведенные выше аргументы, приведя в качестве контрпримера квантовую теорию, где физические вероятности выражают наиболее фундаментальные законы физики. Попробуем объяснить, почему это еще один случай «рассуждений по кругу» (circular reasoning). Необходимо понять, что нынешняя квантовая теория использует совершенно иные стандарты логики, чем остальная наука.

В биологии или медицине, если мы отметим, что некоторый эффект E (например, сокращение мышц, фототропизм или переваривание белка) не возникает пока не будет выполнено условие C (наличие нервного импульса, освещения, пепсина), кажется естественным сделать вывод, что C является необходимой причиной для E . Большая часть того, что известно во всех областях науки, является следствием такого рода рассуждений. Но предположим, что условие C не всегда приводит к эффекту E – какие тогда выводы должен сделать ученый? И вот тут варианты рассуждений биологии и квантовой теории резко расходятся.

В биологических науках считается само собой разумеющимся, что помимо C должен существовать еще какой-то дополнительный причинный фактор F , пока еще не идентифицированный. И его начинают искать, отслеживая предполагаемую причину путем исключения возможностей. Порой это бывает чрезвычайно утомительным. Но настойчивость окупается: в конечном итоге будут достигнут успех, важный с медицинской точки зрения и впечатляющий с интеллектуальной, а предполагаемый неизвестный причинный фактор, наконец, идентифицировался как определенное химическое соединение. Большинство ферментов, витаминов, вирусов и других биологически активных веществ обязаны своим открытием именно этому процессу.

В квантовой теории рассуждают иначе. Рассмотрим, например, фотоэлектрический эффект: освещаем металлическую поверхность и обнаруживаем, что из нее вылетают электроны. Экспериментальный факт состоит в том, что электроны не появляются, если нет света. Итак, свет должен быть причинным фактором. Но свет не всегда приводит к вылету электронов: даже при использовании одномодового лазера с абсолютно постоянной амплитудой электроны появляются только в определенные моменты времени, которые не определяются никакими известными параметрами света. Почему же тогда мы не делаем очевидного вывода, что помимо света должен существовать второй причинный фактор, пока еще не идентифицированный, а работа физика состоит в его поиске?

На самом деле в современной квантовой теории поступают прямо противоположно: если нет очевидной причины, человек просто постулирует, что причины не существует – а следовательно, законы физики не детерминированы и могут быть выражены только в вероятностной форме. Центральная догма состоит в том, что свет определяет не появление фотоэлектрона, а только вероятность его появления. Математический формализм современной квантовой теории – неполный в том же смысле, в каком неполны наши нынешние знания – даже не дает словаря, в котором можно было бы задать вопрос о реальной причине события.

Биологи же создают механистическую картину мира, потому что, будучи приучены верить в причины, они продолжают использовать всю мощь своего интеллекта для их поиска – и поэтому они их находят. У квантовых физиков есть только вероятностные законы, потому что на протяжении двух поколений нам внушали не верить в причины – и поэтому мы перестали их искать. В самом деле, любая попытка поиска причин микроявлений встречает презрение и обвинение в профессиональной некомпетентности и «устаревшем механистическом материализме». Следовательно, чтобы объяснить неопре-

деленность современной квантовой теории, нам не нужно предполагать, что в Природе есть неопределенность: это гарантирует ментальная установка самих квантовых физиков.

Это важно подчеркнуть еще и потому, что большинство людей, не изучавших квантовую теорию в полном техническом объеме, относятся с недоверием, когда им сообщают, что она не занимается причинами. В действительности же, квантовая теория даже не признает понятие «физическая реальность». Ее современная интерпретация принадлежит Нильсу Бору, руководившему Институтом теоретической физики в Копенгагене; поэтому ее стали называть «копенгагенской интерпретацией».

Как неоднократно подчеркивал Бор в своих трудах и лекциях, нынешняя квантовая теория может ответить только на вопросы вида: «Если этот эксперимент будет проведен, каковы возможные результаты и их вероятности?» А ни на один вопрос типа: «Что на самом деле происходит, когда...?» она в принципе ответить не может. Опять же, математический формализм нынешней квантовой теории, как и оруэлловский новояз, даже не дает словаря, в котором можно было бы задать такой вопрос. Эти моменты были подробно рассмотрены в работах Jaunes (1986d, 1989, 1990a, 1992a).

Таким образом, мы предполагаем, что те, кто пытается оправдать концепцию «физической вероятности», указывая на квантовую теорию, попадают в ловушку рассуждений по кругу, по сути ничем не отличающуюся от разобранных выше примеров с подбрасыванием монетами и раздачей карт. В квантовой теории вероятности выражают неполноту человеческого знания – ровно так же, как и в классической статистической механике.

В классической статистической механике вероятностные распределения отражали наше незнание (ignorance) истинных микроскопических координат. Незнание, которого, в принципе, можно было бы избежать, но неизбежного на практике. Однако это не мешает нам предсказывать воспроизводимые явления просто потому, что эти явления не зависят от микроскопических деталей.

В современной квантовой теории вероятности выражают наше собственное неведение из-за нашей неспособности искать настоящие причины физических явлений; и, что еще хуже, нашу неспособность даже серьезно подумать о проблеме. Это неведение может быть неизбежным на практике, но в нашем нынешнем состоянии знаний мы не знаем, неизбежно ли оно в принципе; «центральная догма» просто утверждает это и делает вывод, что вера в причины и их поиски философски наивны. Если бы все приняли и соблюдали эту догму, никаких дальнейших успехов в понимании физического закона не было бы. Действительно, после Сольвеевского конгресса 1927 года, на котором этот образ мышления закрепился в физике, такого прогресса и не было сделано. (Конечно, физики продолжали открывать новые частицы и методы вычислений – точно так же, как астроном может открыть новую планету и новый алгоритм для вычисления ее орбиты, не продвинувшись вперед в своем базовом понимании небесной механики.) По нашему мнению, такая установка – премия за глупость; нам просто не хватает изобретательности придумать рациональное физическое объяснение вместо того, чтобы поддерживать сверхъестественную точку зрения.

Для многих людей эти идеи почти невозможно принять, так как они радикально отличаются от того, чему нас всех учили с детства. Поэтому давайте покажем, как такая же ситуация могла бы произойти в «эксперименте» с подбрасыванием монеты, если бы классические физики использовали те же стандарты логики, которые сейчас используются в квантовой теории.

A.1.9 10.9. Механика под облаками

Нам повезло, что законы механики Ньютона могут быть открыты и проверены с большой точностью путем изучения астрономических явлений, в которых трение и турбулентность

не усложняют то, что мы видим. Но предположим, что Земля, как Венера, была бы постоянно окружена толстым слоем облаков. В течение долгого времени само существование внешней вселенной было бы неизвестно, а обнаружение законов механики зависело бы от наблюдений, которые мы могли бы сделать только локально.

Поскольку подбрасывание мелких предметов – одно из первых занятий каждого ребенка, довольно рано можно было бы заметить, что предметы не всегда падают одной и той же стороной вверх, а попытки контролировать результат тщетны. Естественная гипотеза заключалась бы в том, что именно желание (volition) брошенного объекта, а не желание бросающего, определяет результат. На самом деле, именно так и отвечают маленькие дети, когда их спрашивают об этом.

Но продолжим. После того, как были бы изобретены монеты, большим бы открытием стал тот факт, что они имеют тенденцию проявлять обе стороны («орла» и «решку») примерно одинаково часто; и равенство частот, по мере увеличения количества бросков, кажется, только улучшается. Равенство частот выпадения орла и решки будет рассматриваться как фундаментальный закон физики: симметричные объекты обладают симметричным же желанием падать (как, действительно, кажется, полагали Крамер с Феллером).

Основываясь на этом, мы смогли бы создать математическую теорию подбрасывания объектов, открыв биномиальное распределение, отсутствие временных корреляций, предельные теоремы, комбинаторные законы для частот подбрасывания нескольких монет одновременно, и обобщив на более сложные симметричные объекты, такие как игральные кости и т.п. Все экспериментальные подтверждения теории будут состоять из все новых и новых экспериментов, измерений частот во все более сложных сценариях. Из таких экспериментов никогда не будет найдено ничего, что ставило бы под сомнение существование этого желания брошенного объекта; они только позволят подтвердить наличие этого желания и измерять его все точнее и точнее.

Однако допустим, что нашелся кто-то, кто был настолько глуп, что предположил, что движение брошенного объекта определяется не его собственным желанием, а законами, подобными законам ньютоновской механики, которые регулируются его начальным положением и скоростью. Его встретили бы насмешками и презрением, ибо во всех существующих экспериментах нет ни малейшего доказательства такого влияния. Истеблишмент заявит, что, поскольку все наблюдаемые факты объясняются теорией желания, с философской точки зрения наивно, а с профессиональной – не компетентно, предлагать или искать что-то более глубокое. В этом отношении учебники элементарной физики можно было бы читать так же, как наши нынешние учебники по квантовой теории.

В самом деле, любой, кто попытается проверить механическую теорию, не добьется успеха: как бы осторожно он ни подбрасывал монету (не зная того, что мы знаем), она одинаково часто продолжала бы выпадать орлом и решкой. Чтобы найти какие-либо доказательства причинной связи вместо статистической закономерности, потребовалось бы контролировать начальные условия броска гораздо точнее, чем кто-либо может добиться с помощью ручного подбрасывания. Так можно продолжать до бесконечности, удовлетворяясь законами физической вероятности и отрицая существование причин для отдельных подбрасываний (внешних по отношению к брошенному объекту) – точно так же, как квантовая теория делает сегодня – потому что эти вероятностные законы правильно объясняют все, что мы можем воспроизводимо наблюдать с помощью используемых нами же технологий.

Наконец, после тысячелетнего триумфа статистической теории кто-то создал устройство, подбрасывающее монеты в абсолютно неподвижном воздухе с очень точным контролем начальных условий. И тут, волшебным образом монета начинает давать неравное

количество орлов и решек: оказывается частота выпадения орла частично регулируется устройством. С созданием все более и более точных машин, наконец, достигается такая степень контроля, когда результат подбрасывания может быть предсказан со 100% точностью. И вера в «физические вероятности», выражающие желание монеты, наконец признана необоснованным суеверием. Существование лежащей в основе этого механической теории уже не подлежит сомнению, а длительный успех предыдущей статистической теории рассматривается как результат отсутствия контроля над начальными условиями подбрасывания.

Из-за недавних впечатляющих достижений в технологии экспериментов со все более детальным контролем над начальными состояниями отдельных атомов, нам представляется, что пройдет не так уж и много лет, и то же самое случится и с квантовой теорией. Через столетие истинные причины микроявлений будут известны каждому школьнику, и, перефразируя Сенеку, они будут недоумевать, как такие ясные истины могли ускользнуть от нас на протяжении всего 20 века. (Да и 21-го тоже.)